Automated Writing Assistance: Grammar Checking and Beyond
Topic 5: Beyond the Sentence

Robert Dale
Centre for Language Technology
Macquarie University

Outline
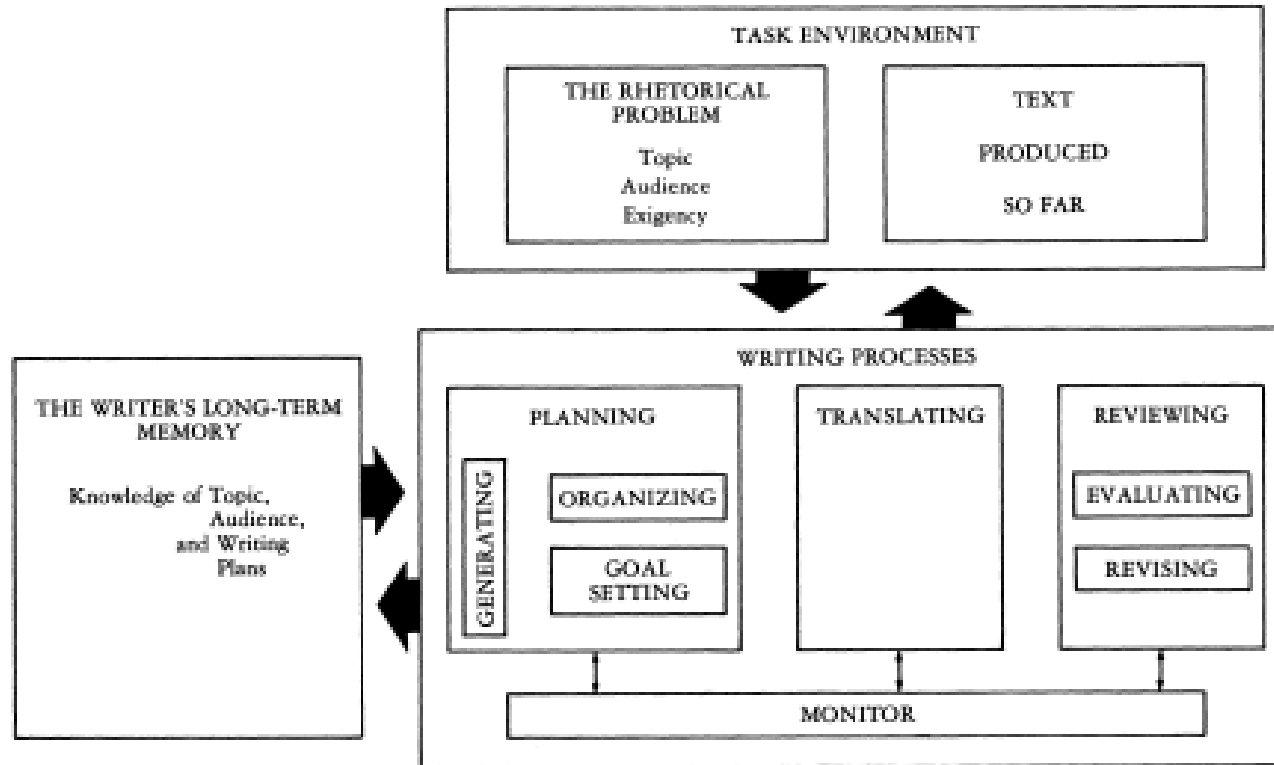
- **Frontiers: Where We Might Want to Go**
- **The View from Natural Language Generation**
- **Closing Remarks**

Frontiers

- **Writing assistance . . .**
  - **Beyond the sentence**
  - **Beyond syntax**
  - **Beyond revision**

