Selbstähnlichkeit in Texten

Märchenhafte Beiträge zur Computererzähltheorie

Cosimo Palma

Master in Computerlinguistik

Fachbereich II

Universität Trier

Betreuer:

Dr. Sven Naumann

Vivian Boyer, M.A.

Trier, 27.03.2018

Danksagung

Ich möchte hiermit bei den Leuten mich persönlich bedanken, die dem Erfolg meines Abschlusses und besonders meiner These wesentlich beigetragen haben.

Meiner Mutter, die mich von unten unterstützt hat. Vielen Dank. Meiner Frau, Quynh, die immer neben mir geblieben ist, die das Literaturverzeichnis in Ordnung gesetzt hat, und dazu extra Geschirre spülen haben soll: vielen Dank meine Blume, bald wirst du belohnt, und nicht durch eine Scheidung. Der oberen Jungfrau, die mich vom oben geführt hat. Vielen Dank, vielen Dank. Mir ist immer bewusst, dass die gute Ideen, die diese Abschlussarbeit durchqueren, nicht mein Verdienst sind. Mein Schuld bleibt aber, sie ab und zu arm ausgedrückt zu haben. Dem Pasquale, der einfach toll ist: mehrmals Dank, für zu viele Sachen. Und meinem Professor und Betreuer der Dr. Sven Naumann, dessen Geduld, Freundlichkeit und Hilfsbereitschaft nicht nur während des Niederschreibens der These mich begleitet haben: vielen Dank. Hoffentlich können auch die andere ausländische Studenten der Welt seinen eigenen Naumann immer finden. Zuletzt aber nicht als letzte möchte ich mich auch bei Julie, Jörg, Filiz und Desiree für ihre sorgfältiges und kurzfristiges Durchlesen meiner These. Die letzte dieser letzten verdient für ihren stetigen Beistand in diesen schwierigen und schönen Semestern zusätzliche Danksagungen. *Eres un fenómeno, bruja!* Gut geschehen.

Abstract

Die folgende Studie setzt sich als Ziel die Suche und die Messung der Selbstähnlichkeit in Texten. Denkweisen und Werkzeuge werden hierfür auch aus anderen wissenschaftlichen Gebieten entlehnt und in unserem Kontext sorgfältig gepfropft. Für das Experiment werden der *Hurst Exponent* (als Selbstähnlichkeitsmaß) und der *Dependenzabstand* (als Komplexitätsmaß) einiger Märchen berechnet, um eine vergleichende Analyse zwischen den Textebenen durchführen zu können.

Inhaltsverzeichnis

1 Deíxis **1**

1.1 Einleitung und Problemstellung 1

1.2 Selbstähnlichkeit: Rekursivität, Fraktale und der goldene Schnitt 3

1.3 Der Text als teleologischer Gegenstand 4

1.3.1 Folge und Struktur I: Der Text als Vereinigung von Synchronie und Diachronie 7

1.3.2 Ascombes Quartett und Isoparametrische Texte: eine wissenschaftstheoretische Parenthese 8

1.3.3 Ausflug in den Fibonacci Zahlen 9

1.4 Transversale Konzeptualisierung 14

1.4.1 Erzählung als Schachspiel 17

1.5 Eine (rückwärts) Geschichte des Problemstellung 20

1.5.1 Selbstähnlichkeit in Texten 20

1.5.2 Noch eine Überlegung über die Folge 26

1.5.3 Die Universität Saarland 27

1.5.4 Die Computererzähltheorie 30

1.5.5 Finlayson 32

1.5.6 Ein bündiger Standpunkt der Forschung 33

1.6 Kurzer aber nötiger Ausflug in der Erzähltheorie 34

1.6.1 Strukturalismus und Poststrukturalismus 34

1.6.2 Propp, Lévi-Strauss, Greimas 35

1.7 Text- und Sprachmodelle 38

1.7.1 Aristoteles 38

1.7.2 Hjelmslev 39
1.7.3 Chomskys generatives Modell 40
1.7.4 Semantik als Physiognomik des Textes 42

2 Katastrophé **45**
2.1 Korpusauswahl 45
2.1.1 Warum das Märchen? Eigenschaften des Märchens 45
2.1.2 Vorläufige Analyse der Grimms Märchen 46
2.1.3 Die *Dramatis Personae*: ein letzter Schritt zur Adoption und Anpassung des Propps Modelles 47
2.1.4 Auflistung der Propps Funktionen 48
2.1.5 Beschreibung der wichtigeren Eigenschaften 51
2.1.6 Widersprechende und ergänzende Paare 52
2.1.7 Zusammensetzungsregeln: die Erzählung als ein L-System 54
2.1.8 Folge und Struktur II: Thue-Morsefolge und UID 58
2.2 Erwartete Reihenfolge und Topic Modeling 61
2.3 Folge und Struktur III: die magischen Quadrate 62
2.3.1 Position und Wert 64
2.3.2 Wert als Selbstinformation in den magischen Quadraten 68
2.3.3 Sentiment Analysis als Wert in der magischen Quadraten 70
2.4 Folge und Struktur IV: Hurst-Exponent und Dependency Distance 71

3 Lýsis **75**
3.1 Zusätzliche Parameter des Experiments 75
3.2 Durchführung des Experiments 76
3.2.1 Korpusanalyse durch alle obengenannte Maße 77
3.2.2 Ergebnisse 77
3.3 Überlegungen über die Ergebnisse 80
3.3.1 Zusammenfassung der eigenen Leistung dem Forschungsbedürfnis 82
3.4 Schluss 83

4 Anhang **a**

4.0.1 Tafel der Propps Funktionen von 16 Grimms Märchen . . . . . a

Literaturverzeichnis . . . . . e

1 Deíxis

1.1 Einleitung und Problemstellung

Die Grundidee dieser Masterarbeit findet seine Wurzel im Dezember 2013. Mein Bachelorstudium in Philosophie wurde noch nicht abgeschlossen, und seit kurzer Zeit hatte ich die Prüfung in Sprachphilosophie bestanden. Unter anderem stellte bei dieser Prüfung *Die Morphologie des Märchens* von Vladimir Propp einen Brennpunkt des Interesses dar. Ich wendete die Erzählfunktionen Propps auf ein chinesisches Märchen an und bemerkte, dass das Muster, welches bei den von Propp bearbeiteten russischen Erzählungen normalerweise vorkam (d.h. die Wiederholung der selben Funktionsketten, jeweils mit einigen wenigen Variationen), bei dem Fernostmärchen jedoch nicht erschien. Es stellte sich mir die Frage, warum dies nicht so war. Weitere ähnliche Überprüfungen mit anderen Märchen haben dasselbe Ergebnis gezeigt, und mich noch mehr davon überzeugt, dass es eine Korrelation zwischen der Erzählung und der Sprache entsteht, in der die Erzählung verfasst ist. Allgemeiner gesagt habe ich immer Neigungen gehabt, die mehr der Chomskys universellen Grammatik als der Sapir-Whorf Theorie[1] geähnelt haben. Ohne das Bedürfnis zu einem dieser Extremen sich religiös zu bekennen (und tatsächlich sind beide Theorien mittlerweile in Frage gestellt worden), kommt es natürlich vor, sich die folgenden Fragen zu stellen: entsteht eine Ähnlichkeit zwischen die syntaktische Struktur einer Sprache und der allgemeinen Ordnung, die die Entwicklung einer elementaren

[1] Die Sapir-Whorf Theorie, auch als 'Theorie des linguistischen Relativismus' gekannt, nimmt eine starke Korrelation zwischen der Sprache des Sprechers und seiner Denkweise, wobei die letzte direkt durch die erste beeinflusst wird. Laut Chomsky sind die Strukturen des Gedankens universell, und sie erzeugen die Sprache durch generative Zusammensetzungsregeln (Vgl. Paragraph 'Zusammensetzungsregeln').

Erzählung in derselben Sprache regelt? Falls ja, könnte die gleiche Struktur auch auf den anderen Ebenen der Sprache wieder gefunden werden? Diese und ähnliche Fragen, die sich durch den allgemeineren Begriff von *Selbstähnlichkeit* zusammenfassen lassen, verfolgen mich seit vier Jahren. Ein lustiger Vortrag des *Sci-Fi* Schriftstellers Kurt Vonnegut[2], trotz seines unwissenschaftlichen und fast parodistischen Geschmackes, hat er in mir den Wunsch wieder erweckt, sowohl eine feste und formelle Notation zu entwickeln, welche die Struktur einer Erzählung ohne Ambiguitäten darstellen kann, als auch ein Modell zu finden, das die echte Beziehungen zwischen Satzbau und Erzählstruktur hervorheben und vernünftig begründen kann. Genau wie ein Märchen und die Mehrheit aller Texte besteht auch diese Abhandlung aus drei Hauptteilen. Das erste einleitende Kapitel behandelt die Themen, die wie ein *roter Faden* die ganze Studie durchqueren, und bringt den Leser auf den aktuellen Stand der Forschung. In diesem Teil werden auch kritische Punkte hervorgehoben, die zu unseren Ansatz und womöglich zu unsere Lösung führen werden, was genau das Thema des zweiten Teils ist. In diesem werden die Neigungen, die im Vorwort aufgeworfen worden sind, endlich Sinn und Ergänzung finden. Welche sind die Formen der Erzählung? Inwiefern sind sie automatisch extrahierbar? Inwieweit sind die Strukturen der Erzählung ähnlich zu den Strukturen der Sätze, die die Erzählung selbst ausmachen? Welche quantitative linguistische Werkzeugen könnten für beide Ebene (Satz/Text) benutzt werden, um Selbstähnlichkeitsspuren zu entdecken? In welcher Weise könnte sich die eventuelle Erfindung dieser geheimen Strukturen auf die Begriffe von Information und *Selbstinformation* beziehen? Auf diese und andere implizite bzw. konsequente Fragen versucht man in dieser Abhandlung zu antworten, oder mindestens zu bestimmen, ob sie sinnvoll und gut formuliert sind. Schließlich lohnt es sich zu erwähnen, dass meine umfassende Literaturrecherche doch nicht ganz vollständig gewesen sein könnte, und deshalb sind mir einige wichtige Beiträge entwischt, noch andere sind mir andererseits wichtiger ausgesehen, als Sie in der Tat sind. Trotz allem bin ich genug davon überzeugt, dass die von mir vorgelegenen Ergebnisse nicht schon woanders erschienen sind.

[2] Es geht um seine berühmte *Shapes of the Stories*, worüber er die Masterarbeit schrieb, die kurze Zeit später abgelehnt wurde. Im Folgenden das Link des Videos:
https://www.youtube.com/watch?v=oP3c1h8v2ZQ

1.2 Selbstähnlichkeit: Rekursivität, Fraktale und der goldene Schnitt

Die Ontogenese ist eine kurze und schnelle Rekapitulation der Phylogenese, bedingt durch die physiologischen Funktionen der Vererbung (Fortpflanzung) und Anpassung (Ernährung).

Ernst Haeckel

Die Selbstähnlichkeit ist die Eigenschaft eines Systems, dessen Bestandteile der Gestalt seiner allgemeinen Struktur gleichen. Wenn die Gestalt der allgemeinen Struktur nicht ganz geglichen, sondern nur geähnelt wird, spricht man genauer gesagt von *Selbstaffinität*. Im Rahmen dieser Studie werden trotzdem diese Wörter oft ausgetauscht, und als Synonyme behandelt. Der Begriff von Selbstähnlichkeit ist unmittelbar mit der *Rekursion* verbunden, die das Hauptverfahren darstellt, um ein selbstähnliches System feststellen zu können. Ein geometrisches selbstähnliches System heißt Fraktal. In der Natur kommen fraktalische Gegenstände häufig vor, wie zum Beispiel der Blumenkohl. *Die Ontogenese rekapituliert die Phylogenese*, wie auch auch im vorangestellten Zitat ausführlicher erklärt ist, ist der Begriff der Selbstähnlichkeit zur Biologie angewendet. In der Kunst und überhaupt in der Architektur findet immer wieder ein Verhältnis statt, das *Aurea Sectio* (goldener Schnitt) genannt wird, welches die Elemente der Struktur in einer Verhältnis von 1.618 stellt. Die Aurea Sectio ist auch das Verhältnis zwischen jedem Element der Fibonacci-Folge mit seinem Vorherigen.

Die Polymerketten entwickeln sich gleichmäßig rekursiv aus einer Grundgestalt durch die Ausführung von Grundoperationen.

Jüngst hat sogar die neurologische Forschung aufgefunden, dass auch das Gehirn, und genauer gesagt der Aufbau der neuronalen Netzwerken, einem selbstähnlichen Modell folgt. Es ist gerade kurz gezeigt worden, inwiefern viele Beispiele der Selbstaffinität überall in der Natur auffindbar sind. Machen diese Gründe allein eine genügende Motivation

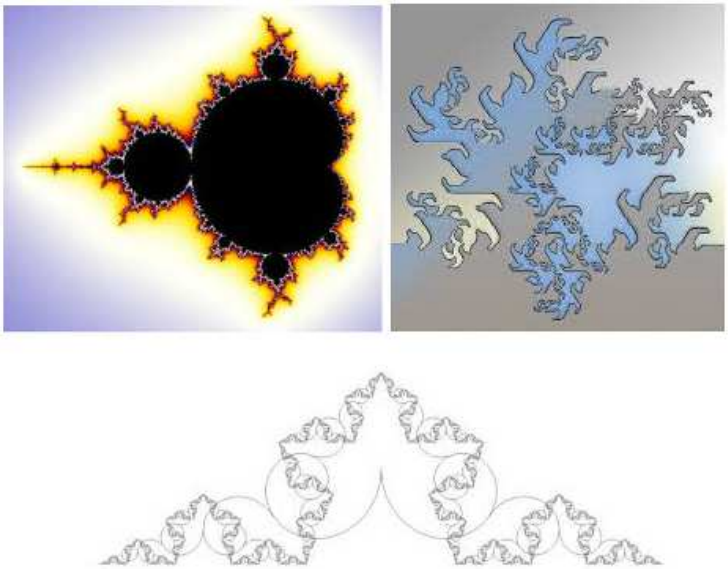


Abbildung 1.1: Die Mandelbrots- und Kochs Kurven gehören zu den berühmtesten Fraktalen.

aus, um vernünftig annehmen zu können, dass auch Texte Selbstähnlichkeitseigenschaften tragen? Welcher Anspruch führt uns zur Annahme, dass ein Text und seine zugrundeliegende Struktur besser verstanden werden kann, wenn man sie in diesem besonderen Modell einrahmt? Beachte man die Ähnlichkeiten der atomaren Struktur mit dem Planetensystem. Nicht jeder Schicht lässt sich gleichmäßig aufbauen. Die Erde selbst, Teil dieses Makro-Systems, ist zum Beispiel nicht wie ein Atom oder wie ein Planetensystem: sie scheint doch ähnlicher dem Kern des Atoms zu sein. Aus diesem und wahrscheinlich mehreren Gründen sollten wir uns nicht von den metaphorischen Vergleichen faszinieren und beeinflussen, sondern nur inspirieren lassen.

1.3 Der Text als teleologischer Gegenstand

Die Texte werden bei der Behandlung unseres Themas als Produkte angesehen, und nicht als rein fremde Objekte, wie es zum Beispiel im Rahmen der Physik geschieht. Das

bedeutet, dass die Texte ein teleologisches Prinzip enthalten, weil sie von Menschen für Menschen erzeugt werden, um einen Zweck zu erfüllen, welcher theoretisch entweder ästhetisch oder informativ sein kann, aber praktisch immer gewissermaßen beides ist, mit unterschiedlichen Proportionen. Selbstverständlich werden auch die ästhetischen Ziele durch die Information erreicht, genauer gesagt durch eine besondere Streuung der Information[3]. Bei diesem letzten Fall ergibt sich, mittels einer tieferen Analyse, die folgende Rekursion: *das ästhetische Ziel wird nur dann erreicht, wenn der Text gelingt, dieselbe Strukturen herauszufordern, die für sein eigenes Verständnis benutzt werden*[4].

Diese Einleitung besteht aus Beobachtungen, die jeder, der manchmal kritischer Weise gelesen oder geschrieben hat, sehr wahrscheinlich schon gemacht hat. Es handelt sich um die Natur des Literaturphänomens, für dessen Studium notwendig und unverzichtbar ist, einige Axiome anzunehmen (oder besser gesagt: zu definieren). Während die linguistischen Elementen und ihre Beziehungen aus einer reinen quantitativen Sicht analysierbar sind, könnte das gleiche jedoch auf der Ebene der Episoden nicht erlaubt werden. Während die Buchstaben und die grammatikalischen Regeln nicht subjektiv sind, und deshalb als natürliche Phänomene der Physik zu betrachtend sind (d.h. als Dinge, die außerhalb unser passieren), kann eine Wissenschaft der Erzählung (sei diese ein Roman, eine Geschichte, ein Bericht, ein Märchen, ein Mythos, oder irgendeine Art von realen, realistischen und unrealistischen Nebeneinanderstellungen von Ereignissen) auf den gleichen Bausteinen *nicht* anlehnen. Das von uns explorierten Universum (der Text) wurde aus einem intelligenten Wesen für einen bestimmten Zweck kreiert. Es ist besonders wichtig diesen impliziten Finalismus des Textes festzuhalten. Er hat einen Anfang, eine Entwicklung und ein Ende, und folgt bestimmten ästhetischen Regeln.

Obwohl sie vom Mensch erzeugt werden, und sich mit einigen Variationen nach einem grundsätzliche Gesetz ausdrücken, stellen sie auch *unbewusste Formen* dar. Die Bienen wissen auch nicht, wie man einen Bienenstock baut, oder besser gesagt, lernen sie es nicht. Darüber hinaus haben sie keine Kenntnisse über Geometrie, Architektur, Chemie,

[3]Diese Behauptung wird im Nachstehenden genauer bewiesen.

[4]Diese Behauptung ist nicht axiomatisch, obwohl sie intuitiv nachvollziehbar aussieht. Was sie wirklich bedeutet, wird im Laufen der These besser erläutert (Siehe Ende des Paragraphen Ërzählung als Schachspiel").

Statik und Fraktaltheorie, die aber nötig scheinen, um einen Bienenstock aufbauen zu können.

Im Gegensatz dazu war sich damals Phidias sicher des goldenen Schnittes bewusst, aber die Leute, die das Parthenon bewundernd beobachteten, wussten höchstwahrscheinlich nicht, dass diese Schönheit- und Harmoniegefühle tatsächlich aus einem mathematisch bestimmbaren Verhältnis stammen könnten.

Die ästhetische Schöpfung besteht deshalb daraus, dass sie eine gewöhnliche Darstellungsmethode verwendet, um tatsächlich intra-textuelle Verbindungen zwischen Ereignisse vorzuschlagen, die sich auch fern voneinander ergänzen und auslösen[5].

[5] Offensichtlich gilt diese Behauptung nur für die diachronische Kunstwerke. Auch in der Musik, wobei man nicht genau von Texten reden kann, gilt ein ähnlicher Prinzip.

1.3.1 Folge und Struktur I: Der Text als Vereinigung von Synchronie und Diachronie

Was durch einen Text (und in jeder sprachlichen Tat) versucht wird, ist durch eine Folge eine Strukturabbildung mitzuteilen. Unser Gehirn entziffert die Folge, um sie in einer strukturierten konzeptuellen Darstellung wiederaufzubauen. Es ist genau diese Eigenschaft der Selbstähnlichkeit, die uns zwingt, die Sprache (in diesem Fall die Texte) als ein selbstähnliches Phänomen zu untersuchen. Was bis jetzt ein Hindernis zur Suche der Selbstähnlichkeitseigenschaften in Texten verursacht hat, ist die Tatsache, dass ein Text nicht ohne Probleme als strukturierter Gegenstand zu betrachtend ist. Darüber hinaus ist ein Text schwierig in einer Gestalt darzustellen, die anders als der Text selbst ist: die Selbstähnlichkeit ist einfacher zu bemerken, wenn man mit einer Gestalt zu tun hat. Schließlich ist die Anwesenheit selbstähnlicher Eigenschaften in vielen Bereichen der Wissenschaft und in vielen Gebieten des Lebens keine genügende Motivation, um den gleichen Pfad in einem sprachwissenschaftlichen Kontext unternehmen zu können: die Forschung selbst, wenn nicht phänomenologisch durchgeführt, kann selbstverständlich gefälschte Ergebnisse liefern. Je tiefer man die Erzählung zerlegt, desto mehr nehmen die Elemente eine rein quantitative Konnotation an:

> La poétique définira nécessairement son trajet entre deux extremes, le trés particulier et le trop général. [48]

Man hat am Anfang behauptet, dass der Text ein diachronisches Mittel ist, um eine synchronische Darstellung im Empfänger hervorzurufen; andererseits wird der Text selbst als *Ergebnis* behandelt, d.h. als Resultat eines teleologischen Verfahrens. Etwas als Mittel und gleichzeitig als Ergebnis zu behandeln, ist kein Widerspruch in sich, überhaupt falls die Domäne unterschiedlich sind, wobei diese zwei verschiedenen Heuristiken angenommen werden. In unserem Fall stellt es eine wunderschöne Funktion der Sprache dar, die Roman Jakobson *metasprachlich* bezeichnet hat. Da die Bedingung der Selbstaffinität die *Selbstbezogenheit* ist, stellt die bereits bewiesene Selbstbezogenheit der Sprache ein Grund mehr dar, um die Selbstaffinitäten in ihr zu suchen.

1.3.2 Ascombes Quartett und Isoparametrische Texte: eine wissenschaftstheoretische Parenthese

Das Ascombes Quartett stellt ein typisches Beispiel dafür dar, dass Funktionen, die die gleiche Regression teilen, in der Tat grundverschieden sein können.

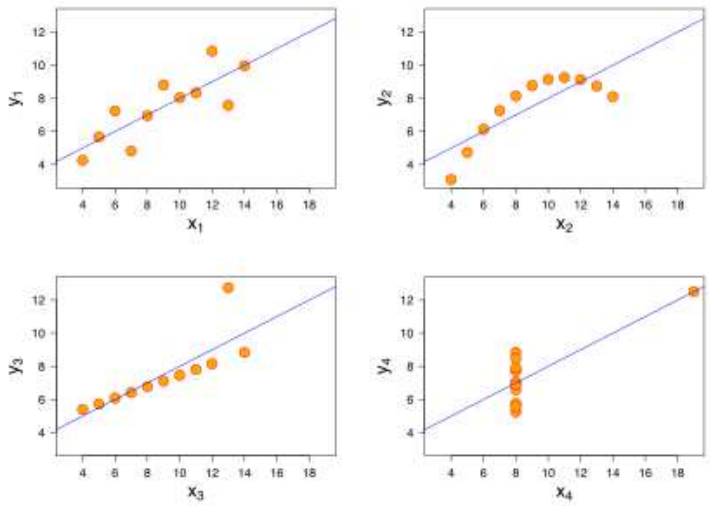


Abbildung 1.2: Ascombes Quartett (Quelle: Wikipedia.)

Das gleiche kann bei jedem Modell außer der Regression und bei jeder Datensammlung außer den Zahlen gelten, wie zum Beispiel bei Textkorpora: Es wäre fehlerhaft einer Menge von Parametern zu erlauben, einen Objekt zu definieren. Einige Texte, die gleiche oder sehr ähnliche Werte z.B. in Entropie, TTR [6], Verbabstände und Hapax-Prozentsatz teilen, könnten überhaupt nicht verglichen werden. Eine realistischere Analyse könnte sogar durch ihre Überlappung geschehen, um lexikalische Distanzen durch die Anwendung der Levensthein-Damerau Distanz auf der Topic-Liste, die zuerst durch *LDA* oder andere *Topic Modeling* Verfahren hergestellt worden sind. Schließlich lohnt es sich nun zu überlegen, dass sogar die von Propp bestimmten Erzählfunktionen nicht genug sind, um ein Märchen zu definieren: wenn ähnliche Eigenschaften bei der Erzählung eines histori-

[6] Die TTR (Type-Token Ratio) ist ein Maß der Polylexie, d.h. der lexikalische Reichtum eines Textes.

schen Ereignisses gefunden werden, bedeutet es nicht das das Geschehen märchenhaft, deshalb phantastisch (bzw. falsch) ist, sondern nur, dass die Funktionen mit dem *Inhalt* oder der Wahrhaftigkeit der Erzählung nichts zu tun haben, sondern nur mit der Formen der progressiven Verwirklichung unserer Gedanken.

1.3.3 Ausflug in den Fibonacci Zahlen

Das häufigste Beispiel, das man zur Erklärung des Begriffes von *Rekursion* bringt, ist zweifellos die Fibonacci Folge. Ich möchte gerne dieser Tradition folgen, sodass gezeigt werden kann, dass die Selbstähnlichkeit sich propagiert, wie eine Resonanz, auf allen den Ebenen, die die finale Struktur ausmachen. Diese Beobachtung beweist gleichzeitig auch, dass in einem semiotischen bzw. mathematischen System die Folge nur die oberflächliche Gestalt darstellt, woraus aber die Analyse der Beziehungen zwischen den Elementen ganz andere *Strukturen* vorausahnen lässt. Die Fibonacci Folge:

1 1 2 3 5 8 13 21 34 55 89 144
233 377 610 987 1597 2584 4181 6765 ...

stellt als Grundelemente zweimal '1' vor, und als zugrundeliegende Operation die Summe eines Elementes mit dem vorherigen, um das darauffolgende Element der Folge zu bestimmen. Wenn man diese Reihe in einer Matrize wieder ordnet, kann man sowohl interessante als auch elementare Beobachtungen durchführen.

