Language Learning and Language Universals

Jeffrey Lidz

This paper explores the role of learning in generative grammar, highlighting interactions between distributional patterns in the environment and the innate structure of the language faculty. Reviewing three case studies, it is shown how learners use their language faculties to leverage the environment, making inferences from distributions to grammars that would not be licensed in the absence of a richly structured hypothesis space.

Keywords: first language acquisition; statistical inference; universal grammar

1. Introduction

What does it mean to learn? Within the cognitive sciences, learning is treated as the creation of a system of mental representations in response to a collection of experiences (Chomsky 1975, Gallistel 1990). The learning organism's task is to infer from data the system that produced that data. In the relatively simple case of learning a word, say 'dog', the learner's job is to collect observations about the use of that word to infer what people in their speech community intend when they say 'dog'. Learning has occurred when the learner knows what thoughts people are having (and intend him to have) when they say 'dog'. We can tell what meaning the learner has acquired for that word on the basis of whether it judges new objects to be dogs or non-dogs. In general, the learning organism's responses to new situations reveal the inferences made on the basis of experience. These inferences, in turn, reveal the properties of the mental representations that underlie learning and use.

The acquisition of syntax is parallel in this respect to the acquisition of words or any other cognitive structure. We assess the representations of the learner by examining how he responds to new sentences. Indeed, nearly every psychologically oriented discussion of syntax begins with these two observations: (i) that we can produce and understand sentences we have never heard before, and (ii) that of the sentences we have never heard before, we can recognize that some are possible but others are not. Just as we can categorize animals we have never seen before as dogs or non-dogs, we can categorize strings of words we have never heard before as sentences or non-sentences.[1] This ability implies a learner

[1] There is no loss of generality if what the learner acquires is a system for assigning

ISSN 1450–3417
Biolinguistics 4.2–3: 201–217, 2010
http://www.biolinguistics.eu

that responds to experience not simply by memorization, but by mapping experience onto representations that make predictions about what other sentences are possible and impossible.

The leading idea of Chomsky's early discussions of learning (Chomsky 1959, 1965, 1975; cf. Lees 1957: 406ff) is that the representations built by the language learner imply a non-obvious metric of similarity between the experienced sentences and the possible but as yet unencountered sentences. The dimensions we use to judge new sentences as possible or impossible appear to be highly abstract and removed from experience. Because these particular dimensions are not the only imaginable dimensions that the learner might have used to construct linguistic representations, Chomsky's argument was that these dimensions (i.e. the dimensions of linguistic analysis) must be supplied not by experience, but by the innate endowment of the child.

Take, for example, the empty category principle (Chomsky 1981, 1986). If it is a true description of our knowledge of English syntax that non-pronominal empty categories must be properly governed, then the learning theory must supply the tools out of which this generalization can be identified or constructed.[2] One of the primary sources of linguistic nativism is the observation that the explanatory pieces of linguistic representation (like proper government and the primitives out of which it is built, for example, c-command, theta-marking, barrier, etc.) are in a vocabulary so far removed from experience that it becomes implausible that both this vocabulary and the complex relations built out of it are induced from experience. If it is further true that a generalization stated over this vocabulary (like the ECP) holds of the syntax of every language, then it becomes a reasonable hypothesis that this generalization reflects a universal feature of linguistic representation, a property that every language must exhibit as a consequence of biological design.

And here is the hypothesized connection between learning and universals. Because the universals reflect constraints on possible representations, learners simply do not consider representations outside of the space defined by these constraints. In the context of a learning theory, identifying the range and limits of possible languages is tantamount to identifying the immanent structure of the child's language acquisition device. A single piece of explanatory machinery would account for both the range of possible linguistic variation and the language learner's initial hypothesis space about what a language can be.

Critically, however, identifying the learner's initial hypothesis space is not equivalent to providing a model of how the learner maps the input onto the appropriate representations (Fodor 1966, Pinker 1984). Such a model requires a procedure for mapping experience (as it is experienced by the learner) onto the representations that generated that experience, i.e. the grammar of the language.

---

convergent derivations to strings of words, mappings from sound to meaning, or triples of sounds, meanings and derivations. The point is just that whatever system the learner acquires for representing the language must be such that it can deal with the unencountered sentences appropriately.

2 I take no stance on the status of the ECP in grammatical theory. The point is that if the theory posits some kind of grammatical knowledge, the learning theory must either be able to construct that knowledge out of some more basic primitives or else it is a primitive built into the learner directly.

