Natural Language Generation

An Overview

Stephan Busemann

DFKI GmbH

Saarbrücken, Germany

Stephan.Busemann@dfki.de

Acknowledgement: *Part of this presentation was inspired by Robert Dale's and Ehud Reiter's tutorial on Applied NL Generation at ANLP '97, Washington D.C, 1997*

Natural Language Generation

AN OVERVIEW

- ***What is NL Generation?***

  a definition, the roots, and scientific directions

- ***What must/should/can a NLG system do?***

  content selection, linguistic planning, realization

- ***How do its components depend on each other?***

  pipelined, integrated, and interacting architectures

- ***Where is the field moving?***

  applications. application areas, and prototypes

- ***Where can I find more information?***

  workshops, books, software, the Web

What is NL Generation?

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

[McDonald 1992]

- **Goal**
  - computer software which produces understandable text in a human language
- **Input**
  - a communicative goal, including
  - a non-linguistic representation of information
- **Output**
  - a text, either plain ASCII or formatted (LaTeX, HTML, RTF), either solo or combined with graphics, tables etc.
- **Knowledge sources required**
  - knowledge of communication, of the domain, and the language

Source: Stephan Busemann

Why is NL Generation Needed?

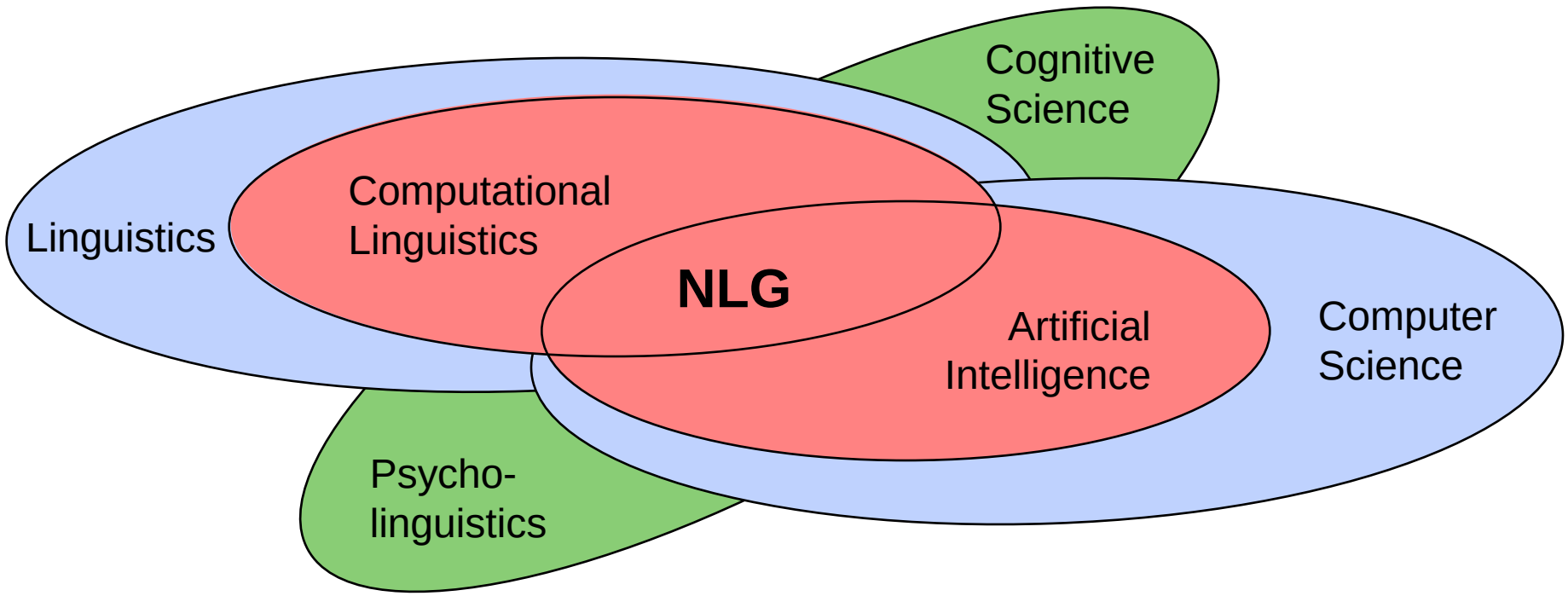
- **Information of interest is stored on the computer in ways which are not comprehensible to the end user.**
- **NLG systems can present this information to users in an accessible way.**

