How Nature Meets Nurture: Universal Grammar and Statistical Learning

Jeffrey Lidz*
Department of Linguistics
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742
jlidz@umd.edu

Annie Gagliardi
Department of Linguistics
Boylston Hall, 3rd floor
Harvard University
Cambridge, MA 02138
gagliardi@fas.harvard.edu

* Corresponding Author

Article Table of Contents

How Nature Meets Nurture: Universal Grammar and Statistical Learning

Abstract

Evidence of children's sensitivity to statistical features of their input in language acquisition is often used to argue against learning mechanisms driven by innate knowledge. At the same time, evidence of children acquiring knowledge that is richer than the input supports arguments in favor of such mechanisms. This tension can be resolved by separating the inferential and deductive components of the language learning mechanism. Universal Grammar provides representations that support deductions about sentences that fall outside of experience. In addition, these representations define the evidence that learners use to infer a particular grammar. The input is compared against the expected evidence to drive statistical inference. In support of this model, we review evidence of (a) children's sensitivity to the environment, (b) mismatches between input and intake, (c) the need for learning mechanisms beyond innate representations, and (d) the deductive consequences of children's acquired syntactic representations.

Keywords: Language Acquisition, Syntax, Statistical Inference, Input, Intake

1. Introduction

The theoretical frameworks used to approach the problem of language acquisition have traditionally taken two forms. In the input-driven tradition, learning is essentially a form of memory. Learners track patterns of co-occurrence in the environment and store them in summarized format. The stored

summary representation of experience allows for the production and comprehension of sentences beyond those that have been experienced (Elman et al 1996; Tomasello 2000). While researchers vary in whether they see the key to generalization as essentially statistical (e.g Aslin & Newport 2012, Elman et al 1996, Thelen & Smith 2006) or based in extralinguistic conceptual cognition (Goldberg 2007, Lieven & Tomasello 2008), these approaches share the perspective that generalizing beyond the input is driven by properties of the learner that are not fundamentally linguistic in nature, but rather rely on cognitive capacities that are applicable across domains and form the basis for understanding learning and social interactions more generally.

The alternative, knowledge-driven, tradition views learning as inference (Chomsky 1965, Pinker 1979, Gleitman 1990, Lightfoot 1991, Crain 1991). The learner, on this view, is endowed with a system of knowledge, typically called Universal Grammar, that specifies a space of possible grammars. Experience provides information that allows for the selection of features from that space in the acquisition of a particular grammar. Domain specific representations provide the foundation for generalization beyond experience. Experience provides the information relevant to the identification of those representations.

These two perspectives differ in two fundamental respects. First, they differ in how they see linguistic representations developing. On the input-driven view, abstract linguistic representations are arrived at by a process of generalization across specific cases. Abstract representations are therefore the output of a learning process that goes from specific to general (Tomasello 2000, Goldberg

2007). On the knowledge-driven view, abstract linguistic representations are part of the inherent linguistic capacity of the learner. Learners do not need to arrive at abstract representations via a process of generalization, but rather use their input to identify the realization of those abstract representations in the surface form of the exposure language (Chomsky 1965, Pinker 1989, Viau & Lidz 2011).

Second, the two views differ with respect to the role that input plays in learning. On the input-driven view, what is learned is a recapitulation of the inputs to the learner (Elman et al 1996). The acquired representations are a compressed memory representation of the regularities found in the input. Thus, new inputs evoke past experiences via some metric of similarity. New sentences are possible to the extent that they are similar to past experiences. On the knowledge-driven view, the learner searches the input for cues to help choose an abstract representation. The abstract representations themselves are innate and define the cues that learners use to identify them (Fodor 1998, Lightfoot 1999). Once those representations are identified, what the learner carries forward is not a recapitulation of the input. Rather, it is a representation that may bear no obvious relation to the input that triggered it. This representation thus allows for specific generalizations that may not be at all similar to the experience of the learner (Chomsky 1981, Snyder 2007).

Where the two approaches do *not* differ is in acknowledging that input plays a significant role in language acquisition. In both theories, acquisition of a particular grammar is guided to a large extent by experience. Where we expect differences will be in the relation between the features of experience and the

acquired linguistic knowledge.

In what follows, we present a model that incorporates key insights from both traditions, highlighting the link between input and outcomes while maintaining the essential insights of the knowledge-driven perspective on language acquisition. This model separates language acquisition into three components. The intake component identifies the features of the information that learners are sensitive to. The UG component identifies the class of representations that shape the nature of human grammatical systems. And the inference component identifies how learners combine the intake with UG in selecting a particular grammar.

After briefly describing this model, we review data that shows the very powerful role that input plays in shaping language outcomes. While this data is often used to underwrite input-driven theories (e.g., Huttenlocher et al 2010, Goldberg 2007), once we understand the components of a UG-based knowledge-driven learning mechanism, we can see why they are fully compatible with knowledge-driven approaches. Next, we show that the statistical sensitivities of the learner are sometimes distinct from ideal-observer measures of informativity, and that these differences may be revealing about the role learners play in selecting relevant input to drive learning. Then, we show that the presence of UG does not remove the need for a learning theory that explains the relation between input and the acquired grammar. Finally, we show that the dissimilarity between the input to the learner and the acquired representations support a UG-based knowledge-driven approach to language acquisition. In the end, we believe that deconstructing a language acquisition device into separate

components for input sensitivity, rational statistical inference and a rich space of representations, can explain many key results from both traditions of research in language acquisition.

2. Inside the Language Acquisition Device

2.1 The poverty of the stimulus and the motivations for Universal Grammar

Language learners acquire knowledge that falls outside of what they experience. This knowledge is evident in the capacity to produce and understand novel sentences, to distinguish possible from impossible sentences and to distinguish possible from impossible interpretations of novel sentences (Chomsky 1959, 1965). Consider, for example, the interpretation of the reflexive pronouns in (1).

(1) a. Sally knew that Jane put a picture of herself on the wall

b. Sally knew that Jane was very proud of herself

c. Sally knew which picture of herself Jane put on the wall

d. Sally knew how proud of herself Jane was

In (1a), *herself* is contained in an argument Noun Phrase and must be interpreted as referring to Jane and not to Sally. Similarly, in (1b), where *herself* is contained in a predicate adjective, only Jane is a possible antecedent. When *herself* is contained in a fronted argument wh-phrase, as in (1c), its interpretation is less restricted, allowing either Sally or Jane as antecedents. Finally, in (1d), *herself* is contained in a fronted predicate wh-phrase and is again restricted to taking only Jane as antecedent (Huang 1993, Heycock 1995, Barss 2001). In sum, we see an interaction between the phrase containing the reflexive (argument vs. predicate) and the position of that phrase (fronted vs. in-situ) whereby the

interpretation is less restricted only in the fronted argument environment (see Leddon & Lidz 2006 for experimental evidence confirming these facts in adults). Where does the knowledge of this interaction come from?

Leddon and Lidz (2006) examined 10,000 wh-phrases in child-directed speech in the CHILDES database (MacWhinney 2000) and found none containing a reflexive pronoun. Consequently, they concluded that whatever English speakers come to know about the possible interpretation of reflexive pronouns in embedded questions like (1c-d), it must come from a projection beyond their experience. Moreover, there are many ways that one could imagine projecting beyond experience in these cases. It might have turned out that both (1c) and (1d) were ambiguous or that neither was. Leddon and Lidz (2005) go on to show experimentally that by age 4, knowledge of this asymmetry is in place. The uniform acquisition of this pattern, despite the lack of directly relevant evidence, calls for explanation.

