1. **Ambiguity in Language**
   - Structural Ambiguity
   - Types of Ambiguity

2. **The Lexicon**
   - Closed vs. Open Classes
   - Parts of Speech
   - Lexical Ambiguity

3. **Semantic Ambiguity: Scope and Reference**

   **Readings:**
   - J&M (1st edition) ch. 8 (pp. 287–298), ch. 10 (pp. 372–376)
   - J&M (2nd edition) ch. 5 (pp. 1–11), ch. 13 (pp. 7–8); NLTK Tutorial: Elementary Language Processing.

   **Reminder:** NLTK labs start next week.

---

### Structural Ambiguity

- In both formal and natural languages, **meaning** is derived (in part) from the **structure** underlying its strings (Weeks 8 & 9).
- **Parsers** are procedures for recovering this structure, as a basis for computing meaning and/or mapping to another form (Weeks 5–7).
- Recall from Week 2 that for CFGs, the structure of a string can be represented as a **tree diagram**.
- The tree diagram represents the set of **derivations** used in producing a string from **S** (the start symbol of the grammar).

---

### Example: Newspaper Headlines

- **stolen painting found by tree**
- **lung cancer in women mushrooms**
- **dealers will hear car talk at noon**
- **miners refuse to work after death**
- **juvenile court to try shooting defendant**
Structural Ambiguity

The designers of formal languages (e.g., XML) or programming languages try to eliminate or reduce structural ambiguity.

**Example**

Python’s use of indentation to indicate embedding and of no indentation to indicate sequence.

Spoken Natural Language has means of disambiguating structural ambiguity that are absent in written language.

**Example**

lung cancer in WOMEN | mushrooms
dealers will hear CAR_TALK at noon

Intonation is one of the reasons we don’t normally see how ambiguous natural language is.

Types of Ambiguity

- Given a string from a language, we want a parser to deliver either its most likely structure or all possible structures (for another procedure to examine further).
- If a string doesn’t belong to the language, what the parser should do depends on the application.
- There are several types of ambiguity we will discuss:
  - global structural ambiguity (see examples above)
  - local structural ambiguity (important for human parsing: Lecture 30)
  - lexical ambiguity (examples coming up)
  - scope ambiguity (examples coming up)
  - referential ambiguity (examples coming up)

Some languages produced by CFGs are inherently ambiguous.

**Example**

Consider the language produced by $G_1 \cup G_2$

$$a^n b^n c^m \cup \alpha b^k c^k \quad m, n, j, k \geq 1$$

How many distinct non-equivalent derivations (tree structures) are there for the simple string $abc$?

How many for the string $aabcc$?

How about any string of the form $a^n b^n c^n$?
Closed vs. Open Classes

Σ: set of terminal symbols
Even in formal languages, we can distinguish two subsets within Σ:
- **Closed class** symbols associated with particular productions and their meaning.
  - **FOL**: \( S \rightarrow (\forall|\exists) \text{ Variable Formula} \)
  - **Python**: \( S \rightarrow \text{for Var in ListOrDictionary : } S^+ \)
  - \( S \rightarrow \text{from Module import Namelist} \)
- **Open class** symbols, which are fully productive: users continually introduce new instances.

Any language has relatively few closed class tokens, and users rarely, if ever, introduce new ones.

Informatics 2A: Lecture 11 Ambiguity and the Lexicon in Natural Language 9

Lexicon in Natural Languages

NL lexicons contain tens of thousands of words, with new ones constantly being created.

Grammars or models for NL can be largely specified in terms of the classes that words belong to, rather than words themselves.

Word classes found in all Indo-European languages and in many other language families:
- nouns
- verbs
- adjectives
- adverbs

All are open classes. New open class words enter English all the time, e.g., **blogger** (N), **ping** (V), **google for** (V), **entrepreneurial** (Adj).

Informatics 2A: Lecture 11 Ambiguity and the Lexicon in Natural Language 10

Parts of Speech

How do we tell what part of speech (word class) a word belongs to?
To make this decision, linguists have developed a set of criteria:
- **Notional** (semantic) criteria: What does it refer to?
- **Distributional** (syntactic) criteria: Where is it found?
- **Formal** (morphological) criteria: What does it look like?

We will look at each of the parts of speech (POS) in turn.

