Linguistic Research 35(1), 29-74
DOI: 10.17250/khisli.35.1.201803.002

An acquisition-based evaluation metric for competing syntactic theories*

Isaac Gould
(Ewha Womans University)

Gould, Isaac. 2018. An acquisition-based evaluation metric for competing syntactic theories. *Linguistic Research* 35(1), 29-74. This paper addresses the puzzle of choosing among competing syntactic theories, which are contrasting theories that can nevertheless all properly account for the utterances of some language. To do so, I develop an evaluation metric that is based on language learnability and acquisition. In particular, such an acquisition-based metric considers speed of acquiring a target grammar, as well as relative accuracy of acquisition. In applying such a metric, I propose using the kinds of novel data that can be generated via computational modeling for syntactic learning. I then demonstrate how this evaluation metric can be used effectively by reporting on learning simulations for a case study, which focuses on two competing theories that arise in the literature for verb raising in Swedish. **(Ewha Womans University)**

Keywords evaluation metric, syntactic modeling, Occam's razor, parameter mis-setting, verb raising, Swedish

1. Introduction

A significant puzzle in syntactic theory comes from adjudicating among competing theories, each of which can account for the full range of standard syntactic data for a particular language. In this scenario, under each syntactic theory, for a particular language L there is some target grammar(s), and each target grammar has the generative capacity to both (a) generate all the well-formed utterances of L,

---

* I thank the following people for their help with various aspects of the research reported here: Kim Ahlström, Hye-Won Choi, Kajsa Djärv, Michael Yoshitaka Erlewine, Elyse Jamieson, Paul Kiparsky, Peter Kipka, David Pesetsky, Gertjan Postma, Clifton Pye, Jeroen van Craenenbroeck, Sten Vikner, Marit Westergaard, Charles Yang, and Jie Zhang. This material has also benefitted from various anonymous reviewers and from being presented at The University of Kansas, Ewha Womans University, and the workshop Dealing with Bad Data in Linguistic Theory at The Meertens Institute. This work was supported by the CORE Program funded by the Korean government (MOE).

and (b) generate none of the ill-formed utterances of L. The puzzle, then, is which theory we should choose. Each of the theories would seem to be adequate in accounting for the syntax of the language in question. What is needed is an evaluation metric to choose between theories (cf. Chomsky 1965). The choice is significant assuming it constitutes a hypothesis about what any child language learner's possible set of grammatical representations (i.e. a set of hypotheses about grammar) can be.[1]

In this paper, I propose a novel evaluation metric based on two familiar ingredients from linguistic theory − speed of acquisition and accuracy of acquisition − and illustrate how this metric can be applied, with a model for language learning, to a case study of competing theories involving verb movement in Swedish. In this section, I first present some brief background on the central puzzle of this paper, introduce the case study, and highlight some contributions of this paper. Section 1.1 introduces the evaluation metric, and Section 1.2 contains discussion relating this metric and the modeling simulations here to earlier work.

Although questions having to do with the puzzle of competing theories, as well as its relation to language learning, are foundational ones for linguistic theory, work focusing on such questions has been increasingly scarce in the last half century, in part because of what Lidz and Dudley (2017) describe as the practical challenges in responding to questions of this sort. It could thus be highly advantageous to consider the puzzle of competing theories in light of modern tools for computational modeling. This is especially true given the renewed interest in questions of theory evaluation as found in work such as Pearl (2017).[2] This paper thus adds to a small but growing body of recent work on related issues. This paper also shines a light on the application of acquisition-related considerations to the puzzle, and on how to do so by harnessing these computational tools with proof-of-concept modeling simulations of the learning process that are grounded in recent work on language

1 Thus the question this paper focuses on is different from Chomsky's (1965) concern with an evaluation measure. For Chomsky (1965) the question is how a learner can choose between competing sets of hypotheses. The puzzle of competing theories in this paper does not concern the learner's choice between competing sets of hypotheses, but rather which set of hypotheses is under the learner's consideration.

2 See also Yang (2017), which pursues an approach somewhat related to the one in this paper – albeit one pertaining to morphology and not syntax – but ultimately one that focuses on a different question, namely that of Chomsky's (1965) evaluation measure from note 1.

learning. To that end, the learning simulations in this paper are run using the probabilistic learning model in Gould (2017).

Indeed, it is worth highlighting that even Pearl's (2017: 144-145) discussion, which is suggestive of what an evaluation metric could be for the puzzle of competing theories, takes an approach that is not directly tied to developmental considerations. Pearl by and large follows an approach that can be found more generally in the syntactic literature (to the extent that the puzzle is considered at all, either implicitly or explicitly), which is to use Occam's razor to help evaluate the competing theories. For our purposes, this approach can be implemented by ranking the competing theories according to some notion of complexity, with the simplest theory being the preferred theory. This way of choosing among theories involves a complexity-based evaluation metric. Indeed, such an evaluation metric reflects more generally how using Occam's razor is embedded as an often unexplored component of theory development in much of the theoretical literature on syntax.

However, there is an important empirical question that a complexity-based evaluation metric does not by itself address: it does not necessarily tell us which of the competing theories (if any) is learnable. In line with Chomsky (1986), I assume that a fundamental goal of theory creation is to develop theories that lead to grammars that can be acquired. Accordingly, if there is evidence suggesting that a simpler theory is less likely to, or even might not, allow a language's grammar to be acquired, then we have an argument for dispreferring such a theory.

Because of the centrality of this question, I propose using learnability as a core, motivating component of a different kind of evaluation metric − an acquisition-based one − to help address the puzzle of competing theories. Further, given this emphasis on acquisition, I propose including a related component in the metric: a way of tying children's developmental trajectories to syntactic theory. Thus a theory that can be more tightly connected to these trajectories is to be preferred over its competitors, as such a connection provides us with an empirical basis in accounting for the nature of these developmental paths. I introduce this metric in Section 1.1 and discuss how applying it with a learning model allows us to generate novel kinds of data that can shed light on the puzzle of competing theories.

Turning now to the literature on Swedish and the case study at hand, two alternative approaches to verb movement can be found that are similar in their overarching theoretical framework, but that crucially differ in some of the mechanics of the verb movement that results in Verb Second (V2) word order. A crucial point of debate in these approaches revolves around whether V2 verb movement in Swedish must proceed via what can be called a process of V-to-C movement, or rather whether it must proceed via a sequence of distinct movement processes. The former approach can be found in works such as Holmberg and Platzack (1995) and Vikner (1995). The latter approach, which lacks V-to-C movement, can be found in works such as Platzack (1986a) and Lightfoot (1991, 1995). Given their similarities and their ultimate point of divergence, these approaches provide a suitable domain for testing the kind of acquisition-based evaluation metric envisioned here.

Further, when we consider the broader literature on verb movement, it is possible to embed these alternative approaches within competing syntactic theories, which I present in Section 2.[3] The discussion in this paper will thus be framed in the context of these competing theories, which are based in part on the alternative approaches from the literature on Swedish. The first, which I call the Stepwise theory (which lacks V-to-C movement), can be aligned with the basic approach found in Platzack (1986a) and Lightfoot (1991, 1995). The second, which I call the Swooping theory (which has V-to-C movement), can be aligned with the basic approach to verb movement found in Holmberg and Platzack (1995) and Vikner (1995). Although the competing theories (and the differing approaches to V2 they are based on) are similar, learning simulations can provide a clear way of empirically distinguishing them: modeling results point to the conclusion that the availability of a process of V-to-C movement is crucial for faring well according to the evaluation metric proposed here.

A more narrow focus of this paper is to shed light on the debate regarding the syntax of verb movement in Swedish. As the learning results provide support for the Swooping theory, and thus a theoretical approach with V-to-C movement, this paper provides a novel argument that is (a) in support of the basic approach in the Swooping-related literature cited above, and (b) against the basic approach in the Stepwise-related literature.

---

3 I note early a limitation in the scope of this work. In Section 2, I discuss how looking at Swedish utterances by themselves does not help in determining whether our theory should allow for the possibility of V-to-C movement. This discussion is limited to Swedish, and I do not consider whether other languages can be used to provide an argument for the inclusion of V-to-C movement in a theory of verb movement. Such considerations go beyond the scope of this work, but for introductory discussion of the debate concerning V-to-C movement outside Scandinavian, see Ackema and Čamdžić (2003).

Returning to the issue of a complexity-based metric, an outcome of this paper is an illustration of the use of learning results as evidence against always adopting a simpler theory. As I discuss, the Stepwise theory is demonstrably less complex than the Swooping theory. Nevertheless, we will see that it is the more complex Swooping theory that fares better according to the acquisition-based evaluation metric. I take this as evidence in favor of dispreferring the simpler theory in favor of the more complex theory.

An implication of this conclusion as regards theory development is that considerations of complexity or parsimony, although often deemed significant in building a theory, are perhaps best viewed as being of secondary importance. They are perhaps most applicable in situations where we do not have access to other sources of data stemming from work in other empirical domains of linguistic study. More broadly, then, by incorporating modeling results that are grounded in syntactic theory, the paper contributes to the diverse sources of data (including acquisition and psycholinguistic processing evidence) that can not only bring us to a deeper understanding of language that extends beyond parsimony, but that can help us address the issue of competing theories, and, more generally, inform formal syntactic theory.

This paper also contributes to the growing discussion in language learnability and acquisition about the role of ambiguous and unambiguous evidence during the learning process. Unambiguous evidence is linguistic input to the learner that with respect to some property of grammar is compatible with only one structural analysis. Accordingly, ambiguous evidence is input that underdetermines the correct structural analysis with respect to that property. In contrast to works such as Fodor (1998) and Sakas and Fodor (2001), which propose that the learner learns only from unambiguous evidence, I build on work such as Gould (2017) in identifying a novel key role for ambiguous evidence in the learning process. In Section 3 I discuss how the interaction between ambiguous evidence and the Swooping theory plays an important role in leading us to expect the latter to fare better per the evaluation metric here. In contrast, such a comparable interaction is not found with the Stepwise theory, leading us to expect it to fare worse.

The structure of this paper is as follows. In the remainder of Section 1, I introduce the evaluation metric and contrast the work here with earlier work. In Section 2, I present the two competing theories for verb movement in Swedish and discuss their relation to earlier theoretical approaches that have been proposed in the literature on Swedish. Section 3 reviews the learning model from Gould (2017) and presents a concrete implementation of this model for verb movement in Swedish. Section 3 also contains discussion of why the more complex Swooping theory is expected to fare better per the evaluation metric. Section 4 presents results and discussion of the learning simulations, and Section 5 concludes.

1.1 The evaluation metric

In this section I introduce my proposal for an acquisition-based evaluation metric and discuss how modeling simulations can generate novel kinds of data that can be used to apply the metric. The first of two ingredients of the evaluation metric is the average speed (measured in tokens of input) in which a learner can learn a target grammar. Children must acquire their target grammar in a finite, relatively limited period of time. Thus speed of learning is relevant because children cannot take too long to learn a target grammar; if a learner needs additional time that extends beyond this limit, then we can say that a target grammar is not always learnable. Given this learning goal, if we have data that speak to such a learning goal being much less likely to be met under a particular theoretical representation of the learner's hypotheses, then we can use that data as evidence against that representation. In other words, the kind of metric that I lay out below would allow us to choose a theory that gives the learner a better likelihood of learning a target grammar. How, then, can the notion of speed be applied to give us the sort of data necessary for an evaluation metric?

