University of Pennsylvania Working Papers in Linguistics

Volume 11 | Issue 1 Article 18

1-1-2005

Addressing acquisition from language change: A modeling perspective

LISA PEARL

Addressing acquisition from language change: A modeling perspective

Addressing Acquisition from Language Change: A Modeling Perspective

Lisa Pearl*

1 Introduction

The syntactic structure of a language is something which is acquired very early on and usually not mutable once adulthood has been reached.[1] Because of this, there is an intimate relationship between early acquisition of certain syntactic structures and language change for those syntactic structures—acquisition is what drives language change. The "change" itself happens during acquisition, where individual children acquire grammars which are slightly different from the adult grammars. As these children grow up and older adults in the population die off, the small changes spread until they are manifested as somewhat larger population level changes.

Interestingly, acquisition and change have conflicting demands. Language acquisition entails rapid convergence to certain parts of the adult grammar by age 2 and so requires clean data as input. Language change, conversely, seems to entail slight misconvergence and so requires dirtier data as input. To be viable, an acquisition proposal must satisfy both these demands. It must allow for "clean enough" data in the input so that children acquire a grammar which allows them to communicate successfully. It must also allow for "dirty enough" data so that small individual changes can occur during acquisition which then result in population change. This places an additional, and quite valuable, constraint on acquisition theory, which we explore here.

Modeling becomes quite useful when we examine acquisition proposals which deal with what data children are learning from (Lightfoot, 1999; Dresher, 1999; Fodor, 1998; Lightfoot, 1991). Such proposals are decidedly difficult to test in a natural setting because we cannot restrict the data set that children are learning from for years on end. However, we can easily manipulate this in an artificial setting. Language change then acts as a population level metric of correct acquisition.

---

*I am immensely grateful to Norbert Hornstein, Amy Weinberg, David Lightfoot, Stephen Crain, Rosalind Thornton, Charles Yang, Partha Niyogi, Tony Kroch, Beatrice Santorini, Cedric Boeckx, and Michelle Hugue for wonderfully useful questions, unwavering encouragement, and very good advice. Additionally, this work has been supported by an NSF Graduate Research Fellowship.

[1]Unlike the lexicon, for instance, which can be increased throughout life.

We will use our quantified model to demonstrate that the restriction of the child's attention to data that are represented by *unambiguous triggers* (Lightfoot, 1999; Dresher, 1999; Fodor, 1998) in *degree-0* clauses (Lightfoot, 1991) can explain the loss of a strong Object-Verb (OV) distribution in Old English between 1000 and 1200 A.D. (based on data from the YCOE corpus of Old English (T,W,P&B 2003) and the PPCME2 corpus of Middle English (Kroch and Taylor)[2]).

2 The Acquisition Proposals

The first acquisition proposal is set in a principles and parameters framework where the adult language consists of a specific set of parameter values (Chomsky, 1981) and acquisition is the process of determining what these parameter values are. An *unambiguous trigger* (Lightfoot, 1999; Dresher, 1999; Fodor, 1998) is a piece of data in the input which unambiguously signals one value over another for a particular parameter. Simply put, an unambiguous trigger is compatible with only one value for a particular parameter, despite whatever other parameters might also be in effect.

A tantalizing benefit for acquisition is the ability to bypass the combinatorial explosion problem of determining values for a set of parameters which interact with each other. With unambiguous triggers, a child is able to separate parameters out one by one since each parameter has its own unambiguous trigger. A potential drawback to this proposal is that the quantity of data which fits this stringent criterion could be very small for a given parameter.

The second proposal restricts attention to *degree-0 input*[3] (Lightfoot, 1991). This proposal is based on the idea that local grammatical relationships provide a lot of information from an acquisition standpoint. The particular local domain of interest is the degree-0 domain, as defined in Lightfoot (1991), which is the matrix clause[4].

