

---

# Intro to Natural Language Generation

*Ehud Reiter*

*(Abdn Uni and Arria/Data2text)*

Background read: Reiter and Dale, *Building  
Natural Language Generation Systems*

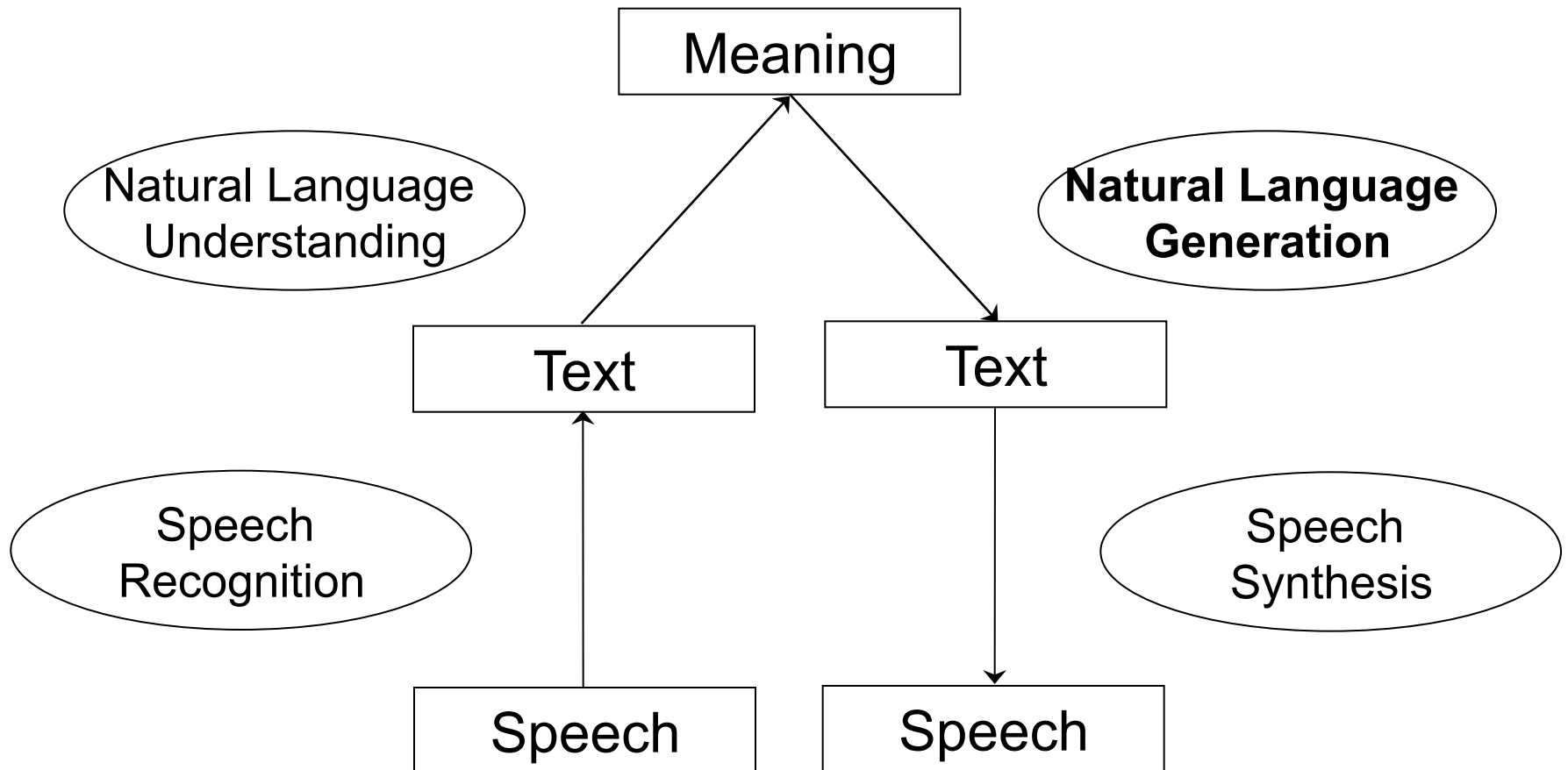
# What is NLG?

---

- NLG systems are computer systems which produces understandable and appropriate texts in English or other human languages
  - » Input is data (raw, analysed)
  - » Output is documents, reports, explanations, help messages, and other kinds of texts
- Requires
  - » Knowledge of language
  - » Knowledge of the domain

