Natural Language Generation

Scott Farrar
CLMA, University of Washington
farrar@u.washington.edu

March 3, 2010

Today's lecture

1. Demos
2. Basics of NLG
   - NLG concepts
   - Issues in NLG
   - NLG subtasks
3. Architecture of NLG systems
   - Two-step architecture
   - Three-step architecture
4. Hw7

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Personals

Fun, loving woman looking to be a seat warmer on occasion this spring and summer. You be over 45, not married, and not a gang type biker. Want someone safe, sane and responsible. I like Harley's or a decent rice grinder. No crotch rockets. Me: 5'6" HWP 120lb. love to ride, love the sun, love life, love to laugh. I really like men with beards/mustaches/goatees. :o)

Personals

Fun, loving woman looking to be a seat warmer on occasion this spring and summer. You be over 45, not married, and not a gang type biker. Want someone safe, sane and responsible. I like Harley's or a decent rice grinder. No crotch rockets. Me: 5'6" HWP 120lb. love to ride, love the sun, love life, love to laugh. I really like men with beards/mustaches/goatees. :o)

I am educated, stable, employed, 6ft tall, 200 lbs, brown eyes and brown hair. I work long hours at times, im sensitive at times, like good music, food, and conversations. I am a good friend, a good person, and easy to talk to. I'm looking for a slender asian girl who is educated and motivated/passionate. I like a girl who is educated. If you are out there, I would like to get to know you. Your pic gets mine.

NLG demo for pollen count[1]

Area	Pollen Value
North	1
North West	1
Central	1
North East	1
South West	1
South East	1

Grass pollen levels for Saturday remain at the low levels of recent weeks with values of 1 across the whole country. Therefore only those most sensitive to pollen will be affected.

[1] http://www.aclweb.org/aclwiki/index.php?title=Online_NLG_demos

NLG demo for pollen count

Area	Pollen Value
North	1
North West	1
Central	1
North East	1
South West	1
South East	10

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

NLG demo for pollen count

Area	Pollen Value
North	1
North West	1
Central	1
North East	1
South West	1
South East	10

Grass pollen levels for Saturday remain at the low levels of recent weeks with values of 1 across most parts of the country. However, in South Eastern areas, pollen levels will be very high with values of 10.

NLG demo for pollen count

Area	Pollen Value
North	1
North West	1
Central	1
North East	1
South West	1
South East	10

Grass pollen levels for Saturday remain at the low levels of recent weeks with values of 1 across most parts of the country. However, in South Eastern areas, pollen levels will be **very high** with values of 10.

NLG demo for pollen count

Area	Pollen Value
North	10
North West	10
Central	10
North East	10
South West	10
South East	10

NLG demo for pollen count

Area	Pollen Value
North	10
North West	10
Central	10
North East	10
South West	10
South East	10

Grass pollen levels for Saturday have increased from the low levels of recent weeks with values of 10 across the whole country. Therefore making Saturday a particularly unpleasant day for hay fever sufferers.

NLG demo for pollen count

Area	Pollen Value
North	10
North West	10
Central	10
North East	10
South West	10
South East	10

Grass pollen levels for Saturday have increased from the low levels of recent weeks with values of 10 across the **whole country**. Therefore making Saturday a particularly unpleasant day for hay fever sufferers.

Today's lecture

1. Demos
2. Basics of NLG
   - NLG concepts
   - Issues in NLG
   - NLG subtasks
3. Architecture of NLG systems
   - Two-step architecture
   - Three-step architecture
4. Hw7

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

What is NLG?

Definition

Natural language generation is a CL subfield with the aim of producing meaningful, grammatical utterances in natural language from some *non-linguistic input*.

The NLG process is based on some **communicative goal** (e.g., refute, describe, agree), and according to some larger discourse plan.

