

THE ECOLOGY OF LANGUAGES*

(A Ecologia das Línguas)

David LIGHTFOOT
(Georgetown University)

ABSTRACT: This paper discusses the life-cycle of languages: languages die, new languages are born, and languages undergo radical changes in form and structure. This paper considers three changes in the history of English: loss of split genitives, introduction of new inflectional categories, and loss of verb movement. The proposal is that these changes are the result of children's reanalysis during language acquisition, based on the interaction between primary linguistic data and universal grammar. These processes of I-language reanalysis lead to the gradual emergence of new E-languages.

KEY-WORDS: language change; language acquisition; history of english; syntax.

RESUMO: Este artigo discute o ciclo de vidas das línguas. Algumas línguas morrem, outras nascem e outras sofrem profundas mudanças na forma e na estrutura. Consideraremos aqui três mudanças na história do Inglês: perda de genitivo cindido, introdução de novas categorias flexionais e perda de movimento do verbo. De acordo com a nossa proposta, essas mudanças são resultado de reanálise durante o processo de aquisição devido à interação entre os dados lingüísticos primários e a gramática universal. Esses processos de reanálise na linguagem conduzem o nascimento gradual de novas línguas.

PALAVRAS-CHAVE: mudança lingüística; aquisição; história do inglês; sintaxe.

* This paper is dedicated to Lucia Lobato, who was a syntactician with deep political concerns for the indigenous people of Brazil and their languages. The paper seeks to bring together the familiar concerns of syntacticians with concerns about the dramatic changes in the distribution of the world's languages. Lucia was also concerned with how linguists might better communicate their work to non-professionals and I have written this paper for a wider audience than syntacticians usually address, perhaps even being understandable in parts for her sons Leandro and Thiago. The paper has been written while I have been working at the National Science Foundation but the views expressed here are my own and not those of the Foundation.

Introduction

Languages come and languages go. A hundred years ago Hebrew was not spoken much outside synagogues. Jews in Europe had their own languages, Yiddish in the north and Ladino by the Mediterranean, but Hebrew was a language used mostly in religious contexts, like Latin and Sanskrit. That Hebrew, Biblical Hebrew, has now been transformed and a new Hebrew language is spoken widely in Israel and beyond, including by very young children. Once Latin was spoken by powerful people throughout much of Europe but now we have the new languages of Portuguese, Spanish, French, Sardinian, Italian, and Romanian, new Latins, all descended from earlier forms of Latin and not existing as separate languages 2,000 years ago. In 1950 there were a hundred or so languages spoken in the Arctic but several languages of forty years ago now have no speakers at all and many more have numbers that indicate that they will not survive more than a few years (Krauss 1988).

In fact, languages tend not to be around very long. People know this if they have gone back and read some Shakespeare. One can read it but it's a little odd in places and there are words and phrases that an untrained reader does not understand, because the language has changed. We have different languages now. Go back two hundred years further and it is harder to read Chaucer, because his language is more different. If one goes back to Old English *Beowulf*, one may as well be reading German. English has changed at a number of points over the last 1,000 years. There is no single entity English existing over that period. There is Old English, Middle English, Early Modern English, Present-Day English, many varieties of each, many other languages in between, and the dividing lines are murky and imprecise. Like French and Portuguese, English didn't exist in any recognizable form 2,000 years ago.

The diversity among the world's languages has emerged over a period of something between 50,000 and 100,000 years. The human language capacity seems to have evolved once, in East Africa, and not long ago in evolutionary terms. In addition, we know that thousands of languages came and went before the modern ones, now extinct. We have direct evidence for classical Greek, Sanskrit, Old English and so on. Also there are languages that must have existed even though we have no direct evidence for them: for example, the common ancestor language of the

modern Indo-European languages from Icelandic to Hindi, what we call Proto-Indo-European, and the ancestor of the Semitic languages from Hebrew to Moroccan Arabic, Proto-Semitic.

1. How do languages rise and fall? What lies behind the ecology of languages?

When one looks across a field towards the edge of a forest, sometimes one sees a deer only when it moves, when there is a change. As new languages emerge, we can learn things about the genetically based human language capacity, about what is represented in the brain when a person speaks some form of Japanese or English. Sometimes change in languages can tell us about the biological capacity that humans have for language and that constitutes the interest in language change for some linguists.

Before we consider how new languages are born, let us think about the other end of their lives, their death. To do that, we must ask how many languages there are.

1.1. *How many languages?*

SIL, the old Summer Institute of Linguistics based in Dallas, counts languages and their fascinating website Ethnologue.com tells us that 6,912 languages were spoken in 2005. If there are over six billion people in the world, that would mean an average of about one million speakers for each language, but in fact there is enormous variation. There are 239 languages in Europe, 2,092 in Africa, and Papua New Guinea, with a population of only 5.5 million, has an astonishing 820 languages. In Europe, English, French, German and Italian have many millions of speakers, Norwegian has 4.5 million, Breton and Basque just over half a million each, and all the varieties of Saami, spoken in northern Norway, Sweden and Russia, total about 20,000. Some of Ethnologue's languages are sign languages, some with small communities of a few hundred users, and with the same fundamental properties as oral languages, just occurring in a different modality.

