

# Uphill Battles in Language Generation: Sentence Planning and Lexical Choice

Kathleen McKeown  
Columbia University

# Generating Sentences

- Original Goals
  - We were interested in the constraints on language generation: why do we say things the way we do?
  - Fluency and flexibility that comes with knowledge about syntax: we used grammars
  - Ability to paraphrase: we addressed lexical choice

# My Favorite Oldies

- Floating Constraints on Lexical Choice
  - Elhadad, McKeown, Robin
  - Constraints on lexical choice come from many sources: lexical, syntax, semantic, pragmatic
  - Understanding how and when these constraints play a role interesting
- FUF, Penman, OpenCCG (not an oldy ^^ )
  - The use of syntax enable a simple proposition to be realized in multiple ways

# Example

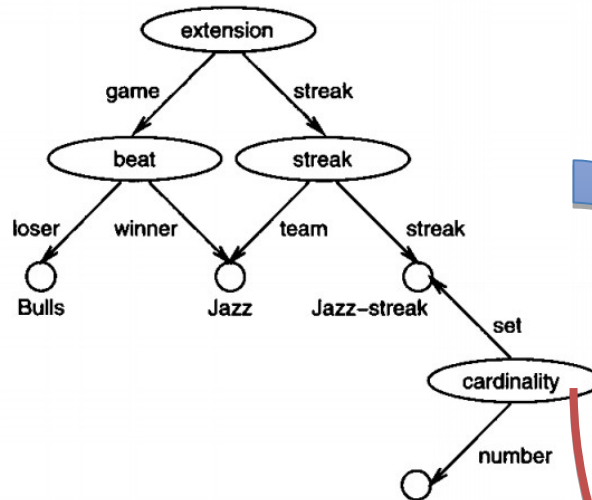
- The Denver Nuggets *beat* the Celtics with a *narrow margin*, 102-101.
  - Result in verb, manner in PP
- The Denver Nuggets *edged out* the Celtics, 102-101.
  - Result and manner in verb

# Problems

- How do we get from what we *want to say* to a sentence?
- Map from concepts to words and from propositions to sentences
- Resources Needed
  - Lexicon
  - Lexico-semantic sentence frame mappings

# Example

Elhadad, McKeown, and Robin



The Jazz defeated the Bulls.

They extended their winning streak to three games.

The Jazz defeated the Bulls for their third straight win.

# Today's methods

- Language modeling takes care of syntax

Ranking:

The Jazz defeated the Bulls for their third straight win.

The Jazz extended their winning streak to three games with a victory over the Bulls.

The Jazz defeated the Bulls.

They extended their winning streak to three games.

\*The Jazz defeated the Bulls to a third straight win.

# Today's Methods

- Learning templates = lexico-semantic frame
- Training data = box scores, basketball reports
- The <WT> defeated the <LT> for their <STREAK> straight win.
- The <WT> extended their winning streak to <STREAK> games with a victory over the <LT>.
- The <WT> defeated the <LT>.
- <WT> extended their winning streak to <STREAK> games.



# Today's Methods: Neural nets

Lebret et al, EMNLP16

**Barack Obama**



**44th President of the United States**

**Incumbent**

**Assumed office**

January 20, 2009

**Vice President Joe Biden**

**Preceded by George W. Bush**

**United States Senator**

**from Illinois**

**In office**

January 3, 2005 – November 16, 2008

Barack Hussein Obama II (US Listeni/bə' rɑ:k hu:'seɪn oʊ 'bɑ:mə/;[2][3] born August 4, 1961) is an American politician who is the 44th and current President of the United States.

Andreas&Klein, EMNLP16



(a) *the sun is in the sky*  
[contrastive]

Ghazvininejad & al, EMNLP16

XCVIII.  
The lads desire some juice and gin to drink,  
but I shall quaff a far less pow'rful sort  
of ale to keep a clearer head to think;  
Firtation is for me a noble sport.  
O Monica, thy liveliness inspires –  
O Erica, resplendent at my side –  
O Rita, with the love my heart requires –  
O Tina, such a beauty I've espied –  
O Sandra, shining like the brilliant Sun –  
O Mary, with our nights beyond compare –  
O Jessica, our romance just begun –  
and lo! To any other maiden fair  
— If thou wilt make with me some am'rous plan,  
a little bit of thee makes me thy man.

*Low Regen, "Mambo No. 5 (A Little Bit of...)"*



Poetry Generation

# What Have We Gained?

- Robustness
- Ability to work in purely statistical environments
- Works well with text-to-text summarization
- Works well in some question answering, image caption generation and explanation tasks

# What Have We Lost?

- Ability to reason about the *constraints* on language usage
- Why do we say things the way we do?
- When should we use different words and different styles?

Can we marry the two?