- **NL dialogue interfaces to application systems**
  - NL DB access, explanations of inferences in XPS, game characters, corrections (false user implicatures)
- **Machine translation**
  - target language text based on result of source language analysis and transfer
- **Text generation**
  - documents, reports, summaries, help messages, etc.

Source: Stephan Busemann

NL Generation is an Interdisciplinary Research Field

- **Artificial Intelligence**
- **Psycholinguistics**
- **Computational Linguistics**



Source: Stephan Busemann

NL Generation in Artificial Intelligence

What are the decision-making and planning processes needed for NL generation?

Research on knowledge-based approaches to developing computer systems capable of human language production

- **Scientific issues**
  - which types of knowledge are necessary, and how should they be represented?
  - how can inferences be modelled and controlled?
  - which representations and interfaces allow efficient processing?
- **Methods**
  - deep modelling for small classes of examples
  - implementation of complex systems
- **Implementations for theory validation or for building research prototypes**

Source: Stephan Busemann

NL Generation in Psycholinguistics

How does human language production work?

Research on human linguistic capabilities (spoken language)

- **Scientific issues**
  - which processes are required for a speaker to produce an utterance?
  - in which order are these processes scheduled?
  - which representations does a speaker access during language production?
- **Methods**
  - experiments with human speakers to retrieve data and to test hypotheses
- **Implementations for theory validation**

Source: Stephan Busemann

NL Generation in Computational Linguistics

Given a semantic representation and a grammar - what are the sentences admitted by the grammar?

Research on the use of modular, linguistically well-founded theories for the mapping between logical formulae and terminal strings

- **Scientific Issues**
  - which semantic and syntactic phenomena should be described by the grammar?
  - which control strategies are suitable for the grammar formalism at hand?
  - under which conditions are the processes reversible?
- **Methods**
  - integrated treatment of semantic and syntax
  - use of constraint-based formalisms (features structures)
- **Implementations for theory validation and as test beds**

Overview (2)

What is NL Generation?

a definition, the roots, and scientific directions

What must/should/can a NLG system do?

content selection, linguistic planning, realization

How do its components depend on each other?

pipelined, integrated, and interacting architectures

Where is the field moving?

applications, application areas, and prototypes

Where can I find more information?

workshops, books, software, the Web

What Must a Generation System Do?

TASKS IN NL GENERATION

- **Content determination**
- **Discourse planning**
- **Sentence aggregation**
- **Lexicalization**
- **Referring expression generation**
- **Surface realization**

