Natural Language Generation and Data-To-Text

Albert Gatt

Institute of Linguistics, University of Malta

Tilburg center for Cognition and Communication (TiCC)

Department of Computing Science, University of Aberdeen

Natural Language Generation (NLG)

Natural language generation (NLG) is the process of **deliberately constructing a natural language text** in order to **meet specified communicative goals**.

(McDonald 1992)

Aims of these lectures

- To introduce the field of NLG
  - Main sub-tasks
  - Architectures
- To introduce some of the challenges in going from non-linguistic data to relatively long, “narrative” texts
  - Document planning
  - Dealing with time
- To look at methodological issues in NLG evaluation using different methods:
  - Evaluating large-scale systems with real users
  - Evaluating Referring Expression Generation algorithms

As a running example, we’ll often be looking at a recent family of systems developed as part of the BabyTalk Project.

Part 1

NLG AND CHOICE

What is NLG?

- Goal:
  - computer software which produces understandable and appropriate texts in human language
- Input:
  - Some non-linguistic representation of information (structured KB, numerical data…)
  - The goal that the system has to satisfy (e.g. inform the user)
- Output:
  - documents, reports, explanations, help messages, and other kinds of texts
- Knowledge sources required:
  - Knowledge of language (lexicon, grammar)
  - Knowledge of the domain (types of entities and events, their properties, etc)

Uses of NLG (examples)

- Informational:
  - Automatic generation of weather reports
  - Producing medical summaries from patient histories.
- Entertainment:
  - Automatic generation of stories, jokes etc.
- Interaction:
  - Dialogue agents that help users achieve a task.
- Assistive technology:
  - Aids for people with special communication needs.

What is involved?

- Suppose someone produces the following utterance:

  →*Today the temperature will be 25 degrees on average.*

- What are the steps required to produce this utterance?

Steps in language production

Today the temperature will be 25 degrees on average.

1. Communicative goal/intention
   - *Inform the reader*
2. Formulation (conceptualisation)
   - Deciding what to say
3. Encoding:
   a. Choice of words
   b. Choice of grammatical construction

Steps in language production

Today the temperature will be 25 degrees on average.

4. Realising the message
   - Ensuring syntactic correctness
   - Applying morphological rules

5. Mapping to speech output (or writing)

The main task

- The main problem for a speaker (and for an NLG system) is **choice**

- Let's unpack some of the choices.

Choice in language generation

Content determination

- Given a communicative intention, the generator needs to decide:
  - What is relevant
  - What the addressee knows (or not)
  - What is easily inferrable (and can therefore be left unsaid)

Choice in language production

Lexicalisation: Choice of words

- Humans have a vast lexicon at their disposal. Word choice can be impacted by a host of factors:
  - The addressee (will he understand a particular word?)
  - The genre and context (is this appropriate in the context?)
  - The speaker's personality.

Choice in language production

Lexicalisation: Choice of words

- *Today the temperature will be **25 degrees on average**.*
- *Today the temperature will be **roughly 25 degrees**.*
- *Today, the temperature will **average around 25 degrees**.*
- *Today, it will be **warm**.*

Choice in language production

Realisation: choosing grammatical structures

- Partly dependent on the words chosen.
- But also on many other factors, including desired style, context etc.
  - *Today the temperature will average 25 degrees.*
  - *Today the average temperature will be 25 degrees.*
  - *Today it'll be warm.*

How are choices made?

- Some choices are simply consequences of the language or grammar (cf. Reiter 2010).
  - *I kicked myself.*
  - **I kicked me.*
  - The choice of pronoun is directly predictable from principles governing the use of reflexives.

- Other choices are less straightforward:
  - *I kicked him. Vs. I kicked Pete.*
  - Depends, among other things, on how salient the person called Pete is in the hearer's memory.

NLG and choice

- Like human speakers, NLG systems too are designed to make choices at every level.
  - Systems differ in their degree of complexity.
  - It is possible to hardwire “shortcuts” into the system to avoid complex choices.
  - But full-blown NLG systems can become extremely complex.

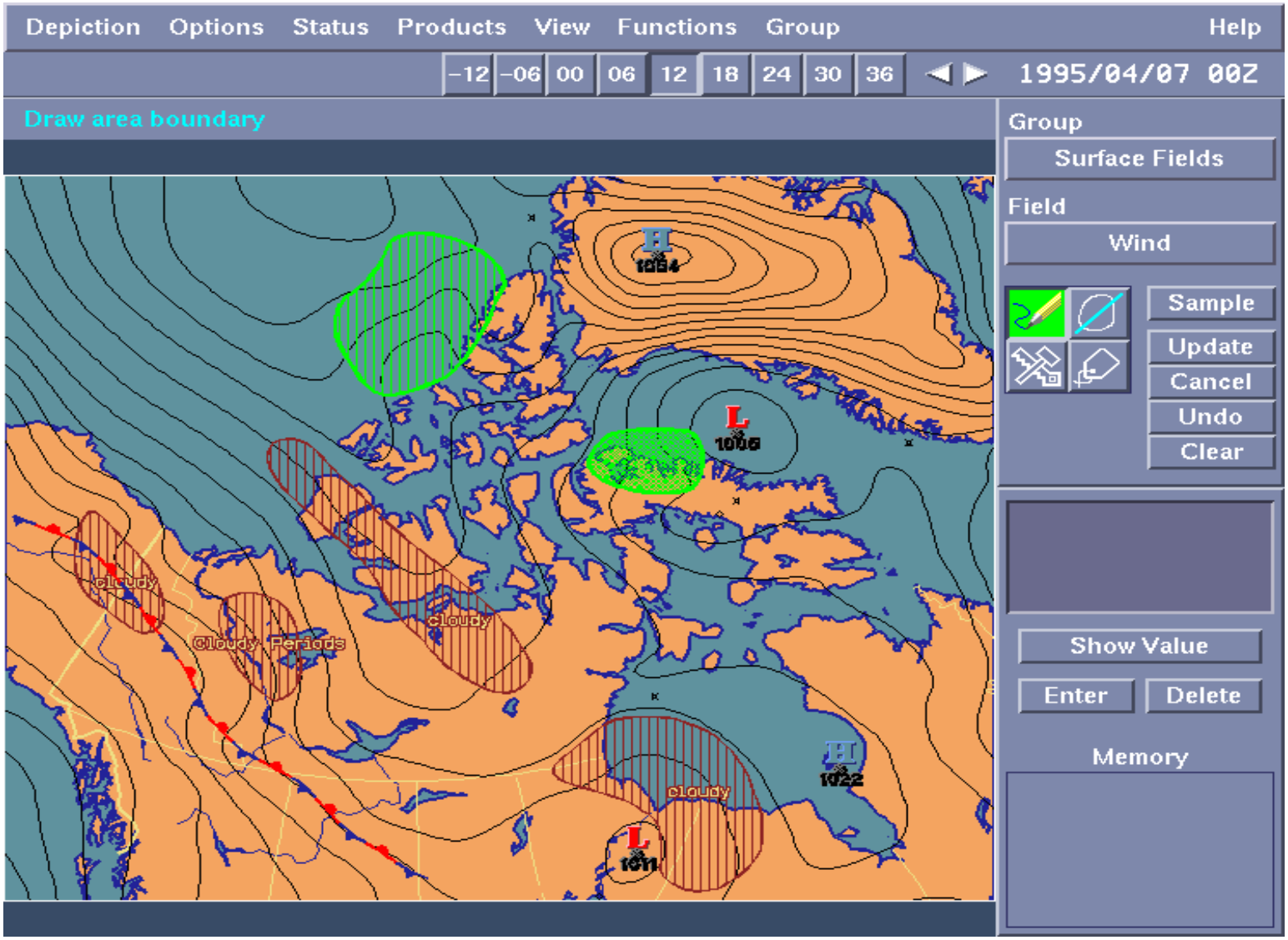
Part 2

SOME EXAMPLES OF NLG SYSTEMS

Example System #1: FoG

- Function:
  - Classic system that produces textual weather reports in English and French
- Input:
  - Graphical/numerical weather depiction
- User:
  - Environment Canada (Canadian Weather Service)
- Developer:
  - CoGenTex Inc.
  - See Goldberg *et al* (1994)

FoG: Input


Depiction
Options
Status
Products
View
Functions
Group
Help
-12
-06
00
06
12
18
24
30
36
1995/04/07 00Z
Draw area boundary
Group
Surface Fields
Field
Wind
Sample
Update
Cancel
Undo
Clear
Show Value
Enter
Delete
Memory
H
1024
L
1006
H
1022
L
1011
cloudy
Cloudy Periods
cloudy
cloudy

FoG: Output


FPCN20 Status: CURRENT-NOT RELEASED
FPCN20 CWEG 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 TONIGH
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

Example System #2: STOP

- Function:
  - Produces a personalised smoking-cessation leaflet
- Input:
  - Questionnaire about smoking attitudes, beliefs, history
- Developer:
  - University of Aberdeen
  - See: Reiter *et al.* (2003)

STOP: Input

SMOKING QUESTIONNAIRE

Please answer by marking the most appropriate box for each question like this: ☒

Q1 Have you smoked a cigarette in the last week, even a puff?