Bustrophedon-Matrize In dieser Matrize-Variante wird die Zahl, die M(i, n-1) folgt[7], in M(i+1, n-1) und nicht in M(i+1, 1) dargestellt, wie es natürlich geschieht. Später wird man auch ein Beispiel mit ganz normalen Matrizen behandeln.

Beobachtung Die ungerade Reihe einer Bustrophedon-Matrize gleichen diejenige einer normalen Matrize. Die Ergebnisse, die rechts erscheinen, stellen nichts anders als

[7] Gleichmäßig der Konvention stellen eben hier 'i' und 'j' die Buchstaben dar, die die Position eines Elementes in der Matrize bezeichnen (Reihe und Spalte), wobei 'm' und 'n' die Zahl der Elementen jeder Reihe und Spalte der Matrize repräsentieren.

der Durchschnitt aller Multiplikationen [M(i, j)* M(i+1, j)] dar. Der Durchschnitt stellt keinen richtig zusammenfassenden Wert dar, da die Abweichung absinkt, je mehr man *nach unten* geht, wie in einer sichtbareren Weise bei der Matrize mit kleinerer Ordnung gezeigt wird.

Ordnung 6-Matrize

$$M = \begin{bmatrix} 1 & 1 & 2 & 3 & 5 & 8 \\ 144 & 89 & 55 & 34 & 21 & 13 \\ 233 & 377 & 610 & 987 & 1597 & 2584 \end{bmatrix} \tag{1.1}$$

$$R_j = \frac{1}{n-1} * \sum_{i=0}^{n-1} M(i,j) * M(i-1,j) \tag{1.2}$$

$$\frac{R_2}{R_1} = \frac{33557}{109} \simeq 307.86 \tag{1.3}$$

Ordnung 5- Matrize

$$M = \begin{bmatrix} 1 & 1 & 2 & 3 & 5 \\ 55 & 34 & 21 & 13 & 8 \\ 89 & 144 & 233 & 377 & 610 \\ 6765 & 4181 & 2584 & 1597 & 987 \end{bmatrix} \tag{1.4}$$

Durch die Verwendung von (1.2) bekommt man die Werte:

42, 4893, 602072

und durch die Verwendung von (1.3):

116.5, 123.04

Der zweite Ratio konvergiert nach 123, obwohl der Durchschnitt der obengenannten Ergebnissen 119.77 ist. Aus diesen Gründen wird im Nachstehenden nicht mehr der Durchschnitt genommen, sondern nur die Konvergenznummer.

Ordnung 4- Matrize

$$M = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 21 & 13 & 8 & 5 \\ 34 & 55 & 89 & 144 \\ 987 & 610 & 377 & 233 \\ 1597 & 2584 & 4181 & 6765 \end{bmatrix} \tag{1.5}$$

Durch die Verwendung von (1.2) bekommt man die Werte:

$$16.25,\ 715.25,\ 33553,\ 1576242.5$$

und durch die Verwendung von (1.3)

$$44.01,\ 46.91,\ 46.97$$

Die Konvergenz geht weiter nach 46.97.

Jetzt ist es schon möglich etwas zu zeigen, dass vielleicht in einer einfacheren und eleganteren Weise durch schärfere mathematische Werkzeuge vor allem allgemein gezeigt worden wäre. Die Relationen zwischen den gepaarten Elementen bleiben fest, und können auch durch eine Funktion zusammengefasst werden:

$$\frac{46.97}{123} \approx \frac{123}{307.86} \approx 0.38$$

Im Allgemeinem:

$$x(z) = \frac{[x(z-1)]^2}{[x(z-2)]} \tag{1.6}$$

$\forall i > 1 \wedge i \in N$,
wobei *x(z)* das R_j in 1.2 entspricht.
In unserem Beispiel:

$$\frac{33353^2}{715.25} \simeq 1576242.5 \tag{1.7}$$

Normale Matrize Diesmal stellen wir die Zahlen in der gewöhnlichen Reihenfolge dar. Zwischen den so identifizierten Paaren führen wir diesmal eine Division durch. Die (1.2) wird:

$$R_j = \frac{1}{n-1} * \sum_{i=0}^{n-1} \frac{M(i,j)}{M(i-1,j)} \tag{1.8}$$

Ordnung 6-Matrize

$$M = \begin{bmatrix} 1 & 1 & 2 & 3 & 5 & 8 \\ 13 & 21 & 34 & 55 & 89 & 144 \\ 233 & 377 & 610 & 987 & 1597 & 2584 \end{bmatrix} \tag{1.9}$$

R_j tendiert nach 17.94.

Ordnung 5-Matrize

$$M = \begin{bmatrix} 1 & 1 & 2 & 3 & 5 \\ 8 & 13 & 21 & 34 & 55 \\ 89 & 144 & 233 & 377 & 610 \\ 987 & 1597 & 2584 & 4181 & 6765 \end{bmatrix} \tag{1.10}$$

R_j tendiert nach 11.09.

Ordnung 4-Matrize

$$M = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 5 & 8 & 13 & 21 \\ 34 & 55 & 89 & 144 \\ 233 & 377 & 610 & 987 \\ 1597 & 2584 & 4181 & 6765 \end{bmatrix} \tag{1.11}$$

R_j tendiert nach 6.854.

Beobachtung

$$6.854 = \sqrt{46.97}$$

$$11.09 = \sqrt{123}$$

$$17.6 = \sqrt{307.86}$$

d.h. die Werte, die durch die normale Matrize gefunden worden sind, gleichen dem quadratischen Wurzel der Werten, die durch die Methode des Bustrophedons gefunden worden sind.

Merkwürdig ist auch die Tatsache, dass
$\frac{6.854}{11.09} \simeq \frac{11.09}{17.33} = 0.618$,
das als ϕ^{-1} bekannt ist, wobei $\Phi = 1.618$.

Schlussfolgerungen Die Wörter sind keine Zahlen, und falls eine geniale Weise auffindbar wäre, um die erste in den zweiten umwandeln zu können, würde ohnehin das Problem bleiben, dass die Erzählfunktionen gleichermaßen den Wörtern direkt miteinander nicht vergleichbar sind. Mit einem Text kann man nicht derartige Operationen

durchführen, um etwas ähnliches nachzuweisen.

Der Sinn dieses Beweises ist deshalb nur zu zeigen, dass nicht ausschließlich in der Natur, ein selbstähnliches System bei den höheren Ebenen ähnliche Relationen beibehält, die bei den unteren Ebenen vorkommen (was bedeutet, dass eine Ebene so in einer Beziehung mit dem anderen sich stellt, dass ähnlich derjenigen erscheint, die zwischen seinen selben Elementen auch vorkommt. Nach der Theorie der universellen Grammatik ist die Syntax jeder Sprache aus rekursiv abgeleiteten Regeln ausgemacht, die die Formen unseres Gedanken darstellen. Da es bewiesen worden ist, dass die neuronale Netze rekursiv hergestellte Bäume sind, was die Theorie Chomskys stark fördert, wird dementsprechend auch die Hypothese unterstützt, die behauptet, dass die Struktur einer Rede bzw. einer Erzählung den gleichen grundsätzlichen Regeln folgt, die nötig zum Aufbau eines Satzes erscheinen.

Aus diesem Grund lassen wir bei unserem Ansatz nur quantitative Werte gelten, die die Beziehungen zwischen den Elementen (oder zwischen den Ebenen) in einer objektiven Art darstellen.

Die besonderen Regeln werden eine Untermenge des allgemeinen Gesetzes repräsentieren.

1.4 Transversale Konzeptualisierung

Da die Zahlen und die Sprachzeichen nicht die gleiche Natur teilen, kann es schwierig und ziemlich fehlerhaft resultieren, sie zur gleichen Kategorie irgendwie zusammenzuführen. Es gibt trotzdem zahlreiche Aspekte der Sprache, die sich durch quantitative Methode analysieren lassen. Das ist der Zweck der quantitativen Linguistik, die seit mehr als fünfzig Jahren wichtige Ergebnisse liefert. Die heutige Richtung der Forschung in diesem Feld scheint aber zu trocknet zu sein: die Mehrheit der Papers über das Thema bieten keinen synthetische Hinsicht, keine praktischen Anwendungsmöglichkeiten und vor allen Dingen keine Chance den Profanen zu gewährleisten, die behandelten Themen für sich selbst als prägnant überlegen zu können. Die gewinnenden Ansätze zur linguistischen Datenverarbeitung sind heutzutage diejenigen, denen es gelingt, mehrere Felder zu ver-

binden. Man kann unter anderen die Computerkreativität und die *Computererzähltheorie* erwähnen, die die Ebene der Sprache überwinden, um zu versuchen, *Gedankenprozesse* zu formalisieren. Die Sprache selbst wird hierbei als Ausdruck des Gedanken überlegt, und nicht als ein rein fremdes Objekt[3]. Bis jetzt hat man (vernünftig) geglaubt, dass um *auf wissenschaftlicher Weise* etwas untersuchen zu können, muss man einen Abstand halten, was bis jetzt nur dazu geführt hat, die Objekte zu neutralisieren und in Teilen zerlegen, die letztendlich mathematisch untersuchbar sind. Die Tatsache bleibt, dass die Benutzung von Heuristiken, Verallgemeinerungen und Modellen, die immer nützlich aber nie wahrhaftig sind, eine Verlangsamung der wissenschaftlichen Fortschritte verursacht haben, obgleich es scheint, wegen der Anzahl von Ergebnissen die nacheinander schnell veröffentlicht werden, dass das Gegenteil gilt. Die transversale Konzeptualisierung [8] kann andererseits viele Fehler verursachen, überhaupt falls sie zu unvernünftig und phantasievoll wird, wohingegen sie schneller und besser zu wichtigen Ergebnissen führen kann.

> *As a research tool, this analogy poses the following questions: 1. what is the narrative equivalent for a fibre, and what is the corresponding process of heckling 2. what is the narrative equivalent for a yarn, and what is the corresponding process of spinning 3. what is the narrative equivalent for a strand, and what is the corresponding process of twisting.* [27]

Wie schon früher hervorgehoben worden ist, trotz ihrer zahlreichen Vorteilen stellt die *Analogie* ein nicht zu unterschätzendes Problem dar: wie könnte man mit Sicherheit behaupten, dass die *Vorsorgemaßnahme*, die man verwendet hat, formell korrekt und kohärent sind, um Begriffe und Methode zwischen unterschiedlichen wissenschaftlichen Gebieten umzusetzen? Zum Glück existieren schon Versuche, dieses Problem zu lösen. Das prägnanteste wurde von Tarek Besold und Enric Plaza durch den Begriff von *Amalgam* vorgeschlagen [22] [9].

[8] die Verwendung von Modellen in einem bestimmten Feld, die normalerweise in anderen (manchmal grundverschiedenen) Feldern benutzt werden.

[9] Grundsätzlich werden bei diesem Ansatz die Ontologien verschmolzen, zu denen die zwei Felder gehö-

Man kann diese Methode besser durch zwei Beispiele erläutern. Zuerst überlege man das Böhm-Jacopini Theorem, und inwieweit es im Rahmen der Computererzähltheorie angewendet werden kann.

Das Böhm- Jacopini Theorem: die natürliche Sprache als Programmierungsprache Das Böhm- Jacopini Theorem, wichtiges Ergebnis der Programmierungstheorie, behauptet, dass alles durch nur drei Kontrollstrukturen programmierbar ist:

- Sequenz
- Selektion
- Wiederholung

Die *Sequenz* ist eine Folge von Durchführungen konsekutiver Subprogramme. Die *Selektion* ermöglicht nur das Subprogramm durchzuführen, das eine bestimmte Bedingung erfüllt, die ihrerseits durch einen boolschen Operator gezeigt wird. Die *Wiederholung* (oder *Schleife*) ist die Durchführung derselben Operation, bis wann eine bestimmte Bedingung nicht mehr erfüllt werden kann.

Ein derartiger Lehrsatz gleicht in der Erzähltheorie der Behauptung von Vladimir Propp,

> *Aber genau so wie alles, was lebt, erzeugt das Märchen nur was ähnlich zu sich selbst ist. Wenn eine Zelle des Oragnismus des Märchens ein kleines Märchen in ihm wird, bestätigt es sich, wie später besser gezeigt wird, nach den gleichen Regeln eines Zaubermärchens.*[2]

Wie die hohe Niveau-Programmierung überhaupt durch komplexe Funktionen fortschreitet, die schon früher aufgebaut worden sind, gleichermaßen verhaltet sich die moderne Literatur, dank seines mythologischen und märchenhaften Substrates, die wir

ren. Durch die gemeinsame Notation der *Description Logic* erfolgt die Umsetzung in eine Domäne, wo Widersprüche und Fehler eindeutig erkennbar und deshalb einfacher korrigierbar sind.

schon in uns vermengen[10]. Die Grundoperationen sind aber auch in diesem Fall, wie bei der Programmierung, die gleiche, die auch die *Grundprogramme* charakterisieren, und sie heißen *Erzählfunktionen*.

Der folgende Paragraph nimmt als Vergleichselement nicht eine Regel, sondern eine Umgebung, die dazugehörigen Variablen und Regeln: das Schachspiel.

1.4.1 Erzählung als Schachspiel

Das Schachspiel besteht aus einer formalisierten Menge von Regeln und Elementen, die aus den Ereignissen hervorbrechen, die eine erzähltheoretisch- konnotierte Interpretation bekommen können.

- Weiße Stücke: Charakter (nach der Erwartung des Lesers);
- schwarze Stücke: (tatsächlicher) Charakter;
- Schachbrett: Länge des Textes , Zeit und Raum der Erzählung;
- Regeln: Ontologie[11], Ziele;
- Bedrohungen: Spannungen;
- Entscheidende Züge: Anfang des Konfliktes, Ende des Konfliktes, Ende der Episode, Anfang der darauffolgenden Episode.

[10] Diese Behauptung könnte aus der Psychologie-Lehre von Carl Gustav Jung über die Archetypen abgeleitet werden, aber um praktisch zu sein, reiche es zu erwähnen, dass Mythen und Märchen, sind kürzer, einfacher, bedeutungsvoller und mehrmals im Leben erfahren. Die Kinder selbst hören oder lesen zuerst Geschichte die märchenhaft oder mythologisch sind. Nicht nur in anderen Büchern (Romanen usw.) sondern auch im alltäglichen Leben (Werbungen, Verweise, Sprichwörter usw.) kommen Bezüge auf dem Mythos so häufig vor, dass sie unvermeidbarer Weise in der Person sich wurzeln (siehe Seite 23).

[11] mit *Ontologie* versteht man die Klasse der logischen Regeln, die das ganze Universum (in diesem Fall der Erzählung) beeinflussen. Zum Beispiel kann ein Muggel im *Harry Potter*s Universum keine Zauberei machen, was ein Regel der Ontologie darstellt.

Gleichermaßen dem Schachspiel, könnte man die Narration als ein Bericht überlegen, der von einem einsamen Spieler verfasst worden ist. Auch bei den Fällen, wo eine Veröffentlichung nicht vorausgesetzt ist, schreibt der Schriftsteller stetig mit der Annahme, dass jemand seine Schrift lesen wird. Durch die Erzählung möchte man nicht eine Information vermitteln, sondern ein ästhetisches Gefühl hervorrufen.

Darüber hinaus geschieht es bei der Erzählungen, wie ausführlicher später nachgewiesen wird, dass am Anfang, in der deklarativen Phase, die Ereignisse kurzer und langsamer sind; keine Konflikte entstehen. In der zweiten Phase beobachtet man ein wichtiges Materialverlust, eine Vermehrung der Möglichkeiten und der Konflikten und, da mehrere Raum verfügbar ist, ist dementsprechend auch die Flugbahn der *Züge* breiter, was bedeutet, dass die verschiedene Situationen im Spiel miteinander stärker verbunden sind. Bei der Finale gibt es andererseits einen kleineres Materialverlust (die Konflikte sind fast alle aufgelöst), aber die Flugbahn der Züge ist noch länger, da noch mehrere *Plätze* verfügbar sind. Die starke Korrelation zwischen Raum und Anzahl von Elemente, die im Schachspiel entsteht, entspricht bei der Erzählung die Geschwindigkeit der Tätigkeit: je mehrere Konflikte, desto kürzere und langsamere Aufhebungen der Konflikten es gibt. Alle diese Eigenschaften bezeichnen auch das Schachspiel, und in beiden Fällen, ist es möglich durch eine quantitative Analyse dies zu beweisen, entweder mithilfe einer Schachdatenbank oder eines Textkorpus. Diese Korrelation zwischen Anzahl von Stücken, verfügbarem Raum und Flugbahn der Züge im Schachspiel entspricht in den Erzählungen der Korrelation zwischen Anzahl der Charaktere, Größe des Universums der Erzählung und Maße des Zusammenhangs der unterschiedlichen Situationen bzw. Episoden, die im Laufen einer Erzählung vorkommen.

Außerdem bietet das Schachspiel eine eindeutige Notation an, die nur aus zwei Elementen besteht: Stück und Position. Eine zusätzliche Bezeichnung lässt uns verstehen, ob ein Zug einen Schach oder einen Schlag verursacht hat. Aus dieser knappen Bezeichnung kann man jeder Art von möglichen Erzählungen ableiten, die mit dem genannten Charakter und Bewegungen kohärent sind.

> *The black queen was four squares north of the centre of the board. The third black pawn was to the right. (...) The black queen saw the third black pawn*

> *leaving to the right. (...) Three days later, the black queen moved southeast. The third white pawn remained behind. (..) The black queen saw the white queen appearing ahead. The black queen attacked the white queen. A month earlier three squares northwest, the white queen was three squares south of the centre of the board. (...) The white queen saw the black queen arriving. The black queen attacked the white queen. The white queen died. The black queen saw the white right bishop arriving. The white right bishop attacked the black queen. The black queen died.*[27]

Um uns kurz von diesem faszinierendem Vergleich zu trennen, wäre es vernünftig letztendlich sich zu fragen: was für ein Spiel ist somit die Erzählung? *Die Erzählung ist ein Nullsummenspiel*[12]*, das nur aus einer Person gespielt wird (Widerspruch: ein derartiges Spiel setzt normalerweise nur zwei Spieler voraus!), dessen Inhalt annotiert/aufgenommen wird, um wieder der Aufmerksamkeit eines Zuschauers/Lesers vorgeschlagen zu werden.* Ziel des schizophrenen Schriftstellers ist es, eine Schizophrenie (oder irgendwelche andere Art von geistlichen Änderungen) auch im Leser zu verursachen. Der Leser möchte verstehen und nicht verstehen, gleichfalls möchte der Schriftsteller verstanden und nicht verstanden werden[13]. Diese und andere Widersprüche, die jedes Definitionsversuch der Narration charakterisieren, werden durch die Selbstähnlichkeit aufgehoben. Der implizite Widerspruch dient dem gleichen Zweck der Selbstähnlichkeit, d. h. der Rückzug des Objektes auf sich selbst, die Vereinigung von *Form* und *Inhalt* als Hauptmittel, um das ästhetische Gefühl zu erreichen. Wie schon früher gesagt worden ist, *diese Ziel wird nur dann erreicht, wann der Text gelingt, dieselbe Strukturen herauszufordern, die für sein selbes Verständnis benutzt werden.*

[12] Ein Nullsummenspiel ist ein Spiel, wobei immer ein Verlierer für jeden Sieger es gibt (Summe null). Normalerweise sind die Spieler zwei. Go und Schachspiel sind Beispiele für Nullsummenspiele, wie auch das Fußball (22 Spieler, aber 2 Teams).

[13] Dies bedeutet, dass ein Kunstwerk *nicht* als Problem wahrgenommen wird. Der Ansatz ist total anders. Es bleibt ja ein leichtes Konflikt zwischen Zuschauer und Kunstwerk, eine Auseinandersetzung, die nur zur Erfüllung des ästhetischen Zweckes zugewiesen ist. Genau in diesem Sinn möchte der Benutzer passiv bleiben, während und obgleich er aktiv seine Verständnis und Empfindlichkeit benutzt.

1.5 Eine (rückwärts) Geschichte des Problemstellung

Man hat bis jetzt die Grundbegriffe und die Elementen unserer Studie erläutert und man hat schon einige wichtige Definitionen vorgeschlagen und eigene Ideen entwickelt. Im Nachstehenden wird eine Geschichte des Problems stattfinden, wobei Ansätze und Methode unterschiedlicher Forscher gezeigt werden. Dieser knappe Überblick wird ab und zu durch Paragraphen unterbrochen, die zum Verständnis der danach erläuterten Ansätze nötig sind.

1.5.1 Selbstähnlichkeit in Texten

Trotz meiner intensiven Recherche ist es mir gelungen nur zwei Artikel zu finden, die explizit mein Thema behandeln. Aus diesem Grund finde ich es bedeutungsvoll einen Abschnitt beiden zu widmen, sodass später meine eigene Leistung besser geschätzt werden kann.

Fraktale Schönheit Im Paper *Fractal Beauty in Text* J. Cordeiro, P. R. M. Inácio, D. A. B. Fernandes nehmen als Maß der Selbstähnlichkeit in einem Text das Hurst Parameter. [14]

Für die Analyse wird ein gemischter, [15] englischer Textkorpus benutzt. Fünf Attribute werden bestimmt, die Bereiche von Wortlänge bezeichnen:

> *For the first principle we have considered six lexical features that are measured for a given block (amount) of text: the number of* (A_0) *non-words;* (A_1) *small words* $(|w| < 3)$*;* (A_2) *medium words* $(3 \leq jwj < 7)$*;* (A_3) *long words* $(|w| > 6)$*;* (A_4) *sentences; and* (A_5) *the lexical diversity.*

[19]

[14] Das Hurst Parameter wird als lange-Zeit-Gedächtnis- Maß von Zeitreihen benutzt, genauer gesagt von ihrer Selbstähnlichkeit. Eine ausführlichere Behandlung wird im Sektion 'Folge und Struktur III: Dependency Distance und Hurst-Exponent' stattfinden.

[15] d.h. ein Korpus, das aus unterschiedlichen literarischen Genren besteht.

Obwohl die Ergebnisse durch weitere Untersuchungen bestätigt werden sollen (empfehlenswert wäre die Nutzung von Textkorpora anderer Sprache, das Hinzufügen anderer Attributen, die Verwendung eines Selbstähnlichkeitsmaß unterschiedlich als das Hurst-Exponent), scheinen sie positiv zu sein.

Besonders wichtig für den Zweck unserer eigenen Forschung sind die Schlussfolgerungen, die am Ende abgeleitet werden.

> *The paper shows that there is fractal beauty in text.* It is naturally not a property that is consciously included by the authors of the text, which make the results even more surprising. They suggest that there might be a relation between what is perceived as text quality and the self-similarity degree[16], *although a more complete (more attributes) and exhaustive (more corpora) analysis is required.[19]*

Selbstähnlichkeit im Koran G. Lancioni, R. Villano und F. R. Romani sind die Authoren von *The Self-Similar Qur'an Project: Automatic Detection of Internal Similarity in Classical Arabic Texts.* Obwohl man von genügenden Gründen verfügt, auf die Jagd der Selbstähnlichkeit in Texten zu gehen, ist in diesem Artikel der Grund durch das Koran selbst begründet: im Koran wird tatsächlich behauptet, dass der Koran selbst selbstähnlich ist.

> *One historical approach to the defining features of the Qur'anic text was prompted by the text itself, which in a couple of passages calls itself mutašābih. This expression, sometimes interpreted as ambiguous[...] means literally ßimilar to itself": in other passages where the same participle mutašābih or other derivatives from the same root are present, interpreters are unanimous in reading it as "(self-)similar", which is also the reading found in some early exegetes.[16]*

[16]Meine Emphase.

Um die Distanz zu evaluieren, wird der Abstand [17] zwischen jeder Verse (V_j) mit allen anderen ($V_{j \neq i}$) berechnet. Um das Problem zu beheben, das aus dem Vergleich von Versen unterschiedlicher Länge vorkommt, benutzt man die Gleichung:

$$\frac{lev_{V_i,V_j} L(V_i), L(V_j)}{max(L(V_i), (L(V_j)) - |L(V_i) - L(V_j)|} \tag{1.12}$$

, die im besten Fall dazu führt, dass der Vergleich zu einem Wert, dass der absoluten Differenz der Versen entspricht, und im schlechtesten Fall (wann die verglichene Versen total unterschiedlich sind) die Länge des längeren Verses. Hierdurch ein *Mapping* NxN der Verse wird als eine dreieckige Matrize dargestellt, deren Werte im Bereich (0,1) liegen.

