The Input for Syntactic Acquisition: Solutions from Language Change Modeling

Lisa Pearl
Linguistics Department, University of Maryland
1401 Marie Mount Hall, College Park, MD 20742
llsp@wam.umd.edu

Abstract

Sparse data is a well-known problem for any probabilistic model. However, recent language acquisition proposals suggest that the data children learn from is heavily restricted - children learn only from **unambiguous triggers** (Fodor 1998, Dresher 1999, Lightfoot 1999) and **degree-0 data** (Lightfoot 1991). Surprisingly, we show that these conditions are a necessary feature of an accurate language acquisition model. We test these predictions indirectly by developing a mathematical learning and language change model inspired by Yang's (2003, 2000) insights. Our logic is that, besides accounting for how children acquire the adult grammar so quickly, a viable acquisition proposal must also be able to account for how populations change their grammars over time. The language change we examine is the shift in Old English from a strongly Object-Verb (OV) distribution to a strongly Verb-Object (VO) distribution between 1000 A.D. and 1200 A.D., based on data from the YCOE Corpus (Taylor et al. 2003) and the PPCME2 Corpus (Kroch & Taylor 2000). Grounding our simulated population with these historical data, we demonstrate that these acquisition restrictions seem to be both sufficient and necessary for an Old English population to shift its distribution from strongly OV to strongly VO at the right time.

1 Introduction

Empirically investigating what data children attend to during syntactic acquisition is a difficult task. Traditional experimental methods are not feasible on logistical and ethical grounds – we can't simply lock a group of children in a room for two years, restrict their input to whatever we want, and then see if their syntactic acquisition matches normal patterns. However, when we have a simulated group of language learners who follow a quantified model of individual acquisition, this is exactly what we can do – restrict the input to syntactic acquisition in a very specific way and then observe the results.

The individual acquisition model we use is inspired by Yang's (2003, 2000) model of probabilistic learning for multiple grammars. By using this model in a simulated population of individuals, we provide empirical support for two acquisition proposals that restrict children to only heed data that are **unambiguous triggers** (Dresher 1999, Lightfoot 1999, Fodor 1998) and that appear in **degree-0** clauses (Lightfoot 1991). We use language change as a metric of "correct" acquisition, based on the following idea: if the simulated population that has these restrictions behaves just as the real population historically did, the simulated acquisition process is fairly similar to the real acquisition process. Language change is an excellent yardstick for acquisition proposals for exactly this reason – any theory of acquisition must not only be able to account for how children converge to a close approximation of the adult grammar, but also how they can "misconverge" slightly and allow specific types of grammatical change over time. The nature of this "misconvergence" is key. Children must end up with an Internal Language ("grammar") that is close enough - but not *too* close - to the Observable Language (O-Language) in the population so that change can happen at the right pace.

The language change we use as our metric is the shift in Old English from a strongly Object-Verb (OV) distribution to a strongly Verb-Object (VO) distribution between 1000 and 1200 A.D.

Proceedings of the Second Workshop on Psychocomputational Models of Human Language Acquisition, pages 1–9,
Ann Arbor, June 2005.

The sharpest part of this shift occurs between 1150 and 1200 A.D., based on data from the YCOE Corpus (Taylor et al. 2003) and the PPCME2 Corpus (Kroch & Taylor 2000). We use this corpus data to estimate the initial OV/VO distribution in the modeled population at 1000 A.D. and to calibrate the modeled population's projected OV/VO distribution between 1000 and 1150 A.D. Then, we demonstrate that the restrictions on acquisition seem both sufficient and surprisingly necessary for the simulated Old English population to shift its distribution to be strongly VO by 1200 A.D – and thus match the historical facts of Old English. In this way, we provide empirical support that we would be hard-pressed to get using traditional methods for these acquisition proposals.