YES ☒ — Please complete the following questions

NO ☐ — Please return the questionnaire unanswered in the envelope provided. Thank you.

Please read the questions carefully. If you are not sure how to answer, just give the best answer you can.

Q2 **Home situation**:

Live alone ☐ | Live with husband/wife/partner ☒ | Live with other adults ☐ | Live with children ☒

Q3 **Number of children** under 16 living at home …… boys ……1……. girls

Q4 **Does anyone else in your household smoke?** *(If so, please mark all boxes which apply)*

husband/wife/partner ☒ | other family member ☒ | others ☐

Q5 **How long have you smoked for?** …10… years

Tick here if you have smoked for less than a year ☐

STOP: Output

Dear Ms Cameron

Thank you for taking the trouble to return the smoking questionnaire that we sent you. It appears from your answers that although you're not planning to stop smoking in the near future, you would like to stop if it was easy. You think it would be difficult to stop because *smoking helps you cope with stress, it is something to do when you are bored, and smoking stops you putting on weight*. However, you have reasons to be confident of success if you did try to stop, and there are ways of coping with the difficulties.

Example System #3: STORYBOOOK

- Function:
  - Produces a fairy tale for entertainment purposes
- Input:
  - An abstract story plan
- Developer:
  - Charles Callaway (part of a PhD project)
  - See: Callaway and Lester (2002)

STORYBOOK: input

```
(actor-property exist-being woodman001)
(refinement and-along-with woodman001
                           wife001)
(refinement belonging-to wife001
                         woodman001)
(specification exist-being
               process-step-type
               once-upon-a-time)
(...)
```

STORYBOOK: Output

Once upon a time, there was a woodman and his wife who lived in a pretty cottage on the borders of a great forest. They had one little daughter, a sweet child, who was a favourite with everyone.

She was the joy of her mother's heart, and to please her, the good woman made her a little scarlet cloak and hood. She looked so pretty in it that everyone called her Little Red Riding Hood.

Part 3

NLG TASKS AND ARCHITECTURES

The architecture of NLG systems

- There is widespread consensus on what NLG systems do (the tasks).

- There's more variation in how NLG systems are organised (the modules, input and output).
  - We often speak of a "consensus" architecture which has been adopted in many systems.
  - But many systems, especially statistical NLG systems, do not conform to this architecture.

NLG core tasks

RAGS project (Mellish *et al*, 2006)

Identified several important tasks which are carried out by many systems.

Based on an exhaustive survey of the state of the art.

But how are these tasks organised?

- **Lexicalisation:**
  - The choice of content words to appear in the final output text.
- **Aggregation:**
  - The combination of several structures (e.g., sentences) into a single, more complex, structure.
- **Rhetorical structuring:**
  - The determination of rhetorical relations and their scope.
- **Referring expression generation:**
  - Selection of content for referring expressions;
  - Decision on the form of these expressions (pronoun, definite description).
- **Ordering:**
  - The choice of linear ordering of the elements of the text.
- **Segmentation:**
  - The dividing up of information into sentences and paragraphs.

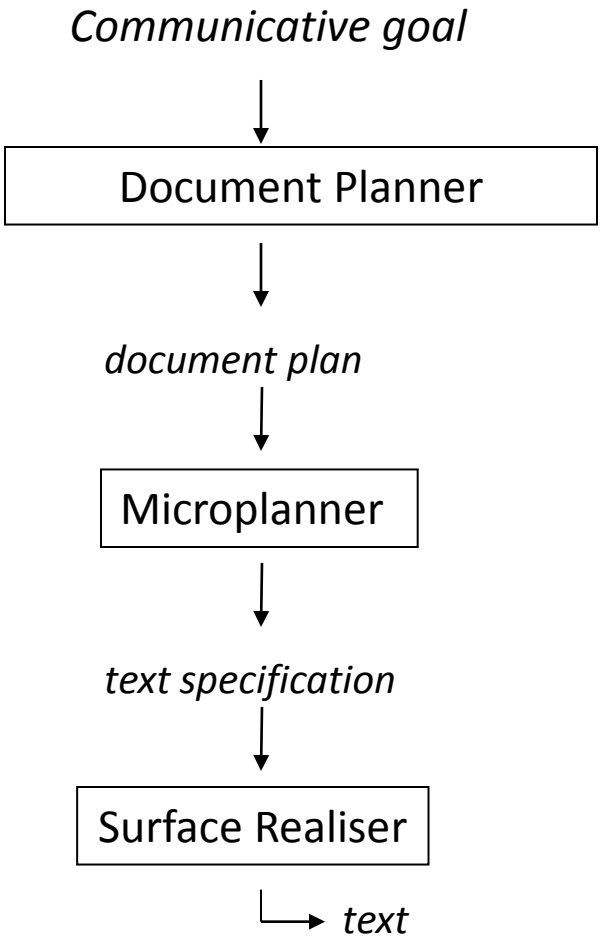
A consensus architecture?

- Reiter (1994) and Reiter and Dale (2000) argued that the various tasks can be grouped in a **three-stage pipeline**.

- Their architecture represents a "consensus" view.

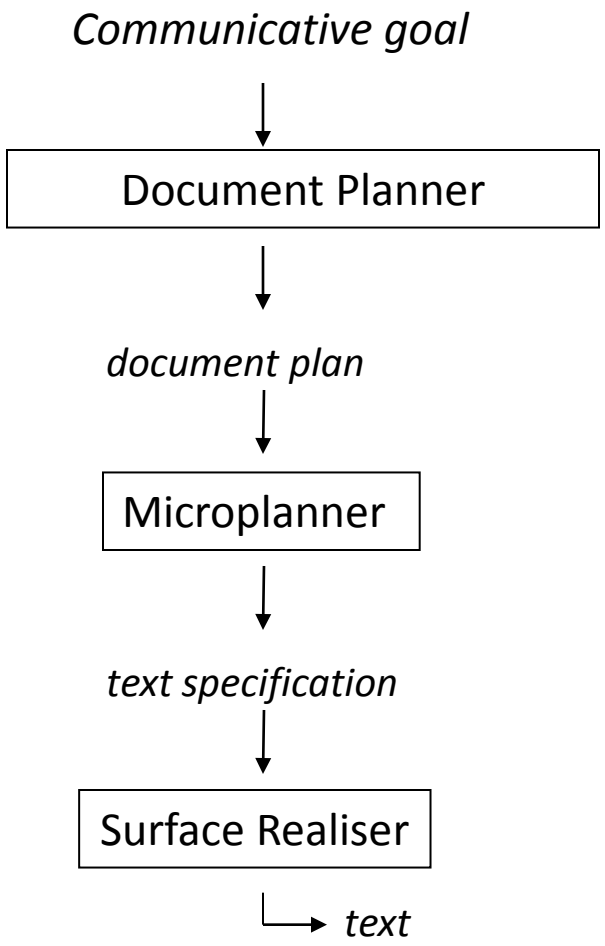
- But note that several systems do not conform to it.
  - This is especially true of statistical NLG systems.

