

Seminar Natural Language Generation (NLG) — Part 1

Introduction to Natural Language Generation

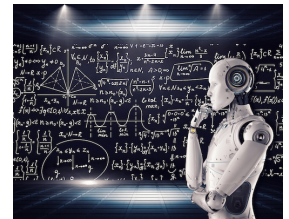
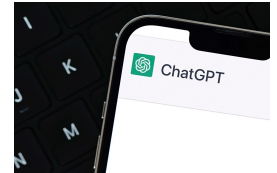
Henning Wachsmuth

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



Outline

- **Motivation**
- **Natural language generation (NLG)**
- **NLG research of the NLP Group**
- **NLG in this seminar**



Motivation

Example: Conversational AI <https://chat.openai.com> (April 13, 2023)

H

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



There are several research topics that could be covered in a master seminar on natural language generation, depending on the interests of the students and the expertise of the instructor. Here are some potential topics:



1. **Neural Language Models:** A foundational topic for natural language generation, this would cover the basics of how neural networks can be used to generate natural language text.
2. **Transformer Models:** This topic could delve deeper into the architecture of transformer models, which have been very successful in natural language processing tasks like language generation and machine translation.
3. **Style Transfer:** This is the process of modifying the style of an existing piece of text to match a desired style. It can be used for a variety of applications, from generating creative writing to improving the readability of scientific papers.
4. **Controllable Language Generation:** This topic would cover techniques for controlling various aspects of generated text, such as sentiment, tone, or topic.

Example: Machine translation <https://www.deepl.com> (April 13, 2023)



Translate text

31 languages



Translate files

.pdf, .docx, .pptx

English (detected) ▾



German ▾

Options ▾

There are several research topics that could be covered in a master seminar on natural language generation, depending on the interests of the students and the expertise of the instructor. Here are some potential topics:



Neural Language Models: A foundational topic for natural language generation, this would cover the basics of how neural networks can be used to generate natural language text.

Es gibt mehrere Forschungsthemen, die in einem Master-Seminar über die Generierung natürlicher Sprache behandelt werden könnten, je nach den Interessen der Studenten und dem Fachwissen des Dozenten. Hier sind einige mögliche Themen:

Neuronale Sprachmodelle: Ein grundlegendes Thema für die Generierung natürlicher Sprache, das die Grundlagen der Verwendung neuronaler Netze für die Generierung natürlichsprachlicher Texte abdecken würde.

back

Example: Project Debater https://www.youtube.com/watch?v=UeF_N1r91RQ

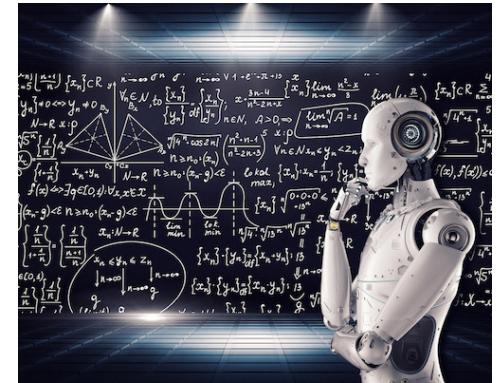


Natural Language Generation (NLG)

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

Analysis
Synthesis

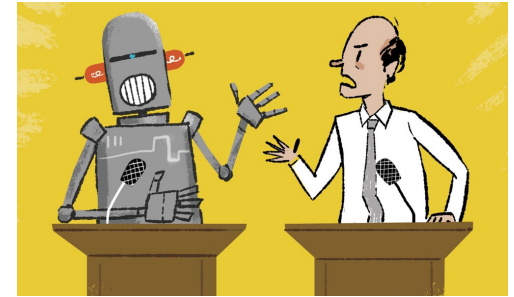
■ Goals of NLP research

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

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.

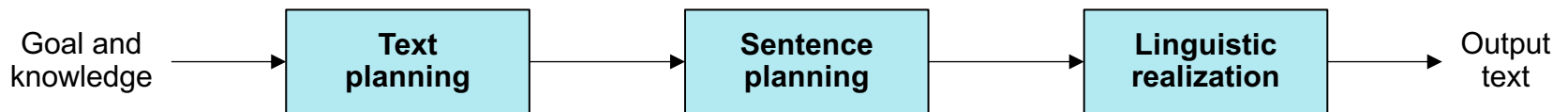
- **General techniques in NLG**

- **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 main steps (and far from all sub-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

