MORPHOLOGY

By:

B.Senthil Kumar, Asst. Prof

CSE, SSN College of Engineering

Natural Language Processing

Agenda

- Morphology – An Introduction

- Survey of English Morphology

    – Inflectional Morphology

    – Derivational Morphology

Morphology – An Introduction

- Plural of Fox ?

  Fox**es**

- Plural of Lion ?

  Lion**s**

- Plural of Butterfly ?

  Butterfl**ies**

- Plural of Fish ?

Morphology – An Introduction

- Two kinds of knowledge to search for singular and plural

  - Spelling(Orthographic) rules: words ending in *-y* are pluralized by changing the *-y* to *-i-* and adding an *-es* . Ex: butterfly --> butterfl*ies*

  - Morphological rules: *fish* has null plural, plural of *goose* is (*geese*) formed by changing the vowel

- Morphological parsing is the process of finding the constituent morpheme in a word

- The problem of recognizing that *foxes* breaks down into the two morphemes *fox* and *-es* is called *morphological parsing*

- Parsing means taking an input and producing some sort of linguistic structure for it

Morphology – An Introduction

- Morphology parsing applies to many affixes other than plurals.

- Given the surface or input form *going*, we might want to produce the parsed form: VERB-go + GERUND-ing {verbal stem+*ing* morpheme}

- Why not, just store all the plural forms of English nouns and *-ing* forms of English verbs in a dictionary and do parsing by look up?

- 1. It is quite inefficient to list all forms of noun and verb in the dictionary because the productivity of the forms

- Productive suffix like *-ing* applies to every verb, similarly *-s* applies to every noun. Ex: *fax* in *-ing* form*: faxing*

- A productive is a one that automatically includes any new words that enter the language

Morphology – An Introduction

- 2. The plural form of new nouns depends on the spelling/pronunciation of the singular form:

    – Noun ends in *-z* then plural form is *-es* rather than *-s*.

- 3. Can not list morphological variants of every word in morphological complex languages like Turkish

    – On an average Turkish word has three morphemes

    – A Verb in Turkish can have 40,000 possible forms without derivational suffixes.

- Key algorithm for morphological parsing is finite-state transducers (FST)

Survey of English Morphology

- Morphology is the study of the way words are built up from *smaller meaning-bearing units*, [morphemes](#)

- Two broad classes of morphemes:

  - The stems: the "main" morpheme of the word, supplying the main meaning

  - The affixes: add "additional" meanings of various kinds

Survey of (Mostly) English Morphology

- Affixes are further divided into prefixes, suffixes, infixes, and circumfixes.

    – Prefix: precede the stem **->** *un-buckle*

    – Suffix: follow the stem -> *eat-s*

    – Circumfix: both precede and follow stem

    past participle of verb *sagen (to say)* is *ge-sag-t (said) (in German)*

    – Infix: inserted inside the stem – two morphemes *hingi* and *um*

    hingi (borrow) hum ingi (the agent of an action) in Philippine language

    Tagalog

Survey of (Mostly) English Morphology

- Prefixes and suffixes are often called concatenative morphology

- A number of languages have extensive non-concatenative morphology

  - The Tagalog infixation example (*hingi* and *um* are intermingled)

  - Templatic morphology or root-and-pattern morphology, common in Arabic, Hebrew, and other Semitic languages

  - In Hebrew: a verb = root (CCC) + template

  - Root: *lmd* means 'learn' or 'study'

  - Active voice template: CaCaC --> *lamad* means 'he studied'

  - Intensive template: CiCeC --> *limed* means 'he taught'

  - Intensive passive voice template: CuCaC --> *lumad* means 'he was taught'

Survey of (Mostly) English Morphology

- A word can have more than one affix.
    - Rewrites – stem *write* and two affixes (*re-, -s*)
    - Unbelievably – stem *believe* and three affixes (*un-, -able, -ly*)

- English does not tend to stack more than four or five affixes

- Languages that tend to string affixes together like Turkish (words with nine or ten affixes) are called **agglutinative** languages

- Four ways to combine morphemes to create words:

- Inflection: word stem + morpheme resulting in a word of same class

- Derivation: word stem + morpheme resulting in word of a different class

- Compounding: combination of multiple word stems together. Ex: milk+man

- Cliticization: word stem + clitic. Ex: *I* + *'ve = I've*

Survey of (Mostly) English Morphology

- Two broad classes of ways to form words from morphemes:

  - **Inflection:** the combination of a word stem with a grammatical morpheme, usually resulting in a word of the *same class* as the original stem

  - Ex: *-s* for marking the plural on nouns and *-ed* for marking past tense on verbs

  - **Derivation**: the combination of a word stem with a grammatical morpheme, usually resulting in a word of a *different* class, often with a meaning hard to predict exactly

  - Ex: verb *computerize* with suffix *-ation* produce the noun *computerization*

Inflectional Morphology
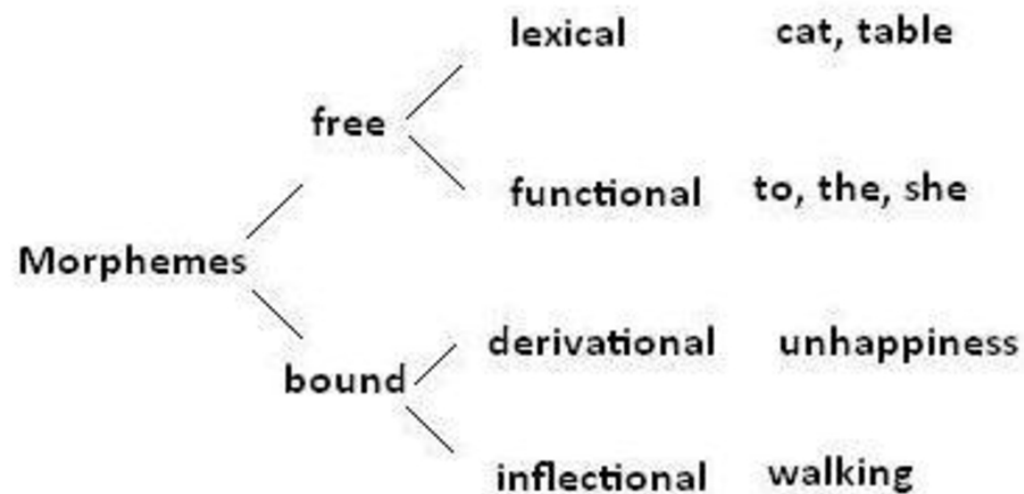
Inflectional Morphology

- In English, only nouns, verbs, and sometimes adjectives can be inflected, and the number of affixes is quite small.

- Two kinds of inflections in English nouns:

  - An affix marking **plural**,

    - Regular nouns: cat(-s), thrush(-es)

    - Irregular nouns: ox (oxen), mouse (mice)

    - most nouns affix is -s → boy(s), apple(s), book(s)

    - Words ending in *-s*, *-z*, *-sh*, *-ch*, *-x* + affix *-es*
      class(-es), waltz(-es), bush(-es), inch(-es), box(-es)

    - Words ending in *-y* change *-y* to *-i* + affix *-es*
      butterfly(-i)-es → butterflies, baby(-i)-es → babies

Inflectional Morphology

- Two kinds of inflections in English nouns:

  – An affix marking **plural**

  – An affix marking **possessive**

    - Apostrophe + -s for regular singular nouns (Ex: Rama's), plural not ending in -*s*, children'*s*

    - Lone apostrophe after regular plural nouns (Ex: Student*s*' guide), some names ending in -*s*, Euripide*s*' comedies

Inflectional Morphology

- Verbal inflection is more complicated than nominal inflection.

    - English has three kinds of verbs:

        - **Main verbs**, *eat*, *sleep*, *impeach*

        - **Modal verbs**, *can will, should*

        - **Primary verbs**, *be, have, do*

    - Main and Primary verbs have inflectional endings

    - Regular verbs: all verbs of this class have same endings marking the same function

stem	walk	merge	try	map
-*s* form	walks	merges	tries	maps
-*ing* principle	walking	merging	trying	mapping
Past form or –*ed* participle	walked	merged	tried	mapped

Inflectional Morphology

- These regular verbs and forms are significant because of their *majority* and being *productive*.