The “consensus” architecture



- A pipeline architecture
  - highly modular
- NB: Diagram does not show knowledge sources!
  - Domain knowledge
  - Lexical/grammatical knowledge
  - Model of the user
  - …

The “consensus” architecture



Tasks:
- Selecting the messages to be included
- Rhetorical structuring
- Ordering
- Segmentation

Tasks:
- Lexicalisation
- Referring Expression Generation
- Aggregation

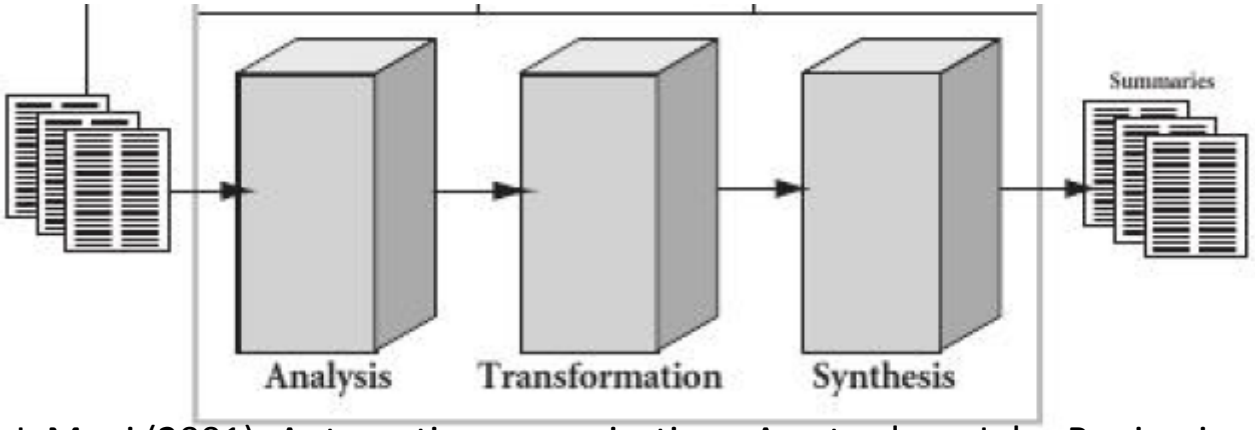
Tasks:
- Choosing syntactic structures
- Applying morphological rules
- Rendering the text as a string

NLG vs Automatic Summarisation

- Automatic summarisation systems generate summaries of one or more input documents.

- Most systems work by:
  - Analysing the input documents to extract important sentences
  - Carrying out some transformation on the result to render the text coherent
  - Synthesising the sentences into a summary

NLG vs Automatic Summarisation



I. Mani (2001). Automatic summarization. Amsterdam: John Benjamins

- There are some similarities between NLG and summarisation.
  - But summarisation systems take text as a starting point.
  - Extractive summarisers perform quite limited linguistic processing.
  - Some of the transformation/synthesis tasks done by summarisation systems are the same as those done by NLG systems during microplanning.

Part 4

CASE STUDY: THE BABYTALK SYSTEMS

Extending the architecture

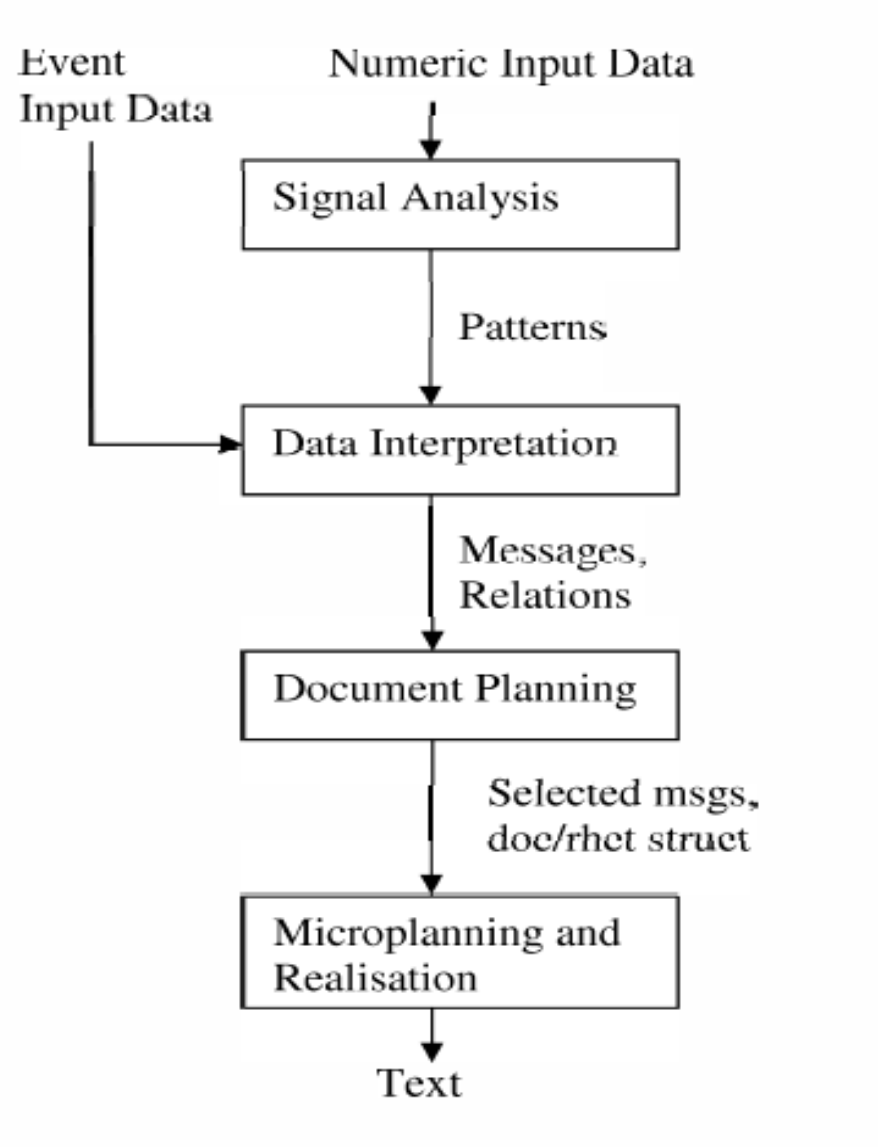
- Some NLG systems have to deal with **raw, unstructured data**.

- This means that prior to generating text, the data has to be analysed in order to:
  - Identify the important things and filter out noise
  - Map the data to appropriate input representations
  - Perform some reasoning on these representations

Extending the architecture

Reiter (2007) proposed to extend the "consensus" architecture to deal with preliminary stages of:

1. Signal analysis: to extract patterns and trends from unstructured input data;
2. Data interpretation: the perform reasoning on the results



BabyTalk

Context

- Neonatal Intensive Care Unit (NICU), where patients:
  - are typically preterm infants (e.g. 27 weeks)
  - are monitored continuously

  … so data is collected all the time.

