# **Das konzeptuelle Modell des volkstümlichen Märchens und seine maschinelle Umsetzung**

## **Die Struktur des Märchens**

### **Einführung**

Basierend auf den nachfolgend aufgeführten axiomatischen Thesen verfolge ich das Ziel, einen zuverlässigen Ansatz zur Formalisierung des Märchens zu entwickeln und dadurch die automatische Analyse von Märchen 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 konzeptioneller Begriff für das kleinste, untrennbare Inhaltselement im Text, entspricht einer Handlung, die im Verlauf eines solchen Auftritts oder einer Szene dargestellt wird. Ein Motiv darf nicht größer als die Szene sein.
5. Die Markierung der objektiv erkennbaren Szene im Text entspricht der Markierung der Handlungsabschnitte und damit auch der Markierung der Motive.
6. Die Kategorisierung der Handlungsabschnitte sowie der beteiligten Figuren entspricht der Festlegung von inhaltlichen Konstanten für das gesamte Genre.
7. Die Schematisierung der gesetzmäßigen Wechselwirkungen zwischen diesen Konstanten entspricht der Modellierung einer universellen Struktur des Genres.
8. Nach der formalen Beschreibung der Inhaltskonstanten und ihrer Wechselwirkungen im Märchen ist es möglich, jeden Vertreter dieses Genres automatisch zu annotieren und zu bewerten sowie ein maschinell erzeugtes Bild der Ähnlichkeiten und Unterschiede zwischen den verglichenen Daten zu erstellen.

Die wichtigste Erkenntnis dieser Arbeit ist die Identifizierung eines bisher fehlenden objektiven Kriteriums zur Erfassung des Motivs als kleinstes narratives Inhaltselement im Märchen (Punkte 1 – 4 oben).

Der erste Teil dieser Arbeit beschreibt das auf diesen Erkenntnissen basierende System der Handlungskategorien und der handlungstragenden Figuren. Anschließend wird der Versuch unternommen, die universelle Struktur des Märchens zu schematisieren.

Der nächste Abschnitt der Arbeit widmet sich der Bereitstellung von Markup als Mittel zur standardisierten Kodierung des Inhalts. 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 vollständig über die GitHub-Veröffentlichung zugänglich.

### **Forschungsstand**

Im Allgemeinen präsentiert diese Arbeit eine digitale Weiterentwicklung der Methode der Vergleichenden Märchenforschung. Dabei stehen die Errungenschaften der historisch-geographische Methode[[1]](#footnote-1) von Finnischen Schule sowie die von V. J. Propp[[2]](#footnote-2) entwickelte strukturalistische Methode im Vordergrund.

Bei der historisch-geographischen Methode wird die Struktur des Märchens als eine hierarchische Ordnung betrachtet, die aus folgenden Elementen besteht: Motiv, Episode und Typ.

Die Terminologie zur Beschreibung der Strukturelemente von Märchen ist oft uneinheitlich, es sei denn, es liegt eine einheitliche Gliederungsmethode zugrunde. Die oben dargestellte Struktur geht auf den Systematisierungsversuch von Aarne zurück. Nach seinem Modell setzt sich das Märchen aus folgenden Elementen zusammen (Aarne, 1913: 65):

1. Erzählung = Element der obersten Kategorie,
2. Hauptteil = mittleres Element,
3. Hauptzug = das kleinste Element.

Diese Struktur erkennt man noch besser bei seiner praktischen Umsetzung. Bei der Analyse des Märchens *ATU 670 - Tiersprachenkundiger Mann* geht Aarne mit der Segmentation des Inhalts wie folgt vor:

*„Den Inhalt des Märchens können wir in drei Hauptteile zerlegen: der erste handelt von dem Erlernen der Tiersprache, der zweite von dem durch die Kenntnis verursachten Streit zwischen dem Manne und der Frau, der den Mann zu den Todesvorbereitungen führt und der dritte von der Beendung des Streites.*

*Die Hauptzüge des ersten Teiles des Märchens sind der tiersprachenkundige Mann oder die Hauptperson des Märchens und die Gelegenheit zum Erlernen der Sprache nebst deren Lehrer“* (Aarne 1914:24).

Die Hauptteile entsprechen hier den Erzählzügen und stellen somit die Elemente der mittleren Kategorie dar. Die Hauptzüge der Teile entsprechen ihrerseits den Bestandteilen der mittleren Elemente und repräsentieren somit die Elemente der niedrigsten Hierarchiestufe.

Eine wesentliche Präzisierung dieser Klassifikation in der entsprechenden Terminologie geht auf Kaarle Krohn zurück. Seiner Empfehlung nach werden die Teile der Erzählung als "Episode" und die Teile der Episode als "Momente" bezeichnet (Krohn, 1926: 29). Trotz seines Arguments, den Begriff "Moment" anstelle von "Motiv" zu verwenden, wurde letzterer später bevorzugt. Schließlich haben sich die Termini 'Typ', 'Episode' und 'Motiv' als etablierte Begriffe herausgebildet, um die hierarchisch segmentierten Erzählungsteile in der vergleichenden Märchenforschung differenziert darzustellen. (vgl. Anderson, 1934/40: 515).[[3]](#footnote-3)

Ein empirisch beobachtbares Merkmal, auf das man sich bei der Definition der genannten Elemente stützen könnte, fehlt in der Beschreibung der Struktur. Eine Ausnahme bildet die gesamte Erzählung, die als Grundlage für die Definition des Märchentyps dient. Wenn eine Geschichte erzählt wird, lässt sich zumindest ein Typ darauf beziehen.

Der formalistische Ansatz in der Märchenforschung betont die Möglichkeit der Verallgemeinerung einzelner ähnlicher Handlungen im Märchen und spricht über eine gewisse Anzahl sogenannter Funktionen, die die gesamte Komposition des Märchens bilden. Es werden jedoch keine physischen Merkmale im Text festgelegt, anhand derer das Vorhandensein dieser Elemente erkannt werden kann. Hinweise darauf, dass diese Funktionen Prädikaten entsprechen oder mit konkreten Protagonisten verbunden sind, können auf kein objektives Zeichen verweisen. Die Verallgemeinerung der von einer Figur ausgeübten Handlungen kann wiederum davon abhängen, welcher Teil oder welche Facette der Handlung vom Forscher priorisiert wird.

Die Bestätigung des Wechsels zwischen den vermeintlichen Strukturelementen im Text durch ein objektives, in jedem einzelnen Text regelmäßig vorkommendes Zeichen unterscheidet sich von allen vorherigen Lehren über die Struktur des Märchens und legt somit andere grundlegende Voraussetzungen für die Diskussion über dasselbe Problem fest.

### **Handlungstragende Figuren im Märchen**

Anmerkung: Das Charakteristikum der handelnden Figuren, ergibt sich aus der Analyse der Handlungsabschnitte im Märchen und nicht umgekehrt, wie es in der Gliederung des vorliegenden Aufsatzes dargestellt wird. Die Strukturierung der vorliegenden Arbeit trotzdem auf verkehrter Weise zielt darauf ab, die Diskussion über die Handlungsabschnitte im Märchen vor der Notwendigkeit zu bewahren, zahlreiche neue Begriffe zu kommentieren. Dies wäre unvermeidlich, wenn diese Abhandlung exakt dem tatsächlichen Forschungsprozess folgen würde.

Die Identifizierung und Kategorisierung der handlungstragenden Figuren im Zaubermärchen erforderten zunächst die Auslegung der im Märchen funktionierenden Weltordnung. Basierend auf den Erkenntnissen empirischer Forschung lässt sich die bisher bekannte Weltordnung im Zaubermärchen als hierarchisches System folgender Akteure beschreiben:

**Oberhaupt:**

Das Oberhaupt ist in den meisten Fällen ein König oder das Familienoberhaupt; selten tritt an dieser Stelle auch eine Gottheit auf.

**Nachfolger:**

Als Nachfolger gilt das handlungsfähige Kind des Oberhauptes.

**Nachwuchs:**

Der Nachwuchs bezieht sich auf die Vertreter der dritten Generation, zunächst das Kind des Nachfolgers.

**Ebenbürtiger des Nachfolgers, der von außerhalb des Clans bzw. der Familie stammt:**

Der Ebenbürtige ist eine Person mit denselben Rechten wie der Nachfolger, gehört jedoch nicht dem Clan oder der Familie des Nachfolgers an.

**Priester / Magier / Totem:**

Diese Figur kann durch besondere Merkmale, Fähigkeiten oder Fertigkeiten hervorgehoben werden, wie zum Beispiel ein Bart, Alter oder Hässlichkeit. Ein Magier kann auch der Geist eines verstorbenen Familienmitglieds sein. Häufig besitzen Totems oder personifizierte Naturerscheinungen wie Wind, Frost oder Sonne magische Kräfte. Vertreter dieser Kategorie stehen außerhalb der Macht anderer Ordnungssubjekte derselben Welt.

**Untertan / Berufstätiger / Beamte:**

Typische Untertanen umfassen Vertreter unehrlicher Berufe wie Müller, Hirt, Türmer oder Portier, aber auch andere Berufstätige wie Goldschmied (wobei der Schmied, genauso wie Heilkundiger, eher mit einem Priester verglichen wird), Wirt, Kutscher, Kleinhändler; ebenso wie Beamte, zum Beispiel Minister, Offizier, Berater, usw. Wir sollten Untertanen / Berufstätige / Beamte als Vertreter des Standes vom Nachfolger mit demselben beruflichen Attribut unterscheiden. Ein gutes Beispiel wäre die häufige Verwendung eines Soldaten sowohl als Held als auch als Untertan im Märchen.

**Besitz / Gut / Sache:**

Dies umfasst materielles und / oder immaterielles Erbe, das als Objekt der Begierde, als Werkzeug oder als Zaubermittel fungieren kann. Manchmal kann es auch ein Helfer sein, wie zum Beispiel ein Ring oder eine Lampe.

Jedes Mitglied dieser Ordnung kann durch **männliche** oder **weibliche** Wesen repräsentiert werden, was jedoch nicht automatisch für Besitz, Gut und Sache gilt, es sei denn, sie sind personifiziert.

Diese Ordnung gilt für beide Welten des Märchens. Diese werden aus der Perspektive des Helden als **eigene** und **fremde** Welten bezeichnet.

**Die Funktion** der handlungstragenden Figur im Märchen werden durch ihre Zugehörigkeit zu verschiedenen Welten (eigene vs. fremde) sowie durch ihre Status in diesen Welten bestimmt. Insgesamt lassen sich folgende drei Kategorien von handlungstragenden Figuren identifizieren:

A. Protagonisten der eigenen Welt

B. Protagonisten der fremden Welt

C. Protagonisten der neutralen Welt (Figuren, die in beiden Welten zuhause sind).

Das Fremde bedeutet dabei nicht zwangsläufig Feindseligkeit; dennoch spiegelt jeder Konflikt im Zaubermärchen die Auseinandersetzung zwischen den Vertretern dieser beiden Welten wider.

Jeder Protagonist kann sich gegenüber seiner Welt sowohl freundlich als auch feindlich verhalten. Bei Protagonisten ohne feste Zugehörigkeit zu einer der Welten ist entscheidend, ob sie auf der Seite der Vertreter der eigenen Welt oder auf der Seite der fremden Welt stehen. Somit ergeben sich zwei Antipoden für jeden Protagonisten: das richtige (***r***) und das falsche (***f***).

In der folgenden Matrix versuchen wir, die beschriebene Weltordnung und ihre konkreten Vertreter im Märchen zu präsentieren und zu erklären:

Die obere Hälfte der x-Achse entspricht der eigenen Welt (die Welt des Helden).

Der untere Bereich der x-Achse entspricht der fremden Welt (die Welt des Antagonisten).

Rechts von der y-Achse befindet sich das Feld für richtiges Handeln. Das bedeutet:

* Vertreter der eigenen Welt handeln zugunsten ihrer eigenen Welt.
* Vertreter der fremden Welt handeln zugunsten ihrer eigenen Welt.
* Vertreter der neutralen Welt handeln zugunsten der Welt des Helden.

Links von der y-Achse befindet sich das Feld für falsches Handeln. Das bedeutet:

* Vertreter der eigenen Welt handeln zugunsten der fremden Welt.
* Vertreter der fremden Welt handeln zugunsten ihrer eigenen Welt.
* Vertreter der neutralen Welt handeln zugunsten der fremden Welt.

Die y-Achse zeigt die Entitäten, die an beiden Welten beteiligt sind.

Die x-Achse entspricht der Liste der Antipoden der Vertreter der Weltordnungsmitglieder im Märchen. Die Namen dieser Figuren sind jeweils mit zwei Großbuchstaben abgekürzt. Die Kleinbuchstaben ***r*** bzw. ***f*** vor den Abkürzungen weisen auf die Attribute *richtig* bzw. *falsch* des entsprechenden Protagonisten hin.

Die Initialen in Großbuchstaben sind wie folgt aufzulösen:

HH: Herr des Helden

HD: Held

RE: Rückeroberungsobjekt

HF: Helfer

ST: Stifter

VB: Verbindende Person

ZM: Zaubermittel

BZ: Besitzer des Zielobjektes

ZO: Zielobjekt

HP: Herr des potenziellen Partners

PP: Potenzieller Partner

AN: Antagonist

Die markierten Zellen in den Feldern und auf der x-Achse zeigen den Bezug zwischen den Mitglieder der im Märchen vorhandenen Weltordnung und den handlungstragenden Figuren sowie ihre Zugehörigkeit zu eigenen, fremden oder neutralen Welten.

|  |  |
| --- | --- |
|  | Vertreter der eigenen Welt |
|  | Vertreter der neutralen Welt |
|  | Vertreter der fremden Welt |

Im Folgenden betrachten wir diese Figuren anhand ihrer *richtigen* und *falschen* Pendants.

Anmerkung: Es gibt einige Vorbehalte für die gesamte Nomenklatur der handlungstragenden Figuren. Die Berücksichtigung dieser Konventionen ist essenziell notwendig, um sprachliche Unstimmigkeiten zu vermeiden. Handlungstragende Figuren (Personen, Wesen, Gegenstände) stehen überall im Singular und werden durch das grammatische Genus des Terminus im Gegensatz zum Genus und Numerus der Instanz bezeichnet. Zum Beispiel wird für die weibliche Aschenputtel der Terminus rHD (richtiger Held) verwendet; die beiden älteren Brüder werden als fHD (falscher Held) bezeichnet, d. h. im Singular und im Maskulin, usw.