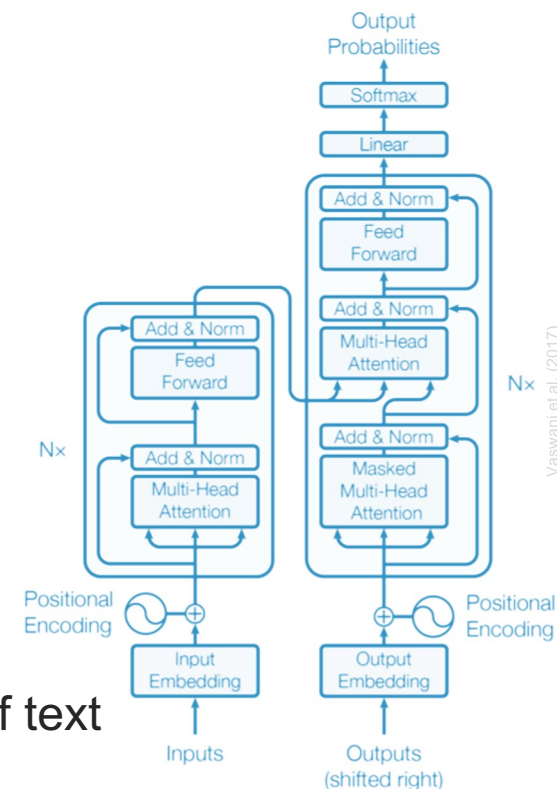
- A probability distribution over a sequence of words
- Can be used to predict the most likely next words

■ 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 @ LUH|AI



▪ Covered topics

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

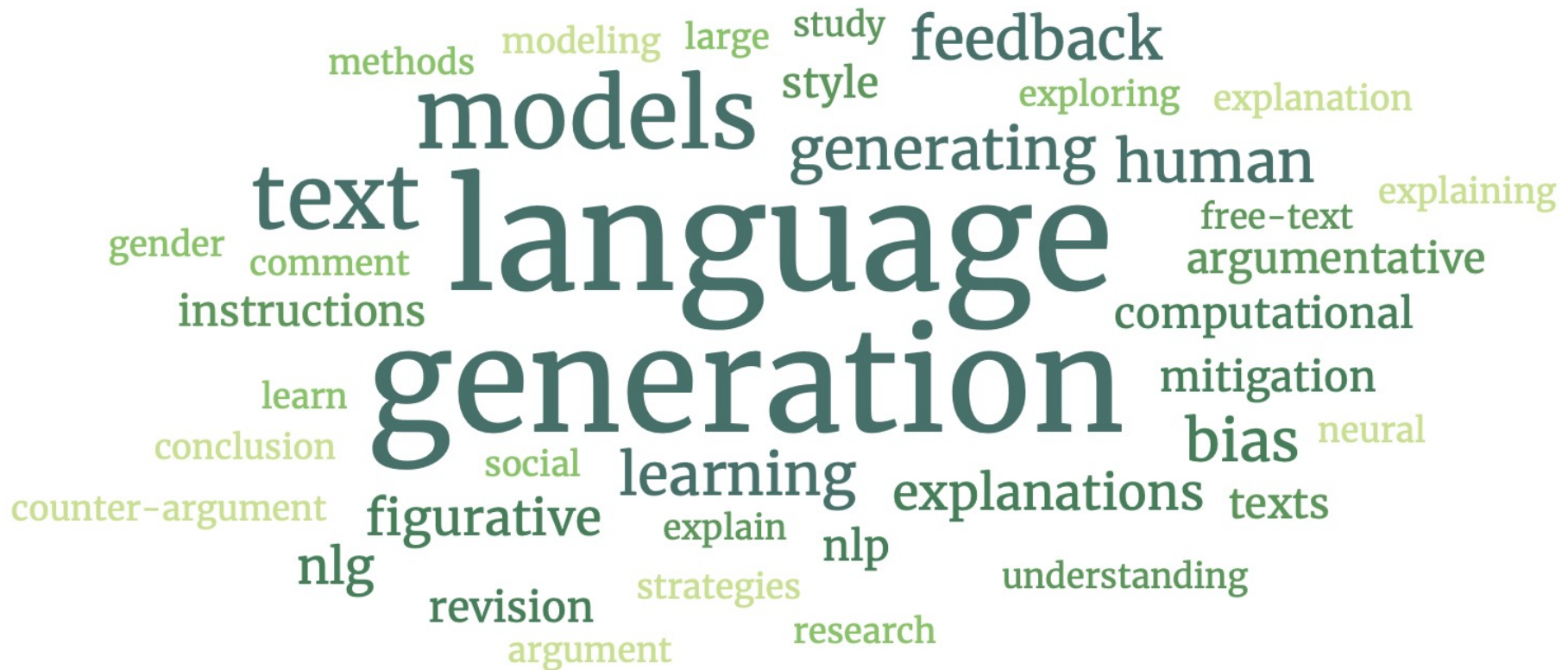


▪ Notice

- 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



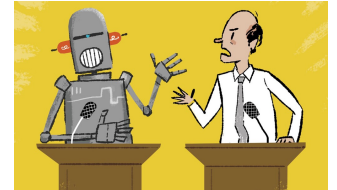
Concrete seminar topics: Next week



Conclusion

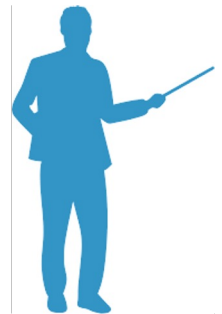
▪ Natural language generation (NLG)

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



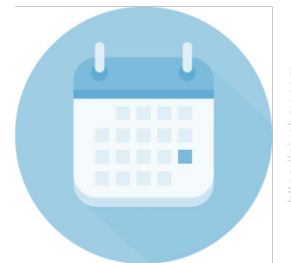
▪ This seminar

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



▪ Next up

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



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.