SUMMER SCHOOL OF COMPUTATIONAL HUMANITIES
BOOK OF ABSTRACTS

Heidelberg, July 15 – 19, 2019
## Contents

<table>
<thead>
<tr>
<th>Organization Committees</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Organizers, Board of Experts</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>List of Speakers</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures, Workshops, Panel</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Schedule</th>
<th>5</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Rooms</th>
<th>7</th>
</tr>
</thead>
</table>

### Day 1 – Monday: Computational Literary Analysis | 9

**Morning Sessions: Lectures**
- Fiction’s functions: Three data-driven hypotheses *Andrew Piper* 9
- Qualitative and quantitative computational metaphor analysis *Berenike Hermann* 10
- CLiC 2 – Digital humanities and the corpus linguistic study of fiction *Michaela Mahlberg* 10

**Afternoon Sessions: Workshops**
- Quantitative Drama Analytics *Benjamin Krautter, Janis Pagel, Nils Reiter, Marcus Willand* 11

### Day 2 – Tuesday: Computational Humanities in the Social Sphere | 13

**Morning Sessions: Lectures**
- Data harms and democratic futures *Joanna Redden* 13
- Assessing algorithmic impact in practice *Henriette Cramer* 14
- Defining “hate speech” *Claudia von Vacano* 14

**Afternoon Sessions: Workshops**
- Quantitative approaches to discourse on social media *Tatjana Scheffler* 15

**Evening Lecture and Panel**
- Working with social media data: Methodological, legal, and ethical issues *Cornelius Puschmann* 16
- Mixed Methods in the Making: Social Network Analysis Methodological and ethical challenges 17

### Day 3 – Wednesday: Methodology of Computational Humanities I | 19

**Morning Sessions: Lectures** 19
Computational humanities—digitized, digitalized, or digitally transformed humanities? Michael Piotrowski

Hidden biases. Ethical issues in working with text, and what to do about them Dirk Hovy

Afternoon Sessions: Workshops

Annotation of audio-visual data Matthias Arnold, Ina Buchweiz

Day 4 – Thursday: Computational Humanities in the Cultural Sphere

Morning Sessions: Lectures

The Case for 3D Cuneiform Adam Anderson

Interactives or interaction? The challenge of visitor engagement using digital methods Claire Warwick

Exploiting audio data for musicological research Christof Weiß

Afternoon Sessions: Poster Session and Workshop

Interdisciplinary Forum of Digital Textual Sciences (InFoDiTex) Stefan Karcher, Christopher Nunn

Europe as a discourse community: Language, emotion and argumentation Bettina Fetzer, Annika Straube

Not my president: How names and titles frame political figures Esther van den Berg, Katharina Korfhage, Josef Ruppenhofer, Michael Wiegand, Katja Markert

Nepal Heritage Documentation Project Matthias Arnold, Ina Buchholz

Discrimination in online recruitment process Faiz Ahamad

The “Redewiedergabe-Korpus” Lukas Weimer, Annelen Brunner, Ngoc Duyen Tanja Tu

Exploring network of comparative structures and comparative opinions in newspaper cuts Erdal Ayan

Mapping the historical monuments in Republican China: Travel, culture and modernity Qiuzi Guo

Big data, black boxes and bias. Janine Aldous Arantes

Formulaic sequences in Early Modern English: A corpus-driven historical pragmatic study Ding Huang

Metapragmatic positioning in social media. A comparison of the German and French body discourse. Vanessa Münch

Cross-lingual information retrieval for multilingual access to archaeological data Giulia Speranza

Study on historical animal diversity and distribution pattern of Beijing-Tianjin-Hebei region Ding Manni

Gamifying a Croatian web dictionary Josip Mihaljević

Interrogating identities: Case study of hashtag movements in India Nanditha Narayananamoorthy

Twitter, politics and society in Iran Hossein Kermani

Emergent collectives via language - Decisions in networks and deliberation in democratic public Jörn Landschoff

Language and discourses of Forensic Psychiatry Agnieszka Karlińska
From data to new languages: towards a multilingual fruition of culture through the social networks Carola Carlino 

Raise your voice - The relation of loudness in direct discourse and women’s emancipation Svenja Guhr 

Cultural practices and digital piracy in Yoruba society in Nigeria Abdul-Rasheed Lanre Sulaiman 

Arctic Tamils: A new chapter in the anthropocene of the Far North Rahul Pillai Swashanmugham 


Fictional characters at the syntax-semantics interface Anna Schmitt 

How sentimental are quotation expressions? A sentiment analysis Ngoc Duyen Tanja Tu 

Lemmatization for under-resourced languages with sequence-to-sequence learning: The case of Early Irish Oksana Dereza 

Combining Cultural Educational Theory and Corpus Linguistics Anna Moskvina 

Virtual spaces and cultural transformations: Deconstructing YouTube (Rewind) Hans Bouchard 

Onigiri: a collaborative data integration software for the Humanities and the Social Sciences William Diakité 

The political rise of the European far-right? A data-driven examination Ivo Bantel 

Automatic processing of context-sensitive elements in task-specific speech interactions Lisa Nußbaumer 

Semantic change and emerging tropes in a large corpus of New High German poetry Thomas Haider, Steffen Eger 

The process of institutionalisation of the term ‘State-Protecting Buddhism’ between 1945-2018 Hyein Lee 

How do the state functions change in the datafication era? Analysis of digital transformation strate Weronika Przecherska 

Multilingual ontologies for the representation and instantiation of annotation schemes for folk tale Thierry Declerck 

Pushing for Insurgent Mobile Citizenship and Resisting the Pitfalls of Humanitarianism Carolin Müller 

Investigating an Internationalism: Corpus-based Studies on OKAY in the Wikipedia Laura Herzberg 

Locally Compiled (Learner) Corpora in Foreign Language Teaching: Research- and Usage-Based Software Ingo Kleiber 

Natural language processing techniques applied to medieval charters studies Sergio Torres Aguilar 

Towards Arabic Text Simplification System Nouran Khallaf 

The Do Re Mi at the workplace Yasmin Soni 

Introduction to 3D Modeling Techniques Hubert Mara
Day 5 – Friday: Methodology of Computational Humanities I 71

Morning Sessions: Workshops .................................................. 71
Providing new views on textual data with knowledge graphs Leo
Born, Juri Opitz ................................................................. 71
Afternoon Sessions: Lectures and Panel ........................................ 72
Visualisation critique and critical visualisations in the context of dig-
ital cultural heritage Katrin Glinka ........................................... 72
Visual interactive analysis of verbatim text transcripts Mennatallah
El-Assady ................................................................. 72
Closing Panel “Science for Society: Transfer, Innovation, Relevance”
Moderation: Eva Wolfangel ................................................. 73

Getting around and Maps 75
Overview Map ................................................................. 75
Query public transportation schedules .................................. 76
Connections from the B&B hotel to the venue ...................... 77
Connections from the old town to the conference venue .......... 79
From Altstadt to conference dinner venue ......................... 79
Immediate neighborhood of Mathematikon ......................... 80
Panel venue ................................................................. 81
Dinner venue ................................................................. 82

Author Index 83
Organization Committees

Local Organizers

Beatrix Busse, Heidelberg University
Anette Frank, Heidelberg University
Ines Rehbein, IDS Mannheim
Josef Ruppenhofer, IDS Mannheim
Aline Schmidt, Heidelberg University
Antonina Werthmann, IDS Mannheim

Board of Experts

Christiane Brosius
Heidelberg Centre for Transcultural Studies; Visual and Media Anthropology
Andreas Dreuw
Interdisciplinary Center for Scientific Computing; Computational Chemistry
Ekkehard Felder
Department of German Language and Literature; German Linguistics
Michael Gertz
Institute for Computer Science
Günter Leypoldt
Department of English/Heidelberg Center for American Studies; English Literary and Cultural Studies
Winrich Löhr
Department of Theology; Historical Theology
Katja Markert
Department of Computational Linguistics
Barbara Mittler
Heidelberg Centre for Transcultural Studies; Sinology
Christiane von Stutterheim
German as a Foreign Language Philology
Michael Winckler
HGS MathComp; Interdisciplinary Center for Scientific Computing

Andreas Witt
Institute for the German Language; University of Mannheim; Department for Computational Linguistics
List of Speakers

Lectures, Workshops, Panel

Lectures

Adam Anderson, Berkeley
Henriette Cramer, San Francisco
Mennatallah El-Assady, Konstanz
Katrin Glinka, Berlin
Berenike Herrmann, Basel
Dirk Hovy, Milano
Michaela Mahlberg, Birmingham
Cornelius Puschmann, Hamburg
Michael Piotrowski, Lausanne
Andrew Piper, Montreal
Joanna Redden, Cardiff
Claudia von Vacano, Berkeley
Claire Warwick, Durham
Christof Weiß, Erlangen-Nuremberg

Workshops

Matthias Arnold, Heidelberg
Ina Buchholz, Heidelberg
Leo Born, Heidelberg
Benjamin Krautter, Stuttgart
Hubert Mara, Heidelberg
Juri Opitz, Heidelberg
Janis Pagel, Stuttgart
Nils Reiter, Stuttgart
Tatjana Scheffler, Potsdam
Marcus Willand, Heidelberg

Panel

Dirk Hovy, Milano (Panel Chair)
Michael Piotrowski, Lausanne
Cornelius Puschmann, Hamburg
Nils Reiter, Stuttgart
Caroline Sporleder, Göttingen
Claudia von Vacano, Berkeley

Closing Panel

Eva Wolfangel (Panel Chair)
Participants of the Summer School
<table>
<thead>
<tr>
<th>Time</th>
<th>Monday, July 15</th>
<th>Tuesday, July 16</th>
<th>Wednesday, July 17</th>
<th>Thursday, July 18</th>
<th>Friday, July 19</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00</td>
<td>Registration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:00</td>
<td>Welcome</td>
<td></td>
<td></td>
<td></td>
<td>Workshop: “Knowledge Graphs” (Just Optiz/Leo Born)</td>
</tr>
<tr>
<td></td>
<td>(Beatrix Busse, Anette Frank)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:30</td>
<td>“Fiction’s Functions” (Andrew Piper)</td>
<td>“Social Impact of New Technologies and Big Data” (Joanna Redlich)</td>
<td>“The Case for 3D Cuneiform” (Adam Anderson)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td>“Computational Literary Analysis” (Benjamin Hermann)</td>
<td>“Data and Algorithmic Bias” (Hentie Cramer)</td>
<td>“Visitor Engagement Using Digital Methods” (Claire Warwick)</td>
<td></td>
<td>Workshop: “Knowledge Graphs” (Just Optiz/Leo Born)</td>
</tr>
<tr>
<td>10:30</td>
<td>Coffee</td>
<td>Coffee</td>
<td>Coffee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00</td>
<td>“DH and the Corpus Linguistic Study of Fiction” (Michaelis Mühlig)</td>
<td>“Defining ‘Hate Speech’” (Claudia von Vacano)</td>
<td>“The Social Impact of NLP” (Dirk Hovy)</td>
<td></td>
<td>Workshop: “Knowledge Graphs” (Just Optiz/Leo Born)</td>
</tr>
<tr>
<td>11:30</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>12:00</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>13:00</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>13:30</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:00</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>15:00</td>
<td>Workshop: “Discourse on Social Media” (Tanja Scheffler)</td>
<td>Workshop: “Annnotation of Audio-Visual Data” (Matthias Arnold)</td>
<td></td>
<td>Poster Session &amp; Coffee</td>
<td>Workshop: “Visual Interactive Analysis of Verbatim Text Transcripts” (Menna El-Aswany)</td>
</tr>
<tr>
<td>15:30</td>
<td>Coffee</td>
<td>Coffee</td>
<td>Coffee</td>
<td>Coffee</td>
<td>Coffee</td>
</tr>
<tr>
<td>17:30</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>18:00</td>
<td>Opening Reception</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:00</td>
<td>“Working with Social Media Data” (Cornelia Puschmann)</td>
<td></td>
<td></td>
<td></td>
<td>Conference Dinner</td>
</tr>
<tr>
<td>19:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20:00</td>
<td>Panel Discussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20:30</td>
<td>“Mixed Methods in the Making: SNA” (Chair: Dirk Hovy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rooms

Locations of HCH19

- Lectures, Workshops, etc.: Mathematikon (INF205, Im Neuenheimer Feld 205)
- Public Lecture and Panel: Heidelberger Center for American Studies (HCA, Hauptstraße 120, located in the Curt and Heidemarie Engelhorn Palais)
- Conference Dinner: HUGO Wine & Dine (Rohrbacher Straße 47)

Rooms in the Mathematikon

- Opening Reception (Monday): roof terrace or common room, 5th floor
- Lectures: conference room on the 5th floor
- Workshops: on Monday, Tuesday, Wednesday are split into two sessions:
  - 1st session: conference room, 5th floor
  - 2nd session: PC pool, 3rd floor
- Workshops on Thursday, Friday: PC pool, 3rd floor
- Poster Session on Thursday: basement
- Panel Session on Friday: conference room, 5th floor
The study of literature abounds in theories. Since the time of Plato, scholars have developed theories about the nature and function of literature. And yet until recently, we have had no way of empirically testing these theories, nor developing new theories based on large-scale observations of literary texts. In this talk I will present research that aims to motivate three emerging hypotheses about the social function of fictional storytelling based on the use of computational approaches to the study of literature. These hypotheses thus represent the first steps in a data-driven theory of literature. I call them the coherence hypothesis, immutability hypothesis and the phenomenological hypothesis. Coherence refers to the degree of semantic and stylistic distinctiveness of fictional versus non-fictional discourse. Immutability refers to the transhistorical (and potentially transcultural) continuity of such distinctiveness. And phenomenological refers to fiction’s unique investment in the fictional subject’s sensing, testing and wondering relationship to the world. The talk will explore the appropriateness of different kinds of data for such analysis, the choice and interpretation of different feature spaces, as well as the available methodologies that can be used to support the development of such hypotheses.
Metaphors are versatile figures of speech and thought, referring to something in terms that usually refer to something else. Through embodied and culturally situated cross-domain mappings, metaphorical expressions are prime tools for creating clarity for complex and abstract phenomena, as well as for expressing emotion and perspectives. They may also be used to hide motives and even manipulate readers.

In an increasingly digitized world, text-based “frames” of reality often tap into metaphor’s world-shaping potential – as cultural artefacts have done for millennia. The bulk of metaphorical language consists in highly conventional structures that may be revitalized through contextual cues.

To capture the cognitive and communicative potential of metaphor, it is important to analyze its rich situated context as well as large scale patterns of figurative language. In qualitative analysis, reliable identification is traditionally a challenge – but a widely accepted solution is the empirically validated Metaphor Identification Procedure VU Amsterdam (MIPVU). In quantitative analysis, vector-space models and machine learning are promising venues, but much further research is needed. My lecture provides an overview of the current state of the field in quantitative and qualitative computational metaphor identification by means of a mixed-methods case study.

CLiC 2 – Digital humanities and the corpus linguistic study of fiction

Michaela Mahlberg
Department of English Language and Linguistics, University of Birmingham, UK

Along with developments in the digital humanities more widely, there is an increasing interest into the corpus linguistic study of fictional texts - sometimes referred to under the umbrella term ‘corpus stylistics’ (Semino and Short 2004). In order to be able to account as fully as possible for properties of literary texts, we need to create tools and develop methodologies that are tailored to the task at hand. Such tools also illustrate overlapping concerns in the digital humanities and in corpus linguistics. In this paper, I will illustrate key functionalities of the web application CLiC and its latest release CLiC 2. CLiC has been specifically designed for the corpus linguistic study of narrative fiction. The case studies that I will present look at textual patterns that contribute to the creation of fictional characters. The examples will be drawn from the CLiC corpora. The CLiC corpora comprise over 140 books and 16 million words across four subcorpora: the corpus of Dickens’s Novels, the 19th Century Reference Corpus (19C), the Corpus of 19th Century Children’s Literature (ChiLit) and the Corpus of Additional Requested Texts (ArTs). For all CLiC texts, direct speech and specific places around speech have been marked up
(Mahlberg et al. 2016). Hence, CLiC can run searches across defined textual subsets and support the analysis of features of narrative fiction. An important question is how a range of features and patterns in fiction can be brought together in a coherent theoretical framework. My suggestions towards such a framework focus on a lexically-driven approach to fictional speech and body language and raise more fundamental questions about how far corpus linguistics can change our theoretical perspective on fiction and connects with broader concerns in the digital humanities.

