In Defence of UG

B. Elan Drescher, University of Toronto

Although I don't find the notion of Universal Grammar to be controversial enough to need a special defence, I gather that not everyone feels the same way. I have sometimes been puzzled at why a notion that seems so reasonable, at least to me, should provoke such opposition from various quarters. If there are any persons here who have difficulty with the notion, experience suggests that nothing I say is likely to change their mind. So, rather than enter directly into debate about the merits of the notion, and the research program of which it forms a part, I thought it might be useful to try to sketch how I understand this enterprise, and what appeals to me about it. In the second half, I will try to address some objections to this approach that I have come across.

What, then, is meant by the term Universal Grammar? The goal of research in generative grammar is to answer the two questions in (1):¹

(1) Questions Addressed by the Theory of Generative Grammar

1. What constitutes knowledge of language?
2. How is it possible to acquire this knowledge?

The answer to (1.1), for any particular language language, will be a specification of the grammar, G, of that language. Since languages differ, there will be many different G's.

A more difficult problem is to answer (1.2), how each G can be acquired. The nature of our problem can be illustrated schematically as in (2):

(2) The Projection Problem

D --> UG --> G
A person, upon being exposed to data, D, from some language, eventually acquires a grammar of that language, G. This ability that humans are endowed with, i.e. to acquire a grammar of a language, works for any human language: any person can learn any human language. Therefore, the cognitive principles which allow this to happen must be universally applicable. Since these principles, moreover, enable one to acquire a grammar, we may call them, collectively, Universal Grammar, or UG. Thus, while the answer to question (1) will be some particular G, the answer to (2) will be a theory of UG, a set of principles which enable any person to acquire any natural language.

Up to here, there isn't anything too controversial about UG: so far, it is just a cover term for whatever it is that enables one to acquire a grammar. However, the type of principles proposed as being part of UG, and their status with regard to issues of innateness and development, have tended to arouse some opposition. I would like, then, to consider some of these issues in the context of a concrete example, namely the grammar of stress.

For the past few years, I have been working, together with Jonathan Kaye of UQAM, on a project, the aim of which is to write an explicit set of procedures, of the sort that can be implemented on a computer, that could account for how someone might be able to learn the stress system of an arbitrary natural language. In other words, if D is data of some language relevant to stress, and G is grammar of stress for that language, our aim has been to develop a working model of UG in the domain of stress, so that a computer supplied with this UG would actually be able to acquire G from exposure to D. I would now like to briefly consider the components of D, G, and UG in the domain of stress.

We assume, for purposes of this project, the prior operation of rules that
convert the input signal into words and segments. We assume also that the various acoustic cues which indicate phonological stress are mapped into one of three degrees of stress. After this preliminary processing, the data relevant to learning word stress consists of words with vowels marked as bearing main stress (2), secondary stress (1), or no stress (0). Some sample forms are given in (3):

(3) Sample Input Data

\begin{align*}
\text{a. } & \text{va}
\text{nco2uv}\text{e0r} & \text{(Vancouver)} \\
\text{b. } & \text{a2lge0brazil} & \text{(algebra)} \\
\end{align*}

Forms like those in (2) serve as the initial input into our model. However, they do not yet represent the input to the stress component. Stress is sensitive to representations built on projections from syllable structure. Moreover, in many languages, stress is sensitive to syllable weight, or quantity. Therefore, the input words are parsed into syllables before becoming the immediate input to the stress learning system.

Let us turn now to the grammar of stress: how should stress be represented in a grammar? Research on stress suggests that it is fruitful to represent stress in terms of metrical structures— for our purposes, we will assume metrical trees—which indicate the relative prominence of the syllables in a word. For example, in the noun récord, the first syllable is more prominent, or stronger, than the second; we can represent this by a tree in which the left node is strong and the right node is weak. By contrast, stress in the verb recórd can be represented by a tree which is right-dominant. Borrowing terms from the study of meter, we can call a group of syllables in which one syllable is more prominent than the others a foot: in the noun récord, we have a trochaic foot, while the verb recórd is an iambic foot:
(4) Representation of Stress in Terms of Metrical Trees

Metrical theory is not just a handy representational system: it is also a theory about possible human stress systems. Observation of the stress systems of many languages reveals that certain patterns crop up over and over again, while other patterns, from a neutral perspective just as simple, or simpler, hardly ever, perhaps never, occur.

