The Story of Human Language
Part II

Professor John McWhorter

THE TEACHING COMPANY ®

John McWhorter, Ph.D.

Senior Fellow in Public Policy, Manhattan Institute

John McWhorter, Senior Fellow at the Manhattan Institute, earned his Ph.D. in linguistics from Stanford University in 1993 and became Associate Professor of Linguistics at UC Berkeley after teaching at Cornell University. His academic specialty is language change and language contact. He is the author of *The Power of Babel: A Natural History of Language*, on how the world's languages arise, change, and mix. He has also written a book on dialects and Black English, *The Word on the Street*. His books on creoles include *Language Change and Language Contact in Pidgins and Creoles*, *The Missing Spanish Creoles*, and an anthology of his creole articles called *Defining Creole*. Beyond his work in linguistics, Dr. McWhorter is the author of *Losing the Race* and an anthology of race writings, *Authentically Black.* He has written on race and cultural issues for *The New Republic*, *The Wall Street Journal*, *The Washington Post*, *The Chronicle of Higher Education*, *The National Review*, *The Los Angeles Times*, *The American Enterprise*, and *The New York Times*. Dr. McWhorter has appeared on *Dateline NBC*, *Politically Incorrect*, *Talk of the Nation*, *Today*, *Good Morning, America*, *The Jim Lehrer NewsHour*, and *Fresh Air* and does regular commentaries for *All Things Considered*. His latest book is *Doing Our Own Thing: The Degradation of Language and Music in America and Why We Should, Like, Care*.

Table of Contents

The Story of Human Language
Part II

The Story of Human Language

Scope:

There are 6,000 languages in the world, in so much variety that many languages would leave English speakers wondering just how a human being could possibly learn and use them. How did these languages come to be? Why isn't there just a single language?

This course answers these questions. Like animals and plants, the world's languages are the result of a long "natural history," which began with a single first language spoken in Africa. As human populations migrated to new places on the planet, each group's version of the language changed in different ways, until there were several languages where there was once one. Eventually, there were thousands.

Languages change in ways that make old sounds into new sounds and words into grammar, and they shift in different directions, so that eventually there are languages as different as German and Japanese. At all times, any language is gradually on its way to changing into a new one; the language that is not gradually turning upside-down is one on the verge of extinction.

This kind of change is so relentless that it even creates "languages within languages." In separate populations who speak the same language, changes differ. The result is variations upon the language—that is, dialects. Often one dialect is chosen as the standard one, and when it is used in writing, it changes more slowly than the ones that are mostly just spoken because the permanency of writing has an official look that makes change seem suspicious. But the dialects that are mostly just spoken keep on changing at a more normal pace.

Then, the languages of the world tend to mix together on various levels. All languages borrow words from one another; there is no "pure" vocabulary. But some borrow so much vocabulary that there is little original material left, such as in English. And meanwhile, languages spoken alongside one another also trade grammar, coming to look alike the way married couples sometimes do. Some languages are even direct crosses between one language and another, two languages having "reproduced" along the lines of mitosis.

Ordinarily, language change is an exuberant process that makes languages develop far more machinery than they need—the gender markers in such languages as French and German are hardly necessary to communication, for example. But this overgrowth is checked when history gets in the way. For example, when people learn a language quickly without being explicitly taught, they develop a pidgin version of it; then, if they need to use this pidgin on an everyday basis, it becomes a real language, called a *creole*. Creoles are language starting again in a fashion—immediately they divide into dialects, mix with other languages, and start building up the decorations that older languages have.

Just as there is an extinction crisis among many of the world's animals and plants, it is estimated that 5,500 of the world's languages will no longer be spoken in 2100. Globalization and urbanization tend to bring people toward one of a few dozen politically dominant languages, and once a generation is not raised in a language, it no longer survives except in writing—if linguists have gotten to it yet. As a language dies, it passes through a "pidgin" stage on its way to expiration. This course, then, is both a celebration and a memorial of a fascinating variety of languages that is unlikely to exist for much longer.

Lecture Thirteen

The Case For the World's First Language

Scope: Most linguists' reception of the Proto-World work has been less skeptical than hostile, and as often in such cases, there is more truth to the theory than many admit. For example, there is increasing evidence that many of the world's families do trace to "mega-ancestors," even if evidence for a Proto-World remains lacking. The Proto-World school's reconstruction of features of the Native American proto-language are promising, and one of these linguists has recently discovered a likely valid link between languages whose speakers have had no contact for 50,000 years.

Outline

I. Smaller superfamilies: Eurasiatic.

A. Greenberg and Ruhlen follow in a tradition that traces back to the early 20th century in noticing crucial similarities between Indo-European languages and other families across the Eurasian landmass. A group of Russian scholars' version of this refers to a grand *Nostratic* family; Greenberg and Ruhlen differ in exactly which families they include but agree in broad outline.

B. Their *Eurasiatic* family includes Indo-European, Uralic (including Finnish and Hungarian), Altaic (stretching across Asia and including Turkish and Mongolian), Korean and Japanese, the Chukchi-Kamchatkan group spoken in far eastern Russia, and the Eskimo-Aleut languages spoken across the Bering Strait in northern North America.

C. Evidence that these families had a common ancestor comes from similarities such as those outlined below.

Evidence for the Eurasiatic mega-family:

	I, me, my	you (sing.)	who	what
Indo-European	*mē	*tu	*kwi	*ma
Uralic	*-m	*te	*ke	*mi
Turkic	men		*kim	*mi
Mongolian	mini	*ti	ken	*ma
Korean	-ma		-ka	mai
Chukchi-Kamchatkan	-m	-t	*kina	*mi
Eskimo-Aleut	-ma	-t	*kina	*mi

D. Note that words for "I" beginning with *m* and words for "you" beginning with *t*—a pattern we are familiar with from Spanish (*me*/*te*)—are common across Asia and in the Arctic. Importantly,

similarities between aspects of grammar, rather than concrete words, are considered more indicative of a historical relationship because grammatical items change more slowly than concrete ones. For example, Russian's noun and verb endings are similar to Latin's in both their shape and function, while its vocabulary is extremely different.

II. Smaller superfamilies: Amerind.

A. Of the dozens of language groups spoken by Native Americans in the New World, Greenberg, supported by Ruhlen, classified them into just three groups: two small ones in the north, Eskimo-Aleut and Na-Dené, and an enormous one encompassing all of the others, which he called *Amerind*.

B. One piece of evidence for Amerind is a particular word shape referring to family members of the same age or younger than oneself, **t—na*, with the vowel changing according to sex. Variations on this pattern are found throughout the New World languages and are unlikely to be accidental.

Evidence for the Amerind family:

**t'ina* "son, brother"	**t'una* "daughter, sister"	**t'ana* "child, sibling"
Iranshe *atina* "male relative"	Iranshe *atuna* "female relative"	
Tiquie *ten* "son"	Tiquie *ton* "daughter"	
Yurok *tsin* "young man"	Salinan *a-t'on* "younger sister"	Nootka *t'an'a* "child"
Mohawk *-tsin* "male, boy"	Tacana *-tóna* "younger sister"	Aymara *tayna* "first-born child"

C. Lately, genetic evidence has supported an Amerind family, showing that Native Americans' genetic patterns differ exactly according to the three groups Greenberg identified.

D. Specialists in Native American languages have objected that the evidence for Amerind as a language group is a collection of chance correspondences and that anyone could find a similar range of chance correspondences to "prove" any classification. Ruhlen objects that it would be impossible to make a case for a **t—na* root with these vowel changes from the world's languages beyond Amerind—and he has a point.

III. How much does time bury?

A. Recently, Ruhlen has documented close affinities between an obscure language of Nepal, Kusunda, generally classified as related to Chinese (Sino-Tibetan) and the language family of Papua New Guinea, called *Indo-Pacific*. Here are some common features between Kusunda and one of the languages of this group, Juwoi.

Evidence for the relationship between Kusunda and Indo-Pacific languages:

	KUSUNDA	JUWOI
I	t^{s}i	tui
my	t^{s}i-yi	tii-ye
you	nu	ŋui
your	ni-yi	ŋii-ye
give	ai	a
this	(y)it	ete
knee	tugutu	togar (“ankle)
unripe	katuk	kadak (“bad character”)

B. This relationship is crucial because humans are known to have traveled from southern Asia to New Guinea at least 50,000 years ago, with recent evidence suggesting as long as 75,000 years ago. Thus, these words may represent the oldest documentable historical relationship between words and show that many linguists’ claim that no relationship between languages can be documented beyond 6,000 or so years is untenable.

IV. Final verdict.

A. Ruhlen’s point that comparative reconstruction is not the only way to show that languages have a common ancestor is valid in itself. He observes that linguists posited the Indo-European group long before Proto-Indo-European itself had been worked out by working backward from the languages. The similarities between language families are close enough that his point is likely valid for mega-groups, such as Amerind and Eurasiatic.

B. A question still remains, however, as to how realistic even this approach is for Proto-World. The issues could be resolved as more proto-languages are reconstructed, although work of this kind is done increasingly less by modern linguists, and for reasons we will see in later lectures, it may be entirely impossible to reconstruct proto-languages for many families.

Essential Reading:

Ruhlen, Merritt. *The Origin of Language: Tracing the Evolution of the Mother Tongue*. New York: John Wiley & Sons, 1994.

———. "Taxonomic Controversies in the Twentieth Century," in *New Essays on the Origin of Language*, edited by Jürgen Trabant and Sean Ward, pp. 97–214. Berlin: Mouton de Gruyter, 2001.

Questions to Consider:

1. Why do grammatical items, such as prefixes and suffixes, change more slowly than separate words do? Along the same lines, what kinds of words do you think might change more slowly than others?
2. The debate between the Proto-World school and other linguists is partly the product of the age-old divide between "lumpers" (attuned to broad patterns) and "splitters" (attuned to fine details). Is it the job of the academic to be a "splitter" and leave the lumping to laymen, or do you think that "lumping" has a place in academic thought as well?

Lecture Fourteen
Dialects—Subspecies of Species

Scope: When the process that turns one language into a number of new ones has not yet gone far enough to create new languages per se, then the variations are considered *dialects* of the original language. This is what dialects are: variations on a common theme, rather than bastardizations of a "legitimate" standard variety. England is home to a number of variations on English, and importantly, Standard English is just another dialect that developed alongside these and happened to be chosen as the "show" dialect. The Parisian dialect of French was anointed in similar fashion. Often, what is considered the "proper" dialect today is a mere "dialect" tomorrow, such as Provençal in France.

Outline

I. Variety within languages.

A. "Language" is, strictly speaking, an artificial, arbitrary concept. Not only has the first language developed into 6,000, but almost all of these languages are, viewed close up, bundles of variations on a theme. These are dialects of the languages.

B. Here are some British dialects of English. Note that many are different enough from Standard English that they require translation.

STANDARD	The government has today decreed that all British beef is safe for consumption.
SCOTS	Efter he had gane throu the haill o it, a fell faimin brak out i yon laund.
	"After he had gone through all of it, a great famine broke out in the land."
LANCASHIRE	Ween meet neaw ta'en a hawse steyler at wur mayin' off with'tit.
	"We have just now taken a horse stealer who was making off with it."
NOTTINGHAMSHIRE	Tha mun come one naight ter th' cottage, afore tha goos; sholl ter?
	"You must come one night to the cottage before you go, will you?"

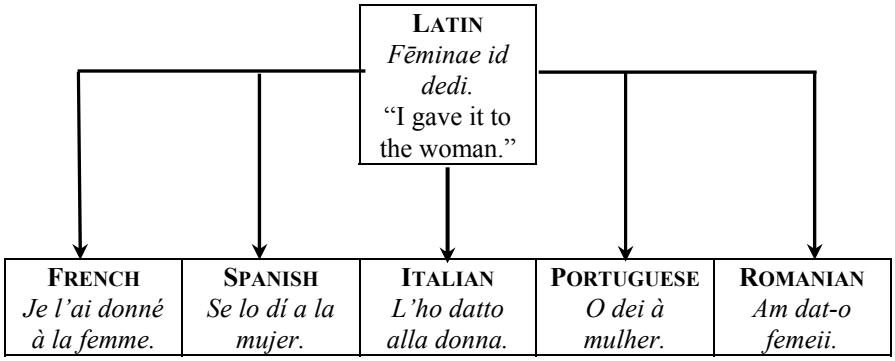
CORNWALL Aw bain't gwine for tell ee.

"He isn't going to tell you."

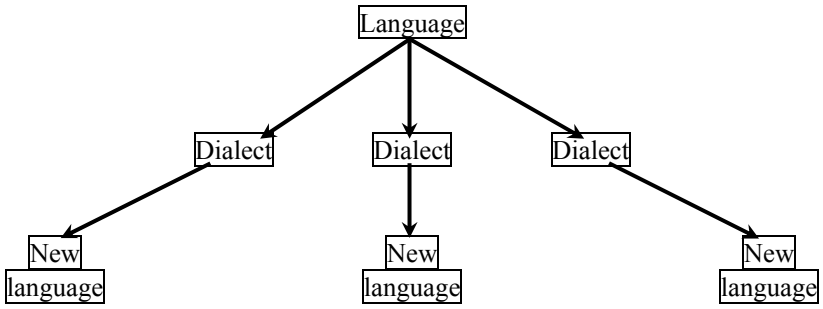
II. Ordinary language change creates dialects.

A. We can understand what dialects are only by shedding the common misconception that a dialect is a degraded version of the standard language. What creates dialects is not sloth but simple language change.

B. Recall how several languages can develop from one, as the Romance languages did from Latin.



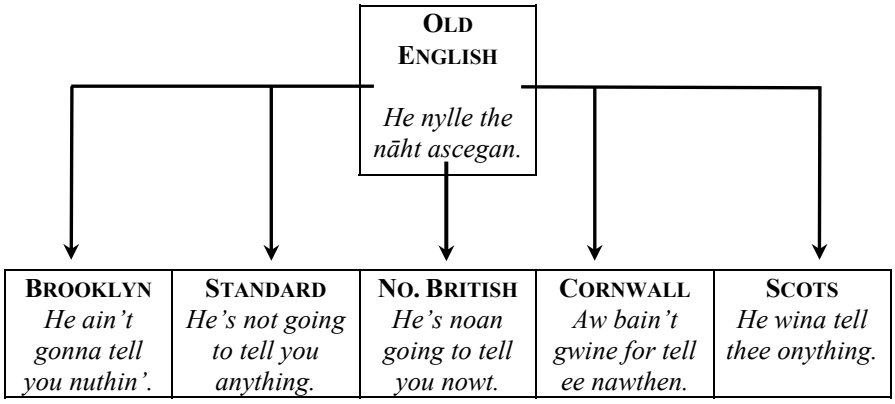
C. Dialects are simply the intermediate stage in this process: at a certain point, a language has changed in several directions into new varieties that are not divergent enough to be different languages altogether but are obviously on their way.



D. We have records of French, for example, at an intermediate stage between Latin and its current state. At that point, one writer complained in 63 A.D.:

Spoken Latin has picked up a passel of words considered too casual for written Latin, and the grammar people use when speaking has broken down. The masses barely use anything but the nominative and the accusative... it's gotten to the point that the student of Latin is writing in what is to them an artificial language, and it is an effort for him to recite in it decently. (Monteilhet, H. *Neropolis: Roman des temps néroniens*. Paris: Éditions du Juillard, 1984.)

