CS674 Natural Language Processing

- Topics for today
  - Need for morphological analysis
  - Basics of English morphology
  - Finite-state morphological parsing

Morphology

- Studies how words are constructed from sub-word units

  "A writer is someone who writes, and a stinger is something that stings. But fingers don't fing, grocers don't groce, haberdashers don't haberdash, hammers don't ham, and humdingers don't humding."   -Richard Lederer

- **Morphological parsing** is the process of finding the constituent morphemes in a word
  - » foxes → fox + es
  - » foxes → fox-N + es-PL
  - » killer → kill + er
  - » killer → kill-V + er-N
  - » going → go-V + ing-GER
  - » going → go-V + ing-PROGRESSIVE

Need for morphological analysis

- Information retrieval: **search**
  - Systems benefit from being able to search for singular and plural forms of search terms
  - Generally fairly easy in English
  - Complications
    - » Irregular plurals handled via morphological rules
      - ◆ goose → geese
      - ◆ fish → fish
      - ◆ ox → oxen
    - » Spelling rules needed
      - ◆ fox +PL → foxes
      - ◆ fly +PL → flies

Need for morphological analysis

- Information retrieval: **stemming**
  - Useful to map all of *walks, walking, walked* to *walk.*
  - Why?
  - Similar, but not identical to morphological parsing…how?

Need for morphological analysis

- Efficiency
  - Listing all of the plural forms of English nouns, all of the verb forms for a particular stem, etc…is a waste of space (and time if the entries are being made by hand).
    - » Suffixes are **productive**

  - Situation is much worse in other languages, e.g. **agglutinative** languages like Turkish

- Other uses for morphological parsing?

Basics of English morphology

- Morpheme – minimal meaning-bearing unit in a language
  - Stems – central meaning-bearing morpheme of the word
  - Affixes – supply "additional" meanings
    - » Prefixes – precede the stem
    - » Suffixes – follow the stem
    - » Circumfixes – precede and follow the stem
    - » Infixes – inserted inside the stem
  - Non-concatenative – morphemes are intermingled rather than concatenated
  - Root-and-pattern morphology – e.g. Hebrew

Basics of English morphology

- Inflection
  - Combination of a word stem with a grammatical morpheme, usually resulting a word of the same class, and usually filling some syntactic function.
  - Nouns
    - » Suffixes for *plural* and *possessive*
  - Verbs
    - » Suffixes for *–s* form, *-ing* participle, past form or *–ed* participle
      - ◆ watch, watch(e)s, watching, watched
  - Adjectives
    - » Suffixes for *comparatives*
      - ◆ cold, colder, coldest

Basics of English morphology

- Derivation
  - Combination of a word stem with a grammatical morpheme, usually resulting in a word of a *different* class, often with a meaning that's hard to predict exactly.
  - Nominalization
    - » organize (V) + -ation
    - » grant (V) + -ee
    - » kill (V) + -er
    - » silly (ADJ) + -ness
  - Creating Adjectives
  - Less productive than inflection
    - » Can't add –ation to every verb that ends in –ize, e.g. re-size

CS674 Natural Language Processing

- Topics for today
  - Need for morphological analysis
  - Basics of English morphology
  - Finite-state morphological parsing
    » Representing the lexicon and morphosyntactics

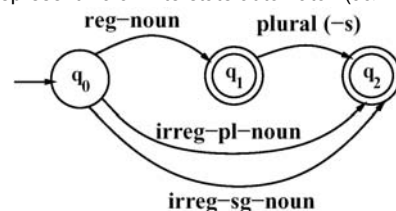# Goal

- Input: surface form
- Output: stem plus morphological features
- Focus: productive nominal plural (-*s*)
        verbal progressive (-*ing*)
  - foxes → fox +N +PL
  - geese → goose +N +PL
  - eating → eat +V +PRES-PART
  - goose → (goose +N +SG) or (goose +V)
  - gooses → ??

What knowledge sources will we need?

- Lexicon
  - List of stems and affixes with basic information about each
- Morphotactics
  - Model of morpheme ordering
  - Explains which classes of morphemes can follow others
- Spelling rules
  - Orthographic rules
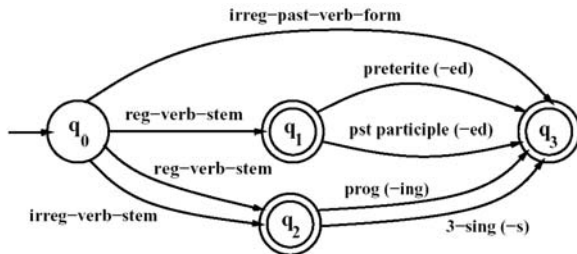 - Model the spelling changes that occur in a word when two morphemes combine

The lexicon

- Usually not represented as a list of words
- Structured as
  - List of stems and affixes
  - Representation of the morphosyntactics
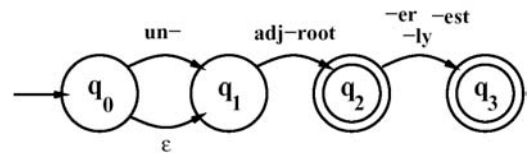- Represent via a finite-state automaton (J&M Ch. 2)

J&M Fig 3.2

Verbal inflection

FSA's for derivational morphology

- Much more complex
- Often use CFG's instead
- Consider adjective morphology…what's the problem?

FSA's for morphological recognition

- Goal: Use the FSA's to determine whether an input string of letters makes up a legitimate English word
  - Combine the list of stems with the FSA
  - Expand each arc with all of the morphemes that comprise the class