Ferdinand de Saussure

Swiss linguist (1857-1913), founder of structuralism, which sees language as a system of interrelated elements that derive their meaning from the relations that hold between them.

- Introduced three important linguistic distinctions:
  1. synchronic vs. diachronic linguistics
  2. langue (language system) vs. parole (individual speech)
  3. syntagmatic vs. paradigmatic relations between units

Two types of lexical relations

1. **Syntagmatic** relations concern position, sequence, and possibilities of combination of words: X and Y and Z

2. **Paradigmatic** relations concern contrast, differentiation, substitution, and selection among senses: X or Y or Z

Examples include synonymy (same meaning), antonymy (opposite meaning), and hyponymy (X is a kind of Y)

John Sinclair: Idioms vs. open choice

- Sinclair (1996) distinguishes two conflicting principles of language: the *idiom principle* and the *open choice principle*

- The open choice principle holds that language is based on grammatical rules, and words are selected slot by slot
  
  This view is associated with generative grammar (e.g., ‘lexical insertion’)

- The idiom principle holds that speakers make use of lexicalized and semi-lexicalized phrases, which are stored whole in long-term memory and retrieved as single items
  
  This view is associated with usage-based approaches like Construction Grammar
Noam Chomsky

- Born 1928 in Philadelphia; affiliated with MIT since 1955
- Chomsky’s mentalistic, generative approach to language revolutionized linguistics and cognitive science in the 20th century

The ‘Cognitive Revolution’ of the 1960s

- Empirical and statistical methods were popular the 1950s, dominating fields from psychology (behaviorism) to electrical engineering (information theory)
- They faded in the 1960s under the ‘cognitive revolution’
- Seminal events:
  - Chomsky’s *Syntactic Structures* (1957) and *Aspects of the Theory of Syntax* (1965)
  - Chomsky’s review of *Verbal Behavior*
  - Chomsky & Miller’s critiques of statistical language models
  - Minsky and Papert’s criticism of neural networks

Syntactic Structures (Chomsky 1957)

From now on I will consider a *language* to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements.

The fundamental aim in the linguistic analysis of a language $L$ is to separate the *grammatical* sequences which are the sentences of $L$ from the *ungrammatical* sequences which are not sentences of $L$ and to study the structure of the grammatical sequences.

One what basis do we actually go about separating grammatical sequences from ungrammatical sequences?

I would like to point out that several answers that immediately suggest themselves could not be correct.

Syntactic Structures (Chomsky 1957)

...First, it is obvious that the set of grammatical sentences cannot be identified with any...finite and somewhat accidental corpus of observed utterances...

Second, the notion “grammatical” cannot be identified with “meaningful” or “significant” in any semantic sense. Sentences (1) and (2) are equally nonsensical, but any speaker of English will recognize that only the former is grammatical.

(1) Colorless green ideas sleep furiously.
(2) Furiously sleep ideas green colorless.
Third, the notion “grammatical in English” cannot be identified in any way with the notion “high order of statistical approximation to English.” It is fair to assume that neither sentence (1) nor (2)...has ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally ‘remote’ from English. Yet (1), though nonsensical, is grammatical, while (2) is not.

Evidently, one’s ability to produce and recognize grammatical utterances is not based on notions of statistical approximation and the like... I think that we are forced to conclude that grammar is autonomous and independent of meaning, and that probabilistic models give no particular insight into some of the basic problems of syntactic structure.

### Chomsky vs. probabilities

- Chomsky’s arguments against probabilistic models of language were very influential
- From the 1960s through the end of the 1980s, Chomsky’s generative grammar framework was dominant
- Theoretical linguists eschewed corpus data, which was hard to obtain and use anyway
- Today statistical methods are popular again, and Chomsky’s arguments against probabilities are no longer accepted

### Refuting Chomsky’s arguments

- Chomsky (1957:15-16) writes:
  1. Colorless green ideas sleep furiously.
  2. Furiously sleep ideas green colorless.

  ... in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally ‘remote’ from English.

- This claim is false—all modern statistical models of language can assign probabilities to previously-unseen utterances

  That’s how speech recognition works!