The poverty of the stimulus resides in the mismatch between experience and the acquired grammar. The argument is an invitation for future research aimed at explaining that mismatch. The typical response to this invitation is to define the innate representations from which the observed interaction could follow. For example, Huang (1993) relates the predicate-argument asymmetry in (1c-d) to the nature of the subject-predicate relation, claiming that predicates carry information about their subjects with them in the syntax. Other responses to this invitation might focus less on the innateness of the representations, but on general learning mechanisms that could give rise to the relevant interactions,

though we are aware of no proposals from outside the knowledge-driven tradition.

Universal Grammar represents one possible response to the invitation from the poverty of the stimulus, linking diverse sets of facts to a small set of highly abstract representations. These abstract representations drive the language learner's capacity to project beyond experience in highly specific ways. The emphasis of most work in generative linguistics has been the specification of these representations, with less focus on how learners use their experience. But this emphasis should not be confused with a claim that experience is irrelevant to language acquisition.

2.2 Beyond a specification of possible grammars

Universal Grammar is thus taken to be one component of a knowledge-driven learning mechanism. It defines the character of the acquired representations, which in turn allows the learner to have knowledge of the structure and interpretation of sentences that fall outside of their experience. Importantly, UG must also must be embedded in an architecture that allows learners to extract information from the input. This information is used to identify which of the resources defined by UG is applicable to any given sentence or sentence-type. Such a model is illustrated in Figure 1.

FIGURE 1 ABOUT HERE

This model of the language acquisition device has several components, which we explore in more detail below. The dashed line represents the division between what happens in the external world and what happens internal to the

child's learning mechanism. The input feeds into a perceptual encoding mechanism which builds an intake representation. This Perceptual Intake is informed by the child's current grammar along with the linguistic and extralinguistic information processing mechanisms through which a representation from that grammar can be constructed (Omaki & Lidz 2014). To the extent that learners are sensitive to statistical-distributional features of the input, as discussed below, that sensitivity will be reflected in the perceptual intake representations.

The developing grammar contains just those features for which a mapping has been established between abstract representations and surface form. A learner's perceptual intake representation does not contain all of the information that an adult represents for the same sentence. If it did, then there would be nothing to learn (Valian 1990, Fodor 1998). What it does contain, however, is information relevant to making inferences about the features of the grammar that produced that sentence. This perceptual intake feeds forward into the inference engine.

The inference engine is based on the idea that every possible grammar defined by UG makes predictions about what the learner should expect to find in the environment (Gleitman 1990, Lightfoot 1991, Fodor 1998, Tenenbaum & Griffiths 2001, Yang 2002, Regier & Gahl 2003, Pearl & Lidz 2009). The arrow from UG to acquisitional intake represents that predictive process. The acquisitional intake compares those predicted features against the perceptual intake representation, driving an inference about the grammatical features

responsible for the sentence under consideration. This inference updates the grammatical knowledge, which is added into the developing grammar. This updated grammar supports linguistic behavior. For the purposes of learning, the updated grammar feeds the perceptual processes through which subsequent sentences are analyzed.

In the next sections, we discuss evidence for each of these components of the learning mechanism, demonstrating (a) that language learners are sensitive to statistical features of their environment, but that these statistical sensitivites do not always match objective measures of informativity, (b) that UG does not by itself define a learning mechanism, and (c) that some of the statistical sensitivities of the learner are defined by UG, allowing for the inferences from which learners build a grammar of the target language, licensing knowledge about phenomena that fall outside of experience.

3. The sensitivities of the learner

3.1 Individual Differences in Input and Outcomes

A growing body of research has identified substantial individual differences in the quality and quantity of caregiver speech and corresponding individual differences in children's speech (Huttenlocher et al 1991, Hart & Risley 1995, Hoff 2003, Rowe 2012). Hart and Risley (1995) examined the in-home conversations in families with children under 2 years of age. They found that in an average week, children of high (professional) socio-economic status (SES) parents hear roughly 215,000 words, whereas children of middle (working class) SES parents hear 125,000 words and children of low SES (public assistance) parents hear only

62,000 words. The higher SES children heard not only more total words than the lower SES children, but also a greater variety of words. Correspondingly, SES-related differences in vocabulary size were evident at the earliest measurements and grew over time. High-SES 3-year olds had an average cumulative vocabulary of about 1000 words, whereas Low-SES 3-year olds had an average cumulative vocabulary of about 500 words. Moreover, Hart and Risley (1995) found that although SES predicted both the parental input and children's vocabulary growth, actual measures of input were stronger predictors of child outcome than SES.

Similarly, Hoff (2003) examined vocabulary growth as a function of maternal speech characteristics in High- and Mid-SES children, finding that SES-related differences in children's vocabulary growth was fully explained by variability in parents' speech. The diversity of word types and the mean length of utterance of the parents accounted for the variability in the rate of vocabulary growth in the children. (see also Huttenlocher et al 1991, Naigles & Hoff-Ginsburg 1998, Hoff and Naigles 2002)

While these studies clearly demonstrate a role for input quantity in vocabulary development, quantity is also not the whole story. Rowe 2013 shows that (controlling for differences in SES) the quantity of speech is the strongest predictor of vocabulary growth in the 2nd year of life, but that the diversity of vocabulary used by parents predicts vocabulary growth in the 3rd year of life, and that the complexities associated with narratives and decontextualized speech predict vocabulary growth in the 4th year of life. These findings, in particular,

suggest that input has different effects on language development as a function of what the child already knows. Consequently, models of language development must be careful to recognize that children extract different information from the input at different points of development, highlighting the variable nature of perceptual intake across developmental time (Gagliardi & Lidz 2014).

Relatedly, Weisleder and Fernald (2013) show that variability in child-directed speech, but not overheard speech, is a significant predictor of vocabulary growth in low-SES Spanish speaking children in the US (see also Schneidman & Goldin-Meadow 2012, Schneidman et al. 2013). However, they also show that the effects of input are mediated by differences in language processing. Differences in input have an impact on children's processing speed, which in turn contributes to the rate of vocabulary growth. This result suggests that the information processing mechanisms through which children understand their input play a critical role in distinguishing the input from the usable input (Omaki & Lidz 2014).

In the domain of syntax, we find similar effects of input on outcomes. Children from higher SES backgrounds produce more complex utterances in spontaneous speech as toddlers (Arriaga et al., 1998) and at age 5 (Snow, 1999), and perform significantly better on measures of productive and receptive syntax at age 6 (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). Moreover, variation in the syntactic complexity of maternal speech substantially explains SES-related differences in the syntactic complexity of 5-year-old children's speech (Huttenlocher et al. 2002). Huttenlocher et al (2010) found that higher

diversity in caregiver speech predicted several properties of their children's speech, including lexical diversity, diversity within constituents and diversity of clause types.

This research establishes a link between specific characteristics of the speech that children hear and the speech that they come to produce. At the same time, however, effects of input seem to be related more to the frequency with which complex structures are used than to the ability to produce them (Huttenlocher et al 2002), suggesting that perhaps the effect of input has more to do with the mechanisms through which knowledge of grammar is deployed, rather than with that knowledge *per se* (Weisleder & Fernald 2013).

It is important to recognize, however, that these effects are all equally compatible with input-driven and knowledge-driven accounts of language acquisition. For the input-driven theorist, what gets acquired is essentially a compressed memory representation of the exposure language. Consequently, properties of children's speech recapitulate properties of the input. For the knowledge-driven theorist, the learner's task is to use the input to identify the surface expression of certain abstract features supplied by UG. Consequently, recognition of those abstract features and the ability to process them efficiently are predicted by their presence in the input.