Die Ergebnisse sind sehr interessant:

> *They show that the number of similar couples at the threshold 0.4 is much higher (about 4 times) in the Qur'an than in the Diatessaron, a preliminary demonstration of the pertinence of self-similarity in the internal organization of the Qur'an also in relation to texts that are quite similar from other points of view.* [16]

[17] Diese Abstand wird durch die bekannte *Damerau-Levenshtein Distanz* berechnet, derer Wert die minimale Anzahl von Operationen ist, die nötig sind, ein Satz in einen anderen zu verändern. Die Operationen sind: *Einfügung, Löschung, Ersatz* und *Umsetzung* zwei anliegender Wörter. Diese Maß wird normalerweise auf der Ebene des Wortes verwendet, d.h. die oben genannte Operationen betreffen die Buchstaben, nicht die Wörter.

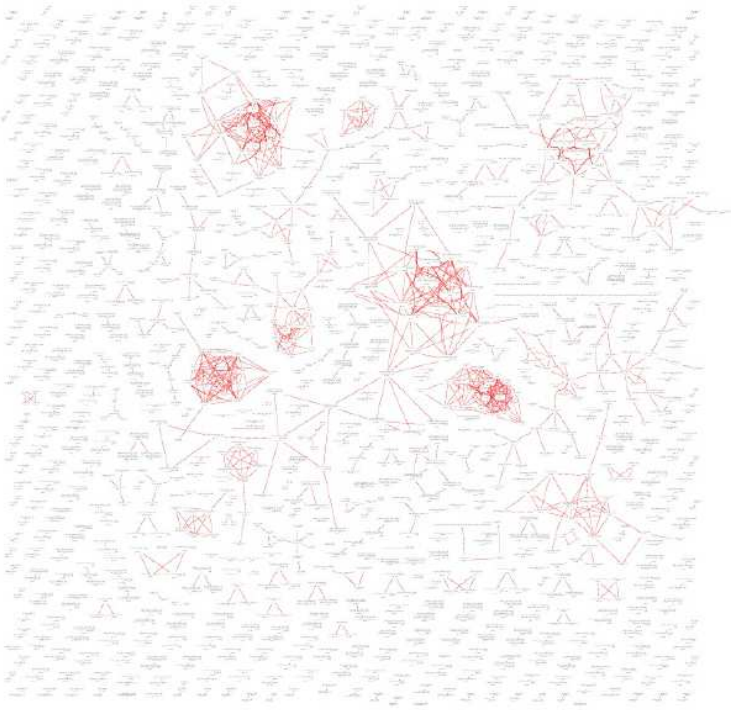


Figure 1: The self-similarity graph for the Qur'an with distance threshold=0.4.

[16]

Die Evolution der Erzählung

Wichtig zu merken ist auch, dass wenn die Grundregeln in unserem Gehirn befestigt sind, fühlt der Leser die Not (wie es für jede Art von Spielen geschieht) nach ein komplexeres bzw. stimulierenderes *Spiel* zu überspringen. Von daher könnte man fast behaupten, dass eine gute Frage, mit der ein kreatives Verfahren beginnen kann, lautet: *Welche Umstände, die von allen als unmöglich irgendwelcher weiteren Entwicklung jenseits der Mittelmäßigkeit überlegt werden, könnten doch überraschenderweise entwickelt werden?* Beispielsweise erscheinen alle rhetorische Stilmittel nichts anders als ein Umkippen der natürlichen/gewöhnlichen Ordnung der Elemente eines Satzes, entweder in einem syntaktischen oder in einem semantischen Sinn. Tragödie und Komödie sind die Hauptderivaten des Überraschenden. Obwohl man die Schönheit nicht zur Überraschung zwingen kann, weil die Überraschung auch ohne Schönheit entstehen kann, gibt

es andererseits keine Schönheit ohne ein gewisses Maß von Überraschung (zumindest für die Künste, die sich in einem diachronischen Weise ausdrücken, wie die Literatur und die Kinokunst). Aus diesen Gründen kann man in einem formalen Modell, das auf der Selbstinformation begründet ist, die Überraschung mit der Schönheit (oder um es besser umzuformulieren: *ästhetische Wirksamkeit*) zweifellos verbinden.

Ganz interessant anzumerken ist, dass bezüglich der Komplexität auch die Länge der Narration steigt. Dasselbe Verhältnis zwischen Länge des Textes und Größe des Schachbrettes hat man schon im Abschnitt *Erzählung als Schachspiel* markiert. Um Märchen zu genießen, benötigt man nur den Verstand von (durch elementare Wörter ausgedrückten) Grundbegriffen, wohingegen das Verständnis komplizierterer Erzählungen auch schwieriger wird, da die Menge allgemeines Wissens dementsprechend höher ist.

$$Mythos \downarrow$$

$$Epos \Rightarrow Märchen \Rightarrow Erzählung \Rightarrow Novelle \Rightarrow Roman \Rightarrow ErgodischeLiteratur$$

Zeit und Räume verkomplizieren sich auch durch die Verwendung von Analepsen, Prolepsen, Zeitreisen, andere Welten, andere Dimensionen, usw. im Laufen der Evolution der literarischen Genren.

Irgendwie liegt das *Exploit* der Filmkunst bzw. der Karikaturen Sammlung auch in der Fähigkeit des Benutzers, innerhalb der Zeiteinheit einen größeren Informationsinhalt zu verbrauchen, mit dem Ergebnis, dass um den *ästhetischen Orgasmus* zu erreichen, man maximal 3 Stunde braucht, und nicht minimal 3 Tage, wie es zum Beispiel mit einem Roman passieren kann.

Die Korrelation zwischen Länge und Informationsinhalt [18] repräsentiert selbstverständlich keine Äquivalenz. Der vorgeschlagene Entwurf soll nur dazu dienen, einen Ansatz vorzuschlagen, der tragbar für Textkorpora ist, die nicht nur aus mehr, sondern auch aus weniger strukturierten Texten bestehen. Großer Teil unserer Mühe soll in dieser Richtung gehen, um zukünftig eine Erweiterung und nicht wieder einen Ersatz dieses Modelles zu verursachen. Im Rahmen der *Information Retrieval* wurde es schon

[18] Vgl. [35]

festgestellt, dass einige Modellen, die aufgrund ihrer Einfachheit in sehr begrenzten und verallgemeinerten Fällen angewendet worden sind, sich nicht effizient mit größeren Datenmengen später verratet haben.

Definition von *Surprisal* und *Kullback-Leibler Divergenz* Das *Surprisal* bzw. die Überraschung oder *Selbstinformation* ist eine Maßeinheit der Informationstheorie, derer Ziel die Messung der Überraschung eines Ereignisses ist, das in einem bestimmten Kontext gesetzt ist. Je erwartbarer ein Ereignis ist, desto kleiner wird sein Überraschungswert. Offensichtlich ist das Surprisal invers proportional der Wahrscheinlichkeit, mit der ein Ereignis vorkommt. Deshalb könnte man in einer vereinfachter Weise schreiben:

$$wow_i = log_2 \frac{1}{(p_i|Kontext)} [26]$$

Diese Abhängigkeit des Erwartungswertes eines Ereignisses mit dem vorherigen Modell setzt das Surprisal im Rahmen der bedingten Wahrscheinlichkeit.

> In this framework, the new data observation D carries no surprise if it leaves the observer beliefs unaffected, that is, if the posterior is identical to the prior; conversely, D is surprising if the posterior distribution resulting from observing D significantly differs from the prior distribution. Therefore we formally measure surprise elicited by data as some distance measure between the posterior and prior distributions. This is best done using the relative entropy or Kullback-Leibler (KL) divergence [29]. Thus, surprise is defined by the average of the log-odd ratio:

$$KL(P(M|D), P(M)) = \int P(M|D) log \frac{P(M|D)}{P(M)} dM \tag{1.13}$$

[33]

Was uns verhindert, im Rahmen dieses besonderen Traktates, einen statistischen Ansatz anzunehmen, ist die Schwierigkeit der Vermessung dieser *Wahrscheinlichkeiten*: soll das Wort die Wahrscheinlichkeit betragen, die in einem kanonischen Textkorpus bekommen hätte, oder die Wahrscheinlichkeit, die nur im Rahmen aller Märchen der Gebrüder

Grimm bekommen würde? Diese zwei sind nicht die einzigen verfügbaren Möglichkeiten; aus diesem Grund wird man bei unserem (kleinen) Korpus und bei unserem Experiment auf diesem Ansatz verzichten. Wie schon erwähnt, und wie später nochmals besser festgestellt wird, sollen die Daten der Märchen ein bedeutendes Basis repräsentieren können, um sinnvolle Untersuchungen über längere Erzählungen durchzuführen.

1.5.2 Noch eine Überlegung über die Folge

Um den Begriff von *Selbstinformation* praktisch visualisieren zu können, benutzt man wieder das Beispiel der Zahlenfolge.

Gegeben sei die Serie

1 3 5 7

Man könnte argumentieren, dass die Bedeutung der Serie so erscheint:

$$\forall i \geq 1, \quad S(i) = S_{(i-1)} + 2$$

Wenn man merkt, dass alle Zahlen der Folge Primzahlen sind, dann kann auch '11' ein gültiger Nachfolger darstellen. Stellt der Prüfer demnächst '9', dann ist die *Primzahltheorie* automatisch gelöscht, und die Option "+2"verstärkt sich. Man fügt weiter '16' hinzu, was dazu führt, unsere Annahmen wieder stark zu überlegen. '16' ist bis jetzt das Element der Folge mit der höchsten Selbstinformation. Was kann es im Rahmen der Literatur bedeuten? Auch wenn zwischen den Elementen keine numerische Beziehung entsteht, versteht man die Bedeutung einer Reihe von Wörter nur am Ende des Satzes, die Bedeutung eine Reihe von Sätze versteht man nur am Ende der *Episode*, und so weiter. Ohnehin ändert sich zwischen jedem Element unsere Meinung, unsere Annahme über welches das darauffolgende Element sein könnte. Es wäre natürlich hier zu widersprechen, dass das 'Raten welche das nachherige Ereignis ist' keine echte Aufgabe des Lesers ist, der tatsächlich sogar bei Krimis diese Annahme nicht bevorzugt. Trotzdem hat jeder Leser ein Gedächtnis, das immer, auch unwillkürlich benutzt wird.

Jede der oben genannten Texteinheiten sind Kohärenzdomänen. Man setzt auch voraus, dass eine Korrelation zwischen *wow*s (oder *shannons*, die Maßeinheit der Selbst-

information) und der neuronalen Tätigkeit besteht[19]. Wir möchten aber hervorheben, das nicht immer (tatsächlich fast nie) was Kompliziertes ist, auch schön ist. In diesem kleinem schriftlichem Beispiel:

> Der Mond scheint
> der Mond bleibt
> still sitze ich fein
> und schaue den Mond an.

habe ich durch die strukturelle Wiederholungen und Alliterationen der ersten drei Versen eine Erwartung aufgebaut, die durch den letzten zwei Versen gestört worden ist. Die Spannung lässt sich durch die Banalisierung einer Struktur mittels ihrer Iteration oder mittels der Adoption einer schon bekannten Struktur als 'Start-/Ausgangssituation'. Die Spannung löst sich dank ihres Hervorbrechens aus der Wiederholung nach den gleichen Methoden, die passend mit der herrschenden vorausgesetzten Ontologie ist, die genau durch die Startsituation aufgebaut wird. Dieser Bruch der Ordnung resultiert nicht störend und sinnlos, weil es schrittweise passiert. Der dritte Vers bereitet es schon vor: $bleibt \rightarrow fein \rightarrow an$ Darüber hinaus ist er zusammenhängend mit dem Rest des Gedichtes, obwohl er länger als den anderen Versen ist. Je größer der Raum ist, desto mehrere und schönere Effekte können kreiert werden, weil mehr Raum mehrere Möglichkeiten zur Erschaffung neuer Kontexte darstellt, und diesbezüglich ist die Bedeutung der schon geschriebenen Sätze *gesetzmäßig* änderbar. Es geht um den Unterschied zwischen *Hervorgehen* und *Hervorbrechen. Es ist wichtig den Zauberwürfel zu lösen, obwohl seine Zerlegung und Wiederzusammenbau das gleiche Ergebnis liefert.*

1.5.3 Die Universität Saarland

> Die Information ist Information, nicht Energie oder Materie.
>
> Norbert Wiener

[19] Vgl. [6]

Unserer Ausflug durch die Herangehensweisen an die Computererzähltheorie lässt man mit der Universität Saarland anfangen. Sogar bei der Universität Saarland gibt es eine Forschungsrichtung, die mehr mit unserer Abhandlung zu tun hätte, die auf dem Aufbauen von Ontologie-basierten Annotierungen der Märchen sich fokussiert.[18] [23]

Hoffentlich scheint die Verweisung an die Artikeln genug, um an dieser Stelle auf ihre Erklärung verzichten zu können. Was man hier zu behandeln bevorzugt, ist die Forschungslinie, die auf den vor kurzem erwähnten Begriffen der Informationstheorie sich basieren[6]. Das Grund dieser Wahl liegt darin, dass nur die Integration statistischer Ansätze mit dem von uns formalisierten Modell ein neuer und wesentlicher Beitrag darstellen kann, was im zweiten Teil der These ausführlicher erklärt wird.

Projekt A3: Surprisal, Scripts und Psycholinguistik

> 'Project A3 proposes to build a model of how expected or surprising an event is within a textual description of an event sequence, and will test whether textual references to expected events differ systematically from textual references to surprising events, in a way predicted by the uniform information density hypothesis. The key resource for estimating event sequence surprisal will consist of the probabilistic script automata developed in A2. The challenge then consists of developing an incremental alignment model that maps text spans of a discourse with appropriate events in script automata. While the project will begin with the narrations from a restricted set of domains, the longer term goal is to apply the model to natural texts. The development of this model will be informed by psycholinguistic experiments carried out within the project, as well as in A1.'

Uniform Information Density Unter der Annahme, dass die Kommunikation als Informationsübertragung über eine Leitung begrenzter Kapazität verstanden werden kann, resultiert eine optimale Strategie der Spracherzeugung, ein uniformer Informationsgrad nah dem höchste Kapazitätsgrenze zu halten. Nach dieser Theorie soll der Sprecher seine Äußerungen so planen, dass die Elemente mit höherer Information verlängert werden, und andererseits Elemente mit kleiner Information verkürzt werden.[26] Diese Hypothese, die durch mehrere Experimente bewiesen worden ist, ist darüber hinaus kohärent mit dem Köhlers synergetischen Modell der Sprache. [35] Im Abschnitt *Folge und Struktur*

II wird man eine zusätzliche interessante Anpassung dieser Theorie betrachtet, die eine engere Beziehung mit dem Thema dieser Forschung besitzt.

Scripts Ein anderer Ansatz, die sehr empirisch ist, ist durch die *Scripts* dargestellt.

> A script is "a standardized sequence of events that describes some stereotypical human activity such as going to a restaurant or visiting a doctor" (Barr and Feigenbaum, 1981). [43]

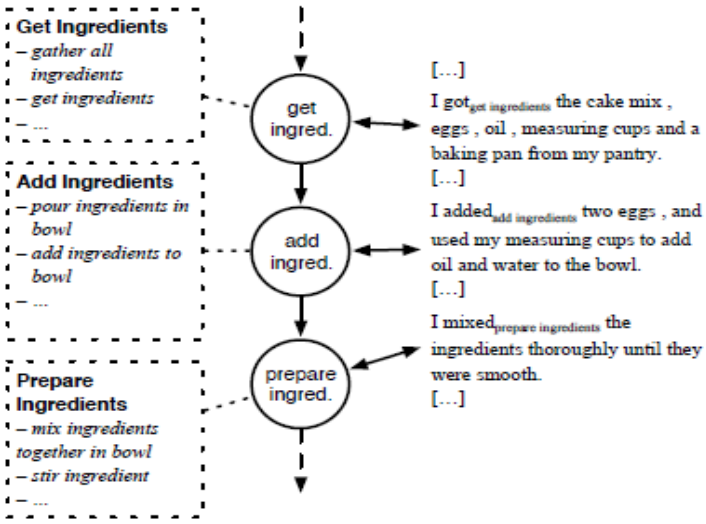


Figure 6: Connecting DeScript and InScript: an example from the BAKING A CAKE scenario (InScript participant annotation is omitted for better readability).

> Script events describe an action/activity along with the involved participants. For example, in the script describing A VISIT TO A RESTAURANT, typical events are ENTERING THE RESTAURANT, ORDERING FOOD or EATING. Participants in this scenario can include animate objects like the WAITER and the CUSTOMER, as well as inanimate objects such as CUTLERY or FOOD. [...] DeScript covers 40 scenarios, and also contains the 10 scenarios from In-Script. This corpus contains texts that describe scripts on

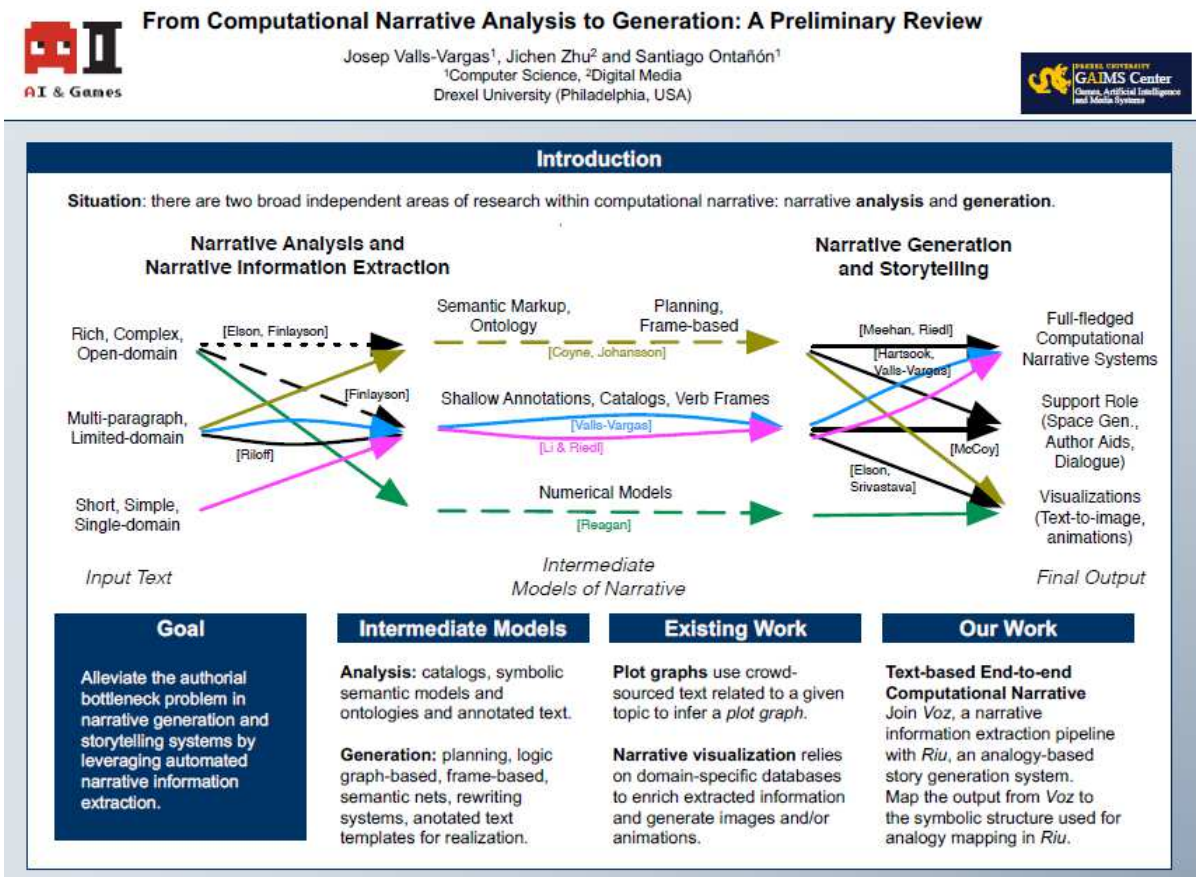
an abstract, generic level, while InScript contains instantiations of scripts in narrative texts. Script events in DeScript are described in a very simple, telegram-style language. [43]

1.5.4 Die Computererzähltheorie

Computational narratology is the study of narrative from the point of view of computation and information processing. It focuses on the algorithmic processes involved in creating and interpreting narratives, modeling narrative structure in terms of formal, computable representations. Its scope includes the approaches to storytelling in artificial intelligence systems and computer (and video) games, the automatic interpretation and generation of stories, and the exploration and testing of literary hypotheses through mining of narrative structure from corpora. [40]

Universität Drexel: Voz

Computational narrative systems usually require knowledge about the story world and narrative theory to be encoded in some form of structured knowledge representation formalism, a notoriously time-consuming task requiring expertise in both storytelling and knowledge engineering. [...] we present an approach that combines supervised machine learning with narrative domain knowledge toward automatically extracting such knowledge from natural language stories, focusing specifically on predicting Proppian narrative functions. Our experiments on a dataset of Russian fairy tales show that our system outperforms an informed baseline and that combining top-down narrative theory and bottom-up statistical models inferred from an annotated dataset increases prediction accuracy with respect to using them in isolation. [17]


From Computational Narrative Analysis to Generation: A Preliminary Review
Josep Valls-Vargas[1], Jichen Zhu[2] and Santiago Ontañón[1]
[1]Computer Science, [2]Digital Media
Drexel University (Philadelphia, USA)
AI & Games
GAIMS Center
Introduction
Situation: there are two broad independent areas of research within computational narrative: narrative analysis and generation.
Narrative Analysis and Narrative Information Extraction
Narrative Generation and Storytelling
Rich, Complex, Open-domain
Multi-paragraph, Limited-domain
Short, Simple, Single-domain
[Elson, Finlayson]
[Finlayson]
[Riloff]
Semantic Markup, Ontology
Planning, Frame-based
[Coyne, Johansson]
Shallow Annotations, Catalogs, Verb Frames
[Valls-Vargas]
[Li & Riedl]
Numerical Models
[Reagan]
[Meehan, Riedl]
[Hartsook, Valls-Vargas]
[McCoy]
[Elson, Srivastava]
Full-fledged Computational Narrative Systems
Support Role (Space Gen., Author Aids, Dialogue)
Visualizations (Text-to-image, animations)
Input Text
Intermediate Models of Narrative
Final Output
Goal
Alleviate the authorial bottleneck problem in narrative generation and storytelling systems by leveraging automated narrative information extraction.
Intermediate Models
Analysis: catalogs, symbolic semantic models and ontologies and annotated text.
Generation: planning, logic graph-based, frame-based, semantic nets, rewriting systems, anotated text templates for realization.
Existing Work
Plot graphs use crowd-sourced text related to a given topic to infer a plot graph.
Narrative visualization relies on domain-specific databases to enrich extracted information and generate images and/or animations.
Our Work
Text-based End-to-end Computational Narrative
Join Voz, a narrative information extraction pipeline with Riu, an analogy-based story generation system.
Map the output from Voz to the symbolic structure used for analogy mapping in Riu.

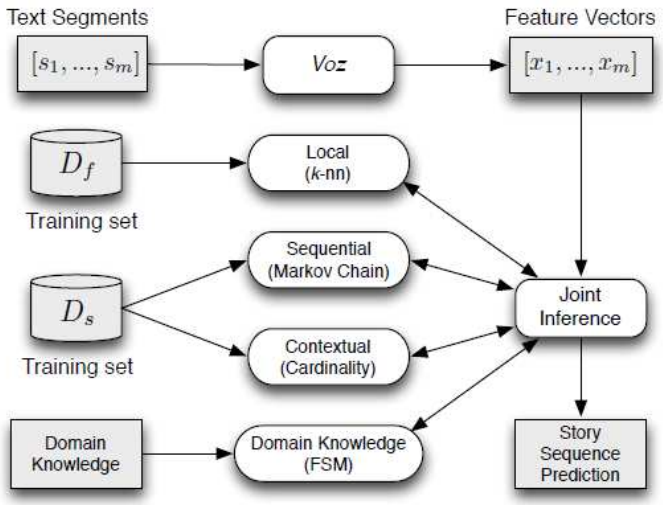


Figure 1: Overall system diagram. Both databases indicate automatically learned training sets from our annotated dataset. *Voz* is our previous narrative information extraction system which we use to automatically compute feature vectors from text segments in a given story. The output is a sequence of narrative function predictions such as: $\langle \alpha, \beta, \delta, A, B, C, \uparrow, H, I, K, \downarrow, W \rangle$ [17]

1.5.5 Finlayson

Vladimir Propp's Morphology of the Folktale is a seminal work in folkloristics and a compelling subject of computational study. I demonstrate a technique for learning Propp's functions from semantically annotated text. Fifteen folktales from Propp's corpus were annotated for semantic roles, co-reference, temporal structure, event sentiment, and dramatis personae. I derived a set of merge rules from descriptions given by Propp. These rules, when coupled with a modified version of the model merging learning framework, reproduce Propp's functions well. Three important function groups—namely A/a (villainy/lack), H/I (struggle and victory), and W (reward)—are identified with high accuracies. This is the first demonstration of a computational system

learning a real theory of narrative structure. [20]

1.5.6 Ein bündiger Standpunkt der Forschung

Der aktuelle Zustand der Forschung zeigt, dass letztendlich die Ebene der Ereignisse, welche in einem Text beschrieben bzw. *erzählt* werden, immer wesentlicheres Gewicht und mehr Aufmerksamkeit bekommt. Viele Forscher hatten die Neigung, das bedeutende Erbe von Vladimir Propp zu übernehmen, um seine Theorie mit den modernen Werkzeugen der Computerlinguistik zu untersuchen, und womöglich zu bestätigen. Der Fokus der Forschung scheint sich definitiv auf den Inhalt bewogen zu haben, der diesmal nicht durch abstrakte Kategorien behandelt wird, sondern mit einem stark empirischen Ansatz, der schon bedeutende Ergebnisse im Rahmen der *Computererzähltheorie*, der Computerkreativität und der *Interaktiven Storytelling* erreicht konnte.