Consider, for example, the stress system of Passamaquoddy:

(5) Passamaquoddy Primary Stress

a. If the penultimate syllable contains a full vowel, it receives stress.
b. Otherwise, if the antepenultimate syllable contains a full vowel, it receives stress.
c. Otherwise, primary stress falls on whichever one of these two syllables is separated by an odd number of syllables from the last preceding full vowel in the word, or—in the absence of a preceding full vowel—the beginning of the word.

This seems quite complex. Compare it now to the pattern in (6), which looks quite a bit simpler:

(6) Imaginary Stress Pattern

a. If a word begins with a voiced segment, stress the penultimate syllable.
b. Otherwise, stress the final syllable.

Nevertheless, this pattern does not, to my knowledge, occur. By contrast, patterns of the Passamaquoddy type occur quite frequently, with smaller or larger differences of detail. This sort of fact calls for an explanation. One way of accounting for this phenomenon is to suppose that metrical structures can be assembled in only a limited number of ways. We might posit, for example,
that UG contains a relatively small number of parameters which control the construction of metrical structure. A learner’s job is to figure out how to set these parameters for the particular language he or she (or it, in the case of a computer) is exposed to.

We have been working with the parameters given in (7):

(7) Parameters of Metrical Theory

P1: The word-tree is strong on the [Left/Right]
P2: Feet are [Binary/Unbounded]
P3: Feet are built from the [Left/Right]
P4: Feet are strong on the [Left/Right]
P5: Feet are quantity sensitive (QS) [Yes/No]
P6: Feet are QS to the [Rime/Nucleus]
P8A: There is an extrametrical syllable [No/Yes]
P8: It is extrametrical on the [Left/Right]
P7: A strong branch of a foot must itself branch [No/Yes]
P9: A weak foot is defooted in clash [No/Yes]
P10: Feet are noniterative [No/Yes]

We will briefly illustrate the effects of some of these parameters, starting with P1. Main stress in a word is controlled by an unbounded word tree, in which either the leftmost or the rightmost node is labelled Strong. For example, the word tree in (8a) has been constructed with P1 [Left]; this gives initial stress, as in languages like Hungarian, or Latvian. Setting P1 [Right] gives fixed final stress, as in French and Farsi:

(8) a. Word-tree with P1 [Left] Constructed on Rime Projections

```
           R
         /   \  
        /     \ 
       /       
      /         
     /           
    /             
   /               
  /                 
 S | W | W | W |   (Fixed initial stress; e.g.
|   |   |   |   | Hungarian, Latvian)
Ri Ri Ri Ri Ri
```

Main stress is not necessarily confined to a peripheral syllable, however, since P1 can interact with other parameters to produce different results. For
example, a peripheral syllable may be designated as extrametrical by PBA and P8, meaning it does not participate in the construction of the word-tree. Extrametricality can result in main stress falling on the second or penultimate syllable; in (8b), it falls on the second syllable:

(8) b. P1 [Left], Leftmost Syllable Extrametrical (PBA [Yes], P8 [Left])

```
R
/ \   
/   \ 
/     
S W W W
/   
(Ri) Ri Ri Ri Ri
```

(Fixed second stress; e.g. Lakota, Araucanian)

In languages such as these, only one syllable in each word is marked Strong, while all the rest are Weak. In many languages, however, syllables are first grouped together into feet, and then the word-tree is constructed on the feet. Every foot receives a stress; hence, languages with feet also exhibit secondary stresses. If a language has feet, a number of other parameters come into play.

P2 allows feet to be at most binary, or else unbounded. Suppose we choose binary feet, which give rise to an alternating pattern of weak and strong syllables. We must then choose P3, the direction of construction, which may be either from left to right or from right to left. In addition, we must select P4, which allows each foot to be left-dominant or right-dominant. We present two illustrative examples---Maranungku, in (9), has P3 [Left] and P4 [Left]---i.e. left-dominant feet constructed from the left; and Warao, in (10), has left-dominant feet constructed from the right---i.e. P3 [Right], P4 [Left];
(9) P3 [Left], P4 [Left]: Left Dominant Feet Built from the Left (Maranungku)

a. F F
   /
   \\ |
   S W .
   | | |
Ri Ri Ri
   | | /
   \ \ \ \ \ 
merepet yangarmata
2 0 1

b. F F
   /
   \\ |
   S W S W S W
   | | | |
Ri Ri Ri Ri Ri Ri Ri
   | | | | | |
   \ \ \ \ \ \ \
2 0 1 0 0 0 0