**HH –** Der Herr des Helden ist das Oberhaupt in der Familie. Wenn es sich um eine königliche Familie handelt, ist der Herr der König. Dieser muss vom König unterschieden werden, der der Herr des potenziellen Partners ist und nicht mit dem König gleichzusetzen ist, dem der Held dient.

Wenn sich der Herr des Helden unverhältnismäßig gegenüber dem Helden verhält, wird er als falscher Herr des Helden (fHH) definiert. Beispiele hierfür sind: Ein Vater begehrt seine eigene Tochter; er liefert seinen eigenen Sohn dem Verderben aus; oder er lässt seine Kinder in der Wildnis zurück. In einem anderen Szenario beauftragt ein König den jüngeren Sohn aufgrund des bösen Rats seiner älteren Söhne mit einer lebensgefährlichen Aufgabe. In all diesen Fällen übernimmt der Familienvater die Rolle eines falschen Herrn des Helden (fHH) und gleicht somit dem Antagonisten (siehe unten).

**HD –** Der Held gilt als Maßstab für die Identifikation aller anderen Figuren im Märchen. In der Regel wird der Held als Nachfolger betrachtet, es sei denn, im Märchen gibt es keinen Herrn des Helden (HH). In umgekehrtem Fall, wenn im Märchen ein Protagonist dem Oberhaupt gleichgestellt ist, ist es notwendig, den Helden unter den anderen Protagonisten zu suchen. Des Weiteren kann es sowohl einen suchenden als auch einen vertriebenen Helden geben. Nicht selten agieren im Märchen zwei Helden. Als solche gelten: Geschwister (auch Halbbruder oder Halbschwester), ein Ebenbürtiger sowie das Kind eines Heldopfers.

Als falsche Helden (fHD) gelten üblicherweise die Geschwister des Helden, sofern sie nicht vor dem Beginn der Heldentätigkeit geboren wurden. Die Rolle des fHD kann auch vom Ebenbürtigen übernommen werden. Dieser beginnt gewöhnlich mit der Schließung der Brüderschaft und endet mit dem Verrat gegenüber dem Helden.

**PP –** Der potenzielle Partner repräsentiert denselben Stand wie der Held, stammt jedoch aus der gegenüberstehenden Welt. Gewöhnlich hat der Partner einen eigenen Herrn (HP). Die Beziehung zwischen dem potenziellen Partner (PP) und dem Helden (HD) beginnt immer aufgrund der Initiative einer der beiden Figuren, meistens durch die männliche Figur, selbst wenn die männliche Figur ein potenzieller Partner ist und nicht der Held, wie zum Beispiel im ATU 425. Daher ist es schwierig, eine weibliche potenzielle Partnerin von einem weiblichen Zielobjekt (ZO, siehe unten) zu unterscheiden.Letztere kann den Helden zwar heiraten, gilt jedoch nicht als potenzielle Partnerin für den Helden, sondern eher als Beute, die üblicherweise für den Auftraggeber, den falschen Herrn des Helden (fHH) oder den richtigen Antagonisten (rAN) bestimmt ist. Die Initiative, eine solche Figur zu gewinnen, geht vom rAN oder vom fHH aus. Es ist ebenfalls schwierig, einen potenziellen Partner aus dem Rückeroberungsobjekt (RE) zu unterscheiden. Ein Identifikationsmerkmal ist die Heiratstüchtigkeit des Objekts, es muss also nicht der Familie des Helden angehören. Zum Beispiel gilt eine befreite Frau im ATU 301 als potenzielle Partnerin, auch wenn sie zuvor geraubt wurde und der Held den Auftrag oder die Erkundigung des Königs befolgt, um die entführte Frau zurückzuholen. Dieses Objekt wird weiterhin als potenzielle Partnerin betrachtet, selbst wenn der Held sie aus verschiedenen Gründen, einschließlich der oben genannten, nicht heiraten möchte.

Im Fall eines falschen potenziellen Partners (fPP) handelt es sich in der Regel um die gleichgeschlechtlichen Geschwister (manchmal auch Halbbruder oder Halbschwester) des richtigen potenziellen Partners oder seine Ebenbürtigen.

**HP –** Der Herr des potenziellen Partners repräsentiert denselben Stand wie der Herr des Helden, stammt jedoch aus der fremden Welt. Seine Hauptfunktion besteht darin, die Hochzeitsaufgaben und allgemein die Voraussetzungen zu stellen, und wenn nötig, die Freier zu verfolgen. Der falsche Herr des potenziellen Partners (fHP) handelt freundlich gegenüber dem Helden.

**RE –** Das Rückeroberungsobjekt (RE) kann sowohl eine Person als auch ein Gegenstand, ein Tier oder ein Zaubermittel sein. Es handelt sich um das Objekt, das vor dem Beginn der Aktivitäten des Helden in seiner Welt vorhanden war und nun zurückgeholt werden muss. Ist das Objekt nach dem Beginn der Aktivitäten des Helden weggenommen, verschollen oder verschwunden, behält es den Status bei, den es vor der Krise hatte, und muss nicht als RE betrachtet werden. (Beispielsweise bleibt der Schlange-Jüngling im ATU 425 auch dann potenzieller Partner, wenn er verschwunden ist und von der Gattin (HD) gesucht wird.) Das Vorhandensein des Rückeroberungsbedarfs bereits vor Beginn der Aktivitäten oder der Geburt des Helden oder des Helfers ist wichtig, um das gesuchte Objekt als Rückeroberungsobjekt (RE) zu definieren. Das Rückeroberungsobjekt verhält sich in den meisten Fällen sehr aktiv. Der Held wird beauftragt, die entführte Mutter zurückzuholen. Sie hilft dem Helden dabei, den Aufenthaltsort der externen Seele des Antagonisten auszuspionieren. Das Rückeroberungsobjekt gehört zur eigenen Welt und kann mit verschiedenen Ständen vertreten sein: Nachfolger, beispielsweise die Geschwister, die vom Antagonisten bereits vor der Geburt des Helden festgenommen wurden (ATU 328). Oberhaupt, wenn es sich beispielsweise um die Rückeroberung der entführten Mutter handelt, die jedoch ebenfalls vor dem Beginn der Aktivitäten des Helden entführt sein muss. Jedoch nicht vor der Geburt des Helden, wie das im Typ ATU 650A - Der starke Hanns vorkommt. In einigen Fällen ist es schwer, das RE von der potenziellen Partnerin zu unterscheiden. Das Mädchen aus dem Typ ATU 301 kann sowohl Braut als auch geraubtes Objekt sein, das zurückgeholt werden muss. In solchen Fällen ist auf den Bezug zwischen dem betroffenen und anderen Protagonisten zu achten. Wenn ein Protagonist einen eigenen Herrn hat und sich in der fremden Welt befindet, handelt es sich um die potenzielle Partnerin.

Das RE kann dann falsch sein, wenn es sich gegen den Helden wendet. Wenn der Held seine, vom AN festgehaltenen Brüder befreit und diese ihn dann in den Brunnen werfen, sind die Brüder die Falschen Rückeroberungsobjekte (fRE) und nicht fHD (diese können sie sowieso nicht sein, wenn sie vor dem Beginn der Aktivitäten des Helden verschollen sind).

**ZO & BZ –** Das Zielobjekt und der Besitzer des Zielobjekts. Das Zielobjekt kann sowohl eine Person als auch ein Gegenstand, ein Tier oder ein Zaubermittel sein. Dieses Objekt gehört zu einer anderen Welt als der Held und wurde zuvor nicht erwähnt. Das Zielobjekt kann sich sowohl aktiv als auch passiv verhalten. Zum Beispiel wird der richtige Held (rHD) vom falschen Herrn des Helden (fHH) beauftragt, einen Löwen (ZO) zu holen. Der Löwe leistet Widerstand, wird jedoch besiegt und mitgebracht. Der rHD wird vom richtigen Herrn des Helden (rHH) beauftragt, ein Pferd (ZO) zu holen. Der Held stiehlt das Pferd, das zwar inaktiv ist, aber dem rHD später andere Aufgaben lösen hilft, indem es als Helfer fungiert.

In den meisten Fällen haben diese Objekte ihre Besitzer. Der beauftragte Held kommt oft bei der Eroberung des Objekts unmittelbar mit dem Besitzer in Kontakt. Es ist üblich, dass der Besitzer selbst als Zielobjekt angefordert wird, was anschließend zur Bestrafung des Antagonisten führt (typische Beispiele ATU 328, 531). Die weibliche Figur des Zielobjekts ist von der potenziellen Partnerin (PP) auch dadurch zu unterscheiden, dass sie keiner Herrschaft untersteht. Nach ihrem Stand soll die weibliche Figur des Zielobjekts eine Magin bzw. ein Gegenstand oder Besitz des Magers sein. Dies bedeutet jedoch nicht, dass alle Figuren im Märchen, die mit magischen Eigenschaften ausgestattet sind, der Kategorie Zielobjekt angehören. Ein Beispiel hierfür ist das Märchen ATU 313, in dem der Jüngling (rHD/rPP) beim Zauberer (rHP/rHH) gerät und sich in dessen Tochter (rPP/rHD) verliebt, mit der er später vor dem rHP/rHH flieht. Ein Zielobjekt kann auch eine abstrakte Aufgabe sein, wie Acker pflügen, Holz holen, Zoll erheben usw., die nicht als Protagonisten des Märchens betrachtet werden.

Ein falsches Zielobjekt liegt vor, wenn das Objekt zugunsten des Helden handelt. Beispielsweise wird nach dem Organ eines Tieres gesucht, und das Tier schenkt dem Helden seinen Jungen. Die Schöne folgt dem Helden, ohne dabei Widerstand zu leisten, usw. Dass diese Objekte später dem Helden helfen, ist jedoch üblich. Im Fall von Zaubergegenständen könnte es sich um ein falsches Objekt handeln, wenn es ausgetauscht wird, zum Beispiel das Lebenswasser.

**ZM –** Ein Zaubermittel ist nur dann als solches zu definieren, wenn es nicht als Zielobjekt fungiert. Es muss also nicht gezielt gesucht und erworben werden, sondern kann geschenkt, bestellt, gefunden oder entwendet werden. Der Zauberring, den der Held als Belohnung für seine guten Taten erhält, ist kein Zielobjekt, sondern ein Zaubermittel (ZM). Im Gegensatz dazu ist das Lebenswasser, das der Held im Auftrag seines Vaters beschafft, das Zielobjekt (ZO) und nicht das Zaubermittel, obwohl es tatsächlich magische Eigenschaften hat. Wenn ein echter Ring durch einen falschen ausgetauscht wird, handelt es sich um das falsche Zaubermittel (fZM). Gleiches gilt für den Fall des ausgetauschten Lebenswassers (fZO, siehe oben).

ZM kann auch als eine besondere Art von HF gelten. Das kann dann der Fall sein, wenn das ZM vom ST geschenkt wird. Das Geschenk vom ST ist gewöhnlich dann als ZM zu betrachten, wenn es nach der Aushändigung des Gegenstandes nicht die Interaktion zwischen dem ZM und dem HD entwickelt wird, sondern das ZM verschiedene Befehle erfüllt (Zauberring vs. Geist), sich in verschiedene Hindernisse umwandelt usw.

**ST & HF –** Der Stifter und der Helfer. Der Stifter kann sowohl ein Mensch als auch ein Tier, eine Pflanze oder ein mythisches Wesen sein. Der Stifter liefert dem Helden einen notwendigen Gegenstand, eine Fähigkeit oder Informationen, hat jedoch selbst keine direkte Wirkung auf den Helden. Beispielsweise kann der Stifter den Helden nicht beleben oder ihm Asyl bzw. Schutz gewähren. Dabei ist es wichtig, dass der Stifter dem Helden lediglich unterstützende und keinesfalls belastende Informationen liefert. Statt beispielsweise zu sagen, dass ein Ungeheuer eine Königstochter entführt hat, wäre es passender, dem Helden den Weg zum Entführer zu weisen oder ihm eine Waffe zu übergeben oder einen Trick zu zeigen, wie man das Ungeheuer besiegen kann.

Der Helfer (HF) ist jemand, der den Helden begleitet und gemeinsam mit ihm handelt. Das gemeinsame Wirken mit dem Helden bei der Lösung der Aufgabe oder Behebung der Notlage ist ein wichtiger Faktor, um den Helfer vom Stifter zu unterscheiden. Die Funktion des Stifters besteht darin, dem Helden zu helfen, aber so, dass er dabei nicht als Mitwirkender agiert. Manchmal können die Funktionen von Stifter und Helfer doch in einer Figur verflochten sein, wie zum Beispiel ein Tier-Schwager aus ATU 552. Er zeigt dem Helden zunächst den Weg zum Antagonisten und übernimmt somit die Funktion des Stifters. Dann jedoch, als der Held im Zweikampf mit dem Antagonisten getötet wird, erweckt ihn derselbe Schwager wieder zum Leben und erfüllt dadurch die Funktion des Helfers. Nach seinem Stand kann ein Helfer ein Mager oder ein Angehöriger des Magers sein. Stiften und helfen können fast alle Protagonisten im Märchen, einschließlich des Antagonisten. Der Stifter oder Helfer selbst kann jedoch nichts anderes tun als zu stiften und zu helfen. In der Welt der Mager und Totem gibt es keine Stände und Hierarchie: Eine junge Schlange wird gerettet, und der Vater der Schlange schenkt dem Retter einen Wunschstein. Eine andere Schlange wird gerettet, und sie beschenkt den Retter mit der Fähigkeit, die Tiersprache zu verstehen.

Der falsche Stifter (fST) bzw. falsche Helfer (fHF) ist jemand, der dem Bösen hilft oder ihm etwas stiftet, zum Beispiel die von einem falschen Helden beauftragte Hexe. Auch die Wesen, die beim bösen Zielobjekt die Wache halten (ATU 551), gelten als falsche Helfer.

**VB –** Die verbindende Person wird durch ein menschliches oder mythisches, jedoch eher anthropomorphes Wesen dargestellt. Ihre Hauptfunktion besteht darin, die Vermittlung zwischen dem Helden und einer anderen handlungstragenden Figur zu gewährleisten oder den Helden über die bestehende Krise bzw. notwendige Belastungen zu informieren. Typischerweise werden in der Rolle der verbindenden Person Untertanen eingesetzt, wie etwa eine alte Gastgeberin, ein Wirt oder ein Hirt. Oftmals ist auch der Herr des Helden, meistens die Mutter des Helden, in dieser Rolle anzutreffen.

Eine verbindende Person wird zur Falschen, wenn sie zu Gunsten des Antagonisten handelt. Häufig finden sich in dieser Rolle Diener, Berater usw., die beim bösen König beschäftigt sind. Die Handlung der falschen verbindenden Person ist nicht immer das Gegenteil dessen, was die Richtige tut. Als typisch gilt das Denunzieren und die Rolle eines Denunzianten oder die Initiierung böswilliger Aufträge.

