

Lecture 9: February 8, 2007

*Lecturer: Partha Niyogi**Scribe: Sravana Reddy*

9.1 The Phenomenon of Language

The study of language includes *Phonetics*, the study of sounds, *Phonology*, the study of sound-patterns in languages, *Morphology*, the study of the structure of words, *Syntax*, the science of word patterns in sentences, *Semantics*, the science of word and phrase meaning, and *Pragmatics*, the study of language in speech acts.

The main treatment of language in the course will be in language learning.

9.2 Overview of Language Learning

What are the distinct features of a language and how do we learn them?

For instance, consider learning the morphophonological rules of English. Given these nonsensical words: timp, wug, heez – what is the most natural way to pluralize them?

- timp → timps
- wug → wugz
- heez → heezes

We seem to be able to guess the pronunciation of plural forms of nonsense words (words we have certainly not learnt about). That is, we have internalized a rule for pluralization by generalizing from a finite set of examples.

What are the natures of these rules?

For pluralization, we may look at the last phoneme¹ of the singular form, and the phonemes of the plural suffix, and generalize accordingly.

Let's say the learner is exposed to the words below and their plurals. It soon begins to group them into equivalence classes as shown.

Equivalence class of words (singular)

place, adz, porch, cabbage, ambush
lip, telegraph, bit
clubs, nerds, colleagues

Corresponding plurals

placiz, adziz, porchiz, cabbagiz, ambushiz
lips, telegraphs, bits
clubz, nerdz, colleaguez

¹A *phoneme* can be thought of as the smallest category representing an equivalence class of acoustic signals.

From the above classes, the learner can generalize to form these rules:

1. If the last phoneme in singular form is s, z, g, ch, sh: pluralize by adding **iz**
2. If the last phoneme is p, f, t: pluralize by adding **s**
3. else pluralize by added **z**.

9.3 More on Phonemes

We can thus conclude the following two facts: 1. Phonemes group themselves into equivalence class w.r.t to certain rules, and 2. There exists a rule for pluralization that has a concise description.

Or formally: Given a set of p of phonemes, a subset s is a natural class of p if it forms an equivalence class w.r.t to a rule in some language.

Examples of other such rules: tense changes, etc.

What are some natural classes of phonemes?

Vowels and consonants form a natural class with respect to syllabification (the number of syllables = the number of vowels).

Can any subset of the n phonemes be a natural class? That is, are there 2^n natural classes?

Theoretically, yes. However, only approximately $n^{\log_2 3}$ are observed.

One explanation for this can be got from considering Jakobson's, Halle's, Troubetzkoy's or Ladefoged's formulations of distinctive features (vocalic, sonorant, constituent, etc). Broadly, all their theories indicate that we can define k features f_1, f_2, \dots, f_k , where each f_i is a binary-valued (or sometimes an n -valued) vector, such that every phoneme is then a unique point in the k -dimensional space defined by these vectors.

Pick $s \subset \{1, 2, \dots, k\}$, thus defining 2^s natural classes.

9.4 Syntax

We now look at the behavior of languages where words (rather than phonemes) are the basic units.

What is a word?

In a stream of speech, there are no distinctive breaks between words. Then, how do learners of the language discover the constituent words in a sentence? (It appears more natural that they would learn one of the extremes, i.e. that the entire sentence is a words, or that the phonemes or syllables (natural units in speech) are words.

Thus, to identify words in sentences, speakers must rely on memory to find the smallest unit of each sentence that carries meaning.

Syntax is the study of well-formedness of sentences. Note the distinction between syntactic and semantic accuracy: a sentence can be meaningless but still have a valid parse in a given language. (Note, for example, how $1+1=2$ and $1+1=3$ are both syntactically correct arithmetic expressions, although the latter is semantically incorrect. In contrast, something like $1+1+1=$ is syntactically ill-formed. Equivalently, one can have sentences like *Colorless green ideas sleep furiously* as permissible by the grammar of English.)

9.5 Learning a Language

A language L is a set of strings. With that in mind, note:

- $L_{English}$ (the set of all English sentences) $\in \sigma^*$ (the set of all possible sentences)
- Knowing $L_{English} \equiv$ knowing English
- $L_{English}$ has infinite size.
- \exists a grammar G that uniquely and completely describes $L_{English}$

What is the nature of G ? How can it be learnt? What is the place of G in the Chomsky hierarchy?

For starters, $G =$ the class of all possible g s, where g is a grammar of English. If G is learnable, $\forall g \in G, g$ is learnable.

More on learnability in the next class.

9.6 Related Reading

The Sound Pattern of English by Chomsky and Halle.

Fundamentals of Language by Jakobson and Halle.