Natural Language Generation

CSCI 544 – Fall 2016
10/19/2016

Kallirroi Georgila

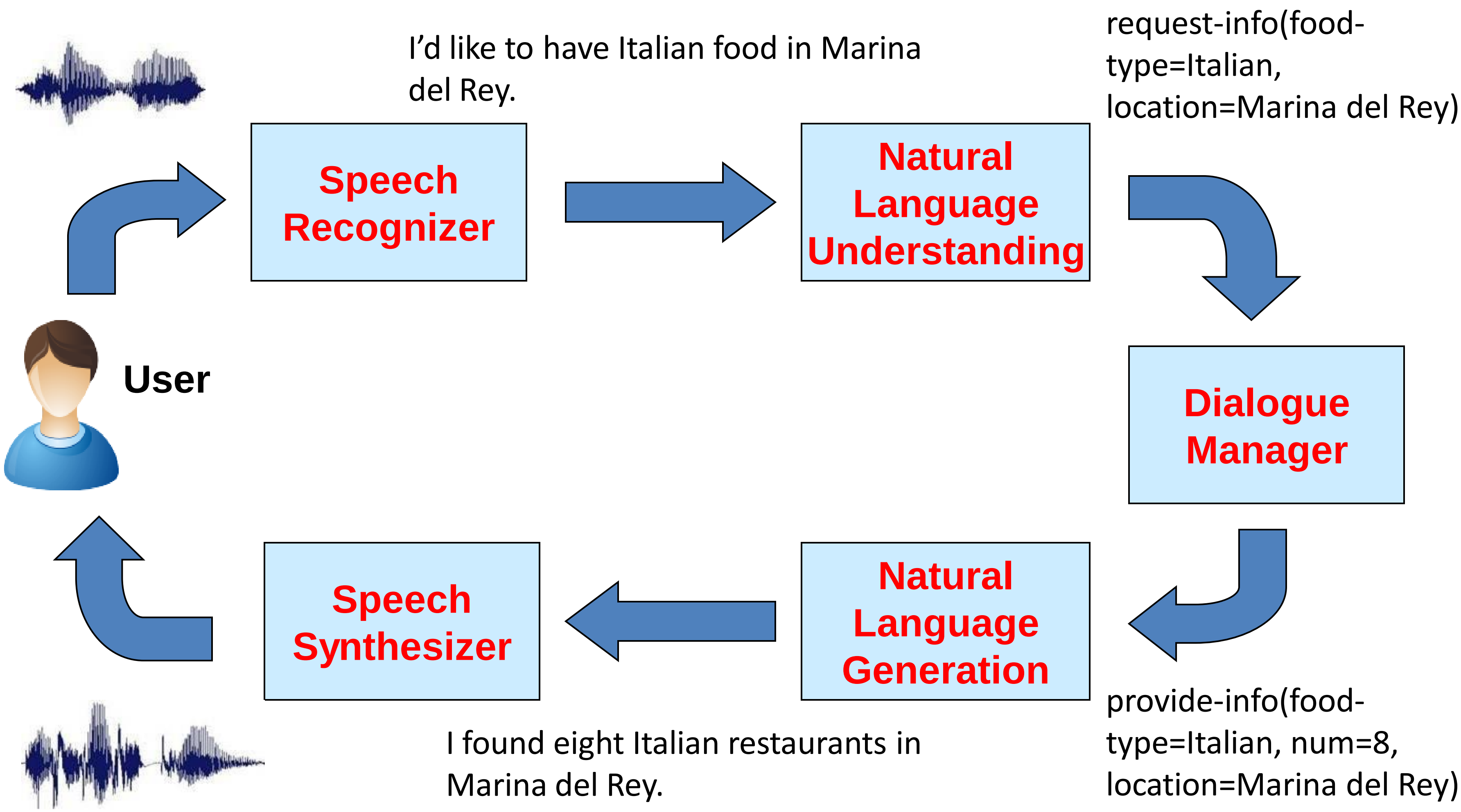
What is natural language generation?

- Natural language generation (NLG) means producing understandable and appropriate text in English or other human languages
- Input: data
  - raw, e.g., numbers in tables
  - analyzed, e.g., semantic representations
- Output: text (reports, summaries, help messages, dialogue system prompts, etc.)