This kind of learning model can be seen as a kind of analysis by synthesis, in which the language learner approximates the grammar that generated his experience (Halle & Stevens 1964, Townsend & Bever 2001).[3]

It is standardly held that having a highly restricted hypothesis space makes it possible for such a learning mechanism to successfully acquire a grammar that is compatible with the learner's experience and that without such restrictions, learning would be impossible (Chomsky 1975, Pinker 1984, Jackendoff 2002). In many respects, however, it has remained a promissory note to show how having a well-defined initial hypothesis space makes grammar induction possible in a way that not having an initial hypothesis space does not (see Wexler 1990 and Hyams 1994 for highly relevant discussion).

The failure to cash in this promissory note has led, in my view, to broad skepticism outside of generative linguistics of the benefit of a constrained initial hypothesis space. Despite the fact that deep insights about the range and limits of syntactic variation have been achieved through the methods of comparative syntax (e.g., Kayne 2000, Richards 2001, Baker 2005, *inter alia*), researchers in adjacent areas of cognitive science have been less impressed with the idea that abstracting out the universal formal properties of natural language from the study of individual grammars would lead to progress in explaining language acquisition.

This skepticism derives from several sources. First, it is not clear how these formal theories make contact with developmental data from children learning their first language. Second, the constantly growing sophistication of computational data-mining techniques seems to undercut the premise that the input does not contain the information relevant to building grammatical representations (Elman *et al.* 1996, Christiansen & Chater 1999, Klein & Manning 2004). Finally, a host of research showing that even young infants are sophisticated statistical learners seems to further raise the possibility that learners can extract more from the input that was assumed by standard arguments from the poverty of the stimulus (e.g., Saffran *et al.* 1996, Gomez & Gerken 2000)

Recent work in our laboratory is beginning to link formal theories of language universals with statistical approaches to language learning. This work explicitly examines the kinds of information that is available to a distributional learner and how this information is used in the course of language acquisition (Pearl 2007, Sneed 2007, Syrett 2007, Viau 2007, Pearl & Lidz 2009, Syrett & Lidz 2009, Takahashi 2009, Viau & Lidz 2009,). Our hope is that this research will both remove skepticism of nativist approaches to language acquisition generally and, more positively, show how the inferences that learners make from input distributions are constrained by antecedent knowledge of the universal features of linguistic representation and the range of possible linguistic variation. This program aims to make good on the promise that a constrained hypothesis space helps learners to use the input effectively in acquiring a particular language.

---

[3] In the current paper, we emphasize the nature of the mapping between the input and the acquired grammar by considering the information that learners use and the conclusions that they can reasonable draw from that information. An important open issue in this context is the role of on-line parsing mechanisms in implementing these kinds of inferences in real time. See Baier & Lidz (2009) and Lidz *et al.* (2010) for discussion of the role of on-line algorithms in making inferences in acquisition.

2. Sensitivity to Input Distributions

In the last 15 years, there has been a resurgence of interest in infants' sensitivity to statistical features of the input language. A wide range of studies with infants, children and adults have demonstrated their ability to track statistical features of an artificial language and to use these features to learn generalizations about those grammars (for reviews, see Gomez & Gerken 2000 and Saffran 2003). However, it is important to emphasize that sensitivity to the statistical features of the exposure language leaves open the question of how this sensitivity contributes to the acquisition of a grammar. In every theory of language acquisition, the learner must be sensitive to features of the environment. The fundamental issue in the domain of distributional learning is what kind of learning mechanism this sensitivity feeds into.

In a learning framework in which the child brings the space of possible grammars to bear on the acquisition of a particular language, sensitivity to statistical features of the environment functions as input to a selective learning mechanism. Such a mechanism, restricted by the child's innate endowment to representing only those relations that grammars can represent, provides an algorithm for selecting the appropriate representation of the input. Given linguistic experience plus the (probably infinite) set of possible grammars, the learning mechanism selects that grammar which provides the best description of the input from within that space (see, e.g., Miller & Chomsky 1963, Fodor 1966, Yang 2002, and Pearl 2007).

So, to best understand the role that the learner's sensitivity to the environment plays in language acquisition, it is important to identify the deductive consequences of this sensitivity. As noted in opening, we learn about the acquired representations by examining how the learner approaches new situations. Thus, by exploring the range of new situations that are taken by the child as compatible with their experience, we can infer something about the content of the acquired representations above and beyond the information contributed by experience.