**AN –** Antagonist (Gegenspieler, Schädling): Der Antagonist gehört zur bösen Welt (fremde Welt) und kann praktisch durch alle Stände vertreten werden. Wie oben angedeutet, kann seine Funktion auch von einem falschen Protagonisten, jedoch aus der eigenen Welt, übernommen werden. Besonders verbreitete Fälle sind der falsche Herr des Helden (fHH) und der falsche Held (fHD).

Ein falscher Antagonist, also eine Böse der zu Gunsten des Helden handelt ist nicht anzutreffen aber theoretisch denkbar. In einem Märchen können mehrere Antagonisten auftreten. Wenn der Drache getötet wird und seine Schwester versucht, Rache zu nehmen, ist auch die Schwester des Drachen ein Antagonist.

### **Handlungsabschnitte im Märchen und ihre Organisation**

Die Segmentierung des Märchens basiert auf zwei empirisch erkennbaren Entitäten: der Vollgeschichte oder Story und den Motiven, die durch den Auftritt handlungstragender Figuren erzeugt werden. Diese Motive fungieren als die kleinste Teilgeschichte innerhalb der Vollgeschichte.

Im Kontext der Segmentierung von Märchen ist die Bedeutung der zweiten Entität, nämlich der Motive, von entscheidender Bedeutung. Wenn wir berücksichtigen, dass volkstümliche Märchen keine detaillierten Porträts, Naturbeschreibungen oder Gefühlsschilderungen enthalten, wird die universelle Anwendbarkeit des Auftritts handlungstragender Figuren als Metrik für die Segmentierung von Märchen unbestreitbar.

Auftritt der handlungstragenden Figur im Märchen ist den im Drama explizit markierten Szenen gleichzustellen. Wenn wir Märchen als Dramen aufzeichnen würden, könnten wir den Text in segmentierte Teile aufteilen, die den Motiven entsprechen. Ein Beispiel dafür ist die segmentierte Zusammenfassung des Märchentyps *ATU 551 - Das Wasser des Lebens* (siehe unten).

Der Auftritt selbst kann im Text in verschiedenen Formen dargestellt werden, wie zum Beispiel:

1. Bericht über das, was ein Protagonist gerade erlebt oder worunter er leidet, zum Beispiel: Der König wird blind (König).
2. Bericht darüber, was gerade zwischen mehreren Protagonisten geschieht, zum Beispiel: Der König beauftragt seinen Sohn, nach seinem Tod sein Grab zu bewachen (König, jüngster Bruder, älteren Brüder).
3. Bericht über die Umgang von zwei Parteien mit dem Dritte, ohne dass diese daran aktiv beteiligt ist. Zum Beispiel: Der jüngste Bruder gibt dem Freier seine Schwester zur Frau (Freier, jüngster Bruder, die Schwester kommt hier gar nicht zum Vorschein).
4. Bericht darüber, wie zwei Figuren gegenseitig interagieren, zum Beispiel: Der Jüngling tötet im Zweikampf das Ungeheuer (Jüngling, Ungeheuer).
5. Ein besonderer Fall ist die Erzählung innerhalb der Erzählung, wie zum Beispiel in *ATU 449 -* *Sidi Numan*.

Die Liste lässt sich erweitern.

Die Motive bevor sie in die Vollgeschichte integriert werden, lassen sich im mittleren Gebilde organisieren. Dieses Gebilde ist traditionell unter dem Namen Episode bekannt.

Durch den Auftritt einer handlungstragenden Figur im Märchen kann eine Episode eingeleitet, fortgesetzt oder abgeschlossen werden. Jede Episode stellt Einheit von folgenden Handlungsphasen dar:

Krise ist ein Zustand, der eine Lösung braucht. Jedes Motiv bezieht sich auf eine Krise und stellt eine Phase ihres Verlaufs dar.

– Krise auslösen) – und – Krise nachvollziehen – und – (Krise bekämpfen – oder – Krise missachten) – und – (Krise lösen – oder

Jedes Motiv ist nach seinem **Wert**, **Sinn** und **Folge** zu bestimmen.

Nach dem **Wert** kann ein Motiv positiv, negativ oder ambivalent sein.

Die **Folge** des Motivs bestimmt seiner Zugehörigkeit zu bestimmter Episodenphase und somit die gegenseitige Spannung zwischen zwei Motive, die miteinander Konsequenz (kausal) verbunden sind: *Offenkundigkeit* vs. *Handlung* (mit zweierlei Ergebnissen: *Lösung* vs. *Auslösung der Krise*).

Was den **Sinn** des Motivs angeht, so kann ein Motiv entweder einen *Bonus*- oder *Lastenausgang* haben. In einigen Fällen kann ein Ausgang auch beide Sinne haben (darüber unten). Ein Motiv mit Bonus-Ausgang dient der Förderung der handelnden Figur (es stellt ihr Ausrüstung, Gegenstände, Informationen, Versprechen oder Helfer zur Verfügung, die ihr in der Not helfen können). Ein Motiv mit Lastenausgang fordert hingegen den Verbrauch der gesammelten Bonus-Ressourcen, um die Behebung der Krise näher zu bringen.

Der Etalon für die Bestimmung vom **Wert** eines konkreten Motivs ist allein der Faktor des Helden und nicht der handelnden Figur (gemeint wird die Handlung ohne Beteiligung des Helden). Die Beteiligung des Helden an einem Motiv ermöglicht es, den Wert des Motivs anhand seiner Auswirkungen eindeutig zu bestimmen, problematisch sind Motive ohne Teilnahme des Helden. In solchen Fällen hängt die Bewertung des Motivs ausschließlich von den Ergebnissen der Handlung ab. Hat die Handlung negative Ergebnisse für den Helden, dann wird das Motiv als negativ bewertet. Hat das Motiv kein direktes Ergebnis für den Helden, dann wird der Wert des Motivs nach seinem absoluten Wert beurteilt. Z. B. Scheitert ein falscher Held bei der Prüfung des potenziellen Stifters, handelt es sich um ein Bonus-Motiv mit negativem Wert (Grund für diese Bewertung dient die Perspektive des richtigen Helden: für ihn wäre dieselbe Ausgang einem Bonus-Motiv mit einer negativen Folge gleich).

Vor der **Offenkundigkeit** der **Krise** muss sie erst einmal ausgelöst worden sein. Die Auslösung der Krise ist das Resultat einer **Handlung**. Davon geht aus, dass eine Handlung, die eine vorhandene Krise löst, kann gleichzeitig eine neue Krise auslösen oder sie offenkundig machen.

Folgendes Beispiel: der vor der schweren Aufgabe stehende Held wendet sich den Stifter an, dem er einmal geholfen hat und bittet ihn um die Gegenleistung. Hiermit handelt es sich um die Offenkundigkeit der Krise für deren Lösung neben dem Helden auch der Stifter engagiert werden muss. Der Stifter händigt dem Helden die Handlungsanweisung aus, mit deren Hilfe er zur Lösung der vor ihm gestellte Aufgabe näherkommen kann. Z. B. Trick für die Erbeutung eines Reittieres. Der Held folgt der Handlungsanweisung des Stifters und ergreift das Objekt, das er später zu seinem Auftraggeber übergibt.

Schauen wir uns diese Handlungskette genauer an so wird es uns auffallen, dass das Ausrüsten des Helden mit Handlungsanweisungen nicht nur als Lösung der Krise (Behebung des Mangels an die Erbeutungsmethode) gilt, sondern auch als Offenkundigkeit einer bevorstehenden Krise. Für solche Krise ist die Erbeutung des Zielobjektes (hier eines Pferdes) zu halten.

Eine Krise existiert nicht als eigenständiges Motiv, sondern ist entweder die Folge einer Handlung oder die Offenkundigkeit eines solchen Handlungsergebnisses.

**Die Offenkundigkeit** der Krise ist die Situation, in der die Krise bekannt gemacht oder erkannt wird. Wie alles andere kann die Offenkundigkeit der Krise negativ, positiv oder ambivalent sein. Beispiele:

* Der Zwerg prüft die Höflichkeit der handelnden Figur, um ihr weiterzuhelfen. Hier liegt eine positive Absicht vor, die darauf abzielt, jemandem zu helfen.
* Ein Mädchen erkennt seinen Makel und wird neidisch auf die Schönheit der Schwester. Dies stellt eine negative Absicht dar, die darauf abzielt, jemanden zu verderben.
* Auf dem Schild bei einem Kreuzweg steht die Warnung, die ignoriert werden muss, um Erfolg erzielt zu werden. Also der Schild spielt die Funktion einer ambivalenten Offenkundigkeit, das, was richtig ist wird als falsch bezeichnet und umgekehrt.

Der Grund für die Krise und die offenkundig gemachte Krise können manchmal ganz verkehrt erscheinen. Ein Beispiel: Der König fühlt sich durch den Helden bedroht (Grund der Krise) und will ihn beseitigen. Deshalb erteilt er dem Helden eine unlösbare Aufgabe, bei deren Erfüllung er sterben soll. Die Offenkundigkeit der Krise lautet jedoch nicht so, wie die Krise heißt, etwa: „Du musst vom Löwen gefressen werden“, sondern so: „Du musst ein Löwenfell beschaffen“. Weil die Offenkundigkeit das Ziel hat, den Helden zu verderben, ist sie trotz ihres Inhalts negativ.

Der Kenntnisnahme der Krise wird durch eine entsprechende **Handlung** gefolgt. Eine positive Handlung schafft entweder einen Bonus für weitere Handlungen oder ermöglicht die Wiedergutmachung der Folgen einer negativen Handlung (Behebung der vorhandenen Krise). Eine negative Handlung lässt hingegen die vorhandene Krise bestehen oder löst eine neue Krise aus. Handlungen mit dilemmatischen Werten können beides tun, wie das Füttern des Vogels mit eigenem Fleisch oder die Tötung eigener Kinder zur Rettung des versteinerten Freundes.

Die Handlung kann verschoben werden. In solchen Fällen wird die Offenkundigkeit einer Krise durch die Offenkundigkeit einer nächsten Krise gefolgt. Es kann auch vorkommen, dass im Rahmen einer Krise geführte Handlungen durch eine neue Handlung aus dem Rahmen derselben Krise gefolgt werden. Gewöhnlich stellt eine solche Handlung einen **Mehrwert** der vorherigen Handlung dar, insbesondere in sogenannten Bonusepisoden. Obwohl der Mehrwert der Handlung als ein autonomes Motiv etabliert ist, ist er nach seinem Wert und Sinn von dem vorherigen Motiv nicht zu unterscheiden (daher auch sein Name – Mehrwert). Anders gesagt, das Resultat dessen, was und wie gehandelt wurde, ist vollkommen von der eigentlichen Handlung abhängig oder sogar in sie inkludiert. Ein Beispiel: Der ältere Bruder verhält sich frech gegenüber dem Zwerg und verpasst den Bonus von ihm. Das Verhalten des falschen Helden kann hier sowohl die Handlung als auch deren Ergebnis wahrgenommen werden. In einer anderen Variante wird dem falschen Helden nicht nur die potenzielle Hilfe untersagt, sondern er kann gelähmt oder anderweitig verhindert werden (hier gilt die Lähmung als Mehrwert der negativen Handlung).

Die Ergebnisse der Segmentierung des Märchens in Handlungsabschnitte lassen sich wie folgt zusammenfassen: Die kleinste empirisch erkennbare Entität für die strukturelle Segmentierung des Märchens entspricht dem Auftritt der handlungstragenden Person(en), d.h., der Szene. Der Wechsel zwischen den Szenen in der Erzählung entspricht dem Wechsel zwischen den Segmenten des Märchens. Die kleinste Kategorie dieser Segmente sind Motive, Motive bilden Episoden, die wiederum für die Bildung der Vollgeschichte zuständig sind. Ein Motiv ist entweder eine Offenkundigkeit oder der Handlung bzw. den Handlungen der in einer Episode thematisierten Krise gleich und beginnt und endet im Laufe eines Auftritts bzw. einer Szene.

Ein Auftritt (eine Szene) besteht wenigstens aus einem Motiv und kann die Motive von mehr als einer Episode enthalten. Eine Episode setzt sich aus den Motiven zusammen, die Anfangs- und Lösungsphasen der in ihr thematisierten Krise widerspiegeln; das sind die Motive der Kategorien 'Offenkundigkeit' und 'Handlung'.

Motive einer Episode können durch die Motive aus anderen Episoden getrennt werden, daraufhin kann eine Episode entstehen, die praktisch die ganze Geschichte umfasst. Die schematisierte Darstellung einer solchen Konstellation könnte wie folgt aussehen:

[Geschichte [Episode1 [Motiv1.1 + Motiv1.2]] + [Episode2 [Motiv2.1 + Motiv2.2 + Motiv2.3]] + [Episode1[Motiv1.3]]]

Wobei jedes Pluszeichen optional, und jede durch das Pluszeichen getrennte eckige Klammer bindend auf die Grenze zwischen zwei verschiedenen Szenen hinweist.

## **Markup**

### **Tupel *m***

Das Markup wird uns helfen, Märchen so zu annotieren, dass alle ihre inhaltlichen Eigenschaften allein anhand dieser Annotationen vergleichbar sind. Die Entwicklung eines solchen Markups ist aufgrund der maximalen Verdichtung der inhaltlich-strukturellen Informationen möglich. Dabei dürfen die Anforderungen an die höhere Ausdruckspotenzial, Anwendungsfreundlichkeit sowie die Kompaktheit des Markups nicht aus den Augen verloren werden.

Zwecks der restlosen Erschließung der von uns oben erfassten Eigenschaften des Motivs (kurz ***m***) versuchen wir, es als ein 4-Tupel von folgenden Attributen darzustellen:

Attribut ***a*** erfasst die Zugehörigkeit des Motivs zum Typ der Geschichte (ATU Index), ***b*** erfasst die Zugehörigkeit des Motivs zu einer Folge (Handlung oder Offenkundigkeit) sowie zum dessen Wert (negativ, positiv, ambivalent), ***c*** fasst den Inhalt des Motivs in einem Wortlaut zusammen, und ***d*** erfasst die Nomenklatur der im Motiv beteiligten Figuren.