more language dependency

more decision-making

Content Determination Selects the Information to be Communicated

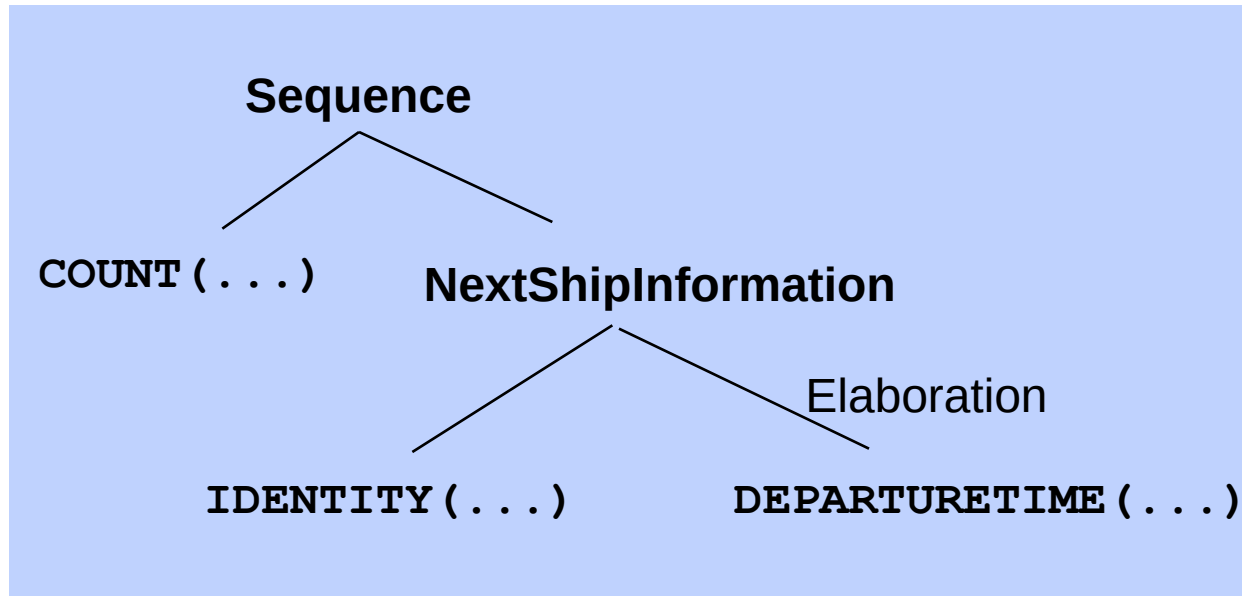
- **Construct a set of *messages* from the underlying data source**
- **Messages are aggregations of data that are appropriate for verbalization**
- **A message may correspond to a word, a phrase, a sentence**
- **Messages are based on domain entities (concepts, relations)**

```
IDENTITY(NEXTSHIP, MS-LILLY)
        ;The next ship is the MS-LILLY.
DEPARTURETIME(MS-LILLY, 1000)
        ;The MS-LILLY departs at 10am.
COUNT(SHIP, SOURCE(HAMBURG),
        DESTINATION(COPENHAGEN), 5, PERDAY)
        ;There are five ships daily from Hamburg to
        Copenhagen.
```

Discourse Planning Organizes Messages into a Coherent Text Plan

- **A text is not just a random collection of sentences**
- **Texts have an underlying structure relating the parts together**
- **Two related issues**
  - conceptual grouping
  - rhetorical relationships

There are five ships daily from Hamburg to Copenhagen. The next ship is the MS-LILLY. It departs at 10am.



Sentence Aggregation Distributes Messages Onto Sentences

- **A one-to-one mapping from messages onto sentences may result in disfluent text**
- **Messages need to be combined to produce larger and more complex sentences**
- **The result is a *sentence plan***

Without aggregation	**With aggregation**
The next ship is the MS-LILLY. It leaves Hamburg at 10am. It has a restaurant. It has a snack bar.	*The next ship, which leaves Hamburg at 10am, is the MS-LILLY. It has a snack bar and a restaurant.*

Source: Stephan Busemann

Lexicalization Determines the Content Words to be Used

- **Knowledge sources include**
  - communicative intention, concepts and relations, focus, user model
- **A variety of subtasks may become critical**
  - consider/choose the discourse focus - *buy* vs *sell*
  - use collocations - *exert influence* vs *administer punishment*
  - consider lexical semantics - *male unmarried adult* vs *bachelor*
  - by default use basic level categories[1] - *dog* vs *poodle*
  - consider underlying situation - *the pole is thick and sufficiently high*
  - consider/choose the attitude - *house* vs *home, father* vs *dad*
  - know about idioms - *kick the bucket*
- **Lexical choice is a mapping from concepts and relations onto lexemes**
- **Lexical choice determines (part of) the syntactic structure**

[1] Basic level categories represent the level of abstraction first named and understood by children (cf work by Eleanor Rosch). People remember and name things more readily at basic level. Languages have simpler names at basic level (e.g. furniture – *chair* – kitchen chair).