Modeling simulations provide us with a practical way of trying to measure speed. The proposal is to use proof-of-concept simulations of the learning model in Gould (2017) to generate novel kinds of data, which can then be used in a language development-based evaluation metric. Accordingly, the first kind of data concerns the average speed in which the target grammar(s) of a given language L, as proposed by a particular theory, can be learned with the learning model. Just as children do not have an indefinite period of time in which to learn, models cannot be run for indefinitely long periods of time, again measured in tokens of input (Yang 2002:45). Thus we are interested in seeing how likely a target grammar can be learned within some critical period of time. As modeling simulations are simplifications of, and accelerated versions of children's actual acquisition task, we are currently faced with a certain arbitrariness as to the boundaries of this critical period. Nevertheless, modeling simulations give us a concrete way of measuring relatively how much time it would take to acquire a certain grammar. Thus we can investigate the following: can a target grammar be learned upon exposure to 10,000 tokens of input?; can it be learned upon exposure to 20,000?; what about 30,000?; and so on. In this way we can assess the relative performance of learners under different theoretical sets of hypotheses. If we see a difference in how long it takes the model on average to learn a target grammar across theories, then this can be taken as a type of evidence in favor of the theory that results in faster learning. All things being equal, a theory that results in slower learning trajectories is less likely to have a learnable target grammar because it is more likely to approach or go beyond the boundary of this critical period. Such differences in speed are important because as we scale up the size and complexity of the learning task to more precisely match that of children, our models will require tokens of input that are orders of magnitude greater in number, thus significantly increasing the likelihood that a target grammar in a slow-learning theory takes too long to learn. (Accordingly, the larger the difference in speed, the stronger the evidence becomes.) Thus even with simplified modeling simulations, we can evaluate competing theories with respect to the likelihood of their target-grammar learnability: given a particular learning procedure and two competing theories, all else being equal the faster-learning theory on average is to be preferred over the slower-learning theory, which is less likely to have a learnable target grammar.

As noted above, such a proof-of-concept illustration with a model necessarily involves various simplifications of the learning task. Nevertheless, the implementation of the model here is sufficiently detailed to capture the basics of verb movement in the case study, and thus provides us with a solid starting point on which to base the conclusions of this paper. Further, it provides a foundation for further refinements and expansions of the model in future work. At various points, I will consider examples of such expansions. Importantly, I discuss how our expectations are that including such refinements would not lead to modeling results that are substantively alter the overall conclusions from the ones reported here. Accordingly, to the extent that the results here can be taken as representative of further enrichment of the model, this paper provides strong evidence in support of the acquisition-based evaluation metric proposed here and its application in the case study at hand. Motivated by this discussion of these expansions, then, my working assumption throughout this paper will be that the results here can be representative of such enrichment.

In addition to speed, another developmental consideration is accuracy; thus the second ingredient of the evaluation metric I propose is the relative accuracy each theory has in acquiring a target grammar. Again, modeling simulations provide us with a concrete way of measuring this. By accuracy I do not only mean whether the model arrives at a target grammar, but also some degree of accuracy with regard to how well the model's learning trajectory matches what is known of children's learning paths (cf. Yang 2002: 6-8). Identifying children's learning trajectories is in fact one of the practical challenges Lidz and Dudley (2017) identify for work of this sort, but given the current availability of child data for certain grammatical phenomena we perhaps can now begin to make the following novel kind of evaluation. Given a grammatical phenomenon for a particular language where we do not see widespread evidence of child errors suggestive of their firmly adopting particular non-target hypotheses − which, in the context here, is mis-setting particular linguistic parameters (cf. Chomsky 1981) − then all things being equal, a theory that results in fewer learners mis-setting such parameters on their way to learning a target grammar is to be preferred over a theory that results in more learners mis-setting such parameters during the course of learning. From a modeling perspective we can measure this by running a model under a particular syntactic theory a number of times, so as to simulate a population of learners. We can then look at what proportion of this population undergoes parameter mis-setting. To be clear, we are not looking for the elimination of errors *per se*. Child learners make errors, and models should attempt to capture these errors (cf. Yang 2002; Gould 2017).

For example, in the case study here we will consider the possibility of parameter mis-setting with respect to verb movement for Swedish learners. Child Swedish errors involving verb placement are sometimes robustly attested, but have been observed to be relatively infrequent for a number of children (Waldmann 2011; see Gould 2017: 291-296 for further discussion). Given the attested errors but also the absence of widespread high-frequency errors across child learners, I propose that we can evaluate a theory that results in relatively less parameter mis-setting as the preferred theory, allowing for the possibility that this theory might result in some parameter mis-setting. In addition to speed, then, looking at the relative rate of parameter mis-setting thus constitutes a second and new kind of data that can be used to evaluate competing theories.

In sum, the evaluation metric for competing theories that I propose is as follows.

All things being equal, Theory α is to be preferred to competing Theory β if (a) on average α learns a target grammar faster than β (i.e. requires fewer tokens of input, including input for learners that never learn a target grammar), and (b) α has more accurate learning trajectories across learners than β (i.e. has a more accurate proportion of learners with parameter mis-setting; for the case at hand this is a smaller proportion).

1.2 Relation to earlier work

To my knowledge no work on syntactic theory has investigated the relation between these new kinds of data (speed and accuracy in terms of parameter mis-setting) and different syntactic theories in any detail, and none has attempted to evaluate competing syntactic theories with actual simulations of the learning process. Thus a contribution of this paper is to shine a light on how such modeling data can play a role in the development of syntactic theory, and in particular to show how an acquisition-based evaluation metric that incorporates such data can be used in addressing the challenge of competing theories.

Relatedly, the idea that learning simulations can help shed light on disputed areas of syntactic theory is mentioned only in passing in Yang (2002) and only briefly in Gould (2017). In contrast, Pearl (2007: Chapter 4) contains learning simulations of different proposals with the aim of supporting the proposal that allows for most accurately modeling diachronic changes in word order. However, none of Pearl's proposals is about syntactic theory *per se*; rather they are concerned with the type of linguistic evidence that a learner or child might actually learn from.

As for the evaluation metric itself, the notion of speed as desirable for acquisition is well recognized in the theoretical literature, but it is not fully developed more generally as part of an evaluation metric for competing theories. As regards modeling, speed is largely considered with respect to what the design of the learning procedure should be (cf. Sakas and Fodor 2001: 195-199; Yang 2002: 45-53), or whether there should be filters on what kind of input the model learns from (Pearl 2007: Chapter 4); this kind of focus is thus not on the current puzzle of competing theories of syntactic hypotheses. Perhaps closest to the use of speed here as part of a metric is the role that parameters can play in theory creation in

Chomsky (1981) and many related works (see also the discussion in Pearl and Lidz 2013). In these works, a theory with a particular parameter (a macro-parameter for Smith and Law 2009) whose setting allows the learner to learn about multiple properties of a language is to be preferred to an otherwise equivalent theory in which those same properties are learned by setting a larger set of parameters, or micro-parameters for Smith and Law (2009) (cf. Safir 1987). Underpinning an evaluation metric that prefers a theory with macro-parameters to one with micro-parameters is speed, in that a target grammar can in principle be learned more quickly with a macro-parameter. This is because with a macro-parameter, learning from input containing any one of the relevant properties entails learning about all the other properties (even if they are not found in that input). In contrast, each micro-parameter might require a different set of input (some of which may be exceedingly or vanishingly rare) in order to set it. A theory with a macro-parameter for a set of properties can conceivably require less input to learn a target grammar, then, as it can in principle learn more from a single token of input than can an equivalent theory equipped with micro-parameters.

As such, macro-parameters are tightly connected to speed of acquisition and are motivated in a spirit that is very much like the one of this paper. However, comparing competing theories on the basis of macro- and micro-parameters is just a particular case of incorporating speed into an evaluation metric. For example, it is not clear that either of the competing syntactic theories of verb movement in the case study of this paper is equivalent to the other in terms of macro- or micro-parameters. Nevertheless, because of its generality, speed can, regardless of the shape of a syntactic theory or its parameters, be used as part of an evaluation metric for the reasoning discussed above. Thus in this paper I develop more fully the notion of speed as part of an evaluation metric by showing how it can be applied more generally to the puzzle of competing theories.

2. Competing theories of verb movement

In this section I introduce the two competing theories (Stepwise and Swooping) of verb movement in Swedish that form the focus of this study. In Section 2.1, I review some of the core theoretical components of these proposals and discuss how both analyses can account for the same set of core utterances in Swedish. In Section 2.2 I consider a prominent potential obstacle for a Stepwise approach that emerges in the literature. I discuss how this objection presents no significant obstacle to the Stepwise approach, and that both theories are seemingly adequate in accounting for the core utterances. In Section 2.3 I discuss more formally how the Swooping approach can be considered the more complex approach.

2.1 Introducing the two theories

I now introduce the two competing theories for verb movement and discuss how they can account for the core linguistic data in matrix and embedded clauses. That the two theories both have such empirical robustness can be taken as support that they are indeed competing theories: they both generate exactly the same, desired set of utterances in Swedish. To see this, I focus on the position of the finite verb in canonical matrix and embedded clauses. I begin by reviewing the core data involving these clauses.

Swedish exhibits V2 word order in matrix clauses. In the well formed (1a), the finite auxiliary occurs in second position, immediately following the clause-initial constituent, but (1b) is ungrammatical, as the finite verb now appears in third position.

(1) a. Erik hade verkligen köpt boken.
Erik had really bought the.book
'Erik had really bought the book.'
b. *Erik verkligen hade köpt boken.
Erik really had bought the.book
(Platzack 1986b: 27-28)

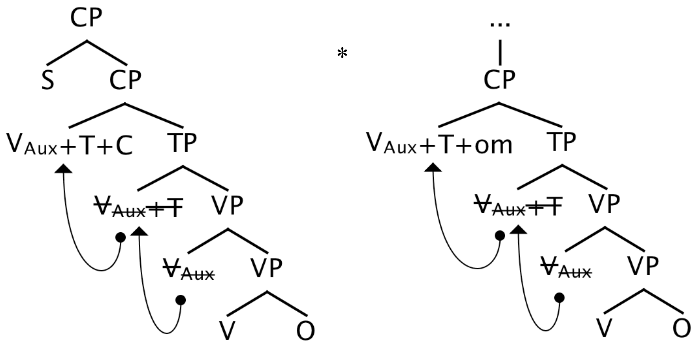
In contrast, canonical embedded clauses are not V2: the presence of complementizers such as *om* 'if' results in the finite verb appearing a position that is farther from the left-edge of the clause, where it must follow the embedded subject and clause-medial adverbials, as in (2).

(2) Vi frågade [om författaren inte skrev någen bok i år].
we asked if author.the not wrote any book in year
'We asked if the author didn't write any book this year.'
(Waldmann 2011: 332)

An influential approach to the contrasting positions of the finite verb in the core data in (1) and (2) has been to assume that the position of the finite verb in matrix clauses is the same as the overt complementizer in embedded clauses (cf. Bach 1962; Koster 1975; den Besten 1977, 1989; Platzack 1986a; Holmberg and Platzack 1995). In this approach, the finite verb moves to the head C in matrix clauses, and V2 word order results from some other constituent occupying SpecCP. This is illustrated schematically in (3a) for a simple transitive utterance with a finite auxiliary. Under the assumption that C and SpecCP are the two highest available structural positions for constituents in matrix clauses, V3 orders such as (1b) are blocked as desired. In embedded clauses, however, the presence of an overt complementizer such as *om* in C blocks this movement of the verb, as illustrated by the starred (3b), which abstracts away from the position of the subject (see below). As the verb is not moving to C in embedded clauses, it does not appear in a clause-peripheral position in these clauses. Note that as indicated in (3), I assume throughout that verb movement is highly local: to move to C, the verb must first move to the next highest head T, after which the V+T complex moves higher and attaches to C (cf. Travis 1984).