One should be wary of these numbers and a lot depends on how one counts. Other counts show over 300 languages spoken in London alone. Are Norwegian and Swedish distinct languages? If two teachers from Oslo

and Stockholm meet at a convention, they may have quite satisfactory conversation. However, somebody drew a red line down through the middle of Scandinavia and people to the left of that line are said to be Norwegians and people to the right are Swedes, speakers of different languages, despite the fact that they communicate successfully over a large domain. Conversely, we in the West tend to think of Chinese as one language. We count it as one language because it is one country, but there are eight varieties of Chinese, as different as French and Italian. When Yugoslavia was one country, no matter how fractious, there was a major language Serbo-Croatian; after Yugoslavia came apart, we now have Serbian, Croatian, and Bosnian. Nothing changed linguistically, only the political boundaries.

What is the difference between a language and a dialect? Often we can do no better than Max Weinreich, who famously declared in 1945: A language is a dialect with an army and a navy. There are no good definitions.

From another perspective, one could count differently and say that there is just one language, Human. Humans have communication systems that are radically different from what one finds in other species. Indeed, when human ethologists look at other species, they discover how herring gulls or honey bees communicate. Honey bees indicate the direction and distance to sources of nectar by facing in the appropriate direction and wiggling their rear ends at different rates. It turns out that there are “dialect” differences and Austrian bees use slightly different frequencies than Swiss bees, but those differences are trivial matters of fine-tuning and are as nothing compared to the differences between honey bees and herring gulls. That is how a Martian ethologist might see the differences between English and Italian, as matters of fine-tuning.

That is an old view, going back at least to Wilhelm von Humboldt (1836/1971). He thought that the form of all languages must be fundamentally identical - they differ as physiognomies differ.

Another way of counting would say that there are six billion languages, one per person. Everybody's language is as distinct as their thumbprint and, indeed, that individuality has been used for forensic purposes. It takes my mother only a second to know whether it is one of my brothers or me on the telephone. Personally, I speak David Lightfoot, and my sons speak Eric Lightfoot and Alex Lightfoot.

1.2. *Language death*

However we count, Ethnologue tells us that of the 6,912 languages, 497 are spoken only by small numbers of elderly people. “Small numbers” means numbers like 1, 5 or 9. Those languages will die out in a few years, as those people die. Ethnologue estimates that another 3,000 languages will not be spoken within a generation, and many more by the end of the century. We are looking at a major shift in human history.

North America is typical. There were approximately 700 languages in 1492. Now there are about 165. 75 are spoken by a handful of older people and only a dozen or so by large numbers of people.

There have been earlier periods of mass extinction. As Latin spread through Europe, we have evidence of languages in Italy and elsewhere that did not survive. About 8-9,000 years ago, when people gave up foraging for farming and began to live in larger and less isolated communities, there may have been a significant reduction in the number of languages but we have no direct evidence. Whatever the comparability of previous extinctions, we are in a dramatic development in human history.

In the face of this, the National Science Foundation, in collaboration with the National Endowment for the Humanities, has undertaken a major new project to document endangered languages and there are comparable efforts in Australia, Germany and the UK.

Humanists and scientists have different, complementary reasons to document languages. Humanists know that languages encode elements of culture. In *Othello*, Shakespeare wrote

If I do prove her haggard,
 Though that her jesses were my dear heart-strings,
 I'd whistle her off, and let her down the wind,
 To prey at fortune.

In Shakespeare's time falconry was an important element of English culture. Everybody knew that jesses were the little strips of leather tied to the ends of falcon's legs. This kind of thing is encoded in languages and humanists document information about cultures, medical practices, myths and poetry revealed in language.

Scientists, on the other hand, are interested in the diversity of languages because of what it reveals about the potential of the human brain. They do a kind of Mendelian genetics, reasoning from observed diversity to necessary internal properties, as Gregor Mendel observed his pea plants and deduced internal factors, what he called “gens,” which determined aspects of their growth and development.

For example, many languages are like English in having a basic Subject-Verb-Object order. The verb phrase has the verb preceding its complement:

- (1) Kim could [_{VP} visit Berlin]

Many other languages are like German in having basic Subject-Object-Verb, the verb following its complement:

- (2) Kim kann [_{VP} Berlin besuchen]

And many languages are like Welsh and have basic Verb-Subject-Object:

- (3) Fe welais I Megan
PRT saw I Megan
 ‘I saw Megan’

Each of these word-order types comes with harmonic properties of various kinds.

A very unusual word order is Object-Verb-Subject. An endangered language in the Amazon has this word-order type: Piratapuyo. We need to record and analyze this language before we lose crucial information about the Object-Verb-Subject word-order type, about what the harmonic properties are. Kristine Stenzel at the University of Colorado is doing just that.

The Documenting Endangered Languages project exploits new computational facilities to record people speaking, where video recordings display muscle movements and even nerve firings. Information collected goes on to the web and is immediately available to other researchers for checking and analysis. Scientists funded by the National Science Foundation have developed an electronic template for the recording of endangered languages. This is E-MELD, the Electronic Metastructure for Endangered Language Data, and GOLD, the General Ontology for Linguistic Descriptions.

This is a far cry from the way we used to do things. We would train graduate students to go out and live with remote people, take notes, learn what they could, and in due course publish significant findings in one of the journals. Only selected data would be published, the data needed to substantiate the theoretical points, but now the descriptions themselves and all data collected are publicly available immediately and checkable.

