Institute on Multimodality 2022

Minds. Media. Technology.

In the context of the ZiF Research Group
Multimodal Rhetoric in Online Media Communications

28 August – 6 September 2022

Convenors Mehul Bhatt (Örebro, SWE) | John Bateman (Bremen, GER) | Kay O'Halloran (Liverpool, GBR)
Organizing Chair Vasiliki Kondyli (Örebro, SWE) | Tel +46 19 302167 | vasiliki.kondyli@oru.se

CoDesign Lab EU > Multimodality. Cognition. Society. (2022) https://codesign-lab.org/multimodality.html

STEERING AND ORGANISATION

The Institute on Multimodality 2022 is fully financed and supported by:
the Center for Interdisciplinary Research /Zentrum für interdisziplinäre Forschung (ZiF)
https://www.uni-bielefeld.de/(en)/ZiF/

The Institute is also part of the **Multimodality. Cognition. Society.** initiative,
most closely synergising with the Training School on Representation Mediated Multimodality
www.codesing-lab.org/multimodality

INSTITUTE CHAIRS

MEHUL BHATT
Örebro University
Sweden

JOHN BATEMANN
University of Bremen
Germany

KAY O'HALLORAN
University of Liverpool
United Kindom

ORGANIZING CHAIR

VASILIKI KONDYLI
Örebro University
Sweden

LOCAL ARRANGEMENTS

TRIXI VALENTIN
ZiF Bielefeld
Germany

SCIENTIFIC COMMUNICATION & PR

MANUELA LENZEN
ZiF Bielefeld
Germany

INSTITUTE ON MULTIMODALITY 2022 >
ZiF - Center for Interdisciplinary Research
Methoden 1, 33615 Bielefeld, Germany (28 Aug - 6 Sept 2022)
institute2022@codesign-lab.org

INSTITUTE 2022 / MULTIMODALITY » MINDS. MEDIA. TECHNOLOGY.

The INSTITUTE ON MULTIMODALITY 2022 is an initiative of the Center for Interdisciplinary Research / Zentrum für interdisziplinäre Forschung (ZiF, Bielefeld) through its Research Group on "Multimodal Rhetoric in Online Media Communications". The institute hosts an invited faculty delivering lectures, tutorials, and keynotes; key participants include young researchers, senior and early career faculty, and industry participants.

From a scientific viewpoint, the Institute is also part of the "Multimodality. Cognition. Society." Initiative, most closely synergising with the EU COST Action Multi3Generation Training School on Representation Mediated Multimodality (www.codesign-lab.org/multimodality).

RESEARCH-BASED EDUCATIONAL AGENDA

The Institute on Multimodality provides a consolidated perspective on the theoretical, methodological, and practical understanding of multimodality research. The institute offers a novel interdisciplinary view of recent advances centralising multimodality research from diverse viewpoints, including:

- Artificial Intelligence - Machine Learning
- Cognition - Media Reception
- Multimodal Interaction - Semiotics
- Media and Communications
- Data Science - Media - Practice

Through a confluence of perspectives from Minds, Media, and Technology, the institute will position and enable an understanding of the manner in which multimodality shapes the socio-semiotic interpretation and propagation of interactional, communicative, and collaborative acts/artefacts in everyday life and work. The Institute on Multimodality 2022 features the following:

- Dissemination Event: Summer School + Doctoral Colloquium
  – consisting of keynotes, lectures, tutorials, student presentations
- Institute 2022 Workshop:
  "Multimodality and Media Studies - Emerging Perspectives"
- Panel discussion, also open to public, co-located as an evening session during the institute

One of the central agenda of the institute pertains to dissemination; through the summer school, for instance, this will be achieved through focussed lectures and tutorial sessions devoted to the topic under consideration. The institute –through its multidisciplinary faculty and technical programme– presents a comprehensive and outward-looking view of multimodality research at the interface of Minds, Media, and Technology. The scientific agenda of the institute aims to explicate the relationships between multimodality studies across areas such as cognition, artificial intelligence, visuo-auditory media, and interaction studies, particularly highlighting the significance of multimodality research towards the design and engineering of next-generation digital technologies.

FACULTY

INSTITUTE ON MULTIMODALITY 2022 / FACULTY

KEYNOTES

- Charles Forceville (University of Amsterdam)
- Asli Özyürek (Radboud University)
- Barbara Tversky (Stanford University)

LECTURERS

- John Bateman (University of Bremen)
- Mehul Bhatt (Örebro University)
- Johanna Björklund (Umeå University)
- Ralph Ewerth (University of Hannover)
- Paul Hemeren (University of Skövde)
- Inés Olza (University of Navarra)

TUTORIAL PRESENTERS

- John Bateman (University of Bremen)
- Mehul Bhatt (Örebro University)
- Ralph Ewerth (University of Hannover)
- Jakob Suchan (German Aerospace Center)

JOHN BATEMAN

University of Bremen
Department of English Language
Germany
bateman@uni-bremen.de
http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/zfn/index.html

BIOGRAPHY

John Bateman (PhD in Artificial Intelligence, Edinburgh) has been professor of applied linguistics in the Faculty of Linguistics and Literary Sciences at the University of Bremen since 1999. His main fields of research range over functional linguistics, semiotics, computational linguistics (particularly natural language generation, discourse and dialogue), formal ontology, and the theory and practice of multimodality. He has published extensively in all these areas, with articles appearing in international journals ranging from the Artificial Intelligence Journal to the Journal of Terrorism and Political Violence. Recent books include monographs on multimodality and genre (2008, Palgrave), film (with Karl-Heinrich Schmidt, 2012), text and image (2014, Routledge), and an introductory textbook to the field of multimodality and its study as a whole (with Janina Wildfeuer and Tuomo Hiippala (2017), de Gruyter). His current research focuses on the further development and application of theories of multimodal communication and their empirical evaluation with respect to ever broader ranges of media, artefacts and performances.