In the core data set, we can summarize the position of finite verbs in Swedish as follows: the finite verb is in C in matrix clauses, whereas the finite verb occupies some head-initial position lower than C in embedded clauses with an overt complementizer. I assume that any grammar under either of the competing theories that can account for our data set and capture these verb placement properties is a target grammar for Swedish.

(3) a. Finite V raises to C in matrix clauses b. Overt complementizers (e.g. *om* 'if') block V raising to C in embedded clauses



The central question here is how the verb moves to C, and relatedly, what happens to the verb when it does not move to C. (3b) is intentionally vague as to the position of the finite verb: the tree simply shows that the verb does not move to C, but leaves open the possibility that the verb may have moved to T. To see this question more clearly, I will introduce three parameters of verb movement. The competing theories differ only in which of these parameters are included in the learner's hypothesis space. The Stepwise theory contains only the first two of these parameters; we can align the approach found in Platzack (1986a) and Lightfoot (1991, 1995) with the Stepwise theory because of its inclusion of just these two relevant parameters. The Swooping theory incorporates the third parameter; we can align the approach found in Holmberg and Platzack (1995) and Vikner (1995) with the Swooping theory because of its inclusion of a parameter that is comparable with this new parameter. The two verb movement parameters of the Stepwise theory find independent motivation in the literature, and I will assume that the Swooping theory includes all three parameters.[4] Section 2.2 provides a review of some of the core issues that emerge from the Swedish literature as regards the inclusion of this new parameter.

I first consider the Stepwise theory. Its two parameters are presented fairly descriptively in (4) and (5) in terms of attraction requirements on a head, and are amenable to various formalizations, such as those considered in Chomsky (1995, 1998).[5]

(4) V-to-T movement parameter:
[+VT]: If finite T is a bound morpheme, it attracts V
[−VT]: Finite T does not attract V

4 The discussion and the numbers of parameters here are based on the simplification that there is a small number of syntactic heads in a clause. The numbers of parameters could be increased, though, if the number of syntactic heads is assumed to be larger. What is crucial is at least (a) all things being equal, the Swooping approach will contain a greater number of parameters; and (b) as we will see below, a fundamental difference between the two approaches is that the set of Swooping parameters allows for more grammars in which the verb moves to C.

5 A reviewer questions the parameters in (4) and (5) under the assumption that as a set, they are incompatible with Chomsky's (2008) Feature Inheritance (FI). Contra Chomsky (2008), I assume that FI is not an obligatory ingredient in the grammar. Support for this view comes from Haegeman and van Koppen (2012) and Diercks (2013), who provide strong empirical arguments against the existence of FI in several languages. Consequently, I assume that theories with no FI, such as those in this paper, are in principle viable analytical possibilities.

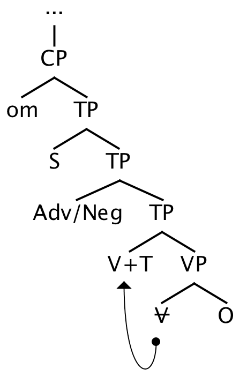
(5) T-to-C movement parameter:
[+TC]: If C is a bound morpheme, it attracts T
[−TC]: C does not attract T

If the V-to-T movement parameter is not positively set, then the verb will remain in-situ (but see below for how this can change with (7)). V will leave VP under a [+VT] grammar, but a [+VT] setting is not sufficient for V to move to C: it is also necessary for the T-to-C movement parameter to be positively set. Under a [+TC] grammar, the head T, along with whatever has moved to T such as V, will move to C. Thus in a Stepwise framework, it is only under a [+VT, +TC] grammar that the finite verb can move to C. We can think of this kind of theory as involving a "Stepwise" approach to movement because the verb ends up in C as a result of distinct parameters that are positively set and allow the verb to raise incrementally higher (i.e. step-by-step).

Before considering (1) and (2) further, it is helpful in pinning down the position of the verb to add parameters regarding head-complement order (which can all be set independently) to the set of parameters in (4) and (5). I include three further parameters for CP, TP, and VP-headedness, each of which has values of [X-init(ial)] or [X-fin(al)], according to which a given phrase with head X will be head-initial or head-final respectively.

We can see how the Stepwise theory accounts for the core data in (1) and (2). In matrix clauses, under a [+VT, +TC] grammar the structure will be as in (3a) with head-initial CP, TP, and VP. The verb moves cyclically through T to C, and another constituent occupies SpecCP as appropriate. Because the verb is moving to such a structurally high position, it will regularly precede adverbials unless they occur in SpecCP. Note that the parameters in (4) and (5) are general in that they are applicable in both matrix and embedded clauses. In embedded clauses, under a [+VT, +TC] grammar the finite verb will move to head-initial TP in embedded clauses, but will not raise to C given a complementizer such as *om* 'if'. We can account for the verb following negation and the subject in (2) if we assume under this approach that clause-medial adverbials attach at the TP level in matrix and embedded clauses in the target grammar, and that subjects likewise raise to the TP level:

(6) Stepwise embedded clause under a [+VT, +TC] grammar (with post-subject adverbial); also a Swooping embedded clause under (8a-b, e) below



In summary, there is a single target grammar under the Stepwise theory. This grammar is [+VT, +TC] and has head-initial phrase markers.

I now consider the Swooping theory, under which, there is a new parameter that can be responsible for verb movement to C. The inclusion of this parameter is the only formal difference between the two theories. This new parameter is given fairly descriptively in (7).[6]

(7) V-to-C movement parameter:
[+VC]: If C is a bound morpheme, it attracts V
[−VC]: C does not attract V

Note that (7) does not mean there is non-cyclic verb movement. If this parameter is positively set, the verb moves cyclically through T and ends up with V+T in C. What is crucial is that it is the only parameter that needs to be positively set for the verb to move to C. This contrasts with the Stepwise theory. Thus we can think of this as

6 A reviewer asks if we could say that the Stepwise theory also has parameter (7), but that it is necessarily set as [−VC]. However, there is no clear motivation for saying this. Linguistic parameters by definition must be able in principle to be set to different values by the learner; stipulating an obligatory [−VC] setting would amount to an invariant principle in which C does not attract V. The role of such a principle is not clear, though, as it is seemingly subsumed by a more general principle that I assume operates in both theories: movement that is not motivated for interpretive purposes (e.g. the verb movement in this paper) must be driven by an attracting head.

a “Swooping” approach to verb movement because by simply setting one parameter we can, with one fell swoop, ensure that the verb raises to C.

Although a V-to-C parameter is sufficient for the verb to raise to C, I assume that the parameters in (4) and (5) are considered by the learner under the Swooping approach. They have been independently motivated for other Germanic languages in Zwart (1997), and I assume that these parameters are included in a learner’s hypothesis space and are set independently by the learner. Under the null hypothesis that the initial state of the hypothesis space of all learners is uniform, then, a theory with (7) will also have parameters (4) and (5). Including all three verb movement parameters under the rubric of the Swooping theory will have important consequences first for how many grammars are compatible with the Swedish data, and ultimately for learning results.

Under the Swooping theory we can account for the core Swedish data as follows. In matrix clauses, under a [+VC] grammar we capture V2 effects in a way that is essentially the same as what we saw above for the Stepwise theory: the finite verb will move to the highest head position, and some other constituent can occupy SpecCP. Thus the tree in (3a) is also applicable under the Swooping approach. It is also possible for the verb to move to C under a [−VC] grammar, so long as the V-to-T and T-to-C movement parameters are positively set, as in the Stepwise theory. In what follows, though, I will largely focus on discussing [+VC] grammars under the Swooping theory so as to highlight the difference between the theories.

An important difference between the theories is that under Swooping there are multiple grammars in which the verb moves to C. Because the verb will still move to C with a [+VC] setting regardless of how the V-to-T and T-to-C parameters are set, it means that the partial grammars in (8) are all in principle compatible with the Swedish data in (1).[7]

---

7 Note that the grammars in (8b-e) are not contradictory, with the V-to-C movement parameter being set independently of other parameters. This is because the parameters are formulated in terms of the attracting properties of particular heads. For example, the [−VT] value in (8c), does not proscribe verb raising to T from ever happening – it states that T has no requirement for such movement. Thus, with the [+VC] setting in (8c) and local head movement (cf. Travis 1984), the verb will end up raising to, and then higher with, T – but not because of any attracting requirement of the T-head itself. Conversely, if a complementizer blocks raising to C under (8c), then the [−VT] setting will preclude raising to T in embedded clauses.

(8) Some partial target grammars under the Swooping theory for verb movement
a. [+VC, +VT, +TC] c. [+VC, −VT, −TC] e. [−VC, +VT, +TC]
b. [+VC, +VT, −TC] d. [+VC, −VT, +TC]

Thus under the Swooping approach there are multiple possible target grammars a learner could acquire and data such as (1) are ambiguous input for the learner; as we will see in Section 3, this ambiguity will be an important factor in our expectations regarding learning results. Indeed, data such as (1) are further ambiguous under both the Stepwise and Swooping theories, as it is not clear given only (1) whether the verb has raised to T or all the way to C.

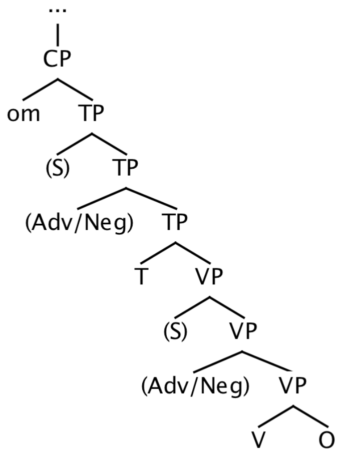
In embedded clauses of the sort in (2) under the Swooping approach, again, overt complementizers such as *om* 'if' will block movement of the verb to C, even when the grammar has a [+VC] setting. Note that if C is an available landing site for the verb, then the verb must move to C (stopping off in T along the way) under [+VC]. But if movement to C is blocked, then (setting aside other movement parameters for the moment) the verb remains in-situ. Because of the various potential target grammars in (8), though, it is still possible for the verb to raise to T in embedded clauses. If the learner adopts any of the [+VT] grammars in (8a-b, e), with adverbials uniformly attaching at the TP level in matrix and embedded clauses in such target grammars and the subject raising to the TP level, then the verb can still move to head-initial TP in embedded clauses as in (2). We can represent such a structure using the same tree in (6) above, which illustrates an embedded clause with a post-subject adverbial. I return to the claim that Swedish could indeed have V-to-T movement in embedded clauses (which has been disputed in the literature) in Section 2.2. In contrast, if the learner adopts either of the grammars in (8c-d), the verb remains in-situ in embedded clauses (with affixal T necessarily attaching to V via some other mechanism such as affix lowering or Morphological Merger; cf. Chomsky 1957, 1981; Marantz 1988):[8, 9]

---

8 These mechanisms are last resort operations in the absence of verb movement. For the learning procedure in Section 3.1, only if the learner has not sampled appropriate parameter values that result in verb movement for an input token, will these operations be available in order to attempt to construct an input-compatible grammar.

9 Thus only under the Swooping theory is it predicted that in a natural language the verb could move to C in matrix clauses but remain in-situ in embedded clauses; cf. Section 2.2 for some

(9) Swooping structure of embedded clauses under (8c-d) (with post-subject adverbial)



Note that the tree in (9) leaves open the possibility that the subject may not have undergone subject raising from its base position (as well as the possibility for adverbials to attach at some fixed level, either VP or TP, in embedded and matrix clauses; cf. note 11). I have illustrated this in the tree by indicating various possible positions with parentheses. Nevertheless, the [−VT] structure in (9) is consistent with (2) under the Swooping theory, although it would be non-target under the Stepwise theory. I return to these points below, but for the moment we can say the following. Under one set of target Swooping grammars, the verb raises in embedded clauses, and the subject will appear at the TP level, but under another set of target Swooping grammars, the verb does not raise in embedded clauses, and the subject may appear at the TP level or lower.[10]

Further, although the focus of this paper is on verb movement in Swedish and not the position of the subject, the position of subjects does bear on the position of the verb. For instance, if the learner concludes that the subject always raises to the TP level, then the learner could infer something about the position of the verb based on where it appears in relation to the subject. Consequently, I discuss subjects

---

Swedish-related discussion.