---

**Afternoon Sessions: Workshops**

**Quantitative Drama Analytics**

Benjamin Krautter, Janis Pagel, Nils Reiter, Marcus Willand  
QuaDramA, Stuttgart and Heidelberg University, Germany

In the tutorial, we will investigate dramatic texts using tools developed within the QuaDramA project. One of the core properties of dramatic texts is their structured nature which allows for straightforward structural analysis (if TEI-encoded corpora, such as theatre classique or GerDraCor are available). In addition to structural analysis (for instance, in the form of character networks), we will cover NLP-based text analysis for the character speech, and combine both types to enrich each other.

Practically, the tutorial covers three aspects: a) a basic introduction into the programming language R and the development environment RStudio, b) a hands-on introduction into the R package DramaAnalysis, used for testing hypotheses derived from drama history, and c) a guide for the operationalization of drama-related research questions, and the subsequent interpretation of quantitative results.
Day 2 – Tuesday: Computational Humanities in the Social Sphere

Morning Sessions: Lectures

Data harms and democratic futures
Joanna Redden
Data Justice Lab, Cardiff University, UK

In this talk I argue that as data systems are integrated into decision-making and information systems across sectors as well as public and private spaces, more attention must be directed to better understanding data harms. Doing so provides us with a better appreciation of where we are heading and where we may want to change course.

The ‘Investigating Data Harms’ project involves recording concrete examples of harms already caused by uses of data systems as well as interviews with practitioners across the fields of law, social work, and education who are investigating and challenging the data harms they see on the ground through their lines of work. As argued by Eubanks (2018) and Barocas and Selbst (2014) new big data systems do not treat everyone equally, the already marginalized in society are more likely to be negatively affected by big data practices. The Data Harm Record shows that harms are already happening at individual and societal levels as people are targeted based on perceived vulnerabilities, have their personal information used in ways that disadvantages them, and can be subject to discriminatory treatment and social sorting (unintentionally and intentionally) in ways that affect access to services and opportunities (Citron and Pasquale 2014). Collectively, we are now all too aware of how online information environments can be manipulated (Woolley and Howard 2017). Real-world harms are also caused by poor data quality, data errors, and algorithm and machine bias (Angwin et al. 2016). Interviews with practitioners demonstrate the need to enhance current understandings of ‘harm’ and for system wide changes. The political and legal challenges people are facing in trying to redress data harms make clear the limitations of current democratic systems. I conclude by highlighting and discussing the range of ideas being put forward about how our democratic systems should change to ensure greater transparency, accountability, and means for citizen intervention.
Assessing algorithmic impact in practice
Henriette Cramer
Spotify, San Francisco, California, US

Algorithmic bias, and potential negative outcomes of machine learning, have gained deserved attention. However, relatively few standard processes exist for industry practitioners to assess and address algorithmic bias and unintended outcomes. This lecture will discuss challenges encountered in practice and at scale, and adds domain-specific examples from the music domain. We describe the ‘translation’ of existing literature frameworks; and practical efforts in understanding data, modeling and measurement decisions’ potential impact. We’ll share technical and organizational lessons learned, data challenges and interpretation pitfalls.

Defining “hate speech”
Claudia von Vacano
D-Lab, University of California, US

In this talk, I will discuss UC Berkeley D-Lab’s online hate speech as a prolific social disease. Hate speech directly harms vulnerable populations, including women, people of color, religious minorities, immigrants, and people with disabilities. Targets of hate speech can be driven away from public forums, sometimes being forced to delete social media accounts to avoid abuse. Hate speech can cause emotional trauma, including depression, fear, and isolation. It may culminate in offline violence through swatting and doxxing, the encouragement of suicide, or terror attacks. Niche communities form reservoirs of evolving hate speech. Self-reinforcing discussions lead to the normalization of abuse, celebrate dehumanization of minority groups, and can foment violent radicalization.

Our scientific understanding is comparatively meager: we have little empirical knowledge of the problem’s true scale or the causal mechanisms involved. Research is hampered by the complexity of defining the term, as a result, many available datasets are unreliable. Data and AI models to detect hate speech are often proprietary. Keyword searches and dictionary methods are often imprecise and overly blunt tools for detecting the nuance and complexity of hate speech. Without the tools to identify, quantify, and classify hate speech, we cannot even begin to consider how to address the causes and consequences of it.

To overcome these challenges and others, this study attempts to develop a methodology for the identification and analysis of incidents of online hate speech. In partnership with Google Jigsaw, D-Lab’s solution sets a new standard for the data science of hate speech: it 1) establishes a theoretically grounded definition of hate speech inclusive of research/policy/practice, 2) develops and applies a multi-component labeling instrument, 3) creates a new crowdsourcing tool to scalably label comments, 4) curates an open, reliable multi-platform labeled hate speech corpus, 5) grows existing data and tool repositories within principles of replicable and
reproducible research, enabling greater transparency and collaboration, 6) creates new knowledge through ethical online experimentation (and citizen science), and 7) refines AI models. Ultimately, we seek to understand the causal mechanisms for intervention and evaluation. All of these innovations are guided by an advisory group, our consortium and a new open-source platform with tools that will make these resources available. Policy recommendations and advocacy organizations will educate and grow the larger community.

---

**Afternoon Sessions: Workshops**

**Quantitative approaches to discourse on social media**

Tatjana Scheffler  
Discourse Research Lab, Potsdam University, Germany

Social media such as Twitter, Facebook, blogs, forums, etc. are an abundant data source for texts generated by a diversity of users online. This provides unique opportunities and challenges for researchers in the humanities and social sciences working with textual data. In this workshop, we address some of the specific challenges posed by social media data: For one, the large amount of data necessitates automatic methods for collecting and storing texts, as well as quantitative approaches to analyzing the resulting corpora. In addition, the language in social media contains many non-standard features which on the one hand, may prevent the use of established tools for natural language processing, and on the other hand, may themselves constitute exciting opportunities for research. In particular, the conversational nature of many kinds of social media draws attention to our lack of theoretical and practical knowledge about how to model dialog and discourse (as opposed to monogical texts).

In this workshop, we will present methods from computational linguistics that enable the collection and analysis of large corpora of social media data, with a particular focus on interactive language. The workshop is aimed at young researchers who want to start working quantitatively with social media data. Since we do not assume programming abilities, the focus will be on available tools and methods for computational linguistic analysis that are approachable for researchers in the humanities and social sciences and can be immediately applied to your next research project. In addition, we will discuss state-of-the-art analyses of the nature and variability of language on social media and approaches to using social media data as a sensor for non-linguistic social data (e.g., health, human well-being, or politics).

Topics covered will include:

- Collecting social media corpora
- Working with non-standard language
- Computational social science: detecting user properties
Hands-on analyses will be carried out using and adapting existing scripts in Python. (In preparation, you may want to install the Python3 distribution through Anaconda.) Finally, we will give pointers to tutorials that allow you to implement even more powerful analyses.
Mixed Methods in the Making: Social Network Analysis

Methodological and ethical challenges

Mixed methods approaches have a great potential for investigating research questions in the humanities. They can lead to novel approaches that allow us to examine complex research questions by both tapping into large amounts of data and shifting between macro- and micro-level perspectives. One research approach that offers great opportunities for mixed methods is Social Network Analysis (SNA), because "networks are both structure and process at the same time, and therefore evade simple categorisation as either quantitative or qualitative phenomena" (Edwards 2010).

SNA has been applied to research questions in many different areas of computational humanities, most prominently in the field of literary studies, historical studies and in the social sciences where quantitative methods for network analysis have been combined with qualitative approaches. This combination offers new possibilities for the humanities. Quantitative methods can reduce the dimensionality of the data so that underlying patterns can emerge, leading to new insights about the properties of the network that would not have been visible otherwise. Qualitative approaches to SNA, on the other hand, can guide the construction of such networks, based on prior theoretical knowledge, and can thus ensure that the questions we ask are meaningful.

In the panel, we will discuss the merits and disadvantages of Social Network Analysis with experts from different fields of computational humanities. We will ask about area-specific challenges for applying SNA to research questions in the humanities, and about proper methods for the evaluation of research findings resulting from quantitative, data-driven approaches. The discussion will also address possible biases and ethical issues especially for applying SNA to social media data.

Panel speakers:

Dirk Hovy, Computer Science, Bocconi University, Milano (panel chair)
Michael Piotrowski, Digital Humanities, Lausanne
Cornelius Puschmann, Hans Bredow Institute for Media Research, Hamburg
Nils Reiter, Institute for Natural Language Processing, Stuttgart
Caroline Sporleder, Digital Humanities, Göttingen
Claudia von Vacano, D-Lab, Berkeley

References

Day 3 – Wednesday: Methodology of Computational Humanities I

Morning Sessions: Lectures

Computational humanities—digitized, digitalized, or digitally transformed humanities?

Michael Piotrowski
Digital Humanities, University of Lausanne, Switzerland

What is digital humanities? Is it "the use of computing technologies to allow humanities research that would otherwise prove impossible" (Melissa Terras), and thus primarily a matter of processing more data more quickly? Or can it "mean anything from media studies to electronic art, from data mining to edutech, from scholarly editing to anarchic blogging, while inviting code junkies, digital artists, standards wonks, transhumanists, game theorists, free culture advocates, archivists, librarians, and edupunks under its capacious canvas" (Stephen Ramsay)? Or is it a question you’re not supposed to ask, as "we will never know what digital humanities is because we don’t want to know nor is it useful for us to know" (Matthew Kirschenbaum)?

Some people are (quite understandably) tired of the discussion, but I contend that it is not only "useful" to explicate digital humanities, but crucial: the creation of academic positions, departments, and programs requires a consensus around an explicit definition—otherwise, how would one ensure the relevance and quality of research, the comparability of degree programs (and thus student mobility), or the adequate evaluation of research programs and thus their financing?

I firmly believe that the digital humanities are more than digitizing sources, using digital research infrastructures and tools, and publishing research results online—and I also believe the problem of defining digital humanities is unnecessarily exacerbated by confounding a number of related, but actually distinct issues. In this talk I will present some fundamental thoughts on this question and derive a definition of digital humanities that is both concise and precise.
Hidden biases. Ethical issues in working with text, and what to do about them

Dirk Hovy
Computer Science, Bocconi University, Milano, Italy

Language is probably the most human endeavor: through language, we fundamentally express who we are. Precisely because we express ourselves through language, we can use language to infer information about the authors of texts. This property makes text a fantastic resource for research into the complexity of the human mind, from social sciences to humanities. The recent introduction of large-scale statistical models has made this research even easier and more powerful.

However, it is exactly that human property of text that also creates some ethical problems. While we can explore the property of text to reflect the authors biases, they can also have unintended consequences for our analysis, which get magnified by statistical models. If our data is not reflective of the population we want to study, if we do not pay attention to biases enshrined in language, we can easily draw the wrong conclusions, and create disadvantages for our subjects.

In this talk, I will talk about four types of biases that affect statistical analysis of text, their sources, and potential counter measures. First, I will cover bias stemming from data, i.e., selection bias (if our texts do not adequately reflect the population we want to study) and label bias (if the labels we use are skewed). We will then look at biases deriving from the models themselves, i.e., their tendency to amplify any imbalances that are present in the data. Finally, we will look at design bias, i.e., the biases arising from our (the researchers) decisions which topics to analyze, which data sets to use, and what to do with them. For each bias, I will provide examples and discuss the possible ramifications for a wide range of applications.

Over the last few years, though, there has been an increasing body of work that not only uncovered such biases, but that has also shown various ways to address and counteract these biases, ranging from simple labeling considerations to new types of models.

I hope to leave the audience with a better, more nuanced understanding of the possible pitfalls in working with text, but also with a sense of how effectively these biases can be addressed with a little bit of forethought.

Afternoon Sessions: Workshops
Annotation of audio-visual data
Matthias Arnold, Ina Buchweiz
Centre for Advanced Transcultural Studies, Heidelberg University, Germany

Within the last decade, audio-visual material received more and more attention in academic research. A/V material is used in very different disciplines and with different research interests. This led to the development of a broad variety of digital tools and platforms, ranging from archives of broadcast stations, linguistic analysis tools, software for visual storytelling, to more general annotation platforms, to name just a few specializations.

In the first part of this workshop, we will provide a short overview of video annotation platforms and discuss different research approaches. Together with the participants we will try to identify and formulate requirements these platforms should meet. In the second part we introduce the Pan.do/ra video annotation platform, a service provided in the Centre for Asian and Transcultural Studies’ (CATS) digital research infrastructure. We will introduce three different use cases and the lessons we’ve learnt from them. In a hands-on session, participants will learn how to interact with the Pan.do/ra platform.

Sample material for the hands-on session will be provided. However, participants are invited to bring samples of their own a/v material and discuss it with the group. If you wish to provide your own material, please kindly provide it beforehand until Monday, 15.7. evening (e.g. USB drive or online sharing platform).
Day 4 – Thursday: Computational Humanities in the Cultural Sphere

Morning Sessions: Lectures

The Case for 3D Cuneiform
Adam Anderson
D-Lab, University of California, US

Thanks to recent high-profile scandals in the news, many of us are already familiar with the present state of affairs involving the collections of artifacts with writing from the ancient Near East. By applying computational tools for analyzing and 3D scanning collections of cuneiform tablets, we can better utilize these collections to address the endless cycle of the destruction of our cultural heritage in this region. By sharing concrete examples of how new computational methods can be used to restore the missing context of ancient archives, I hope to address the most pressing issues regarding the preservation of our cultural heritage, as it continues to be pulled from the ground through illicit looting and illegal excavations.

Interacts or interaction? The challenge of visitor engagement using digital methods
Claire Warwick
Department of English Studies, Durham University, UK

The museum and cultural heritage sector has been pioneering in its use digital technologies, including social media; digitised collections; online exhibitions; 3D scanning and printing of artefacts; and the use of digital art installations in gallery. However, the dominant mode of interaction between museums and cultural heritage sites and their visitors has remained very much the same: curators provide information - whether in the form of printed labels or digital 'interactives' and displays- and visitors consume it relatively passively. Museums that have experimented with web-based exhibitions and digitised collections have realised that, however compelling, such content does not replace the experience of visiting a museum or heritage site. Current developments in Virtual and Augmented Reality techniques offer exciting
new possibilities for cultural heritage. But we are only beginning to understand the range of ways in which such technologies may be used to engage with visitors. This presentation will therefore argue that to make the best use of the potential of digital technologies, museums and cultural heritage organisations may need to reconsider the true meaning of visitor interaction and engagement. This will entail a re-examination of the potential and properties of digital information spaces, and how they relate to the physical museum space.

---

**Exploiting audio data for musicological research**

Christof Weiß  
Audio Labs, University of Erlangen-Nuremberg, Germany

In recent years, quantitative and technical approaches to the humanities and cultural sciences have gained importance. This also affects musicology. Besides digitized sheet music documents, audio recordings of musical performances play an increasing role for this discipline. To analyze music recordings regarding relevant characteristics, signal processing and machine learning techniques need to be used in a suitable way. The application of these techniques to large music collections (corpus analysis) shows the high potential of computational audio analysis for musicological research.

In this talk, the characteristics and complexity of audio data shall be conveyed along with basic concepts of music processing algorithms. Their applications for musicology will be demonstrated with several examples and corpora of Western classical music recordings.
Afternoon Sessions: Poster Session and Workshop

Poster Session

- Location: basement of the Mathematikon
- Time: Wednesday, July 17, 2:30-4:30pm

Interdisciplinary Forum of Digital Textual Sciences (InFoDiTex)

Stefan Karcher, Christopher Nunn
University of Heidelberg, InFoDiTex, Germany

The Interdisciplinary Forum of Digital Textual Sciences at the University of Heidelberg is an open meeting for (junior) researchers in all fields of Digital Humanities. It was founded by doctoral students who meet every month during the semester turn for an informal exchange about theories and methods of digital text analysis and their own projects. As a Junior Research Infrastructure we met every month to...reflect on methods of digital humanities (including the questions what, how to and why at all), ...discuss theories of digital analyses, ...talk with “Big Data” professionals of industry, economy and media and ...present research projects of our members and guests.

