

Seminar Natural Language Processing (NLP) — Part 1

Introduction to Computational Sociolinguistics

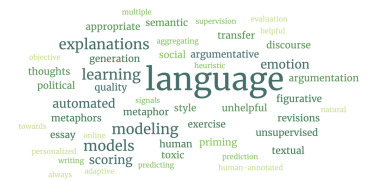
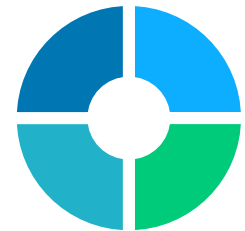
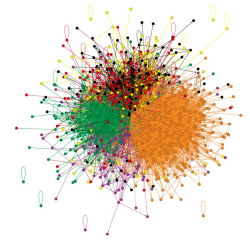
Henning Wachsmuth

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



Outline

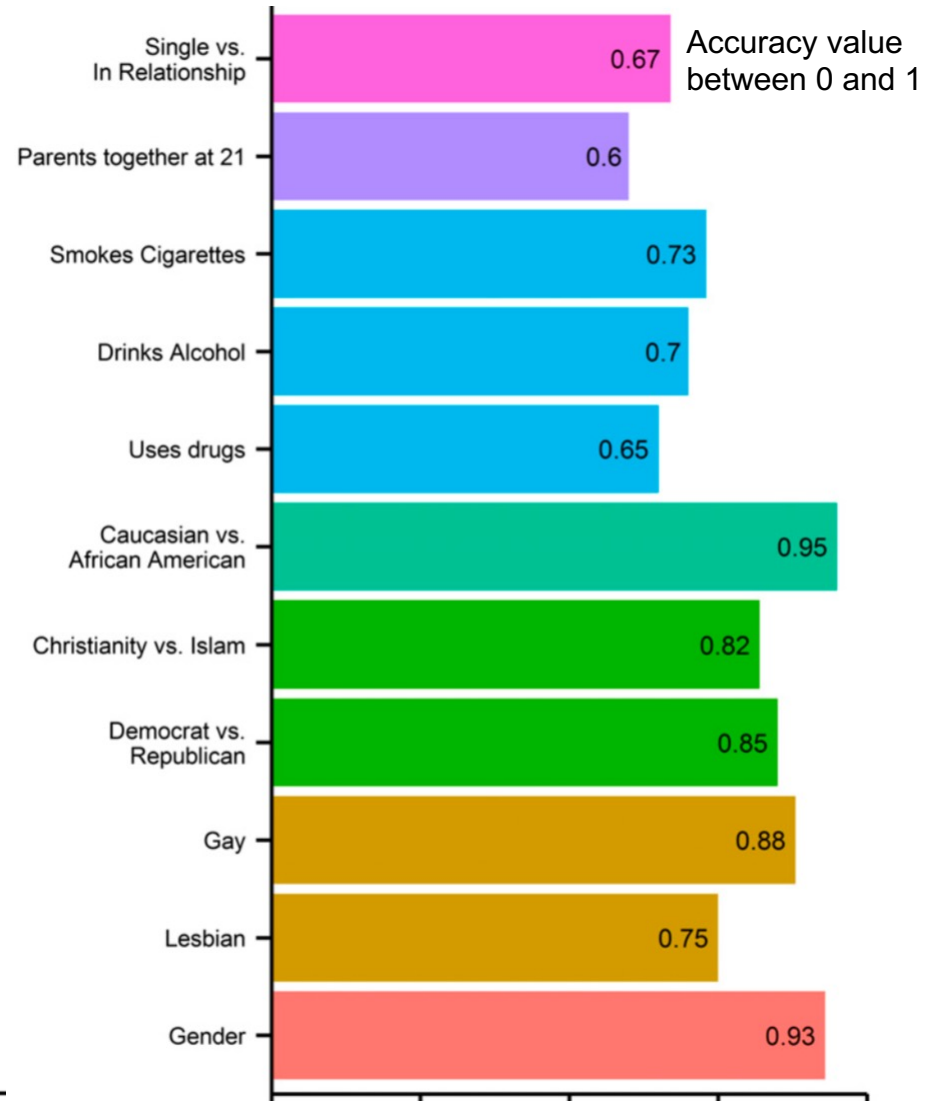
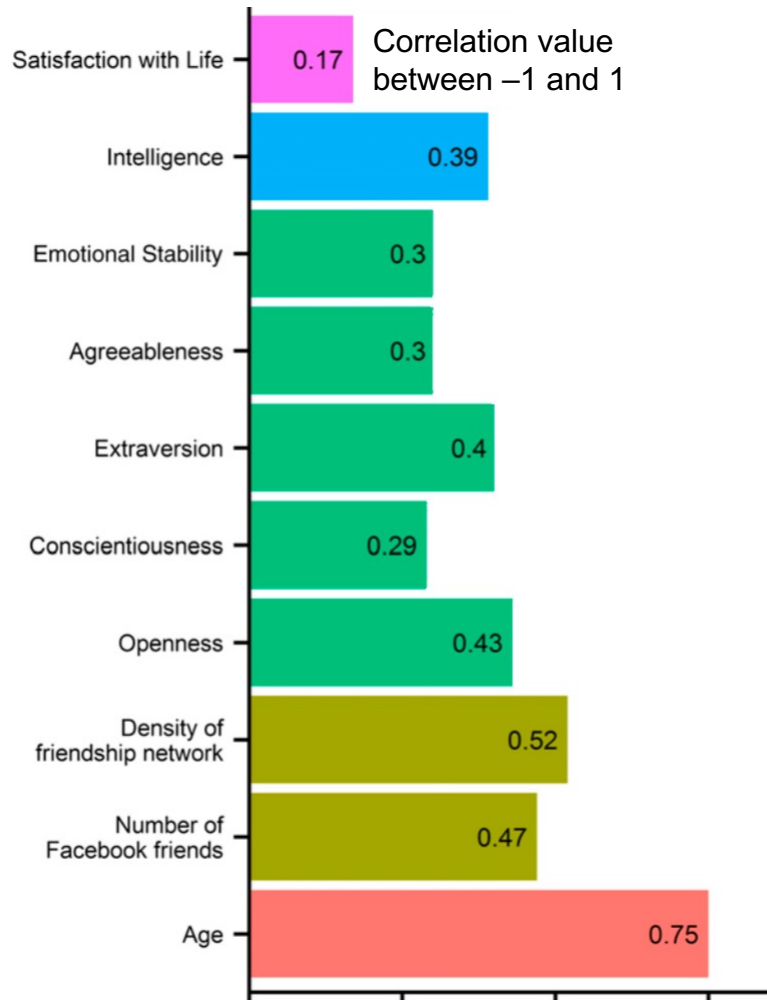
- **Motivation**
- **Computational sociolinguistics (CSL)**
- **CSL research of the NLP Group**
- **CSL in this seminar**



