

What's in a name?

Die Rolle der Sprache zur Kultivierung von inklusiven Zugängen zu Kulturerbe

High-Steskal, Nicole

nicole.high-steskal@donau-uni.ac.at
Donau-Universität Krems, Austria

Seit geraumer Zeit wird bereits darauf hingewiesen, dass die digitalen Geisteswissenschaften trotz internationaler Ausrichtung ein von westlichen Ländern geprägtes Feld sind. Dies zeigt sich einerseits in der Kartierung von DH-Netzwerken (Russell 2014; Fiormonte 2015, 2017), andererseits auch im Methodenkoffer der digitalen Geisteswissenschaften, der hauptsächlich auf westliche Sprachen ausgelegt ist (Dombrowski 2020; Galina Russell 2014). Diesen Umstand versuchen mehrere Initiativen auf unterschiedliche Weisen sichtbar zu machen und zu beheben, u.a. das Netzwerk "multilingual dh", Global Outlook::Digital Humanities, Programming Historian oder auch das Forschungsprojekt "New Languages for NLP. Building Linguistic Diversity in the Digital Humanities" der Princeton University. Diese Initiativen besitzen unterschiedliche Schwerpunkte, doch sprechen vor allem GO::DH und Programming Historian eine wesentliche Hürde in der Zugänglichkeit von digitalisiertem Kulturerbe und der internationalen Vernetzung digital-tätiger Wissenschaftler*innen an, nämlich Sprache. Sowohl GO::DH als auch Programming Historian haben sich zur Aufgabe gemacht, sämtliche Texte und Unterlagen in möglichst viele Sprachen zu übersetzen, um den Zugang zu Wissen im Bereich der digitalen Geisteswissenschaften zu verbessern. Im vorliegenden Beitrag wird diskutiert, welche Voraussetzungen erfüllt werden müssen, damit die Zugänglichkeit von Daten durch Mehrsprachigkeit erhöht werden kann, und welche Rolle dabei offene Systeme, wie etwa Wikidata, spielen können. Es gilt dabei Wege zu finden, wie speziell die Dokumentation von Kulturerbe, und damit auch das kulturelle Gedächtnis, besser zugänglich gemacht und zur inklusiven und interkulturellen Zusammenarbeit zwischen heterogenen, internationalen Wissenschaftsgruppen eingesetzt werden kann, wie dies etwa auch in den CARE-Prinzipien zum ethischen Umgang mit Kulturdaten gefordert wird (Carroll et al. 2020).

Der Beitrag beruht auf Erfahrungen, die in Folge des CELSUS-Projektes gemacht wurden, das von der Autorin 2018 am Österreichischen Archäologischen Institut der Österreichischen Akademie der Wissenschaften in Wien begonnen wurde und nun die Grundlage eines an der Donau-Universität Krems in Beantragung befindlichen Projektes bildet. Ausgangspunkt des Projektes war es, die weitgehend deutsche Literatur zur archäologischen Stätte Ephesos (Türkei) als open access digital zur Verfügung zu stellen, damit dieses Gedächtnis auch türkischen Partner*innen besser zugänglich gemacht werden kann. Es wurde jedoch klar, dass eine technologische Aufbereitung allein nicht ausreicht, sondern der Faktor Sprache wesentlich ist, um Zugänge inklusiv zu gestalten.

Inklusion durch Sprache

Sprache bestimmt Teilhabe. Auch wenn sich die englische Sprache zur lingua franca der Digital Humanities entwickelt hat, kann

Sprache trotzdem noch eine Barriere darstellen (Dombrowski 2020). Gerade um die Zusammenarbeit mit Wissenschaftler*innen aus anderen Sprachgruppen zu stärken, ist es wichtig, Sprache und damit einhergehende Hürden zu bedenken. Für den Abbau von Sprachhürden hat Isabel Galina Russell zwei Alternativen vorgeschlagen: entweder man beginnt Daten in mehreren Sprachen zu publizieren oder Englisch als lingua franca der digital humanities inklusiver zu gestalten (Galina Russell 2014: 314). Dieser Vorschlag mag für die wissenschaftliche Bearbeitung von manchen Themen funktionieren, doch sind häufig kulturelle Daten im Fokus geisteswissenschaftlicher Forschung, wo es wichtig sein kann, Daten in ihrer ursprünglichen Sprache zu publizieren. Beispielsweise hat Roopika Risam (2018, 2019) zuletzt sehr eindrücklich die Verbindung zwischen Postkolonialismus und den digitalen Geisteswissenschaften aufgezeigt, wo sie die Teilhabe lokaler Bevölkerungen an der Aufarbeitung ihres Kulturguts und somit auch technischen und sprachlichen Zugang zu ihren Kulturdaten und ihrem digitalen Gedächtnis fordert.