Source: Stephan Busemann

Referring Expressions Allow the Hearer to Identify Discourse Objects

- **Task: Avoid ambiguity, but also avoid disfluency**
  - *? the deer next to the two trees on the left of the house*
- **Kinds of referring expressions**
  - Proper names - *Hamburg, Stephan, The United States of America*
  - Definite descriptions - *the ship that leaves at 10am, the next ship*
  - Proforms - *it, later, there*
- **Initial reference**
  - use a full name - *the MS-LILLY*
  - relate to an object that is already salient - *the ship's snack bar*
  - specify physical location - *the ship at pier 12*
- **Choosing a form of reference**
  - proform > proper name > definite description

How should definite follow-on descriptions look like?

Source: Stephan Busemann

Surface Realization Generates Grammatically Correct Text

- **Converts sentence plans into text**
- **Subtasks include**
  - insert function words - *he wants to book a ticket*
  - word inflection - *like+ed → liked*
  - ensure grammatical word order
  - apply orthographic rules
- **Techniques of defining grammatical knowledge**
  - declarative bidirectional grammars, mapping between semantics and syntax
  - grammars tuned for generation, widely used in practice
  - templates, easy and fast to implement

Source: Stephan Busemann

Overview (3)

What is NL Generation?

a definition, the roots, and scientific directions

What must/should/can a NLG system do?

content selection, linguistic planning, realization

How do its components depend on each other?

pipelined, integrated, and interacting architectures

Where is the field moving?

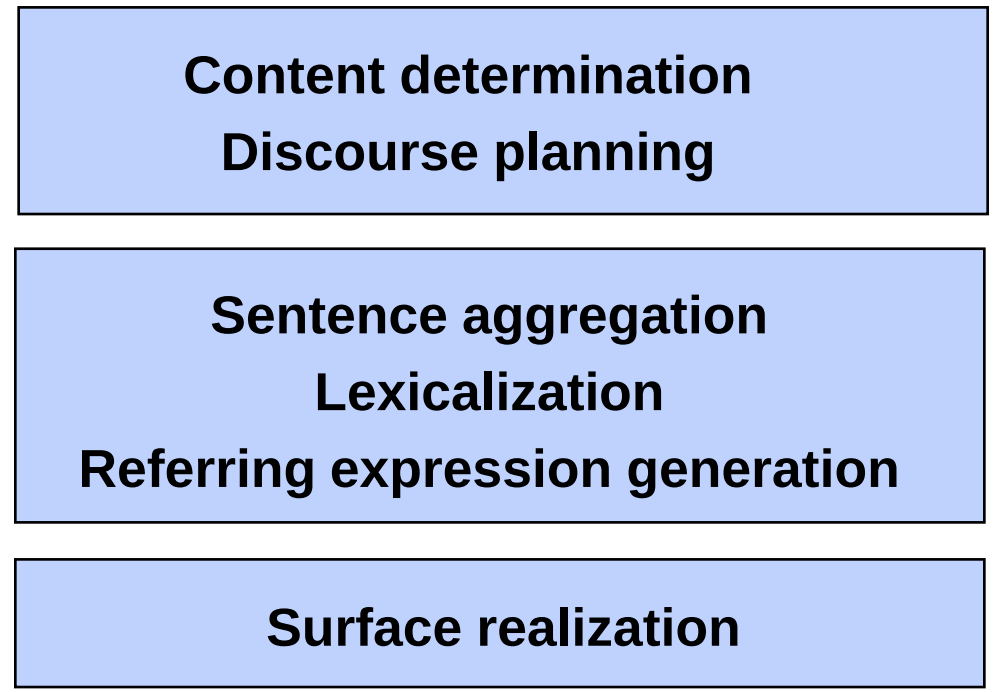
applications. application areas, and prototypes