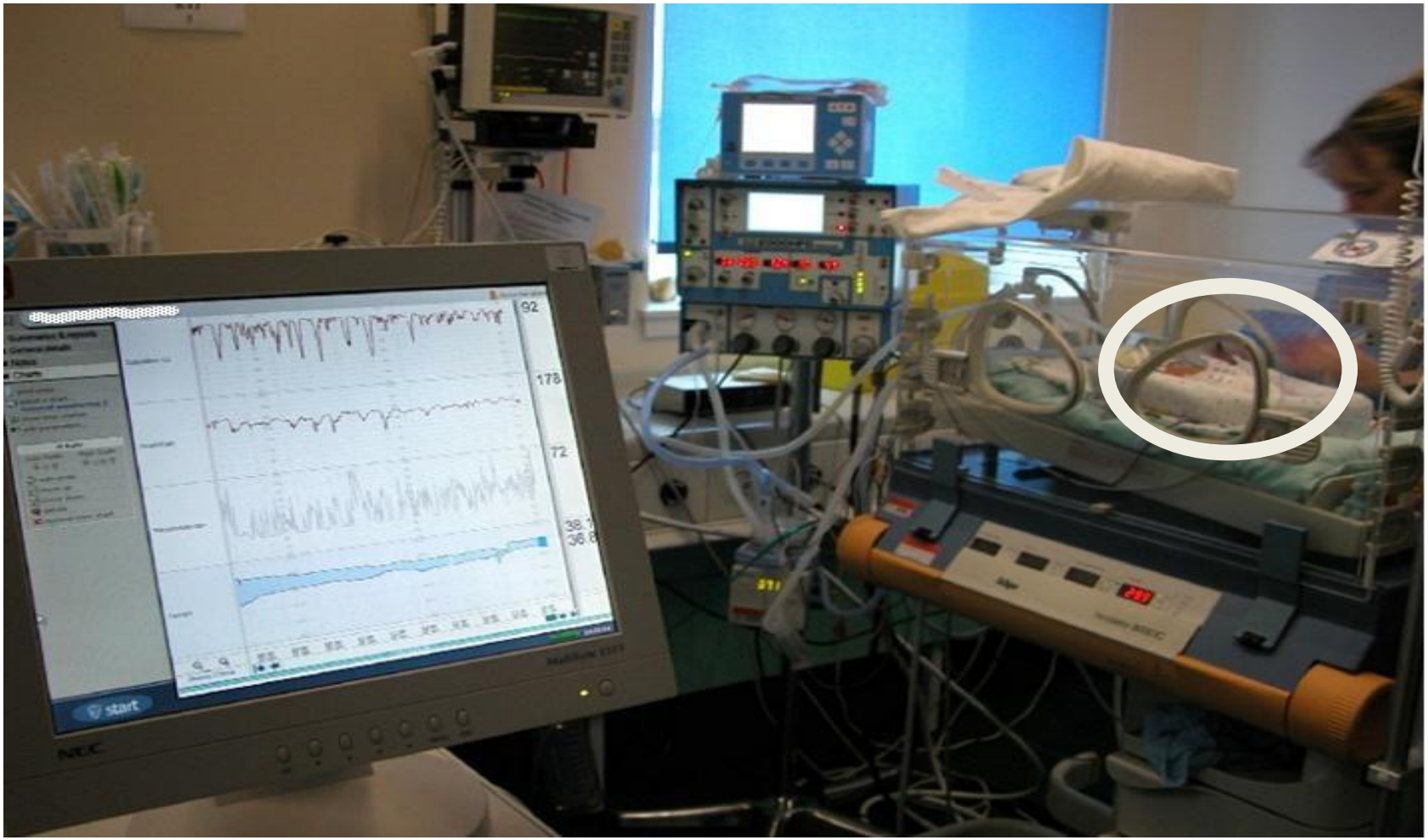
Problem

- A doctor or nurse needs to process the relevant parts of the data to make clinical decisions.

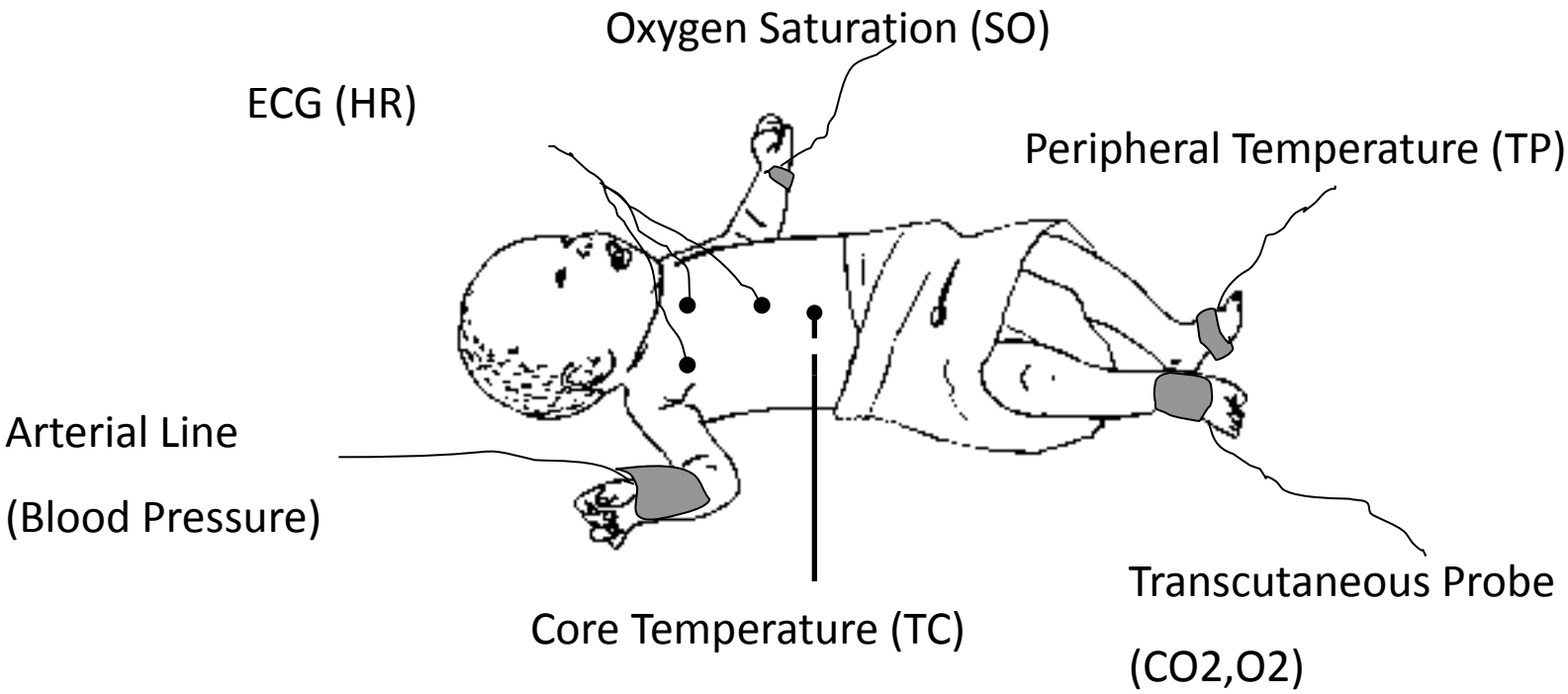
Typical scenario

- Large datasets are investigated using some form of **visualisation** (graphs, time series, etc).

