Natural Language Processing
CMSC 723 (spring, 2001)

March 5, 2001

- English Morphology
- Finite State Transducers (FSTs)
- Orthographic Rules and FSTs
- Combining FST Lexicon and Rules
- FSTs and Ambiguity
- (If time:) Comments about upcoming midterm

Review of FSAs

- FSAs are formal devices used to recognize strings in a language.
- FSAs can be specified using any of a number of formalisms.
- FSAs can be combined in various ways to create FSAs for new regular languages.

Morphology

Morphology is the study of the way words are built up from smaller meaningful units called **morphemes**.

- Stems — The core meaningful units in a lexicon.
- Affixes — Bits and pieces that combine with stems to modify their meanings and grammatical functions.

Word Class

By **word class** we have in mind notions like “noun” and “verb” with which you are probably familiar.

Two Types of Morphology

- Inflectional morphology
- Derivational morphology

Inflectional Morphology

- same word class as the stem
- can serve grammatical role that the stem could not.

Nouns and Verbs (in English)

Nouns have a simple inflectional morphology: markers for plural and markers for possessives.

Verbs complex inflectional morphology: markers appropriate to the function the verb is being asked to serve.

Regulars and Irregulars

- Mouse/Mice, Ox/Oxen, and Goose/Geese
- Go/Went, Fly/Flew

Regular Verbs

Morphological Form	Irregularly Inflected Verbs			
stem	walk	merge	try	map
-s form	walks	merges	tries	maps
-ing participle	walking	merging	trying	mapping
Past form or -ed participle	walked	merged	tired	mapped

Irregular Verbs

Morphological Form	Irregularly Inflected Verbs		
stem	eat	catch	cut
-s form	eats	catches	cuts
-ing participle	eating	catching	cutting
Past form	ate	caught	cut
-ed participle	eaten	caught	cut

Spanish: Amar

[Figure 3.1]

Derivational Morphology

Derivational morphology refers to the stuff that you know but no one ever told you.

Characterized by:

- Quasi-systematicity
- Irregular meaning changes
- Changes of word class

Derivational Morphology Examples

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

More Derivational Morphology Examples

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

English Adjective Morphology

[Figure 3.4]

English Adjective Morphology (continued)

[Figure 3.5]

English Derivational Morphology

[Figure 3.6]

Morphological Parsing

Morphological Parsing: Process of taking an input word and producing a stem plus associated morphological features.

Input	Morphological Parsed Output
cats	cat +N +PL
cat	cat +N +SG
cities	city +N +PL
geese	goose +N +PL
goose	(goose +N +SG) or (goose +V)
gooses	goose +V +3SG
merging	merge +V +PRES-PART
caught	(catch +V +PAST-PART) or (catch +V +PAST)

Building a Morphological Parser

- Lexicon: A list of stems and affixes together with basic info about them.
- Morphotactics: The model of morpheme ordering that explains which classes of morphemes can follow other classes of morphemes inside a word.
- Orthographic Rules: Spelling rules.

FSA Example: English Nominal Inflection

[Figure 3.2]

Embedding the Lexicon

[Figure 3.7]

Who cares?

Should search engines make use of morphological information?

- How might they make use of it?
- What happens if they do?
- What happens if they don't?

Finite State Transducers

- Add an extra tape to normal FSA.
- Add extra symbols to transitions of a normal FSA.

Tapes and Transitions

[Figure 3.8]

Finite State Transducer (FST)

- As recognizer—takes pair of strings as input and outputs accept or reject.
- As generator—outputs pairs of strings of the language.
- As translator—reads a string and outputs a string
- As set relater—computes relations between sets

Some Terminology and Notation

- Upper = lexical tape
- Lower = surface tape.
- Characters correspond to pairs, written a:b.
- Two-level lexical entries.
- If a=b, write a for shorhand.
- # = word boundary
- ˆ = word boundary
- Other = "any feasible pair that is not in this tranducer"

Nominal Inflection FST

[Figure 3.11]

Lexical and Intermediate Tapes

[Figure 3.12]

Spelling Rules

Name	Description of Rule	Example
Conson Double	Double letter before -ing/-ed	beg/begging
E deletion	Silent e dropped before -ing/-ed	make/making
E insertion	added after s,z,x,ch,sh before -s	watch/watches
Y replacement	-y changes to -ie before -s, -i before -ed	try/tries
K insertion	verbs ending with vowel+c, add -k	panic/panicked

Multi-Level Multi-Tape Machines

[Figure 3.13]

Chomsky and Halle notation (on homework!)

$$\epsilon \rightarrow \mathsf{e} \;/\; \left\{ \begin{array}{c} \mathsf{x} \\ \mathsf{s} \\ \mathsf{z} \end{array} \right\} \wedge \;\_\_\; \#$$

Intermediate to Surface Transducer

[Figure 3.14]

State Transition Table

[Figure 3.15]

Two Level Morphology

[Figure 3.16]

Tracing Foxes

[Figure 3.17]

FSTs and Ambiguity

Parse: `unionizable`

What to do?

- Accept first successful structure.
- Run parser through all possible paths.
- Bias the search in some manner.

Ambiguity (continued)

Parse: `assess`