Learning Parameter Setting from Ambiguous Evidence: Parameter Interaction and the Case of Korean

Isaac Gould

1. Introduction

In this paper my goal is to investigate the following acquisition questions. Broadly, what can be learned from ambiguous evidence? Ambiguous evidence is input to the learner that is compatible with multiple grammars or parameter settings. More specifically, what learning outcomes or parameter settings do we expect if (essentially) all of the learner's input is ambiguous? That is, what similarities do we expect to see across grammars for learners of the same language? And what points of variability do we expect to see, and under what conditions?

I will illustrate the puzzle of ambiguous evidence by showing that there is a high degree of ambiguity in canonically verb-final languages. In particular, I will show that there is ambiguity for two kinds of parameters: parameters concerning head-complement order and parameters concerning verb raising (i.e. head movement). By focusing on Korean, I will claim that this ambiguity can be found in essentially all of the learner's input. Such ambiguity poses a problem for learning models such as Sakas and Fodor (2001), in which only unambiguous evidence can be used to set parameters.

I will address the puzzle of ambiguous evidence with a probabilistic learning model. This model will provide a proof-of-concept illustration of how the acquisition questions above can be answered. We will see that the model will learn the grammar(s) of best fit to the input. Further, learning is systematically affected by *parameter interaction*. Parameter interaction will be illustrated in detail below, but roughly it amounts to the choice of some particular parameter value affecting the choice of some other parameter value. In particular we will see that the model is highly successful at learning a consistently head-final grammar, but that there is variability as to whether there is verb movement. Sometimes the model learns a verb raising grammar, while other times it does not. This model builds on the probabilistic work of Yang (2002), but goes beyond Yang's work in several ways, two of which I mention here. The reader is referred to Gould (2015) for a more detailed comparison with Yang's model. First, the model here is designed to address the learnability problem of subset languagess (Gold 1967; Wexler and Manzini 1987), which as I discuss in Gould (2015), remains a problem for Yang's model. Second, Yang does not explore any examples illustrating the effects of parameter interaction. In sum, the discussion of parameter interaction here is a novel approach to learning.

The empirical focus of this paper will be on Korean, although in theory the results of the model could be extended to any verb-final language. The reason for focusing on Korean is because support for the model's results concerning the variability of verb raising come from the recent experimental work in Han et al. (2007) and Han et al. (*to appear*). Indeed, additional experimental work in Han et al. (2008) has revealed similar results in Japanese.

The structure of this paper is as follows. I begin with a schematic example of how the learning model sets parameters for the case at hand, and illustrate this in a hypothesis space with 3 parameters. I then scale up the complexity of the model to more closely simulate Korean and show we can obtain similar learning results in a hypothesis space with 5 parameters. I conclude by drawing a connection between the modeling results here and the experimental results of Han et al. (2007), as well as by

---

* Isaac Gould, Massachusetts Institute of Technology, igould@mit.edu. This paper is an abbreviated version of parts of Gould (2015). I thank the following people for their help on this project: Adam Albright, Michel DeGraff, Michael Yoshitaka Erlewine, Michelle Fullwood, Chung-hye Han, Kwang-sup Kim, Myung-Kwan Park, David Pesetsky, Ayaka Sugawara, and Hedde Zeijlstra.

 Proceedings of the 33rd West Coast Conference on Formal Linguistics, ed. Kyeong-min Kim et al., 157-166. Somerville, MA: Cascadilla Proceedings Project.

discussing how a model that does not learn from ambiguous evidence has no clear way of achieving similar results.

2. Schematic illustration of how the model learns

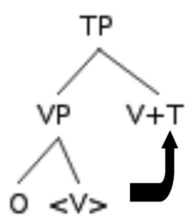
In this section I present the key insight of how the model sets parameters for the case study at hand and illustrate this with a schematic 3-parameter implementation of the model. I begin by discussing the basic ambiguity we see in verb-final languages by focusing on SOV input to the learner.