# Language Technology

---

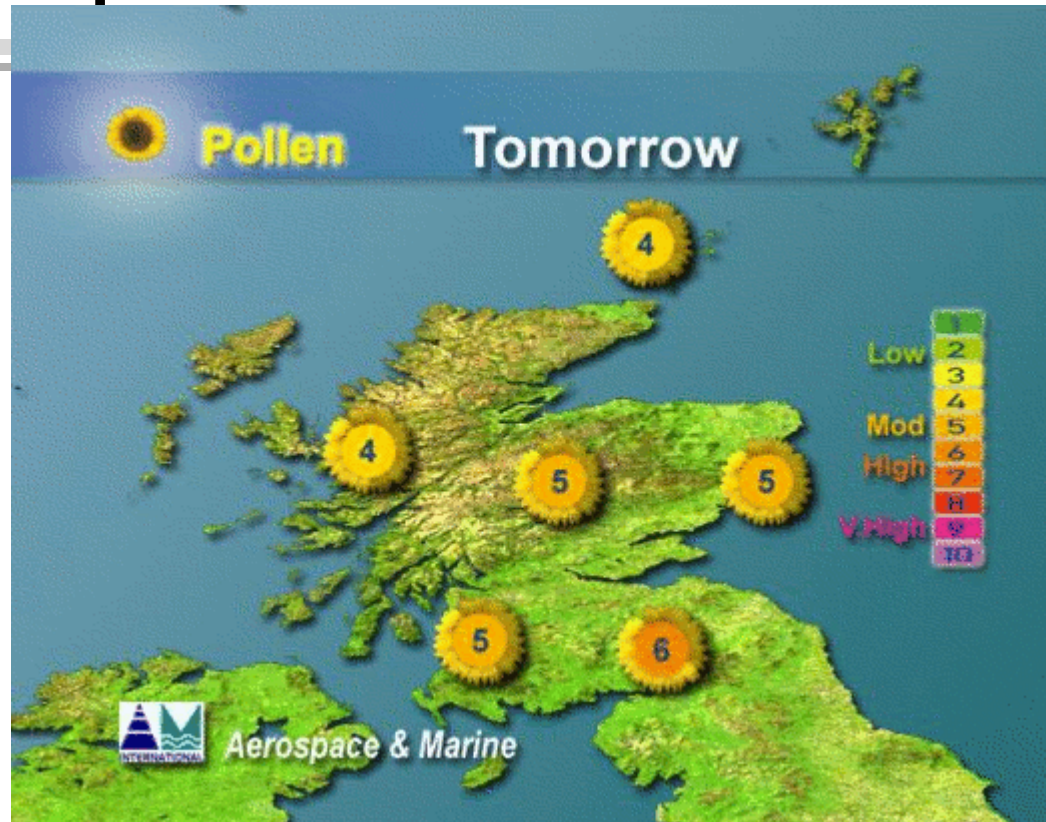


# First Example: Weather Forecasts

---

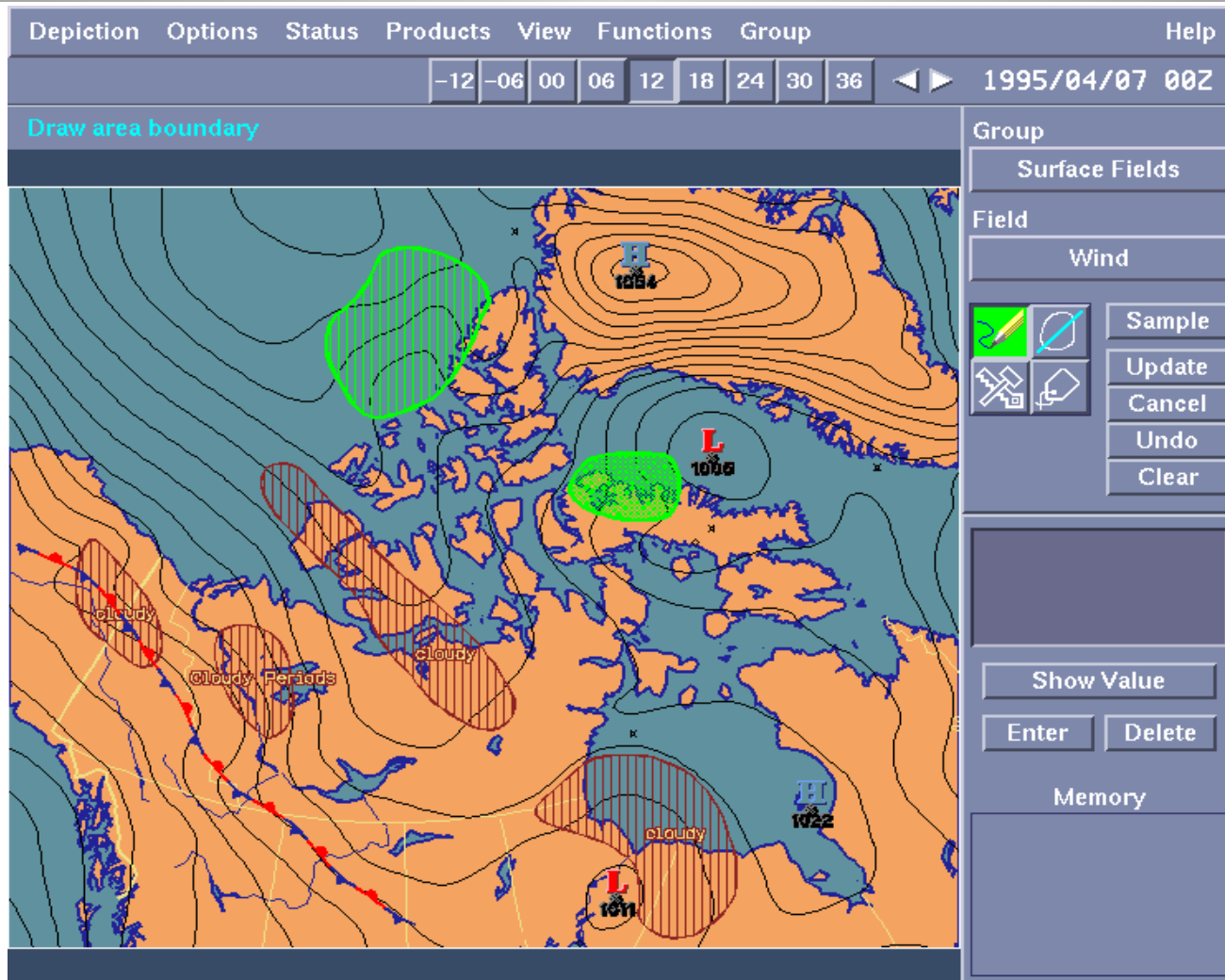
- Input: numerical weather predictions
  - » From supercomputer running a numerical weather simulation
- Output: textual weather forecast
  - » Users prefer some gen texts to human texts!
    - More consistent, better word choice
- <http://www.metoffice.gov.uk/public/weather/forecast-data2text>