Forschungsbedürfnis Was könnte man sich in diesem optimistischem Szenario noch mehr wünschen? Obwohl der Facettenreichtum der möglichen Anwendungen im Alltag viel hoffen lässt, bleibt noch ein nicht zu unterschätzendes Thema, das nicht vergessen werden darf: was Vladimir Propp mit seinem Werk meinte[20] ist, dass die Kompositionsregeln, die die Märchensgenerierung ermöglichen, mit dem Märchen *wenig* zu tun haben. Die Pflanzen, die Märchen, die Sätze, die Menschen und alles, was von den Menschen beobachtet und verstanden wird, folgt ähnlichen Regeln. Einmal, dass die *Information* als größter gemeinsamer Teiler zwischen Wörtern, Sätzen, und Ereignissen anerkannt worden ist, wird die Suche der Selbstähnlichkeit in Texten eine Notwendigkeit sein, um einer zusammenfassenden Theorie der Sprache verfügen zu können. Die Anwendungsmöglichkeiten dieses neuen Bereiches sollen uns deshalb nicht vergessen lassen, dass diese wesentliche Erweiterung dazu beitragen soll, eine *unifizierte Theorie der Sprache*[21] zu

[20] Was noch mehr durch die zahlreiche Zitate aus dem Goethes 'Die Metamorphose der Pflanzen' bewiesen wird, die sein ganzes Buch durchqueren.

[21] Gleichmäßig der Physik, wobei man gerade versucht, die Mikrophysik der Partikeln mit der Makrophysik der Körpern zu verbinden, (genauer gesagt versucht man die verschiedene Kräfte, die auf jeder Dimension wirken, unter einem allgemeinen Gesetzt zu vereinigen), geschieht es in der Geist-

erzielen.

1.6 Kurzer aber nötiger Ausflug in der Erzähltheorie

1.6.1 Strukturalismus und Poststrukturalismus

> *Die Form definiert sich im Gegensatz zu einer Materie, die ihr fremd ist; aber die Struktur [...] Was für irgendein Sprachsystem gilt, gilt noch viel mehr für die Mythen und Märchen, weil die Grammatik und das Lexikon in diesem Fall nicht nur eng miteinander verbunden sind, wie wohl sie auf unterschiedlichen Stufen wirken: sie haften mit ihrer ganzen Fläche ananeinder und decken sich vollständig. Im Unterschied zur Sprache, wo sich noch das Problem des Vokabulars stellt, enthält die Metasprache keine Ebene, aus der die Elemente nicht aus genau determinierten Operationen resultieren und noch Regeln ausgeführt werden. In diesem Sinn ist es hier alles Syntax.* [38]
>
> Claude Lévi-Strauss

Die philosophische Bewegung, auf die wir uns beziehen möchten, ist der Strukturalismus. Laut dieser Weltanschauung löst sich jede Dichotomie in einer Struktur, wo die Polarität zwischen Subjekt und Objekt eine theoretische Weise ist, um eine Einheit zu erreichen. Die Elemente, die die Struktur aufbauen, gewinnen seine Bedeutung nur innerhalb einer Struktur, das heißt in einem geordneten Netz von Beziehungen. Die strukturalistische Untersuchung strebt eine perfekte Durchdringung des linguistischen Stoffes und der Gestalt an, sodass kein Stoff ohne Gestalt existieren würde, und umgekehrt. Die Struktur ist den Inhalt selbst, den in einer logischen Organisation erfasst wird, die als Eigenschaft der Wirklichkeit begriffen wird.

wissenschaften, dass eine Vereinigung zwischen Linguistik und Narratologie durch den Begriff von Information möglich scheint.

1.6.2 Propp, Lévi-Strauss, Greimas

> *Der Begriff „Funktion“, so erläutert er in der Einleitung in das Werk von Marcel Mauss, soll in dieser Formulierung im mathematischen Sinne verstanden werden, nach dem Vorbild der Algebra. Der Ausdruck y = f(x), „y ist eine Funktion von x“, meint, dass es zwischen zwei Wertebereichen, dem der x-Werte und dem der y-Werte, eine bestimmte konstante Beziehung gibt. Die symbolische Funktion besteht demnach darin, dass zwischen verschiedenen Bereichen, etwa zwischen den Verwandtschaftsverhältnissen und den Wohnweisen, eine konstante Relation bestimmter Art herrscht und dass diese Beziehung durch das Symbolische hergestellt wird.* Jacques Lacan

Lévi-Strauss führte als erster den Begriff von *Beziehungensbündel* als großes aufbauende Element des Mythos ein. Solche wichtiges Vorwort kommt aus der Überlegung, dass die mythische Narration sowohl diachronisch als auch synchronisch ist. Nach der Zerlegung in *Mythemen*, kann der Mythos behandelt werden, um eine befriedigende Streuung in den verschiedenen Beziehungensbündeln zu erzielen. Die Mythemen werden in einer Aufstellung auf Fächer verteilt, in der die Reihen die diachronische Abwicklung der Erzählung darstellen, während die Spalten die Beziehungensbündel zeigen, die durch die Mythemen aufgebaut worden sind[37].

Dieses Verhältnis stellt auf der Ebene des Ereignisses die gleiche Bedeutung dar, die auf der Ebene des Satzes durch die syntagmatische und paradigmatische Achse bei der strukturalistichen Linguistik von De Saussure angedeutet worden sind. [22]

> Die *Mytheme* sind große konstitutive Einheiten, die auf der Ebene des Satzes zu suchend sind. Zerlegt man den Mythos in kürzen Sätze und überträgt sie dementsprechend auf einzelne Kanten, dann wird man bestimmte Funktionen unterscheiden können, und gleichzeitig feststellen, dass die Mytheme den Charakter von Relationen tragen (jede Funktion ist einem bestimmten Subjekt zugeordnet).[41]

Diese Idee, die Mythos-Einheiten in einer Matrize einzuordnen, wird später durch einen anderen wesentliche Begriff ergänzt, was überraschende Ergebnisse liefern wird.

[22] Vgl. [41]

Kadmos sucht seine von Zeus entführte Schwester Europa			
		Kadmos tötet den Drachen	
	Die Spartoi rotten sich gegenseitig aus		
			Labdakos (Vater von Laios) = »hinkend« (?)
	Ödipus erschlägt seinen Vater Laios		Laios (Vater von Ödipus) = »linkisch« (?)
		Ödipus bringt die Sphinx um	
Ödipus heiratet Jokaste, seine Mutter			Ödipus = »geschwollener Fuß« (?)
	Eteokles tötet seinen Bruder Polyneikes		
Antigone beerdigt Polyneikes, ihren Bruder, und übertritt das Verbot			

[37]

Vladimir Propp ist im Laufen unserer Rede mehrmals genannt worden, aber nie wurde eine offizielle Präsentation gemacht, ein Mangel, den nun man beheben will. Der russische Linguist der zwanzigsten Jahrhunderts hat durch seine 'Morphologie des Märchens' gezeigt, dass alle von ihm analysierten russischen Märchen durch eine gemeinsame Formalisierung sich beschreiben lassen. Die Anzahl der unterschiedlichen möglichen Erzählfunktionen, die in jedem Märchen vorkommen darf, ist nicht höher als 30. Eine Funktionsmenge kann wiederholt werden, mehrere Funktionen können überhaupt nicht vorkommen, aber die *Reihenfolge*, mit der die Funktionen nacheinander vorkommen, *darf nicht ändern*. Einige Jahrzehnte später wird A. J. Greimas die Idee von Propp an alle Arten von Narration anpassen. Das resultierende Schema ist offensichtlich einfacher und flexibler, bestehend aus weniger und allgemeineren Funktionen, welches *kanonische narratives Schema* genannt wird.

$\overline{p}\,\overline{A}\,\overline{C_1}\,\overline{C_2}\,\overline{C_3}\,\overline{p}\,A_1 p\,(A_2 + F_2 + \text{non}\,c_2)\,\text{d}\,\overline{\text{non}\,p_1}\;(F_1 + c_1 + \text{non}\,c_3)$
$\text{non}\,p_1 d\,F_1\,p_1\;(A_3 + F_3\,\text{non}\,c_1)\;C_2\,C_3\,A\;(\text{non}\;c_3)$

A = Vertrag (Befehl-Annahme) p = Anwesenheit
F = Kampf (Angriff – Sieg) d = rasche Verlagerung
C = Mitteilung (Absendung – Erhalt)

Die Verletzung eines Vertrages (zu Beginn des Konflikts) $\bar{A}$ besteht aus der paarweisen Funktion Verbot – Verletzung eines Verbots $\bar{a}$ *vs* (non $\bar{a}$), die in Korrelation steht zu der Funktion Abschluß eines Vertrages A (Befehl – Annahme a *vs* non a). Die Hochzeit löst endgültig den Konflikt und stellt den ursprünglichen Vertrag wieder her (der Überbringer gibt dem Empfänger, d. h. dem Subjekt, das Wunschobjekt). Die Symbole haben folgende Bedeutung:

A_1 Tätigkeit eines Vermittlers – Einsetzen der Gegenhandlung
A_2 erste Funktion des Schenkers – Reaktion des Helden
A_3 dem Helden wird in der letzten Probe eine Aufgabe übertragen.

Die negative Serie $\overline{C_1}\,\overline{C_2}\,\overline{C_3}$ zu Beginn des Schemas hat folgende Bedeutung: Erkundung – Herausgabe, hinterlistige Falle – Mithilfe, Tätigkeit eines Schädlings – Fehlelement; sie verteilt sich auf die Koordinaten: *Nachricht*, d. h. Frage – Antwort (1), *Kraft* (2, es handelt sich hier sozusagen um den Verlust der Heldenkraft) und *Wunschobjekt* (3, der Erwerb der Zarentochter bildet die Aufhebung des Fehlelements). Die positive Serie bilden die Glieder C_1, C_2, C_3. Kennzeichnung – Erkennen stehen in Korrelation zu dem Paar Erkundung – Herausgabe als Form einer Mitteilung (C_1 *vs* $\overline{C_1}$). Entlarvung – Transfiguration als Offenbarung der Kräfte des Helden stehen in Gegensatz zu hinterlistige Falle – Mithilfe (C_2 *vs* $\overline{C_2}$). Außerdem steht der Erwerb des Zaubergegenstands in Gegensatz zu dem Verlust der Heldenkraft, ausgedrückt durch die Funktion Mithilfe (non c_2 *vs* $\overline{\text{non}\,c_2}$). Der Tätigkeit eines Schädlings entspricht auf der positiven Seite Bestrafung eines Schädlings; das Fehlelement wird nicht nur durch direkte Liquidierung aufgehoben, sondern auch durch die Hochzeit, die den Helden entschädigt (C_3 *vs* $\overline{C_3}$).

Abbildung 1.3: Ein Beispiel der Greimas´ Funktionenbearbeitung.

1.7 Text- und Sprachmodelle

> *[...] die Grammatik keine Substrat der Sprache, sondern ihre Abstraktion darstellt, und es zweifelhaft erscheine, aus der Beschreibung der Märchensfunktionen ihren Archetypus abzuleiten.* Eleazar Meletinskij

1.7.1 Aristoteles

Bevor die Aristoteles' *Poetik* erwähnt wird, lohnt es sich hervorzuheben, dass das Schema, das Aristoteles benutzt, um den Begriff von *Kraft* zu analysieren, die theoretische Einteilung der Erzählung (Deíxis, Katastrophé, Lýsis) widerspiegelt, die später auch von Hjelmslev verwendet wird:

- *δύναμις* (potentieller Kraft, Fähigkeit ein Tat zu machen),
- *ενέργεια* (tatsächlicher Kraft, geübter Kraft),
- *μεταβολή* (Veränderung, Ergebnis des Kraftes).

Auch die epistemologische Dialektik von Hegel (These, Antithese, Synthese) scheint diese triadische Einteilung zu verkörpern. Man hat sich daran gewöhnt, zu glauben, dass eine Episode beendet ist, wenn die Zeit oder der Raum oder beides geändert wird, und die Aktion der Erzählung woanders stattfindet. Anders formuliert ist eine Episode eine Texteinheit, die auch eine Zeit- und Raumeinheit behält: der Aufbau einer Episode folgt der aristotelischen Poetik. Daraus kann man ableiten, dass die Episode auch die kleinste Texteinheit darstellt, in der die Eigenschaft der Selbstähnlichkeit beobachtet werden kann. In seiner *Poetik* behauptet Aristoteles, dass nicht die Episode, sondern die ganze Erzählung eine Raum- und Zeiteinheit berücksichtigen soll, was dadurch erklärt werden kann, dass die Erzählungen entweder Tragödien oder Komödien waren, die inszeniert und angeschaut werden sollten. Eine derartige Theorisierung kann jedoch davon abhängen, dass früher wenige materiellen Ressource es gab, um Szenen häufig zu ändern. Die

Situation sollte nur mittels der Charakter verstanden werden können, um Verwirrungen und Ambiguitäten zu vermeiden.

1.7.2 Hjelmslev

Der dänische Linguist Louis Trolle Hjelmslev, Vertreter des Strukturalismus, stellte seine theoretischen Werkzeuge *aristotelisch* auf, d. h. durch die *Zusammensetzung* weniger allgemeinen Begriffe, die die Mannigfaltigkeit der besonderen Fälle erzeugen.

Auf der *syntagmatischen* Achse wird das Wort mit den darauffolgenden Wörtern in Beziehung gesetzt (*relatio in presentia*), wohingegen auf der *paradigmatischen* Achse das Wort mit allen Wörtern bezogen wird, die es ersetzen können (*relatio in absentia*).

Unabhängig von der überlegten Achse, kann jedes Gliederpaar nur aus drei verschiedenen Funktionen verbunden werden:

- *Interdependenz* (K↔K): Funktion zwischen zwei Konstanten, in der die Anwesenheit eines Gliedes die Anwesenheit des anderen voraussetzt;
- *Feststellung* (K←V): Funktion zwischen einer Konstante und einer Variable, in der die Variable die Anwesenheit der Konstante bestimmt, aber nicht vice versa;
- *Konstellation* (V|V): Funktion zwischen zwei Variablen, in der die Anwesenheit jedes Termen die Anwesenheit des anderen nicht voraussetzt.[31].

Aus der Vereinigung dieser Klassifizierungen, kann man sechs unterschiedliche Funktionstypen bekommen:

- *Solidarität* ($K \sim K$): Interdependenz auf der syntagmatischen Achse;
- *Komplementarität* (K⊥K): Interdependenz auf der paradigmatischen Achse;
- *Selektion* (V→K): Feststellung auf der syntagmatischen Achse;
- *Spezifikation* ($V \vdash K$): Feststellung auf der paradigmatischen Achse;
- *Kombination* (V-V): Konstellation auf der syntagmatischen Achse;

- *Autonomie* ($V?\dagger V$): Konstellation auf der paradigmatischen Achse;[31].

Hjelmslev ist auch bekannt für seine Abstraktion, die die Sprache selbst aus vier möglichen Standpunkten sichtbar macht. Interessant ist, dass sie aus den Standpunkten zweier Grundbegriffe herkommen, die sich miteinander überlappen: Form und Substanz.

- Inhaltssubstanz;
- Inhaltsform;
- Ausdrücksubstanz;
- Ausdrückform[31].

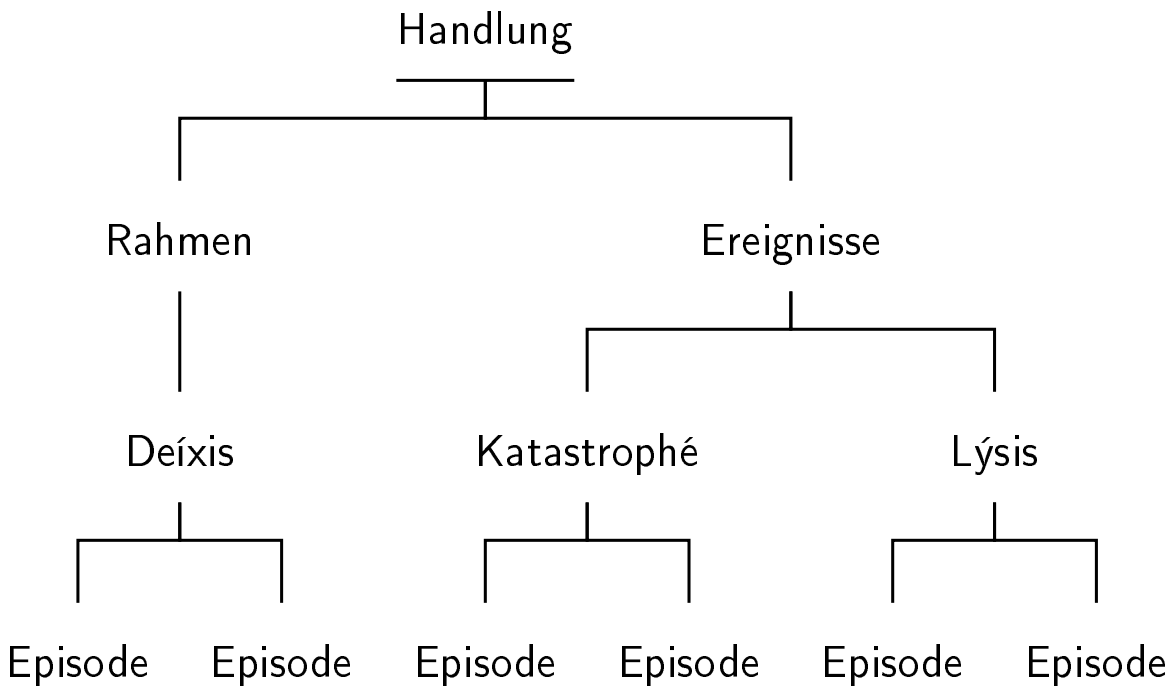


Abbildung des Van Dijks Modelles, als Vereinigung der Modelle von Hijelmslev und Aristoteles. [1]

Das Aristoteles-Hijelmslevs Modell wird sowohl in der Auflistung unserer Version der Propps Funktionen wieder verwendet, als auch im Paragraph "Widersprechende und ergänzende Paare".

1.7.3 Chomskys generatives Modell

Aristoteles und später Hjelmslev haben gezeigt, dass eine Erzählung in drei Bestandteile zerlegbar ist: *Deixis, Katastrophé, Lýsis.* Viele Texte unterschiedlicher Art können in diesem Kodierungsstil beschrieben werden. Beispielsweise besteht eine Programme aus Eingabe, Erklärung der Variablen, Schleifen, Verfahren und Ausgabe. Nicht nur im Rahmen der poetischen Schriften, sondern auch der Zeitschriften- und wissenschaftlichen Artikeln, wird diese *Triade* als Tatsache anerkannt.

Sogar im Bereich des Satzes geschieht dies durch die Gliederung in Subjekt, Prädikat, Objekt. Im Allgemeinen ist *drei* die kleinste Anzahl von Elementen, um aus einer Reihe eine *Folge* bzw. eine *Bedeutung* identifizieren zu können. Aus einem Dreieck kann man sich zahlreiche Dreieckstypen vorstellen, die als erste mögliche Figur der zweidimensionalen Welt alle anderen möglichen Gestalten vorausahnen (und teilweise aufbauen) lässt. Was hat bis jetzt Selbstähnlichkeit in der Sprache bedeuten können? Laut der Definition von Noam Chomsky lautet die Antwort: 'Die Bildung einer unendlichen Menge von Sätzen aus einem endlichen Inventar sprachlicher Elemente und Regeln'[23]. Chomsky kommt schon zum Schluss, dass die gleiche rekursiven Regeln, die die Struktur der Sprache abstrahieren, sogar in unserem Gehirn widergespiegelt sind. Diese holistische epistemologische Theorie, derer bedeutendster Befürworter Goethe selbst war, hat nur in den letzten Jahren wissenschaftliche Beweise gefunden[24].

[23] Vgl. [44]

[24] Am Anfang ist es gesagt worden, dass sowie die Theorie des Innatismus von Noam Chomsky als auch die Theorie der linguistischen Relativität von Sapir und Whorf in Frage gestellt worden sind. Trotzt ihres Antagonismus gibt es zahlreiche wissenschaftliche Ergebnisse, die die beide Theorie fördern. Dies kann bedeuten, dass beide Theorien gewissermaßen nicht widersprüchlich sind, wie sie aussehen. Darüber hinaus zeigt die Ideengeschichte, dass viele Theorien, die in einem Jahrhundert unterlassen werden, in den folgenden wieder auftauchen, und erneut durch neue Werkzeuge ergänzt und wieder etabliert werden.

1.7.4 Semantik als Physiognomik des Textes

Die Semantik steht zum Text, wie die Physiognomik zum Gesicht. Die kanonische Lehre über die Linguistik baut die zugrundeliegende Theorie des Sprachphänomens auf einer *Zwiebelstruktur*, in der die Semantik die letzte, d.h. die äußerste Ebene darstellt. Diese Gestaltung ist nicht falsch in sich, sondern für das gefälschte Verständnis, das früher oder später den Linguist beherrscht, das heißt: die Sprache entspricht verschiedene Struktur-Ebenen, und die Semantik ist etwa mysteriös, lebendig, schwierig zu formalisieren, die wie ein Geheimnis direkt aus der Ebene der Syntax entsteht. Die Folge Phonetik-Morphologie-Syntax-Semantik ist nicht chronologisch, sondern konzeptuell, was häufig nicht festgehalten ist. Das bedeutet, dass auch auf der Ebene der reinen Phonetik (bzw. Phonologie) man *in nuce* eine ähnliche 'vierfaltige' Struktur feststellen kann, wovon aber aus praktischen Gründen die Beispiele immer schwieriger werden.

Offensichtlich ist die semantische Ebene komplexer und subtiler als die syntaktische, was aber keine zureichende Annahme darstellt, um abzuleiten, dass die Semantik aus der Syntax hervorgeht. Im Unterschied zu den übrigen linguistischen Phasen, ist die Semantik in Elementen nicht zerlegbar. Wenn ein Vergleich mit der Geometrie mindestens nützlich sein kann, um einen derartigen Abstand greifen zu können, wie eine Fläche, die in einer komplexen geschlossenen Kurve beinhaltet ist, kann die Semantik auch nur durch Integralen erschöpfend bestimmt werden. Je komplexer die *Exo-struktur* ist, desto facettenreicher wird auch die Semantik. Über Semantik kann man schon bei der morphologischen Ebene reden, und gemäß einiger mutigen Dichter und Linguisten des vorherigen Jahrhunderts sogar auf der *phonologischen*.

	Phonetik	*Morphologie*	*Syntax*	*Semantik*
Phonetik	**Phonetik**	**Phonologie**	**Phonotaktik**	**Phonosymbolismus**
Morphologie	**Phonologie**	**Morphologie**	**Morphosyntax**	**Morphosemantik**
Syntax	**Phonotaktik**	**Morphosyntax**	**Syntax**	**Semantik**
Semantik	**Phonosymbolismus**	**Lexikologie**	**Semantik**	**Semantik/Pragmatik**

In der Linguistik wird ein Regel anerkannt und niedergeschrieben, sobald ihre Gültigkeit abgelaufen ist, oder zumindest anfangen hat, abzusinken. Diese Lemma ist einfach

beweisbar. Die Anerkennung eines Motivs führt unmittelbar dazu, es vernichten zu wollen. Die Vernichtung, die zunächst in der alltäglichen Umgangssprache stattfindet, erfolgt durch den Verlust von Bedeutung, die irgendwie durch neue Regel und neue linguistische Spiele ausgeglichen werden muss. Die Aufhebung von Struktur und Semantik kann nicht durch die Verschmelzung der zwei Ebenen erfolgen, weil sie nicht konsequent zueinander sind. Was sie vereinen kann, ist nur ein allgemeines Gesetz, das aber keine Verwandtschaft nachweist, da die Semantik eine *Struktur*, oder anders gesagt eine Erscheinung eines anderen Substrat ist, welche noch tiefer als die Bedeutung geht. Eine Bedeutung gibt es nur, wo eine Struktur es gibt. Die Struktur setzt die Bedeutung voraus.

Aus diesen Gründen kann man verstehen, dass bei der Entwicklung einer dynamischen Struktur ein bestimmter (Zeit)Punkt es gibt, wobei einen Objekt sich gestaltet, den aus einem theoretischen Ansicht durch ein zusätzliches Ebene sich analysieren lässt.