Motivation

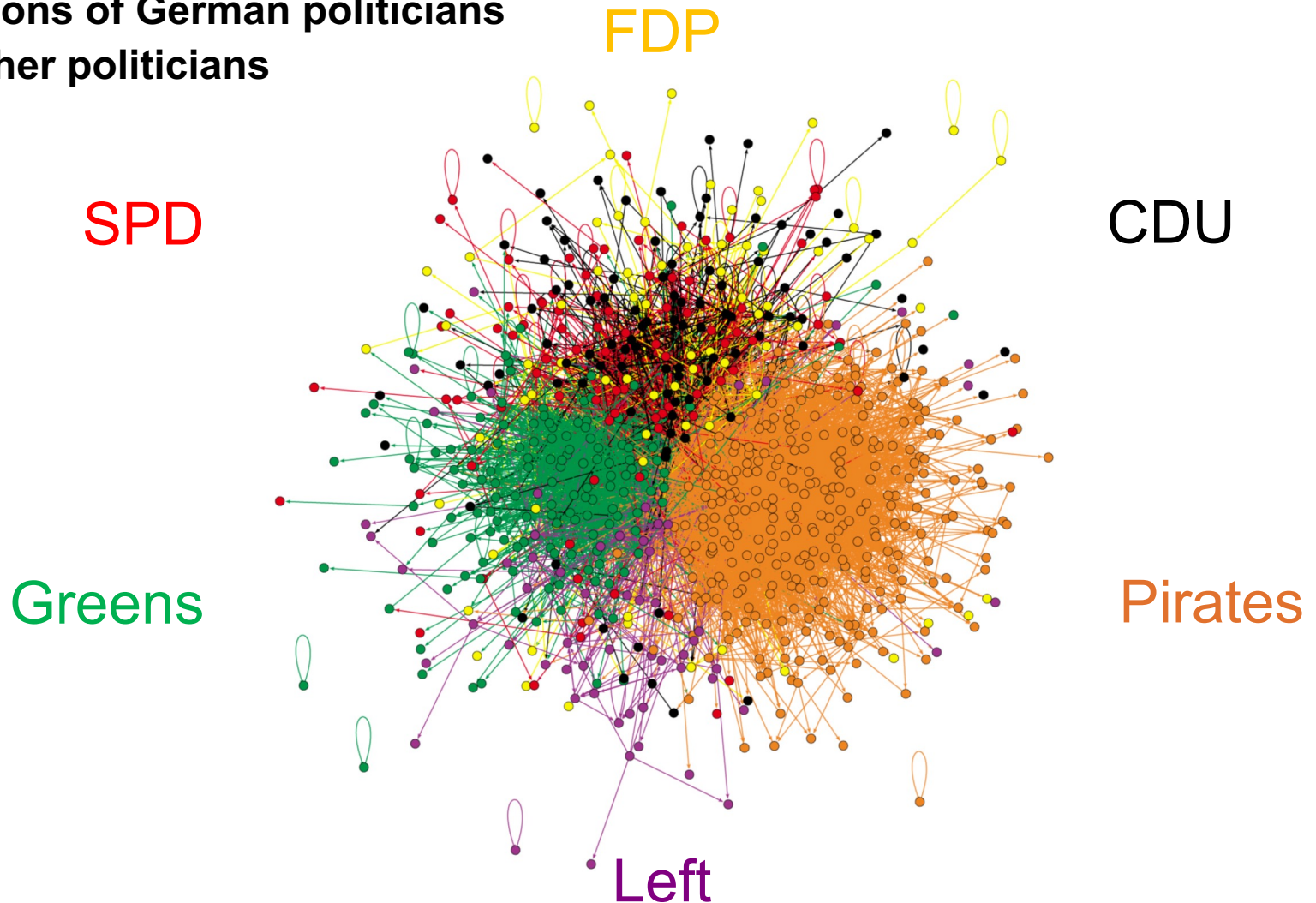
Example: Predictiveness of Facebook likes (Kosinski et al., 2013)

