

Sprachliche Innovation – lokale Ursachen und globale Wirkungen

Das „Dynamische Sprachnetz“

Von der Carl von Ossietzky Universität Oldenburg – Fakultät III – Sprach- und
Kulturwissenschaften – zur Erlangung des Grades eines

Doktors der Philosophie (Dr. phil.)

genehmigte Dissertation

von Herrn Michael Schenke

geboren am 6.4.1957 in Itzehoe

Referent: Prof. Dr. Gerd Hentschel
Korreferent: Prof. Dr. Alexander Lavrov

Tag der Disputation: 23.5.2008

Inhaltsverzeichnis

1	EINLEITUNG.....	8
1.1	Das Ziel der Arbeit	8
1.2	Die Genese der Arbeit.....	9
1.3	Die Gliederung der Arbeit	11
1.4	Die Ergebnisse der Arbeit	12
1.5	Abschließende Bemerkungen.....	13
2	WISSENSCHAFTSTHEORETISCHE VORBEMERKUNGEN	14
2.1	Theorie, Experiment und Computer in der Wissenschaft.....	14
2.2	Computer und Exaktheit in der Wissenschaft	16
2.3	Nutzen dieser Arbeit für die Linguistik	17
2.4	Lokale und globale Betrachtungen.....	19
2.5	Zellulare Automaten als Modelle von Sprechergemeinschaften.....	21
3	ZELLULARE AUTOMATEN	24
3.1	Gittergeometrie	26
3.2	Nachbarschaft	28
3.3	Zustandsmenge	30
3.4	Zustandsänderung	30
3.5	Randbedingungen	32
3.6	Das Spiel des Lebens.....	32
3.7	Klassische zellulare Automaten	33
4	MODELLIERUNG VON SPRACHWANDEL DURCH ZELLULARE AUTOMATEN.....	35
4.1	Metaregeln.....	35
4.2	Monotonie.....	36
4.3	Angemessenheit.....	37
4.4	Symmetrie.....	38

4.5	Eindimensionale Geometrien.....	38
4.6	Mehrdimensionale Geometrien	40
5	DAS PIOTROVSKIJ-GESETZ UND VERWANDTE ANSÄTZE.....	45
6	SPRACHLICHE NEUERUNGEN IM MODELL DES ZELLULAREN GRAPHEN	48
6.1	Wie schnell kann sich eine sprachliche Änderung ausbreiten?	48
6.2	Elemente der Graphentheorie	49
6.3	Netze endlicher Automaten.....	52
6.4	Der Vollständige Binäre Baum als Gittergeometrie	54
6.5	Fraktale Gittergeometrien	56
7	KOMPLEXE NETZWERKE.....	60
7.1	Beispiele für komplexe Netzwerke	61
7.1.1	Soziale Netzwerke.....	61
7.1.2	Informationsnetzwerke.....	62
7.1.3	Technische Netzwerke	62
7.1.4	Biologische Netzwerke	62
7.2	Gradverteilungen.....	63
7.2.1	Mathematische Verteilungen.....	63
7.2.2	Beispiele für uns wichtiger Verteilungen.....	64
7.3	Die Verteilung des Zipfschen Gesetzes (Pareto-Verteilungen)	65
7.4	Andere statistische Kenngrößen für Netzwerke.....	69
7.4.1	Clustering	69
7.4.2	Hierarchie und Gradkorrelationen.....	73
7.4.3	Der „small-world“-Effekt.....	74
7.5	Die zentrale verbundene Komponente.....	75
7.6	Weitere statistische Kenngrößen	75
7.7	Charakteristika von Netzwerken.....	75
8	ZUFALLSNETZWERKE	77
8.1	Das configuration-Modell („Igel-Modell“)	78
8.2	Poisson-Verteilung	78
8.3	Exponential-Verteilung	80
8.4	Wer hat, dem wird gegeben (Pareto-Verteilungen)	80
8.5	Die Anwendung des Matthäus-Effekts.....	80

8.6	Noch einmal das Piotrovskij-Gesetz.....	83
8.7	Zusammenhang der Sprechergemeinschaft (Community structure).....	84
9	PROZESSE AUF NETZWERKEN	88
9.1	Aufrechterhaltung der Kommunikation.....	89
9.2	Sprachliche Neuerungen als epidemiologische Prozesse	90
9.3	Momente und erzeugende Funktion.....	92
9.4	Die Epidemische Schwelle oder: Wieso gewinnt nicht jede sprachliche Neuerung?	93
9.5	Analytische Ergebnisse.....	95
9.6	Änderungsregeln.....	96
9.6.1	Die Allgemeine einfache Mehrheitsregel.....	96
9.6.2	Die Regel des höchsten Lohnes.....	97
9.7	Auswirkung von Alterungsprozessen auf Veränderungen der Sprache.....	97
10	ANWENDUNGEN VON NETZWERKEN IN DER LINGUISTIK	99
10.1	Sprachliche Universalien.....	99
10.2	Netzwerke in einem existierenden Sprachsystem.....	101
10.3	Das word web	102
10.4	Semantische Netzwerke	103
10.5	Thesaurus-Netzwerke	104
10.6	Das WordNet.....	105
10.7	Wort-Assoziations-Netzwerke.....	105
10.8	Netzwerke syntaktischer Abhängigkeiten.....	106
10.9	Netze von Sprachnutzern	108
10.10	Das Zipfsche Gesetz	110
10.11	Sprache und Netzwerke: Weitere Aspekte	111
11	DAS „DYNAMISCHE SPRACHNETZ“	113
11.1	Designentscheidungen für das Dynamische Sprachnetz.....	115
11.2	Anforderungen an Statische Zustandsüberführungsregeln	118
11.3	Detaillierte Analyse der Zustandsüberführungsregeln.....	120
11.4	Die Umschaltfunktion.....	122
11.4.1	Konstante Umschaltfunktionen	125
11.4.2	Lineare Umschaltfunktionen	126

11.5	Spezielle Zustandsüberföhrungsfunktionen	126
11.6	Sonderregeln für kleine n	128
11.7	Dynamische Zustandsüberföhrungsregeln	131
11.8	Anfangsphase der Ausbreitung.....	131
11.9	Latenz.....	133
11.10	Formale Definition eines Dynamischen Sprachnetzes	134
12	DIE MONTE CARLO METHODE	136
12.1	Die Anwendung der Monte Carlo Methode.....	137
12.2	Dimensionsreduktion	137
13	DIE IMPLEMENTIERUNGEN DES MODELLS.....	140
13.1	Programme für Eindimensionale Geometrien	140
13.2	Grobdesign des Dynamischen Sprachnetzes.....	142
13.2.1	Objekte	142
13.2.2	Struktur des Hauptprogramms.....	143
13.3	Dynamik beim Aufbau des Netzes	144
13.4	Die Dynamik auf dem Netz	144
13.4.1	rein zufällige Werte.....	145
13.4.2	empirische Daten.....	146
13.5	Ausbreitung der sprachlichen Neuerung	146
13.6	Die Anfangsphase der Ausbreitung.....	147
13.7	Übertragungsregeln	147
13.8	Alterung im Dynamischen Sprachnetz.....	148
13.9	Sprachliche Erneuerung oder alter Bestand?	149
14	VERFEINERUNGEN DES MODELLS	150
14.1	Zerfall des Netzes	150
14.2	Umstrukturierung des Netzes	152
14.3	Änderung der Zielgröße: Verteilung auf Kategorien	154
14.4	Dimensionsreduktion	154
14.5	Beispiel eines Sprachnetzes	160
15	UNSICHERHEITEN DIE ÜBERNAHME VON LEHNWÖRTERN BETREFFEND	162

15.1	Der Begriff „Lehnwort“	162
15.2	Die benutzten Daten.....	165
15.3	Der deutsche Einfluß auf das Polnische	165
16	DAS MODELL SPIEGELT DIE WIRKLICHKEIT WIDER	167
16.1	Die Experimente.....	168
16.2	Die Ergebnisse	170
16.3	Der Kolmogoroff-Smirnoff-Test	171
16.4	Die Auswertung der Ergebnisse.....	173
16.5	Beigefügte Dateien und Programme.....	175
17	FAZIT UND AUSBLICK	179
17.1	Was wurde erreicht?.....	179
17.2	Was ist für die Zukunft zu erwarten?	180
17.2.1	Ausnutzung des Modells	180
17.2.2	Erweiterungen des Modells.....	182

Verzeichnisse:

Literaturverzeichnis	185
Abbildungsverzeichnis	196
Index	197

1 Einleitung

1.1 Das Ziel der Arbeit

In der vorliegenden Arbeit sollen sprachliche Innovationen, insbesondere allgemeine Muster ihrer Ausbreitung, sowie auch allgemeine Muster des Schwindens sprachlicher Phänomene, sofern ein solches vorliegt, untersucht werden. Gibt es allgemeine Gesetze oder Regeln, die diese Entwicklungen steuern? Wie es einer der üblichen Wege in der Naturwissenschaft ist, soll zunächst ein theoretisches Modell konstruiert und dieses an vorhandenen tatsächlichen Daten überprüft werden. Diesem Modell habe ich den Namen das „Dynamische Sprachnetz“ gegeben.

Als Beispiel, das jedoch vornehmlich nur eine Grundlage für weitere Abstraktionen darstellt, soll die Übernahme von Lehnwörtern aus einer Sprache in eine andere dienen. Die Daten, die in dieser Untersuchung die Rolle des experimentellen Befundes übernehmen, entstammen einer in der Slavistik der Universität Oldenburg existierenden Datenbank, [WDLP], die die deutschen Lehnwörter im Polnischen dokumentiert und als Grundlage für ein Wörterbuch dienen soll. In dieser Datenbank sind unter anderem für jedes Wort das Datum des Erstbeleges (also ungefähr der Zeitpunkt der Übernahme des Wortes) und gegebenenfalls das Datum des letzten Beleges (also ungefähr der Zeitpunkt, zu dem das Wort wieder aus dem Sprachbestand ausgeschieden ist) verzeichnet. Als experimenteller Befund soll die mathematische Verteilung dieser Zeitdauern dienen. Zur Überprüfung des Modells soll diese Verteilung mit der Verteilung der Zeitdauern verglichen werden, wie sie das Modell vorhersagt.

Bisher gibt es bereits einige Untersuchungen zu dem Thema der Übernahme von Lehnwörtern, etwa [OS65], [WLH68], [Bai73], [PP74] und die Arbeiten von Altmann [ABRS83], [Alt83], [Alt85]. Auf das für die Beschäftigung mit dieser Frage grundlegende *Piotrovskij-Gesetz*, einem Gesetz, das die Übernahme eines Lehnwortes in eine fremde Sprache quantitativ beschreibt, und seine Weiterungen werde ich noch ausführlich eingehen. Die hauptsächlichen Effekte, die dort untersucht werden, sind beispielsweise die globale Ausbreitungsgeschwindigkeit und die „Sättigung“, das heißt die Frage, wann es Grenzen für die weitere Ausbreitung eines Lehnwortes gibt. In jedem Falle beziehen sich aber alle Variablen im Modell von Piotrovskij auf globale Phänomene. Das bedeutet, daß nur Größen betrachtet werden, die sich auf die Menge *aller* Sprecher beziehen, etwa die Größe der Sprachgemeinschaft oder die Rate, mit der sich ein neues Wort ausbreitet.

Die ursprüngliche neue Idee meiner Untersuchung ist es, einen zumeist übersehenen Aspekt in den Mittelpunkt der Betrachtungen zu stellen, nämlich denjenigen sogenannter lokal meßbarer Größen. Dabei geht es besonders um Kommunikationsbeziehungen zwischen Individuen. Eine sprachliche Innovation kann sich nur ausbreiten, wenn eine Kommunikation stattfindet. Kommunikation ist aber im Gegensatz zu den bei Piotrovskij verwandten Größen nichts Globales und kann durch globale Variablen nur ungenügend repräsentiert werden, da ja gerade nicht alle Sprecher auf gleiche Art am sprachlichen Austausch beteiligt sind. Ein wesentlicher Bestandteil der vorliegenden Arbeit ist demzufolge die Existenz eines Netzwerkes von Kommunikationsbeziehungen. Dabei hat offensichtlich nicht jedes Individuum die gleiche Möglichkeit, eine Neuerung von anderen zu übernehmen. Die Anzahl sozialer Kontakte kann stark variieren, und die mathematischen Mittel, mit denen ihr Auftreten beschrieben wird, spielen eine bedeutende Rolle.

Die Untersuchungen mit Hilfe von Netzen, die im folgenden eine exakte mathematische Bedeutung erhalten werden, haben gegenüber den streng analytischen, aber nur global ausgerichteten und insgesamt nur recht groben Ansätzen, wie sie beispielsweise den Arbeiten zum Piotrovskij-Gesetz zugrunde liegen, einige Vorteile:

Es scheint so zu sein, daß die durch die Mathematik eingeforderten, in Kapitel 5 erläuterten, aber durch die Anwendung nicht zu erklärenden Konstanten im Piotrovskij-Gesetz durch die Verknüpfungsstruktur im Kommunikationsetzwerk wesentlich mitbestimmt sind. Welcher Art diese Zusammenhänge aber genau sind, das ist ein mathematisches Problem, das in absehbarer Zeit bestimmt nicht gelöst werden wird, auch nicht in der vorliegenden Arbeit. Dazu ist die Mathematik einfach noch nicht weit genug.

Im Verlauf der Forschungen zu Netzwerken in den letzten Jahren hat es sich gezeigt, daß die Struktur des Netzwerks eine erhebliche Rolle für die Resultate von Kommunikationen, die darauf vor sich gehen, spielt. Es gibt, und wir werden ein Reihe solcher Beispiele sehen, viele Netze, die sich in den am häufigsten betrachteten globalen Konstanten (also etwa in der Individuenzahl) nicht unterscheiden, bei denen sich aber die Dynamik von Kommunikationen erheblich unterscheidet. Dadurch können völlig andere Verhaltensmuster entstehen. Diese bei einer rein globalen Betrachtung unumgängliche Unschärfe der Ergebnisse läßt sich vermeiden, sofern man in der Lage ist, geeignete Werte für die Struktur des Kommunikationsnetzwerkes zu bestimmen.

1.2 Die Genese der Arbeit

Zur Untersuchung des geisteswissenschaftlichen Forschungsgegenstandes „Sprache“ sollen in dieser Arbeit auch naturwissenschaftliche Methoden herangezogen werden¹. Eine Grundannahme der Arbeit ist also, wie in der Wissenschaft üblich, die Existenz abstrakter allgemeiner Gesetze, die den Untersuchungsgegenstand, den Sprachwandel, im allgemeinen steuern, aber natürlich nicht ausschließlich sondern im Zusammenspiel mit kontingenten Geschehnissen. Der Ansatz schließt die Aufgabe der Modellbildung ein. Aber selbst zu einer rein naturwissenschaftlichen Untersuchung gehört nicht nur einfach die Konstruktion irgendeines Modells. Deshalb habe ich mich, gerade weil meine Untersuchungen tief in den Geisteswissenschaften wurzeln, und im Hinblick auf das Diktum von Albert Einstein

„Modelle sollen so einfach wie möglich sein, aber nicht einfacher.“

darum bemüht, auch alternative Wege zu prüfen, um einem möglichst verständlichen aber dennoch hinreichend ergiebigen Pfad zu folgen.

Die ursprüngliche Absicht war es daher, einen auch bei mathematischen Laien sehr populären und übersichtlichen Ansatz zu verfolgen, der nicht nur bis in die Unterhaltungsmathematik sondern auch in viele andere Wissenschaften, gerade auch Geistes- und Sozialwissenschaften vorgedrungen ist. Es handelt sich dabei um die *zellularen Automaten*, siehe Kapitel 3. Sie entsprechen einer sehr abstrahierten und daher leichter durch präzise mathematische Methoden zu untersuchenden Beschreibung zahlloser Phänomene.

¹ Ich halte die besonders auf Descartes zurückgehende Trennung zwischen Natur- und Geisteswissenschaften angesichts vieler neuerer Forschungsergebnisse ohnehin für nicht mehr so bedeutend, wie das lange Zeit gesehen worden ist.

Durch Komplexitätstheoretische Überlegungen konnte ich aber zeigen, daß alle auf regulären zellularen Automaten beruhenden Modelle im Widerspruch zum Piotrovskij-Gesetz stehen, siehe Abschnitt 6.1.

Daher bot sich der Übergang zu komplexeren Netzwerken an. Die Forschung an diesen hat seit etwa zehn Jahren einen gewaltigen Schub erhalten. Diese Untersuchungen sind seitdem ebenfalls in einer ganzen Reihe von Wissenschaften nutzbar gemacht worden: Mathematik, Informatik, Physik, Wirtschaftswissenschaften, theoretischer Biologie, Soziologie, Medizin und anderen, auch schon in der Linguistik (siehe Kapitel 10). In Fachkreisen sind solche Forschungen mittlerweile so beliebt, daß Netzwerke in den nächsten Jahren auch in weiten Kreisen außerhalb der Wissenschaft äußerst populär werden dürften, ähnlich wie zelluläre Automaten. Ich bin daher davon überzeugt, in meiner Arbeit einen vielversprechenden und zukunftsweisenden Ansatz gefunden zu haben, der gerade zur Zeit beginnt, seine Kapazität zu entfalten und für die Zukunft noch viel erwarten läßt.

In den letzten Jahren hat es Unmengen von Büchern, Artikeln und Konferenzen zu Netzwerkthemen gegeben. An Hintergründen dieser Arbeit noch tiefer interessierten Lesern kann ich aus einer gewaltigen Fülle von Werken, von denen mehrere in der Literaturliste zu finden sind, einige Bücher ans Herz legen. Insbesondere möchte ich drei Titel nennen²:

(i) *Albert-László Barabási*: *Linked. How everything is Connected to Everything else and what it means for Business, Science, and for Everyday Life.* [Bar03]

Dieses Buch ist von den mathematischen Ansprüchen her das bescheidenste der von mir hier vorgestellten. Es ist im Reportagestil geschrieben und spricht vor allem interessierte Laien an. Das Ziel des Buches ist die Popularisierung des Themas. Der Autor ist aber einer der bedeutendsten Forscher auf dem Gebiet, und es wird ein guter Überblick ohne jeglichen Formalismus geboten.

(ii) *Stephen Wolfram*: *A new kind of Science.* [Wol02]

Dieses Buch kann seine Leser erschlagen. Nach 1200 Seiten hört die Numerierung der Seiten (aber noch nicht das Buch) auf. Der Inhalt sind vor allem Wolframs eigene Forschungen zu zellularen Automaten. Es wird also nur ein Teilaspekt der für uns relevanten Überlegungen geschildert, der dafür aber mit Tiefe (vom Niveau zwischen den beiden übrigen hier von mir genannten Büchern). Viele Graphiken, Beispiele und Bilder wecken Leselust, und das Buch ist brilliant. Der Verfasser ist einer der bekanntesten Mathematiker der Gegenwart, der auch „Mathematica“³ erfunden hat. Mir erscheint manchmal ein gewisser Fanatismus und eine fast missionarische Haltung von Wolfram, mit denen er seine Arbeit als Lösung für fast alles anpreist, störend. Das Buch zeigt aber auch, daß der Formalismus tatsächlich zur Modellbildung in fast allen Wissenschaften geeignet ist.

² Der Markt ist in den Jahren, in denen diese Arbeit entstanden ist, geradezu mit Büchern zu Netzwerken gesättigt worden. Teils enthalten diese Bücher neue Ergebnisse, teils handelt es sich um erweiterte Auflagen veralteter Werke. Aus der Fülle seien beispielhaft [NBW06], [Wat04], [Buc03] genannt. Diese Bücher bieten aber keine für die vorliegende Arbeit relevanten Inhalte, die über die sonst zitierte Literatur hinausgehen.

³ Mathematica ist ein Softwaresystem zur Lösung von Problemstellungen, in denen Berechnungen und Simulationen aller Art notwendig sind. Es gibt eine nahezu unendliche Vielfalt von mathematischen Funktionen zur Lösung von Aufgaben aus allen denkbaren Bereichen vom Ingenieurwesen über Finanz- und Versicherungsmathematik bis hin zu allen Naturwissenschaften. Die herausragende Eigenschaft ist die Integration sowohl der mathematischen als auch der anwenderbezogenen Möglichkeiten des Systems.

(iii) *Sergej N. Dorogovtsev, José F. F. Mendes: Evolution of Networks. From Biological Nets to the Internet and WWW.* [DM03]

Für dieses Buch sind gründliche mathematische Kenntnisse nötig. Es stellt einen Teil des Standes der Wissenschaft im Jahre 2003 dar. Einige der Grunderkenntnisse der letzten Jahrzehnte mit Schwerpunkt auf dem neuen Jahrhundert werden einem zur Mitarbeit bereiten Leser in recht gedrängter Weise geboten. Angesichts der stürmischen Entwicklung des Forschungsgebietes kann das Buch eigentlich nur als Vorbereitung zur Lektüre von Originalartikeln verstanden werden. Die Autoren sind theoretische Physiker und haben selber bedeutende Beiträge zum Feld geliefert. (264 Seiten)

Auch in unserem Fall wird die Übertragung einiger Ergebnisse eine Rolle spielen, besonders aus der theoretischen Physik und der Epidemiologie. Es können viele Eigenschaften bestimmt werden, die Netzwerke haben müssen, damit sie als Modelle für für uns relevante Fragestellungen dienen können. Mit Hilfe eines erstmals 2002 bewiesenen Satzes konnte ich außerdem deutlich machen, daß Ausbreitungsphänomene in Netzwerken im allgemeinen keinen Widerspruch mehr zum Piotrovskij-Gesetz hervorbringen.

1.3 Die Gliederung der Arbeit

Diese Arbeit versteht sich als interdisziplinär, wobei die Interdisziplinarität darin besteht, daß das Ziel sprachwissenschaftlich ist, eine Erklärung, warum Sprachwandel so abläuft, wie er abläuft, daß zu diesem Zweck aber in hohem Maße Mittel aus der Mathematik und der Informatik herangezogen werden, mehr als es in rein philologischen Arbeiten sonst üblich ist. Eine Folge davon ist, daß mathematischen Betrachtungen viel Raum gegeben werden muß. Das spiegelt sich auch in der Gliederung wider.

Nach einigen wissenschaftstheoretischen Vorbemerkungen (Kapitel 2) wird der mathematische Formalismus eingeführt, mit dessen Hilfe ich zumindest im Ansatz den Sprachwandel modellieren werde, die *zellularen Automaten* (Kapitel 3). Danach (Kapitel 4) wird erörtert, wie das Problem der Modellierung für den speziellen Fall des Sprachwandels mit Hilfe des allgemeinen Werkzeuges „zellulärer Automat“ angegangen werden kann. Nach einem kurzen Rückblick auf die klassische einschlägige Methode der Modellierung in der Linguistik, das „Piotrovskij-Gesetz“ (Kapitel 5), wird durch einen Vergleich der Modellierungsmethoden mit zellularen Automaten und den Ergebnissen des Piotrovskij-Gesetzes bewiesen werden, daß Automaten noch eine zu eingeschränkte Struktur haben, um den Sprachwandel adäquat modellieren zu können (Kapitel 6). Wird eine bestimmte Einschränkung zellulärer Automaten weggelassen, die in Unterkapitel 3.1 eingeführte *Isotropie*, so entsteht der Formalismus der „Netzwerke“ oder „dynamischen Graphen“, der dann unter verschiedenen Gesichtspunkten beleuchtet wird: den grundlegenden Eigenschaften von Netzwerken (Kapitel 7), die Art, wie Netzwerke in Simulationen angewandt werden können (Kapitel 8), und der Frage, wie Prozesse auf Netzwerken ablaufen (Kapitel 9), insbesondere Sprachwandelprozesse. Diese beiden letzten Kapitel stellen dabei vor allem die für diese Arbeit benötigten mathematischen Ergebnisse anderer Autoren aus den letzten zehn Jahren vor. Danach werden Ansätze aus der Theorie der Netzwerke vorgestellt, die in der Linguistik schon früher angewandt worden sind (Kapitel 10). Mit Hilfe einer Kombination sprachwissenschaftlicher Überlegungen, mathematischer Ansätze und allgemein logischer Schlüsse wird in Kapitel 11 das Dynamische Sprachnetz entwickelt, das grundlegende Modell zur Analyse von Sprechwandel in dieser Arbeit. Dieses Modell wird in mehreren Schritten

verfeinert und auf dem Computer implementiert (Kapitel 13 und 14). Die danach folgenden Kapitel sind der Verifikation des Dynamischen Sprachnetzes gewidmet. Im Kapitel 15 wird die Frage aufgeworfen, inwieweit Unsicherheiten bei der Übernahme von Lehnwörtern in der realen Welt, in unserem Fall im Polnischen aus dem Deutschen seit dem 13. Jahrhundert, eine abstrakte auf die Entdeckung allgemeiner Gesetze zielende Untersuchung erschweren. Im darauf folgenden Kapitel werden dann die Ergebnisse des Vergleiches zwischen Theorie/Modell und den tatsächlichen Daten vorgestellt, ebenso die statistischen Methoden, mit deren Hilfe die Untersuchungen durchgeführt werden. Ein Fazit mit Ausblick schließt die Arbeit ab.

1.4 Die Ergebnisse der Arbeit

Das theoretische Hauptergebnis der Arbeit ist die Konstruktion des Dynamischen Sprachnetzes. In diesem Modell werden neueste Erkenntnisse, zu einem großen Teil erst aus diesem Jahrhundert stammend, aus anderen Wissenschaften nutzbar gemacht, so daß die Art und Weise von sprachlichen Änderungen mit Hilfe des Netzes erklärt werden können. Dabei wird ein weiter Weg gegangen: Die Grundidee hat zunächst wenig mit Sprache zu tun. Es handelt sich um einen generischen Formalismus, wie er immer wieder auch anderswo nutzbringend verwendet wird. Dieser ist mittlerweile in der Wissenschaft allgemein fest etabliert.