Der Sinn einer tieferen Untersuchung der Bedeutung von *Semantik* liegt darin, sie als zugrundeliegende Dimension der Sprache zu setzen. Der Sinn einer Feststellung der Selbstähnlichkeit liegt darin, alle syntaktischen und (im Allgemeinem) formellen Aspekte der Sprache in eine semantische Dimension zurückzuführen. In diesem Sinn spiegeln alle Elemente einer Struktur sich einander in einer rekursiven Weise auf den unterschiedenen Ebenen der Sprache. Die poetische und die metasprachliche Funktionen der Sprache [25] wirken deshalb auf der Grenze des Widerspruches. Die lebendige Sprache errichtet sich mit vieler Mühe aus dem Leichnam der grammatikalischen Regeln, nicht dank sondern trotz derer. Unsere Verschärfung der Bedeutung von Bedeutung war notwendig, zwecks dem Zeigen der Richtung unser Forschung, die die Selbstähnlichkeit als einen Grundzug eines dynamischen Systems erkennen will.

[25] Nach der Lehre von Roman Jakobson.

2 Katastrophé

2.1 Korpusauswahl

Die Gründe der Auswahl der Märchen als unser Textkorpus hängen von einigen Eigenschaften des Märchens ab, aber überhaupt nicht von der Tatsache, dass die Erzählfunktionen ursprünglich für Erzählungen bzw. Märchen ausgedacht worden sind.

2.1.1 Warum das Märchen? Eigenschaften des Märchens

Das Märchen wird nicht nur wegen seiner einfachen Struktur gewählt, sondern auch wegen der Abwesenheit von Bezügen auf Zeit und Räume, die in unserer Realität existieren (was sehr häufig in Erzählungen und Romanen geschieht). Um sie verstehen zu können, braucht man so wenige Vorkenntnisse, dass man auf dem Aufbau semantischer künstlichen Gedächtnissen/Ontologien verzichten kann. Auf dem Ergebnis wird demnächst die Analyse der Erzählung sich aufbauen, für die die Märchenfunktionen als Erwartungswerten gelten werden. Aus einem rein methodologischen Standpunkt ist das Märchen zu bevorzugen, weil in einem Gebiet wie der Text, das so schwierig formalisierbar ist, das Märchen die Möglichkeit bietet, ein konkretes Ergebnis zu bekommen.

Noch ein Grund für die Wahl eines Märchenkorpus liegt immer noch in der Psycholinguistik: da die Märchen ein archetypisches Substrat unserer Verständnis der komplexeren Texten sind, erscheinen die Verben, die darin benutzt werden, so einfach, dass sie das ganze semantische Gebiet eventueller spezifischeren Wörter darstellen, die später bei 'überlegenen' literarischen Formen verwendet werden. Wenn die benutzten Methoden und Benennungen nicht widersprüchlich mit anderen Textarten sind, können die komplexeren Fälle und die reiferen literarischen Genres schichtweise durch Ergänzungen in

dem Modell später eingeschlossen werden. Außer dem methodologischen Ansatz, der bei irgendwelchen Überlegungen in einem kulturellen und anthropologischen Rahmen angeführt werden kann, gibt es noch eine Ebene, die zu den gleichen Schlüssen führt. Die Verbesserung der Erlesenheit der literarischen Produkte kann nur durch mehr Komplexität erfolgen. Um das zu erzielen, sollte man *Konflikte* hinzufügen, die irgendwie die finale Aufhebung verschieben können. Das geschieht innerhalb des gleichen Formats weil, anders als seine intellektuellen Bedürfnisse, die intellektuelle Fähigkeiten des Menschen im Laufen der Jahrhunderten ungefähr unverändert bleiben[1].

Schließlich stimmen bei den Märchen *fabula* und *syuzhet* [2] überein. Diese wichtige Eigenschaft ermöglicht die Anpassung der Funktionen dem Modell der magischen Quadraten ohne weitere Schwierigkeiten[3].

2.1.2 Vorläufige Analyse der Grimms Märchen

Vladimir Propp hat seine Funktionen empirisch abgeleitet, was bedeutet, dass sie keine perfekte Abstraktion einer Testeinheit darstellen, sondern nur die Vereinigungsmenge der spezifischen Fälle seiner behandelten Sammlung, d.h. die russische Märchen des Afanasevs. Trotzdem können sie durch eine leichte Verallgemeinerung den deutschen Märchen gut angepasst werden. Die Grimms Märchen zeigen eine Abwesenheit großer und bestimmter Konflikte, sodass Proppsfunktionen wie 'Krieg' tatsächlich nie benutzt werden. Alle Fälle, wobei diese bestimmte Funktion vorkommt, bezeichnen ja die Auseinandersetzung von Hauptcharakter und des Gegners, nachdem der Gegner niedergeschlagen

[1] Der sogenannte Flynn-Effekt, der ein stetiges Wachstum des Intelligenz-Quotientes über die Jahrhunderte beschreibt, scheint gegen unsere Behauptung zu sprechen. Wenn man doch überlegt, dass diese Beobachtung nur die standardisierten IQ-Tests der letzten fünf Jahrzehnten betrifft, und dass sie auch davon abhängt, dass in der Zwischenzeit die Ausbildung und das allgemeine Wohlbefinden der getesteten Probanden umso mehr gestiegen ist, könnte man ableiten, dass die Flynns Ergebnisse noch nicht als definitiv gültig zu überlegen sind. Darüber hinaus wäre jetzt mein Intelligenz-Quotient laut seiner Behauptung etwa dreimal höher als das von Leibniz und zehnmal höher als das von Cicero.

[2] In der Erzähltheorie stellt das *Syuzhet* (Subjekt auf russisch) die Folge der Ereignissen dar, so wie sie im Text erzählt werden, wohingegen das *Fabula* die chronologische Folge repräsentiert.

[3] Siehe Paragraph 'Folge und Struktur III'.

wird, aber fast nie in einer gewalttätigen Weise. Eine weitere Eingenschaft der deutschen Märchen, die wirklich sehr oft vorkommt, ist die Ambiguität einiger Charakter, die in ihrer Rolle schwierig zu erkennen sind. Nicht selten ist ein Schenker gleichzeitig Gegner oder ändert er sich in Gegner in einem späteren Moment. Die gleiche Ambiguität gilt auch für einige Funktionen: nicht immer sind sie einfach anzuerkennen, und auch in diesem Fall ist es möglich, einige Abschnitte zu finden, die durch zwei oder sogar mehrere Funktionen anzudeuten sind. Zum Glück gibt es bei diesen Fällen immer ein Prinzip, ohne Zweifeln wählen zu können: statt was beim ersten Anblick erscheint, soll man die *echte* und *tiefe Funktion* des Ereignisses bezeichnen. Zum Beispiel: die Gottheit sagt, dass wenn der Krieg gewonnen wird, wird der Held seine geliebte Frau verlieren. In beiden Fälle, Sieg und Niederlage, gibt es etwas zu verlieren. Aus diesem Grund wird der Krieg später nicht als *Krieg* bezeichnet, sondern als *vorläufiges Unglück*. Wie schon festgestellt, besteht die Aufgabe des Schriftstellers aus der Erzeugung neuer Modelle, die Überraschung im Leser verursachen. Nichtsdestotrotz kann ein solches Umkippen der Erwartungen in einer chaotischen Weise nicht erfolgen. Selbstverständlich stellt eine chaotische Reihenfolge von Ereignisse keine Überraschung dar, weil laut ihrer Definition die Überraschung stark ist, wenn der Erwartungswert ganz unterschiedlich vom Ergebnis ist, *aber trotzdem das Ergebnis kohärent mit der* Ontologie *des Märchens bleibt.*

2.1.3 Die *Dramatis Personae*: ein letzter Schritt zur Adoption und Anpassung des Propps Modelles

Wie von Propp selbst in seinem Meisterwerk hervorgehoben wird, ist fest und unveränderbar in einem Märchen nicht nur die Reihenfolge der Funktionen, sondern auch diejenige der Charakter. Es gibt sieben Arten *Dramatis Personae* (d.h. Charakter), die immer vorkommen, die mehr oder weniger immer gleich charakterisiert werden, und denen eine bestimmte Untermenge der Funktionen gehört (zum Beispiel Betrugsmanöver, Schädigung, Verfolgung usw. gehören offensichtlich zum *Gegner*).

Im Grimms Märchen 'Die Goldene Gans' kann man kaum zwischen Schadenstifter und Schenker entscheiden, und in 'Rumpelstilzchen', könnte der Zwerg, dessen Namen den Titel dem Märchen gibt, sowie als Held als auch als Gegner überlegt werden. Es

gibt zahlreiche ähnliche Beispiele, aber für die Ziele unserer These, wäre es besser sofort fortzufahren, und abschließen, dass in unserer Version der Erzählfunktionen die *Dramatis Personae nicht* überlegt werden, nicht nur wegen ihrer Ambiguität, sondern weil sie automatisch aus dem Gewebe der Tätigkeiten hervorbrechen, damit sogar nutzlos im Sicht der Verständnis der Erzählungen zu verstehen ist, wer was getan hat.

Da die Märchen der Gebrüder Grimm ungefähr zu dem gleichen Zeitraum der Afanasevs Märchen gehören, ist die Ursache dieser schwachen Charakterisierung der *Dramatis Personae* entweder in einer vermutlichen 'deutschen Tendenz', oder in den Autoren selbst aufzufinden.

Die Ambiguität im Bezug auf den *Dramatis Personae* betrifft auch die Funktionen selbst: viele Ereignisse können oft als zwei oder sogar mehrere Funktionen interpretiert werden, und manchmal entwickelt sich eine Funktion in mehrere *Unterfunktionen*, obwohl es klar scheine, dass man noch in der Kohärenz-Domäne der Hauptfunktion sich befindet. Diese Fragestellung zwingt uns eine bestimmte Notation zu verwenden, die uns ermöglicht, diese zusammenstehende Eigenschaften zu schätzen, und gleichzeitig ihnen das passende Gewicht zuzuweisen, sodass ein automatisiertes Verfahren sie richtig verarbeiten kann.

2.1.4 Auflistung der Propps Funktionen

Die folgende 36 Funktionen gleichen im größeren Teil den Propps Narrathemen. Eventuelle Änderungen überlegen die Beobachtungen, die in den vorherigen zwei Paragraphen erklärt worden sind.

- $d \rightarrow$ Deíxis.
- $k \rightarrow$ Katastrophé.
- $l \rightarrow$ Lýsis.

DEÍXIS

- i: Ausgangsituation (d) 0

- a: Entfernung (k) -
- b+ : Befehl (l) +
- b- : Verbot (l) -

deixis

- c+ : ausgeführter Befehl (d) +
- c- : übertretenes Verbot (d) -
- d : Erkundigung (k) 0
- e : Verrat (l) -

katastrophé

- f : Betrugsmanover (d) -
- g : Mithilfe (k) 0
- x : Vorläufiges Unglück (l) -

Lýsis

KATASTROPHÉ

- A : Schädigung (d) -
- α : $Mangelsituation(k)$ -
- B : Vermittlung (k) 0
- C : Einsetzende Gegenhandlung (l) +

2 Katastrophé

- $\rightarrow$: *Abreise*(l) 0

deixis

- Pr : Prüfung (d) -
- $\lambda+$: Positive Reaktion des Heldes (k) +
- $\lambda-$: Negative Reaktion des Heldes (k) -
- Z : Empfang eines Geschenkes/Zaubermittel (l) +

katastrophé

- W : Raumvermittlung/Wegweisung (d) 0
- K : Kampf (k) -
- M : Markierung (l) 0
- S : Sieg (l) +

Lýsis

LÝSIS

- L : Aufhebung des Unglücks (d) +
- $\leftarrow$: *Rckkehr* (d) 0
- V : Verfolgung (k) -
- R : Rettung (l) +

deíxis

- X : Unerkannte Ankunft (d) 0
- U : Unrechtmäßige Ansprüche (k) -
- P : Prüfung (k) -
- Lö : Lösung (l) +

katastrophé

- E+ : Erkennung des Heldes (d) +
- E- : Überprüfung des falsches Heldes (d) +
- T : Transfiguration (k) +
- St : Strafe (k) +
- H* : Hochzeit (Preis, Gutes Ende) (l) +
- od : Ende, Motivierungen (l) 0

Lýsis

2.1.5 Beschreibung der wichtigeren Eigenschaften

Durch die Zusammenfügung mit dem Hjelmslevs Textmodell hat man gezeigt, dass ein Märchen durch die rekursive Anwendung dreier Elementen komponiert werden kann.

Jede Funktion kann so umformuliert werden, dass an erster Stelle die Hauptteil des Märchens gezeigt wird, wo die Funktion erscheint, am zweiten die Hauptteil der Hauptteil und am dritten die Hauptteil der Hauptteil der Hauptteil. Zum Beispiel:

Die *Transfiguration* ist die *Katastrophé - Teil* der *Lysis - Teil* der *Lysis- Teil.* Deshalb können wir sie so ausdrücken:

LLK

Das ist ein gutes Beispiel, um noch eine Fragestellung dieser Notation hervorzuheben: es gibt Funktionen, die in dieser Weise die gleiche Benennung bekommen würden. Wie behebt man diese Problem?

Es wäre bereits möglich zu schätzen, dass in der Hälfte der Fälle irgendwelche Funktion automatisch die Funktion ausschließt, die die gleiche Benennung bekommt (wie zum Beispiel: Befehl und Verbot).

Andererseits kann man überlegen, dass alles in allem ohne großen Bedeutungsverlust die homonymen Funktionen gleich benannt werden können, da sie in der *Ökonomie* des Märchens tatsächlich gleich sind. Zum Schluss kommen bei den übrigen Fällen die homonymen Funktionen direkt nacheinander vor; sie stellen so zu sagen eine Molekül dar (z. B.: Mangelsituation/Vermittlung und Unrechtmäßige Ansprüche/Prüfung).

Wir möchten jetzt nochmal hervorheben, wie diese 'triadische' und rekursive Verfügung der Textteile nicht nur die Märchen, sondern auch die die Zeitschriftartikeln, die Werbungen, die Haikus, Masterarbeiten, usw. wohl charakterisieren, obgleich diese Einteilung in den unterschiedlichen Gebieten jeweils unterschiedliche Benennungen bekommt[4].

2.1.6 Widersprechende und ergänzende Paare

Ein oberflächlicher Überblick auf den Propps Funktionen ermöglicht uns zu verstehen, dass einige als positiv, andere als neutral, und noch andere als negativ sich bewerten lassen. Unsere Domäne ist jedoch nicht die Sentiment Analyse, wobei diese Terminologie schon bekannt und benutzt wird, sondern etwas, dass näher zur Chemie zu überlegen ist: die Oppositionen, die dadurch aufgebaut werden, beinhalten eine Rennstrecke, worin Begriffen und Ereignissen sich wieder wie beim Triade-System anordnen lassen. Die Wörter üben einen Drück, derer Richtung und Stärke von ihrer Position im Text abhängt. Je breiter ist der Abstand zwischen den Elementen eines Opposition-Paares, desto großer wird den Selbstinformationswert, den die zweite Teil des Paares annehmen wird. Die

[4]Dieser Satz wurde nicht, und wird nicht im Nachstehenden bewiesen. Wie auch bei der Durchführung des Experimentes erscheint, stellt das *Haiku* ein zu kurzer Text, um mathematisch darin die Selbstähnlichkeit messen zu können. Trotzdem lassen die Ergebnisse der These die Glaubwürdigkeit dieser Annahme vorausahnen.

Funktionen bezeichnen die Ereignissen. Es scheint überflüssig zu erwähnen, dass auf der Ebene des Satzes das *Dependenzabstand* Modell genau dasselbe festgestellt hat: komplexere Texteinheiten tragen im Durchschnitt mehr Information.

$2-3$	$\rightarrow$	$Verbot - Verstoß$
$4-5$	$\rightarrow$	$Erkundigung - Erringung$
$6-7$	$\rightarrow$	$Betrug - Mitwisserschaft$
$8/9-19$	$\rightarrow$	$Beschädigung - Entfernung$
$11-20$	$\rightarrow$	$Abreise - Heimkehr$
$17-27$	$\rightarrow$	$Kennzeichnung - Identifizierung$
$21-31$	$\rightarrow$	$Verfolgung - Krönung$
$23-29$	$\rightarrow$	$InkognitoAnkunft - Vernderung$
$24-30$	$\rightarrow$	$unbegründeterAnspruch - Bestrafung$[5]

Jede Zahl entspricht der Position einer Funktion innerhalb der theoretischen Folge. Aus der oberen Spalte kann man ableiten, dass in der ersten Phase einer Erzählung (deklarative Phase) alle Ereignissen ein nacheinander vorkommen, ohne weitere meta-textuellen Beziehungen, welche Verbindungen auch außerhalb der Zeit- und Ursache/Folgeketten herstellen. In der zweiten Phase der Erzählung sind die Oppositionen am gespanntesten, um sich gegen das Ende wieder zu reduzieren. Immer wieder kann unsere persönliche Erfahrung und unsere Intuition diese Behauptungen unterstützen, was uns erlaubt, die schon existierende Kennzeichnung besser zu formalisieren, sodass das System auch mathematisch modelliert werden kann.

Auflistung der Abstände einiger ergänzenden Paaren

$$
\begin{aligned}
i - od &\cong 30 \\
\leftarrow - \rightarrow &\cong 20 \\
b\breve{}c &\cong 2 \\
f\breve{}g &\cong 1 \\
A\breve{}St &\cong 22 \\
x\breve{}L &\cong 12 \\
\alpha\breve{}Z &\cong 6 \\
K\breve{}S &\cong 2 \\
V\breve{}R &\cong 1 \\
X\breve{}E\breve{}T/H* &\cong 3/5 \\
U\breve{}\ddot{U}/St &\cong 3/5
\end{aligned}
$$

2.1.7 Zusammensetzungsregeln: die Erzählung als ein L-System

Schon im Paragraph 'Auflistung der Propps Funktionen' hat man bemerkt, dass einige Funktionen die Polarität der ergänzenden Paare entkommen. Die neutrale Funktionen (wie beispielsweise *Erkundigung* und *Vermittlung*) können auch mehrmals vorkommen, und meistens hängen sie nicht von dem Anordnung an. Die übrige Funktionen können aber nicht zufällig gestellt werden, wie auch im Rahmen der Sätze... und der Evolution der Pflanzen geschieht. In diesem Paragraph möchte man versuchen, die Zusammensetzungsregeln der Erzählung als ein Lyndenmeyer-System zusammenzufassen. Unter *Lindenmeyer-System* versteht man einen logischen Formalismus, der ursprünglich als Grundlage einer axiomatischen Theorie der biologischen Entwicklung von dem Biologe Aristid Lindenmyer vorgeschlagen worden ist. Es besteht aus einem Quadrupel $G = (V, S, \omega, P)$, wobei

- V die Zeichen enthält, die die Variable darstellen;

- S die Zeichen enthält, die die Konstanten darstellen. Die Zeichen aus V und S bilden das Alphabet des L-Systems;

- ω ein Wort über dem Alphabet ist, welches als Startwort oder Axiom des L-Systems bezeichnet wird;

- P eine Menge von geordneten Paaren aus Wörtern über dem Alphabet ist, welche Ersetzungsregeln definieren [6].

Ein Beispiel davon ist aus der Hilberts Kurve repräsentiert:

$$
\begin{aligned}
V &\rightarrow & A, B;\\
S &\rightarrow & f, l, r;\\
\omega &\rightarrow & A;^{7}\\
P &\downarrow & \\
A &\rightarrow & lBfrAfArfBl;\\
B &\rightarrow & rAflBfBlfAr^{8}.
\end{aligned}
$$

Merkwürdig ist eben, dass die Zusammensetzungsregel für B genau gegensätzlich zu A ist. Diese Eigenschaft bildet die Grundoperation der Thue-Morse Kette, die in der folgenden Sektion behandelt wird.

Es ist bekannt, dass Propps Funktionen immer die gleiche Reihenfolge folgen. Einige können fehlen, eine Menge davon kann sich mehrmals wiederholen, aber sicherlich kommt keine *Transfiguration* vor dem *Verrat*, und kein *Sieg* vor dem *Kampf*.

Im Folgenden wird es versucht, eine theoretische Erzählung durch die Verschmelzung zwischen der Aristoteles-Hjelmslev Notation und dem L-System aufzubauen.

[6] Vgl. [24]

[7] wobei f als 'forward' bzw. 'zeichne vorwärts', l als 'left' bzw. 'nach links 90° abbiegen' und r als 'right' bzw. 'nach links 90° abbiegen' gelten.

[8] Vgl. [24]

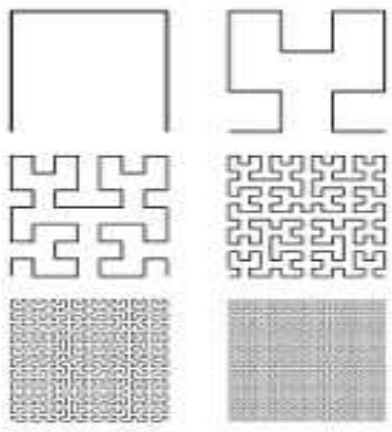


Abbildung 2.1: Abbildung einer Hilberts Kurve, die vom oben beschriebenen Formalismus erzeugt wird.

$$E = (C, P, F, \alpha, \omega, Z)^{9};$$

$$C = (d, k, l)^{10};$$

$$C^1 = (1, 2, 3);$$

$$C \vDash C^1;$$

$$P = \bigcup_{i=1}^{size(C)^{size(C)}} {p_i}^{11};$$

$$p \in P = ccc,\ c \in C^1;$$

$$F = vec(p_1 \xrightarrow{z \in Z} p_2 \xrightarrow{z \in Z} ... \xrightarrow{z \in Z} p_n);$$

$$f \in F = subvec(F);$$

$$\alpha = [ddd] = [111]^{12};$$

$$\omega = [lll]] = [333];$$

$$Z :$$

$$[c * 1] \rightarrow [c * (2 \vee 3)];$$

$$[c * 2] \rightarrow [c * 3];$$

$$[c * 3] \rightarrow [c * (c + 1) * 1] \vee \omega^{13};$$

$$f \rightarrow x \times f, x \in N^{14};$$

Das gerade durchgeführte Versuch, einen Formalismus aufzubauen, um die Zusammensetzungsregeln eines Märchens zu bestimmen, scheint eher schwach in seinem zweiten Teil (der sowieso in seinen Absichten korrekterweise bei den Fußnoten beschrieben wird). Der bedeutendste Erfolg dieses Versuches scheint die Berechenbarkeit zu sein: die Konstanten wurden leicht in Zahlen verwandelt, was eine einfachere eventuelle algorithmische Implementierung fördert. Obgleich es um eine erweiterte Version eines L-Systems sich handelt, weist ohnehin dieser Entwurf nach, dass die Erzählung durch rekursive Regeln *erzeugbar ist.*

[9] Konstanten, Narratemen, Narratemenliste, Startsnarrateme, Endsnarrateme, Zusammensetzungsregeln.

[10] Die Bezeichnungen *d, k, l* entsprechen *Deíxis, Katastrophé, Lýsis.*

[11] Ein Narrateme besteht aus eine Element aus der Liste aller Narrateme, die man aus der Variablenmenge durch eine Permutation mit Wiederholung bekommt, die ihrerseits aus der Länge der Menge hoch der Länge der Menge besteht.

[12] α und ω bezeichnen in diesem Fall die zwei obligatorische Anfangs- und Endekonstanten.

[13] Nach einer *Lýsis* (Nummer 3 in der Abbildung c) wird die obere Domäne von 1 aufgewertet, und die darauffolgende Zahlen werden Reset vorgenommen. Falls eine Aufwertung nicht mehr möglich ist (es gibt kein oberes Domäne mehr zur Verfügung), bedeutet es, dass die Erzählung beendet ist (ω).

[14] Irgendwelches Funktionenvektor kann beliebig viele Malen wiederholt werden.

2.1.8 Folge und Struktur II: Thue-Morsefolge und UID

Eine berühmte Folge, derer Zusammensetzungsregeln auch ein sehr bekanntes Beispiel des L-Systems ist, wird durch die Thue-Morsefolge dargestellt.

L-SYSTEM FÜR DIE MORSEFOLGE

- Variablen 0, 1.
- Konstanten Keine.
- Anfang 0.
- Regeln $(0 \to 01)$, $(1 \to 10)$

Das Ergebnis scheint beim Schritt 6 ungefähr wie folgt:
0110100110010110100101100110100101101001100101101001011001101001,
und beim Schritt 4 so:
0110100110010110.

Man beobachtet, dass jeder gerade Schritt der Morsefolge palindrom ist. Wenn man diese Folge zeichnen möchte, sodass jede 0 als einer Bleistifts Zug konstanter Länge und jedes 1 als eine 60°-Drehung der Richtung des Zuges darstellen, bekommt man die Kochs Kurve 1.1.