(10) P3 [Right], P4 [Left] Left Dominant Feet Built from the Right (Warao)

a. F F F F
   /
   \\ |
   S W S W S W S W
   | | | | | | | |
Ri Ri Ri Ri Ri Ri Ri Ri Ri
   | | | | | | | |
   \ \ \ \ \ \ \ \ \ 
ypurukitanehase yiwaranae
1 0 1 0 1 0 2 0

Word trees have been omitted from the examples; however, they would be constructed on the feet, with main stress devolving upon the strongest vowel in the strong foot. Depending on how various parameters are set, the strong syllable of the strong foot may be fairly far from the periphery; in this way, main stress can be quite variable, despite the limited choices in word-tree construction. Note one additional fact about the Warao word in (10b): the first syllable, being alone in a foot, ought to receive a secondary stress. It's stresslessness is due to the setting of the defooting parameter, P9, to [Yes]. Warao apparently does not tolerate stress clashes, wherein two adjacent syllables are stressed; hence, the non-branching foot is defooted, and the first syllable does not receive a stress.

The feet in (9) and (10) are not affected by the internal structure of the rimes on which they are constructed; in such languages, foot construction, and hence stress, is said to be insensitive to quantity—select P5 [No]. However, many
languages have quantity sensitive stress systems. Quantity sensitivity, in the
theory being adopted here, means that a branching rime (or a branching nucleus,
depending on the setting of P6) may not occupy a weak position in a foot; hence,
the configurations in (11) are not allowed:

(11) Quantity Sensitivity

a. P6 [Rime]:

\[ \star F \backslash \star W \backslash R \]

b. P6 [Nucleus]:

\[ \star F \backslash \star W \backslash N \]

A branching Rime must be Strong A branching Nucleus must be Strong

Syllables containing branching rimes or nuclei are called heavy; those that do
not are called light. It follows from (9) that in quantity-sensitive stress
systems, all heavy syllables are stressed. The presence of heavy syllables can
considerably disrupt the smooth alternation of stresses we have observed up to
here.

In addition to these parameters, UG must also incorporate a learning theory
which specifies to the learner how to set the parameters. For example, what in
the data tells us if a system is quantity sensitive or not, or how to build
feet? This is not a trivial question. But, after working out a learning
procedure of this type, we have been able to write a computer program which can
succeed in correctly setting the parameters of Passamaquoddy, complex though
that system seems to be. The same computer, however, is unable to learn the
apparently simple though nonexistent pattern in (6).³

We have thus arrived at a certain hypothesis concerning the nature of UG, at
least in the domain of stress. Positing that a series of metrical parameters
and the learning theory associated with them form part of UG accounts for a number of things: we get a nice account of particular grammars, i.e. we can elegantly describe the facts of stress in many different languages; second, we can account for why certain patterns are common, while others do not occur; and finally, we can construct models which can learn the grammar of stress for a variety of languages, given only sample data, and of course, a number of simplifying assumptions (such as no exceptions). We can thus make some headway in answering the central questions we began with.

That is why I find this type of research strategy appealing. Yet, many who concern themselves with cognitive science are, apparently, not as enthusiastic about this line of investigation as I am. So, perhaps this is a good time to begin the more overt defence of UG. To keep things concrete, I will continue referring, wherever relevant, to the example at hand.

One objection sometimes raised is that it is begging the question to put so many specific principles into UG; that by doing so we are somehow putting an end to true scientific research, which ought to focus on how the principles are learned, presumably by means of general strategies. A number of considerations point against this sort of objection, however.

First, it is not true that it is so easy to propose that some principle is part of UG, and that such a proposal represents a backing away from hard problems. If, in order to solve a problem in one language, we propose that some principle is in UG, then we are stuck with that principle for all languages. To take a trivial example, suppose we tried to help out our computer learner by telling it that some special fact about English stress (say, the fact that final syllables are extrametrical) is part of UG. While the computer would then have an easier
time learning English, it would be unable to learn many other stress systems, and the proposed universal would have to be rejected. Or, if every time we encountered a new possibility we simply were to add a parameter to UG, we would end up with a large and unwieldy list, making every language hard to learn, defeating our purpose.

In fact, positing some principles of UG derived from study of one language, and then studying the consequences of this move with regard to other languages, has proven to be an enormously productive research strategy, especially in the last few years. Consider, for example, the history of WH-movement, involved in question formation in English. In the earliest work in generative grammar, it was proposed that questions such as (12a:i) were derived from an underlying structure (12a:ii) by a transformation called WH-Movement, which moves the WH word to the front of the sentence (and triggering inversion of the subject and auxiliary).

