

Seminar Natural Language Generation (NLG) — Part 1

# Introduction to Natural Language Generation

---

Henning Wachsmuth

<https://ai.uni-hannover.de>



Leibniz  
Universität  
Hannover



---

# Motivation

# Example: Conversational AI <https://chat.openai.com> (April 4, 2024)

 You

What research topics would you cover in a master seminar on natural language generation?

 ChatGPT

A master seminar on natural language generation (NLG) could cover a wide range of research topics, depending on the interests and focus of the seminar and the expertise of the instructor. Here are some potential research topics that could be covered:

1. **Fundamentals of NLG:** Understanding the basic principles, algorithms, and techniques used in NLG systems, including rule-based, template-based, and machine learning approaches.
2. **Neural NLG Models:** Exploring state-of-the-art neural network architectures for NLG, such as sequence-to-sequence models, transformer models, and their variants, including GPT (Generative Pre-trained Transformer) models.
3. **Evaluation Metrics:** Discussing various metrics for evaluating the quality and effectiveness of NLG systems, including fluency, coherence, relevance, and human evaluation methods.
4. **NLG Applications:** Examining applications of NLG across different domains, such as conversational agents, summarization, machine translation, content generation, and

# Example: Machine translation <https://www.deepl.com> (April 4, 2024)

The screenshot shows the DeepL Translator website interface. At the top, there is a navigation bar with the DeepL logo, a 'Translator' dropdown menu, and links for 'DeepL Pro', 'For Business', 'Why DeepL?', and 'API'. A 'Start free trial' button is also present. Below the navigation bar, there are three main service buttons: 'Translate text' (32 languages), 'Translate files' (.pdf, .docx, .pptx), and 'DeepL Write' (AI-powered edits, BETA). The main content area shows a translation from English (detected) to German. The source text on the left is: 'A master seminar on natural language generation (NLG) could cover a wide range of research topics, depending on the interests and focus of the seminar and the expertise of the instructor. Here are some potential research topics that could be covered: Fundamentals of NLG: Understanding the basic principles, algorithms, and techniques used in NLG systems, including rule-based, template-based, and machine learning approaches. Neural NLG Models: Exploring state-of-the-art neural network architectures for NLG, such as sequence-to-sequence models, transformer models, and their variants, including GPT (Generative Pre-trained Transformer) models.' The translated text on the right is: 'Ein Master-Seminar über die Generierung natürlicher Sprache (NLG) könnte ein breites Spektrum an Forschungsthemen abdecken, je nach Interesse und Schwerpunkt des Seminars und der Expertise des Dozenten. Hier sind einige mögliche Forschungsthemen, die behandelt werden könnten: Grundlagen der NLG: Verstehen der grundlegenden Prinzipien, Algorithmen und Techniken, die in NLG-Systemen verwendet werden, einschließlich regelbasierter, vorlagenbasierter und maschineller Lernansätze. Neuronale NLG-Modelle: Erforschung moderner neuronaler Netzwerkarchitekturen für NLG, wie z. B. Sequenz-zu-Sequenz-Modelle, Transformer-Modelle und ihre Varianten, einschließlich GPT-Modelle (Generative Pre-trained Transformer).' The interface includes language selection dropdowns for 'English (detected)' and 'German', and an 'Options' dropdown.

# Example: Project Debater [https://www.youtube.com/watch?v=UeF\\_N1r91RQ](https://www.youtube.com/watch?v=UeF_N1r91RQ)



---

# Natural Language Generation (NLG)

# Natural language processing

## ■ Computational linguistics (CL)

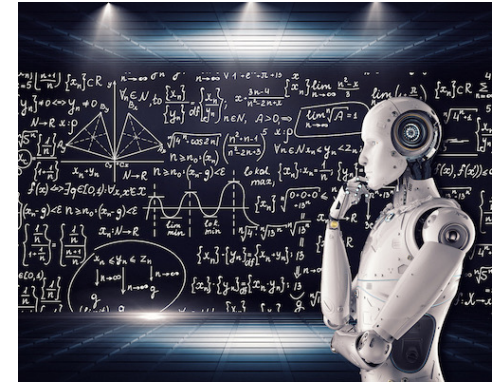
- Intersection of computer science and linguistics
- **Models** for linguistic phenomena, based on knowledge and statistics (machine learning)
- **Technologies** for natural language processing tasks