Flower and Hayes' Cognitive Process Model

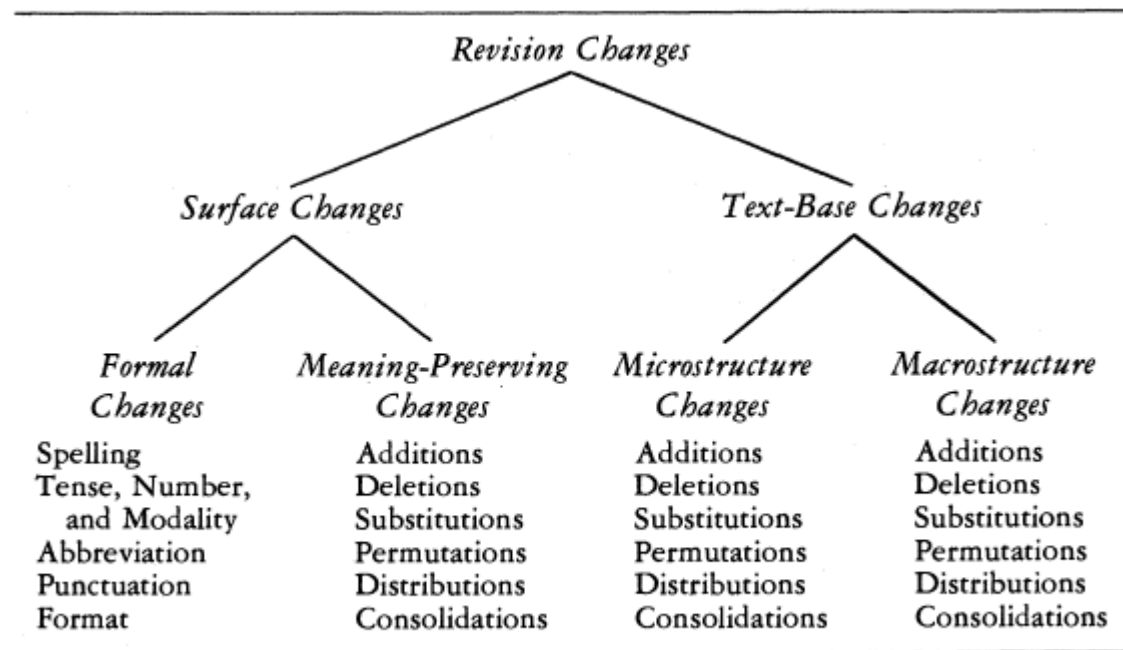**Flower and Hayes 1981**

The Nature of the Revision Process

Figure 1. A Taxonomy of Revision Changes

Faigley and Witte 1981

Outline

- **Frontiers: Where We Might Want to Go**
- **The View from Natural Language Generation**
- **Closing Remarks**
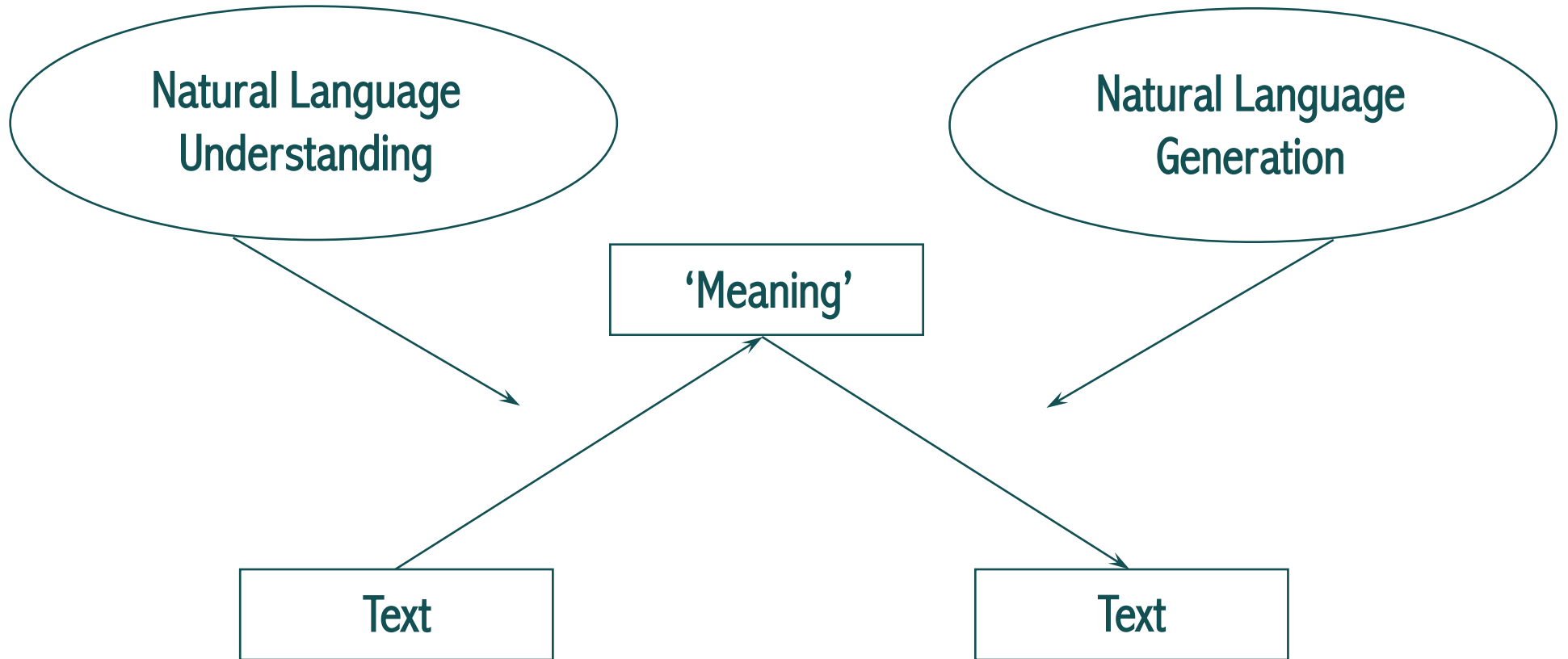
What is NLG?