Hier geben wir die Beschreibung der Variationen jedes dieser Attribute im Einzelnen:

#### **Attribut a**

Jedes Motiv ist Teil einer Episode. Episoden bilden die Vollgeschichte. Also sind jede Episode und damit auch jedes Motiv Bestandteile eines Geschichtentyps. Diese sind, wie bekannt, in einem internationalen Typenkatalog erfasst und entsprechend indiziert. Wir bezeichnen diese Eigenschaft des Motivs mit dem Kleinbuchstaben ***a*** (für ATU) und der darauffolgenden Typennummer, zum Beispie,l

*a300* für den Typ *ATU 300 – Drachentöter*.

Dieses Mal beschränken wir die Menge ***a*** auf die Indizes der Zaubermärchen,

#### **Attribut b**

Markierung der Zugehörigkeit des Motivs zu dem bestimmten Folgen ist durch entsprechende konventionellen Zeichen Möglich:

|  |  |
| --- | --- |
| Zeichenname | Die Folge (Art des Motivs) |
| *Ef* | Offenkundigkeit der Krise |
| *Ha* | Handlung in der Krise |

Die Variationen nach dem unterschiedlichen Wert jeder von diesen Folgen lassen sich wie folgt unterscheiden:

|  |  |
| --- | --- |
| Zeichen | Der Wert der Folgen (Art des Motivs) |
| F (großes *Ef*) | Offenkundigkeit der Krise mit positivem Wert |
| f (kleines *Ef*) | Offenkundigkeit der Krise mit negativem Wert |
| Ff (beide *Ef*) | Offenkundigkeit der Krise mit ambivalentem Wert |
| H (großes *Ha*) | Handlung in der Krise mit positivem Wert |
| h (kleines *Ha*) | Handlung in der Krise mit negativem Wert |
| Hh (beide *Ha*) | Handlung in der Krise mit ambivalentem Wert |

Zusammengefasst sieht die Menge der Variationen des Elements *b* wie folgt aus:

Wir verzichten auf die extra Markierung der Motive mit Doppelfunktionen wie ***Ha*** *und* ***Ef*** und bevorzugen sie nach ihre primäre Funktionen ***Ha*** *oder* ***Ef*** bezeichnet zu werden.

#### **Attribut c**

Inhalt der Aktion stellt einen Wortlaut dar und zählt somit zum informativsten Bestandteil des Markupelements. Dieser Wortlaut entspricht dem inhaltlichen Wert des im Rahmen einer Episodenphase stattgefundener Aktion, z.B. "Heilmittel\_beschaffen".

Im Gegensatz zu anderen Attributen des Tupels, deren Variationen bereits vorbestimmt worden sind, ist eine endgültige Fertigstellung der Wortlautliste, also Variationen für das Attribut ***c****,* eher undenkbar. Es kann immer Variationen geben, die neue Formulierung fordern. Bereits aus diesem Grund zählt das Konzipieren eines Modells für die vorschriftliche Herstellung der Wortlautwerte zu wichtigste Aufgabe bei dem Bau des Markups. Ohne Lösung dieser Aufgabe wäre standardmäßiger Einsatz des Markups zum Scheitern verurteilt.

Dank der Nutzung der Tupel-Konstruktion des Markupelements ist es möglich, einen und denselben Wortlaut in bestimmten Fällen mehr als einmal zu benutzen. Diese Kategorie von Wortlauten nennen wir die geklonte Variante.

Klonen oder unveränderte Anwendung eines Wortlautes ist gewöhnlich möglich beim Wechsel zwischen der Offenkundigkeit als Beauftragung und dem Resultat der danach geleisteten Handlung (falsch oder richtig). Die Variation des Wortlauts ist dabei durch den Wechsel von ***b*** Attribut zu erzielen. Die Anwendung durch die ***d*** Attributwerten ist außerdem möglich der Unterschied zwischen den durch den einen und denselben Wortlaut erfasste Inhalte weiter zu kontextualisieren.

Folgendes Beispiel:

*F:Heilmittel\_beschaffen:rHD\_rHH*

Das Element erfasst den vom Herrn des Helden zum Helden mit guter Absicht erteilten Auftrag (***F***), ihm ein dringend benötigtes Heilmittel zu beschaffen.

Durch den Wechsel ***F*** durch ***f*** und ***rHH*** durch ***rAN***, hätte derselben Wortlaut einen wesentlich anderen Inhalt zusammenfassen können. Das wäre nämlich der vom Antagonisten mit einer schlechten Absicht erteilte Auftrag, der Held bei der Erfüllung der unlösbaren Aufgabe, hier bei der Beschaffung des Heilmittels, verderben lassen. Das entsprechende Markupelement sieht wie folgt aus:

*f:Heilmittel\_beschaffen:rHD\_rAN*

Durch den weiteren Wechsel der Werten vom Attribut ***b*** ist außerdem möglich der Mehrwert der positiv, negativ oder ambivalent ausgegangenen Handlung zu bezeichnen. Das folgende Markupelement kann z. B. wie folgt aufgelöst werden: Der Held endet seine Handlung zur Suche des Heilmittels mit positiven Folgen und händigt seinem Beauftragten das benötige Zielobjekt aus:

*H:Heilmittel\_beschaffen:rHD\_rHH\_rZO*

Kämme ein Zwischenfall vor, bei dem das richtige Lebenswasser vom falschen Helden (hier von neidischen Brüdern) durch das einfache ausgewechselt wird, so könnte auch diesen Fall durch denselben Wortlaut, mit gewechselten Werten von ***b*** und ***d*** Attributen erfasst werden:

*Hh:Heilmittel\_beschaffen:rHD\_rHH\_fZO*

Ein anderer Einsatz der Klon-Methode ist für die Zusammenfassung des ersten Treffens zwischen dem Helden und seinem Stifter oder dem künftigen Weggefährten möglich. Das Letzte kann sowohl ein Helfer als auch ein falscher Held sein. Bei solchen Fällen handelt es sich gewöhnlich um die Probestellung der handelnden bzw. reisenden Figur und deren Reaktion. Die älteren Brüder antworten frech auf die Frage des ihnen auf dem Weg begegneten Männleins und ernten Misserfolg bei der Erfüllung der Aufgabe. Der jüngste Bruder benimmt sich höflich gegenüber dem Männlein und sichert somit seine Hilfe, die in unterschiedlicher Form auftreten kann. Die Begegnung mit dem Zwerg und deren Vorgang kann man z. B. mit Hilfe folgendes Wortlauts erfassen *Höfliches\_Verhanten* Präfixe ***Ef*** und ***Ha*** hätten dabei auf die Phase des Treffens hinweisen können: *F:Höffliches\_Verhalten* als [verschleierte] Offenkundigkeit der Voraussetzung für die potenzielle Hilfe und *H:Höffliches\_Verhalten* als positive Reaktion des Helden auf diese Voraussetzung oder auch *h:Höffliches\_Verhalten* als negative, unhöfliche Reaktion des falschen Helden auf derselben Probe.

Nicht alle, miteinander konsekutiv verbundene Phasen einer und derselben Handlung, lassen sich durch einen Wortlaut, also durch den Klon erfassen. Um den Zusammenhang zwischen diesen Handlungen bzw. Handlungsphasen auch in diesen Fällen trotzdem beibehalten zu können wird eine **Ableitungsmethode** verwendet. Dabei übernimmt ein charakterisierendes Stichwort die Funktion einer Achse, wodurch mehrere, inhaltlich nacheinander folgende Wortlaute trotz ihrer Unterschiede in Verbindung bleiben, z. B.

h:Schweigepflicht\_erhängen > H:Schweigenpflicht\_beheben

Am häufigsten ist diese Methode bei der Zusammenfassung solcher Handlungsphase einzusetzen wie die Auslösung und der Lösung der Krise, oder die Offenkundigkeit einer vorbestimmten Krise und deren Eintritt, z. B.

*h:****Kontakt****\_abbrechen* > *F:****Kontakt****\_wiederherstellen* > *H:* ***Kontakt****\_wiederherstellen*

*h:Auslösung\_von\_****Krankheit*** > *F:****Krankheit****\_Beheben* > *H:****Krankheit****\_Beheben*

Als nächste Methode zum kompakten Aufbau des Markups gilt die **Serialisierung** einige Wortlaute. Dabei ist entscheidend, dass die Handlung zur gleichen Handlungsklasse gehört. Am weitesten sind folgende Handlungsklassen verbreitet:

* Tabu – Wortlaute mit dem Stichwort Tabu bilden eine Liste der Offenkundigkeiten von einzelnen Verboten z. B. *„f:Tabu Tür nicht öffnen“*. Sie alle haben eine einzige Fortsetzung, die lautet: *„h:Tabubruch folgt“*
* Handlungsanweisung – Die zur Hilfe engagierte Figur teilt dem Helden eine Anweisung mit, wie er sich in einer Situation verhalten muss. Am häufigsten geht es dabei darum wie der Held etwas beheben, ergreifen oder erhalten kann.[[4]](#footnote-6)
* Erklärungsbedürftiges Verhalten – Das Stichwort gilt für die Klon Wortlaute, die fragwürdige Verhalten von Figuren beschreiben. Das sind gewöhnlich die Figuren die dem reisenden Helden begegnen und ihn auf die Suche nach der entsprechende Erklärung anregen, z. B. *F:Erklärungsbedürftiges\_Verhalten\_von\_Betrügern* & *H:Erklärungsbedürftiges\_Verhalten\_von\_* *Betrügern*.
* Vortäuschung – Wie das Tabu ist die Vortäuschung im Märchen nur insofern geführt, dass sie unbedingt nachgefolgt werden muss. Also einerseits haben wir die Liste von verschiedene Vortäuschungen als Offenkundigkeiten und eine einzige Wortlaut der als Folge für alle diese Vortäuschungsfälle gilt. Diese lautet wie folgt: *H:Vortäuschung\_nachgehen.*
* Missetat - Das Stichwort bildet die Wortlaute für Erschließung von verschiedene Verratsfälle. Es kann vier Kategorie von Wortlaute verbinden: (1) Missetat in der Planungsphase, (2) Begehen der Tat, (3) Handlungsanweisung wie diese Tat beheben werden kann und (4) die Behebung der Tat. Inhalt der Missetat wird dabei nur bei der 2. Kategorie wiedergegeben, alle andere weisen dann auf den Lauf der Handlung in ihre verschiedenen Phasen. Folgendes Beispiel:
* beheben – gilt als gemeinsames Stichwort für die Lösung der belastenden Situation (der Krise). Folgendes Beispiel:
* erhalten – gilt als gemeinsames Stichwort für die Mehrwert der zum Bonusbescherten Ausgang führende Interaktion zwischen dem Helden und dem Stifter. Folgendes Beispiel:
* ergreifen – fasst die Handlung zusammen, bei der das Zielobjekt oder sein Besitzer erstmal erfasst werden. Folgendes Beispiel:
* beschaffen – wird als gleiches Stichwort verwendet, die Mitteilung und Erfüllung jeweils eines Auftrages erfassen (jedoch nicht das erstmalige Ergreifen des Zielobjektes oder seines Besitzers). Folgendes Beispiel:
* nachgehen – gilt als generalisierte Form solcher Verhalte wie: Akzeptieren oder Ablehnen eines Angebotes, sowie eine richtige oder falsche Reaktion auf die Gefahr, Verführung, Provokation oder Vortäuschung. Folgendes Beispiel:

Die Wortlaute mit den Stichworten *beheben*, *beschaffen*, *erhalten*, *ergreifen*, *nachgehen* können bereits geklonte oder abgeleitete Variante sein.

Ein weiteres, verhältnismäßig wichtiges Teil des Markups besteht aus sog. Monowortlaute. Sie bezeichnen gewöhnlich Ergebnisse oder Zwischenergebnisse der Handlungen, die durch die geklonten oder abgeleiteten Wortlaute erfasst sind. oder den sog. Mehrwert von Aktionen darstellen, die im Laufe der Geschichte keine Ausklang bekommen Z. B.:

*H:Geboren\_werden*

*H:Eigene\_Lebensgeschichte\_Erzählen*

*H:Heiraten,*

*H:Bestrafung,*

*h:Sterben\_eines\_Familienmitgliedes*

*h:Ungeheurlichkeit usw.*

Die Methoden wie Klonen und Ableitung der Wortlaute fördern die einheitliche und benutzerfreundliche Anwendung des daraus entwickelten Markups und gewähren eine hohe Qualität der damit annotierten Daten. Die einheitliche Anwendung der Markups wird außerdem durch den Wortlautkatalog unterstützt, der dem User ermöglicht die in der Wortlautliste Erfasste Werte gegenseitig vergleichbar und leicht auffindbar zu machen.[[5]](#footnote-7)

#### **Attribut d**

Wie bereits bekannt, kann es im Märchen-Genre insgesamt 24 handlungstragende Figuren geben. In einem Auftritt können gewöhnlich eins bis drei solchen Figuren agieren. Agieren bedeutet, dass die Figur in der Erzählung auftritt bzw. erscheint und nicht das, dass eine solche Figur von einer in der Episode agierende Figur erwähnt, gemeint oder bestellt wird).

Unter der Berücksichtigung der Höhenzahl der Kombination zwischen den Figuren, haben wir darauf verzichten müssen, die Reihenfolge der Figuren nach einer syntaktischen Funktion, wie z.B. das Subjekt / direktes Objekt / indirekte Objekt zu erfassen. Stattdessen haben wir überall die Interaktionen mit unterschiedlich kombinierten und / aber gleichen Figuren auf jeweils eine Kombination reduziert.

(ab|ba) = (ab)

(abc|acb|cab|aba|bac|bca) = (abc)

Dabei sind die handlungstragenden Figuren nach einer festen Reihenfolge (s. die Tabelle) kombiniert. D.h.: für die zweigliedrige Kombination wird wie folgt verfahren:

* Als erstens wird der rHD mit allen anderen kombiniert,
* Zweitens wird der rHH ebenfalls mit allen anderen kombiniert, ausgenommen der bereits eingesetzter rHD,
* Drittens wird das rRE mit allen anderen kombiniert, ausgenommen rHD und rHH usw.

