12

Three Aspects of the Relation between Lexical and Syntactic Knowledge

T. G. BEVER

Carol Chomsky's work on language presents a balance of empirical, theoretical, and applied research. In this brief essay, I outline three areas of current research that reflect different emphases in her work. First, language learning can proceed over a long period, possibly into adolescence. That is, the syntactic affordances of individual words can be acquired slowly, sometimes not until early adolescence. Second, language learning is robust despite many individual and environmental differences: the acquisition of basic syntactic patterns follows the same general patterns despite wide variation in individuals and linguistic environment. Third, (psycho)linguistic science can be usefully applied to such problems as reading; in particular, fluent reading involves integrating lexical and phrasal levels.

Each of these areas involves a balance between processing of the two major kinds of information one has about one's language: the lexicon and the syntax. Knowing and using a language necessarily requires both kinds of information: C. Chomsky was somewhat unusual in recognizing immediately that a central problem for understanding language is the relation between these two kinds of information.

12.1 How long does language learning really take?

A dominant explicit and implicit assumption of today's language science is 'the biolinguistic assumption': that language learning is paced by internal maturational factors. The apparent formal similarities of all languages initiates the idea that biological linguistic universals underlie linguistic structure, and hence, language learning. For many years, the matter seemed open and shut to many: the notionally available alternative model of language 'learning', 'associative stimulus-response training',

is hopeless in the face of several facts. First, language learning appears to proceed without direct feedback; second, the child is exposed to a very small number of grammatical utterances, usually without any didactic intent by its caretakers; finally, the similarity of stages in normal language learning in different languages—even sign languages—attest to universal constraints and computational stages that all children bring to language-learning experiences (for recent discussions, see Hauser et al. 2002; N. Chomsky 2007a).

The emerging biolinguistic program defined research on language acquisition as the close study of specific stages and of the prefigured typological dimensions of language (aka 'parameters') that a child must set for his or her native language (see, e.g., Hauser et al. 2002; Lightfoot 1991; Fodor 2001; Fodor and Sakas 2004). A strong empirical corollary of this research approach is that the critical features of each language are acquired by mid-childhood, certainly by age six years: children not only have mastered intricacies of syntactic patterns within, they understand the structure of remote relations between clauses despite frequent patterns that seem identical but are not: for example, the difference between 'John told Bill to go' in which 'Bill' should 'go', and 'John promised Bill to go' in which 'John' should 'go'.

C. Chomsky started her graduate research on language acquisition within this historical framework. But her eventual dissertation famously mitigated the categorical and punctate interpretation of how language learning proceeds (C. Chomsky 1969). She used innovative methods to show that while children seem to have mastered complex verb distinctions such as the difference between 'tell' and 'promise', they actually can systematically misunderstand 'promise' sentences as being like 'tell' sentences until at least age ten or later. That is, a child of nine might interpret 'the monkey promised the dog to leave' as meaning 'the monkey told the dog to leave'. Similarly, children confuse so-called 'tough' constructions with corresponding actives: if a child is asked to make a blindfolded monkey 'easy to see' s/he might simply remove the blindfold. To account for such data, Chomsky formulated a version of a 'minimum distance' principle, on which language-learning children apply a principle that the agent of a verb is its nearest leftward noun phrase. This principle is often discussed even today in comprehension models, generally without attribution to Carol Chomsky. In the language acquisition world, this led to a series of studies, actively pursued today, showing that many aspects of linguistic structure are mastered over a much longer period of childhood than was earlier believed (e.g., see articles collected in Frazier and DeVilliers 1990). C. Chomsky later broadened the evidence for the impact of experience by showing that extensive exposure to written English is associated with more sophisticated mastery of complex constructions (C. Chomsky 1980).

Chomsky's findings should have shattered supportive corollaries of the biolinguistic program—that language acquisition should be rapid and categorical stage by stage, and largely unaffected by experiential variables. But such results can be ignored by committed biolinguists, at least as having no deep implications for the biolinguistically

paced model of language acquisition. If the period is longer than originally thought, it does not in itself deny the possibility of prefigured universals that channel the child's mastery of a native language into setting specific parameters—it remains the case that the incremental associationist model of language learning cannot account for structural phenomena, whether final language learning takes six or twelve years: either way, the language experience of the child is too impoverished, and often totally lacks any corrective feedback to account for a purely incremental accumulation of linguistic structures. If enriched language experience facilitates fluent mastery of complex constructions and subtle vocabulary nuances, that does not undercut the fundamental core of acquisition as dependent on universal maturationally emergent mechanisms. A recent example is evidence that certain neurological fiber tracts connecting different language-processing areas of the brain are not fully functional even at age five, and await further development (see Friederici 2009). This is a modern physiological confirmation of some facts that may underlie the gradual acquisition of syntax in a way consistent with the biolinguistic maturational hypothesis. Linguistically relevant neurological maturation itself may be much slower than had been usually thought.

