

Generation of Referring Expression

Regina Barzilay

February 25, 2004

Today's Topics

- Overview of Natural Language Generation
- Psychological Evidence regarding Generation of Referring Expressions
- Selection of Referring Expressions
 - Symbolic Approaches (Dale&Reiter)
 - Corpus-Based Approaches (Radev, Nenkova&McKeown)

What is NLG?

- Program which produces texts in natural language.
- Input: some underlying non-linguistic representation of information.
- Output: documents, reports, help messages and other types of texts.
- Knowledge sources required: knowledge of language and of the domain.

Why to Use NLG?

- Important information is stored in ways which are not comprehensible to the end users — databases, expert system, log files.
- NLG systems can present this information to users in an accessible way.
- Data presented in textual form can be searched by IR systems.

Example: Summary of basketball games

```
scoring((Shaquille, O'Neal), 37)
time(Friday, night)
team((Shaquille, O'Neal), (Orlando, Magic))
win(Orlando, Magic), (Toronto, Raptors)
score(101, 89)
...
```

Output

Orlando, FL — Shaquille O’Neal scored 37 points Friday night powering the Orlando Magic to a 101 89 victory over the Toronto Raptors, losers of seven in a row.

NLG System Architecture

- Content Determination
- Discourse Planning
- Sentence Aggregation
- Lexicalization
- Syntactic and morphological realization

Content determination

- Input: Knowledge base.
- Schemata and Inference mechanism.
- Output: predicates to be conveyed in the text.

Example:

Game statistics: <code>win(Orlando, Magic), (Toronto, Raptors)</code>
Player's records: <code>team((Shaquille, O'Neal), (Orlando, Magic))</code>
Team's record: <code>lost(7, (Toronto, Raptors))</code>

Discourse Planning

- Text have an underlying structure in which parts are related together.
- Rhetorical relationships.
- Conceptual grouping.

Sentence Planning

- A one-to-one mapping from predicates to messages is bad.
- Sentence planning groups information into sentences (aggregation).

Before aggregation: Shaquille O'Neal scored 37 points. The game was on Friday night. Orlando Magic defeated Toronto Raptors. Raptors lost seven games in a row.

After aggregation: Shaquille O'Neal scored 37 points Friday night powering the Orlando Magic to a 101 89 victory over the Toronto Raptors, losers of seven in a row.

Lexicalization + Syntactic Realization

- This stage determines the particular words to be used to express domain concepts and relations.
- Constraints: discourse focus, style constraints, syntactic environment.
- Implementation: decision tree.

Example: predicate: `win(X, Y)`

Verb: X defeated Y, Y was defeated X, X won in the game against X, X won the game
Noun:: victory of X over Y, victory of X, defeat of Y

Realization task

- Insert function words.
- Choose correct specification of content words.
- Order words.

FUF/SURGE input for the sentence “John likes Mary now.”

```
((cat clause)
  (proc ((type mental) (tense present) (lex "like")))
  (partic ((processor ((cat proper) (lex "John")))
          (phenomenon ((cat proper) (lex "Mary")))))
  (circum ((time ((cat adv) (lex "now"))))))
```

Basics

- A noun phrase is considered to be a referring expression iff its only communicative purpose is to identify an object too the hearer
- Context is the set of entities that the hearer is currently assumed to be attending to
- Referring expression satisfies the referential communicative goal if it is a distinguishing description in the given context

Example: small black dog, large white dog, small black cat

Impact of Conversational Implicature

Redundant Information causes violations of
Conversation Implicature (Grice, 1975)

Sit by the table.

Sit by the brown wood table.

Inappropriate Modifiers

Overly specific or unexpected classifiers violates principles of Conversation Implicature

Look at the dog.

Look at the pitt bull.

Psychological Data

(Levelt, 1989)

People do include unnecessary modifiers in referring expressions

a white bird, a black cup and a white cup

- Incremental Processing Helps in Understanding
- Redundancy Helps in Understanding

Transcript Analysis

(Reiter&Dale,1992) (“assembly task” dialog)

- Which attribute should be used?
gender vs color vs shape
- Is it preferable to use modifier or to use a more specific head noun?
the small dog vs the chihuahua
- Should relative or absolute adjectives be used?
the small dog vs the one foot high dog

Results of Transcript Analysis

- Preference for adjectives that communicate size, shape or color?
the black dog vs the male dog
- The use of specific head nouns depends on audience expertise
the small dog vs the chihuahua
- Preference for relative adjectives in speech, and for absolute adjectives in writing
the small dog vs the one foot high dog

Algorithm: Representation

(Reiter&Dale,1992)

- Input is organized in (attribute, value) pairs
- Type is one of the attributes
- Attributes are organized in taxonomy

Object₁: (type, chihuahua), (size, small), (color, black).