MEHUL BHATT

Örebro University
School of Science and Technology
Sweden
mehul.bhatt@oru.se
http://www.mehulbhatt.org

BIOGRAPHY

Mehul Bhatt is Professor within the School of Science and Technology at Orebro University (Sweden), and a Guest Professor at the University of Skövde (Sweden). His basic research focusses on formal, cognitive, and computational foundations for AI technologies with a principal emphasis on knowledge representation, semantics, integration of commonsense reasoning & learning, explainability, and spatial representation and reasoning. Mehul Bhatt steers CoDesign Lab (www.codesign-lab.org), an initiative aimed at addressing the confluence of Cognition, Artificial Intelligence, Interaction, and Design Science for the development of human-centred cognitive assistive technologies and interaction systems. Since 2014, he directs the research and consulting group DesignSpace (www.design-space.org) and pursues ongoing research in Cognitive Vision (www.codesign-lab.org/cognitive-vision) and Spatial Reasoning (www.spatial-reasoning.com).

Mehul Bhatt obtained a bachelors in economics (India), masters in information technology (Australia), and a PhD in computer science (Australia). He has been a recipient of an Alexander von Humboldt Fellowship, a German Academic Exchange Service award (DAAD), and an Australian Post-graduate Award (APA). He was the University of Bremen nominee for the German Research Foundation (DFG) Award: Heinz Maier-Leibnitz-Preis 2014. Prior to moving to Sweden, Mehul Bhatt was Professor at the University of Bremen (Germany). Further details are available via: www.mehulbhatt.org.

» Research Statement (Artificial and Human Intelligence) / http://codesign-lab.org/hcc/agenda.html

JOHANNA BJÖRKLUND

Umeå University
Department of Computing Science
Sweden
johanna.bjorklund@umu.se
https://www.umu.se/en/staff/johanna-bjorklund/

BIOGRAPHY

Johanna Björklund is an Associate Professor at the Department of Computing Science at Ume University. Her research is on semantic, or human-like, analysis of multimodal data, incorporating, e.g., images, audio, video, and text. She is also a co-founder of the media tech companies Codemill, Adlede, and Accurate Player, which deliver products and services for the media supply chain and count ITV, BBC, and ProSieben among their customers. Johanna received her PhD in Computer Science at Umeå University in 2007. The topic of her PhD thesis was the theory and application of tree languages. After her dissertation, she worked for a period of time at Dresden University as a research assistant at the chair of Prof. Vogler, before returning to Umeå and a position as first a junior and later a senior lecturer, becoming a Docent in 2016. Her work is supported by the Swedish Research Council, the Swedish Defence Research Institute, Vinnova, and various EC funding programs. She is the director of the Wallenberg Research Arena for Media and Language, which part of the research program Wallenberg AI, Autonomous Systems and Software Program (WASP).

RALPH EWERTH

Leibniz University Hannover
TIB Leibniz Information Centre of Science & Technology, and L3S Research Center
Germany
ralph.ewerth@tib.eu
www.tib.eu/en/research-development/research-groups-and-labs/visual-analytics/staff/ralph-ewerth

BIOGRAPHY

Ralph Ewerth is the Head of the Visual Analytics Research Group, at TIB Leibniz Information Centre for Science and Technology and University Library. His main research areas includes visual analytics, information and multimedia retrieval, and search as learning: semantic (automatic) annotation of visual data (images, video, 3D etc.), automatic understanding of multimodal information, digital library as a virtual place of learning and study, informal learning on the web with multimedia data, deep learning and adaptive machine learning methods, interactive exploration of media archives, usability aspects. Ralph Ewerth studied computer science with a minor in psychology in Frankfurt am Main and Marburg. After completing his doctorate in Marburg (2008) on the subject of "Robust video content analysis via transductive learning methods", he headed the research area of multimedia computing at the Chair of Distributed Systems at the Philipps University of Marburg from 2008 to 2012. From 2012 to 2015 he was a professor for digital image processing and media technology at the Ernst Abbe University in Jena, and since the beginning of 2015 he has also been Vice Dean of the Department of Electrical Engineering and Information Technology.

CHARLES FORCEVILLE

University of Amsterdam
Faculty of Humanities, Dep. of Media studies
Netherlands
C.J.Forceville@uva.nl
https://www.uva.nl/c.j.forceville.html

BIOGRAPHY

Charles Forceville is an Associate Professor in the Media Studies department of the Universiteit van Amsterdam. He authored the monograph Pictorial Metaphor in Advertising (Routledge 1996) and co-edited Multimodal Metaphor (Mouton De Gruyter 2009, with Eduardo Urios-Aparisi); The Agile Mind (Mouton de Gruyter 2013, with Tony Veale and Kurt Feyaerts); and Multimodal Argumentation and Rhetoric in Media Genres (Benjamins 2017, with Assimakis Tseronis). In 2020, Forceville published Visual and Multimodal Communication: Applying the Relevance Principle (Oxford University Press), which takes as its point of departure that relevance theory (RT) provides the key to all communication. That said, RT is no less, but also no more than a model which, he argues, needs to be complemented and fleshed out by the best that other disciplines and approaches (semiotics, narratology, stylistics, rhetoric, quantitative social science research) have to offer. Considering modes, media and genres central concepts in studying discourse, Forceville has written on advertising, documentary, animation, cartoons, and children s picture books. He is eager to explore how the study of multimodality can feed into (I) solving problems facing society at large; and (2) a better understanding of cognition. Forceville regularly uploads pre-prints of papers and chapters on his Researchgate.net and Academia.edu profiles.