NICU Environment

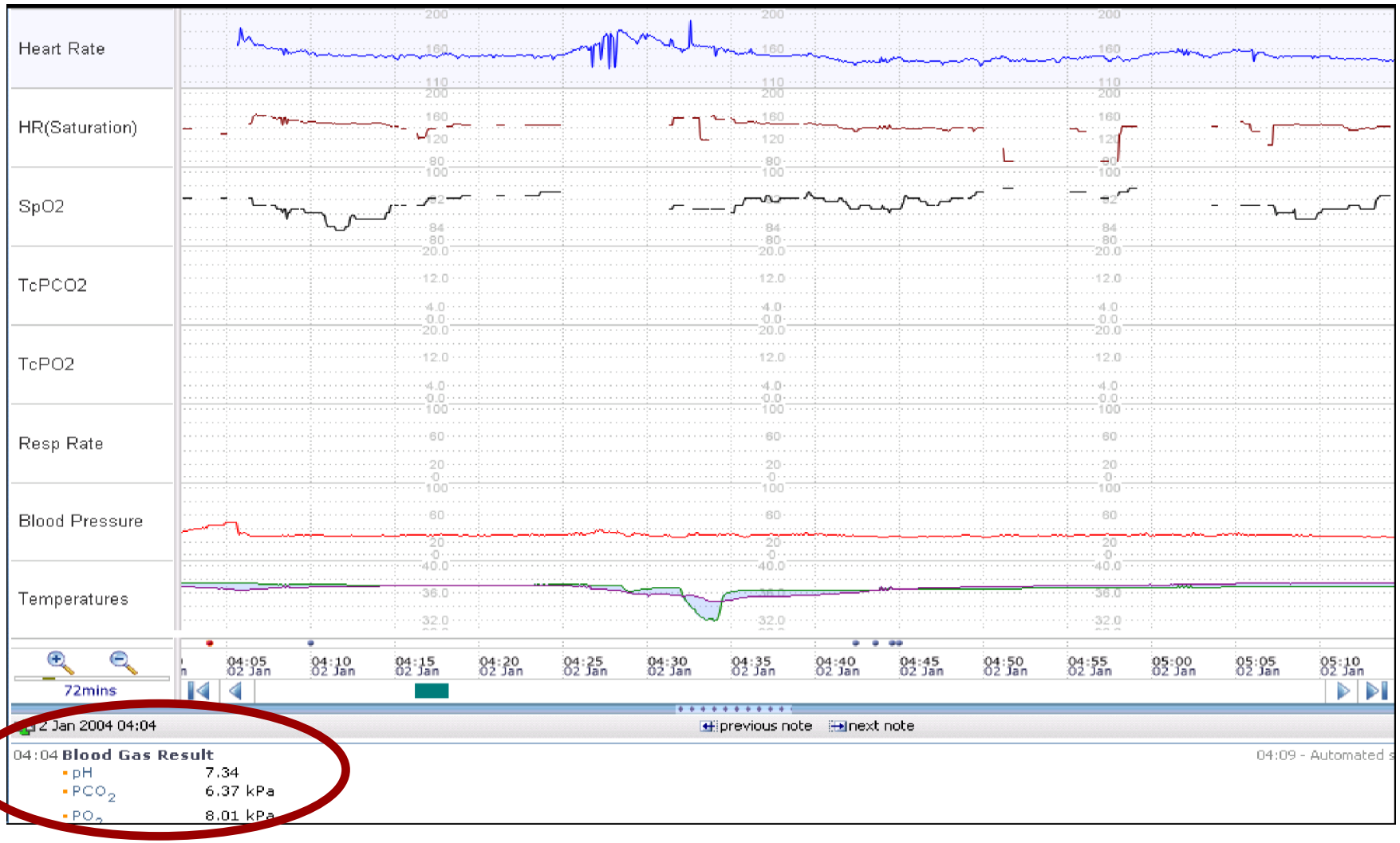


Data (I): continuous



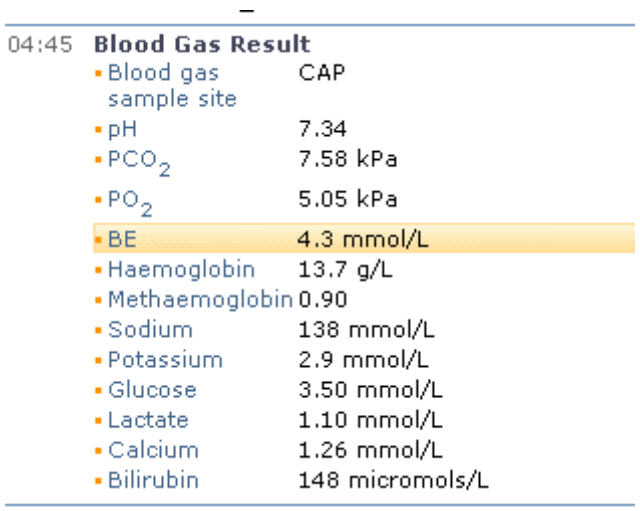
- Each channel sampled at 1Hz
  - 86,400 samples/channel/patient/day

Data (I): Continuous



Data (II): Sporadic (manual)

- Manually entered values
  - Numeric: results of lab tests, etc
  - Symbolic: type of respiratory support, drugs…



04:45	Blood Gas Result	
	Blood gas sample site	CAP
	pH	7.34
	PCO_2	7.58 kPa
	PO_2	5.05 kPa
	BE	4.3 mmol/L
	Haemoglobin	13.7 g/L
	Methaemoglobin	0.90
	Sodium	138 mmol/L
	Potassium	2.9 mmol/L
	Glucose	3.50 mmol/L
	Lactate	1.10 mmol/L
	Calcium	1.26 mmol/L
	Bilirubin	148 micromols/L

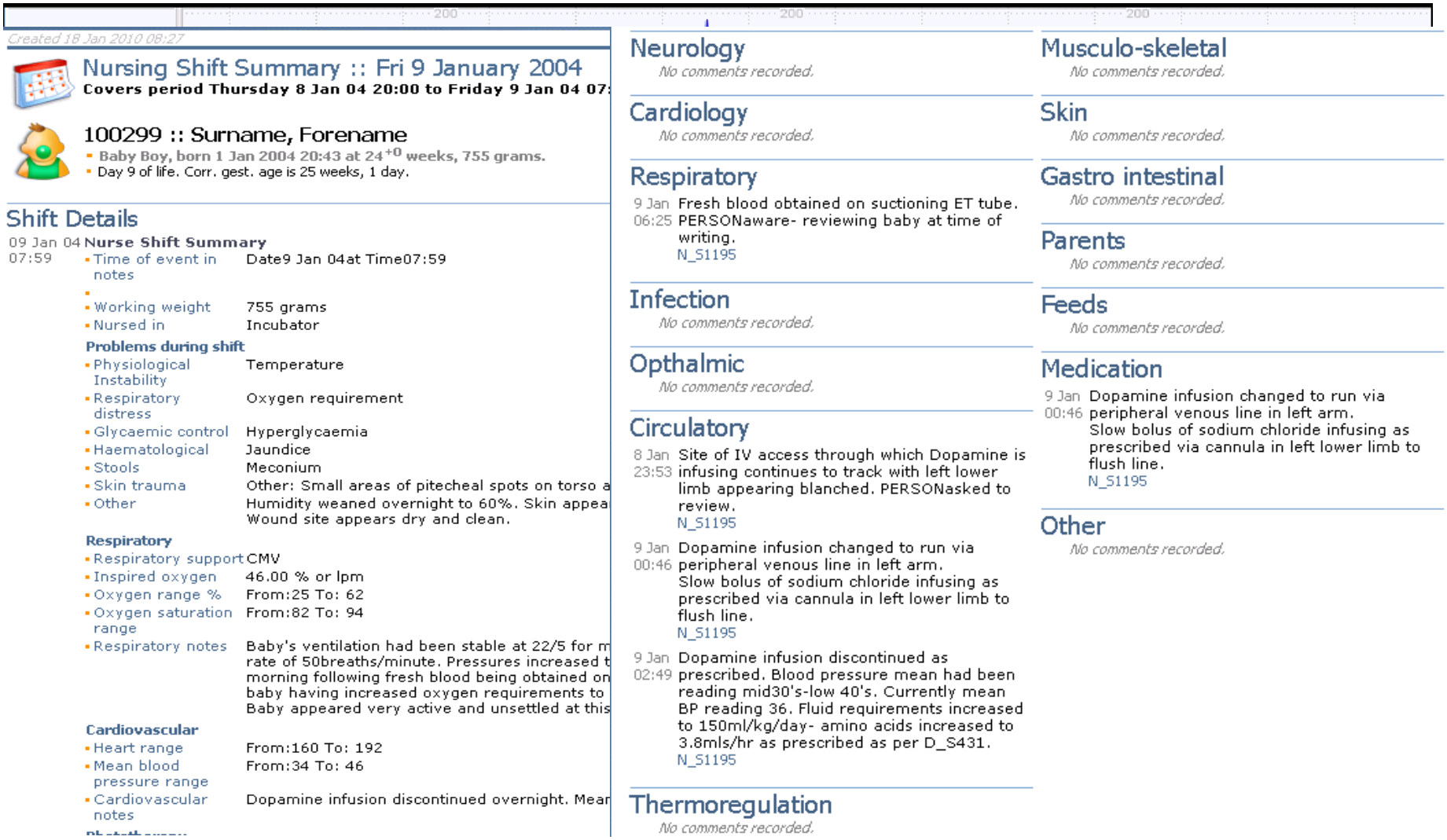
Currently displayed as a structured entry.

Data (III): Free text

- Notes entered periodically by nurses
  - No structure

> “Stable overnight Handling slightly better TPN ordered with extra potassium today Balance = +35mls, UO 2.7ml/kg/hr Ventilation unchanged - good gas overnight Morning bloods sent”

A “shift summary report”



Created 18 Jan 2010 08:27

Nursing Shift Summary :: Fri 9 January 2004

Covers period Thursday 8 Jan 04 20:00 to Friday 9 Jan 04 07:

100299 :: Surname, Forename

- **Baby Boy, born 1 Jan 2004 20:43 at 24^{+0} weeks, 755 grams.**
- Day 9 of life. Corr. gest. age is 25 weeks, 1 day.

Shift Details

09 Jan 04 07:59 **Nurse Shift Summary**

- Time of event in notes: Date9 Jan 04at Time07:59
- Working weight: 755 grams
- Nursed in: Incubator

Problems during shift

- Physiological Instability: Temperature
- Respiratory distress: Oxygen requirement
- Glycaemic control: Hyperglycaemia
- Haematological: Jaundice
- Stools: Meconium
- Skin trauma: Other: Small areas of pitecheal spots on torso a
- Other: Humidity weaned overnight to 60%. Skin appea Wound site appears dry and clean.

Respiratory

- Respiratory support: CMV
- Inspired oxygen: 46.00 % or lpm
- Oxygen range %: From:25 To: 62
- Oxygen saturation range: From:82 To: 94
- Respiratory notes: Baby's ventilation had been stable at 22/5 for m rate of 50breaths/minute. Pressures increased t morning following fresh blood being obtained on baby having increased oxygen requirements to Baby appeared very active and unsettled at this

Cardiovascular

- Heart range: From:160 To: 192
- Mean blood pressure range: From:34 To: 46
- Cardiovascular notes: Dopamine infusion discontinued overnight. Mear

Neurology

No comments recorded.