# Simple ex: Pollen forecasts



- Grass pollen levels for Tuesday have decreased from the high levels of yesterday with values of around 4 to 5 across most parts of the country. However, in South Eastern areas, pollen levels will be high with values of 6.

# Medium ex: marine forecasts



# FoG: Output

**FPCN20 Status: CURRENT-NOT RELEASED**

FPCN20 CHEG 152300  
MARINE FORECASTS FOR ARCTIC WATERS ISSUED BY THE ARCTIC WEATHER CENTRE OF ENVIRONMENT CANADA AT 05.00 PM MDT SATURDAY 15 APRIL 1995 FOR TONIGHT AND SUNDAY WITH AN OUTLOOK FOR MONDAY.  
THE NEXT SCHEDULED FORECAST WILL BE ISSUED AT 05.00 AM MDT.  
WINDS ARE IN KNOTS.  
FOG IMPLIES VISIBILITY LESS THAN 5/8 NM.  
MIST IMPLIES VISIBILITY 5/8 TO 6 NM.

GREAT SLAVE LAKE.  
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING NEAR MIDNIGHT. VISIBILITIES NEAR 2 NM IN SNOW.  
OUTLOOK FOR MONDAY... LIGHT WINDS.

GREAT BEAR LAKE.  
FREEZING SPRAY WARNING ISSUED.  
WINDS EAST 20 TO 25 TONIGHT AND SUNDAY. FREEZING SPRAY.  
OUTLOOK FOR MONDAY... WINDS EASTERLY 20 TO 25.

MACKENZIE RIVER FROM MILE 0 TO MILE 100.  
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING THIS EVENING. VISIBILITIES NEAR 2 NM IN SNOW.  
OUTLOOK FOR MONDAY... LIGHT WINDS.

MACKENZIE RIVER FROM MILE 100 TO MILE 300.  
WINDS LIGHT STRENGTHENING TO SOUTHEAST 15 SUNDAY AFTERNOON. SNOW ENDING EARLY THIS EVENING. VISIBILITIES NEAR 2 NM IN SNOW.  
OUTLOOK FOR MONDAY... WINDS SOUTHEASTERLY 15.

**Forecasts**

- Marine--
- \* ARWC \*\*
- FPCN20**
- FPCN21
- FPCN22/74
- FPCN23/75
- FPCN24/76
- FPCN25/77
- UL 22/83
- Public--
- FPCN15

Set Element Priority ...

Set Active Areas ...

**Source**

- Working Version
- Official Release
- Forecast Rollup

**Language**

- English
- French

Generate Update Edit ... Release Print Close Help

# Complex example: road maintenance

---

- Forecasts for gritting and other winter road maintenance procedures
- Input is 15 parameters over space and time
  - » Temperature, wind speed, rain, etc
  - » Over thousands of points on a grid
  - » Over 24 hours (20-min interval)



# Points



# Generated Text

---

**Overview** Road surface temperatures will reach marginal levels on most routes from this evening until tomorrow morning.

**Wind (mph)** NW 10-20 gusts 30-35 for a time during the afternoon and evening in some southwestern places, veering NNW then backing NW and easing 5-10 tomorrow morning.