What your Facebook likes reveal



Example: Politicians' Twitter practices (Lietz et al., 2014)

- Mentions of German politicians by other politicians



Example: Ethnicity-related police behavior (Voigt et al., 2017)

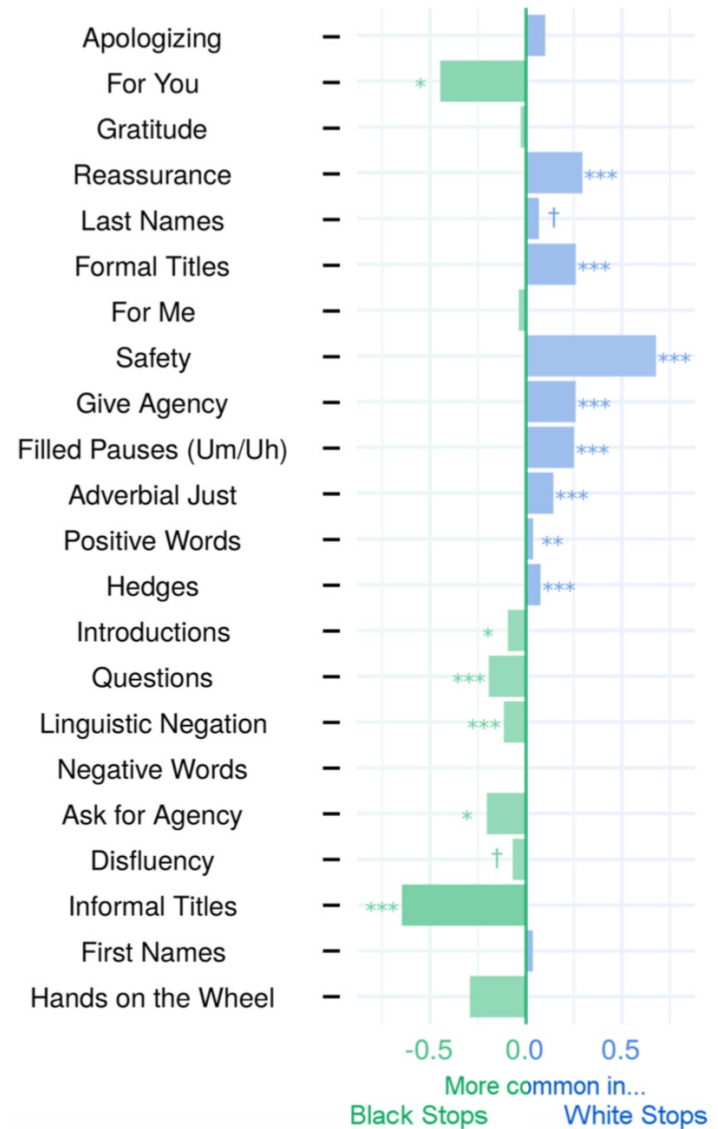
Language of US police officers towards black and white car drivers

FIRST NAME ASK FOR AGENCY QUESTIONS
 [name], can I see that driver's license again?
 It- it's showing suspended. Is that- that's you?
 DISFLUENCY NEGATIVE WORD DISFLUENCY

APOLOGY INTRODUCTION LAST NAME
 Sorry to stop you. My name's Officer [name]
 with the Police Department.