The potential benefit of a degree-0 restriction on the input is that matrix clause data is generally messier than embedded clause data[5]—and so we get

[2]Though a sharp shift has been noted in the literature (Lightfoot, 1991, among many others), it was thought to occur earlier, by around 1122 A.D. However, the recent YCOE and PPCME2 corpora suggest the sharpest part of the shift most likely occurring later, between 1150 and 1200 A.D.

[3]Degree refers to the level of embedding—so degree-0 is the matrix clause, degree-1 is the first-level embedded clause, etc.

[4]Though occasionally the front of the embedded clause, as well.

[5]Matrix clause data is dirtier because grammatical processes that take effect during derivation seem to happen more often "upstairs" in the matrix clause than "downstairs" in the embedded clause, e.g. Verb-Second movement in German.

the "dirtier" data which allows language change to fall out.[6] This benefit, however, comes at the cost of acquisition since we restrict the relevant input to only the *degree-0 unambiguous triggers*. Thus, we have reduced the size of the data set that is relevant for acquisition and made our sparse data even sparser. Whether or not the data is, in fact, *too* sparse will be one question we can explore with our quantified model.

3 Language Change: Old English OV Order and Triggers

The language change we use as our population level metric of "correct acquisition" is the shift in Old English between 1000 and 1200 A.D. from a strong OV distribution to a strong VO distribution.[7] By a strong OV distribution at 1000 A.D.,[8] we mean that most of the utterances had the Object before the Verb as in (1). By a strong VO distribution by 1200 A.D., we mean that many of the utterances had the Object after the Verb as in (2).

(1) OV Order (Beowulf, 625)
he **Gode þancode**
[He **God thanked]**
(2) VO order (Blickling Homilies, 187.35)
þa ahof Paulus up **his heafod**
[then **lifted** Paul up **his head]**

Because acquisition drives language change, the loss of Old English's strong OV distribution would be due to a change in the portion of the input which is relevant for acquisition. In this paper, we examine the plausibility of this "relevant portion" being the degree-0 unambiguous triggers.

The structure of the unambiguous triggers themselves is outlined below.

(3) OV order triggers
$_{VP}$[*Object Verb*] or $_{VP}$[*Object Verb-Marker*]
(4) VO order triggers
$_{VP}$[*Verb Object*] or $_{VP}$[*Verb-Marker Object*]

An unambiguous trigger must have the appropriate surface order of the

---

[6]An additional motivation for using only degree-0 input might be that such data is within a child's processing capability, given complexity considerations.

[7]See footnote 2 regarding the literature on the exact time period of this shift.

[8]While originally completely OV order, Old English began to have VO order appear after the Scandinavian settlers came in the 9th century (Kroch and Taylor, 1997) By 1000 A.D., there was a small amount of VO order in Old English.

Object adjacent to either a Verb or a Verb-Marker, which marks the position the Verb occupied before derivation. In addition, an unambiguous trigger must also have an *unambiguous derivation* to produce the surface order. Because there are grammatical processes which move words around during derivation, some surface orders are ambiguous as to which word order they began with before derivation, as in the SVO order in (5).

(5) a. Subject Verb Object t_{Verb}
b. Subject Verb t_{Verb} Object

Here, because of Verb-Second movement[9], an SVO surface order is ambiguous as to which word order it had before derivation—the Verb could have begun on either side of the Object and produced the surface order. This SVO utterance therefore has the correct surface order to be a VO trigger, but lacks an unambiguous derivation[10] to produce that surface order.

To reiterate, an unambiguous trigger has two properties: the correct surface order and an unambiguous derivation to produce that surface order.

Verb-Markers can be crucial to marking whether a Verb was base-generated in a particular position or moved there. Verb-Markers are words associated with the Verb which mark the original position of the Verb before derivation, as we see with the Particle *up* in (6) and (7). The Particle indicates the position from which the Verb moved.

(6) OV order before derivation

þa	*ahof*	Paulus	***his heafod***	***up***
[then	*lifted*	Paul	***his head***	***up***]

---

[9]Verb-Second movement takes the Verb and moves it over other constituents to a projection higher up (i.e. C in German or I in Yiddish (Kroch and Taylor, 1997)).