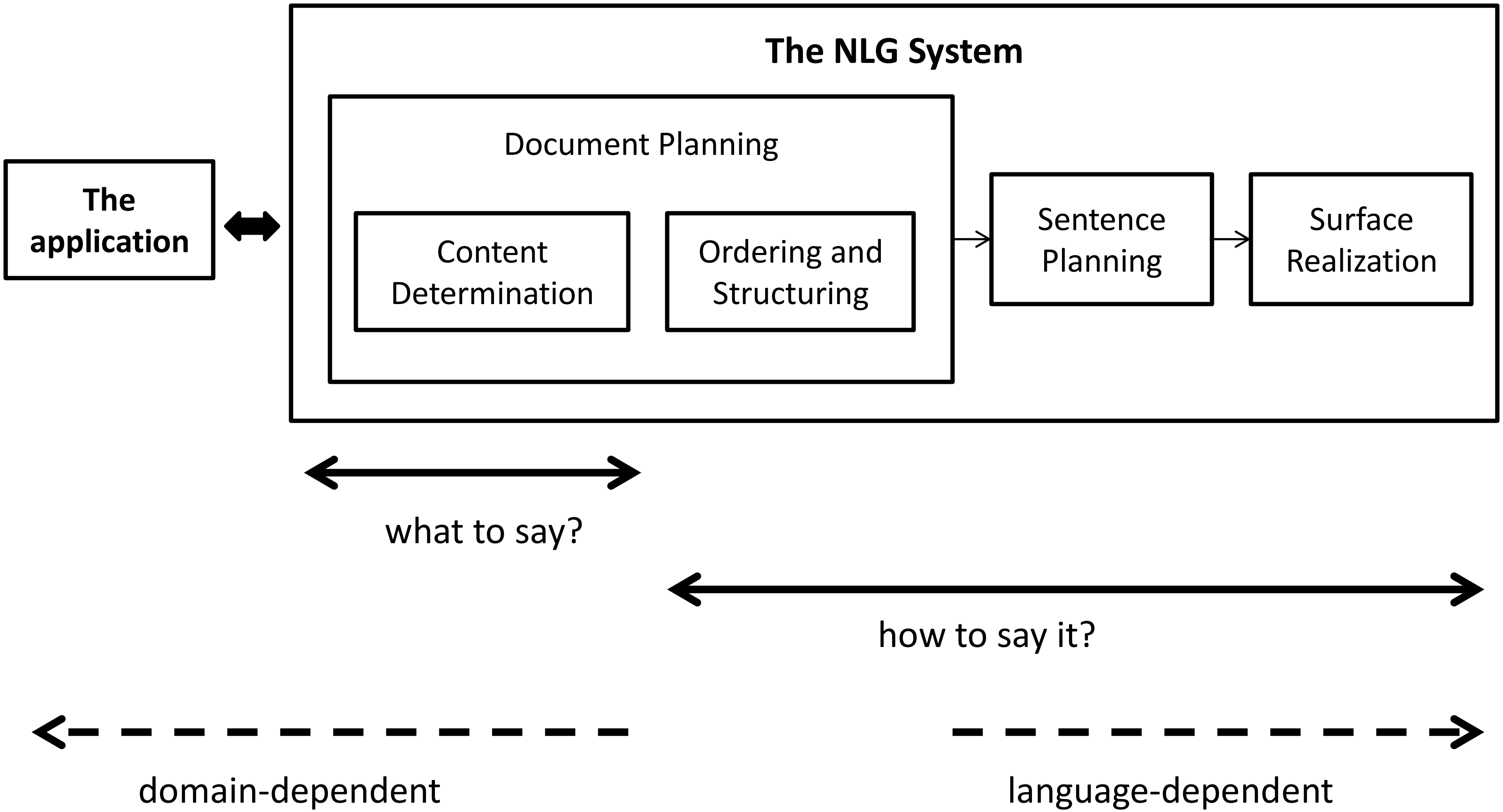
Applications of NLG

- Generate weather reports from numerical weather-related measurements
- Summarize clinical data from sensor data and records of actions/observations by medical staff
- Generate dialogue system prompts based on the user's input and the dialogue context
- Help people improve their writing skills
- …

NLG as part of a spoken dialogue system



Reiter and Dale (2000) NLG pipeline



NLG stages

- Document planning (or macroplanning)
  – Decide on content and structure of text
- Sentence planning (or microplanning)
  – Decide how to linguistically express text, i.e., which words, sentences, etc. to use
- Surface realization
  – Produce text conforming to rules of grammar

Document planning

- Decide on content and structure of text
- Input: initial data
- Output: document plan
- Content determination: determine what information to communicate
  - what is important?
  - what is easy to say?
- Structuring and ordering: organize the content as a text
  - what order do I say things in?
  - how do I make sure that the resulting text is coherent?
- Document planning is heavily domain-dependent

Why is content determination hard?