Im Hinblick auf postkolonialen Argumentationen und der besseren Einbindung von unterrepräsentierten Gruppen wurde im Projekt daher entschieden die primäre Organisationsform der Archäologie, nämlich Toponyme, mehrsprachig aufzuarbeiten, damit die deutschen Begriffe für Kolleg*innen in der Türkei einfacher auffindbar sind. Die deutschen Toponyme waren bereits weitgehend bekannt, doch gab es keine strukturierte Liste der Begriffe in ihren türkischen und englischen Varianten, weshalb es notwendig war nicht nur die deutschen Toponymenlisten ins Türkische zu übersetzen, sondern zusätzlich auch die lokal verwendeten, aber in der Wissenschaft nicht rezipierten, Begriffe zu suchen. Ephesos bietet eine zusätzliche Schwierigkeit: als UNESCO-Kulturerbestätte und beliebtes Touristenziel publizieren nicht nur Wissenschaftler*innen zu diesem Ort, sondern es existiert auch eine Vielzahl an populärwissenschaftlichen Publikationen. Man befindet sich somit im Spannungsfeld zwischen eingebürgerten, touristisch verwendeten Begrifflichkeiten und wissenschaftlich fundierten – aber sonst unbekannten – Termini. Um hier größtmögliche Teilhabe zu ermöglichen, wurde der Entschluss gefasst, alle Begriffe, sowohl wissenschaftliche als auch touristische Toponyme, zu dokumentieren.

Das Projekt fokussierte zunächst auf die Abstimmung der deutschen, englischen und türkischen Begriffe. Für eine möglichst einheitliche Grundlage wurde der im Ephesos-Führer von Peter Scherrer publizierte archäologische Plan herangezogen (Scherrer 1995, Scherrer - Bier 2000, Scherrer 2000; Sun et al. 2020: 5), da er besonders weit verbreitet ist, von Wissenschaftler*innen und Tourist*innen gleichermaßen verwendet wird und in englischer und türkischer Übersetzung vorliegt. Die Kartenlegenden der unterschiedlichen Übersetzungen wurden gescannt, mittels OCR digitalisiert – wobei dies nur für die deutschen und englischen Legenden zu guten Ergebnissen geführt hat, türkische Begriffe mussten manuell verbessert werden – und mit OpenRefine normalisiert. Die entstandene Liste wurde mit weiteren Karten abgeglichen, die seitdem entstanden sind und ebenfalls auf Deutsch, Englisch und Türkisch publiziert wurden, wodurch die Liste teilweise erweitert werden konnte. Ein automatisierter Abgleich mit Geonames, Wikidata, und dem Pleiades Gazetteer hat keine nennenswerte Erweiterung des Datensatzes erbracht. Auch eine Suche nach georeferenzierten Toponymen auf Wikidata war erfolglos. Das Endergebnis war eine Liste mit 117 Toponymen in deutscher, englischer und türkischer Sprache.

Inklusion durch Netzwerke

Sprache bestimmt Netzwerke. Sprache führt nicht nur dazu, dass Hürden im Verständnis entstehen können, sondern Sprache bestimmt oft auch, mit wem man kommuniziert und welche Stimmen man erfassen kann. Eine Auswertung von Gil und Ortega (2016: 23-5) hat etwa ergeben, dass Publikationen von Personen außerhalb der Ballungszentren von Wissenschaftler*innen im "global north" oft schlichtweg nicht wahrgenommen werden und dadurch manche Forschungsbereiche und Fragestellungen aus anderen Sprach- und Kulturregionen nicht rezipiert werden. Das Netzwerk GO::DH versucht, durch eine offene Publikationsplattform interkulturellen und transdisziplinären Ansätzen aus unterrepräsentierten Regionen eine Bühne zu bieten und dadurch verstärkt die Bildung von Netzwerken zu unterstützen. In manchen Fällen kann hier aber bereits durch die Einbindung mehrsprachiger Normdaten eine gewisse Hilfestellung geboten werden.

Innerhalb des Projektes führte der zufällige Fund einer finnischen Übersetzung der Ephesos-Karte von Scherrer auf einen reichen Fundus an zusätzlichen ephesischen Toponymen in anderen Sprachen auf Wikipedia und Wikidata. Die unterschiedlichen Artikel waren zu einem großen Teil nicht miteinander verlinkt, hatten keine zusätzlichen Informationen und sind dadurch schwer auffindbar. Eine tineye-Suche nach der Ephesos-Karte von Scherrer hat eine weitere französische Übersetzung der Legenden zutage gefördert, die ebenfalls über OpenRefine in die Liste eingepflegt werden konnte. Innerhalb kürzester Zeit konnte so eine Liste an deutschen Toponymen mit türkischen, englischen, finnischen und französischen Begriffen ergänzt werden. Der zufällige Fund weiterer Begriffe hat Gils Erfahrung bestätigt und aufgezeigt, dass oft eingeschränkt innerhalb von kleinen Sprachgruppen gearbeitet wird und diese Arbeit – trotz aufwendiger Suche – sehr schwer zu finden sein kann.

