Some Properties of Language in General, and English in Particular

Howard Lasnik
Distinguished University Professor
Department of Linguistics
University of Maryland
lasnik@umd.edu

Children are complete geniuses at acquiring the ability to speak a language. It is impossible to prevent them from developing language, as long as they are exposed to it. Teaching is irrelevant. In this way, for a child 'learning' language is like 'learning' to walk. It happens completely naturally. (Unlike learning an additional language as an adult. There are some 'geniuses' at this, but relatively few. All young children are geniuses at learning to speak.)

Some examples:
Noone is ever instructed that coordination is unbounded.
(1)   Juan and Bill left
        Maria and Juan and Bill left
        Susan and Maria and Juan and Bill left
        ......
Children just unconsciously know this is possible, even though the language they hear does not, in general, contain more than two or three coordinated elements.
Where does this tacit knowledge come from? Noam Chomsky's answer, which I subscribe to, is that it is one aspect of human nature, part of our shared genetic endowment.
This, of course, makes a strong prediction: That **all** human languages have this property. And the prediction is conformed. Noone has yet found a language (spoken or written) where coordination is required to stop at, say, 4 conjuncts.

Another example of unboundedness:
(2)   Juan left
        Bill thinks [Juan left]
        Maria said [Bill thinks [Juan left]]
        Susan claims [Maria said [Bill thinks [Juan left]]]

This kind of 'embedding' of one sentence inside another is a fundamental property of human languages. No known language (spoken or written) lacks the property. And once again, it is extremely unlikely that children learning language are presented with examples with more than one or two levels of embedding. Knowledge without learning! All human languages can make infinite use of finite means. And all children know this without instruction.

Here's a somewhat more complicated example: Many, many languages have displacement processes, what Chomsky called movement transformations. In such processes, some portion of the sentence is pronounced in one place, yet understood in another, as in

(3)   Who should they hire ~~who~~

Who is understood direct object of the verb, just as *someone* is in:

(4)   You should hire someone

It is striking that such movement transformations are generally 'unbounded', able to displace an interrogative expression a great distance from its understood 'deep structure' position.

(5)   Who will they say [they should hire ~~who~~]
        Who do you believe [they will say [they should hire ~~who~~]]
             ...

However, sometimes even quite short movement is impossible. (I use * to indicate an unacceptable sentence.)

(6)   John will see Maria and someone (else)
        *Who will John see Maria and ~~who~~

This is one among several 'island' constraints discovered by John Robert Ross in the mid-1960s, and investigated by scores of researchers since. Like the others properties considered so far, this one is universal, as far as we know. Noone has found any languages that allow violations of the Coordinate Structure Constraint. Further, children never make mistakes about this. They know, and surely with no instruction, that items cannot be moved out of coordinate structures.

Notice that the example questions just above actually display two movement transformations: 'wh-movement', the one we already discussed; and another one moving an auxiliary verb to the position just in front of the subject. This second process also happens in 'yes-no' questions:

(7)   Should they ~~should~~ hire someone

The proper characterization of 'Subject-Auxiliary Inversion' has been a major research question for over 50 years. A very simple account might be as in (8).

(8)   Interrogative inversion process - structure independent (1st attempt)
        Beginning with a declarative, invert the first and second words to construct an interrogative.

This procedure is about as simple as can be. All that it demands is segmentation of sentences into words, surely a minimal criterion for any syntactic account of any linguistic phenomenon. However, despite its adequacy over a huge range of cases, and its obvious simplicity, (8) immediately fails on very slightly more complicated examples:

(9)   Declarative   Interrogative
   a. The woman must leave.   *Woman the must leave?
   b. A sailor can swim.   *Sailor a can swim?
   c. No boy has read the book.   *Boy no has read the book?
   d. My friend is sleeping.   *Friend my is sleeping?

Compare these with the correct pairings:

(10)   Declarative   Interrogative
   a. The woman must leave.   Must the woman leave?
   b. A sailor can swim.   Can a sailor swim?
   c. No boy has read the book.   Has no boy read the book?
   d. My friend is sleeping.   Is my friend sleeping?

It seems clear that what might be simple in some abstract sense is not necessarily simple for the human language faculty. Though I only look at English here, no human language has yet been discovered that has a process like (8). To accommodate the exceptions to (8), we might complicate it to (11), which correctly covers all of the data so far considered.

(11)   Interrogative inversion process - structure independent (2nd attempt)
         Beginning with a declarative, move the auxiliary verb to the front to construct an interrogative.

This version requires, in addition to segmentation into words, labeling of the categories that the words belong to. But what if there are two auxiliary verbs?

(12)   Declarative   Interrogative
   a. Bill could be sleeping.   Could Bill be sleeping?
                                                  *Be Bill could sleeping?
   b. Mary has been reading.   Has Mary been reading?
                                                  *Been Mary has reading?
   c. Susan should have left.   Should Susan have left?
                                                  *Have Susan should left?

Modifying the basic rule one final time, we might try:
(13)   Interrogative inversion process - structure independent (3rd attempt)
         Beginning with a declarative, move the first auxiliary verb to the front to construct an interrogative.

This works for the overwhelming majority of cases found in any corpus, and likely for all of the cases that the child learning the language will encounter. Thus it is particularly striking that it too fails.

(14)   Declarative   Interrogative
   a. The man who is here can swim.   *Is the man who here can swim?
   b. The boy who will play has left.   *Will the boy who play has left?

 For these examples, fronting the second auxiliary verb gives the correct form:

(15)   Declarative   Interrogative
  a. The man who is here can swim.   Can the man who is here swim?
  b. The boy who will play has left.   Has the boy who will play left?