[10]Because of the many parameters available in languages, one might worry that there is no such thing as an unambiguous derivation. For this reason, we believe that the unambiguous triggers proposal ought to be considered with some notion of "parameter priority" that would order which parameter values are determined when. Presumably, the first set of parameter values determined would be those which are acquired earliest—such as Verb-Second, basic word order, and Verb-Raising. We note that we don't include post-position (which, for instance, would shift the Object after the Verb) in our pool of "earliest parameters" since we are not aware of any evidence that post-position is acquired particularly early. The only parameters considered in order to determine if an utterance has an unambiguous derivation are those in that initial set. It would be entirely possible for an utterance which has an ambiguous derivation with respect to parameters *outside* the initial pool to be considered unambiguous with respect to parameters *inside* the initial pool—and counted during early acquisition as an unambiguous trigger.

(7) VO order before derivation

þa	*ahof*	Paulus	***up***	***his heafod***
[then	lifted	Paul	**up**	**his head**]

We use the class of Verb-Markers described in Lightfoot (1991)—particles, negatives, some closed-class adverbials, and nonfinite complements. Old English Verb-Markers, however, did not always remain adjacent to the Object, so (8) might result.

(8) Unreliable Verb Marker—Unknown order before derivation

þa	***up***	*ahof*	Paulus	***his heafod***
[then	**up**	lifted	Paul	**his head**]

When both the Verb and Verb-Marker move away from the Object during derivation, an utterance which could have been an unambiguous trigger becomes ambiguous as to which word order it denotes. This is *unambiguous trigger destruction*. Because of unambiguous trigger destruction, the distribution of OV and VO order in the portion of the data the child is sensitive to—the degree-0 unambiguous triggers—does not necessarily reflect the distribution of OV and VO order in the general population. Because of this disparity, children would be able to acquire parameter values that are slightly different from those of their parents—small changes that happen during acquisition. As these small changes add up and spread through the exponentially growing population, the sharper changes of population-level language change emerge.

4 The Model: Multiple Grammars and Acquisition

4.1 Multiple Grammars, Data Sparseness, and Questions

The model we use is founded on several ideas explored in previous modeling and historical work. Firstly, grammars can compete during acquisition (Clark and Roberts, 1993) and within a population over time (Pintzuk, 2002, among others) since acquisition is driving the linguistic composition of the population over time. Secondly, because change is happening during acquisition and acquisition is an individual process, population-level change results from individual linguistic behavior—and specifically, individual "misconvergence" to the target parameter value (Niyogi and Berwick, 1997, among other work). Thirdly, individual linguistic behavior can be represented as a statistical distribution of multiple grammars (Yang, 2003), which is the direct result of multiple grammars competing *during acquisition* coupled with the option of still having multiple grammars *at the end of acquisition.*[11]

Multiple grammars in a single individual are instantiated as an individual accessing g grammars with some probability p_g allotted to each (Yang, 2003). In our model, $g = 2$ since we have an OV grammar competing with a VO grammar[12]. A stable system with $g = 1$ has the properties in (9):

(9) Stable System, g = 1
  a. $p_1 = 1$ (this grammar occupies the entire probability space)
  b. all unambiguous triggers come from this one grammar alone

A competitive system, such as the language change scenario in Old English, has the properties in (10):

(10) Competitive System, g = 2
  a. $p1 = \text{Prob}(g1)$, $p2 = \text{Prob}(g2) = 1 - \text{Prob}(g1)$
  b. each grammar leaves some portion of the unambiguous triggers

Because each grammar has some unambiguous triggers in the input and these triggers are for conflicting values, it is likely that seeing approximately equal numbers of each type of trigger would cause them to cancel each other out. Thus, the relevant quantity of triggers for acquisition is *how many more unambiguous triggers* one grammar has in the input than the other—the "advantage" (Yang, 2003) one grammar has over another. If there was a data sparseness problem before, it is certainly much worse now. Moreover, because the acquisition proposal we are pursuing restricts the relevant data set to the degree-0 clauses, we have further exacerbated the sparseness of the data—as we can see from the OV order advantages in Table 1.

