



universität
wien



romanistisches
kolloquium XXXIV

Digitale romanistische Sprachwissenschaft. Stand und Perspektiven

21.-23.11.2019

Universitätscampus / Hof 8 / ROM 14



1. Universitätscampus mit Tagungsort (ROM14) und Restaurant Universitätsbräuhaus

www.romanistisches-kolloquium.de



Tagungsort:
 Universität Wien
 Institut für Romanistik
 Garnisongasse 13, Campus AAKH,
 Hof 8, Erster Stock, Rom 14
 A-1090 Wien

Team des Romanistischen Kolloquiums

Prof. Dr. Lidia BECKER

Prof. Dr. Dr. h. c. Wolfgang DAHMEN

Prof. Dr. Günter HOLTUS

Prof. Dr. Johannes KRAMER

Prof. Dr. Julia KUHN

Prof. Dr. Dr. h. c. mult. Michael METZELTIN

Prof. Dr. Christina OSSENKOP

Prof. Dr. Claudia POLZIN-HAUMANN

Prof. Dr. Elton PRIFTI

Prof. Dr. Dres. h.c. Wolfgang SCHWEICKARD

Prof. Dr. Otto WINKELMANN

Editoren des Romanistischen Kolloquiums

Prof. Dr. Lidia BECKER

Prof. Dr. Julia KUHN

Prof. Dr. Christina OSSENKOP

Prof. Dr. Claudia POLZIN-HAUMANN

Prof. Dr. Elton PRIFTI

Koordination und Realisierung des RK XXXIV

Prof. Dr. Elton PRIFTI

elton.prifti@univie.ac.at

Mag. Margit SEEBACHER

margit.seebacher@univie.ac.at

Digitale romanistische Sprachwissenschaft. Stand und Perspektiven

Programmübersicht

Uhrzeit	Donnerstag, 21. November 2019
14.00-14.30	Begrüßungen Elton PRIFTI (Wien): Einleitung in das Thema
14.30-15.15	Sabine TITTEL (HEIDELBERG): Kulturerbe, historische Linguistik und Semantic Web: Eine Einführung mit Fallstudie zu französischen linguistischen Ressourcen
15.15-16.00	Alexandra N. LENZ (Wien): Digitale Sprachwissenschaft - Herausforderungen und Perspektiven aus germanistischer Sicht
16.00-16.30	KAFFEPAUSE
16.30-17.15	Nanette RIßLER-PIPKA (Siegen): Verbindungen von digitaler romanistischer Sprach- und Literaturwissenschaft am Beispiel der Stilometrie
17.15-18.00	Vera NEUSIUS/Julia MONTEMAYOR GRACIA (Saarbrücken): Digitale romanistische Linguistik als Ort methodologischer Paradigmendiskussion? Kritische Überlegungen, Bedarfe und Potenziale
19.30	GEMEINSAMES ABENDESSEN

Uhrzeit	Freitag, 22. November 2019
9.00-9.45	Elmar SCHAFROTH/Riccardo IMPERIALE (Düsseldorf): Gebrauchsbasierte Phraseologie des Italienischen
9.45-10.30	Thomas SCHARINGER (Jena): Die Verbreitung der romanischen Sprachen im Spiegel des frühneuzeitlichen Buchdrucks. Zum Wert digitaler Kataloge für die Sprachgeschichte am Beispiel des USTC
10.30-11.00	KAFFEPAUSE
11.00-11.45	Florian ZACHERL (München): Automatisierte Erschließung und Vernetzung von Daten aus Wörterbuchtexten
11.45-12.30	Alessandro NANNINI (Wien): La marcatura semantica del <i>Lessico Etimologico Italiano</i> (LEI)
12.30-14.00	MITTAGSPAUSE
14.00-14.45	Sarah KREMER/Benjamin HUSSON (Nancy): Les outils de l'informatisation du FEW
14.45-15.30	Beatrice COLCUC (München): Klassische Forschungsfragen, neue Methoden: Digitale Dialektforschung in den Dolomiten
15.30-16.00	KAFFEPAUSE
16.00-16.45	Sam MERSCH (Esch-sur-Alzette): Erste Schritte zu einer digitalen Chrestomathie des Altrumänischen
16.45-17.30	TEAMSITZUNG
19.00	GEMEINSAMES ABENDESSEN

Uhrzeit	Samstag, 23. November 2019
8.30-9.15	TEAMSITZUNG
9.30-10.15	Ursula REUTNER (Passau): <i>Digital Humanities</i> auf dem Prüfstand. Analysemethoden für digitale Korpora von E-Mails über Internetseiten bis zu <i>Wikipedia</i>
10.15-11.00	Robert HESSELBACH (Kassel): Digitale Diskurslinguistik am Beispiel des Machtkampfes in Venezuela (Arbeitstitel)
11.00-11.30	KAFFEPAUSE
11.30-12.15	Mathieu CONSTANT (Nancy): Construction, exploitation et exploration de ressources d'expressions polylexicales
12.15-12.30	Elton PRIFTI (Wien): Schlusswort

ZUSAMMENFASSUNGEN DER VORTRÄGE

Donnerstag, 21. November 2019

14.00-14.30 Eröffnung

Begrüßung

Universität Wien

Michael METZELTIN: Begrüßung

Universität Wien, Mitbegründer des Romanistischen Kolloquiums

Elton PRIFTI: Thematische Einleitung

Universität Wien

14.30-15.15 Sabine TITTEL

Heidelberger Akademie der Wissenschaften (DEAF)

Kulturerbe, historische Linguistik und Semantic Web: Eine Einführung mit Fallstudie zu französischen linguistischen Ressourcen

Es wird zunehmend anerkannt, dass die Erforschung und Erhaltung des europäischen Kulturerbes von großer Bedeutung für unsere Gesellschaft sind. Dies führt zu vielen Anstrengungen, Wissen über unser Kulturerbe zu generieren und es öffentlich digital zugänglich zu machen. Es gibt viele Arten von Kulturgut, die Licht auf die Wurzeln der europäischen Moderne und die Entwicklung der europäischen Identität werfen: materielles wie Alltags- und Ritualgegenstände, Architektur etc., und immaterielles, das uns insbesondere in Form von Textzeugnissen überliefert ist.

