Nachtrag zum Exposé „Das konzeptuelle Modell des volkstümlichen Märchens und seine maschinelle Umsetzung“ eingereicht von Elguja Dadunashvili

Die vergleichende Märchenforschung befasst sich mit dem Phänomen der erstaunlichen Nähe der Märchensujets in verschiedenen, sowohl räumlich als auch kulturell oft weit voneinander entfernten, ethnischen Gruppen. Die Wirkung dieses Phänomens ist anhand der in verschiedenen nationalen Repertoires stabil wiederholbaren Textelemente nachweisbar. Diese Elemente sind die gesamten Geschichten, bekannt als Märchentypen, Episoden, also relativ kleinere Teile der Geschichte, und Motive als kleinste Elemente des Märchens. Die vergleichende Märchenforschung interessiert sich für folgende zwei Fragen:

1. Wie sieht der eigenständige Typenbestand jedes Märchenrepertoires aus?
2. Wie lautet die stabile Struktur der in diesem Repertoire als populär geltenden Märchentypen?

Die Abhandlung der beiden Fragen ist unmittelbar von einer präzisen Segmentierung des Forschungsobjektes abhängig. Die Wahrhaftigkeit dieser Segmentierung erfordert die empirische Nachweisbarkeit der für die Segmentierung verwendeten Einheiten. Nur im Falle der Anwendung eines solchen Kriteriums ist es möglich, über die Beschreibung oder Modellierung eines selbstregulierenden Systems zu sprechen. Entsprechend der neuesten, erstmals in der vorliegenden Publikation bekannt gegebenen Erkenntnisse besteht ein solches Kriterium im Wechsel der in der Erzählung auftretenden Figuren. Das Finden eines Schlüsselelements zur Struktur des Forschungsgegenstandes ermöglicht eine umfassende Beschreibung der gesamten Struktur in einem konzeptionellen Modell.

Nach der Entwicklung eines konzeptuellen Modells des Forschungsgegenstandes wird ein Markup für die Erfassung und Etikettierung seiner einzelnen Fälle und Variationen erstellt. Jedes Markupelement stellt ein 4-Tupel dar. Die Attribute des Tupels beschreiben alle Eigenschaften des Motivs, einschließlich seiner Zugehörigkeit zum Typ, zur Episode, zum Handlungsinhalt und zur Nomenklatur der handelnden Figuren. Die meisten Attribute basieren auf dem bereits etablierten Vokabular der Märchenforschung.

Bei Übereinstimmung zwischen der Etikettierung und der objektiven Struktur des Märchens sollte es möglich sein, diese Struktur auch für die maschinelle Analyse erkennbar zu machen. Obwohl ein vollständiger Workflow für die automatische Etikettierung von Märchen noch entwickelt werden muss, lässt der aktuelle Stand der Forschung dieses Ziel in greifbare Nähe rücken. Besonders vielversprechend ist die Fähigkeit eines künstlichen Assistenten, der für die "One vs. All"-Klassifikation entwickelt wurde.

Die Funktionsweise dieses Assistenten basiert auf zwei aufeinander folgenden Verfahren. Zunächst wird ein Machine-Learning-Modell der logistischen Regression erstellt und mit manuell etikettierten Trainingsdaten gefüttert. Der Algorithmus erstellt ein Modell positiv etikettierter Motive, das aus positiven und negativen Merkmalen besteht.

Bei der Vorhersage der gesuchten Motive kann dieses Modell scheitern, da in den vorhandenen Textabschnitten der Forschungsdaten sowohl positive als auch negative Merkmale nebeneinander auftreten können. Um dieses Problem zu lösen, wird ein weiteres Verfahren angewendet: Nur die positiven Merkmale mit den höchsten Koeffizienten werden aus dem erstellten Modell extrahiert und als synthetisches Metamotiv gespeichert. Anschließend wird das Metamotiv mit allen anderen Textteilen der Forschungsdaten anhand der Metrik Cosinus-Ähnlichkeit verglichen. Dieses Verfahren ermöglicht den Zugriff auf Textabschnitte, die nicht nur das gesuchte Motiv, sondern auch andere Motive enthalten.

Diese Methode bietet zwei wesentliche Vorteile: Erstens übertrifft sie die Präzision der Vorhersagefunktion der logistischen Regression und zweitens ist die Erstellung eines solchen Metamotivs nur anhand weniger Trainingsdaten möglich, während die Menge der Forschungsdaten praktisch irrelevant ist.

Der Assistent wird als zuverlässiges Instrument im Rahmen des Projektes "Kaukasisches Erzählgut im Diskurs der vergleichenden Märchenforschung"[[1]](#footnote-1) eingesetzt. Derzeit wird intensiv daran gearbeitet, die im Rahmen des Projekts bearbeiteten Forschungsdaten durch das TextGrid-Repositorium zugänglich zu machen. In der Umsetzungsphase befindet sich die Erstellung eines TextGrid-Corpus namens "Kaukasische Folklore", das Originaltexte in mehr als zwei Dutzend Sprachen sowie Übersetzungen ins Russische, Deutsche und Georgische enthält und somit für die vergleichende Analyse bereitstellt.

Die Freischaltung des im GitHub-Repositorium angelegten Codes und der gesamten digitalen Infrastruktur kann als Ergänzung des geplanten Aufsatzes nur als parallele Veröffentlichung erfolgen.

Die Einreichung mit dem Schwerpunkt auf die maschinelle Umsetzung des konzeptuellen Modells kann natürlich als eine Projektvorstellung veröffentlicht werden. Jedoch stellt der dabei entwickelte theoretische Hintergrund – die Struktur bzw. das konzeptuelle Modell des Märchens – eine tiefgreifende Revision in der Theorie der vergleichenden Märchenforschung dar und lässt sich somit eher als einen Fachartikel wahrnehmen.

1. URL: <https://gepris.dfg.de/gepris/projekt/501928518> [↑](#footnote-ref-1)