A Case Study in Applied NLG

- Each month an institutional newsletter publishes a summary of the month’s weather
- The summaries are based on automatically collected meteorological data
- The person who writes these summaries will no longer be able to
- The institution wants to continue publishing the reports and so is interested in using NLG techniques to do so

A Weather Summary

MARRIFIELD (Macquarie University No 1)
On Campus, Square F9

TEMPERATURES (°C)
Mean Max for Mth: 18.1 warmer than average
Mean Max for June (20 yrs): 17.2
Highest Max (warmest day): 23.9 on 01
Lowest Max (coldest day): 11.1 on 12
Mean Min for Mth: 8.2 warmer than ave
Mean Min for June (20 yrs): 6.6
Lowest Min (Coldest night): 02.6 on 09
Highest Min (Warmest night): 13.5 on 24
RAINFALL (mm) (24 hrs to 09:00)
Total Rain for Mth: 90.4 on 12 days.
Wettest Day (24h to 09:00): 26.4 on 11
Average for June (25 yrs): 109.0 on 10
Total for 06 mths so far: 54.0 on 72
days.
Very depleted.
Average for 06 mths (25 yrs): 762.0 on 71
days
Annual Average Rainfall (25 yrs): 1142.8 on 131
days

WIND RUN (at 2m height) (km) (24 hrs to 09:00)
Total wind run for Mth: 1660
Windiest Day (24 hrs to 09:00): 185 on 24,
172 on 27
Calmest Day (24 hrs to 09:00): 09 on 16

SUNRISE & SUNSET
Date Sunrise Sunset Difference
01 Jun 06:52 16:54 10:02
11 Jun 06:57 16:53 09:56
21 Jun 07:00 16:54 09:54
30 Jun 07:01 16:57 09:56
(Sunset times began to get later after about June 11)
(Sunrise times continue to get later until early July)
(Soon we can take advantage of the later sunsets)

SUMMARY
The month was warmer than average with average rainfall, but the total rain so far for the year is still very depleted. The month began with mild to warm maximums, and became cooler as the month progressed, with some very cold nights such as June 09 with 02.6. Some other years have had much colder June nights than this, and minimums below zero in June are not very unusual. Very strong winds blew on 23, 24 and 26, 27. Fog occurred on 17, 18 after some rain on 17, heavy rain fell on 11 June.
Output: A Weather Summary

The month was warmer than average with average rainfall, but the total rain so far for the year is still very depleted. The month began with mild to warm maximums, and became cooler as the month progressed, with some very cold nights such as June 09 with 02.6. Some other years have had much colder June nights than this, and minimums below zero in June are not very unusual. The month was mostly calm, but strong winds blew on 23, 24 and 26, 27. Fog occurred on 17, 18 after some rain on 17, heavy rain fell on 11 June.

The Input Data

- A set of 16 data elements collected automatically every 15 minutes: air pressure, temperature, wind speed, rainfall ...
- Preprocessed to construct DailyWeatherRecords:

```
((type dailyweatherrecord)
  (date ((day ...)
         (month ...)
         (year ...)))
  (temperature ((minimum ((unit degrees-centigrade) (number ...)))
                (maximum ((unit degrees-centigrade) (number ...)))))
  (rainfall ((unit millimetres) (number ...))))
```

Other Available Data

- Historical Data:
  - Average temperature and rainfall figures for each month in the Period of Record (1971 to present)
- Historical Averages:
  - Average values for temperature and rainfall for the twelve months of the year over the period of record

Inputs and Outputs

The month was cooler and drier than average, with the average number of rain days, but ...
The Architectural View

Document Planning

Goals:
- to determine what information to communicate
- to determine how to structure this information to make a coherent text

Two Common Approaches:
- methods based on observations about common text structures (schemas)
- methods based on reasoning about discourse coherence and the purpose of the text (entity and relation coherence)

Content Determination

Based on MESSAGES, predefined data structures which:
- correspond to informational elements in the text
- collect together underlying data in ways that are convenient for linguistic expression

Core idea:
- from corpus analysis, identify the largest possible agglomerations of informational elements that do not pre-empt required flexibility in linguistic expression

Content Determination in WeatherReporter

- Routine messages
  - MonthlyRainFallMsg, MonthlyTemperatureMsg, RainSoFarMsg, MonthlyRainyDaysMsg
- Always constructed for any summary to be generated
Content Determination in WeatherReporter