- Hard to develop reusable approaches
  - Multiple domains
  - Multiple input data formats
- Constraints
  - May not be possible to fit all the required information in one page or paragraph or tweet or sentence, etc.

Why is content determination hard? (cont.)

- Hard to choose between alternatives that are equivalent in application but differ significantly in terms of language, e.g., the problem of logical form equivalence (Schieber, 1993)

(*Item*)	String *Canonical Logical Form*
(i)	John threw a large red ball. $\exists x.throw(j,x) \wedge large(x) \wedge red(x) \wedge ball(x)$
(ii)	John threw a red ball that is large. $\exists x.throw(j,x) \wedge red(x) \wedge ball(x) \wedge large(x)$
(iii)	John threw a large ball that is red. $\exists x.throw(j,x) \wedge large(x) \wedge ball(x) \wedge red(x)$

(*Item*)	String *Canonical Logical Form*
(iv)	Clapton was the leader of Derek and the Dominos. $the(x, leader\text{-}of(dd,x), c = x)$
(v)	The leader of Derek and the Dominos was Clapton. $the(x, leader\text{-}of(dd,x), x = c)$
(vi)	Clapton led Derek and the Dominos. $led(c, dd)$

Top-down vs. bottom up content determination

- Top-down (goal-driven, backward): find content to support a particular text type
  - Good when there are strong conventions about what a text should be like
  - Satisfy communicative goals
- Bottom-up (data-driven, forward): look at what is available and how text can be generated from it
  - Good when there needs to be variation in the form of text
  - Pick the most important things

Content determination should take context into account

- Targeted genre: report, narrative, etc.
- Targeted audience: expert, lay person, etc.
- Communicative goal: advice, persuasion, information providing, etc.
- User profile: user preferences, previous knowledge about the user, etc.

Schema-based content determination

- Based on “messages”, i.e., predefined data structures which
  - Correspond to informational elements in the text
  - Collect together underlying data in ways that are convenient