(12) WH-Movement in English

(a) i. Who will John see?
    ii. John will see who

(b) i. *Who will John see the man that knows?
    ii. John will see [the man [that knows who]]

(c) i. *Who will John like the thought that will resign?
    ii. John will like [the thought [that who will resign]]

It was then noticed that there are a variety of structures from which movement of WH is prohibited, such as those in (12b) and (12c). At first, the various restrictions were simply incorporated into the rule of WH-Movement. It was then noticed, however, that a variety of other movement rules in English were also subject to the same restrictions. It was proposed by Chomsky that these conditions be factored out of particular rules, and even out of the particular
grammar of English—he proposed rather that these conditions were universal conditions on syntactic movement. Thus, it was proposed that NP, S and S' are 'bounding nodes', and that no element can move across more than one such node. But if this principle is part of UG, we then expect to see it show up in every language that has syntactic movement rules (of course, it is not necessary for a language to have such rules). In particular, Italian and French, which have WH-movement, should share these restrictions. Interestingly enough, there are restrictions on WH-Movement in these languages, but they are not exactly the same as in English. For example, in French one can say sentences like (13):

(13) WH-Movement in French

i. Voilà une liste des gens à qui on n’a pas encore trouvé quoi envoyer. 'Here’s a list of the people to whom we’ve not yet found what to send'

ii. [ [on n’a pas encore trouvé [ [envoyer quoi à qui]]]]

S' S
S' S

Does this invalidate the hypothesis that bounding nodes are part of UG? Not completely, since there still are restrictions on French and Italian movement. Rather, it was proposed by Rizzi for Italian, and by Sportiche for French, that the facts of these languages could be explained if it is assumed that their bounding nodes are NP and S', but not S. That is, the original proposal was wrong in requiring all languages to have the same bounding nodes. If we hypothesize a parametrized version of this principle, we can account for the facts of English, French, and Italian. This hypothesis proposes, then, that a child learning a language does not have to learn the principle of bounding nodes, but only has to select which nodes are bounding nodes in his or her language. How a child does this, of course, is not immediately obvious, and work on this sort of question is still in its infancy. It is clear, though, that such proposals do not put an end to research on acquisition, but suggest
paths that such work could take.\textsuperscript{4}

Moving now to the offensive, it is not out of place to observe here that more traditional approaches to acquisition have not had much success in this domain; indeed, it is fair to say that questions of the sort mentioned above have scarcely been raised outside of work inspired by notions of UG. Turning to our computer program, someone interested in the general problem of learning might think that we have provided our computer with a UG that is far too specific and richly articulated: aren’t we making the computer’s task too easy? All the computer has to do is select from a menu of choices. Doesn’t real learning, someone might object, involve much more than that? Well, anyone who thinks that is free to create their own computer program, one which would attempt to learn the grammar of stress without any specific parameters of the sort we supply. We doubt that such an attempt could succeed (of course, we may be wrong). I think it is fair to say that there has been virtually no progress in attempting to explain anything in terms of learning by induction. Consider, as a case in point, typical work in machine learning, currently the subject of much activity. The usual pattern is that problems that appear to be problems of induction are converted into problems of deduction, or selection from preassigned choices. For example, a computer program that is supposed to learn the concept of ‘large ball’ (in a world of large and small objects of various shapes and colours), is usually given in advance all the concepts it will need to arrive at the answer. And even then, it is often presented with both positive and negative examples, and is generally coddled in a manner that is quite unrealistic for the case of a child learning a language. VanLehn (1987) describes a program called SIERRA that can learn to subtract. I quote from the abstract:
The key to SIERRA's success lies in supplying a small amount of extra information with the examples. Instead of giving it a set of examples, under which conditions correct learning is provably impossible, it is given a sequence of "lessons," where a lesson is a set of examples that is guaranteed to introduce only one [new BED] subprocedure. This permits unbiased learning, i.e., learning without a priori, heuristic preferences concerning the outcome.

Since the structured sequences required for unbiased learning are not present in the case of a child learning language, it is reasonable to opt for biased learning, i.e. a richly articulated UG.

Attempts to get a computer to learn by induction lead to remarks such as the following, from Winston (1970), of the MIT AI Laboratory, who argues (cited by VanLehn) for "the importance of good training sequences prepared by good teachers. I think it is reasonable to believe that neither machines nor children can be expected to learn much without them."

