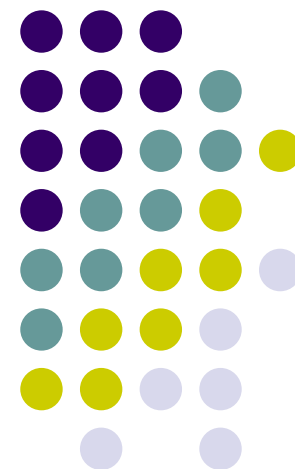


# Context-Sensitive Statistical Language Modeling

Alexander Gruenstein  
Chao Wang  
Stephanie Seneff

<http://www.sls.csail.mit.edu>



# Context-Sensitive Statistical Language Modeling: *Outline*



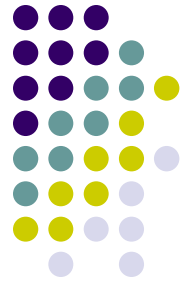
- Motivation and Current Approaches
- A New Approach: Context-Sensitive Dynamic Classes
- Evaluation
- Summary



# Motivation

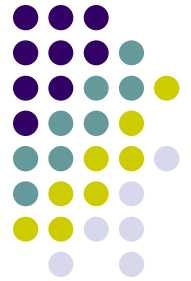
- Spoken dialogue systems often track detailed context via an **information state**
  - Dialogue history
  - Domain-specific context
    - flight reservation systems: partially-formed reservation, available flights
    - robot control: the position, plan, *etc* of the robot (Gruenstein 2002)
- Information state used for *e.g.*
  - **NL interpretation**: anaphora resolution, fragment interpretation, question/answers,...
  - **NL generation**: anaphora, aggregation, ...
- Not often used to constrain speech recognizer's **language model**

# A Common Current Approach: *N*-gram Interpolation



- Dialogue-state specific *n*-gram language models trained on state-specific utterances
  - **Interpolated** with *n*-gram from all in-domain data
  - see (Solsona, et al, 2002), (Aaron, et al, 2001), (Visweswariah & Printz, 2001), (Wessel, et al, 1999), (Xu & Rudnicky, 2000)
- Drawbacks:
  - Arbitrary **partitioning** of dialogue states into equivalence classes of dialogue history
  - Must tweak interpolation **weights**, but perplexity on held-out set poor predictor

# Background: MERCURY Flight Reservation Dialogue System



- Work presented here builds on the **MERCURY** spoken dialogue system (Seneff, 2002)
    - Mixed-initiative flight-reservation system
    - Corpus of 26,886 **transcribed** utterances in **log files** containing records of the **information state**
    - SUMMIT speech recognizer with dynamic classes (Chung, *et al* 2004)
- U1:* I'd like to fly from Boston to San Francisco on the third
- S1:* Okay, from Boston to San Francisco on March third. Can you provide an approximate departure time or airline?
- U2:* United in the afternoon.
- S2:* I have a flight at three p.m. would that work? Or how about five thirty p.m.?
- U3:* The first one.

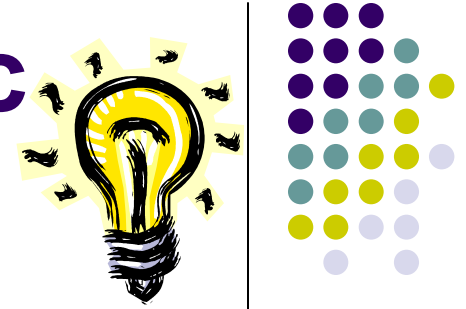
# Context-Sensitive Statistical Language Modeling: *Outline*



- Motivation and Current Approaches
- A New Approach: Context-Sensitive Dynamic Classes
- Evaluation
- Summary

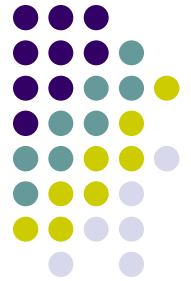
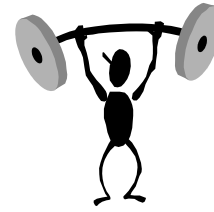


# Context Sensitive Dynamic Classes: Overview



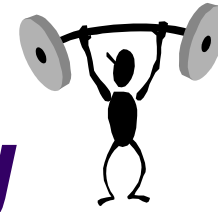
- **Single** class  $n$ -gram language model
  - **Dynamic** classes with expansions populated at run time, based on the current dialogue-system **information state**.
    - Particular classes active only in certain contexts: yielding **context-dependent  $n$ -gram counts** for those classes
  - “**Strong**” dynamic classes populated with **specific** entity expansions that are *strongly* contextually relevant
    - *e.g.* A specific airline or city name we strongly expect to be uttered
  - “**Weak**” dynamic classes populated with the entire set of possibilities, but only one of  $N$  classes are **active**, depending on the **context**
    - *e.g.* All possible airlines, but following a particular prompt