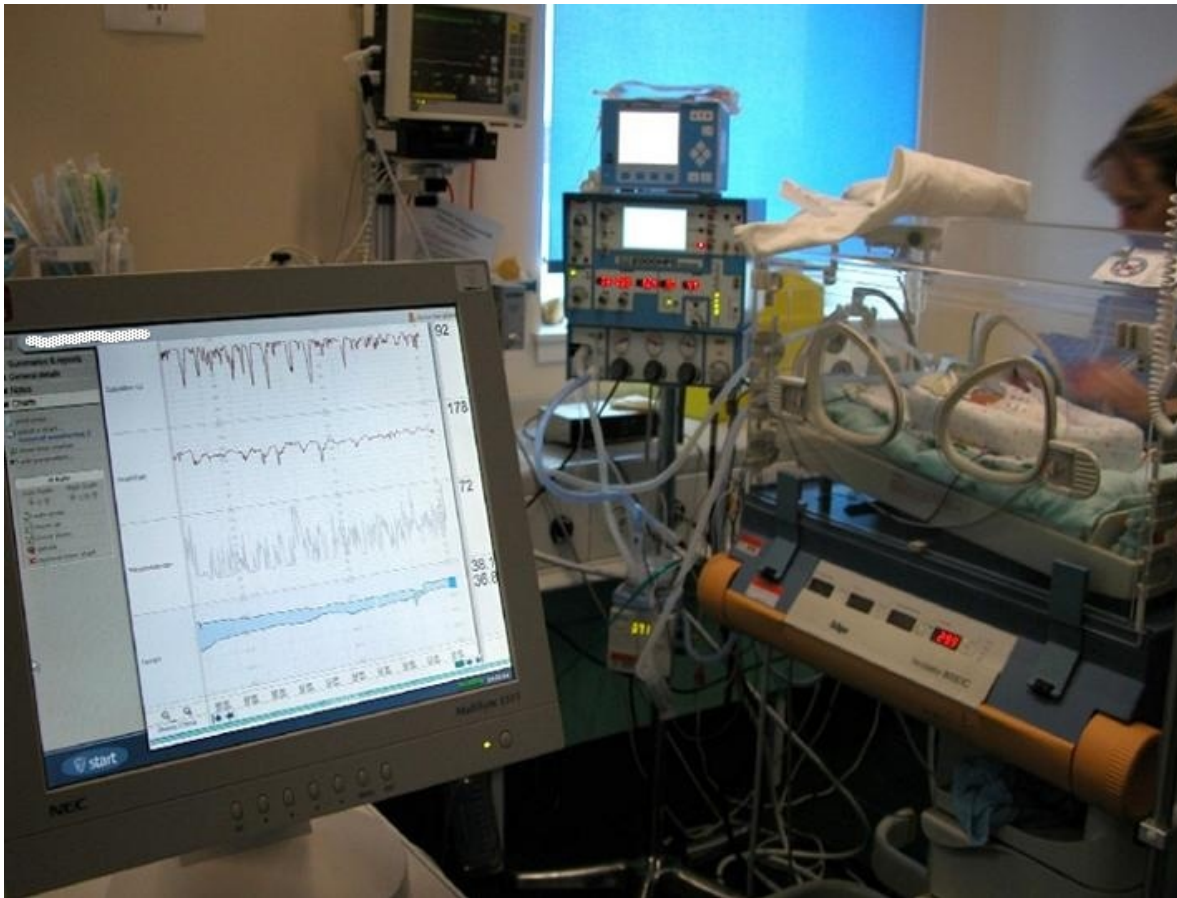
**Weather** Light rain will affect all routes this afternoon, clearing by 17:00. Fog will affect some central and southern routes after midnight until early morning and light rain will return to all routes. Road surface temperatures will fall slowly during this afternoon until tonight, reaching marginal levels in some places above 200M by 17:00.

