

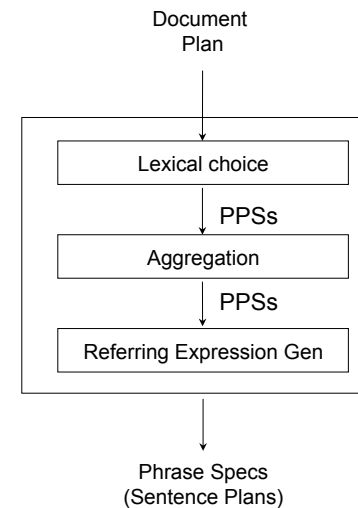
# Sentence Planning 2: Aggregation

Lecture 10

March 15, 2013

*Reading: Chapter 5, Reiter and Dale*

## Pipelined Microplanning

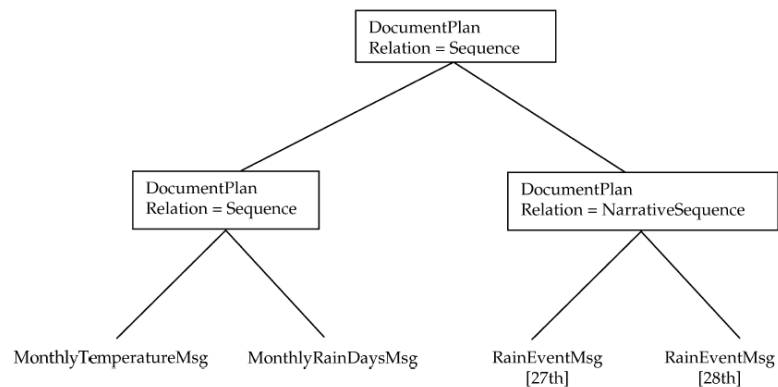


**Lexical choice:** selects words and syntactic structures to express messages. Result is proto-phrase-spec. PPSs may contain refs to domain entities

**Aggregation:** combines multiple PPSs into single PPS

**REG:** takes each PPS and replaces references to domain entities with a phrase spec corresponding to a noun phrase that will uniquely identify that entity to the reader

## Example: Without Aggregation



*The month was slightly warmer than average. It had the average number of rain days. Heavy rain fell on the 27th. Heavy rain fell on the 28th.*

## Aggregation

- Process for determining how messages are distributed across sentences
- Combines multiple proto-phrase specs into a single proto-phrase spec on the basis of
  - information content
  - possible forms of realization
- Basic aggregation mechanisms:
  - Simple conjunction
  - Conjunction via shared participants (ellipsis)
  - Conjunction via shared structure (set introduction)
  - Embedding
- Microplanner must decide
  - When aggregation should be performed
  - What mechanism to use

## Simple conjunction

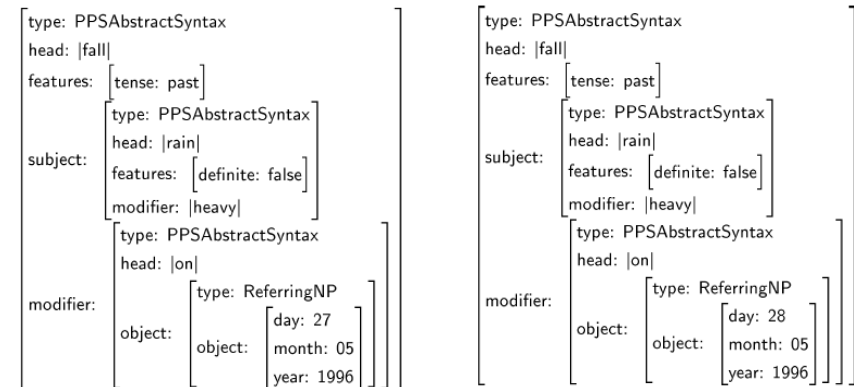
- Combine two or more informational elements with a connective, such as **and**, **but**, ...
- Choice of connective depends on discourse relation that holds between the elements

- Example:

1a) *Heavy rain fell on the 27th.*  
*Heavy rain fell on the 28th.*

1b) *Heavy rain fell on the 27th and heavy rain fell on the 28th.*

## 2 Proto-Phrase Specs

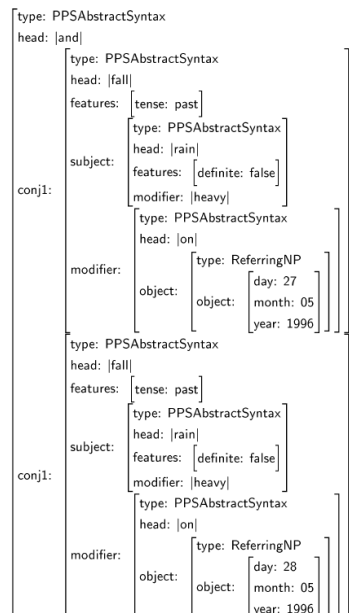


*Heavy rain fell on the 27th.*

*Heavy rain fell on the 28th.*

Proto-phrase spec  
after aggregation by  
simple conjunction

*Heavy rain fell on the 27th  
and heavy rain fell on the  
28th.*



## Conjunction via shared participants

Use when 2 or more informational elements share argument positions with the same content