Again, the goal is to document the languages. It is not the business of the NSF nor of linguists to persuade people to keep speaking their mother's native tongue. People have all kinds of reasons for adopting one language over another. However, it has not escaped notice that recording and understanding a language, providing a writing system, even a computer-friendly writing system, increases the usability of a language and therefore increases its chances of survival. Bear in mind that of Ethnologue's 6,912 languages, only a few hundred are actually written. The vast majority exist only in oral form and many have small numbers of speakers, and that limits their usability.

There are some good books about languages dying out. Nettle & Romaine (2000) point out that about 4% of the world's population speaks at least 60% of its languages. Also 'the greatest linguistic diversity is found in some of the ecosystems richest in biodiversity inhabited by indigenous people;' that biodiversity is encoded in the languages and that is what humanists are so keen to document.

Dying languages look different from other languages. They vary more from speaker to speaker and they are less stable, more subject to rapid change over short periods of time. This is shown in Harrison (2007), which is the basis for a movie *The Linguists*, shown at the 2008 Sundance Film Festival, to great acclaim.

When languages die, we can in principle identify the exact date, the date on which the last native speaker died. So Cornish became extinct as a native language with the death of Dolly Pentreath in Mousehole in 1777, perhaps on 17 October at 11pm. Other people have used forms of Cornish and revivalists have invested energy in teaching the language, but nobody has acquired it as a first, native language for more than two hundred years, since Dolly Pentreath. So with the 497 languages that Ethnologue identifies as about to die; if we watch carefully, we may be able to document the precise moment at which they die as native languages.

1.3. *Language birth*

Language birth is different and we cannot point to a precise moment when a new language emerges as a native language, and the reasons are interesting. Two thousand years ago, Latin was spoken differently around the rivers Tagus, Seine and Tiber and eventually distinct new languages emerged there: Portuguese, French and Italian. When these different forms of Latin, different dialects, became distinct languages, nobody knows. We don't have a clear definition between dialects and languages, as Max Weinreich pointed out.

During the colonial period, often primitive and very limited communication devices would develop in trading contexts, so called pidgins. Sometimes those pidgins became part of what children heard and affected the development of new systems in those children. The first languages that emerged in that way were creoles and, if we had good records from those contexts, we might be able to pinpoint the time when children first converged on different, pidgin-influenced systems. In fact, there is a recent, spectacular development in Nicaragua, which I will discuss later, where we know quite a lot. This involves a signed language emerging among the deaf community there in a creolization process and the emergence of that language adds support to the view that signed languages manifest the same principles and parameters as oral languages.

1.3.1. *So how do new languages emerge?*

To understand the emergence of new languages, we need to distinguish external language out there in the world and we need to think of internal languages that develop in children's brains. This distinction between external and internal languages has been invoked by Noam Chomsky (1986) but, like many modern ideas, it can be found in the writings of Wilhelm von Humboldt. He distinguished the languages of nations from the particular languages of individuals (Humboldt 1836/1971).

External languages like English represent ill-defined social notions, not what children acquire. For example, is *She might could see it* a sentence of English? Yes, if one is raised in Tennessee or Arkansas but not if one was raised in Cornwall. Conversely *Bin her happy?* exists as a sentence of English

for people raised in Cornwall. No child acquires English as such but rather her own particular, private language, sometimes a new language. We will see that there is a genetic component that interacts with environmental facts; nature combines with nurture.

2. Our language capacity

The first thing to recognize is that everybody's capacity ranges over infinity. One's language capacity is represented in one's brain and must be finite but there is an infinite number of things one can say and understand. That is because every language has three recursive devices that permit sentences to go on *ad nauseam* or even *ad mortem*. We can use sequences of relative clauses (cf. (4)) and in English we even do this to bore young children to sleep. Second, we can use one complement clause after another (cf. (5)) and, third, we can coordinate, for as long as one pleases (cf. (6)). And one can combine these three devices (cf. (7)).

– *Relativization:*

- (4) This is the cow that kicked the dog that chased the cat that killed the rat that caught the mouse that nibbled the cheese that lay in the house that Jack built

– *Complementation:*

- (5) Ray said that Kay said that Jay thought that Fay said that Gay told me that Clay reported that there was hay on the way

– *Coordination:*

- (6) Ray and Kay went to the movie and Jay and Fay to the store, while Gay and May and Clay worked where Shay and Jack were watching, but Zach and Mack and Shaq slept

– *Combined:*

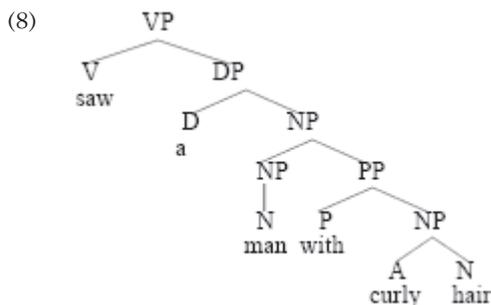
- (7) Fred and the woman I met in New York thought that Chicago was hot

That's a technical way of showing something fundamental: virtually everything we say is novel. It may be quite trivial, *I think that the Ivory Coast will give Argentina a tough time tonight in Hamburg*, but we say it because we want to express that thought, not because we heard somebody else say

this some time ago. In that way, the language capacity is infinitely creative and that makes humans different from other animals. This isn't learned (no child hears a sentence of indefinite length; they all end) but it is built into the human language capacity, which is fine-tuned differently in Tokyo and Toronto.