A MonthlyRainfallMsg:

((message-id msg091)
 (message-type monthlyrainfall)
 (period ((month 04)
   (year 1996)))
 (absolute-or-relative relative-to-average)
 (relative-difference ((magnitude ((unit millimeters)
   (number 4)))
   (direction +))))

Content Determination in WeatherReporter

- Significant Event messages
  - RainEventMsg,
  - RainSpellMsg,
  - TemperatureEventMsg,
  - TemperatureSpellMsg
- Only constructed if the data warrants their construction: e.g., if rain occurs on more than a specified number of days in a row

Content Determination in WeatherReporter

A RainSpellMsg:

((message-id msg096)
 (message-type rainspellmsg)
 (period ((begin ((day 04)
   (month 02)
   (year 1995)))
   (end ((day 11)
     (month 02)
     (year 1995)))
   (duration ((unit day)
     (number 8))))
 (amount ((unit millimetres)
   (number 120))))

Alternative strategies:
- Build all possible messages from the underlying data, then select for expression those appropriate to the context of generation
- Identify information required for context of generation and construct appropriate messages from the underlying data

The content determination task is essentially a domain-dependent expert-systems-like task
Document Structuring via Schemas

Basic idea (after McKeown 1985):
- texts often follow conventionalised patterns
- these patterns can be captured by means of ‘text grammars’ that both dictate content and ensure coherent structure
- the patterns specify how a particular document plan can be constructed using smaller schemas or atomic messages
- can specify many degrees of variability and optionality

Implementing schemas:
- simple schemas can be expressed as grammars
- more flexible schemas usually implemented as macros or class libraries on top of a conventional programming language, where each schema is a procedure
- currently the most popular document planning approach in applied NLG systems

Deriving Schemas From a Corpus

Using the Target Text Corpus:
- take a small number of similar corpus texts
- identify the messages, and try to determine how each message can be computed from the input data
- propose rules or structures that explain why message x is in text A but not in text B — this may be easier if messages are organised into a taxonomy
- discuss this analysis with domain experts, users, and other stakeholders, and iterate
- repeat the exercise with a larger set of corpus texts

Document Planning in WeatherReporter

A Simple Schema:

WeatherSummary →
  MonthlyTempMsg
  MonthlyRainfallMsg
  RainyDaysMsg
  RainSoFarMsg
Document Planning in WeatherReporter

A More Complex Set of Schemata:

WeatherSummary →
   TemperatureInformation RainfallInformation
TemperatureInformation →
   MonthlyTempMsg [ExtremeTempInfo] [TempSpellsInfo]
RainfallInformation →
   MonthlyRainfallMsg [RainyDaysInfo] [RainSpellsInfo]
RainyDaysInfo →
   RainyDaysMsg [RainSoFarMsg]

Schemas in Practice

Tests and other machinery are often made explicit:

(put-template maxwert-grenzwert "MV01"
 (:PRECOND (:CAT DECL-E
   :TEST ((pred-eq 'maxwert-grenzwert)
          (not (status-eq (theme) 'no))))
   :ACTIONS (:TEMPLATE (:RULE MAX-AVG-VALUE-E (self))
               "As a result,"
               (:RULE EXCEEDS-THRESHOLD-E (self))
               ".")
   )

Schemas: Pros and Cons

Advantages of schemas:
- Computationally efficient
- Allow arbitrary computation when necessary
- Naturally support genre conventions
- Relatively easy to acquire from a corpus

Disadvantages
- Limited flexibility: require predetermination of possible structures
- Limited portability: likely to be domain-specific

Document Structuring via Explicit Reasoning

Observation:
- Texts are coherent by virtue of relationships that hold between their parts — relationships like narrative sequence, elaboration, justification ...

Resulting Approach:
- segment knowledge of what makes a text coherent into separate rules
- use these rules to dynamically compose texts from constituent elements by reasoning about the role of these elements in the overall text
**Document Structuring via Explicit Reasoning**

- Typically adopt AI planning techniques:
  - Goal = desired communicative effect
  - Plan constituents = messages or structures that combine messages (subplans)
- Can involve explicit reasoning about the user’s beliefs
- Often based on ideas from Rhetorical Structure Theory

**Rhetorical Structure Theory**

D1: You should come to the Northern Beaches Ballet performance on Saturday.
D2: I’m in three pieces.
D3: The show is really good.
D4: It got a rave review in the Manly Daily.
D5: You can get the tickets from the shop next door.