Let us begin by considering the hypothesis space in the schematic version of the model. This is presented in (1). I note that the model currently assumes that parameters are learned independently of each other. Thus with 3 binary parameters, there are 8 logically possible grammars that the learner could adopt. We will see that the learner will systematically converge on only 2 of these grammars.[1]

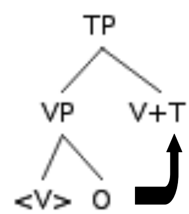
(1) *Hypothesis space (3 parameters)*
 a. 2 projections along clausal spine { TP, VP }
 b. Fix positions of S and O
  1. S is in SpecTP
  2. O is sister of V
 c. Leftward attachment of all specifiers.
 d. 3 binary parameters
  1. TP-headedness: [T-init(ial)] or [T-fin(al)]
  2. VP-headedness: [V-init(ial)] or [V-fin(al)]
  3. V-to-T movement: [+VT] or [–VT]

The ambiguity is as follows. SOV input is compatible with verb movement. With verb movement, though, head-complement order in the VP is underdetermined, as shown in (2).

(2) *Verb raising grammars compatible with SOV input*
 a. [+V-to-T, V-final, T-final]



 b. [+V-to-T, V-initial, T-final]

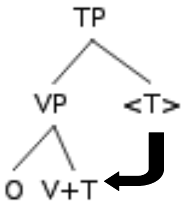


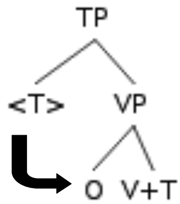
SOV input is also compatible with the verb not moving. If the verb stays in-situ, though, head-complement order for TP is underdetermined, as shown in (3). This is because T is an affix and attaches to the verbal complex; by attaching to the verb, we can no longer tell whether TP is head-initial or head-final. I simply note there that there are various analytical possibilities as to how tense could attach to the verb. These include some process of affix lowering (Chomsky 1957, 1981) or some post-syntactic operation, such as Morphological Merger (Marantz 1988). I will remain agnostic as to what the correct analysis should be, but will assume that if there is no verb raising, one of these processes that attaches verbal suffixes takes place.

[1] The current formulation of the model thus contains no principle that would preclude any of these 8 grammars being learned. We might wonder whether UG does indeed contain some such principle. This is especially relevant given that recent work such as Biberauer et al. (2014) has claimed that some parameter combinations are unattested cross-lingusitically, such as a head-final TP that embeds a head-initial VP. I note that in the model's simulations reported here, the model is always able to learn some grammar that *is* attested cross-linguistically. Nevertheless, the broader question remains as to whether these unattested grammars are learnable at all. In other words, these grammars are unattested because they are unlearnable, and some principle of grammar precludes their being learned. In Gould (2015) I speculate on how the model could be augmented with such a principle, such that these grammars cannot be learned.

(3) *Non-verb raising grammars compatible with SOV input*

a. [–V-to-T, V-final, T-final] b. [–V-to-T, V-final, T-initial]





In sum, SOV input is ambiguous for all 3 parameters in (1). For every parameter value *x*, there is a grammar compatible with the input that contains value *x*. Ambiguity for verb raising in a verb-final langue such as Korean is widely recognized in the literature. However, there is no consensus at to the correct analysis, and there is not much clear evidence in favor of Korean being a language in which the verb always raises or always stays in-situ. I refer the reader to Han et al. (2007) for more detailed discussion of this. In contrast, ambiguity for head-complement order is often implicitly taken to not exist. It is worth emphasizing, though, that head-complement ambiguity is a natural consequence of the different analytical possibilities in linguistic theory. Indeed, the head-complement ambiguity in (2) and (3) is in part a result of the ambiguity regarding verb movement. Thus we see that SOV input can be characterized by pervasive parametric ambiguity.

Despite the high degree of ambiguity with SOV input, certain patterns emerge when we look more closely at the grammars that are input-compatible. A summary of the input-compatible grammars in (2) and (3) is given in (4).

(4) *Grammars compatible with SOV input*

a. [+VT, *T-fin*, V-in] c. [–VT, T-in, V-fin]

b. [+VT, *T-fin*, V-fin] d. [–VT, *T-fin*, V-fin]

In (4) we see that among input-compatible grammars (a) a [+VT] grammar must be [T-fin]; and (b) a [–VT] grammar must be [V-fin]. These relations among parameters values represent *parameter interaction*. The choice of one parameter value affects what the choice of another parameter value can be. Given such parameter interaction, we arrive at the following summary in (5): a majority of grammars that are input-compatible are T-final, and a majority of grammars that are input-compatible are V-final.