## ■ Natural language processing (NLP)

- Methods for understanding and generating speech and human-readable text
- Targets various syntactic, semantic, and pragmatic tasks
- From natural language to structured information, and vice versa

## ■ Goals of NLP research

- **Creativity.** Novelty of developed models and methods
- **Accuracy.** Effectiveness in tackling tasks
- Empirical research is often seen as stronger than theory



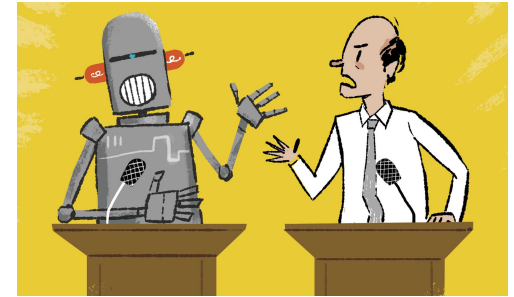
Analysis  
Synthesis



# Natural language generation

- **Natural language generation (NLG)**

- Methods for the synthesis of natural language (text)
- The goal is to encode structured or semi-structured information in an unstructured text



- **What makes NLG challenging?**

- NLG requires to choose and create a specific textual representation from many potential representations.
- Challenges: Grammaticality, coherence, naturalness, and many more

- **General types of NLG**

- **Data-to-text.** Phrase a new text with data from some knowledge base.
- **Text-to-text.** Rewrite a given text into another text.

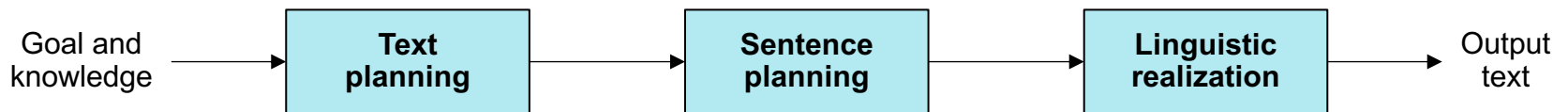
- **General techniques in NLG**

- **Traditional.** Pipeline-like text planning and realization, often knowledge-based
- **End-to-end.** Fine-tuned language modeling, mostly using neural methods

# Traditional NLG

## ▪ **Conceptual NLG process** (Reiter and Dale, 1997)

- **Input.** A goal of what to generate, and knowledge represented in some way
- **Output.** A natural language text



## ▪ **Main modeling steps**

- **Text planning.** Select content, arrange the discourse structure of sentences
- **Sentence planning.** Aggregate sentence content, make lexical choices, build referring expressions, ...
- **Linguistic realization.** Orthographic, morphological, and syntactic processing

Not all steps are always needed.

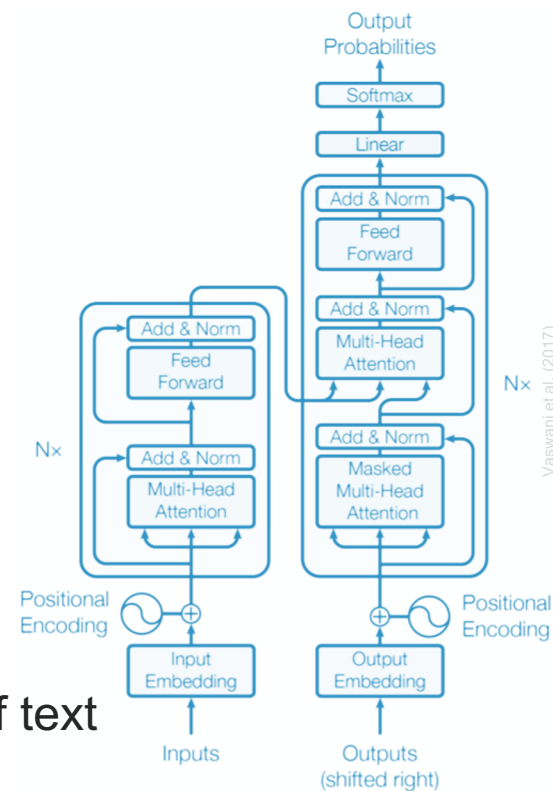
**Template.** "I am  
<stance> <issue>,  
because <reason>."