Natürlich mußte dann die konkrete Anwendung von generischen Netzen auf spezifisch sprachliche Fragestellungen ermöglicht werden. Hier war an manchen Stellen eine detaillierte Begründung notwendig. Was dann entstanden ist, ist eine Klasse von parametrisierten Modellen. Bei solchen Modellen ist es ein übliches Verfahren, „Monte Carlo Methoden“ (siehe Kapitel 12) zu verwenden, mit deren Hilfe der Einfluß der Parameter bestimmt werden kann. Diese Modelle habe ich mit Hilfe der Sprache C++ implementiert (für eine Beschreibung der Implementierung siehe Kapitel 13).

Die damit durchgeführten Simulationen brachten im statistisch abgesicherten Vergleich zu den Werten in der oben erwähnten Oldenburger Datenbank bei verschiedenen Fragestellungen (Experimenten) eine außerordentlich hohe Übereinstimmung, so daß die Angemessenheit des Modells gut untermauert werden konnte. Es sei schon an dieser Stelle auf Abbildung 40 und Abbildung 41 verwiesen. Diese Abbildungen werden im Verlauf der Arbeit genauer erläutert werden. Sie zeigen sehr anschaulich, wie gut, besser als ursprünglich erhofft, die Übereinstimmung zwischen modellgemäßen Vorhersagen und tatsächlichem Befund aus der Datenbank ist.

Insgesamt ist aus den Ergebnissen zu sehen, daß der Sprachwandel nicht gravierend anders verläuft als ein Zustandswechsel in den meisten vergleichbaren zustandsbasierten Systemen, die in anderen Wissenschaften benützt werden. Man mag es als aufregendes Ergebnis ansehen, daß die Sprache von den gleichen Gesetzen bestimmt wird wie viele scheinbar völlig anders geartete Systeme in völlig anderen Bereichen des Lebens. Ich jedenfalls empfinde diesen Befund als begeisternd. Insbesondere dürfte mit dem hier präsentierten Ansatz die Grundlage sowohl für eine Übertragung der vorliegenden Ideen auf andere sprachwissenschaftliche Phänomene als das der Ausbreitung von Lehnwörtern wie auch für eine Übernahme anderer Ergebnisse aus der Theorie der Netzwerke für sprachwissenschaftliche Zwecke gelegt worden sein.

1.5 Abschließende Bemerkungen

Es gibt heute in der wissenschaftlichen Literatur einige Konventionen, denen ich mit Vorbehalten gegenüberstehe. Deshalb sollen jetzt einige Anmerkungen zum Sprachgebrauch in der Arbeit folgen:

- Ich verwende in der Arbeit weder den pluralis majestatis noch den Plural der Bescheidenheit, sondern ich schreibe „ich“, wenn ich mich meine. Das gilt insbesondere dann, wenn ich über die Abfassung der Arbeit schreibe. Wenn ich hingegen über die Lektüre der Arbeit und das Verständnis für die Arbeit schreibe, ziehe ich vor, „wir“ zu schreiben, weil ich dort die Leser mit einbeziehe.

- Ich beflöße mich nicht der „politisch korrekten“ Verwendung von Doppelformen hinsichtlich des Genus (meist maskulin \emptyset vs. feminin *-in*) sondern halte fest an der Verwendung von natürlich entwickelten („natürlich“ hier im Sinne „natürlicher Sprachen“) unmarkierten Formen von Bezeichnungen bei Menschen und höheren Lebewesen. Dafür sprechen aus meiner Sicht Lesbarkeit und Ökonomie.

- In der deutschen Sprache des täglichen Lebens greifen vielerorts amerikanisch-englische Neologismen immer mehr um sich. Ich habe versucht, diese Praxis zu vermeiden. Da aber viele der benutzten Originalarbeiten auf Englisch erschienen sind und sich, weil es sich um sehr neue Artikel handelt, noch kein deutsches Wort etablieren konnte, habe ich eine Reihe von Fachwörtern in ihrer ursprünglichen englischen Form belassen.

- In den letzten Jahren hat es mehrfach Bemühungen zur Reformierung der deutschen Rechtschreibung gegeben. Die Texte dieser Arbeit sind während der Zeit entstanden, zu der die Reform wiederholt reformiert wurde. Um meine Texte nicht beständig durch neue Korrekturen der jeweils aktuellen Form anpassen zu müssen, benutze ich die alte deutsche Rechtschreibung.

In Übereinstimmung mit einer Tradition, der ich mich gerne anschließe, möchte ich einer Reihe von Leuten danken, die in der einen oder anderen Weise an verschiedenen Aspekten der Arbeit beteiligt waren, von der Diskussion mathematischer Fragen über Hilfsarbeiten an der Datenbank und das von mir angeleitete Erstellen von Abbildungen bis hin zum Korrekturlesen. In der alphabetischen Reihenfolge sind dies:

U.Borchert, C.Brandt, A.Kilian, F.Koch, E.Liebscher, S.Münch, J.Seeländer.

Ganz besonders möchte ich natürlich Herrn Professor G.Hentschel danken. Sein Beitrag zu der Arbeit geht über eine einfache Betreuung weit hinaus. Er hat mir die wissenschaftliche Beschäftigung mit der Sprachwissenschaft nahegebracht. Diese Kenntnis eines für mich neuen Gebietes der Wissenschaft hat mein ganzes Denken in einer Weise beeinflusst, die überhaupt nicht überschätzt werden kann.

2 Wissenschaftstheoretische Vorbemerkungen

2.1 *Theorie, Experiment und Computer in der Wissenschaft*

Das herausragende Kennzeichen der (natur)wissenschaftlichen Methode ist traditionell das Zusammenspiel zwischen Theorie und Experiment. Die Aufgabe der Theorie ist dabei zwiefältig:

- Zum einen soll sie die Ergebnisse von Experimenten durch Bezugnahme auf eine höhere Abstraktionsebene erklären.
- Zum anderen soll die Theorie die Ergebnisse neu zu entwickelnder Experimente vorhersagen.

Die dabei benutzte Methode sind traditionell Regeln auf der Grundlage mathematischer Gleichungen, die die Welt, oder besser einen kleinen Teil davon, beschreiben. Durch die Beschränkung auf einige wesentliche Gesichtspunkte lassen sich genaue Zusammenhänge zwischen verschiedenen Entitäten beschreiben, und zwar, soweit wie Mathematik involviert ist, mit absoluter Genauigkeit und Sicherheit.

Jedoch hat dieser Ansatz auch ein entscheidendes Problem - die Schwäche der Mathematik: Nur die einfachsten Systeme lassen sich auf irgendeine Weise durch Gleichungen beschreiben, die anschließend explizit mit mathematischen Mitteln gelöst werden können. Das können beispielsweise durch Differentialgleichungen beschriebene Systeme [Bos94] oder Systeme sein, die sich durch Differenzgleichungen darstellen lassen [KN99], oder allgemeine Modelle der diskreten Mathematik. Solche Abstraktionen, die mit absoluter Genauigkeit und Sicherheit Experimente beschreiben, gründen, soweit es in der Wissenschaft bekannt ist, immer in der Mathematik oder in verwandten Bereichen, etwa in der Logik.

Seit der Entwicklung des Computers stehen aber auch andere Methoden zur Verfügung. In den letzten Jahrzehnten haben numerische Methoden einen erheblichen Aufschwung genommen [Schw97]. Diese liefern im allgemeinen bekanntlich keine exakten Lösungen, aber die Fehler lassen sich zumindest oft exakt abschätzen. Erinnerung sei ferner an die weit verbreitete Nutzung von Simulationen. Viele Mathematiker halten heute ein beherrschbares Verfahren mit minimalen, abschätzbaren und daher für die Praxis irrelevanten Fehlern des Ergebnisses für genauso befriedigend wie die geschlossene Lösung einer Gleichung. Das gilt insbesondere dann, wenn das Verfahren auf einem Computer implementiert werden kann⁴. Durch die Nutzung des Computers können gewaltige Mengen von Details für komplexe Fragestellungen behandelt werden, wie sie kein Mensch, keine Gruppe von Menschen, ohne seine Hilfe bewältigen könnte. Und es können möglicherweise Konsequenzen aufgedeckt werden, die sich der Intuition des Betrachters entziehen, insbesondere wenn es sich um rückgekoppelte Zusammenhänge handelt. Somit entsteht neben, besser: zwischen, Theorie und Experiment eine dritte Ebene des Wissenschaftsbetriebs. Einerseits können

⁴ Es ist durchaus keine Selbstverständlichkeit, daß numerische Verfahren auf einem Rechner ablaufen können, selbst wenn sie aus mathematischer Sicht vernünftige Eigenschaften haben. Schuld daran kann beispielsweise ein hoher Bedarf an Zeit oder Speicher sein.

Computermodelle dazu dienen, Theorien (im Popperschen⁵ Sinne) zu falsifizieren oder wenigstens experimentell zu konkretisieren. Andererseits können sie als Wegweiser für experimentelle Fragestellungen an die reale Welt dienen, so wie es bisher nur die Theorie konnte. Musterbeispiele für dieses neuartige Paradigma bilden die moderne Astronomie und die moderne Kosmologie, wo eine experimentelle Bestätigung extrem schwierig ist, und komplexe(st) Computermodelle fast unentbehrlich werden. Natürlich befruchten sich Theorie und Computermodellierung gegenseitig, ebenso wie die Modellierung und die Fragen, die im Experiment an die reale Welt gestellt werden. An dieser Stelle soll dabei nicht darüber philosophiert werden, wie „real“ und tatsächlich erkennbar die „reale Welt“ wirklich ist. Insgesamt entsteht so das in Abbildung 1 gezeigte Bild der wissenschaftlichen Erkenntnis.

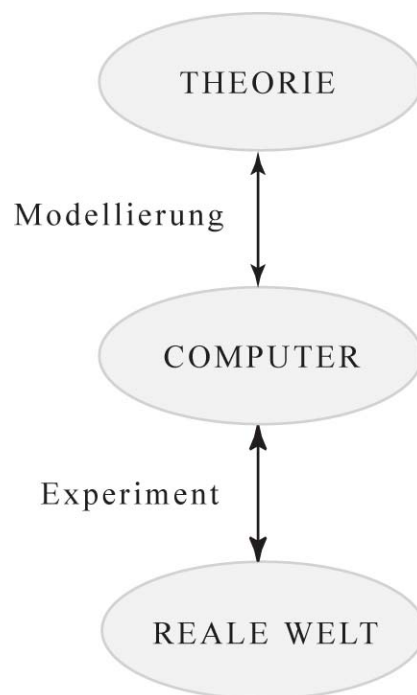


Abbildung 1: Drei Ebenen wissenschaftlicher Erkenntnis

In der Linguistik, wo Theorien vielfach bestenfalls als halbformale Gebilde konstruiert werden, so daß eine präzise empirische Überprüfung ohnehin schwierig ist, kann eine mathematisch fundierte Computermodellierung eine Alternative sein.

⁵ *Sir Karl Popper* (1902-1994), österr.-engl. Philosoph, begründete eine neue wissenschaftstheoretische Methodenlehre. Eine der Hauptideen davon ist das Prinzip der permanenten Fehlerkorrektur im Bereich der Theoriebildung. Eine Theorie kann nicht verifiziert werden. Es können nur fehlerhafte Aussagen der Theorie durch Experimente falsifiziert werden.

2.2 Computer und Exaktheit in der Wissenschaft

Durch die Benutzung des Computers gehen die absolute Genauigkeit und Sicherheit im Vergleich zur Theorie verloren, und zwar nicht nur dadurch, daß man auf geschlossene Lösungen verzichten muß, sondern auch durch die zusätzliche Fehlerquelle des Computers selbst. Soll der Computer die Rolle einer mittleren, historisch gesehen dritten Ebene zwischen Theorie und Experiment einnehmen, so muß man diese Fehlerquelle beherrschen können. Die Sicherheit vor Computerfehlern kann dabei auf mehrfache Weise erzielt werden. Das schwächste, aus wissenschaftlicher Sicht völlig unbefriedigende Mittel sind umfangreiche Tests mit zwar ausgeklügelten aber trotzdem keine Sicherheit bietenden Methoden der Informatik. Ein geflügeltes Wort sowohl bei professionellen Testern als auch bei Entwicklern abstrakter theoretischer Testmethodiken, das diesen Mangel von Tests unterstreicht, lautet

„Tests können immer nur die Anwesenheit von Fehlern zeigen, niemals deren Abwesenheit.“ (Edsger Dijkstra, ursprünglich in [Dij70]).

Aus wissenschaftstheoretischer Sicht befriedigender sind in den letzten drei Jahrzehnten entwickelte Methoden für mathematische Korrektheitsbeweise (siehe beispielsweise [AO94] und viele andere, darunter [Sche97]). Auf diese Art entsteht ein grobes Bild von drei Abstraktionsebenen:

- unten die Mathematik,
- gegründet auf diese dann formal abgesicherte Computerprogramme, die durchzuführende oder bereits durchgeführte Experimente erklären und auch sinnlich wahrnehmbar machen,
- und darüber das Experiment selbst oder anderweitig erhaltene Daten.

Ein damit verwandter Ansatz ist der, Computerprogramme nicht in kommerziell verfügbaren Sprachen zu implementieren, sondern sie direkt als einfache mathematische Objekte zu betrachten. Der bekannteste und am besten untersuchte derartige Ansatz ist die *Turing-Maschine*⁶. Die Überlegungen von Alan Turing zur maschinellen Berechenbarkeit gelten heute vielfach als der Ausgangspunkt der theoretischen Informatik insgesamt. Aufgrund der intensiven Untersuchungen von Turing-Maschinen und äquivalenter Konstrukte wird heute allgemein die *„Churchsche These“* akzeptiert, die natürlich ihrerseits nicht bewiesen werden kann und daher keine mathematische Aussage ist:

„Alles, was im intuitiven Sinne berechenbar ist, das läßt sich auch mit Hilfe von Turing-Maschinen berechnen.“ (Alonso Church)

Solche für Außenstehende eher esoterischen, für die Informatik aber grundlegenden Überlegungen haben der Akzeptanz des Computers *als Mittel der Erkenntnis* und nicht nur als allgemeines Hilfsmittel in allen Wissenschaften lange entgegen gestanden. Für die Mathematik wurden solche Fragen brandaktuell, als es 1978 Kenneth Appel und Wolfgang

⁶ Sowohl die Arbeiten zu Turing-Maschinen als auch die Überlegungen von Turings Doktorvater Church sind grundlegend für die theoretische Informatik. Sie sind jedem Informatiker bekannt und können in jedem Buch zum Thema nachgelesen werden. Ein klassisches Lehrbuch der theoretischen Informatik ist [HMU01].

Haken gelang, die seit über hundert Jahren bestehende Vierfarbenvermutung⁷ unter massivem Computereinsatz zu beweisen [AH78]. Eine hervorragende Übersicht über die philosophischen Folgen für die Mathematik bietet [DH86]. In vielen anderen Wissenschaften hat aber ebenfalls ein Umdenkprozeß, was den Computer als Mittel der Welterkenntnis angeht, eingesetzt, am meisten in den Naturwissenschaften, aber auch in den Gesellschaftswissenschaften. Ein Überblick etwa für die Soziologie kann [Klü95] entnommen werden.

Auch in der Sprachwissenschaft wird der Computer bisher leider eher als Hilfsmittel, beispielsweise der Datenhaltung und –verarbeitung oder bei statistischen Berechnungen wahrgenommen. Wegen der engen Verflechtung zwischen Linguistik und der theoretischen Informatik im Bereich der Sprachverarbeitung gibt es in Teilen der Linguistik sicher eine höhere Akzeptanz des Computers als Mittel der Erkenntnis. Jedoch dienen selbst die Anwendungen der Computerlinguistik normalerweise eher der praktischen Nutzbarmachung vorher ausgearbeiteter Theorien und weniger dem Erkenntnisgewinn für die Sprachwissenschaft. Wie weit andererseits aber doch schon Computernutzung und Linguistik miteinander verwoben sind, zeigen neuere Bücher zur Computerlinguistik, etwa [Bre01] und [Dal02].

2.3 Nutzen dieser Arbeit für die Linguistik

In der vorliegenden Arbeit sollen Gesichtspunkte und Schlußfolgerungen des eben geschilderten mathematisch-naturwissenschaftlichen Ansatzes konsequent für die Linguistik nutzbar gemacht werden. Mathematisch-naturwissenschaftlich fundierte Ansätze haben in der Sprachwissenschaft eine lange Tradition. Ich will gar nicht darauf eingehen, daß meines Erachtens mathematisch inspiriertes Denken für einen Großteil der Linguistik grundlegend ist, nicht nur dort, wo explizit mathematische Methoden benutzt werden. Aber auch durch die ausdrückliche Anwendung mathematischer Modellierung, naturwissenschaftlicher Verfahren ist in der Sprachwissenschaft eine reiche Ernte eingefahren worden. Eigentlich heißt es, Eulen nach Athen zu tragen, wenn ich nur versuchen wollte, einen planmäßigen Überblick über solche Ansätze zu geben. Stattdessen erinnere ich an dieser Stelle nur an einige Klassiker, die ich ohnehin im Verlauf der Arbeit zitieren werde.

Von der Thematik her steht meine Arbeit den Untersuchungen zum schon mehrfach erwähnten „Piotrovskij-Gesetz“ am nächsten. Die Arbeiten von Piotrovskij und anderen, beispielsweise [PP74], wurden von G.Altmann, etwa [Alt83], erweitert. Diese Ansätze werde ich in Kapitel 5 und in Unterkapitel 8.6. ausführlicher beschreiben. Auch die Arbeiten eines anderen Klassikers der quantitativen Linguistik haben Beziehungen zu meinen Forschungen: G.K.Zipf. Er ist, so weit ich weiß, der erste, der in der Linguistik Phänomene mit solchen mathematisch-statistischen Methoden beschrieben hat, wie sie bei mir intensiv benutzt werden. Damit meine ich statistische Untersuchungen über Wortverteilungen, besonders die „Pareto-Verteilung“, die in Unterkapitel 7.3 genauer beschrieben wird. In den Rahmen dieser Untersuchungen fällt insbesondere das berühmte (und umstrittene) „Zipfsche Gesetz“. Dazu ist ebenfalls mehr in den Unterkapiteln 7.3 und besonders 10.10 zu erfahren.

In den letzten Jahrzehnten hat sich die quantitative Linguistik zu einem kraftvollen Zweig am Baum der Linguistik entwickelt. Als sehr verdienstvoll bei der Verbreitung solcher

⁷ Die Aussage des Vierfarbensatzes ist es, daß auf einer Karte von zusammenhängenden Gebieten, beispielsweise Ländern, nur vier verschiedene Farben zur Markierung der Gebiete nötig sind, damit trotzdem keine Grenze existiert, an der zwei gleich gefärbte Gebiete zusammenstoßen.

Methoden hat sich dabei die Reihe „Quantitative Linguistik“ (im Studienverlag Brockmeyer, Bochum, und später im Wissenschaftlichen Verlag, Trier) erwiesen. Zu dieser Arbeit habe ich etwa ein Dutzend Bände dieser Reihe herangezogen. Es seien beispielhaft [Gui82] (zum Zipfschen Gesetz), [Emb86] (zur Statistik in der Linguistik), [PPL90] (zu den Arbeiten von Piotrovskij) und [Tul95] (zu neueren Methoden in der quantitativen Linguistik) genannt, die gewisse Hintergründe meiner Arbeit verdeutlichen. Einen tieferen Einblick in diesen Zweig der Sprachwissenschaft liefert das Buch „Quantitative Linguistik“ [Best03]. Eine Arbeit, bei der in tiefer Weise etwas andere mathematische Methoden, als ich sie benutzen werde, zur Modellierung lexikalischer Evolutionsprozesse herangezogen werden, ist [Leo98].

Die vorliegende Arbeit steht in der eben kurz geschilderten Tradition. Wenn man voraussetzt, daß die Nutzung quantitativer Methoden in der Linguistik sinnvoll ist, bleibt die Frage, was meine Arbeit an Neuem bietet. Um es kurz zu sagen: *Ich möchte mithelfen, eine, wie ich finde, faszinierende Theorie für die quantitative Linguistik nutzbar zu machen, die Theorie der dynamischen Netzwerke*. Ich muß gestehen, daß ich bei weitem nicht der Erste bin, der dies tut. Wegmarken für die Straße hin zu einer Nutzung dynamischer Netzwerke in der Sprachwissenschaft sind bereits in den Arbeiten von R.Ferrer y Cancho und R.V.Solé ([FS01], [FS03]) sowie in [DM01] aufgestellt worden. Diese Autoren untersuchen die „Interaktion“ zwischen Wörtern – ein, wie noch ausgeführt werden wird, nur schwer in exakter und gleichzeitig befriedigender Weise definierbares Konzept – und stellen diese Interaktionen mit Hilfe eines dynamischen Netzes dar, das sie „word web“ nennen. Auch in der Arbeit jener Autoren geht es um die Evolution von Sprache. Jedoch sind sowohl ihre Methoden als auch die Ziele ihrer Forschungen ganz anders als bei mir. Dort wird beispielsweise eine mathematisch exakt begründete Definition des Kernbestandes des Wortschatzes einer Sprache gegeben und untersucht, wie er sich im Laufe der Sprachevolution verhält.

In den letzten etwa sechs Jahren hat es einige weitere Publikationen gegeben, wenn auch noch nicht allzu viele, die auf Möglichkeiten hinweisen, wie die Theorie der Netzwerke für die Linguistik nutzbar gemacht werden kann. Viele der wesentlichsten sind in Kapitel 10 kurz skizziert.

Worum geht es nun bei der Theorie der dynamischen Netzwerke und wie gedenke ich sie im folgenden anzuwenden? Es wäre verfrüht, die Theorie schon hier detailliert darzulegen. Ich muß auf die entsprechenden Teile in meiner Arbeit, insbesondere auf die Konstruktion des „Dynamischen Sprachnetzes“ in Kapitel 11 verweisen. Aber, kurz gesagt, geht es darum, bisher fast ausschließlich in anderen Wissenschaften benutzte Methoden zur Beschreibung sich dynamisch verändernder Systeme auch auf Aspekte der Linguistik anzuwenden. Die hier eingesetzten Mittel entstammen besonders der Physik. Dort dienen sie zur Beschreibung von Vorgängen wie Magnetisierung oder Phasenübergängen, etwa der Vereisung. Mein Ansatz ist der folgende: So wie sich eine Änderung von Zuständen in Weißschen Bereichen oder Wassermolekülen ausbreitet, so breiten sich auch sprachliche Neuerungen in den Köpfen von Sprechern aus. Dabei ist die vorliegende Arbeit in vielem nur exemplarisch. Entsprechend habe ich an mancher Stelle auf Fragen hingewiesen, die auf der Hand liegen und weitergehende Forschungen rechtfertigen. Ich bin überzeugt, daß viele der Methoden der bisher anderswo durchgeführten Forschungen zu dynamischen Netzwerken sich mit Gewinn auf sprachwissenschaftliche Fragen ausweiten lassen. In diesem Sinne habe ich erst einen ersten tastenden Schritt getan in einen hoffentlich blühenden Garten, der viele Früchte bringen kann. Ich habe mich hier sowohl in den Methoden als auch in den Zielen beschränkt: So bilden die Änderungen des Wortschatzes nur einen Aspekt unter vielen möglichen sprachlichen Änderungen. Ohne bisher tiefer in die Materie eingestiegen zu sein, vermute ich

stark, daß die neueren Ergebnisse der Theorie der dynamischen Netze auch weitere interessante Konsequenzen in der Linguistik zeigen dürften, beispielsweise die Resultate zum Zerfall des Netzes oder zur epidemischen Schwelle. Das kann ich in dieser Arbeit nur andeuten, etwa in den Unterkapiteln 9.1 und 9.4, sowie im Schlußkapitel.

Was in dieser Arbeit lediglich geleistet worden ist, das ist die Konstruktion eines sehr eleganten theoretischen Modells mit einer überschaulichen und gut verständlichen Implementierung auf dem Computer. Durch die Einbeziehung lokaler Interaktionen von Sprechern, ein in Unterkapitel 2.4 genauer erklärtes Problem, werden jetzt genauere Einsichten in Mechanismen der Entwicklung von Neuerungen gewonnen. Diese lokalen Einflüsse sind so gravierend, daß ein theoretisches Modell, welches sie negiert, notwendigerweise irgendwann in Erklärungsnot kommt.

Mit dieser Arbeit wird, soweit ich weiß, erstmals ein Modell vorgelegt, das den Sprachwandel nicht nur in der Anfangsphase sondern auch ein mögliches erneutes Ausscheiden der Änderung und auch einen unvollständigen Sprachwandel *aus sprachlichen Motivationen*, insbesondere aus den lokalen Einflüssen, heraus beschreibt. In Kapitel 5 werde ich kurz einige Erweiterungen des Piotrovskij-Gesetzes vorstellen, die reversiblen oder unvollständigen Sprachwandel dadurch ermöglichen, daß *mathematisch notwendige* Konstanten willkürlich, also nur durch das Modell motiviert, herangezogen werden. Auch das Dynamische Sprachnetz kommt nicht ohne Konstanten aus. Da mit den meisten Konstanten hier aber eine sprachliche Bedeutung verbunden ist, sollte zumindest deren Bedeutung leichter erklärt werden können als bei den Konstanten in den klassischen Modellen.

Wesentlich in dieser Arbeit ist ferner, daß es sich bei der eben skizzierten Theorie nicht um eine reine Kopfgeburt handelt, sondern daß die Oldenburger Datenbank umfangreiches Material für eine erfolgreiche experimentelle Überprüfung liefert.