Diese Kurve interessiert uns nun nicht wegen ihrer faszinierenden Gestalt, sondern für ihre enge Verwandtschaft mit unserem Thema, genauer gesagt mit der Informationstheorie und mit der *Uniform Information Density*, zu den früher eine Sektion gewidmet worden ist. Der *trait d´union* ist durch *Recursive Binary Sequences of Differences* repräsentiert, ein Artikel von Robert M. Richman, worin ein sehr interessantes und bedeutendes Ergebnis beinhaltet ist, das weltweit als das *coffee pot problem* bekannt geworden ist:

> There is a flavor concentration gradient in a carafe of coffee: less and less flavor is available for extraction from the coffee grounds as the brewing process progresses, so the coffee is stronger at the bottom of the carafe than at

> the top. Swirling produces horizontal rather than vertical convection, so it does not eliminate the concentration gradient. In pouring two cups of coffee of equal volume from this carafe, how can one pour the coffee to most nearly equalize the solute concentrations in the two cups? Let c be the concentration of interest and x be the height from the bottom of a cylindrical carafe, in units such that the range of x is from 0 to 1. It is likely that the concentration function c satisfies the conditions $c(x) > 0$, $c'(x) < 0$, and $c''(x) > 0$ over the entire interval [0, 1] [...] Let us apportion the coffee from the carafe equally into two cups, A and B, and let the notation AB designate a binary sequence of two pours, the top half of the carafe into cup A, then the bottom half into cup B. If one has the patience to make four pours of equal volume, the possible pouring sequences are AABB, ABBA, and ABAB. The objective is to make the difference in the two concentrations as close as possible to zero. Subtracting the concentration in cup B from the concentration in cup A is equivalent to multiplying the concentration of cup B by -1 and adding the two. Representing the sequences as step functions, they correspond to the FCN functions with $n = 2$. The best pouring sequence is obtained by minimizing the absolute value of the integral of the product of the concentration function times the sequence step function. That is, the concentration difference between cups A and B is minimized by using the step function FCN(2, i, x) that minimizes $|\int_0^1 FCN(2, i, x)c(x)dx|$ over all i. The sequence that best minimizes the difference is ABBA, which is DIF(2, j), whenever c(x) is monotonic, a condition one would normally expect. [...] This result may be applicable to more important problems. For example, it is often difficult to distribute pigment uniformly through paint. These functions could be applied to such situations, where homogenization of inhomogeneous mixtures is difficult. It is likely that uniformity from one sample to the next can best be achieved by distribution following the pattern of the DIF functions. [47]

Wie man diesem Abschnitt entnehmen kann, folgt auch die Funktion der uniformen Streuung eines Stoffes in einem Lösemittel die Rekursive Morsefolge. Wenn zum Beispiel acht Tassen erfüllt werden sollten, wäre die optimale Reihenfolge 'ABBAABBA' gewesen, und so weiter. Die *Uniform Information Density* ist nichts anders als die informationstheoretische Version dieser gleichen Fragestellung. Auch im Rahmen der Information soll man festhalten, dass die Uniformität der Streuung des Signals durch die Kommunikationsleitung mit einem anderen (teilweise widersprechenden) Bedingung der sprachlichen

Kommunikation sich auseinandersetzen soll, und zwar *die stetige Erhaltung oder Erhöhung der Informationsinhalt.* Im folgenden kann man beobachten, dass auch in diesem Fall die Einteilung der Zahlen in einer Matrize interessante Hinsichten bietet:

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \end{bmatrix}$$

Genau wie in einem*magischen Quadrat* gleicht die Summe der Elemente jeder Reihe der Summe der Elemente jeder Spalte. Man beobachtet dass die Reihen immer zwei Muster folgen, die bei der rechten Kante durch die Buchstaben A und B symbolisiert werden. Dadurch kann man leichter erkennen, dass die Palindromie, die auf der Ebene der einzelnen Reihen gilt, auch auf der Ebenen der ganzen Matrize widergespiegelt wird. Durch diese besondere Streuung der Elemente erzielt man die wichtige Eigenschaft der *Selbstähnlichkeit*, die statistisch durch die Skaleninvarianz [15] vermesst werden kann, und die die Erfüllung der UID-Theorie auch auf der Ebene des Textes in der elegantesten Weise ermöglicht.

[15] Der Begriff beschreibt die Eigenschaft eines Verhältnisses, bei dem eventuelle Veränderungen der Skalierung die inklusive Eigenschaften seiner Eckwerte weitestgehend nicht anfechten, sodass ein *selbstähnlicher* Zustand gegeben ist, der meistens gewisse Universalitäteigenschaften zeigt(Vgl. https://de.wikipedia.org/wiki/Skaleninvarianz).

2.2 Erwartete Reihenfolge und Topic Modeling

Eine Funktion wirkt im Rahmen einer Episode, derer das *Ereignis* das Gipfel darstellt, d.h. die hauptsächliche Weise, in der eine Episode sich verwirklicht. Gegeben sei ein Textkorpus, das aus einer Sammlung von Erzählungen ausgemacht ist, die mit Propps Narratemen annotiert worden sind, ist es möglich ein Programm zu trainieren, das künftige nicht annotierte Erzählungen in Episode zerlegt, und jede Episode mit einer Propps Funktion bezeichnet.

- Man weißt schon welche Wörter in einer Funktion meistens vorkommen.
- Die Anwendung der Interpunktion als Grenze zwischen den Episoden ist zum Glück fast eindeutig, und deshalb zweifellos verwendbar.
- Nicht nur die Stoppwörter sollen zur Liste jeder Funktion nicht gehören, sondern auch die Wörter die zu neutral erscheinen [16].
- Man kann die theoretische Reihenfolge, die Liste der ergänzenden Paaren mit dem dazugehörigen Abstand, und die Zusammensetzungsregeln verwenden, um ein sinnvolles statistisches Modell aufzubauen, sodass eine effiziente Disambiguierung durchgeführt werden kann;
- *Germanet*[17] könnte auch sehr erfolgreich für die Disambiguierung und für die Bereicherung der Funktionslisten verwendet werden.

Mit diesen Annahmen kann man aus einer anfänglichen manuell-annotierten Märchensammlung beliebig viele Märchen automatisch annotieren[18].

[16] Unter *neutrales Wort* versteht man ein Wort, die keinem eindeutigen semantischen Bereich gehört, und/oder alle Wörter die in mehrere Funktionen vorkommen.

[17] Germanet ist die deutsche Version von *Framenet*, ein Tool für die semantische Segmentierung. Es besteht aus einer erweiterten lexikalischen Datenbank, wobei jedes Wort mit anderen verbunden ist, die dem gleichen kognitiven Netz angehören (die so genannte *synsets*).

[18] Dieses Thema kann leider in dieser Behandlung keinen Platz zurzeit finden, aber es wird zweifellos in zukünftigen Studien behandelt.

2.3 Folge und Struktur III: die magischen Quadrate

Mithilfe der Fibonacci-Reihe wird es verständlich, dass es geeigneter wäre, eine zwei- oder dreidimensionale Darstellung zu verwenden, um die Beziehungen zwischen den Funktionen der Geschichte besser darstellen zu können. Man kann untersuchen was passiert, wenn man die Erzählfunktionen als Werte bzw. Zahlen in einer Matrize behandelt. Wenn die Anzahl der Funktionen quadratisch ist, kann man sie sogar in einer magischen Quadrat einbetten.

Die magische Quadrate sind besondere quadratische Matrize, wobei die Summe jeder Spalte, jeder Reihe und jeder Diagonale immer den gleichen Wert trägt, und zwar

$$\frac{n * (n^2 + 1)}{2}$$

, wobei n die Anzahl der Elemente jeder Reihe bzw. Spalte ist.

Diese Struktur ist *selbstähnlich*, nicht nur wegen der gerade erwähnten Eigenschaft, sondern auch wegen der Weise, in der sie aufgebaut werden. Zu diesem Punkt ist es nötig zu erwähnen, dass kein magisches Quadrat eindeutig ist (außer für n=3): es ist möglich viele Lösungen und Strategien zu verwenden, um ein magisches Quadrat aufzubauen, denen genau so viele Einbettungen entsprechen. Was jedoch merkwürdig ist, ist die Tatsache, dass jede davon auch die Bedingungen der Selbstähnlichkeit erfüllt. Wenn die Zahlen von 1 zu n progressiv mit einer Gerade verbunden werden, wird man herausfinden, dass die resultierende geometrische Gestalten selbstähnlich und symmetrisch sind. Eine andere sehr wichtige Eigenschaft der magischen Quadrate gerader Ordnung, die ich selbst gefunden habe, ist:

$$\forall i \in M = n^2 \wedge x \in X = n^2 \Rightarrow M[i] = x \wedge M[x] = i \rightarrow M[M[x]] = x$$

Dies bedeutet: wenn in der Position i der Wert x gefunden wird, wird in Position x der Wert i erscheinen.

Da ein magisches Quadrat eine Matrize ist, soll das Wert jeder Position durch die Zahl der Spalte und der Reihe, wo er platziert ist, dargestellt werden d.h. $x = M(i, j), i \wedge j \leq n$.

Wenn man aber die Positionen durch eine Zahl nur bezeichnet, sodass die erste Position der Position oben links und die letzte die Position unter rechts entspricht, und bei neuen Reihen fängt man immer links zu nummerieren, bekommt man eine einzählige Nummerierung, die umso mehr von der Schönheit der oben genannten Eigenschaft begründet wird.

Ist die Selbstähnlichkeit der magische Quadrate und die angenommene Selbstaffinität der Erzählungen ein genügendes Grund, um sie ohne weitere Kriterien kombinieren zu können?

Selbstverständlich nicht. Dieses 'etwas mehr', das man benötigt, um sich eine derartige Operation leisten zu können, erklären wir im folgenden Abschnitt. Wir haben ohnehin gerade festgestellt, wie Position und Wert bei den magischen Quadraten eng verbunden sind. Das bedeutet genau die Vereinigung von Folge und Struktur, die ab dem Anfang dieser Studie gesucht worden ist.

Das magische Quadrat der Sonne

$$\begin{bmatrix} 6 & 32 & 3 & 34 & 35 & 1 \\ 19 & 7 & 11 & 27 & 28 & 8 \\ 30 & 14 & 16 & 15 & 23 & 24 \\ 13 & 20 & 22 & 21 & 17 & 18 \\ 25 & 29 & 9 & 10 & 26 & 12 \\ 36 & 2 & 33 & 4 & 5 & 31 \end{bmatrix}$$

2.3.1 Position und Wert

Durch die transversale Konzeptualisierung wird es versucht sich zu vorstellen, wie das Modell der magischen Quadraten zur Erzählung angepasst werden kann. Was stellt die Position im Märchen dar? Wie schon im Sektion *Warum das Märchen? Eigenschaften des Märchens* wurde es gezeigt, dass *fabula* und *syuzhet* übereinstimmen, deshalb findet man sich nicht in der unbequemen Position zwischen den beiden zu wählen. Wenn man die Schemata, die Lévi-Strauss zur Darstellung der Beziehungen zwischen den Mythen benutzt hat, als Beispiel annimmt, kann man die Reihenfolge der Funktionen in einem magischen Quadrat einbetten. Die folgende Abbildung zeigt die Verschmelzung zwischen unserer Anpassung der Proppsfunktionen den Grimms Märchen und dem magischen Quadrat der Sonne (Kante= 6). Aus allen möglichen Einbettungen, die ohnehin ganz leicht voneinander unterschiedlich sind, wird selbstverständlich diejenige gewählt, die die wichtige Eigenschaft der magischen Quadraten gerader Ordnung entsprechen, die von mir gefunden worden ist:

Δ DEÍXIS	1 i ddd 6	2 a ddk 32	3 b- ddl 3	4 b+ ddl 34	5 c dkd 35	6 d dkk 1
δύναμις	7 e dkl 7	8 f dld 11	9 g dlk 27	10 x dll 28	11 A kdd 8	12 α kdk 30
K KATASTROPHÉ	13 B kdk 19	14 C kdl 14	15 ↑ kdl 16	16 Sch+- kkd 15	17 λ+- kkk 23	18 Z+- kkl 24
ενέργεια	19 W kld 13	20 K klk 20	21 M kll 22	22 S kll 21	23 L ldd 17	24 ↓ ldd 18
Ω LÝSIS	25 V ldk 25	26 R ldl 29	27 X lkd 9	28 U lkk 10	29 P lkk 26	30 Lö lkl 12
Μεταβολή	31 E lld 36	32 U lld 2	33 T llk 33	34 St llk 4	35 H* lll 5	36 # lll 31

Abbildung 2.2: Einteilung der theoretischen Erzählfunktionen im magischen Quadrat der Sonne (6 x 6).

Eine wichtige Eigenschaft des oben gezeigten Quadrates ist, dass die drei Hauptteile mehr oder weniger die gleiche Anzahl von Elementen beinhaltet. Bei allen unseren

Beobachtungen ist aber erschienen, dass tatsächlich die Einteilung der Funktionen das folgende Verhältnis spiegeln:

$$deixis = \frac{1}{4}$$
$$katastrophe = \frac{2}{4}$$
$$lysis = \frac{1}{4}$$

, was uns an die Lösung des "Coffee pot"Problems erinnert.

Die folgende zwei Quadrate zeigen die Märchen *Frau Holle* und *Aschenputtel*. *Frau Holle* scheint genau 36 Funktionen zu haben. Andere mögliche Andeutungen könnten es auch in 30 oder sogar in 40 Funktionen zerlegen, aber man hat für unsere Zwecken offensichtlich die 36-Einteilung bevorzugt. Obwohl *Aschenputtel* ein 8 x 8 Quadrat ist[19], wird es auch in diesem Fall jeder Funktion den Wert zugewiesen, der im theoretischen Reihenfolge 2.2 erscheint.

Der Durchschnitt der Werten jeder Reihe und jeder Spalte ist $\simeq$ 112 (111 ist die Summe der Werte jeder Spalte und jeder Reihe der magischen Quadrat der Sonne.)[20]

Der Durchschnitt der Werten jeder Reihe und jeder Spalte ist in diesem Fall $\simeq 150$[21]. Es ist leicht zu überlegen, dass für längere beobachtete Reihenfolge (wie beispielsweise *Aschenputtel*), soll der erwartete Wert, der der Schnitt der Summe der Elementen jeder Reihe und Spalte entspricht, durch einen besonderen Koeffizient justiert werden. Das Ergebnis kann durch die folgende Gleichung zusammengefasst werden:

$$\sum_{i=1}^{n=\sqrt{F}} \sum_{j=1}^{m=\sqrt{F}} f_{ij} = \frac{37}{2} F \tag{2.1}$$

,wobei $f \in F$ bezeichnet das Narratem, das in der Position ij vorkommt.

[19]Diesmal wird hier keinen Krampf angewendet, die 64 Narratemen sind sauber und eindeutig. Ein schöner Zufall.
[20]Siehe S. 65.
[21]Siehe S. 66.

1 i 6	2 A 8	3 α 30	4 B 19	5 b+ 34	6 C 35	132
7 ↑ 16	8 W 13	9 Sch 15	10 λ+ 23	11 Sch 15	12 λ+ 23	105
13 Sch 15	14 λ+ 23	15 α 30	16 B 19	17 Z 24	18 L 17	128
19 ↓ 18	20 L 17	21 W 13	22 B 19	23 H* 5	24 B/i 19	91
25 A 8	26 α 30	27 C 35	28 ↑ 16	29 W 13	30 Sch 15	117
31 λ- 23	32 Sch 15	33 λ- 23	34 Sch 15	35 λ- 23	36 St 4	100
86	106	146	111	114	113	

Abbildung 2.3: Einteilung der Erzählfunktionen des Märchens *Frau Holle*.

Es scheint intuitiv, dass wenn man die Elemente einer beliebig langen Reihenfolge zufälliger Zahlen von 1 bis 36 summiert, scheint der Schnitt der Summe 18,5 zu sein, mit einer Varianz von 2. Dieses Ergebnis gilt sowohl mit als auch ohne mögliche Wiederholung der Zahlen in der Reihenfolge. Wie kann denn die Theorie der 'magischen' Einteilung der Narratemen in einem Märchen noch 'gerettet' werden? Um eine Antwort zu bekommen, soll man die Fragestellung umkehren: wie kann es sein, dass eine hoch strukturierte Reihenfolge von Elementen[22] wie das Märchen, so entropisch wie eine randomisierte Reihenfolge scheint? Sogar in der Physik ist nachgewiesen worden, dass die Entropie kein Maß des Chaos eines Systems ist, sondern ein Maß seiner Tendenz, wahrscheinlichere Zustände anzunehmen, d.h. Zustände, den eine höhere Anzahl mikroskopischer Zustände entspricht[34]. Auch in der Natur geschieht das gleiche, z.B. bei der Verfestigung der untergekühlten Flüssigkeiten und einigen anderen Phänomenen [34]. Man hat schon festgestellt, dass in der Informationstheorie die Entropie die Größe der Information entspricht, die eine Reihe von Zeichen/Signalen trägt. Obwohl weitere Untersuchungen nötig sind, scheint durch diese besondere (und zufällige) Beispiele unsere Modell stark gefördert. Nicht nur die oben erwähnten Fragestellungen, sondern auch

[22]Siehe den Kapitel *Zusammensetzungsregeln*.

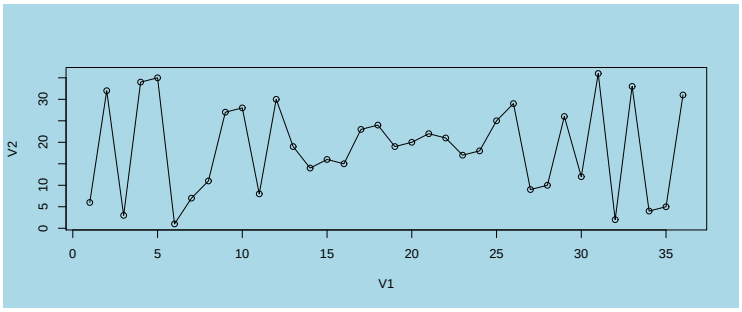
1 i 6	2 Sch 15	3 a 32	4 λ+ 23	5 x 28	6 A 8	7 α 30	8 d 1	143
9 b+ 34	10 a 32	11 c+ 35	12 Z 24	13 λ+ 23	14 B 19	15 b+ 34	16 b- 3	204
17 Sch 15	18 λ+ 23	19 Z- 24	20 Sch 15	21 λ 23	22 Z- 24	23 α 30	24 λ 23	177
25 Z+ 24	26 L 17	27 X 9	28 ↓ 18	29 B 19	30 E 36	31 X 9	32 λ 23	155
33 Z+ 24	34 L 17	35 X 9	36 ↓ 18	37 B 19	38 E 36	39 X 9	40 λ 23	155
41 Z+ 24	42 L 17	43 X 9	44 M 22	45 α 30	46 B 19	47 U 10	48 f 11	142
49 g 27	50 e 7	51 U 10	52 W 13	53 U 10	54 f 11	55 g 27	56 e 7	112
57 U 10	58 W 13	59 P 26	60 E 36	61 Lö 12	62 W 13	63 H* 5	64 St 4	119
164	141	154	169	164	166	154	95	

Abbildung 2.4: Einteilung der Erzählfunktionen des Märchens *Aschenputtel*.

unsere ganze Rede über die Informationstheorie zwingt uns jetzt, eine weitere Interpretation des Wertes zu suchen, die unsere Theorie erweitern kann.

2.3.2 Wert als Selbstinformation in den magischen Quadraten

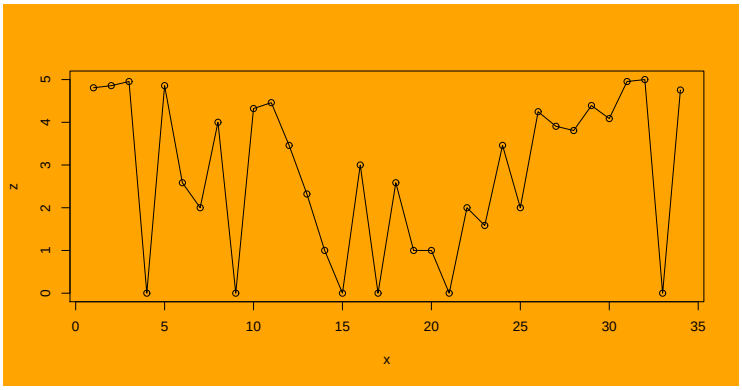
Wenn man die Zahlen der magischen Quadrate in einem kartesischen Koordinatensystem anordnet, sodass auf der Abszisse die Positionen gestellt werden, und auf der Ordinate die entsprechende Werte, kann man die Zeitreihe der Werten besser beobachten[23].



Man bildet danach einen ähnlichen Graph, wobei auf der Ordinate nicht die Werte, sondern das absolute Wert der Differenz jedes Wertes mit dem darauffolgenden, bekommen wir eine Darstellung der Selbstinformation jedes Ereignisses[24]. Diese Andeutung des Informationsinhaltes ist anders als früher theoretisiert worden ist, weil wir diesmal noch keines Erwartungswertes verfügen. Durch den Algorithmus, der in *Erwartete Reihenfolge und Topic Modeling* skizziert worden ist, kann man aus unseren annotierten Märchensammlung schon die Wahrscheinlichkeitswerten für jede Funktion ableiten, die als Basis für eventuelle Untersuchungen auf komplexeren Korpora dienen können.
[25]

Mit welchem Objekt der Erzähltheorie stimmt der Wert überein? Was bedeutet im Rahmen der Erzähltheorie, dass die Summe der Werte jeder Reihe und jeder Spalte immer gleich ist? Was symbolisiert die Reihe? Und die Spalte? *Eine Spalte stellt die paradigmatische Achse der Narration dar.* Obgleich sein durchschnittlicher Wert dem

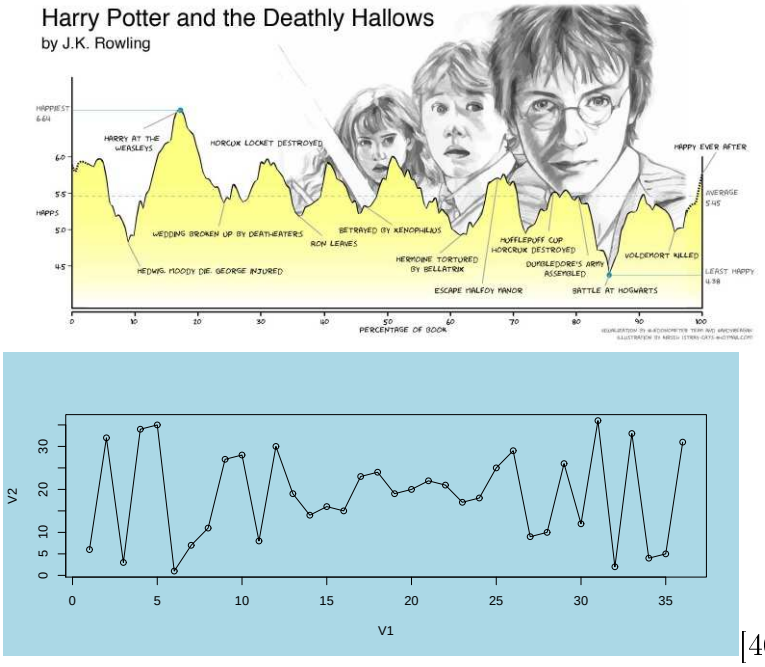
[23]Siehe S. 67, blauer Grafik.
[24]Siehe S. 67, gelber Grafik. Für diese zweite Darstellung wird eine bessere Erklärung benötigt. In dem zweiten Grafik wird offensichtlich die Überraschung (Selbstinformation) als eine Abstand dargestellt. Es geht um eine vereinfachte Version der Differentialrechnung, die später nicht weiter verwendet wird, welche jedoch beiträgt, das Verständnis und die Visualisierung unserer Funktion zu ergänzen.
[25]Jeder Abstand wurde logarithmisiert, um den Bereich der Abbildung begrenzt zu halten.



Wert der Reihen gleicht, sind sie ontologisch grundverschieden, deshalb nicht austauschbar. In einem normalen magischen Quadrat stellt die Spalte nur die vertikale Sicht einer Reihe. Diese Verallgemeinerung kann im Rahmen der Erzählung leider nicht stattfinden. *Eine Reihe stellt eine Episode dar, die nach dem aristotelischen Theorie und nach unserer Theorie der Selbstaffinität der Texten, eine konsistente Texteinheit ist*, die durch die gleiche Typen zerlegbar ist, die das ganze Märchen durchqueren: *deíxis, katastrophé, lýsis*, was heißt: die Episode ist ein Märchen in einem größeren Märchen. Allgemeiner gesagt stimmt unser mathematische *Wert* innerhalb der Erzählung zum Gewicht des besonderen Ereignissen überein. Dieses Gewicht kann entweder statistisch (durch die Selbstinformation), oder psycholinguistisch (durch die *Sentiment analysis*) gemessen werden. Außer den schon erwähnten Methoden, könnte man folgendes überlegen, um wesentlich und *rekursiv* die statistische Herangehensweise durch die Psycholinguistik zu verbessern: der Erwartungswert könnte nur die Spannung überlegen, die dem durchschnittlichen Arbeitsgedächtnis entspricht. Die Erwartungswert wäre zum Beispiel auf den drei vorherigen Ereignissen aufzubauen, die durch ein Parameter justiert werden, das aus den drei vorherigen *Erwartungswerten* abgeleitet wird, d.h. aus den neun vorherigen Ereignissen. Selbstverständlich sollte diese Anzahl durch die Psycholinguistik (oder warum nicht, die Psychoerzähltheorie!) mittels eines empirischen Ansatzes festgestellt werden. Im folgenden Paragraph wird man kurz überlegen, welche Art von Wert bzw. Gewicht die *Sentiment Analysis* zur Lösung unseres Problem anbieten kann, und inwieweit sogar diese Variante unserem allgemeinen Modell passt.