(5) *Summary of grammars in (4)*

a. T-init: 25%	b. V-init: 25%	c. +VT: 50%
T-fin: 75%	V-fin: 75%	–VT: 50%

We can now begin to get a sense of the answers to the acquisition questions I began with. A probabilistic learner learns the grammar(s) of best fit to the input. In the case at hand a best-fit grammar has the parameter values [±VT, T-fin, V-fin]. In brief the insight of the model is that parameter interaction makes a head-final value (for either TP or VP) more likely to be compatible with the input. The model can capitalize on this likelihood, such that we do not expect to see variability across learners with respect to head-finality. In contrast, there is insufficient parameter interaction to favor either verb raising or non-verb raising; both values are attested in half the compatible grammars. Accordingly, we mght expect to see variability across speakers in learning whether there is verb raising. Let us now see how the learning procedure of the model can capitalize on the tendencies in (5) to learn parameter values.

Before giving an overview of the learning model, I note that what is presented here is a simplification of a more complex model introduced in Gould (2105). As a general characterization, the implementation of the model adopted here is discriminative: the model tests out some vector of parameter values to see whether they are compatible with a particular token of input; if they are compatible, then those values are probabilistically reinforced. This has the effect of the learner converging on values that are more likely to be input-compatible, e.g. [V-fin]. In its broad outlines, this implementation is highly parallel to Yang's (2002) model. However, several comments are in order. First, the model here is embedded in a more complex generative model in Gould (2015). The full version of the model in Gould (2015) differs from Yang's in that different grammars assign different probabilities to individual sentences of input. The full model converges on the grammar that maximizes the likelihood of the entire corpus of input. The full model thus goes beyond merely discriminatively evaluating the compatibility of different grammars with the input. As discussed in Gould (2015), this allows the full model to address the problem of learning a subset language (Wexler and Manzini 1987), which Yang's model is not able to do. Second, the work here goes beyond Yang by shining a light on how parameter interaction results in particular learning outcomes via ambiguous evidence. Yang discusses no empirical examples of parameter interaction in language learning. For expedience, in this paper I will follow the simplified implementation of the model below, but see Gould (2015) for more discussion on the full model and how it differs from Yang (2002).

The learning procedure of the model can be sketched as follows. Each parameter value has a weight associated with it, and these weights are used to generate probabilities for parameter values. The stronger the weight, the closer to 1 the probability is likely to be.[2] The model receives a token of input and attempts to find a grammar that is compatible with that input datum by sampling a value for each parameter using the parameters' probabilities. If the grammar sampled is compatible with the input, then the weights of that grammar's parameter values are reinforced (i.e. increased). Reinforcing parameter weights strengthens them, which in turn increases their associated probabilities. Reinforcement thus has the effect of increasing the likelihood of sampling those values again when encountering similar input. If the grammar sampled is not compatible with the input, then the model samples again until a compatible grammar is sampled. This process iterates with all subsequent input. As certain values are reinforced more than others, given sufficient input, the learner will be pushed to very strong weights for some parameter values. We can say that a very strong weight for a particular parameter value indicates the model has learned that parameter setting.[3]

We can now see more precisely how this learning procedure leads to certain expected learning outcomes for the model. With sufficient parameter interaction, the model can be pushed toward learning certain parameter values. First consider head-complement direction. Given the parameter interaction in (4), on average the model will sample and reinforce [T-fin] and [V-fin] much more than [T-init] or [V-init]. In the long run, given sufficient SOV input, the model will be pushed toward increasingly stronger weights for head-final values. This results in head-final parameter settings. Suppose we run the model a number of times and that each time we run it corresponds to a different learner in a population. The expectation is that we will see consistent head-finality across the population of learners: all runs of the model have strong weights for [T-fin] and [V-fin]. Next consider verb movement across the population. Neither raising nor non-raising is favored via parameter interaction. Given a learner that is not too tentative (and that can randomly reinforce one value sufficiently more frequently than the other value), the model could learn either value (cf. Pearl 2007). The expectation, then, is that we will see variability across the population for verb raising: some runs of the model will have strong weights for [+VT], and some will have strong weights for [–VT].