Die Einzeldisziplinen der Linguistik haben bereits eine Fülle von Wissen anhand dieser Textzeugnisse erarbeitet und in einer Vielzahl von linguistischen Ressourcen gespeichert: in Texteditionen, Wörterbüchern, Sprachatlanten, Sekundärliteratur. Die Zahl der linguistischen Ressourcen, die dem Nutzer online zur Verfügung gestellt werden, wird dabei immer größer. Aber die meisten Daten sind in Datensilos im World Wide Web (WWW) veröffentlicht, das heißt, öffentlich verfügbar, aber nur in einem proprietären Format oder mit eingeschränktem Zugriff. Aus diesem Grund ist es kaum möglich, einen Zugriff zu den wertvollen Daten zu erhalten, der über Ressourcen, Sprachen, historische Sprachstufen, inhaltliche Ausrichtungen und über Datenformate hinweg funktioniert. Aber ge-

rade diese Art von Zugriff ist nötig, um die Entwicklungen unserer multikulturellen, vielsprachigen Gesellschaft besser beleuchten zu können und Zusammenhänge zu erkennen.

Die Defizite des Datenzugriffs im WWW können von Technologien des Semantic Web (SW) überwunden werden. Das SW verfolgt das Ziel, die prinzipiell unstrukturierten Informationen des WWW so mit inhaltlichen Markierungen zu versehen, dass sie maschinenlesbar und informatisch-semantisch interpretierbar werden. Das Paradigma von *Linked Open Data* (LOD) ist in den letzten zehn Jahren zu einem der wichtigsten Eckpfeiler des SW geworden. Es umfasst eine Gruppe von Standardmethoden, mit Hilfe derer Daten strukturiert, publiziert und so mit semantischen Informationen versehen werden, dass sie mit anderen Daten verknüpft und gemeinsam abgefragt werden können. Dazu gehören das Graphdatenformat Resource Description Framework (RDF) für die Modellierung der Daten, der Einsatz von Ontologien für die Abbildung von Wissensstrukturen und URIs als eindeutige Identifizierungen für die Daten.

Anhand eines Fallbeispiels mit historischen linguistischen Daten des Französischen zeigen wir, wie Ressourcen als LOD in RDF modelliert werden. Wir demonstrieren, wie damit eine strukturelle (durch Formate und Modelle) und konzeptuelle Interoperabilität (durch semantische Annotierung) der Daten erzielt wird. Dies wiederum erfüllt das Desiderat der ressourcen-, sprach- und inhaltsübergreifenden Abfragemöglichkeit.

15:15-16:00 Alexandra N. LENZ

Österreichische Akademie der Wissenschaften (Austrian Center for Digital Humanities), Universität Wien

Digitale Sprachwissenschaft - Herausforderungen und Perspektiven aus germanistischer Sicht

Der Vortrag setzt sich zum Ziel, die Herausforderungen und Möglichkeiten digitaler Zugänge aus der Perspektive der „Digitalen Sprachwissenschaft“ zu beleuchten. Unter *Digitaler Sprachwissenschaft* wird dabei ganz allgemein Sprachwissenschaft verstanden, die sich bei ihren Forschungsvorhaben digitaler Methoden und Werkzeuge bedient. Dies beinhaltet erstens Forschungsprozesse der Generierung und Erschließung von sprachwissenschaftlich relevanten Daten, zweitens die Aufbereitung und Anreicherung der Daten (z. B. in Form von Transkriptionen und Annotationen), drittens die Analyse und Interpretation von Forschungsfragen sowie viertens die digitale Bereitstellung der erhobenen, aufbereiteten und analysierten Daten. Neben mehr theoretisch-methodologischen Überlegungen werden im Vortrag konkrete und aktuelle Forschungsprojekte

und -ergebnisse der germanistischen Digitalen Sprachwissenschaft vorgestellt und diskutiert.

16.00-16.30 Kaffeepause

16.30-17.15 Nanette RIBLER-PIPKA

Universität Siegen

Verbindungen von digitaler romanistischer Sprach- und Literaturwissenschaft am Beispiel der Stilometrie

Die Romanistik zeichnet sich durch ihre Gesamtbetrachtung der romanischen Sprachen und Literaturen aus, zerfällt aber schon seit vielen Jahren mehr und mehr in die Betrachtung der Einzelsprachen und auch der sprach- und literaturwissenschaftlichen Disziplinen. Die neueren, starken Impulse aus den *Digital Humanities* dienen nun überraschenderweise als ein erneutes Bindeglied, das sowohl die einzelnen Sprachen, Literaturen aber auch die Methoden der Sprach- und Literaturwissenschaft wieder näher zusammen bringt. Der Beitrag erläutert am Beispiel der Stilometrie aus literaturwissenschaftlicher Perspektive die Bezüge zur Stilistik sowie die Chancen einer neuerlichen Annäherung zwischen romanistischer Sprach- und Literaturwissenschaft auf der Basis gemeinsamer digitaler Methoden. Umgekehrt besteht in den *Digital Humanities* großer Bedarf, die vor allem an englischsprachigen Korpora erprobten Methoden sprachübergreifend auszuweiten. Hier kann die Romanistik ihre lange Erfahrung im Vergleich verwandter Sprachen und Literaturen einbringen.

17.15-18.00 Julia MONTEMAYOR GRACIA/Vera NEUSIUS

Universität des Saarlandes

Digitale romanistische Linguistik als Ort methodologischer Paradigmendiskussion? Kritische Überlegungen, Bedarfe und Potenziale

Der Verknüpfung qualitativer und quantitativer Methoden kommt heute gerade in interdisziplinär angelegten wissenschaftlichen Arbeiten eine international wachsende Anerkennung zu. Ausgehend von Forschungsarbeiten im angloamerikanischen Raum erfuhr der *Mixed-Methods*-Ansatz in Deutschland zunächst Beachtung in den empirischen Sozialwissenschaften, von wo aus er nun nach und nach auf andere Disziplinen übergreift. In der romanistischen Linguistik, wo die Auseinandersetzung mit einer systematischen Methodenkombination bislang nur marginal und vorrangig im Rahmen der nach wie vor dominierenden

Paradigmendebatte um qualitative vs. quantitative Vorgehensweisen stattfindet, besteht hier im Vergleich zu Nachbardisziplinen ein Nachholbedarf. Dieser spiegelt sich auch in der bisher eher zögerlichen Anwendung computergestützter Analyseverfahren.

In diesem Spannungsfeld der romanistischen Forschung zwischen Tradition und Innovation wird im Rahmen grundlegender theoretisch-methodischer Überlegungen beleuchtet, inwiefern die methodologische Grundsatzdiskussionen Anlass bietet, durch die Nutzung computergestützter Methoden neue Wege zu gehen und welche Bedarfe und Potenziale hierbei in unserem Fach existieren. Abschließend sollen durch exemplarische Einblicke in die Praxis Möglichkeiten der softwarevermittelten qualitativen Kategorisierung von Sprachdaten sowie die partielle Adaption korpuslinguistischer Verfahren bei der Analyse sprachlicher Daten vorgestellt werden.