Another thing that isn't learned, which is built into the system, is that language, everybody's language, is compositional, consisting of units consisting of smaller units. In an expression *I saw a man with curly hair*, *man with curly hair* is a unit but *man with* is not. What that means is that units may undergo computational operations and non-units may not. We'll see what that means in a moment.

– *Compositional:*



Things get more interesting and people know many things subconsciously about their language that they are not aware of and for which they had no evidence in their childhood experience. This is where we see the genetic component being revealed through what is called the “poverty of the stimulus,” a central element of the way that many linguists think.

Consider some of the secret things that English speakers know subconsciously and how they are so much smarter than they think.

2.1. *What children learn and what they don't learn*

English speakers may say *Kim is taller* or *Kim's taller*, with *is* reduced. One can think of this as an operation *is* → 's. Children hear both the full

and reduced form and can learn the operation on exposure to external data.

Now comes the poverty of stimulus problem. In (9) the underlined *is* never reduces but this cannot be learned exactly.

- (9) a. Kim's taller than Jim is
 b. I don't know what the problem is with this solution
 c. I wonder where the concert is on Wednesday
 d. She asked what that is up there

The stimulus that children have does not convey this kind of negative information. Children hear things but they are not instructed in what does not occur and therefore they do not LEARN the limitation. Helicopter parents may try to correct the occasional *goed* or *taked*, but they don't tell children that a reduced *is* does not occur here or there. That is partly because they don't know and partly because children do not misuse the reduced and full forms and there is no need for correction – a lot of ingenious experimental work has shown how sophisticated children's language capacities are (Crain & Thornton 1998).

This seems mysterious but we now understand the mystery and understand some aspects of the brain basis of language. Children are exposed to simple speech, what linguists call "primary linguistic data." That acts as a triggering experience and the initial genetic inheritance (what linguists call Universal Grammar) blossoms into a mature phenotypical capacity (a biological "grammar"), depending on whether the children are raised in Toronto or Tokyo, in Bali or Baltimore. Nurture interacts with nature. Humans have very different communicative possibilities from other species, simply by virtue of the biological fact that they are humans, and they are influenced by the external language around them. The goal is to tease apart internal and external factors, the contributions of the genetic inheritance from the contributions of environmental factors. We find that both internal, genetic factors and external, environmental elements shape a child's internal language system and that much of what children come to know is intrinsic, not learned.

Primary Linguistic Data (Universal Grammar —> grammar)

Triggering experience (genotype —> phenotype)

3. The poverty of the stimulus problem

3.1. "That" deletion

Take another poverty-of-stimulus problem. English embedded clauses may be introduced by a sentence introducer, a word like *that*.

- (10) a. Peter said [that Kay left]
 b. The book [that Kay wrote] arrived
 c. It was obvious [that Kay left]

Those words may be omitted and there may be an operation *that* \rightarrow \emptyset . Again, this is learnable: children hear the full forms (10) and the corresponding reduced forms (11). French children and Dutch children do not hear equivalent reduced forms and learn no comparable operation.

- (11) a. Peter said [Kay left]
 b. The book [Kay wrote] arrived
 c. It was obvious [Kay left]

Here is the poverty of stimulus problem. The operation deleting *that* may not apply to (12) and the equivalent forms without *that* are not what English speakers would say (13).

- (12) a. Ray said yesterday in Chicago [that Kay had left]
 b. The book arrived yesterday [that Kay wrote]
 c. Fay believes, but Kay doesn't, [that Ray is smart]
 d. [that Kay left] was obvious to all of us
- (13) a. *Ray said yesterday in Chicago [Kay had left]
 b. *The book arrived yesterday [Kay wrote]
 c. *Fay believes, but Kay doesn't, [Ray is smart]
 d. *[Kay left] was obvious to all of us

Again, children have NO EVIDENCE for this limitation. They sometimes hear forms with *that* (10), sometimes without *that* (11), but they are not informed and have no evidence that the forms of (13) do not exist. Somehow they deduce that limitation, and we now know how.

Virtually every surface generalization about language breaks down in this way and we understand why: it is because of the internal factors that influence our language.

Here is what is involved. There is a simple principle of the human language capacity, part of our genetic inheritance and encoded somehow in our genetic material. The crucial information is:

**Something is deleted if it is (in) the complement
of an adjacent, overt word.**

In the simple forms, the bracketed clause completes the meaning of *said*, *book* and *was obvious*. *That* is adjacent to those words, is in the complement, and so may be deleted.

- (14) a. Peter said [~~that~~ Kay left]
 b. The book [~~that~~ Kay wrote] arrived
 c. It was obvious [~~that~~ Kay left]

The forms of (13) are not English, because the bracketed clauses of (12) do not complete the meaning of the adjacent *Chicago*, *yesterday* or *doesn't*. And in (12.d) there is nothing preceding it. Therefore, in these cases *that* may not be deleted.

That simple principle of our language capacity solves this poverty-of-stimulus problem and accounts for a lot of apparently odd things.