E. Here is an example of the same sentence in several different English dialects:



F. Most languages are bundles of dialects like this.

1. English borrowed *warrant* from French, but in Standard French, the word is *garant*. *Warrant* is borrowed from the Normandy dialect, which often had *w* where Standard French has *g*.

2. Italian dialects are so different from one another that the dialect of Sicily is essentially a different language from the standard.

3. The situation is similar in Germany. In Standard German, "You have something" is *Du hast etwas*; in a southern dialect, Schwäbisch, it is *De hesch oppis*.

III. The standard is just lucky.

A. When a language is a written one, one of the dialects is usually chosen as the standard dialect, used in writing and public contexts. But an important thing to notice is that standard dialects usually develop alongside nonstandard ones, rather than the nonstandard ones developing from the standard.

B. "A standard is a dialect with an army and a navy"—standards become standard because they have "the juice" in some way. Francien French became predominant because the national courts settled in its region; Castillian Spanish because it was spoken by the armies who advanced

southward to defeat the Moors; Tuscan Italian because that region produced Dante, Petrarch, and Boccaccio.

C. Standard English is the dialect that happened to be spoken in the region where London was. Before this, England was a patchwork of very different dialects. In the late 1400s, printer William Caxton told a story of a Londoner who had barely been able to make himself understood in Kent, the region just next door, because he had asked for *eggs* instead of using the Kentish dialect word, *eyren*.

D. France was also once home to many distinct dialects. This was seen as a problem as France coalesced from a patchwork of feudal duchies into a nation. The Abbé Grégoire, a Catholic priest and revolutionary, worried in 1789 that:

> France is home to perhaps 8 million subjects of which some can barely mumble a few malformed words or one or two disjointed sentences of our language: the rest know none at all. We know that in Lower Brittany, and beyond the Loire, in many places, the clergy is still obliged to preach in the local patois, for fear, if they spoke French, of not being understood. (Grillo, Ralph. *Dominant Languages: Language and Hierarchy in Britain and France*. Cambridge: Cambridge University Press, 1989, p. 31.)

The dialect of French that had developed in the Paris area was imposed on the population for practical reasons.

E. *Standard today, dialect tomorrow*. Ukrainian and Russian are similar enough that for a Russian, learning Ukrainian straddles the boundary between learning a new language and adjusting to a variety of Russian itself. Indeed, before the Ukraine was cordoned off as a separate region in the Soviet Union, it was a region within Russia, and the speech of the Ukraine was considered a kind of "Russian." When the center of power in Russia was Kiev, the speech of the Ukraine was considered the "best" Russian. After this, however, Ukrainian was dismissed as the speech of peasants. Then, when the Ukraine became a political entity, Ukrainian again became a "language." The difference had been in culture and politics, not in the speech variety itself.

IV. The standard seems "better" only because of accident. Dialects are equivalent to subspecies in the animal and plant kingdoms. Scots, Brooklyn English, and Standard English are to "English" as cocker spaniel, dachshund, and collie are to "dog." Just as there is no "default" or unequivocally "best" dog, there is no "real" dialect of a language. Rather, dialects are evidence of the variety-within-the-variety among the descendants of the first language.

Essential Reading:

Crystal, David. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 1995 (especially chapter 5: "Early Modern English").

McWhorter, John H. *Word on the Street: Debunking the Myth of a "Pure" Standard English*. New York: Perseus, 1998.

Supplementary Reading:

Grillo, Ralph. *Dominant Languages: Language and Hierarchy in Britain and France*. Cambridge: Cambridge University Press, 1989.

Questions to Consider:

1. Often, people come away from a lecture like this one nevertheless still quite convinced that certain ways of speaking are just "incorrect." If by chance you feel this way, explore the difference between "nonstandard" and "incorrect" and how this justifies your sense of proper and improper language.
2. Generally, even speakers of nonstandard dialects consider their way of speaking "not real language." But if the way they speak evolved alongside the standard variety based on the same processes, then what conditions this sense of what "real" language is? Do you agree or disagree with this sense?

Lecture Fifteen

Dialects—Where Do You Draw the Line?

Scope: The labels *language* and *dialect* are, in practice, arbitrary, and necessarily so. Dialects of one language can be called separate languages simply because they are spoken in different countries, such as Swedish, Norwegian, and Danish. Different languages can be called dialects because they are spoken in the same country and written in the same system, such as Chinese "dialects," which are as different as French and Spanish. Often, dialects change slightly from region to region until people at one end of the chain cannot converse with people on the other end; where one draws the line between dialect and language here becomes meaningless.

The truth is that there is no such thing in any definable sense as a "language." Tens of thousands of dialects are spread across the globe, many of them akin enough to be perceptible as variations on "the same thing"—but even here, only in variable degrees.

Outline

I. Dialects as "languages": Often what begins being considered a dialect of one language is recast as a separate "language" of its own when its speakers are incorporated into a new nation.

A. *Scandinavian*. Swedish, Norwegian, and Danish are official languages of Sweden, Norway, and Denmark. But speakers of them can manage a conversation, and on the page, they reveal themselves as minor variations on a pattern, rather like Scots, Cornwall English, and Standard English.

The Danes initially ruled Sweden and Norway, and there was no such thing as a Swedish "language" until Sweden became independent in 1526 or a Norwegian "language" until Norway became independent in 1814. Until their independence, Sweden and Norway's speech varieties were simply considered dialects of Danish.

B. *Moldovan*. Romania used to extend eastward into a little hump of land called Moldova. At first, the speech of Moldova was considered one of many nonstandard dialects of Romanian. But after Moldova was incorporated into the Soviet Union, the Soviets directed Moldovan linguists to write grammars of a new Moldovan "language," even though many of these were just grammars of Romanian translated into Russian.

C. *Different culture, different language?* Hindi is spoken in India and written in the Devanagari script, while Urdu is spoken in Pakistan and written in Arabic script. Because of this and the religious and political

tensions between the countries, Hindi and Urdu are treated as separate "languages" when they are, in fact, the same one. Hindi has more Sanskrit borrowings, while Urdu has more from Arabic, but these impede communication little more than the differences between American and British English.

D. *Indigenous languages*. The continuum nature of the language/dialect distinction is clear even when the speech varieties are not adopted as written languages and assigned by nations as single official ones.

1. *Malinke, Bambara, and Dyula in West Africa*. The "languages" Malinke and Bambara are spoken in a vast region spread across such West African countries as Senegal, Mali, and Guinea, alongside dozens of other languages in each country. But speakers of these languages can understand one another, as well as speakers of the Dyula "language" in Côte d'Ivoire. Only cultural affiliations determine what this one "language" is called from place to place.

2. *Tourai and Aria in New Guinea*. On the island of New Britain near New Guinea, there are two groups called the Tourai and the Aria. What the two groups speak appears to be the same language with minor differences on the page, and other peoples in the area learn the same language to speak to both. But while the Tourai think of the Aria as speaking a different language, the Aria think of themselves as speaking the same thing as the Tourai.

II. Languages as "dialects": In other cases, separate languages are treated as dialects of one because they are all spoken in one nation or by the same cultural group.

A. *Chinese*. As we have seen, Chinese "dialects," such as Mandarin and Cantonese, are actually as different as the Romance languages are from one another.

Obviously these are separate languages, and the five other main Chinese "dialects" are just as different from one another, such as Taiwanese and Shanghainese. But all of the languages are written with the same system, which uses symbols for whole words instead of for sounds. This means that the languages look quite similar to one another on the page, since, for example, the word for *man* is the same symbol in all of the languages even though the spoken word is quite different. Then, the sense that all of the languages' speakers have of being united as "Chinese" completes the impression that there is a single Chinese "language."

B. *Arabic*. The varieties of what is called "Arabic" in various nations are as different as the Romance languages as well.

nothing in Arabic "dialects":

Algerian	ši
Tunisian	šay
Nigerian	še
Moroccan	wálu
Saudi	walašay
Egyptian	dilwa'ti
Libyan	kān lbarka

But these languages are largely used only for speaking. Modern Standard Arabic, based on the language of the Koran, is used in writing and formal language in most of these countries, and the spoken variety is considered a bastard version of the standard rather than as a separate "language" in its own right. Hence, there is a sense that one language, "Arabic," is spoken across the Arab world, rather than several different languages.

III. Dialect continua: The distinction between language and dialect is ever more hopeless when we see that in many parts of the world, one dialect shades into another one from region to region until people on one end of the chain speak a different "language" than the ones at the other, but there has been no single point along the chain where a new language can be seen as beginning.

A. *Gurage*. Gurage is the name of a dialect continuum of the Semitic family, spoken in Ethiopia. Here is "He thatched a roof" in several of the varieties, shading gradually from one "language" to another.

He thatched a roof in Gurage dialects

Soddo	kəddənəm
Gogot	kəddənəm
Muher	khəddənəm
Ezha	khəddərəm
Chaha	khədərəm
Gyeto	khətərə
Endegen	həttərə

People speaking one variety can converse with people speaking the one next door, have a harder time with the one spoken two regions away, and so on. Soddo and Endegen seem easily identifiable as "languages," but whether, for example, Chaha in the middle is a different "language" from either of them is as arbitrary an issue as whether purple is more red or more blue.

B. *Turkic varieties*. Turkish is one of a litter of languages stretching from Turkey east across the new "stan" countries into western China. These "languages" vary in the same way as what are called "dialects" of

many other languages and form a continuum. Here is the word for *eight* stretching from west to east.

eight in Turkic languages

Turkish	sekiz
Azerbaijani	səkkiz
Turkmen	sekiz
Uzbek	sakkiz
Kazakh	segiz
Kirghiz	segiz
Uighur	säkkiz

Yet the Gurage varieties are thought of as "dialects," while these are "languages"—the terminology is arbitrary, based largely on the fact that the Turkic ones are spoken in separate political entities.

IV. Dialect of A or new language B?

A. Even when there is no continuum of this kind, the question of whether one speech variety is a dialect of one language or a new language entirely is often undecidable.

B. Scots English can test the comprehension of an English speaker. Consider that *auld lang syne* means *old long since*. But hearing Scots spoken at speed in casual situations, an English speaker is often confronted with what feels like an ill-tuned radio signal. This is an experience typical of speakers of most languages: English is unique in how few speech varieties straddle the line between it and other languages.

Essential Reading:

The topic of this lecture is not generally covered in sources for a general audience. With all due humility, I believe that the most pertinent survey of the topic of this lecture is my own: McWhorter, John. *The Power of Babel*. New York: HarperCollins, 2001 (chapter 2).

Supplementary Reading:

These are some language area surveys that those interested might find useful:

Arabic: Versteegh, Kees. *The Arabic Language*. New York: Columbia University Press, 1997.

Chinese: Norman, Jerry. *Chinese*. Cambridge: Cambridge University Press, 1988.

Moldovan: Dyer, Donald L. *The Romanian Dialect of Moldova*. Lewiston, NY: Mellen Press, 1999.

Scots: Crystal, David. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 1995 ("Middle Scots," pp. 52–53; "Variation in Scotland," pp. 328–333).

Questions to Consider:

1. After this lecture, do you perhaps have a sense that there is any salvaging of the distinction between language and dialect that rises above the messy reality?
2. Arabic as spoken from country to country differs as much as the Romance languages do, but the writing system helps all of the peoples in question see themselves as speaking one "Arabic." The Chinese "dialects" are similar. Are speech varieties that share a writing system "the same language"? How important is writing to defining what a "language" is?

Lecture Sixteen
Dialects—Two Tongues in One Mouth

Scope: In most Arabic-speaking countries, the Arabic of public use (the media, speeches, writing) is essentially a different language from the one used casually and learned from parents. This phenomenon is called *diglossia* and is common worldwide. Swiss German speakers only occasionally see the language they speak on the page, where High German is required. Different languages are also often used in diglossic relationships: the Tanzanian often uses English and Swahili at work and a local native language at home. Diglossia is the template within which 6,000 languages and countless dialects share space on a planet with only 200-odd nations.

The nonstandard dialect and the standard one often coexist in a structured relationship in a society. The standard or "high" (H) variety is used in formal situations, while the nonstandard or "low" (L) variety is used in informal ones. This is called *diglossia*, Greek for "two tongues."

Outline

I. Typical examples: Modern Standard Arabic versus Egyptian Arabic; High German versus Swiss German in Switzerland; Katharévousa versus Dhimotikí in Greece.

II. Typical traits of diglossia.

- **A.** *Writing versus speaking*. People read the paper in H and discuss the issues in L. Speeches are given in H; conversations are conducted in L.
- **B.** *Acquisition*. H is learned in school; L is learned at home.
- **C.** *Standardization*. H is standardized with official "rules," while nonstandard varieties are described systematically only by academic linguists, missionaries, and similar researchers. This lack of standardization often encourages several L's to arise, such as the many nonstandard dialects of German.
- **D.** *Prestige*. People tend to disown that they speak L, or do not consider L a "real language" (there were riots in Greece in 1903 over the publication of the New Testament in Dhimotikí).

III. Typical examples.

- **A.** *Egyptian Arabic*. In Egyptian Arabic, "now" is *dilwa'ti*; in Standard Arabic, it is *'al'āna*. Egyptian Arabic for "nose" is *manaxīr*; in Standard Arabic, it is *'anf*. In other cases, the Egyptian is a variation

on the standard: "many" is *kathirah* in Standard*, kətir* in Egyptian. An Egyptian learns to speak, essentially, a whole new language in school.

B. *Swiss German*. In German-speaking Switzerland, to be a functioning person requires being bilingual in two forms of "German" that are as different as Spanish and Portuguese. High German for "drink" is *trinken*; Swiss German has *suufe*. High German has *kein* for "not one"; Swiss German has *ke*.

C. *Triglossia*. In particularly hierarchical societies, there can be three levels of language according to context. In Javanese, for example, there is a "middle rung" between the "highest" and "lowest" forms. Here is "Are you going to eat rice and cassava now?" on all three levels.

HIGH	menapa	pandjenengan	baḍé	ḍahar	sekul	kalijan	kaspé	samenika?
MIDDLE	napa	sampéjan	adjeng	neḍa	sekul	lan	kaspé	saniki?
LOW	apa	kowé	arep	mangan	sega	lan	kaspé	saiki?
	Are	you	going	to eat	rice	and	cassava	now

D. The closest equivalent to diglossia in English is the difference between such words as *dine* and *eat*, *children* and *kids*, or *parcels* and *bags*. Imagine if differences like these applied to most of the words in the language!

IV. Diglossia of languages.

A. There are about 6,000 languages in the world and only 200-odd countries; this shows that multilingualism in nations is a norm.

B. The appearance otherwise is explained by the fact that only a quarter of the world's countries recognize two languages officially, and only four recognize three or more. India recognizes Hindi, English, and 14 regional languages; Singapore: Chinese, Malay, Tamil, and English; Spain: Spanish, Catalan, and Basque; and Luxembourg: French, German, and the local German dialect Letzebuergesch.

C. Languages typically share space in a country in diglossic relationships. An example is Paraguay, where the official languages are Spanish and the Native American language Guaraní. But the two languages are not simply used side by side in all contexts. Guaraní is used as the L language and Spanish as the H one.

D. In fact, where there is extensive bilingualism, diglossia is almost inevitable.

1. In Quebec before 1974, English was the H language and French the L one. But in the 1970s, a law was enacted that made French the province's official language and required the use of French in the government and on public signs. This has been a delicate and charged situation, imposed rather than emerging by itself.

2. Although extensive bilingualism without diglossia is rare, diglossia can exist among an elite in a society even when most of the society's people are not bilingual. In Czarist Russia, upper-class people often spoke French among themselves, especially on formal occasions. French was the H and Russian was the L.