References

InFoDiTex: What We Do and What We Want (2018, February 26), Retrieved from https://infoditex.hypotheses.org/6

Europe as a discourse community: Language, emotion and argumentation

Bettina Fetzer, Annika Straube
Institute for Translation and Interpreting, Heidelberg Uni, Germany

Europe is not only a political and economic community sharing a legal framework, but also a discourse space, the European Discourse Community: People in different countries with various linguistic and cultural backgrounds simultaneously discuss a wide range of issues relevant to society (European integration, migration, balance between security and civil rights etc.). Analysing discourse across different countries requires a multilingual and comparative approach to discourse analysis that combines qualitative and quantitative methods within a corpus-based, contrastive methodology. The aim of our project is to develop a gradually expandable platform providing methods for the contrastive analysis of cross-lingual discourses and to test it on specific discourse. For our pilot study, we chose vaccination for the following reasons: The argumentations of vaccine opponents and supporters are highly emotional, the topic combines general knowledge and specialist knowledge
and the settings in different countries differ strongly in some aspects (legal framework, common vaccination practices, actors, etc.); Based on a small journalistic corpus, we identify and analyse relevant discourse parameters in German, English, French, and Italian and develop an annotation scheme combining annotations on an auto-semantic micro level and an abstract, interpretation-based macro level. Our highly abstract parameters cover argumentation structure, emotion, main points of controversy, and discourse roles of the actors involved. ; We employ the annotation tool CATMA and work on the basis of comprehensive guidelines that we constantly revisit and develop. The annotations are then exported and statistically evaluated using the programming languages Python and R. For our evaluation, we correlate single parameters, but also use multivariate methods such as Principal Component Analysis and Multiple Correspondence Analysis. Correlations are evaluated for each language individually as well as across languages and shed light on the principal convergences and divergences between the different languages and cultures analysed.

References


Collins, L. (2015). How can semantic annotation help us to analyse the discourse of climate change in online user comments? Linguistik online, 70(1), 43-60.


Not my president: How names and titles frame political figures

Esther van den Berg, Katharina Korfhage, Josef Ruppenhofer, Michael Wiegand, Katja Markert
Leibniz ScienceCampus, ICL Heidelberg/IDS Mannheim, Germany

Naming and titling have been discussed in sociolinguistics as markers of status or solidarity. However, these functions have not been studied on a larger scale or for social media data. We collect a corpus of tweets mentioning presidents of six G20 countries by various naming forms. We show that naming variation relates to stance towards the president in a way that is suggestive of a framing effect mediated by respectfulness. This confirms sociolinguistic theory of naming and titling as markers of status.
On April 25 and May 12, 2015, Nepal was hit by strong earthquakes with magnitudes of up to 7.8Mw (moment magnitude scale) and 7.3Mw. These disastrous events killed and injured thousands of people, and destroyed buildings and infrastructure. They also had a devastating effect on the rich cultural heritage in the area. Together with the vivid and recent urbanization processes, Nepalese heritage sites are under constant threat and their sheer existence is at risk.

Among the many projects to support the people in Nepal, the “Nepal Heritage Documentation Project” (NHDP) focuses on the documentation of endangered historical monuments. It aims at developing and implementing the first comprehensive inventory of endangered monuments, especially monastic courtyards and arcaded resthouses.

Based in Heidelberg and Kathmandu, NHDP is a truly collaborative effort. The team includes experts from a broad range of fields: architects and monument conservationists, historians, linguists and anthropologists, as well as IT and Digital Humanities specialists. During the documentation process, in-situ data as well as architectural and historical data are combined with aspects of the monument’s cultural, religious, and social context.

Overall, NHDP will document about 400 monuments within the first two years. The data form the basis for a new Open Access “Digital Archive of Nepalese Architecture and Monuments” (DANAM). Findings from the rich historical and anthropological research carried out during the last decades, provided by academic partners, completes individual records. DANAM is an implementation of the ARCHES platform, an open source management system for cultural heritage (jointly developed by Getty Conservation Institute and World Monuments Fund). It is a hybrid system based on the CIDOC-CRM ontology.

NHDP hopes that its commitment to systematic documentation and data management, up-to-date standards of documentation, and Open Access will set the stage for sustainable preservation of monuments. Intense co-operations with Nepalese institutions, especially the Department of Archaeology of the Government of Nepal,
will create a new cultural heritage repository designed specifically for the management, preservation and protection of national and communal heritage.

In Germany, NHDP is formed through the cooperation of the Heidelberg Centre for Transcultural Studies (HCTS) and the Heidelberg Academy of Sciences and Humanities (HAdW). In Nepal, the Saraf Foundation of Himalayan Traditions and Culture is the key project partner and representative. NHDP is also supported by an international network of partners, including the Kathmandu Valley Preservation Trust, i3mainz at the Mainz University of Applied Sciences, and UNESCO Nepal, among others. The project was made possible by a generous grant from the charitable British foundation Arcadia.

References

Discrimination in online recruitment process

Faiz Ahamad
Tata Institute of Social Sciences, Mumbai, India

Any ideology has a profound impact on an individual’s attitudes, and beliefs. Political ideology, described as the set of beliefs about the proper and suitable order of society and the process to accomplish this order, is broadly categorized on a liberal-conservative continuum. People with different political ideology have different assumptions regarding organizational process and policies, e.g., liberals support hiring and the inclusion of women in the leadership role, etc., whereas conservatives do not. Numerous studies have found that conservatives are more pro self, whereas liberals are more pro-social in term of goal defining, and they have high concern for others. Hence liberals generally show a high level of cooperation, whereas conservatives do not. Similarly, studies in political psychology have found a relationship between political ideology and stereotypes, discrimination, inequality toward dissimilar group such as women, religious minorities who are generally perceived as value violators etc. In organizations, the political ideology of recruiter has shown association with discrimination during recruitment, promotion etc. which ultimately affects whole organization work culture and diversity. In comparison to a conservative, the liberals have been found to display no or very less support for discrimination or in-group favoritism in organizational process such as recruitment which is attributed to their ideology toward other gender and groups that everyone can succeed as a leader. Furthermore, with the rise of web 2.0 applications and internet technology, the electronic recruitment process is not free from discrimination, where many incidents of biases and discriminatory behavior have been found. But till yet no study has explored the impact of recruiter ideology on the perceived suitability of job applicants during the online job application process. Hence, the main objective of this study is to find out the impact of the online platform and ideological difference on discrimination faced by job seekers who are ideologically different from recruiter.
Research questions: The study aims to answer if the digital platform is used to discriminate against job seekers? How do conservatives and liberals take recruitment decisions in case of a job seeker who is ideologically different from them? Does online platform increase the chance of discrimination?

Methodology: First, we will do the sentiment analysis of online reviews data will be done. Data will be collected from a various social site such as twitter, facebook etc. In the second stage, an experimental study will be done to find if there are any biases against job seekers.

The “Redewiedergabe-Korpus“
Lukas Weimer, Annelen Brunner, Ngoc Duyen Tanja Tu
University of Würzburg, Würzburg, Germany

In the poster session, I present the “Redewiedergabe-Korpus“, a historical corpus consisting of German texts from 1840-1920 which was created by the DFG-funded project “Redewiedergabe” (http://redewiedergabe.de/). It is manually labelled for speech, thought and writing representation (STWR). With the help of the corpus, it should be possible, among others, to answer research questions about the development of different forms of STWR especially in the 19th century.

The corpus is available open access as a beta release since spring 2019 (https://github.com/redewiedergabe/corpus) and will have its final release in spring 2020. The corpus contains fictional and non-fictional texts. The source for the fictional part are narratives of Digitale Bibliothek edited by TextGrid (https://textgrid.de/digitale-bibliothek), for the non-fictional part newspapers and magazines, collected and edited by the University of Bremen, the Leibniz Institute for the German Language Mannheim and The German Text Archive. We created a corpus that is representative of many styles of STWR by pulling samples with an average length of 500 words from the text material. To enable various analyses, we assigned metadata to each sample.

To guarantee high-quality annotations, two primary annotators annotated each corpus sample individually and then a third person created a consensus annotation as the gold standard. This gold standard is the basis for all our evaluations. Similar to Semino/Short 2004 and Brunner 2015 we not only labelled the three media speech, thought and writing, but also four different types of representation: direct (He said: “What’s for lunch?”), indirect (He asked, what there was for lunch.), free indirect (What would it be good for lunch?) and reported (He talked about lunch.). Furthermore, we annotated speaker, expressions that introduce the STWR and some attributes like embedding or negations. We counted these annotations in total and per decade and thus gained a first overview of the distribution of STWR. The beta release of our corpus contains about 360,000 tokens – about 45,000 per decade – and almost 9,500 instances of STWR.

My poster shows these statistics visualized by graphics. In addition, I present first conclusions that emerge from the evaluations: For example, direct representation seems to be a preferred form in fictional texts, whereas the medium writing is
used more frequently in newspaper articles. In my poster, I document these results with data and try to justify them from different perspectives. Currently, we are creating an automatic recognizer for STWR based on both machine learning and rule-based approaches to automatically label texts of different time periods to get a broader view of the phenomena and provide even more in-depth insights.

References


Exploring network of comparative structures and comparative opinions in newspaper cuts

Erdal Ayan
Computational Linguistics, Kassel University, Germany

Topic modeling, comparative opinion mining and lexical semantic networks have lately been promising research areas that deserve closer attention in terms of their functionality and usability in descriptive research on big qualitative data. According to Varathan et. al. (2017), comparative opinion mining is still in its infancy. They state that, apart from some studies in Chinese, Korean and English, there has not been much research carried out on comparative sentence detection and learning in other languages by using supervised and unsupervised techniques in the field. They report that “more than 60% of the researchers who worked on comparative opinion mining are Chinese researchers.” (828). The studies on comparative opinion mining focus more on customer or consumers’ opinions on certain products (Jindal & Liu, 2006a, 2006b; Xu, Liao, Li, & Song, 2011) but there are very few studies (Balahur & Steinberger, 2009; Pollak, Coesemans, Daelemans, & Lavrač, 2011) on opinion mining presented in newspaper articles even though there are several descriptive studies (Burton & Langer, 1955; Gaur, Chand, Gaur, & Yadav, 2013; Lipinsky, 2013, 2014a, 2014b) conducted on forms, functions and representations of press clippings. The online content which produced in blogs, forums and social media sites might be easier to extract topic models, comparative opinions and networks when compared to newspaper articles which are more complex in terms of language used, the length of content created, more opinion focused and more effective in creating public opinion in the past. Authors may tend to use particular types of words and comparative forms to create topics while writing their articles. The types of words and comparative forms used should necessarily be context specific, namely fall within the scope and content of the newspaper in this sense, and embody the intentional representation of main topics and ideas produced in the body of the article. The current study is a descriptive and comparative media content analysis, based on qualitative data and corpus-based detection methodologies. Natural Language Processing are going to be used as two approaches in this comparative opinion mining research. In this regard, word types, comparative forms and topics may appear in a framework
of lexical-semantic networks frequently used by the researchers. They may also provide evidence for certain ideological differences among different countries. The main purpose of the research is to investigate topic models and comparative opinions, and explore the lexical-semantic relationships in the newspaper clippings digitized by the Herder Institute. Therefore, it has been hypothesized that exploring word types, comparative forms and lexical-semantic networks may help researchers understand how and what types of topics are used, and have emerged among the authors in terms of their selection of lexical items and topics for their articles during the Cold War Era. More importantly, the results of the study will explore how topic models in comparative opinions comparatively change over time in western and eastern newspapers during the Cold War era.

References


Mapping the historical monuments in Republican China: Travel, culture and modernity

Qiuzi Guo

Institute for East Asian Art History, Heidelberg University, Germany

My project intends to recreate the story of discovering the historical monuments during cultural investigation tour to Northwestern China in the Republican period utilizing GIS mapping technology and resources found in Shaanxi Art Museum and private collections. Digital tools help me to find a new approach to visualize my research question: which geographic locations that European-American explorers and Chinese explorers had discovered the monuments and what connections between the monuments and the specific locations in the map. I also explore some new questions from the building process: how to identify the differences between wartime investigation and foreign ethnographic expeditions in the early twentieth century. Through mapping the textual-visual journey during wartime, we cannot only explore the geographic locations that Chinese explorers had never visited before the war, but also better understand the cultural implications of the historical monuments and its connections with the specific geographical spots. The narrative map puts the documents, maps and photographs into a dynamic conversion that yield a new story of the wartime travel.

References


Big data, black boxes and bias.

Janine Aldous Arantes
University of Newcastle, Australia

The unprecedented amount of educational data being collected in and around teaching and learning in K-12 contexts has attracted significant interest across multiple domains. Commercial apps that use big data to ‘personalize’ advertising to teachers and develop insights on academic and behavioral outcomes in the classroom are increasingly common in Australian classrooms. This brings both opportunity and challenge. Approaching the research, from a perspective of exploring immanently ethical differences, the researcher aims to identify any interconnections (or lack thereof) between the commercialization of K-12 education, personalization (predictive analytics) and the changing role of the teacher. This is timely and significant, as although K-12 teachers across Australia are using such technology as part of their general practice (Selwyn, Nemorin, Bulfin, & Johnson, 2017), the implications of commercial predictive analytics have yet to be sufficiently debated (Hogan, Thompson, Sellar, & Lingard, 2018; Human Rights and Technology Issues Paper, July, 2018). The research focuses on two potential implications. Firstly, how teachers are aware of and understand implications associated with platform capitalism (Srnicek, 2017) and secondly, how teachers are negotiating potential impacts of algorithmic bias (Cheney-Lippold, 2011). Acknowledging that predictive analytics offer K-12 settings multiple benefits (Papamitsiou & Economides, 2014, Wolff, Zdrahal, Herrmannova, Kuzilek, & Hlosta, 2014), these implications were chosen as some suggest that there may be a ‘black boxed’ (Pasquale, 2015) perpetuation of social and historical inequity (Crawford & Schultz, 2014) in the insights provided to teachers, and that schools and teachers may be somewhat exploited (Williamson, 2017). When combined, these two implications suggest that there may be a phenomena occurring whereby rapid up-scaling and participatory online cultures (platform capitalism) may therefore result in apps that use or provide ‘black boxed’ inequitable insights being rapidly adopted and ‘trusted’ by teachers. To explore this phenomena, data was collected between November 2018 and April 2019, via 215 online surveys and 23 hour long semi structured interviews with Australian K-12 teachers as part of the ‘Apps in Australian Classrooms’ project. The data was thematically analysed alongside data collection. The next phase of analysis, sees the researcher participating in the HCH19 to discover ways to quantitatively analyse discourse on selected edtech’s social media. By analyzing selected edtech social media and blogs, the researcher aims to illuminate to what extent teacher behavior may or may not be modulated in relation to using such analytics and to what extent the apps raise awareness regarding challenges associated with algorithmic bias. The findings will then be used in combination with the themes identified in the interviews to guide a second round of interviews with the K-12 teachers. Therefore, the researcher is interested in talking with other students and experts in relation to quantitative approaches to discourse on social media and how to quantify social and personal impacts of predictive analytics and big data.
Formulaic sequences in Early Modern English: A corpus-driven historical pragmatic study

Ding Huang
Anglistisches Seminar, Heidelberg University, Germany

Even though language is creative, people also use many fixed, clichéd, situation-based multi-word units, or Formulaic Sequences (FSs), to say things. FSs in Present-Day English (PDE), being studied exhaustively, are found to make up a large proportion of both spoken (58.6%) and written (52.3%) discourse in PDE (Erman and Warren, 2000). So what about FSs in English of earlier periods?

Since FSs in historical texts are still an understudied topic, this PhD project aims at making further contribution to research on FSs in Early Modern English (EModE) and intends to determine what pragmatic functions FSs served in EModE and how FSs characterized different EModE text-types. More specifically, with a corpus-assisted approach, the project compares the use of FSs in face-to-face communication (both authentic and constructed dialogues) and written communication (letters).

Moreover, the project discusses the relationship between FSs and construction grammar (e.g. Goldberg, 2003), pattern grammar (e.g. Hunston and Su, 2017) and collocations, i.e. they are similar to and/or different from each other in two aspects: syntactical structure and the relationship between form, meaning and function. Generally speaking, grammar patterns contains both lexical units and grammatical elements (e.g. it v-link ADJ that), and ‘they relate to form only’ (in Hunston and Su, 2017). In construction grammar, constructions, a mapping of form and meaning, can be as fully lexical as words, phrases and sentences, and as abstract as grammatical
structures (e.g. DET-ADJ-N) and units containing words and grammar categories (e.g. a large amount of + N), etc. FSs, as well as collocations, are also constructions (Buerki, 2016). They are also constructions, are fully lexical and larger than one word. However, my project suggests that FSs are different from collocations because FSs not only map form to meaning, but also map form and meaning to functions. Therefore, it is a triangle relationship, i.e. a form-meaning-function mapping.