2.4 Lokale und globale Betrachtungen

Die klassische Modellierung (und damit die Gewinnung von theoretischen Erkenntnissen) ging meistens so vor sich, daß einzelne Teilbereiche des zu untersuchenden Systems identifiziert und durch Gleichungen, zumeist Differentialgleichungen, verbunden wurden. Ein solcher Ansatz für das uns hier besonders interessierende Gebiet der lexikalischen Evolution, in dem detaillierte wahrscheinlichkeitstheoretische Untersuchungen die Hauptrolle spielen, wird zum Beispiel in [Leo98] entwickelt. Sollte im Rahmen solcher mit klassischer Modellierung beschäftigter Forschungen ein Computer herangezogen werden, so geschieht dies heutzutage im allgemeinen zur Suche nach numerischen Lösungen solcher globaler Differentialgleichungen, quasi als Ersatz für die meistens nicht erreichbare exakte mathematische Lösung, oder als direkte Simulation der globalen Parameter.

In den vergangenen Jahrzehnten hat sich jedoch auch eine Fülle zusätzlicher Ansätze herausgebildet. Vielen dieser Ansätze ist eines gemein: *Es wird lokales Verhalten (Mikroverhalten) modelliert, daraus werden Schlüsse für das globale Systemverhalten (Makroverhalten) gezogen*. Schon unsere Intuition sagt uns, daß komplexen Systemen Organisationsprinzipien zu Grunde liegen müssen, die in irgendeiner Form durch die Topologie der Systeme bedingt sind.

Ist der Computer als erkenntnisbildendes Hilfsmittel erst einmal akzeptiert, so können mit seiner Hilfe außerordentlich komplexe Systeme auch auf eine völlig andere und neue Art untersucht werden. Diese Bemerkung ist natürlich nicht als uneingeschränkte Wahrheit sondern nur als Tendenz zu verstehen:

Auch die computergestützte Modellierung von Systemen muß oft globale Parameter einbeziehen.

Ferner muß angemerkt werden, daß die Dichotomie von Mikro- und Makroverhalten nicht erst durch den Computer heraufbeschworen wurde. Sie spielt seit vielen Jahrzehnten in der Physik eine große Rolle und konnte durch systemtheoretische Überlegungen auf andere Wissenschaften ausgedehnt werden. Man vergleiche [Hak81] für soziale Prozesse oder [PS81] für die Chemie.

Mit Hilfe des Computers kann jedoch die oben schon angesprochene Schwäche der Mathematik überwunden und der systemtheoretische Ansatz erst fruchtbar gemacht werden. Dabei sollte beachtet werden, daß auch die Aktionen innerhalb des Mikroverhaltens so komplex sein können, daß es unmöglich ist, sie mit mathematischen Methoden exakt zu analysieren oder auch sie nur mit dem Computer exakt nachzubilden. In diesem Fall ist es angebracht, die altbekannte Methode der Simulation zu verwenden, indem das Verhalten von Individuen mit statistischen Methoden nachgebildet wird. Dies ist oft möglich, da es sich ja gerade um ein Massenphänomen handelt. Was wird also für einen solchen systemtheoretischen Ansatz nötig sein? In [Klü95] und ähnlich in [GS95] werden genannt:

- die Einheiten, aus denen das System besteht,
- die Beziehungen zwischen den Einheiten (daraus folgend die Systemstruktur),
- Randbedingungen.

Darauf wird später noch genauer eingegangen werden. Der amerikanische Systemtheoretiker C.Langton kennzeichnet in [Lan88] den „*Micro-Macro-Link*“ für sein Forschungsgebiet „*Künstliches Leben*“ mit den folgenden Worten, die sich auch auf seine Arbeit mit zellularen Automaten beziehen und die auch für unsere eigenen Überlegungen hätten richtungweisend sein können:

„Artificial life involves the realization of lifelike behavior on the part of man-made systems consisting of populations of semi-autonomous entities whose local interactions with one another are governed by a set of simple rules. Such systems contain no rules for the behavior of the population at the global level, and the often complex high-level dynamics and structures observed are emergent properties which develop over time from out of all the local interactions among low-level primitives.“

Um noch einen Unterschied zwischen dem von mir anvisierte Ansatz mit Hilfe eines Micro-Macro-Links und der klassischen Modellierung gerade auch in der Linguistik deutlich zu machen, möchte ich noch einmal auf das Beispiel von [Leo98] zurückkommen: Dort ergibt sich ein rein globales Bild des Systems, das allerdings in seiner Art detailliertere Phänomene erfassen kann als der in dieser Untersuchung vorliegende Ansatz. Zur Zeit! Auf Netzen (wie dem Dynamischen Sprachnetz) beruhende Arbeiten sind im Vergleich zu klassischen Modellierungstechniken noch sehr jung, so daß die Methoden noch nicht so spezialisiert sind und daher auch nicht so detaillierte Erkenntnisse bringen können. Dafür gibt es noch eine Reihe unerforschter potentieller Anwendungsmöglichkeiten in verschiedenen Gebieten, die den Rahmen dieser Arbeit sprengen und auf die ich immer wieder hinweisen werde. Mein Hauptziel hier ist es eigentlich nur, einen faszinierenden neuen Ansatz für die Linguistik zu

nutzen und durch einige als exemplarisch zu bezeichnende Ergebnisse seine Anwendbarkeit dort zu demonstrieren.

2.5 Zellulare Automaten als Modelle von Sprechergemeinschaften

Welches Modell könnte nun geeignet sein, den Micro-Macro-Link, also in unserem Falle das Zusammenspiel zwischen einem individuellen Sprecherverhalten und der globalen Sprachentwicklung, abzubilden?

Der Nachteil der Turing-Maschine, des üblichsten mathematischen Modells, bei dem der Computer beteiligt ist, ist der, daß schon die einfachsten Probleme, die auf ihrer Grundlage berechnet werden, eine schier unübersichtliche Komplexität der zugehörigen Turing-Maschinen-Programme erfordern. Damit geht aber gerade der Vorteil der Mathematik im naturwissenschaftlichen Paradigma verloren, die Rückführbarkeit „der Welt“ auf einige einfache, intuitiv einsichtige, leicht behandelbare Prinzipien. Soll die Erklärung von Phänomenen der alltäglichen Welt auf der Grundlage von Mathematik aber mit Hilfe einer Abstraktionsschicht, die formale Computermodelle benutzt, durchgeführt werden, so ist es daher ratsam, mit einem Mechanismus zu arbeiten, der äquivalent zu Turing-Maschinen ist. Solche Mechanismen gibt es wie Sand am Meer, was ja gerade die Rechtfertigung der Churchschen These ist. Zusätzlich sollte dieser Mechanismus noch zusätzlich die Eigenschaft haben, die oben im Zusammenhang mit Simulationen genannt wurde: die *sinnliche Erfahrbarkeit*. Der vielleicht vielversprechendste Vorschlag in dieser Richtung, auf alle Fälle leicht auf dem Rechner implementierbar, wurde vom ungarischen Mathematiker John von Neumann⁸ und von Stanley Ulam⁹ gemacht: die zellularen Automaten. Diese haben sich als sehr fruchtbringend erwiesen. Auf der Grundlage umfangreicher vorhergehender Überlegungen vieler Mathematiker und Computerwissenschaftler hat der Amerikaner Stephen Wolfram das oben schon erwähnte Grundwerk [Wol02] mit fast allen Inhalten zum Thema geschrieben. Soweit zellulare Automaten betroffen sind, werde ich mich auf dieses Buch beziehen.

Bevor näher auf mathematische Grundlagen der zellularen Automaten eingegangen wird, sollen an dieser Stelle einige Bemerkungen gemacht werden zu der Frage, was ein Modell zu einem guten wissenschaftlichen Modell macht.

Es ist eine Binsenweisheit, daß ein Modell nicht alle Details des zu untersuchenden Systems widerspiegeln soll. Aber es gilt sogar mehr: Das Modell soll nur eine ganz abstrakte

⁸ John (ursprünglich János) von Neumann (1903 in Budapest - 1957 in Washington, DC) war einer der genialsten und vielseitigsten Mathematiker des 20. Jahrhunderts. Anfangs arbeitete er auf dem Gebiet von axiomatischer Mengenlehre und mathematischer Logik. Spätere Ideen führten zu einem seiner originellsten Entwürfe, der Spieltheorie. So bewies von Neumann das Minimax-Theorem für die Existenz einer optimalen Strategie in "Nullsummenspielen". Nach seiner Übersiedlung in die USA wurde er zum Freund und Kollegen Albert Einsteins in Princeton. Von Neumann arbeitete ab 1943 am Manhattan-Projekt in Los Alamos. Er war ebenfalls Verfasser des ersten mathematisch durchdachten Buches zur Quantenmechanik. Von Neumann gilt als einer der Väter der Informatik. Nach ihm wurde die so genannte Von-Neumann-Architektur benannt, ein Computer, in dem unter anderem Daten und Programm binär codiert im selben Speicher liegen. So gut wie alle modernen Rechner beruhen auf von Neumanns Ideen. 1953 entwickelte er auch die Theorie selbstreproduzierender Automaten, für die er ein kompliziertes Beispiel angab. Konsequenzen daraus sind in unserem Zusammenhang bedeutend, siehe Kapitel 3.

⁹ Stanley (ursprünglich Stanislaw) Ulam ist besonders bekannt durch seine Arbeit an der Wasserstoffbombe. Im Gegensatz zu dem bekannteren, eher praktisch ausgerichteten Edward Teller war er der theoretische Kopf und gilt als „mathematischer Vater der Wasserstoffbombe“.

Darstellung des Systems liefern. In der Regel gibt es keine direkte Verbindung des Modells mit dem Systemaufbau und dem Systemverhalten selbst. Nehmen wir als Beispiel die Beschreibung des Planetensystems durch Differentialgleichungen. Es soll hier nicht erörtert werden, welcher ontologische Status diesen Gleichungen zukommt, seien sie menschengemacht, gottgegeben oder „in der Natur gelegen“. Wichtig ist mir nur ein Punkt: Sie sind im System selber nicht unmittelbar angelegt. Nirgends im Planetensystem werden Differentialgleichungen gelöst, damit das System durch sie gesteuert werde. Was solche Modelle zu einem guten wissenschaftlichen Modell macht, ist nicht, daß sie Handlungen einzelner Elemente genau beschreiben, sondern das ist, daß sie eine möglichst genaue Korrespondenz zwischen den wichtigen Effekten des Modells und der Natur zeigen. Also muß hier unterschieden werden, welche Effekte als wichtig angesehen werden und welche nicht.

Neben dieser geforderten Übereinstimmung im prinzipiellen Verhalten, nicht aber in den Einzelheiten, ist die Einfachheit des Modells ein Desideratum. In einem Modell, das ähnlich komplex ist wie die Erscheinung, die es eigentlich nachbilden soll, ist es sehr schwer herauszuarbeiten, woher eine bestimmte Eigenschaft eigentlich kommt. In der Geschichte der Naturwissenschaft haben mathematische Modelle zumeist die Form von Gleichungssystemen angenommen, die aber durch ständiges Hinzufügen einzelner Eigenschaften immer komplizierter wurden.

Was macht es nun zu einer guten Idee, Sprecherverhalten zunächst durch zellulare Automaten (und später durch zellulare Graphen) zu modellieren?

Da ist zum einen die leichte Erweiterbarkeit. Mit Blick auf die bisher genannten Punkte scheint es mir vernünftig, in einem ersten Ansatz Sprachverhalten durch zellulare Automaten zu modellieren, und später dann zu den komplexeren Graphen überzugehen. Dabei gilt diese zusätzliche Komplexität nur für eine strikte mathematische Analyse aber nicht für das intuitive Verständnis des Modells selber und nicht für die Implementierbarkeit.

Ein weiterer Punkt ist der der sinnlichen Erfahrbarkeit. Man stelle sich ein Netz vor, dessen Knoten Sprecher symbolisieren und dessen Kanten Kommunikationsbeziehungen darstellen! Wenn man sich vorstellt, eine Neuerung werde von einem Sprecher zum nächsten weitergegeben, bekommt man einen Eindruck, wie die Neuerung sich in der ganzen Gemeinschaft ausbreitet und sich dann vielleicht überall festsetzt, steckenbleibt oder wieder verschwindet¹⁰. Das ist sicher dichter an der Vorstellung normaler Menschen als die sonst üblichen Differentialgleichungen.

Es seien noch einmal die wesentlichen Punkte zur Begründung des künftig untersuchten Ansatzes aufgeführt:

- die Einfachheit der Modelle (im eben geschilderten Sinne),
- Abstraktion von Details des individuellen Verhaltens,
- Paradigma des Computers als neue Zwischenschicht zwischen Theorie und beobachteter Wirklichkeit,

¹⁰ Diese Dynamik muß nicht unbedingt nur vor dem geistigen Auge ablaufen. Es gibt verschiedene Computerimplementierungen, mit deren Hilfe Systemabläufe in zellularen Automaten graphisch sichtbar gemacht werden können. Da mich vor allem das eigentliche Modell interessiert, habe ich allerdings auf solche zusätzlichen schwer implementierbaren Spielereien verzichtet.

- leichte Implementierbarkeit,
- leichtere Analysierbarkeit als sie bei rein mathematischen Modellen gegeben ist. Hier muß allerdings auch bekannt werden, daß die Analyse eines rein mathematischen Modells meistens tiefergehende Ergebnisse liefert.

3 Zellulare Automaten

Ausgangspunkt für die Arbeit mit zellularen Automaten waren, wie schon erwähnt, vermutlich Überlegungen des ungarischen Mathematikers John von Neumann zu einem Konzept einer sich selbst vervielfältigenden Maschine. Zusammen mit Stanislaw Ulam entwickelte er ein mathematisches Modell einer solchen Maschine, deren einzelne Komponenten gewisse, jedoch nur endlich viele Zustände annehmen können. Diese Komponenten erlauben nur Wechselwirkungen mit Komponenten in ihrer Nachbarschaft, wobei diese wechselseitigen Beeinflussungen getaktet sind¹¹. Das heißt, eine Entwicklung des Gesamtsystems kann nur zu bestimmten Zeitpunkten, dann aber parallel für alle Komponenten, stattfinden. Die Vorstellung, von der von Neumann und Ulam ausgingen, war, daß das Gesamtsystem aus einzelnen Komponenten, im weiteren Verlauf von ihnen „Zellen“ genannt, in Form eines Gitters angeordnet waren. Das Grundprinzip eines zellularen Automaten ist, abstrakt gesehen, die Übersetzung logischer oder durch mögliche Anwendungen motivierter Regeln in geometrische.

Dieses einfache mathematische Modell erwies sich im Verlauf der folgenden Jahrzehnte als fruchtbringend in vielen Feldern der Wissenschaft. Zunächst standen Fragen nach der Selbstreproduktion von theoretischen Maschinen im Mittelpunkt – der Zusammenhang mit der theoretischen Biologie ist evident – sowie theoretische Fragen der allgemeinen Berechenbarkeit. Obwohl hier schon erste Überlegungen zu dem heute wichtigen Gebiet der Parallelrechner durchgeführt wurden, führten die Automaten lange ein Dasein als Mauerblümchen, wurden dann aber Anfang der siebziger Jahre durch einen Geniestreich nicht nur im Bereich der Wissenschaft bekannt sondern in weiten Kreisen der Bevölkerung geradezu populär. Dieser Geniestreich ist das „Spiel des Lebens“, das weiter unten (Unterkapitel 3.6) wegen seiner Einfachheit und Wichtigkeit in einem kurzen Abschnitt näher erläutert werden soll. Schlagartig wurden zellulare Automaten zunächst selber ein beliebter Gegenstand mathematischer Forschungen. Wenige Jahre später bemächtigten sich ihrer auch andere Wissenschaften, zumeist als inzwischen wohletablierte mathematische Formalismen, die in dem oben geschilderten Sinne komplexes Verhalten durch einfache Regeln erzeugen. Beispielhaft seien hier die im Rahmen der „künstlichen Intelligenz“, einem Teilgebiet der Informatik, in den achtziger Jahre sich in Windeseile ausbreitenden neuronalen Netze, siehe [Roj93], genannt. Als ein Beispiel für die Breite der Anwendungen von zellularen Automaten in vielen Bereichen der Wissenschaft sei hier ferner ein extremer Untersuchungsgegenstand, die Formalisierung der Theorien des Marxismus durch Klüver et al., vgl. [Klü95] durch zellulare Automaten, erwähnt.

Da das Modell des zellularen Automaten, genauer gesagt: Weiterungen davon, grundlegend für die folgende Arbeit ist, seien hier zunächst seine zentralen Bestandteile informell aufgeführt:

- eine Menge von Zellen, die nur endlich viele Zustände annehmen können,

¹¹ Wir laufen hier in ein grundsätzliches Problem der theoretischen Informatik, ja sogar der Philosophie insgesamt, die Frage nach der Natur der Zeit. Ist sie kontinuierlich oder diskret? Die von mir im Dynamischen Sprachnetz benutzten Modelle haben genau wie Turing-Maschinen und zellulare Automaten eine schrittweise Entwicklung also eine diskrete Zeit. Auf die theoretischen Fragen wie auch auf die praktischen Probleme, die in der Informatik durch diese Dualität der Zeit entstehen, bin ich an anderer Stelle, in der Einleitung zu meiner informatischen Habilitationsschrift [Sche97], eingegangen.

- eine Gittergeometrie, die die Lage der Zellen zueinander bestimmt,
- alle Zellen sind identisch,
- Zustandsänderungen erfolgen zu diskreten Zeitpunkten,
- es gibt eine feste, für alle Zellen gültige Form der (lokalen) Nachbarschaft,
- der Folgezustand hängt nur von den Zuständen der Zelle selbst und der Zellen in der so definierten unmittelbaren Nachbarschaft ab.

Zellulare Automaten gehen also, wie wir oben gesehen haben, davon aus, daß die Welt diskret in Raum und Zeit ist. Auch heute noch stehen manche Mathematiker solchen Modellen reserviert gegenüber. Deshalb sollen einige kurze Begründungen geliefert werden:

- Als naturwissenschaftliche Rechtfertigung für diesen Ansatz wird vielfach die Quantentheorie herangezogen, die, abhängig von der mathematischen Behandlung und der Interpretation, ja gerade fordert, daß Zeit, Raum, Impuls, Energie usw. gequantelt, also eben diskret, auftreten. Diese Rechtfertigung ist zwar passend aber in Wirklichkeit nachgeschoben. Das ergibt sich schon aus den mannigfaltigen Anwendungen zellulärer Automaten, die sich viel höhere Abstraktionsebenen zu eigen machen und überhaupt nicht auf quantentheoretische Überlegungen rekurren. Es sei allerdings darauf hingewiesen, daß Wolfram zeigt, wie auch quantentheoretische Vorgänge, ja sogar die Eigenschaften von Zeit und Raum selber, durch zellulare Automaten nachgebildet werden können [Wol02].
- Eine bessere Rechtfertigung zellulärer Automaten liefert meines Erachtens die Geschichte von Mathematik und Physik. Mathematik war eben bis hin zur Entwicklung der Differentialrechnung durch Leibniz und Newton diskrete Mathematik. Die Entwicklung der nicht-diskreten Mathematik durch diese beiden Wissenschaftler entsprach gerade dem Bedürfnis, die Welt mit scheinbar angemessenen neuen Methoden zu erklären. Als reine Denkmodelle stehen zellulare Automaten also in einer uralten Tradition.
- Außerdem gibt es heutzutage keine so strenge Trennung zwischen diskreter und nicht-diskreter Mathematik mehr in den Anwendungen, wie es sie in den Köpfen einiger Fanatiker früher gegeben hat. Als Beispiel seien nur die vielfältigen Anwendungen der Theorie endlicher Gruppen bei der Analyse von Symmetrien genannt.

Die Bestandteile eines zellulären Automaten entsprechen mit gewissen Erweiterungen den in Unterkapitel 2.3 genannten Bedingungen für den Micro-Macro-Link. Sie sollen nun zusätzlich zu einigen noch nicht genannten Bedingungen aus der Praxis anhand von Beispielen illustriert werden.

3.1 Gittergeometrie

Die erste Entscheidung ist die, ob der Raum *isotrop* sein soll. Das heißt, daß der Raum von jeder Zelle aus gesehen die gleiche Form haben soll, daß das Netz also, anschaulich gesagt, „regelmäßig“ sein soll. Diese Gittergeometrien werden auch „regulär“ genannt. Um dies in mathematischen Worten zu präzisieren, muß ich zunächst einen weiteren Begriff einführen:

Definition: Sei K eine Gittergeometrie. Zu einer Zelle $z \in K$ sei $N(z)$ die Menge der Nachbarn von z .

- Sei L eine weitere Gittergeometrie. Eine Abbildung $\rho: K \rightarrow L$ heißt *Isomorphismus* (von Gittergeometrien), wenn gilt:
 1. Die Abbildung ρ ist bijektiv¹².
 2. Für alle $z \in K$ ist $\rho(N(z)) = N(\rho(z))$ ¹³.
- Die Gittergeometrie K heißt *regulär*, wenn es zu je zwei Zellen z_1 und z_2 von K einen Isomorphismus $\rho: K \rightarrow K$ gibt, also einen Isomorphismus mit identischer Urbild- und Bildmenge, mit

$$\rho(z_1) = z_2$$

Das heißt dann in der Tat, daß die Welt von z_1 und von z_2 aus gleich aussieht. Einige Beispiele mögen diese Begriffe verdeutlichen:

Eine *Translation* (*Verschiebung*) ist eine Abbildung, bei der jede Zelle um den jeweils gleichen Wert in x - und in y -Richtung verschoben wird. In Abbildung 2 ist eine Verschiebung um 1 in x - und um 2 in y -Richtung gezeigt. Translationen auf dem unbegrenzten zweidimensionalen Gitter sind Isomorphismen.

¹² Eine Abbildung $\rho: K \rightarrow L$ heißt *bijektiv*, wenn jedes Element von L Bild genau eines Elementes aus K ist. Damit wird durch ρ also jedem Element von K genau ein Element von L zugeordnet (und umgekehrt).

¹³ Es soll also für jeden Punkt das Bild seiner Nachbarschaft gleich der Nachbarschaft seines Bildes sein.

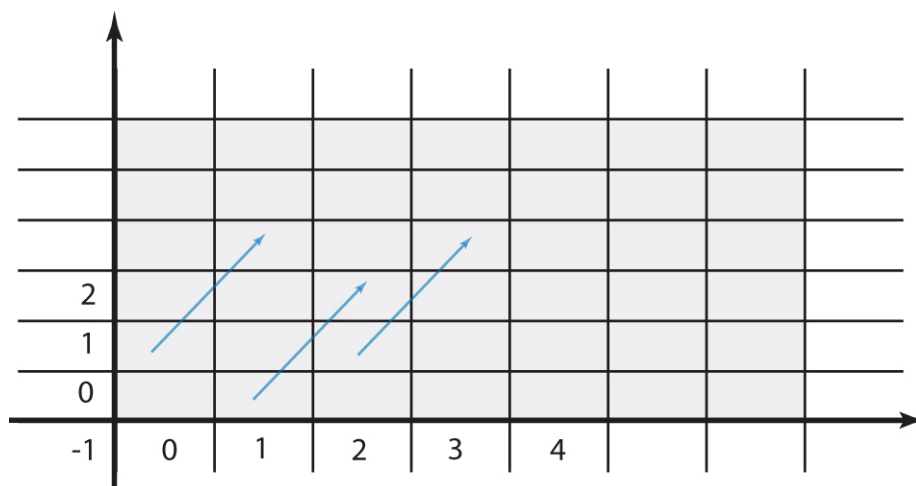


Abbildung 2: Translation auf einem regulären Gitter

Eine andere Form des Isomorphismus ist in Abbildung 3 zu sehen, die *Rotation*. Die unterliegende Gittergeometrie sei beispielsweise die regelmäßige Bedeckung der Ebene mit gleichseitigen Dreiecken. Die Rotation finde um 60° um den Punkt Z statt. Beispielhafte Wirkungen werden durch die roten Pfeile angedeutet.

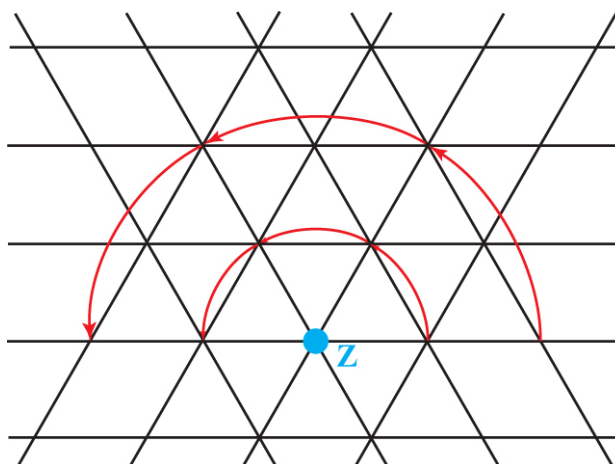


Abbildung 3: Rotation in einem regulären Gitter

Da die Behandlung anderer isotroper Gittergeometrien in unserem Zusammenhang nichts Neues bringt, soll, wenn ausnahmsweise von isotropen Geometrien die Rede ist, immer das normale quadratische Gitter einer endlichen Dimension d gemeint sein.

Wird auf die Bedingung verzichtet, daß der Raum isotrop sein soll, wird das Modell natürlich komplexer und schwieriger überschaubar, aber andererseits normalerweise realistischer. Die mathematische Theorie, mit der unter anderem solche Gittergeometrien untersucht werden, die Graphentheorie, siehe Unterkapitel 6.1 oder Bücher wie z.B. [Jun91],

ist völlig unabhängig von zellularen Automaten entstanden und hat sich in vielen Anwendungsbereichen als sehr erfolgreich erwiesen.

