

EIN ALPINER „WORTSCHATZ“

Die Zeitschrift des Deutschen und Österreichischen Alpenvereins bildet ein einzigartiges Textkorpus, das von Claudia Posch und Gerhard Rampl digital aufbereitet und der wissenschaftlichen Gemeinde zugänglich gemacht wird.



42.000 BUCHSEITEN wurden eingescannt, nun werden die „alpinen“ Texte in ihre Bestandteile – Sätze und Wörter – zerlegt.

Manchmal gibt es scheinbar triviale Gründe für ein Forschungsprojekt. „Wenn die Schweizer so etwas haben, dann brauchen wir das auch“, dachten sich Claudia Posch und Gerhard Rampl, als sie bei Recherchen auf das eidgenössische Projekt „Text und Berg“ stießen. Ein Projekt, in dem die seit 1864 publizierten Jahrbücher des Schweizer Alpen-Clubs digitalisiert und in Form eines annotierten linguistischen Textkorpus verfügbar gemacht wurden. Posch und Rampl machten sich also auf den Weg und stießen auf einen Schatz, der für eine fachgerechte Hebung schon gut vorbereitet war – die Zeitschrift des Deutschen und Österreichischen Alpenvereins.

Erstmals erschien die Zeitschrift im Jahr 1872 nach dem Zusammenschluss der beiden Alpenvereine. In dem Jahrbuch publizierten die führenden Forscher und Bergsteiger ihrer Zeit, ab 1884 auch Wissenschaftlerinnen und Alpinistinnen. Thema war die Erforschung und Erschließung der Berge – zuerst jene der Alpen, später auch jene anderer europäischer und außereuropäischer Gebirge. „Die Zeitschrift beinhaltet eine große inhaltliche Bandbreite“, sagt Rampl, „sie reicht von Erstbesteigungsbeschreibungen über geologische, geografische und biologische Abhandlungen bis hin zu historisch-sprachwissenschaftlichen Aufsätzen.“ Doch Rampl, so wie Claudia Posch Forscher am Bereich Sprachwissenschaft

des Instituts für Sprachen und Literaturen, ist an etwas anderem interessiert, beinhalten doch die rund 42.000 Buchseiten einen Schatz, genauer genommen einen alpinen Wortschatz, dem sich Posch und Rampl nun aus sprachwissenschaftlicher Sicht nähern wollen.

Gescannte Vorarbeit

„Ein großer Vorteil für uns ist, dass die Basisarbeit schon vor rund zehn Jahren geleistet wurde“, sprechen Rampl und Posch den Umstand an, dass die Zeitschriften der Jahrgänge 1872 bis 1998 schon vor etwas mehr als zehn

Für den nächsten Projektschritt können Posch und Rampl auf Schweizer Programmier-Know-how zurückgreifen. Das gesamte Textkorpus wird in Sätze zerlegt, diese wiederum in Wörter tokenisiert, sprich aufgeteilt, diese anschließend annotiert, sprich mit der jeweiligen Wortart versehen. Ende 2016 soll dieser Projektteil beendet sein, mit Hilfe von NER (Named Entity Recognition) soll das Textkorpus auch noch auf Eigennamen durchforstet werden. Eigennamen, die in älteren Ausgaben der Alpenvereinszeitschrift teilweise zum ersten Mal in schriftlicher Form auftauchen. „In



„Ein großer Vorteil für uns ist, dass die Basisarbeit schon vor rund zehn Jahren geleistet wurde. Damals wurde die Jahrgänge von 1872 bis 1998 eingescannt. Das gesamte Korpus wurde nun mit OCR neu erfasst, als erstes Teilergebnis des Projekts liegt nun ein ‚relativ gut‘ erkannter Text vor.“

Claudia Posch und Gerhard Rampl

Jahren an der Universität Innsbruck eingescannt worden waren (und auch online unter anno.onb.ac.at bzw. www.literature.at einsehbar sind). Nicht optimal war allerdings – dem damaligen Stand der Technik geschuldet – die Texterkennung, speziell bei den Ausgaben in Fraktur, aber auch bei bestimmten Worten: die Osttiroler Gemeinde Kals taucht oft als Kais auf, das Wort Wasser, erzählt Posch, wurde in einem Band immer als Master gelesen.

Das gesamte Korpus wurde daher mit OCR (Optical Character Recognition) neu erfasst sowie automatisch und per Hand kontrolliert, als erstes Teilergebnis des – von der Österreichischen Akademie der Wissenschaften geförderten – Projekts „Alpenwort“ liegt nun ein, so Rampl schmunzelnd, „relativ gut erkannter Text“ vor. Eine Textsammlung über 42.000 Buchseiten hinweg, die via Textsuche nicht nur den Archivaren und Forscherinnen des Alpenvereins, sondern der gesamten alpin-interessierten Wissenschaftscommunity neue Forschungsfelder und -möglichkeiten eröffnet.

dieser Zeit war die Benennung der Berge, von Wänden und Routen noch nicht abgeschlossen“, weiß Rampl, der sich aus der sprachwissenschaftlichen Analyse dieser Namen neue Einsichten in Benennungsmotive erwartet.

Claudia Posch wiederum möchte sich aus feministisch-diskursanalytischer Sicht dem Textkorpus annehmen, möchte Fragen nachgehen, „wie Bergsteigen konzipiert und beschrieben wurde, was daran spezifisch genderbezogen ist und ob sich in den Texten spezielle Vorstellungen von Maskulinität, Stärke oder Kameradschaft finden lassen“. Auch wiederkehrende Muster des Sprachgebrauchs, z.B. n-gramme, der Gebrauch von Personalpronomina, Intensivierer, „Geocollokationen“ u.v.m. können somit systematisch erforscht werden. Bei letzterem, so Claudia Posch, gehe es darum, in welchem Zusammenhang und wie über bestimmte fremde Gebiete geschrieben wurde. Wobei es durchaus sein kann, so die Sprachwissenschaftlerin, dass man dabei mehr über den Schreiber als über das Fremde erfahre. ah 