

## Computergestützte Inhaltsanalyse in der Migrations- und Integrationsforschung

Bericht zum MigTex-Projektworkshops „Mining Migration Debates“ am 29.-30. November 2018 in Duisburg



Teilnehmer\*Innen des MigTex-Workshops während der Begrüßung durch Prof. Dr. Andreas Blätte

In zwei intensiven Workshoptagen konnten am 29. und 30. November mehr als 20 Forscher\*Innen aus verschiedenen europäischen Ländern in einem Workshop des Projektes „MigTex – Textressourcen für die Migrations- und Integrationsforschung“ ihre Kenntnisse im Bereich des Text-Mining vertiefen und praktisch in die Arbeit mit den aktuellen Forschungsergebnissen des Projektes eingeführt werden. Organisiert und konzipiert wurde der Workshop mit dem Titel „Mining Migration Debates“ von Prof. Dr. Andreas Blätte, der Projektmitarbeiterin Merve Schmitz-Vardar sowie dem Projektmitarbeiter Christoph Leonhardt.

Das MigTex-Projekt hat das Ziel, grundlegende und relevante Datenbestände für den Forschungsbereich zu identifizieren, zu analysieren, in maschinenlesbarer Form aufzubereiten und nachhaltig bereitzustellen. MigTex wird in einer Kooperation zwischen der Universität Duisburg-Essen (Interdisziplinäres Zentrum für Integrations- und Migrationsforschung / InZentIM) und dem Wissenschaftszentrum Berlin (WZB) durchgeführt und durch das Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) gefördert. Das MigTex-Projekt baut dabei auf der Infrastruktur, der Codebasis und den Zielsetzungen des [PolMine-Projekts](#) auf. Als Strukturprojekt trägt es zur Forschungsdateninfrastruktur des Deutschen Zentrums für Migrations- und Integrationsforschung (DeZIM) bei.

Ein besonderer Fokus des MigTex-Workshops „Mining Migration Debates“ lag auf der Vielfalt und Interaktivität der angebotenen Formate: Im Rahmen von praktischen Übungen in Kleingruppen, Einführungs- und Vertiefungsvorträgen, Reflektionseinheiten, einer Keynote sowie eines Hackathons konnten die Teilnehmenden intensive Einblicke in den aktuellen Arbeitsstand des Projektes sowie einen Überblick über den Forschungsstand im Bereich des Text-Mining erlangen und ihre zuvor individuell eingereichten und betreuten Forschungsfragen gemeinsam bearbeiten.

Zur Begrüßung der Gäste des Workshops hielten Prof. Dr. Andreas Blätte (Projektleiter) und Prof. Dr. Michael Kaeding (Prodekan für Forschung und wissenschaftlichen Nachwuchs an der Fakultät für Gesellschaftswissenschaften der Universität Duisburg-Essen) eine kurze Rede, in welcher beide die Wichtigkeit der Vernetzung von Forschenden gerade im Bereich der *eHumanities* bzw. *Computational Social Sciences* betonten.

Nach einer kurzen Vorstellung des Programmablaufs sowie einer Einführung in MigTex-Projektstruktur und das Projekt-Datenpaket „MigParl“, konnten die Teilnehmenden im Rahmen einer Vorstellungsrunde ihr Forschungsprofil und ihre eingereichten Proposals erläutern. In einer Reflektionswerkstatt wurden die bisherigen Erfahrungen der Teilnehmenden des zehnwöchigen Onlineworkshops zur Einführung in die Arbeit mit der Analysesoftware 'polmineR' abgefragt, welcher dem „Mining Migration Debates“-Workshop vorrausgegangen war. Weitere Programmpunkte des ersten Veranstaltungstages umfassten das Forschungsdatenmanagement und die Replizierbarkeit und Transparenz in der textbasierten Migrations- und Integrationsforschung sowie die praktische Anwendung unterschiedlicher Methoden textanalytischer Sozial- und Sprachwissenschaft in Kleingruppen.



Keynote von Dr. Gregor Wiedemann

Als besonderes Programmhilighlight und zum Abschluss des ersten Workshoptages referierte Dr. Gregor Wiedemann in seiner Keynote zum Thema „Computergestützte Modellierung von Textbedeutung: Geschichte und aktuelle Trends“. Hierbei identifizierte er methodische Spannungsfelder in den Bereichen der Datengenerierung und des Trainings von Machine Learning-Anwendungen.