Die Datengrundlage von vielen Toponymen in Wikidata hatte einen weiteren Fehler: sie waren großteils nicht georeferenziert und auch nicht definiert (fehlende Beschreibung, Geokoordinaten und "instance of - P31"-Felder). Popescu et al. (2009: 58) haben in ihrer Arbeit festgelegt, dass mehrsprachige Gazetteers drei Elemente unbedingt benötigen, damit große Toponym-Datensätzen über Sprachgrenzen hinweg zusammengeführt werden können: 1. eine Bezeichnung, 2. GPS-Koordinaten, 3. einen Typ. Die Liste wurde daher mit GPS-Daten, geographischer Zuordnung und Kurzbeschreibung weiter ergänzt. Da viele der Grunddaten bereits – zwar verteilt – auf Wikidata zur Verfügung standen, wurde beschlossen, dass die Ergänzungen und Änderungen auf Wikidata eingespielt werden und die Plattform zur Normalisierung von unterschiedlichen Gazetteers eingesetzt werden kann. Mittels OpenRefine konnten die Änderungen automatisiert in Wikidata eingespielt werden, wodurch nicht nur die Datengrundlage auf Wikidata verbessert wurde, sondern auch etliche unverknüpfte Wikipedia-Seiten zueinander in Beziehung gesetzt wurden. Von den ursprünglich 117 Begriffen, die im Scherrer-Führer genannt werden, konnten somit 56 Begriffe auf Wikidata eingespielt, verknüpft und in fünf Sprachen ergänzt werden. Wikipedia und Wikidata wurde dadurch dazu genutzt, um die Vernetzung von Wissen in anderen Sprachen zu verbessern.

Inklusion durch Technologie

Zur Teilhabe in den digitalen Geisteswissenschaften sind infrastrukturelle Voraussetzung notwendig, wie etwa verlässliche Stromversorgung und stabile Internetverbindung, sowie Compu-

terzugänge und Lizenzen für bestimmte Programme. Zugleich müssen Wissenschaftler*innen, die mit unterrepräsentierten Sprachen arbeiten, sehr viel Grundlagenarbeit leisten, um überhaupt Datensätze und digitale Methoden für Fragestellungen in unterrepräsentierten Sprachen anwenden zu können. Gerade im Umgang mit Kulturdaten sind Ansätze des minimal computing hilfreich, wie von Gil und Ortega (2016: 26) vorgeschlagen, aber auch offene Systeme, die durch die lokale Bevölkerung ohne aufwendige IT-Infrastruktur und technisches Vorwissen ergänzt werden können. Unserer Erfahrung nach hat sich gerade Wikidata und Wikipedia für diese Umsetzung ausgezeichnet geeignet. Dies zeigt sich auch dadurch, dass seit der Bereinigung der Datenlage im Dezember 2020 die Begriffe von anderen Nutzer*innen weiter bearbeitet wurden. Es sind für manche Bereiche sowohl sprachliche Ergänzungen (etwa Arabisch und Russisch) als auch Verweise auf Normdaten in anderen Sprachen hinzugekommen. Diese Ergänzungen wurden hauptsächlich bei Toponymen durchgeführt, die sehr bekannt sind, etwa Artemistempel oder House of Virgin Mary, eröffnen wiederum neue Möglichkeiten für das Projekt und machen es möglich, Publikationen und Quellen in Türkisch aber auch anderen (noch nicht antizipierten) Sprachen zu Ephesos zu erfassen. Zusätzlich hat sich auch gezeigt, dass einige Änderungen durch User mit mobilen Endgeräten durchgeführt wurden und somit ein niederschwelliger Zugang gerade für User mit alternativen Internetzugängen möglich war (siehe z.B. <https://www.wikidata.org/w/index.php?title=Q43018&action=history>; letzter Zugriff: 1. Dezember 2021). Die Anzahl der Änderungen ist leider nicht quantifizierbar.