3. Deductive Consequences of Phrase Structure: Constraints on Movement

Every syntactic theory recognizes that sentences in a human language are not simply linear strings of words. Rather, words in a sentence are arranged in nested hierarchical structures (Chomsky 1957, Jackendoff 1977). These structures make it possible, for example, to derive multiple interpretations from a single string. Consider the well-worn example:

(1) ancient history teacher

There are two possible meanings for this string. On one interpretation, this string refers to a very old person who teaches history. On the other, it refers to a teacher of ancient history. The ambiguity of the string supports the idea that a single string can be structured in multiple ways, as shown in (2).

(2) a. [ancient [history teacher]]
b. [[ancient history] teacher]

Constituent structure representations provide explanations for (at least) three kinds of facts. First, constituents provide the units of interpretation, as just seen. 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, as in (3):

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

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

(4) 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 & Newport (2007) make a very interesting 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 kind of statistical signature of the structure in the surface form of the language. The continued co-occurrence 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.

They go on to argue that this statistical footprint could be used by learners in the acquisition of phrase structure. And they show that adult learners are able to use this statistical footprint in assigning constituent structure to an artificial language. But again, showing that learners are sensitive to the statistical features of the environment does not yet provide information about the acquired representations. It is impressive that learners learned about the constituent structure of an artificial language given only statistical information about that structure. But this demonstration remains silent about the character of the acquired representations and the inferences that these representations license.

In order to determine whether the acquired representations have properties that derive from the structure of the learner, it is important to identify their deductive consequences. Do learners know things about constituent structure (even if this structure is acquired using statistical features of the environment) that are not evident in the statistics themselves?

In order to answer this question, Eri Takahashi and I constructed a miniature artificial grammar containing internally nested constituents. In addition, the grammar contained rules which allowed for the repetition of constituents of a certain type, the movement of certain constituents and substitution of certain constituents by pro-forms. We then created a corpus of sentences from this language in which these rules applied often enough to provide statistical evidence for the constituent boundaries. In other words, the language provided statistical cues to the internal structure of the sentences.

Our first question, using this artificial language, was whether adults and infants could acquire constituent structure using only statistical information. The language was presented in contexts that did not provide any referential information, so that no meaning could be assigned to any of the words. And, there was no prosodic or phonological information of any kind that could serve as a cue to the phrase structure. So, to the extent that learners could acquire the phrase structure, they would have to do so through the statistical features of the exposure. In order to test whether the learners acquired the phrase structure, we asked whether they could distinguish novel sentences containing either moved constituents or moved non-constituents. Since only constituents can move in natural languages, we reasoned that if learners could distinguish moved constituents from moved non-constituents, it must be because they had learned the constituent structure of the artificial language. We found that both adults, after 36 minutes of exposure, and 18-month-old infants, after only 2 minutes of exposure, were able to do so (Takahashi & Lidz 2007, Takahashi 2009). Thus, the statistical footprint of constituent structure is detectable by learners and is usable in the acquisition of phrase structure.

Now, the exposure provided to the learners in this experiment 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. We thus went on to ask whether the inference that only constituents can move derives from the learner's exposure to movement rules which apply only to constituents or whether this inference derives from the child's antecedent knowledge about the nature of movement rules in natural language.

To ask this question, we created a new corpus of sentences from our artificial language. In this novel corpus we included sentences in which (i) certain constituents were repeated in a sentence, (ii) certain constituents were optionally absent from a sentence, and (iii) certain constituents were replaced by pro-forms. This combination of operations created a statistical signature of the phrase structure of the language such that it was possible to identify the constituent boundaries in the language. However, in this input corpus we included no examples of movement. This made it possible for us to identify the locus of the learner's knowledge that only constituents can move. If this knowledge derives from the learner's experience in seeing movement rules, then we would expect learners to be unable to distinguish moved constituents from moved non-constituents. On the other hand, if the learner brings knowledge about what kinds of movement operations are possible in natural language to the learning task, then we would expect learners to correctly distinguish moved constituents from moved non-constituents.

We found that both adults and 18-month-old infants displayed knowledge of the constraint that only constituents can move, even when their exposure to the artificial language contained no instances of movement whatsoever. Thus, we can conclude that some of what is acquired on the basis of statistical information is not itself reflected in the statistics. Since the learners in this experiment had seen no examples of movement, their knowledge of the constraint that only constituents can move could not have come from the exposure language but rather must have come from the learners themselves.

In sum, identifying the constituency of a language has consequences for novel sentences with structures never before encountered. These deductive consequences reveal the structure of the learner over and above any role of distributional learning. Distributional learning therefore functions as part of a process of mapping strings onto the grammar that generated them. But some properties of the identified grammar are contributed by the learner's antecedent knowledge of the class of possible grammars.