# Example 2: BabyTalk

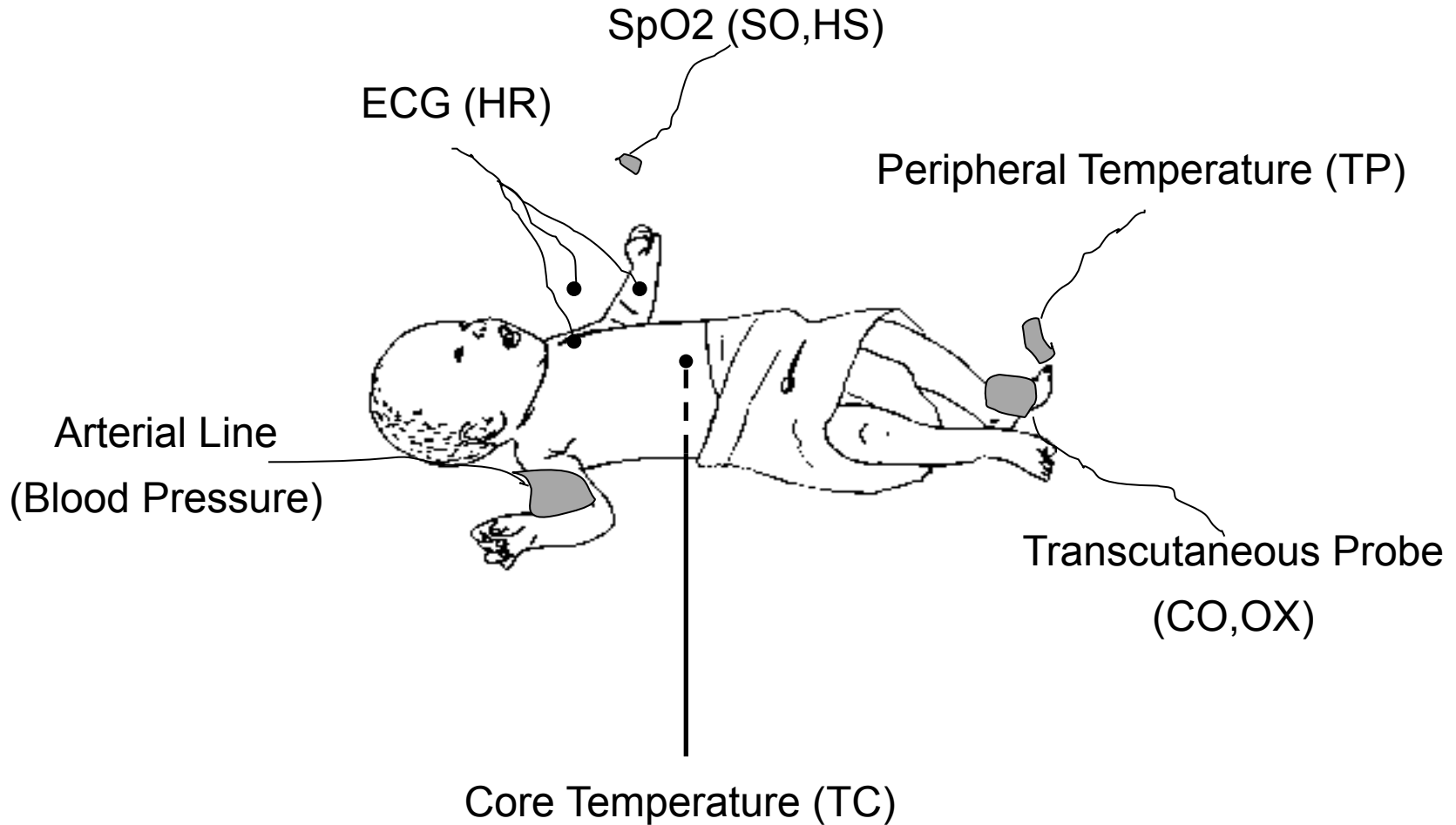
---

- Goal: Summarise clinical data about premature babies in neonatal ICU
- Input: sensor data; records of actions/ observations by medical staff
- Output: multi-para texts, summarise
  - » BT45: 45 mins data, for doctors
  - » BT-Nurse: 12 hrs data, for nurses
  - » BT-Family: 24 hrs data, for parents

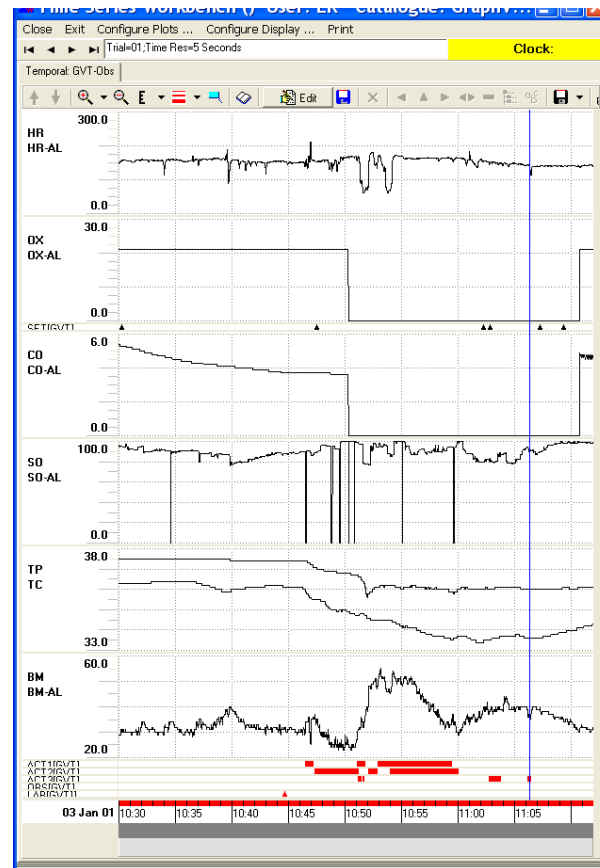
# Neonatal ICU



# Baby Monitoring



# Input: Sensor Data



# Input: Action Records

FullDescriptor	Time
SETTING;VENTILATOR;FiO2 (36%)	10.30
MEDICATION;Morphine	10.44
ACTION;CARE;TURN/CHANGE POSITION;SUPINE	10.46-10.47
ACTION;RESPIRATION;HAND- BAG BABY	10.47-10.51
SETTING;VENTILATOR;FiO2 (60%)	10.47
ACTION;RESPIRATION;INTUBATE	10.51-10.52

# BT45 texts (extract)

---

## *Computer-generated text*

- By 11:00 the baby had been hand-bagged a number of times causing 2 successive bradycardias. She was successfully re-intubated after 2 attempts. The baby was sucked out twice. At 11:02 FIO<sub>2</sub> was raised to 79%.

## *Human corpus text*

- At 1046 the baby is turned for re-intubation and re-intubation is complete by 1100 the baby being bagged with 60% oxygen between tubes. During the re-intubation there have been some significant bradycardias down to 60/min, but the sats have remained OK. The mean BP has varied between 23 and 56, but has now settled at 30. The central temperature has fallen to 36.1°C and the peripheral temperature to 33.7°C. The baby has needed up to 80% oxygen to keep the sats up.



# BT-Nurse text (extract)

---

## **Respiratory Support**

### **Current Status**

Currently, the baby is on CMV in 27 % O<sub>2</sub>. Vent RR is 55 breaths per minute. Pressures are 20/4 cms H<sub>2</sub>O. Tidal volume is 1.5.

SaO<sub>2</sub> is variable within the acceptable range and there have been some desaturations.

...

### **Events During the Shift**

A blood gas was taken at around 19:45. Parameters were acceptable. pH was 7.18. CO<sub>2</sub> was 7.71 kPa. BE was -4.8 mmol/L.

...

# BT-Family text (extract)

---

John was in intensive care. He was stable during the day and night. Since last week, his weight increased from 860 grams (1 lb 14 oz) to 1113 grams (2 lb 7 oz). He was nursed in an incubator.