19.30 Abendessen

Freitag, 22. November 2019

9.00-9.45 Elmar SCHAFFROTH/Riccardo IMPERIALE

Heinrich-Heine-Universität Düsseldorf

Gebrauchsbasierte Phraseologie des Italienischen

Seit August 2018 wird an der Universität Düsseldorf das auf drei Jahre angelegte DFG-Forschungsprojekt *Gebrauchsbasierte Phraseologie des Italienischen* (<https://gephri.phil.hhu.de/>) durchgeführt. Hauptziel von GEPHRI ist es, die auf der Basis von Frequenzanalysen ermittelten 600 häufigsten verbalen Idiome des Italienischen mit Methoden der Korpuslinguistik in einer Datenbank ganzheitlich zu beschreiben und dazu beizutragen, eine in dieser Weise nirgends dokumentierte verstehensrelevante Darstellung des „Funktionierens“ eines Phrasems (auch in pragmatischer und diskursiver Hinsicht) umzusetzen.

Die Beschreibung der formalen und inhaltlichen Besonderheiten von Phrasemen erfolgt auf der Basis italienischer Sprachkorpora wie *PAISÀ*, *La Repubblica*, *CO-RIS* und *itTenTen16* sowie nach den gängigen zweisprachigen und einsprachigen Wörterbüchern (auch phraseologischen Spezialwörterbüchern) des Italienischen. Zu den Deskriptoren gehören ferner begleitende Gesten und prosodische Merkmale, die, wann immer dies möglich ist, durch Videomaterial illustriert werden.

Der theoretische Impetus zu diesem Projekt erfolgte durch die Fillmoresche Verstehens- und Framesemantik und durch gebrauchsbasierte Ansätze der Konstruktionsgrammatik.

Da GEPHRI wissenschaftliche und didaktische Ziele verfolgt, wird dem L2-/LS-Aspekt (Italienisch als Zweit- oder Fremdsprache) ausgiebig Rechnung getragen: z.B. durch die Formulierung leicht verständlicher „Gebrauchshinweise“ zu einem Phrasem, die Differenzierung des modellierten sprachlichen Phrasemwissens in eine Basis- und eine Detailsbene und durch eine kontrastive Perspektive für Deutschsprachige (eigene, korpusbasierte Bedeutungsparaphrasen).

9.45-10.30 Thomas SCHARINGER

Friedrich-Schiller-Universität Jena

Die Verbreitung der romanischen Sprachen im Spiegel des frühneuzeitlichen Buchdrucks. Zum Wert digitaler Kataloge für die Sprachgeschichte am Beispiel des USTC

Im Vortrag soll der Nutzen von digitalen Verzeichnissen frühneuzeitlicher Drucke für die sprachgeschichtliche Forschung am Beispiel des USTC (*Universal Short Title Catalogue*) aufgezeigt werden (vgl. z.B. Scharinger 2017).

Der im *Open Access* zugängliche USTC, in den die Ergebnisse vorangegangener Großprojekte wie EDIT16 eingespeist worden sind, stellt das bisher umfangreichste digitale Verzeichnis von Drucken, die im Europa des 16. und 17. Jh. entstanden sind, dar. Aufgrund der vielfältigen, miteinander kombinierbaren Suchoptionen (Land, Stadt, Zeitraum, Genre und v.a. Sprache) lassen sich zahlreiche Daten gewinnen, die über die Verbreitung einer bestimmten romanischen Sprache außerhalb ihres angestammten Territoriums Auskunft geben. So ist es z.B. möglich, herauszufinden, in welchem Land bzw. in welcher Stadt im 16. Jh. (oder auch nur in den 1520er Jahren) die meisten italienischen Drucke außerhalb Italiens produziert wurden. Während die Quantität der Drucke zunächst nur Rückschlüsse auf den ökonomischen Stellenwert einer Sprache in einem bestimmten Gebiet zulässt, können anhand der Qualität der Drucke (Genres) auch die Rezipienten ermittelt werden. Handelt es sich bei der Mehrheit der Drucke um Belletristik, spricht dies dafür, dass die Nachfrage von einer gebildeten einheimischen Elite, die die Sprache als Fremdsprache beherrscht, ausgeht, wohingegen eine hohe Anzahl an Titeln, die der alltäglichen Gebrauchsliteratur oder Verwaltungsschriftlichkeit zuzuordnen sind, eher auf die Existenz von Migranten schließen lässt. Anhand der so gewonnenen Ergebnisse lassen sich auch etablierte An-

nahmen zur Verbreitung einer romanischen Sprache als Kultur- oder Migrantensprache überprüfen (zum *italiano fuori d'Italia* vgl. z.B. Stammerjohann 2013, Banfi 2014, Baglioni 2016).

Des Weiteren sind die Daten des USTC auch im Hinblick auf den Stellenwert einer romanischen Sprache innerhalb ihres angestammten Gebiets von Interesse. So kann z.B. ermittelt werden, in welchem romanischsprachigen Land zum ersten Mal mehr volkssprachliche als lateinische Drucke produziert wurden und in welchem Genres die Volkssprache besonders stark vertreten war. Schließlich kann auch das Verhältnis zwischen zwei romanischen Sprachen in ein und demselben Gebiet (z.B. Spanisch und Katalanisch) näher beleuchtet werden. Im Vortrag werden Fallstudien zum Französischen, Italienischen und Spanischen vorgestellt.

Literatur

Baglioni, Daniele (2016): „L'italiano fuori d'Italia: dal Medioevo all'Unità“, in: Lubello, Sergio (ed.): *Manuale di linguistica italiana*, Berlin, de Gruyter, 125-145.

Banfi, Emanuele (2014): *Lingue d'Italia fuori d'Italia. Europa, Mediterraneo e Levante dal Medioevo all'età moderna*. Bologna, Il Mulino.

EDIT16 = *Censimento nazionale delle edizioni italiane del XVI secolo*. [http://edit16.iccu.sbn.it/web_iccu/ihome.htm]

Scharinger, Thomas (2017): „L'italiano fuori d'Italia im Spiegel des frühneuzeitlichen Buchdrucks - Zu Marktwert und Verbreitung italienischer Drucke im Europa des 16. Jahrhunderts“, in: Ambrosch-Baroua, Tina/Kropp, Amina/Müller-Lancé, Johannes (eds.), *Mehrsprachigkeit und Ökonomie*, München, Open Access LMU, 161-188.

Stammerjohann, Harro (2013): *La lingua degli angeli: italianismo, italianismi e giudizi sulla lingua italiana*. Florenz, Accademia della Crusca.