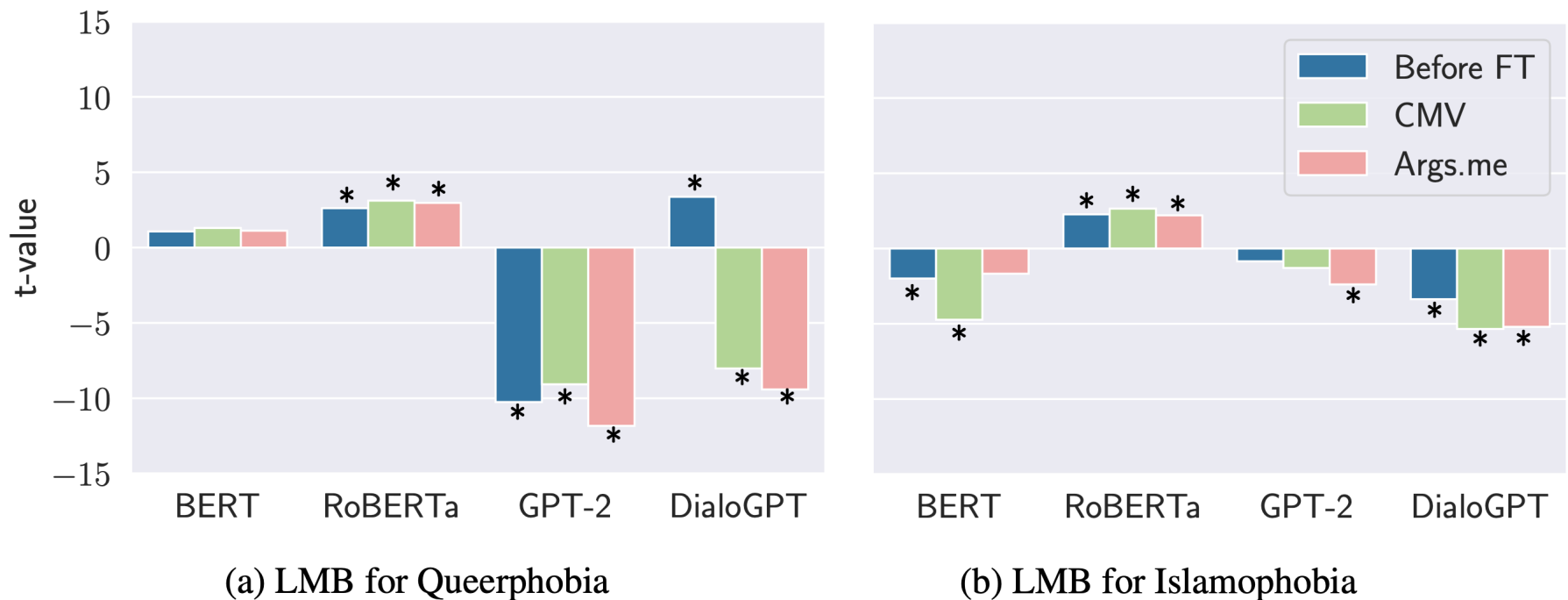
INFORMAL TITLE ASK FOR AGENCY ADVERBIAL "JUST"
 All right, my man. Do me a favor. Just keep your
 hands on the steering wheel real quick.
 "HANDS ON THE WHEEL"

FORMAL TITLE SAFETY PLEASE
 There you go, ma'am. Drive safe, please.



Example: Social bias in language models (Holtermann et al., 2022)