- Goal:
  - computer software which produces understandable texts in English or other human languages
- Input:
  - some underlying non-linguistic representation of information
- Output:
  - documents, reports, explanations, help messages, and other kinds of texts

NLP = NLU + NLG

Inputs and Outputs

The inputs to NLG:

- A knowledge source

- A communicative goal

- A user model

- A discourse model

The output of NLG:

- A text, possibly embodied as part of a document or within a speech stream

Component Tasks in NLG

1   **Content determination**

2   **Discourse planning**

3   **Sentence aggregation**

4   **Lexicalisation**

5   **Referring expression generation**

6   **Syntactic and morphological realization**

7   **Orthographic realization**

1   Content Determination

- The process of deciding what to say

- Can be viewed as the construction of a set of MESSAGES from the underlying data source

- Messages are aggregations of data that are appropriate for linguistic expression: each may correspond to the meaning of a word or a phrase

- Messages are based on domain entities, concepts, and relations

2   Discourse Planning

- **A text is not just a random collection of sentences**

- **Texts have an underlying structure in which the parts are related together**

- **Two related issues:**
    - **conceptual grouping**
    - **rhetorical relationships**

3   Sentence Aggregation

- A one-to-one mapping from messages to sentences results in disfluent text

- Messages need to be combined to produce larger and more complex sentences

- The result is a sentence specification or SENTENCE PLAN

4   Lexicalisation

- So far we have determined text content and the structuring of the information into paragraphs and sentences, but the raw material is still assumed to be in the form of a conceptual representation

- Lexicalisation determines the particular words to be used to express domain concepts and relations

5   Referring Expression Generation

- Referring expression generation is concerned with how we describe domain entities in such a way that the hearer will know what we are talking about

- Do we use a proper name? A definite or indefinite description? A pronoun?

6   Syntactic and Morphological Realization

- **Every natural language has grammatical rules that govern how words and sentences are constructed**
  - **Morphology: rules of word formation**
  - **Syntax: rules of sentence formation**

7 Orthographic Realization

- **Orthographic realization is concerned with matters like casing and punctuation**