Before looking at results for the schematic simulation with 3 parameters, I first present a refinement to the input corpus. In anticipation of the 5-parameter version of the model in Section 3, I will include SV input in the schematic corpus. Recall that if the subject is in SpectTP, SV input is compatible with all 8 logically possible grammars. The learner will be randomly presented with SV and SOV tokens according to their overall corpus frequency. Crucially, then, the learning expectations for head-finality depend on there being sufficient SOV input in the corpus. We saw that it was SOV input that favored head-finality. To estimate the frequency of the 2 types of input, I extrapolated the proportions in (6) from analyses of child-directed Korean.

[2] The weights associated with parameter values are pseudo-count totals of the dirichlet probability distribution. Using the dirichlet distribution allows us to model a learner's uncertainty at any particular point in the learning process. This constitutes another difference with Yang (2002). For Yang the learner is never uncertain at any point in the learning process regarding its current belief as to the nature of the target grammar it is trying to acquire. Modeling uncertainty with the dirichlet distribution can be considered a more psychologically plausible approach to learning (cf. Kemp et al. 2007).

[3] More technical details concerning the implementation of the model (e.g. the values of the weights and how the weights are reinforced) can be found in Gould (2015).

(6) *Randomly generated input frequencies (based on frequencies in child directed Korean)* [4]
   a. SV   p. = .75
   b. SOV   p. = .25

I reiterate that the entire corpus in (6) is ambiguous for all parameter values, and that multiple grammars are compatible with the entire corpus. Nevertheless, we shall see that there is indeed an ample proportion of SOV input in the corpus to confirm the learning expectations from above.

Simulations of the model were run using the Church programming language (Goodman et al. 2008). The model was run 15 times for an average of 1,850 tokens of input per run. Results of the 3-parameter model are given in Table 1.

Table 1. Average proportions of weights for parameter values in 3-parameter space (Average of approximately 1850 tokens of input per run)

Parameter Value	**Proportion**	**# Runs**
[V-fin]	.9777	15
[V-init]	.0233	
[T-fin]	.9655	
[T-init]	.0345	
[–VT]	.9442	11
[+VT]	.9015	4

In all runs, the model reaches a point where we see strong weights for both [V-fin] and [T-fin]. Thus the model has a high degree of success in learning a consistently head-final grammar. Further there is little variability in head-finality during the course of learning. The weights for the head-initial values generally go steadily down. On only 1 run does the proportion of the weight for [T-init] go above .75 before later dropping below .25. We also see that at the point during learning at which there is consistent head-finality, there is a non-zero probability of learning a raising or non-raising grammar. In 4 runs, the model learned that there is obligatory verb movement, whereas in 11 runs it learned that the verb remains in-situ in the VP. [5]

In sum, we see that despite all the input being ambiguous, the model systematically learns head-finality, while there is variability concerning verb raising.

3. Scaling the model up: A simple simulation of Korean

In this section I discuss how the conclusions from the 3-parameter model are more general. I consider two ways of expanding the model so as to more closely approximate actual Korean. First, I will increase the number of parameters to 5 by including the functional projection *v*P in the hypothesis space. Second, I will consider enriching the input corpus by adding in auxiliaries. Ultimately, though, I will maintain the same input corpus from Section 2, as the inclusion of auxiliaries does not clearly change the basic ambiguity of the learning challenge.

I begin by introducing the expanded hypothesis of the 5-parameter model in (7) below. This is the same as (1), but (a) adds a head-directionality parameter for *v*P; and (b) divides verb raising between movement from VP to *v*P, and movement from *v*P to TP. Differences with (1) are in boldface in (7). Just as in 3-parameter model, all parameters are learned independently. With 5 binary parameters, there are thus 32 possible grammars that the learner can construct by sampling different parameter values during the course of learning.