1a) *The month was colder than average. The month was relatively dry.* [Shared subject]

1b) *The month was colder than average and relatively dry.*

2a) *January was colder than average. February was colder than average.* [Shared object]

2a) *January and February were colder than average.*

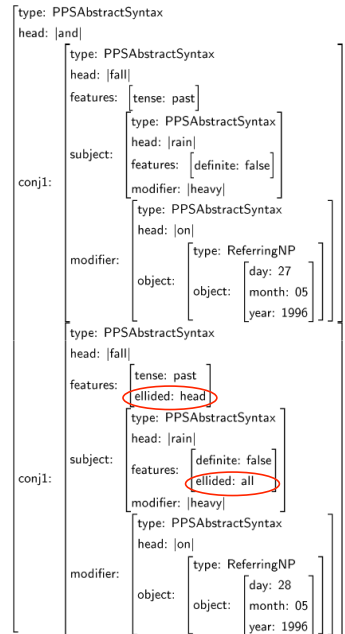
3a) *John gave a book to Mary. John gave a book to Fred.*

3b) *John gave a book to Mary and a book to Fred.*

3c) *John gave books to both Mary and Fred.* [Shared subj and obj]

By elision: Indicate what parts of proto phrase spec should be elided in realization

*Heavy rain fell on the 27th and [] on the 28th.*



## Conjunction via shared structure

Conjunction at other levels of grammatical structure

1a) *Heavy rain fell on the 27th. Heavy rain fell on the 28th.*

1b) *Heavy rain fell on the 27th and 28th.*

2a) *January was drier than average. January was colder than average.*

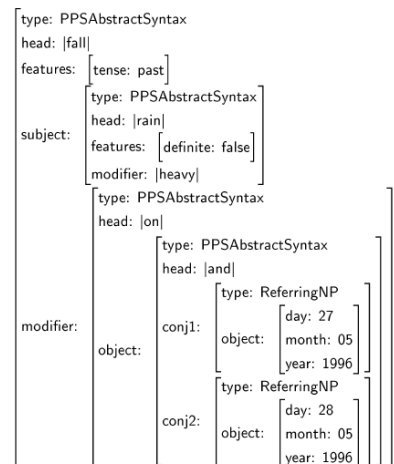
2b) *January was drier and colder than average.*

## Conjunction via shared structure

Put appropriate conjunction inside the proto-phrase spec, instead of top-level

*Heavy rain fell on the [27th and 28th].*

Sometimes called set introduction



## Question

Is microplanning the best way to deal with all apparent cases of aggregation?

An alternative:

- content selection builds a single message specifying all the days on which there was a lot of rain;
- i.e., produce one RainSpellMessage asserting that a lot of rain occurred from 27/5/96 to 28/5/96, instead of two separate RainEventMessages

## Aggregation needs access to meaning!

- 3a) John gave a book to Mary. John gave a book to Fred.
  - 3b) John gave a book to Mary and a book to Fred.
  - 3c) John gave books to both Mary and Fred.
  - 3d) \* John gave a book to Mary and Fred.
- Whether two constituents are considered equivalent for the purposes of aggregation requires knowledge of their underlying meanings
  - Not sufficient to know just the surface lexemes that will be used in the realization

## Aggregation in Weather Reporter: Summary

Without aggregation:

- *Heavy rain fell on the 27th.*
- *Heavy rain fell on the 28th.*

With aggregation via **simple conjunction**:

- *Heavy rain fell on the 27th **and** heavy rain fell on the 28th.*

With aggregation via **ellipsis**:

- *Heavy rain fell on **the 27th and [] on the 28th.***

With aggregation via **set introduction**:

- *Heavy rain fell on **[the 27th and 28th].***

## Aggregation via Syntactic Embedding

**Hypotactic aggregation:** informational element that might have been realized as separate major clause is instead realized as a constituent subordinated to some other realized element

Example: As an unrestricted relative clause

- 1a) *March had 120mm of rainfall. It was the wettest month.*
- 1b) *March, which was the wettest month, had 120mm of rainfall.*

## Choice Heuristics

- Many ways to aggregate a given message set, so how do we choose when to aggregate, and what mechanism to use?
- Many sources of heuristics
  - conform to genre conventions and rules (as gleaned from corpus)
  - consult style manuals
  - observe structural properties of document plan
    - e.g., only aggregate messages that are siblings in the document plan tree
  - take account of pragmatic goals
  - consult psycholinguistic evidence
- But it depends on your application!