- **This also extends into typographic issues: font size, column width …**

Tasks and Architecture in NLG

- **Content determination**
- **Discourse planning**

Document Planning

- **Sentence aggregation**
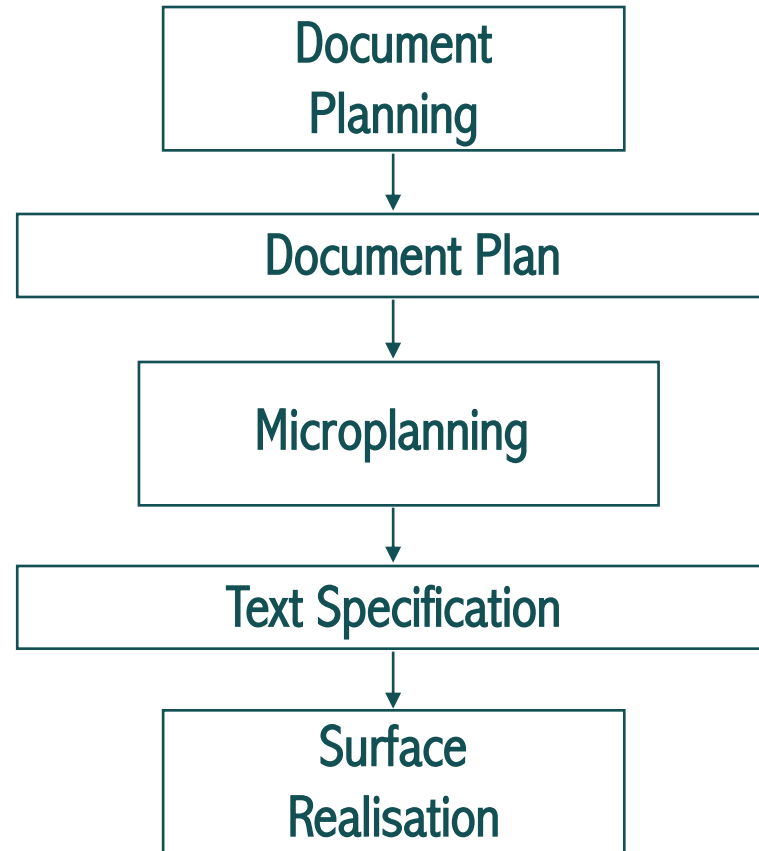- **Lexicalisation**
- **Referring expression generation**

Micro Planning

- **Syntax + morphology**
- **Orthographic realization**

Linguistic Realization

A Pipelined Architecture

```
┌─────────────────────┐
│      Document       │
│      Planning       │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    Document Plan    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    Microplanning    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  Text Specification │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│      Surface        │
│    Realisation      │
└─────────────────────┘
```

Microplanning Help

- **Paraphrase**
- **Sentence simplification via summarisation techniques**

Aggregation

Combinations can be on the basis of

- information content
- possible forms of realisation

Some possibilities:

- Simple conjunction
- Ellipsis
- Embedding
- Set introduction

Some Examples

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].`

An Example: Embedding

Without aggregation:

- `March had a rainfall of 120mm.`
  `It was the wettest month.`

With aggregation:

- `March, which was the wettest month, had a`
  `rainfall of 120mm.`

Rhetorical Structure Theory

- **Basic idea:**

  - **The elements of a text are connected together by rhetorical relations**

  - **A text is coherent by virtue of the presence of these relations---if the text cannot be analysed in these terms then it is not coherent.**
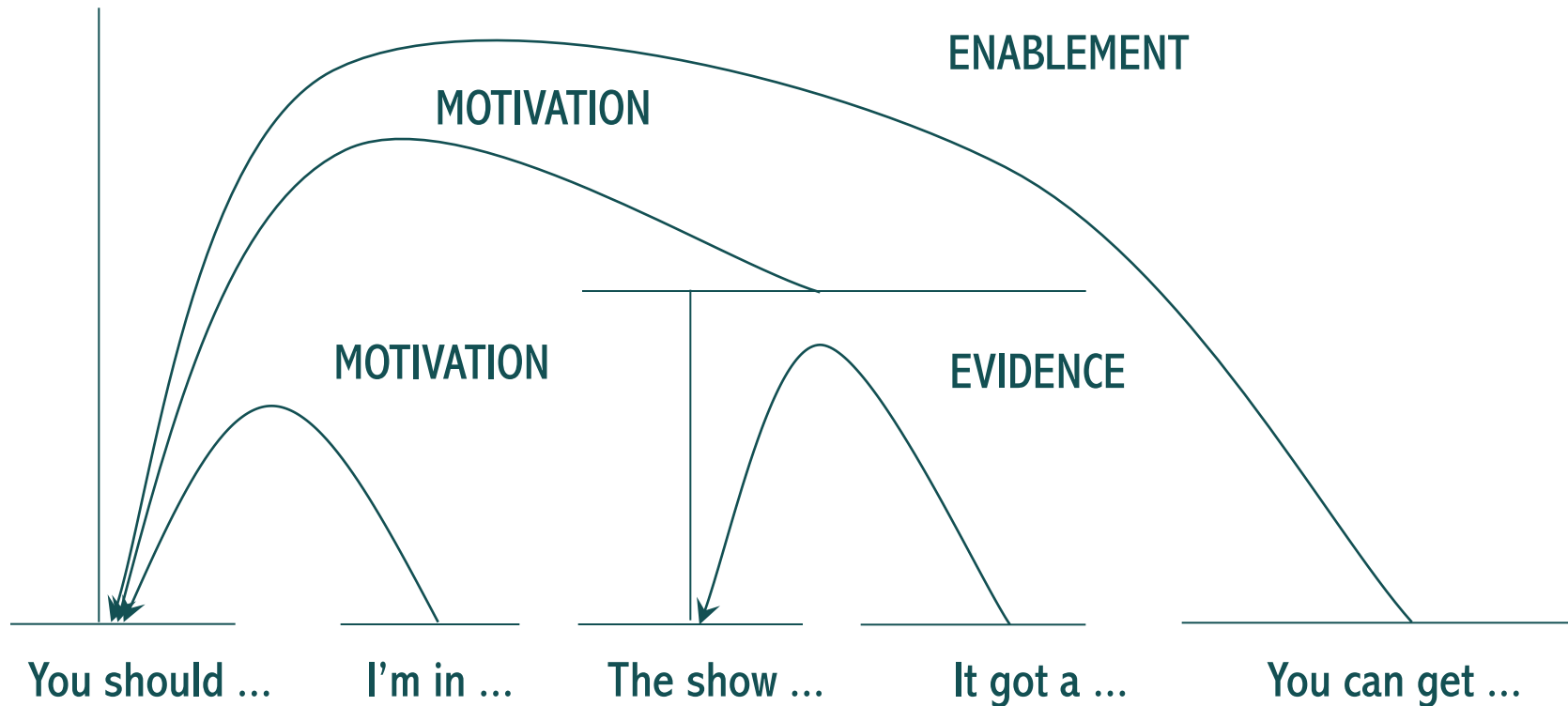
Text Structure

You should come to the Northern Beaches Ballet performance on Saturday. I'm in three pieces. The show is really good. It got a rave review in the Manly Daily. You can get the tickets from the shop next door.

Beyond Pairs of Sentences

S1: You should come to the Northern Beaches Ballet performance on Saturday.

S2: I'm in three pieces.

S3: The show is really good.

S4: It got a rave review in the Manly Daily.

S5: You can get the tickets from the shop next door.

The Ballet Text

An RST Relation Definition

Relation name: Motivation

Constraints on N:

   Presents an action (unrealised) in which the hearer is the actor

Constraints on S:

   Comprehending S increases the hearer's desire to perform the action presented in N

The effect:

   The hearer's desire to perform the action presented in N is increased

Outline

- **Frontiers: Where We Might Want to Go**
- **The View from Natural Language Generation**
- **Closing Remarks**

Conclusions

- Current technology only scratches the surface in terms of the kinds of support we would like to give to authors

- Almost any aspect of NLP technology can be pressed into service to support authors

- NLG techniques provide a rich source of ideas for how to build symbiotic systems that take advantage of the knowledge and capabilities of both human and machine

Who Today's Main Players Are

- Google

- Microsoft

- Educational Testing Service

- Activities around the University of Cambridge

Finding Out More

- ACL Workshops on Innovative Use of NLP for Building Educational Applications: 2011 was the sixth in the series

- Relevant material often found in journals outside the normal 'ACL space':

  CALICO Journal
  College Composition and Communication
  Computers and Composition
  Computer Assisted Language Learning,
  Journal of Second Language Writing