3.2. *Wh* questions

English speakers form questions by displacing the interrogative word to the front of its clause and deleting the original element in the position in which it is understood; there is an operation *Copy wh-*.

It doesn't have to be that way. The equivalent word in Chinese is not displaced. One says *Zhangsan jiao shui?*, literally "Zhangsan teaches who" and the "who" word stays in the position in which it is understood. Similarly in Japanese with its object-verb order: *Taro wa dare o oshiemasu ka?*, literally "Taro who teaches?". So the English way of asking questions has to be learned.

For English speakers, the simple expression *Who did Jay see?* has a representation in which *who* is copied to the front of the clause and the original *who* is deleted (15). It is the complement of *see* and the deletion conforms to our principle.

- (15) Who_i did Jay see ~~who_i~~?

The slightly more complex *Who did Jay say that Fay saw?* has the representation (16): *who* is copied to the front of its own clause, then to the front of the next clause up. Each element is deleted in conformity with our principle: Each of the *whos* with the strike-through is the complement of the word immediately to the left or in the complement of the word immediately to the left and therefore gets deleted.

- (16) Who_i did Jay say [~~who_i~~ that Fay saw ~~who_i~~]?

3.3. Gapping verbs

Now consider another operation, whereby the second of two identical verbs may be deleted: *Gap V*. There may be an understood, empty verb in the second part. So alongside (17a) we find (17b), perfectly normal, comprehensible speech, which has a representation with an empty verb (17c).

- (17) a. Jay introduced Kay to Ray and Jim introduced Kim to Tim
 b. Jay introduced Kay to Ray and Jim Kim to Tim
 c. Jay introduced Kay to Ray and Jim _ve Kim to Tim

(18a) is another example of verb gapping, a good, A+ sentence for all of us. However, we do not gap a verb AND delete the sentence introducer *that* (18b), which would have the representation (18c). Again, our principle has the explanation: *that* may not delete at the front of its clause (hence boldface), if it is not (in) the complement of an adjacent, overt word. Here the adjacent verb is not overt.

- (18) a. Fay said Ray left and Tim that Jim stayed
 b. *Fay said Ray left and Tim Jim stayed
 c. Fay said Ray left and Tim _ve [that Jim stayed]

Similarly, we do not gap verbs AND displace some *wh*- elements. (19a) has a representation (19b). Each deleted *who* conforms to our principle: it is adjacent to an overt verb and either the complement of that verb or contained in its complement. However, if we gap the second verb, we do not say (19c), which would have the representation (19d). There the boldface *who* fails to delete, because there is no adjacent, overt verb.

- (19) a. Who did Jay think Kay hit and who did Jim think Kim hit?
 b. Who_i did Jay think [~~who_j~~ Kay hit ~~who_i~~] and who_j did Jim_v think [~~who_j~~ Kim hit ~~who_i~~]?
 c. *Who did Jay think Kay hit and who did Jim Kim hit?
 d. Who_i did Jay think [~~who_j~~ Kay hit ~~who_i~~] and who_j did Jim_ve [~~who_j~~ [Kim hit ~~who_i~~]]?

Nor (20a), which would have the structure (20b), where the boldface *which woman* fails to delete, because there is no adjacent, overt verb.

- (20) a. *Which man did Jay introduce to Ray and which woman did Jim to Tim?
 b. Which man_i did Jay introduce **which man_i** to Ray and which woman_j did Jim_ve **which woman_j** to Tim?

Things are getting a bit complex.¹ However, the point is that nothing complex is learned by children. One's language is a complex system but the complexity can be understood in terms of an interaction between some simple principles at the genetic level and some simple generalizations that are triggered in children on exposure to certain kinds of speech.

3.4. Back to reducing "Is"

Now we can return to our first example and see that the same deletion principle accounts for the distinctions noted. A reduced *is* is absorbed into the preceding word and becomes an integral part of it. It is pronounced differently, depending on the last segment of the word it attaches to, as a voiceless 's' in *Pat's*, as a voiced 'z' in *Doug's* and as an extra syllable in *Alice's* (21a). Plural markers, possessives and third person singular forms of verbs (21b-d) show the same differences.

- (21) a. Pat's happy, Doug's happy, and Alice's here
 b. Cats, dogs, and chalices
 c. Pat's dog, Doug's cat, and Alice's crocodile
 d. Commits, digs, and misses

Now we can see why we don't reduce *is* in certain contexts. (22a) has a representation (22b), where *tall* is deleted, adjacent to the verb *is*, of which it is the complement. However, (22c) does not exist, because *tall*

¹ For more details, see Lightfoot (2006b).

has no adjacent verb. The representation would be (22d) and the reduced *is* has been absorbed into *Tim* and therefore is no longer a separate, complement-taking word to license the deletion of *tall*.

- (22) a. Kim is taller than Tim is
 b. Kim is taller than Tim is ~~tall~~
 c. *Kim is taller than Tim's
 d. Kim is taller than Tim's tall

Similarly one finds (23a), which has the representation (23b) and *what* deletes, licensed by the adjacent verb whose meaning it completes. On the other hand, we do not have (23c), which would have the representation (23d), where the reduced *is* has been absorbed into *that* and cannot license the deletion of *what*.