Object₂: (type, chihuahua), (size, large), (color, white).

Object₃: (type, cat), (size, small), (color, black).

Algorithm

- Check Success: see if the contracted description picks up one entity from the context
- Choose Property: determine which properties of the referent would rule out the largest number of entities
- Extend Description: add the chosen properties to the description being constructed and remove relevant entities from the discourse.

Text-to-Text Generation

- Input: text (lack of semantic information)
- Applications: summarization, question-answering, machine translation

Supervised Approach to Referent Selection

Goal: Select the best entity description in a given corpus
(Radev, 1998)

Elections (1996): “Bill Clinton, the democratic presidential candidate”

False bomb alert in Little Rock, Ark (1997): “Bill Clinton, an Arkansas native”

Key Idea

- Semantic constraints imposed on lexical choice are reflected in contextual indicators
- This correlation can be learned automatically from a large collections of texts, given a feature vector and a referent

Entity Profile

Collection of entity descriptions (automatically constructed)

Example: Profile of Ung Huot

a senior member, Cambodia's, Cambodian foreign minister, co-premier, first prime minister, foreign minister, MR., new co-premier, new first prime minister, newly-appointed prime minister, premier

Distributional Properties of Profile

- 11, 504 entities from 178 MB of newswire
- 9, 053 have a single description
- 2, 451 vary from 2 to 24 descriptions

Features

- Context — a bag of words surrounding the entity
- Length of the article — an integer
- Name of the entity — e.g., “Bill Clinton”
- Profile — set of all the descriptions
- Wordnet — WordNet extension for Profile members

Experimental Results

- Training: 10, 353 Testing: 1,511
- Precision: 88.87%, Recall 63.39%
- Steep learning curve
(500: 64.29%, 2.86% → 50,000: 88.87%, 63.39%)
- Positive impact of WordNet extension — 10% increase on average

Rule Examples

IF inflation IN CONTEXT, THEN “politician”

IF detective IN PROFILE AND agency in CONTEXT,
THEN “policeman”

IF celine IN CONTEXT, THEN “north american”

Unsupervised Referent Selection

Statistical Model of Syntactic Realization
(Nenkova&McKeown, 2003)

Hypothesis: There is a regularity in lexical realization of referent chain

Example

Many years ago, there was an Emperor, who was so excessively fond of new clothes, that he spent all his money in dress. He did not care to go either to the theatre or the chase, except for the opportunities then afforded him for displaying his new clothes. He had a different suit for each hour of the day; and as of any other king or emperor, one is accustomed to say, "he is sitting in council," it was always said of him, "The Emperor is sitting in his wardrobe." One day, two rogues, calling themselves weavers, made their appearance. They gave out that they knew how to weave stuffs of the most beautiful colors and elaborate patterns, the clothes manufactured from which should have the wonderful property of remaining invisible to everyone who was extraordinarily simple in character. "These must, indeed, be splendid clothes!" thought the Emperor.

Target NP Features

- Premodifiers:
 - Titles “President George W. Bush”
 - Name-external modifiers “Irish Flutist James Galway”
- Postmodifiers:
 - Apposition
 - Relative Clause
 - Prepositional Phrase Modificiation

Types of Referents

- Is the target named entity the head of the phrase or not?
- If it is the head what kind of pre- and post- modifier does it have?
- How was the name itself realized in the NP?

HHM Construction

- Each state of HMM corresponds to one syntactic realization
- Transitions are estimated based on corpus counts (anaphoric expressions are not resolved!)

Fragment of HMM

	Modification	No Modification
Initial	0.76	0.24
Modification	0.44	0.56
No modification	0.24	0.75

Evaluation

Rewriting rules based on HMM improve the performance of summarization system:

Preferences: 89% rewrite, 9% original, 2% no preference