Yesterday, John was on a ventilator. The mode of ventilation is Bilevel Positive Airway Pressure (BiPAP) Ventilation. This machine helps to provide the support that enables him to breathe more comfortably. Since last week, his inspired Oxygen ( $FiO_2$ ) was lowered from 56 % to 21 % (which is the same as normal air). This is a positive development for your child.

During the day, Nurse Johnson looked after your baby. Nurse Stevens cared for your baby during the night.

# Other NLG projects

---

- Blogging birds: generate “blogs” from red kites based on location data
- Standup: help children with learning disabilities tell jokes
- Skillsum: give adults feedback on literacy/ numeracy assessment
- Thomson-Reuters: Automatically generate newswire articles
- Etc, etc

# How do NLG Systems Work?

---

- Usually three stages
  - » Not including data analysis
- *Document planning (content determination)*: decide on content and structure of text
- *Microplanning*: decide how to linguistically express text (which words, sentences, etc to use)
- *Realisation*: actually produce text, conforming to rules of grammar

# NLG as choice-making

---

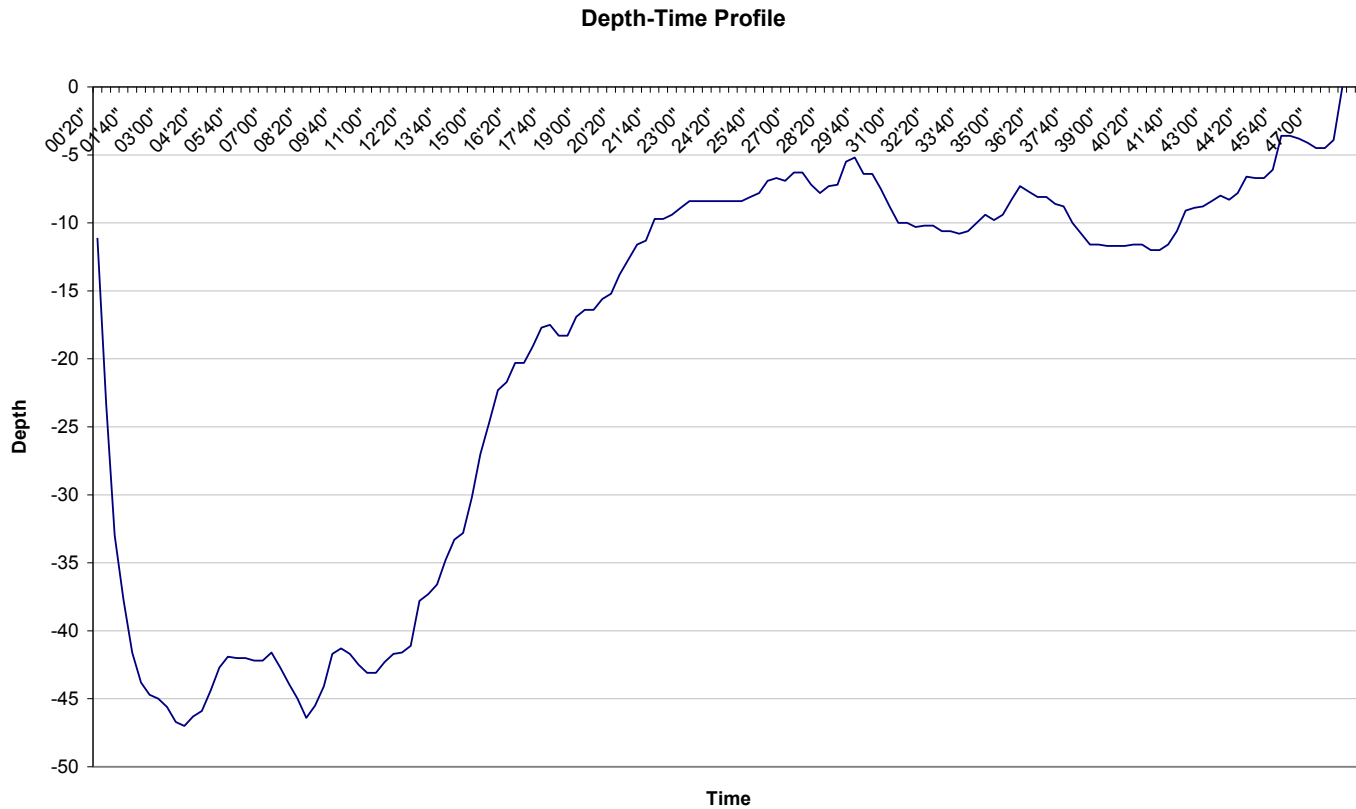
- Need to make *choices* about the generated text
  - » Content
  - » structure
  - » Packaging information into sentences
  - » Words
  - » Syntax
  - » etc

# Scubatext example

---

- Demo system (Dr Sripada) for scuba divers
- Input is *dive computer data*
  - » Depth-time profile of scuba dive
- Output is feedback to diver
  - » Mistakes, what to do better next time
  - » Encouragement of things done well

# Scuba - input



# Scuba – output

---

- Risky dive with some minor problems. Because your bottom time of 12 min exceeds no-stop limit by 4 min this dive is risky. But you performed the ascent well. Your buoyancy control in the bottom zone was poor as indicated by ‘saw tooth’ patterns.