Indeed, while the relevant individual-difference work has not been done within the knowledge-based tradition, the idea that the input should predict the timeline of acquisition is well represented in that literature (Newport, Gleitman & Gleitman 1978, Valian & Casey 2003, Yang 2002, Legate & Yang 2007).

Legate and Yang (2007) examined the acquisition of tense in the productions of child learners of English, Spanish and French. Children learning these (and many other) languages have been shown to go through a stage during which they produce “root infinitives,” i.e., main clauses lacking in tense, unlike in the target language (Poeppel & Wexler 1993, Phillips 1995). However, children learning languages with richer verbal morphology tend to have shorter root infinitive stages than children learning languages with poorer verbal morphology (Phillips 1995, Guasti 2002).

Legate and Yang (2007) propose to link this variability with UG and the idea that different languages express the same syntactic function differently. In particular, some languages have a morphosyntactic expression of tense whereas others (e.g., Chinese, Japanese) do not. The learner’s task from this perspective is to determine which kind of grammar the language they are exposed to exhibits. Within the languages that express tense in the morphosyntax, there is variability in the degree to which tensed clauses can be distinguished from untensed clauses morphologically. Thus, it is predicted that the age of acquisition of a grammar with tense is a function of the degree to which the input unambiguously supports that kind of grammar over one without tense.

Legate and Yang show that roughly 60% of child-directed utterances in Spanish unambiguously support a grammar with tense, whereas 39% of child-directed utterances in French do, and only 6% of child-directed utterances in English do. Correspondingly, the root infinitive stage appears to end in Spanish children at around 2;0, in French children at around 2;8, and in English children

after 3;5. Root infinitives are, on this view, an expression of children maintaining the possibility that the target grammar is tenseless. The more the input is consistent with this possibility, the longer these sentences persist in children's speech.

In sum, the existence of correspondences between properties of the input and properties of the acquired grammar are perfectly consistent with both input-driven and knowledge-driven accounts of language acquisition. Both theories predict tight correlations between the exposure language and the timeline of acquisition, though for different reasons. Consequently, we should be careful not to take such data as prima-facie evidence for one approach over another. Rather, where the approaches differ concerns the way that learners project beyond their experience, as we will discuss below. More generally, one can view children's sensitivity to frequency and statistical-distributional properties of the input as an expression of the intake mechanisms that feed forward for inferences about grammatical structure.

3.2 Statistical Sensitivity in Artificial Language Paradigms

A large body of work explores infants' abilities to extract generalizations about the grammars of artificial languages based on statistical patterns in the data. In particular, learners have been shown to track statistical information in uncovering dependencies among segments, syllables, words and phrases.

Saffran, Aslin & Newport (1996) examined the use of adjacent dependencies in word segmentation by exposing infants to sequences of words from an artificial language made up of 3 syllable "words". While there were no

pauses between words in the speech stream, words could be segmented based on transitional probabilities between syllables, i.e, the probability of one syllable given the previous one. The language was designed such that transitional probabilities were high within words, but low across word boundaries. Eight-month-old infants exposed to these stimuli learned to segment words, evident in their differentiation of words from nonwords during a test phase.

While this ability is impressive, it appears to be somewhat limited to the extremely simplified 'toy' quality of the artificial language used. Johnson & Tyler (2001) carried out a similar experiment but with words that varied in syllable number. They found that infants were not able to perform word segmentation in this more realistic situation.

Not all added complexity thwarts infants' ability to extract statistical patterns, however. Gomez & Gerken (1999) showed infants' sensitivity to transitional probabilities between words at the sentence level. In their experiment, infants exposed to an artificial language were able to distinguish grammatical strings from ungrammatical ones after only two minutes of exposure.

Infants also appear able to go beyond tracking the adjacent dependencies recoverable from transitional probabilities, extracting generalizations about nonadjacent dependencies as well. Gomez (2002) shows that infants can learn dependencies between nonadjacent words, but only when dependencies between adjacent words are not detectable. Further probing this ability, Gomez & Maye (2005) found that including higher variability in the space between the elements in a nonadjacent dependency promotes better learning. Similarly, Lany

& Gomez (2008) showed that first acquiring adjacent dependencies makes acquiring nonadjacent dependencies possible.

This body of work highlights children's sensitivity to statistical patterns in their linguistic environment. However, because this work fails to address the question of how the relevant dependencies are represented and how learners might project beyond their input, it is silent with respect to distinguishing input-driven and knowledge-driven models.

Consider, for example, constituent structure in syntax and the data that learners use to acquire it. Constituent structure representations provide explanations for (at least) three kinds of facts. First, constituents provide the units of interpretation. Second, the fact that each constituent comes from a category of similar constituents (e.g., NP, VP, etc.) makes it such that a single constituent type may be used multiple times within a sentence:

(2) $[_{IP}\ [_{NP}$ the cat$]\ [_{VP}$ ate $[_{NP}$ the mouse$]]]$

Third, constituents provide the targets for grammatical operations such as movement and deletion:

(3) a. I miss [the mouse]$_i$ that the cat ate $\_\__i$.
b. The cat ate the mouse before the dog did $[_{VP}$ ~~eat the mouse~~].

Thompson and Newport (2007) make a powerful observation about phrase structure and its acquisition: because the rules of grammar that delete and rearrange constituents make reference to structure, these rules leave a statistical signature of the structure in the surface form of the language. The continued cooccurrence of certain categories and their consistent appearance and disappearance together ensures that the co-occurrence likelihood of

elements from within a constituent is higher than the co-occurrence likelihood of elements from across constituent boundaries.

Thompson and Newport (2007) show that adult learners can use this statistical footprint in assigning constituent structure to an artificial language with a flat constituent structure. This result was extended by Takahashi & Lidz (2008) & Takahashi (2009) who showed that this statistical footprint could be used by both adults and 18-month-old infants to acquire a grammar with hierarchically nested constituent structure (see also Morgan & Newport 1981, Saffran 2001). One probe of this knowledge came from participants' ability to distinguish moved constituents from moved non-constituents. Because only constituents move in natural language, evidence that learners learned which substrings of a sentence could enter into movement relations provided evidence of their knowledge of which substrings were constituents.

But still these observations do not illuminate the role that the statistical footprint plays in acquisition. Does statistical sensitivity feed a memory representation that supports the use of new sentences that are appropriately similar to those in the input, as under an input-driven theory of acquisition? Or does it feed an inferential process whereby a learner with expectations about how phrase structure is organized compares those expectations against the statistics to select a representation, as under a knowledge-driven theory?

Critically, the exposure provided to the learners in Takahashi's experiments included sentences containing movement. Although the particular sentences tested were novel, they exhibited structures that had been evident

during the initial exposure to the language. But where did the knowledge that only constituents can move come from? Did it come from the training, or did it come from expectations about the link between constituency and movement?

To ask this question, Takahashi (2009) created a corpus of sentences from an artificial language. This corpus exhibited a statistical signature of the constituent structure of the language, but provided no evidence of the possibility of movement rules. After exposure to this corpus, both adults and 18-month-old infants successfully distinguished sentences with moved constituents from those with moved nonconstituents, thus displaying knowledge of the constraint that only constituents can move, Given that the exposure did not include any evidence of movement, some of what participants acquired on the basis of statistical information was not itself reflected in the input statistics. This finding is consistent with a knowledge-based approach to learning because the constraint that only constituents can move was not reflected in the exposure language.

While there are many possible explanations for this effect, among the contenders is the idea that learners know antecedently that human languages exhibit movement relations and that these relations are restricted to applying to constituents. Thus, identifying a string as a constituent makes it a possible target of movement rules. And, for the purposes of this section, the availability of such an explanation illustrates the independence of statistical sensitivity and Universal Grammar.