Schema-based content determination in Weather Reporter

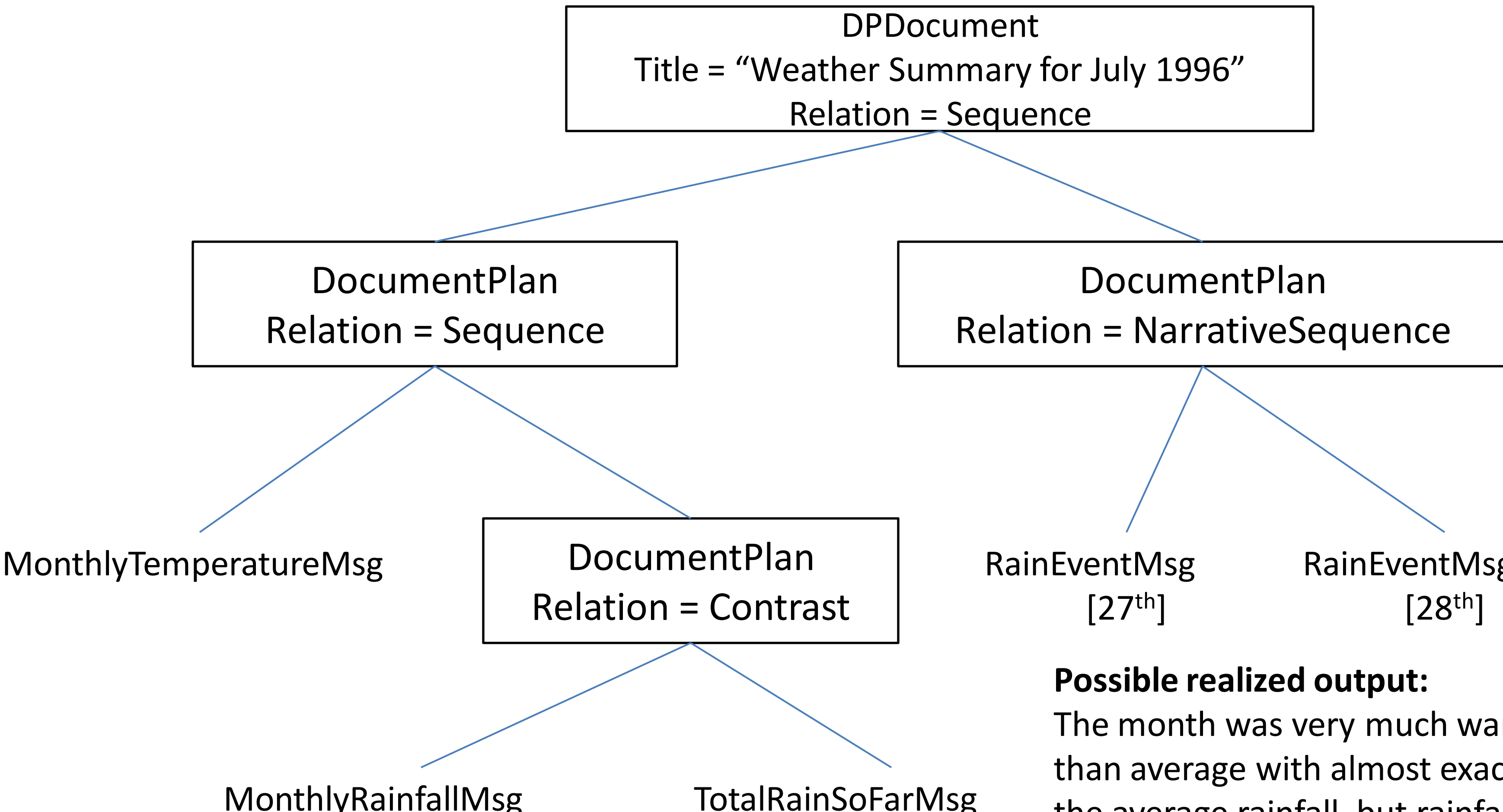
- Routine messages: always included
  - MonthlyRainfallMsg
  - MonthlyTemperatureMsg
  - TotalRainSoFarMsg
  - MonthlyRainyDaysMsg
- Significant event messages: only constructed if the data warrants it: e.g., if rain occurs on more than a specified number of days in a row
  - RainEventMsg
  - RainSpellMsg
  - TemperatureEventMsg
  - TemperatureSpellMsg

Schema-based content determination in Weather Reporter (cont.)

- type: MonthlyTemperatureMsg
- period:
  - month: 07
  - year: 1996
- temperature:
  - type: RelativeVariation
  - magnitude:
    - unit: degreesCentigrade
    - number: 2
  - direction: +

- type: RainEventMsg
- period:
  - month: 07
  - year: 1996
- date:
  - day: 27
  - month: 07
  - year: 1996
- rainType: heavy

Schema-based document planning in Weather Reporter



Possible realized output:
The month was very much warmer than average with almost exactly the average rainfall, but rainfall for the year is still very depleted. Heavy rain fell on the 27th and 28th.

A simple surface realizer in Weather Reporter

Sets of sentence templates, e.g.,

For the MonthlyTemperatureMsg:

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

What would happen if we realized each sentence of a document plan?

Result:

The month was cooler than average.

The month was drier than average.

There was the average number of rainy days.

The total rain for the year so far is well below average.

There was rain every day for 8 days from 11^{th} to 18^{th}.

Rainfall amounts were mostly small.

What we want is something like:

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

Other types of document planning

- Automated planning
  - Find sequence of actions to satisfy a goal
  - Domain knowledge is modeled using planning languages (STRIPS, ADL, PDDL)
- Automated reasoning
  - Start from a Knowledge Base encoding knowledge about the domain
  - Reasoning requires explicit symbolic representations of how data are connected (rules, ontologies, etc.)
- Statistical methods
  - Construct a model that given inputs assigns probabilities to outputs
  - Model can be trained from corpora of human-authored texts aligned with contents

Sentence planning

- Decide how to linguistically express text, i.e., which words, sentences, etc. to use
- Input: document plan
- Output: sentence plan
- Aggregation: How information should be distributed across sentences and paragraphs
- Lexicalization: Which words and linguistic structures should we use?
- Reference: How should the text refer to objects and entities

Aggregation

- How information should be distributed across sentences and paragraphs
- Without aggregation:
  - Heavy rain fell on the 27th.
  - Heavy rain fell on the 28th.
- With aggregation via simple conjuction:
  - Heavy rain fell on the 27th and heavy rain fell on the 28th.
- With aggregation via ellipsis:
  - Heavy rain fell on the 27th and [] on the 28th.
- With aggregation via set introduction:
  - Heavy rain fell on the 27th and 28th.