The rest of the paper is laid out as follows: section 2 elaborates on the two acquisition proposals of **unambiguous triggers** and **degree-0 data**; section 3 gives specific implementations of these proposals for Old English; section 4 describes the model used to simulate the population of Old English speakers and how the historical corpus data was used; sections 5 and 6 present the results and conclusion.

2 The Acquisition Proposals

The first proposal is set in a Principles and Parameters framework (Chomsky 1981) where the adult grammar consists of a specific set of parameter values and the process of acquisition is figuring out what those values are. An unambiguous trigger (Fodor 1998, Dresher 1999, Lightfoot 1999) is a piece of data from the O-language that unambiguously signals one parameter value over another for a given parameter. Crucially, an unambiguous trigger for value P_1 of parameter P can be parsed only with value P_1 (and not P_2), no matter what *other* parameter values (A, B, C, …) might also be affecting the O-language form of the data. Because an unambiguous trigger corresponds to exactly one parameter P and thus can alter the value of P only, this proposal would allow children to bypass the Credit Problem noted by Dresher (1999), which is the problem of deciding which parameters to update given a particular piece of input. In addition, unambiguous triggers allow the learner to bypass the combinatoric explosion problem that could occur when trying to set **n** parameters. Instead of having to test out $\mathbf{2^n}$ different grammars on the input in the O-languages, the child's language acquisition mechanism simply tests out the n parameters separately by looking for unambiguous triggers for these n parameters in the input from the O-language. Thus, this proposal aids the process of acquiring the adult grammar quickly and correctly. A potential pitfall of this proposal is data sparseness: the quantity of data that fits this very specific restriction might be very small for a parameter P and the child just might not see enough of it for it to have an effect[1].

The second proposal is that children only heed data in **degree-0** clauses (Lightfoot 1991) when they first begin to set their syntactic parameter values. "Degree" refers to the level of embedding, so a degree-0 clause corresponds to a main clause[2].

(1) Jack thought the giant was easy to fool.
[--Degree-0-]
[---------Degree-1---------]

The basis for this proposal is that while local grammatical relationships (such as those in degree-0 clauses) provide a lot of information to the learner, degree-0 data tends to be "messier" grammatically – that is, more grammatical processes seem to apply to degree-0 clauses than to degree-1 clauses. The messier status of this data allows the child to converge to a grammar that is *not* exactly the same as the adult grammar. Thus, this proposal focuses on how to allow small grammatical changes to occur in individuals so that larger changes can happen to the population over time. The cost of combining this proposal with the previous one is that the child is now restricted to learn only from **degree-0 unambiguous triggers**, thereby compounding the potential data sparseness problem that unambiguous triggers already have.

1 In fact, it may well be necessary to restrict the set of parameters relevant for determining if a trigger is unambiguous to some initial pool in order to get any unambiguous triggers at all. A candidate set for the initial pool of parameters might be derived from a hierarchy of parameters along the lines of the one based on cross-linguistic comparison that is described in Baker (2001, 2005).

2 The exact domain of a degree-0 clause is defined as the main clause and the front of the embedded clause for theory-internal reasons. For a more detailed description and explanation, see Lightfoot (1991).

3 Old English Change

Allowing language change to occur as it historically did is a mark of "correct" acquisition, especially for change involving syntactic parameters that can only be altered during acquisition - any change that builds up in the population must be due to changes that occur during acquisition. The parameter we use in this work is OV/VO word order and the change is a shift in Old English from a strongly OV distribution between 1000 and 1150 A.D. to a strongly VO distribution at 1200 A.D. A strongly OV distribution has many utterances with OV order (2). A strongly VO distribution as many utterances with VO order (3).

(2) he Gode þancode
he God thanked
'He thanked God'
(*Beowulf,* 625)

(3) þa ahof Paulus up his heafod
then lifted Paul up his head
'Then Paul lifted his head up'
(*Blickling Homilies,* 187.35)

Because change can occur only during acquisition, the data children are heeding in their input during acquisition has a massive effect on the population's linguistic composition over time. In this work, we explore the possibility that the data children are heeding during acquisition are the degree-0 unambiguous triggers. For Old English, the unambiguous triggers have the form of (4a) and (5a). Examples of unambiguous triggers of each kind are in (4b-c) and (5b-c).