# Scuba: data analytics

---

- Look for trends and patterns in data
  - » Trends: eg, depth increases fairly steadily over first 3 minutes
  - » Patterns: eg, sawtooth between 3 and 15 minutes
- Will not further discuss here

# Document Planning

---

- *Content selection*: Of the zillions of things I could say, which should I say?
  - » Depends on what is important
  - » Also depends on what is easy to say
- *Structure*: How should I organise this content as a text?
  - » What order do I say things in?
  - » Rhetorical structure?

# Scuba: content

---

- Probably focus on patterns indicating dangerous activities
  - » Most important thing to mention
- How much should we say about these?
  - » Detail? Explanations?
- Should we say anything for safe dives?
  - » Maybe just acknowledge them?
  - » But encouragement also important

# Scuba: structure

---

- Mention most dangerous thing first?
  - » Or should we just order by time?
  - » Start with overview?
- Linking words (cue phrases)
  - » Also, but, because, ...

# Document planning

---

- Content-determination is very domain dependent
  - » Based on knowledge about what is important to mention in text
- Structure is also genre-dependent
  - » Conform to existing conventions

# Microplanning

---

- *Lexical/syntactic choice*: Which words and linguistic structures to use?
- *Aggregation*: How should information be distributed across sentences and paras
- *Reference*: How should the text refer to objects and entities?

# SCUBA: microplanning

---

- Lexical/syntactic choice:
  - » *Risky vs dangerous vs unwise vs ...*
  - » *Performed the ascent vs ascended vs ...*
  - » *12 min vs 720 sec vs 700 sec vs 714.56 sec*
- Aggregation: 1 sentence or 2 sent?
  - » “Because your bottom time of 12 min exceeds no-stop limit by 4 min this dive is risky, but you performed the ascent well.”

# Scuba: Microplanning

---

- Aggregation (continued)
  - » Phrase merging
    - “Your first ascent was fine. Your second ascent was fine” vs
    - “Your first and second ascents were fine.”
  - » Reference
    - Your ascent vs
    - Your first ascent vs
    - Your ascent from 33m at 3 min



# Realisation

---

- Grammars (linguistic): Form legal English sentences based on decisions made in previous stages
  - » Obey sublanguage, genre constraints
- Structure: Form legal HTML, RTF, or whatever output format is desired

# Scuba: Realisation

---

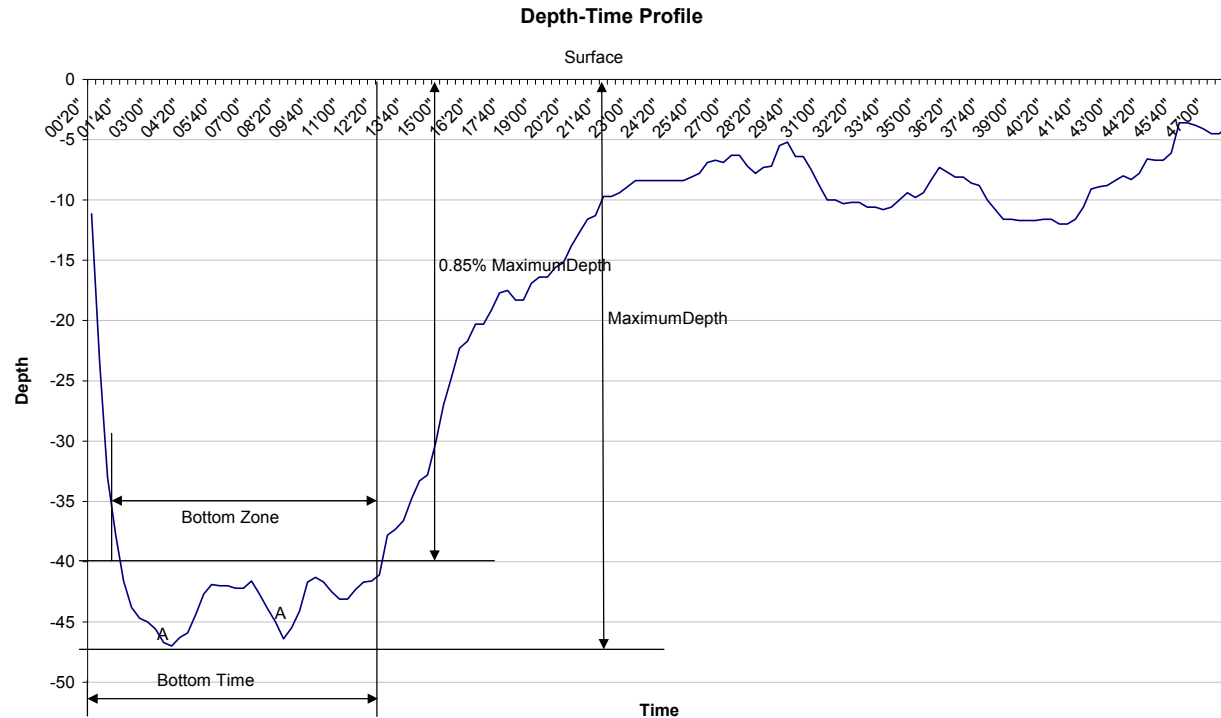
- Simple linguistic processing
  - » Capitalise first word of sentence
  - » Subject-verb agreement
    - Your first ascent was fine
    - Your first and second ascents were fine
- Structure
  - » Inserting line breaks in text (pouring)
  - » Add HTML markups, eg, <P>

# Multimodal NLG

---

- Speech output
- Text and visualisations
  - » Produce separately, OR
  - » Tight integration
    - Eg, text refers to graphic, OR
    - graphs has text annotations

# Combined (Preferred)



Risky dive with some minor problems. Because your bottom time of 12.0min exceeds no-stop limit by 4.0min this dive is risky. But you performed the ascent well. Your buoyancy control in the bottom zone was poor as indicated by 'saw tooth' patterns marked 'A' on the depth-time profile.