Eine andere Entscheidung, die im allgemeinen ebenso zu einem Gegensatz zwischen einfachen und überschaubaren lokalen Modellen einerseits und komplexen aber leistungsfähigeren globalen Modellen andererseits führt, ist die bei der Wahl der Dimension. Da es sich hier nicht um eine mathematische Arbeit handelt, soll an dieser Stelle nicht dargelegt werden, inwieweit der Dimensionsbegriff für allgemeine Graphen sinnvoll ist. Am einfachsten und, wie die Arbeiten von Wolfram zeigen, überraschend leistungsfähig ist natürlich die eindimensionale Gittergeometrie. Unter den zweidimensionalen Geometrien sind diejenigen die wichtigsten, die durch eine lückenlose Überdeckung der Ebene mit regelmäßigen n -Ecken entstehen, insbesondere das regelmäßige Gitter aus Quadraten. Natürlich können auch dreidimensionale Geometrien und sogar Geometrien noch höherer Dimensionalität sinnvoll sein, sobald man sich von der Vorstellung löst, Dimensionen müßten immer räumliche Dimensionen bedeuten. Eine solche konkretere Sicht ist in den Wissenschaften üblich, die mit diesen aus der Mathematik entlehnten Begriffen arbeiten.

3.2 Nachbarschaft

Wie sieht allgemein die Menge der Nachbarn einer Zelle in einer Gittergeometrie aus? Hat die Gittergeometrie die Form eines beliebigen Graphen, so ist der Begriff der Nachbarschaft durch die Graphentheorie eindeutig bestimmt: Für jede Zelle bilden genau diejenigen Zellen die Nachbarschaft, die direkt mit ihr verbunden sind. Für isotrope Geometrien müssen jedoch spezielle Vereinbarungen getroffen werden.

Im eindimensionalen Fall bilden natürlich im allgemeinen die beiden unmittelbaren Nachbarn die Nachbarschaft.

Für die häufigste derartige zweidimensionale Geometrie, das quadratische Gitter, im folgenden mit L bezeichnet, sind zwei verschiedene Nachbarschaftstypen üblich:

- die *von-Neumann-Nachbarschaft*

Zu jeder Zelle besteht die *von-Neumann-Nachbarschaft* aus den vier Zellen, die direkt daneben liegen. Formal besteht die Nachbarschaft N_{ij} einer Zelle (i,j) aus der Menge

$$N_{ij} = \{(k,l) \in L \mid |k-i| + |l-j| = 1\}$$

Die $(k,l) \in L$ sind also die nahe bei (i,j) liegenden Zellen. In Abbildung 4 besteht die von-Neumann-Nachbarschaft der schwarzen Zelle also aus den vier hellblauen Zellen.

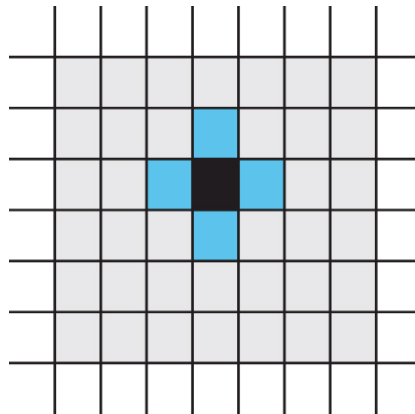


Abbildung 4: Die von-Neumann-Nachbarschaft

- die *Moore-Nachbarschaft*

Zu jeder Zelle besteht die *Moore-Nachbarschaft* aus den vier Zellen der von-Neumann-Nachbarschaft und zusätzlich aus den vier Zellen, die diagonal Kontakt dazu haben.

$$N_{ij} = \{(k,l) \in L \mid \max(|k-i|, |l-j|) = 1\}$$

Wieder bezeichnen in Abbildung 5 die hellblauen Zellen die Nachbarschaft der schwarzen.

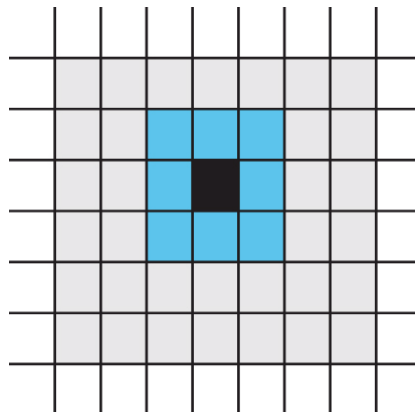


Abbildung 5: Die Moore-Nachbarschaft

Diese beiden Definitionen sind Spezialfälle von aus anderen Gebieten der Mathematik wohlbekanntem Definitionen, aus diesem Grund die natürliche Wahl und leicht auf höhere Dimensionen erweiterbar. Es werden in der Literatur mitunter auch Nachbarschaften betrachtet, zu denen zusätzlich weiter entfernte Zellen gehören. In unserem Zusammenhang spielt all das aber keine Rolle.

3.3 Zustandsmenge

Die Menge der Zustände, die eine Zelle annehmen kann, ist endlich. Welche Werte in einer Untersuchung tatsächlich verwandt werden, hängt von der Anwendung ab. Häufig reichen zwei Zustände, die etwa mit 0 und 1 bezeichnet werden. Genauere Untersuchungen erfordern jedoch oft Werte, die die Eigenschaften der untersuchten Objekte präziser darstellen, also größere Wertemengen. Es ist natürlich auch bei solchen Betrachtungen nötig, das Modell so einfach wie nötig zu halten. Das zeigt nicht nur die allgemeine Erfahrung mit der Modellierung sondern auch die folgende Überlegung:

Gegeben sei ein zweidimensionales Gitter der Größe $i*j$ aus Quadraten mit k Zuständen. So beträgt die Mächtigkeit der Zustandsmenge Z des gesamten Automaten¹⁴

$$|Z| = k^{i*j}$$

Selbst für einen kleinen Automaten mit $k = 2$, $i = j = 20$ ergibt sich daraus $|Z| = 2^{400}$, eine Zahl mit über 120 Nullen. Diese schnell anwachsende Komplexität zellulärer Automaten zeigt aber auch die Leistungsfähigkeit schon einfacher Modelle.

Historisch ist es so, daß der erste zellulare Automat von von Neumann 29 Zustände hatte, heute verwandte Automaten haben in der Naturwissenschaft selten mehr als zwei oder drei. Da in den Sozialwissenschaften, die Entitäten (Menschen) komplexer sind, kommen dort oft auch höhere Zustandszahlen vor.

3.4 Zustandsänderung

Der Zustand des Automaten ändert sich simultan zu diskreten Zeitpunkten, wobei für eine Zelle z sich der Nachfolgezustand zu einem Zeitpunkt $t+1$ abhängig von den Zuständen von z und allen Nachbarn von z zum Zeitpunkt t errechnen läßt. Seien also der Zustand der Zelle z zum Zeitpunkt t mit $z(t)$ und die Nachbarn von z mit z_1, \dots, z_n bezeichnet, so wird die Zustandsänderung durch die Funktion f beschrieben mit

$$z(t+1) = f(z(t), z_1(t), \dots, z_n(t))$$

Es ist also insbesondere so, daß für alle Zellen und alle Zeitpunkte die gleiche Übergangsfunktion f gilt. Die Wahl dieser *Zustandsüberföhrungsfunktion* (oder *Zustandsüberföhrungsregel*) f ist *das* Herzstück des Modells. Dem gegenüber entsprechen die Wahl der Zustandsmenge und die Wahl der Nachbarschaft mathematisch nur der Wahl von Werte- und Definitionsbereich. Die Leistungsfähigkeit des gegenwärtig vorgestellten Modellansatzes soll noch einmal durch eine einfache Überlegung illustriert werden:

Wir gehen zunächst von den einfachsten Strukturen aus, einem eindimensionalen zellulären Automaten mit zwei Zuständen, wobei die Nachbarschaft jeder Zelle durch ihre

¹⁴ Jedes der $i*j$ Quadrate kann jeden der k Zustände unabhängig annehmen.

beiden direkten Nachbarn definiert sei. Das bedeutet, daß in der Zelle und ihrer Nachbarschaft $2^3 = 8$ verschiedene Belegungen möglich sind, wie in Abbildung 6 zu sehen.



Abbildung 6: Belegungen im eindimensionalen Gitter

Für jede dieser acht Belegungen bestehen zwei Wahlmöglichkeiten bei der Konstruktion von f , insgesamt sind also $2^8 = 256$ verschiedene Übergangsfunktionen möglich. Für diesen unrealistisch einfachen, für weitergehende theoretische Untersuchungen unergiebigem Fall lassen sich sogar noch alle möglichen Funktionen einschließlich der möglichen Entwicklung von zellularen Automaten explizit aufführen. Der Leser sei auf [Wol02], S.54-56, verwiesen, wo eine vollständige Aufzählung zu finden ist.

Aber schon für den in der Forschung ebenfalls häufig angewandten Fall einer Moore-Nachbarschaft im zwei-dimensionalen quadratischen Gitter mit zwei Zuständen sind für jede Zelle und ihre acht Zellen umfassende Nachbarschaft $2^{8+1} = 2^9 = 512$ verschiedene Belegungen möglich, und das heißt, daß mithin insgesamt 2^{512} verschiedene Übergangsfunktionen möglich sind, eine unvorstellbare und selbst mit modernen Rechnern nicht mehr zu bewältigende Zahl.

Bei theoretischen Untersuchungen wird die Zahl der möglichen Übergangsfunktionen dadurch erheblich vermindert, daß f nicht mehr von der Belegung der Zelle und ihrer Nachbarschaft insgesamt abhängig gemacht wird, sondern nur von der Anzahl der Zellen in der Nachbarschaft, die einen bestimmten Zustand annehmen, sogenannte „totalistische“ oder „semitotalistische“ Regeln.

Definition:

- Eine Regel heißt *totalistisch*, wenn der Nachfolgezustand einer Zelle nur von der Anzahl der benachbarten Zellen (einschließlich der Zelle selbst) in bestimmten Zuständen abhängt aber nicht von deren Position. Die Zelle selber wird also genauso behandelt wie jede ihrer Nachbarzellen.
- Bei *semitotalistischen* Regeln hängt der Nachfolgezustand einer Zelle von der Anzahl der benachbarten Zellen in bestimmten Zuständen ab und unabhängig davon auch vom Zustand der Zelle selbst.

Der Zustand der Zelle selber spielt also im Gegensatz zu totalistischen Regeln bei semitotalistischen Regeln eine herausgehobene Rolle. Die letzte Regelform gilt beispielsweise beim Spiel des Lebens, wo, wie noch genauer dargelegt werden wird, nicht auf die Belegung geschaut wird sondern nur auf den Zustand der sich wandelnden Zelle und die Anzahl der toten oder lebendigen Zellen in der Umgebung. Da in der vorliegenden Arbeit keine grundlegenden Untersuchungen über zelluläre Automaten durchgeführt werden sollen, betrachten wir nur Übergangsfunktionen die im konkreten Modell benötigt werden. Diese sind, wie ich in Kapitel 11 noch begründen werde, im Dynamischen Sprachnetz alle semitotalistisch.

3.5 Randbedingungen

Soll ein zellulärer Automat nur als theoretisches Modell behandelt werden, so kann man ihn sich als (potentiell) unendlich groß vorstellen, das heißt, daß zwar zu jedem Zeitpunkt nur endlich viele Zellen besetzt sind, daß die Zahl der besetzten Zellen aber mit der Zeit über jede Grenze hinauswachsen kann. Für alle praktischen Fälle, in denen der Automat tatsächlich implementiert werden soll, und auch für viele theoretische Untersuchungen wird jedoch eine endliche Gittergeometrie angenommen. Dann erhebt sich die Frage, was am Rand geschehen soll, da dort die Zellen keine vollständige Nachbarschaft besitzen. Einer der häufigsten Tricks, wie pathologisches Randverhalten vermieden werden kann, ist es, den linken mit dem rechten Rand und den oberen mit dem unteren Rand zu verkleben. Im eindimensionalen Fall entsteht dadurch ein Ring, im zweidimensionalen Fall ein Torus.

Zum Abschluß dieses Kapitels sollen zwei fachlich wichtige Details den Lesern nicht verheimlicht werden: eine kurze Ausführung zum Spiel des Lebens und eine mathematisch exakte Definition zellulärer Automaten, zitiert nach [DM99].

3.6 Das Spiel des Lebens

Das Spiel des Lebens wurde erstmals 1968 vom britischen Mathematiker John Horton Conway beschrieben. Die Akteure sollten ursprünglich das Wachstum biologischer Zellkulturen auf einem zweidimensionalen rechteckigen Feld nachbilden. Es gelten die Festlegungen

- Die Gittergeometrie wird durch das isotrope zweidimensionale Gitter gebildet.
- Die Nachbarschaft einer Zelle ist die Moore-Nachbarschaft.
- Die Zustandsmenge hat zwei Elemente, die die Eigenschaften „lebendig“ und „tot“ widerspiegeln und meistens mit 1 oder 0 bezeichnet werden.
- Die Übergangsfunktion ist gegeben durch die Vorschrift, daß der Nachfolgezustand einer Zelle im Zustand 1 dann 1 ist, wenn sie zwei oder drei Nachbarn im Zustand 1 hat. Der Nachfolgezustand einer Zelle im Zustand 0 ist dann 1 , wenn sie genau drei Nachbarn im Zustand 1 hat. In allen anderen Fällen ist der Nachfolgezustand 0 . Das bedeutet im Falle eines Zustandswechsels ein Sterben an Einsamkeit oder an Überbevölkerung.

An dieser Definition ist zu ersehen, daß wir es hier mit einer semi-totalistischen Regel zu tun haben. Beim Übergang ist eben nicht wesentlich, wo die Zellen im Zustand 0 oder 1 nun genau liegen. Wichtig ist ihre Zahl. Das trifft sowohl auf totalistische wie auf semi-totalistische Regeln zu. Semi-totalistisch ist die Regel im Spiel des Lebens, weil es für den Fall, daß die Zelle, deren Nachfolgezustand zu berechnen ist, im Zustand 0 ist, andere Übergangsgesetze gibt, als wenn sie im Zustand 1 wäre.

Es entstehen beim Spiel des Lebens vielfältige und überraschende Zellmuster; und es ist

nicht vorhersehbar, wie sich diese Zellmuster im Laufe der Zeit entwickeln. Wir haben hier also ein Musterbeispiel dafür, wie sich aus einfachen lokalen Wechselwirkungen einfacher Bausteine eine Welt selbst organisieren kann, die aus globaler Sicht ungeheuer komplex ist. Der zu einer Behandlung der globalen Abstraktionsebene notwendige theoretische Apparat ist beträchtlich. Er kann in [Wol02] studiert werden. An dieser Stelle soll das in Abbildung 7 angeführte Beispiel des „Blinkers“ reichen.

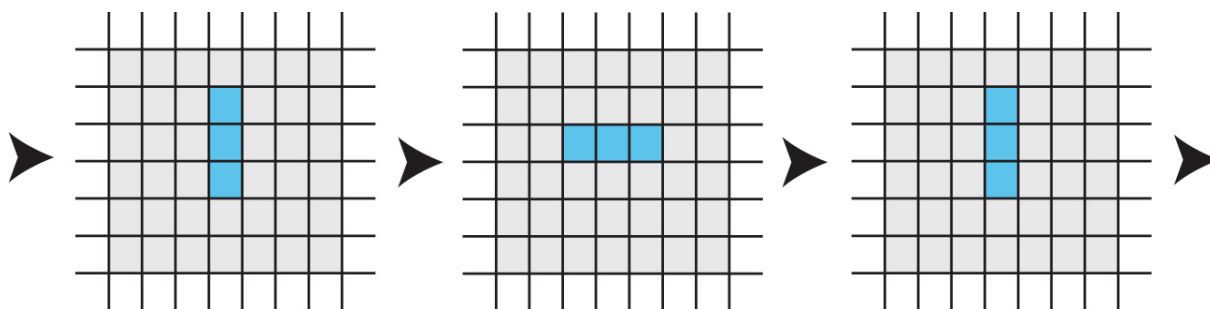


Abbildung 7: Spiel des Lebens – Der „Blinker“

Beim Blinker handelt es sich um die Formation, bei der drei benachbarte Felder besetzt, das heißt im Zustand 1 , sind. Sind in der Nähe keine weiteren Nachbarn im Zustand 1 , die die weitere Entwicklung stören könnten, dann entwickelt sich der Blinker so, daß seine zentrale Zelle belebt bleibt, während die beiden Randzellen absterben. Dafür haben die beiden anderen Zellen, die der zentralen Blinkerzelle benachbart sind, ja genau drei besetzte Nachbarn. Sie werden also im nächsten Schritt besetzt, so daß dann ein neuer Blinker entsteht, der zu dem ursprünglichen orthogonal ist. Im nächsten Schritt entsteht dann wieder der ursprüngliche Blinker usw. ad infinitum.

In [Wol02] wird auch nachgewiesen, daß Conways Übergangsregeln in einem strengen Sinne die einzigen sind, die auf dem isotropen zweidimensionalen Gitter zu interessanten Strukturen und einer interessanten dynamischen Entwicklung führen.

Es kann außerdem bewiesen werden, daß das Spiel des Lebens nicht nur das Wachstum von Zellkulturen nachmodellieren kann sondern auch die logischen Grundbausteine moderner Computer. Das heißt, das Verhalten dieser Computer kann ebenfalls imitiert werden. Damit ist das Spiel des Lebens so mächtig wie die Turing-Maschine.

3.7 Klassische zellulare Automaten

Definition: (nach [DM99]) Ein *zellulärer Automat* A ist ein 4-Tupel $A = (d, S, N, f)$. Dabei sind:

- d eine natürliche Zahl, die *Dimension*,

- S eine endliche Menge, die *Zustandsmenge*,
- N ein endlicher Vektor mit Elementen in \mathbb{Z}^d , die *Nachbarschaft*, wobei Z die Menge der ganzen Zahlen bezeichnet,
- $f: S^{n+1} \rightarrow S$, die lokale *Übergangsfunktion* oder *Regel* von A . Die Zahl n ist genau die Länge von N . Durch f wird also jeder möglichen Kombination Zuständen in der Zelle und in der Nachbarschaft der Nachfolgezustand zugeordnet.

Wie man sieht, definieren Delorme und Mazoyer zellulare Automaten hier nur mit Gittergeometrien, die aus (d -dimensionalen) ganzzahligen Gittern bestehen. Wenn sie im weiteren Verlaufe auch mit anderen Geometrien arbeiten, ziehen sie je nach Bedarf weitere Definitionen hinzu.

Beispiel: Für das Spiel des Lebens sind

- $d = 2$,
- $S = \{0,1\}$,
- N ein Vektor der Länge 8 von Paaren ganzer Zahlen, also $n=8$. Zum Punkt $(3,4)$ etwa ist
$$N = \{(2,3), (2,4), (2,5), (3,3), (3,5), (4,3), (4,4), (4,5)\}.$$
- Die Übergangsfunktion ist schon oben erklärt worden. Für den Punkt $(3,4)$ etwa wird sie mit Hilfe der Werte auf den $n+1=9$ Punkten von N zuzüglich $(3,4)$ selber definiert.

4 Modellierung von Sprachwandel durch zellulare Automaten

Die Modellbildung für sprachlichen Wandel soll dahingehend erfolgen, daß eine Zelle des Automaten einer elementaren Einheit im Sprachgeschehen entspricht. Das können einzelne Sprecher, Cluster von Sprechern oder andere Handelnde, etwa Massenmedien, sein. Der Zustand der Zelle soll das Maß widerspiegeln, in dem die Innovation von der entsprechenden Einheit verwandt wird.

4.1 Metaregeln

Einer der Hauptgrundsätze, die auch in [Wol02] für die Modellierung mit Hilfe zellulärer Automaten aufgestellt werden, ist die naheliegende Forderung, das Modell solle so einfach wie möglich sein. Es wird dort auch gezeigt, wie aus sehr einfachen Überföhrungsfunktionen ein überraschend komplexes Verhalten des Gesamtsystems resultiert. Wie wir eben gesehen haben, wächst die Zahl der grundsätzlich möglichen Überföhrungsfunktionen sehr schnell über jedes Vorstellungsvermögen hinaus. Um in diesen Überfluß von Möglichkeiten etwas Ordnung zu bringen, stelle ich daher im folgenden einige *Metaregeln für sinnvolle Überföhrungsregeln* in dem uns interessierenden Zusammenhang auf¹⁵. Die Idee einer Modellbildung allgemein ist einfach die, daß man versucht, das Verhalten der beobachteten Objekte, hier das sprachliche Verhalten von individuellen Sprechern, so zu abstrahieren, daß etwas mathematisch Sinnvolles entsteht. Dabei hat es sich für diese Arbeit als nützlich erwiesen, daß zunächst einige notwendige Bedingungen für den Begriff des „Sinnvollen“, vorwiegend unter sprachlichem Aspekt, aufgestellt werden. Diese „Regeln für Regeln“ formuliere ich hier zunächst für zellulare Automaten; sie werden später auch etwas allgemeiner für zellulare Graphen gelten. Dann kann ich mich später, ausgehend von allen mathematisch möglichen Regeln, auf die Untersuchung der metaregelkonformen Regeln beschränken. Zunächst werden einige Begriffe eingeföhrt:

Es sei Z die Menge der möglichen Zustände einer Zelle. Dann läßt sich auf Z in natürlicher Weise eine Anordnung definieren. Für zwei Zustände z_1 und z_2 sei $z_1 \leq z_2$, wenn z_1 die Innovation höchstens so häufig benutzt wie z_2 . Im Falle einer zweielementigen Zustandsmenge gölte also beispielsweise $0 \leq 1$, wenn 0 bedeutet, daß die Innovation nicht verwandt wird, 1 aber deren durchgängige Nutzung anzeigt. Der Vollständigkeit halber sei angemerkt, daß auch $0 \leq 0$ und $1 \leq 1$ gelten. Eine andere Art der Modellierung, die im folgenden kurz verwandt werden wird, ist eine dreielementige Zustandsmenge $\{0, 1, 2\}$ mit der intuitiv klaren Ordnungsrelation. In dem jetzt zunächst betrachteten Modell bedeutet 0 , daß

¹⁵ Weder die Überlegung, daß Metaregeln generell sinnvoll sein können, um eine Ordnung in die Vielzahl der möglichen Übergangsfunktionen zu bringen, noch die konkrete Bestimmung, welche konkreten Regeln sowohl aus mathematischer als auch aus linguistischer Sicht fruchtbringend sein können, habe ich irgendwo sonst gefunden, insbesondere auch nicht bei Wolfram. Es sollte mich aber dennoch nicht wundern, wenn vergleichbare Gedanken schon irgendwo publiziert sein sollten.

Auf jeden Fall werden wir sehen, daß die einfachen Metaregeln, die ich in diesem Kapitel vorstelle, bei der Konstruktion des Dynamischen Sprachnetzes in Kapitel 11 ungemein nützlich sind. Mit ihrer Hilfe läßt sich die Menge der möglichen Regeln gut in den Griff bekommen.

die Innovation nicht verwandt wird, 1 zeigt die teilweise Nutzung und 2 ihre durchgängige Nutzung an. Später werden wir erneut zu einem dreielementigen Modell zurückkehren. Dann wird die Interpretation der Zustände aber anders sein.

Diese Ordnung läßt sich sofort auf Vektoren von Zuständen hochheben.

Definition: Für zwei Vektoren von Zuständen v_1 und v_2 sei $v_1 \leq v_2$, wenn für alle i die i -te Komponente von v_1 mit der i -ten Komponente von v_2 in der \leq -Relation steht.

So ist beispielsweise $(0,1,0) \leq (0,1,1)$. Denn jede Komponente des linken Vektors ist kleiner als die entsprechende Komponente des rechten Vektors oder gleich. Aber es gilt andererseits nicht $(0,1,0) \leq (1,0,1)$, da zwar die Werte in der ersten und der dritten, nicht aber in mittleren Komponente in der \leq -Relation stehen.

Man sollte beachten, daß im letzten Fall auch nicht gilt $(1,0,1) \leq (0,1,0)$. Hier handelt es sich um eine Struktur, die in der Mathematik als Halbordnung¹⁶ bezeichnet wird. Die Betrachtung von Vektoren von Zuständen ist deshalb nötig, weil die Zustandsüberföhrungsfunktion auf Zuständen von Nachbarschaften einer Zelle, also auf Vektoren von Zuständen, definiert ist. Schließlich sei für einen Zustandsvektor $v = (z_1, \dots, z_n)$ das Maximum $\max(v)$ definiert als der maximale der Zustände z_1, \dots, z_n . Analog sei das Minimum $\min(v)$ definiert. Wir sind jetzt in der Lage, die Metaregeln zu definieren.

4.2 Monotonie

Eine Übergangsregel f heißt *monoton*, wenn für alle Zustandsvektoren v_1, v_2 in Nachbarschaften von Zellen z_1, z_2 gilt:

$$v_1 \leq v_2 \Rightarrow f(v_1) \leq f(v_2)$$

Das ist die ganz normale mathematische Definition monotoner Funktionen auf Halbordnungen und bedeutet in unserem Fall anschaulich, daß, wenn zu einem bestimmten Zeitpunkt die Innovation in einer bestimmten Umgebung häufiger verwandt wird als das in einer Vergleichsmessung der Fall ist, dann wird sich dieses Verhältnis zum Folgezeitpunkt zumindest nicht umkehren¹⁷.

Beispiel: Es sei eine dreielementige Zustandsmenge gewählt. Wir gehen von einem eindimensionalen Netz aus; die Nachbarschaft eines Knotens bestehe aus den beiden angrenzenden Knoten. Damit ist f auf Tripeln mit Einträgen aus der Menge $\{0,1,2\}$ definiert. Dabei stehen zwei Einträge für die Zustände der beiden Nachbarn und einer für den

¹⁶ In einer Halbordnung gelten zwar die üblichen Axiome der beispielsweise von Zahlen her bekannten Ordnungen, also

1. Reflexivität: $a \leq a$,
2. Antisymmetrie: $a \leq b \wedge b \leq a \Rightarrow a = b$
3. Transitivität: $a \leq b \wedge b \leq c \Rightarrow a \leq c$

Es muß in einer Halbordnung aber nicht notwendigerweise für alle Paare a, b entweder $a \leq b$ oder $b \leq a$ gelten.