Specific Goals

Goal: *describePerson SAM*

```
(subclass Family Collection)
(subclass Human LivingThing)
(subclass MaleHuman Human)
(instance Sam MaleHuman)

(inFamily Sam f1)
(familyName f1 Smith)
(age Sam 32)
(maritalStatus Sam single)
(know Sam Jill)
```

Specific Goals

Goal: *describePerson SAM*

```
(subclass Family Collection)
(subclass Human LivingThing)
(subclass MaleHuman Human)
(instance Sam MaleHuman)

(inFamily Sam f1)
(familyName f1 Smith)
(age Sam 32)
(maritalStatus Sam single)
(know Sam Jill)
```

Sam Smith is 32 years old and is single.

Specific Goals

Goal: *refute* $PROP45$
given: $like(BILL, SALLY)$

Specific Goals

Goal: *refute PROP*45
given: *like*(*BILL*, *SALLY*)

- *It's not true that Bill likes Sally.*
- *Bill doen't like Sally.*
- *Bill likes Sally, right!*

Specific Goals

Goal: $(compare\ WA,\ \ VA)$
Given: $loc(WA, NORTHWEST) \wedge loc(VA, EASTCOAST)$

Specific Goals

Goal: $(compare\ WA,\ VA)$
Given: $loc(WA, NORTHWEST) \wedge loc(VA, EASTCOAST)$

- *Washington State is located in the northwest,* **while** *Virginia is on the east coast.*

Specific Goals

Goal: (*compare WA, VA*)
Given: $loc(WA, NORTHWEST) \wedge loc(VA, EASTCOAST)$

- *Washington State is located in the northwest,* **while** *Virginia is on the east coast.*
- *Washington State is located in the northwest* **and** *Virginia on the east coast.*

Some generation applications

Some generation applications

- The output component of a machine translation system

Some generation applications

- The output component of a machine translation system
- Interface to a database system

Some generation applications

- The output component of a machine translation system
- Interface to a database system
- Interface to an expert system (math tutor)

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Some generation applications

- The output component of a machine translation system
- Interface to a database system
- Interface to an expert system (math tutor)
- Autogeneration of help pages for a software system

Some generation applications

- The output component of a machine translation system
- Interface to a database system
- Interface to an expert system (math tutor)
- Autogeneration of help pages for a software system
- Robot-human communication

Some generation applications

- The output component of a machine translation system
- Interface to a database system
- Interface to an expert system (math tutor)
- Autogeneration of help pages for a software system
- Robot-human communication
- Data summarization (stock/weather reports)

Some generation applications

- The output component of a machine translation system
- Interface to a database system
- Interface to an expert system (math tutor)
- Autogeneration of help pages for a software system
- Robot-human communication
- Data summarization (stock/weather reports)
- Text summarization

Comparison to NLU

Similarities

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Comparison to NLU

Similarities

- both utilize a lexicon, grammar, and knowledge representation

Comparison to NLU

Similarities

- both utilize a lexicon, grammar, and knowledge representation
- have same "endpoints" (internal computational representation, natural language utterances).

Comparison to NLU

Similarities

- both utilize a lexicon, grammar, and knowledge representation
- have same “endpoints” (internal computational representation, natural language utterances).
- both symbolic and stochastic (corpus-based) methods apply

Comparison to NLU

Differences

Comparison to NLU

Differences

- NLU is about hypothesis management, where as NLG is about **choice**; When producing NL utterances from non-linguistic material, everything bit of information to be encoded in the output has to be chosen along the way.

Comparison to NLU

Differences

- NLU is about hypothesis management, where as NLG is about **choice**; When producing NL utterances from non-linguistic material, everything bit of information to be encoded in the output has to be chosen along the way.
- parsing, or most shallow NLU tasks, ignore certain aspects of meaning; NLG utilizes **elaborated meaning representations**

Comparison to NLU

Differences

- NLU is about hypothesis management, where as NLG is about **choice**; When producing NL utterances from non-linguistic material, everything bit of information to be encoded in the output has to be chosen along the way.
- parsing, or most shallow NLU tasks, ignore certain aspects of meaning; NLG utilizes **elaborated meaning representations**
- research in NLG has focused more on **texts** than has NLU

Texts

For NLG, we often need to deal with structures larger than the sentence.