4. The Deductive Consequences of (In)Definiteness: The Interpretation of Bare Plural Subjects

It is important to recognize that this kind of argument is not limited to learning artificial grammars. For example, Sneed (2007) made an argument of exactly this form in examining children's interpretations of indefinite NPs. She showed, first, that there is a distributional difference between indefinite and definite NPs that could be used in a process of categorizing NPs as either definite or indefinite, and second, that this categorization licenses inferences about interpretation that are not themselves supported in the input.

Because indefinites are generally used to introduce discourse referents and definites are generally used to identify existing discourse referents (Heim 1982, Kamp 1982), indefinites are significantly more likely to be used on the first mention than definites are. Similarly, because old information is more likely to occur earlier in a sentence than new information (Prince 1992), definites are significantly more likely than indefinites to occur in subject position.

Sneed showed that there is ample evidence of these asymmetries in speech to children learning English so that determiners can be accurately classified as either definite or indefinite simply by tracking the relative likelihood of their occurring in first mention contexts and by tracking the relative likelihood of their occurring in subject position. Sneed argued (among other things) that bare plurals (e.g., *dogs*) can be classified as indefinites using this procedure.

As is well-known, bare plurals do occur in subject position some of the time (though as Sneed showed, not nearly as often as they occur in object position). When they do occur in subject position, they are often ambiguous between a generic and an existential interpretation (Diesing 1992). Consider the sentence in (5) and its interpretations in (6):

(5) Crocodiles live in the swamp.

(6) a. It is generally true of crocodiles that they live in the swamp.
b. There are some crocodiles that live in the swamp.

In examining a corpus of child-directed speech, Sneed (2007) found that bare plural subjects were uniformly used generically (i.e. the interpretation (6a)) and were never used existentially.

This observation thus raises the question of whether children know that bare plural subjects can be used with an existential interpretation despite the fact that speech to children simply does not contain examples illustrating this fact. If learners simply acquired the distributional features of the exposure language, then we would expect them to learn that bare plural subjects are obligatorily interpreted generically. On the other hand, if the interpretive properties of bare plurals follow from their meaning in concert with their syntactic position (as, for example, in Diesing 1992), then we might expect that once children have identified an expression as indefinite, its interpretive profile follows automatically.

Indeed, Sneed found that 4-year-old children were equally able to interpret a bare plural subject existentially and generically, despite the fact that they were apparently never exposed to the existential interpretation of such expressions.

There are two important lessons to be drawn from this work. First, while it is certainly true that there are statistical cues to an NPs classification as definite or indefinite, these cues are informative only to the extent that they are antecedently connected to a representation. That is, it is only because of their interpretive properties that indefinites are relatively less likely to occur in subject position than definites are. The conclusion that an expression occurring less likely as a subject than an object is an indefinite is valid only if the learner is using this asymmetry to feed a decision process about preexisting categories. The learner could not use this asymmetry to draw an inference about (in)definiteness unless this asymmetry derived from a fundamental feature of the representation that predicted this asymmetry to exist.

Second, to the degree that learners use distributional information in language acquisition, it must be that such information is used only when there is a question of how to represent a given part of the language. If learners simply tried to reproduce the distribution that they were exposed to, then we would not have found that children allow bare plural subjects to be interpreted existentially. In other words, learners must not be trying to determine whether bare plurals can be interpreted existentially. If they were, then they should have drawn the inference that they cannot. Rather, learners must simply be trying to classify nominal expressions as definite or indefinite. Any additional interpretive properties follow as a matter of grammar, independent of the distribution of these interpretations in the input.

Again, the deductive consequences of distributional learning reveal the contribution that the learner makes to language acquisition. And again, evidence of distributional learning is not evidence against the learner having a highly constrained hypothesis space. Rather, in this case, as in the previous one, distributional learning can be seen as feeding a selective process by which learners use the data to identify the grammar that generated that data (see also Syrett & Lidz 2009, Pearl & Lidz 2009).

5. Selective Learning in the Acquisition of Ditransitives

The inferences from surface form to grammatical representation can also be significantly less direct. Consider, for example, the range of ditransitive constructions in English, Spanish and Kannada.

(7) *English*

a. John sent the book to Mary.

b. John sent Mary the book.