Time	Degree-0 Clauses	Degree-1 Clauses
1000 A.D.	6.6%	16.2%

Table 1. OV order advantage, based on data from the YCOE. OV grammar advantage over VO grammar = ([# of unambiguous OV triggers] – [# of unambiguous VO triggers])/([# of clauses total in the input]).

[11]As we will show, this does not necessarily mean that each grammar has equal prominence—one may well be much more in use than the other(s).

[12]Technically, we actually have two *sets* of grammars in our model (those with OV order and those with VO order) since we aren't concerned here with the other parameters in the initial pool. Thus, the OV set = (OV, *V2 value*, *VRaising value*, ...) and the VO set = (VO, *V2 value*, *VRaising value*, ...).

A 6.6% advantage in the degree-0 unambiguous triggers for the OV order grammar means that only about 7 out of every 100 sentences in the input are actually doing any work from an acquisition standpoint. The degree-1 clauses have a much higher OV advantage, but children's input primarily consists of degree-0 clauses.[13]

Given this, we have two questions for our quantified model, using the language change metric which specifies that the population must be strongly OV between 1000 and 1150 A.D. and more strongly VO by 1200 A.D.

(11) *Sufficiency*: Is the restricted data of degree-0 unambiguous triggers sufficient to satisfy the language change metric? That is, are there enough unambiguous triggers in the degree-0 clauses to cause individuals in the population to be strongly OV between 1000 and 1150 A.D. and more strongly VO by 1200 A.D.?

(12) *Necessity*: Are the restrictions themselves necessary?[14]

a. *Unambiguous triggers*: Surface Order *and* Unambiguous Derivation. Unambiguous derivation is more costly to compute. Can a population behave as the Old English population historically did if unambiguous triggers are determined by surface order alone?

b. *Degree-0 Input Only*: Because the degree-1 clauses have a much higher OV order advantage, seeing a sufficient quantity of degree-1 data in the input would prevent a population from being "VO enough" by 1200 A.D. But there is only a small quantity of degree-1 data generally available in children's input—is this small quantity sufficient to prevent an Old English population from being "VO enough" by 1200 A.D.? Is it necessary for children to be restricted to the degree-0 data?

4.2 Acquisition Implementations

The model of individual acquisition is based very strongly on a probabilistic access function of binary values (Bock and Kroch, 1989). For example, an individual could have an access function like the following: 30% of the time use an underlying VO order and 70% of the time use an underlying OV order.[15,16] The acquisition effect is fairly straightforward—because children

[13] 91.2% based on a 4K sentence sample from the CHILDES database (MacWhinney and Snow 1985) and 76.1% based on a 4K sentence sample of young children's stories.

[14] In the sense that this language change pattern would be unexplained if the modeled learner was not restricted by these conditions.

[15] This is a good example of how multiple grammars can exist concurrently and

will see unambiguous OV triggers as well as unambiguous VO triggers, children must also acquire a probabilistic access function in order to account for all the data.

Individuals in the model have a probabilistic access function which is represented as a single value ranging between 0.0 (all OV order) and 1.0 (all VO order). The value represents what percentage of the probability space the VO grammar occupies, e.g. a value of 0.3 = 30% VO usage. There is no default bias towards either word order,[17] so the child begins with the unbiased value of 0.5.

The model also includes two methods which provide a limited memory[18], since we want the probabilistic access function a child acquires to reflect, at least to some extent, the distribution of the unambiguous triggers in the input,[19] To do this, we employ two indirect memory mechanisms: the Noise Filter and the Batch Learner method (Yang, 2003).