**An RST Relation Definition: Motivation**

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
Example: Evidence

Relation name: Evidence
Constraints on Nucleus:
  The reader may not believe N to a degree satisfactory to the writer
Constraints on Satellite:
  The reader believes S or will find it credible
Constraints on combination of N+S
  The reader’s comprehending S increases their belief of N
Effect (the intention of the writer)
  The reader’s belief of N is increased

Document Structuring in WeatherReporter

Three basic rhetorical relationships:
- SEQUENCE
- ELABORATION
- CONTRAST

Applicability of rhetorically based planning operators determined by attributes of the messages

Message Attributes

status=primary
  MonthlyTempMsg
  MonthlyRainMsg
  RainyDaysMsg
  RainSoFarMsg
  RainAmountsMsg
status=secondary
  TempEventMsg
  TempSpellMsg
  RainSpellMsg
significance=routine

Document Structuring in WeatherReporter

SEQUENCE
  Two messages can be connected by a SEQUENCE relationship if both have the attribute message-status = primary
ELABORATION
Two messages can be connected by an ELABORATION relationship if:
- they both have the same message-topic
- the nucleus has message-status = primary

CONTRAST
Two messages can be connected by a CONTRAST relationship if:
- they both have the same message-topic
- they both have the feature absolute-or-relative = relative-to-average
- they have different values for relative-difference:direction

Select a start message
Use rhetorical relation operators to add messages to this structure until all messages are consumed or no more operators apply
Start message is any message with message-significance = routine

The algorithm:
DocumentPlan = StartMessage
MessageSet = MessageSet - StartMessage
repeat
- find a rhetorical operator that will allow attachment of a message to the DocumentPlan
- attach message and remove from MessageSet
until MessageSet = {} or no operators apply
Target Text #1

The month was cooler and drier than average, with the average number of rain days, but the total rain for the year so far is well below average. Although there was rain on every day for 8 days from 11th to 18th, rainfall amounts were mostly small.

Document Structuring in WeatherReporter

The Message Set:

- MonthlyTempMsg ("cooler than average")
- MonthlyRainfallMsg ("drier than average")
- RainyDaysMsg ("average number of rain days")
- RainSoFarMsg ("well below average")
- RainSpellMsg ("8 days from 11th to 18th")
- RainAmountsMsg ("amounts mostly small")

More Complex Algorithms

Adding complexity, following [Marcu 1997]:

- Assume that multiple DocumentPlans can be created from a set of messages and relations
- Assume that a desirability score can be assigned to each DocumentPlan
- Determine the best DocumentPlan
One content plan, many text plans

- The month was cooler and drier than average, with the average number of rain days, but the total rain for the year so far is well below average. Although there was rain on every day for 8 days from 11th to 18th, rainfall amounts were mostly small.

- The month was cooler than average. The month was drier than average. There was the average number of rainy days. The total rain for the year so far is well below average. There was rain on every day for 8 days from 11th to 18th. Rainfall amounts were mostly small.

- The month was cooler and drier than average. The total rain for the year so far is well below average, even though there was an average number of rainy days this month. There was rain on every day for 8 days from 11th to 18th, but rainfall amounts were mostly small.

The Document Plan

- Result is a DOCUMENT PLAN: a tree structure populated by messages at its leaf nodes
- Next step: realising the messages as text

A Simple Realiser

- We can produce one output sentence per message in the document plan
- A specialist fragment of code for each message type determines how that message type is realised

Example: For the MonthlyTemperatureMsg:

TempString = case (TEMP - AVERAGE_TEMP)
  [2.0 ... 2.9]: 'very much warmer than average.'
  [1.0 ... 1.9]: 'much warmer than average.'
  [0.1 ... 0.9]: 'slightly warmer than average.'
  [-0.1 ... -0.9]: 'slightly cooler than average.'
  [-1.0 ... -1.9]: 'much cooler than average.'
  [-2.0 ... -2.9]: 'very much cooler than average.'
endcase
Sentence = 'The month was' + TempString
Research Issues

- The use of expert system techniques in content determination -- for example, case based reasoning
- Principled ways of integrating schemas and relation-based approaches to document structuring
- A better understanding of rhetorical relations
- Knowledge acquisition -- e.g., methodologies for creating content rules, schemas, and relation applicability conditions for a particular application