(8) *Kannada*

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

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

(9) *Spanish*

a. Carmen envió el libro a su profesora.
Carmen sent the book to her professor

b. Carmen **le** envió el libro a su profesora.
Carmen ***CL*** *sent the book to her professor*

c. Carmen envió a su profesora el libro.
Carmen sent to her professor the book

d. Carmen **le** envió a su profesora el libro.
Carmen ***CL*** *sent to her professor the book*

Whereas English has two surface forms for ditransitives, Kannada and Spanish have four. Viau & Lidz (2009), building on earlier work by Harley (2002), Bleam (2003), and Lidz & Williams (2005), argue that despite these surface differences, there is a coherent mapping of ditransitive structures across languages. In particular, languages make available two kinds of ditransitives: Those with the theme asymmetrically c-commanding the goal and those with the goal asymmetrically c-commanding the theme. In English, these correspond to the prepositional dative (7a) and the double object construction (7b), respectively, with word order functioning as a surface correlate of the syntactic structure. In Kannada and Spanish, however, word order is not an expression of the underlying configurational structure. Rather, the structure with the theme c-commanding the goal is the morphologically unmarked form whereas the structure with the goal c-commanding the theme is the morphologically marked form. In Kannada, this morphological form is realized through the verbal auxiliary *koDu* (8b,d). In Spanish it is realized through the dative pronominal clitic *le* (9b,d).

The argument for this way of carving up the data comes from two kinds of facts. First, the goal argument functions as a kind of possessor in those configurations where it is argued that the goal c-commands the theme (cf. Oehrle 1976). Second, patterns of binding from one argument into the other also support this classification.

The basic pattern of judgments with a quantified dative argument in shown in (10), where DAT indicates the indirect object, marked with dative case, ACC indicates the direct object, marked with accusative case, and Q- indicates which of these noun phrases contains a quantifier.

(10) a. Q-DAT$_x$ ACC$_x$ BEN

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

b. Q-DAT$_x$ ACC$_x$ unaffixed

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

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

d. *ACC$_x$ Q-DAT$_x$ unaffixed

* 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.'

Descriptively speaking, when the dative-marked object comes first (10a–b), it can bind into the accusative-marked object, whether or not the benefactive affix is present. In contrast, when the accusative-marked object comes first (10c–d), the dative can bind into it only in the presence of the benefactive affix.

If the quantificational phrase is the accusative argument and the pronominal is contained in the dative argument, however, a different pattern emerges.

(11) 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. DAT$_x$ Q-ACC$_x$ 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.'

c. 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'.

d. Q-ACC$_x$ DAT$_x$ 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'.

Here we see that when the accusative-marked object comes first (11c–d), it can bind into the dative-marked object, regardless of whether the benefactive affix is present on the verb. However, when the dative-marked object comes first (11a–b), the accusative-marked object can bind into it only when the benefactive affix is absent. The relevant binding possibilities for quantified dative and accusative arguments are summarized below.

(12)

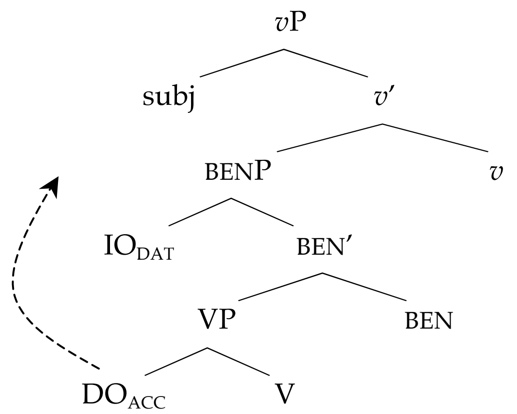
a.	√ Q-DAT$_x$ ACC$_x$ V-**BEN**	e.	* DAT$_x$ Q-ACC$_x$ V-**BEN**
b.	√ Q-DAT$_x$ ACC$_x$ V	f.	√ DAT$_x$ Q-ACC$_x$ V
c.	√ ACC$_x$ Q-DAT$_x$ V-**BEN**	g.	√ Q-ACC$_x$ DAT$_x$ V-**BEN**
d.	* ACC$_x$ Q-DAT$_x$ V	h.	√ Q-ACC$_x$ DAT$_x$ V

Lidz & Williams (2005) argue that the above asymmetries arise from there being two distinct underlying structures for ditransitives in Kannada. When the benefactive affix is present, the DAT–ACC order is the underlying order, with the ACC–DAT order derived by A-movement. Thus, according to Lidz & Williams (2005), (12c) is derived from (12a) by movement of the accusative argument past the dative. The appearance of backward binding in (12c) is due to the fact that A-movement of the accusative over the dative does not destroy the binding relation established in the underlying order (12a), in which the quantified dative NP c-commands and thereby grammatically binds into the accusative. Similarly, since the DAT–ACC order is underlying, a quantificational accusative-marked object cannot bind into the dative (12e) unless A-movement has occurred, introducing a new configuration to license binding (12g).