The Noise Filter is designed to separate "signal" from "noise", where what counts as "noise" depends on what the current hypothesis is for the probabilistic access function. A trigger for the minority grammar, for instance, will get construed as noise much more often than a trigger for the majority grammar. The Batch Learner method (Yang, 2003) has to do with how many triggers it takes to alter the current hypothesis for the probabilistic access function—in essence, how big a "batch" is required. The batch size depends on what the current hypothesis is so that more triggers from the minority grammar are required to make a batch.[20]

The current hypothesis, on which these two mechanisms are based, is a direct reflection of the relative quantities of triggers which have been seen up until the current point. For example, a hypothesis of 0.3 (the VO grammar occupies 30% of the probability space—down from the initial 50%) can only come about if relatively more OV triggers have been seen than VO triggers. Individual acquisition is implemented using the following algorithm:

---

not be equal—here, the OV grammar is more prominent than the VO grammar.

[16]We note that this is the distribution of the speaker's utterances *before* derivation. Grammatical processes, such as Verb-Second movement, will distort the distribution so that the distribution a child has access to (the surface distribution) would not necessarily reflect a 30–70 split.

[17]In line with theories which suppose no default value for certain parameters (Sugisaki, 2002).

[18]We want only a *limited* memory so that it doesn't require significant resources.

[19]A memoryless model, while nicely low on resource usage, can't tell us anything about the distribution seen up to the current point because it can only give information about the very last thing it saw.

[20]Details of the memory mechanisms are in Pearl (2003).

(13) Individual Acquisition Algorithm

```
Initial setting = 0.5
While in critical period
    Get one datum from input
    If datum contains unambiguous trigger
      Increase relevant batch learner counter
          If counter reaches batch threshold
          Alter current hypothesis
      End if
    End If
End while
```

Because individual acquisition drives the linguistic composition of the population, we can now go straight to the population level implementation.

(14) Population Algorithm

```
Initialization:
    PopulationAgeRange = 0 to 60; PopulationSize = 50,000
At year 1000 A.D.
    Initialize all members to initial population value
    Increment year by 2
While year < 1200 A.D.
    Members age 59 to 60 die
    Increment all other members' age by 2 years
    Create new members in age range 0 to 1
    Determine new members' probabilistic access function values
    // input comes from remaining members age 2 to 60
    Increment year by 2
End While
```

The "population value" is the average distribution of the OV and VO order grammars before derivation in the population. Now, historical data does not reflect the distribution *before* derivation—it reflects the distribution *after* derivation, and importantly, the distribution after trigger destruction has already occurred. If we examine Table 2, we can see that the degree-0 clauses at 1000 A.D. have considerable trigger destruction since so many triggers have become ambiguous after derivation.

Time	Ambiguous	Unambiguous
1000 A.D.	11630	6033

Table 2. Degree-0 triggers, as derived from the YCOE corpus data.

In order to determine the distribution before derivation (in which all triggers are unambiguous), we must decide which of the ambiguous triggers began as OV order and which began as VO order before derivation. To do this, we note that while the degree-0 clause and degree-1 clause distribution are both skewed from the underlying distribution before derivation, the degree-0 distribution is skewed more than the degree-1 distribution.[21] We then use the skew between the degree-0 and degree-1 distribution to estimate the skew between the degree-1 and underlying distribution. In this way, we can step back to the underlying distribution which produced the two surface distributions we actually see in the historical data.[22] Once we do this, we get the population values in Table 3.

Time	1000 A.D	1050-1150 A.D.	1200 A.D.
Value	*0.162*	*0.235*	*0.549*

Table 3. Average population values, derived from YCOE and PPCME2 data.

Thus, we begin our population members at a probabilistic access function value of 0.162 at 1000 A.D. and run our population algorithm until 1200 A.D., when we can check if the population is "VO enough" to match the historical data—i.e. at an average value of 0.549.