---

[4] According to Fukuda and Choi (2009) approximately half of all verbs spoken to children in their sample are transitive, and according to Kim (2000: 345) on average there is object drop with approximately half of all transitive verbs in child directed Korean. If half of all clauses in the corpus are transitive, and if half of clauses contain an overt object, then assuming independence between the two, the likelihood of a clause that is transitive with an overt object (i.e. is SOV) is .25.

[5] A binomial test was run under the hypothesis that there is a .5 probability of learning either a raising or non-raising grammar. Under such a hypothesis, a success rate of 4/15 for learning a raising grammar has a two-tailed p-value of .1185.

(7) *Hypothesis space (5 parameters)*
a. 3 projections along clausal spine: { TP , ***v*P** , VP }
b. Fix positions of S and O
1. S is in SpecTP
2. O is sister of V
c. Assume leftward attachment of all specifiers and adjuncts
d. 5 binary parameters

1. TP-headedness:	[T-init] or [T-fin]
2. *v*P-headedness:	**[*v*-init] or [*v*-fin]**
3. VP-headedness:	[V-init] or [V-fin]
4. V-to-*v* movement:	**[+V-*v*] or [–V-*v*]**
5. *v*-to-T movement:	**[+*v*-T] or [–*v*-T]**

When we consider which grammars are compatible with SOV input, we see that the parameters interact in a way that is similar to that found in the 3-parameter model. SOV input is ambiguous for every parameter value. The verb is free to move or not, but wherever the verb is, that projection must be head-final. (8) summarizes the frequencies at which different parameter values are attested in grammars that are compatible with SOV input, and is comparable with (5).

(8) *16 grammars compatible with SOV input*

a. V-fin grammars: 75% V-init grammars: 25%	d. +V-*v* grammars: 50% –V-*v* grammars: 50%
b. *v*-fin grammars: .625% *v*-init grammars: .375%	e. +*v*-T grammars: 50% –*v*-T grammars: 50%
c. T-fin grammars: .625% T-in grammars: .375%	f. [+V-*v* , +*v*-T] grammars: 25%

Again we see that for each of the directionality parameters, a majority of compatible grammars contains a head-final value. To assess variability in learning verb movement, I focus on whether there is any movement at all, i.e. the value of [±V-*v*]. Again, the 2 values for this parameter are found evenly among all compatible grammars. Thus, the grammars of best fit given the input are consistently head-final, but may or may not have verb raising.

If we focus on [±V-*v*], we can now ask what an instructive way of expanding the schematic corpus might be. Given this focus, the input that is most relevant concerns the c-command domain of *v*, i.e. the VP. There are 2 kinds of prototypical input in this domain: input in which V does or does not have a complement, viz. SV and SOV input. This is precisely the kind of input we have already seen in Section 2.

My claim, then, is that a corpus of { SV, SOV } input, albeit coarse, is highly representative of the major patterns of evidence in Korean for setting the parameters under discussion. Moreover, the claim is that (essentially) all input in Korean is ambiguous for all the parameters considered here.[6]

There are various kinds of input in Korean that one might consider as being unambiguous for the parameters here. I will consider one prominent example here and will present evidence that it is in fact ambiguous (see Gould 2015 for discussion of additional types of input). Might auxiliaries be a source of unambiguous evidence? Consider the structures in (9) below, which contain an auxiliary. Suppose the presence of the auxiliary blocks the main verb from raising to the Aux-head or higher. There are 2 hypotheses we might then consider regarding the structural position of the verb. Either the auxiliary takes the VP as its complement (9a), in which case the VP is unambiguously head-final, or the auxiliary takes some other projection XP as its complement. Under hypothesis (9b), as XP contains the VP, we are left with the original puzzle concerning ambiguity in (2) and (3). In other words, if the auxiliary takes a larger complement, then there is ambiguity regarding VP-headedness: for example,

[6] A possible exception to this is the input that Han et al. (2007) use to identify the height of the verb (see §4). This is input clearly showing the scope of negation relative to a quantified object. Following Han et al., I assume this input is sufficiently rare in the input to the learner that it has little to no role in the average learner's development.

the VP could be head-initial if the verb raises to a head-final XP, or the VP could be head-final if the verb remains in-situ.