Texts

For NLG, we often need to deal with structures larger than the sentence.

Definition

Text: A unit of language larger than the individual clause (spoken or written). This is referred to as a *discourse* in J&M.

Texts

For NLG, we often need to deal with structures larger than the sentence.

Definition

Text: A unit of language larger than the individual clause (spoken or written). This is referred to as a *discourse* in J&M.

- a phone conversation

Texts

For NLG, we often need to deal with structures larger than the sentence.

Definition

Text: A unit of language larger than the individual clause (spoken or written). This is referred to as a *discourse* in J&M.

- a phone conversation
- a recipe

Texts

For NLG, we often need to deal with structures larger than the sentence.

Definition

Text: A unit of language larger than the individual clause (spoken or written). This is referred to as a *discourse* in J&M.

- a phone conversation
- a recipe
- a weather report

Texts

For NLG, we often need to deal with structures larger than the sentence.

Definition

Text: A unit of language larger than the individual clause (spoken or written). This is referred to as a *discourse* in J&M.

- a phone conversation
- a recipe
- a weather report
- a paragraph in War and Peace

Texts

For NLG, we often need to deal with structures larger than the sentence.

Definition

Text: A unit of language larger than the individual clause (spoken or written). This is referred to as a *discourse* in J&M.

- a phone conversation
- a recipe
- a weather report
- a paragraph in War and Peace
- an entire novel

Text coherence

Such texts exhibit a certain structure and we can speak of their well-formedness, just like any other unit of language.

Text coherence

Such texts exhibit a certain structure and we can speak of their well-formedness, just like any other unit of language.

Definition

A **coherent** text is one whose parts are interrelated in meaningful way. An incoherent text is one whose parts do not bind together in a naturalistic manner.

Text coherence

Such texts exhibit a certain structure and we can speak of their well-formedness, just like any other unit of language.

Definition

A **coherent** text is one whose parts are interrelated in meaningful way. An incoherent text is one whose parts do not bind together in a naturalistic manner.

Text coherence

Such texts exhibit a certain structure and we can speak of their well-formedness, just like any other unit of language.

Definition

A **coherent** text is one whose parts are interrelated in meaningful way. An incoherent text is one whose parts do not bind together in a naturalistic manner.

- Pronominalization

Text coherence

Such texts exhibit a certain structure and we can speak of their well-formedness, just like any other unit of language.

Definition

A **coherent** text is one whose parts are interrelated in meaningful way. An incoherent text is one whose parts do not bind together in a naturalistic manner.

- Pronominalization
- Temporal expressions

Text coherence: pronominal expressions

James Riddle "Jimmy" Hoffa (born February 14, 1913, disappeared July 30, 1975), was an American labor leader. As the president of the International Brotherhood of Teamsters from the mid-1950s to the mid-1960s, Hoffa wielded considerable influence. After he was convicted of attempted bribery of a grand juror, he served nearly a decade in prison. He is also well-known in popular culture for the mysterious circumstances surrounding his unexplained disappearance and presumed death. His son James P. Hoffa is the current president of the Teamsters.

Text coherence: pronominal expressions

James Riddle "Jimmy" Hoffa (born February 14, 1913, disappeared July 30, 1975), was an American labor leader. As the president of the International Brotherhood of Teamsters from the mid-1950s to the mid-1960s, Hoffa wielded considerable influence. After he was convicted of attempted bribery of a grand juror, he served nearly a decade in prison. He is also well-known in popular culture for the mysterious circumstances surrounding his unexplained disappearance and presumed death. His son James P. Hoffa is the current president of the Teamsters.