Cardiology

No comments recorded.

Respiratory

9 Jan 06:25 Fresh blood obtained on suctioning ET tube. PERSONaware- reviewing baby at time of writing.
N_S1195

Infection

No comments recorded.

Opthalmic

No comments recorded.

Circulatory

8 Jan 23:53 Site of IV access through which Dopamine is infusing continues to track with left lower limb appearing blanched. PERSONasked to review.
N_S1195

9 Jan 00:46 Dopamine infusion changed to run via peripheral venous line in left arm. Slow bolus of sodium chloride infusing as prescribed via cannula in left lower limb to flush line.
N_S1195

9 Jan 02:49 Dopamine infusion discontinued as prescribed. Blood pressure mean had been reading mid30's-low 40's. Currently mean BP reading 36. Fluid requirements increased to 150ml/kg/day- amino acids increased to 3.8mls/hr as prescribed as per D_S431.
N_S1195

Thermoregulation

No comments recorded.

Musculo-skeletal

No comments recorded.

Skin

No comments recorded.

Gastro intestinal

No comments recorded.

Parents

No comments recorded.

Feeds

No comments recorded.

Medication

9 Jan 00:46 Dopamine infusion changed to run via peripheral venous line in left arm. Slow bolus of sodium chloride infusing as prescribed via cannula in left lower limb to flush line.
N_S1195

Other

No comments recorded.

Why isn't that enough?

Previous research

- Nurses and doctors have trouble identifying important patterns in data;
- Long-term trends in a patient's health difficult to detect;
- Shift reports consist of "disjointed" items of information.

(McIntosh et al `00; Alberdi et al `01)

Law et al (2005):

- Off-ward experiment with doctors and nurses;
- 45 minutes of patient data, presented visually or in written form;
- Preference expressed for visualisations (the standard presentation format);
- Better decisions with summaries.

Data to Text

Our question

Can data-to-text NLG be used to summarise patient data in the NICU for decision support?

The answer

We think it can. This claim is based on the development and evaluation of the systems in the BabyTalk project.

BabyTalk Systems

BabyTalk developed systems to summarise patient data.

BT45

- Prototype system to summarise 45 minutes' worth of data from the NICU.

BT-Nurse

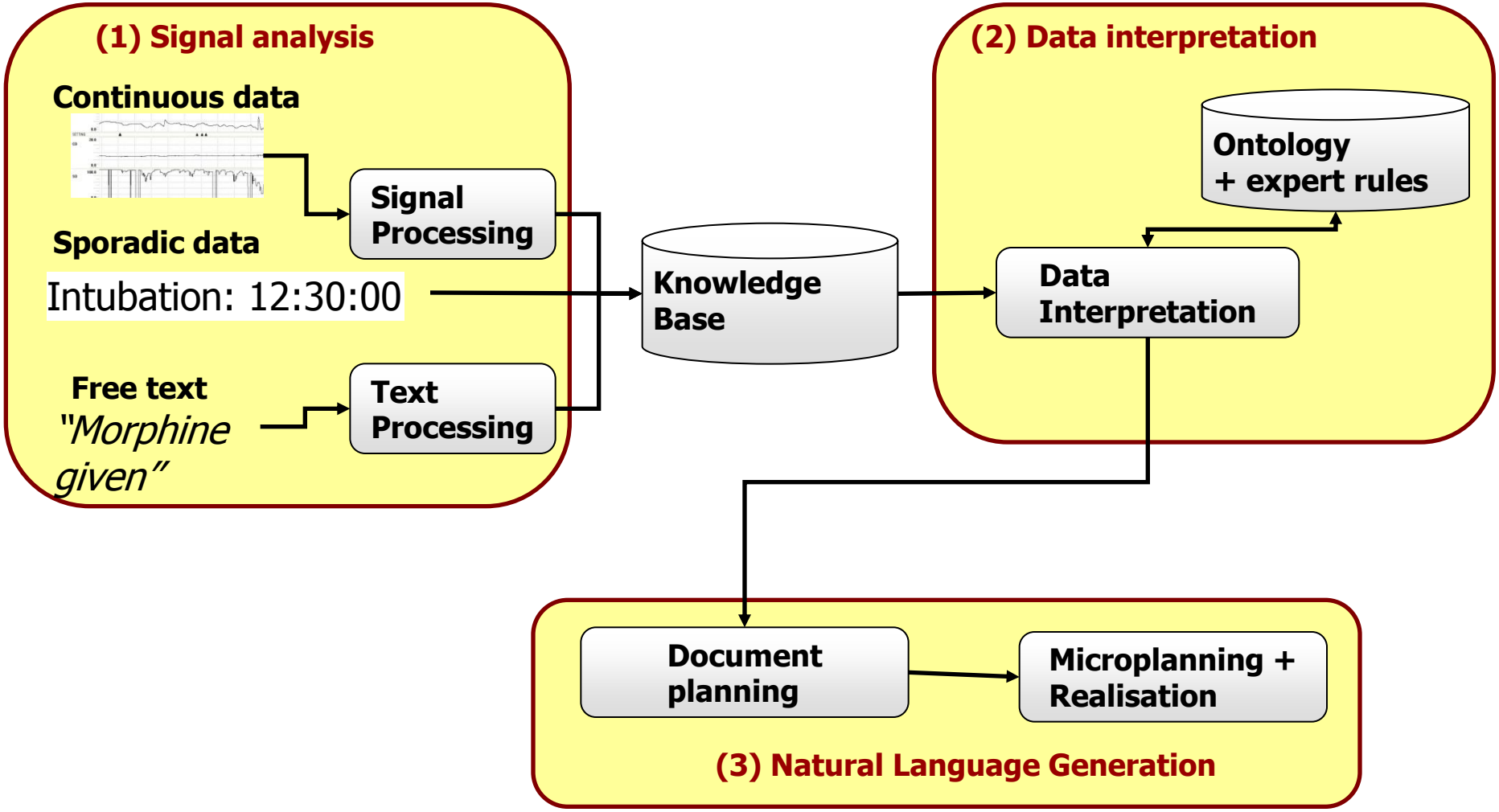
- Large-scale system to summarise a whole shift (12 hrs) of data to help nurses.

BT-Family/BT-Clan

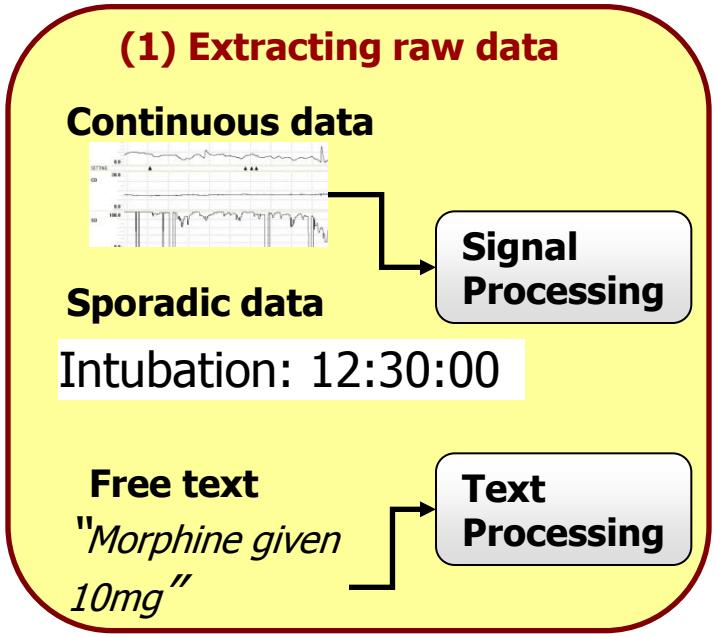
- Summarises patient data for family members and friends who want to know how their loved one is doing.