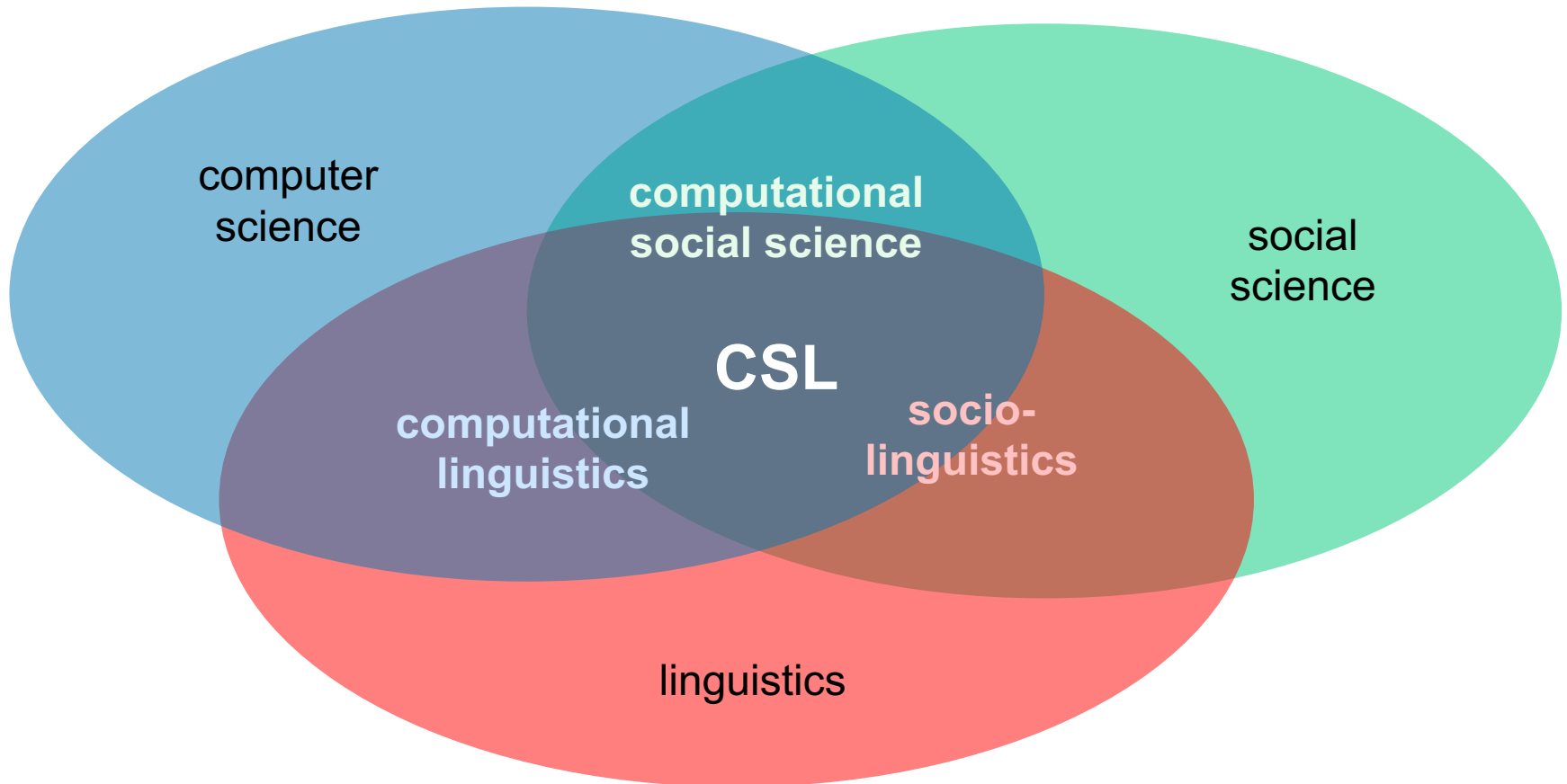
- **Social bias in language models before and after fine-tuning (FT) on subjective language**



Computational Sociolinguistics (CSL)

An interdisciplinary research area

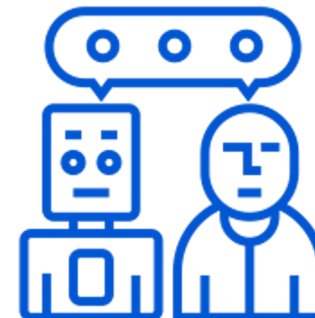
- **Two views of computational sociolinguistics (CSL)**
 - The intersection of computational linguistics and sociolinguistics
 - Computational social science on language data



Computational linguistics based on Tsujii, (2011)

■ Computational linguistics (CL)

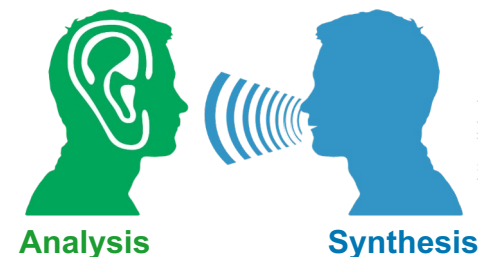
- Studies the intersection of computer science and linguistics
- **Models** for linguistic phenomena, based on knowledge and statistics (machine learning)
- **Technologies** for natural language processing tasks



<https://thenounproject.com>

■ Natural language processing (NLP)

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



<https://pixabay.com>

■ Goals of research

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

Sociolinguistics based on Nguyen et al. (2016)

▪ Sociolinguistics (SL)

- Studies the mutual interaction of society and language
- **Relations** between social variables and language use
- **Language variation** across social groups, social contexts, and communicative situations



<https://thenounproject.com>

▪ Language as a social phenomenon

- Social identity of speakers and listeners (gender, age, ...) are inherently connected to language use
- People can choose how to use language to achieve their goals
- Analyzing language often requires to consider the people

▪ Goals of research

- **Validity.** Extent to which research design isolates an issue to be studied
- **Reliability.** Reproducibility of a result
- Empirical research is seen as a means to support theory

Computational social science

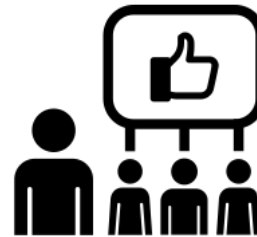
■ Computational social science (CSS)