# ***Strong* Context-Sensitive Dynamic Classes**

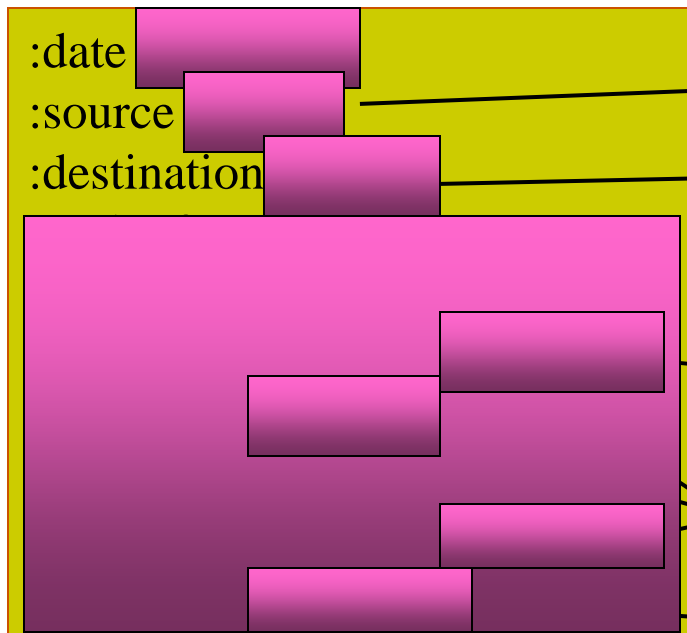


- In *addition* to **default** *n*-gram classes, add to the language model **context-sensitive** classes with expansions dependent on **specific** contextual cues
- **Default** classes:
  - *\$city, \$state, \$airport, \$airline, \$digit, ...*
  - These expand to e.g. **all cities**
- **Strong** context-sensitive classes:
  - *\$dynsource, \$dyndestination, \$dynairport, \$dynairline, \$dyntime*
  - Expansions on **next slide**

# Strong Context-Sensitive Dynamic Classes: Tagging

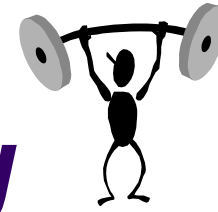


Information State Snippet  Class Expansions



<i>\$dynsource</i>	→	“austin”
<i>\$dyndest</i>	→	“san francisco”
<i>\$dyntime</i>	→	“two”, “two o’clock”, “two p_m”, “two o’clock p_m”, “four thirty”, “four thirty p_m”
<i>\$dynairline</i>	→	“united”, “united airlines”, “american”, “american airlines”
<i>\$dynnthflight</i>	→	“the first one”, “the first flight”, “the second one”, “the second flight”

# Strong Context-Sensitive Dynamic Classes: Tagging



UI: I'd like to fly from Austin/**\$city** to San Francisco/**\$city** on the **third/\$digit**.

## Information State Snippet

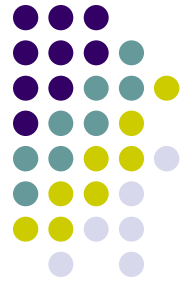
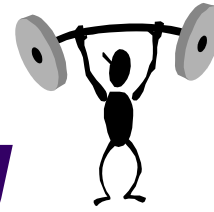
```
:date ""  
:source ""  
:destination ""  
:reply_frame { }
```



## Class Expansions

<i>\$dynsource</i>	[]
<i>\$dyndest</i>	[]
<i>\$dyntime</i>	[]
<i>\$dynairline</i>	[]
<i>\$dynnthflight</i>	[]

# Strong Context-Sensitive Dynamic Classes: Tagging



*U1*: I'd like to fly from **Austin/\$city** to **San Francisco/\$city** on the **third/\$digit**.

*S1*: Okay, from **Boston** [*misrecognized*] to San Francisco on March third. Can you provide an approximate departure time or airline?