**Issue.** Death penalty  
**Stance.** Con  
**Reason.** "It kills people"

"I am *con* death penalty,  
because *the death penalty*  
*kills people.*"

# Neural language models in NLG

- **Language model**
  - A probability distribution over a sequence of words
  - Can be used to predict the most likely next words
- **Neural language models**
  - Generalize sequences from words to embeddings
  - Mostly based on the neural transformer architecture
  - Generate output sequences to input sequences
- **Main modeling steps**
  - **Pretrain** general language model on huge amounts of text
  - **Fine-tune** model task-specific input-output pairs



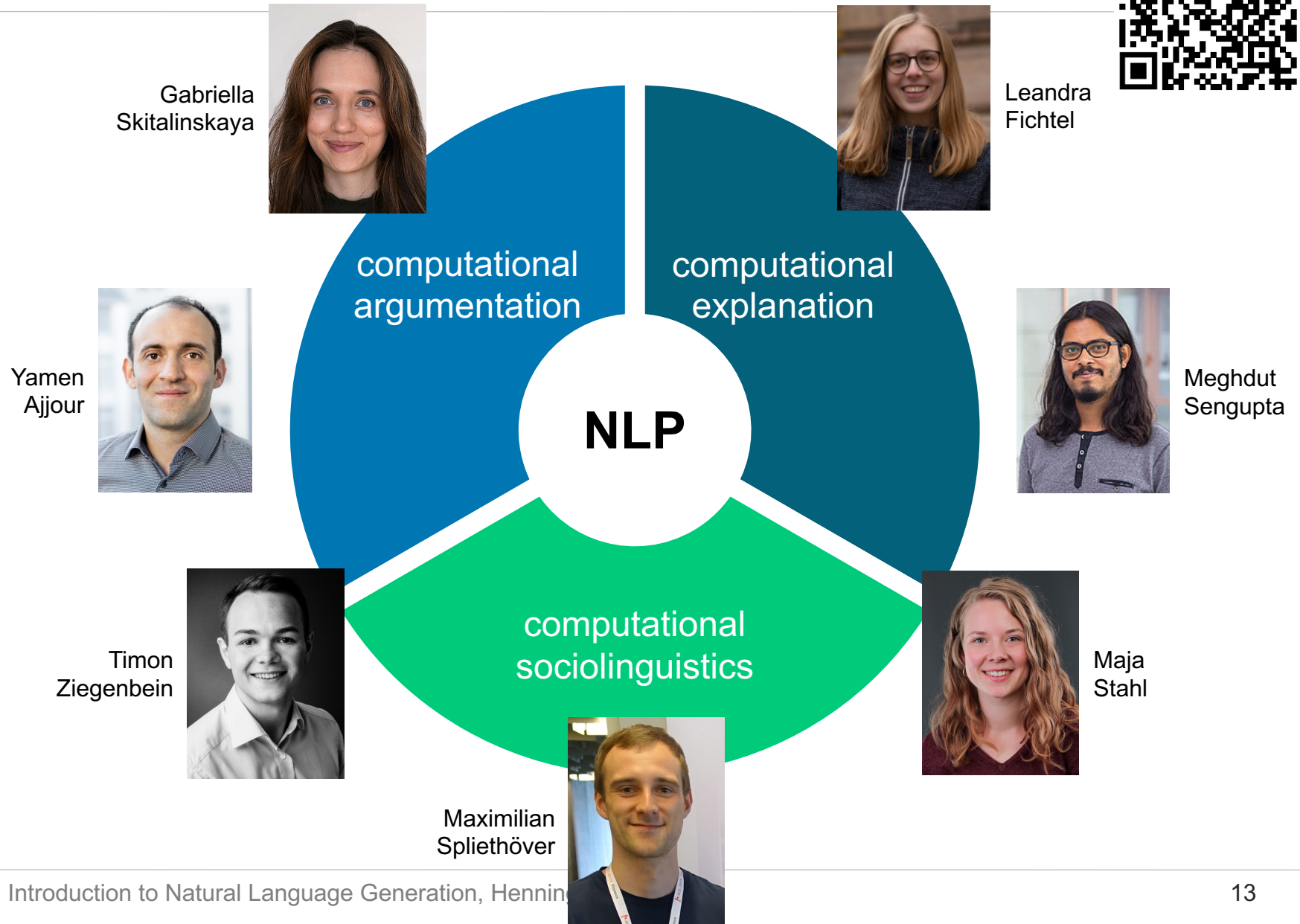
*“We spent one night at that hotel. Staff at the front desk was very nice, the room was clean and cozy, and the hotel lies in the city center... but all this never justifies the price, which is outrageous!”*