# Building NLG Systems

---

- Knowledge and corpus analysis
- Statistical/learning techniques
- Systems

# Building NLG Systems: Knowledge

---

- Need knowledge
  - » Which patterns most important?
  - » What order to use?
  - » Which words to use?
  - » When to merge phrases?
  - » How to form plurals
  - » Etc
- Where does this come from?

# Knowledge Sources

---

- Imitate a *corpus* of human-written texts
  - » Most straightforward, will focus on
- Ask domain experts
  - » Useful, but experts often not very good at explaining what they are doing
- Experiments with users
  - » Very nice in principle, but a lot of work

# Scuba: Corpus

---

- See which patterns humans mention in the corpus, and have the system mention these
- See the words used by humans, and have the system use these as well
- etc



# Systems

---

- Ideally should be able to plug knowledge into NLG framework
  - » Unfortunately good NLG frameworks not available publicly to students and researchers

# Statistical techniques

---

- Learn knowledge from corpus
  - » Just text (language)
    - Zillions of these around
  - » Parallel data-text corpora
    - Input data and corresponding target text
    - Many created for specific projects
    - Only a handful used more generally
      - SumTime, Tuna (Aberdeen)

# Learning from Text Corpora

---

- Specific choices

- » “a” vs “an”

- Bigram freq: “a university” vs “an university”

- » Adj order (“big red ball” vs “red big ball”)

- Need semantic category, eg <colour>

- Global choice

- » generate candidate texts

- » use language model to choose one of these

# Learn from parallel corpora

---

- Specific choices
  - » Choosing words to express data
    - What time does “by evening” mean?
  - » Choosing content
    - Should Babytalk text mention morphine?
- Global choice
  - » Case-based reasoning

# Statistical NLG

---

- Statistical techniques very successful in other areas of NLP
- Still not clear how they can be most effectively used in NLG
- Better resources would help
  - » Especially parallel data-text corpora

# Evaluating NLG Systems

---

## Type

- Metric (eg, BLEU)
  - Human ratings
  - Task performance
- 
- Controlled vs real world?

# Example: BT45 Evaluation

---

- Controlled evaluation based on task perform
- Showed 35 medical professionals 24 scenarios in 3 conditions (8 of each)
  - » Visualisation of medical data
  - » Textual summary (manually written)
  - » Textual summary (from BT45)
- Asked to make a treatment decision
  - » Limited to 3 minutes
  - » Measured correctness (against gold stan)
- Off-ward, using historical data
  - » So no other knowledge about baby

# Results

---

- No sig difference in time taken
- Avg decision-quality (scale -1 to 1)
  - » Human texts: 0.39
  - » Computer texts: 0.34
  - » Visualisation: 0.33
- Human texts especially good for junior nurses (ie, least experienced subjects)
- Computer texts good in some scenarios, poor in others



# Example: BT-Nurse eval

---

- Real-world eval based on human rating
- Deployed BT-Nurse on-ward
  - » Running on cot-side system, using live data from babies in ward
- Asked nurses to read BT-Nurse texts
  - » For babies they were looking after
  - » Questionnaire: understandable, accurate, helpful
  - » Free text comments

# Results

---

- 165 trials
  - » 90% nurses said understandable
  - » 70% said accurate
  - » 60% said helpful
- Free-text comments
  - » More information wanted
  - » Many software bugs
  - » Only a few comments about language

# Evaluation

---

- No consensus about best technique
  - » Lots of people (including me) distrust evaluations based on metrics
- Active area of research

# Commercial NLG

---

- Arria/Datatext: U Abdn spinout company
  - » Monitoring equipment on oil platforms
  - » weather forecasts
  - » Agricultural information
  - » Financial summaries

# Others

---

- Narrative Science - Builds bespoke “automatic narrative generation” systems
  - » Academic roots in computational creativity
- Automated Insights - writes “insightful, personalized reports from your data”
  - » Non-academic roots
- Yseop - “Smart NLG” software that “writes like a human”
  - » Chief scientist, Alain Kaeser did NLG in 1980s

# Others

---

- Lots of small young startups, I lose track of them
  - » OnlyBoth “Discovers New Insights from Data. Writes Them Up in Perfect English. All Automated”
  - » InfoSentience “Developers of the Most Advanced Automated Narrative Generation Software”
  - » Text-on (German) “Aus abstrakten Daten werden so Texte”
- NLG projects at large companies.
  - » INLG 2012 panel - Thomson-Reuters, Agfa
  - » More secretive

# Common Themes

---

- Almost all claim to generate narratives/stories from data
- Financial reporting is most commonly mentioned use
- Companies still quite small
  - » Fewer than 100 employees, compared to 12,000 at Nuance or 400,000 at IBM
  - » But large compared to earlier NLG companies
  - » Also lots of them!

# Questions?

---