## Heuristics for aggregation

Based on psycholinguistic studies of reading comprehension

- **When to aggregate:**
  - When messages are linked by a rhetorical relation that is expressed by an explicit cue word
  - Do not express more than one rhetorical relation within a single sentence
- **How to aggregate:**
  - Embedded clauses should only be used for Elaboration relations

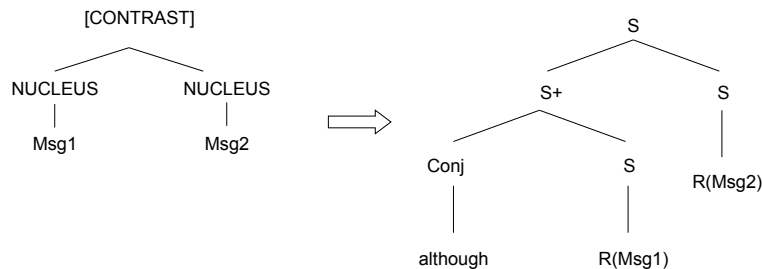
(Scott and de Souza, 1990)

## Aggregation in WeatherReporter

Sensitive to rhetorical relations:

- If two messages are in a **Sequence** relation they can be conjoined at the same level
- If one message is an **Elaboration** of another it can either be conjoined at the same level or embedded as a minor clause or phrase
- If one message is a **Contrast** to another it can be conjoined at the same level or embedded as a minor clause or phrase

## An Aggregation Rule



## Tradeoff: Conciseness vs. syntactic complexity

- MAGIC system produces spoken output in a real-time environment (McKeown et al., 1997)
  - texts must be brief, so they can be spoken quickly
  - sophisticated aggregation techniques used to minimise length of texts
- STOP system generates personalized letters to encourage people to stop smoking (Reiter et al., 2003)
  - letters must be easily readable by wide variety of people, including those with limited education and reading ability
  - use short and simple sentences
  - only simple aggregation is done

## Aggregation in Magic

Domain: Generating spoken healthcare briefings

- 2a) The patient's last name is Jones.  
The patient is female.  
The patient has hypertension.  
The patient has diabetes.  
The patient is 80 years old.  
The patient is undergoing CABG.  
Dr. Smith is the patient's doctor.
- 2b) Ms. Jones is an 80 year old, hypertensive, diabetic, female patient of Dr. Smith undergoing CABG.

*From (McKeown et al, 1997)*

## Pragmatics: STOP Example

- Making the text friendlier by adding more empathy:

*It's clear from your answers that you don't feel too happy about being a smoker and it's excellent that you are going to try to stop.*

- Making the text easier for poor readers:

*It's clear from your answers that you don't feel too happy about being a smoker. It's excellent that you are going to try to stop.*

## Aggregate with care!

- Complex and little-understood semantic constraints on embedding.
- Example:

3a) *John is tall. John is a man.*

3b) *John is a tall man.*

4a) *John is lazy. John is a pianist.*

4b) *John is a lazy pianist.*

## More examples: Syllepsis

*You are free to execute your laws, and your citizens, as you see fit.*

(Star Trek: The Next Generation)

*All my friends are getting Firsts and married.*

(Stephen Fry)

*I live in shame . . . and the suburbs.*

(Uncle Fester in "Addams Family Values", 1993)

*I notice with some relief, that the Icelandic group staying on my floor have either checked or passed out.*

(Agent Cooper: Twin Peaks)

*"Miss Bolo [...] went straight home, in a flood of tears and a sedan-chair."*

(Charles Dickens, The Pickwick Papers, Chapter 35)

## More examples: Syllepsis



*And he said, as he hastened to put out the cat,  
The wine, his cigar and the lamps:  
“Have some Madeira, m'dear...”*

*She lowered her standards by raising her glass,  
Her courage, her eyes and his hopes.*

*When he asked “What in Heaven?” she made no reply,  
Up her mind, and a dash for the door.*

(Michael Flanders “Have Some Madeira M'Dear”)

## References



- K. McKeown et al (1997). “Language generation for multimedia healthcare briefings.” *Proc of the 5th Conference on Applied Natural Language Processing*, pp: 277-282.
- E. Reiter, R. Robertson and L. M. Osman (2003). “Lessons from a failure: generating tailored smoking cessation letters.” *Artificial Intelligence* **144**(1-2): 41-58.
- D. R. Scott and C. Sieckenius de Souza, (1990). “Getting the message across in RST-based text generation.” In *Current Research in Natural Language Generation*, R. Dale, C. Mellish, and M. Zock (Eds.) Academic Press Cognitive Science Series, vol. 4. pp 47-73.