2.3.3 Sentiment Analysis als Wert in der magischen Quadraten

Bei der folgenden Abbildung wird eine Graphik gezeigt, die aus dem Artikel 'The emotional arcs of stories are dominated by six basic shapes' geliehen worden ist[46]. In diesem Fall entspricht der Wert bei jeder Position (d.h. die Seite des Buches, das dadurch behandelt wird) die Glücklichkeit [26]. Die Ähnlichkeit mit dem kartesischen Plan des magischen Quadrates (6x6), der die allgemeine Struktur der allgemeinen Erzählung beschreibt, scheint unbestreitbar zu sein.



[46]

Trotz der großen Anzahl der Teilnehmer diesem Experiment, scheint diese Ansatz nicht ganz allgemein zu sein, da die Reaktionen der Probanden durch eine hedonische Analyse ihrer Gesichtsausdrücke gemessen worden ist. Die Mehrheit der Leuten zeigen wesentlich übertriebene Reaktionen, deshalb wäre es notwendig, um dieses Modell zu fördern, dass dasselbe Experiment durch eine *direkte* Messung der Gehirnstätigkeit erfolgt [27].

[26] Letztendlich ist der Scherz Vonneguts Wissenschaft geworden!
[27] Man könnte dafür entweder Stichproben des Speichels der Probanden vor und nach dem Experi-

2.4 Folge und Struktur IV: Hurst-Exponent und Dependency Distance

Je voudrais avoir écrit une lettre plus courte, mais je n'en ai pas le temps.

Blaise Pascal

Der Hurst-Exponent Das Hurst-Parameter, das bereits in der Sektion 'Selbstähnlichkeiten in Texten' genannt worden ist, wird in zahlreichen wissenschaftlichen Bereichen verwendet, um die Selbstähnlichkeit einer Zeitfolge zu messen. Der Hurst-Exponent ist auch als *Dependenz-Index* anerkannt. Seine Symbole (H), das nicht mit der H der Entropie verwirrt werden soll, ist zwischen 0 und 1 enthalten. $H = 0.5$ zeigt eine randomisierte Folge, wohingegen $H > 0.5$ eine Trend-verstärkende Folge. Je höher ist der H-Wert, desto stärker der Trend ist[28].

Bei der Durchführung unseres Experiments wird dieser Wert für jedes analysierte Märchen automatisch mithilfe der Programmiersprache R berechnet.

Dependenzabstand Als Maß der Komplexität eines Satzes stellt der Dependenzabstand zweifellos eine der wichtigsten dar. Dieses Parameter ist von Haitao Liu eingeführt worden[9]. Nach dieser Heuristik ist die Komplexität eines Satzes durch den Durchschnitt der Dependenzabstände aller Wörter repräsentiert, die den Satz ausmachen. Der Abstand wird jeweils mithilfe des Chomskys hierarchischen Modelles berechnet. Beispielsweise beim Satz

Da gingen sie weiter und kamen an einen See, auf dem schwammen viele, viele Enten.

scheint die Folge der Dependenzabstandswerte wie folgt:

0 0 1 2 0 3 1 1 3 1 2 0 2 1 3

ment nehmen, um die Änderung des Prozentsatz von Serotonin empirisch zu vermessen, oder durch Elektroden, die direkt auf dem Kopf des Probanden gesetzt werden.

[28] Vgl. [45]

Sie und *weiter* hängen von *gingen* ab, *an* und *See* von *kamen*, *einen* und *dem* von *See*, *Enten* von *schwammen*, und *viele* von Enten. Der Durchschnitt aller Werten ergibt die *MDD* (Mean Dependency Distance), die die Komplexität des Satzes messt:

$$MDD = \frac{1}{n-s} \sum_{i=1}^{N} DD_i$$

Bei unserem Beispiel beträgt der durchschnittliche Dependenzabstand 1,33.

Der Dependenzabstand ist wegen seiner Berechnungseinfachheit für unseres Experiment gewählt worden. Auch in diesem Fall spricht die Psycholinguistik eine große Rolle zur Feststellung des Modelles: die Komplexität wird durch das Gehirn als eine Belastung empfunden, die ihm mehr Arbeitsgedächtnis zu verwenden zwingt. In diesem Sinn resultiert die Feststellung eines Maßes der Komplexität in einer Folge einfacher als in einer Struktur, wobei im Gegensatz mehrere Parameter angenommen werden sollen.

The Hurst exponent can be calculated by rescaled range analysis (R/S analysis). For a time series, $X = X_1, X_2, \ldots Xn$, R/S analysis method is as follows:

(1) Calculate mean value m.

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(2) Calculate mean adjusted series Y

$$Y_t = X_t - m, \quad t = 1, 2, \ldots, n$$

(3) Calculate cumulative deviate series Z

$$Z_t = \sum_{i=1}^{t} Y_i, \quad t = 1, 2, \ldots, n$$

(4) Calculate range series R

$$R_t = max(Z_1, Z_2, \ldots, Z_t) - min(Z_1, Z_2, \ldots, Z_t)$$
$$t = 1, 2, \ldots, n$$

(5) Calculate standard deviation series S

$$S_t = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(X_i - u)^2} \quad t = 1, 2, \ldots, n$$

Here u is the mean value from X_1 to X_t.

(6) Calculate rescaled range series (R/S)

$$(R/S)_t = R_t/S_t \quad t = 1, 2, \ldots, n$$

Note $(R/S)_t$ is averaged over the regions $[X_1, X_t]$, $[X_{t+1}, X_{2t}]$ until $[X_{(m-1)t+1}, X_{mt}]$ where m=floor(n/t). In practice, to use all data for calculation, a value of t is chosen that is divisible by n.

[45]

3 Lýsis

3.1 Zusätzliche Parameter des Experiments

In diesem Sektion wird eine Liste aller Parameter gezeigt, die beim Experiment vorkommen. Vor der Erfüllung dieser Aufgabe lohnt es sich nun ein Wort über die Beziehung zwischen Komplexität und Information zu widmen. Man hat bereits früher erwähnt, dass je wahrscheinlicher ein Ereignis ist, desto weniger (Selbst)information trägt. Man hat auch hervorgehoben, dass nach der synergetischen Linguistik und der UID-Theorie die Ausdrückseinfachheit ein wesentlicher Trend der Sprache darstellt, sodass eine Korrelation zwischen Länge und Informationsinhalt der Texteinheit entsteht. Man hat auch gezeigt, dass die Länge mit der Komplexität korreliert ist, woraus man ableiten kann, dass Information und Komplexität verwandt sind[1]. Die Komplexität bereitet den Boden der Information, die einfacher in komplexeren Umständen *hervorbrechen* kann. Einfache Strukturen sind alle zueinander ähnlich, wohingegen komplexe Strukturen können auch grundverschieden voneinander sein. Die Tatsache, dass die komplexe Strukturen höchstwahrscheinlich eine *Umkehrung der gewöhnlichen Ordnung der Sprache* darstellen, macht Komplexität und Information zwei Begriffe, die im Zusammenhang verwendet werden können. Schließlich scheint es uns auch sehr sinnvoll als Maß des Informationsinhaltes eines Satzes die Zahl der Narratemen hinzuzufügen. Wenn eine Funktion zum Beispiel durch 3 Sätze gestreut ist, dann wird jeder Satz den Wert 0.333 bekommen. In diesem Sinn ist ein Satz, der allein eine Funktion beinhaltet, wesentlicher aus einem informativen Standpunkt. Diese knappe Parenthese ermöglicht uns implizit die Parameter I und

[1] Vgl. [29]

F zu verstehen, die im Nachstehenden unter anderen gelistet werden:

- SS: Satzgefüge.
- #S: Anzahl der Sätze im Satzgefüge.
- #W: Anzahl Wörter im Satzgefüge.
- DWL: Durchschnittliche Wortlänge.
- DSL: Durchschnittliche Satzlänge
- ADD: Durchschnittliche Dependenzabstand.
- #P: Anzahl Narratemen im Satzgefüge.
- I: Vereinigung von ADD und #P durch Multiplikation.
- O: Verhältnis von I mit dem Wörterzahl.
- F: Vereinigung von ADD und #P durch Addition.
- A: Verhältnis von F mit dem Wörterzahl.

3.2 Durchführung des Experiments

Für unseres Experiment sind zwei kurze (*Der alte Großvater und sein Enkel* und *Das Bäuerlein im Himmel*) und eine lange Erzählung (*Die Bienenkönigin*) ausgewählt worden. Sie sind wegen ihrer Kürze manuell bearbeitet worden. Die Berechnung aller Variablen ist durch *Excel* erfolgt. Für größere Datensammlungen könnte man vernünftig überlegen, die Texte automatisch zu verarbeiten, was für alle oben genannte Maße sogar trivial scheint. Problematisch bleibt jedoch die Berechnung der Dependenzabstand, den nicht so naiv algorithmisiert werden kann. Nach der Bestimmung der Parameter für jedes Märchen, hat man zunächst eine Graphik herstellt, um unmittelbar visuell merken

zu können, ob einige merkwürdige Verhältnisse zwischen Parameterpaaren es gibt[2]. Der Selbstähnlichkeitsgrad der Zeitfolgen jedes Parameters wurde durch den Hurst-Exponent berechnet.

3.2.1 Korpusanalyse durch alle obengenannte Maße

Die berechnete Werte der Märchen scheinen wie folgt[3]:

Um die Anpassung derselben Graphik zu ermöglichen, hat man die Werte, die die *durchschnittliche Satzlänge* und die *Wörteranzahl* entsprechen, durch 10 dividiert, und diejenige die O und A entsprechen, durch 10 multipliziert.

3.2.2 Ergebnisse

Hurst-Parameter für die Reihenfolge der Fibonacci Zahlen

- **Simple R/S Hurst estimation: 0.6333257**
- Corrected R over S Hurst exponent: 0.8223557
- Empirical Hurst exponent: 0.8223557
- Corrected empirical Hurst exponent: 0.6776776
- Theoretical Hurst exponent: 0.6464496

Hurst-Parameter für die Reihenfolge der Wertdifferenzen beim magischen Quadrat 6 x 6

- **Simple R/S Hurst estimation: 0.6817705**
- **Corrected R over S Hurst exponent: 1.01137**
- **Empirical Hurst exponent: 1.01035**

[2]Dies bedeutet, dass die gezeigte Grafiken nicht alle mögliche Permutationen zwischen den Werten zeigen, sondern nur die Paaren, die bei der Abszissenachse die Reihe der Sätze tragen.

[3]Siehe Seite 84.

- **Corrected empirical Hurst exponent: 0.892741**
- Theoretical Hurst exponent: 0.6005503

Hurst-Parameter für die Reihenfolge der durchschnittlichen Dependenzabstände in *Der alte Großvater und sein Enkel*

- Simple R/S Hurst estimation: 0.4971043
- Corrected R over S Hurst exponent: NA
- Empirical Hurst exponent: 0.9348219
- Corrected empirical Hurst exponent: 0.6583566
- Theoretical Hurst exponent: 0.7017126

Dieses Ergebnis zeigt, dass leider *Das Bäuerlein im Himmel* und *Der alte Großvater und sein Enkel* zu kurz sind, um einen präzisen H-Exponent aus ihren Variablen ableiten zu können, deshalb wird unsere Untersuchung durch den Hurst-Exponent ausschließlich auf *Die Bienenkönigin* sich begrenzen. Im folgenden werden nur die bedeutendste Ergebnisse gelistet, die ein Hurst Exponent größer gleich 0,6 nachweisen können:

Hurst-Parameter für die Reihenfolge der A-Werte in *Die Bienenkönigin*

- **Simple R/S Hurst estimation: 0.5992169**
- **Corrected R over S Hurst exponent: 0.8142752**
- **Empirical Hurst exponent: 1.083673**
- **Corrected empirical Hurst exponent: 0.9163009**
- **Theoretical Hurst exponent: 0.6005503**

Hurst-Parameter für die Reihenfolge der Satzlänge in *Die Bienenkönigin*

- ***Simple R/S Hurst estimation: 0.6163787***

- *Corrected R over S Hurst exponent: 0.9291284*
- *Empirical Hurst exponent: 1.206166*
- *Corrected empirical Hurst exponent: 1.026695*
- Theoretical Hurst exponent: 0.597666

Hurst-Parameter für die Reihenfolge der durchschnittlichen Wortlänge in *Die Bienenkönigin*

- **Simple R/S Hurst estimation: 0.6186447**
- Corrected R over S Hurst exponent: 0.7395739
- Empirical Hurst exponent: 0.7256099
- Corrected empirical Hurst exponent: 0.6346316
- Theoretical Hurst exponent: 0.597666

Hurst-Parameter für die Reihenfolge der durchschnittlichen Satzlänge in *Die Bienenkönigin*

- **Simple R/S Hurst estimation: 0.6191298**
- Corrected R over S Hurst exponent: 0.8182548
- Empirical Hurst exponent: 0.9031126
- Corrected empirical Hurst exponent: 0.7828386
- Theoretical Hurst exponent: 0.597666

Die folgende Abbildungen bieten andererseits ein Überblick über die Verteilungen der Variablen in jedem Märchen[4].

[4] siehe Seite 85 (erste Abbildung: Das Bäuerlein im Himmel; zweite Abbildung: der alte Großvater und der Enkel) und Seite 86 (Die Bienenkönigin).

3.3 Überlegungen über die Ergebnisse

Die Ergebnisse, die in diesem Abschnitt diskutiert werden, betreffen ausschließlich das finale Experiment, und nicht die andere Überlegungen, die die ganze These durchqueren. Am Anfang beweist der Hurst-Exponent effizient zu sein. Es würde getestet auf zwei Zeitreihen, derer Selbstähnichkeit schon früher durch andere Methoden bewiesen worden ist: die Fibonacci-Folge und das magische Quadrat. In beiden Fälle scheint der Hurst-Parameter unsere vorherigen Ergebnisse zu bestätigen. Es wurde ferner behauptet, dass das Verhalten der Elementen eines Satzgefüges (oder, besser gesagt, seiner Werte) nach dem Selbstähnlichkeitsprinzip auf der Ebene des Textes widergespiegelt wird: der H-Wert der durchschnittlichen Wortlänge, der durchschnittlichen Satzgefügelänge und der Satzlänge sind tatsächlich identisch[5]. Man hat erfunden, dass das A-Wert wesentlich bessere H-Werte als der O-Wert liefert, was bedeutet, dass zwei Informationsmaße besser summiert statt miteinander multipliziert werden sollen. Unsere Datensammlung ist zu klein, und meine Kenntnisse der Statistik noch ungenügend, um tiefer dieses Verhalten anzudeuten. Aus diesem Grund gilt die vorherige Behauptung nur für ähnliche Experimente. Bei dem Märchen *die Bienenkönigin* kann man bemerken, dass die durchschnittliche Satzlänge und der A-Wert durch den H-Exponent eng verbunden sind, was das UID-Modell fördert. Darüber hinaus scheint es bei der Verteilung vieler seiner Variablen sehr ähnlich der Verteilung des magischen Quadrat zu sein. Wie schon erwähnt stellt diese Beobachtung kein Zufall: was ein magisches Quadrat ist, wurde ausführlich und präzis beschrieben, als auch seine Verwandtschaft mit der Morsefolge (gleiche Summe der Reihen und der Spalten in der Matrize, nachgewiesene Palindromie für die gerade Ordnungen), die wiederum auf der Informationstheorie bezogen ist. Diese Ähnlichkeiten und Verwandtschaften sind ohnehin nicht genug untersucht worden, um besser das allgemeine Ziel und Thema dieser Studie zu berücksichtigen. Schließlich möge man hier überlegen, dass das Experiment widersprüchlich mit dem Paragraph Äscombes Quartett und Isoparametrische Texte: eine wissenschaftstheoretische Parentheseäussehen könnte, da wir eine rein analytisch konnotierte Analyse durchgeführt haben. Es scheint vielleicht

[5]Dies stellt die Verwirklichung der *Skaleninvarianz* in unserer Datensammlung.

nötig genauer zu bestimmen, dass man dafür kein willkürliches Modell gewählt hat, und keine Verallgemeinerung erfolgt ist: die vergleichende Analyse ist nur im Sinn der Messung der Selbstaffinität auf einer quantitativen Ebene durchgeführt worden.

Verbesserungsmöglichkeiten Das letzte Experiment könnte durch mehrere Daten wesentlich verbessert werden. Die Ähnlichkeit zwischen den Verteilungen sollte mittels eines statistischen Testes für die Übereinstimmung zweier Verteilungen (wie zum Beispiel das Kolgmorov-Smirnov Test) mathematischer Weise überprüft. Auch dafür verfügt die Programmiersprache R einer Funktion, die das Test automatisch durchführen kann. Ihre Verwendung ist in diesem Fall vermieden worden, weil eine kompetente Interpretation des Ergebnisses von mir nicht geführt werden konnte. Ähnliche vergleichende Analyse könnte man auch auf den Werten durchführen, die durch die Verwendung der anderen Information- und Komplexitätsmaßen bekommen worden sind, die im Laufen dieses Traktates betrachtet worden sind. Darüber hinaus könnte man eine vergleichende Analyse nicht zwischen den Variablen, sondern zwischen den unterschiedliche Methoden tun, sodass besser verstanden werden kann, inwieweit Information, Gefühl[6], und theoretischer 'magische' Wert miteinander verbunden sind. Hierfür folgt ein kleines Beispiel, das unsere Erwartung bereits erhöht:

Hurst-Exponent der Verteilung der theoretischen Werte[7] in *Aschenputtel*

Simple R/S Hurst estimation: 0.613473
Corrected R over S Hurst exponent: 0.8066844
Empirical Hurst exponent: 0.5014202
Corrected empirical Hurst exponent: 0.549992
Theoretical Hurst exponent: 0.5174121

[6]Das hier erwähnte Gefühl spiegelt die gleiche Andeutung, die im Paragraph SSentiment Analysis als Wert in der magischen Quadraten"besprochen worden ist.
[7]Werte der Proppfunktionen, die in dem magischen Quadrat der Sonne eingeteilt sind.

3.3.1 Zusammenfassung der eigenen Leistung dem Forschungsbedürfnis

In diese Abhandlung habe ich der Computererzähltheorie durch die folgende Leistungen beigetragen:

- Jede Propps Funktion ist tatsächlich in 3 Elementen weiter zerlegbar, die einer der Variablen der Erzählung entsprechen: *Deíxis, Katastrophé, Lýsis* (Hjelmslev Variablen).

- Die Zusammensetzungsregeln dieser Variablen können effizient durch ein Lyndenmeyer System ausgedrückt werden.

- Die theoretische Funktionenfolge beinhaltet 36 Elementen. Wenn jeder Funktion den Wert der entsprechenden Position in einem magischen Quadrat der Sonne zugewiesen wird, findet man in tatsächlichen Beobachtungen heraus, dass der Durchschnitt der Summe der Reihen und der Spalten $\simeq 111$ ist, genau wie im magischen Quadrat der Sonne geschieht (Vgl. *Frau Holle*).

- Längere beobachtete Funktionenfolgen (Vgl. *Aschenputtel*) behalten auch eine Äquivalenz zwischen den Durchschnitten der Summe jeder Reihe und Spalte, wenn man jeder Funktion den Wert zuweist, der in der theoretische Funktionenfolge vorkommt. Dies bedeutet, dass das magische Quadrat der Sonne als Modell auch der Erzählungen größerer Länge dienen kann.

- Der Hurst-Exponent der längeren Erzählungen (Vgl. *Die Bienenkönigin*) über quantitative Texteinheiten (Satzgefügelänge, Satzlänge, Wortlänge) scheint gleich zu sein, was unsere anfängliche Annahme beweist.

- In *Die Bienenkönig* kann man noch bemerken, dass alle die obengenannte Variablen in einer 'magischen' Art sich verteilen. Dieses Ergebnis bestätigt die *Uniform Information Density* Theorie und fügt auch etwas hinzu: die Verteilung der Information durch die Leitung soll nicht nur 'aufpassen', uniform zu sein. Sie soll auch

alle die Bedingungen erfüllen, die Selbstinformation am höchstens zu halten. Diese beide Bedingungen sind durch das Modell des magischen Quadrates erfüllt, da es als eine Erweiterung der Morsefolge angesehen werden kann, die auch Zahlen außer 0 und 1 beinhaltet.

Außer dem Forschungsbedürfnis sind wegen Zufall(sentdeckungen) auch zwei Leistungen erzielt worden, die die Fibonacci-Folge und die magischen Quadrate betreffen:

$$x(z) = \frac{[x(z-1)]^2}{[x(z-2)]} \tag{3.1}$$

$\forall i > 1 \wedge i \in N$
und

$$\forall i \in M = n^2 \wedge x \in X = n^2 \Rightarrow M[i] = x \wedge M[x] = i \rightarrow M[M[x]] = x$$

Die neue Eigenschaften, die in diesen bekannten mathematische Strukturen erfunden worden sind, haben uns geholfen, mit passenden mathematischen Werkzeugen die Forschung auszustatten.

3.4 Schluss

Angesichts der jüngsten Entwicklungen des NLPs[8] scheint es vernünftig sich zu erwarten, dass die quantitative Sprachforschung weitere Terminologie einschließen wird, die bis jetzt ausschließlich qualitative Bedeutung getragen hat. Ein Beispiel dafür wäre *lustig*, das als eine bestimmte Spezifikation von *überraschend* gelten kann, und diesbezüglich mathematisch modelliert werden könnte. Wenn ein Software, das das Lustiges in einem Text zu erkennen fähig ist, überhaupt nutzlos aus ästhetischen und menschlichen Gründen ist (der Mensch braucht keine maschinelle Unterstützung, um zu *genießen*), wäre eine derartige Technologie aus dem kreativen Standpunkt sehr interessant, da sie dem Mensch die Möglichkeit vorstellen würde, bei der Schöpfung eines Textes auf allen

[8]Natural Language Processing, d.h. 'linguistische Datenverarbeitung'.

trivialen Aspekten wie die Organisation, die Anordnung, und die Darstellung der Ereignissen, die schon oder noch nicht gebaut worden sind, verzichten zu können. In anderen Wörtern:

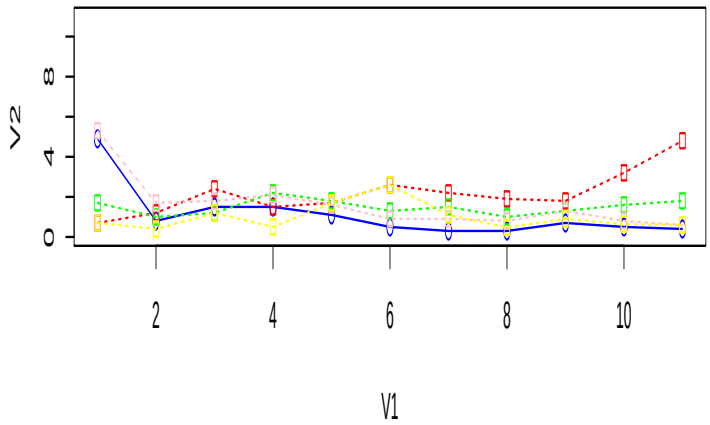
> Alleviate the authoral bottleneck problem in narrative generation and storytelling systems by leveraging automated narrative information extraction.[9]

Startend aus der Analyse der Funktionen Propps wäre es auch möglich einen Graph herstellen zu können, mithilfe dessen eine Geschichte *synchronisch* visualisierbar ist.

Aus praktischen Gründen stellt die Informationsvisualisierung unter anderem eine wichtige Anwendungsgelegenheit. Funktionen, die im Graph eine höhere Anzahl von Pfeile bekommen, würden höchstwahrscheinlich zum zweiten Teil der Erzählung gehören, d. h. zum Teil, die zur Entwicklung der Beziehungen zwischen Charakteren und Ereignissen zuständig ist. Wichtig wäre auch die Möglichkeit zu haben, *rückwärts* von einem Graph zu einem Text zurückzugehen, was wesentlich die kreative Prozesse unterstützen und beschleunigen kann. Die *Computerpoetik* wird eine neue künstlerische Ära feierlich eröffnen. Hiermit haben wir gezeigt, aus welchen Bausteinen man eine unifizierte Theorie der Sprache zu begründen beginnen kann, um definitiv das alte *Form-Inhalt Modell* zu überwinden. Der Begriff von *Struktur* allein stellt dafür vielleicht kein Schlüssel, aber bestimmt eine Tür dar.