USTC = *Universal Short Title Catalogue. A Digital Bibliography of Early Modern Print Culture*. St. Andrews. [<https://www.ustc.ac.uk>]

10.30-11.00 Kaffeepause

11.00-11.45 Florian ZACHERL

Ludwig-Maximilians-Universität München

Automatisierte Erschließung und Vernetzung von Daten aus Wörterbuchtexten

Gedruckte sprachwissenschaftliche Werke enthalten eine Vielzahl von Informationen, die aber selbst beim Vorliegen von Digitalisaten oft nur unter sehr großem Aufwand ausgewertet oder mit anderen Quellen verglichen werden können. Der Vortrag beleuchtet am Beispiel des "Romanischen Etymologischen Wörterbuchs" in der Auflage von 1935 einen Weg, diese Informationen auszulesen

und in einen strukturierten, digitalen Datenbestand zu überführen. Der Prozess kann dabei in vier wesentliche Schritte unterteilt werden:

Im ersten Schritt werden die Scans der Printausgabe mit Hilfe von Texterkennung eingelesen und zeilenweise in eine relationale Datenbank importiert. Darauf folgt eine Erkennung der Struktur der einzelnen Artikel des Wörterbuchs. Dies wird mit einer teilautomatisiert erstellten "Parsing Expression Grammar" [1] durchgeführt. Das Resultat dieses zweiten Schritts ist eine baumartige Struktur, welche die verschiedenen Entitäten des Artikels widerspiegelt. Diese entsprechen beispielsweise Abschnitten oder Listenstrukturen bis hinunter zu elementaren Bestandteilen wie einzelnen Lemmata oder Sprachabkürzungen. Der folgende Schritt transformiert diese weiter in ein generalisiertes Modell für lexikographische Daten, wobei vorhandene implizite Konventionen der entsprechenden Quelle aufgelöst werden. Der letzte Teil behandelt die Publikation der Daten, einerseits in einem speziellen Online-Portal, andererseits über technische Schnittstellen, was eine Zusammenführung mit anderen lexikographischen Datenbeständen sowie eine Einbettung in das Semantic Web ermöglicht.

Ausgehend von exemplarischen Einträgen wird detailliert auf die einzelnen Arbeitsschritte eingegangen, wobei jeweils mögliche Werkzeuge und spezifische Herausforderungen vorgestellt werden.

Bibliographie

[1] Ford, Bryan (2004): *Parsing expression grammars: a recognition-based syntactic foundation*, in: SIGPLAN Not. 39, 1, 111-122. DOI: <https://doi.org/10.1145/982962.964011>

11.45-12.30 Alessandro A. NANNINI

Universität Wien

La marcatura semantica del Lessico Etimologico Italiano (LEI)

Da alcuni anni il progetto pluridecennale Lessico Etimologico Italiano (LEI), fondato da Max Pfister nel 1969, ha intrapreso un complesso percorso di digitalizzazione, che consiste, a grandi linee, nell'informatizzazione del processo redazionale, del prodotto lessicografico nonché della sua rappresentazione e utilizzo. Il processo di digitalizzazione apre nuove prospettive anche nell'elaborazione del materiale lessicografico, contribuendo all'aumento della qualità del prodotto. In linea con le recenti tendenze della lessicografia digitale ci si è posto l'obiettivo di fornire un metodo per una mappatura semantica del LEI, che permetterà la ricerca semasiologica di alta precisione e di ampia portata. La nuova funzione, combinata con le altre possibilità di ricerca (e.g. diacronica, diatopica,

morfologica), metterà a disposizione un potente strumento per lo studio linguistico del panorama italo-romanzo.

Il sistema concettuale di partenza è il Begriffssystem di Rudolf Hallig e Walther von Wartburg (Hallig/Wartburg 1963), già utilizzato da altre opere lessicografiche, tra le quali menzioniamo il DAG e il DAO. Il primo passo da compiere consiste nella trasformazione del Begriffssystem in una collezione di dati informatici, al fine di renderlo collegabile agli altri dati presenti nel database del LEI. Questo sistema di base, opportunamente aggiornato e integrato con altre basi di conoscenza (e.g. Wikidata, DBpedia), potrà essere utilizzato per mappare i materiali già pubblicati, ma anche in fase di redazione dei nuovi articoli del vocabolario.

In questo modo, il nodo dell'informazione concettuale si aggiungerebbe agli altri nodi del grafo informatico che rappresentano le diverse informazioni (e.g. formali, morfologiche, bibliografiche) reperibili nel LEI (Prifti 2019, 6). Data questa connessione tra i dati, appare evidente la possibilità di compiere ricerche incrociate tra più parametri. La query di un concetto o di un'area concettuale di interesse restituirebbe all'utente le forme semanticamente corrispondenti, corredate delle altre ricche informazioni lessicografiche presenti nel LEI.

Il mio intervento mira a mettere in evidenza le possibilità, i metodi e le prospettive legate alla marcatura semantica delle forme registrate nel database del LEI. Nella prima parte verranno presentate l'architettura dei dati del LEI-Digitale e le idee circa l'aggiunta di un codice semantico al grafo lessicografico. Saranno presi in considerazione gli sviluppi della lessicografia digitale, così come quelli del Web semantico con gli strumenti e le opportunità che esso offre alla lessicografia e alla ricerca linguistica. Ci si concentrerà, quindi, sull'analisi di progetti simili, per individuare le best practices e i punti critici che potrebbero interessare anche il nostro lavoro. Nell'ultima parte verrà presentato un prototipo del prodotto che si intende sviluppare.

Bibliografia

DAG = Baldinger, Kurt, *Dictionnaire onomasiologique de l'ancien gascon*, Berlin/Boston, De Gruyter, 1975ff.

DAO = Baldinger, Kurt, *Dictionnaire onomasiologique de l'ancien occitan*, Tübingen, Niemeyer, 1975ff.

Hallig, Rudolf/Wartburg, Walther von, *Begriffssystem als Grundlage für die Lexikographie = Système raisonné des concepts pour servir de base à la lexicographie*, Berlin, Akademie Verlag, 1963.

LEI = Prifti, Elton/Schweickard, Wolfgang (edd.), *LEI. Lessico Etimologico Italiano. Fondato da Max Pfister*, Wiesbaden, Reichert, 1979ff.

Prifti, Elton, *Lo stato della digitalizzazione del LEI. Un resoconto*, in: Leonardi, Lino/Squillaciotti, Paolo (eds.): *Italiano antico, italiano plurale. Testi e lessico del Medioevo nel mondo digitale*, Firenze, 2019 (in corso di stampa).