(4a) Unambiguous OV Trigger
[**Object Verb/Verb-Marker**]$_{VP}$

(4b) he$_{Subj}$ [**hyne**$_{\mathbf{Obj}}$ **gebidde**$_{\mathbf{VerbFinite}}$
Subj **Obj** **Verb**
[mid ennum mode]$_{PP}$]$_{\mathbf{VP}}$
PP
(*Ælfric's Letter to Wulfsige,* 87.107)

(4c) we$_{Subj}$ sculen$_{VerbFinite}$ **[[ure yfele +teawes]**$_{\mathbf{Obj}}$
Subj Verb **Obj**
forl+aten$_{\mathbf{Verb\text{-}Marker}}$]$_{\mathbf{VP}}$
Verb-Marker
(*Alcuin's De Virtutibus et Vitiis,* 70.52)

(5a) Unambiguous VO Trigger
[**Verb/Verb-Marker Object**]$_{VP}$

(5b) & [mid his stefne]$_{PP}$ he$_{Subj}$ [**awec+d**$_{\mathbf{VerbFinite}}$
PP Subj **Verb**
deade$_{\mathbf{Obj}}$ [to life]$_{PP}$]$_{VP}$
Obj PP
(*Saint James,* 30.31)

(5c) þa$_{\mathbf{Adv}}$ ahof$_{\mathbf{VerbFinite}}$ Paulus$_{\mathbf{Subj}}$ [**up**$_{\mathbf{Verb\text{-}Marker}}$
Adv Verb Subj **Verb-Marker**
[his heafod]$_{\mathbf{Obj}}$**]** $_{\mathbf{VP}}$
Obj
(*Blickling Homilies,* 187.35)

The Object is adjacent to either a Verb or a Verb-Marker on the appropriate side – the correct O-language order. In addition to this correct "surface order" in the O-language, an unambiguous trigger must also have an **unambiguous derivation** to produce this surface order. This means that no other combination of parameters with the alternate word order value could produce the observed surface order. For example, a *Subject Verb Object* utterance could be produced more than one way because of the Verb-Second (V2) movement parameter which was also available in Old English (as in modern Dutch and German). With V2 movement, the Verb moves from its "underlying" position to the second position in the sentence.

(6a) V2 Movement Ambiguity
Subject Verb Object t_{Verb}. (OV + V2)
Subject Verb t_{Verb} *Object.* (VO + V2)

(6b) Subject Verb Object example
heo$_{\mathbf{Subj}}$ cl+ansa+d$_{\mathbf{VerbFinite}}$
Subj Verb
[+ta sawle +t+as r+adendan]$_{\mathbf{Object}}$
Obj
(*Alcuin De virtutibus et vitiis*, 83.59)

(6c) *Subject Verb t_{Subj}* ***Object t_{Verb}***.
(parsed with OV + V2)

heo$_{Subj}$ cl+ansa+d$_{VerbFinite}$ t_{Subj}
Subj Verb
[[+ta sawle +t+as r+adendan]$_{Obj}$
Obj
***$t_{VerbFinite}$*]$_{VP}$**
trace-Vb

(6d) *Subject Verb t_{Subj}* ***t_{Verb} Object***.
(parsed with VO + V2)

heo$_{Subj}$ cl+ansa+d$_{VerbFinite}$ t_{Subj}
Subj Verb
[*$t_{VerbFinite}$* [+ta sawle +t+as r+adendan]$_{Object}$]$_{VP}$
trace-Vb **Obj**

Because of this, a *Subject Verb Object* utterance can be parsed with either word order (OV or VO) and so cannot unambiguously signal either order. Thus, correct surface order alone does not suffice – only an utterance with the correct surface order and that cannot be generated with the competing word order value is an unambiguous trigger[3].