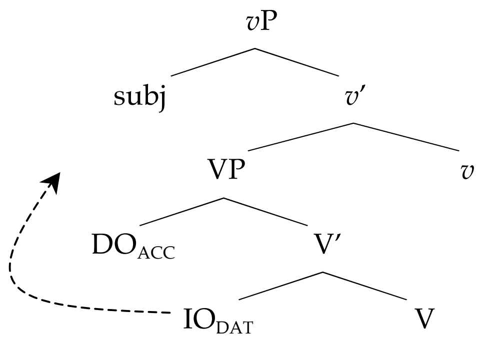
When there is no benefactive affix, the ACC–DAT order reflects the underlying structure, and the DAT–ACC order is derived by A-movement. Thus, (12b) is derived from (12d). The quantified dative NP in (12d) cannot bind into the accusative because it does not c-command the accusative; only after moving above the accusative, as in (12b), can it grammatically bind into the accusative. By the same logic, since the accusative is underlyingly higher than the dative, the binding of the dative by the accusative can be established over this representation (12h) and subsequent A-movement will not destroy it (12f).

We will assume this analysis as well as the syntactic representations that it entails, shown in (13) with the benefactive affix and in (14) without. Optional A-movement is marked with a dashed arrow.

(13) Subject DAT ACC V-BEN



(14) Subject ACC DAT V



The conclusion that there are two distinct underlying structures in Kannada, each of which can be transformed by A-movement of the lower NP past the higher one straightforwardly captures the binding asymmetries discussed above.

The variable binding facts just presented have exact analogs in Spanish (Bleam 2003) with the clitic doubled ditransitives functioning exactly as the benefactive ditransitives in Kannada, and the morphologically unmarked ditransitives functioning exactly alike in the two languages.

In sum, so far, English, Kannada, and Spanish all utilize essentially the same two structures in ditransitives. In the ‘prepositional dative’, the Accusative argument c-commands the Dative underlyingly. This structure is expressed in English as the prepositional dative, in Spanish as the non-clitic-doubled ditransitive, and in Kannada as the non-benefactive ditransitive. In the ‘DO-Dative’, the Dative argument c-commands the Accusative underlyingly. This structure is expressed in English as the DOC, in Spanish through dative clitic doubling, and in Kannada through the benefactive verbal affix. Importantly, despite this fundamental structural symmetry across the three languages, the surface manifestation of these structures is distinct in each language. DO-datives show a distinct surface word order in English but not in Kannada and Spanish. Both Kannada and Spanish, unlike English, have a morphological distinction which correlates with the choice of DO-dative or prepositional dative. In Kannada, the DO-dative variant is marked by a benefactive verbal affix, while in Spanish this variant is marked via clitic doubling of the dative argument.

Given this characterization of the facts, learners cannot rely either on the word order or the morphological form as evidence in determining which of the two abstract structures underlies a ditransitive sentence in the language they are learning. Neither cue is cross-linguistically reliable (Haspelmath 2005). Thus, to the extent that learners can be shown to identify the appropriate structure we are faced with an interesting puzzle. The fact that the very same structures are exhibited in languages with such divergent surface syntax points towards just the sort of cross-linguistic commonality that a selective learning theory is intended to explain, since the pieces of explanation are identical across languages. On the other hand, the fact that the surface realizations of these structures diverge across languages would appear to make it difficult to use the surface form as a cue for the underlying structure (see Hyams 1986 and Snyder 1995 for related problems in other syntactic domains).

Viau & Lidz (2009) demonstrate first that four year old learners of Kannada already command the variable binding facts just described, placing an upper bound on how much time it takes learners to acquire these facts. Moreover, it seems highly unlikely that learners could have acquired the full range of binding possibilities simply by being exposed to positive examples of the relevant sort. The kinds of sentences that exhibit these asymmetries are exceedingly rare. And, even if these sentences did occur, there is no guarantee that the learner would know what the intended interpretation was or whether other interpretations were possible but simply not yet encountered. Thus, Viau & Lidz looked for a more indirect source of evidence that learners could use to identify the correct representation. We argued specifically that the relevant source of evidence resides in the distribution of animate and inanimate goal arguments.