Learning mechanisms based in statistical association as well as those based in inferential hypothesis testing predict learners to be sensitive to

distributional features in their environment (Frank & Tenenbaum 2011). Thus, the existence of statistical sensitivity plays no role in distinguishing these models (contra Saffran et al 1996, Huttenlocher 2010). Rather, in order to distinguish these approaches, we must be careful to distinguish statistical sensitivity from the consequences of that sensitivity. It is only through the consequences for sentences outside of the training data that we can draw conclusions about the role that the input plays in language acquisition (Frank, Mathis & Badecker 2012). When we see learners generalizing to structures that are sufficiently dissimilar from their input experience, we find clear candidates for the contribution of UG, while still recognizing the statistical sensitivities of language learners.

4. Distinguishing Input from Intake: Noun Classes in Tsez

The model in Figure 1 distinguishes the input from two levels of intake. The input is passed through the perceptual mechanisms of the learner in order to build a perceptual intake representation. This representation feeds into the inference engine, which selects specified features of that representation (the acquisitional intake) to derive conclusions about the grammatical representations underlying the target language. In this section, we present evidence that the information used to drive learning is only a subset of what the learner can perceive in the ambient language. Hence, we find a mismatch between objective measures of informativity in the input and the weighting assigned to different cues by children during the process of language acquisition (see also Lidz, Gleitman & Gleitman 2003, Hudson-Kam & Newport 2009).

The relevant phenomenon comes from the acquisition of noun classes in Tsez. Noun classes (grammatical gender) can be characterized in two ways: by distributional information *external* to the noun and by distributional information *internal* to the noun. Noun external distributional information is mainly expressed through other words (e.g., verbs, adjectives) that agree with an NP. The form of the agreement varies depending on the class of the head of the agreeing NP. Such information is highly regular, as it is this behavior that forms the basis for assigning nouns to classes, and consequently is highly informative. If a noun triggers a particular agreement pattern, this is diagnostic of the class of that noun.

Noun internal distributional information, on the other hand, consists of properties of the nouns themselves; semantic and phonological features that can (but do not necessarily) correlate with class. This information works probabilistically, as some features provide very good predictions of class while others may make weaker predictions or no predictions at all. Gagliardi & Lidz (2014) measured noun internal distributional information in Tsez by tagging all of the nouns in a corpus of Tsez child directed speech for potentially relevant semantic and morphophonological cues. They used decision tree modeling to determine which features were most predictive of each of the four noun classes (cf. Plaster, Polinsky & Harizanov, 2013). While the features vary in terms of how well they predict a class, there is at least one feature that can predict each class. In general, biological semantic features, which perfectly predict the class of a noun, make the strongest predictions about class. Other semantic features, such

as being clothing or made of paper, can also predict the class, but less reliably. Some phonological features, in particular the first and last segment, are also predictive of class, but to a lesser degree than the biological semantic features.

Because the noun internal distributional information was measured from a corpus of Tsez child directed speech, it serves as a good characterization of the information Tsez learners are exposed to. By comparing the reliability of an information source to the learners' sensitivity to this information in noun class acquisition, we can determine how learners encode their experience and hence how the inference engine uses this perceptual encoding to drive learning. To the extent that learners are biased to attend to certain information sources, we see the contribution of intake mechanisms. Further, because the information perceived at different stages of development changes, we can see the residue of this perceptual intake as a bias towards sensitivity to certain information sources over others. And, among the information that the learner can perceive, a bias in favor of one source over another may reflect the contribution of knowledge-driven acquisitional intake mechanisms.

Gagliardi and Lidz conducted a classification experiment to determine the degree to which Tsez speakers (and learners) are sensitive to noun internal distributional information. The words used for classification were either real words that had the predictive features discovered in the corpus (or certain combinations of the features) or nonce words invented to have these features (as well as a set of nonce words without any predictive features). Speakers were tested on nouns with each of the single features, as well as certain feature

combinations making different predictions about class (e.g., a biological semantic cue for class 2 but a phonological cue for class 4). Speakers (adults, older children (8-12 yrs), and younger children (4-7yrs)) were introduced to test nouns in the form of labeled pictures of objects, and sentences containing agreeing verbs were elicited. As noun class is evident in verbal agreement, speakers' classification of the test nouns was visible through such agreement.

For real words, all groups of speakers tended to correctly classify each word type. However, there were two exceptions to this pattern. First, words with weak (nonbiological) semantic features were less likely to be correctly classified by both groups of children. Second, when the real words' class conflicted with the class predicted by a phonological noun internal feature, children were more likely to assign the word to the incorrect class, relying inappropriately on the phonological cue.

In classifying nonce words, participants generally displayed appropriate sensitivity to the relevant features. However, this was not the case for all features. First, neither group of children used the weak semantic features at all. Adults were only mildly sensitive to these features, but more so than children. Second, children, especially in the younger group, were more likely to use phonological features instead of semantic features when the two made conflicting predictions, despite the higher informativity of the biological semantic features.

Overall, although speakers were sensitive to both semantic and phonological features that predict class, they used them out of proportion with their statistical reliability, even ignoring certain highly predictive semantic

features. The intuition that children's classification behavior does not match the predictions based on the statistical data available in the input is confirmed by the predictions of an optimal naïve Bayesian classifier (Gagliardi, Feldman & Lidz, 2012). This model predicts classification in line with the semantic features when both semantic and phonological features are present, and provides a representative estimate of adult behavior.

The preference for using phonological information over semantic information likely reflects perceptual intake in the initial stages of noun class learning. Phonological information can be tracked before the learner has assigned a meaning to a given word-form, but a learner can be sure of the presence of semantic information only after acquiring the meaning. These differences in how reliably a learner can perceive the different types of features may be reflected in the perceptual intake, causing the learner to more reliably represent phonological information than semantic when building a lexicon. Alternatively, the acquisitional intake may put greater weight on phonological information in forming noun classes, perhaps reflecting an innate bias to treat formal categories as formally (and not semantically) conditioned. This bias in the acquisitional intake would account for the learner's tendency to draw inferences from each feature type unequally.

The apparent blindness to nonbiological semantic features, despite their high reliability, may also stem from the acquisitional intake. Given that noun class systems nearly universally implicate semantic features such as animacy, humanness, or natural gender, it is likely that the learner has expectations about

the role of such features in structuring the noun lexicon. The other semantic features implicated in noun class systems vary greatly across languages and thus may be more difficult for learners to treat as relevant.

In summary, Tsez acquiring children are sensitive to predictive noun internal features out of proportion with their predictive reliability in the input. This finding highlights the need to distinguish both input from intake, and perceptual intake from acquisitional intake in our model of language acquisition.

5. Why UG is not sufficient to drive learning

It has been recognized at least since Chomsky 1965 that UG is only one piece of a language acquisition mechanism (Pinker 1979, Wexler & Culicover 1980, Pinker 1989, Gibson & Wexler 1991, Fodor 1998, Yang 2002, Sakas & Fodor 2012). In some sense this should be obvious because the elements of UG are abstract, but the form of a language is concrete. Thus, learners must discover a mapping between the abstract principles of grammar and their realization in the surface form of the particular language they are acquiring.

Suppose, for example, that UG defined an innate category of reflexive pronoun, paired with constraints on interpretation. These constraints would require that a reflexive pronoun must take an antecedent, that the antecedent must c-command it, and that there is some domain of locality for finding that antecedent, with room for cross-linguistic variation in defining that domain (Koster & Reuland 1991, Cole, Hermon & Huang 2001). Even with this innate structure, the learner would have to identify which forms in the language fall into that category, be able to identify structures over which c-command relations

obtain and to define alternative locality domains (Sutton, Fetters & Lidz 2012, Pearl & Sprouse 2013). This identification may require a complex set of statistical inferences based on partial information about intended interpretations and the syntactic environment of the pronoun. (Orita, McKeown, Feldman, Lidz & Boyd-Graber 2013).