- Studies questions from the social science through empirical data analysis
- **Insights** into social phenomena and dynamics (primary)
- **Technologies** to support social context (secondary)



■ Data (among others)

- Sociocultural key indicators
- Social network structures
- Online activities
- **News and social media texts**



■ Analyses (among others)

- Statistical correlation analyses
- Data mining
- **Natural language processing**



Computational sociolinguistics based on Nguyen et al. (2016)

■ Computational sociolinguistics (CSL)

- Studies relations between language and society computationally based on data
- **Questions** emerging from theory in sociolinguistics
- **Methods** from computational linguistics



<https://thecourproject.com>

■ NLP in the context of CSL

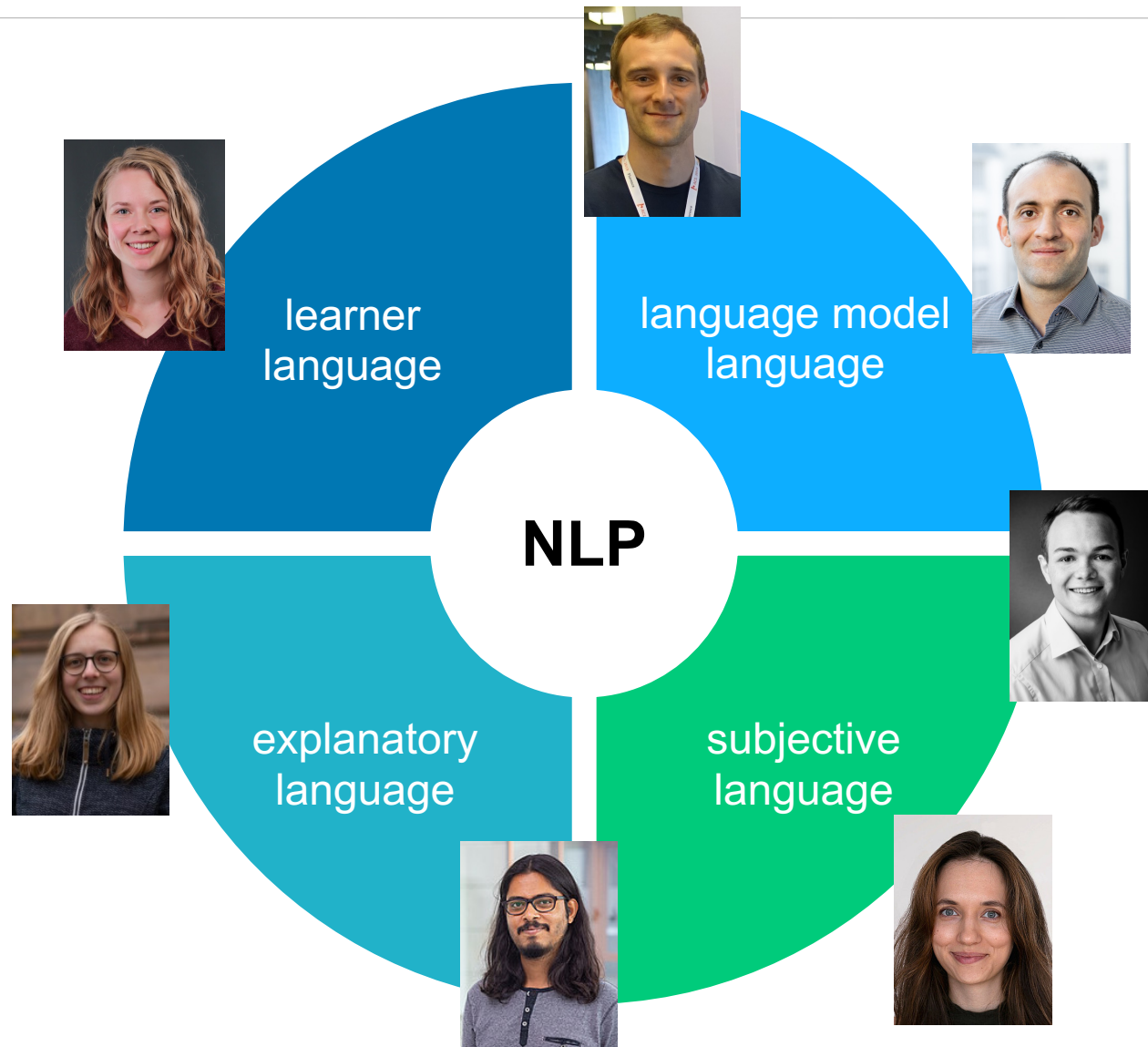
- **Data.** Natural language texts, along with sociocultural meta-information
Texts often come from news or from discussions and posts on social media.
- **Methods.** Primarily analysis (classification, regression, clustering, ...), but also text generation may be involved
- **Applications.** Tools with social dimensions (chatbots, writing support, ...)