Somit setzt sich die Menge der Variationen von ***d*** Elemente aus den Untermengen von Einzeln, zu zweit oder zu dritt agierenden Figuren zusammen:[[6]](#footnote-8)

{

*{rHD, rHH, rRE, rHF, rST rVB, rZM, rBZ, rZO, rHP, rPP, rAN, fAN, fPP, fHP, fZO, fBZ ... }*

*{rHD\_rHH, rHD\_rRE, rHD\_rHF, rHD\_ST, rHD\_rVB, rHD\_rZM, rHD\_rZO, rHD\_HP, rHD\_rPP ... }*

*{rHD\_rHH\_rRE, rHD\_rHH\_rHF, rHD\_rHH\_rST, rHD\_rHH\_rVB, rHD\_rHH\_rZM, rHD\_rHH\_rBZ ... }*

}[[7]](#footnote-9)

### **Musterbeispiel**

Im Folgenden versuchen wir, den Zusammenhang zwischen dem empirischen Objekt und dem darauf basierenden Modell anhand eines Beispiels zu veranschaulichen. Dazu werden wir einen klassischen Märchentext nach seinen Szenen segmentieren. [[8]](#footnote-10) Jedes Segment wird zusammengefasst und mit dem entsprechenden Markupelement versehen.

|  |  |  |  |
| --- | --- | --- | --- |
| 1. | Auftritt: |  | Der König |
|  | Inhalt: |  | Ein kranker König liegt im Sterben. |
|  | *Annotation:* | 1 | *a551h:Krankheit:rHH* |
| 2. | Auftritt: |  | Die drei Söhne und der alte Mann |
|  | Inhalt: |  | Ein weiser alter Mann weist die verzweifelten Königssöhne auf das schwer zu beschaffende Lebenswasser hin. |
|  | *Annotation:* | 2 | *a551:* *F:Heilmittel\_beschaffen:rHD\_rVB\_fHD* |
| 3. | Auftritt: |  | Der Vater und der älteste Sohn |
|  | Inhalt: |  | Der älteste Sohn bittet den Vater um Erlaubnis und macht sich auf den Weg, das Lebenswasser zu finden. |
|  | *Annotation:* | 3 | *a551:F: Werben\_um\_den\_Auftrag:rHH\_fHD* |
| 4. | Auftritt: |  | Der älteste Sohn und der Zwerg |
|  | Inhalt: |  | Der ältere Sohn zeigt Arroganz gegenüber dem Zwerg. Dieser wird zornig auf den ältesten Sohn. |
|  | *Annotation:* | 4  5 | *a551:F: Höfliches\_Verhalten:rST\_fHD*  *a551:h: Höfliches\_Verhalten:rST\_fHD* |
| 5. | Auftritt: |  | Der älteste Sohn und der Zwerg |
|  | Inhalt: |  | Der älteste Sohn wird in einer Felsspalte eingesperrt. |
|  | *Annotation:* | 6 | *a551:h: Festnahme\_bzw\_Sperren:rST\_fHD* |
| 6. | Auftritt: |  | Der Vater und der mittlere Sohn |
|  | Inhalt: |  | Der mittlere Sohn bittet den Vater um Erlaubnis und begibt sich auf die Suche nach Lebenswasser. |
|  | *Annotation:* | 7 | *a551:F:Werben\_um\_den\_Auftrag:rHH\_fHD* |
| 7. | Auftritt: |  | Der mittlere Sohn und der Zwerg |
|  | Inhalt: |  | Der mittlere Sohn zeigt ebenfalls Arroganz gegenüber dem Zwerg. Dieser wird wütend auf den mittleren Sohn. |
|  | *Annotation:* | 8  9 | *a551:F: Höfliches\_Verhalten:rST\_fHD*  *a551:h: Höfliches\_Verhalten:rST\_fHD* |
| 8. | Auftritt: |  | Der mittlere Sohn und der Zwerg |
|  | Inhalt: |  | Auch der mittlere Sohn wird in einer Felsspalte eingesperrt. |
|  | *Annotation:* | 10 | *a551:h: Festnahme\_bzw\_Sperren:rST\_fHD* |
| 9. | Auftritt: |  | Der Vater und der jüngste Sohn |
|  | Inhalt: |  | Der jüngste Sohn bittet den Vater um Erlaubnis und macht sich auf den Weg, Lebenswasser zu suchen. |
|  | *Annotation:* | 11 | *a551:F:Werben\_um\_den\_Auftrag:rHH\_fHD* |
| 10. | Auftritt: |  | Der jüngste Sohn und der Zwerg |
|  | Inhalt: |  | Der jüngste Sohn verhält sich höflich gegenüber dem Zwerg. Dieser gibt dem jüngsten Sohn notwendige Anweisungen. |
|  | *Annotation:* | 12  13  14  15 | *a551:F: Höfliches\_Verhalten:rHD\_rST*  *a551:H:Höfliches\_Verhalten: rHD\_rST*  *a551:H:* *Handlungsanweisungen\_Wache\_besänftigen:rHD\_rST*  *a551:h:Kontakt\_abbrechen:rHD\_rST* |
| 11. | Auftritt: |  | Der jüngste Sohn und die Wache |
|  | Inhalt: |  | Der jüngste Sohn besänftigt die Wache, die ihn passieren lässt. |
|  | *Annotation:* | 16  17 | *a551:H:Wache\_besänftigen:rHD\_fHF*  *a551:H:Wache\_entkommen:rHD\_fHF* |
| 12. | Auftritt: |  | Der jüngste Sohn und die verwunschenen Wesen |
|  | Inhalt: |  | Der jüngste Sohn nimmt die Zaubergegenstände. |
|  | *Annotation:* | 18 | *a551:H:Zaubermittel\_erhalten:rHD\_rZM* |
| 13. | Auftritt: |  | Der jüngste Sohn und die Besitzerin des Lebenswassers |
|  | Inhalt: |  | Es entwickelt sich eine Liebesbeziehung zwischen dem jüngsten Sohn und der Besitzerin des Lebenswassers. |
|  | *Annotation:* | 19  20 | *a551:F:* *Handlungsanweisung\_Objekt\_ergreifen:rHD\_fBZ*  *a551:h:Kontakr\_abbrechen: rHD\_fBZ* |
| 14. | Auftritt: |  | Der jüngste Sohn und das Lebenswasser |
|  | Inhalt: |  | Der jüngste Sohn erlangt das Lebenswasser. |
|  | *Annotation:* | 21 | *a551:H:* *Objekt\_ergreifen:rHD\_rZO* |
| 15. | Auftritt: |  | Der jüngste Sohn und der Zwerg |
|  | Inhalt: |  | Der jüngste Sohn bittet den Zwerg um die Freilassung seiner Brüder, ignoriert dabei aber dessen Warnung. |
|  | *Annotation:* | 22  23  24 | *a551:F:* *Kontakt\_wiederherstellen:rHD\_rST*  *a551:F:* *Verrat\_vermeiden:rHD\_rST*  *a551:h:* *Verrat\_vermeiden\_missachten:rHD\_rST* |
| 16. | Auftritt: |  | Der jüngste Sohn und seine Brüder |
|  | Inhalt: |  | Der jüngste Sohn trifft seine befreiten Brüder und teilt ihnen von seinem Erfolg mit. |
|  | *Annotation:* | 25 | *a551:Hh:Festnahme\_bzw\_Sperre\_beheben:rHD\_fHD* |
| 17. | Auftritt: |  | Die Könige aus fernen Ländern und der jüngste Sohn |
|  | Inhalt: |  | Auf dem Rückweg nach Hause unterstützt der jüngste Bruder die Herrscher von drei unterschiedlichen Ländern mithilfe seiner magischen Gegenstände. |
|  | *Annotation:* | 26  27  28 | *a551:F:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*  *a551:H:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*  *a551:h:Kontakt\_abbrechen:rHD\_rHF* |
| 18. | Auftritt: |  | Die Könige aus fernen Ländern und der jüngste Sohn |
|  | Inhalt: |  | Auf dem Rückweg nach Hause unterstützt der jüngste Bruder die Herrscher von drei unterschiedlichen Ländern mithilfe seiner magischen Gegenstände. |
|  | *Annotation:* | 29  30  31 | *a551:F:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*  *a551:H:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*  *a551:h:Kontakt\_abbrechen:rHD\_rHF* |
| 19. | Auftritt: |  | Die älteren Söhne beim Schlaf ihres jüngsten Bruders |
|  | Inhalt: |  | Die älteren Brüder werden neidisch und tauschen heimlich das Lebenswasser gegen gewöhnliches Wasser aus. |
|  | *Annotation:* | 32  33 | *A551:f:* *Missetat\_planen:rZM\_fHD*  *a551:h:* *Missetat\_Aneignen:rZM\_fHD* |
| 20. | Auftritt: |  | Der jüngste Sohn und ihr Vater |
|  | Inhalt: |  | Das beschaffte Heilmittel macht den König noch kränker. |
|  | *Annotation:* | 34  35 | *a551:* *h:Heilmittel\_beschaffen:* *:rHD\_rHH\_fZO*  *a551:h:Krankheit\_beheben:rHD\_rHH\_fZO* |
| 21. | Auftritt: |  | Die älteren Söhne der Vater und das Heilmittel |
|  | Inhalt: |  | Die älteren Söhne heilen den Vater und verleumden den jüngsten Bruder. |
|  | *Annotation:* | 36  37  38 | *a551:* *h:Heilmittel\_beschaffen:* *:* *rHH\_rZO\_fHD*  *a551:Hh:Krankheit\_beheben:* *rHH\_rZO\_fHD*  *a551:h:Verleumdung\_von\_Untreue:* *rHH\_rZO\_fHD* |
| 22. | Auftritt: |  | Die älteren und der jüngste Bruder |
|  | Inhalt: |  | Die älteren Brüder bringen den jüngsten Bruder zum Schweigen. |
|  | *Annotation:* | 39 | *a551:h: Zum\_Schweigen\_bringen:rHD\_fHD* |
| 23. | Auftritt: |  | Der König und der Scharfrichter |
|  | Inhalt: |  | Der König befiehlt den Tod seines jüngsten Sohns. |
|  | *Annotation:* | 40 | *a551:f:Todesauftrag:rVB\_fHH* |
| 24. | Auftritt: |  | Der ältere Sohn und der Scharfrichter |
|  | Inhalt: |  | Der Scharfrichter weigert sich, den Prinzen zu töten, und dieser flieht. |
|  | *Annotation:* | 41  42 | *A551:H:* *Auslösen\_von\_Mitleidsgefühl:rHD\_rVB*  *a551:H:* *Todesauftrag\_entkommen:rHD\_rVB* |
| 25. | Auftritt: |  | Der Vater, die drei fremden Könige |
|  | Inhalt: |  | Die drei Könige, denen der Prinz einst half, besuchen das Königreich. Der König erfährt die Wahrheit und vergibt seinem jüngsten Sohn. |
|  | *Annotation:* | 43  44 | *a551:H:Kontakt\_wiederherstellen:rHH\_rHF*  *a551:H:* *Missetat\_enthüllen: rHH\_ rHF* |
| 26. | Auftritt: |  | Der König und der Scharfrichter |
|  | Inhalt: |  | Der König erfährt vom Scharfrichter, dass sein jüngster Sohn freigelassen wurde, und erkennt die Verleumdung. |
|  | *Annotation:* | 45 | *a551: H:Verleumdung\_beheben:rHH\_rVB* |
| 27. | Auftritt: |  | Die Besitzerin des Zaubermittels |
|  | Inhalt: |  | Gleichzeitig bereitet sich die Besitzerin des Lebenswassers auf die Suche nach ihrem Geliebten vor. |
|  | *Annotation:* | 46 | *a551:F:Kontakt\_wiederherstellen:fBZ* |
| 28. | Auftritt: |  | Der älteste Sohn und der Diener der Besitzerin des Lebenswassers |
|  | Inhalt: |  | Der älteste Sohn scheitert ebenfalls beim Identitätstest. |
|  | *Annotation:* | 47 | *a551:h:Identität\_testen:fHD* |
| 29. | Auftritt: |  | Der mittlere Sohn und der Diener der Besitzerin des Lebenswassers |
|  | Inhalt: |  | Der mittlere Sohn scheitert ebenfalls beim Identitätstest. |
|  | *Annotation:* | 48 | *a551:h:Identität\_testen:fHD* |
| 30. | Auftritt: |  | Der jüngste Sohn und die Besitzerin des Lebenswassers |
|  | Inhalt: |  | Der jüngste Sohn besteht den Test und wird erkannt. |
|  | *Annotation:* | 49  50 | *a551:H:Identität\_testen:rHD\_fBZ*  *a551:H:Kontakt\_wiederherstellen:rHD\_fBZ* |
| 31. | Auftritt: |  | Vater und der älteste Sohn |
|  | Inhalt: |  | Der älteste Sohn gesteht dem König die Wahrheit. Die Verbrecher fliehen und kehren nie zurück. |
|  | *Annotation:* | 51  52 | *a551:H: Schweigenpflicht\_beheben:rHD\_rHH\_fHD*  *a551:H:Bestrafung:rHD\_rHH\_fHD* |

*.*

### **Motiv, Episode, Märchen**

Die Organisation der inhaltlichen Elemente im Märchen sind auf zwei Strukturebene empirisch zu beobachten: das sind die Motivebene und der Ebene der Vollgeschichte. Das erste lässt sich durch den Anfang und Ende der Szenen in der Geschichte erkennen, das andere durch den Anfang und Ende des gesamten Textes.

Die Vollgeschichte als oberste Instanz der Analyse entspricht einer Makrostruktur, die aus den in den mesoskopischen Strukturelementen organisierten Mikroelementen besteht. Bei der Annahme dieser Definition für die Vollgeschichte ist das Motiv hingegen als kleinstes und somit mikrostrukturelles Element des Systems zu definieren. Was zwischen der Vollgeschichte und dem Motiv vorkommt, gehört zum mesoskopischen Strukturelement. Als solches sind die Episoden zu zählen.

Die mesoskopische Natur der Episode beinhaltet automatisch die Repräsentationsfähigkeit der Charakter von Mikro- und Makroelemente. Daher ergibt sich die Möglichkeit, dass eine Episode sowohl ein Baustein als auch ein Gebilde sein kann, das selbst aus solchen Bausteinen besteht.

Anhand unseres Musterbeispiels versuchen wir den Zusammenhang zwischen den Strukturelementen ***Motiv > Episode > Märchen*** zu präsentieren. Dafür werden wir die Annotation des Textes als Motivliste darstellen (siehe die Werte Annotation in der Tabelle 2) und die Zusammenhänge der Motive in Episoden erfassen.

In einer Episode wird die Geschichte einer großen oder kleinen Krise erzählt. Die Episode folgt einer Struktur, die die Auslösung / Offenkundigkeit, Handlung mit der Behebung der vorhandene oder mit der Auslösung einer neuen Krise behandelt.

Wie bereits bekannt, wird die Behebung der Krise durch das Motiv mit den Attributwerten ***b=H/Hh*** dargestellt. Die Auslösung oder Offenkundigkeit der Krise wird hingegen durch ein Motiv erfasst, das dem letzten vorabgeht, und einen vergleichbaren Wortlaut (Attributwert ***c***[[9]](#footnote-11)) hat. Diese Verbindung zwischen den Motiven qualifizieren wir als Kausalität ersten Grades.