PAUL HEMEREN

University of Skövde
Department of Information Technology
Sweden
paul.hemeren@his.se
https://www.his.se/mot-hogskolan/medarbetare/paul.hemeren/

BIOGRAPHY

Paul Hemeren is an associate professor in informatics at the University of Skövde and was recruited as a cognitive scientist to the University of Skövde where the subject area of informatics includes cognitive science for the development of computational models of intelligence and to improve the interaction between humans and different kinds of technology. His Ph.D. in Cognitive Science is from Lund University. In his research, he looks at how we form concepts for things. More specifically, he examines how we perceive other people's movements in the form of different actions. Humans and other creatures have an amazing ability to quickly perceive what others are doing. How does this happen? What information is used to perceive another person's intentions in an action or movement? One side of his research is about how we organize our knowledge of other people's and our own actions. The other side is on how movement information in connection with the actions of others is processed in the human brain. An important aim of his research is to integrate these two sides to gain a broader and deeper understanding of how we perceive our interaction with others. These questions turn out to be important inputs for more applied projects about how we can create safer situations for cyclists and how we can achieve a better interaction between humans and robots.

INÉS OLZA

University of Navarra
Institute for Culture and Society, MultiNeg and Multimodal Pragmatics Lab
Spain
iolzamor@unav.es
https://sites.google.com/site/inesolza/

BIOGRAPHY

Inés Olza is a Tenured Researcher in Linguistics at the Institute for Culture and Society (ICS) of the University of Navarra (UNav), an interdisciplinary research center for the Humanities and Social Sciences. Within ICS, she is affiliated with the Emotional Culture and Identity Project (CEMID). At UNav she leads the Multimodal Pragmatics Lab (MuPra Lab) and the Knowledge Generation Project MultiNeg on multimodal patterns for negation and disagreement, funded between 2019 and 2022 by the Spanish Ministry of Science (PGC2018-095703-B-I00). She is also a member of the Red Hen Lab, a distributed laboratory for research on multimodal communication. Her research focuses on figurative language, gesture and phraseology from the perspective of Pragmatics, Cognitive Linguistics and Multimodality.

ASLI ÖZYÜREK

Radboud University
Center For Language Studies (Faculty of Arts) and
the Donders Institute for Brain, Cognition and
Behavior (Faculty of Social Sciences)
Netherlands
asli.ozyurek@ru.nl
https://www.ru.nl/english/people/ozyurek-h/

BIOGRAPHY

Asli Özyürek is a full Professor at the Center For Language Studies (Faculty of Arts) and the Donders Institute for Brain, Cognition and Behavior (Faculty of Social Sciences) at Radboud University and the Director of the Multimodal Language and Cognition lab. Asli Özyürek's research in general investigates the relations between cognition, language, communication and development. She is interested in embodied and situated approaches to language and in particular to what extent our knowledge and use of bodily actions interact with language, its processing and learning. Asli Özyürek investigates this question in two domains of human communicative behavior in which body and language are closely related: A) gestures that speakers use along with speech B) sign languages (established or emerging) She has received NWO-VIDI grant (5-year), ASPASIA Award and ERC Starting Grant (5-year) and hosted three Marie-Curie Individual Postdoctoral Fellowships. She is also an elected (2019) member of Academia Europea. Currently Asli Özyürek is the PI of a 5-year NWO-VICI Grant for a research proposal entitled "Giving cognition a hand: Linking spatial cognition to linguistic expression in native and late learners of sign language and bimodal bilinguals (until 2021).

JAKOB SUCHAN

German Aerospace Center
(DLR)
Germany
jakob.suchan@dlr.de
https://codesign-lab.org/cognitive-vision

BIOGRAPHY

Jakob Suchan is a researcher based at the German Aerospace Center (DLR); previously, Jakob was based as doctoral researcher at the Human-Centred Cognitive Assistance Lab at the Faculty of Mathematics and Informatics, University of Bremen, Germany. His research is in the area of (computational) Cognitive Vision (https://codesign-lab.org/cognitive-vision/), particularly focussing on the integration of Vision and AI (specifically, KR) from the viewpoint of computational cognitive systems where integrated (embodied) perception and interaction are involved.

BARBARA TVERSKY

Stanford University, and Columbia University
Department of Psychology
USA
btversky@stanford.edu
https://profiles.stanford.edu/barbara-tversky

BIOGRAPHY

Barbara Tversky studied cognitive psychology at the University of Michigan, where she focused on the then neglected topic of spatial memory and imagery. The work continued and expanded at the Hebrew University in Jerusalem and Stanford University to include categorization, memory, cognitive maps, spatial mental models, spatial language and memory, eyewitness testimony, biased visual and verbal memory, HCI, design, diagrammatic thinking, gesture, event perception and cognition, and creativity. She is currently Professor of Psychology at Columbia Teachers College and Professor Emerita of Psychology at Stanford. She has received awards for teaching and for a computer laboratory for teaching cognitive psychology, is a fellow of the American Academy of Arts and Sciences, the American Psychological Society, the Cognitive Science Society, and the Russell Sage Foundation, and was elected to the Society of Experimental Psychology. She has served on the governing boards of many professional organizations, on the editorial boards of many journals, and on the organizing committees of nearly 100 international interdisciplinary conferences. She has enjoyed collaborations with linguists, philosophers, computer scientists, neuroscientists, biologists, chemists, engineers, architects, designers, and artists.