Essential Reading:

Ferguson, Charles A. *Language Structure and Language Use* (essays selected and introduced by Anwar S. Dil). Palo Alto, CA: Stanford University Press, 1971 ("Diglossia" essay).

Supplementary Reading:

Geertz, Clifford. "Linguistic Etiquette," in *Sociolinguistics*, edited by John Pride and Janet Holmes, pp. 167–179. Harmondsworth, England: Penguin, 1972.

Questions to Consider:

1. Is extreme diglossia, such as in Arabic-speaking countries, a problem? Would it be better if spoken languages were used in formal contexts as well, or is there an advantage to the existence of a "common coin" that unites all such countries?
2. What are some words or expressions that we regularly say but rarely write, such as "whole nother"? You will find that there is more diglossia in modern English than we are often aware of.

Lecture Seventeen

Dialects—The Standard as Token of the Past

Scope: Languages typically change quite quickly: there are cases where linguists examine a language at one point only to find that 60 years later, it has morphed into practically a brand new one. However, when a dialect of a language is used widely in writing and literacy is high, the pace of change is artificially slowed because people come to see "the language" as on the page and inviolable. This helps create diglossia: standard Arabic is based on the language of the Koran, while the colloquial Arabics went on with natural change.

Outline

I. The normal speed of language change.

A. When linguists studied the northern Australian language Ngan'gityemerri in 1930, they found a language with sentences similar to the following:

1930:
Dudu dam, dam dudu, kinji dinj parl.
Track poke poke track here he-sat camp

"He poked along, tracking it along here to where it made its camp."

1990:
Damdudu, damdudu, kinyi dinyparl.
Poke-track poke-track here he-sat-camp

Notice that in 1930 the speaker could give the order of *dudu* and *dam* (*track* and *poke*) in either order; they were separate words. But when linguists returned to the language in 1990, its entire grammar had changed. Now, *dudu* had grammaticalized into a prefix of *dam*, such that there was one word *dududam*, meaning roughly "pokingly tracked." This had happened with all verbs in the language. Ngan'gityemerri had moved along the path toward becoming a language like Yupik Eskimo, which packs a sentence's worth of meaning into one word. (Recall the Yupik Eskimo word for "He had not yet said again that he was going to hunt reindeer": *Tuntussuqatarniksaitengqiggtuq*.)

B. But English has changed more slowly in the time after the Middle Ages. Shakespeare speaking 500 years ago would have sounded strange to us, but we could converse with him. However, Shakespeare would have found an Old English speaker from 500 years earlier almost as incomprehensible as a German.

C. This phenomenon can be explained by the fact that when a language is written and standardized and literacy becomes widespread, the written form comes to be seen as "The Language," and it affects people's speaking habits enough that the language changes more slowly than it would naturally. Standardized languages are "frozen in aspic," as it were.

D. A contrast: we can easily read presidential addresses from the late 18th century, but a speaker of Saramaccan Creole in Suriname would find the speech of a chief in 1789 extremely peculiar. For example, at that time, the way to say "not" was *no*, but today, it is just *a*.

II. Standard languages and diglossia.

A. When a standard language is "frozen" in place while the spoken language develops naturally, often the result is diglossia between the standard and the colloquial variety.

B. This was the case with Arabic. For example, the regional Arabic dialects are the result of natural changes Arabic went through over time in each place, while the standard reflects the archaic language of the Koran.

1. Notice that the contrast between standard *kathirah* and Egyptian *kətir* shows the erosion of sounds at the ends of words, just as we pronounce name as "NEIGHM" rather than "NAH-muh," the earlier form of the word that the spelling preserves.

2. Modern Standard Arabic has three case endings: "house" is *bayt**u***, "of the house" is *bayt**i***, and when "house" is used as an object, it is *bayt**a***. But in Egyptian, these endings have disappeared, because sound erosion wore off final vowels, as it does so often in language change.

C. Notice also that the words for "are" in the levels of Javanese from the previous lecture show the same kind of development:

"Are you going to eat rice and cassava now?"

HIGH	menapa	pandjenengan	baḍé	ḍahar	sekul	kalijan	kaspé	samenika?
MIDDLE	napa	sampéjan	adjeng	neḍa	sekul	lan	kaspé	saniki?
LOW	apa	kowé	arep	mangan	sega	lan	kaspé	saiki?
	Are	you	going	to eat	rice	and	cassava	now

The word for *now*, *samenika* in the high variety, becomes *saniki* and *saiki*.

D. Standard French versus colloquial French.

1. Although Standard French has a double-negative marking, as in *Je **ne** marche **pas***, "I do not walk," in spoken French, the *ne* is almost always dropped: *Je marche pas* has been good spoken French since the Middle Ages. Small words, such as *ne*, that are not

accented tend to erode and even disappear in languages, just as sounds at the ends of words do. Spoken French has developed "naturally," while written French preserves a past stage.

2. French has a pronoun *on* used generically, equivalent to the *se* in *Aquí se habla español*, or *one* in English. But over the centuries, although *nous* has been the standard form for "we," *on* has been used in its place in casual speech. We are taught to say *nous parlons* for "we speak," but French people at all levels of society actually say *on parle*.

   That is the only thing that we do not do.

   STANDARD FRENCH:
   C'est la seule chose que **nous ne** faisons pas.

   SPOKEN FRENCH:
   C'est la seule chose qu'**on __** fait pas.

3. This means that to learn to speak French, we must learn a different dialect than the one taught in school—there are two Frenches, the standard that reflects what French was like centuries ago and the spoken version that has evolved since then.

III. The standard is not always more complex.

A. Because nonstandard dialects lose material over time, it can appear that the standard must really be the "better" version because it retains these things, and thus is "larger" than the nonstandard dialects.

B. But actually, languages complexify as they evolve while they are simplifying. This has happened in regional Arabic dialects, such as Egyptian. For example, Standard Arabic is fairly simple in terms of showing differences in time conceptions. Basically, there is a past and a present: "he wrote" is *kataba*; "he writes" is *yaktubu*. The future, the progressive, and so on are usually left to context.

C. But Egyptian, like other regional Arabic varieties, has developed markers to indicate time distinctions. For example, in Saudi Arabic, one places *b-* before a verb to indicate the future: *aguul*, "I tell"; *baguul*, "I will tell." *Kaan* before a verb means "used to": *kaan aguul*, "I used to tell."

Essential Reading:

The most pertinent exposition on this subject for the general reader is, honestly: McWhorter, John. *The Power of Babel*. New York: HarperCollins, 2001 (chapter 6).

Questions to Consider:

1. Is language change a bad thing? It often seems so in real life as we live it ("Why are people using *impact* as a verb?"). But Shakespeare played a major part in changing our language. Are there good changes versus bad ones, and what is the difference?
2. In Black English, *be* is used to indicate a habitual action: "She be goin to the store every Tuesday." This is a more explicit way of marking habituality than Standard English's simple "She goes to the store every Tuesday." Indeed, the *be* is "unconjugated," but in your opinion, does that render this usage of *be* "wrong" even if it also lends the dialect some clarity?

Lecture Eighteen
Dialects—Spoken Style, Written Style

Scope: We often see the written style of language as how it really "is" or "should be." But in fact, writing allows uses of language that are impossible when a language is only a spoken one, which all but about 200 of the world's languages effectively are. Writing allows the preservation of a massive vocabulary in dictionaries: spoken languages have some tens of thousands of words at most. Writing allows longer, more elaborate sentences than are typical of speech anywhere in the world. Early writing, such as the Hebrew Bible with its brief phrases, represents speech rather than the artifice of writing.

A main reason that standard varieties appear to be "realer" than nonstandard ones is that they have a richer vocabulary and more elaborated syntax. But it is important to realize that this trait is an artificial imposition from technology on the natural history of human language.

Outline

I. Spoken language: Raggedy but effective.

A. In the lecture, we hear part of a speech by Congressman Adam Clayton Powell, Jr. from the late 1960s, in the fundamentalist preaching style. As majestic as this passage is, its structure and language are rather simple. Sentences are short and repetitive. A composition teacher, if presented with the passage in writing, would likely advise the writer to use some graceful transitional words to knit the sentences together, such as *although*, *seeing that*, etc.

B. But this is how language is spoken casually worldwide. Standard English often comes in prose of this kind from Gibbon's *The Decline and Fall of the Roman Empire*:

The whole engagement lasted above twelve hours, till the gradual retreat of the Persians was changed into a disorderly flight, of which the shameful example was given by the principal leaders and the Surenas himself. They were pursued to the gates of Ctesiphon, and the conquerors might have entered the dismayed city, if their general, Victor, who was dangerously wounded with an arrow, had not conjured them to desist from such a rash attempt, which must be fatal if it were not successful. (Edward Gibbon, *The Decline and Fall of the Roman Empire*, 1776 [Volume I, chapter 24].)

Here, a single sentence stretches endlessly, in elaborate structure that a composition teacher would approve of.

C. But this kind of language is possible only because there is writing. Writing is conscious and slow, allowing the writer to carefully compose long sentences and the reader to process them. Spoken language occurs in real time and generally occurs in packets of, on average, seven words.

D. If language had existed for 24 hours, then writing would have existed only since about 11:08 P.M. Only about 200 out of the 6,000 languages are "written" in the true sense of being used in official documents and having a literature. The elaborate traits of written language are a historical accident.

II. Spoken versus written language.

A. *Vocabulary*. Spoken language makes use of a more limited vocabulary than written language. This is partly because writing allows the preservation of words over time. In spoken—that is, normal!—languages, old words die away.

1. The Lokele of the Democratic Republic of the Congo use a talking-drum language that has many words no one recalls the meanings of. There is no dictionary to preserve them the way *ruth*—the root of *ruthless*—is preserved in English dictionaries.

2. Spoken English makes use of a small subset of all the words in the language. Linguists Wallace Chafe and Jane Danielewicz have shown that even educated Americans use hedges to compensate for the difficulty of making maximal use of English vocabulary when speaking in real time, such as in this quote:

> She was still young enough so I… I just… was able to put her in an… uh—*sort of*… sling… I mean one of those tummy packs… you know.

Languages only used orally tend to have thousands or maybe tens of thousands of words—not the hundreds of thousands that written languages hoard in dictionaries for eternal reference.

B. *Syntax*. Spoken language uses shorter, simpler sentences than written language. This is part of a folktale narrated by a speaker of Saramaccan Creole. Because this is spoken language, the sentences are rather short.

Anasi dɛ a wã kɔndɛ.
Anancy [the spider] was in a village.

Nɔɔ hɛ̃ wɛ wã mujɛɛ bi dɛ a di kɔndɛ nããndɛ.
And a woman was in the village there.

Nɔɔ di mujɛɛ, a pali di miii wã daka.
And the woman bore a child one day.

Nɔɔ di a pali, nɔɔ dee oto sɛmbɛ u di kɔndɛ, de a ta si ɛ̃ u soni.
And when she gave birth, the other people in the village didn't want to have anything to do with her.

Hɛ̃ wɛ a begi Gadu te a wei.
Then she prayed to the gods fervently.

Hɛ̃ wɛ a go a lio.
Then she went to the river.

Nɔɔ di a go a lio, dee Gadu ko dɛ̃ɛ̃ wã mujɛɛmii.
And when she went to the river, the gods gave her a girl-child.

III. Language goes from spoken to written.

A. Even in early written English, it is clear that the writers are still writing with significant influence from how a language is used in speech. Here is a passage from the first English printed book, namely William Caxton's prologue to *The Recuyell of the Historyes of Troy*. Note that it is structured rather like the Saramaccan passage, with short phrases following one after the other:

And afterward whan I remeberyd my self of my symplenes and vnperfightnes that I had in bothe langages, that is to wete [wit] in Freshe and in Englisshe, for in France was I neuer, and was born and lerned myn Englissh in Kente in the Weeld, where I doubte not is spoken as brode and rude Englishh as is in ony place of Englond; & haue contynued by the space of xxx yere for the most parte in the contres of Braband, Flandres, Holand, and Zeland;… (William Caxton's prologue to *The Recuyell of the Historyes of Troy*, cited in: Crystal, David. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 1995, p. 57.)

B. A passage like this reflects the almost sobering reality of how we speak English, rather than write it. This is an exchange between two students in the 1970s, and one must admit that this indeed reflects casual spoken English, as opposed to how we write it:

A. On a tree. Carbon isn't going to do much for a tree really. Really. The only thing it can do is collect moisture. Which may be good for it. In other words in the desert you have the carbon granules which would absorb, collect moisture on top of them. Yeah. It doesn't help the tree but it protects, keeps the moisture in. Uh huh. Because then it just soaks up moisture. It works by the water molecules adhere to the carbon moleh, molecules that are in the ashes. It holds it on. And the plant takes it away from there.

B. Oh, I have an argument with you.

A. Yeah.

B. You know, you said how silly it was about my, uh, well, it's not a theory at all. That the more pregnant you are and you see spots before your eyes it's proven that it's the retention of the water.

A. Yeah, the water's just gurgling all your eyes.

(Carterette, Edward C., and Margaret Hubbard Jones. *Informal Speech*. Berkeley: University of California Press, 1974, p. 390.)

C. Only by "translation" can we transform spoken English into written, a form that would never emerge from any human being speaking any language naturally, as with this passage as presented by linguist M.A.K. Halliday:

Spoken version:

I had to wait, I had to wait till it was born and till it got to about eight or ten weeks of age, then I bought my first dachshund, a black-and-tan bitch puppy, as they told me I should have bought a bitch puppy to start off with, because if she wasn't a hundred percent good I could choose a top champion dog to mate her to, and then produce something that was good, which would be in my own kennel prefix.

Hypothetical written version:

Some eight or ten weeks after the birth saw my first acquisition of a dachshund, a black-and-tan bitch puppy. It seems that a bitch puppy would have been the appropriate initial purchase, because of the possibility of mating an imperfect specimen with a top champion dog, the improved offspring then carrying my own kennel prefix.

(Halliday, M.A.K. "Spoken and Written Modes of Meaning," in *Comprehending Oral and Written Language*, edited by Rosalind Horowitz and S. Jay Samuels. New York: Academic Press. 1987, p. 59.)

D. The roots of written language in spoken language can be seen in the earlier written documents of many languages.

1. Here is the way the opening passage of the Bible is often written:

In the beginning, when God created the heavens and the earth, the earth was a formless wasteland, and darkness covered the abyss, while a mighty wind swept over the waters. Then God said, "Let there be light," and there was light.

2. But the original Hebrew version does not scan this way at all. Instead, it is written in short sentences, reflecting spoken language:

Bereshit bara Elohim et hashamayim ve'et ha'arets.
Veha'arets hayetah tohu vavohu
vechoshech al-peney tehom veruach.

Elohim merafechet al-peney hamayim.
Vayomer Elohim yehi-or va-yehi-or.

In the beginning God created the heavens and the earth.
And the earth was formless and empty
with darkness on the face of the depths.
God's spirit moved on the water's surface.
God said, "There shall be light" and light came into existence.

The Hebrew Bible was written at a time when writing was relatively new, and the writer was still inclined to simply transcribe language as it was spoken.

IV. What we are conditioned to view as the "real" type of language is actually a technological luxury, allowed by the transcription of language onto the page. All but a few languages are used orally only, and as complex as they tend to be, they are spoken in small "word packets," juxtaposed with a certain freedom that relies on context as much as structure to convey meaning and with relatively small vocabularies. The *Oxford English Dictionary* and the prose of Milton are historical curiosities, departures from the "natural," similar to dogs that bring in the newspaper.