Note: For simplicity, examples are shown using realized sentences, but in reality aggregation is applied to more abstract representations, e.g., messages

Lexicalization

- Which words and linguistic structures should we use?
- If several lexicalizations are possible, consider:
  - User knowledge and preferences
  - Consistency with previous usage
  - Pragmatics: emphasis, level of formality, personality, …
  - Interaction with other aspects of microplanning
- Example:
  - rainfall was very poor
  - a much worse than average rainfall
  - much drier than average

Note: For simplicity, examples are shown using realized sentences, but in reality lexicalization is applied to more abstract representations, e.g., messages

Reference

- How should the text refer to objects and entities?
- Initial introduction of an object vs. subsequent references to an already salient object
- Referring to months:
  - June 1999
  - June
  - the month
  - next June
- Referring to temporal intervals:
  - 8 days starting from the 11^{th}
  - from the 11^{th} to the 18^{th}

Note: For simplicity, examples are shown using realized sentences, but in reality reference is applied to more abstract representations, e.g., messages

What would happen if we realized a sentence plan?

Many different results are possible:

Result 1:

The month was cooler than average. It was also drier than average, even though there was an average number of rainy days this month. Although there was rain every day for 8 days from the 11^{th} to the 18^{th}, rainfall amounts were mostly small. The total rain for the year so far is well below average.

Result 2:

The month was cooler and drier than average, with the average number of rainy days. The total rain for the year so far is well below average. Even though there was rain every day from the 11^{th} to the 18^{th}, rainfall amounts were mostly small.

Surface realization

- Produce text conforming to rules of grammar, i.e., convert text specifications (e.g., in a sentence plan words are unordered and not inflected) into actual text
- Input: sentence plan
- Output: natural language text
- Hide the peculiarities of English (or another language) from the rest of the NLG system

Types of surface realization

- We can use grammars to transform sentence plans to natural language
- Sometimes the output of these grammars can be stilted or robotic or even wrong, or there could be multiple possible outputs
- Statistical models (n-grams) can help pick the most natural output (that was seen in a corpus)

Example of over-generating and re-ranking – Combination of grammar and n-grams

A simple grammar:	Outputs of the grammar:	
S $\rightarrow$ NP VP	the child cooked the child	WRONG
NP $\rightarrow$ DET N	the child cooked the cake	WRONG
VP $\rightarrow$ V NP	the child baked the child	WRONG
DET $\rightarrow$ the	the child baked the cake	CORRECT
N $\rightarrow$ child	the cake cooked the child	WRONG
N $\rightarrow$ cake	the cake cooked the cake	WRONG
V $\rightarrow$ cooked	the cake baked the child	WRONG
V $\rightarrow$ baked	the cake baked the cake	WRONG

By applying a language model (n-grams) trained from a corpus to the outputs of the grammar, we can prune the ones that do not make sense, e.g., the cake baked the child, the child cooked the cake (we bake cakes, we do not cook cakes)

NLG and user modeling (summarize and refine approach)

- Standard information presentation approach vs. summarize and refine approach (Paksima, Georgila, and Moore, SIGDIAL 2009)

- Example of standard information presentation in a restaurant recommendation domain:

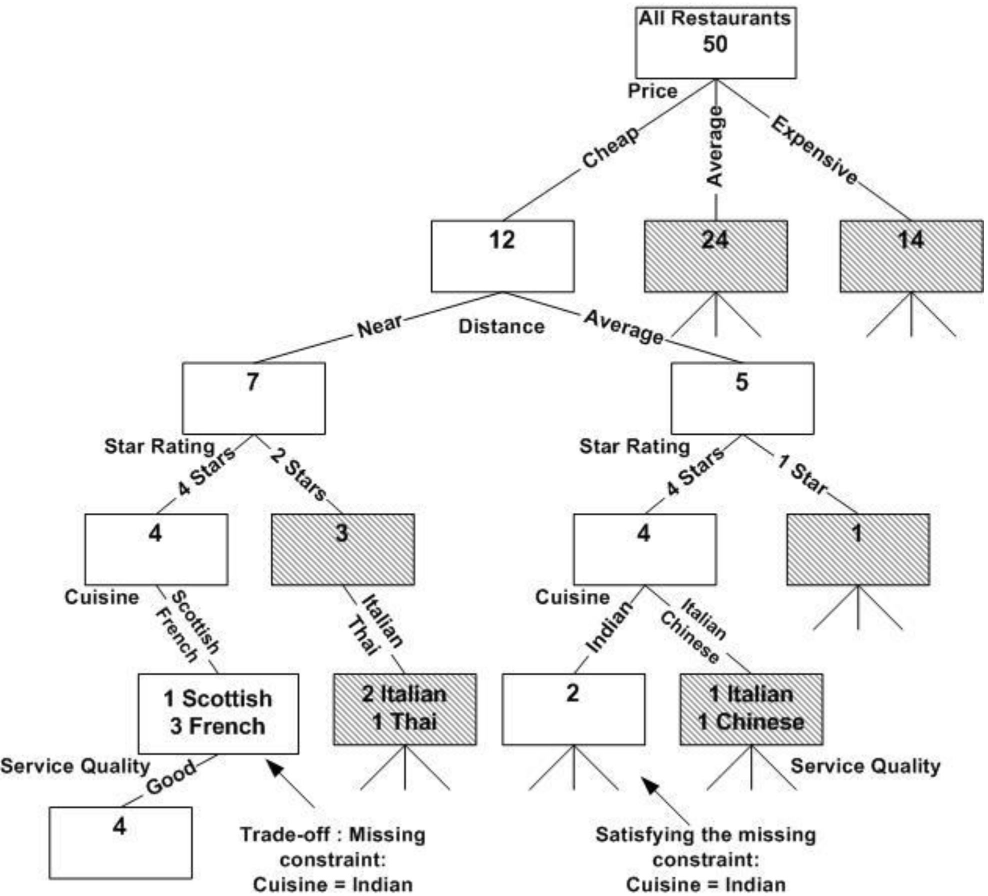
  There are six restaurant options matching your query.

  Number 1: Voujon offers a bright, airy and contemporary dining area, with simple floral displays and leather seating. It serves Indian cuisine. It is located in the city centre. The average price is £24 per person.

  Number 2: Saffrani's decor is modern, the dining room wee, though the menu is enormous, and the atmosphere charming. It offers new Indian dishes never before seen in Edinburgh. It serves Indian, seafood cuisine. It is located in the city centre. The average price is £28 per person.

  Number 3: Britannia Spice . . .

Option tree, and summarize and refine output



U: Tell me about the restaurants you have.

S: I found 12 restaurants which are inexpensive. Only 7 of them are near the city centre, out of which only 4 have a high star rating. I also have information about cuisine and service quality.

U: Tell me about the cuisine.

S: There is only 1 restaurant that serves Scottish food, and 3 restaurants that serve French food. They all provide good service quality. All of them are cheap, are near the city centre, and have a high star rating. For a restaurant that serves Indian cuisine you would have to go farther from the city centre.

U: Okay, tell me about the Scottish one.

S: The Jam House’s menu includes Scottish dishes sparked with international flavour, with a decent choice of vegetarian dishes. It is located in the centre. The average price is £22 per person.

U: What about Indian restaurants?

S: I found 2 restaurants that serve Indian food. Both of them are cheap, within average distance from the city centre, and have a high star rating. Only one of them provides good service quality.

U: Okay, tell me about this one.

S: The Raj on the Shore’s theme extends from the soldier figurines behind the exotic blue bar to the predominantly Bangladeshi, Indian menu. It serves Indian and Bangladeshi cuisine. It is located in the north. The average price is £20 per person.

Types of NLG evaluation

- Task performance
  - Measure whether the system achieves its communicative goal (e.g., help the user learn, entertain the user, etc.)
- Human ratings
  - Readability (linguistic quality)
  - Accuracy (content quality)
  - Usefulness
- Metric (gold standard)
  - Create a gold-standard reference text (usually human-authored)
  - Compare output to reference text
  - Widely used in machine translation

References

- Ehud Reiter and Robert Dale. Building natural language generation systems. Cambridge University Press. 2000
- Johanna D. Moore. Slides from the Natural Language Generation course http://www.inf.ed.ac.uk/teaching/courses/nlg/
- Summer School on Natural Language Generation, Summarisation and Dialogue Systems http://nlgsummer.github.io
- Taghi Paksima, Kallirroi Georgila, and Johanna D. Moore. Evaluating the effectiveness of information presentation in a full end-to-end dialogue system. SIGDIAL 2009