KEYNOTES

RELEVANCE: THE KEY PRINCIPLE OF ALL COMMUNICATION

CHARLES FORCEVILLE

UNIVERSITY OF AMSTERDAM, THE NETHERLANDS

MULTIMODALITY AS DESIGN FEATURE OF HUMAN LANGUAGE CAPACITY

ASLI ÖZYÜREK

RADBOUD UNIVERSITY, THE NETHERLANDS

COMMUNICATION FACE-TO-FACE AND ON THE PAGE: HOW GRAPHICS WORK

BARBARA TVERSKY

STANFORD UNIVERSITY, UNITED STATES

RELEVANCE: THE KEY PRINCIPLE OF ALL COMMUNICATION

CHARLES FORCEVILLE

University of Amsterdam
THE NETHERLANDS

ABSTRACT.

The humanities are in need of a general, all-encompassing model of communication. The contours of such a model actually already exist: Relevance Theory: Communication and Cognition (Blackwell 1995 [1986]). In this monograph anthropologist Dan Sperber and linguist Deirdre Wilson claim that Paul Grice s four maxims of conversation (of quantity, quality, relation, and manner) can actually be reduced to a single one: the maxim of relevance. RT s central claim is that each act of communication comes with the presumption of its own optimal relevance to the envisaged audience. Hitherto, however, RT scholars (virtually all: linguists) have almost exclusively analysed face-to-face exchanges between two people who stand next to each other. The type of communication studied is thus predominantly verbal (perhaps supported by gestures and facial expressions).

In order to fulfil RT s potential to develop into an inclusive theory of communication, it is necessary to explore how it can be adapted and refined to account for (I) communication in other modes than (only) the spoken verbal mode; and for (2) mass-communication. In Visual and Multimodal Communication: Applying the Relevance Principle (Oxford UP 2020) I take a first step toward this goal by proposing how RT works for mass-communicative messages that involve static visuals. In my presentation I will first provide a crash course in classic RT for non-linguists, and go on to show what the theory can contribute to visual and multimodal communication by discussing some examples.

Importantly, RT is no less but also no more than a model, and has little to contribute to the analysis of specific instances of communication. Therefore, RT cannot replace other theories and approaches that provide analytical models for interpreting specific discourses, such as (social) semiotics, narratology, and stylistics. It only aims to provide an all-encompassing communication model within which the insights from other approaches can be put to optimal use.

MULTIMODALITY AS DESIGN FEATURE OF HUMAN LANGUAGE CAPACITY

ASLI ÖZYÜREK

Radboud University
THE NETHERLANDS

ABSTRACT.

One of the unique aspects of human language is that in face-to- face communication it is universally multimodal (e.g., Holler and Levinson, 2019; Perniss, 2018). All hearing and deaf communities around the world use vocal and/or visual modalities (e.g., hands, body, face) with different affordances for semiotic and linguistic expression (e.g., Goldin-Meadow and Brentani, 2015; Vigliocco et al., 2014; Özyürek and Woll, 2019). Hearing communities use both vocal and visual modalities, combining speech and gesture. Deaf communities can use the visual modality for all aspects of linguistic expression in sign language. Visual articulators in both cospeech gesture and sign, unlike speech, have unique affordances for visible iconic, indexical (e.g., pointing) and simultaneous representations due to use of multiple articulators. Such expressions have been considered in traditional linguistics as being external to the language system. I will however argue and show evidence for the fact that both spoken languages and sign languages combine such modality-specific expressions with arbitrary, categorical and sequential expressions in their language structures in cross-linguistically different ways (e.g.,Kita and Özyürek, 2003; Slinmska, Özyürek and Capirci, 2020; Özyürek, 2018; 2021). Furthermore they modulate language processing, interaction and dialogue (e.g., Rasenberg, Özyürek, and Dingemanse, 2020) and language acquisition (e.g., Furman, Kuntay, Özyürek, 2014), suggesting that they are part a design feature of a unified multimodal language system. I will end my talk with discussion on how a multimodal (but not unimodal one) view can actually explain the dynamic, adaptive and flexible aspects of our language system enabling optimally to bridge the human biological, cognitive and learning constraints to the interactive, culturally varying communicative requirements of face-to-face context.

COMMUNICATION FACE-TO-FACE AND ON THE PAGE: HOW GRAPHICS WORK

BARBARA TVERSKY

Stanford University
USA

ABSTRACT.

Everyday face-to-face communication is inherently multimodal. It involves not just the words coming forth from our mouths but also the ways the voice is modulated, it involves the movements of our faces and bodies, it involves the world around us, including those we are interacting with, it involves the history of communication with our partners and with others like, and unlike, our partners. Communication entails not just ideas, but the ways the ideas are strung together (or not). I will discuss some of the many multimodal ways ideas are expressed and some of the ways they are strung together.

Creating communicative marks on spatial surfaces goes back far in time, to ancient humans though not to other primates. Marks create meanings in many ways, through resemblance, literal and metaphoric, but also through their abstract features, notably dots, lines, arrows, and boxes. Ancient graphics across the world represented space, time, and number. Contemporary graphics do the same, but have expanded poetic means of creating meaning, especially in comics. Graphics serve to expand the mind, to structure thought, to communicate to self and other.

