

	DAY 1: THURSDAY 28 OCTOBER [room MM 02.110]	DAY 2: FRIDAY 29 OCTOBER [rooms MM 02.110 + MM 00.130 / MM 02.130]	
8:00-8:30	Registration with coffee		
8:30-9:00	Opening & welcome	Coffee	
9:00-10:00	<b>Keynote 1: The perks and perils of Natural Language Processing</b> <i>Malvina Nissim</i>	<b>Keynote 2: Scaling it up on Reddit: From ‘small’ to ‘bigger’ social media data in research on visual prosody</b> <i>Jannis Androutsopoulos</i>	
10:00-10:15	Break	Break	
	<b>Session 1 [MM 02.110]:</b> <i>Sociolinguistics</i>	<b>Session 4a [MM 02.110]:</b> <i>Corpus compilation and coding</i>	<b>Session 4b [MM 00.130]:</b> <i>Multimodality</i>
10:15-10:45	Which conversation partners impact teenagers’ online writing style? An overview of accommodation in teenagers’ instant messaging <i>Lisa Hilte, Reinhild Vandekerckhove &amp; Walter Daelemans</i>	Look back without anger: Recapitulation of the corpus <i>What’s up, Switzerland?</i> <i>Simone Ueberwasser</i>	A database of North American multiple modals from YouTube <i>Steven Coats</i>
10:45-11:15	Parallels between spoken and CMC language: Do tweets reflect spoken language choices? <i>Adriana Picoral, Elisa Stumpf, Larissa Goulart, Isabella Calafate de Barros, Bruna Sommer-Farias, Marine Laísa Matte, Marina Carcamo Garcia &amp; Mariana Centanin Bertho</i>	Working with socially unacceptable discourse online: Researchers’ perspective on the distressing data <i>Kristina Pahor de Maiti &amp; Darja Fišer</i>	Snap-by-Snap! How to analyze identity work of young female footballers on Snapchat <i>Melanie Fleischhacker &amp; Eva-Maria Graf</i>
11:15-11:45	Social media data as a naturalistic test bed for studying sociolinguistic and psycholinguistic patterns in verb spelling errors <i>Hanne Surkyn, Dominiek Sandra &amp; Reinhild Vandekerckhove</i>	Public perception and usage of environmental vocabulary: Building and exploiting a thematic social media corpus	The world’s emotions in emojis <i>Peter Bußwolder</i>

		<i>Tomara Gotkova &amp; Nikolay Chepurnykh</i>	
11:45-12:05	<b>Poster pitches</b>	<b>Poster pitches</b>	
12:05-13:00	Lunch	Lunch	
13:00-14:00	<b>Poster session: 6 posters</b> <ul style="list-style-type: none"> <li>• Language style accommodation in webcare conversations – <i>Lieke Verheijen &amp; Christine Liebrecht</i></li> <li>• Webcare in building virtual rapport management: Apologising in hotel management responses to negative online reviews – <i>Ly Wen Taw, Shamala Paramasivam, Alan Libert &amp; Christo Moskovsky</i></li> <li>• Metapragmatic discourses in Instagram accounts on “heterocringe” – <i>Alice Cesbron</i></li> <li>• Register variation in Reddit comments: A multidimensional analysis – <i>Axel Bohmann, Kyla McConnell, Hanna Mahler, Gustavo Maccori Kozma &amp; Rafaela Tosin</i></li> <li>• Familiarisers and taboo vocatives in Twitter discourse – <i>Ignacio M. Palacios Martínez</i></li> <li>• Political Internet memes and digital activism on Facebook: A multimodal corpus-based analysis of China’s Diba Expedition to Hong Kong – <i>Zhe Liu</i></li> </ul>	<b>Poster session: 7 posters</b> <ul style="list-style-type: none"> <li>• Buzz or change: How the social network topology conditions the fate of lexical innovations on Twitter – <i>Louise Tarrade, Jean-Pierre Chevrot &amp; Jean-Philippe Magué</i></li> <li>• SUNCODAC: A corpus of students’ forum discussions in higher education – <i>Mario Cal-Varela &amp; Francisco Javier Fernández-Polo</i></li> <li>• Emotion recognition and sarcasm mining using rule-based and deep neural networks – <i>Abdelouafi El Otmani, Trang Lam &amp; Julien Longhi</i></li> <li>• Recontextualization strategies in medical YouTube videos – <i>Anna Tereszkiewicz &amp; Magdalena Szczyrbak</i></li> <li>• Using keystroke logs to analyze CMC – <i>Erin Pacquetet</i></li> <li>• Annotation of attack speech acts in Wikipedia talk pages – <i>Céline Poudat</i></li> <li>• Online grief corpora: Data collection and questions – <i>Niclas Bodenmann &amp; Karina Frick</i></li> </ul>	
	<b>Session 2 [MM 02.110]:</b> <i>Text mining</i>	<b>Session 5a [MM 02.110]:</b> <i>Business communication</i>	<b>Session 5b [MM 02.130]:</b> <i>Interactional linguistics</i>