The selective learning account works as follows. The child comes to the learning task with the knowledge that natural languages use at least two ways to configure ditransitives: A possession-based structure in which the IO occurs higher than the DO, and a location-based structure in which the DO occurs higher than the IO. If the child is faced with two distinct types of ditransitive clauses (e.g., DO- vs. prepositional dative in English, benefactive vs. non-benefactive in Kannada, clitic-doubled vs. not clitic doubled in Spanish), she must then identify which of these to associate with which underlying configuration. To do so, the child relies on the distribution of animate IOs. The construction in which IOs are more likely to be animate than inanimate has the possession configuration, since inanimates are highly unlikely possessors. And the construction in which IOs are more likely to be inanimate than animate has the location configuration, for the same reason.

Importantly, once the learner correctly identifies the underlying configuration, the variable binding asymmetries that we have observed in our experimentation follow directly. Thus, the learner requires no experience with particular binding configurations in order to acquire the variable-binding asymmetries we have observed in our experimentation.[4]

---

[4] This reasoning is exactly parallel to the discussion of the 'compounding parameter' in Snyder 1995. The learner can observe productive root compounding as a surface correlate of a particular structure from which many apparently unrelated facts follow.

This is not to say, of course, that there is no learning involved in the acquisition of ditransitives. Our account is a learning-theoretic account in which the child, armed with a set of possible configurations for ditransitives and faced with the data, is able to use certain patterns of distribution to identify a mapping between surface forms and innate configurations. The innate guidance comes from the set of configurations and their semantic properties. Knowing these semantic properties enables the learner to track appropriate distributional information in the surface forms in order to learn which surface forms map onto which of the innate configurations (cf. Hyams 1986). Again, the configurations, in concert with basic structural requirements on variable binding and knowledge of how word order can be manipulated in the target language (which surely is at least partially learned), directly determine the binding possibilities for ditransitives.

The critical feature of this account in the current context is that the asymmetry in animacy of dative arguments across constructions in a useful cue to the underlying structure only if the learner comes equipped with knowledge of the class of possible ditransitive structures. Without that knowledge, the asymmetry in animacy is completely uninformative about the hierarchical structure of the clause.

6. Conclusion

In this paper, I have sketched some results pointing towards an integrated theory of language acquisition that strongly bridges theoretical work on grammatical structure with developmental work on first language acquisition. It has long been held that the theory of language universals is equivalent to the theory of the language learner. This equivalence is only partial, however, since it only relates language universals to a specification of the initial state of the language learner. It is silent with respect to the algorithms that learners use to map their experience onto particular grammars within the space defined by universal grammar. The work described in the current paper shows how it is possible to find evidence about the structure of the learner's initial state from developmental data revealing how learners use distributional, statistical facts about the language they experience to acquire a particular grammar.

In addition, the work I have described demonstrates the convergence between statistical approaches to language learning with traditional nativist approaches. The fact that learners are highly sophisticated when it comes to identifying statistical regularities in the environment does not by itself provide evidence either for or against a learning mechanism driven by innate knowledge of the space of possible representations. It is only when we identify the deductive consequences of statistical learning that we begin to see how statistical learning works in the service of grammatical inference. In the cases reviewed here, the deductive consequences of statistical learning are very rich, pointing to a highly articulated hypothesis space over which statistical inference can be carried out in language acquisition.

References

Baker, Mark. 1996. *The Polysynthesis Parameter*. Oxford: Oxford University Press.

Baier, Rebecca & Jeffrey Lidz. 2009. Predictive parsing impedes word learning in 19-month-olds. Poster presented at the *34th Boston University Conference on Language Development (BUCLD 34)*, Boston. (Boston University, 6–8 November 2009)

Bleam, Tonia. 2003. Properties of the double object construction in Spanish. In Rafael Núñez-Cedeño, Luis López & Richard Cameron (eds.), *A Romance Perspective on Language Knowledge and Use*, 233–252. Amsterdam: John Benjamins.

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

Chomsky, Noam. 1959. A Review of B. F. Skinner's Verbal Behavior. *Language* 35, 26–58.

Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.

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

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Christiansen, Morten & Nick Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23, 157–205.

Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.

Elman, Jeffrey, Elizabeth Bates, Mark H. Johnson & Domenico Parisi. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Fodor, Jerry A. 1966. How to learn to talk: Some simple ways. In Frank Smith & George Miller (eds.), *The Genesis of Language*, 105–122. Cambridge, MA: MIT Press.

Gallistel, Charles R. 1990. *The Organization of Learning*. Cambridge, MA: MIT Press.

Gomez, Rebecca & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Science* 4, 178–186.