LECTURES

TOWARDS A GENERAL ACCOUNT OF MULTIMODAL MEANING RELATIONS AND ITS APPLICATION TO NEWS MEDIA

JOHN BATEMAN

UNIVERSITY OF BREMEN, GERMANY

REPRESENTATION MEDIATED MULTIMODALITY: A CONFLUENCE OF AI AND COGNITION

MEHUL BHATT

ÖREBRO UNIVERSITY, SWEDEN

SEMANTIC PARSING OF MULTIMODAL DATA

JOHANNA BJÖRKLUND

UMEÅ UNIVERSITY, SWEDEN

COMPUTATIONAL MODELS FOR MULTIMODAL INFORMATION

RALPH EWERTH

LEIBNIZ UNIVERSITY HANNOVER, GERMANY

APPLYING THE RELEVANCE PRINCIPLE TO VISUAL AND MULTIMODAL COMMUNICATION

CHARLES FORCEVILLE

UNIVERSITY OF AMSTERDAM, THE NETHERLANDS

MULTIMODALITY IN ACTION RECOGNITION AS A BASIS FOR DEVELOPING COMPUTATIONAL MODELS IN RELATION TO HUMAN COGNITION

PAUL HEMEREN

UNIVERSITY OF SKÖRVDE, SWEDEN

MULTIMODAL APPROACHES TO ORAL CORPORA

INÉS OLZA

UNIVERSITY OF NAVARRA, SPAIN

COMMUNICATION FACE-TO-FACE AND ON THE PAGE: HOW GESTURES WORK

BARBARA TVERSKY

STANFORD UNIVERSITY, AND COLUMBIA UNIVERSITY, USA

TOWARDS A GENERAL ACCOUNT OF MULTIMODAL MEANING RELATIONS AND ITS APPLICATION TO NEWS MEDIA

JOHN BATEMAN

University of Bremen
GERMANY

ABSTRACT.

In this talk I introduce a general framework for analysing multimodal meaning making that incorporates a variety of theoretical constructs and methodological principles for engaging with complex media combining diverse forms of expression, such as text, moving images, static images, diagrams and so on. Particular attention will be paid to the close link drawn between the structure of the theory and practical decisions for annotating complex media for further study, both corpus-oriented and experimental. Illustrative examples will be drawn on throughout and some recommendations for further lines of development proposed.

SELECT PUBLICATIONS

Bateman J. Multimodality, where next? Some meta-methodological considerations. Multimodality & Society. 2022, 2(I):41-63.

Pflaeging, J., Wildfeuer, J., & Bateman, J. A. Empirical Multimodality Research: Methods, Applications, Implications. 2021, DeGruyter.

Bateman, J. Growing theory for practice: Empirical multimodality beyond the case study. Multimodal Communication. 2021, II (I), 63-74.

REPRESENTATION MEDIATED MULTIMODALITY: A CONFLUENCE OF AI AND COGNITION

MEHUL BHATT

Örebro University
SWEDEN

ABSTRACT.

In this lecture, I will address three questions relevant to computationally "making-sense" of (embodied) multimodal interaction:

1. What kind of relational abstraction mechanisms are needed to perform (explainable) grounded inference, e.g., question-answering, qualitative generalisation, hypothetical reasoning, relevant to embodied multimodal interaction?
2. How can such abstraction mechanisms be founded on behaviourally established cognitive human-factors emanating from naturalistic empirical observation? and
3. How to articulate behaviourally established abstraction mechanisms as formal declarative models suited for grounded (computational) knowledge representation and reasoning (KR) as part of large-scale hybrid AI or (computational) cognitive systems.

I will contextualise (1-3) in the backdrop of key results at the interface of (spatial) language, knowledge representation and reasoning, and visuo-auditory computing. The lectures will focus on summarising recent and ongoing research towards establishing a human-centric foundation and roadmap for the development of (neurosymbolically) grounded inference about embodied multimodal interaction. Here, intended functional purposes addressed encompass diverse operative needs such as explainable multimodal commonsense understanding, multimodal summarisation, multimodal interpretation guided decision-support, analytical visualisation. Through the lectures and tutorial, the overall purpose is to highlight the significance of "Representation Mediated Multimodality" as a foundation for next-generation, human-centred AI applicable in diverse domains where multimodality or multimodal interaction are crucial, e.g., in Digital Visuo-Auditory Media (e.g., news, movies), Autonomous Vehicles, Social and Industrial Robots, User Experience and Interaction Design.

SELECT PUBLICATIONS

Bhatt, M., and Suchan, J. (2022). Artificial Visual Intelligence: Perceptual Commonsense for Human-Centred Cognitive Technologies. In: Advanced Course on Artificial Intelligence (Human-Centred AI). Lecture Notes in Artificial Intelligence (LNAI) - Tutorial Lecture Series. (to appear; preprint available upon request, or upon release, via: https://codesign-lab.org/cognitive-vision/

Bhatt, M., Suchan, J. (2020): Cognitive vision and perception. In: ECAI 2020 - 24th European Conference on Artificial Intelligence. Frontiers in Artificial Intelligence and Applications, vol. 325, pp. 2881–2882. IOS Press (2020).

Bhatt, M., Schultz, C., Freksa, C. (2013). The 'Space' in Spatial Assistance Systems: Conception, Formalisation and Computation. In: Representing space in cognition: Interrelations of behaviour, language, and formal models. Series: Explorations in Language and Space. 978-0-19-967991-1, Oxford University Press (2013).

SEMANTIC PARSING OF MULTIMODAL DATA

JOHANNA BJÖRKLUND

Umeå University
SWEDEN

ABSTRACT.