Where can I find more information?

workshops, books, software, the Web

The NLG Tasks Can be Grouped Into Modules With Interfaces

- **Text planning (Macroplanning)**
- *Text Plan*
- **Sentence planning (Microplanning)**
- *Sentence Plan*
- **Linguistic realization**



Applicable techniques include planning and rule-based – or constraint-based – systems

Source: Stephan Busemann

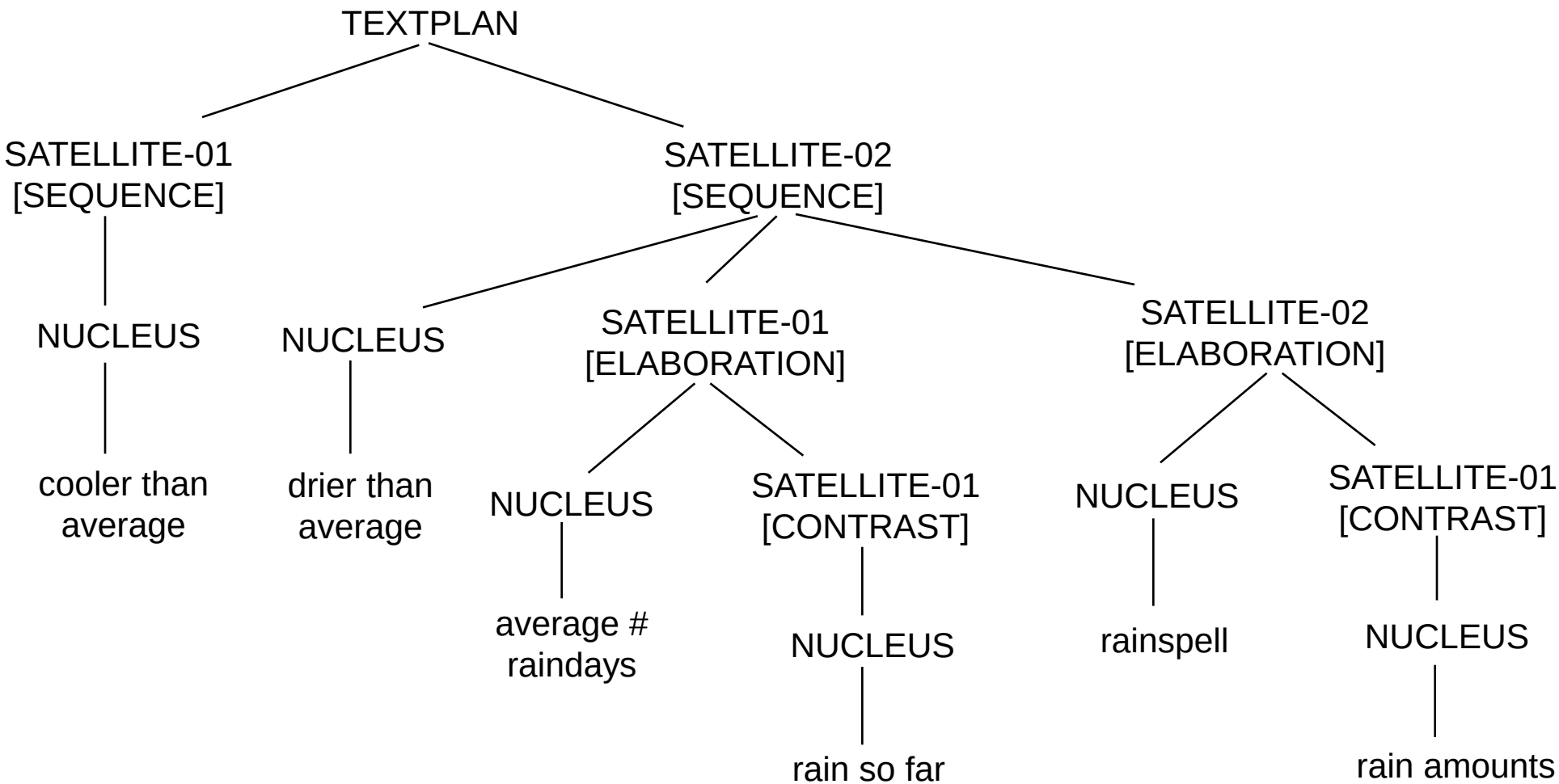
A Generated Target Text

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.

msg1 msg2, msg3, BUT msg4. ALTHOUGH msg5, msg6.

Source: Stephan Busemann

A Sample Text Plan



- **Rhetorical Structure Theory is a basis for discourse planning**

Source: Stephan Busemann

A Sample Sentence Plan (Sentence Plan Language, SPL)

```
(l / greater-than-comparison
   :tense past
   :exceed-q (l a) exceed
   :domain (m / one-or-two-d-time :lex month :determiner the)
   :standard (a / quality :lex average :determiner zero)
   :range (c / sense-and-measure-quality :lex cool)