Similarly, suppose that UG defines a narrow class of possible dependency relations, including for example, the relations (a) between a wh-phrase and the variable it binds, (b) between a quantifier and a pronoun it binds, (c) between an agreeing head and the trigger for agreement, or (d) between a head and its complement, each with different properties. While certain properties of these relations will be a consequence of the UG principles that define them, their expression in the surface form of the language will not be so defined. Consequently, the learner must still use surface-based evidence from the input to identify possible dependency relations and to determine which type of dependency must be invoked for representing any particular case (Gagliardi, Mease & Lidz 2014, Omaki, Orita & Lidz 2014).

An extreme example of the need for a learning mechanism that includes more than UG comes from the interpretation of quantificational sentences in Korean. Han, Lidz and Musolino (2007) examined the relative scope of a universal quantifier in object position in a clause containing negation, as in (4).

(4) a. Khwuki monste-ka motun khwuki-lul mek-ci ani ha-yess-ta
cookie monster-NOM every cookie-ACC eat-CI NEG do PST-DECL
'Cookie Monster didn't eat every cookie.' (long negation)

b. Khwuki monste-ka motun khwuki-lul an mek-ess-ta
cookie monster-NOM every cookie-ACC NEG eat-PST-DECL

'Cookie Monster didn't eat every cookie.' (short negation)

In principle, such sentences could be compatible with two interpretations. If the universal takes scope over negation, then this would mean that Cookie Monster ate none of the cookies. If negation takes scope over the universal, then it would mean that Cookie Monster ate fewer than all of the cookies. The latter interpretation is compatible with Cookie Monster eating some cookies, whereas the former is not.

Han, Lidz and Musolino relate the interpretations of these sentences to theories of verb placement in the clause, but what is important here is that interpretive judgments in the literature about Korean grammar vary considerably. To probe the source of this variability, Han and colleagues tested adult and 4-year-old speakers of Korean in order to determine which interpretations of such sentences were licensed.

Consistent with the variability in the literature, they found that people differed systematically in their judgments. Roughly half of both populations allowed only the interpretation where the universal takes scope over negation. And roughly half allowed the interpretation where negation takes scope over the universal. This was true for both long and short negation. Moreover, this variability was not simply due to the difficulty of computing interpretations; when the universal was in subject position, all speakers allowed only the interpretation where the universal takes wide scope. In addition, in other languages, judgments about these kinds of sentences are stable in both adults and children (Musolino, Crain & Thornton 2000, Lidz & Musolino 2002, Lidz & Musolino 2006). Finally

Han, Lidz and Musolino (2014) found that speakers were consistent in their judgments across multiple testing sessions, and that they were consistent in their scope assignments for both kinds of negation in Korean. This pattern supports the view that each speaker controls only one grammar, with the variability in scope judgments following from which grammar that is.

Han et al. argue that this variability results from a sparseness of relevant evidence. Sentences like (4) are extremely rare (Gennari & MacDonald 2005), and even if they did occur frequently, their intended interpretation in context is unlikely to be transparent without knowledge of the grammar. Thus, even if UG defines a narrow space of possible grammars, the input to the learner may still be insufficient for deciding which grammar is responsible for the target language. In this case, the absence of evidence leads learners of Korean to pick a grammar at random. In the vast majority of sentences that learners encounter, this choice has no consequences. But in the case of the relative scope of an object quantifier and negation, the effects of this choice can be seen.

To explore this possibility further, Han, Lidz and Musolino 2014 tested a group of children with their parents. They found again that roughly half of both populations allowed the universal to scope over negation. However, they found no correlation between parents and their children, consistent with the view that grammar selection in this case is not driven by the environment.

With respect to the model in Figure 1, the Korean scope data highlight the need for a mechanism distinct from UG that allows learners to select a grammar. In most cases, the selection of an alternative will be driven by the input, and

acquistion will be driven by the degree to which learners can identify evidence in favor of one grammar over another (Gibson & Wexler 1991, Frank & Kapur 1996, Yang 2002, Pearl & Weinberg 2007, Pearl 2011, Fodor & Sakas 2012). In the extreme case that there is no relevant evidence, this mechanism chooses a grammar at random from the set of structures licensed by UG. This result illustrates (a) the potential independence of the acquired grammar and the language of the environment and (b) the independence of UG from the grammar selection mechanism, since in this case UG specifies the options but fails to determine the acquired grammar.

6. On the dissimilarity of intake statistics and acquired knowledge

A critical feature of the knowledge-driven learning model illustrated in Figure 1 is that it separates (a) the ability to track distributional features in the input and use them to select a grammatical representation from (b) the deductive consequences of those representations. One part of the learning mechanism is inferential, using distributional evidence to license conclusions about the abstract representations underlying the language. The second part is deductive, yielding a wide range of empirical consequences from the acquired abstract representations. To the extent that the empirical consequences of the representations do not resemble properties of the input that gave rise to those representations, we have evidence in favor of this kind of model. In the current section, we examine such a case, capitalizing on structural features of ditransitive constructions crosslinguistically.

6.1 A poverty of the stimulus problem in Kannada

Kannada ditransitives exhibit a flexible word order, allowing either order of the dative and accusative arguments. In addition, Kannada also optionally displays a benefactive affix on the verb in ditransitives. Putting these two features together allows for four possible ditransitives, illustrated in (5).

(5) a. hari rashmi-ge pustaka-vannu kalis-id-a
hari rashmi-DAT book-ACC send-PST-3SM
'Hari sent the book to Rashmi.'

b. hari rashmi-ge pustaka-vannu kalis-i-**koTT-**a
hari rashmi-DAT book-ACC send-PP-**BEN.PST**-3SM
'Hari sent the book to Rashmi.'

c. hari pustaka-vannu rashmi-ge kalis-id-a
hari book-ACC rashmi-DAT send-PST-3SM
'Hari sent the book to Rashmi.'

d. hari pustaka-vannu rashmi-ge kalis-i-**koTT-**a
hari book-ACC rashmi-DAT send-PP-**BEN.PST**-3SM
'Hari sent the book to Rashmi.'

In addition, we find asymmetries with respect to binding across these constructions (Lidz & Williams 2006; Viau & Lidz 2011). In the benefactive construction, the dative argument can bind into the accusative argument independent of word order, (6a-b). However, in the nonbenefactive construction, the dative can bind into the accusative only when the dative comes first (6c-d).

(6) a. Q-DAT$_x$ ACC$_x$ V+BEN
Rashmi pratiyobba hudugan-ige avan-a kudure-yannu tan-du-**koTT**-aLu
Rashmi every boy-DAT 3SM-GEN horse-ACC return-PPL-**BEN.PST**-3SF
'Rashmi returned every boy his horse'

b. ACC$_x$ Q-DAT$_x$ V+BEN
Rashmi avan-a kudure-yannu pratiyobba hudugan-ige tan-du-**koTT**-aLu
Rashmi 3SM-GEN horse-ACC every boy-DAT return-PPL-**BEN.PST**-3SF
'Rashmi returned his horse to every boy'

c. Q-DAT$_x$ ACC$_x$ V

Rashmi pratiyobba hudugan-ige avan-a kudure-yannu tan-d-aLu
Rashmi every boy-DAT 3SM-GEN horse-ACC return-PST-3SF
'Rashmi returned every boy his horse'

d. *ACC_x Q-DAT_x V
*Rashmi avan-a kudure-yannu pratiyobba hudugan-ige tan-d-aLu
Rashmi 3sm-GEN horse-ACC every boy-DAT return-PST-3SF
'Rashmi returned his horse to every boy'

With respect to the accusative argument binding into the dative, the pattern is reversed. Here, in the benefactive construction the accusative can bind into the dative only when the accusative comes first (7a-b). But in the nonbenefactive construction the accusative can bind into the dative independent of word order (7c-d).