5 Results

As we can see from Figure 1, a population constrained to learn from only degree-0 unambiguous triggers is able to support the loss of a strongly OV distribution and become "VO enough" by 1200 A.D. The restricted data set is, in fact, sufficient to satisfy the language change metric. There is not a data sparseness problem. Moreover, we can see sharper change falling out for free from the dynamics of the population itself—the small misconvergences add up over time, spreading through the exponentially growing population.

We next turn to the necessity of each of the restrictions. Our current definition of unambiguous triggers requires the proper surface order and an unambiguous derivation to produce that surface order. Do we need to have the unambiguous derivation criterion, particularly since it requires some resources to determine if a derivation is unambiguous or not? Suppose we simply defined unambiguous triggers as those with the proper surface order.

[21]This is apparent since the degree-0 clauses always have more trigger destruction than the degree-1 clauses, according to the YCOE data.

[22]For a detailed description of how this is done, see Pearl (2003).

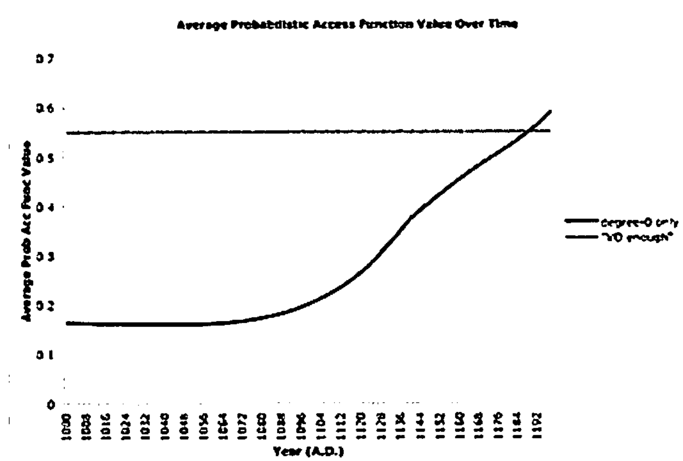

Figure 1. The trajectory of a population constrained to learn from only the degree-0 unambiguous triggers.

Then, an utterance of the form in (13) would be an unambiguous VO order trigger, even though it has an ambiguous derivation.

(13) Subject Verb Object
= Subject Verb t_{Verb} Object?
= Subject Verb Object t_{Verb}?

If we use this definition of unambiguous triggers, we get the advantages for the *VO* grammar listed in Table 4.

Time	*VO* Advantage
1000 A.D.	*1.8%*
1050-1150 A.D.	*6.7%*

Table 4. Advantages for the *VO* grammar at different time periods.

The problem we have now is that it is the *VO* grammar, *not* the OV grammar, with an advantage early on—and a population of individuals learning from input with this VO advantage would have lost a strong OV distribution much earlier than the Old English population historically did. Thus, it seems we need to include unambiguous derivation as part of our definition of unambiguous trigger in order to get the historical facts right. Not only is the restricted data sufficient, but the restriction itself seems to be necessary.

We turn now to the degree-0 restriction. Recall that the degree-1 data has a much higher OV advantage than the degree-0 data, so if enough of the degree-1 data was available in children's input, a population might not be able to shift away from a strongly OV distribution quickly enough to match the historical facts. However, also recall that only a small quantity of degree-1 data is actually available in the input—and so that small quantity of degree-1 data might not have very much effect. What we can determine with our quantified model is how much degree-1 data in the input it takes to have an effect, i.e. to cause an Old English population to not be "VO enough" by 1200 A.D. We see these results in Figure 2.