10 Note that if the verb never raises only as high as T, as in a [±VC, −VT] grammar, it is also possible for a target grammar of Swedish to have a head-final TP. Otherwise, just as in the Stepwise theory, the target value for all the headedness parameters is [X-init].

briefly below, and I discuss subjects again in Section 3 in the context of ambiguity. In this section I wish to highlight two related points. The first is that from the learner's perspective, just as there is ambiguity regarding verb movement parameters, the position of a subject that is lower than C in Swedish can also be ambiguous. As I discuss in Section 3, such ambiguity will complicate making the kind of inference mentioned above.

The second point is that although it is often assumed (e.g. Holmberg and Platzack 1995; Lightfoot 1995) that subject raising in Swedish does occur, the ambiguity illustrated in (9) is viable in that it is consistent both with the empirical facts of Swedish and earlier theoretical work. Various researchers (e.g. Platzack 1986b; Svenonius 2002; Engdahl *et al.* 2004) have noted that subjects lower than C can precede or follow clause-medial adverbials (cf. (9)). Further, multiple structural subject positions in Swedish have been suggested in works such as Svenonius (2002). The order of subjects and adverbials in Swedish has been linked to interpretive factors related to information structure (Svenonius 2002; Engdahl *et al.* 2004), but I follow Richards (2016) in assuming that these factors do not clearly indicate the actual structural position of the subject, or whether the subject has undergone any raising.

The crucial point that emerges from the discussion of subjects is that the position of the subject − as well as whether the verb has raised − in examples such as (2) is ambiguous, and this is true under both theories.[11] Of course, if the verb raises to T, as in (6), then there must be some subject raising. And given that the target grammar under the Stepwise theory involves the V-to-T movement in (6), it follows that there must be subject raising in the target grammar under the Stepwise theory. However, input such as (2) does not by itself tell the learner that either the subject or the verb has undergone raising. Under either theory, the scenarios in (6) and (9) are consistent with (2). The ambiguity with regard to raising that we see in (2) is pervasive throughout Swedish, and in Section 3.2, I will discuss how because of this

11 The discussion throughout this paper touching on the ambiguity of the position of the subject (and how that relates to the position of the verb) also applies in its essentials to the position of a clause-medial adverbial, with the vast majority of the learner's input being compatible with it attaching at either the VP or TP level (cf. discussion of (13)) under either theory. For reasons of space, I do not discuss learning the position of adverbials, but I note that incorporating such a step of learning into learning simulations is not expected to alter the overall conclusions based on the learning results reported here.

ambiguity, it is not clear that the relative position of the subject can be used with high reliability to learn target values for verb movement parameters.

In sum, we have now seen how both theories (with the movement and headedness parameters from above) can account for verb placement in the core data in (1) and (2).

The claim, then, is that these theories are competing theories. Again I focus on the core verb placement patterns in (1) and (2), but the claim holds more generally. Both theories allow for target grammars that account for exactly the set of finite verb positions that is possible in Swedish. Under these target grammars only the following verb placement possibilities obtain: (a) the verb raises to head-initial C in matrix clauses, and (b) the verb does not raise to C in embedded clauses with overt complementizers, but rather occurs in some lower, head-initial position. The data set in (1) and (2) is admittedly small, but the claim of competing theories is further substantiated with the schematic corpus in Section 3 that I use for the learning simulations. This corpus provides a diverse range of data that represents many of the core word order patterns in Swedish, as well as much of the core of the input we expect a learner of Swedish to be exposed to (cf. Yang 2002; Westergaard 2006). Crucially, the input types in the schematic corpus follow the same core verb placement patterns that we have already seen in (1) and (2): the verb can be analyzed as being in C in matrix clauses, whereas the verb is in some lower head-initial position in embedded clauses. Accordingly, we can expand our core data set that our competing theories are meant to capture from data of the sort in (1) and (2) to now include the input types found in the schematic corpus introduced in Section 3. For the purpose of this paper, then, I will take data of the sort in (1) and (2), along with the input types of the schematic corpus, as being representative of Swedish.

As target grammars under both theories can account for exactly the learner's input, both theories have the potential to present a learner with a target grammar that is in principle viable. This exemplifies the puzzle of competing theories, and the question I will focus on is how well these theories can result in learning such target grammars as per the evaluation metric I have introduced. The conclusion I reach is that a V-to-C movement parameter is crucial for faring better. This supports the Swooping theory, and in turn supports proposals such as Holmberg and Platzack (1995) and Vikner (1995), which are aligned with the Swooping theory.

Finally, the claim that the Stepwise and Swooping theories are competing theories for Swedish stands despite the fact that one theory has more parameters than the other. Even though, all else being equal, including a V-to-C parameter results in a Swooping parameter space whose members form a proper superset of the members of the Stepwise parameter space, this does not mean that the utterances of a language acquired under the Swooping theory necessarily form a proper superset of those of a language acquired under the Stepwise theory. The generative capacity of the two approaches for Swedish is the same: they both ostensibly account for the same set of utterances.

2.2 A potential problem for the Stepwise theory?

In Section 2.1, I discussed how both the Stepwise and Swooping theories can be considered competing theories. An important component of this discussion was that it is possible for there to be V-to-T movement in embedded clauses in Swedish. However, this claim has been challenged in the literature. In advocating a V-to-C movement parameter, various researchers, such as Holmberg and Platzack (1995) and Vikner (1995), have claimed that the finite verb in Swedish does not raise to T in embedded clauses. In the terms of the parameters from Section 2.1, this would mean that Swedish is [−VT]. In Section 2.1 we saw how Swedish could be [−VT], but only if the grammar is also [+VC]. Recall, though, that under the Stepwise theory, Swedish must be [+VT]. If the claim about Swedish not having embedded V-to-T movement is correct, then we would have an argument against the Stepwise theory, and for a theory with a V-to-C parameter. In other words, despite the *prima facie* adequacy of the Stepwise theory that we saw in Section 2.1, the Stepwise theory could not be a competing theory if it can be independently argued that there is no embedded V-to-T movement in Swedish. The question, then, is whether we have a problem for maintaining the Stepwise theory as a viable theory for Swedish because of its reliance on embedded verb movement.

In this section, I briefly review perhaps the most prominent argument against embedded V-to-T movement in Swedish. This argument involves the Rich Agreement Hypothesis (RAH). I conclude that the RAH cannot be taken as an argument against embedded V-to-T movement in Swedish. Consequently, the

Stepwise theory remains *a priori* a viable theory for Swedish, and we are still faced with the puzzle of competing theories.[12]

According to the strongest version of the RAH, V-to-T movement can be found only in languages with paradigms of subject agreement that are morphologically rich. One way to define richness is in terms of the number or kinds of morpho-syntactic contrasts that exist in the paradigm (cf. Bobaljik 2000 for a review of several proposals of this sort). This view is adopted in Koeneman and Zeijlstra (2014), which is perhaps the most recent and comprehensive work in which the strong version of the RAH is widely tested. According to Koeneman and Zeijlstra, Swedish is not a language with rich agreement, as it lacks any morpho-syntactic contrasts in its subject agreement paradigm. The RAH makes the claim, then, that there cannot be V-to-T movement in Swedish because it does not have rich subject agreement. From a learning perspective, a learner equipped with the RAH could use the RAH as a diagnostic for acquiring parameter settings for verb movement. As learners of Swedish consistently encounter input lacking rich subject agreement, after sufficient input the learner could conclude that Swedish is a [−VT] language. If correct, the RAH could be used as an independent argument against the linguistic viability of the Stepwise theory.

However, various researchers have identified cross-linguistic evidence that the strong version of the RAH is not correct. Thus, Baptista (2000) discusses several Creoles, which are not considered by Koeneman and Zeijlstra, that do not have rich agreement, but that are analyzed as having verb movement to T (cf. Klingler 2003:295, 321 for further data). Further, in a response to Koeneman and Zeijlstra, Heycock and Sunquist (2017) argue against the RAH by presenting diachronic evidence that V-to-T movement occurred for several centuries in Early Modern Danish despite the language not having rich agreement.

---

12 Another argument against embedded V-to-T movement in Swedish concerns the lack of Object Shift (OS). Holmberg and Platzack (1995) claim that OS in Swedish is possible only if the verb moves high enough in the clause. Given that OS is possible in matrix clauses but impossible in canonical embedded clauses, one might conclude, as Holmberg and Platzack do, that there is no verb movement in Swedish embedded clauses. However, an alternative analysis is available. Fox and Pesetsky (2005) present a framework for analyzing OS, in which the impossibility of OS in embedded clauses can be accounted for independently of whether the verb moves. (For a brief review of Fox and Pesetsky, see Lee 2009.) I conclude, then, that the absence of embedded OS in Swedish does not necessarily bear on the issue of whether Swedish can be a [+VT] language, or on the viability of the Stepwise theory in general.

I take this as important evidence against the RAH as a cross-linguistic generalization. Consequently, it is not clear that a Swedish learner could infer Swedish is a [−VT] language on the basis of its agreement paradigm, and thus a [+VT] analysis of Swedish − and by extension, the Stepwise theory − remain viable in principle.

2.3 The relative complexity of the competing theories

Here I discuss slightly more formally in what sense the Swooping theory is more complex than the Stepwise theory. In the terms of information theory, any complete description of the Swooping theory will contain more information than that of the Stepwise theory; this greater amount of information is the basis for calling the Swooping theory more complex.

Let us now consider what a complete description of the two theories should contain. Recall that the only point of difference between the theories was the Swooping theory's inclusion of a V-to-C parameter. Everything else in the theories − including all other parameters and principles/constraints − is assumed to be identical. I will treat it as a constant variable *K* for theory description, but for exposition will not include in *K* the verb movement parameters summarized in (10). As regards describing the difference between the theories, we can focus on (10). A description of the Stepwise theory should encode in discrete bits of information exactly (10a) and *K*, whereas a description of the Swooping theory should encode exactly (10b) and *K*.

(10) a. Stepwise movement parameters: { [±VT] , [±TC] }
b. Swooping movement parameters: { [±VC] , [±VT] , [±TC] }

As the null hypothesis, I assume that whichever parts are shared by the two theories will be encoded with equal amounts of information. As descriptions of both languages must contain (10a) and *K* (because of the superset-subset relation between (10a-b)), this means that with regard to the content of (10a) and *K*, for any particular language of encoding, the theories will be encoded with equal amounts of information > 0. Let us label the amount of information used to encode (10a) and *K* for any particular language as I_1.

We can now say that the Kolmogorov complexity (see e.g. Solomonoff 1964a, b) of the Swooping theory is greater than that of the Stepwise theory. Roughly speaking, the Kolmogorov complexity of a theory can be measured with the amount of information required to encode a description of the theory. The more information required, the more complex the theory. Crucially, for any given language used to encode a description, a description of the Swooping theory in that language must contain some extra information than a description of the Stepwise theory in that language. This is because whereas the Stepwise theory must be described exactly with I_1, the Swooping theory must be described exactly with I_1 plus some other amount I_2, where I_2 represents the amount of information (which is greater than 0) necessary to encode the V-to-C movement parameter in (10b). As $I_1 + I_2 > I_1$, it follows that the Swooping theory is more complex than the Stepwise theory.[13]

3. Implementing the model and expectations for the learning simulations

This section looks at how we can use new kinds modeling data to choose between the two theories according to the acquisition-based evaluation metric. Section 3.1 reviews some of the basics of the probabilistic learning model in Gould (2017) and provides details of a concrete implementation of the learning model for the Swedish case at hand. Section 3.2 then considers how well each theory is expected to fare with respect to the evaluation metric.