12.30-14.00

Mittagspause

14.00-14.45

Sarah KREMER/Benjamin HUSSON

ATILF

Les outils de l'informatisation du FEW

L'informatisation du FEW menée aujourd'hui à l'ATILF a posé de nombreux défis tant théoriques que pratiques. Linguistes, informaticiens et typographes ont ainsi collaborés pour développer un ensemble d'outils permettant de mener à bien cette entreprise. Cette présentation sera l'occasion d'aborder (1) les motivations et les perspectives liées à l'informatisation du FEW, (2) les problèmes pratiques rencontrés et les techniques mises en œuvre pour les résoudre, (3) la nécessité et l'intérêt du développement d'outils typographiques pour apporter des solutions visuelles aux particularités de rédaction et de consultation du FEW.

14.45-15.30

Beatrice COLCUC

Ludwig-Maximilians-Universität München

Klassische Forschungsfragen, neue Methoden: Digitale Dialektforschung in den Dolomiten

Die Dolomiten sind eine Gebirgsgruppe der südlichen italienischen Ostalpen, die geographisch durch das Eisack- und das Etschtal im Westen, das Pustertal im Norden, den Fluss Piave im Osten und das Valbelluna im Süden abgegrenzt werden.

In Bezug auf die aktuell im Dolomitengebiet gesprochenen Varietäten wird in der Romanistik ununterbrochen darüber diskutiert, welche Lokalvarietäten genau welchen Dialektfamilien ('Sprachen') zugeordnet werden müssen bzw. wie genau die Varietäten innerhalb des Gebietes klassifiziert werden müssen: Einerseits wird die Idee einer klaren Abgrenzbarkeit der ladinischen Idiome vermittelt (vgl. Kattenbusch 1988; Goebel 2003); Andererseits werden die dort gesprochenen Varietäten als ein Teil des romanischen Dialektkontinuums angesehen (vgl. Pellegrini/Sacco 1984). Die hier vorgestellte Studie möchte versuchen, diese Frage auf der Ebene der Sprecherwahrnehmung zu stellen. Einerseits möchte die Studie herausfinden, wie die Sprecher aufgrund ihrer Perzeption die jeweils

benachbarten dialektalen Varietäten wahrnehmen und auf Grund konkreter Hörerfahrung in einem Test zuordnen. Außerdem soll herausgearbeitet werden, wie die autochthonen Sprecher den Sprachraum strukturieren und welchen Grad an Nähe oder Distanz sie in Bezug auf ihre Varietät dem einen oder dem anderen gesprochenen Dialekt zuschreiben. Diese Herangehensweise wird von der Theorie der perzeptiven Varietätenlinguistik (Krefeld/Pustka 2010) inspiriert und ermöglicht ein ideologiefreies Bild des Sprachgebietes aus der Perspektive der Sprecher zu erlangen.

Auf der anderen Seite zielt die Studie auf die Repräsentationen der Sprecher über den Sprachraum. Dabei sollen insbesondere Sprechermeinungen und Spracheinstellungen erhoben werden, die durch außersprachliche Elemente gefärbt, wenn nicht motiviert sind. Die Methoden, die hierfür verwendet werden, gehören zur sog. *perceptual dialectology* (Preston 1982, 1989; Iannàccaro/Dell'Aquila 2001). Der Vergleich zwischen den perzeptiv und den außersprachlich basierten Repräsentationen von Sprechern soll dazu beitragen, das Bild, das Sprecher vom Dolomitengebiet haben, empirisch abzubilden.

Die Daten werden auf zweierlei Art erhoben: Einerseits durch ein Wahrnehmungsexperiment mit Sprachproben, worauf online zugegriffen werden kann; Andererseits durch face-to-face Interviews. Die Studie wird im digitalen Forschungsparadigma durchgeführt: Die im Wahrnehmungsexperiment präsentierten Audio-Sprachbeispiele werden im Sinne der *digital humanities* aus bereits existierenden Forschungsprojekten wie dem Sprachatlas VIVALDI (Kattenbusch 1998-2016) übernommen. Die aus der Untersuchung resultierende Dissertationsarbeit wird auf einer Internetplattform verfasst, die die Visualisierung von Daten auf interaktiven Karten, das direkte Einbauen von Audiobeispielen und weiteren Medien, weiterführenden Links und externen Verweisen ermöglicht. Die hier zusammengefasste Forschungsarbeit folgt somit dem open-access Gedanke zur Nutzung der Daten durch Dritte und ist den FAIR-Prinzipien verpflichtet (vgl. *Open Access Germany*; Krefeld 2019).

Diese Studie will einen komplementären Beitrag zur Diskussion über die Klassifikation der Varietäten im Dolomitengebiet leisten, denn es wird eine neue Perspektive eingenommen, in der die Sprecherperzeption und das Sprecherwissen bisher vernachlässigte Elemente zur Beschreibung des sprachlichen Raumes liefern. Diese Studie versucht also, eine bislang noch fehlende Dimension in der sprachlichen Beschreibung des betreffenden Bereichs aufzuarbeiten: um einen Sprachraum unverkürzt darzustellen, müssen nämlich beide epistemologische Niveaus, das der Laien (Sprecher) und das der Experten (Linguist), berücksichtigt werden (vgl. Krefeld/Pustka 2010, 23).

Der Vortrag soll die oben geschilderte Diskussion im Detail vorstellen und auf die Forschungsfrage, die Hypothesen und die Erhebungs- und Arbeitsmethoden tiefer eingehen.