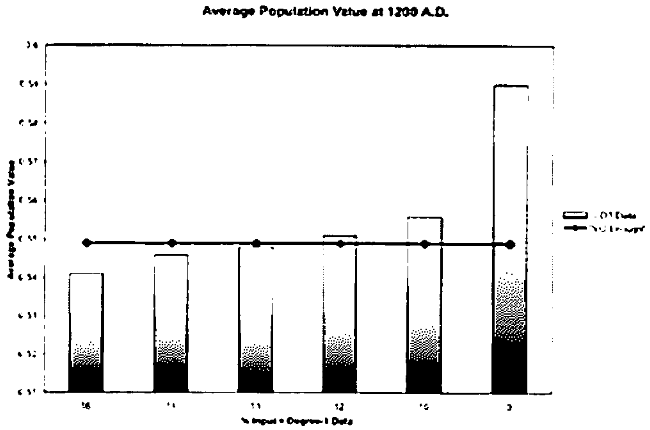

Figure 2. The average population value at 1200 A.D. for six populations with degree-1 data comprising different amounts of the input.

The threshold of permissible degree-1 data in the input appears to be around 13%. However, the average amount of degree-1 data available to modern English children seems to be about 16%.[23] If we assume the amount of degree-1 data available in the input doesn't change over time, then Old English children also would have had about 16% of their input made up of degree-1 data. However, as we can see from Figure 2, a population which sees 16% degree-1 data in their input is unable to be "VO enough" by 1200 A.D. This would suggest that Old English children must not have looked at the degree-1 data available to them. Thus, not only *can* we have the degree-0

[23]8.8% based on a 4K sentence sample from the CHILDES database (MacWhinney and Snow, 1985) and 23.9% based on a 4K sentence sample of young children's stories. We take the average to get ~ 16.4%.

restriction, but in fact we *must* have it in order to satisfy the language change metric.

6 Conclusions

Using a quantified model, we have been able to provide some evidence for both the viability and the necessity of learning from only degree-0 unambiguous triggers. This acquisition proposal appears to be able to satisfy the language acquisition demands, which require the child to get close enough to the target grammar to communicate effectively, while also satisfying the language change demands, which require the child *not* to get the *exact* target grammar. In future work, we hope to test this acquisition proposal with other language change metrics, such as the loss of a strong Verb-Raising distribution in Middle English (Lightfoot, 1999, among others). In addition, we can use the current quantified model to refine the notion of what an early unambiguous trigger is by including or excluding possible parameters from the initial pool considered to determine if a derivation is unambiguous.

References

Bock, Kathryn and Anthony Kroch. 1989. The isolability of syntactic processing. In G. Carlson and M. Tannenhaus, eds., *Linguistic Structure in Language Processing*. Boston: Kluwer.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Clark, Robin and Ian Roberts. 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24:299–345.

Dresher, Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30:27–67.

Fodor, Janet. 1998. Unambiguous triggers. *Linguistic Inquiry* 29:1–36.

Kroch, Anthony and Ann Taylor. 1997. Verb movement in Old and Middle English: Dialect variation and language contact. In A. van Kemenade and N. Vincent, eds., *Parameters of Morphosyntactic Change*. Cambridge: Cambridge University Press, 297–325

Kroch, Anthony and Ann Taylor. The Penn-Helsinki Parsed Corpus of Middle English, second edition. (PPCME2).

Lightfoot, David. 1991. *How to Set Parameters*. Cambridge, MA: MIT Press.

Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution*. Oxford: Blackwell.

MacWhinney, Brian and Catherine Snow. 1985. The Child Language Data Exchange System. *Journal of Child Language* 12:271–96.

Niyogi, Partha and Robert Berwick. 1997. Evolutionary consequences of language learning. *Linguistics and Philosophy* 20:697–719.

Pearl, Lisa. 2003. What Language Change Modeling Can Do for Acquisition Theory. Ms, University of Maryland.

Pintzuk, Susan. 2002. Verb-Object Order in Old English: Variation as Grammatical Competition. *Syntactic Effects of Morphological Change*. Oxford: Oxford University Press.

Sugisaki, Koji. 2002. Parameters and their default values: A case study in the acquisition of preposition stranding. *The 20th Annual Meeting of the English Linguistics Society of Japan*. Tokyo: Japan.

Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose. (YCOE).

Yang, Charles. 2003. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Department of Linguistics
University of Maryland
College Park, MD 20742
llsp@wam.umd.edu