Andere Motive, die sich auf dieselbe Krise beziehen, sollen sich in unmittelbarer Nähe zu den oben genannten Motiven befinden und gleiche ***d*** Attributwert haben. Diese Verbindung nennen wir die Kausalität der zweiten Grad. Als letztes kommt die mit den Motiven der letzten Kategorie angehängte Elemente mit gleiche ***c*** Attributwert. Diese wird als Kausalität der dritten Grad bezeichnet.

Die Feststellung der Kausalität zwischen den Motiven einer Episode bzw. Episodencluster beginnt von unten nach oben. Dem letzten Motiv mit dem Attribute ***b=H/Hh*** soll in oberen Bereich der Liste ein Motiv entsprechen, das einen vergleichbare ***c*** Attributwert hat.

Als nächster Schritt sollen die Motive erfasst werden, die in der unmittelbaren Nähe von solchen Motiven stehen und gleiche ***d*** Werte haben.

Die Tatsache, dass durch diese Recherche komplexe Episodencluster herausgetragen werden oder einige Motive gleichzeitig in mehreren Episoden auftauchen, ist vollkommen normal und weist auf den Sachverhalt hin, demzufolge der konsekutive Abschluss einer Krise und somit auch einer Episode den Beginn einer neuen Krise bzw. Episode bedeuten kann. Folgendes Beispiel:

Der Held steht vor der Wahl, gegenüber dem Zwerg entweder genau so frech zu verhalten wie seine älteren Brüder *(h:Höfliches\_Verhalten:rST\_fHD)* oder im Gegensatz zu denen, sich höflich benehmen*.* Der Held verhält sich höflich, also wählt positive Lösung der Krise aus *(H:Höfliches\_Verhalten:rHD\_rST)*. Als Mehrwert der Lösung erhält er Handlungsanweisung darüber, wie er die Pförtner besänftigen kann, damit er einen freien Zugang zum Zielobjekt beschafft. Das Motiv lautet wie folgt: *H:Handlungsanweisung\_Pförtner\_besänftigen:rHD\_rST*

Zwar ist die Zugehörigkeit des Motivs zu der Episode mit Teilnahme des Stifters unumstritten, zeigt es anderseits genauso eindeutige Koppelung mit der ihm anschließend einzutretenden Episode. Diese entwickelt sich an einem anderen Ort und mit der Teilnahme ganz anderen Figuren. Der mit der Handlungsanweisungen des Stifters (hier des Zwerges) ausgerüstete Held gewinnt die Sympathie der leidende Wache und verschafft einen freien Zugang zum Zielobjekt (hier zum Heilmittel):

*H:Pfortner\_besänftigen:rHD\_fHF*

*H:Pfortner\_entkommen:rHD\_fHF*

Hier einige Beispiele der nach dieser Regel festgestellte Episoden anhand der Motivliste des oben annotierten Musterbeispiels:

1. Beispiel:

**Kausalität der ersten Grad**

**Kausalität der zweiten Grad**

**Kausalität der dritten Grad**

*H:Handlungsanweisung\_Objekt\_ergreifen:rHD\_fBZ*

*h:Kontakt\_abbrechen:rHD\_fBZ*

*H:Objekt\_ergreifen:rHD\_rZO*

*F:Kontakt\_wiederherstellen: fBZ*

*h:Ientität\_testen:fHD\_fVB*

*h:Ientität\_testen:fHD\_fVB*

*H:Ientität\_testen:rHD\_fBZ*

*H:Kontakt\_wiederherstellen:rHD\_fBZ*

1. Beispiel:

**Kausalität der ersten Grad**

**Kausalität der zweiten Grad**

**Kausalität der dritten Grad**

*h:Krankheit:rHH*

*F:Heilmittel\_beschaffen:rHD\_rVB\_fHD*

*F:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*

*H:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*

*F:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*

*H:Rettung\_der\_Hilfsbedürftigen:rHD\_rHF\_rZM*

*h:Heilmittel\_beschaffen:rHD\_rHH\_fZO*

*h:Krankheit\_beheben:rHD\_rHH\_fZO*

*Hh:Heilmittel\_beschaffen:rHH\_rZO\_fHD*

*Hh:Krankheit\_beheben:rHH\_rZO\_fHD*

*h:Verleumdung\_von\_Untreue:rHH\_rZO\_fHD*

*F:Zurückgreifen\_Rettung\_der\_Hilfsbedürftigen:rHH\_rHF*

*H:Verleumdung\_beheben:rHH\_rHF*

Der Aufgabe, wie diese Regel Zwecks der vollautomatischen Analyse des Märchens eingesetzt werden kann ist noch zu erledigen.

## **Maschinelle Analyse des Märchens**

### **Einführung**

Angesichts ihrer einfachen Formen und Strukturen sind Genres der mündlichen Überlieferung beliebte Objekte für die Demonstration der Effektivität einiger analytischer und generativer Modelle der künstlichen Intelligenz. Zwar beziehen sich viele dieser Versuche auf das wissenschaftliche Erbe der Finnischen Schule, insbesondere auf die ATU- und MIT-Klassifikationssysteme, schränken sich jedoch generell mit der Automatisierung dieser Instrumentarien ein, ohne ihre Grundmethode (Arbeitsmethode der Finnischen Schule) entsprechend den Anforderungen des neuen Ansatzes (Methoden der digitalen Geisteswissenschaften) grundsätzlich zu reformieren (Declerck u. a., o. D.; D'Huy 2019: passim).[[10]](#footnote-12) Noch populär zu sein scheint es, die durch die strukturalistische Analyse des Märchens erfasste Textbausteine sog. Proppsche Funktionen zu formalisieren und ihre maschinelle Erkennbarkeit zu prüfen.[[11]](#footnote-13)

Im Gegensatz zu diesen Versuchen handelt es sich in der vorliegenden Arbeit nicht um die Anwendung digitaler Methoden in der Märchenforschung, sondern um den digitalen Ausbau der Methode der Märchenforschung. Die wichtigste Aufgabe des neuen Ansatzes besteht in der formalisierten Erschließung des Forschungsobjektes (die Struktur des Märchens). Dass ein formalisiertes Forschungsobjekt maschinell nachvollziehbar sein muss, ist eigentlich selbstverständlich.

Um den artikulierten Unterschied zwischen den auf diese Strategie beruhte Forschung und den Ergebnissen der bereits veröffentlichten Abhandlungen im Bereich der elektronischen Märchenforschung zu verdeutlichen, versuchen wir hier beispielhaft, eine der jüngsten Publikationen zu diesem Thema zu besprechen. Es handelt sich um den Aufsatz von J. Eklund und anderen unter dem Titel: *"Teaching Tale Types to a Computer: A First Experiment with the Annotated Folktales Collection"* (Eklund et al. 2023: passim).

Neues Instrument alte Methode vs. Entsprechend dem neuen Wissen entwickelte Methode.

Wie aus dem genannten Artikel hervorgeht, setzen sich die Autoren zum Ziel, den Computer dazu zu bringen, die Zugehörigkeit der Texte zu den bekannten Märchentypen vorherzusagen. Als Basis für die Klassifikation dient der internationale Märchenkatalog (ATU) mit annotierten Typen. Die maschinelle Erkennung des Typs beruht auf den inhaltlichen Eigenschaften des Textes, die wiederum durch die im Motivkatalog von Thompson (MIT) erfassten inhaltlichen Einheiten identifiziert werden. Zusammenfassend ergibt sich, dass der Computer in der Lage ist, die Automatisierung der bei der Erstellung der Märchenkataloge erledigte Handarbeit zu gewährleisten.

Zum Beispiel in Bezug auf den konkreten Text folgendes zu sagen: Er besteht aus einer Kette von Motiven wie etwa[[12]](#footnote-14) B312.2, B11.2.3.1, B11.10, T68.1, D1975, D1978.2, B11.11, H105.1, K1933, K1932, H151.2 und H105.1 und sollte daher mit höherer Wahrscheinlichkeit dem Typ *ATU 300 – Der Drachentöter* zugeordnet werden.

Die Effizienz eines solchen Unternehmens ist nicht zu unterschätzen. Dennoch ist zu hinterfragen, ob es wirklich den Anforderungen und Leistungen der angewandten Klassifikation übertrifft.[[13]](#footnote-15) Eine automatisierte Erkennung des Textes ohne kritische Auseinandersetzung mit dem Erbe der vergleichenden Märchenforschung aus der Zeit vor den digitalen Geisteswissenschaften kann nichts anderes sein als die Anwendung digitaler Methoden in der Märchenforschung (siehe oben). Wie bereits angedeutet, distanzieren wir uns von dieser Herangehensweise und versuchen, die Schwächen der bisherigen Methode der vergleichenden Märchenforschung durch den Ausbau dieser Methode aufgrund der computergestützten Lösungsmodelle zu bewältigen.

Der Grund für die Entstehung des vorliegenden Vorhabens liegt genau im Bedarf nach einem, auf philologischer Analyse basierenden Modell, das für die maschinelle Analyse von Märchentexten verwendet werden kann. Formalisierte Beschreibung der Struktur des Märchens eröffnet den Weg zu maschineller Bearbeitung des Forschungsobjektes. Diese meint Automatisierung sowohl die Segmentation des Märchens in Strukturelemente von unterschiedlicher Größe als auch Erzeugen eines zusammenfassenden Bildes aufgrund des Vergleichs der je nach der Fragestellung ausgewählte Forschungsobjekten, wie etwa: vergleichende Analyse des Typs X anhand der Repertoires Y und Z. Kombination von Motiven innerhalb der Episode X. Status der Handlungstragende Figuren im Motiv Y und vieles andere.

Die Lösung der Maschinelle Analyse des Märchens als nächster Schritt nach der Formalisierung des Forschungsobjektes ist mit der Entwicklung der entsprechenden digitale Infrastruktur verbunden. Im Folgenden werden wir einige wichtigste Elemente dieser Infrastruktur betrachten.

### **Textkorpus**

Das Textkorpus ist eine XML-Datei, die auf dem TEI-Standard basiert und das Wurzelelement <teiCorpus> hat.[[14]](#footnote-16) Die Texte innerhalb des Korpus können in verschiedenen Sprachen vorliegen, müssen jedoch gemäß den Open-Data-Anforderungen standardisiert und in Unicode kodiert sein. Der gezielte Zugriff auf die Texte aus verschiedenen Sprachen oder Gruppen kann anhand der im <TEI>-Element erfassten Metadaten erfolgen. Hierbei ist die Anwendung des xml:id-Attributs beim Element <Text> empfehlenswert. Dieses Attribut besteht aus vier Teilen, die Informationen über die Lizenzbedingungen[[15]](#footnote-17), über die Zugehörigkeit des Textes zu einem bestimmten Repertoire, die Sprache des Textes und die laufende Nummer des Textes unter anderen Materialien mit gleicher Zugehörigkeit zum Repertoire enthalten.

Folgende Beispiele

* *deu\_deu\_1* = 1. Text aus dem deutschen Repertoire in deutsche Sprache
* *deu\_eng\_1* = Derselbe 1. Text aus dem deutschen Repertoire in englische Sprache
* *deu\_eng\_2* = 2. Text aus dem deutschen Repertoire in englische Sprache usw.

Für die Benennung der Namen der Repertoires und Sprachen werden Abkürzungen gemäß dem ISO-639-3-Standard verwendet.

Im TEI-Korpus sind die Märchen einzeln in separaten TEI-Elementen erfasst. Für jedes Märchen liegt eine codierte oder zur Codierung bereitgestellte Textfassung vor. Die Basistexte befinden sich in verschiedenen Repositorien und sind untereinander sowie mit dem codierten Text entsprechend referenziert. Die Texte, die nicht den OpenData-Forderungen entsprechen, werden im Textkorpus in abgeleiteter Form dargestellt und mit bibliographischen Angaben zur primären Textquelle versehen.

Die Hauptaufgabe des Korpus besteht darin, eine standardisierte Segmentierung der Daten sowie einen gezielten Zugriff auf diese Daten zu gewährleisten. Das entwickelte Datenmodell ermöglicht es, Szenen im Text auseinander zu differenzieren und in ihr eingebettete Motive zu erfassen. Dafür sind das Element ***<seg>*** und in ihm eingebettete Tupel von Attributen ***a, b, c*** und ***d*** zuständig. Die optimale Anzahl von Motiven innerhalb einer Szene beträgt fünf. Dies bedeutet, dass für jedes Motiv jeweils zwei Attribute ***b*** und ***c*** erfasst werden können, sowie weitere zwei gemeinsame Attribute ***a*** und ***d*** für alle fünf Tupel.

Ein Beispiel für die Annotation des 10. Abschnitts in unserem Text wäre wie folgt (N gilt für die Null):

*<seg*

*a1:ana = “a551“*

*b1:ana = “F“ c1:ana = “ Höfliches\_Verhalten“*

*b2:ana = “H“ c2:ana = “ Höfliches\_Verhalten“*

*b3:ana=“H“ c3:ana=“ Handlungsanweisungen\_Wache\_besänftigen“*

*b4:ana=“N“ c4:ana=“ N“*

*b5:ana=“N“ c5:ana=“ N“*

*d:ana=“ rHD\_rHF“ >* Die Begegnungsszene zwischen dem Held und dem Stifter

*</seg>*

### **Künstlicher Assistent für die semiautomatische Annotation der Texte**

Ein zentraler Bestandteil der maschinellen Analyse von Märchen ist ein künstlicher Assistent mit Vorhersagefunktion.[[16]](#footnote-18) Dieser Assistent ist dafür zuständig, in einer umfangreichen Menge von Daten die Textabschnitte zu erkennen, die über die gesuchten inhaltlichen Eigenschaften verfügen.

Die Entwicklung eines solchen Assistenten erfordert die Erfüllung der folgenden drei Aufgaben:

1. Auswahl einer geeigneten Vorgehensweise für die maschinelle Textanalyse.
2. Anpassung dieser Vorgehensweise an die Forschungsfrage.
3. Umformulierung des gewonnenen Wissens über das Forschungsobjekt in die Sprache der ausgewählten Vorgehensweise.

Besonders wichtig bei der letzten Aufgabe ist die Anpassung zwischen den konzeptuellen und empirisch beobachtbaren Elementen. Eine Szene innerhalb der Geschichte wird dabei als empirisch beobachtbare Grenze für den Anfang bzw. das Ende eines Motivs betrachtet. Für die Maschine bedeutet dies eine Zeichenkette, bei der das kleinste erkennbare Element ein Wort und das größte ein Absatz ist. Auf inhaltlicher Ebene entspricht dies dem Textabschnitt mit veränderten Figuren, ihren Attributen, Ortschaften und Handlungen. Wir können zwar der Maschine beibringen, diesen Wechsel zu erkennen, jedoch müssen wir zuvor deutlich machen, welcher Zusammenhang zwischen den Roh-, Trainings-, Test- und Zieldaten besteht.