Text coherence: temporal expressions

During Buck's time in jail, Clyde had been the driver in a store robbery. The wife of the murder victim, when shown photos, picked Clyde as one of the shooters. On August 5, 1932, while Bonnie was visiting her mother, Clyde and two associates were drinking alcohol at a dance in Stringtown, Oklahoma (illegal under Prohibition). When they were approached by sheriff C.G. Maxwell and his deputy, Clyde opened fire, killing deputy Eugene C. Moore. That was the first killing of a lawman by what was later known as the Barrow Gang, a total which would eventually amount to nine slain officers.

Text coherence: temporal expressions

During Buck's time in jail, Clyde had been the driver in a store robbery. The wife of the murder victim, when shown photos, picked Clyde as one of the shooters. On August 5, 1932, while Bonnie was visiting her mother, Clyde and two associates were drinking alcohol at a dance in Stringtown, Oklahoma (illegal under Prohibition). When they were approached by sheriff C.G. Maxwell and his deputy, Clyde opened fire, killing deputy Eugene C. Moore. That was the first killing of a lawman by what was later known as the Barrow Gang, a total which would eventually amount to nine slain officers.

Evaluation

Problem: evaluation of NLG systems is always a great challenge (no gold standard)

Scott Farrar
CLMA, University of Washington far-rar@u.washington.ed

Evaluation

Problem: evaluation of NLG systems is always a great challenge (no gold standard)

Evaluation techniques:

Evaluation

Problem: evaluation of NLG systems is always a great challenge (no gold standard)

Evaluation techniques:

- Turing Test (subjective)

Evaluation

Problem: evaluation of NLG systems is always a great challenge (no gold standard)

Evaluation techniques:

- Turing Test (subjective)
- task-oriented (expensive)

Evaluation

Problem: evaluation of NLG systems is always a great challenge (no gold standard)

Evaluation techniques:

- Turing Test (subjective)
- task-oriented (expensive)
- statistical comparison with real texts (untrustworthy)

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Turing Test for evaluation of an NLG system

5 Indistinguishable from human
4 Most likely human
3 Maybe human or machine
2 Most likely machine
1 Definitely machine

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Grice's Conversational Maxims

Grice's Conversational Maxims

- **Quantity**: Make your contribution as informative as is required

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Grice's Conversational Maxims

- **Quantity**: Make your contribution as informative as is required
- **Quality**: Do not say what you believe is false. Do not say that for which you lack adequate evidence.

Grice's Conversational Maxims

- **Quantity**: Make your contribution as informative as is required
- **Quality**: Do not say what you believe is false. Do not say that for which you lack adequate evidence.
- **Relation**: Be relevant.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Grice's Conversational Maxims

- **Quantity**: Make your contribution as informative as is required
- **Quality**: Do not say what you believe is false. Do not say that for which you lack adequate evidence.
- **Relation**: Be relevant.
- **Manner**: Be perspicuous, avoid obscurity of expression, avoid ambiguity, be brief, be orderly.

Statistical evaluation scenarios

Comparison with humanly produced texts:

Statistical evaluation scenarios

Comparison with humanly produced texts:

- text length, mean length of utterance (MLU)

Statistical evaluation scenarios

Comparison with humanly produced texts:

- text length, mean length of utterance (MLU)
- average number of embedded clauses (and other complex structures)

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Statistical evaluation scenarios

Comparison with humanly produced texts:

- text length, mean length of utterance (MLU)
- average number of embedded clauses (and other complex structures)
- diction (number of word types)

Statistical evaluation scenarios

Comparison with humanly produced texts:

- text length, mean length of utterance (MLU)
- average number of embedded clauses (and other complex structures)
- diction (number of word types)
- number of long distance dependencies

General evaluation criteria

Despite the challenges, we can posit some fundamental criteria for evaluating NLG systems:

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

General evaluation criteria

Despite the challenges, we can posit some fundamental criteria for evaluating NLG systems:

- **content**: does the output contain appropriate/enough information?