Literatur

- Goebel, Hans (2003): "Externe Sprachgeschichte der romanischen Sprachen im Zentral- und Ostalpenraum/Histoire externe des langues romanes: les Alpes centrales et orientales." In: *Romanische Sprachgeschichte. Ein internationales Handbuch zur Geschichte der romanischen Sprachen = Histoire linguistique de la Romania*, Gerhard et al. Ernst (Hrsg.). Berlin, New York: de Gruyter Band 1, 747–773.
- Iannàccaro, Gabriele/Dell'Aquila, Vittorio (2001): "Mapping languages from inside: notes on perceptual dialectology." *Social&Cultural Geography* 3, 265–280.
- Kattenbusch, Dieter (1988): "Rätoromanisch oder Ladinisch? Dolomitenladinisch = Sellaladinisch = Zentralladinisch = Zentralrätoromanisch? Einige Bemerkungen zu einem terminologischen Streit." *Ladinia* 12, 5–16.
- Kattenbusch, Dieter (1998-2016): "Vivaio Acustico delle Lingue e dei Dialetti d'Italia (VIVALDI)." <<https://www2.hu-berlin.de/vivaldi/>> [Zugriff am 20/08/2019].
- Krefeld, Thomas (2019): "I principi FAIR nel progetto VerbaAlpina, ossia il trasferimento della geolinguistica alle Digital Humanities." <<https://www.verba-alpina.gwi.uni-muenchen.de/?p=8212&db=191#rf1-8212>> [Zugriff am 22/08/2019].
- Krefeld, Thomas/Pustka, Elissa (2010): "Für eine perzeptive Varietätenlinguistik." In: *Perzeptive Varietätenlinguistik*, Thomas Krefeld/Elissa Pustka (Hrsg.). Frankfurt am Main: Peter Lang, 9–28.
- Open Access Germany: "Was bedeutet Open Access?" <<https://open-access.net/informationen-zu-open-access/was-bedeutet-open-access>> [Zugriff am 20/08/2019].
- Pellegrini, Giovan Battista/Sacco, Sergio (Hrsg.) (1984): *Il ladino bellunese. Atti del convegno internazionale, Belluno 2-4 giugno 1983 (Studi ladini 3)*. Belluno: Istituto Bellunese di Ricerche Sociali e Culturali.
- Preston, Dennis R. (1982): "Perceptual Dialectology: Mental maps of United States dialects from a Hawaiian perspective." *Working Papers in Linguistics* 14, 5–49.
- Preston, Dennis R. (1989): *Perceptual Dialectology. Nonlinguists' Views of Areal Linguistics (Topics in Sociolinguistics 7)*. Berlin: de Gruyter Mouton.

15.30-16.00 Kaffeepause

16.00-16.45 Sam MERSCH

Université du Luxembourg

Erste Schritte zu einer digitalen Chrestomathie des Altrumänischen

In dem Beitrag wird auf das Digitalisierungsprojekt zu Gasters Rumänischer Chrestomathie eingegangen, das aus der Zusammenarbeit von Wolfgang Dahmen und Sam Mersch erwachsen ist.

Die rumänische Chrestomathie von Moses Gaster (*Chrestomatie română*, 1891) ist und bleibt nach wie vor ein wichtiger Bestandteil für die rumänische diachrone Sprach- und Literaturwissenschaft, wenngleich der Text schon 1891 publiziert und auch späterhin nicht ergänzt oder verändert wurde. Johannes Kramer hat ein Jahrhundert später, im Jahr 1991 diese Chrestomathie nochmals herausgegeben, damit das Werk der Öffentlichkeit zugänglicher gemacht werden konnte.

Die direkte Zugänglichkeit zum Material ist heutzutage, also über 125 Jahre später, aber absolut gewährleistet dadurch, dass das Original Copyright längst abgelaufen ist und die Chrestomathie (die erste Ausgabe, nicht die Wiederauflage von J. Kramer) von mehreren Seiten digitalisiert und auf öffentlichen Plattformen zur Verfügung gestellt wurde bzw. wird.¹ Es handelt sich hier aber lediglich um eine digitalisierte Bereitstellung des analogen Buches und hat dadurch kaum einen Mehrwert, wenn man von der vereinfachten Zugänglichkeit absieht.

Der Zugang zum Werk als (digitalisiertes) Buch mag zwar dadurch vereinfacht sein, jedoch nicht der Zugang zum eigentlich dargebrachten Material.

Ziel des Projektes nicht, ein einfaches Digitalisat zu liefern, sondern der Gattung der Chrestomathie ein zeitgemäßes Update zu geben, so dass die dort enthaltenen Texte auch philologisch bearbeitet werden können.

Es werden vor allem methodische und technische Probleme erläutert, die mit einer solchen Textgattung in altkyrillischem Alphabet einhergehen.

19.00 Abendessen

Samstag, 23. November 2019

09.30-10.15 Ursula REUTNER

Universität Passau

Digital Humanities* auf dem Prüfstand. Analysemethoden für digitale Korpora von E-Mails über Internetseiten bis zu *Wikipedia

Die digitale Revolution hat unser Leben in vielerlei Hinsicht verändert. In den Wissenschaften zeigt sich dies in der Entstehung ganz neuer Wissenschaftsdisziplinen sowie in neuen Untersuchungsgebieten und Analysemethoden für bereits etablierte Disziplinen. Zugleich drängen sich Fragen nach dem zusätzlichen

¹ U.a. bei Wikimedia Commons <https://commons.wikimedia.org/wiki/Category:Books_by_Moses_Gaster?uselang=de>, aber auch bei archive.org <<https://archive.org/details/chrestomatie-rom00unkngoog/page/n11>>, beide zuletzt abgerufen am 11.06.2019.

Erkenntnisgewinn bei der mit digitalen Mitteln erfolgten Auswertung auf: Sind die Methoden der *Digital Humanities* eher ein Hilfsmittel, um zuvor mühevoll ermitteltes Wissen einfacher und sicherer zu generieren? Oder werden dank dieser Methoden auch Ergebnisse erzielt, die ohne sie nicht denkbar wären? Diese Fragen können an dieser Stelle weder abschließend noch generell im Hinblick auf alle möglichen Einsatzgebiete beantwortet werden, wohl aber exemplarisch an einigen Beispielen. Sie sind daher Anlass genug, einmal die eigene Forschung im Hinblick auf den Erkenntnisgewinn durch *Digital Humanities* zu betrachten und zugleich auf den Prüfstand zu stellen. In dem Beitrag werden die angewandten Methoden zur Auswertung originär digitaler Korpora aus E-Mails, Internetseiten und Wikipediaartikeln gemeinsam mit den erzielten Ergebnissen vorgestellt und im Anschluss jeweils kritisch hinterfragt.

10.15-11.10 Robert HESSELBACH

Universität Kassel

Digitale Analyse webbasierter Kommunikation: Sprachgebrauchsmuster in den Tweets von Juan Guaidó und Nicolás Maduro

Mit der Ausrufung Juan Guaidós zum legitimen Präsidenten Venezuelas im Januar 2019 entzündete sich ein politischer Machtkampf in dem südamerikanischen Land, der zu großen Teilen auch im Internet geführt wird. Sowohl Juan Guaidó (JG) als (Parlaments)Präsident als auch der bisherige Amtsinhaber Nicolás Maduro (NM) nutzen die kommunikativen Möglichkeiten, die ihnen Online-Dienste wie Twitter bieten, um den politischen Diskurs in Venezuela zu gestalten bzw. zu dominieren, wie die beiden Beispiele zeigen:

(1) Los venezolanos no mendigamos nuestros derechos. Por eso pronto vamos juntos a Miraflores a rescatar la oficina a todos los venezolanos, a exigir nuestros derechos. Salgamos a organizarnos cuadra por cuadra para la #OperaciónLibertad. #OperaciónLibertadAnzoátegui (JG, 23.3.2019).