Again, we are not dealing with a peculiarity of English. No known human language has a transformational process that would produce pairings like those in (14). Further, the incorrect forms in (14) (like the incorrect forms in (9) and (12), are not attested in any of the voluminous literature documenting the errors young children make in learning their language. In fact, experiments specifically designed to determine whether such incorrect forms are possible for children (even 3-year old children) have invariably shown that they are not. The seemingly simple structure independent computational operations in (8), (11), and (13) are evidently not available to the human language faculty.

 The right generalization is a priori much more complicated, relying on structured hierarchical organization:

(16)   Interrogative inversion process - structure dependent
         Beginning with a declarative, move the first auxiliary verb following the subject to the
         front to construct an interrogative.

Does the child have evidence that would determine the correct process and exclude the incorrect ones? Example dialogues like those in (15) surely are not uniformly available to the child learning language. Even more significantly, even if the child is exposed to (15), that alone does not rule out the other possibilities as options. This line of reasoning is a model of the classic 'poverty of the stimulus' argument for innateness of some aspects of language ability.

So languages are similar in important respects. But of course they also differ. And those of us involved in language instruction, especially second language instruction, are particularly sensitive to the differences. Further, even internal to individual languages, there are different dialects, and different registers. Often written language has its own special rules that are not necessarily part of the spoken language. Some very familiar ones are listed here in (17).

(17)   a.   "Don't end a sentence with a preposition"
        b.   "Don't split an infinitive"
        c.   "Use *whom* instead of *who* in certain circumstances"
        d.   "Don't use contractions [!]"

It is very interesting that students have been receiving explicit instruction on these rules for centuries, but they consistently fail to learn them. (This is just the opposite of most linguistic properties, where there is no explicit instruction and perfect learning.) The spoken language does not follow these rules, and, you may be surprised to hear, probably never did. 'Stranding' a preposition has been possible since Old English. At least since Shakespeare's time, we find *who* in sentences like (18).

(18)   Who will you appoint

Where do these rules come from? For the most part. they can be traced back a few centuries to language teachers and language 'mavens' concerned about how English was degenerating (Sound familiar?) and hoping to return it to something more like its pristine form (assumed to be Latin). Well, (17)a-c were indeed part of Latin. But that is no reason to assume they are or were part of English. Interestingly, attempts to follow the rule in (17)c lead to more subtle 'mistakes'. In transformational terms, the 'circumstances' mentioned in (17)c are when in deep structure the interrogative is a direct object. But examples like (19) occur frequently in the writing (and speech) of those trying to follow the rule.

(19)   Whom do you think will win the race

Note that in (19) the interrogative word was a deep subject, not an object, as seen in (20).

(20)   You think [someone will win]

This sort of hypercorrection often happens when speakers (or writers) attempt to follow a prescriptive rule that is not truly part of their language.

Spoken languages have rules of their own, almost always unconscious and 'learned' with no training. Some particularly striking examples are based on 'violations' of (17)d. Contraction is very widespread in spoken English. One instance, I will represent as *wanna*, the contraction of *want* when it is immediately followed by *to*:

(21)   You want to visit someone → You wanna visit someone

Is this just a lazy anarchic way of talking? Actually not. The process obeys a highly abstract constraint. First, consider another 'good' case:

(22)   Who do you want to visit → Who do you wanna visit

But now suppose it were not the object of *visit* that you were interested in, but the subject:

(23)   You want someone to visit you
(24)   Who do you want to visit you → *Who do you wanna visit you

Superficially, there is no relevant difference between (24) and the other *wanna* examples, yet native speakers overwhelmingly reject this example as odd, while accepting the others. The difference shows up not on the surface, but in deep structure:

(25)   You want [who to visit you]

Contraction of *want* and *to* is possible only if these two words are adjacent not just in surface structure, but also in deep structure. Totally uninstructed, contraction users are thus performing a highly sophisticated bit of unconscious mental gymnastics, and behaving in a highly rule-governed way (even though noone but a linguist would ever have articulated the rule). Another familiar contraction process, reduction of *is* to *'s,* displays a similar constraint. Look at

the following patterns:

(26)   John is here → John's here

Now suppose I were wondering about John's location:

(27)   I wonder where John is

Speakers have a robust judgment that contraction is not possible in this case:

(28)   *I wonder where John's

And notice that the constraint is not as simple as (29).

(29)   Don't contract *is* if it comes at the end of the sentence.

The inadequacy of (29) is illustrated in (30).

(30)   I wonder where John is now → *I wonder where John's now

The correct generalization, as in the case of *wanna* contraction, relies on deep structure:

(31)   Don't contract *is* if an item followed *is* in deep structure, but does not in surface structure.

And once again, noone ever receives instruction or training on this constraint, just as noone is ever conscious of the fact that he or she is 'obeying' it.

Needless to say, this is only a tiny sample of what linguists have discovered about the human capacity for language. Further phenomena from English, and from scores of other languages, have been analyzed in the extensive literature of theoretical linguistics. I hope that even the tiny sample I have been able to provide here give you some idea of the extent to which human children are geniuses for acquiring language.

Some Readings

Chomsky, Noam. 1988. *Language and problems of knowledge*. Cambridge, Mass.: MIT Press.
Cogswell, David. 1996. *Chomsky for beginners*. New York: Writers and Readers Publishing, Inc.
O'Grady, William, Michael Dobrovolsky, and Mark Aronoff. 1989. *Contemporary linguistics: An introduction*. New York: St. Martin's.
Pinker, Steven. 1994. *The language instinct*. New York: Harper Perennial.
Towell, Richard, and Roger Hawkins. 1994. *Approaches to second language acquisition*. Clevedon, Avon: Multilingual Matters Ltd.