General evaluation criteria

Despite the challenges, we can posit some fundamental criteria for evaluating NLG systems:

- **content**: does the output contain appropriate/enough information?
- **organization**: is the discourse structure realistic?

General evaluation criteria

Despite the challenges, we can posit some fundamental criteria for evaluating NLG systems:

- **content**: does the output contain appropriate/enough information?
- **organization**: is the discourse structure realistic?
- **correctness**: are there grammatical or stylistic errors?

General evaluation criteria

Despite the challenges, we can posit some fundamental criteria for evaluating NLG systems:

- **content**: does the output contain appropriate/enough information?
- **organization**: is the discourse structure realistic?
- **correctness**: are there grammatical or stylistic errors?
- **textual flow**: is the language choppy or smooth?

NLG subtask: Non-linguistic

There are several tasks for a full generation system:

NLG subtask: Non-linguistic

There are several tasks for a full generation system:

- **content determination**: task of deciding what information is to be communicated

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

NLG subtask: Non-linguistic

There are several tasks for a full generation system:

- **content determination**: task of deciding what information is to be communicated
- **discourse structuring**: deciding how to package the 'chunks' of content

NLG subtask: Linguistic

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

NLG subtask: Linguistic

- **lexicalization**: determine the particular words and construction types to use

NLG subtask: Linguistic

- **lexicalization**: determine the particular words and construction types to use
- **aggregation**: decide how much information to include in each sentence/phrase

NLG subtask: Linguistic

- **lexicalization**: determine the particular words and construction types to use
- **aggregation**: decide how much information to include in each sentence/phrase
- **referring expression generation**: determine what expressions should be used to refer to the entities in the domain

NLG subtask: Linguistic

- **lexicalization**: determine the particular words and construction types to use
- **aggregation**: decide how much information to include in each sentence/phrase
- **referring expression generation**: determine what expressions should be used to refer to the entities in the domain
- **linguistic realization**: task of converting abstract representations of sentences to real text; linearization.

NLG subtask: Linguistic

- **lexicalization**: determine the particular words and construction types to use
- **aggregation**: decide how much information to include in each sentence/phrase
- **referring expression generation**: determine what expressions should be used to refer to the entities in the domain
- **linguistic realization**: task of converting abstract representations of sentences to real text; linearization.
- **structure realization**: task of converting abstract structures such as paragraphs and sections into markup symbols understood by the document presentation component (e.g., HTML, LaTeX)

NLG subtask: Linguistic

- **lexicalization**: determine the particular words and construction types to use
- **aggregation**: decide how much information to include in each sentence/phrase
- **referring expression generation**: determine what expressions should be used to refer to the entities in the domain
- **linguistic realization**: task of converting abstract representations of sentences to real text; linearization.
- **structure realization**: task of converting abstract structures such as paragraphs and sections into markup symbols understood by the document presentation component (e.g., HTML, LaTeX)

Need to modularize the above tasks appropriately, given the goal of the NLG system and the available resources.

Lexicalization

(1) a. Mary's car

b. the car owned by Mary

Lexicalization

(3) a. Mary's car

b. the car owned by Mary

(4) a. the ship's cargo hold

b. the cargo hold which is part of the ship

Agrregation

(5) a. I am Ron Paul. I am a rogue Republican.

b. My name is Ron Paul and I'm a rogue Republican.

Agrregation

(7) a. I am Ron Paul. I am a rogue Republican.
    b. My name is Ron Paul and I'm a rogue Republican.

(8) a. The course number is ling571.
    b. The course is difficult. The course is open to undergraduates.
    c. Ling571 is difficult, but open to undergraduates.