The project takes a two-phase procedure to search for FSs from the corpora: computationally retrieving lexical bundles (LBs) and manually identifying FSs based on LBs. The project argues that despite a useful way to investigate FSs in a massive amount of historical texts, LBs are, to some extent, different from FSs by definition: briefly, LBs are only frequency based, and semantically and syntactically incomplete, while FSs are more complicated and diverse in terms of the degree of idiomaticity and syntactic regularity. Therefore, although some recurring patterns could be found via examining LBs, LBs could not describe/account for the form-meaning-function mapping of FSs.

All in all, the project expects to find that certain pragmatic functions are served by different FSs in different registers. Extralinguistic factors such as the profession, age, social rank of the contributors to the texts are also taken into consideration.

References


Metapragmatic positioning in social media. A comparison of the German and French body discourse.
Vanessa Münch
German and French Studies, Heidelberg and Paris University, Germany

Social positioning is an essential basis for communication, negotiation and stabilisation of fundamental values and hierarchies. At the same time, it represents a means to structure and organise society and to constitute social groups, of which actors either want to be part of or from which they want to distinguish themselves (Spitzmüller 2013: 282). Within the scope of the doctoral project, the question investigated is how actors position themselves in a French-German comparison by a certain language usage or multimodal representation mode in the body discourse and how local practices of positioning are linked to discursive ones. By analysing a section of a multimodal body discourse, it is asked how actors position themselves in the public communication in fitness communities in social media.

The corpus of the doctoral project consists of data about the two German fitness stars Sophia Thiel and Anne Kissner as well as the two French fitness stars Sissy
Mua and Justine Gallice and contains different components (Instagram, YouTube, Facebook, Twitter). It is created using data from a period of one week. The investigated period starts on 7th January 2019 and ends on 13th January 2019. Further data will be gathered in the future.

The model of metapragmatic positioning according to Spitzmüller 2013 is used as a starting point and will be combined with approaches and concepts of other scientific disciplines (agonal characteristics in the paradigm of pragma-semiotic text work, enunciative-pragmatic tradition of Discourse Analysis in France, Sociology of the Body, Semiotics of Discourse, Online Discourse Analysis, Contrastive Discourse Linguistics and Corpus Linguistics). Consequently, the project positions itself in Contrastive Discourse Linguistics and at the interface with interdisciplinary research. Using an abductive approach, hypothesis will be generated in a three-level spiral movement. From there, prognoses will be derived which in turn will be examined inductively. This process is repeated until an appropriate explanatory model is found. Accordingly, the doctoral project pursues the aim of developing an interlingual Discourse Analysis model for multimodal positioning in social media theoretically and practically which takes the macro-, meso- and micro-level perspectives into account.

References


Cross-lingual information retrieval for multilingual access to archaeological data

Giulia Speranza
University of Naples ‘L’Orientale’, Naples, Italy

This project aims at investigate cultural heritage (CH) data in the field of archaeology in order to offer a multilingual digital access to reliable information about CH to the general public, thus contributing to make the cultural experience in museums or archaeological sites more interactive and inclusive, overcoming linguistic and cultural barriers in the communication process between cultural institutions and visitors. The starting point of this research is to analyse the multilingual resources available online as open data such as dictionaries, thesauri, term banks and to focus on a multidisciplinary approach, which ranges from the computational linguistics to the use of new technologies for translation in the field of Cross-Language Information Retrieval (CLIR), as well as the principles and methodologies of the Linguistic Linked Open Data (LLOD). An in-depth study of open data, ontologies and the CIDOC Conceptual Reference Model (CRM) for describing the relationships used in cultural heritage documentation, is a crucial aspect of the research project in order to implement a digital repository containing information about cultural objects in Italian, English and German. Based on the prepared terminology, a multilingual
Study on historical animal diversity and distribution pattern of Beijing-Tianjin-Hebei region

Ding Manni
History of science, University of Chinese Academy of Science, Beijing, China

Located on the northeast of China, the Beijing-Tianjin-Hebei region is China’s third largest economy after the Yangtze River delta region and Pearl River delta region. Since New China was born, rapid population growth and industrialization in this region have increased resource consumption and aggravated the contradiction between population, economic development and the ecological environment. This area is facing severe eco-environmental and development pressures including lack of natural resources, eco-environment deterioration and expanding development gaps. Therefore, the government proposes a coordinated development strategy for the Beijing-Tianjin-Hebei region and attaches great importance to the restoration of the ecological environment. This study provides a historical reference for ecological restoration in the region.

The Beijing-Tianjin-Hebei region includes Beijing, Tianjin and other 11 prefecture-level cities of Hebei Province. This area is located between 113°27'E to 119°50'E and between 36°05'N and 42°40'N. The terrain is high in the northwest and low in the southeast. It can be divided into three major geomorphic units: the Bashang Plateau, the Yanshan and Taihang Mountains, and the Hebei Plain. This area is a temperate continental monsoon climate. Most areas have four distinct seasons.

Especially the Beijing-Tianjin-Hebei region is the only region in China with plateaus, mountains, hills, plains, lakes and seashores. Rich landscape environment breeds a wide variety of ecosystems. In the material cycle and energy cycle of various ecosystems, mammals play a very important role. They are sensitive to environmental changes. Changes in the diversity of mammalian communities can reflect the quality of habitat changes and the degree of disturbance of human activities.

At present, most researches on the Beijing-Tianjin-Hebei region focus on environmental governance, urban community development, and ecosystem services. Among the animal-related studies are the aquatic community structure of Hebei Panda Reservoir, the deer and tiger of North China. No relevant research has been found on animal diversity in the historical period of the Beijing-Tianjin-Hebei region. This paper selects the beasts in the Beijing-Tianjin-Hebei region during the Qing Dynasty as a research object, trying to fill the gap in this research field.

I have collected and analyzed records of the mammalian community in the local gazetteers of the Beijing-Tianjin-Hebei region during the Qing Dynasty, mainly using literature retrieval, historical materials research, data mining, GIS. Above all, I set up a small database of historical animals. Through this database, I find the mammals distributed in the Beijing-Tianjin-Hebei region during the the Qing Dynasty were at least 7 orders, 19 families and 42 genera. Most of them are carnivorous species, followed by rodents and artiodactyla. The species of perissodactyla,
primates, erinaceomorpha and lagomorpha are less. The diversity of mammalian communities is relatively rich, and the distribution of species within each community is different. The main large mammals such as Carnivora and Artiodactyls are mainly distributed along the northern and Taihang Mountains. Small mammals such as foxes and rodentia are mainly distributed in the North China Plain.

References

Ding Manni, Cao Zhihong. Construction on Animal Diversity: Historical Database of Beijing-Tianjin-Hebei Region[C]. The 17th International Conference of Historical Geographers, 15.08-20.08.2018, Warsaw, Poland.


Ma D Z. Brief introduction on the compiling of the Qing History[M]. Social Science Front, 2009(11): 123-131


Gamifying a Croatian web dictionary
Josip Mihaljević
Institute of Croatian Language and Linguistics, Zagreb, Croatia

The purpose of my research is to present a model of gamification of the content of web-born dictionaries. This research will be conducted within the project Croatian web dictionary – Mrežnik of the Institute of Croatian Language and Linguistics, supported by the Croatian Science Foundation. Thus, at the beginning of my presentation, I will present briefly the Mrežnik project. The aim of the Mrežnik project is to create the first Croatian corpus-based and web-born dictionary. Mrežnik has three modules: the module for adult native speakers of Croatian, the module for elementary school children and module for foreigners learning Croatian. In all the modules educational web games with certain gamification elements (scoring, levels, time limit, badges, leaderboards) will be implemented within dictionary entries and on the dictionary website. These games will allow users to interact directly with dictionary content through certain game mechanics and elements which can be adjusted for different learning purposes. Educational games mostly consist of quizzes because they are simple and easy to make and there are a lot of web services such as Quizlet and Kahoot! that allow free and easy creation of quizzes. However, in a gamified dictionary, other types of games should be used in addition to language quizzes. On the other hand, language quizzes should also have explanations why something is the correct or wrong answer. The author will analyze the existing gamified content of dictionaries and encyclopedias aimed at foreigners and children to see how they
differ. The results of this research will be used to develop the gamification model of Mrežnik. In Mrežnik the plan is to have, in addition to quizzes, custom made educational games such as crosswords, memory, drag and drop, Tetris, Hangman, Tick-Tack-Toe, etc. Data for the dictionary is taken from the two Croatian corpora and data for the games is taken from dictionary definitions and examples. These games will help players in learning words, meanings, definitions, semantic relations (e.g. synonyms, antonyms, meronyms, hypernyms and hyponyms, feminatives, etc.). The technology used for the creation of the games will also be mentioned in the presentation. In addition, leaderboards can be created where users compete against each other within certain games. The author will also present his plans for research on students who learn Croatian as a foreign language at Croaticum – center for learning Croatian as a second and foreign language. Two groups of students (an experimental and a control group) will write the same test after the experimental group has been exposed to the gamified content. Their results and variance will be calculated with a repeated measure ANOVA in Excel. In the end, hopefully, the created model will be successful and it could be used for other web-born dictionaries and encyclopedias.

References

Interrogating identities: Case study of hashtag movements in India
Nanditha Narayanamoorthy
Department of Humanities, York University, Toronto, Canada

This research situates itself at the intersection of Comparative Cultural Studies, Digital Humanities, and Computational Social Science, and seeks to interrogate the complexities of identity formation on online platforms in the Southern Hemisphere. I explore the construction of collective and personal identities in the South, in the era of Big Data, through the study of digital activist movements in India. My intention with this research is a dialogue on both affordances and constraints of social media movements with a broader discussion of the presence and absence of representation, participation, and access in the Cyber-South.

I explore two potent examples of hashtag movements in India. The first is the queer movement of homosexual empathy and support surrounding the archaic, colonial law, Section 377 of the Indian Penal Code that until August 2018 promoted homosexuality as illegal and unnatural. The online digital activist wave writes itself on social media platforms as a veritable act of resistance to combat traditional heteronormative postcolonial power structures. I collect and analyze tweets that discuss public discourses surrounding the movement beginning in January 2018. I employ hashtags pertaining to the event in order to filter tweets from in and around
New Delhi, India, Toronto, Canada, and San Francisco, California to collect public opinion using tweets that were made publicly available at the time. The hashtags I employ in my analysis are #377quitindia, #article377, #homosexuality, #ipc377, #lgbtqia, #respectforlgbtq to eventually gauge at an understanding of the construction of a collective identity of the marginalized through a close reading of online discourses.

Second, I employ a similar methodology of data extraction using Twitter API wherein I collect and manually annotate tweets with #MeTooIndia beginning October 2018. I delve into the #MeToo Movement in India through the #MeTooIndia account, in the quest for alternative forms of feminist embodiment than those offered by traditional, dominant and masculinist models of identity delineation. I offer the #MeToo as a cyberfeminist and posthumanist solution to arrive at a place of complete breaking of identity categorizations as a model feminist movement in the South. Delving into the particular case study of the #MeToo movement in India enables an exploration of its evolution, its migration to the South and an examination of how women take control of their bodies and sexualities using digital affordances and ultimately reimagine their histories and their futures.

References


Twitter, politics and society in Iran

Hossein Kermani
University of Tehran, Iran

I am trying to analyse the dynamics and procedures of political activism on Persian Twitter. I argue that users are co-authoring and co-working in producing an imaginary political field on Persian Twitter which any battle and struggle on power lay down there. Therefore, my main and primary goal in this research is to unpack and map out this field. Moreover, I will analyse the strategies that various actors on Persian Twitter are employed to gain more power and increase their capitals (as discussed by Bourdieu (1984)). Furthermore, I will identify and assess the links between these communities and political parties and groups in Iran to sift their mutual and interrelated connections. Based on Bourdieu’s field theory (1984), I argue that actors in such a field are not individuals anymore, but networked actors which will be conceptualized as networked publics (Boyd, 2010). These actors are struggling on power in this field by performing networked practices (Van Der Haak, Parks, & Castells, 2012). Two of such practices include networked gatekeeping and framing were analysed in previous research. Having analysed these practices, I will introduce another practice performed by and shaped in networks: networked narrating. Exploring such practices in various and different networked publics on Persian Twitter
paves the way for us to probe those publics to which extent are trying to dominate the field.

In order to do this research, a corpus of 3 million and 476 thousand related tweets were captured by recruiting the Twitter's API from first to 25th of June 2017, during the last presidential election in Iran. I will use a mixed research method approach for the analysis, applying both computational methods (e.g., social network analysis and text mining) and qualitative methods (e.g., content and discourse analyses), to achieve a close and distant reading of tweet corpus. At the first stance, I will identify the different networked publics on Persian Twitter using cluster analysis. Moreover, I will employ a combination of social network methods and qualitative content analysis to assess the networked framing and gatekeeping in these publics. In this stage, I will use some networked centrality’s measures as PageRank to identify the most influential users in each public (Bruns & Stieglitz, 2013). Upon identifying those actors, I will analyze their tweets qualitatively to provide an explanation of their framing and gatekeeping practices. Next, I will apply narrative analysis on each public to explore their strategies in defining events and shape other people’s minds. Finally, I will investigate how these publics fight over power and capitals on this field using both network metrics and qualitative interpretations.

References


Emergent collectives via language - Decisions in networks and deliberation in democratic public

Jöran Landschoff
Heidelberg University, Germany

Political Sciences as well as journalists try to investigate collective political behaviour – especially in the field of the Political, and even more when it comes to election processes in Western Democracies. More often than not, citizens’ voting behaviour appears counter-intuitive, and predictions on the outcome of elections fail, as was seen after the unexpected success of Donald Trump in the last US campaign. Sociology has come to think of the individual to be embedded in a set of network ties that structure not only its social reality, but also its political beliefs as well as notions (Mützel/Fuhse 2010, 15). Political perspectives are therefore to be interpreted as supra-individual phenomena that develop in networks as emergent structures. It is the strong and weak ties within these formations that contribute
to the establishment of groups that share beliefs, perspectives, habitual schemas, and opinions (Stegbauer 2016, 1). It is argued here that these are at some point decided upon during a not necessarily conscious decision-making process of the group. Because the collective must somehow coordinate its preferences, it is assumed that the members of the collective rely (among other things) on speech as a means of homogenisation of divergent views. Because of this, the analysis of language use and its change is an adequate method to unveil the decision-making process as well as the underlying (political) opinions of a collective. This project will use methods of discourse analysis and combine them with network models to find collectives that are constituted by their similar use of semantic, grammatical and lexical elements. Deciding to use a certain lexical item instead of any other valid one is sediment of the choice for a certain perspective on subject matter (Müller 2011, 37). Established semantic content or lexical elements are reflections on established world views and indices for collective embeddings. The analysis will investigate collective decision-making processes in public internet discourse via language by investigating collective decisions for language. Network structures detected in this way will be correlated to networks of actor-referencing, i.e. who speaks and answers to whom. In this way, it is hoped that a method can be developed that adequately describes which actors in public discourse have discursive power as a result of their position within the network ties. Furthermore, cross-collective communication can be analysed and classified, and tested for its deliberative quality as a measurement for democratic quality.

References

Language and discourses of Forensic Psychiatry
Agnieszka Karlińska
Institute of Sociology, University of Warsaw, Poland

The objective of my project is to analyse forensic psychiatric reports as situated social texts and products of discursive practices. I will focus predominantly on their linguistic and discursive form to reconstruct the language patterns and discursive strategies of forensic psychiatrists. The main subject of the analysis will be the term of legal sanity. The evaluation of sanity is one of the main and most common tasks of forensic psychiatrists, which results are of decisive importance for criminal responsibility. In total, the study will cover about 150 reports with a volume of several to several dozen pages each.