(7) a. *DAT_x Q-ACC_x BEN
*sampaadaka adar-a lekhan-ige pratiyondu lekhana-vannu kaLis-i-**koTT**-a
editor it-GEN author-DAT every article-ACC send-PP-**BEN.PST-**3SM
'The editor sent its author every article'

b. Q-ACC_x DAT_x BEN
sampaadaka pratiyondu lekhana-vannu adara lekhan-ige kaLis-i-**koTT**-a
editor every article-ACC it-GEN author-DAT send-PP-**BEN.PST**-3SM
'The editor sent every article to its author'

c. DATx Q-ACCx unaffixed
sampaadaka adar-a lekhan-ige pratiyondu lekhana-vannu kaLis-id-a
editor it-gen author-dat every article-acc send-pst-3sm
'The editor sent its author every article'

d. Q-ACCx DATx unaffixed
sampaadaka pratiyondu lekhana-vannu adar-a lekhan-ige kaLis-id-a
editor every article-acc it-gen author-dat send-pst-3sm
'The editor sent every article to its author'

In sum, we see the interaction of three factors: word order, morphology and the grammatical function of the quantifier. When the benefactive morpheme is present on the verb, the dative argument behaves as if it is syntactically

prominent for binding, hence indifferent to word order. But when the benefactive morpheme is absent, it is the accusative argument that behaves as if it is syntactically prominent for binding, hence indifferent to word order.

With respect to learning, these patterns are not reflected in the input. Viau and Lidz (2011) conducted two large-scale corpus analyses and observed that ditransitive sentences in which one internal argument is a quantifier and the other contains a pronoun that matches that quantifier in phi-features almost never occur. The few cases that do occur would not provide enough variability to license conclusions about which binding configurations are licensed across constructions. Consequently, the data from which children come to acquire these patterns must involve projections from other facts that are more readily available.

6.2 Comparative Syntax as a Window into the Contribution of UG

Harley 2002 (building on Freeze 1992, Kayne 1993) makes three important observations about ditransitive constructions cross-linguistically. First, ditransitives differ cross-linguistically in whether the theme or the goal behaves as though it is syntactically prominent for the purposes of binding. Many languages, like Kannada and English, exhibit both goal-prominent and theme-prominent ditransitives. But some, like Irish and Diné, exhibit only theme-prominent ditransitives. Second, goal-prominence is typically paired with a more restricted interpretation on the goal, such that it must be a possible possessor of the theme. For example, in both English double object constructions and Kannada benefactive ditransitives, the goal argument must be interpreted as a possible possessor of the theme argument, whereas such restrictions do not

apply to the theme prominent ditransitives in these languages (or in languages with only theme prominent ditransitives).

(8) a. * The editor sent New York the book

b. The editor sent the book to New York

(9) a. * hari bangluur-ige pustaka-vannu kalisi-**koTT**-a
hari banaglore-DAT book-ACC send-**BEN.PST**-3SM
'Hari sent Bangalore the book'

b. hari bangaluur-ige pustaka-vannu kalis-id-a
hari bangalore-DAT book-ACC send-PST-3SM
'Hari sent the book to Bangalore'

Finally, goal-prominence occurs only in languages that have possession constructions in which the possessor is syntactically higher than the possessed. For example, Irish possessives do not allow the possessor to bind a pronoun inside the possessed. And, in Irish ditransitives the goal cannot bind into the theme, illustrating the presence of only theme prominent structures. In contrast, Kannada possession constructions allow the possessor to bind into the possessee and that language also displays goal-prominent ditransitives independent of word order, as we have seen.

6.3 Two contributions of UG to learning

In light of these cross-linguistic patterns in ditransitives, Viau and Lidz (2011) identified two potential contributions for UG in the acquisition of the binding facts illustrated in (6-7). One contribution, as discovered by Harley (2002), is in defining the space of possible languages, linking the syntax of possession to the syntax of ditransitives. The second contribution is in allowing that syntax to define the acquisitional intake from which statistical inference can proceed.

Regarding the first contribution, UG makes a complex set of facts follow from a single representational parameter concerning the syntax of possession relations. If a language exhibits possession constructions in which the possessed is higher than the possessor, it can recruit that structure in certain ditransitives, treating the goal argument as a possessor and making it syntactically prominent. This syntactic prominence explains the binding asymmetries. Importantly, the level at which the cross-linguistic generalization applies is highly abstract and thus is not detectable directly in the surface form of the language.

Regarding the second contribution, UG helps to define the kind of information that children should use in determining whether a given ditransitive utilizes the goal-prominent or theme-prominent syntax. The surface realization of ditransitives varies considerably cross-linguistically. In English, the two kinds of ditransitives are distingished in word-order (Oehrle 1976). In Kannada, they are distinguished by an affix on the verb (Lidz & Williams 2006) but not by word order. In Spanish, they are distinguished by clitic doubling of the dative argument but not by word order (Uriagereka 1988, Bleam 2001). But given this divergent surface realization the mapping between the structure and surface form is opaque.

Viau and Lidz (2011) argued that matching the strings with their underlying structures can be achieved by tracking the kinds of NPs that occur as the dative argument in each surface form. Because the dative in a goal-prominent argument is restricted to being a possible possessor of the theme, the kinds of NPs that fill that role is expected to be more restricted. In particular, possessors tend to be

animate and so learners should expect a relatively high proportion of animates as the dative argument in a goal-prominent ditransitive. The perceptual intake must therefore consist of a representation of morphological variability, word-order variability and the grammatical functions of each argument. The acquisitional intake is a representation of the relative proportion of animate to inanimate datives for each of the morphological and word-order variants. If the learner sees a construction that is statistically biased towards animate goals, that skew in the distribution would support the inference that that construction involves the goal-prominent syntax.

Summarizing this section, we have identified two roles of UG in language acquisition: (1) explaining the specific ways that children project beyond their input; and (2) defining the acquisitional intake of the learner, making statistical-distributional evidence informative about grammatical structure. In the case of Kannada ditransitives, the former explains children's knowledge of binding patterns across novel sentence types while the latter explains how observations of the distribution of animate datives can be linked to those binding patterns.

7. Conclusions

In this paper, we have outlined a model of language acquisition with several important properties. The model integrates the statistical sensitivity of learners to a UG-based knowledge-driven approach to language learning. In addition, by recognizing a distinction between input and intake, we can understand why learners' sensitivities do not match up perfectly with observer-neutral measures of informativity. In some cases it is the learner's expectations about how

languages are structured that defines the information in the environment that drives learning (Lidz, Gleitman & Gleitman 2003, Viau & Lidz 2011).

We have also tried to emphasize that a theory of UG is not equivalent to a theory of acquisition. UG defines a space of possible representations that, when mapped to appropriate strings, license a rich set of conclusions about sentences that are highly dissimilar from those of experience. But this space of representations only sets the initial conditions for learning. Beyond that, we must also have mechanisms for mapping sentences onto those representations and for defining the environmental inputs that guide that mapping process.

We hope that the model described here will help dissolve the traditional nature-nurture dialectic that has polarized the cognitive science of language and to allow the field to bridge the contributions of the environment with those of the learner. Input undoubtedly plays a critical role in shaping language development. By the same token, learners are able to project far beyond their experience in acquiring a language. The perspective adopted here makes it possible to unify these observations in a single framework.

