Elguja Dadunashvili

Das konzeptuelle Modell des volkstümlichen Märchens und seine maschinelle Umsetzung

Exposé

Basierend auf den unten aufgeführten axiomatischen Thesen strebe ich einen zuverlässigen Zugang zur Formalisierung des Märchens an und beabsichtigen folglich, die automatische Analyse des Märchens zu ermöglichen.

Die Thesen lauten wie folgt:

1. Der Inhalt eines Märchens setzt sich aus einer Abfolge einzelner Handlungen (Aktionen) zusammen.
2. Jede Handlung im Märchen wird von einer handlungstragenden Figur oder mehreren solcher Figuren geleitet.
3. Jeder Auftritt einer solchen Figur oder mehrerer Figuren löst eine neue Handlung aus.
4. Das Motiv, als konzeptueller Begriff für das kleinste, untrennbare Inhaltselement im Text, entspricht einer Handlung, die im Verlauf eines solchen Auftritts oder anders gesagt, einer solchen Szene dargestellt wird (ein Motiv darf nicht größer sein als die Szene).
5. Die Markierung der objektiv erkennbaren Szene im Text entspricht der Markierung der Handlungsabschnitte und somit auch der Markierung der Motive.
6. Die Kategorisierung der Handlungsabschnitte sowie der dabei beteiligten Figuren entspricht der Definition inhaltlicher Konstanten für das gesamte Genre.
7. Die Schematisierung der gesetzmäßigen und folgerichtigen Reziprozität zwischen diesen Konstanten entspricht der Modellierung einer universellen Struktur des Genres.
8. Nach der formalisierten Beschreibung der inhaltlichen Konstanten und ihrer gegenseitigen Beziehungen im Märchen ist es möglich, jeden einzelnen Vertreter dieses Genres automatisch zu annotieren, auszuwerten und ein maschinell generiertes Bild von Ähnlichkeiten und Unterschieden zwischen den verglichenen Daten anzufertigen.

Die wichtigste Erkenntnis dieser Arbeit besteht in der Feststellung eines bisher fehlenden objektiven Kriteriums für die Erfassung des Motivs als kleinstes erzählerisches Inhaltselement im Märchen (Punkte 1-4 oben).

Der erste Teil der vorliegenden Arbeit beschreibt das aufgrund dieser Erkenntnisse entwickelte System von Handlungskategorien und den Kategorien handlungstragender Figuren. Dies wird mit dem Versuch der Schematisierung der universellen Struktur des Märchens fortgesetzt. Der nächste Abschnitt der Arbeit widmet sich der Bereitstellung des Markups als Mittel zur standardisierten Kodierung bzw. Dekodierung des Inhalts des Märchens. Im Anschluss präsentiert die Arbeit einige Module des künstlichen Assistenten für die semiautomatische Annotation des volkstümlichen Märchens sowie Tools für die visualisierte Darstellung der gegenseitig verglichenen Daten. Die in diesem Abschnitt der Arbeit diskutierte digitale Infrastruktur ist durch die GitHub-Publikation vollständig zugänglich.

Im Allgemeinen präsentiert diese Arbeit eine digitale Weiterentwicklung der Methode der vergleichenden Märchenforschung und somit des wissenschaftlichen Erbes der internationalen Märchenforschung. Zusammenfassend untersucht die vorliegende Arbeit allgemeine Regeln für die Organisation und Kombination von inhaltlichen Elementen auf allen Ebenen des Märchens und liefert eine formalisierte Beschreibung der erkannten Ordnung. Basierend auf diesem konzeptuellen Modell des Genres entwickle ich ein Machine-Learning-Modell, das es uns ermöglicht, die kostspieligen und zeitaufwändigen Aufgaben der vergleichenden Märchenforschung zu automatisieren.

Ein besonderer Mehrwert der erzielten Forschungsergebnisse liegt in der Bereitstellung eines künstlichen Assistenten zur Erkennung inhaltlich-struktureller Elemente innerhalb einer unüberschaubaren Menge von Textdaten. Hervorzuheben ist dabei die Zusammenführung der Leistungen von zwei unabhängigen Algorithmen in einem einzigen. Nur durch diesen Ansatz ist es möglich, die Vorhersagefunktion erfolgreich zu verwenden, selbst wenn die nach dem Modell getrennten positiven und negativen Klassifikationsmerkmale in einem untersuchten Textabschnitt zusammengehalten werden.

**Gliederung**

1. Die Struktur des Märchens
   1. Einführung
   2. Handlungstragende Figuren im Märchen
   3. Handlungsabschnitte im Märchen und ihre Organisation
2. Markup
   1. Tupel m
      1. Attribut a
      2. Attribut b
      3. Attribut c
      4. Attribut d
   2. Musterbeispiel der Anwendung des Markups
   3. Motiv, Episode, Märchen
3. Maschinelle Analyse des Märchens
   1. Textkorpus
   2. Künstlicher Assistent für die semiautomatische Etikettierung der Texte
   3. Tools für die Annotation und Auswertung der Daten
4. Ausblick