Because V2 movement (among other kinds of movement) can move the Verb away from the Object, Verb-Markers can be used to determine the original position of the Verb with respect to the Object. Verb-Markers include **particles** ('*up*'), **non-finite complements** to finite verbs ('shall…*perform*'), some **closed-class adverbials** ('*never*'), and **negatives** ('*not*') as described in Lightfoot (1991).

The curious fact about Old English Verb-Markers (unlike their modern Dutch and German counterparts) is that they were unreliable – often they moved away from the Object as well, leaving nothing Verb-like adjacent to the Object. This turned utterances which potentially were unambiguous triggers for either OV or VO order into ambiguous utterances which could not help acquisition. We term this "trigger destruction," and it has the effect of making the distribution of OV and VO utterances that the child uses during acquisition (the distribution in the degree-0 unambiguous triggers) *different* from the distribution of the OV and VO utterances in the population. It is this difference that "biases" children away from the distribution in the population and it is this difference that will cause small grammatical changes to accumulate in the population until the larger change emerges – the shift from being strongly OV to being strongly VO. Thus, the question of what data children heed during acquisition has found a very suitable testing ground in Old English.

[3] We note that this could potentially be very resource-intensive to determine since all other interfering parameter values (such as V2) must be taken into account. Hence, there is need for some restriction of what parameters must be initially considered to determine if an utterance contains an unambiguous trigger for a given parameter.

4 The Model

4.1 The Acquisition Model & Old English Data

The acquisition model in this work is founded on several ideas previously explored in the acquisition modeling and language change literature. First, grammars with opposing parameter values (such as OV and VO order) compete with each other both during acquisition (Clack & Robert 1993) and within a population over time (Pintzuk 2002, among others). Second, population-level change is the result of a build-up of individual-level "misconvergences" (Niyogi & Berwick, 1997, 1996, 1995). Third, individual linguistic behavior can be represented as a probabilistic distribution of multiple grammars. This is the result of multiple grammars competing during acquisition and still existing *after* acquisition.

Multiple grammars in an individual are instantiated as that individual accessing g grammars with probability p_g each (Yang 2003). In our simulation, there are two grammars (g = 2) – one with the OV/VO order set to OV and one with the OV/VO order set to VO. In a stable system with g=1, g_1 has probability p_{g1} = 1 of being accessed and all unambiguous triggers come from this grammar. In the unstable system for our language change, g=2 and g_1 is accessed with probability p_{g1} while g_2 is accessed with probability $p_{g2} = 1 - p_{g1}$. *Both* grammars leave unambiguous triggers in the input to the child.

If the quantity of unambiguous triggers from each grammar is approximately equal, these quantities will effectively cancel each other out (whatever quantity pulls the child towards OV will be counterbalanced by the quantity of triggers pulling the child towards VO). Therefore, the crucial quantity is *how many more* unambiguous triggers one grammar has than the other, since this is the quantity that will not be cancelled out. This is the *advantage* a grammar has over another in the input. Table 1 shows the advantage in the degree-0 (D0) clauses and degree-1 (D1) clauses that the OV grammar has over the VO grammar in Old English at various points in time, based on the data from the YCOE (Taylor et al. 2003) and PPCME2 (Kroch & Taylor 2000) corpora.

	D0 OV Adv	D1 OV Adv
1000 A.D.	**1.6%**	**11.3%**
1000 – 1150 A.D	**0.2%**	**7.7%**
1200 A.D.	**-0.4%**[4]	**-19.1%**

Table 1. OV grammar's advantage in the input for degree-0 (D0) and degree-1 (D1) clauses at various points in time of Old English.