Essential Reading:

Ong, Walter. *Orality and Literacy: The Technologizing of the Word*. London: Routledge, 1982.

Supplementary Reading:

Chafe, Wallace, and Jane Danielewicz. "Properties of Spoken and Written Language," in *Comprehending Oral and Written Language*, edited by Rosalind Horowitz and S. Jay Samuels, pp. 83–112. New York: Academic Press, 1987.

Goody, Jack, and Ian Watt. "The Consequences of Literacy," in *Literacy in Traditional Societies*, edited by Jack Goody, pp. 27–84. Cambridge: Cambridge University Press, 1968.

Halliday, M. A. K. "Spoken and Written Modes of Meaning," in *Comprehending Oral and Written Language*, edited by Rosalind Horowitz and S. Jay Samuels, pp. 55–82. New York: Academic Press, 1987.

Questions to Consider:

1. Make a tape recording of you and some friends speaking casually, and listen to how choppy and unstructured casual speech actually is. Do you and your friends talk the ways books are written?

2. Listen to a passage of a stand-up comedian and "translate" it into formal, written English. Does the passage lose something in the translation, or

would you rather that the comedian had phrased it the way you have written it?

Lecture Nineteen
Dialects—The Fallacy of Blackboard Grammar

Scope: Understanding language change and how languages differ helps us to see that many of the things that we are taught are "wrong" about speech are misanalyses. Grammarians of the 1600s and 1700s passed many of these conceptions down to us, assuming that all languages should be patterned after Latin and Greek (thus, no *Billy and me went to the store*), that language change is decay (thus requiring the retention of *whom*), and that grammar must make strictly logical sense (thus, a pox on *I ain't seen nothin'*).

Another artificial incursion into the natural history of language is that because of the influence of standard dialects, people who speak written languages are often taught that constructions that they produce spontaneously are "errors" that they must be taught out of. This is a prescriptivist approach to language, in contrast to the descriptivist approach that linguists take.

Outline

I. History of prescriptivism in English: Many of the linguistic habits we are taught to avoid were only identified as "errors" by two influential English grammars.

- **A.** Robert Lowth wrote *A Short Introduction to English Grammar* in 1762, and Lindley Murray followed in its footsteps with his *English Grammar* in 1794.
- **B.** Because English had grown from a lowly vernacular to a language of worldwide influence, Lowth and Murray saw themselves as helping prepare English for its new role by giving it more "rules." But they labored under various illusions that this course teaches us out of. The result was a realm of "blackboard grammar" caveats that, in truth, have no logical foundation.

II. Illusion 1: Latin and Greek are the "best" languages.

- **A.** Lowth and Murray thought that Latin and Greek were "better" than English because of their complex case endings. Actually, languages without endings, such as Chinese, are complex in other ways, including their tones, classifiers, sentence-final particles, and so on.
- **B.** Thus, we are taught that *Billy and me went to the store* is "wrong" because *me* is a subject. However, only sometimes do languages neatly assign pronouns according to the subject/object distinction.
  - **1.** Latin was one of those languages, where the subject *I* was *ego* and the object form was *mē*, and never would *mē* be used as a subject.

2. But in a great many languages, two forms share the subject position, depending on the type of sentence. In French, one would say *Guillaume et **moi** sommes allés au magasin*, with the object form, not *Guillaume et **je** sommes allés au magasin*. No one complains about this in French.
3. Even in English, it is impossible to apply the "subject" rule consistently. If someone asks "Who did that?" and you know that it was two people on the other side of the room, when you point them out you say "Them!" not "They!", even though it is they who did it, and thus, we are dealing with subject form.

III. Illusion 2: Language change is decay.

A. Because Modern English contrasts most immediately with Old English in having lost most of its noun and verb endings, it was natural for Lowth and Murray to suppose that language change always involves loss of features and should be resisted. We tend to harbor a similar feeling today, even though, as we have seen, languages create new material as they lose it.

B. This sense that case distinctions must be retained is why we are still taught to use *whom*.

1. Notice that we must be *taught* to use it, because otherwise, *what* and *who* are no longer marked for three cases (genitive, dative, and accusative) as they once were.
2. But we only retain *whom* because it was still perceptible in English when grammarians began standardizing it. *Whom* was actually a remnant of a full system that had died unmourned. If we are to say *Whom did he see?* then the question arises as to why we do not say *Wham did he give it?* for *Who did he give it to?*, because *wham* was the dative ("to-") form of *who* in Old English.

IV. Illusion 3: Language must be logical.

A. We are often taught that "proper" language is logical in the sense of mathematics. But this is unrealistic: all languages are full of wrinkles that do not make strict logical sense, but whose meaning is clear nevertheless. The influence of such grammarians as Lowth and Murray has sometimes shunted Standard English into unnatural detours.

B. *Double negatives*. Double negatives, such as *She ain't seen nobody*, are common worldwide: the Spaniard says ***Nunca** he visto **nada*** ("never have I seen nothing") for *I have never seen anything*.

1. Old English had double negatives:

Ic ne can noht singan.

I no can nothing sing

"I can't sing anything."

2. But in the region where Standard English happened to be developing, there was an alternative construction using forms with *any*, such as *I haven't seen anything*. Even here, though, double negatives could still be used for emphasis, even in Shakespeare, where Falstaff in *Henry IV (II)* says, "There's never none of these demure boys come to any proof" (IV.iii.97).

3. Lowth, Murray and others, however, decided that "two negatives make a positive," and gave double negatives an air of slovenliness that has been permanent. But notice that every single nonstandard dialect of English uses double negatives worldwide, as do thousands of languages!

C. *You was*. In other cases, applying logic of one sort even works against speakers trying to iron out a wrinkle in the grammar themselves.

1. There is a wrinkle in how Standard English treats *you* with the verb "to be." Why is the plural form *were* used even when *you* is singular?

I was	we were
you **were**	you were
he/she was	they were

2. Many nonstandard English dialects iron this out by using the singular form *was* when *you* refers to one person. This makes for a tidier chart:

I was	we were
you **was**	you were
he/she was	they were

3. Well into the 1800s, this was even a common construction in Standard American English. Here is a letter written by a man to his lady friend in the 1830s; the elegance of the language makes it clear that his *you was* is not a mistake, and he uses it often.

Indeed, I know not one word you did say, for I was so perfectly astonished in the first place, to see you going home without appearing even to think of me, and then when I met you at the door to find out that you **was** angry with me, I knew not what to make of it. There were many people looking at us, and I knew it. (Cohen, Patricia Cline. *The Murder of Helen Jewett*. New York: Vintage, 1998, p. 244.)

4. But Lowth and Murray considered this to be using *you* with the "wrong" form; thus, English speakers are taught out of being logical!

D. Languages simply do not make perfect sense: if we say *I am*, then why do we say *aren't I* instead of *amn't I*?

V. Artful language versus blackboard grammar.

A. Certainly there are grounds for being taught how to structure one's sentences effectively and for being taught the nuances of "written" vocabulary, such as the difference between *uninterested* and *disinterested*. However, a great deal of what we are taught as "proper" or even "better" expression is based on sheer myth.

B. Thus, we must avoid supposing that part of the natural history of language entails that in developed civilizations, decadence, democratization, and overburdened school systems lead to the language "going to the dogs." Constructions that toddlers produce naturally, and that as adults we avoid as a conditioned reflex but often slip into in unguarded moments, are natural language, not mistakes.

Essential Reading:

Bryson, Bill. *The Mother Tongue: English and How It Got That Way*. New York: William Morrow and Co., 1990 (chapter 9: "Good English and Bad").

Crystal, David. *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press, 1987 ("The Prescriptive Tradition").

Supplementary Reading:

Pinker, Steven. *The Language Instinct*. New York: HarperPerennial, 1994 (chapter 12: "The Language Mavens").

Questions to Consider:

1. Since the Middle Ages, English speakers have been using such sentences as "Tell each student that they can hand in their paper at the office," rather than "Tell each student that he can hand in his paper at the office," in formal writing. Yet we are often told that this is "wrong." If Italians use their *lei* to mean both "she" and a formal "you," then can we uphold insisting that *they* must refer to the plural?
2. Given that saying "you was" when referring to one person would technically make more "sense" and be "clearer" than saying "you were," can you identify precisely what conditions our native sense that this would be taking it "too far"?

Lecture Twenty
Language Mixture—Words

Scope: The first language's 6,000 branches have not only diverged into dialects but have constantly been mixing with one another on all levels. The level of words is the first: most of English's vocabulary is borrowed from Viking invaders, French rulers, and Latin and Greek. This is a common situation: 30 percent of Vietnamese's words are from Chinese. Often words are borrowed as "high" versions of native ones: thus English *pig* and French *pork*. This kind of word mixture is the essence of Spanglish today, although seeing the process at close hand often occasions discomfort.

So far, I have implied that the first language has developed like a bush, with a single sprout branching into a mass of twigs decorated with leaves. But this metaphor can take us only so far, because in actuality, languages and dialects have mixed with one another constantly. The relationship between the world's languages is analogous to a stew.

Languages mix to various extents. In this lecture, we will examine how they mix on the level of words (which is only the first, and least transformative, level possible).

Outline

I. The bastard vocabulary of English.

- **A.** *The dictionary experience*. We English speakers are accustomed to finding that words in our language trace to Dutch, Greek, French, Latin, and other languages. It is almost the unexpected case that a word will simply trace directly back to Old English. Yet the Pole, for example, finds that many more of the words in his language proportionately trace back to Proto-Slavic.
- **B.** Indeed, out of all of the words in the *Oxford English Dictionary*, no less than 99 percent were taken from other languages. The relative few that trace back to Old English itself are also 62 percent of the words most used, such as *and*, *but*, *father*, *love*, *fight*, *to*, *will*, *should*, *not*, *from*, and so on. Yet the vast majority of our vocabulary originated in foreign languages, including not merely the obvious "Latinate" items, like *adjacent*, but common, mundane forms not processed by us as "continental" in the slightest.
- **C.** For example, every single word in that last sentence longer than three letters originated outside of English itself!
- **D.** Main sources of borrowed words in English.

1. *Vikings*. Vikings invaded and settled in the northern half of Britain starting in 787; they spoke Old Norse (ancestor of today's Scandinavian languages) and scattered about a thousand words into English, including such staples as *both*, *same*, *again*, *get*, *give*, *are*, *skirt*, *sky*, and *skin*.
2. *Normans*. In 1066, French speakers took over England for roughly the next 200 years and introduced no fewer than about 7,500 words, including such ordinary words as *air*, *coast*, *debt*, *face*, *flower*, *joy*, *people*, *river*, *sign*, *blue*, *clear*, *easy*, *large*, *mean*, *nice*, *poor*, *carry*, *change*, *cry*, *move*, *push*, *save*, *trip*, *wait*, *chair*, *lamp*, *pain*, *stomach*, *fool*, *music*, *park*, *beef*, *stew*, *toast*, *spy*, *faith*, *bar*, *jail*, *tax*, and *fry* that hardly feel "foreign" to us now.
3. *Latin*. The "Latinate" layer, most perceptible to us as a word class apart, came after the withdrawal of the French, with the increasing use of English as a language of learning—hence, *client*, *legal*, *scene*, *intellect*, *recipe*, *pulpit*, *exclude*, *necessary*, *tolerance*, *interest*, et alia.

E. Thus, an English that had developed without these lexical invasions would be incomprehensible and peculiar to us. For this reason, Icelanders can read literature in their language from the 1300s and Hebrew speakers can tackle Biblical Hebrew, but *Beowulf* is opaque to us.

F. Advantages and disadvantages.
1. *Advantage*. Because English is so larded with Latin and French words, we have a good head start on learning the vocabularies of French and other languages descended from Latin. This is especially true of the more formal layers of these languages, because most of our words from French and Latin entered "from above," contributed by rulers and scientists. *Association*, *opportunité*, and *présent* give us little trouble.
2. *Disadvantage*. Because so little of the Old English rootstock remains in English, there is no other language that is close enough to ours to be especially easy to learn, as Portuguese is for Spaniards, Zulu is for Xhosa speakers, and so on. Thus, if a language does not have the Latinate inheritance that Western European languages do, then we must learn both its humble and its formal vocabulary from the ground up. Russian's "bread," "water," and "fish" are *xleb*, *voda*, and *ryba*; its "association," "opportunity," and "the present" are *soedinenje*, *vozmožnost*, and *nastojaščee*.

II. Word sharing is ordinary and inevitable.

A. It is often supposed that this heavy borrowing makes English an especially "flexible" language. But all languages borrow words, usually a lot of them. Cultural disposition makes some languages more

resistant to borrowing words than others, but the space to maneuver is pretty narrow.

B. *"Real" languages as well as written ones*. For example, this borrowing does not require writing or extensive travel. In Australia, it is difficult to trace a family tree among the 260 languages originally spoken there because many have borrowed as much as 50 percent or more of their vocabularies from other Australian languages. This is partly because of widespread intermarriage.

C. *Japanese*. Japan was traditionally one of the most isolated modern cultures in the world, but over the past few decades it has inhaled countless American English words, such as *beisuboru* ("baseball"), *T-shatsu* ("T-shirt"), *sukii* ("ski"), *fakkusu* ("fax:), and *bouifurendo* ("boyfriend").

D. *High and low*. Norman French left many diglossic doublets in English, such as *pig* and *pork* and *help* and *aid*. This is common across languages.

1. *Japanese*. Japanese has thousands of Chinese-derived words, including the numbers one through four, *ichi*, *ni*, *san*, *shi*. The original Japanese numbers—*hitotsu*, *futatsu*, *mittsu*, *yottsu*—are used less, for example when giving children's ages.

2. *Vietnamese*. The Chinese occupied Vietnam for more than a thousand years, and Vietnamese is about 30 percent Chinese in its vocabulary, including doublets such as the written *hoả-xa* for "train" and the spoken native *xe lửa* meaning "train" in casual speech.

III. Word sharing and dialects: Dialects generally borrow from dialects.

A. *Doublets*. This means that a language may get two words from one, borrowing different versions of it from two dialects. *Chant* was borrowed from standard French's verb *chanter*, "to sing." But *cant*, in the sense of platitudinous talk, was borrowed from Norman French's version of the same verb, *canter*.

B. *Different dialects, different borrowings*. Scots English took on some Dutch words that dialects to the south did not. Thus, Standard English has such words as *cruise* and *easel*, but Scots has such words as *callan*, "lad," and *cowk*, "to retch." Because the Norse-speaking Viking invaders settled in what became Scotland, Scots also has a stronger Norse imprint than Standard English, such as *til* for "to," *gie* for "give," and *richt* for "right."

Thus, it is ordinary for languages to share words, and far beyond the level of obvious exoticisms, such as *sushi* and *taco*. Often, the borrowings help to trace the movement of peoples and the history of their languages.

IV. Word mixture in real life: Although it is easy to accept word mixture that happened in times long past, when we see it happening in our lifetimes, it often occasions discomfort, out of a sense that purity is compromised. But we are simply watching a time-honored process taking place.

A. *Spanglish*. When a Latino immigrant in the United States says *brecas* for "brakes," instead of the original Spanish *frenos*; or *carpeta* to refer to a rug rather than, as in original Spanish, a folder; or *Voy a manejar mi troca a la marketa* for "I'm going to drive my truck to the market," instead of *Voy a manejar mi camión al mercado*, speakers of Spanish in Spain, Mexico, and other Latin countries often see this as "polluted" Spanish. But this is as natural, and inevitable, a process as the influx of French words into English under the Norman occupation.