[9] Vgl. [17]

				Das Bäuerlein im Himmel						
SS	#S	#W	DWL	DSL	ADD	#P	I=ADD*#P	O= I/#W	F=ADD+#P	A=F/#W
1	1	7	5	7	1,7	2	3,4	0,49	3,7	0,53
2	3	12	4,1	4	1	1	1	0,08	2	0,17
3	2	24	4,3	12	1,2	3	3,6	0,15	4,2	0,18
4	3	15	4,5	5	2,2	1	2,2	0,15	3,2	0,21
5	1	17	4,4	17	1,8	1	1,8	0,11	2,8	0,16
6	1	26	4,8	26	1,3	1	1,3	0,05	2,3	0,09
7	2	22	5	11	1,5	0,5	0,75	0,03	2	0,09
8	4	19	4,4	4,8	1	0,5	0,5	0,03	1,5	0,08
9	2	18	5	9	1,3	1	1,3	0,07	2,3	0,13
10	5	32	4,1	6,4	1,6	1	1,6	0,05	2,6	0,08
11	8	48	4,4	6	1,8	1	1,8	0,04	2,8	0,06
				Der alte Großvater und sein Enkel						
SS	#S	#W	DWL	DSL	ADD	#P	I=ADD*#P	O= I/#W	F=ADD+#P	A=F/#W
1	3	20	4.1	6.7	1,4	1	1,4	0,07	2,4	0,12
2	4	29	4,1	7,3	1,3	1	1,3	0,04	2,3	0,08
3	3	40	4.7	13.3	1,5	1	1,5	0,04	2,5	0,06
4	1	13	4	13	1,2	0,5	0,6	0,05	1,7	0,13
5	3	16	5	5.3	2,7	0,5	1,35	0,08	3,2	0,2
6	2	11	4.4	5.5	1	1	1	0,09	2	0,18
7	2	16	5	8	2	1	2	0,13	3	0,19
8	2	17	4,5	8.5	2,6	0,5	1,3	0,08	3,1	0,18
9	2	7	4	3.5	1	0,5	0,5	0,07	1,5	0,21
10	3	17	4,6	5,7	1,6	1	1,6	0,09	2,6	0,15
11	7	38	4	5.3	3,15	1	3,15	0,08	4,15	0,11
					Die Bienenkönign					
SS	#S	#W	DWL	DSL	ADD	#P	I=ADD*#P	O= I/#W	F=ADD+#P	A=F/#W
1	3	21	5,2	7	1,5	1	1,5	0,07	2,5	0,12
2	2	13	4,6	6,5	1,6	0,5	0,8	0,06	2,1	0,16
3	5	31	5	6,2	1	0,5	0,5	0,02	1,5	0,05
4	2	11	5	5,5	2	1	2	0,18	3	0,27
5	4	19	5	4,8	1,1	1	1,1	0,06	2,1	0,11
6	1	4	5	4	0,5	1	0,5	0,13	1,5	0,38
7	4	12	4,3	3	1,25	1	1,25	0,1	2,25	0,19
8	3	14	4,1	4,7	1,35	1	1,35	0,1	2,35	0,17
9	4	18	4,4	4,5	1,55	1,5	2,33	0,13	3,05	0,17
10	3	12	4,3	4	1,25	0,5	0,63	0,05	1,75	0,15
11	3	16	4,7	5,3	1,1	1	1,1	0,07	2,1	0,13
12	3	18	4,6	6	1,1	1	1,1	0,06	2,1	0,12
13	2	9	4,4	4,5	1,5	1	1,5	0,17	2,5	0,28
14	3	12	4,6	4	1,25	1	1,25	0,1	2,25	0,19
15	2	16	5	8	1,5	1,5	2,25	0,14	3	0,19
16	5	25	4,1	5	1,2	0,5	0,6	0,02	1,7	0,07
17	2	17	4	8,5	1,6	0,5	0,8	0,05	2,1	0,12
18	2	11	4,1	5,5	1	0,5	0,5	0,05	1,5	0,14
19	2	10	4,2	5	0,6	0,5	0,3	0,03	1,1	0,11
20	4	16	4,4	4	0,8	0,5	0,4	0,03	1,3	0,08
21	2	13	4,5	6,5	1,5	0,5	0,75	0,06	2	0,15
22	3	14	5,3	4,7	1,6	0,5	0,8	0,06	2,1	0,15
23	5	29	5,5	5,8	1,2	0,5	0,6	0,02	1,7	0,06
24	6	38	4,7	6,4	1	0,5	0,5	0,01	1,5	0,04
25	4	21	4,1	5,2	1,2	1	1,2	0,06	2,2	0,1
26	3	13	4	4,2	1	1	1	0,08	2	0,15
27	4	30	4,8	7,5	2,6	1	2,6	0,09	3,6	0,12
28	2	12	4,2	6	1,5	2	3	0,25	3,5	0,29
29	5	22	4,1	4,4	1,6	0,5	0,8	0,04	2,1	0,1
30	6	37	4,6	6,16	1,2	0,5	0,6	0,02	1,7	0,05
31	2	16	5	8	1	1	1	0,06	2	0,13
32	4	24	4,6	6	1,2	0,5	0,6	0,03	1,7	0,07
33	1	7	5	7	1	0,5	0,5	0,07	1,5	0,21
34	2	15	5	7,5	1,8	0,3	0,54	0,04	2,1	0,14
35	4	35	5	7	1,3	0,3	0,39	0,01	1,6	0,05
36	6	41	4,6	6	1,2	0,3	0,36	0,01	1,5	0,04
37	1	5	4	5	1,3	0,5	0,65	0,13	1,8	0,36
38	3	16	5	5	2	0,5	1	0,06	2,5	0,16
39	2	17	4,8	8,5	1,2	0,5	0,6	0,04	1,7	0,1
40	1	9	5,9	9	1,2	0,5	0,6	0,07	1,7	0,19



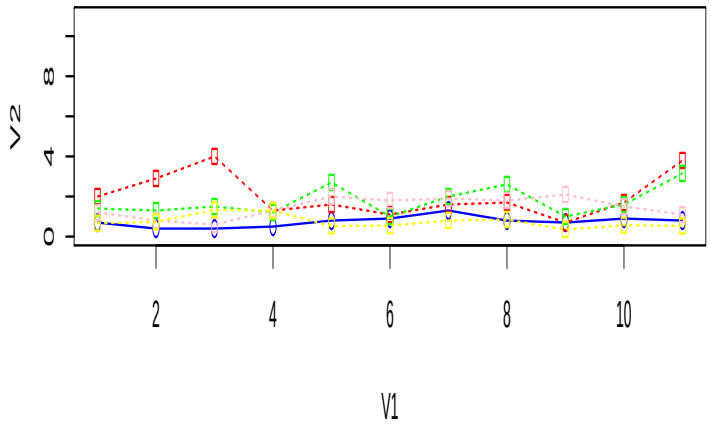


Abbildung 3.1: Die x-Achse stellt die Zeitreihe der Satzgefügen dar. Die rote Gerade zeichnet die Verteilung der Anzahl der Wörter, die gelbe die durchschnittliche Satzlänge, die grüne den durchschnittlichen Dependenzabstand, die blaue das Parameter O und die pinke das Parameter A.

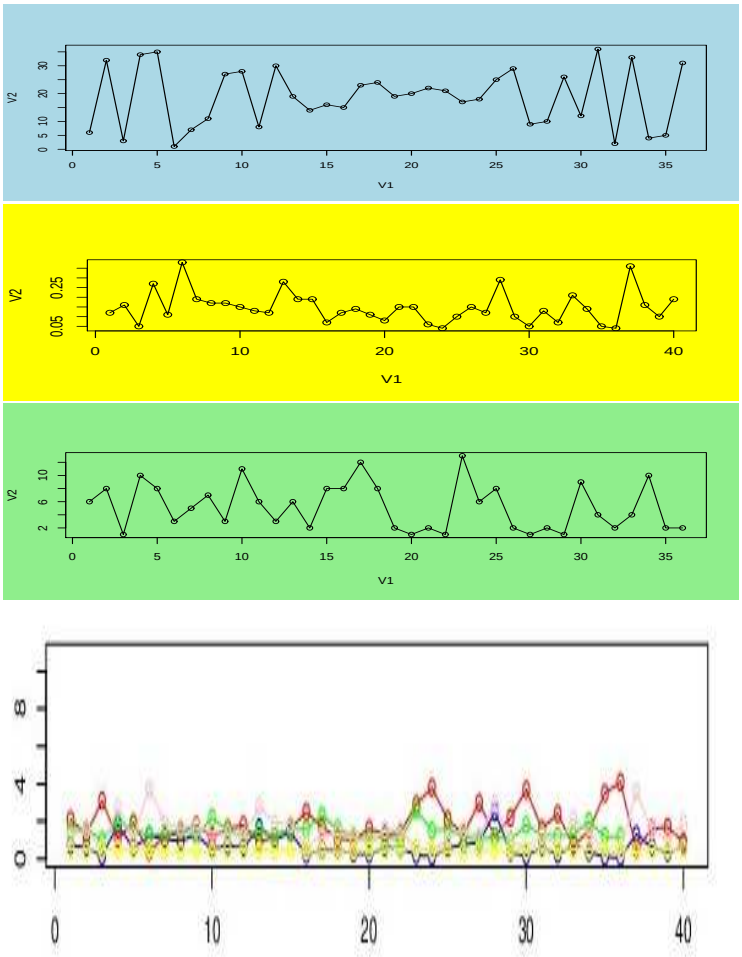


Abbildung 3.2: Die Abbildung zeigt, in Ordnung: das magische Quadrat 6x6, die Verteilung der A in *Die Bienenkönigin*, die Verteilung des Depenenzabstandes in *Die Bienenkönigin* und die Verteilungen der Parameter, die auch für die andere Märchen ausgewählt worden sind.

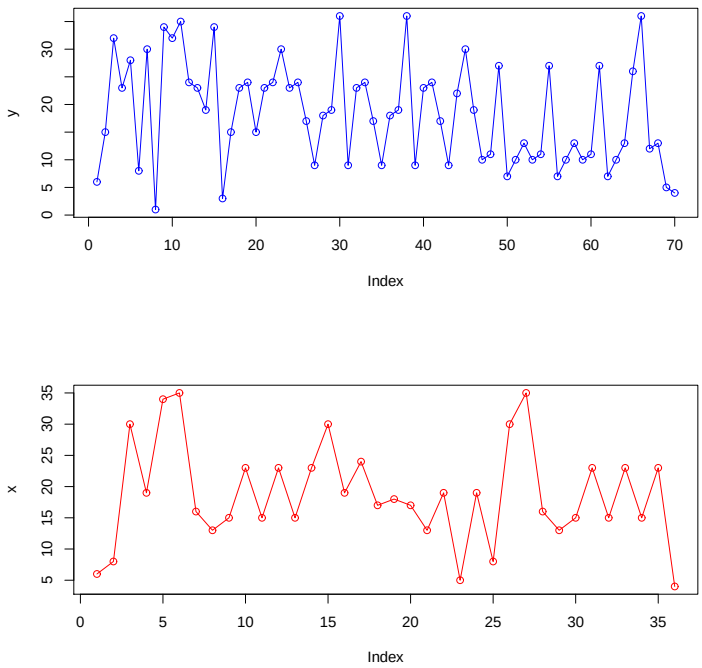


Abbildung 3.3: Verteilung der theoretischen Werte in Aschenputtel und Frau Holle.

4 Anhang

4.0.1 Tafel der Propps Funktionen von 16 Grimms Märchen

fisherfrau	Rotkäppchen	Haselgel	HänselGretel	FrauHolle	Goldenegans	Rumpelstilzc	Fundevogel	Froschkönig	AlteSultan	Bienenkönig	DieklugElse	GroßvaterEn	GestiefeltKat	ArmeReiche	Dornröschen
i	i	i	i	i	i	i	i	i	i	i	i	i	i	i	i
>>	b+	>>	f	A	a	b+	a	a	d	a	X-	x	a	Sch	α
Sch	b-	a	e	B	Z	B	A	x	e	>>	U-	α	α	λ-	Sch
λ+	>>	d	Z	b+	W	b+	R	α	α	B	b+	C	B	W	Z
d	a	λ+	f	C	Sch	α	f	Sch	B	α	c+	Str	Sch	Sch	L
e	f	λ-	S	>>	λ-	Sch	g	B	f	L	α	λ+	λ+	λ+	B
α	g	λ+	f	W	α	λ+	d	Sch	g	W	B	A	Z	Z	α
b+	f	λ-	A	Sch	<<	Sch	B	λ+	Z	α	b+	L	C	Z	Sch
C	g	f	g	λ+	a	λ+	C	L	L	L	c+	P	>>	Z	Z
>>	g	g	α	Sch	Z	Z	>>	<<	H*	W	d	Lö	f	B	x
c+	h	f	L	λ+	W	L	A-	V	i	d	α	H*	g	V	Z
λ+	f	g	<<	Sch	Sch	b+	b+	B	f	f	B	od	x	f	C
Z	g	x	Lö	λ+	λ-	B	c+	V	λ-	g	b+		A	g-	x
<<	f	f	f	α	α	b+	>>	b+	x	Sch	c+		W	Sch	A
H*	e	<<	e	B	<<	α	A-	c+	A-	λ-	d		Sch	λ-	B
α	A	B	Z	L	b	Sch	<<	b+	W	A	α		λ+	St	C
b+	h	C	f	W	b+	λ+	b+	c+	K	Sch	B		f	St	M
C	e	>>	S	<<	Z	Sch	c+	b+	S	λ-	b+		g	St	Pr
>>	f	Z	f	L	>>	λ+	>>	c+	L	A	c+		A	od	Lö
c+	f	W	A	Z+	Sch	Z	A-	b+	E	Sch	d		W		Sch
λ+	f	K	g	W	λ+	L	<<	c-	H*	λ+	α		Sch		λ+
Z	A	Sx73!	α	B	T	b+	b+	T	od	α	B		λ+		Z
<<	α	St	Lö	H*	Z	B	c+	B		P	C		B		T
H*	B	H*	<<	>>	W	b+	>>	X		Sch	>>		d		L
α	E-	od	Lö	W	f	α	A-	E		λ+	d		f		H*
b+	R		x	Sch	A-	Sch	<<	Ü		P	α		g		od
C	B		α	λ-	f	λ+	S	H*		Lö	B		x-		
>>	St		Z	Sch	A-	Sch	L	od		Sch	C		A-		
c+	H*		g	λ-	f	λ+	<<			λ+	>>		f		
λ+	od		f	Sch	A-	Z	H*			L	d		g		
Z			g	λ-	f	L	od			T	B		M		
<<			A	α	A-	b+				H*	α		Sch		
H*			f	B	f	B				od	H*		λ+		
α			g	L	A-	b+					i		f		
b+			b+	W	B	α					a		g		
C			g	<<	Pr	Sch					b+		A-		
>>			b+	L-	λ+	λ+					c+		f		
c+			g	Z-	L	Sch					f		g		
λ+			f	W	Pr	λ+					g		A-		
Z			g	B	λ+	Z					f		W		
<<			L	St	L	L					g		f		
H*			R	od	Pr	H*					f		g		
α			L		λ+	α					g		f		
b+			<<		L	B					B		g		
C			α		Pr	P					S		x		
>>			Sch		λ+	λ-					α		f		
c+			w		L	P					B		g		
λ+			λ+		H*	λ-					λ+		x		
Z			H*		od	Z					A		K		
<<			od			P					St		S		
H*						λ+					od		L		
α						St							T		
b+						od							H*		
C													od		
>>															
c+															
λ+															
Z															
<<															
H*															
St															
od															

a

Abbildungsverzeichnis

1.1 Die Mandelbrots- und Kochs Kurven gehören zu den berühmtesten Fraktalen. 4
1.2 Ascombes Quartett (Quelle: Wikipedia.) 8
1.3 Ein Beispiel der Greimas´ Funktionenbearbeitung. 37

2.1 Abbildung einer Hilberts Kurve, die vom oben beschriebenen Formalismus erzeugt wird. 56
2.2 Einteilung der theoretischen Erzählfunktionen im magischen Quadrat der Sonne (6 x 6). 64
2.3 Einteilung der Erzählfunktionen des Märchens *Frau Holle*. 66
2.4 Einteilung der Erzählfunktionen des Märchens *Aschenputtel*. 67

3.1 Die x-Achse stellt die Zeitreihe der Satzgefügen dar. Die rote Gerade zeichnet die Verteilung der Anzahl der Wörter, die gelbe die durchschnittliche Satzlänge, die grüne den durchschnittlichen Dependenzabstand, die blaue das Parameter O und die pinke das Parameter A. 86
3.2 Die Abbildung zeigt, in Ordnung: das magische Quadrat 6x6, die Verteilung der *A* in *Die Bienenkönigin*, die Verteilung des Depenenzabstandes in *Die Bienenkönigin* und die Verteilungen der Parameter, die auch für die andere Märchen ausgewählt worden sind. 87
3.3 Verteilung der theoretischen Werte in Aschenputtel und Frau Holle. . . . 88

Literaturverzeichnis

[1] Gülich E. / Raible W., *Linguistische Textmodelle. Grundlagen und Möglichkeiten*, München, 1977.

[2] Propp V., *Morphologie des Märchens*; Hg. von Karl Eimermacher, Frankfurt a. M., 1975.

[3] Moretti F., *Distant Reading*, Verso Books, 2013.

[4] Si M., Thue D., André E., Lester J., Tanenbaum J., Zammitto V., *Interactive Storytelling.* 4th International Conference on Interactive Digital Storytelling, ICIDS 2011, Vancouver, Canada, November 28-1 December, 2011, Proceedings. Springer Science & Business Media, 2011.

[5] Blei D. M., *Introduction to Probabilistic Topic Models*, 2012.

[6] Zarcone A., Van Schijndel M.,Vogels J., Demberg V., *Salience and Attention in Surprisal-Based Accounts of Language Processing.* Frontiers in psychology 7. Jg., 2016, S. 844.

[7] Meyer P., *Two semi-mathematical asides on Menzerath-Altmann´s law*; in: Exact Methods in the Study of Language and Text, 2007.

[8] Chambers N., Jurafsky D., *Unsupervised Learning of Narrative Event Chains.* Proceedings of ACL-08: HLT, 2008, S. 789-797.

[9] Liu H., *Dependency Distance as a Metric of Language Comprehension Difficulty*; in: Journal of Cognitive Science 9. Jg., Nr. 2, 2008, S. 159-191.

[10] Jaeger T. F., *Redundancy and reduction: Speakers manage syntactic information density.* Cognitive psychology, 61. Jg., Nr. 1, 2010, S. 23-62.

[11] Todorov T., *Théorie de la littérature*; Seuil, 1966.

[12] Scheidel A., Declerck T., *APftML – Augmented Proppian fairy tale Markup Language*; in: Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts, 2010, S. 95.

[13] Rostagni A., *Aristotele poetica*, Biblioteca di Filologia Classica, 1945.

[14] Eco U., *Interpretazione e sovrainterpretazione.* un dibattito con Richard Rorty, Jonathan Culler e Christine Brooke-Rose. Giunti, 2012.

[15] Propp V., *Fairy Tale Transformations.* Narrative Dynamics: Essays on Time, Plot, Closure, and Frames, 1928, S. 73-93.

[16] Lancioni G., Villano R., Romani F. R., *The Self-Similar Qur'an Project: Automatic Detection of Internal Similarity in Classical Arabic Texts*, in: Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on. IEEE, 2016, S. 355-360.

[17] Valls-Vargas J., Zhu J., Ontañon S., *Predicting Proppian Narrative Functions from Stories in Natural Language*, in: Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference. 2016.

[18] Declerck T., Koleva N., Krieger H., *Ontology-Based Incremental Annotation of Characters in Folktales*, in: Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. Association for Computational Linguistics, 2012, S. 30-34.

[19] Cordeiro J., Inácio P. R. M., Fernandes D. A. B., *Fractal Beauty in Text*, in: Portuguese Conference on Artificial Intelligence. Springer, Cham, 2015, S. 796-802.

[20] Finlayson M.A, *Inferring Propp's Functions from Semantically Annotated Text*, Journal of American Folklore 129. Jg., Nr. 511, 2016, S. 55-77.

[21] Lebowitz M., *Story Telling as Planning and Learning*, Poetics, 14. Jg., Nr. 6, 1985, S. 483-502.

[22] Besold T.R., Plaza E., *Generalize and Blend: Concept Blending Based on Generalization, Analogy, and Amalgams*, in: ICCC, 2015, S. 150-157.

[23] Eisenreich C., Ott J., Süßdorf T., Willms C., Declerck T. , *From Tale to Speech: Ontology-based Emotion- and Dialogue Annotation of Fairy Tales with a TTS Output*, in: International Semantic Web Conference (Posters & Demos), 2014, S. 153-156.

[24] Fernau H., *Iterierte Funktionen, Sprachen und Fraktale*, BI-Wiss.-Verlag, 1994.

[25] Finlayson M.A., *ProppLearner: Deeply annotating a corpus of Russian folktales to enable the machine learning of a Russian formalist theory*, Digital Scholarship in the Humanities, 32. Jg., Nr. 2, 2015, S. 284-300.

[26] Frank A. F., Jaeger T. F., *Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production*; in: Proceedings of the Annual Meeting of the Cognitive Science Society, 2008.

[27] Gervás P., *From the Fleece of Fact to Narrative Yarns: a Computational Model of Composition*, in: Proceedings of the Third Workshop on Computational Models of Narrative, 2012, S. 123-131.

[28] Gervas P., *Computational Approach to Storytelling and creativity*, AI Magazine, 30. Jg., Nr. 3, 2009, S. 49.

[29] Grassberger P., *Randomness, information and complexity*, arXiv:1208.3459, August 2012.

[30] Großman A., Holldöbler S., Skvortsova O., *Symbolic Dynamic Programming within the Fluent Calculus*, in: Proceedings of the IASTED International conference on Artificial and Computational Intelligence. ACTA Press, 2002, S. 378-383.

[31] Hjelmslev L. T., *Prolegomena zu einer Sprachtheorie*, München, Hueber, 1974.

[32] Hofstadter D.R., *Gödel, Escher, Bach*, New York: Vintage Books, 1980.

[33] Itt L., Baldi P., *Bayesian Surprise Attracts Human Attention*, in: Advances in neural information processing systems, 2006, S. 547-554.

[34] Signorini G. F., *L´entropia é disordine?*, 30/10/2009, `http://www.chim.unifi.it/~signo/did/etc/entropia/entropia-split0/#foot16`.

[35] Köhler R., *Synergetic linguistics*; in: Contributions to quantitative linguistics. Springer, Dordrecht, 1993, S. 41-51.

[36] Lakoff G., *Structural Complexity in Fairy Tales*, 1972.

[37] Lévi-Strauss, C., *Structural Anthropology*, Basic Books, New York, 1963.

[38] Lévi-Strauss, C., *Die Struktur und die Form. Reflexionen über ein Werk von Vladimir Propp*, in: ders.: *Strukturale Anthropologie II*, Frankfurt, M.1975,S. 167

[39] Liu H., *Probability distribution of dependency distance*, Glottometrics, 15. Jg., 2007, S. 1-12.

[40] Mani I., *Computational Narratology.* Handbook of narratology, 2014, S. 84-92.

[41] Meletinskij E., *Zur strukturell-typologischen Erforschung des Volksmärchens.* tr. C. Wendt, 1969, S. 179-214.

[42] Meyer P., *Two semi-mathematical asides on Menzerath-Altmann's law*, in: Exact Methods in the Study of Language and Text, 2007, S. 449-460.

[43] Modi A., Anikina T., Ostermann S., Pinkal M., *InScript: Narrative texts annotated with script information*, arXiv preprint arXiv:1703.05260, 2017.

[44] Chomsky, N., *Aspects of the theory of syntax*, MIT Press, 1965.

[45] Qian B., Rasheed K., *Hurst Exponent and financial market predictability*, 2007.

h

[46] Reagan A.J., Mitchell L., Kiley D., Danforth C. M., Dodds P. S., *The emotional arcs of stories are dominated by six basic shapes*, EPJ Data Science, 5. Jg., Nr. 1, 2016, S. 31.

[47] Richman R.M., *Recursive Binary Sequences of Differences*, Complex Systems, 13. Jg., Nr. 4, 2001, S. 381-392.

[48] Todorov T., *Poetique*, Éditions du Seuil, 1973.

[49] Wanzare L.D., Zarcone A., Thater S., Pinkal M., *DeScript: A Crowdsourced Corpus for the Acquisition of High-Quality Script Knowledge*, in: The International Conference on Language Resources and Evaluation, 2016. $http://www.lrec-conf.org/proceedings/lrec2016/pdf/913_{P}aper.pdf$ (Zugriff am 14.02.2018).

Eidesstattliche Erklärung

Hiermit erkläre ich,

Palma Cosimo, geboren am 13.03.1993 in Battipaglia (SA, Italien),
Matrikelnummer: 1262270

an Eides statt, dass die von mir vorgelegte Wissenschaftliche Arbeit mit dem Titel

Selbstähnlichkeit in Texten: märchenhafte Beiträge zur Computererzähltheorie

selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Texten entnommen sind, wurden unter Angabe der Quellen (einschließlich des World Wide Web und anderer elektronischer Text- und Datensammlungen) und nach den üblichen Regeln des wissenschaftlichen Zitierens nachgewiesen. Dies gilt auch für Zeichnungen, bildliche Darstellungen, Skizzen, Tabellen und dergleichen. Mir ist bewusst, dass wahrheitswidrige Angaben als Täuschungsversuch behandelt werden und dass bei einem Täuschungsverdacht sämtliche Verfahren der Plagiatserkennung angewandt werden können.
Trier, 27.03.2018

Cosimo Palma