3.1 Implementation of the learning model

In the implementation of the learning simulations of Swedish, there are three aspects of the model from Gould (2017) to focus on. The first is the hypothesis

13 A reviewer asks whether a theory's having more movement parameters than some other theory means that the former is more complex. Although this holds in the case study here, we cannot necessarily conclude that it is generally true. For instance, we could imagine there is a theory that involves a high degree of complexity without containing any movement parameters, whereas there might be some alternative theory that has a lesser degree of complexity while still containing a number of movement parameters. As regards Kolmogorov complexity more generally in comparing theories, what is necessary is detailed information regarding the description of a theory.

space of the learner, which determines what a possible grammar for any learner can look like. In discussing a hypothesis space I will focus on the set of parameters that the model learns about here. The second is the input to the learner; this will take the form of a schematic corpus. The input is what the learner leverages to set parameters in an attempt to arrive at a target grammar that falls within the hypothesis space. And the third is the learning procedure for setting parameters given some input, which involves probabilistic sampling and probability updating.

It is important to make clear that there are, in fact, two implementations of the model (and thus two sets of learning simulations) that we will consider, one for each competing theory. The only difference between these implementations is in the hypothesis space itself. Thus, the learner is never faced with a choice between a set of hypotheses/parameters from the Stepwise theory and one from the Swooping theory (cf. note 1).

The hypothesis spaces are in (11) and contain the parameters from Section 2, which allow the model to capture the basic facts of finite verb placement in Swedish, and thus in principle allow for the possibility of learning a target grammar for Swedish.

(11) a. Stepwise: {[±VT], [±TC], [C-init/fin], [T-init/fin], [V-init/fin]}
b. Swooping: {[±VC], [±VT], [±TC], [C-init/fin], [T-init/fin], [V-init/fin]}

The second core component of the model is the schematic input corpus that the model learns from. This is given in (12) and represents a broad range of basic types of utterances in Swedish and closely related languages (cf. Westergaard 2006 and Waldmann 2014, as well as Yang 2002). This corpus includes input that corresponds to the core data we saw in (1) and (2), as well as a variety of other input types, all of which have been analyzed (cf. the references in Section 2) as having the same basic verb placement pattern. These input types can be described as follows.[14] (12a-d) are simple subject-initial declaratives that vary in whether they are transitive

14 I note two simplifications here of how the model treats the input. First, all subjects are assumed to be generated in SpecVP. I leave for future research the question of how unaccusatives/passives might play a role in the learning process. Second, the model does not consider the structural relationship between the embedded and matrix clauses (e.g. adjunct vs. argument); see Gould (2017:150-152) for some related discussion regarding the minimal impact such a simplification is expected to have.

(i.e. contain O) or have a finite auxiliary (i.e. contain Aux). In contrast, (12e-h) are subject-initial declaratives that contain embedded clauses, where C represents an overt complementizer such as *om* 'if'. In (12i-l), X is a variable for any initial constituent other than the other initial constituents found elsewhere in (12). (12m-n) are object-initial declaratives; (12o-r) are polar questions. The remaining types in (12s-w) all contain Neg, which is a variable for any adverbial (such as negation) that occurs clause-medially. See the appendix for illustrative examples of the input types in (12).

(12) Schematic corpus of Swedish input types and their relative probabilistic frequencies

a. SV	4.75%	e. SV[C SV]	2.75%	i. XVS	6.5%
b. SVO		f. SV[C SVO]		j. XVSO	
c. SAuxV		g. SV[C SAuxV]		k. XAuxSV	
d. SAuxVO		h. SV[C SAuxVO]		l. XAuxSVO	
m. OVS	1%	o. VS	8.5%	s. SVNeg	1.75%
n. OAuxSV		p. VSO		t. SVNegO	
		q. AuxSV		u. SAuxNegV	
		r. AuxSVO		v. SAuxNegVO	

w. SVNeg[C SNegAuxVO] 0.08%   x. SV[C SNegAuxVO] 0.92%

The model learns in an online fashion, meaning that every time the learner is exposed to a single token of input (i.e. a token of one of the types found in (12)), the learning procedure described below will apply. Input is presented randomly to the learner one token at a time by the model sampling tokens according to the frequencies given in (12). The proportions of the frequencies in (12) correspond to the values of a multinomial that models the probability distribution of the input types. These frequencies are based on the corpus frequencies of different sentence types found in Westergaard (2006) and Waldmann (2014), as well as Yang (2002) and Lightfoot (1997).

Lastly, I briefly review the basics of the third core component of the model, namely the probabilistic learning procedure. Every time the learner receives a token of input, the learner will construct a grammar that is compatible with that input.[15] The

15 The learning simulations here implement the simplified version of the model in Gould (2017). This contrasts with the full version in Gould (2017), in which for each token of input the learner generates a string-meaning pair that matches that token of input. Nevertheless, learning conclusions

learner does so by sampling a set of parameter values and then checks to see whether that set of parameter values is compatible with the input. Sampling is done by using probabilistic weights that are associated with each parameter value. At the start of the learning process, the weights associated with each of a particular parameter's values are equal. These weights are updated each time a grammar is constructed that is compatible with a token of input.[16] This means that the input-compatible parameter values that were sampled are reinforced by increasing their probabilistic weights. Doing so increases the likelihood of sampling in the future parameter values that were previously successful in constructing an input-compatible grammar. Over time, the learner will on average reinforce more and more some set of parameter values that is most compatible with the input corpus. And we can say that a parameter is set to a particular value when the total probabilistic weight associated with that parameter is concentrated on that particular value.[17] In this way the learner acquires a grammar of best-fit to the overall input corpus.

Two important aspects of this learning procedure are first that the weights associated with parameter values are updated only incrementally after each token of input, and second that these weights are updated on the basis of the current token of input without making reference to prior tokens of input. This means that just because one particular input-compatible grammar was constructed for a prior token of input, it does not mean that the learner will necessarily sample the parameter values of that grammar when faced with a new token of input. The learner may be more likely to do so, but the learner is not strongly committed to the parameter values of some particular grammar after a single token of input. It is only after a parameter value's weight has been repeatedly reinforced that the learner reaches a higher level of confidence that such a value is found in the target grammar, and is thus much more likely to adopt that parameter value. This is significant, in part, because with such an online learning procedure, the impact of some input that is unambiguous for a particular parameter value *x* might be highly minimal. With such a learning

---

along the lines of those in this paper are also expected under the full version of the model (cf. Section 3.2 for further discussion of enriching the model).

16 The learning simulations in this paper follow the update procedure detailed in Gould (2017:192-193). For precise specifications on the update procedure, the reader is referred to this source.

17 Given this discussion of variable weights, to (8a-d) we can add further target grammars that do not have settings for [±VT] and [±TC] (cf. Pintzuk 2002).

procedure, the learner certainly cannot compare different input types side by side to reach inferences. In other words, just because one token might be unambiguous for a particular parameter value *x*, it does not mean the learner necessarily adopts such a value in the analysis of some other ambiguous input type. But moreover, this learning procedure means that any conclusions reached from a single token or a small amount of unambiguous evidence by assigning an incremental amount of probabilistic weight for some particular parameter value *x* can be (at least temporarily) neutralized, or drowned out, by a large amount of ambiguous evidence, for which a relatively larger amount of probabilistic weight could be assigned to some competing value *y* for the parameter in question. In Section 3.2 we will see an example of this, where the effect of some unambiguous input − (12w) above − can effectively be drowned out, at least for long stretches of learning. What is especially crucial is the overall shape of the corpus (i.e. how much unambiguous input for a parameter value there is relative to ambiguous input), as well as what the model is able to learn from ambiguous input.

Indeed, it is possible for ambiguous input to be highly informative for the learner, and the extent to which this holds or does not hold in (12) will be an essential ingredient in understanding the learning expectations. To see this, it is necessary to examine the tendencies that are found on average among grammars that are compatible with a particular type of input data. For example, even if a token of input *T* is ambiguous with regard to a particular parameter value *x*, if the preponderance of grammars that are compatible with *T* contain that value *x*, then there is a high likelihood for the learner to sample that value *x*. This means that *T* favors such a parameter value *x*, and the learner will accordingly be pushed on average toward adopting a grammar with a parameter setting of *x*. However, sometimes the ambiguity is such that neither of a particular parameter's values is favored. That is, the balance in input-compatible grammars is equal (or close to equal) for both values of a particular parameter. I next look at the implications of such ambiguity for learning expectations.

3.2 Expectations for the learning simulations

In this section I discuss the learning expectations for the model according to which the Stepwise theory is expected to fare worse than the Swooping theory under the evaluation metric. To see why we, we will consider the nature of the ambiguity of the learner's input and whether the input favors a particular parameter value, as discussed in Section 3.1. The crucial difference in expectations between the competing theories is related to this ambiguity and hinges on whether the model can learn that the verb raises all the way to C. For the Stepwise theory, this will involve a difficulty in learning a target [+TC] parameter setting, whereas for the Swooping theory, there will be no comparable difficulty in learning a target [+VC] parameter setting. As I discuss, the vast majority of the input is ambiguous and favors neither a [+TC] nor a [−TC] value under the Stepwise theory, meaning that there can be a delay in learning a [+TC] setting and that learning a non-target [−TC] parameter setting (from which it will be difficult if not impossible to recover) is sometimes an expected outcome. In contrast, although ambiguous, the vast majority of the input favors a [+VC] value under the Swooping theory, meaning that learning a target [+VC] parameter setting is an expected outcome that can be reached relatively quickly. More broadly, under neither theory is there difficulty in learning a target grammar as regards the other parameters from Section 3.1 (i.e. V-to-T movement and head-complement order), and indeed, learning expectations are consistent with these results. As there is no substantive difference between the theories in this regard, to streamline discussion I will largely focus on learning expectations pertaining to the crucial difference mentioned above − that is, on [±TC] values for the Stepwise theory and on [±VC] values (and to a certain extent T-to-C movement) for the Swooping theory. Let us now consider the ambiguity and expected learning outcomes in more detail, by focusing on illustrative examples of input to the learner. The ambiguity was previewed in Section 2.1 and involves how high the verb raises; here we will see it is a more general phenomenon.

I begin with the Stepwise theory. Consider the OVS input in (12m). Supposing for the moment that the object is in SpecCP, then where could the verb be? Assuming there are only leftward specifiers for subjects (and setting aside the possibility of extraposition), this input is unambiguous evidence for verb movement (cf. note 14). That is, the only way for the lexical verb to precede the subject is to move past it. But verb movement *per se* is not sufficient for the verb to be in C, as this input is ambiguous as to where the verb moves. The verb could raise only as high as T, or it could raise all the way to C. In other words, the learner must analyze this input by sampling a [+VT] grammar, but input of this type is compatible with either value of [±TC]. Further, focusing on embedded clauses, as in (2) or (12x), does not help with the ambiguity: as the complementizer blocks verb movement to C, either value of [±TC] is in principle compatible with the input. Thus input of these types is ambiguous with regard to T-to-C movement.