B. *English in the days of yore*. When the new French words were still processible as "new," there were even English speakers who decried them as "wrong." Man of letters John Cheke instructed in 1561 that "Our own tung shold be written cleane and pure, vnmixt and vnmangeled with borrowing of other tunges," following this with substituting *mooned* for *lunatic* and similar usages. (Interesting that both *pure* and *mangled* came from French!)

Essential Reading:

Bryson, Bill. *The Mother Tongue: English and How It Got That Way*. New York: William Morrow and Co., 1990.

Crystal, David. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 1995.

Stavans, Ilan. *Spanglish: The Making of a New American Language*. New York: HarperCollins, 2003.

Questions to Consider:

1. Choose a sentence or two from a magazine or newspaper, look up the etymology of each word, and see how mixed English's vocabulary is. How does this make you feel about issues of language purity?

2. If you were an official in a foreign country whose language was taking in a great many words from English, would you advise that native words be constructed to substitute for the English ones, as the French Academy does? Or would you simply allow the influx of English words? How would you defend your position in either case?

Lecture Twenty-One
Language Mixture—Grammar

Scope: Languages also mix their grammars. Yiddish is basically a dialect of German, but it has not only many words but even grammatical features from Slavic languages, such as Polish. Indian Indo-European languages, such as Hindi, place their verbs at the end of sentences because the other language family of India has the same feature. In some cases, languages mix so intimately that they become new ones, such as Media Lengua in Ecuador, which uses Spanish words with endings and word order from the local Indian language Quechua. There are no languages without at least some signs of grammar mixture.

Outline

I. Introduction.

A. Words are only the beginning of how languages mix. Languages consist not only of words but of how the words are put together: grammar. In situations where large numbers of people are bilingual, the two languages they speak often come to resemble one another on the level of sounds and sentence structure, as well as exchanging words—rather like married couples who gradually begin to look like each other over the decades.

B. This happens most readily when literacy in the language is not widespread, such that there is relatively little sense that a standard variety is "The Language." For that reason, this kind of grammar mixture has largely occurred beneath the radar screen of writing—before the last several centuries in languages familiar to most of us. Yet its impact has played a major part in determining what the world's languages are like today, especially considering that only about 200 of the world's 6,000 languages are written regularly.

II. Basic examples.

A. *Clicks in Khoi-San and Bantu*. The Khoi-San ("Bushman") languages of southern Africa are not the world's only languages with clicks. For example, some Bantu languages spoken near them have clicks: Miriam Makeba even made the clicks famous in a popular song in her native Xhosa. These Bantu languages inherited the clicks from Khoi-San languages long ago.

B. *Indo-Aryan languages*. We saw that Indo-European languages in India, such as Hindi, place the verb last in a sentence.

Hindi:

Mẽ Apu se mila tha.
I Apu with meet did

"I met Apu."

This is not an accident. Indo-European languages of Europe usually do not place their verbs at the end of the sentence or only do so optionally. Indian Indo-European languages borrowed this word order from languages of another family originally spoken in India, the Dravidian family. Below is a sentence in one of the main Dravidian languages, Kannada:

Kannada:

Avanu nanage biskeṭannu tinisidanu.
he to-me biscuit fed

"He fed me a biscuit."

C. Among linguists, it has always been known that languages regularly exchange words, but until rather recently, grammar mixture has often been treated as marginal, with basic processes of independent change seen as "basic." But it is increasingly clear that all of the languages of the world bear marks from both the words and the grammars of languages spoken close by.

III. Intertwined languages.

A. There are many languages in the world that are so mixed that they cannot be treated as either Language A or Language B; these are hybrids, in the same way that mules are neither horses nor donkeys.

B. Code-switching.

1. These languages begin with an ordinary process called *code-switching*, where speakers regularly alternate between one language and another, often within the same sentence.

2. *Nuyorican*. Here is an example of a Puerto Rican code-switching between Spanish and English in New York:

Why make Carol *sentarse atras para que* everybody
sit in back so that

has to move *para que se salga*?
so that she gets out

Code-switching is common among bilinguals worldwide. Generally, code-switchers are fully competent in both languages but switch back and forth according to topic or when a word they

are more familiar with in one language comes along and sparks a switch into that word's language.

C. *Media Lengua*. In some cases, code-switching becomes so well entrenched that a new language emerges, splitting the difference between the two languages. For example, among men in Ecuador who grew up speaking Quechua but spent long periods working in the capital Quito using Spanish, a new hybrid language called *Media Lengua*—"middle language"—emerged. Media Lengua uses Spanish words with the endings and word order of Quechua:

"I come to ask a favor."

Spanish:

Vengo para pedir un favor.
I-come for ask a favor

Quechua:

Shuk fabur-da maña-nga-bu shamu-xu-ni.
one favor ask come-ing-I

Media Lengua:

Unu fabur-ta **pidi**-nga-bu **bini**-xu-ni.
a favor ask come-ing-I

Media Lengua uses the Spanish words but with the sound system of Quechua (Quechua does not have *e* or *o*) and with its endings and its word order, where the object (here, *favor*) comes before the verb.

D. *Mednyj Aleut*. In the 1800s, Russian traders colonized the Aleut Islands off Alaska and brought Aleuts (Eskimos) to work along with them on one of the islands (Copper Island). The traders and Aleut women produced children who created a language of their own, mixing, of all things, Russian with an Eskimo language.

Languages like this are not just random mixing on the spur of the moment. Mednyj Aleut has rules. Certain verb endings, such as the one in the sentence that follows, are from Russian, as are certain pronouns. Case endings on nouns as well as nouns and verbs themselves are usually from Aleut.

Mednyj Aleut:

Ya tibe cíbux ukaɤla:ɤa:sa:**l**
I to you package bring-ed

"I brought you a package."

E. There are intertwined languages mixing Russian and the Aleut language of Eskimos, English and the Gypsy language Romani, and many others.

IV. Biological analogies.

A. I have analogized language mixture to the mating of a horse and a mule, but this implies that language mixture is exceptional and that its results are somehow deficient. But another biological analogy is more appropriate. Lynn Margulis and other biologists have called attention to the fact that symbiosis—communal, co-dependent living between different species—is central to the existence of life as we know it. Plants derive crucial nutrients via the fungi in their roots that process nitrogen for them; cows could not digest their food without the bacteria filling their stomachs; and even the organelles within cells, such as mitochondria in animals, began as independent bacteria.

B. As Margulis has it:

> In reality the tree of life often grows in on itself. Species come together, fuse, and make new beings, who start again. Biologists call the coming together of branches—whether blood vessels, roots, or fungal threads—anastomosis…. Anastomosis, although less frequent, is as important as branching. Symbiosis, like sex, brings previously evolved beings together into new partnerships. (Margulis, Lynn. *Symbiotic Planet: A New View of Evolution*. New York: Basic Books, 1998, p. 52.)

In broad view, the world's languages comprise tens of thousands of dialects harboring evidence of symbiotic matings in the past. Margulis describes anastomosis as "branches forming nets," and this analogy is so useful that it can replace the one of the flowering bush.

Essential Reading:

McWhorter, John H. *The Power of Babel*. New York: HarperCollins, 2001 (chapter 3).

Supplementary Reading:

Thomason, Sarah Grey. *Language Contact: An Introduction*. Washington, DC: Georgetown University Press, 2001.

Questions to Consider:

1. When you are learning a foreign language, it is natural to occasionally put things in ways that reflect your native language: in Spanish, *I like the book* is "to-me pleases the book": *Me gusta el libro*. But you would be less likely to say *ojo médico* for *eye doctor*. Can you think of reasons why some mistakes like this are more likely than others?

2. Although it is quite common for people to mix two languages when speaking, as code-switchers do, it is much less common for one person to speak in one language and the other to answer in another, even if both people speak both languages. (Only small immigrant children tend to do this as they begin switching from their home language to the national one.) Can you speculate about why adults are so reluctant to do this?

Lecture Twenty-Two
Language Mixture—Language Areas

Scope: When unrelated or distantly related languages are spoken in the same area for long periods, they tend to become more grammatically similar, because of widespread bilingualism. The classic case is Indo-European languages of the Balkans, which share various traits that they did not have originally. But linguists are discovering the same phenomenon across the world: languages of Southeast Asia stem from four different families but share a similar "template." Linguists are finding that the usual situation is that few new languages emerge, but the ones that exist stew together in this way; only invasions and migrations interrupt this process and create brand-new languages.

Outline

I. Grammar sharing does not occur only between pairs of languages.

- **A.** Not only do we see distinct languages with aspects of grammar that one must have borrowed from the other, but even distinct language groups or families that are so similar to one another in structure that it is clear that over time, a certain complex of grammatical traits has been shared and distributed widely, creating what is called a *language area*.
- **B.** Thus, in a language area, although it can appear that all the languages trace back to a single ancestor, in fact, they may trace to several different proto-languages. Their similarity has arisen over time from grammar mixture.

II. The Balkans.

- **A.** A classic example is the Indo-European languages in the Balkans. Romanian is a Romance language. Albanian is a highly distinct branch of its own, as is Greek. Serbo-Croatian, Macedonian, and Bulgarian are Slavic languages.
- **B.** Yet these languages share several grammatical patterns that were not initially present in most of the languages when they emerged. For example, Romance languages usually place the definite article before the noun (Spanish: *el hombre*, Italian: *il uomo*), but Romanian places its definite article after the noun: *om-**ul***.
- **C.** This placement is the result of the development of Romanian in an area where there was once a great deal of bilingualism, partly because of migrations and invasions. Some of the languages placed their definite article after the noun. Bulgarian for "the woman" is *žena-**ta***; Albanian for "the friend" is *mik-**u***. This is why Romanian for "the man" is *om-ul*.

D. Then, it is odd that Bulgarian has a definite article at all because Slavic languages usually do not (Russian has no words for *the* or *a*). Bulgarian inherited this characteristic from such languages as Romanian and Albanian.

E. This is called a *Sprachbund*—a group of languages that have become increasingly similar to one another over time because of heavy bi- or multilingualism.

III. The "Sinosphere."

A. Southeastern Asia contains several distinct language families. The southern Chinese varieties, such as Cantonese, belong to the *Sino-Tibetan* family. Thai and Laotian are members of a different family called *Tai-Kadai*. Vietnamese and Cambodian are members of yet another family, *Austroasiatic*, and there are also scattered small languages, such as Hmong, part of a family called *Miao-Yao*.

B. Yet all these languages are based on a common "game plan." We saw some of it in Cantonese in Lecture Ten, with its particles at the end of sentences that convey attitude and its classifiers used with numbers, such as our *two head of cattle* instead of *two cattle*. But there are many other features typical across these families. A language of this area tends to be tonal, to have no gender marking or case marking, to have most words consist of a single syllable instead of two or more, and so on.

C. This phenomenon can be partly explained by the fact that Chinese speakers conquered and migrated southward, lending parts of their grammar to the languages they encountered. But the process went both ways: Chinese in the south became more like the languages it encountered, as well.

D. As a result, on first glance, Cantonese, Vietnamese, Thai, and Hmong appear to trace to a common ancestor, being so unlike other language families and so similar to one another. But actually, the resemblance is due to millennia of constant grammar sharing. Linguist James Matisoff has termed this language area a "Sinosphere."

IV. The European language area.

A. Even Western Europe is a language area, although when we speak a European language and are most exposed to others, it is easy to suppose that European features are simply "normal."

B. *Articles*. For example, as normal as it seems to us for a language to have words for *a* and *the*, in fact, only about one in five of the world's languages do, with many having neither (such as the ones in the Sinosphere). Proto-Indo-European did not have words for *a* and *the*. Instead, these words developed in a great many of its children and ones of different subfamilies spoken in the same region. In addition, even

Hungarian has *a* and *the*, despite being of a different family altogether, Uralic, which elsewhere tends not to have articles. The prevalence of this feature in Western Europe is due to grammar sharing over time, between subfamilies and even families of language.

C. Another example is the *perfect* construction with *have*. To express the perfect with *have* in a sentence such as *I have sewn this dress* is almost exclusively found in Europe. Again, this was not a feature of Proto-Indo-European, yet as rare as it is in languages of the world, it has developed again and again in various of its descendants.

D. These are a few of many ways in which European languages are similar, even though Proto-Indo-European lacked the feature and the feature often appears in languages outside of Indo-European, including Finnish, Hungarian, or Basque.

V. Equilibrium and punctuation.

A. The linguist R. M. W. Dixon has argued influentially that the development of language areas is a norm. The typical situation worldwide has been that groups of languages spoken by small numbers of people have coexisted for millennia, sharing words and grammar and becoming increasingly alike. This situation is one of what Dixon terms *linguistic equilibrium*, in which it is rare that new languages develop.

B. However, invasions, migrations, and geographical upheavals sometimes lead speakers of a language to move to other regions, replacing the languages of previous inhabitants. The new groups of speakers, separated from the original ones, develop new branches of the original language in each new location. Dixon terms this *punctuation*, modeled on the evolutionary theory of paleontologists Niles Eldredge and Stephen Jay Gould that evolution proceeds in abrupt leaps rather than tiny steps. Under Dixon's theory, the branching of a language into new ones is a special circumstance, a leaping kind of change distinct from the relative stasis of an equilibrium situation.

Essential Reading:

Dixon, R. M. W. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press, 1997.

Questions to Consider:

1. Taking classes in European languages, such as French, Spanish, and German, can make it seem as if languages only vary so much from English in terms of grammar: master some endings and get used to gender marking and the rest is relatively straightforward. But increasingly, Americans are learning languages from further away, such as Chinese, Japanese, and

Arabic. If you have approached one of these, what were some of the differences in how they were put together, showing that there is a rough "European" game plan that most languages are *not* based on?

2. As linguists realize how much languages have shared grammars over the years, it is becoming increasingly clear that the comparative reconstruction method will be of little use in tracing back to proto-languages for many of the world's language families. In your opinion, does this suggest that the enterprise should be given up, or does it perhaps make the methods of the Proto-World school more attractive?

Lecture Twenty-Three
Language Develops Beyond the Call of Duty

Scope: A great deal of a language's grammar is a kind of overgrowth, marking nuances of life that many or most languages do without. Some languages require one to mark how one learned whatever one is saying; others mark possession differently, according to whether one refers to an object or a body part. Even the gender marking familiar from European languages is a frill, absent in thousands of languages. The theme is decoration over necessity.

Outline

I. Introduction.

A. A central aspect of how languages and dialects develop through time is that all of them are replete with features that, in the strict sense, they do not need. This is important to realize not only for the sheer wonder of it, but also because an awareness of it sheds insight on how languages' structure is determined in part by their history, which we will explore in later lectures.

B. For example, the *have*-perfect is not only rare across languages but unnecessary. The perfect merely implies that something that happened in the past is still relevant in the present, and a great many languages leave that semantic shade to context. When Dorn at the end of Chekhov's *The Seagull* says, "Konstantin Gavrilovich has shot himself," in Russian, it simply translates as "Konstantin Gavrilovich shot himself"—that the event has ongoing implications is quite clear from context.

II. Evidential markers.

A. In many languages, when one states something, one must also indicate how one learned the statement, through seeing, hearing, general sense, or the like. In English, we can say *I saw that they are tearing down the building*, but it is quite proper to just say *They are tearing down the building*. In many languages, such a sentence would be as incomplete as *They tearing down building* would be to us.