The corpus data shows a 1.6% advantage for the OV grammar in the D0 clauses at 1000 A.D. – which means that only 16 out of every 1000 sentences in the input are actually doing any work for acquisition (and more specifically, pulling the child towards the OV grammar). The data also show that the D1 advantage is much stronger. However, this does not help our learners for two reasons:

a) Based on samples of modern children's input (4K from CHILDES (MacWhinney & Snow 1985) and 4K from young children's stories (for details on this data, see Pearl (2005)), D1 clauses only make up ~16% of modern English children's input. If we assume that the quantity of D1 input to children is approximately the same no matter what time period they live in[5], then our Old English children also heard D1 data in their input ~16% of the time.

b) Our learners can only use D0 data, anyway.

This leads to two questions for the restrictions imposed by the acquisition proposals - a question of sufficiency and a question of necessity. First, we can simply ask if these restrictions on the data children heed are sufficient to allow the Old English population to shift from OV to VO at the appropriate time. Then, supposing that they are, we can ask if these restrictions are necessary to get the job done – that is, will the population shift correctly even if these restrictions do not hold? We can relax both the restriction to learn only from unambiguous triggers and the restriction to learn only from degree-0 clause data – and then see if the population can still shift to a strongly VO distribution on time.

4.2 The Acquisition Model: Implementation

The acquisition model itself is based around the idea of probabilistic access function of binary parameter values (Bock & Kroch 1989) in an individual. For example, if an individual has a function that accesses the VO order value 30% of the time, the utterances generated by that individual would be VO order 30% of the time and OV order 70% of the time. Note that this is the distribution *before* other parameters such as V2 movement alter the order, so the O-language distribution produced by this speaker is *not* 30-70. However, the O-language distribution will still have some unambiguous OV triggers and some unambiguous VO triggers, so a child hearing data from this speaker will have to deal with the conflicting values. Thus, a child will have a probabilistic access function to account for the OV/VO distribution– and acquisition is the process of setting what the VO access probability is, based on the data heard during the critical period.

The VO access value ranges from 0.0 (all OV access) to 1.0 (all VO access). A value of 0.3, for example, would correspond to accessing VO order 30% of the time. A child begins with this value at 0.5, so there is a 50% chance of accessing either OV or VO order.

Two mechanisms help summarize the data the child has seen so far without using up computing resources: the Noise Filter and a modified Batch Learner Method (Yang 2003). The Noise Filter acts as a buffer that separates "signal" from "noise". An unambiguous trigger from the minority grammar is much more likely to be construed as "noise" than an unambiguous trigger from the majority grammar. An example use is

[4] A negative advantage for OV advantage means the VO grammar has the advantage.

[5] At this point in time, we are unaware of any studies that suggest that the composition of motherese, for example, has altered significantly over time.

below with the VO access value set to 0.3 (closer to pure OV than pure VO):

6) Noise Filter Use
probabilistic value of VO access = 0.3
if next unambiguous trigger = VO
    = "noise" with 70% chance and ignored
    = "signal" with 30% chance and heeded
if next unambiguous trigger = OV
    = "noise" with 30% chance and ignored
    = "signal" with 70% chance and heeded

The initial value of VO access of 0.5, so there is no bias for either grammar when determining what is "noise" and what is "signal". The modified Batch Learner method deals with how many unambiguous triggers it takes to alter the child's current VO access value. The more a grammar is in the majority, the smaller the "batch" of its triggers has to be to alter the VO access value (see Table 2). The current VO access value is used to decide whether a grammar is in the majority, and by how much.

VO Value	OV Triggers Required	VO Triggers Required
0.0-0.2	1	5
0.2-0.4	2	4
0.4-0.6	3	3
0.6-0.8	4	2
0.8-1.0	5	1

Table 2. How many unambiguous triggers from each grammar are required, based on what the current VO access value is for the child.

Below is an example of the modified Batch Learner method with the VO access value set to 0.3:

7) modified Batch Learner method use
probabilistic value of VO access = 0.3
if next unambiguous trigger = VO
    if 4th VO trigger seen,
        alter value of VO access towards VO
else if next unambiguous trigger = OV
    if 2nd OV trigger seen,
        alter value of VO access towards OV

The initial value of 0.5 means that neither grammar requires more triggers than the other at the beginning to alter the current value.