14:00-14:30	<p>Detecting facial emotions to support linguistic analysis of political tweets: A multimodal approach</p> <p><i>Julien Longhi</i></p>	<p>Sustainability communication of Nestlé on Twitter in the German and French context</p> <p><i>Clarissa Glück &amp; Marie Steffens</i></p>	<p>Syntactic variation and interactional coherence in online communication: The German conjunction “weil” in written interactions</p> <p><i>Aivars Glaznieks</i></p>
14:30-15:00	<p>Automatic humor detection on Jodel</p> <p><i>Manuela Bergau</i></p>	<p>Controlling social media data: A case study of the effect of social presence on consumers’ engagement with brand-generated Instagram posts</p> <p><i>Anna Corone, Annemarie Nanne &amp; Emiel van Miltenburg</i></p>	<p>Referring to other participants in asynchronous online discussions: Citation patterns in a higher education context</p> <p><i>Francisco Javier Fernández-Polo &amp; Mario Cal-Varela</i></p>
15:00-15:30	<p>Towards automatic detection of reported press and media freedom violations in Twitter and news articles</p> <p><i>Tariq Yousef, Antje Schlaf, Janos Borst, Andreas Niekler, Gerhard Heyer &amp; Benjamin Bock</i></p>	<p>Webcare across public and private social media channels: How stakeholders and the Netherlands Red Cross adapt their messages to channel affordances</p> <p><i>Christine Liebrecht &amp; Charlotte Van Hooijdonk</i></p>	<p>The speech act of apologising in Japanese computer-mediated discourse: A corpus-assisted approach</p> <p><i>Eugenia Diegoli</i></p>
15:30-16:00	Break	Break	
	<p><b>Session 3 [MM 02.110]:</b></p> <p><b><i>Connecting to society</i></b></p>	<b>Panel discussion</b>	
16:00-16:30	<p>Vitamin B: Bad or beneficial? A social network analysis of Tweets on perceived risks in the food and consumer product safety domain</p> <p><i>Nina Laurant, Lidwien van de Wijngaert &amp; Florian Kunneman</i></p>		

16:30-17:00	Live text coverage of political events: Combining content and corpus-based discourse analysis <i>Hendrik Michael &amp; Valentin Werner</i>	
17:00-17:30	Understanding polarization: A case study of Black Pete in the Netherlands <i>Lidwien van de Wijngaert</i>	Closing
17:30-18:00	Hashtags in English tweets about the New World Order <i>Natalia Mora-López</i>	
19:00-...	Diner @Bistro Bar Ivory (optional)	

Keynote Thursday 28 October:  
**The perks and perils of Natural Language Processing**  
Malvina Nissim (University of Groningen)

Tools based on Natural Language Processing are by now part of daily life, both in the personal realm as well as in the work-related arena. Recommender systems on social media, automatic translators, job application assessors are just some examples. People take many such tools for granted, and oftentimes do not even realise they are playing a role, and which role, in their lives. Developing and using such tools for a variety of studies and applications poses many choices: What data should I use? Which algorithm is most appropriate? Are my results reliable? And how much are my modelling choices influencing my findings and my analyses?

In my talk, I will talk about the relationship between the giant leap forward Natural Language Processing research has made in recent years, and the risks such positive developments might pose. I will zoom in on bias and show how it can be found in models, algorithms, data, society, and - importantly - in ourselves, too. We will also see that the definition of bias is not straightforward, that what counts as bias isn't a universally accepted concept, as it heavily interacts with our own cultural and individual biases, and that it is not obvious what the best strategies are to deal with it from an NLP perspective. Biases that percolate from our society to our data and thus to our widely used language models and to the tools that are built on them are not the only problem, of course. The fact that contemporary language models, such as the GPT family, are able to generate text that can be indistinguishable from human-produced language yields potentially dangerous situations where, for example, large amounts of unaccountable machine-generated hateful comments might be spread to social media, or fake news can be beautifully concocted and massively circulated (at a volume and speed which are obviously unparalleled by human production.)

As researchers who apply computational methods to language analysis and as NLP practitioners who contribute to the development of applications used at large by society, how should we pose ourselves with respect to such issues intrinsic in, and raised by, our work? I will critically reflect on these aspects by means of examples and by interacting as much as possible with the audience.