■ Mutual impact of involved fields

- **SL → CL.** Build more robust and well-grounded computational methods
- **CL → SL.** Refine theoretical models, better understand social dynamics

CSL research of the NLP Group

Computational sociolinguistics in the NLP Group



Research on learner language

▪ **Learner language**

- Exercises given to writing and language learners
- Texts written by learners to solve exercises
- Feedback given to learners in response to their texts



▪ **Analysis**

- Empirical analyses of properties of language in learner texts
- Modeling of the structure and style of the texts
- Computational assessment of the quality of the texts



<https://pixabay.com>

▪ **Generation**

- Computational reconstruction of implicit parts of learner texts
- Computational generation of feedback to learner texts
- Automatic creation of exercises for learners



<https://pixabay.com>

Research on explanatory language

▪ Explanatory language

- Explanations given to describe or justify phenomena
- Linguistic devices used in explanations
- Explanatory dialogues between two or more people



▪ Analysis

- Modeling of structural patterns in explanatory texts and dialogues
- Computational understanding of the functioning of explanations
- Human and machine evaluation of explanation quality



<https://pixabay.com>

▪ Generation

- Computational generation of explanations
- Computational adjustment of explanations to user language
- Explaining machines that lead dialogues with users



<https://pixabay.com>

Research on subjective language

▪ Subjective language

- Statements that express sentiment, opinions, and emotions
- Personal stories, anecdotes, and thoughts
- Claims and arguments that express viewpoints



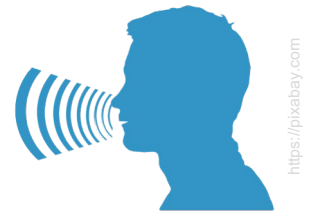
▪ Analysis

- Automatic classification and scoring of sentiment and stance
- Identification of key points in subjective texts
- Computational assessment of claim and argument quality



▪ Generation

- Computational reconstruction of implicit argument parts
- Computational generation of counterarguments
- Computational rewriting and optimization of subjective texts



Research on language model language

▪ Language model language

- Language reflected in the data language models are trained on
- Prompts to communicate with language models
- Answers given by language models in response to prompts



▪ Analysis

- Empirical evaluation of social biases in language models
- Computational analysis of priming effects on language models
- Assessment of the quality of language model outputs



<https://pixabay.com>

▪ Generation

- Instruction tuning of language models towards specific styles
- Automatic generation of counterfactuals for bias mitigation
- Computational mitigation of social bias in texts



<https://pixabay.com>

CSL in this seminar

This seminar

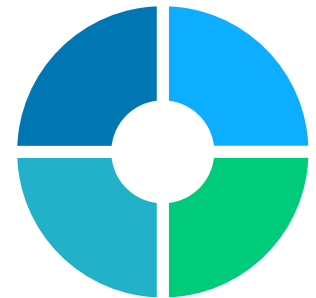
▪ **Frame of this seminar**

- Basics of NLP for computational sociolinguistics
- State-of-the-art NLP research in this area
- Connections to research at Leibniz University Hannover



▪ **Covered topics**

- Assessment and generation of learner language
- Evaluation and generation of explanatory language
- Modeling and rewriting of subjective language
- Prompting and adjustment of language model language