¹⁷ Hier sollte man nicht aus der Monotonie der Regeln auf die Monotonie der Anzahl der Zellen in einem bestimmten Zustand schließen. Letztere kann durchaus schwanken, insbesondere bei komplexeren Gittergeometrien.

ursprünglichen Zustand des betrachteten Knotens. Dann bedeutet etwa anschaulich der Zustandsvektor $(0,1,0)$, daß die Neuerung von dem diesem Knoten zugeordneten Sprecher gelegentlich benutzt wird aber überhaupt nicht von dessen Nachbarn; und die Festlegung der Übergangsfunktion

$$f((0,1,0)) = 1$$

heißt anschaulich: Wird eine Innovation von einem Sprecher gelegentlich benutzt, so wird sie zum nächsten Zeitpunkt immer noch gelegentlich benutzt, auch wenn dies kein Nachbar tut. Aus $(0,1,0) \leq (0,1,1)$ folgt wegen der Monotonie von f auch $f((0,1,0)) \leq f((0,1,1))$. Das bedeutet, daß nur noch gelten kann:

$$f((0,1,1)) = 1 \text{ oder } f((0,1,1)) = 2, \text{ aber nicht } f((0,1,1)) = 0,$$

also anschaulich: Benutzt jetzt zusätzlich einer der Nachbarn die Innovation gelegentlich, so benutzt der Sprecher erst recht zum nächsten Zeitpunkt noch diese Innovation gelegentlich, möglicherweise unter dem Einfluß des Nachbarn sogar immer. Es ist aber unlogisch (und daher durch diese Metaregel ausgeschlossen), daß er unter dem Einfluß eines gelegentlich in der neuen Art sprechenden Nachbarn diese Innovation ablegt, obwohl er sie bei ausschließlich konservativen Nachbarn beibehalten würde.

4.3 Angemessenheit

Eine Übergangsregel f heißt *angemessen*, wenn für alle Zustandsvektoren v in der Nachbarschaft einer Zelle z gilt:

$$\min(v) \leq f(v) \leq \max(v)$$

Das bedeutet anschaulich, daß der Wert des Zustandes einer Zelle z zu einem Zeitpunkt $n+1$ höchstens so groß wie der Wert des höchsten Zustandes in der Umgebung von z zum Zeitpunkt n sein darf, entsprechend für den Wert des niedrigsten Zustandes.

Beispiel: Wie im Beispiel eben sei wieder eine dreielementige Zustandsmenge gewählt. Dann sind

$$f((0,1,1)) = 0 \text{ und } f((0,1,1)) = 1 \text{ angemessene Werte, aber nicht } f((0,1,1)) = 2,$$

also anschaulich annähernd: Wird eine Innovation von einem Sprecher (mittlere Komponente) gelegentlich benutzt, so wird sie zum nächsten Zeitpunkt auch allenfalls gelegentlich benutzt, wenn keiner der Nachbarn sie immer benutzt. Der Sprecher beginnt aber im allgemeinen nicht einfach aus sich heraus ohne Vorbild mit einer durchgängigen Nutzung. Dem Phänomen eines ersten Nutzers, eines „Trendsetters“, wird im Dynamischen Sprachnetz durch Rückgriff auf Methoden der Epidemiologie in Form einer ersten „Infektion“ Rechnung getragen werden.

Ob Angemessenheit in der Realität tatsächlich immer vorliegt, das mag bezweifelt werden. Aber erstens ist das in den meisten Fällen tatsächlich der Fall und zweitens würde ein

dahingehender Einwand grundsätzlich das Wesen der Modellbildung mit zellularen Automaten verkennen, insbesondere soweit es die Erörterungen in [Wol02] angeht. Dort wird immer wieder unterstrichen, wie nötig es ist, sich von Details und Sonderfällen zu lösen. Ich erinnere dazu noch einmal an das Beispiel der Planetensysteme, die lediglich im Modell aber nicht in der Wirklichkeit durch Differentialgleichungen gesteuert werden, aus Unterkapitel 2.5.

4.4 Symmetrie

Eine Übergangsregel f heißt *symmetrisch*, wenn für alle Zustandsvektoren v in der Nachbarschaft einer Zelle z und für alle Permutationen $perm$ der Nachbarschaft gilt:

$$f(v) = f(perm(v))$$

Eine Permutation eines Vektors v ist ein Vektor v' , der dieselben Werte enthält wie v , aber in einer anderen Reihenfolge. Das bedeutet anschaulich, daß der Wert des Zustandes einer Zelle z zu einem Zeitpunkt $n+1$ unabhängig von der genauen Lage der beeinflussenden Nachbarn zum Zeitpunkt n ist; platt ausgedrückt: Es ist gleichgültig, ob der mich beeinflussende Sprecher links oder rechts von mir steht.

Beispiel: Eine Zelle habe in einem eindimensionalen Gitter den Zustand 1 , ihr linker Nachbar den Wert 0 und der rechte Nachbar den Wert 1 . Dann muß sie zum Folgezeitpunkt denselben Zustand haben, als wenn die Zustände ihrer Nachbarn gerade umgekehrt wären, also

$$f((0,1,1)) = f((1,1,0))$$

Dadurch ist selbstverständlich keiner der Werte $f((0,0,1))$, $f((0,1,0))$, $f((1,1,1))$ festgelegt. Es gilt die folgende

Bemerkung: Semitotalistische Regeln gehorchen der Metaregel der Symmetrie.

Die Wahrheit dieser Beobachtung ist unmittelbar einsichtig, da es bei semitotalistischen Regeln ja gerade nicht auf die relative Lage der Knoten zueinander ankommt, sondern nur auf die Anzahl der Nachbarn in bestimmten Zuständen. Die bleibt bei Permutationen aber gerade gleich.

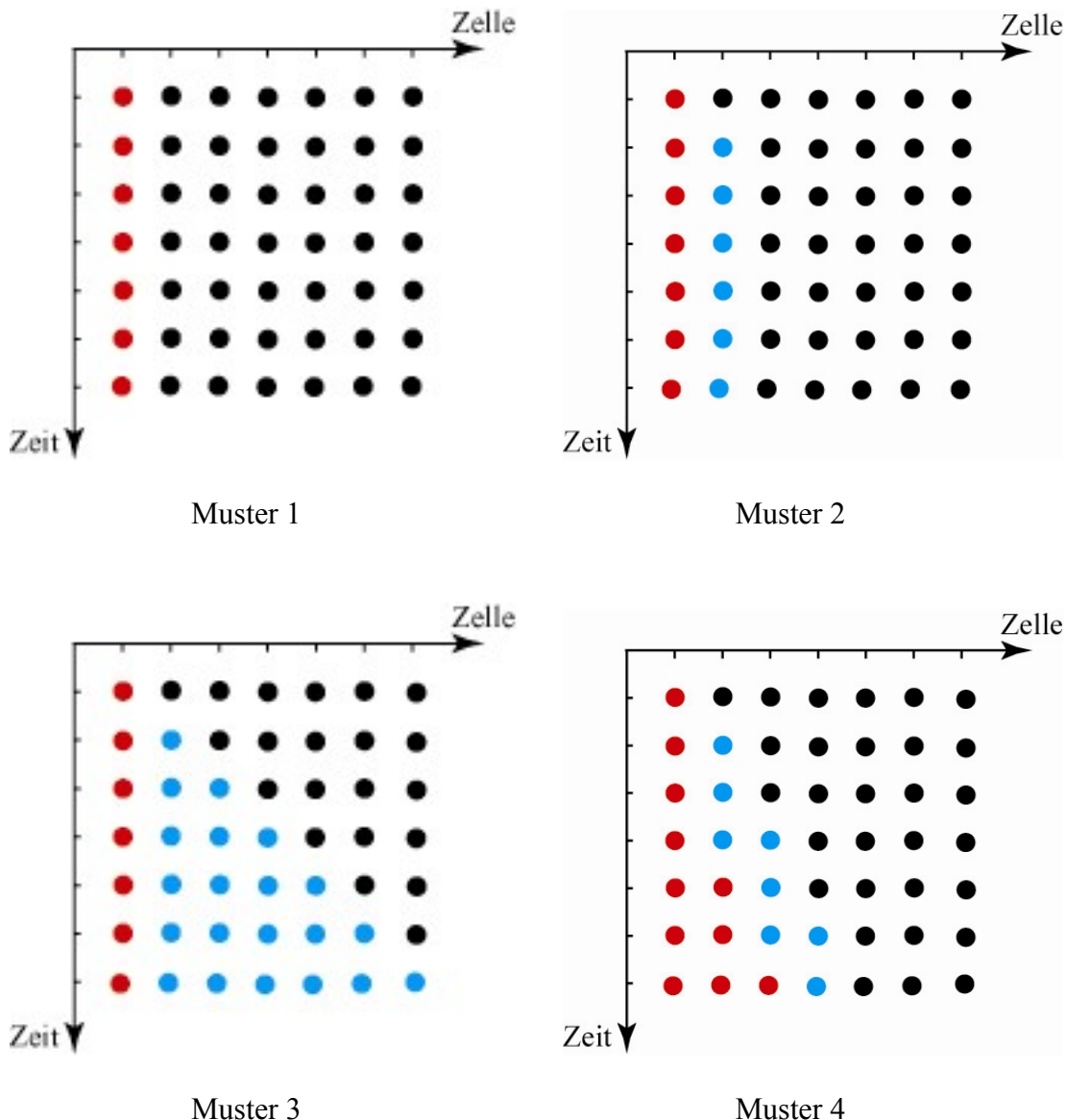
4.5 Eindimensionale Geometrien

Mit Blick auf Einsteins Forderung nach möglichst einfachen Modellen erhebt sich die Frage:

„Wie komplex muß ein sinnvolles Modell für die Ausbreitung einer sprachlichen Innovation mindestens sein?“

Diese Frage wird einer der Kernpunkte unserer folgenden Untersuchungen sein. Zunächst sollen eindimensionale zellulare Automaten betrachtet werden. Für den Fall einer Nachbarschaft, die zu jeder Zelle nur aus den beiden unmittelbar benachbarten Gitterzellen besteht, ergibt sich folgende Situation: Ich übergehe hier den viel zu einfachen Fall einer zweielementigen Zustandsmenge. Der Fall einer dreielementigen Zustandsmenge läßt sich vollständig mit Hilfe des Rechners untersuchen.

Die beigefügten, in Unterkapitel 13.1 näher kommentierten Programme und deren Ergebnis zeigen, daß es in diesem Falle genau 64 Regeln gibt, die den in Kapitel 4 aufgestellten Metaregeln genügen. Keine dieser Regeln läßt irgendeine komplexe Systementwicklung erkennen. In jedem Anfangszustand gibt es genau eine Störung (ein einfaches erstes Auftreten der Innovation). Die Programme zeigen genau sechs Ausbreitungsmuster, die in Abbildung 8 vorgestellt werden.



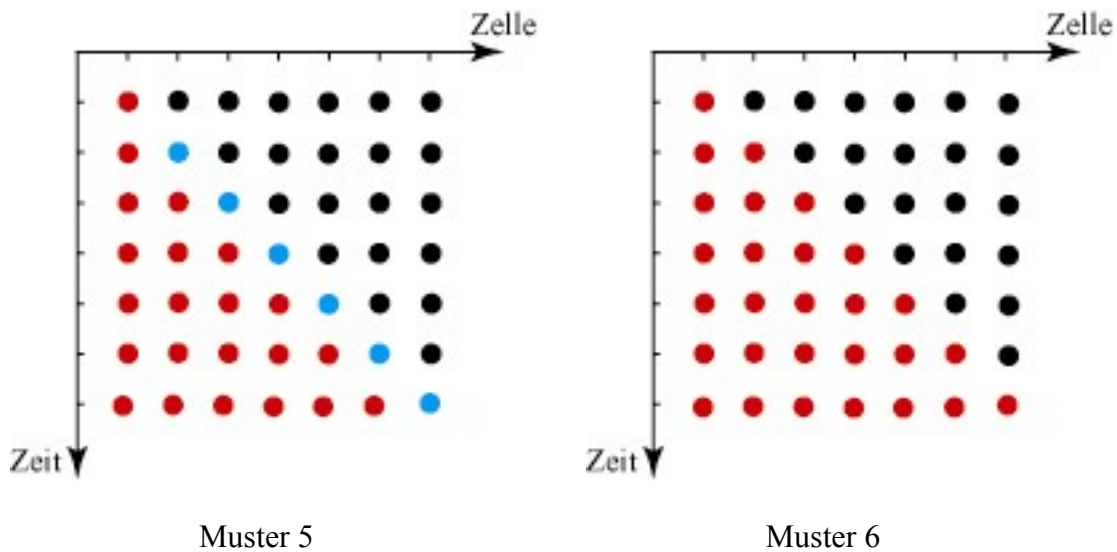


Abbildung 8: Ausbreitungsmuster eindimensional, drei Zustände

Die Farben zeigen den Zustand an: Es bedeuten

- rot: Zustand 2 (ständiger Gebrauch der Neuerung),
- blau: Zustand 1 (gelegentlicher Gebrauch der Neuerung),
- schwarz: Zustand 0 (kein Gebrauch der Neuerung).

In der ersten Zeile (Anfang der Infektion) findet genau ein Zustandswechsel von 0 in 2 statt. Für den weiteren Verlauf zeigen sich in den folgenden Zeilen damit genau drei Möglichkeiten:

- Die Störung pflanzt sich (im wesentlichen) gar nicht fort (Muster 1 und 2).
- Die Störung pflanzt sich linear mit einer halben Zelle je Zeiteinheit fort (Muster 4).
- Die Störung pflanzt sich linear mit einer ganzen Zelle je Zeiteinheit fort (Muster 3, 5 und 6).

4.6 Mehrdimensionale Geometrien

Offenbar sind also die Regeln des letzten Unterkapitels zu simpel, als daß sie das sprachliche Geschehen modellieren könnten. Dazu sind die Ausbreitungsmuster schlicht zu simpel. Das werden wir insbesondere im Zusammenhang mit dem Piotrovskij-Gesetz, Kapitel 5, sehen. Die Frage lautet, ob die zu große Einfachheit von der zu eingeschränkten Zustandsmenge oder von der zu einfachen Gittergeometrie kommt. Eine Antwort liefert der weiter unten in diesem Unterkapitel angegebene Satz, bei dem die sogenannte *o-Notation*¹⁸

¹⁸ Die mathematische Definition der sogenannten *o-Notation* ist die folgende:

verwandt wird¹⁹. Zunächst soll die o-Notation und ihr Zusammenhang mit dem Größenwachstum von Funktionen durch ein Beispiel verdeutlicht werden.

Beispiel: Das Größenwachstum der Funktionen f und g mit

$$f(h) = h^2 \quad \text{und} \quad g(h) = h^2 + h + 1$$

ist gleich, da sich auf lange Sicht der Term h^2 größtmäßig durchsetzt. Das ist an den beiden folgenden Abbildungen zu sehen. Bei Abbildung 9 sind nur kleine Werte zu sehen. Es ist deutlich zu erkennen, daß sich die absolute Differenz zwischen den beiden Funktionen sogar vergrößert.

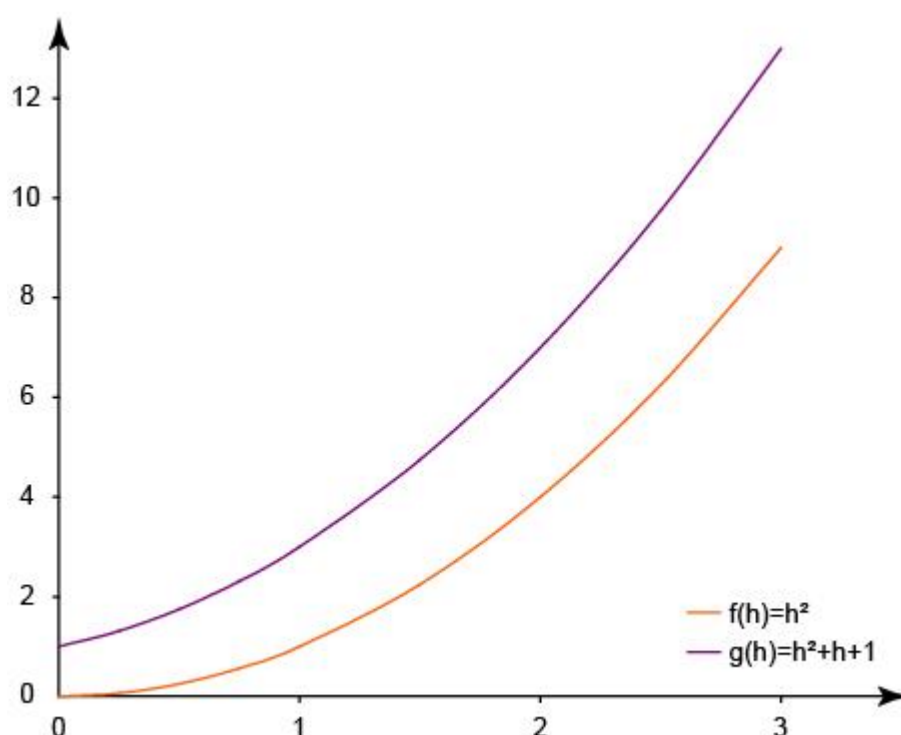


Abbildung 9: Wachstum zweier Funktionen (kleine Werte)

Bei einer Betrachtung des gesamten Verlaufs erkennt man jedoch, daß die beiden Funktionen sich, relativ gesehen, bei immer größeren Werten immer mehr einander annähern. In Abbildung 10 zeigt sich das daran, daß der absolute Unterschied kaum mehr darstellbar ist.

Eine Funktion f ist $o(g)$, wenn gilt: Es gibt eine natürliche Zahl n_0 und eine reelle Zahl c , so daß für alle $n > n_0$ gilt: $|f(n)/g(n)| < c$.

Damit macht die o-Notation nur eine Aussage über das Größenwachstum der beiden Funktionen f und g . Anschaulich bedeutet die Definition, daß das Wachstum der Funktion f bei wachsendem n (hier konkret: immer ab dem festen Wert n_0) durch die Funktion g nach oben begrenzt ist.

Eine duale Definition bezeichnet eine Funktion f , deren Wachstum bei wachsendem n durch die Funktion g nach unten begrenzt ist. Eine Funktion f ist $\Omega(g)$, wenn gilt: Es gibt eine natürliche Zahl n_0 und eine reelle Zahl c , so daß für alle $n > n_0$ gilt: $f(n)/g(n) \geq c$.

¹⁹ Der angepeilte Satz und sein Beweis stammen an dieser Stelle von mir. Obwohl ich nirgendwo etwas Vergleichbares gefunden habe, vermute ich aber, daß ich nicht der erste bin, der solche Gedanken schon veröffentlicht hat.

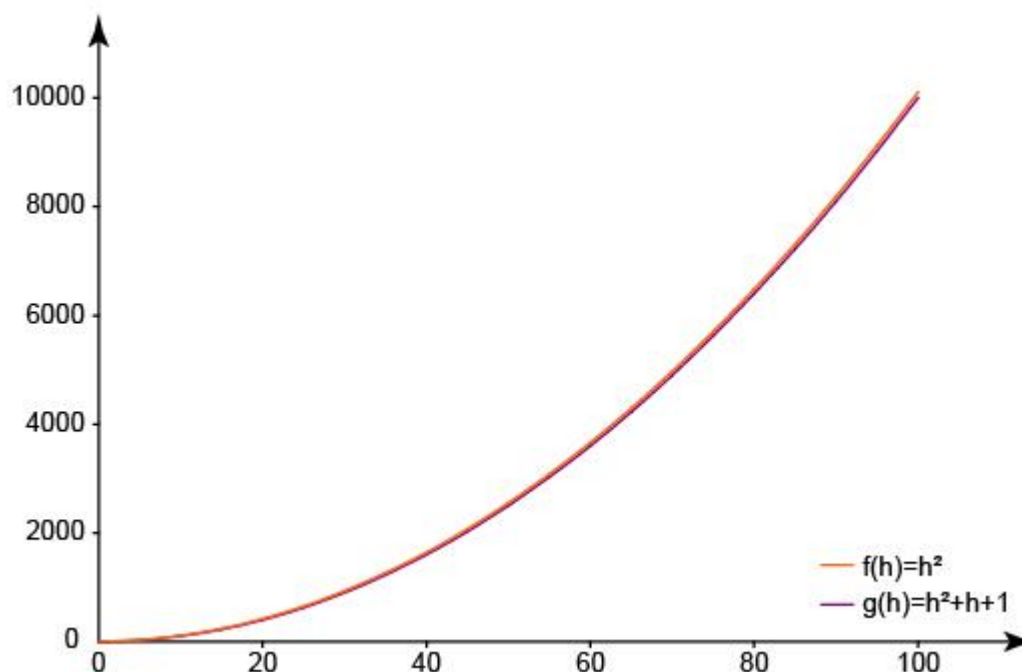


Abbildung 10: Wachstum zweier Funktionen (größerer Bereich)

Für die o-Notation bedeutet dies

$$f \in o(g) \quad \text{und} \quad g \in o(f)$$

Die erste Aussage ist klar, da g das f offensichtlich majorisiert. Für die zweite Aussage wählen wir in der Definition der o-Notation in der Fußnote $c=2$ und $n_0=2$. Dann gilt

$$g(n)/f(n) = (n^2 + n + 1)/n^2 < 2 = c$$

für alle n , die größer oder gleich 2, also gleich n_0 , sind. Daher gilt auch die zweite Aussage.

Zum Beweis des angekündigten Satzes werde ich eine Technik benutzen, die „vollständige Induktion“ genannt wird. Vollständige Induktion ist eine mathematische Schlußweise, bei der eine allgemeingültige Aussage $A(n)$ für alle natürlichen Zahlen n dadurch nachgewiesen wird, daß man zunächst $A(0)$ oder $A(1)$ zeigt und dann beweist, daß für alle i aus der Gültigkeit von $A(i)$ die Gültigkeit von $A(i+1)$ folgt.

Zuerst soll zwecks einfacheren Verständnisses ein anschauliches Beispiel für für den Schluß $A(i) \Rightarrow A(i+1)$ folgen:

Beispiel: Es geht um ein einfaches Beispiel aus der Zahlentheorie. Behauptet wird, daß die Summe der ersten n ungeraden Zahlen gleich n^2 ist, und zwar für alle Zahlen. Da sich die i -te ungerade Zahl durch $2i-1$ ausdrücken läßt, lautet der Satz formal:

$$\sum_i^n (2i-1) = n^2$$

Die Aussage $A(1)$ ist dann $2 \cdot 1 - 1 = 1^2$, eine unbestreitbare Wahrheit. Die Aussage $A(n)$ ist identisch mit dem zu beweisenden Satz. Sie darf als wahr angenommen werden. Zu beweisen ist dann $A(n+1)$. Dafür gilt

$$\sum_{i=1}^{n+1} (2i-1) = \sum_{i=1}^n (2i-1) + 2 \cdot (n+1) - 1 = n^2 + 2 \cdot n + 1 = (n+1)^2$$

Die erste Gleichheit ist die Definition des Summenzeichens, die zweite folgt aus der Wahrheit von $A(n)$, die dritte ist trivial (Binomische Formel). Damit ist $A(n+1)$ nachgewiesen, und der Satz gilt für alle natürlichen Zahlen.

Qed

Der Beweis kann durch Abbildung 11 veranschaulicht werden: Im siebenten Schritt seien schon die $7^2 = 49$ blau umrandeten Kästchen konstruiert. Im achten Schritt kommen die $2 \cdot 8 - 1 = 15$ grün umrandeten (insgesamt also 64) und im neunten Schritt die $2 \cdot 9 - 1 = 17$ rot umrandeten Kästchen hinzu (zusammen also 81), so daß in jedem Falle ein Quadrat entsteht.

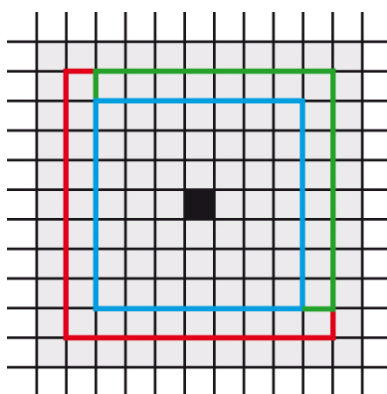


Abbildung 11: Wachstum von Quadratzahlen

Nach dem Beispiel folgt der schon angekündigte

Satz: Sei A ein isotroper zellulärer Automat einer endlichen Dimension d . Die Übergangsregel lasse keine Entstehung einer Störung in einer ungestörten Zelle zu, wenn auch ihre gesamte Umgebung ungestört ist. (Das ist etwa bei allen angemessenen Regeln der Fall.) Es trete in einer Zelle eine Störung auf. Dann pflanzt sich die Störung höchstens mit einer Geschwindigkeit f mit $f(n) \in o(n^d)$ fort.

Beweis: Der Beweis wird geführt, indem wir zunächst festlegen, daß k die größte Entfernung in Richtung irgendeiner der Koordinaten sei, so daß gemäß der Übergangsregel eine Zelle noch durch ihre Umgebung beeinflusst werden kann. (Bei der Moore- und bei der von-Neumann-Umgebung ist also $k = 1$.) Dann wird durch vollständige Induktion die folgende Aussage bewiesen:

- (*) Nach dem n -ten Schritt sind alle gestörten Zellen in einem d -dimensionalen Würfel der Kantenlänge $2kn+1$ enthalten.

Der Beweis ist einfach:

$A(0)$: Für $n=0$ gibt es nach Voraussetzung genau eine gestörte Zelle, die in einem Würfel der Kantenlänge $2k*0+1 = 1$ enthalten ist.