B. Tuyuca is spoken in the Amazon and has several such markers:

Evidential markers in Tuyuca:

Kiti-gï tii	-gí	"He is chopping trees" (I hear him)
	-í	"He is chopping trees" (I see him)
	-hòi	"Apparently he is chopping trees" (I can't tell)
	-yigï	"They say he chopped trees"
chop trees-he	AUX	

C. The way these markers develop is through the grammaticalization that we examined in Lecture Four. For example, in the North American Native American language Makah, to say that from what one sees, the weather is bad, one says, "it's bad weather," then a suffix is added, meaning that the statement is based on seeing something, *-pid*. The *-pid* started out as a separate verb meaning "it is seen" but eroded and grammaticalized into becoming an evidential marker.

Makah:

wikicaxak-**pid**
"It's bad weather—from what it looks like." ("Looks like bad weather.")

D. Importantly, evidential markers like this are not necessary. In all languages, one might specify how one learned something, but it is a frill to have to indicate it as an obligation. Grammaticalization has a way of taking a ball and running with it: what begins as an indication of something concrete and necessary often devolves into a useless habit.

III. Alienable possession.

A. One *has* one's ear in a different way than one *has* a table, and *has* one's relative in a different way than one *has* a car. In English, we use the same word *have* for both conceptions, but just as often, languages mark this subtle difference.

B. In Mandinka in West Africa, for example, to say "your father," one says *i faamaa*, but to say "your well," one says *i* ***la*** *koloŋo*. The *la* particle signals that something is possessed in the "table" way instead of the "ear" way. Linguists differentiate these concepts as *alienable possession* (the table kind) and *inalienable possession* (the ear kind).

Mandinka

i faamaa "your father"

i la koloŋo "your well"

IV. Inherent reflexive marking.

A. To avoid a sense that this is just a trait of exotic languages, we return to a European frill of this kind. In English, one can say *I wash myself*—this is normal *reflexivity*, marking an action that one performs upon oneself. But in other European languages, one pays much closer attention to whether an action occurs upon oneself. In Spanish, *yo* ***me*** *siento* means "I sit myself down." In French, *je* ***me*** *fâche* is "I anger myself"; in German, *ich erinnere* ***mich*** is "I remember myself."

B. Here, the literal kind of reflexivity that English has went a step further. Now, the reflexive marker is grammaticalized as a way of indicating even the slightest degree to which one could conceive of an action as happening to a person rather than being effected by the person on something or someone else.

V. Gender marking.

A. In European and many other languages, nouns are divided into gender classes. Spanish has masculine and feminine, marked with an article and often with the final vowel: ***el*** *sombrer****o***, ***la*** *cas****a***. German has three: "the spoon," "the fork," and "the knife" are ***der*** *Löffel*, ***die*** *Gabel*, and ***das*** *Messer*. This is not necessary in a language: it is an accident of history.

B. *Stage one*. In many languages, we can see how this marking begins. In Dyirbal, spoken in Australia, all nouns must be preceded by a separate word. Which word a noun takes depends on which of four categories it fits into. One is for males and animals, another for female things, another for food that is not flesh, and another is the grab bag.

Dyirbal gender classifiers:

	MARKER	EXAMPLE
masculine, animals	bayi	bayi yaṛa "man"
feminine	balan	balan gabay "girl"
nonflesh food	balam	balam gayga "cake"
grab bag	bala	bala yugu "wood"

C. *Stage two*. Over time, separate words such as these erode and become prefixes or suffixes—grammaticalization again. At first, the new prefixes or suffixes still correspond fairly well to categories. Swahili is at this stage. Swahili has seven "genders" (although because sex is not one of the categories marked, linguists call them *noun classes*). The one with an *m-* prefix contains people: ***m****tu*, "man"; ***m****toto*, "child." The one with an *n-* prefix contains animals: ***n****dege*, "bird"; ***n****zige*, "locust."

D. *Stage three*. But as time goes on, sound change, cultural changes, eccentric semantic switches (such as the one that made the word for *sister-in-law* masculine in Proto-Indo-European), and other processes

make the correspondence between marker and category increasingly vague. European languages are an example of this stage, where only marking actual male beings masculine and actual female beings feminine makes any immediate sense anymore.

VI. Thus, a great deal of what a language's grammar pays attention to is technically a kind of window dressing. Keep in mind that there are actually some languages that do not mark tense at all, and some where *I* and *we* are the same word, *he* and *they* are the same word, and so on, because pronouns mark person but not number! This shows that it is inherent to human language to overelaborate.

Essential Reading:

McWhorter, John H. *The Power of Babel*. New York: HarperCollins, 2001 (chapter 5).

Questions to Consider:

1. To people speaking languages with alienable possessive marking, English appears rather crude in making no distinction between *my mother* and *my chair*. Then to us, a language that cannot distinguish *Elvis left the building* from *Elvis has left the building* seems somehow impoverished. Think of some meanings that a foreign language you are familiar with marks that English doesn't and some distinctions English marks that the foreign language does not.

2. Often, languages differ not in simply whether they express a concept or not, but in how obligatory it is to express the concept. English, for example, does not have obligatory evidential marking, as Tuyuca does, but we do express such things if necessary. Think of some words, expressions, or even intonations English uses to indicate sources of information.

Lecture Twenty-Four
Language Interrupted

Scope: Generally, a language spoken by a small, isolated group will be much more complicated than English. In fact, it is typical for languages to be vastly overgrown in this way. Languages are "streamlined" when history leads them to be learned more as second languages than as first ones, which abbreviates some of the more difficult parts of their grammars. This means that a language such as Tsez of the Caucasus Mountains in Asia is so complex that one wonders how it could be learned, while one variety of Indonesian created by adult speakers of local languages is so streamlined that one wonders how it conveys enough meaning to be useful! Such languages as English and Mandarin Chinese are intermediate cases.

Outline

I. Introduction.

A. Now that we have seen that languages tend naturally to develop beyond what is necessary to communication, we are in a position to begin examining how languages' complexity can differ depending on historical circumstance.

B. One lesson I have tried to convey is that there are no "simple" languages in the world, even when they do not have tables of endings as European languages tend to. A language without endings will usually have tones like Chinese. Overall, there are many ways for a language to be complex beyond even tones, such as the classifiers, evidential markers, distinctions between shades of possession, and other features we have seen in these lectures.

C. Linguists often remind students and readers that all languages are complex, with the implication that all languages are equally complex. This, however, is not quite true. In reality, many languages are more complex than others.

D. It is natural to suppose that the more advanced a society is, the more complex its language will be. And it is true that only a language with a history of writing can amass an enormous vocabulary. But a language is not only its words but also its grammar, and usually, a language spoken by a small, preliterate group is more complicated than English, Spanish, Japanese, or other First World languages.

II. How complex can languages get? The case of Tsez.

A. Tsez is spoken in the Caucasus Mountains in Asia by about 14,000 people. It does not have a large written literature: it is mostly a spoken language.

B. In Tsez, there are four "genders" of noun. There is a masculine class and a feminine one. But the feminine gender also contains objects that are flat or pointed (go figure). Another gender has many animals but also lots of other things, and the fourth one has various other inanimate objects.

C. The gender marker is not attached to the noun but to the verbs, adjectives, adverbs, or prepositions associated with the noun. Here, for example, we see three of the gender markers on verbs following the noun.

eniy **y**-ˤuƛ'-no "the mother feared"

buq **b**-ajnosi "the sun rose"

tatanu ɣudi **r**-oqxo "the day warmed up"

D. But then, there is a bizarre wrinkle—the gender markers are only used when the word begins with a vowel! If it begins with a consonant there is no marker. This means that, in a way, the exception is the rule:

kid **y**-iys "the girl knows"

kid **__**-božizi yoq-xo "the girl believes"

E. Tsez also has many case markers, like Latin. But these are often extremely irregular, as if such differences as *children* versus *child* and *people* versus *person* were typical of hundreds of nouns in English. The word for *fish* is regular, but look what happens when the same endings are added to the words for *tongue* and *water*. These things must simply be learned by rote:

besuro "fish"	giri "tongue"	ɬi "water"
besuro-**s** "the fish's"	giri-**mos** "the tongue's"	ɬ-**ās** "water's"
besuro-**bi** "fishes" (fish)	giri-**mabi** "tongues"	ɬ-**idabi** "waters"

F. In addition, Tsez has a trait common in small languages: a subject takes an ending when it has an object but not when it doesn't. Therefore, to say *The girl knows* is one thing, but to say *The girl washed the dress* means putting a special ending onto the word for *girl*! This is called *ergativity*.

kid y-iys
"the girl knows"

kid-**ba** ged esay-si.
"girl-ERG dress washed"

"The girl washed the dress."

G. Finally, Tsez is full of unusual sounds, many made back in the throat, with fine variations on these to boot, including mixtures of them.

H. And of course, there are, as in all languages, exceptions galore to the rules, plus all kinds of other complications (for example, Tsez has evidential marking). Yet people speak this language without effort every day. This is what "real" languages are like. We find similarly complex grammars in languages spoken by small tribes in the Amazon and many other locations. It has been said that Native American languages, such as Cree and Ojibwa, are so complex that children are not fully competent in them until the age of 10.

III. How simple can languages get? The case of Riau Indonesian.

A. A contrasting case is a dialect of Indonesian spoken in Sumatra, called Riau Indonesian. Standard Indonesian appears "normally" complex to the English speaker, with a certain number of prefixes and endings, a set word order, and so on. But while Tsez makes one wonder how people could speak it without having a stroke, Riau Indonesian makes one wonder how one could speak it and even be understood.

B. This is a dialect spoken by human beings every day that has no endings, no tones, no articles, and no word order at all. Sentences are only placed in time if context alone does not make it clear, and even then, only with such words as *already* and *tomorrow*, not with special endings or words used only to mark tense. There is no verb "to be." The same word means *he*, *she*, *it*, and *they*.

C. This means that a sentence in Riau Indonesian can have endless meanings according to context. For example, *ayam* means *chicken* and *makan* means *eat*. The sentence *ayam makan* can mean, "The chicken is eating," "The chicken ate," "The chicken will eat," "The chicken is being eaten," "The chicken is making somebody eat," "Somebody is eating for the chicken," "The chicken that is eating," "Where the chicken is eating," "When the chicken is eating," "How the chicken is eating," and so on.

D. But this simplicity is not connected to the fact that its speakers are not First Worlders. Riau Indonesian developed among people who spoke various languages related to Indonesian in Sumatra as first languages and learned Indonesian as a second one. Their first languages are "typical" in complexity, with very complex prefixes, and so on. But as

is common among adults, when these people learned Indonesian as a second language, they did not acquire it completely. This is especially common when people learn a language outside of the school setting. Children born into a society where most people are speaking a language incompletely learn that variety and pass it down the generations.

IV. An intermediate case: Mandarin Chinese.

A. It is common worldwide for a language to be streamlined somewhat when at one point, more people learn it as a second language than as a first one. Languages like this are less imposingly complex than a language such as Tsez.

B. This is true of Mandarin Chinese in comparison to other Chinese languages, such as Cantonese. Mandarin has four tones; Cantonese has six (or depending on how one counts, nine). A Mandarin word can end only in *n* or "ng"—there is no such word in Mandarin as *fap* or *fam*. But a Cantonese word can end in six different consonants, *p*, *t*, *k*, *m*, *n*, and "ng." Cantonese has about 30 of the sentence-final particles that convey attitude; Mandarin has only about a half dozen of these. Mandarin is the "easy" language among the Chinese group.

C. In antiquity, the northern part of China where Mandarin is spoken was ruled by people speaking such languages as Mongolian and Manchu. These people learned Mandarin as a second language and passed this "learner's variety" down the generations. Chinese developed "normally" in the south and became such varieties as Cantonese and Taiwanese. In the north, Chinese was, as it were, "semi-Riau-ized."

V. Other cases: Many languages have undergone what Mandarin did. Swahili is one of the only Bantu languages out of more than 500 that has no tones, and this is because only a small number of Muslim people on the east African coast use it as a first language. For centuries, Swahili has been east Africa's main lingua franca, learned by most of its speakers as a second or third language. This has rendered it less Tsez-esque than the other Bantu languages.

VI. Our lesson is that it is normal for languages to be awesomely complex, regardless of the societal level of advancement of their speakers. What is unusual is when languages are less complex than these tribal ones. Languages get "shaved down" when history leads them to be spoken more as second languages than as first ones. We are now in a position to understand some aspects of English better, then to proceed to pidgin and creole languages.

Essential Reading:

McWhorter, John H. *The Power of Babel*. New York: HarperCollins, 2001 (chapter 5).

Questions to Consider:

1. If languages tend to be more complex in smaller, isolated societies, then this suggests that the languages that are spreading to millions of people scattered across vast areas will be increasingly simple in their structure. Is this a good thing or a bad one?
2. Compared to its close relatives, such as German and Swedish, English is rather streamlined: for example, in the present tense, it has but one ending, third person singular *-s*. Some linguists have supposed that this was just an accident, but what alternative analysis might this trait of English suggest?

Timeline

150,000–80,000 B.C.	Estimated time during which human language arose
4000 B.C.	Probable origin of Proto-Indo-European
3500 B.C.	First attested writing
3000 B.C.	Probable origin of Semitic
2000 B.C.	Bantu speakers begin migrations south and eastward
A.D.	
450–480	First attestation of English
787	First Scandinavian invasions of England
mid-1300s	Beginning of the standardization of English
1400	Beginning of the Great Vowel Shift in English
1564	Birth of William Shakespeare
c. 1680	The origin of Saramaccan creole
1786	Sir William Jones gives first account of Proto-Indo-European
1887	Ludwig Zamenhof creates Esperanto
c. 1900	The birth of Hawaiian Creole English
1916	Discovery of Hittite

Glossary

Algonquian: Family of Native American languages spoken in Canada and the northern and northeastern United States, including Cree, Ojibwa, Shawnee, Blackfoot, Fox, and Kickapoo. Much work has been done on the reconstruction of Proto-Algonquian.

alienable possessive marking: Distinguishing things possessed as objects (alienably) from those possessed as parts of one's body or as personal intimates (inalienably), e.g., *my chair* versus *my mother*. Many languages have different possessive pronouns for these two situations or distinguish between them in various other ways.

Amerind: One of the three families into which Joseph Greenberg divided the notoriously variegated hundreds of Native American languages. Amerind is by far the biggest of the families, comprising most of the languages native to the Western Hemisphere.

Areal: Of or pertaining to an area or region.

assimilation: The tendency for a sound to become similar to one adjacent to it: Early Latin *inpossibilis* became *impossibilis* because *m* is more like *p* than *n* is, in requiring the lips to come together.

Austroasiatic: The Southeast Asian language family that includes Vietnamese and Khmer (Cambodian).

Austronesian: The massive Southeast Asian and Oceanic language family that includes Tagalog (Filipino), Indonesian, Javanese, Malagasy, and Polynesian languages, such as Hawaiian and Samoan.

Baltic: The small subfamily of Indo-European today including only Lithuanian and Latvian, the closest languages in the family to the Proto-Indo-European ancestor.

Bantu: The 500 languages spoken in sub-Saharan Africa, of which Swahili and Zulu are the best known; a subfamily of the Niger-Congo family.

Broca's area: The area of the brain, above the Sylvian sulcus on the left side, that is thought to control the processing of grammar.

Celtic: The subfamily of Indo-European including Irish Gaelic, Welsh, and Breton, all now under threat; the family once extended across Europe.

