

CSCI 5832

Natural Language Processing

Lecture 6
Jim Martin

2/1/07

CSCI 5832 Spring 2007

1

Today 2/1

- Review
- Noisy Channel Model
- Basic Probability Review
- Break
- N-Gram language models

2/1/07

CSCI 5832 Spring 2007

2

Review

- **FSA/FSTs can do lots of cool stuff but... they can't do it all.**
 - **In many cases they simply don't have the power to handle the facts (e.g. $a^n b^n$)**
 - *More on this later (e.g. CFGs)*
 - **In the case of global ambiguity, they can't tell us which output is more likely to be the correct one**

2/1/07

CSCI 5832 Spring 2007

3

So...

- **We'll modify finite state machines so they can tell us more about how likely various (correct) outputs are.**
 - **By applying some simple probability theory**

2/1/07

CSCI 5832 Spring 2007

4

Noisy Channel

- An influential metaphor in language processing is the noisy channel model



2/1/07

CSCI 5832 Spring 2007

5

Noisy Channel

- Obvious applications include
 - Speech recognition
 - Optical character recognition
 - Spelling correction
- Not so obvious
 - Semantic analysis
 - Machine translation
 - I.e German to English is a matter of uncorrupting the original signal

2/1/07

CSCI 5832 Spring 2007

6