Several developments in recent years have increased the salience of the finding that at least some linguistic features are acquired slowly. First, there has been a burst of interest in showing that random discourses do contain statistically valid information from which it is possible to extract categorical structures, given the right sort of statistical engine (Cartwright and Brent 1997; Moerk 2000; Mintz 2002, 2006; Yang, 2006) At the same time, studies of infants and young children are showing that they do have pattern-extracting abilities that might interact with statistically valid information to aid, if not completely support language acquisition (Gerken 1996). Finally, we have never actually been restricted to considering only two kinds of learning models: behaviorist associationism vs biolinguistic nativism. There is at least one third kind, which may be receiving renewed support, that integrates both biologically prefigured categories and the statistically valid features of experience: a hypothesis-testing model on which language learning utilizes both innate constraints and human problem-solving strategies (Bever et al. 1984; Bever 2009).

In the last century, the sustained work of gestaltists (especially Wertheimer 1945) outlined several features of how problem solving works. The most important feature is the emotional importance of the so-called 'aha' reaction when a person thinks s/he has found the solution to some problem or task. What this shows is often overlooked in formulating theories of learning—humans experience an intrinsic thrill merely in solving a problem—this is true, whether the problem is an important one or not. Indeed, we often engage in creating otherwise useless problems to solve, just so we can enjoy the experience of solving them. An entire theory of aesthetic experience is based on this principle: music sets acoustic problems for resolution; graphic arts do the same in vision; of course, drama, literature, and poetry are the flagship cases of problem creation and resolution.

Suppose the child treats discovering the syntax of her language as one of the first big life problems to solve. This would explain it as motivated not by the urge to communicate (as in the usual behaviorist explanation), nor as forced by maturation (as in the strong biolinguistic explanation), but as an activity that is cognitively intrinsically thrilling and fun. That is, the child learns the language because it is an exciting, self-stimulating thing to do (Bever 1987). At the same time, current sociolinguistic research reminds us that language variation serves an important group-identifying purpose (see articles in Eckert and Rickford 1995). On this integrated view, children are determined to solve the problem of how their native language works because it helps them be 'just like' the grown-ups around them: the cognitive thrill involved in successive solutions to how the adult system works provides stage by stage feedback and intrinsic reward.

What are the structural features of problem-solving models, and what do they tell us about language and language learning? Miller et al. (1960) rehabilitated the older Gestalt model of problem solving, as 'hypothesis formation and testing'. In the case of language, this requires a set of systems that formulate hypotheses and mechanisms for testing those hypotheses. Recently, we have formulated this in the framework of an analysis by synthesis model of language acquisition (Townsend and Bever 2001; Bever 2009). On this model, children apply both inductive and deductive computations for hypothesis formulation and confirmation. The overall goal is to find a coherent structure for the language experiences that systematizes the relation amongst and between meanings and forms.

This model makes several kinds of predictions:

(a) Languages should exhibit statistically valid patterns, independent from structural constraints. This is a necessity for the inductive component of the analysis by synthesis acquisition model to have data to formulate hypotheses for structural confirmation based on the child's structural, deductive, language component. A simple example of this is the universality of a 'Canonical Syntactic Form' in every language. In English, this appears with a general surface: almost every sentence has the surface form, 'noun phrase' followed by a 'predicate' that agrees with the noun phrase, followed by other material. For a time, this has been thought to motivate the existence of a particular configurational constraint on derivations, the so-called 'extended projection principle' (N. Chomsky 1981; Lasnik 2001; Epstein and Seely 2002; and Svenonius 2002; Richards 2003). Other languages have other canonical forms, sometimes based on word order (e.g., German is 'inflected verb second'), sometimes based on inflectional morphology, sometimes on a combination of linguistic features. In each case, the universality and frequency of the canonical form is unmotivated by universal linguistic architectural constraints—thus, attested languages are a subset of architecturally possible languages, such that they exhibit forms that facilitate the discovery by the child of an initial set of generalizations for test and analysis in structural

terms. The theoretical problem with the EPP is that it is an add-on configurational constraint on derivations. This did not matter so much in the context of GB theory, with its many 'filters'. But in the context of today's Minimalist program it is definitely a stipulated universal, not one that follows from more general principles. Thus if we can explain it as a function of what makes attested languages learnable, we have removed it as a theoretical carbuncle (Bever 2009).