(9) *2 hypotheses for auxiliaries*

a. Aux takes VP as complement
→ Unambiguously head-final

b. Aux takes XP as complement
→ Ambiguity as per (2) and (3)

I now present evidence for (9b) and thus for the persistence of ambiguity in Korean. In (10) we see that there are various kinds of morphology that can appear between the main verb and an auxiliary. First, auxiliaries sub-categorize for a number of linking morphemes, such as *–eya*. Moreover, these linking morphemes sub-categorize for tense morphology below the auxiliary.

(10) *Evidence for functional material between auxiliaries and the VP* (Cho 1993: 16)

a. Mary-ka pap-ul mek-Ø-eya ha-n-ta.
Mary-nom rice-acc eat-pres-EYA must-pres-decl
'Mary must have a meal.'

b. Mary-ka pap-ul mek-ess-eya ha-n-ta
Mary-nom rice-acc eat-past-EYA must-pres-decl

I conclude that there is indeed structure between the main verb and the auxiliary, and that auxiliaries do not present a counterexample to the claim regarding the high degree of ambiguity in Korean. As the null hypothesis, we can assume that the structure immediately dominating all verbs, whether main verbs or auxiliaries, is the same: *T > v > Aux/V*. Auxiliaries thus present us with an ambiguity puzzle that is analogous to the puzzle we began with when we looked at SOV input in (2) and (3). For the purpose of a proof-of-concept illustration, it is now possible to conflate SOV input with input containing auxiliaries.

In summary, for the 5-parameter model I first enriched the hypothesis space by including 2 additional parameters, and we saw that the basic mechanics of parameter interaction work in a similar way. Second, I considered enriching the schematic corpus and decided to maintain the { SV, SOV } corpus, as it is representative of the ambiguity faced by Korean learners.

Given this discussion, expected outcomes for the 5-parameter model are comparable with those of the 3-parameter model. We expect consistent head-finality for all learners, i.e. in all runs of the model. Further, we expect variability in verb raising across learners, i.e. across runs of the model.

The 5-parameter model was also run 15 times for an average of 1,650 tokens of input per run. Results are given in Table 2.

Table 2. Average proportions of weights for parameter values
(Average of approximately 1650 tokens of input per run)

Parameter Value	**Proportion**	**# Runs**
[V-fin]	.9921	15
[V-init]	.0079	
[*v*-fin]	.9338	
[*v*-init]	.0662	
[T-fin]	.9476	
[T-init]	.0524	
[–V-*v*]	.9232	13
[+V-*v*]	.9427	2
[–*v*-T]	.9459	6
[+*v*-T]	.8686	9

Again we see that there is a high degree of success in learning a consistently head-final grammar. In all runs, the model reaches a point where there are strong weights for all head-final values. There is also no variability in head-finality during the course of learning: in general the weights for the head-initial values go steadily down, and the proportion of the weights for any head-initial value never exceeds (or gets close to) .75. Further, we also see that at the point during learning at which there is consistent head-finality there is a non-zero probability in learning a raising or non-raising grammar. In 2 runs, the model learns there is verb movement at least as high as *v*, and in the remaining 13 runs the verb remains in-situ.

The expectations for the learning model are thus confirmed in this more complex hypothesis space. One might ask, though, why learning [+V-*v*] is relatively infrequent. The answer is a more local kind of parameter interaction, something that we can call *secondary parameter interaction*, an example of which is discussed here. In (8) we saw that the model gets pushed most strongly toward [V-fin]: 75% of input-compatible grammars are [V-fin], whereas only 62.5% of input-compatible grammars are [T-fin] or [*v*-fin]. Of input-compatible [V-fin] grammars, 66.6% are in fact [–V-*v*]. The effect on learning is the following. On average, the learner is more likely early in the learning process to have the strongest weighting for [V-fin]. Once the weight for [V-fin] is relatively stronger, then the learner will be pushed more strongly toward a non-raising [–V-*v*] grammar (than toward a [+V-*v*] raising grammar). It is still possible to learn a raising grammar, though, so long as the weight for [V-fin] is not overly strong relative to [T-fin] and [*v*-fin] early in the learning process. Such a scenario is certainly possible given the randomness of the sampling process, and given that [T-fin] and [*v*-fin] are themselves favored, as shown in (8). Due to this secondary parameter interaction, on average we expect a non-raising [–V-*v*] grammar to be learned more frequently than a raising [+V-*v*].[7] This accounts for why, in the 5-parameter model, non-raising is a more common outcome, even though variability exists across the larger population.