We give an accessible, high-level, introduction to semantic parsing of multimodal data. In other words, the translation of composite media items such as a video with audio tracks and subtitles, or a digital news article with text and images, into structured representations that capture central aspects of the combined media. The problem appears in many technological areas. In robotics, it takes the form of language grounding, where the linguistic constituents of a natural language command are linked to real-world objects, attributes, and relations. In media asset management, it appears as automatic captioning of images. It is also of inherent value in machine learning, because it allows us to transfer knowledge between different modalities: Knowledge that we have learnt from text can, e.g., be used to understand images. The focus is on so-called neuro-symbolic methods, that combine the power of deep-learning methods with the transparency and control of rule-based methods.

SELECT PUBLICATIONS

Björklund, J., Cleophas, L. Aggregation-based minimization of finite state automata Acta Informatica, Springer Nature 2021, Vol. 58 : 177-194.

Björklund, J., Drewes, F., Mollevik, I. Towards Semantic Representations with a Temporal Dimension SLTC 2020 The Eighth Swedish Language Technology Conference, Online, November 25 27, 2020.

Dahlgren Lindström, A., Björklund, J., Bensch, S., et al. Probing Multimodal Embeddings for Linguistic Properties: the Visual-Semantic Case Proceedings of the 28th International Conference on Computational Linguistics (COLING), 2020, 730-744.

COMPUTATIONAL MODELS FOR MULTIMODAL INFORMATION

RALPH EWERTH

Leibniz University Hannover
GERMANY

ABSTRACT.

Multimodal information is omnipresent in the Web and comprises, for example, videos, online news, educational resources, scientific talks and papers. Besides Web search, there are other important scenarios and applications for multimodal data, e.g., human-computer interaction, digital humanities (graphic novels, audiovisual data), or learning environments. From a computational perspective it is difficult to adequately automatically interpret multimodal information and cross-modal semantic relations. One reason is that the automatic understanding and interpretation of textual, visual or audio sources themselves is difficult and it is even more difficult to model and understand the interplay of two different modalities. While the visual/verbal divide has been investigated in the communication sciences for years, it has been rarely considered from an information retrieval perspective which we do in this lecture. We present machine learning approaches to automatically recognize semantic cross-modal relations that are defined along several dimensions: cross-modal mutual information, semantic correlation, status, and abstractness. These dimensions are based on own previous work and other taxonomies. The presented approaches utilize deep neural networks, for which typically a large amount of training data is needed. We describe two strategies to overcome this issue. Finally, we outline possible use cases in the fields of search as learning and news exploration.

SELECT PUBLICATIONS

K. Pustu-Iren, G. Bruns, R. Ewerth: A Multimodal Approach for Semantic Patent Image Retrieval In: Proceedings of Workshop on Patent Text Mining and Semantic Technologies, co-located with International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2021, 45-49.

M. Springstein, E. M ller-Budack, R. Ewerth: QuTI! Quantifying Text-Image Consistency in Multimodal Documents. In: Proceedings of 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2021, 2575-2579.

J. Ghauri, S. Hakimov, R. Ewerth: Supervised Video Summarization via Multiple Feature Sets with Parallel Attention. In: Proceedings of IEEE 22nd International Conference on Multimedia & Expo (ICME), 2021.

APPLYING THE RELEVANCE PRINCIPLE TO VISUAL AND MULTIMODAL COMMUNICATION

CHARLES FORCEVILLE

University of Amsterdam
THE NETHERLANDS

ABSTRACT.

In this talk we will interactively examine in more detail what RT can contribute to the study of multimodality, and how the model needs to be complemented by the best insights from semiotics, conceptual metaphor & metonymy theory, blending theory, narratology and other approaches. Key issues to be addressed are (I) the trajectory from catching the prospective audience s attention via succeeding in conveying information to convincing them that this information is indeed relevant; (2) the degrees of certainty with which meaning can be attested: fully explicit, strongly implicit, weakly implicit, symptomatic, paranoiac; (3) the importance of genre; (4) how to accommodate misleading information in RT. Discussion of these issues will be conducted with reference to a number of examples.

SELECT PUBLICATIONS

Forceville, C. J. (2022) Visual and multimodal communication across cultures. In I. Kecskes (Ed.), The Cambridge Handbook of Intercultural Pragmatics Cambridge University Press.

Forceville, C., & Paling, S. (202I). The metaphorical representation of DEPRESSION in short, wordless animation films. Visual Communication, 20(I), I00-I20.

Forceville, C. (2020). Visual and Multimodal Communication: Applying the Relevance Principle. Oxford University Press.

MULTIMODALITY IN ACTION RECOGNITION AS A BASIS FOR DEVELOPING COMPUTATIONAL MODELS IN RELATION TO HUMAN COGNITION

PAUL HEMEREN

University of Skövde
SWEDEN

ABSTRACT.

It is no small secret that human vision is highly sensitive to the movement of other individuals. This sensitivity, however, is not restricted to local motion patterns as such. When we see the movements of others, we do not merely see the independent movement of hands, arms, feet and legs and movement of the torso. Instead, we are able to quickly and accurately identify many motion patterns as meaningful actions (running, jumping, throwing, crawling, etc.). Understanding human action recognition by using point-light displays of biological motion allows us to then compare the accuracy of computational models in relation to human cognitive and perceptual factors. This area can be used to demonstrate some of the modality factors in human action recognition as well as the possible relationship between modality factors and levels of action and event perception. This lecture will present findings about different levels of action and event perception as well as direct comparisons between computational models and human cognition and perception using point-light displays of biological motion. A key question is then to evaluate the similarities and differences between human processing and computational models. To what extent should AI-development using multimodality computation in human-machine interaction be concerned about the relation between processes and results? What role should this comparison (computational models and human cognition) have in understanding human cognition?