(b) In English, for example, it is critical that the canonical form both have a near universal surface appearance, but also have critical differences in some of the mappings of that surface form onto thematic relations. In English *almost* every sentence with the canonical surface form assigns the initial noun phrase 'agent' or 'experiencer' status in relation to the following predicate. But it is critical for the model that not every such sentence is mapped the same way. This variation sets a problem for the child to solve: what is the overall structure that accounts for both the surface features and the variation in the thematic mapping? This calls on application of the structural component of the dialectic involved in building up syntactic knowledge.

(c) The canonical form is learned by the child as an inductive process rather than an initial stage. Numerous studies have confirmed this, that the child starts to rely on the canonical form of its language by age three to four, not initially (Bever 1970; Slobin and Bever 1982).

(d) The problem-solving model can mitigate the 'poverty of the stimulus', by utilizing the canonical form to generate sets of meaning–form pairs that the child has not yet experienced. This helps the language-learning child to be a 'little linguist' (Valian 1999) without having memorized a large number of form–meaning pairs, and without querying the adult world the way grown-up linguists do. A classic reflection of this is in the research of Ruth Weir (1962) showing that children manifestly 'practice' to themselves the paradigms in their language—most important is the apparent fact that they utter sentences in canonical frames that they have never heard.

(e) Certain aspects of language learning may be relatively dependent on induction, and hence may take a longer time to be mastered than others. We can (and do) interpret C. Chomsky's findings of the relatively slow mastery of certain kinds of verb-intrinsic structural constraints are support for this prediction. In the framework of a hypothesis-testing model of acquisition, certain linguistic features will intrinsically emerge as the 'core' of the language, and others will be modifications of the core by virtue of their less frequent appearance. In this way, frequency of a feature in the child's experience can actually explain some aspects of the order of acquisition of different components.

It should be emphasized that this view does not deny nor minimize the critical computational capacities that underlie the successive structural hypotheses that the child formulates to match the empirical generalizations. The model requires the dynamic interaction of both biological constraints *and* statistical features of experience.

12.2 Language learning is robust despite variation in individual biology and experience

C. Chomsky's work showing that some aspects of syntax are acquired slowly has been taken as support for language-learning theories that emphasize experience and induction primarily. However, the implication of Chomsky's other acquisition work showed that children are remarkably resistant to the effects of variation in experience and individual abilities in learning the basic syntactic forms of language (e.g., C. Chomsky 1986a). That is, despite the effects of experience on the final level of sophistication children can reach, they all learn the essential core of language roughly as fast and well. In this sense, Chomsky's discoveries of how experience can affect vocabulary subtleties that interact with syntactic structures actually served to highlight the fundamental similarities in the formative stages of language acquisition.

These studies highlight the contrast between the acquisition of lexical items and the acquisition of syntactic patterns: since language learning involves both kinds of knowledge, it raises the possibility that there might be profound individual differences in even the early stages of how children approach language acquisition. An interesting possibility lies in genetic variation in the nature of cerebral asymmetries for language. More than half a century ago, A. Luria observed that right-handers with left-handed family members (RH-LHF) appear to have more right hemisphere involvement in language than right-handers with only right-handed family members (RH-RHF): RH-LHF aphasics recover language function faster and more fully, and they show a greater incidence of 'crossed aphasia' (aphasia resulting from right-hemisphere damage) (Luria 1948, 1970). These findings have been replicated (Hutton et al. 1977); recently, a direct fMRI study showed that some RH-LHF people actually show no left-hemisphere asymmetry in brain activation language tasks (e.g., Knecht et al. 2000; Khedr et al. 2002).

It is important that the frequency of RH-LHF people is not small: roughly 40 percent of all undergraduates we have studied are RH-LHF. The question then is, is the difference in brain laterality related to a difference in language processing, and hence in neurological organization for language? The emerging answer is yes. Over many years of research, we have found evidence for a major language processing difference in the relative emphasis on accessing separate words and syntactic patterns between the two biologically coded groups of people: right-handed adults with familial left-handers (RH-LHF) access lexical items more readily than syntactic patterns, while right-handed people with no left-handed family members (RH-FRH) access syntactic patterns more readily. Before discussing a possible explanation for this initially strange finding, here are some published facts that support the generalization (Bever et al. 1989).