*U2*: Not **Boston/\$dynsource**; from **Austin/\$city** to **San Francisco/\$dyndest**

```
:date "March3"  
:source "BOS"  
:destination "SFO"  
:reply_frame {  
  : provide_departure_or_airline  
}
```

<i>\$dynsource</i>	"boston"
<i>\$dyndest</i>	"san francisco"
<i>\$dyntime</i>	[]
<i>\$dynairline</i>	[]
<i>\$dynnthflight</i>	[]

# Strong Context-Sensitive Dynamic Classes: Tagging



U1: I'd like to fly from Austin/*\$city* to San Francisco/*\$city* on

the  
 :date "March3"  
 :source "AUS"

S1: Ok  
 :destination "SFO"  
 :reply\_frame {

Ma  
 :best\_departure {  
   :departure\_time "2:00pm"  
   :airline "AA" }

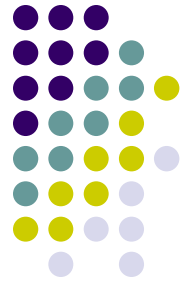
U2: No  
 :second\_departure {  
   :departure\_time "4:30pm"  
   :airline "UA" }}

<i>\$dynsource</i>	"boston"
<i>\$dyndest</i>	"san francisco"
<i>\$dyntime</i>	"two", "two o'clock", "four", "four o'clock"
<i>\$dynairline</i>	"united", "american", ...
<i>\$dynnthflight</i>	"the first one", "the second one", ...

S3: Okay, from Austin to San Francisco on March third. I've got a flight on American at two o'clock, would that work? Or I've got one on United at four thirty.

U3: How about the flight at two/*\$dyntime*.

# Weak Context-Sensitive Dynamic Classes

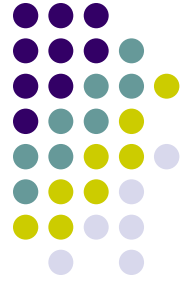


- Replace **default**  $n$ -gram classes with “**weak**” **context-sensitive** classes
  - Each class is populated with the complete set of possible expansions (e.g. all airlines), but the **active class** is determined based on **more general** contextual cues
- 3 different types of classes based on **3** contexts:
  - **C1**: Following a prompt for one of the class members: *Do you have an airline preference?*
  - **C2**: Following prompt: *How may I help you?*
  - **C3**: All other contexts
- Exactly one class is enabled at any given time

# Context-Sensitive Statistical Language Modeling: *Outline*



- Motivation and Current Approaches
- A New Approach: Context-Sensitive Dynamic Classes
- **Evaluation**
- Summary



# Evaluation: Overview

- Corpus of 26,886 **transcribed** utterances in **MERCURY log files**
  - Training: 24,815 utterances
  - Test set: 2,071 utterances
- Class **Trigram** language models
- Base vocabulary size of 1,586 (excluding city names, state names, airlines, airports)



# Evaluation: Conditions

- Manipulated two variables:
  - *Within-class expansion weight:*
    - *corpus* statistics `$city -> boston .153`
    - *uniform* `-> austin .0015`
    - *city population* `-> san jose .013`
  - *Class vocabulary size:* ...
    - *medium:* 516 city names, 329 city\_states, 68 airlines  
(*matches the training corpus*)
    - *large:* 16,956 city names, 25,334 city-states, 68 airlines
- *Resulting in 4 conditions:*
  - *CM, UM, PL, UL*

# Evaluation: Overall Word Error Rate



Word Error Rate			
	Static Baseline	Dynamic: <i>Strong</i>	Dynamic: <i>Weak</i>
CM	17.8	17.7	17.6
UM	25.2	24.4	20.8
PL	27.1	26.7	26.0
UL	46.7	45.0	42.1

**C**: Corpus based class expansion weights

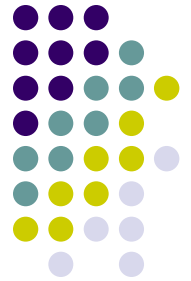
**U**: Uniform class expansion weights

**M**: Medium class vocabulary size

**L**: Large class vocabulary size



# Evaluation: *Utterances Targeted by Strong Classes*



- An utterance is “targeted” by a *strong* dynamic class when that class has non-empty expansions
  - S: I have United and Delta non-stop flights available
  - U: How about the **United** flight / I prefer American.