Halle, Morris & Kenneth Stevens. 1964. Speech recognition: A model and program for research. In Jerry A. Fodor & Jerrold J. Katz (eds.), *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice Hall.

Harley, Heidi. 2002. Possession and the double object construction. In Pierre Pica & Johan Rooryck (eds.), *The Yearbook of Linguistic Variation* 2, 29–68. Amsterdam: John Benjamins.

Haspelmath, Martin. 2005. Ditransitive constructions: The verb 'give'. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 426–429.

Heim, Irene. 1982. The semantics of definite and indefinite noun phrases. Amherst, MA: University of Massachusetts dissertation.

Hyams, Nina. 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Hyams, Nina. 1994. Discreteness and variation in child language: Implications for principle and parameter models of language development. In Yonata Levy (ed.), *Other Children, Other Languages*, 11–40. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jackendoff, Ray. 1977. *X-bar Syntax*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal Methods in the Study of Language*, 277–322. Amsterdam: Mathematisch Centrum.

Kayne, Richard S. 2000. *Parameters and Universals*. New York: Oxford University Press.

Klein, Dan & Christopher D. Manning. 2004. Corpus-based induction of syntactic structure: Models of dependency and constituency. *Proceedings of the 42nd Annual Meeting of the ACL*, 478–485.

Lees, Robert. 1957. Review of Chomsky's Syntactic Structures. *Language* 33, 375–408.

Lidz, Jeffrey, Ann Bunger, Erin Leddon, Rebecca Baier & Sandra Waxman. 2010. When one cue is better than two: Syntactic vs. lexical cues in infant verb learning. Unpublished manuscript, University of Maryland, College Park.

Lidz, Jeffrey & Alexander Williams. 2005. C-locality and the interaction of reflexives and ditransitives. In Keir Moulton & Matthew Wolf (eds.), *Proceedings of NELS 34*, vol. 2, 389–404. New York: Stony Brook University.

Miller, George & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology*, vol. II, 268–321. New York: Wiley.

Oehrle, Richard. 1976. The grammatical status of the English dative alternation. Cambridge, MA: MIT dissertation.

Pearl, Lisa. 2007. Necessary bias in language learning. College Park, MD: University of Maryland dissertation.

Pearl, Lisa & Jeffrey Lidz. 2009. When domain general learning succeeds and when it fails. *Language Learning and Development* 5, 235–265.

Pinker, Steven. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness, and information-status. In Sandra Thompson & William Mann (eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, 295–325. Amsterdam: John Benjamins.

Richards, Norvin. 2001. *Movement in Language: Interactions and Architectures*. Cambridge: Cambridge University Press.

Saffran, Jenny. 2003. Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12, 110–114.

Saffran, Jenny, Richard Aslin & Elissa Newport. 1996. Statistical learning by eight-month-old infants. *Science* 274, 1926–1928.

Sneed, Elisa. 2007. The acquisition of generics. Evanston, IL: Northwestern University dissertation.

Snyder, William. 1995 Language acquisition and language variation: The role of morphology. Cambridge, MA: MIT dissertation.

Syrett, Kristen. 2007. Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives.

Evanston, IL: Northwestern University dissertation.

Syrett, Kristen & Jeffrey Lidz. 2009. 30-month-olds use the distribution and meaning of adverbs to interpret novel adjectives. Unpublished manuscript, Rutgers University, New Brunswick, NJ & University of Maryland, College Park.

Takahashi, Eri. 2009. Beyond statistical learning in the acquisition of phrase structure. College Park, MD: University of Maryland dissertation.

Takahashi, Eri & Jeffrey Lidz. 2008. Beyond statistical learning in syntax. In Anna Gavarró & M. João Freitas (eds.), *Proceedings of Generative Approaches to Language Acquisition (GALA) 2007*, 446–456. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Thompson, Sandra & Elissa Newport. 2007. Statistical learning of syntax: The role of transitional probability. *Language Learning and Development* 3, 1–42.

Townsend, David & Thomas Bever. 2001. *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.

Viau, Joshua. 2007. Possession and spatial motion in the acquisition of ditransitives. Evanston, IL: Northwestern University dissertation.

Viau, Joshua & Jeffrey Lidz. 2009. Selective learning in the acquisition of Kannada ditransitives. Unpublished manuscript, Johns Hopkins University, Baltimore, MD & University of Maryland, College Park.

Wexler, Kenneth. 1990. Innateness and maturation in linguistic development. *Developmental Psychobiology* 23, 645–660.

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

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