$A(i) \Rightarrow A(i+1)$: Seien nach dem i -ten Schritt alle gestörten Zellen in einem Würfel der Kantenlänge $2ki+1$ enthalten. Im $i+1$ -ten Schritt kann die Störung sich nur höchstens um k Zellen in jede Richtung ausbreiten. Damit ist die Ausdehnung der Störung auf jeder Koordinatenachse höchstens gleich $2ki+1+k+k = 2k(i+1)+1$. Ein Beispiel für diesen Schluß wird nach dem Ende des Beweises gegeben.

Damit ist (*) bewiesen. Das heißt aber, daß nach dem n -ten Schritt höchstens $(2kn+1)^d$ Zellen gestört sind. Sei $f(n)$ die Anzahl der gestörten Zellen nach dem n -ten Schritt. Dann ist

$$f(n) \leq (2kn+1)^d \in o(n^d)$$

Die letzte Inklusion ist aus der Komplexitätstheorie wohlbekannt, was den obigen Satz beweist.

Qed

Das wichtige an dem Satz ist weniger die konkrete Aussage sondern seine Interpretation: Im Modell eines isotropen zellularen Automaten pflanzt sich eine Störung höchstens mit polynomieller Geschwindigkeit²⁰ fort. In Unterkapitel 6.1 werde ich einen bedeutenden Schluß daraus ziehen, der für mich sehr überraschend war.

²⁰ Eine Funktion f hat polynomielles Wachstum, wenn es ein Polynom p gibt mit $f \in o(p)$, daß f also etwa so schnell wächst wie ein Polynom. Polynomielle Geschwindigkeit heißt in diesem Zusammenhang, daß die Anzahl $A(n)$ der gestörten Zellen in Abhängigkeit vom Zeitpunkt n durch die Funktion A mit polynomielltem Wachstum beschrieben wird.

5 Das Piotrovskij-Gesetz und verwandte Ansätze

Um in 6.1 deutlich machen zu können, warum die im letzten Abschnitt erzielten Einsichten es ausschließen, daß der Sprachwandel auf eine derartig vereinfachte Art, wie ich es bisher getan habe, erklärt werden kann, möchte ich auf einige ältere Ergebnisse eingehen, insbesondere auf das „Piotrovskij-Gesetz“, das sich mit der speziellen Form der sprachlichen Innovation beschäftigt, die durch das Vordringen eines Lehnwortes gegeben ist. Ich stütze mich dabei im wesentlichen auf die Arbeiten von G.Altmann [ABRS83], [Alt83], [Alt85]. Eine neuere Arbeit, die zeigt, wie weitreichend das Piotrovskij-Gesetz ist, ist [Best03a].

Das Piotrovskij-Gesetz stellt eine hypothetische Aussage über den zeitlichen Verlauf der Veränderung einer sprachlichen Entität dar. Dabei stellt man sich einen Verlauf vor, bei dem die Veränderung langsam beginnt, dann, sobald sie allgemein akzeptiert wird, beschleunigt verläuft und dann wieder langsamer vor sich geht, wenn sie möglicherweise auf einen harten Kern von Ablehnenden trifft oder die gesamte Sprachgemeinschaft erfaßt hat. In [Alt83] werden verschiedene Autoren zitiert, die einen solchen „S-förmigen“ Verlauf postulieren ([OS65], [WLH68], [Bai73]). Diese Beobachtungen und Messungen wurden dann in [PP74] auf eine theoretische Grundlage gestellt, wobei die mathematischen Grundlagen durch die *arctg*- (arcustangens) und die *tanh*-Funktion (tangens hyperbolicus), beide in der postulierten S-Form verlaufend, gegeben werden.

Ein anderer Ansatz, der von einer in der Mathematik lange bekannten Wachstumskurve ausgeht, wird in [ABRS83] vorgestellt. Dieser Ansatz hatte sich auch schon früher in anderen Wissenschaften zur Beschreibung von Wachstumsphänomenen als angemessen erwiesen, etwa Biologie, Wirtschaftswissenschaften oder Epidemiologie. Darum soll er ganz kurz beschrieben werden.

Ausgangspunkt ist die Überlegung, daß die Zunahme einer Population proportional zu ihrer Anzahl ist. Sei also diese Anzahl zu einem Zeitpunkt t mit $p(t)$ bezeichnet, so entspricht der Zuwachs der Ableitung p' dieser Funktion. Die Proportionalitätsaussage wird dann durch eine Differentialgleichung gegeben:

$$(*) \quad p' = b p$$

mit einer reellen Zahl b . Deren Lösungen sind bekanntlich durch die Menge aller Funktionen der Form

$$p(t) = a \exp(bt)$$

mit einer weiteren reellen Zahl a bestimmt. Diese Lösung ist natürlich völlig unbefriedigend, da sie

- unintuitiv ist, weil das gegebene p ungebremstes Wachstum beschreibt, und
- mit den in der Linguistik gemessenen Daten nicht übereinstimmt.

Deshalb machen die Autoren sich Modelle zu gebremstem Wachstum zunutze. Die Idee ist, daß das Wachstum nicht nur von der Größe der Bevölkerung sondern auch von den zur Verfügung stehenden Ressourcen abhängt. Diese Ressourcen hängen ihrerseits ebenfalls von der Größe der Bevölkerung ab, aber nicht proportional. Je größer die Bevölkerung im Verhältnis zu den vorhandenen Ressourcen ist, desto geringer wird das Wachstum sein. Für

den Sprachwandel heißt das: Je mehr Menschen schon zu der neuen Form übergewechselt sind, desto weniger können sie noch in Zukunft neu annehmen. Zusammen mit einer in der Mathematik üblichen Normierung geht (*) durch die Einfügung eines neuen Ressourcenfaktors in die sogenannte „logistische Differentialgleichung“ über:

$$(**) \quad p' = b p (1-p)$$

Deren Lösungen sind durch die Menge aller Funktionen der Form

$$p(t) = 1 / (1 + a \exp(-bt))$$

mit wieder einer geeigneten reellen Zahl a bestimmt. Diese Lösungen sind eine Form des schon von Piotrovskij benutzten tangens hyperbolicus. Die logistische Differentialgleichung ist ebenfalls in der Mathematik seit langem bekannt und in den letzten Jahrzehnten zu einem Fetisch der Chaos-Theorie geworden. Darauf soll aber nicht weiter eingegangen werden, da die von mir betrachteten Systeme nicht chaotischer Natur sind.

Durch die Wahl eines weiteren Proportionalitätsfaktors gelingt es Altmann, mit Hilfe der Funktionenmenge

$$p(t) = c / (1 + a \exp(-bt))$$

eine unvollständige Veränderung der Sprache zu beschreiben. Das c ist also der limes, wenn die Zeit gegen unendlich geht, und beschreibt den Wert, bei dem sich das zahlenmäßige Verhältnis von alter Form und Innovation stabilisiert.

Die herausragende Eigenschaft aller dieser Lösungen ist, daß sie nur den irreversiblen Sprachwandel beschreiben. Um mit dem Problem des reversiblen²¹ Sprachwandels fertig werden zu können, nimmt Altmann in [Alt83] an, daß der Proportionalitätsfaktor b in (**) nicht konstant sondern ebenfalls zeitabhängig, also eine Funktion von t , ist. Er setzt ohne nähere Begründung den neuen Faktor $b-Ct$ mit reellen Zahlen b, C ein und erhält die Differentialgleichung

$$(***) \quad p' = (b-Ct) p (1-p)$$

Deren Lösungen sind durch die Menge aller Funktionen der Form

$$p = 1 / (1 + a \exp(-bt + ct^2))$$

²¹ Um sinnvoll über den Begriff des reversiblen Sprachwandels reden zu können, erinnere ich noch einmal daran, daß es sich bei dem im Augenblick geschilderten Ansatz um eine globale Theorie handelt. Wir brauchen uns also keine Gedanken zu machen, was geschieht, wenn einzelne Individuen einen schwankenden Gebrauch von der Innovation machen. Ganz allgemein ist es aber auch in globalen Modellen wünschenswert, einen möglichen Rückgang des neuen Gebrauchs überhaupt darstellen zu können.

Im übrigen ist der Terminus „reversibel“ nicht ganz unproblematisch: Zwar heißt die Ausbreitung eines Lehnwortes L nicht zwingend, daß ein „Erbwort“ E verdrängt wurde, aber selbst in diesem Fall gilt: Wenn L wieder verschwindet, muß es nicht unbedingt wieder durch E verdrängt werden. In der Regel wird E eher durch ein anderes „Erbwort“ E' , durch ein anderes Lehnwort L' verdrängt werden oder mit der bezeichneten Realie untergehen. Uns interessiert hier nicht, welcher dieser Fälle nun genau auftritt; wir haben es nur mit dem Gegensatz zwischen L und „nicht L “ zu tun. In Übereinstimmung mit der Bezeichnung in [Alt83] benutze ich, wenn immer ich von einem Verlauf rede, bei dem L wieder verschwindet, den Terminus „reversibel“.

mit einer reellen Zahl a und $c = C/2$ bestimmt. Der Rest der Arbeit in [Alt83] ist der statistischen Bestimmung der Konstanten a, b, c und deren Verifikation anhand bekannter Zahlen in der Linguistik gewidmet.

Hier sieht man aber auch die Crux des gesamten Ansatzes überdeutlich: Die Wahl der mathematischen Form des Faktors $b-Ct$ in (***) ist willkürlich. Auch die Konstanten sind rein mathematisch motiviert. Eine linguistisch motivierte Erklärung für deren Werte konnte damals nicht gegeben werden; und eine solche wird es sicher in absehbarer Zeit nicht geben.

Ein ebenfalls globales, auf klassischer analytischer Mathematik basierendes Modell für die Entstehung diatopischer Varianten stellt Altmann in [Alt85a] vor.

In [Leo98] wird die Ausbreitung einer sprachlichen Neuerung als stochastischer Prozeß gedeutet. Das Piotrowskij-Gesetz ist dann der Spezialfall eines seit langem bekannten stochastischen Prozesses, des *Watson-Prozesses* (mit Sättigungsglied).

Ich möchte noch einmal darauf aufmerksam machen, daß es sich bei dem beschriebenen Ansatz um rein globale Untersuchungen handelt. Das lokale Verhalten der Individuen spielt keine Rolle.

6 Sprachliche Neuerungen im Modell des zellularen Graphen

6.1 Wie schnell kann sich eine sprachliche Änderung ausbreiten?

Zunächst soll jetzt begründet werden, warum die im letzten Abschnitt vorgestellten Ergebnisse von Altmann auch die Konsequenz haben, daß zellulare Automaten mit isotroper Gittergeometrie aus Gründen des Funktionenwachstums nicht in der Lage sind, die Übernahme eines Lehnwortes, und daher ex forteriori die Übernahme von sprachlichen Veränderungen ganz allgemein, zu modellieren. Auf den Zusammenhang zwischen der Modellierung von Wachstum in zellularen Automaten, durch Differentialgleichungen (wie bei Piotrovskij und Altmann) und in den in diesem Kapitel eingeführten zellularen Graphen bin ich schon an anderer Stelle, [Sche06], ausführlicher eingegangen und werde meine Leser daher mit allzuviel Mathematik verschonen.

Ich will hier nur durch eine kurze Bemerkung einem Mißverständnis vorbeugen: Ich untersuche hier die Ausbreitung von etwas, aber ich werde im folgenden immer den Terminus „Wachstum“ verwenden. Im Kontext der zellularen Modelle sind diese beiden Begriffe synonym. Mit der Ausbreitung einer sprachlichen Veränderung geht das Wachstum ihrer Verwendung Hand in Hand.

Wir haben schon gesehen, daß sich im Modell eines isotropen zellularen Automaten eine Störung höchstens mit polynomieller Geschwindigkeit fortpflanzt, daß also die Anzahl $A(n)$ der gestörten Zellen in Abhängigkeit vom Zeitpunkt n durch die Funktion A mit polynomielltem Wachstum beschrieben wird. Die Ergebnisse von Altmann zeigen aber, daß zumindest in der Anfangsphase die Ausbreitung der Störung durch die Exponentialfunktion exp annähernd beschrieben werden kann. Es ist aber eine in der theoretischen Informatik wohlbekannte (und auch sonst sehr wichtige) Tatsache, daß für alle Polynome p gilt

$$exp \notin o(p)$$

Die Exponentialfunktion hat also kein polynomielles Wachstum. Insbesondere gilt in den Formeln aus Kapitel 4 für exp und für alle Polynome p , daß $exp \in \Omega(p)$ ist. Anschaulich bedeutet dies, daß die Störung sich in den bisherigen Modellen auf der Basis zellulärer Automaten einfach zu langsam fortpflanzt.

Wir sind also im Hinblick auf die eben geschilderten Überlegungen gezwungen, unser Modell etwas, aber nur wenig komplexer zu machen und werden daher künftig die Forderung der Isotropie fallenlassen und davon ausgehen, daß die Gittergeometrie ein allgemeiner Graph ist. Insbesondere muß eine Forderung aus dem Anfang von Kapitel 3 fallen gelassen werden, nämlich, daß es eine feste, für alle Zellen gültige Form der (lokalen) Nachbarschaft gibt. Solche zellularen Automaten werden auch als *zellulare Graphen* oder *Netzwerke* bezeichnet. Für künftige Untersuchungen mit Hilfe von zellularen Graphen haben wir jetzt zwei Möglichkeiten bezüglich der Gittergeometrie:

- die Benutzung bekannter besonders einfacher Graphen mit bekannten auf analytische Weise mathematisch beschreibbaren Eigenschaften,

- die Benutzung komplexerer Graphen mit gut untersuchten statistischen Eigenschaften. Solche Graphen wurden in den letzten Jahren für den Fall zufällig erzeugter Graphen ausgiebig studiert. Die Forschung steckt aber noch in den Kinderschuhen.

Zunächst soll daher ein wenig Graphentheorie vorgestellt werden.

6.2 Elemente der Graphentheorie

Die Graphentheorie, mit der solche Gittergeometrien behandelt werden, geht zurück auf das Jahr 1736, als der Schweizer Mathematiker *Leonhardt Euler* das „Problem der sieben Brücken von Königsberg“ löste, und den ersten Schritt hin zu einer Theorie tun sollte, die sich heute als eine der anwendungsbezogensten mathematischen Theorien überhaupt erweist. Da hier wenigstens ein traditionelles Beispiel eines Graphen vorgeführt werden soll, tue ich das mit dem klassischen Graphen. Herr Professor Euler hatte auf seinen Spaziergängen durch die Stadt Königsberg das folgende Problem:

Durch die Stadt floß der Pregel und zerlegte diese in Stadtteile, die damals durch sieben Brücken verbunden wurden. Die ungefähre Form des Flusses Pregel, seiner Inseln, der Stadtteile von Königsberg und der vorhandenen Brücken zeigt Abbildung 12.

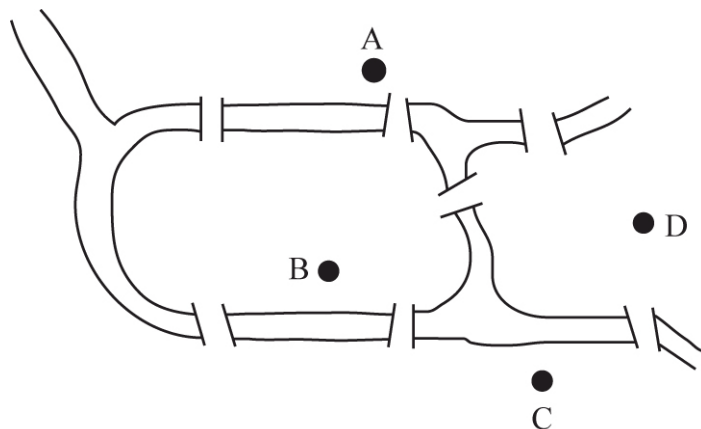


Abbildung 12: Die Stadt Königsberg und ihre Brücken

Herr Euler wollte für seine Spaziergänge einen Weg finden, auf dem er jede Brücke genau einmal überqueren und dann wieder daheim ankommen konnte. Für seine grundlegenden Untersuchungen sah Euler von den zufälligen geographischen Gegebenheiten ab. Das abstrahierte Wegenetz hatte dann diese Form:

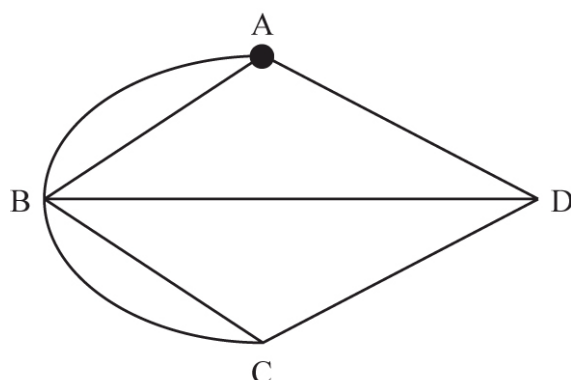


Abbildung 13: Die Brücken von Königsberg als Graph

Ich will hier nicht weiter auf Professor Eulers Problem eingehen, einen Weg von seiner Wohnung in A wieder zurück nach A zu finden, bei dem jede Kante genau einmal durchlaufen wird. Das würde uns zu weit vom Thema wegführen. Insbesondere die Lösung kann jedem Graphentheoriebuch entnommen werden. Nur soviel sei verraten: Die Wünsche unseres Spaziergängers sind unerfüllbar. Statt einer theoretischen Erörterung soll zunächst der Begriff des Graphen formal definiert werden:

Definition: Ein *Graph* ist ein Paar (V,E) aus einer Menge V , deren Elemente *Knoten* (engl. vertices) genannt werden, und einer Menge E von Paaren von Knoten, *Kanten* (engl. edges) genannt²².

Anschaulich stellt man sich unter Kanten Verbindungen zwischen Knoten vor. Die erste Komponente einer Kante ist der Ausgangs-, die zweite Komponente der Zielknoten einer solchen Verbindung. Wir werden künftig eine weitere Eigenschaft unserer Graphen voraussetzen, nämlich die *Symmetrie*. Mathematisch bedeutet das, daß mit jeder Kante $(a,b) \in E$ auch $(b,a) \in E$ sein muß. Es ist also keine Kantenrichtung bevorzugt. Bei allgemeinen Graphen müssen Verbindungen nicht symmetrisch sein. Unsere Vereinbarung spiegelt die Tatsache wider, daß bei sprachlicher Kommunikation im Prinzip keine Richtung bevorzugt ist²³.

Ich führe für spätere Verwendung noch einige graphentheoretische Begriffe ein:

Definition: Sei $G = (V,E)$ ein Graph.

- Ein *Weg* oder *Pfad* durch den Graph ist eine Folge von Kanten $(a_1,b_1), (a_2,b_2), \dots, (a_n,b_n)$ mit der Eigenschaft, daß immer $b_i = a_{i+1}$ für alle i zwischen 1 und $n-1$ gilt.

²² Damit bilden die Kanten eine mathematische Struktur, eine *Relation* auf V . Im allgemeinen ist eine Relation R auf einer Menge M nichts weiter als eine Teilmenge des cartesischen Produktes, also $R \subseteq M \times M$. Bei uns ist diese Relation also durch $E \subseteq V \times V$ gegeben. Relationen kommen fast überall vor. Sie sind in der Mathematik gut untersucht und haben trotz ihrer Einfachheit interessante Eigenschaften.

²³ Auch im Umfeld unserer Untersuchungen gibt es Graphen, bei denen E nicht symmetrisch ist. Das sind dann aber keine linguistischen Anwendungen und daher hier von geringem Interesse. Darüber kann in [DM03] mehr gelesen werden.

- Ein *geodäsischer* Pfad zwischen zwei Knoten a und b ist ein Pfad minimaler Länge zwischen a und b .
- Ein *Kreis* ist ein Weg, bei dem außerdem Anfangs- und Endpunkt übereinstimmen, bei dem also zusätzlich $a_1 = b_n$ gilt.
- Ein Graph heißt *zusammenhängend*, wenn es zwischen je zwei Knoten mindestens einen Weg gibt.
- Der *Durchmesser* von G ist die Länge des längsten geodäsischen Pfades in G , wenn G zusammenhängend ist. Sonst ist der Durchmesser unendlich.
- Ein *Baum* ist ein zusammenhängender, kreisfreier Graph.

Beispiele: Die folgenden Graphentypen sind die einfachsten. Sie gelten als die Musterbeispiele für Graphen.

Eine *Linie* ist ein Graph, der aus einem einzigen Weg besteht.



Abbildung 14: Linie als Graph

Ein *Vollständiger Graph* V_n (der Größe n) ist der eindeutige Graph mit n Knoten, bei dem jeder Knoten mit jedem anderen Knoten, also ausschließlich sich selber, genau einmal durch eine Kante verbunden ist.

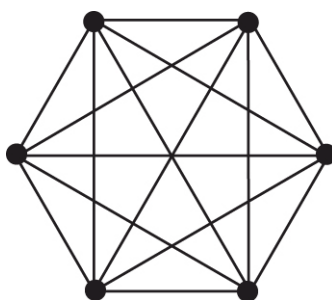


Abbildung 15: Der Vollständige Graph V_6 der Größe 6

Ein *Binärer Baum* ist ein Baum, bei dem die Wurzel, das ist beispielsweise in Abbildung 16 der oberste Knoten, und auch alle anderen Knoten höchstens zwei Nachfolger haben. Bei einem *Vollständigen Binären Baum* sind es genau zwei (im Falle eines endlichen Baumes bis zu einer gewissen Tiefe). Es ist im Prinzip also eine Linie auch ein binärer Baum,

aber kein vollständiger Binäre Bäume treten in vielen Anwendungen auf, zum Beispiel bei Suchverfahren²⁴, und sind dementsprechend gründlich untersucht worden.

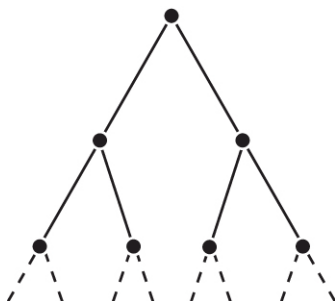


Abbildung 16: Der Vollständige Binäre Baum

6.3 Netze endlicher Automaten

Aus der Sicht der theoretischen Informatik haben wir jetzt eine Kombination zweier vieluntersuchter Formalismen vor uns: endliche Automaten²⁵, die in Form eines Graphen angeordnet sind.

Was Graphen sind, habe ich ja schon erläutert. Für endliche Automaten gilt die

Definition: Ein (*deterministischer*) *endlicher Automat* ist ein 5-Tupel $A = (X, Z, \delta, z_0, F)$ mit

- einer endlichen Menge X von *Eingaben*,
- einer endlichen Menge Z von *Zuständen*,
- einer *Zustandsüberföhrungsfunktion* $\delta: Z \times X \rightarrow Z$,
- einem *Anfangszustand* $z_0 \in Z$,
- einer endlichen Menge $F \subseteq Z$ von *Endzuständen*.

Der Begriff des endlichen Automaten ist zentral in der theoretischen Informatik. Er bildet einen Kern in jeder einföhrenden Theorievorlesung.

Was haben nun die fröher betrachteten zellularen Automaten mit den jetzt neu eingeföhrten Begriffen zu tun? Die Antwort ist einfach: Bei uns wird ein zellulärer Automat als Netz endlicher Automaten betrachtet, wobei die Eingaben der einzelnen endlichen

²⁴ Wie geht eine schnelle Suche vor sich? Man halbiere den Suchraum. Zur Suche in den Hälften werden diese wieder halbiert, usw., und schon ist ein (endlicher) binärer Baum entstanden.

²⁵ Ich möchte noch einmal explizit darauf hinweisen, daß es sich bei „endlichen Automaten“ und „zellularen Automaten“ um zwei verschiedene Konzepte mit ähnlichen Namen handelt.

Automaten spezieller Natur sind. Da die Knoten im Netz ihren Zustand in Abhängigkeit von den Zuständen ihrer Nachbarn ändern, kann das aus Automaten-sicht auch so interpretiert werden, als bestehen diese Eingaben nämlich aus den Zuständen der benachbarten Automaten, also genauer aus Vektoren von solchen Zuständen. Die anderen Komponenten von endlichen Automaten sind entweder schon definiert, so die Zustandsüberföhrungsfunktion, oder in unserem Fall irrelevant, wie die Anfangs- und Endzustände der Automaten. Aus der Sicht der Netze bedeutet das:

- Die Knoten werden von den endlichen Automaten gebildet.
- Zwischen zwei Knoten existiert eine Kante, wenn in der Überföhrungsfunktion des einen auf den Zustand des anderen Bezug genommen wird.

Beispiel: Die Interpretation des zweidimensionalen Gitters mit von-Neumann- und mit Moore-Umgebung als Graph ist in Abbildung 17 zu sehen. Jeder einzelne Punkt wird als endlicher Automat betrachtet. Für jeden Punkt ist in beiden Sichtweisen ein interner Zustand gegeben. Soll dann etwa die Überföhrungsfunktion jeweils des mittleren endlichen Automaten (Knotens) in den beiden Fällen berechnet werden, so werden zunächst die Nachbarn bestimmt. Im linken Bild geschieht dies durch die waagerechten und die senkrechten Nachbarn (von-Neumann-Umgebung), rechts kommt noch die Beeinflussung durch die diagonalen Nachbarn hinzu (Moore-Umgebung). Die internen Zustände der in der Abbildung markierten Nachbarn bestimmen den Wert der Überföhrungsfunktion, also den neuen Zustand des betrachteten Knotens. Genau derselbe Effekt wäre gegeben, wenn man die Zustände der Nachbarn direkt als Eingabe für den endlichen Automaten angesehen hätte.

Durch diese Sicht, die zu der in dieser Arbeit bisher mehrheitlich eingenommenen traditionellen Sicht äquivalent ist, lassen sich auch Erkenntnisse über endliche Automaten zu Untersuchungen über dynamische Netzwerke heranziehen.