We will present the extent to which a computational model based on kinematics can determine action similarity and how its performance relates to human similarity judgments of the same actions. The comparative experiment results show that both the model and human participants can reliably identify whether two actions are the same or not. In another one of our studies, the affective motion of humans conveys messages that other humans perceive and understand without conventional linguistic processing. This ability to classify human movement into meaningful gestures or segments plays also a critical role in creating social interaction between humans and robots. We will also show the effect of levels of processing (top-down vs. bottom-up) on the perception of movement kinematics and primitives for grasping actions in order to gain insight into possible primitives used by the mirror system. We investigated whether or not segmentation was driven primarily by the kinematics of the action, as opposed to high-level top-down information about the action and the object used in the action.

SELECT PUBLICATIONS

Hemeren, P., & Rybarczyk, Y. (2020). The Visual Perception of Biological Motion in Adults. In Modelling Human Motion (pp. 53-71). Springer, Cham.

Hemeren, P. E., & Thill, S. (2011). Deriving motor primitives through action segmentation. Frontiers in Psychology, 1, 243.

Hemeren, P., Veto, P., Thill, S., Li, C., & Sun, J. (2021). Kinematic-based classification of social gestures and grasping by humans and machine learning techniques. Frontiers in Robotics and AI, 308.

MULTIMODAL APPROACHES TO ORAL CORPORA

INÉS OLZA

University of Navarra
SPAIN

ABSTRACT.

The multimodal turn in interactional pragmatics and conversation analysis has opened new paths in understanding the full complexity of human linguistic behavior. This turn calls to integrate words, gesture, prosody and the many shapes of speaker-space and speaker-speaker physical interaction in the analysis of the multilayered linguistic and communicative signals (Wagner, Malisz & Kopp 2014; Mondada 2019; Brown & Prieto 2021). Moreover, the dynamic relationship between the modalities involved in these complex communicative signals needs to be explained in light of high-scale cognitive operations and wider behavioral patterns such as alignment (Rasenberg, Ozyurek & Dingemanse 2020), conceptual integration (Valenzuela et al. 2020) or memory processes (Schubotz et al. 2020), among others.

In parallel, the speed of digital advances now allows to collect, store and access unprecedented amounts of ecologically valid interactional data. Traditional oral corpora are giving the way to multimodal corpora where multilayered signals can be fully described in real situated contexts (Steen et al. 2018). Such data thus incorporate individual, intentional, social (intersubjective) and context-dependent variables that model how multimodality works in face-to-face interaction.

My lectures will explore the interplay between multimodal analyses of interaction and corpus linguistics in several directions, aiming to reflect on how adding the multimodal layer(s) impacts on how we build linguistic corpora, analyze and explain situated data, and support corpus-based theoretical conclusions. I will rely on four case studies, drawn from different languages and various interactional genre, to foster discussion on the necessary multimodal look at oral linguistic data.

Lecture 1. Corpus-based approaches to multimodal interaction. Case study 1: big multimodal data to study negative constructions. Case study 2: small multimodal data to unfold the secrets of simultaneous interpreting.

Lecture 2. Building and managing multimodal corpora. Case study 3: NewsScape, a big multimodal dataset to study human communication. Case study 4: the impact of multimodality in corpus segmentation.

SELECT PUBLICATIONS

Olza, I. (2017), Metalinguistic negation and explicit echo, with reference to English and Spanish, in M. Roitman (ed.), The Pragmatics of Negation, Amsterdam/Philadelphia, John Benjamins, 45-61.

Steen, F. J., A. Hougaard, J. Joo, I. Olza, C. Pag n, A. Pleshakova, S. Ray, P. Uhrig, J. Valenzuela, J. Wozny & M. Turner (2018). Toward an infrastructure for data-driven multimodal communication research. Linguistics Vanguard 4/1.

Valenzuela, J., C. Pagán, I. Olza & D. Alcaraz (2020). Gesturing in the wild: Evidence for a flexible mental timeline. Review of Cognitive Linguistics 18(2), 289-315.

COMMUNICATION FACE-TO-FACE AND ON THE PAGE: HOW GESTURES WORK

BARBARA TVERSKY

Stanford University, and Columbia University
USA

ABSTRACT.

Everyday face-to-face communication is inherently multimodal. It involves not just the words emerging from our mouths but also the actions emerging from our faces and bodies, it involves the world around us. Gesture has its own structure, different from though integrated with language, and similar to the structure of graphics. Like graphics, gesture serves our own thought as well as that of others.

SELECT PUBLICATIONS

Tversky, B.(2021) Mind in Motion: How Action Shapes Thought, Springer Heidelberg.

Emmorey, K., Tversky, B., and Taylor, H. (2000). Using space to describe space: Perspective in speech, sign, and gesture. Spatial Cognition and Computation. 2.

Tversky, B., Lee, P., and Mainwaring, S. (1999). Why speakers mix perspectives. Spatial Cognition and Computation. 1. 399-412.

TUTORIALS

INSTITUTE 2022 TUTORIALS »
VISUO-AUDITORY NARRATIVITY AND THE MOVING IMAGE

Visuo-auditory narratives such as narrative film, online media are now unquestionably a central medium for the negotiation of issues of social relevance in myriad spheres of discourse, including the general public, in education, in transcultural studies, and so on. Yet despite the medium's prominence, exactly how this works is still poorly understood: the gulf between fine grained technical details of visuo-auditory narratives and abstract configurations of social import is still considered by many researchers to be too great for effective research.