*“Nice and central hotel but outrageous price”*

---

# NLG Research of the NLP Group

# Natural language generation in the NLP Group



# Computational argumentation research

## ▪ **Argumentation**

- The usage of arguments, along with rhetoric and dialectic, to persuade or agree with others
- Arguments give reasons for claims on controversial issues
- Involved people affect the language and impact of arguments



## ▪ **Computational argumentation**

- Computational analysis and synthesis of argumentative texts
- Research primarily targets the development of methods for specific tasks
- Important for web search, debating systems, writing support, etc.

## ▪ **Selected NLG tasks**

- Summarizing argumentative texts
- Synthesizing new claims and arguments
- Improving the quality of arguments

*The EU should allow rescue boats ...*

*”... in the Mediterranean Sea.  
Many innocent refugees will die  
if there are no rescue boats.“*

# Computational sociolinguistics research

## ▪ Sociolinguistics

- The mutual interactions of society and language
- Relations between social variables and language use
- Language variation across social groups, social contexts, and communicative situations



## ▪ Computational sociolinguistics

- Analysis and adjustment of datasets and language models in social contexts
- Research primarily targets answers to sociological questions
- Important for human-machine interaction in any sociotechnical system

## ▪ Selected NLG tasks

- Creating counterfactuals for model training
- Mitigating social and political bias in language
- Generating text that fits a target audience

*”Maja dreams of a career as a researcher.“*

*“Maja works for her career as a researcher.“*

# Computational explanation research

## ▪ Explanation

- Explaining is a communicative process between an explainer and an explainee
- Explanations aim at understanding on the explainee's side
- Often has educational or informational purposes



## ▪ Computational explanation

- Analysis and synthesis of explanatory texts and dialogues
- Research primarily targets the development of methods and systems
- Important for explainability of AI, educational applications, etc.

## ▪ Selected NLG tasks

- Generating explanations of specific concepts
- Adjusting an explanation to a specific person
- Leading an explanatory dialogue

*What's blockchain?*

*It's a way that we can trade.  
Do you know what trade is?*

*It's when you take turns  
doing something, right?*



---

# NLG in this Seminar

# This seminar

---

## ▪ **Frame of this seminar**

- Basic ideas of natural language generation
- State-of-the-art NLG research
- Connections to research of the NLP Group



## ▪ **Covered topics**

- Argument mining, assessment, and generation
- Analysis and mitigation of media bias and social bias
- Computer-assisted language learning



## ▪ **Notice**

- We take a broad view on natural language generation
- Topics are selected according to our research interests
- Basics of NLP are expected rather than taught

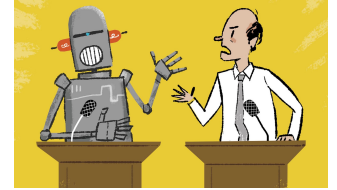




# Conclusion

## ▪ Natural language generation (NLG)

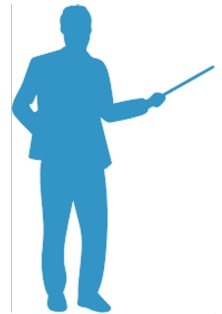
- Computational synthesis of natural language text
- Various methods from templates to neural end-to-end models
- Applications in content creation and human-AI interaction



<https://admgate.com>

## ▪ This seminar

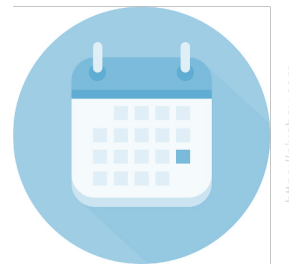
- State-of-the-art NLG research around neural language models
- Talks on argumentation, sociolinguistics, and explanation
- Close connection to research in the NLP Group



<https://www.svgsilh.com>

## ▪ Next up

- Overview of concrete seminar topics with literature pointers
- Topic preference choice and topic assignment
- Basics of scientific presentation



<https://pixabay.com>

# References

---

- **Reiter and Dale (1997).** Ehud Reiter and Robert Dale. 1997. Building applied natural language generation systems. *Natural Language Engineering*, 3(1):57–87.
- **Vaswani et al. (2017).** Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser. Attention Is All You Need. In 31st Conference on Neural Information Processing Systems, 2017.