University of Malta
BSc IT (Hons.) Year III

## CSA305: Natural Language Processing

### Lecture 4

### Linguistic Knowledge

- Scope
- Kinds of Linguistic Knowledge
- Phrases
- Phrase Structure

Department of Computer Science and AI
2000/01

## References/Acknowledgements

See:

The Internet Grammar of English

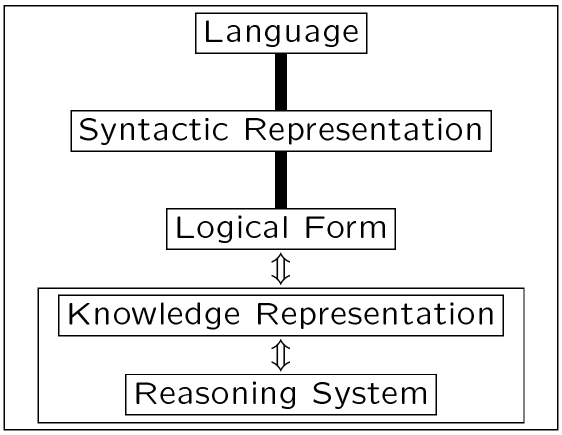
http://www.ucl.ac.uk/internet-grammar

and

James Allen, Natural Language Understanding, Menlo Park: Benjamin Cummings, 1987 (chapter 2)



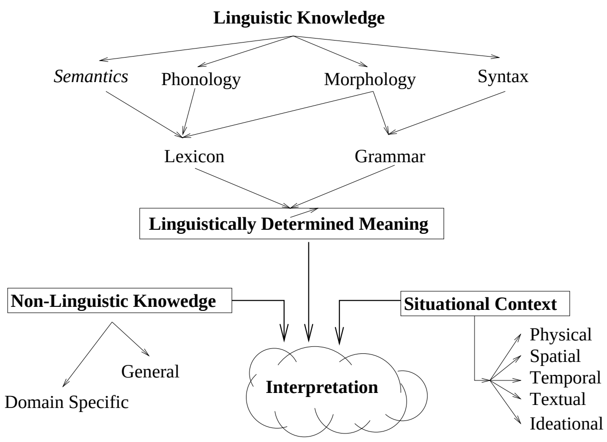
## Scope of Linguistic Knowledge




- Linguistic knowledge determines "translation" to logical form, usually via some kind of syntactic representation.
- Interpretation of logical form is more like a "transaction" with a KR system.



## Language and Knowledge



## Linguists View of Linguistic Knowledge

| LANGUAGE | |
|---|---|
| Sound | 1. Phonology |
| GRAMMAR | 2. Morphology |
| | 3. Lexicon |
| | 4. Syntax |
| Meaning | 5. Semantics |



## Phonetics and Phonology

- **Phonetics** deals with the *physical* building blocks of a language's sound system,

  The word *tieghi* has three phonetic realisations

  tee-eye
  tee-ay
  tee-ee

- **Phonology** deals with *idealised* building blocks called phonemes.

  tieghi = t + ie + ghi (3 phonemes)



## Morphology

**Morphology** studies the structure of words in terms of more primitive units called *morphemes*. Morphemes are the smallest units of language capable of bearing meaning.

- **inflectional morphology**: relates a word to forms of the same word that is appropriate to a particular context. Example - verb endings:

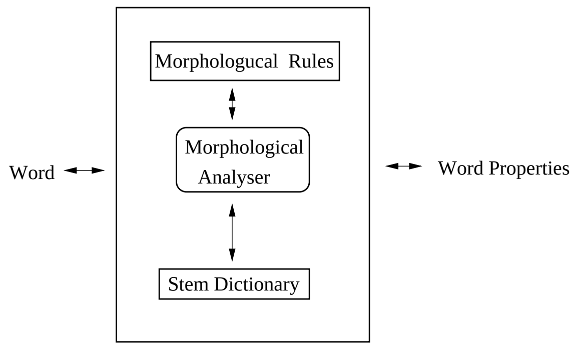
  "puts" = put + s

- **derivational morphology**: relates a word and outputs a different words that are derived from the first. Example: the prefix "un"; the suffix "ness"

  "unhappiness" = un + happy + ness



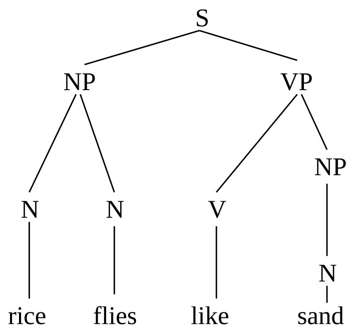
## The Lexicon




- The lexicon is a repository of all the exceptions in the language.
- Division of labour between grammar rules and lexicon.
- Lexicon relates *words* to different kinds of information, as shown by the outer square.
- Full form lexicon *versus* lexicon with morphological analysis.

## Syntax defines Syntactic Constituent Structure

Syntactic structure indicates how the words are related to each other:




- Syntax tree is assigned by a grammar and a lexicon
- Lexicon indicates syntactic category of words.
- Grammar (typically CFG) specifies legitimate concatenations of constitutents.



## Levels of Grammar Organisation

- **Word Classes.** Different parts of speech.
- **Phrases.** Sequences of words inheriting the characteristics of certain word classes.
- **Clauses.** Clauses contain at least a verb phrase:
  - Sentence (= main or matrix clause)
  - Subordinate Clauses (subordinate to the sentence or other higher clause)

    *John said that Bill said that Sandy was sick*
- **Grammatical Functions.** Role played by particular constituents of the sentence.
  - Subject
  - Predicate
  - Direct Object
  - Indirect Object (s)



## Word Classes

### Closed Classes

Fixed number of elements.