(2) Invito a todo nuestro pueblo a seguir movilizados en las calles, en tribunas antiimperialistas para la formación permanente de la conciencia y la ideología Bolivariana. ¡Máxima Conciencia, Máxima Movilización! (NM, 20.3.2019).

Im meinem Vortrag möchte ich auf eine korpusbasierte Untersuchung eingehen, die alle Tweets von Guaidó und Maduró drei Monate vor und nach dem 23. Januar 2019 (dem Tag der Ausrufung Guaidós) berücksichtigt. Ziel der Studie ist es, verschiedene Sprachgebrauchsmuster in der webbasierten Kommunikation der beiden Kontrahenten sichtbar zu machen, wie es allgemein bereits Bubenhofer fordert, wenn er festhält: „Muster im Sprachgebrauch zu entdecken ist die

zentrale Operation einer korpuslinguistischen Diskursanalyse“ (2009, 18). Um dies zu erreichen, werden verschiedene Funktionen (beispielsweise zu Kollokationen und zu n-gram-Analysen) des Tools CorpusExplorer herangezogen und die Ergebnisse präsentiert. Dadurch soll gezeigt werden, wie bedeutsam digitale Analyseverfahren sowohl für die romanistische Sprachwissenschaft im Allgemeinen als auch für die Beschreibung von (politischen) Diskursen in konkreten Fällen sein können.

Literatur

Bubenhof, Noah. 2009. Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin (u.a.): De Gruyter.

11.00-11.30 **Kaffeepause**

11.30-12.15 **Mathieu CONSTANT**

ATILF

Construction, exploitation et exploration de ressources d'expressions polylexicales

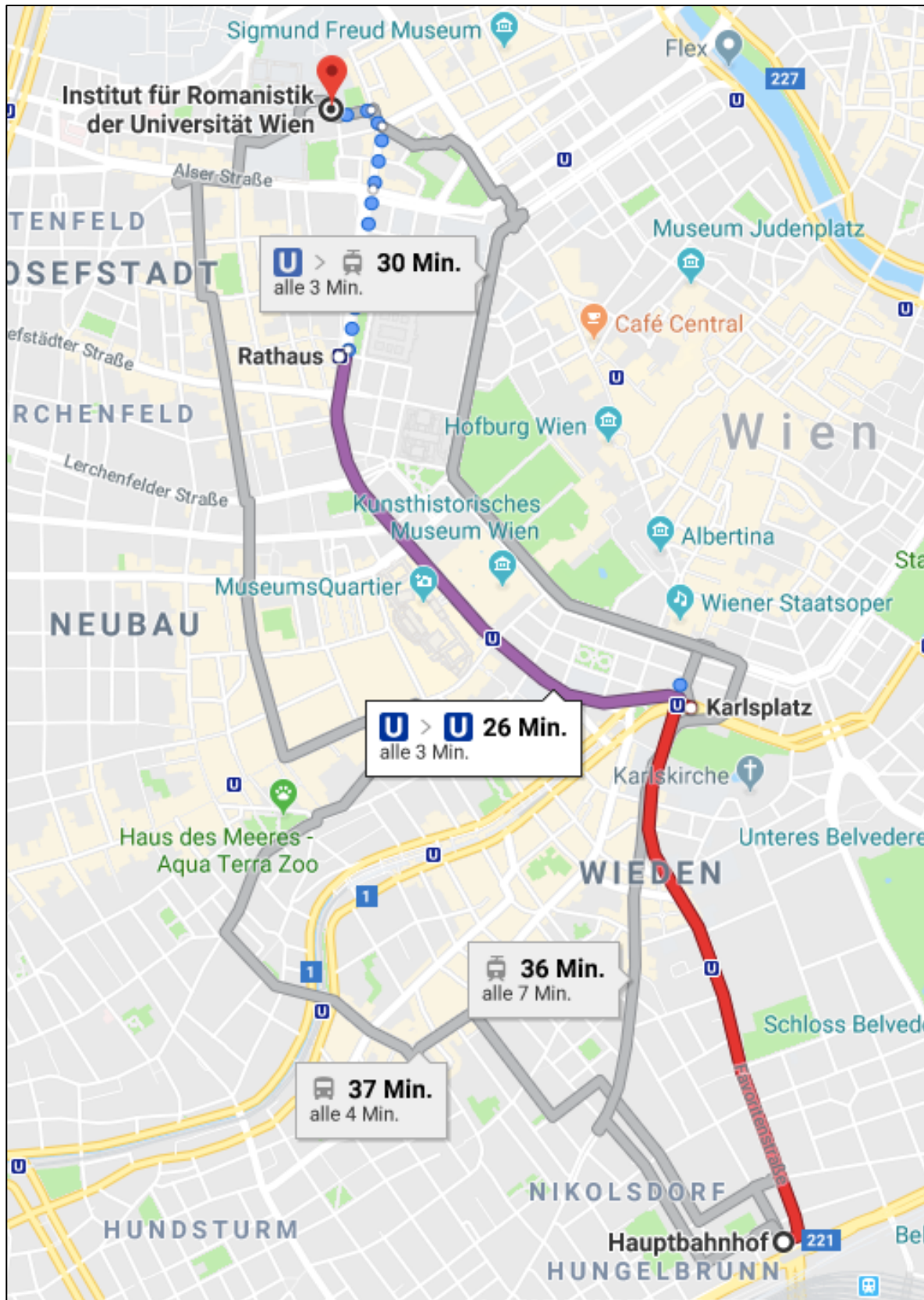
Dans cette communication, nous présenterons les ressources linguistiques développées au sein du projet PARSEME-FR dédié au traitement automatique des expressions polylexicales pour le français. Les expressions polylexicales sont des séquences de lexèmes qui se composent de manière inattendue, et qui posent de sérieux défis pour le traitement automatique des langues. En particulier, nous décrirons la méthodologie que nous avons utilisée pour construire plusieurs corpus annotés de telles expressions, en nous appuyant sur un ensemble de critères formels stricts. Nous verrons ensuite comment ces corpus ont été exploités pour apprendre des outils d'identification d'expressions polylexicales. Nous montrerons enfin un démonstrateur en-ligne permettant à la fois de tester les outils développés et d'explorer un corpus annoté en expressions verbales aligné avec un lexique syntaxique.