Both mechanisms rely on the probabilistic value of VO access to reflect the distribution of triggers seen so far. The logic is as follows: in order to get to a value below 0.5 (more towards OV), significantly more unambiguous OV triggers must have been seen; in order to get to a value above 0.5 (more towards VO), significantly more unambiguous VO triggers must have been seen.

The individual acquisition algorithm used in the model is below:

Initial value of VO access = 0.5
While in critical period
    Get a piece of input from the linguistic environment created by the rest of the population members.
    If input is an unambiguous trigger
        If input passes through Noise Filter
            Increase relevant batch counter
            If counter is at threshold
                Alter current VO access value

Note that the final VO access value after the critical period is over does not have to be 0.0 or 1.0 – it may be a value in between. It is supposed to reflect the distribution the child has heard, not necessarily be one of the extreme values.

4.3 Population Model: Implementation

Since individual acquisition drives the linguistic composition of the population, the population algorithm centers around the individual acquisition algorithm:

Population age range = 0 to 60
Initial population size = 18000[6]
Initialize members to starting VO access value[7]
At 1000 A.D. and every 2 years until 1200 A.D.
    Members age 59-60 die; the rest age 2 years
    New members age 0 to 1 created
        New members use individual acquisition algorithm to set their VO access value

[6] Based on estimates from Koenigsberger & Briggs (1987).
[7] Based on historical corpus data.

4.4 Population Values from Historical Data

We use the historical corpus data to initialize the average VO access value in the population at 1000 A.D., calibrate the model between 1000 and 1150 A.D., and determine how strongly VO the distribution has to be by 1200 A.D. However, note that while the VO access value reflects the OV/VO distribution *before* interference from other parameters causes utterances to become ambiguous, the historical data reflects the distribution *after* this interference has caused utterances to become ambiguous. Table 3 shows how much of the data from the historical corpus is comprised of ambiguous utterances.

Time Period	D0 % Ambig	D1 % Ambig
1000 A.D.	76%	28%
1000-1150 A.D.	80%	25%
1200 A.D.	71%	10%

Table 3. How much of the historical corpus is comprised of ambiguous utterances at various points in time.

We know that either OV or VO order was used to generate all these ambiguous utterances – so our job is to estimate how many of them were generated with the OV order and how many with the VO order. This determines the "underlying" distribution. Once we know this, we can determine what VO access value produced that underlying OV/VO distribution. Following the process detailed in Pearl (2005), we rely on the fact that the D0 distribution is more distorted than the D1 distribution (since the D0 distribution always has more ambiguous triggers). The process itself involves using the difference in distortion between the D0 and D1 distribution to estimate the difference in distortion between the D1 and underlying distribution. Once this is done, we have average VO access values for initialization, calibration, and the target.

Time A.D.	1000	1000-1150	1200
Avg VO	.23	.31	.75

Table 4. Average VO access value in the population at various points in time, based off historical corpus data.

Thus, to satisfy the historical facts, a population must start with an average VO access value of 0.23 at 1000 A.D., reach an average VO access value of 0.31 between 1000 and 1150 A.D., and reach an average VO access value of 0.75 by 1200 A.D.

5 Results

5.1 Sufficient Restrictions

Figure 1 shows the average VO access value over time of an Old English population restricted to learn only from degree-0 unambiguous triggers. These restrictions on acquisition seem sufficient to get the shift from a strongly OV distribution to a strongly VO distribution to occur at the right time. We also note that the sharper population-level change emerges after a build-up of individual-level changes in a growing population.