In der Frage der Datengenerierung wird dieses Spannungsfeld einerseits durch rein datengetriebene, andererseits durch theoriefundierte Ansätze definiert, während bei den für das Training von Machine-Learning-Algorithmen notwendigen Daten zwischen einem Rückgriff auf bereits vorhandene Trainingsdaten gegenüber der eigenen Erstellung solcher Daten unterschieden werden muss. Im weiteren Verlauf erläuterte Dr. Wiedemann zudem sein aktuelles Forschungsprojekt zur Verbesserung qualitativer Analysen mittels sogenannten „semantic embeddings“, indem er aufzeigte, dass Begriffe auf Wort-, Satz- und Diskursebene sehr unterschiedliche Bedeutungen haben können.

### **MigParl-Datensatz als zentrale Textressource für die parlamentarische Migrationsforschung**

Der Schwerpunkt des zweiten Tages lag auf der Anwendung und dem Transfer der erlernten korpuslinguistischen Methoden. Im Rahmen eines Hackathons wurden die eigenen Vorhaben mit dem im MigTex-Projekt erstellten Datenkorpus „MigParl“, welches den Workshopteilnehmer\*Innen in einer Beta-Version zur Verfügung steht, weiterentwickelt.

Das MigParl-Textkorpus wird als eines der Projektergebnisse Forschenden den Umgang mit Parlamentsdebatten perspektivisch in mehrfacher Hinsicht erleichtern. Indem eine induktiv und methodisch geleitete Vorauswahl migrations- und integrationspolitisch relevanter Debatten und Topics aus den Plenarprotokollen sämtlicher deutscher Landtage getroffen wird, steht Interessent\*Innen in Zukunft ein über einzelne Forschungsvorhaben hinweg vergleichbarer Datensatz zur Verfügung. In einer weiteren Übung mit dem Titel „Validität und Intersubjektivität - Forschung in einem umkämpften Forschungsfeld“ konnte die Frage nach der Produktion belastbarer Ergebnisse für die Migrations- und Integrationsforschung durch die Teilnehmenden umfassend erörtert werden. Nach der Vermittlung theoretischer Konzepte fand eine Reflexion über die Übereinstimmung zwischen den theoretischen Überlegungen und geeigneten Methoden für die eigenen Forschungsprojekte in Kleingruppen statt, in denen viel diskutiert wurde.

## Zentrale Ergebnisse aus dem Workshop

Die Programmgestaltung des Workshops „Mining Migration Debates“ hatte das Ziel, den Teilnehmenden eine praktische Einführung und Fortbildung in die Arbeit mit Text-Mining Applikationen anhand eigener Forschungsfragen zu ermöglichen und gleichzeitig Feedback aus der forschungspraktischen Arbeit mit dem speziell für diesen Workshop bereitgestellten Beta-Datenzugang zum MigParl-Datenkorpus zu erhalten, um dieses in die weitere Projektentwicklung mit einfließen zu lassen. Die vielfältigen Rückmeldungen und Bedarfe sowohl an das Analysepaket „polmineR“, welches von Prof. Dr. Andreas Blätte seit Jahren kontinuierlich unter einer freien Lizenz weiterentwickelt wird, als auch an das Datenkorpus 'MigParl', werden in der Nachbereitung des Workshops sukzessiv eingearbeitet. Eine finale Veröffentlichung des MigParl-Korpus ist für den Herbst 2019 geplant.

Die Bereitstellung von Software und Datenpaketen mit möglichst geringen technischen Hürden und in lizenzrechtlich geklärter Form im Sinne eines nachhaltigen Forschungsdatenmanagements ist eines der zentralen Ziele des MigTex-Projektes. Die für den zehnwöchigen Onlineworkshop entwickelten Tutorials, welche den beiden Praxistagen des Workshops vorrausgegangen sind, stehen unter dem Titel „Using Corpora in Social Science Research“ (UCSSR) im Web als [interaktiver Foliensatz](#) sowie als [eigenständiges Paket für RStudio](#) bereit. Mit diesen ist ein einfacher Einstieg in die Arbeit mit Korpusdaten möglich, als Grundlage hierfür dient das „GermaParl“ Korpus - eine linguistisch annotierte Sammlung aller Plenardebatten des deutschen Bundestages von 1996 bis 2016, welcher [im Web kostenfrei](#) verfügbar ist.

Die überaus positive und ergebnisreiche Zusammenarbeit der Workshopteilnehmer\*Innen soll perspektivisch in der gemeinsamen Publikation eines Sammelbandes münden. Zur weiteren Ausarbeitung dieses Vorhabens ist ein Autorenworkshop für 2019 in Planung. Über Teilnahmemöglichkeiten an diesem Workshop wird das MigTex-Projektteam im Frühjahr informieren.