The analysis will consist of two stages. I will combine quantitative and qualitative methods of text analysis by taking the approach called multi-strategy (Bryman,
2006) or mixed methods research (Creswell, 2003; Tashakkori and Teddlie, 2003). Firstly, I will investigate patterns in vocabulary, grammar and text structure to characterize the “genre” of Polish forensic psychiatric reports. Secondly, I will explore the discourses behind their narratives. I will employ the methods and tools of Corpus Linguistics, incl. generating frequency lists, identifying keywords and specifying dominant collocations (see Baker, 2006; Sinclair, 1991, 2003; Stubbs, 2001). I am also planning to use stylometric analysis and topic modelling provided by CLARIN-PL LTC (Topic, WebSty).

My research focuses on the relationship between legal and psychiatric interpretations of the same facts and the adaptation of medical-psychological language to the requirements of the justice system. Forensic psychiatrists have to prepare their texts for a “foreign audience” (lawyers, journalists, etc.) with significantly different education, disciplinary culture, and professional language. Therefore, they can be perceive as a “translators” – their task is to translate from psychiatric into legal language such that the court can align medical diagnosis with legal proceedings (see Verde et al., 2006; Griffith and Baranoski, 2007; Griffith et al., 2010). However, legal language and terms are difficult to reconcile with the diagnostic possibilities of modern psychiatry. Law is concerned with justice, fact-finding and the attribution of guilt. The subject matter of psychiatry concerns itself with human behaviour and mental disorder. During the judicial proceedings, these two distinct categorization systems collide.

Psychiatry often comes into the process of applying the law – the expert’s role is not only to examine the subject, but also to interpret and to draw conclusions regarding legal qualification of the act. Forensic psychiatrists play the role of “interpreters of crime”. They explain criminal act within the framework of a possible mental illness to give meaning to events that go beyond the legal, political, economic or moral order. It should be noted that in Poland there is no one operative model of forensic psychiatric report. Therefore, each forensic psychiatrist create own conceptual apparatus combining medical, psychological, sociological and pedagogical language.

References

From data to new languages: towards a multilingual fruition of culture through the social networks

Carola Carlino
University of Naples 'L'Orientale', Naples, Italy

The research topic is in line with the SNSI ("National Strategy for Intelligent Specialization") referring to the Area 6, Tourism, cultural heritage and creativity industry and in particular to technologies and applications for the conservation, management and enhancement of cultural, artistic and landscape assets. The proposal aims at encouraging the development of an innovative service with a high added value for cultural heritage, that is to say the access to multilingual information related to cultural heritage, thus expanding the potential visitor population.

To this end, a close relationship was established with the Italian society “Smart Apps” to identify the interdisciplinary research themes that could best be used to generate a mutual technology transfer of skills. The project intends to propose a research that carries out a survey on how data and information related to cultural heritage can be effectively processed for a better communication with the public. In the documentation of cultural heritage, the sources of information are formed both by data structured in metadata and by unstructured data and one of the objectives is to make the latter data sharable and available. Furthermore, the data and information available on cultural assets in public or private form are almost heterogeneous and rich, multilingual and interconnected with information in other domains. Cultural institutions could be equipped with appropriate digital applications that allow a more effective communication with the visitors. The research is highly interdisciplinary because it requires the interaction of fields of study such as Cross Language Information Retrieval, text linguistics, translation technologies, digital storytelling applied to social media. We want to develop an innovative methodology for a digital storytelling technology in the field of cultural heritage for Italian and English. The idea, co-designed with SmartApps and the Berlin School of Library and Information Science, Humboldt-Huniversitat, stems from the strong desire to enrich the studies of this sector in Italy.

The project aims at improving access to and dissemination of cultural heritage content and intends to propose a research that allows to relate the three main strategic areas of a cultural institution: cultural heritage with objectives related to conservation, management, care and enhancement up to physical and cultural accessibility; the public to reach, interest and encourage active and conscious involvement; the physical or digital network with individuals, institutions, the community.
Raise your voice - The relation of loudness in direct discourse and women’s emancipation

Svenja Guhr
GCDH, University of Göttingen, Germany

The treatment of narrative forms and discourse is an important criterion of narratology. As a new research field audionarratology deals with sounds experienced in silent reading. There are descriptions of noises made by industry or nature but also by human beings who are primarily indicated by the record of their conversations. Especially in prose, authors use direct speech to assign voices to their characters that can be heard in the reader’s imagination. A new approach to analyse loudness in prose is to focus the used verba dicendi as well as the adjectives and adverbs applied for the description of dialogues. They facilitate the interaction between text and the perceived voice of the characters therein, and show the readers, how they should imagine the designated manner of speaking, e.g. whether the characters are screaming or whispering.

Within a pilot study, a closer investigation is made into the power of digital opportunities in conveying information on loudness signals. The overall objective is to investigate if there is any relation between women’s way of speaking and the growing emancipation of women in Europe’s society from 1870 until 1910.

The study is based on an analysis of a corpus of prose with female characters written in German by male and female authors from 1870 until the beginning of the 20th century. Employing a rule-based method, the verba dicendi, their surrounding adjectives, adverbs and names of the speaking characters were extracted. The applied algorithm deals with the language-specific morphological properties of the verba dicendi regarding their construction and preferred application in past and present tense, as well as their proximity to direct discourse indicated by punctuation (e.g. quotation marks). Subsequently, the data obtained was matched with the results of a preceding online survey study (carried out in 2018). The survey focused on the loudness experienced by readers during silent reading of direct discourse introduced by different verba dicendi graded in generated percentages (from 100% schreien ‘to shout’ to 2% denken ‘to think’). In addition, the results of an investigation into the adjectives and adverbs in a window of four tokens before and after the relevant verb were matched with a loudness-indicating word list. These survey findings provided the basis for the creation of a loudness profile of an examined character.

With the investigation of a potential increase of loudness and quantity of women’s speech with growing emancipation in mind, this poster will focus on two minor parts of the project: The first is a comparison of the quantity of speech and of the loudness of female characters in the presence and absence of men compared to the behaviour of their male counterparts (all in relation to the number of their appearances in a novel). The second is an investigation into potential loudness increases regarding the social class differences between the examined characters.

References

Cultural practices and digital piracy in Yoruba society in Nigeria

Abdul-Rasheed Lanre Sulaiman
Department of Sociology, University of Ilorin, Kwara State, Nigeria

The significance of information and communication technology cannot be overemphasized. ICT is central to globalization, industrialisation and development. However, several factors continue to impede the full development of information technology markets. One of the cogent factors affecting the growth of IT market is digital piracy. Digital piracy could encourage cultural distortion and immoral behaviours. Piracy is a social activity as it involves circulation of videos, songs, computer software etc., among relatives, friends, co-workers and unknown persons. In a collective society, group influence forces its members to share their skills, talents and other resources. Inventors are obliged to share their innovations and improve the society as a whole. They are always conscious of the group. Members of such society would find it interesting and normal in sharing digital goods. While in individualistic society, individual ownership of innovations is the order of the day. Members of such society may not find it interesting and normal in sharing their developments including digital goods. It is therefore imperative to find out the extant cultural norms in the community and how it affects the rife of digital piracy. Numerous studies have been conducted on the causes of digital piracy especially from the legal perspective in Nigeria; the socio-cultural factors affecting digital piracy in Nigeria have not been well investigated. Thus, the study aims at examining the cultural norms affecting digital piracy in Yoruba society in Nigeria. The study will be anchored on social learning theory and Hofstede’s cultural dimensions theory. Mixed methods research design which consists of survey, in-depth interview and observation
methods will be used for the study. Multi stage sampling method will be used to select 384 respondents for the survey, while purposive sampling method will be used to select respondents for the in-depth interview. Questionnaire schedule, in-depth interview and observation guides will be used to elicit data from the respondents. Both quantitative and qualitative analytical methods will be used to analyse the data. MAXQDA software will be used to analyse the quantitative and qualitative data to be obtained. The quantitative analytical method will include univariate, bivariate, and multivariate levels of analyses. The univariate analysis will involve the use of frequency distributions, percentages and modes. The bivariate analysis will involve the use of chi square and contingency co-efficient, while multinomial logistics regression and odds ratio are the multivariate techniques that will be adopted. Contents analysis method will be used to analyze data from the in-depth interviews and observation.

Arctic Tamils: A new chapter in the anthropocene of the Far North
Rahul Pillai Sivashanmugham
Department of Geography, Parvatibai Chowgule College, Margao, Goa, India.

This narrative is about Sri Lankan ethnic Tamil refugees, who fled to the Norwegian Arctic starting late 1980’s and successfully made it their home since almost three decades. In doing so, these pioneering settlers in Arctic played their crucial part in the changing Arctic, bringing far away cultural and social experiences and anthropocene enterprise into the desolated Arctic North of Norway. Not only early Sri Lankan Tamil labor had a significant impact on the Arctic North of Norway, but these new environments so far away from their original tropical homelands had impacted the refugees themselves and their future generations growing-up in the multi-cultural Norway.

Following upon my interest on not just how refugees kept alive their networks, but also of their synergies with the Norwegian Arctic, I have looked at Northern Norway, particularly among the Sri Lankan Tamil Refugees who had settled down in the Finnmark region. The following are my major understanding I developed, and my observations from the many interviews conducted with these Sri Lankan Tamil ex-refugees (now Norwegian citizens) visiting their families, friends and place of birth in Northern Sri Lanka:

Early Sri Lankan Tamil refugees arriving in Norway, found employment in the far north, particularly in the fishing community. They started to settle down in these areas because while it had a thriving potential for fishing and allied industries like fish processing, there was acute lack of manpower. The lack of manpower in these fishing villages and towns in and around Finnmark was caused due to the harsh weather conditions and relative hardship of life, which caused younger ethnic Norwegians to migrate out of the region. Early Sri Lankan Tamil refugees found a foothold in these regions, and took up these professions which gave them assured employment and generated saving, which enabled them to bring their families too
from Sri Lanka into the Norwegian Arctic. Once settled, they learned Norwegian, and adapted their traditional customs into their new situation, creating a unique blend among the communities of Arctic.

After many years now, the people who came as the fleeing Sri Lankan Tamil refugees have generally prospered and blended in well as ‘model immigrants’. However, the very forces that pushed out ethnic Norwegians from the Arctic started coming into play, and the younger generations of these migrants are now leaving for the more comfortable life and facilities found in the cities of Southern Norway, like Oslo. Interestingly, this cycle of refugee adaptation in the Norwegian Arctic has proved to be an enduring pattern, and with each new wave of immigrants arriving in an immigrant-friendly Norway, many of them, be it from Afghanistan or Syria or elsewhere, have followed roughly the same trends of making their beginning in the physically difficult areas of the Norwegian Arctic, much like the early Sri Lankan Tamil refugees.

References


What makes Angela Merkel cry? – How Junk News shaped Europe perception in Chinese social media

Yi Zhu
Institute of Chinese Studies, University of Heidelberg, Germany

Why is Angela Merkel described in Chinese social media as ‘White Left’ and ‘Holy Mother’? Why do many Chinese people believe that ‘Europe is Islamizing at 230% speed’, while others dream of the ‘Golden Passport’ and thus consider Europe as their destination for immigration? There is increasing awareness among Western politicians and academics that social media have become the center stage of information consumption and are shaping public opinion on political issues. The same is true in China. Nowadays, more than 80% of Chinese Internet users consume news via social media. My study aims to explore the perception of Europe in Chinese social media using data derived from WeChat public accounts.

The super app WeChat dominates China’s cyberspace and had more than one billion active users in 2018 (according to its parent company Tencent). Aside from regular accounts for private users, WeChat offers so-called public accounts for media, businesses, government institutions, and individuals who operate ‘self-media’ and struggle hard for the attention of the general public.

Based on a database (provided by the WeChatscope project at the University of Hong Kong) with one million posts from the WeChat’s public accessible accounts in 2018, this project identified more than 11000 posts which refer to Europe or European countries. Through mixed methods, the study tries to answer the following research questions: who are the most active contributors to the discussion of Europe in Chinese social media? How did they frame Europe, what frames and frame
elements did they use to constitute the picture of Europe?

References

Fictional characters at the syntax-semantics interface
Anna Schmitt
Department of English and Linguistics, Mainz University, Germany

Fictional figures, according to scholars in the field, are introduced to stories by referential expressions such as proper names, definite descriptions (i.e. as appositions or adjective-noun patterns) or personal pronouns (Margolin 66). Whereas machines have learned to automatically retrieve references to names (NER) and pronouns (POS-tagging), the automatic retrieval of descriptions is a much more complex task for the machine. One reason for this is that description of fictional figures is rarely submitted in the form of (retrievable) lexemes but appears in more complex clusters spread over the text. Especially modernist writers seem to have resorted to a strategy of “synsemantic” phrasing (i.e. “Sätze ohne Zeigefeld”, Bühler 367). In these sentences, description is decoupled from an origo (cf. Mellman 119) as deictic centre and, resultingly, lacks clear (linguistic) reference. However, in many of these instances semantic objects can be identified instead.

In this proposed contribution, corpus methods will aide me to refine numbers of the strategies mentioned above. I want to gain numbers of frequency of (1) references to names, (2) pronouns and (3) definite descriptions employing a POS-tagged corpus of modernist texts and a reference corpus. Further, I want to retrieve numbers of noun phrases with a semantic object as head. This study will hopefully help me to carve out more finely the significance of two different strategies of writing fictional figures: a mode of explicit reference and a mode of implicit induction (Eibl). It may also further account for the perceived difficulty of modernism’s experiments of representing consciousness.

References
How sentimental are quotation expressions? A sentiment analysis
Ngoc Duyen Tanja Tu
Leibniz Institute for the German Language, Mannheim, Germany

Quotation expressions are expressions which refer to a direct or indirect speech representation (e.g. the verb says in She says: “I am happy.” or the phrase slap on the back in He slaps him on the back: “I am proud of you.”). Michel 1966 detects about 400 different quotation verbs which refer to direct and indirect speech representation with 5000 occurrences in 8 German novels. Jäger 1968 detects about 300 different quotation verbs which refer to indirect speech representation in 14 texts of different text types. The analyzes show us that the vocabulary of quotation expression is large. But what are the factors which determine the use of a certain quotation expression? Why not use only say? The poster presents one of the analyzes I conducted to answer this question. It shows a sentiment analysis on the quotation expression and the direct respectively indirect speech it refers. The goal is to answer the question how both relate to each other semantically. Also, the question how much information is stored in a quotation expression as well as the speech representation they refer. This lexicon was chosen because it does not only distinguish between positive and negative sentiments, but also between 8 emotions: 1) anger, 2) anticipation, 3) disgust, 4) fear, 5) joy, 6) sadness, 7) surprise and 8) trust.

As a data basis I extracted about 2000 quotation expressions from texts of the Redewiedergabe-Beta-Release-Korpus (available on GitHub: https://github.com/redewiedergabe/corpus/tree/v0.1.0-beta). The corpus consists of fictional (narratives) and non-fictional (newspaper and magazine articles) German text samples, each about 500 words long, and written between 1840-1920. The text samples are manually labelled for, among others, quotation expressions and the speech representation they refer.

References
This case study of developing a neural network lemmatizer for Early Irish data is a part of my PhD project ‘Deep Learning for Morphological Analysis of Under-Resourced Languages’.

One of the most serious problems one faces working on NLP tools for under-resourced languages is the lack of data. It is widely known that in machine learning the quality of a model largely depends on the size of the training corpus. The situation becomes even more complicated when it comes to medieval texts, since historical language data is not only sparse, but also very inconsistent.

Old Irish language, often referenced together with Middle Irish as “Early Irish”, has an extremely complicated grammar. Along with many verbal and nominal inflection classes, typical for most old languages, it presents such morphological phenomena as initial mutations and verbal infixation. The latter results into two sets of forms, called absolute and conjunct, in each mode and tense of a verb. It means that an average number of forms for each lemma in Old Irish will be substantially bigger than in many other European languages. Therefore, this task requires a considerable training corpus for any machine learning algorithm to work.

Is there any solution except manually annotating some digitized texts first, and then building ML-based NLP tools, or opting for rule-based systems? It seems like going down from word-level to character-level and using sequence-to-sequence learning (Sutskever, Vinyals, & Le, 2014) might help.

If we reformulate the lemmatization task as taking a sequence of characters (form) as input and generating another sequence of characters (lemma), we can forget about tens of verbal and nominal inflection classes, let alone spelling variation. Moreover, this approach allows us to use the electronic Dictionary of the Irish Language (Toner, Bondarenko, Fomin, & Torma, 2007), which lists some forms for every lemma, as source of data.