Chinook Jargon: The pidgin based on Chinook and Nootka with heavy admixture from French and English, used between whites and Native Americans in the Pacific Northwest, most extensively in the 19th century.

classifiers: Equivalents to *head* in such English expressions as *three head of cattle*, used more regularly in many languages, usually after numerals, and

varying according to shape or type of noun (long, flat, round, and so on). Many languages, such as Chinese ones, have dozens of such classifiers.

code-switching: When speakers regularly alternate between two languages while speaking, including in the middle of sentences.

comparative reconstruction: The development of hypothetical words in a lost proto-language of a family of modern languages through comparing the words in all the languages and deducing what single word all could have developed from. This is also done to reconstruct prefixes, suffixes, and sentence structure.

creole: The result of the expansion of a reduced version of a language, such as a pidgin, into a full language, which usually combines words from a dominant language with a grammar mixing this language and the ones the creole's creators spoke natively.

creole continuum: The unbroken range of varieties of a creole extending from one sharply different from the language that provided its words ("deep" creole) to varieties that differ from the dominant language largely in only accent.

critical-age hypothesis: The observation that the ability to acquire language flawlessly decreases sharply after one's early teens, first explicated by Eric Lenneberg in 1967 but since then referred to extensively by the Chomskyan school as evidence that the ability to learn language is innately specified.

diglossia: The sociological division of labor in many societies between two languages, or two varieties of a language, with a "high" one used in formal contexts and a "low" one used in casual ones. The classic cases are High German and Swiss German, practically a different language, in Switzerland, and Modern Standard Arabic, based on the language of the Koran, and the colloquial Arabics of each Arabic-speaking region, such as Moroccan and Egyptian, which are essentially different languages from Modern Standard and as different from one another as the Romance languages

double negative: The connotation of the negative in a sentence via two negator words: *I ain't seen nothing*.

Dravidian: A family of languages spoken mostly in southern India, including Tamil and Kannada, separate from the Indo-Aryan languages spoken elsewhere in the country.

equilibrium (vs. punctuation): A state when many languages share space in constant contact with one another, with no language threatening any other one to any significant extent over a long period of time. Linguist R. M. W. Dixon proposes this as human language's original state, contrasting with *punctuation* in which speakers of one language migrate and conquer other peoples, spreading their language across large areas.

ergativity: The condition in which a language marks subjects with different prefixes, suffixes, or separate particle words depending on whether the subject

acts upon something (*He kicked the ball*) or just "is" (*He slept*). In ergative languages, if the subject does not act upon something, it takes the same marker as the object, while subjects that act upon something take a different marker. Ergativity is rather as if in English we said *Him saw* instead of *He saw* in a sentence without an object, but then said *He saw her* when there was an object

Esperanto: A language created in the late 19th century by Ludwig Zamenhof, who hoped it would help foster world peace; comprised largely of words and grammar based on Romance languages but made maximally simple. Esperanto has been the most successful of many artificial languages.

Eurasiatic: A "superfamily" proposed by Joseph Greenberg comprising Indo-European, Uralic (e.g., Finnish and Hungarian), Altaic (e.g., Turkish, Mongolian), Dravidian, Kartvelian (of the Caucasus mountains), Afro-Asiatic (e.g., Arabic, Hausa), Korean, Japanese, Chukchi-Kamchatkan (of eastern Russia), and Eskimo-Aleut. The Eurasiatic hypothesis differs from the Nostratic hypothesis in that the latter is based on comparisons of the families' proto-languages while the former is based on more general cross-family comparisons.

evidential markers: Markers that indicate how one learned a fact being stated (i.e., seen, heard, suspected, and so on); all languages have ways of expressing such things, but in some languages, one *must* express them with each sentence.

FOXP2 gene: The gene that is connected to humans' ability to speak, also found in slightly different form in chimpanzees and found to be damaged in a family in which a speech defect (specific language impairment) was common.

gender marking: The distribution of nouns into two or more classes, masculine and feminine usually included; the term usually refers to this as applied to inanimate objects, as well as animate ones, such as German's d*er Löffel*, d*ie Gabel*, and d*as Messer* for the spoon, the fork, and the knife.

Germanic: A subfamily of Indo-European including German, Dutch, Yiddish, Swedish, Icelandic, and English, distinguished by how very close Icelandic is to Proto-Germanic and how strikingly far English is from it.

grammatical words (vs. concrete words): Words that have no concrete essence but perform grammatical functions in a sentence, such as *would* or *then* or, well, *or*. These are as crucial as concrete words in making human language what it is.

grammaticalization: The development of a word from a concrete one into a grammatical one over time, such as French's *pas* from meaning "step" to "not." Grammaticalization is how most grammatical words, as well as prefixes and suffixes, come into being.

Great Vowel Shift: The transformation of many English vowels into other ones in the 1400s, before which many English spelling conventions had already gelled. This is why *made* is spelled as if it were pronounced "MAH-deh," which at a period before the Great Vowel Shift, it was.

Grimm's law: A curious transformation in the consonants of Proto-Germanic, in which Proto-Indo-European *p* became *f* (hence, Latin *pater*, English *father*), *t* became *th* (Latinate *tenuous*, original English *thin*), and so on.

Indo-Aryan: The subfamily of Indo-European including Hindi, Bengali, Gujarati, and other languages descended from Sanskrit.

Indo-European: The language family now occupying most of Europe, Iran, and India, likely originating in the south of present-day Russia; its proto-language has been reconstructed, called Proto-Indo-European.

Indo-Pacific: The family of languages including the several hundred spoken on New Guinea and some others spoken on nearby islands; the group is often termed Papuan. Relationships among the languages have only begun to be worked out.

inherent reflexive marking: The extension of reflexive marking (*I hurt myself*) to verbs indicating emotion, movement, and other processes done to or occurring within one's self: German *ich erinnere mich*, "I remember myself," for "I remember"; similarly, French *je me souviens*. Especially common in Europe.

intertwined language: Languages developed by people with a bicultural identity that neatly combine the grammatical structure of one language with words from another one, in various fashions; e.g., Media Lengua and Mednyj Aleut.

Italic: The subfamily of Indo-European that included Latin and is now represented by the Romance languages; Latin's relatives, such as Oscan and Umbrian, are long extinct.

Khoi-San: The family of languages spoken in regions of southern Africa best known for their click sounds; perhaps the world's most ancient language family.

laryngeals: The breathy sounds reconstructed by Ferdinand de Saussure as having existed in Proto-Indo-European, to explain why many of its reconstructed roots were "open-ended" ones with a long vowel and no final consonant. De Saussure was proven correct when such sounds occurred in the places he predicted in Hittite, an extinct Indo-European language discovered in documents in the early 20th century.

Media Lengua: An intertwined language spoken in Ecuador, with Quechua endings and word order and Spanish words.

Mednyj Aleut ("middle" Aleut): An intertwined language, now basically extinct, spoken by children of Russian traders and Aleut women on one of the Aleutian islands starting in the 19th century.

Miao-Yao: A family of languages spoken by isolated groups in South Asia, including Hmong. Presumably, the family was much more widespread before Chinese peoples migrated southward.

Moldovan: A variety of Romanian spoken in Moldova, a country adjacent to Romania formerly incorporated into the Soviet Union. Only this history leads Moldovan to be considered a separate language from Romanian in any sense.

Normans: The French people who took over England in the 11th century, speaking the Norman dialect of French, which profoundly influenced the English vocabulary. Norman was derived from Norsemen, that is, Vikings.

Nostratic: A "superfamily" proposed by Russian linguists Aron Dolgopolsky and Vladislav Illich-Svitych comprising Indo-European, Uralic (e.g., Finnish, Hungarian), Altaic (e.g., Turkish, Mongolian), Dravidian, Kartvelian (of the Caucasus mountains), and Afro-Asiatic (e.g., Arabic, Hausa). See also Eurasiatic.

particle: A short word that is not an ending or a prefix that has a grammatical function.

perfect construction: A construction separate from the ordinary past one, connoting that a past event still has repercussions in the present. *I have decided not to take the job* implies that the impact of the decision is still ripe; *I decided not to take the job* sounds more like recounting a long-past occurrence. This is especially common in Europe.

pidgin: A makeshift, reduced version of a language used by people with little need or inclination to master the language itself, usually for purposes of trade. If used as an everyday language, a pidgin can become a real language, a creole.

poverty of the stimulus: The Chomskyan argument that actual speech is full of mistakes and hesitations and rarely offers demonstrations of various rules of a language that children nevertheless master early; Chomsky and others argue that this supports the idea of language as an innate faculty.

prescriptivism (vs. descriptivism): The school of thought that proposes how language ought to be (e.g., *Billy and I went to the store* is "better" than *Billy and me went to the store* because *I* is a subject), as opposed to the descriptivist approach, which simply describes how language is naturally (the latter fundamental to academic linguistics).

Provençal: The Romance variety of southern France closely related to French. Formerly the vehicle of the music of the troubadours, now represented by modern relatives, such as Occitan, threatened by French.

rebracketing: The redrawing of boundaries between words or parts of words as a result of plausible mishearings, such as *nickname* developing when speakers heard the original word *ekename* used after an indefinite article: *an ekename* became *a nickname*.

Riau Indonesian: A colloquial dialect of Indonesia spoken on the island of Sumatra with unusually little overt grammatical apparatus, leaving more to context than most known languages.

Russenorsk: A pidgin spoken especially in the 1800s between Russians and Norwegians trading during summers, neatly splitting the difference between Russian and Norwegian.

Sapir-Whorf hypothesis: An idea developed especially by Benjamin Lee Whorf speculating that differences between languages' grammars and vocabularies may channel how their speakers think, creating distinct views of the world.

Saramaccan: A creole language spoken in the Suriname rain forest by descendants of slaves who escaped into the interior and founded their own communities; the creole mixes words from English, Portuguese, Dutch, and the African languages Fongbe and Kikongo and has a grammar highly similar to Fongbe's.

Schwäbisch: A dialect of German spoken in the south of Germany, one of the many that is different enough from High German as to essentially be a different language.

semantic broadening: The development over time of a word's meaning into one more general: *bird* once referred to small birds but now refers to all birds.

semantic drift: The tendency for words' meanings to morph gradually over time to the point that the distance between the original meaning and the current one can be quite striking: *silly* used to mean *blessed*.

semantic narrowing: The development over time of a word's meaning into one more specific: *hound* once referred to all dogs but now refers to only a subset of them.

semi-creole: Languages not quite as different from a standard one as a creole is but more different than the typical dialect of that standard language. The French of Réunion Island, further from French than, for example, Canadian French but hardly as different from it as Haitian Creole, is a typical semi-creole.

Semitic: A language family spoken in the Middle East and Ethiopia including Arabic, Hebrew, and Amharic; most famous for its three-consonant word skeletons (K-T-B means "write" in Arabic; thus, *kataba*, "he wrote"; *maktab*, "office"; and so on).

Sinosphere: Linguist James Matisoff's term for the language area in Eastern and Southeastern Asia, where several separate language families have come to share several structural traits, such as tone, over the millennia because of constant contact.

Sino-Tibetan: A language family including Chinese, Tibetan, Burmese, and many other languages spoken in Southern and Southeast Asia; tone is common in the family.

sound shift: The tendency for sounds to change their articulation gradually and become new ones; the Great Vowel Shift in English is one example, as is the increasingly common pronunciation of *aw* as *ah* in America (*rah fish* instead of *raw fish*).

specific language impairment: The condition discovered in an English family in the 1980s, in which sufferers spoke rather slowly and hesitantly and often made errors usually made by foreigners. Those afflicted were found to have a faulty FOXP2 gene.

Sprachbund: An area where separate languages have come to share many grammatical features as the result of heavy bi- and multilingualism over time. A classic case is found in the Balkans, where Albanian, Romanian, Serbo-Croatian, Macedonian, Bulgarian, and Greek have become a Sprachbund. Of late, the term *language area* is becoming increasingly prevalent.

standard dialect: The dialect out of language's many that happens to become the one used in writing and formal situations, typically developing a larger vocabulary and norms for written, as opposed to spoken, expression.

SVO: The word order subject-verb-object, such as in English; SOV order is actually more common worldwide.

Tai-Kadai: A language family of Southeast Asia including Thai, Laotian, and lesser known languages, such as Shan.

Tocharian: An extinct Indo-European language once spoken by white peoples who migrated eastward to China, known from Buddhist manuscripts discovered in Central Asia.

Tok Pisin: An English pidgin spoken in Papua, New Guinea, now spoken as a native language by many and, thus, a creole; one of the few such languages used commonly in writing and in the government.

Tsez: A language spoken in the Caucasus Mountains in Asia, typical of languages in this area in having an extremely complex system of sounds and grammar.

Volapük: An artificial language created by Johann Schleyer in the 19th century based on a European pattern; initially popular but less user-friendly than Esperanto, which quickly replaced it as the most popular artificial language.

Wernicke's area: The area of the brain, below the Sylvian sulcus, that is thought to control the processing of meaning.

Bibliography

Abley, Mark. *Spoken Here: Travels among Threatened Languages*. Boston: Houghton Mifflin, 2003. Abley subscribes too much to the Sapir-Whorf perspective for my taste, but this book provides vivid descriptions of assorted language revival movements, giving nicely balanced assessments of their likelihoods of success.

Arlotto, Anthony. *Introduction to Historical Linguistics.* Boston: University Press of America, 1972. An especially clear introduction to comparative reconstruction of proto-languages, often assigned in undergraduate courses some years ago. Newer books in the vein have come along, but this one is worth seeking in a library for its conciseness because the newer ones cover the historical linguistics field more broadly.

Bailey, Richard. *Nineteenth-Century English*. Ann Arbor: University of Michigan Press, 1996. A useful examination of how English just a little more than a century ago was more different from today's than one might suppose. The chapter on slang also gives a useful portrait of the "underbelly" of English so difficult to glean from most writings before the 1960s.

Baker, Mark. *The Atoms of Language*. New York: Basic Books, 2001. This complements Steven Pinker's *The Language Instinct* in describing an area of inquiry pursued by syntacticians working in the Chomskyan school in accessible terms. Pinker's classic is, ultimately, somewhat challenging in its length, while this one hews to a more compact point.

Barber, Elizabeth Wayland. *The Mummies of Ürümchi.* New York: W.W. Norton & Co., 1999. An accessible account of the discovery of historical evidence of the Tocharian-speaking people, knitting the linguistic issues into archaeology and history.

Baugh, A. C., and T. Cable. *A History of the English Language*. Englewood Cliffs, NJ: Prentice-Hall, 1978. One of those deathless staple sources, a standard, accessible history of English for those hungry for the details but not the trivia.

Bickerton, Derek. 1990. *Language and Species.* Chicago: University of Chicago Press, 1990. Argues that human language began with a "proto-language" substrate now preserved in the language ability of apes, infant speech, and pidgins, incorporating the author's pioneering theories about the birth of an English creole in Hawaii (there termed "Pidgin").

———. 1995. *Language and Human Behavior.* Seattle: University of Washington Press, 1995. A collection of lectures filling out the author's theory about human "proto-language" and its implications for how language began.

Bodmer, Frederick. *The Loom of Language*. New York: W.W. Norton, 1944 (paperback edition, 1985). Getting a little long in the tooth—not much on Third World languages—but remains a valuable compendium of data on many of the

world's "grand old" languages, with a comparative focus. Still in print after 60 years for a reason.

Bryson, Bill. *The Mother Tongue: English and How It Got That Way*. New York: William Morrow and Co., 1990. Unsurpassed as a jolly, often laugh-out-loud trip through the history of English. Baugh and Cable will give the details, but this is a great introduction.