Nouns

**Notional** criterion: nouns generally refer to living things (**mouse**), places (**Scotland**), things (**projector**), or concepts (**intelligence**).

**Distributional** criterion: nouns can appear after determiners like the or before relative pronouns like that.

**Example**: the blob/mouse/university that ate Chicago

**Formal** criterion: words that end in -ness, -tion, -ity, and -ance tend to be nouns.

**Example**: happiness, exertion, levity, significance

Formal and distributional criteria help people (and machines) recognize the class of unknown words.

**Example**: Within the conurbation, open countryside is limited to a network of corridors.
Verbs

Notional criterion: verbs refer to actions (sleep, wash, give).

Distributional criterion: verbs can be classified by the number of arguments they co-occur with:
- intransitive verbs (1 arg): Smoke rises.
- transitive verbs (2 args): John washed the glass, The cat groomed itself.
- ditransitive verbs (3 args): John served us steak, Mary gave Fred a toothpick.
- verbs with 4 args: Fred transferred the glass from the table to the shelf.

Formal criterion: words that end in -ate or -ize tend to be verbs, and ones that end in -ing are often the present participial of a verb.

Example: automate, calibrate, equalize, modernize; rising, washing, grooming.

Adjectives

Notional criterion: adjectives describe things that are nouns (small, wee, salubrious, excellent).

Distributional criterion: adjectives usually appear before a noun or after a form of be.

Example: wee drop; The food is excellent.

Formal criterion: words that end in -al, -ble, and -ous tend to be adjectives.

Example: formal, invisible, capable, salubrious, parlous

Adverbs

Notional criterion: adverbs describe actions or events (quickly, often, possibly) or adjectives (really).

Distributional criterion: adverbs can appear next to a verb, or an adjective, or at the start of a sentence.

Example: run quickly; often walk; really nice; Actually, she’s invisible.

Formal criterion: words that end in -ly are adverbs.

Closed Class Words

The set of closed classes can vary from language to language. Closed classes in English include:
- determiners: the, any, a, ...
- prepositions: in, of, with, without, ...
- conjunctions: and, because, after, ...
- auxiliaries: have, do, be
- modals: will, may, can, need, ought
- pronouns: I, she, they, which, where, myself, themselves

English doesn’t have clitics (like French le) or particles (like Japanese ga). Russian lacks reflexive pronouns.

The functions performed by closed class words in one language may be performed by the morphology of open class words in another language (e.g., reflexivity in Russian).
Lexical Ambiguity

Two important types of lexical ambiguity:

**Sense Ambiguity**: e.g., *intelligence*:

1. Power of understanding
2. Obtaining or dispersing secret information; also the persons engaged in obtaining or dispersing secret information

**Part of Speech Ambiguity**: e.g., *still*:

1. *Adverb*: at present, as yet
2. *Noun*: (1) silence; (2) individual frame from a film; (3) vessel for distilling alcohol
3. *Adjective*: motionless, quiet
4. *Transitive Verb*: to calm

Next week (Lecture 13), we deal with POS ambiguity in detail.

Scope Ambiguity

In both scope ambiguity and referential ambiguity, the same structures and word senses can still have multiple interpretations. *Not* is a word that operates on the meanings of other words and structures, and hence may produce scope ambiguity:

**Example**

David did **not** take the bus to the zoo. Instead, he drove there. ⇒ *He didn’t take the bus to the zoo.*

he took it to Murrayfield. ⇒ *He didn’t take the bus to the zoo.*

he stayed home. ⇒ *He didn’t take the bus to the zoo.*

Referential Ambiguity

A referring term can be ambiguous as to which of several entities it refers to, producing different interpretations of the same sentence.

**Example**

David decided to hide the keys to Fred’s car because **he** didn’t want him to drive. **he** = ??

had had too much to drink. **he** = ??

was joking around. **he** = ??

Summary

- **Structural ambiguity**: a string can be associated with more than one structure (represented as trees).
- **Lexical ambiguity**: a word can associated with more than one part of speech or with more than one sense.
- **Open classes** are productive, present in many languages, and are often class-preserving under translation (noun, verb, etc.).
- **Closed classes** are not productive and vary across languages.
- To identify the POS of a word, we can use **notional**, **distributional**, or **formal** criteria.
- **Scope ambiguity** and **referential ambiguity**: types of semantic ambiguity, where the same structures and word senses can still have multiple interpretations.