<b>Targeted <i>dyntime, dynairline, dynnthflight</i></b> (13% of test set)		
	Static Baseline	Dynamic: <i>Strong</i>
CM	18.6	16.0
UM	32.0	25.3
PL	30.4	27.6
UL	52.9	47.1

**See paper  
for much  
more  
detailed  
analysis!**



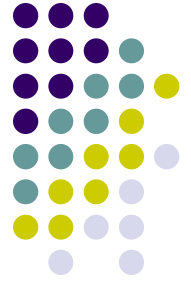
C: Corpus based class expansion weights  
U: Uniform class expansion weights

M: Medium class vocabulary size  
L : Large class vocabulary size

# Context-Sensitive Statistical Language Modeling: *Outline*



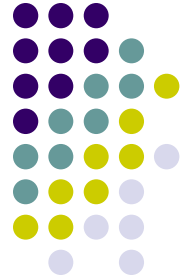
- Motivation and Current Approaches
- A New Approach: Context-Sensitive Dynamic Classes
- Evaluation
- **Summary**



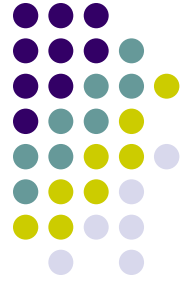
# Summary

- Incorporating context-sensitive dynamic classes improves recognition performance over standard  $n$ -gram language model
- Context sensitivity without **partitioning** the dialogue state space
  - Classes are populated independently based on particular portions of the **information state**
  - $N$ -gram frequency estimates involving the dynamic classes are calculated based on the **corpus**, not tweaked by hand, incorporating precise contextual expectation into the language model in **real time**
  - Scalable

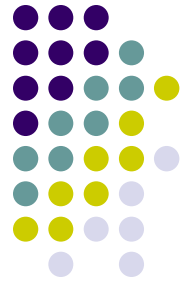
# Thank You



# Extra Slides



# Approaches: State-Specific Context-Free Grammars



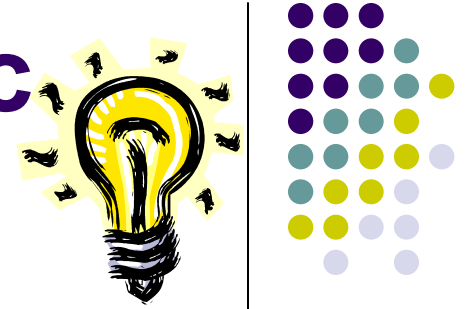
- Dialogue-state specific **context-free** language models
  - Hand crafted (as in many commercial systems)
  - Inferred from unification grammar with **back-off** based on confidence threshold
    - see (Lemon & Gruenstein, 2004)
- Drawbacks:
  - Hand-crafted or hand-partitioned grammars: difficult to maintain principled linguistic approach
  - One grammar per **dialogue state**
  - Sometimes requires multiple recognition passes

S -> Y | N

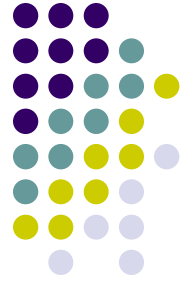
Y -> yes | yeah | yup

N -> no | nah | nope

# Context-Sensitive Dynamic Classes: *Summary*



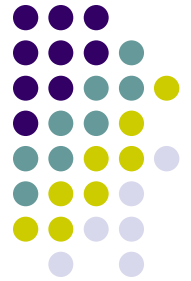
- Class  $n$ -gram is trained by **tagging** class expansions in the corpus based on the dialogue **context** provided by the **log files**
  - Each class is expected to commonly occur in a different context, resulting in **context-specific  $n$ -gram probabilities**
- At run-time, classes are **activated** and **populated dynamically**, or **deactivated**, depending on the context



# Room for Improvement

- Avoid **partitioning** the information state into one of **N** dialogue states
  - Requires forming equivalence classes of dialogue history (e.g. only taking into account the previous prompt)
- Instead, focus on **key phrases** or **syntactic constructions** likely to occur based on cumulative dialogue context

# Evaluation: *Utterances Targeted by Strong Classes*



- An utterance is “targeted” by a *strong* dynamic class when that class has non-empty expansions
  - S: I have United and Delta non-stop flights available
  - U: How about the **United** flight / I prefer American.

<b>Targeted source, destination</b>		
	Static Baseline	Dynamic: <i>Strong</i>
CM	19.5	19.2
UM	29.4	27.8
PL	30.1	29.1
UL	52.5	50.1

<b>Targeted time, airline, nthflight</b>		
	Static Baseline	Dynamic: <i>Strong</i>
CM	18.6	16.0
UM	32.0	25.3
PL	30.4	27.6
UL	52.9	47.1



C: Corpus based class expansion weights  
 U: Uniform class expansion weights

M: Medium class vocabulary size  
 L : Large class vocabulary size