In the course of the experiment, a character-level sequence-to-sequence model was trained during 34,000 iterations but reached minimum loss and maximum accuracy of 69.8 % on a validation set after 10,000 iterations. When the training set (= known words) accuracy reached its maximum of 99.2 %, the validation set (= unknown words) accuracy dropped to 64.9 %; on the test set (= unknown words) the model showed an accuracy of 63.9 %. These results are a serious improvement over the rule-based model described in (Dereza, 2016), which showed only 45.2 % on unknown words. As poor as the scores may seem, they are not very different from those achieved by sequence-to-sequence models on analogous tasks, such as spelling correction and grapheme-to-phoneme translation, and meet the results of other NLP systems for historical data (Kestemont, De Pauw, van Nie, & Daelemans, 2016; Müller, Cotterell, Fraser, & Schütze, 2015).

References


51
Combining Cultural Educational Theory and Corpus Linguistics

Anna Moskvina

IwiSt (Computational Linguistics), Hildesheim University, Germany

In the course of this project we intend to extract textual evidence for cultural education processes from online book reviews. In order to do this, we mainly focus on the cultural education theory which is developed according to selective open axial coding in the spirit of Grounded Theory methodology. We conceive cultural education processes in line with Marotzki’s (1990: 52) description as processes where the conception which an individual has of himself or herself, and/or his/her conception of others and of the world around them undergo a change. We may identify cultural education processes when they are the topic of (parts of) book reviews, as shown by the arbitrarily chosen examples in (1) to (3).

(1) “[...] und dieses Buch hat mir die Augen geöffnet” (“[...] and this book was an eye-opener for me”)

(2) “[...] durch [...] bekommt man eine ganz andere Sichtweise auf solche Themen” (“[but] because of [...] you get a totally different perspective onto such topics”)

(3) “Ein Buch für alle, die [...] unsere emotionalen Reaktionen besser verstehen wollen” (“A book for everyone who wants to better understand [...] our emotional reactions”)

Examples (1) and (2) indicate the change of perspective, (3) the voluntary act.

To seed this search, we used 430 reviews from Amazon book reviews (McAuley, J., Leskovec, J. 2013) that have been manually annotated according to the components of reviews, their contents, the emotions of reviewers, their recommendations etc. (for more information on the annotation scheme see Kutzner et al. 2018). We analyzed sentences with four types of labels, all of which are likely to contain indicators of cultural education processes, and we manually extracted indicators from the respective sentences.

The extraction of indicators or patterns was not limited to the search for single words or word combinations. Instead, morphosyntactic variations (constituent order, passivization, etc.) and lexical variation (synonyms, related items, nominalizations, etc.) were accounted for. Additionally, the extraction must be aware of the context, due to the polysemy and to the broad range of contexts in which some of the indicator expressions can appear. For example, out of 82 occurrences of the word “Ratschlag” (“advice”), only 42 were about books giving advice on a given
topic. Such patterns were designed according to the CQP query language (Evert, 2011). In (1) we give an example of a search pattern for one of the categories:

(1) Pattern: (((lemma = “genauso”) | [lemma = “ebenso”]) [pos = “Adj”] [ ] 0,2 [pos = “Adj”])

This pattern matches any string that consists of either the word “genauso” or “ebenso” followed by several adjectives with a possible distance between them up to two tokens. Some patterns are combined with external sources like Germanet (Hamp, B., Feldweg H. 1997) in order to find more synonyms of a particular word, without losing the connection to the semantic field of the indicator, for example: “lernen” – “einarbeiten” – “nachmachen” – “studieren” – “erwerben” – “schulen” – “weiterlernen” (“learn” – “acquire”). The sentences found in this manner were later added to the manually annotated corpus and used as training and testing data for machine learning algorithms.

References
Hamp, B., & Feldweg, H. (1997). QSpiAccessible::accessibleEvent not handled: "8008" obj: QOjbect(0x0) "invalid interface!" xical semantic resources for NLP applications.

Virtual spaces and cultural transformations: Deconstructing YouTube (Rewind)
Hans Bouchard
Romance Studies, Ruhr-Universität Bochum, Germany

My PhD project aims to develop a spatial methodology of social platforms parting from a cultural, discursive as well as an infrastructural perspective in order to discuss mexican cultural representation and transformation processes on social platforms. This requires a mixed approach disassembling the multidimensional concept of space (cultural and media studies). For this I consider the various spatial dimensions of the social platforms, especially ‘formalized inscriptions’ (Van Dijck, 2013, p. 7), which define the criteria of relevance (likes, shares/retweets etc.) and the processes of ‘dataification of objects’ and ‘objectification of data’ (Hui, 2016, p. 50). However, those configurations overlap with other spatial dimensions and discourses like nation and identity (Anderson, 1991), which do sometimes converge in categories as ‘trending’ topics (‘trending in country X’). How do social platforms reconfigure our concept of space and how do we build spaced on formalized criteria? What are
the transformation processes of Mexican identity and space? Is there a mexicanidad (Paz, 1959) 2.0 and what are its basic criteria?

For the purpose of this summer school, I want to shift my attention on a specific topic in order to avoid language barriers and cultural contextualization (mexican discourse), while maintaining the theoretical framework of my PhD thesis. Here, I will focus on YouTube Rewind and aim to conclude with an ontological configuration: What is YouTube? What does YouTube want to be? What role do 'formalized inscriptions' play in this ontological configuration?

YouTube Rewind is the yearly video 'Review' of YouTube and a rendition of popular videos, creators and topics on YouTube. Nevertheless, this year’s video has become (by user 'rallying’) the most disliked video on YouTube; replacing Justin Biebers 'Baby'. This shows a clear disconnect between user and platform, which paved the way for alternative 'YouTube Rewinds’. One of the videos is PewDiePie’s take on a year review, which was met with incredible positive response. Unsurprisingly, it is also the result from constant user input that structures his reaction/commentary content (subreddit r/PewdiepieSubmissions). This consequently rises the question: Who is more ‘YouTube’? The platform, or one of its most representative content-creator with nearly 97 million subscribers? And what is the relation between platform infrastructure, content and user practices?

Therefore, I will analyze and compare both videopages of 'YouTube Rewind 2018’ by YouTube Spotlight and the adaptation by PewDiePie, to illustrate a mixed approach. What can we see by comparing the datasets of the comments in json format? What are other implicit references either coded within the mediality of the video, the comments and the platform? What are the limits and opportunities by those approaches and how can we trace the implicit references?

This approach consequently includes crawling and analyzing both video comments (directly via the YouTube API and with the YouTube comment scraper by Philip Klostermann), as well as an analysis of the content of both videos and their spatial configuration, highlighting crossreferentiality within videopages, comment sections, the platform and its 'own' culture.

References


Onigiri: a collaborative data integration software for the Humanities and the Social Sciences

William Diakité
Université de Rennes 2, France

While humanities scholars engage more and more into studies that involve computational approaches (i.e. distant reading), researchers face new difficulties when preparing their datasets. While data collection tools and techniques such as crawling and scraping tend to make their way into scholars practices, it becomes harder to create ensembles of data coming from different sources. Data integration is the task of joining such data based on an entity’s attributes (i.e. a person’s name). While common data integration applications are usually used for commercial purposes and use machine learning to rapidly match a large number entities across datasets, Onigiri has been developed to avoid false positives by letting the expert do the match via an easy-to-use interface. It has been designed so scholars can choose the matching method to apply across datasets stored as CSV files and resolve these matches collaboratively. This poster presents the problem that motivated the development of Onigiri and both the design and algorithmic choices that were made during its development.

The political rise of the European far-right? A data-driven examination

Ivo Bantel
University of Zurich, Switzerland

The European far-right achieved major electoral successes in the last decades—thus, its electoral rise is undisputable. However, a second, less obvious potential development is mostly neglected in academic literature: the far-right’s political rise, i.e. the success of far-right arguments, positions, and framings in political discourse. Investigating the political success of the European far-right, this project investigates two related research questions: I firstly ask to what extent has the European political discourse shifted to the right since the year 2000? Secondly, I proceed by comparatively scrutinizing the determinants for political discourse shifting to the right. To address the first question, I employ quantitative text analysis, ordering texts along a scale. This part’s data basis is provided by far-right positions for reference and non-partisan and aggregated positions, such as parliamentary reports and debates (excl. far-right contributions), thus representing the political spectrum as a whole. In detail, I investigate (i) dis/similarities between far-right and aggregated texts over time, (ii) conduct word scoring to compare classified reference texts with other texts to estimate the latters’ positions, and (iii) employ wordfish analysis to quantify whether aggregated positions have shifted to the right. The second part addresses the determinants for change by disaggregating the material and scrutinizing the variation in the DV (“shift to the far-right”) by introducing different IVs.
Investigating two sets of factors, I differentiate between factors which are endogenous to the political environment and exogenous factors. For the former category, I analyse (i) whether certain parties are more likely to shift to the right and (ii) whether shifts are more likely to occur on certain topics. For the latter category, I investigate whether (i) terrorist attack and (ii) refugee presence makes a shift to the far-right more likely.

Thus, the project attempts to estimate in two steps whether, in addition to electoral success, far-right positions have become politically more salient, successful, and efficacious beyond the far-right. The findings’ potential implications are substantial: a change in political discourse as a whole towards the far-right—which exhibits pronounced anti-democratic tendencies—represents a momentous challenge to democracy.

---

**Automatic processing of context-sensitive elements in task-specific speech interactions**

Lisa Nußbaumer  
Linguistics, University of Vienna, Austria

The question of what do autonomous systems (e.g. robots) need to have a human-like understanding of natural language has triggered an ongoing debate in the field of NLP. This interdisciplinary research attempts an integrative analysis of context-sensitive elements, i.e. pronouns, in spoken interactions merging considerations from the fields of formal linguistics and NLP. Pronouns (e.g. we, you, it) constitute an essential part of communication, but they are also subject to linguistic and extra-linguistic conditions that may render them ambiguous and difficult to interpret. In spoken interactions, speakers not only use words to provide information for the addressee, but also quite naturally various multimodal-signs, such as gestures, facial expressions and gaze etc. Those signs, however, are especially difficult to comprehend for nonhuman intelligent systems, but they can contain vital decoding information for the addressee. Without the ability to comprehend not only language input but also extra-linguistic context, spoken communication will at a certain point break down. It is thus crucial to include these multimodal factors in the processing of language information, if in the future machines should learn tasks from human teachers. The main objective of this research project, therefore, is to explore the question of how referring relations between pronouns and their referents are developed and interpreted in spontaneous speech to develop a better theoretical understanding of spoken language. The empirical data for the study will be collected via task-oriented speech experiments with (native German speaking) adult speakers. Findings from this study can then be further used in studies of automated language processing and anticipate potential improvements in resolution accuracy.
Semantic change and emerging tropes in a large corpus of New High German poetry

Thomas Haider, Steffen Eger
MPIEA Frankfurt / IMS Stuttgart / TU Darmstadt, Germany

Following in the footsteps of traditional poetry analysis, Natural Language Understanding (NLU) research has largely explored stylistic variation (Kaplan and Blei, 2007; Kao and Jurafsky, 2015), (over time) (Voigt and Jurafsky, 2013), with a focus on sound devices (McCurdy et al., 2015; Hench, 2017) and broadly canonised form features such as meter (Greene et al., 2010; Agirrezabal; et al., 2016; Estes and Hench, 2016) and rhyme (Reddy and Knight, 2011; Haider and; Kuhn, 2018), as well as enjambement (Ruiz et al., 2017) and noun+noun metaphor (Kesarwani et al., 2017).

However, poetry also lends itself well to semantic change analysis, as linguistic invention (Underwood and Sellers, 2012; Herbelot, 2014) and succinctness (Roberts, 2000) are at the core of poetic production. Poetic language is generally very dense, where concepts / ideas cannot be easily paraphrased. With a distributional semantics model, Herbelot (2014) finds that the coherence of poetry significantly differs from Wikipedia and random text, allowing the conclusion that poetry is – compared to ordinary language – unusual in its word choice, but still generally regarded comprehensible language. Recently, there has been research with topic models on poetry with Latent Dirichlet Allocation. Navarro-Colorado (2018) explores the overarching topical motifs in a corpus of Spanish sonnetts, while Haider (2019) sketches the evolution of topics over time in a German poetry corpus, identifying salient topics for certain literature periods and applying these for downstream learning how to date a poem.

Following in this vein, we offer a new method to explore poetic tropes, i.e. word pairs such as ‘love (is) magic’ that gain association strength (cosine similarity) over time, finding that most are gaining traction in the Romantic period. Further, we track the self-similarity of words, both with a change point analysis and by evaluating ‘total self-similarity’ of words over time. The former helps us to reconstruct literary periods, while the latter provides us with further evidence for the law of linearity of semantic change (Eger and Mehler, 2016) using our new method.

We do this with a model that learns diachronic word2vec embeddings jointly over all our time slots (Bamman et al., 2014), avoiding the need to compute the cosine similarity of two word vector representations on second order to align the embeddings.

Our contributions are: we (1) provide a large corpus of German poetry which consists of about 75k poems, ranging from the 16th to early 20th century with more than 11 million tokens. We then track semantic change in this corpus with (2) two self-similarity experiments and finally (3) by investigating the rise of tropes (e.g. ‘love is magic’) over time.

References

David M Kaplan and David M Blei. 2007. A computational approach to style in american poetry. In Seventh IEEE International Conference on Data Mining (ICDM 2007), pages 553–558. IEEE.

57


The process of institutionalisation of the term ‘State-Protecting Buddhism’ between 1945-2018

Hyein Lee
East Asian Studies, Freie Universität Berlin, Germany

This project investigates how ideas and discourses are circulated and reformed while complicating and shaping the identities of involved actors via a case of military initiatives of institutional Buddhism in South Korea. During the Meiji period (1868-1912), a Japanese Buddhologist used the term ‘state-protecting Buddhism’ to describe the convergence of polity and Buddhism in the modern context. Even though this term was used to disparage the nature of Korean Buddhism during the Japanese Colonial Rule of Korea (1910-1945), it carried increasing significance throughout the 1950s, 1960s, and 1970s, because it aligned closely with the ideologies of the authoritarian regimes of the time. It is still widely in use in media, text books, and especially in the institutions related to military initiatives of the Chogye Order of Korean Buddhism, the most influential Korean Buddhist Order. With ‘state-protecting Buddhism’ as its ethos, the Chogye Order established the Buddhist Chaplaincy Service in 2005, a military institution responsible for educating and selecting aspiring military personnel in addition to running Buddhist monasteries in conjunction with military camps. In this context, this project examines the process of institutionalisation of the term state-protecting Buddhism in South Korea between 1945-2018. Corpus and network analysis of co-occurrence words of ‘state-protecting Buddhism’ will be conducted in order to answer 1) how ‘state-protecting Buddhism’ became the tenet of the Chaplaincy Service 2) how the meanings, implications, and applications of ‘state-protecting Buddhism’ as a concept changed and 3) by which agents such changes were initiated. Print media, governmental documents, and major academic works written by opinion formers in which the term state-protecting Buddhism appears will be analyzed with Python.

How do the state functions change in the datafication era? Analysis of digital transformation strate

Weronika Przecherska
University of Warsaw, Poland

The purpose of this project is to describe how the state functions change in the datafication era. This paper investigates – using methods of quantitative and qualitative text analysis – how Estonia, Germany, Poland, and the United Kingdom define data and perceive its potential in the policy-making processes. Two definitions are of the key importance for the research – state and data.

According to Raciborski, the concept of the state is understood as: ‘[...] Government, consisting of all public authorities, central and local administration, army and police.” (Raciborski 2017: 11). Although Raciborski includes in his definition “army” and “police”, this research centers around the executive branch institutions.
The subject of the study requires also distinguishing tasks of the analyzed institutions. In this case, the reference point will be Zieliński’s (2017) conviction assuming that the state is in the era of the rapid information and communications technologies development “[...] the main broker and a gatekeeper of information [...]” (Zieliński 2017: 271).

Although to describe the vast amount of data the term “big data” is also frequently used, it is argued that the latter will lose its popularity in the future (Marr 2016: 1). Therefore, in the latest publications “data” and “information and communications technologies” began to be used interchangeably (Fabris 2018: 29-30). As the various terms are used in this regard, within this project the definition of data focusing on the process of datafication itself is being used:

“[...] a description of something that allows it to be recorded, analyzed, and reorganized. [...] So let’s call them datafication. To datafy a phenomenon is to put it in a quantified format so it can be tabulated and analyzed” (Mayer-Schönberger, Cukier 2013: 164).