Finally, it is interesting to note that in other biological domains, problems that looked like problems of instruction have usually turned out to involve selection from a prespecified set of choices. Consider the immune system. Because of the enormous number of foreign bodies that the immune system can recognize and combat, it had been widely believed that the antigen teaches the antibody molecule what conformation to take when confronted with a foreign body. Jerne (1967), who won a Nobel Prize for his work on the immune system, showed, however, that antibody formation is a selective process, whereby the antigen selects and amplifies specific antibodies which already exist. He writes, "Looking back into the history of biology, it appears that wherever a phenomenon resembles learning, an instructive theory was first proposed to account for the underlying mechanisms. In every case, this was later replaced by a selective theory." Other examples are: the instructionist Lamarckian view of heredity, as opposed to Darwin's theory of Natural Selection, and theories
accounting for the specific properties of certain proteins (see Jacob 1982, 15ff.). A parametric theory of UG, then, is not the only selectionist theory in biology—on the contrary, the burden of proof is on the advocates of an instructionist theory of learning.⁶

It is sometimes asserted that if there is an innate learning program, it must not be specific to language, but must be part of a more general cognitive faculty; i.e. what we have been calling UG is really just the application to language of UC, a theory of Universal Cognition, which also accounts for how learning takes place in other domains. It may, of course, turn out that this is the case, although there is no a priori reason to insist on it. In any event, there appears to be no other way to find out if this is so or not except by studying a number of domains in the way we have been studying UG, and then seeing if the principles we arrive at in different domains can be related.

Consider, for example, the metrical parameters: can they be related to anything else? There may be connections with perception of rhythm in general. Hayes (1985) notes that it has been shown that a series of evenly spaced pulses are perceived as paired into binary trochees (S W) if alternate pulses differ in intensity (...DUH duh DUH duh...); by contrast, pulses which differ in duration (...duh duhhhhh duh duhhhh...) are perceived as pairs of iambbs (W S). Hayes observes that languages with quantity sensitive stress systems, i.e. languages where there is a systematic difference in duration between heavy and light syllables, tend to have iambic feet, while quantity insensitive languages, in which duration of syllables does not play an important role, tend to have trochaic feet. This perceptual phenomenon is also exploited in music. For example, if a series of eighth notes is articulated tu du tu du... on an instrument like the recorder, the listener will hear them as trochaic pairs (S
the difference between \textit{tu} and \textit{du} being mainly one of intensity of attack; articulating \textit{ru tu ru tu}..., however, where the difference between \textit{ru} and \textit{tu} is mainly one of duration, results in the impression of iambic pairings (W S).

Perhaps these accentual, rhythmic, and musical phenomena all derive from basic notions of constituency. According to Halle and Vergnaud (1987, 47),

A fundamental property of sequences of linguistic units – phonemes, morphemes, words, etc. – is that when such units are concatenated, it is not the case that all elements in the string are treated on a par, like beads on a string. Rather, sequences of linguistic units are composed of one or more constituents in each of which one element is especially marked, or made the head...stress is a surface manifestation of a particular kind of constituent organization imposed on sequences of stressable phonemes.

The argument, then, is not that principles of UG cannot be related to wider cognitive principles; rather, it is that it is legitimate to study UG in the domain of language without necessarily deriving everything from something else, a strategy that simply has not produced results. We can best find out if the metrical parameters are indeed simply one manifestation of a deeper set of principles by studying them, and other domains, in more detail, along the lines sketched above.

It is sometimes suggested that UG is not innate, but develops from more basic principles, perhaps in the manner suggested by Piaget. Again, this may well be the case, but it must be shown, and not merely insisted on without proof. As far as I know, no plausible candidates for a ‘proto-UG’ have been advanced.\footnote{7}

It is sometimes argued that a UG of the type supposed above could not exist because it could not have evolved. Robin Campbell has called my attention to the following syllogisms by T. Bever:
(14) Bever's Syllogisms, via Robin Campbell

<table>
<thead>
<tr>
<th>Chomskyian Syllogism</th>
<th>Anti-Chomskyian Syllogism</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. X can be shown to be in grammar G.</td>
<td>a. X can be shown to be in grammar G.</td>
</tr>
<tr>
<td>b. X cannot have been learned.</td>
<td>b. X cannot have evolved.</td>
</tr>
<tr>
<td>c. Therefore, X is innate.</td>
<td>c. Therefore, X is learned.</td>
</tr>
</tbody>
</table>

The two sides of the syllogism are not equal, however. In the case of learning, a conclusion that 'X cannot be learned' is often well-supported. Thus, it could be shown that the relevant examples are simply not available in the experience of the learner, or that unbiased learning is impossible in certain situations. Here, there are also mathematical results, as well as extensive experimental research.