The challenge that begins to emerge from these examples concerns a way for the learner to acquire a grammar with a target [+TC] parameter setting. Indeed, it can be observed that the only input that provides unambiguous evidence for a [+TC] grammar is input type (12w), with a clause-medial adverbial in both the matrix and embedded clauses. Assuming a fixed, left-adjoined position for adverbials either at TP or VP, the learner must sample a [+VT, +TC] grammar when faced with (12w); the relevant structure is illustrated schematically in (13). As the finite matrix verb precedes the adverbial, it must have raised at least as high as T, meaning there is a [+VT] grammar. Accordingly, the finite embedded verb will also have raised to T (and can raise no higher because of the blocking complementizer). But as the finite embedded verb follows the adverbial, the adverbial must be adjoined to TP. This means that the matrix adverbial is also adjoined to TP. Consequently, it must be the case that the matrix verb has moved higher to C, where it can precede the adverbial.

(13) Schematic structure of (12w) under the Stepwise theory
[$_{CP}$ V+T+C [$_{TP}$ Neg [$_{TP}$ ~~V+T~~ [$_{VP}$...~~V~~ [$_{CP}$ *om*$_{C}$ [$_{TP}$ Neg [$_{TP}$ V+T [...~~V~~...]]]]]]]]

Thus there is unambiguous evidence for [+TC] under the Stepwise theory − namely (12w) − but such input is rare, occurring on average in 0.08% of the learner's input. Given, the incremental online learning procedure described in Section 3.1, a reasonable expectation is that such rare, unambiguous evidence will get drowned out (at least for some long stretches of learning) by the overwhelming preponderance of the input (occurring on average in 99.92% of the learner's input), which is ambiguous evidence for [±TC]. I note here that such unambiguous input being drowned out by ambiguous input is plausible in principle given the evidence in Gould (2017: 149), which suggests that parameter mis-setting is still possible even when over 10% of the input corpus on average is unambiguous for the target value of the parameter in question. If 10% in this sense represents some possible upper bound for when the impact of unambiguous input can be drowned out for some period of time, we can observe that at 0.08% of the input corpus, the relative frequency of unambiguous (12w) falls well within such a boundary. Accordingly, our expectations about learning will be shaped by the overall shape and analysis of the corpus more generally, and not specifically by (12w). The question then becomes: what effect will this ambiguous input have on the learner?

Analysis of the corpus reveals that the ambiguous input in (12) favors neither [+TC] nor [−TC]. For each input type other than (12w), the number of input-compatible grammars splits evenly between [+TC] on the one hand, and [−TC] on the other.

Nevertheless, although neither grammar is favored by the ambiguous input, it is still possible to formulate learning expectations based on this kind of ambiguity. With such an even split among compatible grammars, for the vast majority of the input, the learner could reinforce either parameter value for [±TC]. But as neither value is overly favored, early on in learning it is possible for the learner to move repeatedly back and forth between reinforcing either value. This can result in delays in learning any parameter setting for T-to-C movement. Moreover, given the randomness of the sampling procedure and given a learner that is not overly tentative in the amount of weight that parameter values can be reinforced by, it is also likely that at some point during the learning process the learner will reinforce a single parameter value sufficiently so as to tip the balance, thereby favoring one parameter value over the other, and allowing the model to learn a parameter setting for that value. As regards parameter setting, we would thus expect the following: each time the model is run under the Stepwise theory (called a run, which is intended to simulate a single learner among a population of learners), some runs of the model will end up favoring one value and learning a target parameter setting of [+TC], whereas other runs will end up favoring the other value and learning a non-target parameter setting of [−TC]. Indeed, Gould (2017) shows that in learning simulations for Korean with this kind of ambiguity in the corpus, such variability obtains across learners with regard to learning a grammar with or without verb movement. Further, parameter mis-setting can further delay learning a target grammar. One way a learner could reset a non-target [−TC] parameter setting to a target [+TC] setting is by encountering a stretch of input that contains a sufficient amount of the unambiguous (12w). Given the randomness of the input, such a stretch of input is possible, but as (12w) is relatively rare, we expect the learner on average to have to wait a long time to encounter it. This would result in a delay in learning a target grammar, which in the worst-case scenario would stretch to the limit of time available for language acquisition.

Thus, because of both pervasive ambiguity in the corpus and the even balance of input-compatible grammars for [±TC], we expect both (a) delays in learning a parameter setting for [±TC], as the learner is not initially pushed strongly toward either value; and (b) for it to be likely for the model to acquire a non-target [−TC] parameter setting, which can result in further delays in learning a target grammar. These expectations would seemingly not bode well for the Stepwise theory under our evaluation metric, although we must consider the Swooping theory in more detail to see this more clearly.

Before turning to the Swooping theory, I return to the issue from Section 2.1 of the position of the subject. I mentioned there that there was the possibility that learning something about the position of the subject could in principle allow the learner to make an inference about the position of the verb. Thus, supposing the model has learned the subject must raise to SpecTP, then any time the learner encounters a post-verbal subject, such a subject must be analyzed as being in SpecTP, and the verb must accordingly be analyzed as being in a higher position, namely C. Thus, learning about the position of the subject could in principle provide the model under the Stepwise theory with a path to improving learning outcomes when it comes to the challenge of learning a target [+TC] setting. Although such a path toward a [+TC] grammar seems promising at first glance, enriching the model to include subject movement to TP would not substantively change the overall picture of learning expectations presented so far.

Before discussing such an enrichment, it is important to emphasize that because of the online learning procedure discussed in Section 3.1, the learner cannot make inferences by comparing multiple tokens of input side by side. Suppose that the learner has a [+VT] parameter setting, and is then confronted with embedded clause input in which the subject precedes the finite verb. For such a learner, if the verb raises to T in a grammar that is compatible with such input, then the subject must be at the TP level, and the learner can accordingly reinforce a parameter value for subject raising to TP. However, any conclusions based on this token of input containing an embedded clause do not carry over to the next token of input beyond the incremental amount of parameter reinforcement. Thus, if the next token of input is ambiguous with regard to subject raising to TP, it is possible for the learner to sample and reinforce a grammar in which there is no subject raising. What is crucial is to what extent parameter values for subject raising are reinforced during the learning process, and this is determined in part by the nature of the ambiguous and unambiguous evidence that favors or does not favor such raising.

Thus it is interesting to observe that when the hypothesis space of the Stepwise implementation of the model is further expanded to include hypotheses regarding the position of the subject (as well as adverbials and displaced objects), the corpus in (12) favors learning a grammar that does not have subject raising to TP. To get a sense of how this can be the case, let us focus on the illustrative example of the XVS input in (12i), which is unambiguous for [+VT]. This type of input is ambiguous in such a way as to not favor learning a grammar with subject raising to TP. One ambiguity concerns the subject, which could remain in-situ or could raise to the TP level, depending on how high the verb raises. Turning to X, its position is also ambiguous, as it could attach at the TP or the CP level. Thus there are multiple positions for X without subject raising (and thus multiple ways of constructing an input-compatible grammar), but only one position for X (i.e. at the CP level) with subject raising to TP. Further analysis of this kind of input indicates that there are more input-compatible grammars with the subject not raising to TP than there are input-compatible grammars that have the subject raising to TP. When we factor into consideration the effect of ambiguities of this sort, only 38% of the input corpus favors subject raising to TP, whereas 62% of the corpus favors the subject not raising to TP. Moreover, nearly all of that 62% favors no subject raising substantially more so (i.e. has a greater percentage of input-compatible grammars) or to a highly comparable extent as the 38% that favors subject raising.

A reasonable expectation, then, is that the learner will be pushed sufficiently toward a non-subject raising grammar, such that it is likely for the learner to have parameter mis-setting with regard to obligatory subject raising.[18] This complicates making any inferences about the position of the verb based on learning that there is subject raising.

18 In addition to the challenge of learning a [+TC] grammar, another potential difficulty for the Stepwise theory, then, is learning there is subject raising. Recall that under the Stepwise theory, there is subject raising in the target grammar. As I focus on verb movement, though, I will not consider further this kind of difficulty beyond its implication (discussed in the main text) for verb movement.

As regards our expectations concerning verb movement, then, we can hypothesize the following. Assuming, that learning there is subject raising is likely to be fraught with delays and parameter mis-setting that can (at least temporarily) neutralize it as a helpful source in learning a target grammar, we then have no reason to think that including subject raising to TP will substantively change the nature of the [−TC] parameter mis-setting and its attendant delays in parameter resetting that are expected under the Stepwise theory.[19, 20] Moreover, even if including such subject raising were to slightly speed up learning for Stepwise learners that do not have [−TC] parameter mis-setting, we do not expect to see any improvement as a consequence of this in the Stepwise learning results relative to the Swooping ones. This is because when it comes to learning V raises to C, a learner equipped with the Swooping theory can benefit from learning there is subject raising in a comparable way that a learner equipped with the Stepwise theory could. Further, the learning expectations under the Swooping theory (discussed below) still hold with the inclusion of such subject raising.

I conclude, then, that in expanding the model, considerations of the subject position present no clear obstacle to the earlier expectations regarding the challenge of learning that finite matrix V raises to C under the Stepwise theory. Further, we expect comparable conclusions based on simulation results with such an expanded implementation of the model.

Next, I consider the kind of ambiguity we see with the Swooping theory, as well as expected learning results, with a focus on the V-to-C movement parameter. Recall that input of the sort in (1) and (2) is ambiguous with regard to [±VC] parameter values. Indeed, the same kind of reasoning that we have just seen for the Stepwise theory allows us to understand how the input more generally in (12) is ambiguous for [±VC]. For instance, input such as the OVS input in (12m), which as discussed above is only compatible with a grammar in which there is verb movement, is still ambiguous for [±VC]. The verb movement could be the result of a [+VC] grammar, or it could be the result of a [−VC, +VT] grammar.

---

19 Indeed, preliminary analysis suggests these delays will be sufficiently long so as to allow for fairly robust [−TC] parameter mis-setting given the results in Section 4 and the results on mis-setting in Gould (2017).

20 We can also note that including subject raising to TP is not expected to help with the anticipated delays early on during learning when the learner moves back and forth between reinforcing [−TC] and [+TC] grammars. This is again because even when including such raising, the proportion of input-compatible grammars across input types is highly balanced for the values [±TC].

Given this ambiguity, the challenge for the learner under the Swooping theory is in learning the matrix finite verb raises to C. More precisely, the challenge can be framed as one of learning either a [+VC] or [+TC] grammar. Setting aside input type (12w) for the moment, based on the discussion above, we can say that the rest of the input corpus is ambiguous for [±VC]. Still, as with (13) for the Stepwise theory, (12w) provides unambiguous evidence that the finite verb raises to C for the Swooping theory: for the finite verb to precede the matrix adverbial, an input-compatible grammar must have the verbs move. But because finite verbs both follow and precede the adverbials such movements must end up in distinct head positions, one of which is C. Strictly speaking, (12w) is not unambiguous evidence for either [+VC] or [+TC] because either one can in principle be chosen as a way of getting the verb in C. Still, either value is necessary for learning the verb raises to C, and (12w) forces the learner to end up sampling one of those values. At first glance, then, it might seem that a corpus containing input of type (12w) will ensure learning a target grammar, in which the verb raises to C. However, we had reason above when we considered the Stepwise theory to think that given its low frequency, input (12w) might not be sufficient to learn a target grammar, or at least to do so relatively quickly. What was crucial there was to consider the potential effect of the vast preponderance of the input, which is ambiguous. The same type of question applies here: given such a high percentage of the input corpus that is ambiguous for [±VC], how will the learner fare in acquiring a target grammar with verb movement to C?