Agrregation: other examples

Agrregation: other examples

- John's bicycle is red

Agrregation: other examples

- John's bicycle is red
- Mary's bicycle is yellow

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Agrregation: other examples

- John's bicycle is red
- Mary's bicycle is yellow
- Tom's bicycle is blue

Agrregation: other examples

- John's bicycle is red
- Mary's bicycle is yellow
- Tom's bicycle is blue
- Lisa's bicycle is red

Agrregation: other examples

- John's bicycle is red
- Mary's bicycle is yellow
- Tom's bicycle is blue
- Lisa's bicycle is red

becomes:

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Agrregation: other examples

- John's bicycle is red
- Mary's bicycle is yellow
- Tom's bicycle is blue
- Lisa's bicycle is red

  becomes:
- John and Lisa have red bicycles.

Agrregation: other examples

- John's bicycle is red
- Mary's bicycle is yellow
- Tom's bicycle is blue
- Lisa's bicycle is red

becomes:

- John and Lisa have red bicycles.
- Tom's and Mary's bicycles are blue and yellow respectively.

Lexical aggregation

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Lexical aggregation

- Ericsson made profit in 2004

Lexical aggregation

- Ericsson made profit in 2004
- Nokia made profit in 2004

Lexical aggregation

- Ericsson made profit in 2004
- Nokia made profit in 2004
- Siemens made profit in 2004

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Lexical aggregation

- Ericsson made profit in 2004
- Nokia made profit in 2004
- Siemens made profit in 2004
- ATT made profit in 2004

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Lexical aggregation

- Ericsson made profit in 2004
- Nokia made profit in 2004
- Siemens made profit in 2004
- ATT made profit in 2004

becomes:

Lexical aggregation

- Ericsson made profit in 2004
- Nokia made profit in 2004
- Siemens made profit in 2004
- ATT made profit in 2004

  becomes:
- All telecom companies in the world except Alcatel made profit in 2004

Referring Expressions

(9) a. I know that guy.

b. I know Noam Chomsky.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Referring Expressions

(11) a. I know that guy.

b. I know Noam Chomsky.

(12) a. It's a long-legged, hairy one.

b. The European wolf spider is a long-legged, hairy spider.

Today's lecture

1. Demos
2. Basics of NLG
   - NLG concepts
   - Issues in NLG
   - NLG subtasks
3. Architecture of NLG systems
   - Two-step architecture
   - Three-step architecture
4. Hw7

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Ways to generate

A generation system can be devised for individual utterances or whole texts.

Ways to generate

A generation system can be devised for individual utterances or whole texts.

- canned text (easy, brittle)

Ways to generate

A generation system can be devised for individual utterances or whole texts.

- canned text (easy, brittle)
- template-based generation (easy, more flexible, ad hoc)

Ways to generate

A generation system can be devised for individual utterances or whole texts.

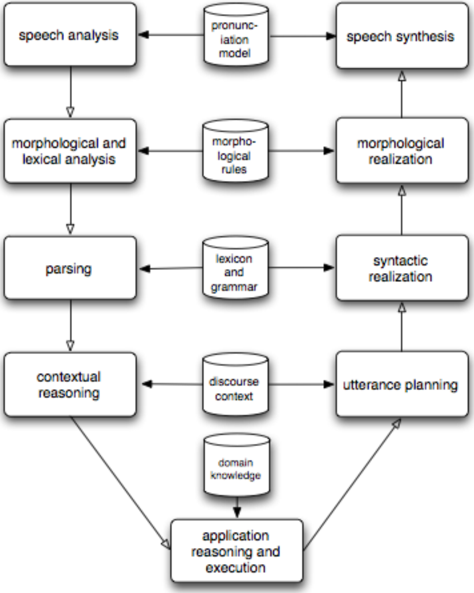
- canned text (easy, brittle)
- template-based generation (easy, more flexible, ad hoc)
- feature collection and linearization (very hard, flexible, theory driven)

Templates from ELIZA program

```
  ( "Perhaps you don't want to %1.",
    "Do you want to be able to %1?",
    "If you could %1, would you?")),
...

  ( "Why do you think I am %1?",
    "Does it please you to think that I'm %1?",
    "Perhaps you would like me to be %1.",
    "Perhaps you're really talking about yourself?")),
```