Der rohe Text, wie er uns gewöhnlich vorliegt, ist durch die folgenden explizit erkennbaren Zeichen wahrnehmbar: Wort, Satz, Absatz und Text. Die Grenzen der somit organisierten Textteile entsprechen jedoch nicht unbedingt dem von uns gesuchten Szenenwechsel im Text.

Die Aufbereitung des rohen Textes zu Trainingsdaten erfordert die Umwandlung der sogenannten Istzeichen in Sollzeichen. Dabei ist es oft notwendig, die Anzahl der Absätze im Text zu korrigieren. Die Absätze können entweder zusammengelegt oder in mehrere Teile aufgeteilt werden. In manchen Fällen kann auch ein einziger Satz einem solchen Absatz entsprechen.

Messdaten sind den Rohdaten gleich, sie enthalten bibliographische Metadaten und stellt wohlgeformte XML Daten dar. Zieldaten sind annotierten Messdaten. Diese können teilweise oder vollständig annotiert werden. Die bereits annotierten Teile der Messdaten werden bei der automatischen Analyse der Texte nicht miteinbezogen. Sie können allerdings als Trainingsdaten verwendet werden.

Für die automatische Erkennung der Motive und Episoden im Text verwenden wir den Algorithmus der logistischen Regression (LogisticRegression). Ein Modell, das von diesem Algorithmus gesteuert wird, versucht, Merkmale zu finden, die für die Erkennung der gesuchten Episoden (positiv etikettierte Datensätze) sowie ihrer Gegenteile (negativ etikettierte Datensätze) anwendbar sind, und ordnet sie auf beiden Seiten einer imaginären Trennlinie an.

Entscheidend für die Einbettung der Elemente in das Modell sind die nach den TF\*IDF-Werte skalierten Merkmale der positiv und negativ bewerteten Datensätzen. Vor der Berechnung der TF\*IDF-Werte findet die Datenbereinigung statt, wobei sogenannte Stoppwörter und andere Wörter, die für den Vergleich irrelevant sind, von Datensätzen entfernt werden.

Aufgrund der üblicherweise begrenzten Anzahl positiv etikettierter Trainingsdaten wird eine mehrfache Kreuzvalidierungsmethode angewendet. Diese Methode teilt die gesamten Trainingsdaten in gleichmäßige Abschnitte auf, die als Validierungs- und Trainingsdaten verwendet werden. Das trainierte Modell sollte nun in der Lage sein, die charakteristischen Merkmale (Wortformen) für positiv und negativ etikettierte Daten zu trennen und sie als eine sortierte Liste der entsprechenden Zeichengruppen darzustellen.

Das trainierte Modell wird durch die Vorhersagefunktion erweitert. [[17]](#footnote-19) Obwohl die Funktion in einigen Fällen scheitert, kann sie dennoch als zuverlässiges Werkzeug betrachtet werden. Der genaue Grund für das Scheitern der Vorhersagefunktion in einigen Fällen ist nicht eindeutig zu ermitteln. Es ist jedoch anzunehmen, dass dies teilweise an der Methode liegt, die Ähnlichkeit zwischen dem Modell und den Messdaten abzugleichen. Die Sigmoid-Funktion, die erfolgreich zur Berechnung der Distanz zwischen dem Modell und den Messdaten eingesetzt werden kann, könnte hierbei möglicherweise behindert werden, da der Datensatz oft beide Klassen (sowohl positiv als auch negativ) repräsentiert. In solchen Fällen könnte die angenommene Ähnlichkeit oder der Unterschied zwischen den verglichenen Inhalten fehlerhaft sein, da im geprüften Textteil sowohl negative als auch positive Merkmale gleichzeitig vorhanden sein können.

Unter Berücksichtigung des oben genannten Faktors scheint es viel effektiver zu sein, die Daten ausschließlich anhand der mit den höchsten Koeffizienten versehenen positiven Merkmale zu untersuchen. Dadurch werden die Datensätze nicht mehr über das Modell verglichen, sondern über das aus dem Modell extrahierte prototypische Metamotiv. Dieses funktioniert wie eine Suchabfrage an die Dokumentensammlung.

Ein Bild, das Screenshot, Reihe, Rechteck, Diagramm enthält.

Automatisch generierte Beschreibung

Abb. 6

Das Extrahieren der Merkmale erfolgt durch die Berechnung des Schwellenwertes für die Topklassifikationsmerkmale. Dieser Wert ist an der Stelle der Merkmal-Kette zu suchen, wo die Gewichte der positiven und negativen Merkmale bzw. Koeffizienten abgeglichen werden.

Nach dem graphischen Verfahren kann man diese Stelle wie folgt lokalisieren: Die Abbildung 6 zeigt die Schwingung zwischen den maximalen und minimalen Koeffizienten der negativ und positiv eingeschätzten Klassifikationsmerkmale. Die für beiden Kategorien geltende maximale Zahl der Merkmale (z) ist eine Variable und kann entsprechend der Situation angepasst werden, diesmal beträgt sie 200. Blaue Spalten stellen den positiven Merkmalen dar, die roten Spalten visualisieren hingegen die negativen Merkmale (Wegen der hohen Anzahl überlappen sich die Merkmale (siehe schwarze Wortwolke im unteren Bereich der Grafik). Für die Visualisierung der gesuchten Stelle wird die vorhandene Grafik an dem Treffpunkt von roten und blauen Spalten geteilt und die linke Hälfte so weit nach rechts verschoben, bis die kleinste rote Spalte unter der größten blauen Spalte platziert wird. Beachtet man die Größe von blauen und roten Spalten kann man leicht feststellen, an welcher Stelle sie übereinstimmen (s. Abb. 6.1).

Ohne graphische Darstellung kann man derselben Schwellenwert anhand folgender Formel berechnen:

Kommt die Gleichung zustande:

, *wobei ,*

so ist der Schwellenwert gleich *.*

Da die Topklassifikationsmerkmale nichts anderes darstellen als eine einfache Liste von positiv bewerteten Elementen bzw. Wörtern, ist ihre Bildung mit Hilfe einer einfachen Stufenfunktion möglich (entspricht der grünen Grafik auf der Abb. 6.1):

Ein Bild, das Screenshot, Reihe, Rechteck, parallel enthält.

Automatisch generierte Beschreibung

Abb. 6.1

Die Liste der auf diese Weise festgelegten Merkmale wird als prototypisches Metamotiv in einem Datensatz gespeichert und bei den Messdaten angelegt.

Die Erkennung des gesuchten Motivs in den Messdaten erfolgt durch die Klassifikation auf Basis der Kosinus-Ähnlichkeit. [[18]](#footnote-20) Dabei wird jeder Datensatz mit einem synthetischen Datensatz verglichen, der von uns oben als das prototypische Metamotiv definiert wurde.

### **Test des Assistenten**

Für den Test der Vorhersagefunktion des Assistenten werden wir folgendes Experiment durchführen. Zunächst erstellen wir ein Textkorpus aus Daten, die mit hoher Sicherheit dem Genre Märchen zuzuordnen sind. Hierfür nutzen wir die im TextGrid-Repositorium veröffentlichten deutschsprachigen Märchensammlungen.

Wir wählen die Märchentexte von Jacob und Wilhelm Grimm (215), Heinrich Pröhle (150), Karl Bartsch (52 + 47) und Ludwig Bechstein (81) aus und fügen sie in einem XML-TEI-Corpus zusammen. Zusätzlich werden 12 weitere Texte aus der Sammlung „Europäische Märchen“ aufgenommen. Aufgrund rechtlicher Gründe sind diese Texte in einer abgeleiteten Fassung dargestellt. Insgesamt umfasst unser Korpus somit 557 Texte.

Nun wählen wir ein Beispiel eines Märchentyp ATU 300 – Der Drachentöter aus der Kollektion von Karl Bartsch aus und annotieren den Text anhand des Markups für die inhaltliche Annotation des Märchens.

Der ausgewählte Text stellt ein komplexes Märchen dar, das aus vier Typteilen besteht. Nach den Anfangsepisoden aus den Typen ATU 567 (vom Essen des außergewöhnlichen Vogelherzes bis zur Flucht) und ATU 303 (Großziehen von zwei ähnlich aussehenden Brüdern bis zur Trennung der Brüder und Austausch der Lebensindikatoren) beginnt die Episode des Typs ATU 300 – Der Drachentöter. Diese umfasst den Inhalt von der Erkundigung der menschlichen Opfergabe bis zur Heirat der geretteten Prinzessin. Der Text wird weiter als Typ ATU 303 fortgesetzt. Dabei wird berichtet, wie der jüngste Bruder von der Hexe versteinert, aber dann mit Hilfe seines Bruders wieder erlöst wird.

Für das Experiment werden wir in diesem Märchen nur den Textteil des Typs ATU 300 erschließen, das heißt mit Hilfe des Markups labeln. Somit wird das Modell nur anhand positiv gelabelter Datensätze trainiert. Im Rahmen dieses Verfahrens werden aus allen gelabelten Datensätzen die Tokens extrahiert und ein Vokabular erstellt.

Dabei wird das Vokabular von sogenannten Stoppwörtern und anderen für den Vergleich unwesentlichen Wörtern bereinigt. Letztere gelten als seltene Wörter, die nur vereinzelt in den Datensätzen vorkommen und daher keine Rolle beim Vergleich dieser Datensätze mit anderen spielen. Für die Bereinigung des Vokabulars von vereinzelt stehenden Wörtern ist der Parameter min\_df=n zuständig. Angesichts der geringen Anzahl der Datensätze sowie der Kürze der Liste der in diesen Datensätzen anzutreffenden Tokens, wird der Wert des min\_df Parameters bei unserem Experiment auf den minimalen Wert (min\_df=2) gesetzt. Das bedeutet, dass jedes im Vokabular erfasste Wort wenigstens in zwei verschiedenen Datensätzen vorkommen muss.

Als nächstes werden die im Vokabular vorhandenen Wortformen nach ihren TF\*IDF-Werten skaliert. Die TF (Term Frequency) eines Tokens wird berechnet als:

Häufigkeit des Tokens in einem Datensatz / Gesamtzahl der Tokens im Datensatz

Die IDF (Inverse Document Frequency) eines Tokens wird berechnet als:

Loge (Gesamtzahl der Datensätze + 1 / Anzahl der Datensätze mit dem gesuchten Token + 1) + 1

Anschließend werden die TF\*IDF-Werte durch das L2-Normalisierungsverfahren umskaliert. Das somit berechnete Koeffizient bestimmt den Wert jedes einzelnen Tokens im Vokabular. Die Rangordnung der Tokens im Vokabular ermöglicht es, die Anzahl der Tokens im Vokabular effektiv zu korrigieren.

Für die Annäherung zwischen dem erstellten Vokabular und dem Modell einer konkreten Geschichte ist die Segmentierung der Trainingsdaten in natürliche Handlungsabschnitte von besonderer Bedeutung. Sie stellt sicher, dass Wörter, die in zwei aufeinanderfolgende Handlungsabschnitte (Offenkundigkeit der Krise – Auseinandersetzung mit der Krise) wiederholt vorkommen, mit getrennten Datensätzen dargestellt werden. Dadurch erhalten sie höhere TF\*IDF-Gewichte oder werden überhaupt im Vokabular erfasst.

Im nächsten Schritt wird aus den restlichen (nicht gelabelten) Datensätzen im Textkorpus ein Messdatenkorpus aggregiert. In diesen Messdatenkorpus wird auch der zuvor erstellte Modell-Datensatz eingefügt. Anschließend werden alle anderen Datensätze aus dem Messdatenkorpus mit dem Modell-Datensatz verglichen, um vorherzusagen, welche von ihnen diesem am nächsten stehen.

Das Verfahren umfasst folgende Schritte:

1. Erstellung und Bereinigung des Vokabulars: Das Vokabular wird von Stoppwörtern und anderen für den Vergleich unwesentlichen Wörtern bereinigt.

2. Berechnung der TF\*IDF-Werte: Für jedes Wort im Vokabular werden die TF\*IDF-Werte berechnet.

3. Erstellung dünnbesetzter Matrizen: Für jeden Datensatz wird eine dünnbesetzte Matrix erstellt, die die TF\*IDF-Werte der enthaltenen Tokens repräsentiert.

4. Vergleich der Matrizen: Jede dieser Matrizen wird mit der Matrix des Modell-Datensatzes verglichen.

Die Berechnung der Ähnlichkeit zwischen den Datensätzen beruht auf der Formel der Cosinus-Ähnlichkeit, die wie folgt definiert ist:

Dabei entspricht ***ai*** einem, in einem TF\*IDF Wert skalierte Merkmal (Token) aus dem Modell-Datensatz, ein ***bi*** entspricht hingegen einem in einem TF\*IDF Wert skalierte Merkmal (Token) aus dem beliebigen Mess-Datensatz. Je näher der Wert der berechneten Cosinus-Ähnlichkeit bei Eins liegt, desto plausibler ist die vorhergesagte inhaltliche Ähnlichkeit zwischen ***a*** und ***b*** Datensätzen. Diese Methode ermöglicht es, die Ähnlichkeit der Datensätze zu quantifizieren und diejenigen zu identifizieren, die dem Modell-Datensatz am nächsten stehen.

Die Datensätze mit den höchsten Koeffizienten, also die wahrscheinlichsten Vorhersagen, stehen an der Spitze der Liste. Anhand des Nachschlagewerks von Hans-Jörg Uther, dem „Deutschen Märchenkatalog“, lässt sich mühelos überprüfen, ob der Algorithmus alle im Textkorpus enthaltenen Texte bzw. ihre Teile gefunden hat, die zur gesuchten Märchenkategorie gehören (gemeint ist der Typ ATU 300 – Der Drachentöter).

Die Recherche im Katalog zeigt, dass in der Tat alle entsprechenden Texte durch den Algorithmus sicher erfasst wurden. Das sind wie folgt:

|  |  |
| --- | --- |
| Text im Katalog | Text ID im XML-Textkorpus |
| Bartsch 1879f. I, 474ff. | @n=“dummy20“ |
| Bechstein/Uther 1997 I, Nr.49 | @n=“dummy122“ |
| Grimm/KHM (1857) Nr. 60 | @n=“dummy484“ |
| Pröhle 1853, Nr. 4 | @n=“dummy234“ |
| Pröhle 1853, Nr. 5[[19]](#footnote-21) | @n=“dummy256“ |

In diesem Experiment gezeigte Verfahren ist insbesondere bei der Erstellung der Trainingsdaten einzusetzen. Mit den so gewonnenen Trainingsdaten ist es möglich, die Vorhersage nicht nur zu präzisieren, sondern auch auf kleinere Inhaltselemente zu richten. Um diese Eigenschaft zu demonstrieren, führen wir ein weiteres Experiment durch.