   :inclusive (r / one-or-two-d-time
     :lex day
     :number plural
     :property-ascription (r / quality :lex rain)
     :size-property-ascription
                          (av / scalable-quality :lex the-av-no-of)))
```

The month was cooler than average with the average number of rain days.

- **SPL input to KPML; wide spread surface-semantic representation level**

Source: Stephan Busemann

Interdependencies of Components

EXAMPLES

- **Discourse planning and sentence aggregation**

  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.

  The month was cooler and drier than average, with the average number of rain days, but the yearly rain so far well below average.

- **Sentence aggregation and syntax**

  (1) likes(dog,cat) (2) black(cat) (3) sad(dog) (4) chase(cat,mouse)

  (1, 2, 3) The dog [that likes the [black] cat] is sad.

  (1, 2, 3, 4) ? The dog [that likes the [black] cat [that is chasing the mouse]] is sad.

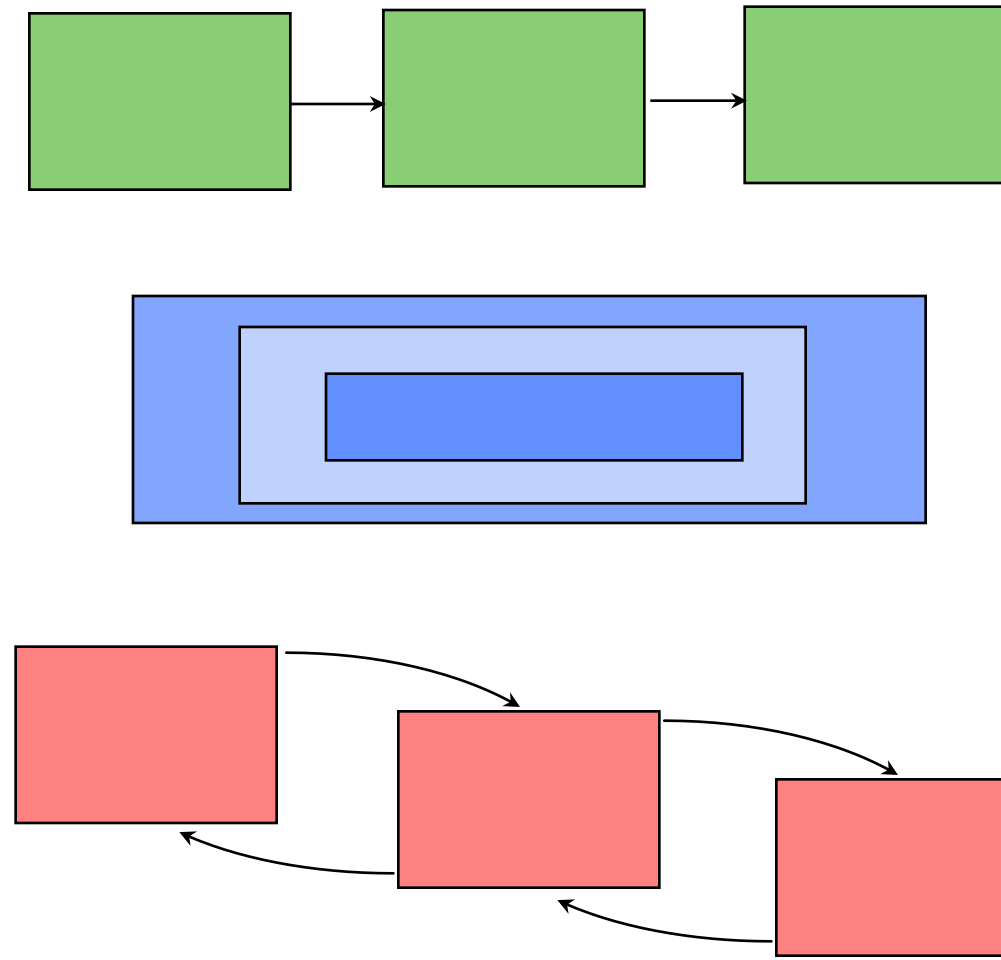
- **Discourse planning and lexicalization**