- (23) a. I wonder what that is up there
 b. I wonder what that is ~~what~~ up there
 c. *I wonder what that's up there
 d. I wonder what that's ~~what~~ up there

Contrast (24a), where the deleted *where* is licensed by the adjacent *is* (24b). But not (24c), where *where* has no adjacent verb and may not delete (24d).

- (24) a. I wonder where the concert is on Wednesday
 b. I wonder where the concert is ~~where~~ on Wednesday
 c. *I wonder where the concert's on Wednesday
 d. I wonder where the concert's ~~where~~ on Wednesday

Again, nobody taught us these distinctions, there was no evidence for them in what we heard as children, and there was no learning. They are just part of our language capacity, a function of a piece of genetic information that linguists have discovered.

3.5. Deleted verb phrases

English allows deleted verb phrases in many different contexts. They are understood by reference to another verb phrase, *leave for Rio* in these examples.

- (25) a. Max left for Rio, although Mary didn't _{vp}[leave for Rio]
 b. Although Max couldn't _{vp}[leave for Rio], Mary was able to leave for Rio
 c. A: Susan left for Rio B: Yes, but Jane didn't _{vp}[leave for Rio]
 d. The man who left for Rio knows the woman who didn't _{vp}[leave for Rio]
 e. Don't _{vp}[leave for Rio]!

However, deleted, empty verb phrases occur only where they are the complement of an adjacent modal auxiliary verb, *didn't* or *couldn't* in these examples.

English allows quantifiers like *all* and adverbs like *often* or *certainly* to occur to the left or right of a modal auxiliary (26a, b). Now notice that we have (26c) but not (26d). In (26d) the verb phrase is not the complement of the adjacent word, *all*, cannot delete so the representation crashes. Similarly, we have (26e), where the deleted VP is the complement of *had*, but not (26f).

- (26) a. We all had left/we had all left
 b. We often have run/we have often run
 c. They denied reading it, although they all had _{vp}[read it]
 d. *They denied reading it, although they had all _{vp}[read it]
 e. They denied reading it, although they often/certainly had _{vp}[read it]
 f. *They denied reading it, although they had often/certainly _{vp}[read it]

Our principle also explains the difference between (27a), where the deleted verb phrase is the complement of *has*. But not (27b), where the reduced *has* has been absorbed into *John* and cannot license a deleted verb phrase, causing it to crash.

- (27) a. I haven't read it, but John has _{vp}[read it]
 b. *I haven't read it, but John's _{vp}[read it]

3.6. *What children learn*

In short, we have five operations, each learnable by children on exposure to the relevant sentence-type:

- (28) a. that → 0 (e.g. *Peter said Kay left*)
 b. copy wh- (e.g. *Who did Jay see?*)
 c. gap V (e.g. *Jay saw Ray and Jim Kim*)
 d. is → 's (e.g. *Kim's happy*)
 e. delete VP (e.g. *Mary didn't*)

And we have one simple principle of the human language capacity, governing how elements are deleted. That principle is the source of many distinctions. The interaction between nature and nurture captures the immense complexity of a person's language capacity, revealing distinctions that most of us are not aware of.

4. New languages

Now we can think about the birth of new languages. Children acquire their internal language under the influence of their biology and their environment, as we have seen. The environment means language out there, the kinds of things that children hear. Sometimes the environment may shift a little, yielding new primary data, and then there may be new internal languages. That is when we have bumpy changes and new systems emerge, new internal languages.

Primary Linguistic Data (Universal Grammar —> grammar)

We noted earlier that people's speech is individual and unique; people have different systems and furthermore they use their systems differently. For example, people differ in how and how often they use tag questions like *It is raining, isn't it?*, in how they use the topic constructions favored by sports commentators: *Taylor, he throws the ball down the middle*. People's use of their system varies, sometimes just randomly and sometimes there are statistical tendencies we can identify.

Because of varying use, all children have different experiences even in relatively homogeneous language communities and hear different things around them with different frequencies. It is those experiences of external language, language out there, that trigger the development of a child's internal language. Since no two children have exactly the same experiences, there is always the possibility of new internal languages emerging.

Consider some well understood examples from the history of English (drawn from Lightfoot (2006a)), structural shifts that have made Shakespeare sometimes difficult for modern Londoners to understand, Chaucer still harder without special training, and *Beowulf* as incomprehensible as German. Each of these shifts has made up a new language.

4.1. *Split genitives*

Middle English, Chaucer for example, shows “split genitive” expressions like (29), which came to cluster differently. They became modern (30), what are sometimes called “group genitives.”

- (29) a. The clerk’s tale of Oxford
 b. The wife’s tale of Bath
 c. King Priam’s son of Troy
- (30) a. The clerk of Oxford’s tale
 b. The wife of Bath’s tale
 c. King Priam of Troy’s son

That was a result of the loss of genitive cases on nouns. That genitive ending was *-es* and it came to be construed as a clitic, attached not to a noun but to a whole noun phrase like *the clerk of Oxford*, *King Priam of Troy*. Case endings are part of nouns but clitics may attach to noun phrases; so the clitic *'s* needed a noun phrase to attach to. Universal Grammar provides the categories available for particular systems acquired by children, along with their definitions and properties, categories like *CASE* and *CLITIC*, and changes like this enable us to see the effects of a change in category and therefore to learn something about the nature of the categories and what triggers them.