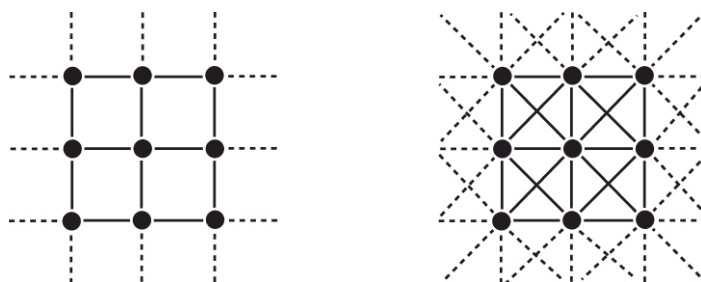


Abbildung 17: Reguläre Gitter als Graphen

6.4 Der Vollständige Binäre Baum als Gittergeometrie

Wir sind, zumindest aus der Sicht der theoretischen Informatik, auf relativ natürliche²⁶ Weise zu einer Situation gelangt, in der es sinnvoll ist, wenn wir als Gittergeometrie gewöhnliche Graphen annehmen. Wieder sollen die oben erwähnten Metaregeln gelten. Wir wenden uns zunächst den oben genannten Graphen zu. Später werden wir auch, damit über den rein mathematischen Sinn hinausgehend, linguistische Reflektionen einfließen lassen, wobei wir aber die mathematische Handhabbarkeit nie aus den Augen verlieren werden. Zunächst zu den obigen Beispielen:

Die Linie ist in einem mathematisch genau faßbaren Sinne äquivalent zu dem eindimensionalen Gitter. Von dem haben wir schon eingesehen, daß es aus Komplexitätsgründen für unsere Zwecke nicht brauchbar ist.

Wie ist die Lage nun beim Vollständigen Graphen? Entweder tritt nach endlich vielen Schritten überhaupt keine Zustandsänderung auf oder ein periodischer Wechsel mit relativ kurzer Periode. Daß das Verhalten periodisch ist, liegt einfach daran, daß das gesamte System nur endlich viele Zustände annehmen kann. Bei beliebig langem Fortschritt muß dann irgendwann ein früher schon angenommener Zustand erneut angenommen werden. Da unsere Regeln strikt deterministisch sind, wird sich fortan das Netz immer wieder identisch fortentwickeln. Die Periode ist auch sehr kurz (im Extremfall, dem Fall ohne Zustandsänderung, eben gleich 1). Das sieht man folgendermaßen: Wir gehen ja davon aus, daß, solange keine Störung auftritt, alle Knoten den gleichen Zustand haben („0“ oder „nicht infiziert“). Auch nach der Infektion ist die Situation in allen Knoten (außer dem Infizierten) identisch, insbesondere auch, was die Situation in den Nachbarknoten und damit die Werte der Überföhrungsfunktion angeht. Damit verhalten sich alle diese Knoten auch im weiteren Verlauf identisch. Folglich verhält sich das Netz für die hierzu untersuchende Frage so, als ob es überhaupt nur zwei Knoten hätte. Genauer: Sei die Mächtigkeit der Zustandsmenge eines Knotens gleich n , so kann das Netz als Ganzes höchstens n^2 Zustände annehmen. Damit ist auch die Periode durch n^2 begrenzt, einen sehr kleinen Wert.

Eine echte Periode (eine, die größer als 1 ist,) ist aber nur bei Zustandsmengen mit mindestens drei Elementen möglich. Ich will das an dieser Stelle aber nicht beweisen.

Beispiel: Ein einfaches Beispiel dafür, wie ein Verhalten mit echter Periode, in diesem Falle 2, auftreten kann, wird durch Regeln mit den Eigenschaften

$$f(x,v) = 0 \quad \text{für } v=0 \text{ (Das heißt } v \text{ ist der Nullvektor.)},$$

$$f(x,v) = 2 \quad \text{falls alle Komponenten in } v \text{ größer als } 0 \text{ sind,}$$

$$f(x,v) = d(v) \quad \text{sonst (Also enthält } v \text{ eine Mischung aus Nullen und anderen Werten.)},$$

und mit der Zustandsmenge $\{0,1,2\}$ hervorgerufen. Dabei sollen x für den Zustand der zu verändernden Zelle stehen und v für den Zustandsvektor der benachbarten Zellen. In der letzten Zeile bedeute $d(v)$ den ganzzahligen Anteil des Durchschnittes der Zustände. In einem Vollständigen Graphen der Mächtigkeit 3 alterniert der Automat lediglich zwischen den beiden Zustandsvektoren $(2,0,0)$ und $(0,1,1)$. Dieses Verhalten ist in Abbildung 18 skizziert.

²⁶ Immerhin werden nur Standardkonzepte verwendet und keine ausgefallenen Konstrukte.

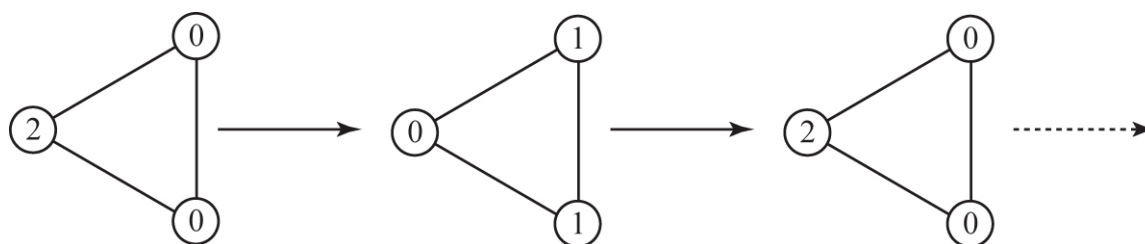


Abbildung 18: Alternierende Zustände

Die mathematische Erklärung ist wie folgt:

- erster Übergang: Für den infizierten Knoten ist die erste Regel entscheidend. Alle seine Nachbarn sind uninfiziert, und deshalb geht der Knoten in den Zustand 0 über. Die Nachbarschaft aller anderen Knoten enthält sowohl Nullen als auch andere Werte. Deswegen ist ihr Nachfolgezustand nach der dritten Regel $d(v)$. Das berechnet sich zu 1.
- zweiter Übergang: Für den ursprünglich infizierten Knoten ist die zweite Regel entscheidend. Alle seine Nachbarn sind im Zustand 1, und deshalb geht der Knoten wieder in den Zustand 2 über. Die Nachbarschaft aller anderen Knoten enthält erneut sowohl Nullen als auch andere Werte. Deswegen ist ihr Nachfolgezustand nach der dritten Regel $d(v)$. Das berechnet sich aber diesmal zu 0, und der ursprüngliche Zustand ist wieder hergestellt.

Wegen des strikten Determinismus der Regeln ist damit die gesamte Zukunft des Netzes festgelegt.

Ein solches Verhalten, bei dem sich die Änderung schlagartig überallhin ausbreitet und wild hin- und heralterniert, widerspricht den Befunden von Altmann (ebenso jeglicher Erfahrung und wäre auch unerklärlich).

Diese letzten Überlegungen bestätigen meine früheren Befunde, daß nämlich die Ausbreitungsgeschwindigkeit einer Veränderung von der Topologie des Netzes abhängt.

Beispiel: Als Kompromiß zwischen den nur lokal verbundenen Gittern in den isotropen Automaten mit ihrer nur linearen Ausbreitungsgeschwindigkeit und den sehr eng gekoppelten Vollständigen Graphen, in denen sich eine Änderung schlagartig ausbreitet, wollen wir nun die Lage in einem Vollständigen Binären Baum untersuchen.

Die Zustandsmenge sei zweielementig, die Veränderung beginne ohne Beschränkung der Allgemeinheit an der Wurzel. Damit diese Veränderung sich überhaupt ausbreiten kann, muß die Regel die sein, daß eine unmarkierte Zelle zu einer markierten wird, sobald mindestens einer ihrer Nachbarn markiert ist. Was mit einer bereits markierten Zelle geschieht, ist unbedeutend, wie der folgende Satz zeigt:

Satz: Sei unter den gegebenen Umständen $A(n)$ die Anzahl der markierten Zellen zum Zeitpunkt n . Dann gilt

$$2^n \leq A(n) \leq 2^{n+1} - 1$$

Der Beweis ist eine einfache vollständige Induktion, die sich folgende Ideen zunutze macht aber hier nicht in den Einzelheiten ausgeführt werden soll: Im n -ten Schritt sind

zumindest alle die Zellen markiert, die *genau* die Entfernung n von der Wurzel haben; davon gibt es genau 2^n Stück. Die Zahl aller Zellen, die zum Zeitpunkt n überhaupt jemals markiert waren, ist dadurch beschränkt, daß solche Zellen *höchstens* die Entfernung n von der Wurzel haben können; davon gibt es genau $2^{n+1}-1$ Stück.

Qed

Der Beweis soll an Hand von Abbildung 19 illustriert werden: Dort sind die Schritte 0, 1, 2 zu sehen. In jedem dieser Schritte sind zumindest die mit dicken schwarzen Kreisen bezeichneten Knoten in den Schichten 0, 1, 2 markiert. Im Schritt n , also in der Schicht n , sind das genau 2^n Stück. Die mit kleinen schwarzen Kreisen bezeichneten Knoten in höheren Schichten sind unmarkiert; sie spielen also keine Rolle. Über die mit kleinen roten Kreisen bezeichneten Knoten in niedrigeren Schichten ist im allgemeinen nichts bekannt. Von diesen gibt es in Schritt n höchstens 2^n-1 Stück. Daraus folgt die Anzahlabschätzung des Satzes.

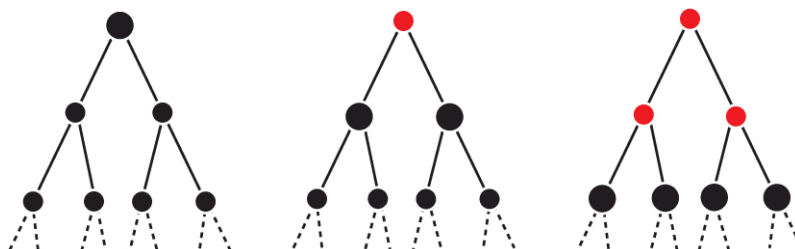


Abbildung 19: Ausbreitung im Vollständigen Binären Baum

Die Bedeutung dieses Satzes liegt in unserem Zusammenhang darin, daß er ein erstes, einfaches Beispiel für exponentielles Wachstum²⁷ der Zahl markierter Zellen vorstellt.

6.5 Fraktale Gittergeometrien

Der binäre Baum hat eine Eigenschaft, die seit den achtziger Jahren des zwanzigsten Jahrhunderts, zu einem Forschungsobjekt in vielen Wissenschaften und zum Gegenstand auch vieler populärwissenschaftlicher Darstellungen geworden ist: Er ist fraktal, [Man67], [PR86], [Fed89]. Ein *Fraktal*²⁸ bezeichnet natürliche oder künstliche Gebilde oder geometrische Muster, die einen hohen Grad von Skaleninvarianz bzw. Selbstähnlichkeit²⁹ aufweisen. Das ist beispielsweise der Fall, wenn ein Objekt aus mehreren verkleinerten Kopien seiner selbst

²⁷ Eine Funktion f hat *exponentielles Wachstum*, wenn gilt $f \in \Omega(d^n)$ für eine geeignete Konstante d . Das namensgebende Beispiel liefert die oben angesprochene Funktion *exp*.

²⁸ lat. fractus: gebrochen, von frangere: brechen

²⁹ Das bedeutet, daß bei einem einfachen Hingucken nicht zu entscheiden ist, welche Größenskala gerade gilt. Ein Beispiel bilden etwa Börsenkurse. Schaut man nur auf eine Graphik, die das typische Auf und Ab wiedergibt, kann man nicht entscheiden, ob man gerade einen Tages-, einen Wochen-, oder einen Jahresverlauf betrachtet. Im Unterschied zu vollständigen Binärbäumen sind allerdings die Teilstrukturen nicht mit der ganzen Struktur bis auf die Größe identisch. Die Verläufe ähneln einander nur. Deshalb wird hier nicht von Skaleninvarianz sondern von Selbstähnlichkeit gesprochen.

besteht. Geometrische Objekte dieser Art unterscheiden sich in wesentlichen Aspekten von gewöhnlichen glatten Figuren. Fraktale werden bei der computergestützten Simulation formenreicher Strukturen wie beispielsweise realitätsnaher Landschaften eingesetzt. Ein Beispiel, wo das Auftreten fraktaler Strukturen in der Sprachwissenschaft durch ein Ökonomieprinzip erklärt wird, ist [FS03].

Fraktale Konzepte finden sich auch in der Natur. Dabei ist jedoch die Anzahl der Stufen von selbstähnlichen Strukturen begrenzt, und beträgt oft nur 3-5. Weit verbreitet sind fraktale Strukturen ohne strenge sondern mit statistischer Selbstähnlichkeit. Dazu zählen beispielsweise Bäume, der Blutkreislauf, Flusssysteme und Küstenlinien.

Beispiel: Der binäre Baum $B=(V,E)$ habe die Knoten mit den Koordinaten

$$V = \{ (k/2^n, 1/2^{n-1}) \mid 0 < k < 2^n, k \text{ ungerade}, n \text{ natürliche Zahl} \}$$

Es gebe Kanten, die wie folgt definiert sind:

$$E = \{ ((k/2^n, 1/2^{n-1}), (2k-1)/2^{n+1}, 1/2^n) \} \cup \{ ((k/2^n, 1/2^{n-1}), (2k+1)/2^{n+1}, 1/2^n) \}$$

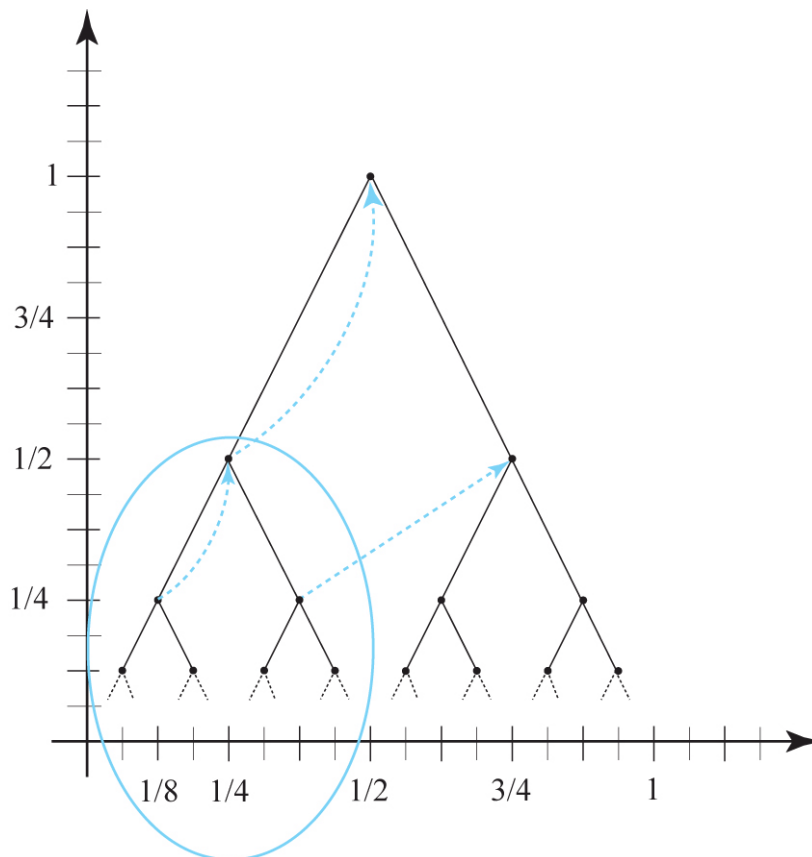


Abbildung 20: Ein Binärer Baum als fraktale Struktur

So ergibt sich etwa die Wurzel für $n=1, k=1$ als $(1/2, 1)$. Sie ist verbunden vermöge der beiden Kanten $((1/2, 1), (1/4, 1/2))$ und $((1/2, 1), (3/4, 1/2))$. Betrachtet man nur den von dem Knoten $(1/4, 1/2)$, mithin für $n=2, k=1$, abhängigen Teilbaum B' von B , also den in

Abbildung 20 durch eine blaue Ellipse markierten Teil, so erhält man den Skalierungsfaktor $\frac{1}{2}$, da B' entsteht, indem die Punkte in B mit $\frac{1}{2}$ in horizontaler und in vertikaler Richtung multipliziert werden. Die in der Abbildung eingezeichneten Pfeile zeigen für ausgewählte Punkte von B' , wie die bijektive Skalierungsabbildung zwischen B' von B etabliert wird.

Es ist leicht einzusehen, daß fraktale Geometrien zu exponentiellem Wachstum führen. Sie haben damit insgesamt mindestens drei Vorteile bei einer Nutzung als Modell für sich:

- Sie bieten die Möglichkeit zu exponentiellem Wachstum einer Veränderung.
- Sie sind mathematisch gut untersucht, ein allgemein akzeptierter Gegenstand der Forschung.
- Sie haben in unserer Perspektive die Logik auf ihrer Seite. Auch in soziologischen Modellen ist die Gesellschaft fraktal gegliedert [Klü95], und damit auch die Gemeinschaft der Sprecher. Die Begründung dafür liefert die hierarchische Gliederung der Gesellschaft in Individuen, Familien, Stadt, Land, Staat usw.



Abbildung 21: Sierpinski-gasket

In Abbildung 21 wird noch einmal ein Musterbeispiel für eine fraktale Struktur angegeben, das *Sierpinski-gasket*³⁰. Dieses entsteht folgendermaßen: Man beginnt mit einem gefärbten gleichseitigen Dreieck (Schritt 0). Dieses wird auf die im linken Bild gezeigte Weise in vier gleich große Dreiecke zerlegt. Das zentrale Dreieck wird wieder entfärbt (Schritt 1). In Schritt i wird wieder aus dem bis dorthin konstruierten Objekt in jedem gleichseitigen Teildreieck das zentrale Dreieck entfernt³¹. Das Sierpinski-gasket ist der Durchschnitt aller so konstruierten Objekte. Es läßt sich zum Beispiel bijektiv mit Hilfe einer Skalierungsabbildung auf diejenigen Teilstrukturen projizieren, die aus den drei gefärbten gleichseitigen Dreiecken nach Schritt 1 entstehen.

Ein entscheidender Nachteil für uns (und in der Praxis ganz allgemein) ist jedoch die Tatsache, daß die Anzahl der Stufen von selbstähnlichen Strukturen in der Realität und insbesondere im Kommunikationsverhalten begrenzt ist auf die oben genannten 3-5 Ebenen.

³⁰ Die wörtliche Bedeutung von gasket ist „Dichtung“ (nicht im Sinne von „Poesie“ sondern im Sinne von „Klempner“). Angesichts dessen habe ich die englische Bezeichnung beibehalten.

³¹ Es handelt sich also auch hier um eine Konstruktion durch vollständige Induktion.

Das bedeutet, daß Betrachtungen über Fraktale in unserer Arbeit keine wesentliche Rolle spielen werden. Eine Ausnahme bilden die lokal baumähnlichen Strukturen, mit deren Hilfe ich in Unterkapitel 8.6 zeigen kann, wie sich das Auftreten von anfänglich exponentiellem Wachstum (wie im Piotrowskij-Gesetz) in geeigneten Netzwerken erklären läßt.

Es dürften sich, gerade weil Fraktale in der Mathematik gut untersucht sind, hier erfolgversprechende Ansätze zu weiteren Forschungen ergeben.

7 Komplexe Netzwerke

Wie wir am Anfang von Kapitel 6 gesehen haben, führt die Konstruktion allzu einfacher auf Geometrie beruhender Modelle zu Unzulänglichkeiten. Auch die traditionelle Graphentheorie, wie sie bis in die neunziger Jahre des zwanzigsten Jahrhunderts fast allgemein üblich war, würde sehr schnell an ihre Grenzen stoßen. In den letzten Jahren hat es jedoch einen erheblichen Schub in der Untersuchung von Graphen gegeben. Es werden nicht mehr wie bis dahin kleine Graphen vorwiegend mit algebraischen Methoden untersucht und die Eigenschaften einzelner Knoten und Kanten bestimmt, sondern es werden große Graphen und deren Eigenschaften mit Hilfe von statistischen Methoden und von Simulationen erforscht. Solche Graphen werden im Rahmen der neuen, zunächst maßgeblich von theoretischen Physikern vorangetriebenen Forschungen meistens als „Netzwerke“ in anderen Wissenschaften auch kürzer als „Netze“ bezeichnet.

Wurden etwa früher Dinge gefragt wie: „Welcher Knoten in dem gegebenen Graphen ist der wichtigste für den Zusammenhang des Graphen?“, so verlagert sich diese individuelle Fragestellung hin zu der wesentlich anwendungsbezogeneren statistischen Frage: „Welcher Anteil von Knoten mit welchen Eigenschaften muß aus dem gegebenen Netzwerk entfernt werden, damit der Zusammenhang gefährdet ist?“ Um die praktische Relevanz solcher Untersuchungen noch etwas zu belegen, möchte ich schon an dieser Stelle ein ganz herausragendes Objekt dafür nennen: das Internet. Es ist das wohl größte menschengemachte Netzwerk; die Knoten sind Rechner, die Kanten sind Verbindungen zwischen ihnen. Für eine Kommunikation zwischen zwei Punkten ist der Zusammenhang des Netzwerks offensichtlich eine notwendige Bedingung. Die eben gestellte Frage läuft also im Grunde darauf hinaus zu ergründen, der Ausfall welcher (wichtiger) Knoten (mit welchen Eigenschaften) die Kommunikation im Weltnetz beeinträchtigt.

Schon aus den bisherigen Bemerkungen wird die Wichtigkeit computergestützter Untersuchungen deutlich. Große Datenmengen lassen sich eben nur in den seltensten Fällen mit analytischen Methoden untersuchen. In [New03] wird ein weiterer Punkt gemacht: In der klassischen Graphentheorie sind die hauptsächlichsten Methoden der Untersuchung neben der Mathematik die Zeichnung und das Auge. Aber jeder, der einmal versucht hat, einen (gar nicht einmal übermäßig komplexen) Graphen so zu zeichnen, daß man ihm seine Eigenschaften ansieht, erkennt sehr schnell, wie schwierig dies Vorhaben ist. Die Schwierigkeit dabei ist in erster Linie nicht die, große Graphen irgendwie „aufs Papier zu bringen“. Da hat es in den letzten Jahren einige erstaunlichen Fortschritte gegeben. Theoretische Grundlagen werden in [KW01], [JM04] geschildert. Ebenso gibt es einige leistungsfähige mathematische Werkzeuge, etwa das von S. Wolfram entwickelte „Mathematica“ oder den „H3Viewer“ [Mun01]. Eine Weiterentwicklung des H3Viewers ist die besonders von Kilian programmierte Werkzeugfamilie Viator [HTK+02]. Zum Beispiel wurde Abbildung 39 mit Hilfe eines Programmes aus Viator erstellt. Haben auch alle Werkzeuge Probleme mit sehr großen Graphen, so ist der Punkt in unserem Zusammenhang doch eher der: Das menschliche Gehirn ist nicht in der Lage, riesige Graphen zu analysieren. Statistik soll also die Analyse durch das Auge und das menschliche Gehirn ersetzen; in den Worten von [New03]:

„Wie kann ich sagen, wie ein Netzwerk aussieht, wenn ich es nicht überschauen kann?“

Die Größe der Netze zwingt also dazu, von den exakten Methoden der diskreten Mathematik zu Statistiken über Netzwerke überzugehen. Was ist die Aufgabe einer solchen Netzwerkstatistik? Unabhängig von der in dieser Arbeit gestellten Aufgabe sollte sie

- die wesentlichen Eigenschaften des Netzwerkes beschreiben,

Dadurch muß es möglich sein, von der speziellen Struktur des Netzwerkes zu abstrahieren und nur mit den statistischen Eigenschaften zu arbeiten.

- die Bildung interessanter Klassen von Netzwerken fördern,

Hier geht es also darum, bei der Untersuchung einer konkret gegebenen Menge von Netzwerken statistische Invarianten zu finden, die allen diesen konkreten Netzwerken gemeinsam sind aber möglichst wenigen Netzen außer der gerade vorliegenden Menge.

- nützlich für algorithmische Behandlung der Netze sein,

Hierhin gehören zwei Aspekte:

- Wie findet man ein Verfahren um Netzwerke zu finden und zu konstruieren, die den Anforderungen genügen?
- Wie kann man das statistische Wissen ausnutzen, um Berechnungen auf dem Netz durchzuführen?

- brauchbare Anwendungen finden.

Dazu werden unten mehrere Beispiele folgen.

Aus einer eher philosophischen Perspektive läßt sich die Untersuchung komplexer Netzwerke auch als ein weiterer Schritt über reduktionistische Ansätze hinaus deuten, hin zu einem Verständnis komplexer Systeme als Ganzes.

7.1 Beispiele für komplexe Netzwerke

Komplexe Netzwerke führen zwar zu interessanten mathematischen Fragestellungen; sie sollten aber keinesfalls nur für ein mathematisches Spielzeug gehalten werden. Außer zu mathematischen Erkenntnissen haben sie auch zu einer angesichts der kurzen Zeit der Untersuchungen ungeheuren Fülle von Anwendungen geführt:

7.1.1 Soziale Netzwerke

Die Tatsache, daß hier viele Individuen vielfältige Beziehungen miteinander eingehen, lassen soziale Kontakte zu einem offensichtlich geeigneten Untersuchungsgegenstand mit den genannten Methoden werden. [Rap57] ist ein erster früher Vorläufer. In den letzten Jahren ist es zu Untersuchungen von wirtschaftlichen Hintergründen bis hin zu Sexualkontakten gekommen. Wenn es sich nicht um eine derartig kleine Stichprobengröße handelt, daß bei einer graphentheoretischen Behandlung überhaupt traditionelle Methoden ausreichen, wobei aber gerade die Kleinheit der Untersuchung dann auch zu Unsicherheiten führt, so werden in den wissenschaftlichen Experimenten keine tatsächlichen Netzwerke rekonstruiert, sondern man versucht, aus idealtypischen Modellen allgemeingültige Schlüsse zu ziehen.