12.15-12.30 **Elton PRIFTI**

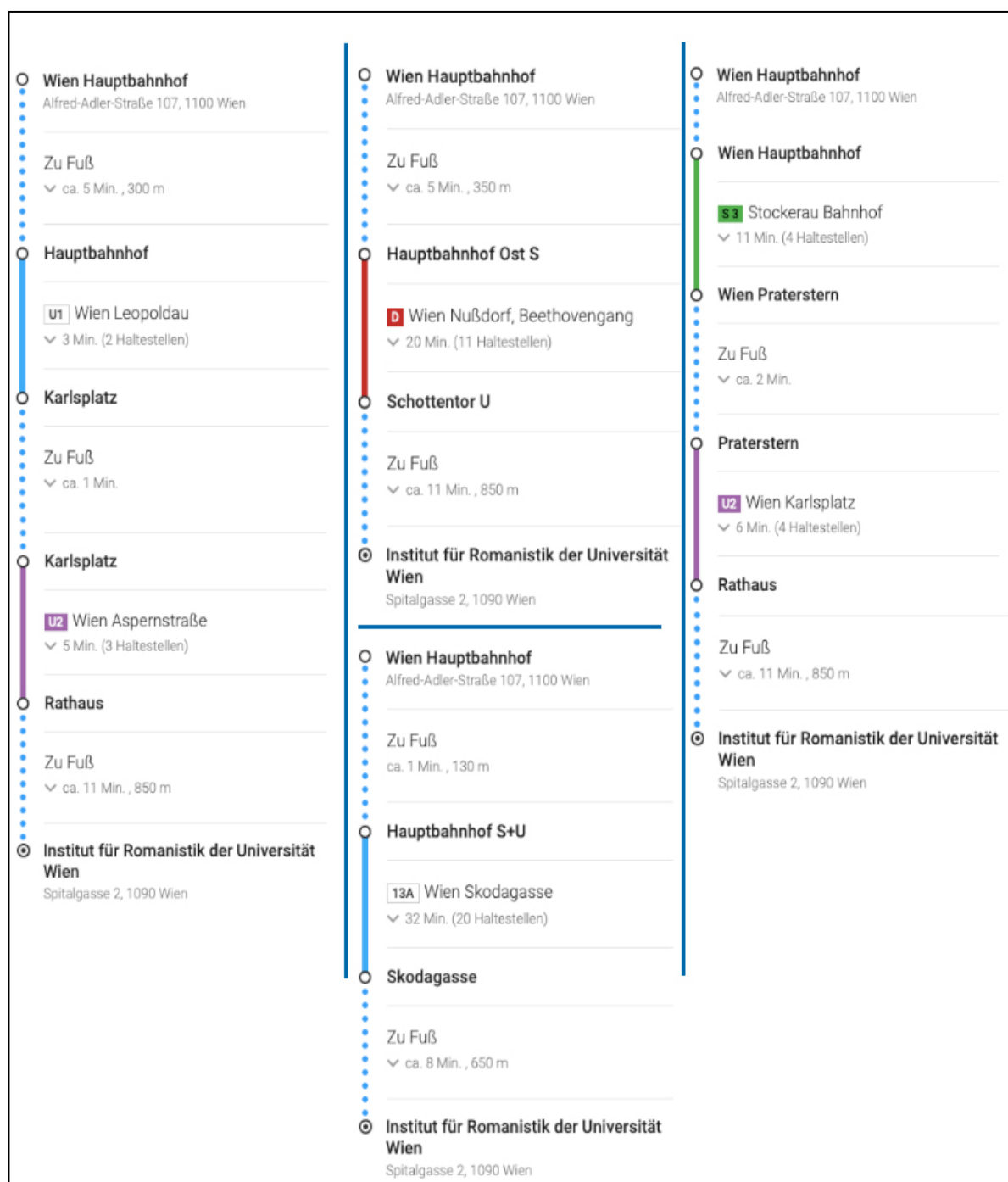
Universität Wien

Abschließende Statements und Informationen zur Publikation

Notizen



2. Wegbeschreibung: Wien Hauptbahnhof – Tagungsort (ROM14)



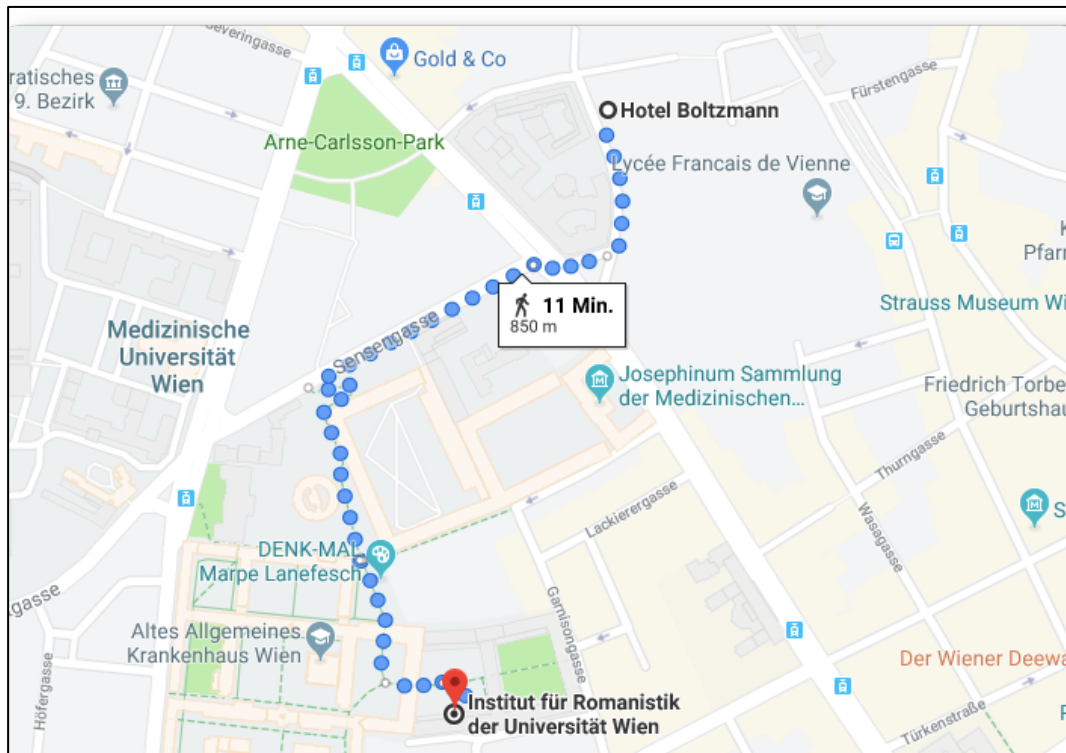
3. Mögliche Verkehrsverbindungen: Wien Hauptbahnhof – Tagungsort (ROM14)



4. Wegbeschreibung: Hotel Bleckmann – Tagungsort (ROM14)



5. Wegbeschreibung: Tagungsort (ROM14) - Restaurant Diana



6. Wegbeschreibung: Hotel Boltzmann - Tagungsort (ROM14)