This loss of genitive cases was part of a much larger, dramatic shift in Middle English: the massive loss of morphological endings. Verbs, for example, had different forms depending on their tense and person. So with a first person subject, I, a verb might be *fremme*, but you *fremst*, he or she *fremþ*, they *fremmaþ*. Another verb was I *seo*, you *siebst*, she *siehb*, they *seop*. And another verb was *ride*, *ritst*, *ritt*, *ridaþ*. The past tense of that verb was *rad*, *ride*, *rad*, *ridon*; and the past tense of the first verb was *fremed*, *fremedest*, *fremede*, *fremedon*. *This* is just a fraction of the complexity; there were strong verbs, weak verbs, and verbs of many classes. All of this disappeared and that entailed that new syntactic structures emerged. The group genitive was one, replacing earlier split genitives.

4.2. Introduction of an inflectional category

English speakers use verbs in the perfect aspect (31a) but not modal auxiliaries (31b). (31b) does not occur, although it is clear what it might mean: he has been able to understand chapter 4 but now cannot.

- (31) a. He has understood chapter 4
 b. *He has could understand chapter 4

Verbs occur with *-ing* forms (32a) but not auxiliaries (32b), which would mean 'being able to understand ...'.

- (32) a. Understanding chapter 4, she went on to chapter 5
 b. *Canning understand chapter 4,

Verbs occur in an infinitival *to* form (33a) but not auxiliaries (33b), which would mean 'wanted to be able to understand.'

- (33) a. He wanted to understand
 b. *He wanted to can understand

A verb may occur with a modal (34a) but not two modal auxiliaries (34b).

- (34) a. He will try to understand
 b. *He will can understand

Verbs may have direct objects (35a) but not modal auxiliaries (35b).

- (35) a. He understands music
 b. *He can music

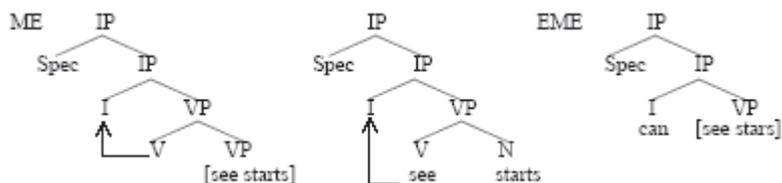
However, this is true of modern English but not of English up to the early sixteenth century. The (b) forms showed up in the texts and we find modal auxiliaries with an *-ing* form (36a), meaning 'a thing being able to or having to be done' and combinations of modal auxiliaries (36b). Sir Thomas More was one of the last writers to use a modal auxiliary in a *to* infinitive (36c) and modal auxiliaries in the perfect aspect (36d).

- (36) a. The potential mode signyfyeth a thing as mayying or owing to be done (c1512, Linacre, *Progymnasmata*)
 b. I fear that the emperor will depart thence, before my letters shall may come unto your grace's hands (1532, Cranmer, *Letters*)

- c. That appered at the fyrste to mow stande the realm in grete stede
(1533, More, *Works* 885 C1),
'appeared at first to be able to stand the realm in good stead.'
- d. If wee had mought convenient come togyther, ye woulde rather haue
chosin to haue harde my minde of mine owne mouthe
(1528, More, *Works* 107 H6)
'if we had been able to come together conveniently, ...'

These changes took place quickly, at the same time, and served to make Early Modern English different from Middle English. It was a single change structurally. In Middle English all verbs could move to a higher inflection position but by the early sixteenth century all speakers of English had classified words like *can*, *must* and *may* no longer as verbs but as a new Inflectional category. Children had developed a new internal language and, from that single fact about people's internal languages, it follows that each of the (31-35b) forms did not exist any more.

(37)



We also know the reason for this shift, and we see again that language change often shows domino effects. The change in category membership was due to prior morphological changes that had the effect of singling out the new modal auxiliary verbs. Middle English saw a massive simplification of morphology: the bewildering range of endings on different classes of verbs that I just described reduced to just one ending in the present tense, *-s* in the third person singular.

The verbs *can*, *may*, *must*, etc, the verbs whose behavior changed, belonged to a particular inflectional class in early English, the so-called "preterit-presents." What was distinctive about that class was that, unlike all other classes of verb, there was no *-eth* or *-s* ending for the third person singular present tense forms. Verbs like *can*, *may*, and *must* never had the *-s* ending. When there were many kinds of inflectional classes, this was just

one fact among hundreds. However, once the morphological system had eroded, the presence of a *-s* ending for the third person singular became the single, defining property of English verb morphology, and these verbs lacked it. As a result, verbs with no *-s* ending became distinctive and evidence shows that they were assigned to a new category. The evidence is the changes just described.

Again, Universal Grammar provides the available structures and sometimes we see changes in the structures and thereby learn about the structures and what triggers them in children.

The new behavior of modal auxiliaries is one feature of the new language of Early Modern English, one way in which Shakespeare's language differed from that of Chaucer. And Shakespeare's language also differed from Jane Austen's because of other structural shifts, bumps in the history of English that gave rise to yet newer forms.

4.3. *Loss of verb movement*

A little later English lost expressions that had been normal and whose equivalents are normal in most modern European languages (38).