In fact, analysis of the ambiguous input shows that such input largely favors learning a target [+VC] grammar under the Swooping theory. This is different from what we saw with Stepwise theory, where there was an even split between [+TC] and [−TC] grammars among input-compatible grammars for each ambiguous input type. In contrast, under the Swooping theory some input types significantly favor a [+VC] grammar. The reason for this is tied to the earlier observation that there are various target grammars that the learner could in principle acquire when learning Swedish. Thus, if the learner samples a [+VC] value that is compatible with a given a token of input, the learner can in principle sample any combination of [±TC] and [±VT] for the entire input corpus. However, if the learner samples a [−VC] value, the options for constructing an input-compatible grammar are considerably more limited (e.g. the learner might be required to end up sampling a [+VT] value in the case of OVS input). The balance of input-compatible grammars favors [+VC] over [−VC]. Thus, when the effects of different types of input favoring [±VC] are factored into consideration, a balance of over 17% of the input corpus favors [+VC] over [−VC] by a ratio of 2 to 1 among the set of input-compatible grammars. On average, then, we expect the learner to sample and reinforce a [+VC] value. Gould (2017) shows that in learning simulations, when ambiguous evidence favors a given parameter value to a comparable degree, the model is highly successful in consistently learning that value.

Reasonable expectations, then, for the Swooping theory are as follows. First, as the learner will be pushed strongly toward a [+VC] grammar from the beginning of the learning process, we are not likely to see the same degree of indeterminancy with respect to [±VC] that we are likely to see for [±TC] early on with the Stepwise theory. The expectation is to see less of a delay early on in learning a target grammar with respect to verb placement. Moreover, as the learner is on average pushed strongly toward a [+VC] grammar by much of the input, we do not expect to see the same degree of parameter mis-setting for verb placement under the Swooping theory. Thus under the Swooping theory fewer learners are expected to have parameter mis-setting with regard to V being in C (viz. a grammar with non-target [−VC] and [−VT]/[−TC] settings), which in turn also results in there being less of a delay in learning a target grammar. Because of this strong push toward [+VC], our expectation more generally is that it is unlikely for a Swooping learner to encounter much in the way of an obstacle in learning a target grammar for verb placement, and thus that it is unlikely to not learn such a target grammar relatively rapidly. This is not to rule out entirely the possibility of parameter mis-setting (nor the impossibility/possibility of parameter resetting of [+VC] or [+TC] along the lines of what was discussed for the Stepwise theory), as the input is highly ambiguous. Rather, we expect parameter mis-setting to be rather unlikely, given how much the input favors a [+VC] parameter setting. Thus the worst-case scenario of not being able to learn a target grammar within the time constraints of acquisition is expected to be much less likely with the Swooping theory.

Assuming all else to be equal across the learning simulations of the two hypothesis spaces in (11), then, we can focus on the following expected simulation results. As regards speed, we expect the average number of tokens required to reach a target grammar (which includes the total amount of input received throughout the course of learning if a target grammar is never learned) each time the model is run (called a run) to be greater under the Stepwise implementation than under the

Swooping one. Similarly, as regards learning trajectories, we expect the proportion of runs that involve parameter mis-setting to be greater under than Stepwise implementation that under the Swooping one. If these expectations are borne out, then according to the acquisition-based evaluation metric we have support for the Swooping theory, and an argument against the Stepwise theory.[21]

Finally, we can note that these expectations crucially rely on learning from ambiguous input. If this approach to the puzzle of competing theories is on the right track, it provides an argument in favor of language learning models that do learn from ambiguous input, and against those such as Sakas and Fodor (2001) that do not learn from ambiguous input.

4. Results and discussion

Results from the learning simulations confirm our expectations. Results here are from the first 20 runs of the model under the Stepwise theory, and the first 75 runs under the Swooping theory. The model was run fewer times under the Stepwise theory owing to the lengthy runtime under this approach. For the sake of concreteness, a parameter was said to be set when 90% or more of the probabilistic weights for that parameter were concentrated on one of its values. Let us first consider speed of acquisition before looking at developmental trajectories. In what follows, when I speak of a run learning a target grammar, I am referring to when it first reached a relatively strong cumulative set of weights (of a target grammar) for all parameters that have been set, after which the run was terminated. The difference in the number of tokens between when a run first learned a target grammar and when it learned stronger weights for parameter settings in that grammar was fairly small. I report results on this “stronger grammar” to illustrate a kind of best-case scenario for a particular run, but this choice does not substantively affect the discussion.

Figures 1 and 2 show the relative speed of the model under the two theories in

21 Note that the expectation that the Swooping theory fares better under the metric does not follow from the Swooping theory’s hypotheses simply being a proper superset of those of the Stepwise theory. Rather this expectation is based on the kind of analysis detailed in this section, which looks at relations between a hypothesis space and an input corpus.

arriving at a target grammar. The figures plot each run of the model, and each run is plotted along the x-axis with respect to the number of tokens it took to learn a target grammar. The y-axis then shows what percentage of the runs have been comparably successful in learning a target grammar given that number of tokens or less. Every time the model was run, it was able to learn a target grammar, but when we look at the plots in Figures 1 and 2, a sharp difference emerges. Whereas around 40% of runs have learned a target grammar within 5,000 tokens of input under the Stepwise theory in Figure 1, Nearly 99% of runs have done so with under half the number of tokens under the Swooping theory in Figure 2. We can attribute this effect to the more prevalent delays early in the learning process for the Stepwise theory, as per the discussion of learning expectations in Section 3.2. Moreover, once we pass the 5,000 token threshold under the Stepwise theory, it becomes noticeably more difficult for the model to learn a target grammar. After 20,000 tokens we are still waiting for around a quarter of the runs to learn a target grammar, and after 30,000 tokens there are still a couple runs that have not done so. It is only when we set a ceiling of 40,000 tokens that all runs have sufficient input to actually learn a target grammar. In contrast, only one run under the Stepwise implementation takes longer than the threshold of 2,500 tokens, and it does not need more than 25,000 tokens of input.

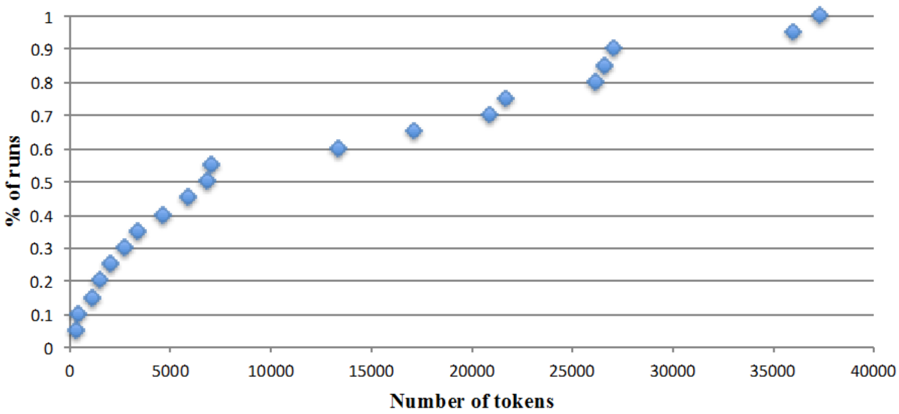


Figure 1. Stepwise theory: % of runs to have strongly learned a target grammar per number of tokens

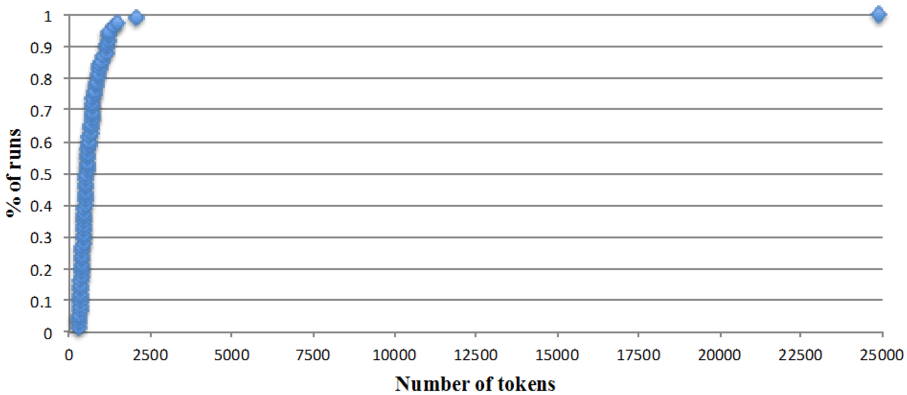


Figure 2. Swooping theory: % of runs to have strongly learned a target grammar per number of tokens

The mean and sample standard deviation of the x-axis values of the two distributions of runs further illustrate how the Swooping implementation learns a target grammar relatively rapidly (as expected) and is the faster of the two implementations. Under the Stepwise theory, the average is 13,101 tokens with a standard deviation of 12,434; indeed close to half the runs exceed this average. In contrast, under the Swooping theory the average is 980 tokens with a standard deviation of 2,820 (the size of which is largely owing to the conspicuous outlier shown in Figure 2). This is a striking illustration of a substantial difference in average speed of learning between the two theories and is a strong confirmation of the expectation that the Swooping theory would learn a target grammar faster. As far as speed is concerned, then, from a learnability perspective we have strong evidence to support the claim that as compared to the Swooping theory, the Stepwise theory is much less likely to result in learning a target grammar because of how long it takes.[22] This evidence thus favors the Swooping theory, but before we can apply the evaluation metric in favor of a Swooping approach, we must first consider accuracy of acquisition with respect to developmental trajectories.

Looking at the learning paths of individual runs, the only parameters that ever

22 Note that the average run-times of the model's simulations in Gould (2017) do not exceed that of the Swooping theory by much more than a thousand tokens, and some of the simulations reported there have shorter average run-times. Thus the Swooping results here are in line with earlier results from the same learning model.

get mis-set are the T-to-C and V-to-C movement parameters. Under the Stepwise theory, a little over half the runs involve a non-target [−TC] parameter setting (14a). Indeed, nearly all these runs exceed the 5,000 token threshold in Figure 1, pointing to a tight link between the slow speed of acquisition under the Stepwise theory and parameter mis-setting. This matches the expectation from Section 3.2 that parameter mis-setting can have a delaying effect on learning a target grammar. This is reinforced when we consider the Swooping theory. As there are multiple target grammars under the Swooping approach, we can say there is parameter mis-setting with regard to V being in C only when we have non-target settings of [−VT]/[−TC] and [−VC] simultaneously. Thus happens in only one run (14b), and this is the one run plotted by the conspicuous outlier in Figure 2.

(14) Proportion of runs with parameter mis-setting across simulations
  a. Stepwise simulations: 11/20 = 55%
  b. Swooping simulations: 1/75 = 1.33%

What we see is a sharp contrast in the developmental trajectories of the two approaches. The Swooping theory results in a much smaller proportion of runs that involve parameter mis-setting, and this again confirms our expectation. Although robust within-speaker errors have been reported for moving V to C in child Swedish (cf. Section 1), this has not been reported as a widespread, robust phenomenon across speakers. Thus the Swooping results more accurately match what is known about verb placement in child Swedish.

I now apply the evaluation metric to the competing theories. Compared to the Stepwise theory, the Swooping theory on average learns a target grammar faster and has a more accurate acquisition trajectory. According to the evaluation metric, this constitutes evidence in favor of the more complex Swooping theory, and against the simpler Stepwise theory. Based on this, then, the Swooping theory is the preferred theory for verb movement in Swedish. As for the debate in the literature over verb movement in Swedish, this supports an approach that involves a parameter for V-to-C movement.

5. Conclusion

In this paper I developed a novel acquisition-based evaluation metric for competing syntactic theories that draws on two kinds of linguistic data from modeling simulations: speed of learning and accuracy of learning trajectories. This metric is empirically grounded as it attempts to address core learnability and acquisition concerns. I showed how this metric can be applied in a case study of verb movement in Swedish to argue for a theory of movement that supports an approach where there is a parameter of V-to-C movement. A promising line of future inquiry thus lies in applying this evaluation metric to shed light on further theoretical questions in areas where traditional syntactic description has proven insufficient.