(a) RH-LHF subjects read texts faster in a self-paced word-by-word reading paradigm when each button press brings the next word. RH-RHF subjects read texts faster in a self-paced clause-by-clause reading paradigm where each button press brings the entire next clause (Bever et al. 1989).

(b) RH-LHF subjects recognize that a probe word occurred in a just-heard sentence faster than RH-RHF subjects. But RH-LHF people recognize a short phrase that is synonymous with part of a just-heard sentence more slowly than RH-RHF subjects (Townsend et al. 2001).

(c) RH-LHF subjects understand short essays better when the text alternates words between the ears than when the text is presented monaurally to one ear or the other: the opposite obtains for RH-RHF subjects (Iverson and Bever, reported in Bever 1989).

These and many other studies support the behavioral distinction in adult language behavior, with some initial implications for language learning. Decades ago, we speculated that the group differences might result from a general difference in the extent to which relevant neurological areas of the right and left hemisphere are more equipotential in people with familial left-handedness. This relative equipotentiality could then result in more widespread and redundant representation for those aspects of language that are not as computationally demanding as syntactic processing. That is, the neurological representation of lexical items in RH-LHF people can be more widespread and hence offer more separately available representations of words.

Recently we tested this hypothesis in a brain-imaging study and found support for it (Bever et al. in preparation). We gave subjects words in random order, and asked them to think of the input in a logical order. In one case, a syntactic task, the words could be ordered into a sentence ('mothers upset daughters'): in the other case, a lexical task, the words could be ordered by class inclusion ('penny, coin, money'). With fMRI imaging, we found relatively faster processing for the lexical task in relevant areas of the right hemisphere than the left for RH-LHF subjects only (Bever et al. in prep.).

A recent study published evidence that this differentiation has implications for differences in how language is acquired. We looked at the effects of the initial age of exposure to ASL in a large population of deaf people: RH-LHF people show a considerably younger 'critical' age for mastery of ASL, roughly at eight years, while the corresponding age for RH-RHF people is at least twelve years (Ross and Bever 2004). We interpreted this difference as reflecting the usual age at which there is a burst of word learning, roughly between four and eight years—if RH-LHF people depend on word learning as a vehicle for language acquisition, then they would be more dependent on the period of rapid lexical learning. Of course, this would also make predictions for individual differences in the phenomena found by C. Chomsky involving later learning of the syntactic restrictions related to specific lexical items. However, that research remains to be done.

The preliminary conclusion from this aspect of our research is that there may indeed be more than one way to access the information critical for language learning, as implied by C. Chomsky's early studies. Of course, there is the further possibility that the neurological representation of some essential features of language may also reflect the biological variables. But that somewhat radical viewpoint remains to be shown.

12.3 Understanding fluent reading as unifying word sequences into phrases

The preceding discussions revolve around acquiring and using the relation between individual words and syntactic patterns as the child learns language. C. Chomsky addressed the corresponding problem in the acquisition of reading: how does the child pass through from the stage of reading word-by-word to fluent phrase-by-phrase reading in which words are grouped into computationally appropriate phrases? How can we help this process? In this goal, C. Chomsky hit on a paradigm which utilizes the child's relatively natural ability to group phrases in normal intonational units, to carry over to his/her reading ability. Chomsky developed the technique that sounds really simple, but like many such ideas is simple only after one has isolated it: *have the child read and then reread the same passage repeatedly* (C. Chomsky 1976a). On rereadings, the word-by-word reading child naturally begins to impose normal intonation patterns on the now familiar word sequence: Chomsky showed that the result was a notable increase in fluent reading of new material—the experience of discovering how to read known passages with normal grouping may play a role in stimulating grouping strategies in general. One interpretation of this is that the child learns to listen to a 'voice in the head' as s/he reads, after training with the 'voice outside the head'. That is, like St Augustine, the child can discover that it is possible to read 'silently', while hearing and utilizing natural intonation patterns generated internally (Bever 2009).

In our research, we have worked out a different kind of application of phrasing knowledge to the reading process. In our case, we systematically increase spaces between phrases.