In the following sessions, we will focus mainly on the challenges with BT-Nurse and BT45

BabyTalk architecture



Data-to-text architecture

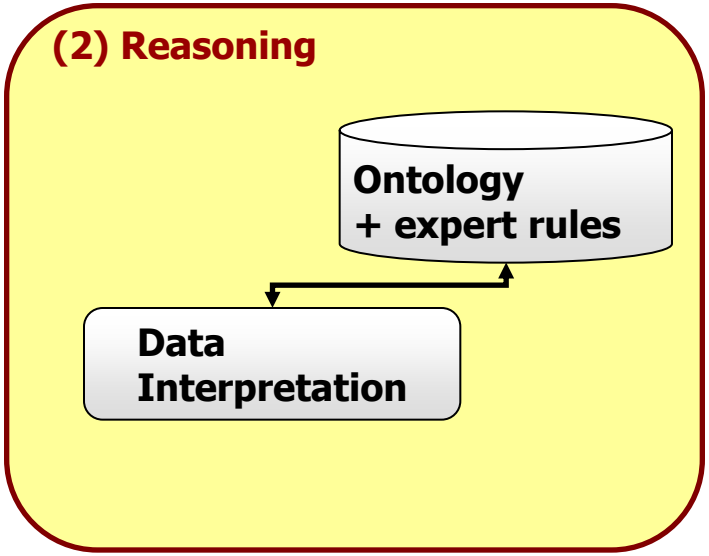


Events

- interval representation;
- at this stage, an unstructured list of all events in the relevant period.

Morphine: 15mg (12:31 – 12:31)
Intubation (12:30 – 12:34)
Channel reading (HR): 69 (12:35:20 – 12:35:40)
Channel reading (HR): 66 (12:36:20 – 12:36:40)
Channel reading (HR): 65 (12:53:10 – 12:53:25)

Data-to-text architecture



Events

- mapped to ontology concepts;
- expert rules to link them;
- importance assigned based on expert rules;
- Abstractions (e.g. channel readings merged into trends);
- some “diagnostic” reasoning

DRUG_ADMIN	
start	12:31
end	12:31
patient	baby_001
drug_given	morphine_001
drug_amount	10mg

INTUBATION	
start	12:30
end	12:34
patient	baby_001
imp	98

cause

TREND	
channel	HR
start	12:35
end	12:53
direction	down
lowest_val	65
imp	100

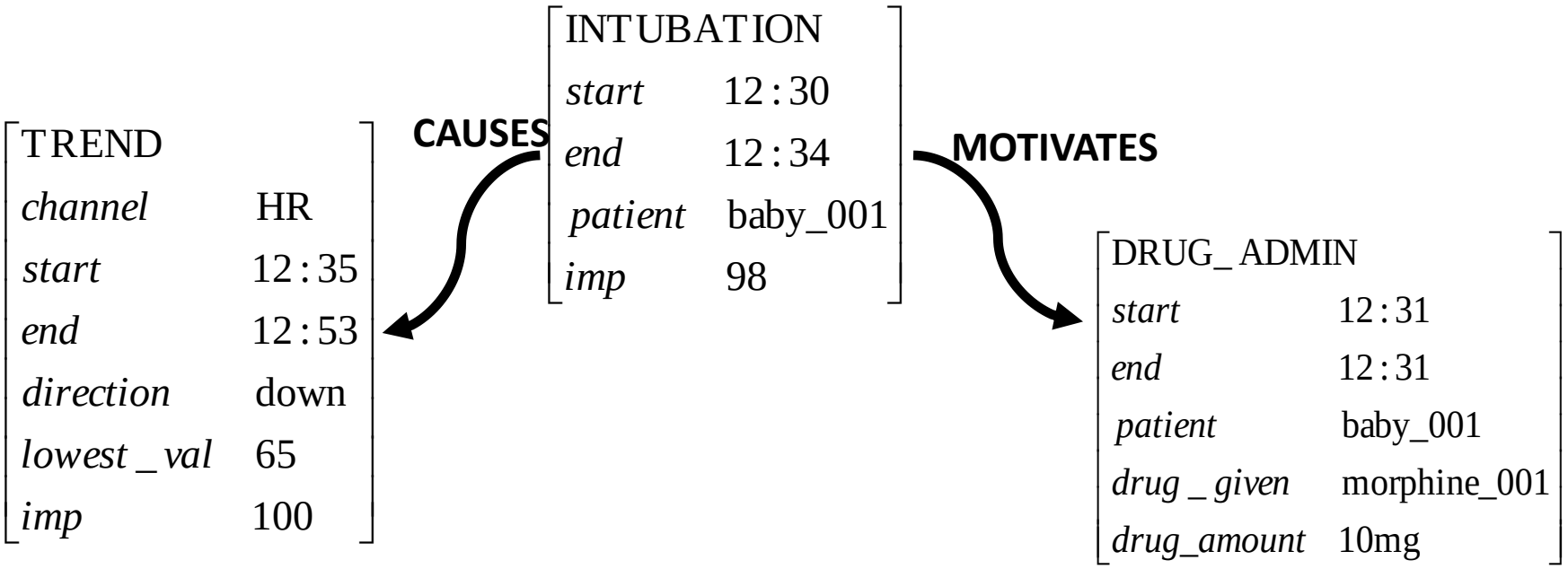
Data-to-text architecture

(3) Natural Language Generation

Document planning

Events

- selected based on importance;
- structured into a document plan (tree) with rhetorical links;
- sections/paragraphs determined.



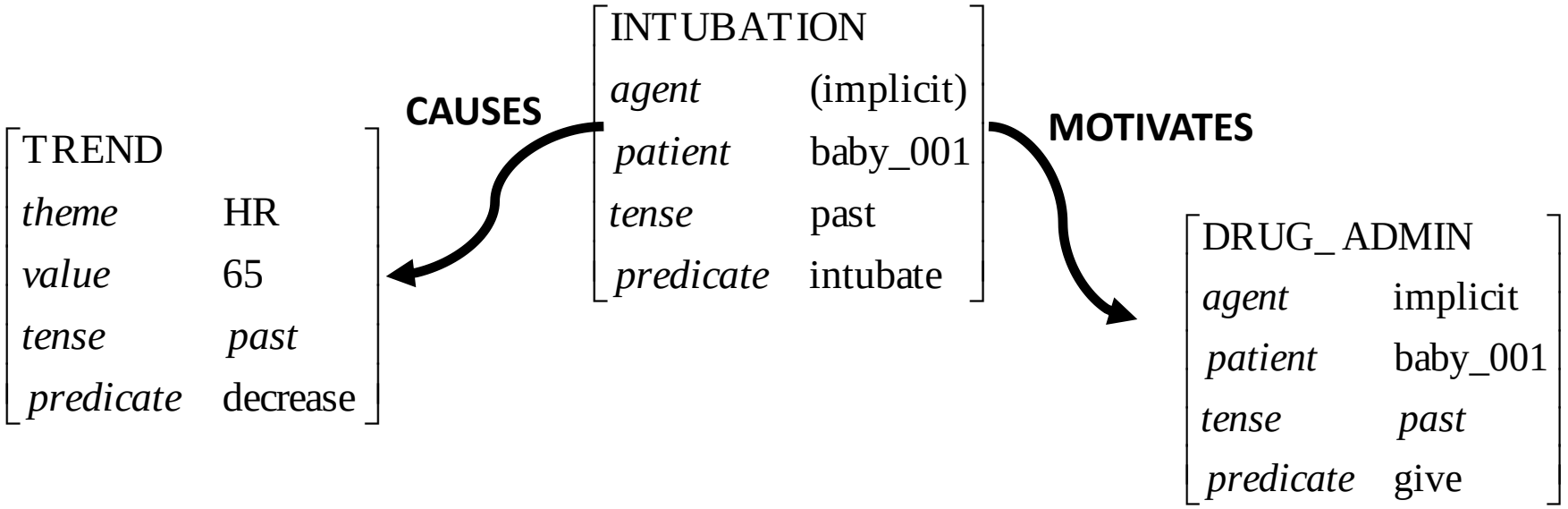
Data-to-text architecture

(3) Natural Language Generation

Microplanning

Events

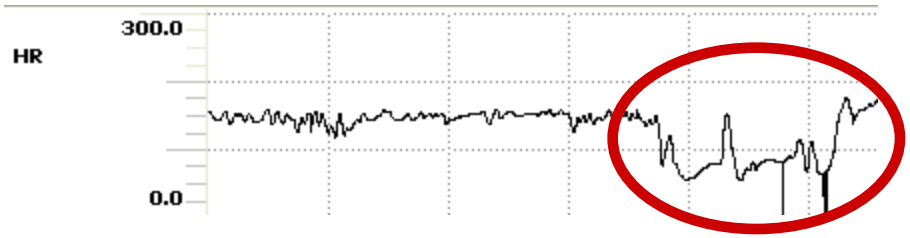
- rules to map events to argument structure;
- lexicalisation;
- referring expressions;
- temporal planning: tense, time modifiers;



Part 5

THE BT ARCHITECTURE: A MICRO EXAMPLE

A micro example



Input data: unstructured raw numeric signal from patient's heart rate monitor (ECG)

There were 3 successive bradycardias down to 69.