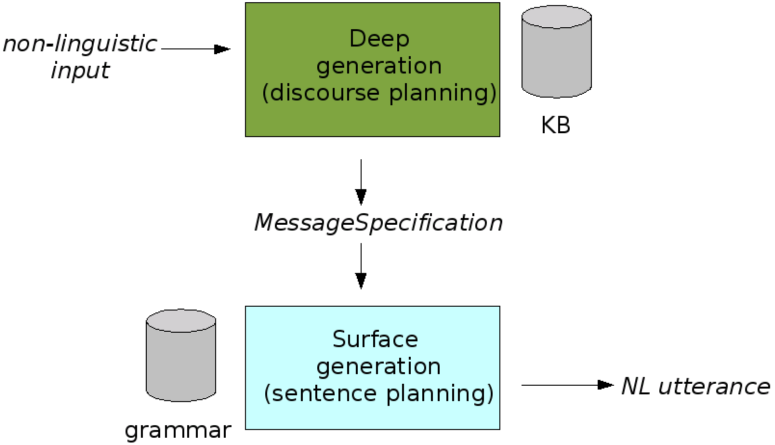
NLP reference architecture



NLG architectures paradigms

- Two common architectures (two- and three- step systems)
- Based on splitting up the NLG subtasks (e.g., textplanner, lexical aggregation).
- Based on how complex the system needs to be.

NLG architecture: Two-step



Two-step: Main components

1. **Deep generation**: determines and structures content of resulting text; insert words (lexicalization); map message to linguistic structure
2. **Surface generation**: fill in grammatical details, some lexicalization

Two-step: Example systems

Used in simpler, early systems:

- FOG System: generates weather reports from numerical weather simulations.
- Peba: generates taxonomic descriptions or comparisons of animals from a knowledge base of animal facts.

Problems in modularization and control over choice. In general, it's really hard to strictly keep the **non-linguistic** choices separate from the **linguistic** choices.

Example from FOG System: weather text

Winds southwest 15 diminishing to light, late this evening. Winds light Friday. Showers ending late this evening. Fog. Outlook for Saturday...light winds.

PEBA system: knowledge base sample

```
(hasprop Echidna (linean-classification Family))
(distinguishing-characteristic Echidna Monotreme
    (body-covering sharp-spines))
(hasprop Echidna (nose long-snout))
(hasprop Echidna (social-living-status lives-by-itself))
(hasprop Echidna (diet eats-ants-termites-earthworms))
(hasprop Echidna (activity-time active-at-dusk-dawn))
(hasprop Echidna (colouring browny-black-coat-paler-coloured-spines))
(hasprop Echidna (lifespan lifespan-50-years-captivity)
```

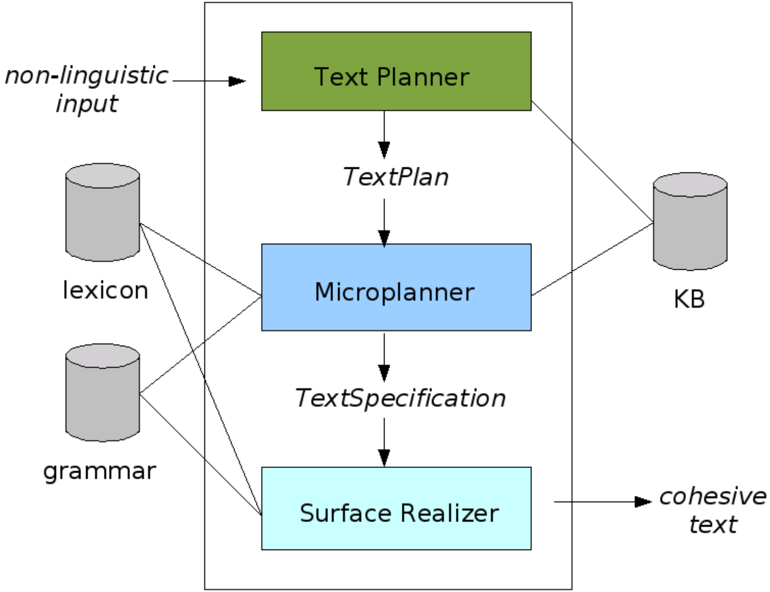
PEBA system: knowledge base sample

```
(hasprop Echidna (linean-classification Family))
(distinguishing-characteristic Echidna Monotreme
    (body-covering sharp-spines))
(hasprop Echidna (nose long-snout))
(hasprop Echidna (social-living-status lives-by-itself))
(hasprop Echidna (diet eats-ants-termites-earthworms))
(hasprop Echidna (activity-time active-at-dusk-dawn))
(hasprop Echidna (colouring browny-black-coat-paler-coloured-spines))
(hasprop Echidna (lifespan lifespan-50-years-captivity)
```