Although public organizations produce huge amounts of data, at the same time they insufficiently use the data that is available to them – especially when compared to the business environments (World Bank 2019). This disproportion is particularly visible when taken into account the fact that the public sector possesses the largest amount of data that could be potentially used to solve social problems, develop artificial intelligence systems and, consequently, the European economy (European Commission 2019).

The usage of data by the state institutions is gradually becoming the subject of scientific research. The preliminary research within this paper enables – using qualitative and quantitative texts analysis – conceptualizing definitions, identifying key threads in opportunities for data usage, and mapping obstacles and challenges within the latter. The subject of the analysis will be countries’ strategies and key documents obtained from the web. Documents (e.g. programs, speeches, etc.) will be selected on the basis of keywords describing datafication processes in three languages. The analysis will be carried out using the program Atlas.ti. Within the quantitative part, the content analysis (e.g., with the usage of word cruncher function in Atlas.ti), correlations between particular elements of the text (e.g., with the usage of co-occurrence tool in Atlas.ti), as well as visualization of data (e.g. with usage of word clouds) are going to be applied (Friese 2013).

References


Multilingual ontologies for the representation and instantiation of annotation schemes for folk tale

Thierry Declerck
Multilinguality and Language Technology Lab, DFKI GmbH, Saarbrücken, Germany

We present a compact overview of the results of a series of Bachelor/Master theses and software projects conducted by students of the Computational Linguistics Department of the Saarland University, all related to the development of annotation schemes and ontologies for the representation and annotation of elements of folktales. More recently we started to consider the description of elements of drama.

In this work we formalized various seminal works in the broader field of folkloristics, like the 'Morphology of the tale’ by Vladimir Propp, the 'Motif-index of folk-literature’ by Stith Thompson and the 'The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson’ by Hans-Jörg Uther. And more recently we worked on the ontologisation of the '36 Dramatic Situations’ by Georges Polti. Those resources are currently used in a software project dealing with the use of neural networks for the classification or generation of multilingual folktales.

References


Matthias Lindemann, Stefan Grünewald, Thierry Declerck. Annotation and Classification of Locations in Folktales. In: Andrew U. Frank, Christine Ivanovic, Francesco Mambrini, Marco Pazzorotti, Caroline Sporleder (eds.) Proceedings of the Second Workshop on Corpus-Based Research in the Humanities, Vienna, Austria, Gerastree Proceedings, GTP 1., Academy Corpora of the Austrian Academy of Science, Sonnenfelsgasse 19, 1010 Wien, Austria, Vienna, 1/2018
Pushing for Insurgent Mobile Citizenship and Resisting the Pitfalls of Humanitarianism

Carolin Müller

Germanic Languages and Literatures, The Ohio State University, Dresden, Germany

Overview: The current age of migration and mobility encourages a discourse that investigates the social transformations resulting from the global transit of people. Cross-border movements remap demographic population patterns, but also the landscapes of cultural production, consumption and exchange. Artists around the globe contribute and also benefit from the capitalization of mobile cultural heritage, on the one hand, through art tourism, and, on the other hand, through collaborative productions. This account will shed light on the ambiguities, chances and challenges involved in cross-border artistic relationships through a case study of the practices and programs of the German musician-collective Banda Internationale and West-African musicians from Burkina Faso. This research draws on my observations during October 2018, when the band participated in joint music-making with West African musicians in Burkina Faso, the home country of one of its members. At the core of the project was the establishment of a collaborative reciprocal culture of gifting and exchange that utilized local resources in a way that would benefit local cultural workers sustainably, and promote a shared culture of music-making. At the same time, this research seeks to provide a critical reflection on the hypocrisies and dilemmas of humanitarian engagement through musical exchanges with postcolonial contexts from the West. By engaging with Didier Fassin’s analysis of “humanitarian reason” – the prioritization of a narrative of suffering to justify humanitarian engagement – this analysis interrogates the processes and structures of humanitarian musical exchange. To what extent can and does the musical exchange address the social issues that lie at the root of experienced inequality? I use ethnographic methods (semi-structured interviews and participant observation) to examine:

- What are the aspirations of musicians from the West and from the postcolonial context for engaging in musical exchanges? How do these differ or overlap?
- In what ways do practices toward achieving these aspirations change in the context of challenges? Can, and if so, how can mutual exchanges of knowledge and skills install frameworks of protection for artistic and artisanal property, and foster mobile musical citizenship?

and c) How does this project utilize the processes of mobility to build lasting international networks between artists from postcolonial societies and the western hemisphere?

With the support from the Mershon Center for International Security Studies I completed one month of on-site fieldwork in Burkina Faso. In addition, I observed preparation and follow-up over one year (2018) in order to observe change in practice and outcomes over time. The analysis identifies themes that organically emerge from the data to provide perspectives on the differently privileged capacity to deal with ambiguity and complex realities of humanitarian musical acts.

Significance of the Project: Different projects of humanitarian activism seek to establish sustainable relationships with underprivileged communities globally. Yet, much of these humanitarian politics interfere and intervene often more than they
work alongside to strengthen lost and deprived rights. A better understanding of the strategies utilized and challenges faced in reflecting on actors’ own privilege may be achieved by turning the lens to the practices of cross-border musical exchange. This research serves to educate the global community that it is imperative to inform artistic engagement across borders on the pitfalls of humanitarian engagement to resist exploitative models of artistic advancement and to subvert the restrictive distribution of funds toward insurgent acts of citizenship.

Investigating an Internationalism: Corpus-based Studies on OKAY in the Wikipedia
Laura Herzberg
German Chair of Linguistics, University of Mannheim, Germany

OKAY has its origin in the American English language as a deliberately misspelled abbreviation for “all correct” (Read 1963). Since its creation in 1839, it has spread into many languages of the world with spellings and pronunciations adapted to the respective languages, e.g. okra in Finnish or okej in Swedish (Read 1963, Metcalf 2011). Over time, OKAY has developed various functions and meanings which are described e.g. in Metcalf 2011. In general, most of the empirical studies investigate OKAY in the spoken English language (e.g. Schegloff/Sacks 1973, Beach 1993, Condon/Cech 2007); there are few other studies on OKAY in other languages, for example in French and German, but still with a strong focus on OKAY in spoken language (e.g. Delahaie 2009, Kaiser 2011).

The presented PhD research project aims at broadening the description of OKAY’s functional range in written Computer-Mediated Communication (CMC), more precisely, on Wikipedia talk pages. On these talk pages in particular, on which users interactively exchange information and negotiate edits of the related articles (Storrer 2018), the functional diversity of OKAY comes to the fore. Building up on Herzberg (2016), who investigated the usage of OKAY in German Wikipedia articles in comparison to its usage in German Wikipedia talk pages, OKAY is further investigated in digital communication. The project sheds light on the practices of the diverse OKAY forms, positions and functions by starting with a profound overview. This overview is developed by analysing OKAY in written CMC. It will be also traced how OKAY disseminated to other digital interaction environments, e.g. to voice control systems such as ‘OK Google’. The aim of this step is to develop a categorical system which displays the forms and functions of OKAY in order to expose OKAY functions specific to the CMC genre. In a next step, the categorical system will be broadened to include OKAY usages found in the English and French CMC data. The goal of this comparative work step is to determine OKAY usages that are specific to a certain language in contrast to language-universal OKAYs. Using the classification of OKAY across languages as a starting point, it will be then zoomed into more specific OKAY patterns, e.g. language-specific co-occurrences of OKAY such as well okay, ja okay, oui ok or the application of OKAY in conflictual discussions in which OKAY is applied as a politeness strategy in order to introduce criticism in a neutral manner.
Besides the Wikipedia corpora, the Research and Teaching Corpus of Spoken German FOLK is used as a comparative corpus resource. The Wikipedia corpus collection is available at the Institute for the German language (IDS). In my investigations, copies of the English, French and German Wikipedia talk pages are used. These corpora are queried via the Corpus Research and Management System COSMAS II which is additionally used for the co-occurrence analyses.

References

Corpora and tools:
Institute for the German language Mannheim.
FOLK: Forschungs- und Lehrkorpus für gesprochenes Deutsch. Institute for the German language Mannheim.
Available via the DGD.

Locally Compiled (Learner) Corpora in Foreign Language Teaching: Research- and Usage-Based Software

Ingo Kleiber
Heidelberg School of Education, Heidelberg University, Germany

Locally compiled (learner) corpora, i.e. language corpora compiled by learners and teachers themselves, have considerable potential for both language teaching and learning as well as for the analysis of learner language in general. However, there is still a wide gap between state-of-the-art (linguistic and didactic) research and current classroom practices with regard to the adoption of corpus linguistic approaches to foreign language teaching.

My research project aims at bringing together modern theory, methods, and approaches from the fields of (corpus) linguistics, computational linguistics, second language acquisition, and (foreign language) teaching. The two main goals of this research project are the development and testing of both a new linguistic and didactic
model as well as the development of software and methodology that allow for the implementation of the model.

In this presentation, following a more general introduction to the topic of locally compiled (learner) corpora, I will present the research- and usage-based software architecture designed for this project. The key question to be answered will be how current linguistic and didactic theory can and must be used as the starting point for the successful development and implementation of a new software platform targeted towards language learners and teachers. In order to this, individual components of the architecture will be discussed based on a small, locally compiled corpus of student essays.

Natural language processing techniques applied to medieval charters studies
Sergio Torres Aguilar
LIMSI, Paris-Saclay, Paris, France

We propose to create an interaction model for the automatic generation of annotations and the indexation of textual structures in the digital edition of medieval documents. Building on recent progress in NER (Named Entities Recognition), a technique of computational linguistics, and using of natural language processing techniques (NPL), we want to recognize the emergence of personal and geographical names in medieval charters and, at same time, make a classification, for the propose of semantic annotation, based on the fixed forms used in the internal structure of texts. For this purpose, we want to build two automatic recognition models: one operating on named entities detection and recognition and another operating on detection of discourse formulaic parts used in medieval charters. We approach supervised machine learning methods using hand-annotated French medieval corpora from 9th to 13th century (CBMA-Cluny charters) including ecclesiastical and seigneurial charters (donations, sales, exchanges...), chancery archives, royal decrees, or notarial documents. These types of texts are quite appropriate because they combine fixed forms and formulae with rich nominal information. We want to explore the capabilities of NER on a new documentary typology rarely treated in the literature in order to generate a more accurate and reproducible process of annotation, but also to analyze the textual structure in order to recover, by means of a search engine, a level of information more advanced that the plain textual offered by an erudite edition. By the means of our results we demonstrate in several studies how both models can be master pieces to build best computer assisted method to date undated charters; to represent exhaustively the evolutions in the vocabularies and in the scriptural uses; to discover macro movements in social and spatial realities or to design better infographics based on topographic reconstructions.

References
Towards Arabic Text Simplification System

Nouran Khallaf
School of languages, University of Leeds, UK

Simplifying a text is the process of reducing its linguistic complexity while maintaining its meaning and original information (Saggion, 2017). Text simplification has very important roles; such as its usage in designing and simplifying the language curriculum for both second language and first language learners, to make text easy-to-read for first language users with cognitive impairments and low literacy language level. Moreover, evidence suggests that Automatic Text Simplification is (i) a fundamental pre-process in NLP applications such as text retrieval, extraction, summarization, categorization and translation (Saggion, 2017), and (ii). a post-process in Automatic speech recognition. Therefore, the major aim of this study is building an automatic Arabic Text Simplification system by applying robust NLP techniques.

Automatic Arabic Text simplification is a very challenging process because, it is a highly morphologically rich language with flexible word order, the multifunctionality of Arabic nouns and the lack of vocalisation diacritics in most text. Unlike English, only a few researchers have been tackling the problems of Arabic text simplification. In this area of research for Arabic, found only a prototype unreleased system by Al-Subaihin and Al-Khalifa (2011) at King Saud University which is inaccessible and anther starting project by Al Khalil, Habash, and Saddiki (2017)at New York University in Abu-Dhabi.

The proposed Arabic Text Simplification system architecture is composed of four main steps, pre-processing stage, lexical and semantic complexity analyzer, syntactic complexity analyzer, and Natural Language Generation Module. It will be implemented using Python with a third party Python libraries such as scikit-learn and NLTK toolkit, MADAMIRA Arabic morphological analyser (Pasha et al., 2014), and CAMeL dependency parser (Shahrour, Khalifa, Taji, & Habash, 2016). For Evaluation process two evaluation measurements would be applied, firstly The Common European Framework of Reference for Languages (CEFR) readability measures to ensure that the produced text is much simpler than the processed text, secondly, by testing the intelligibility of the transformed text with a group of Arabic Second language learners.

This project will provide a text simplification tool that will be used by a wide range of users. The target group will be learners of Arabic as a foreign language since this tool will assist them to understand complex Arabic texts leading them to master the Arabic language. It can help various groups of people, including children, the functionally illiterate, and people with cognitive disabilities, in such case the tool will be an aid for them to simplify the complex Arabic text to be more understandable and easier to read.

References


The Do Re Mi at the workplace

Yasmin Soni
Tata Institute of Social Sciences, TISS, Deonar, Mumbai

Today firms are trying their best to capitalize on their human capital and finding innovative ways to reduce the per unit cost of the product by increasing the productivity of workers. There are various monetary and non-monetary ways to increase productivity which includes incentives, recognition, motivation, felicitation. One of the innovative ways to increase the productivity of workers is by conditioning the work with some form of rhythm. Music gives rhythm to work and it also induces a positive state of mind where the individual becomes more receptive towards the task. Corhan and Gounard (1976) demonstrated that rock music improves performance on a signal detection task compared to relaxing instrumental music. Davies et al. found that during a difficult visual vigilance task, music exposure prevented detection latencies that were evident during no music (1973).

Objectives

- To examine the impact of music listening on shop floor workers.
- To investigate the impact of music listening on the mood of the shop floor workers.
- To examine the impact of the enhanced mood on the productivity of the shop floor workers.
- To investigate if the different genre of music can elicit a differential impact on the productivity of the shop floor workers.
- To investigate what kind of music induces maximum productivity.

Plan of the experiment

It is a quasi-experimental, pre-test-post-test study. It’s a quantitative study which includes observation and group discussion to support the results. Plan of the
Experiment: Experiment phase for Study 1 and Study 2 had three weeks, Study 1 had 6 working days while Study 2 had 7 working days. These days in a week were divided further based on the categories of music. For a music playing week, on day 1 we played reflexive and complex, day 2 we played upbeat and conventional, day 3 we played energetic and rhythmic, for day 4 and 5 we played soundtracks or instrumental music, day 6 and day 7 was kept for the preferred music of the workers, obtained through music preference scale.

Results and findings

The various test was conducted including, T-test, Regression, Hausman Statistics, Quantile regression which derived thought-provoking results. Studies show a similar trend of increased productivity when music was played and lower productivity when music was not played. we can conclude from this analysis that workers generally have a negative change in their mood after a hectic day at the workplace, but if we intervene with any stimulation, which in this case is music listening, then a positive change in the mood can be observed we could establish that a change in mood of the workers can have an impact on the productivity. An interesting outcome of this study is that people who are less productive could not enhance their productivity with a change in mood due to music listening. Though music listening has enhanced their mood, it was not translated into better productivity.

Important themes coming out of observation and group discussion are emotional affiliation, social bonding, religious inclination, mood enhancement.

References

Corhan, C.M., & Gounard, R. B. (1976) Type of music, schedules of background stimulation, and visual vigilance performance, Perceptual and Motor Skills 42, 662.


Introduction to 3D Modeling Techniques

Hubert Mara
Forensic Computational Geometry Laboratory, Heidelberg University, Germany

In museums, collections, libraries and archaeological excavations the daily use of 3D scanners is rapidly increasing to acquire precise 3D models of tangible heritage. This documentation in 3D is the basis of in-depth analysis of objects like cuneiform tablets, roman ceramics, Greek sealings or medieval inscriptions. The visualization and extraction of characters, symbols, traces of wear or usage is key to provide information about e.g ancient trade networks or manufacturing techniques. Therefore we will introduce the core concepts of 3D data processing to machine learning for character recognition on the example of cuneiform script. A second example are similarity metrics for Minoan and Mycenaean sealings. Especially the visualization of a similarity network of Maya glyph s will illustrate its assistance in the Text Database and Dictionary of Classic Mayan (TWKM, housed at the University of
Bonn). Additionally we will have an analytic hermetic discussion of the underlying computational methods.