A micro example: pre-NLG steps

SEQUENCE (BRADYCARDIA)

BRADYCARDIA (16:58:46) Imp: 31.64
BRADYCARDIA (17:01:15) Imp: 79.80
BRADYCARDIA (17:03:57) Imp: 80.21
BRADYCARDIA (17:04:30) Imp: 39.97
BRADYCARDIA (17:05:01) Imp: 34.60
BRADYCARDIA (17:06:03) Imp: 66.24

(1) Signal Analysis (pre-NLG)

- Identify interesting patterns in the data.
- Remove noise.

(2) Data interpretation (pre-NLG)

- Estimate the importance of events
- Perform linking & abstraction

Document planning/Content Selection

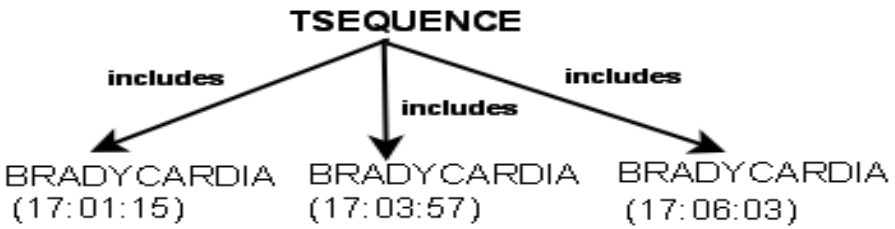
- Main tasks
  - Content selection
  - Information ordering

- Typical output is a document plan
  - tree whose leaves are messages
  - nonterminals indicate rhetorical relations between messages (Mann & Thompson 1988)
    - e.g. *justify*, *part-of*, *cause, sequence…*

A micro example: Document planning

SEQUENCE (BRADYCARDIA)

BRADYCARDIA (16:58:46) Imp: 31.64
BRADYCARDIA (17:01:15) Imp: 79.80
BRADYCARDIA (17:03:57) Imp: 80.21
BRADYCARDIA (17:04:30) Imp: 39.97
BRADYCARDIA (17:05:01) Imp: 34.60
BRADYCARDIA (17:06:03) Imp: 66.24



(1) Signal Analysis (pre-NLG)

- Identify interesting patterns in the data.
- Remove noise.

(2) Data interpretation (pre-NLG)

- Estimate the importance of events
- Perform linking & abstraction

(3) Document planning

- Select content based on importance
- Structure document using rhetorical relations
- Communicative goals (here: *assert something*)

A micro example: Microplanning

- Lexicalisation
  - Many ways to express the same thing
  - Many ways to express a relationship
  - e.g. SEQUENCE(x,y,z)
    - x happened, then y, then z
    - x happened, followed by y and z
    - x,y,z happened
    - there was a sequence of x,y,z
  - Many systems make use of a lexical database.

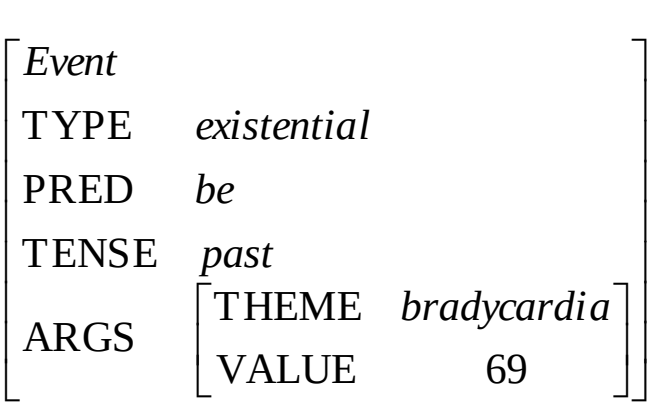
A micro example: Microplanning

- Aggregation:
  - given 2 or more messages, identify ways in which they could be merged into one, more concise message
  - e.g. be(HR, stable) + be(HR, normal)
    - (No aggregation) *HR is currently stable. HR is within the normal range.*
    - (conjunction) *HR is currently stable and HR is within the normal range.*
    - (adjunction) *HR is currently stable within the normal range.*

A micro example: Microplanning

- Referring expressions:
  - Given an entity, identify the best way to refer to it
  - e.g. BRADYCARDIA
    - *bradycardia*
    - *it*
    - *the previous one*
  - Depends on discourse context! (Pronouns only make sense if entity has been referred to before)

A micro example



(4) Microplanning

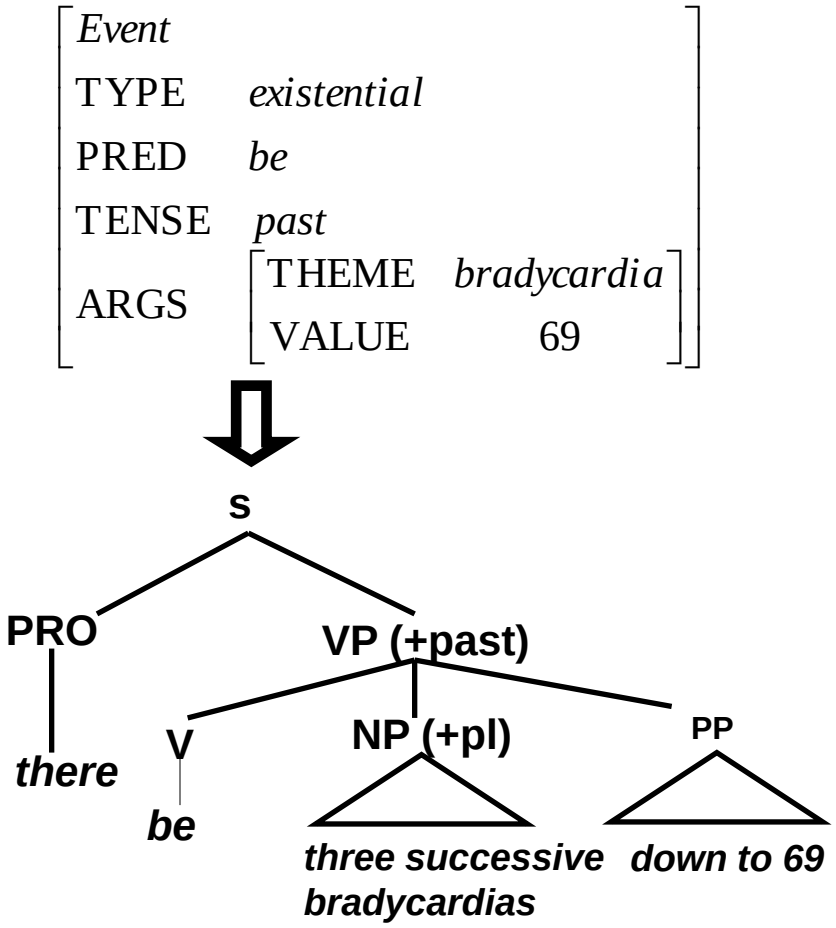
Map events to semantic representation

- lexicalise: *bradycardia* vs *sudden drop in HR*
- aggregate multiple messages (3 bradycardias = one sequence)
- decide on how to refer (*bradycardia* vs *it*)

A micro example: Realisation

- Subtasks:
  - map the output of microplanning to a syntactic structure
  - needs to identify the best form, given the input representation
    - typically many alternatives
    - which is the best one?
  - apply inflectional morphology (plural, past tense etc)
  - linearise as text string

A micro example



(4) Microplanning

Map events to semantic representation

- lexicalise: *bradycardia* vs *sudden drop in HR*
- aggregate multiple messages (3 bradycardias = one sequence)
- decide on how to refer (*bradycardia* vs *it*)
- choose sentence form (*there were…*)

(5) Realisation

- map semantic representations to syntactic structures
- apply word formation rules

Challenges

Document planning

- How to select appropriate content from such a vast input dataset?
- How to structure it according to the users' expectations?

Microplanning

- How to ensure that the (longish) texts generated are coherent?
- How to express events in such a way that their temporal and causal sequence is fully clear to the reader?