Wir trainieren das Modell anhand der zuvor gelabelten Datensätze, darunter die oben erwähnten Texte des Typs ATU 551 „Das Wasser des Lebens“.[[20]](#footnote-22) Die ID-Zeichen dieser Texte im Textkorpus beginnen mit dem Suffix „zyx\_“. Nach der Aggregation eines provisorischen Textkorpus aus diesen und den im ersten Experiment gelabelten Datensätzen (Datensätze mit dem ID-Suffix „dummy“) können wir alle in diesem Korpus vorhandenen Markupelemente abrufen und betrachten.

Als nächstes wählen wir aus dieser Liste ein oder mehrere Markupelemente aus und erstellen eine entsprechende Abfrage. Angenommen, wir suchen nach dem Motiv der Offenkundigkeit der Krankheit des Vaters. Die entsprechende Abfrage besteht aus den Attributen „a551“ und „Krankheit:“. Der Doppelpunkt nach dem Wort „Krankheit“ stellt sicher, dass ein weiterer Wortlaut wie „Krankheit\_beheben“ in der Suche ignoriert wird.

Das Ergebnis der anschließenden maschinellen Analyse bestätigt die hohe Präzision der eingesetzten Vorhersagefunktion. Der erste Treffer entspricht dem Textteil mit dem gesuchten Motiv. Außerdem ist der Unterschied zwischen den Koeffizienten des ersten und der folgenden Datensätze so groß, dass sie zweifelsohne zwei kontrastiv unterschiedliche Vorhersagen darstellen müssen. Tatsächlich, wie der oben erwähnte Katalog bestätigt, sind im Textkorpus keine anderen Texte mit gleichem Inhalt zu erwarten.

### **Tools für die Auswertung der Daten**

Durch die inhaltliche Annotation wird es möglich, eine große Menge von Texten einheitlich und mit präziser Fokussierung auf die konkreten Forschungsfragen auszuwerten und Ergebnisse nachvollziehbar und anschaulich darzustellen. Hier betrachten wir Anwendungsbeispiel mit dem Ziel die Normalform eines Typs in Westeuropäische Kulturraum in der visualisierten Form auszuwerten.

Im Vorfeld der Auswertung muss es vorbestimmt werden (a) nach welchem Typ bzw. Typencluster wir suchen und (b) im Welchem Teil der Daten die Suche durchgeführt werden soll.

Angenommen, interessieren wir uns für die Normalform des Märchentypus *a551 – Wasser des Lebens* im europäischen Märchenrepertoire[[21]](#footnote-23) (Deutsch, Österreichisch, Französisch, Italienisch, Ungarisch, Griechisch, Russisch).

Nach der Eingabe von entsprechenden Suchparameter greift der Assistent auf die annotierten Märchentexten mit passenden Identifikatoren zu und stellt die Auswertungsergebnisse in der Form von Graphen und dabei in der einfachen Liste der Kompositionsbau beteiligten Motive dar.[[22]](#footnote-24)

Der Graph zeigt ein zusammengeführtes Bild der in den verschiedenen Texten gegenseitig geketteten Motiven mit gesuchtem ***a*** Attributwert. An den Peripherien der Kette können auch einige Motiven auftreten, die andere ***a*** Attributwerte haben. Diese stellen ihrerseits die Peripherien der Motivkette dar, die den mit dem gesuchten Typ kombinierten Typen ausmachen.[[23]](#footnote-25) Nach der Eingabe der Werten von solchen ***a*** Attribute ist zu erwarten, dass der Graph der Beziehungen von Motiven innerhalb eines Typs zum Graphen der Beziehungen der Motive innerhalb eines Typenclusters umgewandelt wird.

Die Auswertung der Motivbestand in einem Typ bzw. Typcluster, wird automatisch durch die Aggregation der Bibliothek der dabei involvierten Motive gefolgt. Diese kann, solange es keine entsprechende Aufforderung vorliegt, eine einfache Reihe von Motiven ohne Hinweise auf jegliche Verbindung darstellen. Nach der Eingabe eines konkreten Motivs wird die Liste wie folgt reorganisiert: das eingegebene Motiv bekommt die Nummer Null und wird an seine Stelle in der Reihe von gesamten Motiven platziert, die auf der Zeitachse ihm nach vorne und nach hinter liegende Motive bekommen entsprechende Nummerierung mit Minus- und Pluszeichen und werden den Platz oberhalb bzw. unterhalb des mit Null bezeichneten Motivs übernehmen. Wird die Kette der Motive von einem Typ durch eine oder mehrere Motiven mit einem anderen ***a*** Attribut getrennt so werden sie ohne Hinweis auf ihre Position in der Motivkette erfasst. Somit erhalten wir eine Liste von Motiven, die mit dem zusammengefassten Inhalt des ausgewerteten Typs übereinstimmt.

Angesichts der Unübersichtlichkeit, die durch die Wechsel der handelnden Personen hervorgerufenen Variationen der Motive mit sich bringen, ergibt sich die Empfehlung, das zusammenfassende Bild der Komposition des Märchens auf einige wenige Elemente zu verdichten. In diesem Sinne scheint es effektiv zu sein, das letzte Attribut aus dem annotierten Markupelement zu entfernen.

Beispiel: Im Typ a551 (=Attribut a) kann der Auftrag "F" (=Attribut b) zur Beschaffung des Lebenswassers (=Attribut c) verschiedenen Protagonisten (Attribut d) erteilt werden, sei es dem kranken Vater (rHH) oder dem ratgebenden alten Mann (rST). Dies führt automatisch dazu, dass bei der Zusammenfassung im Prinzip derselben Handlung zwei Markupelemente erstellt werden:

a551:F:Heilmittel\_beschaffen:rHH

a551:F:Heilmittel\_beschaffen:rST

Durch die Abschaffung des Attributs ***d*** bleibt nur ein Markupelement übrig, das trotz des Informationsverlustes in der Lage ist, ein gemeinsames Bild der Handlungskette zu vermitteln. Entsprechender Code findet man in GitHub-Repositorium.[[24]](#footnote-26)

Letztendlich hängt der wissenschaftliche Wert und die Plausibilität einer solchen Auswertung vollständig vom Grad der Repräsentativität der erfassten Forschungsdaten ab. In dieser Publikation beschränken wir uns darauf, die Funktionsweise des Codes zu demonstrieren. Die mit dem Einsatz des Codes erzielten Ergebnisse anhand des kaukasisches Erzählgutes werden separat veröffentlicht.

## **Ausblick**

Die vorliegende Arbeit untersucht allgemeine Regeln für die Organisation und Kombination von inhaltlichen Elementen in allen Ebenen des Märchens: Motiv, Episode, Vollgeschichte und gibt die formalisierte Beschreibung der erkannten Ordnungen. Auf Basis des, dadurch erfassten konzeptuellen Models des Genres entwerfe ich ein Machine\_Learning\_Modell, das uns ermöglicht die kosten und zeitintensivste Aufgaben des vergleichenden Märchens 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.

Ein wichtiger Bestandteil der im Zuge der vorliegenden Arbeit bereitgestellten Infrastruktur ist das Modul zur visualisierten Darstellung der ausgewerteten Normalformen von einzelnen Typen. Besonders effektiv scheint die Anwendung der Visualisierungsmethode durch Digraphen zu sein. Das daraus entwickelte Bild vermittelt eine klare und lückenlose Vorstellung über die Popularität einzelner Typen und die Häufigkeitsfälle ihrer gegenseitigen Kombinationen.

Die Ergebnisse dieser Studie werden gegenwärtig praktisch angewandt, um zur Erschließung der nordkaukasischen Folklore beizutragen. Das von der DFG geförderte Projekt verfolgt das Ziel, ein elektronisches Datenkorpus zu erstellen und durch die Annotation der inhaltlich-strukturellen Eigenschaften der in diesem Korpus gesammelten Texte diese für die international-vergleichende Märchenforschung bereitzustellen. Als besondere Herausforderungen gelten dabei:

1. sprachlichen Vielfältigkeit des Materials[[25]](#footnote-27),

2. Einschränkungen bei der Anwendung der Open-Date Prinzipien.[[26]](#footnote-28)

Nach nur anderthalb Jahren Anwendung der in diesem Buch erfassten theoretischen und praktischen Forschungsergebnisse zeichnet sich ihre Nachhaltigkeit deutlich ab. Besonders bemerkenswert ist dabei die Fähigkeit, sich neuen Entwicklungsanforderungen anzupassen. Dies wird deutlich in der Umstellung von der CSV-Datenkollektion auf ein XML-Datenkorpus sowie im Übergang von grundlegenden Segmentierungsverfahren von Episoden auf die Motivebene.

Bei fortgesetzter Anwendung dieser Arbeitsmethode lässt sich das Erreichen eines besonderen Synergieeffekts gut prognostizieren. Wie bereits angedeutet, entwickelt sich die vergleichende Märchenforschung sukzessiv, nach dem Vorbild eines selbstregulierenden Systems. Dies bedeutet unter anderem, dass jedes neu erschlossene Repertoire die vergleichende Analyse eines anderen Repertoires begünstigt; beide tragen wiederum zur Erschließung des nächsten bei, und so weiter. Unter Berücksichtigung dieses Aspekts werden die Ergänzungen der einzelnen infrastrukturellen Elemente, wie beispielsweise Markups oder synthetische Trainingsdaten, als beiläufiger Beitrag bei jedem neuen Einsatz der Arbeitsmethode betrachtet.

1. Anderson, Walter (1934/40): Geographisch-historische Methode. *Handwörterbuch des deutschen Märchens,* Hrsg. von Lutz Mackensen. Bd. 2, Berlin 1934/40, S. 508-522; Röhrich, Lutz (1987): Geographisch-historische Methode. In: *Enzyklopädie des Märchens.* Bd. 5, Berlin/New York, Sp. 1012-1030. [↑](#footnote-ref-1)
2. Propp, Vladimir J. (1975): *Morphologie des Zaubermärchens*, Hrsg. Karl Eimermacher, Frankfurt am Main, Suhrkamp. [↑](#footnote-ref-2)
3. Nach Anderson zerfällt die Erzählung in Episoden, die man auch Akte oder Kapitel nennen könnte. „Eine jede Episode besteht aus mehreren einzelnen Motiven oder Momenten, die ihrerseits eine Reihe von Zügen enthalten.“ Es ist zu betonen, dass der Begriff Zug, aus dessen Reihen das Motiv besteht, nichts anderes umfasst als Variationen einzelner Motive. Zum Vergleich siehe Anderson 1923: 242-260. [↑](#footnote-ref-3)
4. Handlungsanweisung muss eindeutig von der Probestellung unterschieden werden. Sie kommt erst dann vor wenn die Beziehung zwischen dem Helden und dem Stifter bzw. Helfer geklärt ist, und der Held vom Stifter eine garantierte Unterstützung genießen darf. [↑](#footnote-ref-6)
5. Vollständige Liste der Attribute s. GitHub: edadunashvili/VerMa/kf/vmf\_c1-5.xsd [↑](#footnote-ref-7)
6. vollständige Liste der Elemente s. GitHub [↑](#footnote-ref-8)
7. Vollständige Liste der Attribute s. GitHub: edadunashvili/VerMa/kf/vmf\_d.xsd [↑](#footnote-ref-9)
8. Wasser des Lebens, KHM 97; vgl. Blécourt, Willem de(2014): Wasser des Lebens. Enzyklopädie des Märchens. Sp. 509-514. [↑](#footnote-ref-10)
9. Gemeint werden geklonte und abgeleitete Wortlaute, die die *c* Attributwerte ausmachen (siehe den Abschnitt *Attribut c*) [↑](#footnote-ref-11)
10. zu einem der jüngster [↑](#footnote-ref-12)
11. Als jüngster Versuch sind die Publikationen von Finlayson (2012, 2016) sowie Lendvai u.a. (2010a, 2010b) zu nennen. [↑](#footnote-ref-13)
12. Für die Auflösung der hier aufgezählten Indizien siehe MIT [↑](#footnote-ref-14)
13. Gemeint werden die obengenannte Klassifikationsmodele ATU und MIT [↑](#footnote-ref-15)
14. Github: edadunashvili/VerMa/Textkorpus.xml [↑](#footnote-ref-16)
15. Dafür werden Abkürzungen cc für die freigegebene Daten und cr für die mit dem Copyright geschützte Texte. [↑](#footnote-ref-17)
16. GitHub: edadunashvili/VerMa/erthaos\_11.ipynb [↑](#footnote-ref-18)
17. GitHub: edadunashvili/VerMa/erthaos\_11.ipynb (Zellen #14-1). [↑](#footnote-ref-19)
18. W.o. Zellen # 16 – # 18. Als vorbildlich gilt die Anwendung von Needham, Mark 2016. [↑](#footnote-ref-20)
19. Der Text stellt eine Kombination von Typen 303 und 300 dar. Im Katalog von Uther nur bei der Erschließung des Typs ATU 303 erfasst. [↑](#footnote-ref-21)
20. Europäische Märchen. CD-Ausgabe [↑](#footnote-ref-22)
21. Repertoire entspricht den in der elektronischen (CD) Ausgabe der „Europäische Märchen“ zusammengetragene Texte unter dem Typnummer ATU 551. [↑](#footnote-ref-23)
22. GitHub: edadunashvili/VerMa/Auswertung\_abcd.ipynb [↑](#footnote-ref-24)
23. Stellt ein Mitglied aus den gesuchten Motiven den Anfang oder das Ende eines Textes dar, so wird dieses durch extra Zeichen „Anf-“ und „End-“ gekennzeichnet. [↑](#footnote-ref-25)
24. GitHub: edadunashvili/VerMa/Auswertung\_abc.ipynb [↑](#footnote-ref-26)
25. Auf dem heutigen Territorium der Föderationsrepublik Dagestan werden ungefähr 25 einheimische Sprachen gesprochen. [↑](#footnote-ref-27)
26. Als Ausweg aus dieser Situation bleibt die Umwandlung der Originaltexte in abgeleitete Forschungsdaten. Häufig verwenden wir dabei Texte, die von Stoppwörtern bereinigt wurden. Referenzen zu den Quelldaten sind bei der Freigabe solcher Ableitungen immer routinemäßig nachweisbar. [↑](#footnote-ref-28)