Literature Cited

Arriaga RK, Fenson L, Cronan T, Pethick SJ. 1998. Scores on the MacArthur Communicative Development Inventory of children from low- and middle-income families. *Applied Psycholinguistics* 19:209-25.

Aslin RN, Newport EL. 2012. Statistical learning: From learning items to generalizing rules. *Current Directions in Psychological Science 21*:170-76.

Barss A. 2001. Syntactic Reconstruction Effects. in M. Baltin & C. Collins (eds). *Handbook of Contemporary Syntactic Theory*. Cambridge, MA: Blackwell.

Bleam T. 2001. Properties of the double object constructions in Spanish. In R Nuñez-Cedeño, L López, and R Cameron (eds.), *A Romance Perspective on Language Knowledge and Use*. Amsterdam: John Benjamins.

Chomsky N. 1959. Review of B.F. Skinner's Verbal Behavior. *Language* 35:26-58.

Chomsky N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Cole P, Hermon G, Huang C-T J. 2001. *Long-distance Reflexives*. NY: Academic Press.

Crain S. 1991. Language Acquisition in the Absence of Experience. *Behavioral and Brain Sciences* 14:597-612.

Elman JL, Bates EA, Johnson M, Karmiloff-Smith A, Parisi D, Plunkett K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Fodor JD. 1998. Parsing to learn. *Journal of Psycholinguistic Research* 27:339-74.

Frank R., Kapur S. 1996. On the Use of Triggers in Parameter Setting. *Linguistic Inquiry* 27:623-60

Frank R, Mathis D, Badecker W. 2013. The Acquisition of Anaphora by Simple Recurrent Networks. *Language Acquisition* 20:181-227.

Frank MC, Tenenbaum JB. 2011. Three ideal observer models for rule learning in simple languages. *Cognition* 120:360-371.

Freeze R. 1992. Existentials and other locatives. *Language* 68:553–595.

Gagliardi A, Feldman NH, Lidz J. 2012. When suboptimal behavior is optimal and why: Modeling the acquisition of noun classes in Tsez. *Proceedings of the 2012 meeting of the Cognitive Science Society*.

Gagliardi A, Lidz J. 2014. Statistical insensitivity in the acquisition of Tsez Noun Classes. *Language* 90:1-32.

Gagliardi A, Mease T, Lidz J. 2014. U-shaped development in the acquisition of filler-gap dependencies: Evidence from 15- and 20-month olds. Article submitted for publication.

Gennari, SP, MacDonald MC. 2006. Acquisition of negation and quantification: Insights from adult production and comprehension, *Language Acquisition* 13:125-168.

Gibson E, Wexler K. 1991. Triggers. *Linguistic Inquiry* 25:403-454.

Gleitman, LR. 1990. The structural sources of verb meanings. *Language Acquisition* 1:3-55.

Gómez RL. 2002. Variability and detection of invariant structure. *Psychological Science* 13:431-36.

Gómez RL, Gerken LA. 1999. Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition* 70:109-35.

Gómez RL, Maye J. 2005. The developmental trajectory of nonadjacent dependency learning. *Infancy* 7:183-206.

Guasti, MT. 2002. *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press.

Han, CH, Lidz J, Musolino J. 2007. Verb-movement and Grammar Competition in Korean: Evidence from Quantification and Negation. *Linguistic Inquiry* 38:1-47.

Han, CH, Lidz J, Musolino J. 2014. When Grammar Transmission Fails: Verb-placement and Scope in Korean. Article submitted for publication.

Harley H. 2002. Possession and the double object construction. *Linguistic Variation Yearbook* 2:

Hart B, Risley T. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.

Heycock C. 1995. Asymmetries in Reconstruction. *Linguistic Inquiry* 26:547-570.

Hoff E. 2003. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development* 74:1368-1378.

Hoff E, Naigles L. 2002. How children use input to acquire a lexicon. *Child Development* 73:418-33.

Huang, J. 1993. Reconstruction and the Structure of VP: Some Theoretical Consequences *Linguistic Inquiry* 24:103-38.

Hudson-Kam CL, Newport EL. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology 59*:30-66.

Huttenlocher J, Haight W, Bryk A, Seltzer M, Lyons T. 1991. Early vocabulary growth: relation to language input and gender. *Developmental Psychology* 27: 236-48.

Huttenlocher J, Vasilyeva M, Cymerman E, Levine S. 2002. Language input at home and at school: Relation to child syntax. *Cognitive Psychology* 45:337-74.

Huttenlocher J, Waterfall H, Vasilyeva M, Vevea J, Hedges LV. 2010. Sources of variability in children's language growth. *Cognitive Psychology* 61:343-65.

Johnson EK, Tyler M. 2010. Testing the limits of statistical learning for word segmentation. *Developmental Science* 13:339-45.

Kayne R. 1993. Toward a Modular Theory of Auxiliary Selection. *Studia Linguistica* 47:3-31.

Koster J, Reuland E. 1991. *Long-distance Anaphora*. Cambridge: Camridge University Press.

Lany JA, Gómez RL. 2008. Twelve-Month-Olds Benefit from Prior Experience in Statistical Learning. *Psychological Science* 19:1247-52.

Leddon E, Lidz J. 2006. Reconstruction Effects in Child Language. *Proceedings of the 30th Boston University Conference on Language Development.* Cambridge, MA: Cascadilla Press.

Legate J, Yang C. 2007. Morphosyntactic learning and the development of tense: A new approach to Root Infinitives. *Language Acquisition* 14:315-44.

Lidz J, Gleitman H, Gleitman LR. 2003. Understanding How Input Matters: The Footprint of Universal Grammar on Verb Learning. *Cognition* 87:151-78.

Lidz J, Musolino J. 2002. Children's Command of Quantification. *Cognition* 84: 113-54.

Lidz J, Musolino J. 2006. On the Quantificational Status of Indefinites: The View from Child Language. *Language Acquisition* 13:73-102.

Lidz J, Williams A. 2006. On the lack of reflexive benefactives in Kannada. in T. Battacharya (ed.) *Yearbook of South Asian Languages and Linguistics*, 237-54. Berlin: Mouton de Gruyter.

Lieven EV, Tomasello M. 2008. Children's first language acquisition from a usage-based perspective: In P. Robinson and N. Ellis (eds). *Handbook of Cognitive Linguistics and Second Language Acquisition*. pp. 168-196.

Lightfoot D. 1991. *How to set parameters*. MIT Press: Cambridge.

Lightfoot D. 1999. *The Development of Language: Acquisition, Change and Evolution*. Cambridge, MA: Wiley-Blackwell.

Morgan JL, Newport EL. 1981. The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior* 20:67-85.

Musolino J, Crain S, Thornton R. 2000. Navigating Negative Quantificational Space. *Linguistics* 38:1-32.

Newport EL, Gleitman H, Gleitman LR. 1977. Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. E. Snow and C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109- 150). Cambridge: Cambridge University Press.

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

Naigles L, Hoff-Ginsberg E. 1998. Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language* 25:95-120.

Oehrle R. 1976. *The Grammatical Status of the English Dative Alternation*. Ph.D. thesis, MIT.

Omaki A., Lidz J. 2014. Linking Parser Development to the Acquisition of Syntactic Knowledge. *Language Acquisition*, Article in Press.

Omaki A, Orita N, Lidz J. 2014. Beyond artificial language: Fifteen-month-old infants acquire an English non-adjacent syntactic dependency through statistical learning in the lab. Article submitted for publication.

Orita N, McKeown R, Feldman N, Lidz J, Boyd-Graber J. 2013. Discovering Pronoun Categories using Discourse Information. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*.

