



# Speech Processing 15-492/18-492

---

Speech Synthesis

Overview

Text processing

# Speech Synthesis

- ◆ *From text to speech*
- ◆ *Text Analysis*
  - *Strings of characters to words*
- ◆ *Linguistic Analysis*
  - *From words to pronunciations and prosody*
- ◆ *Waveform Synthesis*
  - *From pronunciations to waveforms*

# Text Analysis

- ◆ *This is a pen.*
- ◆ *My cat who lives dangerously has nine lives.*
- ◆ *He stole \$100 from the bank.*
- ◆ *He stole 1996 cattle on 25 Nov 1996.*
- ◆ *He stole \$100 million from the bank.*
- ◆ *It's 13 St. Andrew St. near the bank.*
- ◆ *Its a PIII 1.5Ghz, 512MB RAM, 160Gb SATA, (no IDE) 24x cdrom and 19" LCD.*
- ◆ *My home page is <http://www.geocities.com/awb/>.*

# Email

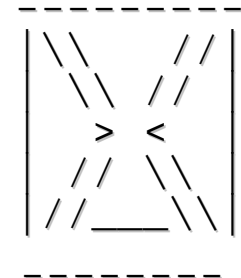
```
from awb@cstr.ed.ac.uk ("Alan W Black") on Thu 23 Nov 15:30:45:
>
> ... but, *I* wont make it :-) Can you tell me who's going?
>
  IMHO I think you should go, but I think the followign are going
    George Bush
    Bill Clinton
    and that other guy
```

Bob

--

```
-----
+-----+
| Bob Beck  E-mail bob@beck.demon.co.uk |
+-----+
```

Alba gu brath



# Text Analysis Tasks

- ◆ *Character encodings:*
  - *Latin-1, iso-8859-1, utf-8 (or special)*
- ◆ *Find tokens*
  - *White space separated*
- ◆ *Chunk into reasonably sized chunks*
  - *Sort of sentences*
- ◆ *Map tokens to words*
- ◆ *Disambiguate token types*
  - *Numbers*

# Chunking

- ◆ *Making reasonable sized sections*
  - *Something to do with full stops ...*

*Hi Alan,*

*I went to the conference. They listed you as Mr. Black when we know you should be Dr. Black days ahead for their research. Next month I'll be in the U.S.A. I'll try to drop by C.M.U. if I have time.*

*bye  
Dorothy  
Institute of XYZ  
University of Foreign Place  
email: dot@com.dotcom.com*

# Text analysis

- ◆ *Normal words*
  - *Homographs, OOVs*
- ◆ *Numbers*
  - *Years, quantities, digits, addresses*
- ◆ *Other standard forms*
  - *Dates, times, money*
- ◆ *Abbreviations and Letter Sequences*
  - *NASA, CIA, SATA, IDE*
- ◆ *Spelling errors (choices)*
  - *Sooooo, ... colour, collar*
- ◆ *Punctuation*
  - *:-) quotes, dashes, ascii art,*
- ◆ *Text layout*

# Finding Words

- ◆ *White space separated tokens*
  - *But---if I may interject---not all word(s) are like that*
  - *Wean-Hall-like architecture*
- ◆ *Some languages don't use spaces*
  - *Chinese, Japanese, Thai*
- ◆ *Some languages use lots of compounding*
  - *unspacedmultiwords*

# Homographs

## ◆ *Homographs*

- *Same writing, different pronunciation*
- *(Homophones: same pronunciation different writing. “to” “two” “write” “right”)*

## ◆ *English: not many:*

- *Stress shift (Noun/Verb)*
  - ⊠ *Segment, project, convict*
- *Semantic*
  - ⊠ *Bass, read, Begin, bathing, lives, Celtic, wind, Reading, sun, wed, ...*
  - ⊠ *Roman Numerals*

# Non-standard Words (NSW)

- Words not in the lexicon

<i>Text Type</i>	<i>%NSW</i>
<i>Novels</i>	<i>1.5%</i>
<i>Press wire</i>	<i>4.9%</i>
<i>Email</i>	<i>10.7%</i>
<i>Recipes</i>	<i>13.7%</i>
<i>Classifieds</i>	<i>27.9%</i>
<i>IM</i>	<i>20.1%</i>

# Distribution of NSW

- 3yrs News text, 2.2M tokens 120K NSWs

<i>Major type</i>	<i>Minor type</i>	<i>% of NSW</i>
<i>Numeric</i>	<i>Number</i>	<i>26%</i>
	<i>Year</i>	<i>7%</i>
	<i>Ordinal</i>	<i>3%</i>
<i>Alphabetic</i>	<i>As word</i>	<i>30%</i>
	<i>As letters</i>	<i>12%</i>
	<i>As Abbrev</i>	<i>2%</i>

# Processing NSWs

## ◆ *How hard are they?*

- *Finding them*
- *Identifying them*
- *Expanding them*

## ◆ *Current processing techniques*

- *Ignored*
- *Lexical lookup*
- *Hacky hand-written rules*
- *(not so) Hacky hand-written rules*
- *Statistically train models (and hacky hand written rules)*

# Homograph Disambiguation (Yarowsky)

- ◆ *Same tokens in different contexts*
- ◆ *Identify target homograph*
  - *E.g. numbers, roman numerals, “St”*
- ◆ *Find instances in large text corpora*
- ◆ *Hand label them with correct answer*
- ◆ *Train a decision tree to predict types*