1. **Determiner:** a, an, my, some, the
2. **Preposition:** to, for, at, in, of, over, with
3. **Conjunction:** and, because, but, if, or
4. **Pronouns**: he, she, it, them, who, which

### Open Classes

Variable number of elements.

4. **Noun:** brother, car, David, house
5. **Verb:** be, drive, grow, sing, think
6. **Adjective:** big, foolish, happy, talented, tidy
7. **Adverb:** happily, recently, soon, then



## Nouns

- NUMBER: Singular, Dual, Plural

  eye - eyes
  ghajn - ghajnejn
- GENDER (NATURAL & GRAMMATICAL): Masculine, Feminine, Neuter (Animate, Vegetable) AND AGREEMENT
- CASE: Nominative (subject); Accusative (object)

  ich liebe dich
  mensa, mensa, mensam, mensae, mensae, mensa
- TYPES: Proper; Common; Mass; Count

  Paris - cow - water - loaf

## Determiners

Nouns are often preceded by the words the, a, or an. These words are called DETERMINERS. They indicate the kind of reference which the noun has.

- definite/indefinite: a, the
- proximity to speaker: this, that
- quantity: all, both, many, several, four

Determiners should not be confused with numerals or pronouns. Find the determiners below:

1. Each team consists of eleven players
2. My lucky number is five
3. Is this yours?
4. These questions are easy



## Adjectives

- Typically describe an attribute of a noun.
- Subtypes: ( e.g. colour, size, age)
- Usually comes before the noun (in English).
- Number/Gender/Case (not English).
- Comparison: young, younger, youngest
  recent, more recent, most recent
- Can take a modifying adverb, such as very, extremely, or less, before them.



## Pronouns

Basic feature: can stand for nouns

| Person | Singular | Plural |
|---|---|---|
| 1 | I | we |
| 2 | you | you |
| 3 | he/she/it | they |

| Person | Singular | Plural |
|---|---|---|
| 1 | me | us |
| 2 | you | you |
| 3 | him/her/it | them |

Personal pronouns are marked for Number, Gender, Case, Person. Other types of pronoun are

| Type | Example |
|---|---|
| Possessive | mine, yours |
| Reflexive | myself, yourself |
| Reciprocal | each other, one another |
| Relative | that, which, who, whom |
| Demonstrative | this, that these |
| Interrogative | who what |
| Indefinite | anybody, anyone |



## Verbs

- Finiteness: to love; he loves; he has loved
- Agreement: person, number, gender

  kont, kont kien, kienet, konna, kontu, kienu
- Tense: past, present, (future)
- Voice: active/passive

  John ate the cake
  The cake was eaten by John
- Mood: Indicative; Imperative; Subjunctive
- Aspect: Perfective, Imperfective

  Masha was writing letters
  Masha wrote the letters
  Masha wrote letters several times
- Transitivity

## Verbs

Forms

| Non-Finite | | | Finite | |
|---|---|---|---|---|
| Infinitive<br>to .. | -ing<br>he is .. | -ed<br>he has .. | Present<br>he .. | Past<br>he .. |
| cook<br>take<br>be | cooking<br>taking<br>being | cooked<br>taken<br>been | cooks<br>takes<br>is | cooked<br>took<br>was |

**Identify the forms:**

1. He plays the piano in a jazz club
2. David is singing in the shower again He was told not to laugh at policemen
3. His arm swelled up after the accident
4. The population has increased by 6since 1970



## Auxiliary Verbs

- the old lady *is writing* a play
- the film *was produced* in Hollywood

Main or Auxiliary?

1. I will *have* the soup
2. Police *are* investigating the incident
3. It *is* very peaceful here
4. Where *does* your brother work?
5. They *have* decided to advertise your job
6. He *does* his homework on the way to school



## Voice: Active/Passive

In a passive construction, a form of the passive auxiliary be is used, and the main verb has an -ed inflection.

Active or Passive?

1. Jane decided to cancel the part
2. The lecturer was impressed by your essay
3. Your horse is walking on our lawn
4. The top floor was destroyed
5. The pills should be taken twice a day



## Adverbs

- Basic function is to modify another consi-tituent which might be
  - Adjective

    *very large*.
  - Clause

    *slowly, he rose*.
  - Can come after the item modified

    *he rose slowly*.

## Phrases

- **Types**: Noun, Verb, Adjective, Adverb, Preposition.
- Consist of one or more words.

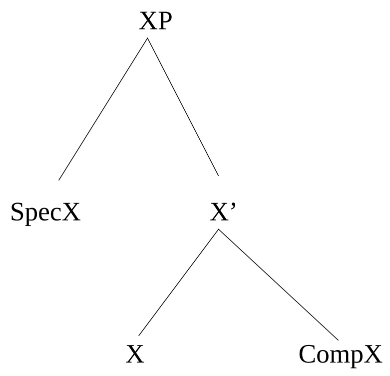
  *Paris*
  *The capital of France*

  *sleeps*
  *ate a juicy hamburger yesterday*
- Structured around
  - A **head** word, and
  - One or more **complements**, i.e. phrases bearing a fixed semantic relationship to the head word. Some complements are obligatory, particularly in case of certain verbs.
  - Any number of modifying **adjuncts**.



## Structure of Phrases

The basic syntactic categories are head, specifier, complemement and adjunct. They form a set of relationships which may be represented in a syntactic 'tree':




The sub-Xs on SpecX and CompX means that the categories of specifiers and complements will differ depending on the category of the XP they occur in. Only nouns allow determiners (a, the, that, etc.) as specifiers and adjectives do not allow noun complements