Bio

Prof. Dr. Malvina Nissim holds a Chair in Computational Linguistics and Society at the University of Groningen, the Netherlands. She has experience in sentiment analysis and author identification and profiling, and is interested in Natural Language Generation, with a focus on style-controlled text. A crucial aspect of her work is the reflection over ethical issues in NLP, and she is a Member at Large of the newly formed ACL Ethics Committee. She is the author of 100+ publications in international venues, is member of the main associations in the field, annually reviews for the major conferences and journals, and organises and/or (co-)chairs large-scale scientific events. She graduated in Linguistics from the University of Pisa, and obtained her PhD in Linguistics from the University of Pavia. Before joining the University of Groningen, she was a tenured researcher at the University of Bologna (2006-2014), and a post-doc at the Institute for Cognitive Science and Technology in Rome (2006) and at the University of Edinburgh (2001-2005). In 2017, she was elected as the 2016 University of Groningen Lecturer of the Year.

Keynote Friday 29 October:  
**Scaling it up on Reddit: From ‘small’ to ‘bigger’ social media data in research on visual prosody**  
Jannis Androutsopoulos (University of Hamburg)

Research on language in social media can be broadly divided into ‘small’ and ‘big data’ approaches. Advantages and disadvantages put aside, there is a lack of theoretical and methodological dialogue between the two approaches. Research on large social media corpora draws on computational methods, while studies that rely on small data sets are becoming more hybrid, blending on and offline communication and different media formats (cf. Androutsopoulos/Staehr 2018, Bolander/Locher 2020). In this talk, I argue that ‘small’ and ‘big’ data approaches would mutually benefit from exchange and synergy-building, and suggest that one way to achieve this is in terms of research designs that start out with ‘smaller’ and move on to ‘larger’ data sets (‘scaling up’) or the other way around (‘scaling down’). Such back and forth between aggregated data and richer situational and sequential contexts can help illuminate language in social media from several perspectives (Ilbury 2020).

This idea is discussed on the example of a case study on visual prosody, a term used here as a bracket term for manipulations of written-language signs that fulfil expressive functions (e.g. emphasis, excitement, a jocular tone), which are mainly conveyed by prosody in speech. The feature in focus is the ‘agitation mark’ (*Aufregezeichen*) in German social media: a sequence of graphs that consists of one or more tokens of the exclamation mark, the digit <1> and occasionally other graphs, for example <!!!1!>, <!!!1!1!>, or <!1!!>. In an explorative study (Androutsopoulos 2022), a small and subjectively collected corpus of agitation marks was used to identify the sign’s semiotic status, its pragmatic meaning, and some patterns of discourse usage. This corpus was ‘subjective’ – in the sense of consisting out of screenshots I made myself over several months on a range of social media platforms, including Facebook, Twitter, and Reddit – and contextually rich, as all examples came with a lot of discursive, sequential and multimodal context and were personally identified and scrutinized by myself. The qualitative analysis of this material found that agitation marks are used as a marker of stylization and distancing, and contextualize social voices – voices of populists, chauvinists, covid deniers, and others – that the writers of the sign distance themselves from. Despite considerable allographic variability, these sign combinations are recognised as tokens of a distinct graphic sign, and have developed a distinct political usage in German digital discourse in the course of the 2010s.

While ‘small’ data take us quite far indeed in some respects, they are not really adequate in others: a more systematic sampling of social media data is required to examine, for example, just how frequent this sign is in social media data and how its use developed over time. To this aim I am currently developing a second data collection from the biggest and oldest German-language community on Reddit, /de. My interest at this stage is to reflect on the methodological implications of this upscaling, which entails not just (obviously) a widening of the data, but also a narrowing, inasmuch as data is now collected on one particular subreddit, and a pre-configured control over sampling and selected metadata.

- Androutsopoulos J. (in press/2022) Interpunktion und Stilisierung im digitalen Diskurs: Struktur, Registrierung und Pragmatik des ‚Aufregezeichens‘. In *Sprachreflexive Praktiken – Empirische Perspektiven auf Metakommunikation*, ed. by F. Busch/ P. Droste / E. Wessels.
- Androutsopoulos, J. / A. Staehr (2018) Moving Methods Online: Researching Digital Language Practices. In *The Routledge Handbook of Language and Superdiversity*. ed. by A. Creese / A. Blackledge.
- Bolander, B. / M.A. Locher (2020) Beyond the Online Offline Distinction: Entry Points to Digital Discourse. *Discourse, Context & Media* 35.
- Ilbury, C. (2020) “Sassy Queens”: Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *Journal of Sociolinguistics* 24:2.

### Bio

Prof. Dr. Jannis Androutsopoulos is professor in German and Media Linguistics at the University of Hamburg, and from 2016-2023 research professor at MultiLing, University of Oslo. His research explores relationships between language, media and society, covering themes such as spelling and script in digital communication, multilingualism online, language ideologies in media discourse, and the role of media in sociolinguistic change. Recent publications include *Digital language and literacies: practices, awareness, and pedagogy* (Guest-edited Special Issue, *Linguistics & Education*, 2021), and *Polymedia in interaction* (Guest-edited Special Issue, *Pragmatics & Society*, 2021). Together with Heike Zinsmeister, he coordinates *DiLCo*, a new research network on “Digital language variation in context”.