Sample 3D-data for the hands-on session will be provided, which will be inspected and visualized with our GigaMesh Software Framework. Linux packages and Windows binaries are freely available at https://gigamesh.eu and can be installed in advance. The sample data will be prepared to be suitable for processing on standard PCs/Laptops. Comprehensive large 3D-datasets for cuneiform script are available in advance via the Heidelberg University Library platforms Hei-DATA (https://doi.org/10.11588/data/IE8CCN) and HeidICON (https://doi.org/10.11588/heidicon.hilprecht). Please note, that these high-resolution real world data-sets require a certain minimum of CPU and GPU computing power.
Day 5 – Friday: Methodology of Computational Humanities I

Morning Sessions: Workshops

Providing new views on textual data with knowledge graphs

Leo Born, Juri Opitz
Computational Linguistics, Heidelberg University, Germany

When (historical) text collections grow on scale, it can become increasingly more difficult to obtain a bigger picture without quantitative analyses. Computational methods and tools can aid us in achieving bigger pictures. To that end, we distinguish between (i) an explorative methodology which aims at helping in formulating research questions by offering new and structured views on text collections, and (ii) a research question-driven methodology which uses computational methods and tools in a task-specific way in order to address one or several research questions.

Practically, the workshop covers two aspects:

- FROM TEXT TO TRIPLES: a hands-on introduction into the Python NLP package spacy by using it to construct a simple knowledge graph from a corpus of historical documents.

- FROM TRIPLES TO THESES: analyses of the resulting knowledge graph using the package networkx and using visualizations of the knowledge graph in a dedicated web application.

In the course of the workshop, particular emphasis will be put on the constraints research questions pose on these methods as well as constraints that these methods pose on research questions. Furthermore, we will raise the following important questions: What do we want to and what can we model? How can we harness the strengths of computational statistical methods? How can we minimize or – more fundamentally – become aware of biases which may distort the perception of the extracted structures in unwanted ways?

Prerequisites: basic knowledge of the Python programming language
Afternoon Sessions: Lectures and Panel

Visualisation critique and critical visualisations in the context of digital cultural heritage

Katrin Glinka
Stiftung Preußischer Kulturbesitz, Berlin, Germany

The talk builds on a common reading of data- and information visualisation as graphic representation of complex data and contextualised knowledge. Visualisations are consequently conceived quite literally as “images” and in this respect related to scientific methods in visual studies and art history where images are treated as visual sources that need to be interpreted or act as visual arguments in scholarly discourses. In this context, the talk illustrates how visualisation research and the humanities could profit from a reciprocal methodological exchange. It emphasises the need to foster critical approaches to visualisation on two levels: first of all, the critical interpretation of visualisations in the context of source criticism and visualisation literacy. Second of all, in terms of the actual creation of visualisations and exploratory interfaces that specifically act as a means to challenge conventions, encourage reflection, and support critical intervention.

Visual interactive analysis of verbatim text transcripts

Mennatallah El-Assady
Data Analysis and Visualization, Konstanz University, Germany

Verbatim text transcripts capture the rapid exchange of opinions, arguments, and information among participants of a conversation. As a form of communication that is based on social interaction, multiparty conversations are characterized by an incremental development of their content structure. In contrast to highly-edited text data (e.g., literary, scientific, and technical publications), verbatim text transcripts contain non-standard lexical items and syntactic patterns. Thus, analyzing these transcripts automatically introduces multiple challenges.

In this talk, I will present approaches developed (in context of the VisArgue project) to enable humanities and social science scholars to get different perspectives on verbatim text data in order to capture strategies of successful rhetoric and argumentation. To analyze why specific discourse patterns occur in a transcript, three main pillars of communication are studied through answering the following questions: (1) What is being said? (2) How is it being said? (3) By whom is it being said?

In addition to reporting on visualization techniques for the analysis of conversation dynamics, I will argue for the importance of tuning automatic content analysis models to unique textual characteristics, appearing, for example, in verbatim text transcripts. In particular, I will present a visual analytics framework for the progressive learning of topic modeling parameters. Our human-in-the-loop process simplifies the model tuning task through intuitive user feedback on the relationship
between topics and documents.

Closing Panel “Science for Society: Transfer, Innovation, Relevance”

Moderation: Eva Wolfangel

Panelists: Participants of the Summer School
Getting around and Maps

Overview Map

The map below shows all the relevant sites for the summer school. As you can see the venue for lectures, the Mathematikon (blue dot), is located north of the Neckar river, whereas the other three sites are located south of it: the hotel (yellow with white star inside) is straight to the south; Hugo, the restaurant for the conference dinner, is in the Weststadt district (red dot); and the Heidelberg Center for American Studies, the venue for the panel (green circle with white cross), is located in the Old Town.
Query public transportation schedules

In the following pages we give instructions on how to get around. We chose what we thought were the best options and not showing all possible ways to make these trips.

If you want or need to find out more about how to get around in Heidelberg on public transportation, one very good source of information is the local public transport system. They provide a free app with an English language interface for your mobile phone: myvrn is available both in Apple’s as well as Google’s app store.

You can also visit the public transport’s website and query schedule information there at https://www.vrn.de/

Alternatively, you can get the schedule also by way of Google Maps.
Connections from the B&B hotel to the venue

We list two attractive options for getting from the hotel to the venue. The relevant bus and tram stops are marked on the map below.

![Figure 1: Neighborhood around B&B Hotel (with stops)](image)

**Option 1: Tram 26 + Tram 24**

This service is available roughly every 10 minutes. It requires one transfer.

- Go to the tram stop **RUDOLF-DIESEL-STRASSE** (red circle at lower right)

  Coming out of the B&B Hotel, you head south east on **Rudolf-Dieselstraße** and then turn left onto **Hebelstraße**: the tram stop is near the corner.

- Board Tram 26 in the direction of **BISMARCKPLATZ**

- Get off at the 4th stop, called **BETRIEBSHOF**

- Transfer to Tram 24 in the direction of **HANDSCHUHSHEIM**

  The direction might also be given as **HANS-THOMA-PLATZ** or **BURGSTRASSE**: that’s also ok, you’re still going in the right direction north across the Neckar.

77
• Get off at the 2nd stop, BUNSENGYMNASIUM.

You will see the Mathematikon across the street from the tram stop.

Option 2: Bus 721

This service runs only once an hour, usually at 15 minutes past the full hour. It is the shortest route though and doesn’t require you to make any transfers.

• Go to the bus stop SCHWETZINGER TERRASSE (yellow circle at upper left of map in Figure 1)

Coming out of the B&B Hotel, you head out north west on Rudolf-Diesel-Straße. Pretty soon you’ll cross a big road called Speyer Straße, on the other side of which Rudolf-Diesel-Straße continues under the name Langer Anger. Head along for a short stretch and you’ll find the stop for the 721 bus.

• Board Bus 721 in the direction of BUNSENGYMNASIUM WEST.

Don’t get on any bus that gives WALLDORF as its direction – it would take you south instead of north. Instead find the bus stop on the other side of the street.

• Get off at the last stop, called BUNSENGYMNASIUM

The stop is actually located in Mönchhofstraße: head west across Berliner Straße, where Mönchhofstraße continues as Im Neuenheimer Feld. You’re now on the university campus and the Mathematikon is right there on the corner of Berliner Straße and Im Neuenheimer Feld.
Connections from the old town to the conference venue

<table>
<thead>
<tr>
<th>Hotel Schnookeloch</th>
<th>Hotel Backmulde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haspelgasse 8</td>
<td>Schiffgasse 11</td>
</tr>
<tr>
<td>69117 Heidelberg</td>
<td>69117 Heidelberg</td>
</tr>
</tbody>
</table>

Figure 2: Neighborhood around Schnookeloch and Backmulde.

The Bus Stop Universitätsplatz is within 5-minute walking distance from both hotels. The Bus Line 31 (direction “Neuenheim Chirurgische Klinik, Heidelberg”) runs every 10 minutes. Get off at the Stop Bunsengymnasium. The Mathematikon is located right across the street on Berliner Straße (building number INF 205).

From Altstadt to conference dinner venue

If you would like to take a stroll down Hauptstraße, the dinner venue is conveniently located in a 20min walking distance from the hotels in the Old Town:

HUGO Wine & Dine  
Rohrbacher Str. 47  
69115 Heidelberg

Another option are the bus lines 31, 32 and 33 from Universitätsplatz.

Get off at Bismarckplatz. Head south-west, past the Galeria Kaufhof department store and cross the big intersection onto Rohrbacher Straße. Continue south on Rohrbacher Straße until you find HUGO on the left side of the street.
Immediate neighborhood of Mathematikon

The map below shows the surroundings of the summer school’s venue.

- The Mathematikon is the set of buildings to the west of Berliner Straße. The part of the Mathematikon where our lectures will take place is marked by the red blob. Note that the more northern parts of the Mathematikon include grocery and drug stores (note the shopping cart symbol).

- The blue dot with the white star inside marks the tram stop where you get off when you arrive from south of the river, i.e. from town. (The tram runs between the lanes for the cars.)

- The green circle with white square inside marks the tram stop where you get on if you want to return south of the river.

- The orange square with white cross inside to the right shows the stop for busses going towards our downtown venues, that is, the panel venue and the dinner venue.
Panel venue

The public panel will be held at the Heidelberg Center for American Studies (HCA). It is located in the Curt and Heidemarie Engelhorn Palais at Hauptstraße 120 in the Old Town. As the overview map shows, you have to go south across the river and then east to get there.

Option 1: Bus 31

This service is available every 10 minutes.

- From the Mathematikon, go to the bus stop BUNSENGYMNASIUM

  Starting from Im Neuenheimer Feld, head east and cross Berliner Straße. On the other side of Berliner Straße, the street changes names to Mönchhofstraße. Very close to the intersection you’ll find the bus stop on the right side of the street.

- Board Bus 31 in the direction of UNIVERSITÄTSPLATZ.

- Get off at the 8th stop, MARSTALLSTRASSE. (Point A on the map below.)

- Follow Marstallstraße south until Hauptstraße, where you take a right. After a short distance, HCA will be on your right. (Point B)

It’s about 200 meters from the bus stop MARSTALLSTASSE to HCA.
Dinner venue

Hugo’s is located at Rohrbacher Str. 47. It’s south of the river in the part of Heidelberg called Weststadt. On the map below, it’s the blue point marked B.

**Option 1: Bus 29**

To get there, the best option is to take Bus 29. In the early evening it runs every 20 minutes at 06, 26, 46 minutes after the full hour.

- Find the bus stop for Bus 29.
  
The stop is located in Mönchhofstraße, which is the continuation of Im Neuenheimer Feld east of Berliner Straße.
- Get on in the direction of BOXBERG
- Get off at the 7th stop, called KAISERSTRASSE, which is located on Rohrbacher Straße (red dot with white square on map)
- Head back north on Rohrbacher Straße for 120 meters: Hugo’s is on the right side of the street.

**Option 2: Bus 31**

- Go to the bus stop BUNSEGYMNASIUM
  
Starting from Im Neuenheimer Feld, head east and cross Berliner Straße. On the other side of Berliner Straße, the street changes names to Mönchhofstraße. Very close to the intersection you’ll find the bus stop on the right side of the street.
- Board Bus 31 in the direction of UNIVERSITÄTSPLATZ.
- Get off at the 5th stop, BISMARCKPLATZ. (point “A” on the map)
- Head south-west, past the Galeria Kaufhof department store onto Rohrbacher Straße.
- Continue south on Rohrbacher Straße until you find Hugo’s on the left side of the street.

It’s about 700 meters from the BISMARCKPLATZ stop to Hugo’s.
## Author Index

<table>
<thead>
<tr>
<th>Name</th>
<th>Age(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aguilar</td>
<td>Sergio Torres, 65</td>
</tr>
<tr>
<td>Ahamad</td>
<td>Faiz, 28</td>
</tr>
<tr>
<td>Anderson</td>
<td>Adam, 23</td>
</tr>
<tr>
<td>Arantes</td>
<td>Janine Aldous, 33</td>
</tr>
<tr>
<td>Arnold</td>
<td>Matthias, 21, 27</td>
</tr>
<tr>
<td>Ayan</td>
<td>Erdal, 30</td>
</tr>
<tr>
<td>Bantel</td>
<td>Ivo, 55</td>
</tr>
<tr>
<td>Born</td>
<td>Leo, 71</td>
</tr>
<tr>
<td>Bouchard</td>
<td>Hans, 53</td>
</tr>
<tr>
<td>Brunner</td>
<td>Annelen, 29</td>
</tr>
<tr>
<td>Buchholz</td>
<td>Ina, 27</td>
</tr>
<tr>
<td>Buchweiz</td>
<td>Ina, 21</td>
</tr>
<tr>
<td>Carlino</td>
<td>Carola, 44</td>
</tr>
<tr>
<td>Cramer</td>
<td>Henriette, 14</td>
</tr>
<tr>
<td>Declerck</td>
<td>Thierry, 61</td>
</tr>
<tr>
<td>Dereza</td>
<td>Oksana, 51</td>
</tr>
<tr>
<td>Diakité</td>
<td>William, 55</td>
</tr>
<tr>
<td>Eger</td>
<td></td>
</tr>
<tr>
<td>Steffen</td>
<td>57</td>
</tr>
<tr>
<td>El-Assady</td>
<td>Mennatallah, 72</td>
</tr>
<tr>
<td>Fetzer</td>
<td>Bettina, 25</td>
</tr>
<tr>
<td>Glinka</td>
<td>Katrin, 72</td>
</tr>
<tr>
<td>Guhr</td>
<td>Svenja, 45</td>
</tr>
<tr>
<td>Guo</td>
<td>Qiuzi, 32</td>
</tr>
<tr>
<td>Haider</td>
<td>Thomas, 57</td>
</tr>
<tr>
<td>Hermann</td>
<td>Berenike, 10</td>
</tr>
<tr>
<td>Herzberg</td>
<td>Laura, 63</td>
</tr>
<tr>
<td>Hovy</td>
<td>Dirk, 17, 20</td>
</tr>
<tr>
<td>Huang</td>
<td>Ding, 34</td>
</tr>
<tr>
<td>Karcher</td>
<td>Stefan, 25</td>
</tr>
<tr>
<td>Karlińska</td>
<td>Agnieszka, 42</td>
</tr>
<tr>
<td>Kermani</td>
<td>Hossein, 40</td>
</tr>
<tr>
<td>Khallaf</td>
<td>Nouran, 66</td>
</tr>
<tr>
<td>Kleiber</td>
<td>Ingo, 64</td>
</tr>
<tr>
<td>Korfhage</td>
<td>Katharina, 26</td>
</tr>
<tr>
<td>Krautter</td>
<td></td>
</tr>
</tbody>
</table>
Benjamin, 11
Landschoff
Jöran, 41
Lee
Hyein, 59
Müller
Carolin, 62
Münch
Vanessa, 35
Mahlberg
Michaela, 10
Manni
Ding, 37
Mara
Hubert, 68
Markert
Katja, 26
Mihaljević
Josip, 38
Moskvina
Anna, 52
Narayananmoorthy
Nanditha, 39
Nußbaumer
Lisa, 56
Nunn
Christopher, 25
Opitz
Juri, 71
Pagel
Janis, 11
Piotrowski
Michael, 17, 19
Piper
Andrew, 9
Przecherska
Weronika, 59
Puschmann
Cornelius, 16, 17
Redden

Joanna, 13
Reiter
Nils, 11, 17
Ruppenhofer
Josef, 26
Scheffler
Tatjana, 15
Schmitt
Anna, 49
Sivashanmugham
Rahul Pillai, 47
Soni
Yasmin, 67
Speranza
Giulia, 36
Sporleder
Caroline, 17
Straube
Annika, 25
Sulaiman
Abdul-Rasheed Lanre, 46
Tu
Ngoc Duyen Tanja, 29, 50
van den Berg
Esther, 26
von Vacano
Claudia, 14, 17
Warwick
Claire, 23
Weiß
Christof, 24
Weimer
Lukas, 29
Wiegand
Michael, 26
Willand
Marcus, 11
Wolfangel
Eva, 73
Zhu
Yi, 48