# NSW: Roman Numerals

- ◆ *Roman Numerals as cardinal, ordinals or letters*
  - *Henry V: Part I Act II Scene XI: Mr X I believe is V I Lenin, and not Charles I.*
- ◆ *Ordinal: Henry V*
- ◆ *Number: Part II*
- ◆ *Letter: Mr X*
- ◆ *Times: 2 X 4 inches*
- ◆ *Word: I am.*

# NSW models

- ◆ *What features help predict class:*
  - *The word form itself*
  - *The word “King” “Queen” “Pope” nearby*
  - *A king/queen/pope name nearby*
  - *Capitalization of nearby words.*
- ◆ *class: n(umber) l(etter) r(ex) t(imes)*
- ◆ *rex rex\_names section\_names num\_digits p.num\_digits, n.num\_digits, pp.cap, p.cap, n.cap, nn.cap*

*n II 0 0 0 11 7 2 3 7 0 0 1 1*

*n III 0 0 0 3 4 3 3 5 0 0 1 1*

*r VII 1 0 0 4 9 3 3 3 1 1 0 0*

*n V 0 0 1 3 1 4 1 2 0 1 0 1*

*...*

# CART Tree

- Automatically find which feature questions give the best answers
- Classification (and Regression) Trees (CART)

```
((p.lisp_tok_rex_names is 0)
  ((lisp_num_digits is 5)
    ((number))
    ((lisp_num_digits is 4)
      ((number))
      ((nn.lisp_num_digits is 13)
        ...
        ((nn.lisp_num_digits is 2)
          ((letter))
          ((n.cap is 0) ((letter)) ((number))))))))
...

```

# Hard cases

- ◆ *Some harder roman numeral cases*
  - *William B. Gates III*
  - *Meet Joe Black II*
  - *The madness of King George III*
  - *He's a nice chap. I met him last year*

# Letters, Abbrevs and Words

- ◆ *How to pronounce an unknown letter sequence:*
  - ◆ *Letters: IBM, CIA, PCMCIA, PhD*
  - ◆ *Words: NASA, NATO, RAM*
  - ◆ *Abbrev: etc, Pitts, SqH, Pitts Int. Air.*
  - ◆ *Hybrids: CDROM, DRAM, WinNT, MacOS*
- Letter language model (letter frequencies)*

# NSW models



- ◆ *Classified ads*

- ◆ *57 ST E/1st & 2nd Ave Huge  
drmn 1 BR 750+ sf, lots of sun &  
clsts. Sundeck & lndry facils. Askg  
\$187K, maint \$868, utils  
incl. Call Bkr Peter 914-428-9054.*




- ◆ *Default model* 

- ◆ *Trained model* 

# Domain Knowledge

- ◆ *Modify text processing for the domain:*
- ◆ *Smith, Bobbie Q, 3337 St Laurence St,  
Fort Worth, TX 71611-5484, (817)839-3689*
- ◆ *Anderson, W, 445 Sycamore Way NE,  
Lincoln, NE 98125-5108, (212)404-9988*
- ◆ *Standard Mode* 
- ◆ *Address Mode* 

# Sometimes need more than text

- ◆ *Different context requires different delivery*
- ◆ *What will the weather be like today in Boston?*
  - *It will be **rainy** today in Boston.* 
- ◆ *When will it be rainy in Boston?*
  - *It will be rainy **today** in Boston* 
- ◆ *Where will it be rainy today?*
  - *It will be rainy today in **Boston*** 

# Mark-up Languages

- ◆ *Add explicit markup to text*
- ◆ *Can be done in machine generated text*
- ◆ *SSML (Speech Synthesis Markup Language)*
  - *Choice voices, languages*
  - *Give pronunciations*
  - *Specify breaks, speed, pitch*
  - *Include external sounds*

# SSML Example

- ◆ 

```
<?xml version="1.0"?>
<!DOCTYPE SABLE PUBLIC "-//SABLE//DTD SABLE speech mark up//EN"
    "Sable.v0_2.dtd"
[]> <SABLE> <SPEAKER NAME="male1">
```

*The boy saw the girl in the park <BREAK/> with the telescope.  
The boy saw the girl <BREAK/> in the park with the telescope.*

*Some English first and then some Spanish.  
<LANGUAGE ID="SPANISH">Hola amigos.</LANGUAGE>  
<LANGUAGE ID="NEPALI">Namaste</LANGUAGE>*

*Good morning <BREAK /> My name is Stuart, which is spelled  
<RATE SPEED="-40%">  
<SAYAS MODE="literal">stuart</SAYAS> </RATE>  
though some people pronounce it  
<PRON SUB="stoo art">stuart</PRON>. My telephone number  
is <SAYAS MODE="literal">2787</SAYAS>.*

*I used to work in <PRON SUB="Buckloo">Bucclench</PRON> Place,  
but no one can pronounce that.*

*By the way, my telephone number is actually  
<AUDIO SRC="http://att.com/sounds/touchtone.2.au"/>  
<AUDIO SRC="http://att.com/sounds/touchtone.7.au"/>  
<AUDIO SRC="http://att.com/sounds/touchtone.8.au"/>  
<AUDIO SRC="http://att.com/sounds/touchtone.7.au"/>.*



# Summary

- ◆ *Text to Speech*
  - *Text analysis*
  - *Linguistic analysis*
  - *Waveform Synthesis*
- ◆ *Text analysis*
  - *Chunk text*
  - *Find tokens and their types*
  - *Convert to standard words*
- ◆ *Non-standard Words (NSW)*