4. Empirical support for the model

In this section I briefly discuss a source of additional support for the learning model I have proposed here. I have shown how a result of the learning model is variability across learners concerning verb raising. Support that the model is on the right track comes from recent experimental and acquisition work in Han et al. (2007). Han et al. found systematic variability in children and adults in the scope of negation with respect to a quantified object. Speakers largely divided into 2 groups: those with low or high scope of negation. Further Han et al. (*to appear*) show this variability is relatively stable within speakers over time. In these papers, the variability in scope is attributed to the absence or presence of verb raising. In brief, negation attaches to the verb and goes wherever the verb goes: if the verb raises, the negation outscopes the object; and if the verb remains in-situ, then the object outscopes negation. To the extent that these experimental results and the results of the learning model point in the same direction, they provide support for the basic approach to modeling variability that I have discussed in this paper.[8]

5. Model comparison

In the introduction I noted how pervasive ambiguity in the input is a challenge for models of parameter setting, such as Sakas and Fodor (2001), which set parameters only on the basis of unambiguous evidence. Let us consider in more detail how such a model would fare with Korean given the discussion in this paper.

[7] I have considered a number of additional parameters with which the model could be augmented, and among input-compatible grammars, the proportion of [–V-*v*] grammars given [V-fin] does not increase. These include parameters for subject raising and T-to-C movement. Thus there is no obvious cause for concern such that secondary parameter interaction would preclude the possibility of the model learning a verb raising grammar.

[8] In Gould (2015), I discuss several differences between the syntactic analysis of Han et al. (2007) and the syntax of the 5-parameter model. These differences are not crucial to the discussion here, though. In Gould (2015) I propose a minor modification to Han et al.'s analysis that is consistent both with their experimental results and the results of the model here.

I have claimed that there is insufficient unambiguous evidence in Korean to set the parameters under consideration.[9] A model that relies on unambiguous evidence is now confronted with a serious problem: how could a learner acquire Korean? One possibility that is available in such a model is a universal set of default values. For example, in the initial state the learner could have default values of [V-fin, T-fin]. The learner would only adopt a head-initial value if presented with unambiguous evidence for head-initiality. As such unambiguous evidence is not found in Korean (nor for head-finality, for that matter, as I have discussed), the learner will never move from the defaults of the initial state. A universal set of defaults could thus also achieve the result of consistent-head finality, but defaults struggle to account for variability with verb raising. Suppose the default is not to move the verb. As the learner does not receive unambiguous evidence for verb movement, we do not predict some learners to adopt a non-default value and acquire a grammar with verb raising. Such a prediction runs counter to the claim in Han et al. (2007). Similarly, if the default were to raise the verb, as the learner does not receive unambiguous evidence to keep the verb in-situ, we do not predict any learners to adopt a non-default grammar that does not have verb movement.

In short, a model that relies on unambiguous evidence is unable to account for variability across learners. In contrast, I have shown how under the model I have proposed variability can emerge when systematically learning from ambiguous evidence.

6. Final remarks

In this paper I set out to see what can be learned from ambiguous evidence. I outlined an approach for learning systematically from ambiguous input. The learning that we saw is shaped by parameter interaction. With sufficient parameter interaction, the learner is pushed toward particular parameter values. With insufficient parameter interaction, there can be variability across learners as to which parameter values are adopted. What gets learned crucially depends on what the parameters are (and how they interact). Looking toward future research, I hope to have shown that to get the details right about modeling actual speaker populations, it is imperative that we have a clear sense of what the parameter space is and that we factor in the role of ambiguous evidence in the learning process.

References

Biberauer, Theresa, Anders Holmberg, and Ian Roberts. 2014. “A Syntactic Universal and Its Consequences”, *Linguistic Inquiry* 45 (2): 169-225.