Pearl LS. 2011. When unbiased probabilistic learning is not enough: Acquiring a parametric system of metrical phonology. *Language Acquisition* 18: 87-120.

Pearl LS, Lidz J. 2009. When domain general learning fails and when it succeeds: Identifying the contribution of domain specificity, *Language Learning and Development 5:* 235-265.

Pearl LS, Sprouse J. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition* 20:23-68.

Pearl LS, Weinberg A. 2007. Input Filtering in Syntactic Acquisition: Answers from Language Change Modeling, *Language Learning and Development 3:* 43-72.

Pinker S. 1979. Formal Models of Language Learning. *Cognition*, 7:217-83.

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

Phillips C. 1995. Syntax at age two: Cross-linguistic differences. *MIT Working Papers in Linguistics 26*:325-82.

Plaster K, Polinsky M, Harizanov B. 2013. Noun classes grow on trees: Noun classification in the North-East Caucasus. *Language typology and historical contingency: In honor of Johanna Nichols*, ed. by Balthasar Bickel, Lenore A. Grenoble, David A. Peterson, and Alan Timberlake, 153-70. Amsterdam: John Benjamins.

Poeppel D, Wexler K. 1993. The full competence hypothesis of clause structure in early German. *Language* 69:1-33.

Regier T, Gahl S. 2003. Learning the unlearnable: The role of missing evidence. *Cognition 93:*147-55.

Rowe M. 2012. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development* 83:1762-74.

Saffran J. 2001. The use of predictive dependencies in language learning. *Journal of Memory and Language* 44:493-515.

Saffran J, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* 274:1926-28.

Sakas WG, Fodor JD. 2012. Disambiguating Syntactic Triggers. *Language Acquisition* 19:83-123.

Shneidman L, Goldin-Meadow S. 2012. Language input and acquisition in a Mayan village: How Important is Directed Speech? *Developmental Science* 15:659-73.

Shneidman L, Arroyo M, Levine S, Goldin-Meadow S. 2013. What counts as effective input for word learning? *Journal of Child Language* 40:672-86.

Snow CE. 1999. Social perspectives on the emergence of language. In B. MacWhinney (Ed.), *The emergence of language* (pp.257-76). Mahwah, New Jersey: Lawrence Erlbaum.

Snyder W. 2007. *Child Language*. Oxford: Oxford University Press.

Sutton M, Fetters M, Lidz J. 2012. Parsing for Principle C at 30 Months. *Proceedings of the 36th Boston University Conference on Language Development*. Cambridge, MA: Cascadilla Press.

Takahashi E. 2009. *Beyond Statistical Learning in the Acquisition of Syntax*. Ph.D. Dissertation. University of Maryland.

Takahashi E, Lidz J. 2008. Beyond Statistical Learning in Syntax. In *Proceedings of GALA 2007: Language Acquisition and Development*. (Eds.) A. Gavarró & M. João Freitas, pp. 444-54. Cambridge, UK: Cambridge Scholars Publishing.

Tenenbaum JB, Griffiths TL. 2001. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences* 24: 629-40.

Thelen E, Smith LB. 2006. Dynamic Systems Theories. In *Handbook of Child Psychology, Volume 1, Theoretical Models of Human Development*, 6th Edition, William Damon (Editor), Richard M. Lerner (Volume editor), pp 258-312.

Thompson SP, Newport EL. 2007. Statistical learning of syntax: The role of transitional probability. *Language Learning and Development* 3:1-42.

Tomasello M. 2000. Do young children have adult syntactic competence. *Cognition* 74:209–53.

Uriagereka, J. 1988. *On Government*. Ph.D. Dissertation, University of Connecticut.

Valian V. 1990. Logical and psychological constraints on the acquisition of syntax. In Frazier & J. de Villiers (Eds.), *Language processing and language acquisition* (pp. 119-45). Dordrecht: Kluwer Academic Publishers.

Valian V, Casey L. 2003. Young children's acquisition of wh-questions: the role of structured input. *Journal of Child Language* 30:117-43.

Viau J, Lidz J. 2011. Selective learning in the acquisition of Kannada ditransitives. *Language*, *87*:679-714.

Weisleder A, Fernald A. 2013. Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science* 24:2143-52.

Wexler K, Culicover P. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

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

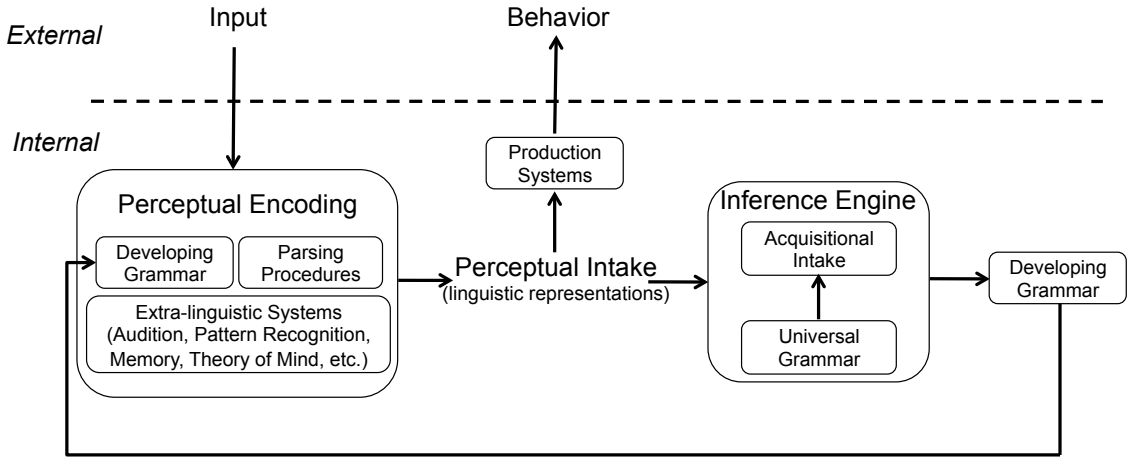

Figure 1: Inside the Language Acquisition Device

Summary points

1) Effects on language acquisition of statistical-distributional patterns in the input are expected under any theory of language acquisition, even those with a large contribution of innate knowledge.

2) The hypothesized existence of Universal Grammar does not remove the need for a theory of learning that explains how experience contributes to language acquisition.

3) The Language Acquisition Device consists of three primary components: Universal Grammar, Intake mechanisms, and Inference mechanisms.

4) Universal Grammar explains the dissimilarity between experience and acquired knowledge. It also allows learners to draw inferences from statistical-distributional evidence to the grammatical representations responsible for producing that evidence.

5) Intake mechanisms explain how learners filter their input to identify critical information to learn from. Input may be filtered by information processing mechanisms, prior knowledge, or expectations associated with particular grammatical hypotheses.

6) Inference mechanisms connect UG with the intake to determine the appropriate mapping from abstract representations to surface form.

Future Issues

1) What is role of on-line information processing mechanisms in shaping perceptual intake?

2) What are the features of Universal Grammar that allow it to shape the acquisitional intake?

3) Can hypotheses about the contents of Universal Grammar be evaluated by examining the predictions they make about the surface features that allow particular grammatical representations to be selected?

4) Does the child's developing, incomplete, grammar allow for the identification of unambiguous information to support grammar selection in a way that overcomes problems of overlapping grammatical hypotheses?

5) Can we build explicit computational models of inference that link hypotheses about Universal Grammar with real input data in order to better understand the contributions of intake, inference and Universal Grammar?

6) How should the models of inference that drive the identification of grammatical features in the target language be evaluated with respect to the algorithms that implement them?