Writing systems in general (but not always historically) have specific ways of indicating segmentation in words, thereby solving a major problem of speech comprehension. Today, we take it as obvious that putting a space between words is a good idea. We also rely on punctuation conventions that can mark major phrases from each other. But what about smaller phrasing such as in the previous sentence, as broken up below:

We also rely

on punctuation conventions

that can mark major phrases

from each other

Numerous published studies, starting in the 1960s have shown that indicating phrase boundaries by some marker improves text comprehension (see review in Bever et al. 1990). This fact remained a laboratory curiosity for many years without practical value, for three reasons: identifying 'phrases' had to be done by actual people; implementing the phrase boundary markers was limited to actual characters or extra whole spaces, which looked odd if not downright ugly; the notion of what counted as a relevant 'phrase' was not well understood or uniform. Modern computer and printing techniques have offered solutions to each of these problems. Printers can be controlled to modify spaces and characters in very small increments that do not result in aesthetic disturbance; 'phrases' can be automatically identified by many algorithms; the algorithms themselves provide precise definition of the phrases.[1]

Why should phrase spacing improve reading? On the traditional view, it is because it reveals to the reader how to segment words together and build the correct surface phrase structure as an initial step in reading comprehension: this follows from the traditional view that the first step in comprehension is to determine the correct syntactic structure. But our phrase-formatting algorithms in fact do not find the syntactically correct phrase structure—rather, they isolate those kinds of phrases that are easily detected, based on distributional patterns of words and phrases in actual texts. For example, our algorithm phrases the two sentences below differently, as shown by extra spaces in them. Yet, from a linguistic standpoint, they have identical phrase structures as shown by the bracketed examples.

The large dog was barking at the small cat
The large dog barked loudly at the small cat

(the (large dog)) ((was barking) (at (the (small cat))))
(the (large dog)) ((barked loudly) (at (the (small cat))))

The different analyses assigned by our algorithms follow from the fact that function words such as /was/ and /at/ are easily learned as beginning phrases, while /barked/ is infrequent and will not be recognized by a model that learns phrase boundary cues from texts. This raises a question of theoretical interest: which kind of phrase boundaries are the best to use for implementing segmentation, syntactically correct ones or those assigned by ReadSmart? With linguistic colleagues to help us assign a correct surface phrase structure to standard font-testing texts, we examined this question carefully. We contrasted the comprehension of phrase-spaced formats based on syntactic vs ReadSmart phrases. The results (published) astounded even us: the

[1] We have been testing the efficacy of a set of automatic programs we have written, called ReadSmart™ (now patented), which incrementally increase space size between phrases. We have shown that comprehension of ReadSmart texts and reading speed improve by roughly 15% each, more for poor readers. (See, e.g., Jandreau and Bever 1992; Bever, 2009.) We have also found that the texts are enjoyed more by readers and found to be more convincing. In one semester-long classroom study, readers using the phrase-spaced format earned significantly more honor grades, and had significantly fewer failures than readers using the normal format.

ReadSmart-phrased texts were far easier to comprehend; in fact the syntactic-phrased texts were *harder* to understand than normal untreated texts (Bever et al. 1990).

This follows from the reconstructive view of reading comprehension, as refined by our consideration of details of the analysis by synthesis model of spoken language (see Goodman 1967 for the original proposal of this idea). That model involves two phases of structure assignment, an initial one based on readily available cues and patterns, and a later one based on a full syntactic analysis (Townsend and Bever 2001). Our results show that basing visually salient phrase information on readily available cues leads to the best comprehension, thereby giving empirical support to our claims about initial phases of reading comprehension itself. It also gives support to the larger claim that like speech comprehension, reading involves several stages of extracting structure and assigning meaning. This notion is now receiving empirical support from neurolinguistics studies. For example, Dikker et al. (2009) have shown that an early *sensory* component of evoked potentials (within 100–200 milliseconds) occurs to local phrase violations only in cases involving explicit function words or morphemes. This is direct support for the initial prospective component of comprehension proposed by us and Goodman, and is consistent with our finding that phrase spacing based on morphologically explicit phrases is most effective in improving comprehension.

These considerations offer some perspective on how readers rapidly create a linguistic representation along with the ghostly voice offering an internal rendering of the text. C. Chomsky's technique of having children read and reread text to inculcate phrasing fluency brings out an initial 'voice outside the head'. Our phrase-based formatting technique aids the reader in discovering the corresponding 'voice in the head'.

12.4 Conclusion

Chomsky's sustained work was an early clarion, reminding us of the importance of input and time in language acquisition, even if 'environment', 'input', and 'incremental learning' cannot account for the fundamental linguistic structures that are universally acquired. This and her work on the teaching of reading were early frameworks for a more integrated theory of language behavior involving both words and syntactic patterns, and the applications of linguistics to practical problems.