Fazit

Der Aspekt der Inklusion und Teilhabe durch türkische Partner*innen in der Wissenschaft und in der Lokalbevölkerung wurde im Projekt anfangs nur im Hinblick auf technologische Zugänglichkeit gedacht. Das Anliegen des Projektes war es jedoch die Zugänglichkeit wissenschaftlicher Dokumentation einer Kulturerbestätte zu erhöhen, weshalb schnell klar wurde, dass mehr Arbeit notwendig war als nur Digitalisate online zu stellen und die Mehrsprachigkeit der nächste logische Schritt war. Obwohl das Augenmerk zunächst nur auf der türkischen Sprache lag, konnten per Zufallsfund weitere Sprachen nach einem einheitlichen Workflow hinzugefügt werden, wobei die weite Verbreitung einer einzelnen Karte die Datenlage unterstützte. Die Erfahrung zeigte, dass teilweise bereits sehr viele Daten vorhanden sind, diese aber erst gesucht und zusammengeführt werden müssen. Wikidata und Wikipedia hatten für das Projekt den Vorteil, dass die bereinigten Datensätze von anderen schnell gefunden werden können, und durch einheitlich Ansprache und Normdaten verknüpft sind. Wikidata hatte den zusätzlichen Vorteil, dass die Wikimedia-Plattform kostenfrei und niederschwellig in der Nutzung und Einpflegung von Daten ist. Die Ergebnisse des Projektes zeigen, dass dieser Prozess zur verbesserten Wahrnehmung von Forschung aus unterrepräsentierten Gruppen führen und dadurch ein multikulturelles und inklusives Gedächtnis entstehen kann.

Bibliographie

Carroll, Stephanie / Russo, Ibrahim / Garba, Oscar L. / Figueroa-Rodríguez, Jarita / Holbrook, Raymond / Lovett, Siameon / Materechera, Mark Parsons / u. a. (2020). "The CARE

Principles for Indigenous Data Governance“. *Data Science Journal* 19 (1): 1-12. <https://doi.org/10.5334/dsj-2020-043>.

Dombrowski, Quinn (2020): “What’s a „Word“: Multilingual DH and the English Default.” <https://www.quinndombrowski.com/?q=blog/2020/10/15/whats-word-multilingual-dh-and-english-default> [letzter Zugriff 15. Juli 2021]

Fiormonte, Domenico (2015): “Towards Monocultural (Digital) Humanities?” in: *Infolet*. <https://infolet.it/2015/07/12/monocultural-humanities/> . [letzter Zugriff 15. Juli 2021]

Fiormonte, Domenico (2017): “Digital Humanities and the Geopolitics of Knowledge.” in: *Digital Studies/Le champ numérique* 7: 1-18.

Galina Russell, Isabel (2014): “Geographical and Linguistic Diversity in the Digital Humanities.” in: *Literary and Linguistic Computing* 29: 307–316.

Gallon, Kim (2016): “Making a Case for the Black Digital Humanities” in: *Debates in the Digital Humanities*. <https://dhdebates.gc.cuny.edu/read/untitled/section/fa10e2e1-0c3d-4519-a958-d823aac989eb> . [letzter Zugriff 15. Juli 2021]

Gil, Alex / Ortega, Élika (2016): “Global Outlooks in Digital Humanities: Multilingual Practices and Minimal Computing” in: Crompton, Constance / Jane, Richard J. / Siemens, Ray *Doing Digital Humanities*. London / New York: Routledge 22-34.

Laurini, Robert (2015): “Geographic Ontologies, Gazetteers and Multilingualism” in: *Future Internet* 7: 1–23.

Laurini, Robert (2017): “Gazetteers and Multilingualism” in: *Geographic Knowledge Infrastructure*. London: ISTE Press Ltd 157–182.

Piller, Ingrid/Takahashi, Kimie (2011): “Linguistic Diversity and Social Inclusion” in: *International Journal of Bilingual Education and Bilingualism* 14: 371–381.

Popescu, Adrian/Grefenstette, Gregory/Bouamor, Houda (2009): “Mining a Multilingual Geographical Gazetteer from the Web” in: *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. Milan, Italy: IEEE. 58–65. <http://ieeexplore.ieee.org/document/5284918/> .

Risam, Roopika (2018): “Decolonizing the Digital Humanities in Theory and Practice” in: Sayers, Jentery (ed.): *The Routledge Companion to Media Studies and Digital Humanities*. New York: Routledge, Taylor & Francis Group, 78–86.

Risam, Roopika (2019): *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston, Illinois: Northwestern University Press.

Scherrer, Peter (1995): *Ephesos - der neue Führer: 100 Jahre österreichische Ausgrabungen; 1895 - 1995*. Wien: Österreichisches Archäologisches Institut.

Scherrer, Peter (2000): *Efes: Rehberi*. Istanbul: Ege Yayınları.

Scherrer, Peter / Bier, Lionel (2000): *Ephesus: The New Guide*. Istanbul: Ege Yayınları.

Sun, Kai / Hu, Yingjie / Song, Jia / Zhu Yunqiang (2020): “Aligning Geographic Entities from Historical Maps for Building Knowledge Graphs” in: *International Journal of Geographical Information Science*: 1–30.