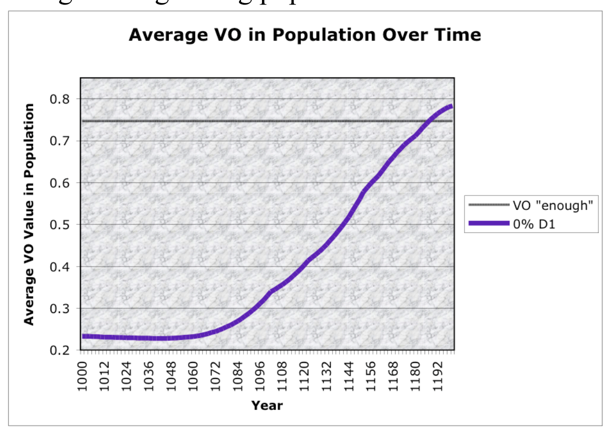

Figure 1. The trajectory of a population restricted to learn only from degree-0 unambiguous triggers.

Thus, we have empirical support for the acquisition proposal since it can satisfy the language change constraints for Old English word order.

5.2 Necessary Restrictions

5.2.1 Unambiguous Triggers

We have shown these restrictions – to learn only from degree-0 unambiguous triggers - are sufficient to get the job done. But are they necessary? We examine the "unambiguous" aspect first – can we still satisfy the language change constraints if we don't restrict ourselves to unambiguous triggers? This is especially attractive since it may be resource-intensive to determine if an utterance is unambiguous. Instead, we might try simply using surface word order as a trigger. This would create many more triggers in the input - for instance, a *Subject Verb Object* utterance would now be parsed as a VO trigger. Using this definition of trigger, we get the following data from the historical corpus:

	D0 **VO** Advantage
1000 A.D.	4.8%
1000 – 1150 A.D.	5.5%
1200 A.D.	8.5%

Table 5. Advantage for the VO grammar in the degree-0 (D0) clauses at various times, based on data from the historical corpus.

The most salient problem with this is that even at the earliest point in time when the population is supposed to have a strongly OV distribution, it is the VO grammar – and *not* the OV grammar – that has a significant advantage in the degree-0 data. A population learning from this data would be hard-pressed to remain OV at 1000 A.D., let alone between 1000 and 1150 A.D. Thus, this definition of trigger will not support the historical facts – we *must* keep the proposal which requires unambiguous triggers.

5.2.2 Degree-0 Data

We turn now to the degree-0 data restriction. Recall that the degree-1 data has a much higher OV advantage before 1150 A.D. (see Table 1). It's possible that if children heard enough degree-1 data, the population as a whole would remain OV too long and be unable to shift to a VO "enough" distribution by 1200 A.D. However, the average amount of degree-1 data available to children is about 16% of the input, based on estimates from modern English children's input. Is this small amount enough to keep the Old English population OV too long? With our quantified model, we can determine if 16% degree-1 data causes our population to not be VO "enough" by 1200 A.D. Moreover, we can estimate what the threshold of permissible degree-1 data is so that the modeled Old English population can match the historical facts. Figure 2 displays the average VO access value in 5 Old English populations exposed to different amounts of degree-1 data during acquisition. As we can see, the population with 16% of the input comprised of degree-1 data is *not* able to match the historical facts and be VO "enough" by 1200 A.D. Only populations with 11% or less degree-1 data in the input can.

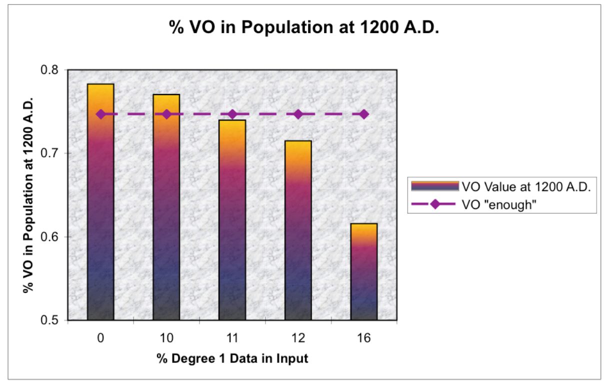

Figure 2: Average VO access values at 1200 A.D. for populations with differing amount of degree-1 data available during acquisition.