The argument that 'X cannot have evolved', however, often rests on much shakier grounds, given our ignorance of the evolution and genetic basis of cognition. According to F. Jacob (1982), "Present-day biology actually has little to say about human behavior and the genetic component of mental abilities." Thus, the evolution of other cognitive faculties is as much a mystery as the evolution of UG, and clearly no conclusion about their existence can be drawn from this fact. Moreover, we will not make much progress in finding out how UG, or any other cognitive faculty, evolved unless we have some idea of what it is. The usual research strategy, even in the case of much simpler cognitive systems, is to work from observed behaviour, positing principles of a rather abstract sort, before getting down to genetics. For example, Gould and Marler (1987) discuss the organization of bee knowledge. Among other things, they show that "bees behave as though they have an appointment book by which they schedule their visits;" they go on to discuss the structure of the entries in this appointment book. A model of bee cognition can be developed in this way, even though its genetic basis and evolutionary history are currently unknown.
It is often assumed that human language evolved for reasons of communication. This assumption does make the evolution of UG appear to be especially mysterious, as it is not easy to see how many of the characteristics of UG are necessary for communication. But there is no reason to assume that language evolved in this way. To quote again from F. Jacob (1982, 58),

The role of language as a communication system between individuals would have come about only secondarily, as many linguists believe. Its primary function would rather have been, as with earlier evolutionary steps in mammals, the representation of a finer and richer "reality," a way of handling more efficiently a greater amount of information...most of the information to be shared with others and concerning immediate features of life could be handled by means of rather simple codes. In contrast, to translate a visual and auditory world so that objects and events can be precisely labeled and recognized weeks or years later requires a much more elaborate coding system.

As with other organisms, our experience is shaped by our cognitive and perceptual systems. Research into UG aims to discover the principles which underlie one of these systems. I would like to conclude with an example from English stress. We observe that a trisyllabic word like Álbany is stressed on the antepenultimate syllable, while agénda is stressed on the penult. This difference can be traced to the fact that the penultimate vowel in Álbany is followed by a single consonant, while the penultimate vowel in agénda is followed by two consonants. As this distinction is quite systematic in English (compare cámara, Canáda, etc. with veránda, Tóronto, etc.), we might propose that a language learner pays attention to the number of consonants that follows a vowel when learning the stress system.

But consider now the word álgebra, which has antepenultimate stress though the penultimate vowel is followed by two consonants. A language learner who is counting consonants would presumably have difficulty with this type of case, and would have to realize that it is not just the number of consonants, but their
identity, that is important: sequences of br (and other obstruent liquid sequences) count as if they are only one consonant.

Our computerized learner, and, by hypothesis, real children, do not have a problem with this case, however. As mentioned above, the input to our stress learner consists of words parsed into syllables. A learner has evidence, both from UG and from language-particular facts, for the parsings al.ba.ni, a.gen.da, and al.ge.bra. UG also tells our stress learner to ignore the onsets of syllables, and to pay attention only to whether the rime (that part of the syllable that remains when the onset is stripped off) contains one element or more than one (i.e. is light or heavy). Replacing segments by X (since their identity is not taken into account), the words in question appear to the stress learner as: XX.X.X (al.ba.ni), X.XX.X (a.gen.da), and XX.X.X (al.ge.bra). Viewed through these lenses, algebra looks just like Albany, and there is no temptation to class it with agenda. As to a bee in a field of flowers, the relevant cues stand out with striking clarity.

Notes

* This is the text of a talk delivered at York University in March 1987. I would like to thank Robin Campbell for much interesting discussion of some of these issues, and Ellen Bialystock, Juan Pascual-Leone, and Peter Roosen-Runge for their comments.

1. These questions were put at the centre of linguistic theory by Noam Chomsky; see Chomsky (1988) for a recent formulation of these and other questions.


4. For a recent discussion of this case, see Lightfoot (1986).

5. See also Jerne (1985).

6. Lightfoot (1986) discusses a few more examples of this type.

7. See Piattelli-Palmarini (1980) for a thorough airing of this issue.

8. See also Chomsky (1988, Chapter 5).

References