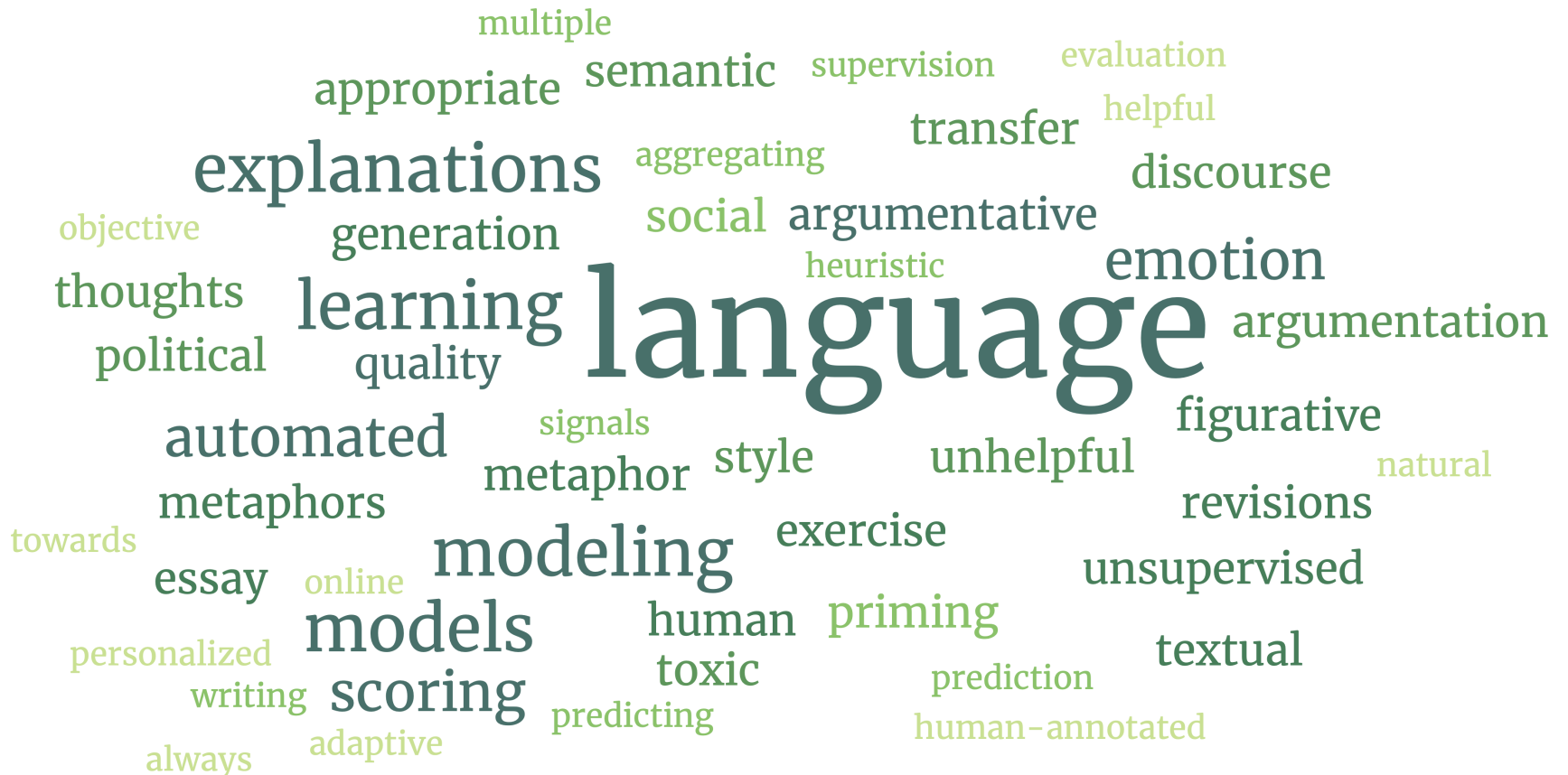


▪ **Notice**

- We take a broad view on computational sociolinguistics
- Topics are selected according to our research interests
- Required basics of NLP to be acquired rather than taught



Concrete seminar topics: Next week



Conclusion

▪ **Computational sociolinguistics (CSL)**

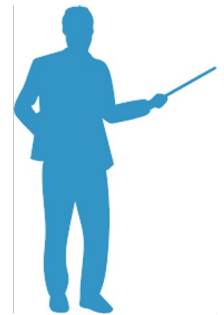
- Studies relations between language and society computationally
- Intersection of computational linguistics and sociolinguistics
- Analysis and synthesis of texts from social contexts



<https://tremounproject.com>

▪ **This seminar**

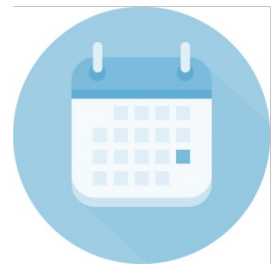
- State-of-the-art NLP research on computational sociolinguistics
- Learner, explanatory, subjective, and language model language
- Close connection to research in the NLP Group



<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

- **Holtermann et al. (2022).** Carolin Holtermann, Anne Lauscher, and Simone Ponzetto. Fair and Argumentative Language Modeling for Computational Argumentation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 7841–7861, 2022.
- **Kosinski et al. (2013).** Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.
- **Lietz et al. (2014).** Haiko Lietz, Claudia Wagner, Arnim Bleier, and Markus Strohmaier. When Politicians Talk: Assessing Online Conversational Practices of Political Parties on Twitter. In Proceedings of the Eighth International Conference on Weblogs and Social Media, 2014.
- **Nguyen et al. (2016).** Dong Nguyen, A. Seza Dođruöz, Carolyn P. Rosé, Franciska de Jong. Computational Sociolinguistics: A Survey. *Computational Linguistics* 42(3): 537–593, 2016.
- **Tsujii (2011).** Jun’ichi Tsujii. Computational Linguistics and Natural Language Processing. In Proceedings of the 12th International Conference on Computational linguistics and Intelligent Text Processing - Volume Part I, pages 52–67, 2011.
- **Voigt et al. (2017).** Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, 114(25):6521–6526, 2017.