This data supports the necessity of the degree-0 restriction since the amount of degree-1 data children hear on average during acquisition (~16%) is too much to allow the Old English population to shift at the right time.

6 Conclusions

Using a probabilistic model of individual acquisition to model a population's language change, we demonstrate the sufficiency and necessity of certain restrictions on individual acquisition. In this way, we provide empirical support for a proposal about what data children are learning from for syntactic acquisition – the degree-0 unambiguous triggers.

Future work will refine the individual acquisition model to explore the connection between the length of the critical period and the parameter in question, including more sophisticated techniques of Bayesian modeling (to replace the current mechanisms of Noise Filter and Batch Learner), and investigate what parameters must be considered to determine if a trigger is "unambiguous". As well, we hope to test the degree-0 unambiguous trigger restriction for other parameters with documented language change, such as the loss of V2 movement in Middle English (Yang 2003, Lightfoot 1999, among others). This type of language change modeling may also be useful for testing proposals about what the crucial data is for phonological acquisition.

Acknowledgement

I am immensely grateful to Amy Weinberg, Charles Yang, Garrett Mitchener, David Lightfoot, Norbert Hornstein, Stephen Crain, Rosalind Thornton, Tony Kroch, Beatrice Santorini, Ann Taylor, Susan Pintzuk, Philip Resnik, Cedric Boeckx, Partha Niyogi, Michelle Hugue, and the audiences of the 28th PLC, DIGS VIII, ICEHL 13, the Cognitive Neuroscience of Language Lunch Talks, and the 2003 Maryland Student Conference. Their good questions, unwavering encouragement, and sensible advice were invaluable. That I may have ignored some of it is my own fault. This work is supported by an NSF Graduate Fellowship.

References

Baker, M. (2001). *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York, NY: Basic Books.

Baker, M. (2005). Mapping the Terrain of Language Learning. *Language Learning and Development*, 1: 93-129.

Bock, J. & A. Kroch. 1989. The Isolability of Syntactic Processing. In *Linguistic Structure in Language Processing* . Edited by G. Carlson and M. Tannenhaus. Boston: Kluwer.

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

Clark, Robin & 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 D. (1998). Unambiguous Triggers. *Linguistic Inquiry,* 29:1-36.

Gibson, Edward & Wexler, Kenneth. (1994). Triggers. *Linguistic Inquiry, 25,* 407-454.

Koenigsberger, H.G. & Briggs, A. (1987). *Medieval Europe, 400-1500.* Longman: New York.

Kroch, Anthony and Taylor, Ann. (2000). The Penn-Helsinki parsed corpus of Middle English. Philadelphia: Department of Linguistics, University of Pennsylvania, 2nd edn. Accessible via http://www.ling.upenn.edu/mideng

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, B. & C. Snow. (1985). The Child Language Data Exchange System. *Journal of Child Language* 12: 271-96.

Niyogi, Partha & Berwick, Robert C. (1995). *The logical problem of language change.* AI-Memo 1516, Artificial Intelligence Laboratory, MIT.

Niyogi, Partha & Berwick, Robert C. (1996). A language learning model for finite parameter spaces. *Cognition, 61,* 161-193.

Niyogi, Partha & Berwick, Robert C. (1997). Evolutionary consequences of language learning. *Linguistics and Philosophy, 20,* 697-719.

Pearl, Lisa. (2005). The Input to Syntactic Acquisition: Answer from Language Change Modeling. Unpublished Manuscript, University of Maryland, College Park. http://www.wam.umd.edu/~llsp/papers/InputSynAcq.pdf

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

Taylor, A., Warner, A., Pintzuk, S., and Beths, F. (2003). The York-Toronto-Helsinki parsed corpus of Old English. York, UK: Department of Language and Linguistic Science, University of York. Available through the Oxford Text Archive.

Yang, Charles. (2000). Internal and external forces in language change. *Language Variation and Change,* 12: 231-250.

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