  - Recently created verb: *fax*

    - *fax* (-es, -ing, -ed) → *faxes, faxing, faxed*

    - *My mom <u>faxed</u> me the note from cousin Everett*

- An irregular verb often have five forms of inflection (*e.g., eat*), or eight (*e.g.,* the verb *be)* or even three (*e.g., cut*)

- Irregular verbs inflect in the past form (***preterite***) by changing:

  - its vowel (*eat/ate, dig/dug*)

  - its vowel and some consonants (*catch/caught*)

  - no change at all (*cut/cut*)
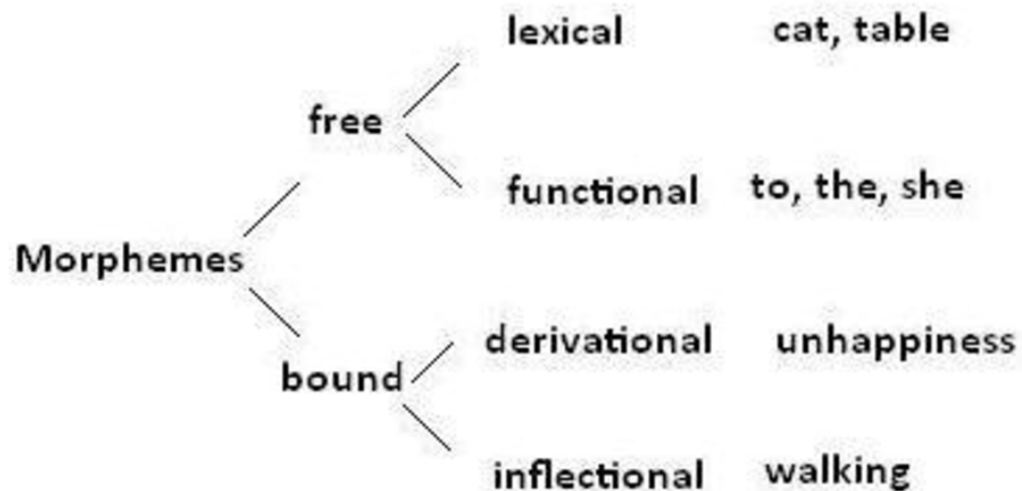
Inflectional Morphology

- The irregular verbs have some more or less idiosyncratic forms of inflection

  - Stem form is used in the infinitive form (I want to _walk_ home)

  - -s form is used to distinguish the third-person singular ending (She _talks_ everyday)

  - _-ing_ participle is used to mark present or ongoing activity (It's _raining_), when the verb is treated as a noun (called _Gerund_) (_Fishing_ is fine)

  - _-ed_ participle is used in the _perfect_ construction (_He's eaten lunch already_) or the _passive_ construction (_The verdict was overturned yesterday_)

stem	eat	catch	cut
_-s_ form	eats	catches	cuts
-ing participle	eating	catching	cutting
Past form	ate	caught	cut
_–ed/-en_ participle	eaten	caught	cut

Inflectional Morphology

- Number of spelling changes occur at these morpheme boundaries.

- A single consonant letter is doubled before adding *-ing, -ed* suffixes
  dig(*-ing*) → *digging, beg → begging / begged*

- If the base ends in a silent -e, it is deleted before adding *-ing, -ed*
  merge(*-ing/-ed*) → *merging, merged*

- Words ending in *-s, -z, -sh, -ch, -x* + affix *-es*
  toss(-es), waltz(-es), wash(-es), catch(-es), tax(-es)

- Words ending in *-y* change *-y* to *-i* + affix *-es*
  try(-i)-es → tries, fly(-i)-es → flies

Derivational Morphology

Derivational Morphology

- Nominalization in English:

  - The formation of new nouns, often from verbs or adjectives

Suffix	Base Verb/Adjective	Derived Noun
-ation	computerize (V)	computerization
-ee	appoint (V)	appointee
-er	clean (V)	cleaner
-ness	fuzzy (A)	fuzziness

  - Adjectives derived from nouns or verbs

Suffix	Base Noun/Verb	Derived Adjective
-al	computation (N)	computational
-able	embrace (V)	embraceable
-less	clue (N)	clueless

Derivational Morphology

- Derivation in English is more complex than inflection because

  - Generally *less productive*

    - A nominalizing affix like *–ation* can not be added to absolutely every verb. *eatation*(*)

  - There are subtle and complex meaning differences among nominalizing suffixes – *sincerity* has a subtle difference in meaning from *sincereness*

plays

played

player

playing

playful

play

play

Hasna

Played

playground

playground

playmate

playdate

THANK YOU