Burgess, Anthony. *A Mouthful of Air: Language, Languages…Especially English*. New York: William Morrow and Co., 1992. Burgess intended his tour of the world's languages as a primer for teaching us how to master them. I fear he was a bit idealistic on that goal, but he was a marvelous tour guide nonetheless and was less obsessed with Europe than writers in his vein back in the day.

Calvin, William H., and Derek Bickerton. *Lingua ex Machina: Reconciling Darwin with the Human Brain*. Cambridge, MA: MIT Press, 2000. A leading neurophysiologist and a specialist in language origins join forces in an engaging discursive exchange about how language began, within the framework of modern syntactic theory. Both are born teachers—a nice ride.

Cavalli-Sforza, Luigi Luca, and Francesco Cavalli-Sforza. *The Great Human Diasporas.* Cambridge, MA: Perseus Books, 1995. A general-public summary of what Luigi Cavalli-Sforza has discovered about human migrations in antiquity, using relationships between language families as support.

Chafe, Wallace, and Jane Danielewicz. "Properties of Spoken and Written Language," in *Comprehending Oral and Written Language*, edited by Rosalind Horowitz and S. Jay Samuels, pp. 83–112. New York: Academic Press, 1987. This article illuminates in clear language the differences—often shocking—between how we actually talk and how language is artificially spruced up in even casual writing, showing that spoken language, despite its raggedness, has structure of its own.

Comrie, Bernard, Stephen Matthews, and Maria Polinsky, eds. *The Atlas of Languages*. New York: Facts on File, 1996. One of several tours of the world's languages now available, especially useful for its maps, charts, and diagrams; attractively laid out. A nice introduction to the Indo-European languages, including the folk tale in full.

Crystal, David. *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press, 1987. An invaluable encyclopedia, lavishly illustrated, on anything one might want to know about language and languages. This selection has been at arm's length from my desk for 10 years now.

———. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 1995. A magnificent, almost imposingly rich trip through English past and present in all of its facets, as beautifully illustrated as the volume described directly above. Captures between two covers a magnificent volume of information, much of it otherwise hard to access.

———. *Language Death*. Cambridge: Cambridge University Press, 2000. The crispest and most down-to-business of the various treatments of this topic released recently, by an author personally familiar with the travails of the Welsh revival movement.

Dalby, Andrew. *Dictionary of Languages*. New York: Columbia University Press, 1998. A feast of information on any language one might want to know about, clearly written and utilizing countless obscure sources. Especially good on writing systems and history.

Deacon, Terrence W. *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: W.W. Norton & Co., 1997. The most detailed account of the neurological perspective on the origins of language, representing a common view among such specialists that language "rides" on more general cognitive abilities. Many generative syntacticians would disagree, but Deacon's is an especially comprehensive argument from the other side.

Diamond, Jared. *Guns, Germs, and Steel*. New York: W.W. Norton & Co., 1997. Diamond's now classic account of why some societies have acquired more power than others incorporates ample information about how languages have spread across the globe, admirably accurate as well as readable.

Dixon, R. M. W. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press, 1997. A muscular little monograph arguing that languages typically stew amongst one another in one place, becoming increasingly similar, and that only post-Neolithic migrations have led some languages to travel and give birth to brand-new offshoots taking root in new lands. The dedicated layman will glean much from the argument, which parallels Stephen Jay Gould's on punctuated equilibrium.

Dyer, Donald L. *The Romanian Dialect of Moldova*. Lewiston, NY: Mellen Press, 1999. A readable account of a "language" that is really just a minor dialectal variant of Romanian and how the confusion arose.

Ferguson, Charles A. *Language Structure and Language Use* (essays selected and introduced by Anwar S. Dil). Palo Alto, CA: Stanford University Press, 1971. Ferguson wrote his seminal work when linguists still wrote important work in a style accessible to interested readers; this essay of 1959 remains the classic introduction to the subject.

Finegan, Edward. *Language: Its Structure and Use*. Fort Worth, TX: Harcourt Brace, 1989. A textbook that combines layman-friendliness with detailed surveys of certain issues, such as the Polynesian languages and their history. I have used this one for years to usher undergraduates into the linguistic frame of mind.

Flodin, Mickey. *Signing Illustrated: The Complete Learning Guide*. New York: Perigee, 1994. This is an especially esteemed introduction to sign language for those stimulated by the subject.

Geertz, Clifford. "Linguistic Etiquette," in *Sociolinguistics*, edited by John Pride and Janet Holmes, pp. 167–179. Harmondsworth, England: Penguin, 1972. A

classic and readable article on layers of language in Java—and, by analogy, the fashion in which a language varies according to social factors, divested of the loaded sociological implications that, inevitably, coverage of this subject referring to dialects closer to home tends to entail.

Goody, Jack, and Ian Watt. "The Consequences of Literacy," in *Literacy in Traditional Societies*, edited by Jack Goody, pp. 27–84. Cambridge: Cambridge University Press, 1968. This is a truly magic piece that shows how the sheer fact of language written on the page transforms consciousness and history. It's long but thoroughly readable and worth the commitment.

Grillo, Ralph. *Dominant Languages: Language and Hierarchy in Britain and France*. Cambridge: Cambridge University Press, 1989. A solid coverage of how standard English and standard French became what they are, rather than the marginal dialects that they were at their inception. For those interested in a closer look at a process usually described in passing, this is a good place to look, although available only in university libraries.

Halliday, M. A. K. "Spoken and Written Modes of Meaning," in *Comprehending Oral and Written Language*, edited by Rosalind Horowitz and S. Jay Samuels, pp. 55–82. New York: Academic Press, 1987. A useful comparison of the spoken and the written, which like the Chafe and Danieliwicz article, highlights a difference that is easy to miss.

Hockett, Charles F. "The Origin of Speech." *Scientific American* 203 (September 1960). This article is still useful in getting down to cases as to what distinguishes human speech from the fascinating but "not quite it" renditions of language seen in parrots, apes, and even dogs. Few have done it better since.

Hopper, Paul J., and Elizabeth Closs Traugott, eds. *Grammaticalization*. Cambridge: Cambridge University Press, 1993. Grammaticalization has been commonly discussed among linguists for only about 20 years, and this is the leading textbook on the subject. It is rather compact and written in terms that will not overly tax the interested layman.

Kaye, Alan. "Arabic," in *The World's Major Languages*, edited by Bernard Comrie, pp. 664–685. Oxford: Oxford University Press, 1990. Kaye writes in a distinctly "human" way in a book intended as drier than what he submitted. This is a nicely readable introduction to Arabic and its structure.

"Languages of the World." *Encyclopedia Brittanica*. 1998 edition. This chapter, nowadays festooned with gorgeous, crystal-clear color maps, has been one of my own staples since I was 13. It covers the language families of the world in admirable and authoritative detail.

Lucy, John A. *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press, 1992. For those with a serious interest in the "Does language channel thought?" hypothesis that so often intrigues laymen, this monograph summarizes and critiques all of the relevant sources and experiments on the Sapir-Whorf

hypothesis up to its publication. (There have been a few studies slightly more promising for the hypothesis since.)

Matisoff, James. "On Megalocomparison." *Language* 66 (1990): 106–120. A cool-headed objection to Proto-World and related theories by a linguist who pulls off the feat of writing academically respectable linguistics papers in prose reasonably accessible to the layman, including a puckish sense of humor.

Matthews, Stephen. *Cantonese: A Comprehensive Grammar*. London: Routledge, 1994. Reference grammars can be forbidding to those outside academia, but this one is relatively accessible, as well as admirably detailed.

McWhorter, John. *The Power of Babel*. New York: HarperCollins, 2001. The basic thesis of this course, that human language is a natural history story, just as the evolution of animals and plants is, is encapsulated in this book. Solely as a result of lack of competition, the book is unique in giving a tour of human language from a modern perspective, including recent developments in the study of language change and how languages color one another.

———. *Word on the Street: Debunking the Myth of a "Pure" Standard English.* New York: Plenum Publishing, 1998. In this book, I attempt an argument that there is no such thing as "bad grammar," using Black English as a springboard but also addressing bugbears of the "Billy and me went to the store" type. There is also a chapter on how Shakespearean language is less accessible to us than we often suppose, useful in illuminating the subtleties of how languages change.

Nettle, Daniel, and Suzanne Romaine. *Vanishing Voices: The Extinction of the World's Languages*. New York: Oxford, 2000. As informed as David Crystal's *Language Death* but also founded on a sober (if, in my view, sadly unlikely) political argument for those interested in this view on the subject.

Norman, Jerry. *Chinese*. Cambridge: Cambridge University Press, 1988. A compact survey of Chinese in its "dialectal" variety, easy to read, trimming most of the fat (although one might skip the details periodically), and in print.

Ong, Walter. *Orality and Literacy: The Technologizing of the Word*. London: Routledge, 1982. A readable and invaluable classic exploration of the impact on the human experience created by something as seemingly mundane as the encoding of speech in written form; truly eye-opening and one of my favorite books.

Oppenheimer, Stephen. *The Real Eve: Modern Man's Journey out of Africa*. New York: Carroll & Graf, 2003. A survey of recent genetic evidence tracing human migrations, including evidence of higher-level mental activity further back in time than traditionally supposed by those pursuing a "Big Bang" 30,000 years ago. This is an updated report on the topic of Cavalli-Sforza's classic book: a bravura detective story, only occasionally tiring the non-specialist a bit.

Pei, Mario. *The Story of Language*. Philadelphia: J.B. Lippincott, 1949. Now available only on the library shelf but worth a read; this grand old "The World's Languages" trip inspired many a linguist (including myself). Put on your

historical-perspective glasses and savor an old-fashioned scholar's best of his many books for the public.

Pepperberg, Irene Maxine. *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*. Cambridge, MA: Harvard University Press, 2002. Battling the skeptics, Pepperberg tells us about her uncannily articulate parrots. Push aside the arcane and the dry and marvel at how human a pop-eyed bird can seem.

Pinker, Steven. *The Language Instinct*. New York: HarperPerennial, 1994. This is the classic introduction to what many linguists do in the modern world, examining whether there is an innately specified ability to use language in our brains. Pinker writes with hipness and wit.

Ramat, Anna Giacalone, and Paolo Ramat, eds. *The Indo-European Languages*. London: Routledge, 1998. This book includes survey chapters for each family, written by experts; it assumes some familiarity with linguistic terminology but will be of use to interested laymen who desire more detail than Dalby, Crystal (1987) or Comrie, Matthews, and Polinsky on this list give in their surveys.

Richardson, David. *Esperanto: Learning and Using the International Language*. El Cerrito, CA: Esperanto League for North America, 1988. This is the best source for learning, or learning about, this fascinating and beautiful experiment.

Rickford, John Russell, and Russell John Rickford. *Spoken Soul: The Story of Black English*. New York: Wiley and Sons, 2000. Most literature on Black English is written from a political and cultural point of view, specifically from the left. This book is no exception, but for those interested in exploring these aspects of the dialect, which will be natural given its charged nature in our times, this book is the most up-to-date and solid and includes some coverage of grammar and history, as well.

Roberts, Peter. *West Indians and Their Language*. Cambridge: Cambridge University Press, 1988. A readable survey of Caribbean creoles, which a great deal of the creolist literature focuses on, despite my aim to give a more global picture in this lecture series. This book also covers the sociological issues that, despite their interest, are not especially germane to the thrust of our story here.

Ruhlen, Merritt. *The Origin of Language: Tracing the Evolution of the Mother Tongue*. New York: John Wiley & Sons, 1994. Merritt Ruhlen and the Proto-World camp's articulate call to arms for the general public. One cannot come away from this book without suspecting that these people are at least on to something.

———. "Taxonomic Controversies in the Twentieth Century," in *New Essays on the Origin of Language*, edited by Jürgen Trabant and Sean Ward, pp. 97–214. Berlin: Mouton de Gruyter, 2001. For those who would like to dig in somewhat more specifically to the Proto-World perspective without being inundated with long lists of words and comparisons only a historical linguist could love, this is the handiest presentation I am aware of.

Sampson, Geoffrey. *Educating Eve: The "Language Instinct" Debate*. London: Cassell, 1997. A gifted rhetorician tears away at the Chomskyan perspective, unique among those making such attempts in having thoroughly engaged the often forbidding literature in question. A valuable counterpoint to Pinker's *The Language Instinct*.

Sebba, Mark. *Contact Languages: Pidgins and Creoles*. New York: St. Martin's Press, 1997. Of the various textbooks on pidgins and creoles, this is the clearest, most up-to-date, and most worldwide in its orientation. Run, don't walk—this one made me decide not to write one of my own.

Simonson, Douglas (Peppo). *Pidgin to da Max*. Honolulu: The Bess Press, 1981. A jocular illustrated glossary of the creole English of Hawaii, focusing on "colorful" vocabulary but giving a good sense of a creole as a living variety.

Stavans, Ilan. *Spanglish: The Making of a New American Language*. New York: HarperCollins, 2003. "Spanglish" has inspired a fair degree of semi-informed musings, but here is finally a more considered and informed piece, also situating the variety sociopolitically.

Thomason, Sarah Grey. *Language Contact: An Introduction*. Washington, DC: Georgetown University Press, 2001. A recent textbook on language mixture—a topic unknown to the textbook until recently—by a linguist with a gift for clarity, as well as relentless good sense. One of my favorite thinkers who has endlessly inspired me—highly recommended.

Versteegh, Kees. *The Arabic Language*. New York: Columbia University Press, 1997. This book includes anything anyone, other than a specialist, would want to know about the awesome cathedral that is Arabic, in accessible language. Details can be bypassed, but this will serve as one's dependable Bible (or Koran) on the subject.

Wallman, Joel. *Aping Language*. Cambridge: Cambridge University Press, 1992. This selection usefully compiles, between two covers, the issues regarding how closely apes approximate human speech. Not too closely, Wallman argues, but the book offers all one needs to know about the field of inquiry as a whole.

Watkins, Calvert, ed. *The American Heritage Dictionary of Indo-European Roots*. Boston: Houghton Mifflin, 1985. This will serve those who want a brass-tacks look at how Indo-Europeanists go about their business. It is a book version of an appendix included in the *American Heritage Dictionary*, aimed at a general readership.

Wells, Spencer. *The Journey of Man: A Genetic Odyssey*. Princeton: Princeton University Press, 2002. An alternative rendition of a story updating Cavalli-Sforza, told more comprehensively by the Oppenheimer book on this list; somewhat lesser on renegade insight and narrative suspense but more compact for those with less time.

Whorf, Benjamin Lee. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, edited by J. B. Carroll. Cambridge, MA: MIT Press, 1956.

The take-home version of Whorf's ideas on how language channels thought. Now available only at university libraries, but a useful way to get the insights at their source without trawling the obscure and scattered venues in which the work originally appeared.

Wright, Robert. "Quest for the Mother Tongue." *Atlantic Monthly* 267 (1991): 39–68. A general-public account of the Proto-World thesis and its notably acrid reception by most other linguists; this is a nice introduction to whet the appetite for Ruhlen's book.

Internet Resources:

http://www2.arts.gla.ac.uk/IPA/index.html. On the Web site of the International Phonetic Association, you will find charts of the International Phonetic Alphabet, many of whose symbols were used throughout this booklet.

http://www.languagehat.com. A feast for language lovers, consisting of essays, comments, and links to dozens of language-related Web sites, including linguablogs, language resources, and more.

http://www.languagelog.org. A composite of language-related essays; some funny, some serious, all thought-provoking.