From an analytical perspective, the results favoring a particular theory were crucially related to the nature of the learner's hypothesis space (i.e. a complex one in which there are multiple potential target grammars), as well as to how the learner can use these hypotheses to learn from ambiguous evidence. Thus in addition to demonstrating a new and helpful application of ambiguous evidence and providing a further argument for its inclusion in a model of what a child can learn from, there is a broader implication we can draw from the analysis and approach adopted here. When it comes to proposing and designing theories, insights from modeling can be helpful to the theoretician. In particular, the theoretician can consider how ambiguous evidence might be used as a tool to acquire a proposed target grammar, as well as how designing a more complex theory might be advantageous to the learner. These insights are not, I believe, widely incorporated into theoretical work, but as we have seen, they have the potential to stand us in good stead in the development of theories.

References

Ackema, Peter and Amela Čamdžić. 2003. LF complex predicate formation: The case of participle fronting in Serbo-Croation. *UCL Working Papers in Linguistics* 15: 131-175.

Bach, Emmon. 1962. The order of elements in a transformational grammar of German. *Language* 38(3): 263-269.

Baptista, Marlyse. 2000. Verb movement in four creole languages: A comparative analysis. In John D. McWhorter (ed.), *Language change and language contact in pidgins and creoles*, 1-33. Amsterdam: John Benjamins Publishing Company.

Bobaljik, Jonathan David. 2000. The Rich Agreement Hypothesis in review. Ms., McGill University.

Brandtler, Johan. 2012. *The evaluability hypothesis: The syntax, semantics, and pragmatics of polarity item licensing*. Amsterdam: John Benjamins Publishing Company.

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

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

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

Chomsky, Noam. 1998. Minimalist inquiries: The framework. *MIT Occasional Papers in Linguistics* 15. Cambridge, MA: MITWPL.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos Otero, and Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory*, 133-166. Cambridge, MA: MIT Press.

den Besten, Hans. 1977. On the interaction of root transformations and lexical deletive verbs. Ms., Massachusetts Institute of Technology and Universiteit van Amsterdam. Appeared in den Besten (1989).

den Besten, Hans. 1989. *Studies in West Germanic syntax*. Amsterdam: Rodopi.

Diercks, Michael. 2013. Indirect agree in Lubukusu complementizer agreement. *Natural Language and Linguistic Theory*. 31(2): 357-407.

Engdahl, Elisabet. 1980. *The syntax and semantics of questions in Swedish*. PhD Dissertation, University of Massachusetts – Amherst.

Engdahl, Elisabet, Maia Andréasson, and Kersti Börjars. 2004. Word order in the Swedish midfield - an OT approach. In Fred Karlsson (ed.), *Proceedings of the 20th Scandinavian Conference of Linguistics*, 1-13. University of Helsinki, Dept. of General Linguistics.

Fodor, Janet Dean. 1998 Unambiguous triggers. *Linguistic Inquiry* 29(1): 1-36.

Fox, Danny and David Pesetsky. 2005. Cyclic Linearization of syntactic structure. *Theoretical Linguistics* 31(1): 1-45.

Gould, Isaac. 2017. *Choosing a grammar: Learning paths and ambiguous evidence in the acquisition of syntax*. Amsterdam: John Benjamins Publishing Company.

Haegeman, Liliane and Marjo van Koppen. 2012. Complementizer agreement and the relation between C^0 and T^0. *Linguistic Inquiry* 43(3): 441-454.

Heycock, Caroline and John D. Sundquist. 2017. Don't rush to rehabilitate: A remark on Koeneman and Zeijlstra 2014. *Linguistic Inquiry* 48(1): 173-179.

Holmberg, Anders and Christer Platzack. 1995. *The role of inflection in Scandinavian syntax*. Oxford: Oxford University Press.

Holmes, Philip and Ian Hinchcliffe. 2013. *Swedish: A comprehensive grammar*. New York: Routledge.

Hörberg, Thomas. 2016. *Probabilistic and prominence-driven incremental argument interpretation in Swedish*. PhD Dissertation, Stockholm University.

Klingler, Thomas A. 2003. *If I could turn my tongue like that*: *The creole language of Pointe Coupee Parish, Louisiana*. Baton Rouge: Louisiana State University Press.

Koeneman, Olaf and Hedde Zeijlstra. 2014. The Rich Agreement Hypothesis rehabilitated. *Linguistic Inquiry* 45(4): 571-615.

Koster, Jan. 1975. Dutch as an SOV language. *Linguistic Analysis* 1: 111-136.

Larsson, Ida. 2009. *Participles in time: The development of the perfect tense in Swedish*. PhD Dissertation, The University of Gothenburg.

Lee, Kwangho. 2009. Successive cyclic movement in English and Scandinavian and order preservation. *Linguistic Research* 26(2): 15-30.

Lidz, Jeffrey and Rachel Dudley. 2017. Evaluation measures and the study of language acquisition. *Language Acquisition* 24(2): 81-84.

Lightfoot, David. 1991. *How to set parameters: Arguments from language change*. Cambridge, MA: MIT Press.

Lightfoot, David. 1995. Why UG needs a learning theory: Triggering verb movement. In Adrian Battye and Ian Roberts (eds.), *Clause structure and language change*, 31-52. Oxford: Oxford University Press.

Lightfoot, David. 1997. Shifting triggers and diachronic reanalysis. In Ans van Kemenade and Nigel Vincent (eds.), *Parameters of morphosyntactic change*, 253-272. Cambridge: Cambridge University Press.

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*, 253-270. San Diego: Academic Press.

Pearl, Lisa. 2007. *Necessary bias in natural language learning*. PhD Dissertation, The University of Maryland.

Pearl, Lisa. 2017. Evaluation, use, and refinement of knowledge representations through acquisition modeling. *Language Acquisition* 24(2): 126-147.

Pearl, Lisa and Jeffrey Lidz. 2013. Parameters in language acquisition. In Cedric Boeckx and Kleanthes K. Grohmann (eds.), *The Cambridge handbook of biolinguistics*, 129-159. Cambridge: Cambridge University Press.

Pintzuk, Susan. 2002. Verb-object order in Old English: Variation as grammatical competition. In David Lightfoot (ed), *Syntactic effects of morphological change*, 276–300. Oxford: Oxford University Press.

Platzack, Christer. 1986a. Comp, Infl, and Germanic word order. In Lars Hellan and Kristi Koch (eds.), *Topics in Scandinavian syntax*, 185-234. Dordrecht: Reidel.

Platzack, Christer. 1986b. The position of the finite verb in Swedish. In Hubert Haider and Martin Prinzhorn (eds.), *Verb second phenomena in the Germanic languages*, 27-48. Dordrecht: Foris Publications.

Richards, Norvin. 2016. *Contiguity Theory*. Cambridge, MA: MIT Press.

Safir, Ken. 1987. Comments on Wexler and Manzini. In Thomas Roeper and Edwin Williams (eds.), *Parameter setting*, 77-90. Dordrecht: Reidel.

Sakas, William Gregory and Janet Dean Fodor. 2001. The Structural Triggers Learner. In Stefano Bartolo (ed.), *Language acquisition and learnability*, 172-233. Cambridge: Cambridge University Press.

Smith, Neil and Ann Law. 2009. On parametric (and non-parametric) variation. *Biolinguistics* 3(4): 332-343.

Solomonoff, Ray J. 1964a. A formal theory of inductive inference, Part 1. *Information and Control* 7(1): 1-22.

Solomonoff, Ray J. 1964b. A formal theory of inductive inference, Part 2. *Information and Control* 7(2): 224-254.

Svenonius, Peter. 2002. Introduction. In Peter Svenonius (ed.), *Subjects, expletives, and the EPP*, 3-27. Oxford: Oxford University Press.

Thunberg, Karin and Johan Sundelöf. 2017. *Palliations praktikan: Del 2–reflektionsmaterial*. Riga: Dardedze Holografija.

Travis, Lisa deMena. 1984. *Parameters and effects of word order variation*. PhD Dissertation, Massachusetts Institute of Technology.

Viberg, Åke, Kerstin Ballardini, and Sune Stjärnlöf. 1991. *Swedish: Essentials of grammar*. Chicago: Passport Books.

Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages*. Oxford: Oxford University Press.

Waldmann, Christian. 2011. Moving in small steps towards verb second: A case study. *Nordic Journal of Linguistics* 34(3): 331-359.

Waldmann, Christian. 2014. The acquisition of Neg-V and V-Neg order in embedded clauses in Swedish: A microparametric approach. *Language Acquisition* 21(1): 45-71.

Westergaard, Marit. 2006. Triggering V2: The amount of input needed for parameter setting in a split-CP model. In Adriana Belletti, Elisa Bennati, Cristiano Chesi, Elisa DiDomenico, and Ida Ferrari (eds.), *Language acquisition and development: Proceedings of GALA 2005*, 564-577. Newcastle upon Tyne: Cambridge Scholars Publishing.

Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang, Charles. 2017. Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition* 24(2): 100-125.

Zwart, Jan-Wouter. 1997. *The morphosyntax of verb movement: A Minimalist approach to Dutch syntax*. Dordrecht: Kluwer.

Appendix

This appendix provides illustrative Swedish examples of the input types in (12). The numbering in (12') corresponds to the numbering in (12).

(12') a. Isen smälte. (Larsson 2009:33)
the.ice melted

b. Han köpte boken. (Holmberg and Platzack 1995:7)
he bought the.book

c. Frida har sevit. (Larsson 2009:111)
Frida has slept

d. Han har köpte boken. (Holmberg and Platzack 1995:7)
he has bought the.book

e. Maja undrar om Sven kommer. (Engdahl 1980:111)
Maja wonders if Sven comes

f. Jag undrar om Paul spelar bas. (Brandtler 2012:94)
I wonder if Paul plays bass

g. Jag undrar om Elsa har ringt. (Kim Ahlström, p.c.)
I wonder if Elsa has called

h. Jag undrar om du har sett Maria. (Brandtler 2012:30)
I wonder if you have seen Maria

i. Nu sticker vi. (Viberg *et al.* 1991:128)
now push.off we

j. I dag lässer han svenska. (Holmes and Hinchliffe 2013:550)
in day reads he Swedish

k. På bara två minuter var kitten klappad. (Larsson 2009:220)
in only two minutes was the.cat petted

l. Troligen hade John köpt boken. (Platzack 1986a:185)
probably had John bought the.book

m. De levande fångade jag. (Hörberg 2016:140)
the living caught I

n. Hård hud lär Bill Clinton behöva. (Hörberg 2016:92)
hard skin be.said Bill Clinton need

o. Arbetar Elsa? (Viberg *et al.* 1991:24)
works Elsa

p. Köpte John boken? (Platzack 1986a:186)
bought John the.book

q. Har någon ringt? (Viberg *et al.* 1991:101)
has anyone called

r. Hade John köpt boken? (Platzack 1986a:186)
had John bought the.book

s. Olle cyklar inte. (Viberg *et al.* 1991:23)
Olle cycles not

t. John köpte inte boken. (Holmberg and Platzack 1995:12)
John bought not the.book

u. Deltagarna har redan anlänt. (Larsson 2009:25)
the.participants have already arrived

v. Erik hade verkligen köpt boken. (= (1))
Erik had really bought the.book

w. En av dem undrar också (Thunberg and Sundelöf 2017:18)
one of them wonders also
om hon inte ska få lite dropp[.]
if she not should get little drip

x. Jag frågade (Platzack 1986b:28)
I asked
om Erik verkligen hade skrivit boken.
if Erik really had written the.book

Isaac Gould
Department of English Language and Literature
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea
E-mail: gould@ewha.ac.kr

Received: 2017. 10. 17.
Revised: 2017. 12. 26.
Accepted: 2017. 12. 26.