  Mary was killed. She was shot by John.

  ? *Mary was shot. She was killed by John.*

Architectures in NLG

- **Pipelined**
  - simplest
  - inadequate
  - most widespread
- **Integrated**
  - all in one formalism
  - elegant
  - inefficient
- **Interacting**
  - psycholinguistically plausible
  - complex
  - impractical



Source: Stephan Busemann

Overview (4)

What is NL Generation?

a definition, the roots, and scientific directions

What must/should/can a NLG system do?

content selection, linguistic planning, realization

How do its components depend on each other?

pipelined, integrated, and interacting architectures

Where is the field moving?

applications, application areas, and prototypes

Where can I find more information?

workshops, books, software, the Web

The Complete NLG System Does Not Exist (Yet)

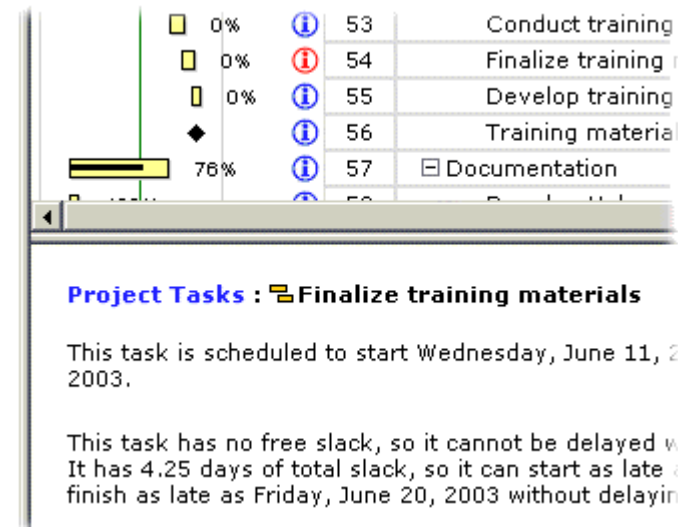
- **Discourse planning**
  - proof of concept for many sample domains
  - relation classes are hard to define
- **Sentence aggregation**
  - techniques quite well understood
  - applicability conditions unknown
- **Lexicalization**
  - methods understood in isolation
  - often shifted aside due to complex interdependencies
- **Referring expression generation**
  - pronominalization well understood
  - initial object characterization difficult
- **Surface realization**
  - scientifically solved in principle
  - reusable application systems being fielded

Source: Stephan Busemann

Some NLG Applications (1)