The tutorial programme expands upon several of the themes covered during the lectures and keynotes towards addressing precisely this central research challenge by focusing on a highly constrained and yet crucial component of the visuo-auditory medium:

> the interpretation and synthesis of emotionally-engaging visuo-auditory narrative media through a confluence of empirically and cognitively well-founded formalisations of narrative and its workings by combining well-specified and mutually complementary approaches concerned with humanities-based analyses of narrative patterns, its recipients and contexts of reception, and fine-grained computational cognitive modelling of visuo-auditory narrative interpretation from the viewpoint of embodied multimodal interaction and formal narrative semantics.

The tutorial session will demonstrate state of the art techniques, as well as discuss opportunities for future research.

AI FOR ANALYSING MEDIA, AND ITS RECEPTION


JAKOB SUCHAN AND MEHUL BHATT

DLR – ÖREBRO UNIVERSITY, GERMANY & SWEDEN

NEW DEVELOPMENTS IN THE AUTOMATIC ANALYSIS OF NARRATIVE STRATEGIES IN AUDIOVISUAL NEWS REPORTING


RALPH EWERTH AND JOHN BATEMAN

LEIBNIZ UNI. – UNIVERSITY OF BREMEN, GERMANY

DOCTORAL COLLOQUIUM

INSTITUTE 2022 | DOCTORAL COLLOQUIUM

The Doctoral Colloquium (DC) at the Institute on Multimodality 2022 is an opportunity primarily (but not exclusively) for early stage doctoral researchers to present ongoing or planned research in one or more of the key themes in scope of the institute. Participating DC members engage with institute faculty and participants throughout the institute through planned lectures, as well as in dedicated poster sessions and a final presentation event devoted solely for DC members.

Contributing doctoral colloquium participants are:

• Hassan Banaruee	University of Bonn – GERMANY
• Jingwen Cai	Umeä University – SWEDEN
• Javier Yániz Ciriza	University of Navarra – SPAIN
• Sandy Ciroux	University of Konstanz – GERMANY
• Sara Cucurachi	Örebro University – SWEDEN
• Pinelopi Ioannidou	University of Cologne – GERMANY
• Vasiliki Kondyli	Örebro University – SWEDEN
• Zhe Liu	University of Leeds – UNITED KINGDOM
• Vipul Nair	University of Skövde – SWEDEN
• Gabriella Souza Oliveira	University of São Paulo – BRAZIL
• Romain Pastureau	Basque Center on Cognition – SPAIN
• Madeleine Pikowsky	University of Cologne – GERMANY
• Rosa Suviranta	University of Helsinki – FINLAND
• Amber te Velde	University of Amsterdam – THE NETHERLANDS
• Jiahao Yang	University of Warwick – UNITED KINGDOM
• Jia Zhangjingwen	University of Amsterdam – THE NETHERLANDS

PANEL DISCUSSION

INSTITUTE 2022 » PANEL DISCUSSION
MULTIMODAL RHETORIC AND MEDIA COMMUNICATIONS

The panel discussion on "Multimodal Rhetoric and Media Communications" addresses issues raised by the diverse ways in which current media commonly apply diverse multimodal strategies –spanning static and moving images, sounds and music, spoken and written language, gesture, immersive/virtual content and more– in order to inform, influence, and engage recipient stakeholders, e.g., general public, special interest groups, policy and decision-makers.

The panel will draw on current states of the art and practice, as well as emerging and outward looking questions, to critically examine the significance of new multimodal approaches for the design and engineering of next-generation technologies for the synthesis, dissemination, and analyses of (multimodal) media content in socially relevant contexts. The panel will also reflect upon the opportunities and threats posed by emerging technologies in AI and Machine Learning particularly in the media & social media context e.g., from the viewpoints of disinformation and fake news. Other areas of impact, such as visuo-auditory media design and user experience design, are also reflected upon.

The panel discussion is also open to the public. We welcome all interest groups to participate and contribute in the discussions.

INVITED PANELISTS

- Prof. Joanna Björklund — Umeå University – SWEDEN
- Prof. Charles Forceville — University of Amsterdam – THE NETHERLANDS
- Dr. Inés Olza — University of Navarra – SPAIN
- Prof. Barbara Tversky — Stanford University, and Columbia University – UNITED STATES

PANEL CONVENERS AND MODERATORS

- Prof. Mehul Bhatt — Örebro University – SWEDEN
- Prof. John Bateman — University of Bremen – GERMANY

WORKSHOP

INSTITUTE 2022 » WORKSHOP
MULTIMODALITY AND MEDIA STUDIES: EMERGING PERSPECTIVES

The final event of Institute on Multimodality 2022 is a perspectives workshop aiming to articulate immediate follow-up actions to further document as well as advance research at the interface of "Multimodality, Cognition. Society" in the specific context of design, media, and communications studies. Invited participants encompass the fields of media and communications, linguistics, cognitive science (perception), and computer science (AI and ML):

- John Bateman — University of Bremen – GERMANY
- Mehul Bhatt — Örebro University – SWEDEN
- Anders Björkvall — Örebro University – SWEDEN
- Charles Forceville — University of Amsterdam – THE NETHERLANDS
- Paul Hemeren — University of Skövde – SWEDEN
- Tuomo Hiippala — University of Helsinki – FINLAND
- Vasiliki Kondyli — Örebro University – SWEDEN
- Gautam Pal — University of Liverpool – UNITED KINGDOM
- Jakob Suchan — German Aerospace Center (DLR) – GERMANY
- Chiaoi Tseng — University of Bremen – GERMANY
- Assimakis Tseronis — Örebro University – SWEDEN

INSTITUTE ON MULTIMODALITY 2022 | Minds. Media. Technology.

ZiF - Center for Interdisciplinary Research
Methoden 1, 33615 Bielefeld, Germany (28 Aug - 6 Sept 2022)
https://codesign-lab.org/institute2022/