- E.g., Pereira’s (2000) statistical model of newspaper text assigns (1) a probability 200,000 times greater than (2)

- Even a simple model based on word classes can do this

### Brilliant white flares swayed eerily

# Search the Brown corpus for patterns of the form
# "\w+/jj \w+/jj \w+/nn\w* \w+/vb\w* \w+/rb"
# as in "Colorless green ideas sleep furiously"

excellent/jj foreign/jj bomb/nn takes/vbz only/rb
brilliant/jj white/jj flares/nns swayed/vbd eerily/rb
prevalent/jj mental/jj disturbance/nn affecting/vbg even/rb
little/jj green/jj biplane/nn struggled/vbd northward/rb
fake/jj therapeutic/jj devices/nns figure/vb prominently/rb
routine/jj vital/jj statistics/nns got/vbd nowhere/rb
appropriate/jj regional/jj office/nn listed/vbn below/rb
autonomic/jj hypothalamic/jj balance/nn occurring/vbg spontaneously/rb
long/jj fluorescent/jj tube/nn suspended/vbn directly/rb
little/jj sweet/jj potato/nn trilled/vbd neatly/rb
looking/jj young/jj officer/nn fell/vbd back/rb

# Search the Brown corpus for patterns of the form
# "\w+/rb \w+/vb\w* \w+/nn\w* \w+/jj \w+/jj"
# as in "Furiously sleep ideas green colorless"
Chomsky vs. corpus linguistics

- Chomsky emphasizes the creativity of language
- This creativity is manifested in recursive generative rules like $S \Rightarrow NP \ VP$, $VP \Rightarrow V \ NP$ (or more recently, “Merge”)
- Empiricists, on the other hand, tend to focus on common patterns (e.g., collocations), and to emphasize the redundancy and predictability of language
- Warren Weaver (1949):
  
  [A]bout half of the letters or words we choose in writing or speaking are under our free choice, and about half (although we are not ordinarily aware of it) are really controlled by the statistical structure of the language.

Autonomy of syntax

- Chomsky (1957) writes:
  
  I think that we are forced to conclude that grammar is autonomous and independent of meaning
  
  [I]t is folly to decouple lexis and syntax, or either of those and semantics. The realization of meaning is far more explicit than is suggested by abstract grammars. The model of a highly generalized formal syntax, with slots into which fall neat lists of words, is suitable only in rare uses and specialized texts. By far the majority of text is made of the occurrence of common words in common patterns. Most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up text.

Chomsky on corpus linguistics: 2004 interview

Corpus linguistics doesn’t mean anything... [S]uppose physics and chemistry decide that instead of relying on experiments, what they’re going to do is take videotapes of things happening in the world and they’ll collect huge videotapes of everything that’s happening and from that maybe they’ll come up with some generalizations or insights. Well, you know, sciences don’t do this. But maybe they’re wrong. Maybe the sciences should just collect lots and lots of data and try to develop the results from them. Well if someone wants to try that, fine. They’re not going to get much support in the chemistry or physics or biology department... We’ll judge it by the results that come out. So if results come from study of massive data, rather like videotaping what’s happening outside the window, fine—look at the results. I don’t pay much attention to it. I don’t see much in the way of results.

Chomsky on corpus linguistics: 2004 interview

My judgment, if you like, is that we learn more about language by following the standard method of the sciences. The standard method of the sciences is not to accumulate huge masses of unanalyzed data and to try to draw some generalization from them. The modern sciences, at least since Galileo, have been strikingly different. What they have sought to do was to construct refined experiments which ask, which try to answer specific questions that arise within a theoretical context as an approach to understanding the world... If you want to understand how bodies fall, Galileo would not have been interested in videotapes of leaves falling and balls going around and rocks rolling down mountains and so on and so forth. What he was interested in is the highly refined abstract conception of a ball rolling down a frictionless plane, which doesn’t even exist in nature... To say that it’s more empirical to just collect and observe data is completely wrong.
Chomsky on corpus linguistics: 2004 interview

People who work seriously in this particular area do not rely on corpus linguistics. They may begin by looking at facts about frequency and shifts in frequency and so on, but if they want to move on to some understanding of what’s happening they will very quickly, and in fact do, shift to the experimental framework. Where you design situations, you enquire into how people will act in those situations. You design them within a framework of theoretical inquiry which has already suggested that these are likely to be important questions and I want the answers to them. But that’s not corpus linguistics. If you want to use hints from data that you acquire by looking at large corpuses, fine. That’s useful information for you, fine. I mean, Galileo might have gotten some hints from looking at events that were happening in the world. In fact, he did. He observed the tides—that’s like corpus linguistics. You’re observing the tides.

Constructing refined experiments

- For Chomsky, “to construct refined experiments” means to ask a native speaker to judge the acceptability of artificially constructed example sentences
- Corpus linguists counter that this approach can be problematic
- Examples of acceptability judgments from van Riemsdijk & Williams (1986), Introduction to the Theory of Grammar
  1. Who did Jo think said John saw him?
  2. John I believe Sally said Bill believed Sue saw.
  3. John wants very much for himself to win.
  4. *What did Sally whisper that she had secretly read?
  5. *The boys read Mary’s stories about each other.

Modern generative grammar