text

The small spined monotreme belongs to the Echidna Family. Its nose is a long snout. The Echidna lives by itself.

NLG architecture: Three-step



Three-step: Example systems

Better systems

More flexible, more modular. More control over the output.

- WEATHERREPORTER: more complex, cohesive weather reports
- KNIGHT System: a biology explanation system, from knowledge base to explanatory paragraphs

More flexible, more modular. More control over the output.

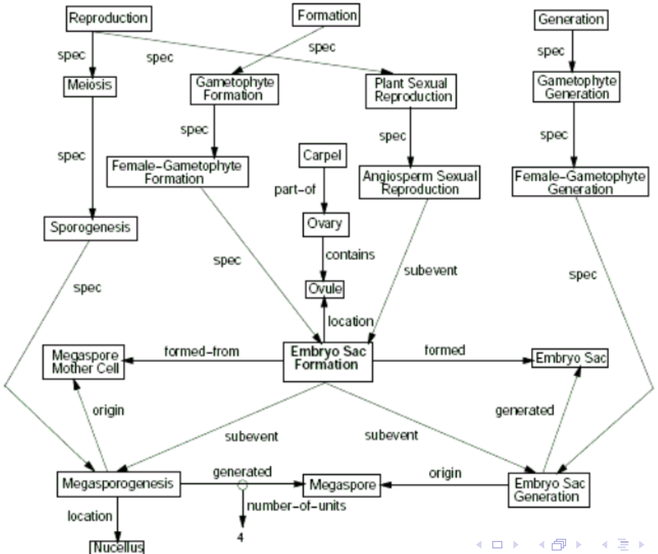
Three-step: Main components

Module	Content task	Structure task
Text Planner	content determination	rhetorical structuring
Microplanner	lexicalization; referring expression generation	aggregation
Surface Realizer	linguistic realization	structure realization

WeatherReporter: weather text

The month was rather dry with only three days of rain in the middle of the month. The total for the year so far is very depleted again.

KNIGHT System KB



KNIGHT System Output

Embryo sac formation is a kind of female gametophyte formation. During embryo sac formation, the embryo sac is formed from the megaspore mother cell. Embryo sac formation occurs in the ovule.

Today's lecture

1. Demos
2. Basics of NLG
   - NLG concepts
   - Issues in NLG
   - NLG subtasks
3. Architecture of NLG systems
   - Two-step architecture
   - Three-step architecture
4. Hw7

Hw7, main points

Tasks

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Hw7, main points

Tasks

1. Create code to process an input text specification (XML)

Hw7, main points

Tasks

1. Create code to process an input text specification (XML)
2. Build a microplanner (in language of your choice)

Hw7, main points

Tasks

1. Create code to process an input text specification (XML)
2. Build a microplanner (in language of your choice)
3. Use the SimpleNLG (v.4) surface realizer (Java)