Cho, Sae-Youn. 1993. “Auxiliary Verb Constructions in Korean”, *Studies in the Linguistic Sciences* 23 (2): 1-24.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

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

Fukuda, Shin and Soonja Choi. 2009. “The Acquisition of Transitivity in Japanese and Korean Children”, in Shoichi Iwasaki, Hajime Hoji, Patricia Clancy, Sung-Och Sohn (eds.), *Japanese and Korean Linguistics 17*. Stanford: CSLI Publications,pp. 613-624.

Gold, E. Mark 1967. “Language Identification in the Limit”, *Information and Control* 10 (5): 447-474.

Goodman, Noah, Vikash Mansinghka, Daiel Roy, Keith Bonawitz, and Joshua Tenenbaum. 2008. “Church: A language for generative models”, in David McAllester and Petri Myllymaki (eds.), *Proceedings of the Twenty-fourth Conference on Uncertainty in Artificial Intelligence (UAI-08)*. Corvallis, Oregon: AUAI Press, pp. 220-229.

Gould, Isaac. 2015. *Syntactic Learning from Ambiguous Evidence: Errors and End-States*. PhD dissertation, Massachusetts Institute of Technology.

Han, Chung-hye, Jeffrey Lidz, and Julien Musolino. 2007. “V-Raising and Grammar Competition in Korean: Evidence from negation and quantifier scope”, *Linguistic Inquiry* 38 (1): 1-47.

Han, Chung-hye, Dennis Ryan Storoshenko, and Yasuko Sakurai. 2008. “An Experimental Investigation into the Placement of the Verb in the Clause Structure of Japanese”, in *Proceedings of the 2007 International Conference on Linguistics in Korea (ICLK-2007)*. Seoul: The Linguistic Society of Korea.

Han, Chung-hye, Jeffrey Lidz, and Dennis Ryan Storoshenko. To appear. “Variation in Negation and Quantifier Scope Judgments in Korean”, in *Proceedings of the 47th Annual Meeting of the Chicago Linguistic Society*. Chicago: The Chicago Linguistic Society.

[9] Again, following Han et al. (2007), I assume that input concerning the scope of negation and a quantified object, which can be used to detect whether the verb has moved, is sufficiently rare in the input to the learner that it has little to no role in the average learner's development.

Kemp, Charles, Amy Perfors, and Joshua Tenenbaum. 2007. "Learning Overhypotheses with Hierarchical Bayesian Models", *Developmental Science* 10 (3): 307-321.

Kim, Young-Joo. 2000. "Subject/Object Drop in the Acquisition of Korean: A cross-linguistic comparison", *Journal of East Asian Linguistics* 9 (4): 325-351.

Marantz, Alec. 1988. "Clitics, Morphological Merger, and the Mapping to Phonological Structure" in Michael Hammond and Michael Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. San Diego: Academic Press, pp. 253-270.

Pearl, Lisa. 2007. *Necessary Bias in Natural Language Learning*. PhD dissertation, The University of Maryland.

Sakas, William Gregory and Janet Dean Fodor. 2001. "The Structural Triggers Learner", in Stefano Bartolo (ed.), *Language Acquisition and Learnability*. Cambridge: Cambridge University Press, pp. 172-233.

Wexler, Kenneth and M. Rita Manzini. 1987. "Parameters and Learnability in Binding Theory", in Thomas Roeper and Edwin Williams (eds.), *Parameter Setting*. Dordrecht: Reidel, pp. 41-76.

Yang, Charles. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Proceedings of the 33rd West Coast Conference on Formal Linguistics

edited by
Kyeong-min Kim, Pocholo Umbal,
Trevor Block, Queenie Chan,
Tanie Cheng, Kelli Finney, Mara Katz,
Sophie Nickel-Thompson, and Lisa Shorten

Cascadilla Proceedings Project Somerville, MA 2016

Copyright information

ISBN 978-1-57473-469-0 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Gould, Isaac. 2016. Learning Parameter Setting from Ambiguous Evidence: Parameter Interaction and the Case of Korean. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, ed. Kyeong-min Kim et al., 157-166. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3235.