- (38) a. Understands Kim chapter 4?
 b. Kim understands not chapter 4
 c. Kim reads always the newspapers

Instead there were new forms with a dummy, meaningless verb *do*:

- (39) a. Does Kim understand chapter 4?
 b. Kim does not understand chapter 4
 c. Kim always reads newspapers

The single structural shift here is that children ceased to acquire the operation that moved verbs to a higher Inflection position, mentioned a moment ago. That operation had yielded the three now-obsolete forms (38).

Shakespeare used both systems. (40) has three lines from *Othello*, which manifest the old system and the new system with the dummy *do* side-by-side. New-old in (40a), old-new in (40b), and new-old in (40c).

- (40) a. Where didst thou see her? - O unhappy girl! - With the Moor, say'st thou?
 b. I like not that. // What dost thou say?
 c. Alas, what does this gentleman conceive? - How do you, madam?

The shift was due to two prior changes and we see another domino effect. The first was the recategorization of modal verbs that we just discussed and the second was the emergence, first in the Westcountry, of “periphrastic” *do* forms as an alternative option for expressing past tense: *John did leave*, *John did not leave*, etc, instead of *John left* and *John left not*. As a result, the inflection position was occupied by modal auxiliaries and by *do* and was not available as a target for verb movement in those instances; verbs did not occur in that position as often as before the days of periphrastic *do* and before modal auxiliaries were no longer verbs.

These are three structural shifts in the history of English, each of them yielding new forms or eliminating older forms and changing the language, giving birth to new languages and making Chaucer’s language very different from that of our students.

5. Nicaraguan sign language

There’s a spectacular case of a new language emerging over the last 25 years: Nicaraguan Sign Language, which has been the subject now of many articles in the technical literature and the popular press, and of a wonderful BBC documentary. The Somoza dictatorship that ruled Nicaragua for forty years had a view of the deaf, that they were subhuman and should not be allowed to congregate; therefore they did not develop a common language. When the Sandinistas came to power just over 25 years ago, they rejected that view, allowed the deaf to congregate and set up a school in Managua. There were about 500 deaf people in Nicaragua and the people who first came to the school were of different ages, ranging from two-year olds to adults. They came together and a common language quickly began to emerge. Linguists and psychologists have been able to watch the emergence of a new common, sign language and to track the developments, particularly very different developments in young people and adults.

In general, work over the last generation has shown that signed languages are structured in pretty much the same way as oral languages and are subject to the same principles. It seems that there is a principle, a constraint on human vocabulary, that there are no verb stems that indicate manner and direction of movement. *Roll* describes the manner of movement and *descend* describes direction, but no verb stem indicates both manner and direction. Senghas, Kita & Özyürek (2004) have shown that sign languages are subject to the same constraint. This is particularly interesting, because one might have thought that it would be easy to gesture a rotating downward motion. In fact, it is easy to make such gestures and they were made in the first years of the emerging sign language in Nicaragua, when the language was used most frequently by older people who had developed their language much later than children typically get their first language system. However, as the language came to be filtered through the usual acquisition processes of younger people from infancy, such gestured forms came to be bleached out of the language. We see the effects of the constraints of the human language capacity in the way in which Nicaraguan Sign Language has developed over its 25-year history.

Again, examining changes from one language to another, we see effects of Universal Grammar, the human language capacity.

6. Conclusions

So people have their internal languages and they use them in different ways. That affects external language, language out there, what young children hear. If language out there changes in certain ways, it may trigger new internal languages in children. Whatever English is, it is a function of the different internal systems that certain people have acquired, whether they live in Tennessee or Cornwall. These external languages are ill-defined social notions. It is not clear what the limits of English are and, as we noted, it seems to depend on who draws political boundaries where. On the other hand, internal languages are biological phenomena that develop in us in ways that we can understand.

What we find is that both external language and internal languages may change, can be born and die out. They come and they go; they are different kinds of objects but they interact. We cannot understand how

new internal languages arise without considering language out there, and language out there is an aggregation of people's internal languages and their use. We have seen domino effects: by virtue of the human language capacity, if one thing changes, other changes follow and new languages entail other new languages. Just as languages can fall into disuse, new languages are always emerging: once new Latins and now new Englishes and new Spanishes.

E-mail: dlightfo@nsf.gov

REFERENCES

- CHOMSKY, N. 1986. *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- CRAIN, S. & R. THORNTON. 1998. *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.
- HARRISON, D. 2007. *When languages die*. Oxford: Oxford University Press.
- HUMBOLT, W. von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Royal Academy of Sciences of Berlin [*Linguistic variability and intellectual development*, transl. G.C. Buck & C. Raven. 1971. Philadelphia: University of Philadelphia Press].
- KRAUSS, M. 1988. Many tongues - Ancient Talespp. In: W. W. FITZHUGH & A. CROWELL (eds.) *Crossroads of continents: Cultures of Siberia and Alaska*. Washington, DC: Smithsonian Institution. 144-150.
- LIGHTFOOT, D. 2006a. *How new languages emerge*. Cambridge: Cambridge University Press.
- _____. 2006b. Minimizing government: Deletion as cliticization. *The Linguistic Review* 23.2: 97-126.
- NETTLE, D. & S. ROMAINE. 2000. *Vanishing voices: The extinction of the world's languages*. Oxford: Oxford University Press.
- SENGHAS, A., S. KITA & A. ÖZYÜREK. 2004. Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science* 305: 1779-1782.