- **FoG**
  - Function: produces textual weather reports in English and French
  - Input: graphical weather depiction
  - User: Environment Canada (Canadian Weather Service)
  - Developer: CoGenTex (http://cogentex.com/)
  - Status: Fielded, in operational use since 1992

- **Chart Explainer**
  - Function: generates fluent NL summaries of charts and tables
  - Developer: CoGenTex (http://cogentex.com/products/chartex/index.html)



- **Project Reporter**
  - Function: generates project information reports
  - Developer: CoGenTex (http://cogentex.com/products/reporter/index.html)

Some NLG Applications (2)

- **PlanDoc**
  - Function: produces a report describing simulation options an engineer has explored
  - Input: simulation log file
  - User: Southwest Bell
  - Developer: Bellcore and Columbia University
  - Status: Fielded, in operational use since 1996

- **STOP**
  - Function: produces a letter encouraging patients to stop smoking
  - Input: patient data, trials, successes and failures, pros and cons
  - User: patients suffering from nicotine-caused diseases
  - Developer: U Aberdeen
  - Status: field trial in hospital, 2002

Source: Stephan Busemann

Evaluation – a Research Topic on Its Own

- **Problem: NLG tasks usually have multiple solutions**
  - Edit distance measures wrt gold standard are therefore inappropriate
- **Intrinsic evaluation**
  - Select aspects of NLG, e.g. REG, and measure their quality
  - Annotations relating REs and their properties to concepts and distractors
    - TUNA corpus, 780 human-produced REs, visual domain of furniture and people
    - Compare system output to reference outputs in the corpus
  - Comparative evaluation through Shared Tasks
    - Train systems on set of entities and REs, and test on set of objects (no REs)
- **Extrinsic evaluation**
  - Define application scenario and measure task performance
  - Direct test, e.g. use generated RE to identify an object in a picture
  - Indirect test by embedding NLG systems: measure added value of NLG to task

Research issue: How can intrinsic and extrinsic evaluations be combined?

Source: Stephan Busemann

Conclusions

What is NL Generation?

a definition, the roots, and scientific directions

What must/should/can a NLG system do?

content selection, linguistic planning, realization

How do its components depend on each other?

pipelined, integrated, and interacting architectures

Where is the field moving?

applications. application areas, and prototypes

Where can I find more information?

workshops, books, software, the Web

Pointers to NLG Resources

- **SIGGEN (ACL Special Interest Group for Generation)**
  - http://www.siggen.org/
  - Who‘s who and what’s where?
  - Papers, bibliographies, conference and workshop announcements, job offers, …
  - Free software, demos
- **Conferences and Workshops**
  - International Conference on NLG (in even-numbered years)
  - European Workshop on NLG (on odd-numbered years)
  - NLG papers at ACL, ANLP, IJCAI, AAAI, ...
- **Research Labs and Companies**
  - U Aberdeen: http://www.csd.abdn.ac.uk/research/nlg2/
  - Saarbrücken: http://www.dfki.de/service/NLG/
  - CoGenTex: http://www.cogentex.com
- **Key Literature**
  - Ehud Reiter and Robert Dale (2000): *Building Natural Language Generation Systems*. Cambridge University Press.

Source: Stephan Busemann

Questions Answered by Slideset

- **What NLG architectures can be distinguished? What are their pros and cons? Which one is most widely used, and why?**
- **What tasks are identified for an NLG system and how can they be grouped into modules?**
- **What is aggregation and why is it needed in NLG?**
- **What must an NLG system reason about when generating a referring expression for a concept „ship376“?**
- **Which NLG tasks are interfering in the following example, and what went wrong during the generation?**
  - The next ship is the MS-Lilly. It, which has a snack bar, leaves at 10 am.
- **What is the role of basic level categories (BLC) when lexicalizing a concept? Give an example of lexical choice using BLC.**
- **What evaluation methodologies are available?**
- **When is the use of gold standards useful for NLG evaluation? Why does it not work in general?**

Source: Stephan Busemann