Eins der wenigen in der Philologie vorkommenden Beispiele für die Untersuchung eines Netzwerkes mit (traditioneller) Graphentheorie ist die Untersuchung des Beziehungsnetzwerkes der Protagonisten in „Anna Karenina“ in [Knu93].

7.1.2 Informationsnetzwerke

Der Urtyp eines solchen Netzwerkes in der wissenschaftlichen Literatur spiegelt den Informationsfluß durch die Zitierung wissenschaftlicher Arbeiten wider. Ein früher Vorläufer von in diese Richtung gehenden Untersuchungen ist die Entdeckung des sogenannten „Gesetzes der wissenschaftlichen Produktivität“ durch A.Lotka schon im Jahre 1926, [Lot26]. Dieses Gesetz werde ich weiter unten noch etwas präzisieren.

Die zur Zeit bedeutendsten Informationsnetzwerke sind das *world wide web* und seine Teilnetze.

7.1.3 Technische Netzwerke

Hierzu zählen beispielsweise Strom- und Telefonnetze, aber auch das *Internet*, das in unserem Zusammenhang nicht mit dem *world wide web* verwechselt werden sollte: Während die Knoten des Internets, als Graph betrachtet, von den miteinander physisch verbundenen Rechnern gebildet werden, besteht das *world wide web* aus den virtuell verbundenen, „verlinkten“, Dokumenten auf diesen Rechnern. Dadurch daß es sich beim Internet um physische Verbindungen handelt, spielen die geographischen Entfernungen für die Errichtung von Kanten eine größere Rolle als beim *world wide web*. Es gibt folglich im Internet tendenziell weniger Knoten mit einer hohen Zahl von Verbindungen. Insgesamt gelten daher auch leicht unterschiedliche Gesetze in den beiden Netzen.

7.1.4 Biologische Netzwerke

Als Netzwerke untersucht wurden etwa der Stoffwechsel, Ernährungsnetze und die Steuerung der Gene. Von größter Bedeutung ist die Erforschung der „*neuronalen Netze*“, die die Tätigkeit von Gehirnzellen nachbilden sollen. Da der Fokus dieses Wissenschaftszweiges aber ein anderer ist als der uns interessierende, möchte ich auf neuronale Netze nicht weiter eingehen.

Zur Illustration der weitgefächerten Bedeutung komplexer Graphen sollen diese im Vergleich zur allgemeinen Forschungslage wenigen Beispiele reichen.

7.2 Gradverteilungen

In diesem und dem nächsten Unterkapitel möchte ich einige der gebräuchlichsten Charakteristika von Netzwerken vorstellen. Da die Gradverteilung beim gegenwärtigen Stand der Forschung die bei weitem wichtigste Kenngröße ist, sollen ihr 7.2 und 7.3 als eigene Unterkapitel gewidmet sein.

7.2.1 Mathematische Verteilungen

Bevor ich jedoch genauer auf diese Eigenschaften von Netzwerken eingehen kann, müssen noch einige ganz allgemeine mathematische Begriffe eingeführt werden:

Definition:

- Der *Grad eines Knotens* ist die Anzahl der Kanten, die mit diesem Knoten verbunden sind. Man beachte, daß dieses nicht gleich der Zahl der benachbarten Knoten ist, da es zwischen zwei Knoten mehrere Kanten geben kann.
- Die *Komponente* eines Graphen G , zu der ein Knoten k gehört, ist die Menge derjenigen Knoten, die von k aus durch einen Pfad in G erreicht werden können³².

Wenn eine Menge S von Beobachtungen statistisch ausgewertet werden soll, bedarf es einer Funktion, im allgemeinen mit P bezeichnet, die gewissen Teilmengen M von S einen numerischen Wert, zuordnet, die Wahrscheinlichkeit dafür, daß eine bestimmte Beobachtung X in dieser Teilmenge M liegt³³.

Definition: Eine *Kumulative Verteilungsfunktion* (oder einfach *Verteilung*) ist eine Funktion, die jeder reellen Zahl x die Wahrscheinlichkeit dafür zuordnet, daß die Beobachtung X einen Wert liefert, der kleiner ist als x oder gleich x .³⁴

Dieser Wert wird mit $P(X \leq x)$ bezeichnet. Hier nimmt also die Menge aller reellen Zahlen, die kleiner sind als x oder gleich x , die Rolle von M ein. In vielen der uns interessierenden Fälle läßt sich eine, dann *diskret* genannte, Verteilung einfach dadurch

³² Anschaulich sind die Komponenten die (maximal großen) miteinander nicht verbundenen Teile des Graphen. Zur Komponente eines Knotens k gehören also sowohl k als auch alle Nachbarn von k und deren Nachbarn und so weiter. In vielen praktischen Fällen ist die Komponente gleich dem ganzen Graphen. Das muß aber nicht so sein. Ein extremes Beispiel dafür ist ein Graph, der nur Knoten enthält, aber keine Kanten. Dort ist jeder Knoten eine eigene Komponente. Nach der Definition ist dies nicht verboten.

³³ Ich appelliere hier eher an die Intuition der Leser, da ich an dieser Stelle nicht die mathematischen Überlegungen über die Existenz solcher Abbildungen rekapitulieren möchte. Auch einige hier eigentlich irrelevante mathematische Fachausdrücke möchte ich vermeiden. Stattdessen rekapituliere ich ein Standardbeispiel, das Werfen eines gewöhnlichen Würfels. Die Menge der möglichen Beobachtungen ist

$$S = \{1,2,3,4,5,6\}$$

Die Wahrscheinlichkeit, eine bestimmte Beobachtung X zu machen, ist für jeden Wert $P(X) = 1/6$.

³⁴ Einem Brauch bei der Benutzung von Gradverteilungen folgend werde ich künftig nicht streng zwischen den Begriffen „Verteilung“ und „Dichte“ unterscheiden. Mathematikern ist der Unterschied klar; Nichtmathematiker wird er nur verwirren.

definieren, daß für alle elementaren Ereignisse x die Wahrscheinlichkeit $P(x)$, in diesem Falle auch P_x geschrieben, bestimmt wird. Für beliebige Teilmengen M läßt sich dann definieren:

$$P(M) = \sum_{x \in M} P(x) \quad {}^{35}$$

Für nicht diskrete Verteilungen wird aus der Summe ein Integral (falls die Integration überhaupt möglich ist). Bei uns ist die Menge der möglichen Beobachtungen meistens die Menge N der natürlichen Zahlen (einschließlich 0). Da die Wahrscheinlichkeit, daß irgendeine beliebige Beobachtung gemacht wird, gleich 1 ist, gilt dann immer

$$(*) \quad \sum_{x \in N} P(x) = 1.$$

Die überragende Wichtigkeit, die die Gradverteilung beim gegenwärtigen Forschungsstand spielt, wird besonders in [GDM03] hervorgehoben. Dort können auch theoretische Begründungen für diese Tatsache nachgelesen werden.

Neben der Gradverteilung sind auch einige verwandte Kenngrößen von Netzwerken untersucht worden, zum Beispiel Entropie und Redundanz. Welche dieser Kenngrößen sich als fruchtbar erweisen werden, das kann erst die Zukunft zeigen. Für uns sind sie unwichtig.

7.2.2 Beispiele für uns wichtiger Verteilungen

7.2.2.1 Poisson-Verteilung

Dieses ist eine diskrete Verteilung. Dabei ist die Wahrscheinlichkeit eines Elementarereignisses x festgelegt durch

$$P_x = e^{-\lambda} \lambda^x / x!$$

Dabei ist λ ein freier Parameter, und der Faktor $e^{-\lambda}$ erzwingt wegen $\sum_{x \in N} \lambda^x / x! = e^\lambda$ die Gültigkeit der obigen Gleichung (*).

Die Poisson-Verteilung zeichnet sich dadurch aus, daß bei dem durchschnittlichen Wert k eine scharfe Spitze P_k existiert. Bei den übrigen Werten x nehmen die P_x überexponentiell, das heißt noch schneller als exponentiell, also sehr schnell, ab. In unserem Zusammenhang erfolgt diese Abnahme unrealistisch schnell.

7.2.2.2 Exponential-Verteilung

Hier haben wir eine nicht-diskrete Verteilung vor uns, deren Kumulative Verteilungsfunktion gegeben ist durch

³⁵ Die Wahrscheinlichkeit, eine durch 3 teilbare Zahl zu würfeln ($M = \{3,6\}$) ist dann

$$P(M) = P(3) + P(6) = 1/3.$$

$$\begin{aligned} P(X \leq x) &= 1 - e^{-x/\beta} && \text{für } x \geq 0, \\ P(X \leq x) &= 0 && \text{für } x < 0. \end{aligned}$$

Dabei ist β ein Parameter für den Mittelwert. Hier gilt das, was auch schon für die Poisson-Verteilung gesagt wurde: Die Verteilungsfunktion nimmt exponentiell, also ebenfalls sehr schnell, ab. Beide Verteilungen genießen für Netzwerke noch großes aber eher historisches Interesse.

7.3 Die Verteilung des Zipfschen Gesetzes (Pareto-Verteilungen)

Dieses Unterkapitel gehört von der Logik her zu Abschnitt 7.2.2. Wegen des Umfangs der nötigen Erläuterungen und der Wichtigkeit der Pareto-Verteilungen (oder Potenzverteilungen) soll diesen jedoch ein eigenes Unterkapitel gewidmet werden.

Die Mitglieder der Klasse der Pareto-Verteilungen sind festgelegt auf den Elementarereignissen durch

$$P_x = cx^{-\gamma} \text{ mit } x > 0.$$

Dabei ist γ ein Parameter, der Exponent der Verteilung. Es muß $\gamma > 1$ sein, da sonst $\sum_{x \in \mathbb{N}} x^{-\gamma}$ nicht konvergiert. Dann ist Gleichung (*) aus Abschnitt 7.2.1 nicht mehr erfüllbar. Falls andererseits $\gamma > 1$ gilt, wird c gleich dem Kehrwert dieser Zahl³⁶ gesetzt und so die Gültigkeit der Gleichung (*) erzwungen.

Gehorcht eine Gradverteilung einer Pareto-Verteilung, so zeichnet sie sich gegenüber der Poisson- und der Exponential-Verteilung dadurch aus, daß der Anteil der Knoten, mit relativ hohem Grad sogenannter „hubs“³⁷, besonders groß ist. Zwar nähert sich die Zahl der Knoten, die einen bestimmten Grad k haben, für alle Verteilungen bei wachsendem k immer mehr der 0 an, aber bei Pareto-Verteilungen geschieht dies langsamer, eben polynomiell und nicht exponentiell³⁸. Dadurch wird der lange rechte Teil der Verteilung dicker, was auch zu der Bezeichnung „fat-tail-distribution“ geführt hat. Ein anderer Name, der auf eine formale Ähnlichkeit mit Fraktalen zurückgeht, nämlich der der Skaleninvarianz, ist „scale-free“. Das soll aber hier nicht weiter erläutert werden.

Die Pareto-Verteilungen³⁹ werden die für uns wichtigsten Verteilungen im Verlauf der Arbeit sein. Deshalb seien hier einige exemplarisch zusammen mit einer Poisson-Verteilung

³⁶ Dieser Wert $\sum_{x \in \mathbb{N}} x^{-\gamma}$ wird auch mit $\zeta(\gamma)$ bezeichnet. Dabei ist ζ eine der berühmtesten Funktionen der Mathematik, die *Riemannsche ζ -Funktion*.

³⁷ Das Wort „hub“ bedeutet im Englischen „Nabe“. Es spielt auf alte Speichenräder an, wo die Nabe ähnlich stark durch Speichen mit dem Rand des Rades verbunden ist wie die hubs in einem Netz mit ihren Nachbarn durch Kanten. Der Begriff „hub“ wird immer wieder in der Arbeit auftauchen.

³⁸ Wir stoßen hier also wieder auf den Unterschied zwischen polynomiell und exponentiellem Wachstum, der auch in Unterkapitel 6.1 eine so bedeutende Rolle gespielt hat.

³⁹ Die Pareto-Verteilung ist nach dem weiter unten erwähnten italienischen Wirtschaftsmathematiker Vilfredo Pareto benannt, der sie erstmals zur Beschreibung realer Phänomene benutzt hat. Der Name Potenz-Verteilung, der besonders in der angelsächsischen Literatur vorherrscht, beruht darauf, daß die Wahrscheinlichkeit eines Elementarereignisses x mit der Potenz γ gegen 0 geht.

zum Vergleich skizziert. Die genauen Werte sind mit Hilfe des Programmes ParetoPoissonTest.cpp erstellt worden und in der Datei verlaufParetoPoisson.txt verzeichnet. Beide Dateien sind auf der CD zu finden.

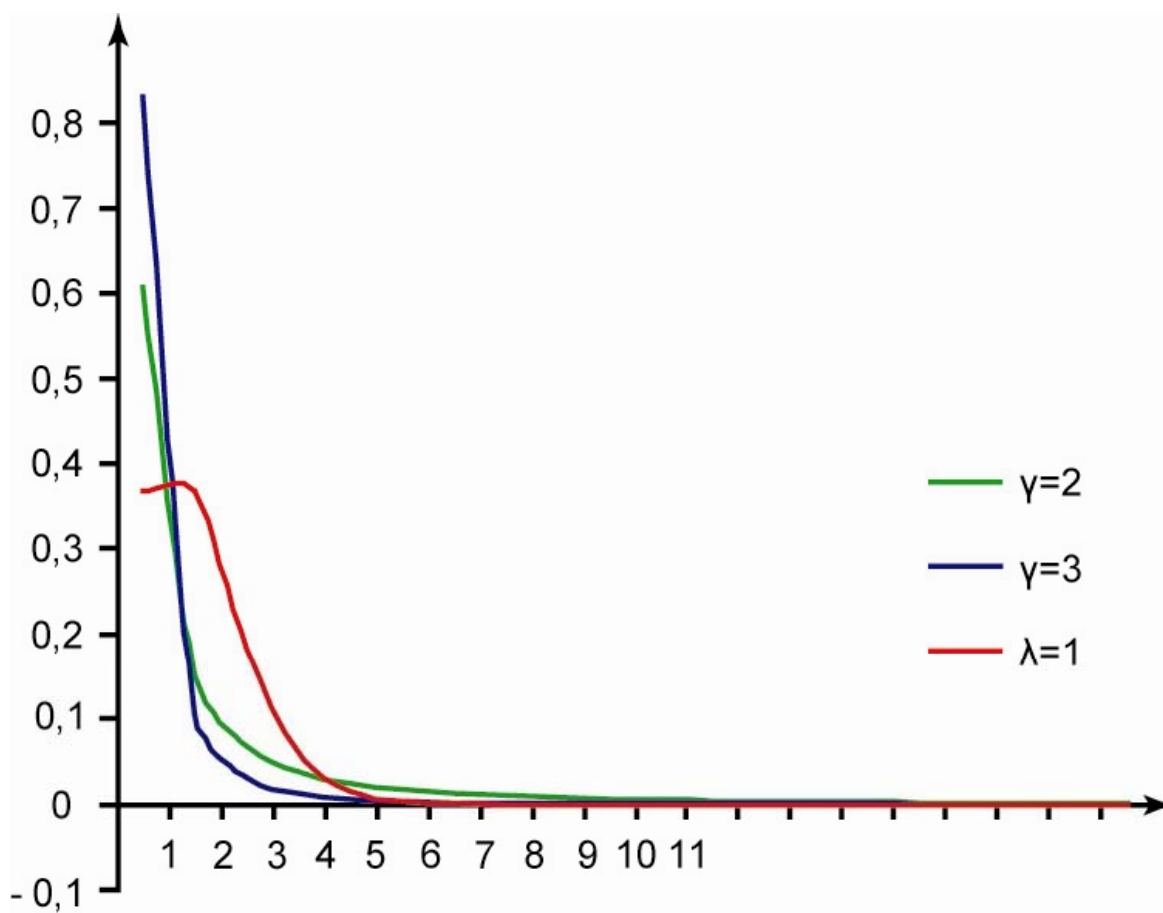


Abbildung 22: Pareto-Verteilungen und Poisson-Verteilungen

Die grüne und die blaue Kurve zeigen den Verlauf einer Paretoverteilung mit dem Parameter $\gamma = 2$ und mit $\gamma = 3$ an (in interpolierter Form). Zum Vergleich ist auch noch die Poisson-Verteilung mit dem Parameter $\lambda = 1$ hinzugefügt. Der y -Wert an der Stelle x bedeutet also die Wahrscheinlichkeit für das Auftreten des Elementarereignisses x . Wie alle Verteilungen gehen auch diese Verteilungen für wachsende x gegen 0 . Es ist aber schwer zu sehen, was für uns das Wesentliche ist: Die Pareto-Verteilung geht wesentlich langsamer gegen 0 , so daß auch für größere x noch eine gewisse Wahrscheinlichkeit für das Auftreten des entsprechenden Elementarereignisses vorhanden ist, jedenfalls um Größenordnungen höher als bei der Poisson-Verteilung. Deshalb habe ich noch Abbildung 23 eingefügt, wo dieselben Kurven logarithmiert zu sehen sind. Jetzt bedeutet also der y -Wert an der Stelle x den Logarithmus (zur Basis 10) der Wahrscheinlichkeit für das Auftreten des Elementarereignisses x . Jetzt sind die unterschiedlichen Geschwindigkeiten des negativen Kurvenwachstums deutlich zu sehen.

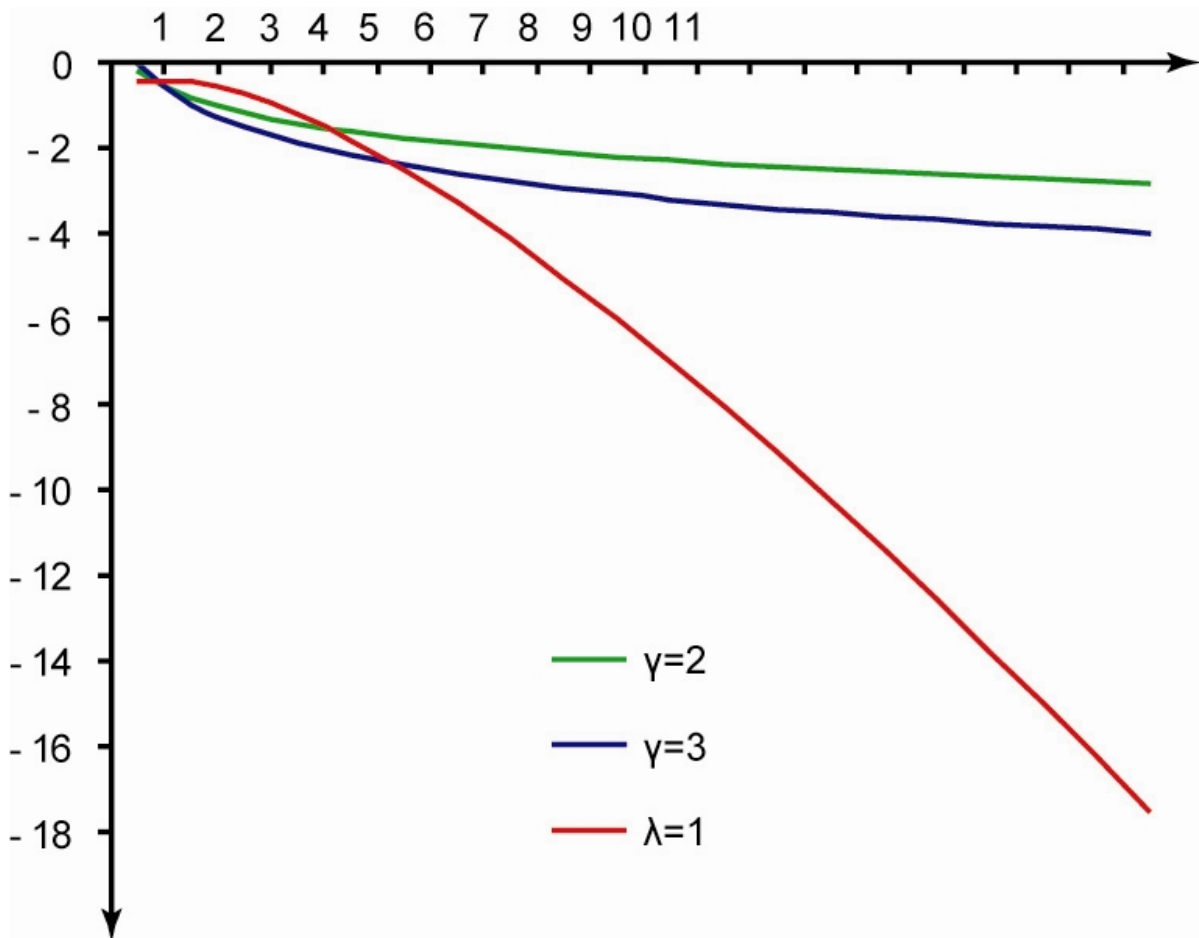


Abbildung 23: Pareto-Verteilungen und Poisson-Verteilungen (logarithmiert)

Wegen $P_x = cx^{-\gamma}$ gilt in Abbildung 23 für die grüne und die blaue Kurve:

$$y(x) = \log(cx^{-\gamma}) = \log(c) - \gamma \log x,$$

und daher liegt hier bei den Pareto-Verteilungen im wesentlichen eine negative logarithmische Funktion vor, also eine Funktion die sehr langsam gegen $-\infty$ strebt. Dagegen nähert sich wegen $P_x = e^{-\lambda} \lambda^x / x!$ und daher

$$y(x) = \log(e^{-1} / (x+1)!) = -\log(e) - (\sum^{x+1} \log(i))$$

die logarithmierte Poisson-Verteilung sogar etwas schneller dem negativen Unendlichen als eine lineare Funktion. Das sieht man in Abbildung 23 daran, daß die rote Kurve im wesentlichen unterhalb einer Gerade mit der negativen Steigung von -1 liegt. (Sie selber ist keine Gerade!)

Messungen in tatsächlich existierenden komplexen Netzwerken haben Pareto-Verteilungen mit Exponenten zwischen 1 und 3 ergeben, siehe [AB02], wobei der weit

überwiegende Teil sogar Exponenten zwischen 2 und 3 besitzt. Dieser scheinbar minimale Unterschied ist besonders wichtig für epidemiologische Untersuchungen im Zusammenhang mit der „Epidemischen Schwelle“, siehe 9.4. Für eine umfangreiche Aufstellung der Parameter in 37 tatsächlich existierenden komplexen Netzwerken siehe [DM03], p.80 f..

Klassische Beispiele für solche Verteilungen:

- Einkommensverteilungen (Pareto) [Par1897]⁴⁰,
- Anzahl der von einem Wissenschaftler publizierten Artikel (Lotka) [Lot26]⁴¹,
- Bevölkerungen in Großstädten (Zipf) [Zip49].

Beispiele für solche Verteilungen in der Linguistik:

Die wohl bekannteste Pareto-Verteilung in der Linguistik ist die der Wortfrequenzen. Das ist gerade die Aussage des Zipfschen Gesetzes [Zip49]. In der die hier geeignetsten Zusammenhänge am deutlichsten machenden Fassung besagt es, daß, wenn die Wörter eines Textes nach absteigender Häufigkeit geordnet werden, die Häufigkeit P_x des x-ten Wortes durch ein Potenzgesetz der Form

$$P_x = cx^{-\gamma} \text{ in diesem Falle mit } \gamma = 1$$

gegeben ist. Ob diesem Verhalten ein Ökonomieprinzip zugrunde liegt, das ist auch heute noch umstritten. In [FS03] zeigen die Autoren, daß man zumindest unter der Annahme eines Ökonomieprinzips mit Hilfe des word webs, siehe Unterkapitel 10.3, das Zipfsche Gesetz ableiten kann. Über dieses Gesetz ist noch mehr in Unterkapitel 10.10 zu sagen.

Es sind aber auch in neuerer Zeit immer wieder einmal Pareto-Verteilungen in der Linguistik aufgetaucht, etwa bei einer Untersuchung über Wortfrequenzen in [Tul95].

Bei den beiden letzten Beispielen handelt es sich schon um Gradverteilungen der in Untersuchungen zu Netzwerken in der Sprachwissenschaft gewünschten Form, wie sie in Kapitel 10 noch genauer beschrieben werden. Insgesamt gibt es verschiedenartige Verteilungen wie Sand am Meer. Eine umfassende Liste sehr vieler in der wissenschaftlichen Modellierung üblicher Verteilungen mit ihren Eigenschaften kann etwa in [LK91] gefunden werden.

Warum interessieren uns gerade die oben angegebenen? Die Antwort ist bedauerlicherweise nicht inhaltlich, also von den zu untersuchenden Gegenständen her, sondern in den Unzulänglichkeiten der Mathematik zu finden: Es ist zu schwierig, Netze mit gewünschten Eigenschaften zu konstruieren.

⁴⁰ Dies ist die oben erwähnte Untersuchung, nach der die Pareto -Verteilung ihren Namen bekommen hat.

⁴¹ Darin findet sich das oben erwähnte Gesetz der wissenschaftlichen Produktivität.

Die Gradverteilung ist eine der bedeutendsten Kenngrößen eines Netzwerkes. Dabei bedeutet P_k gerade die Wahrscheinlichkeit, daß ein zufällig im Graphen ausgewählter Knoten gerade als Ausgangspunkt von k Kanten dient. Es ist für vorgegebene Verteilungen möglich, gezielt einen Graphen zu konstruieren, bei dem die Knotengrade dieser Verteilung in etwa gehorchen, zum Beispiel das unten angeführte „configuration-Modell“. Aber dann ist es sehr schwierig, zusätzliche Eigenschaften zu erzwingen. Bei den drei oben angegebenen Verteilungen geht das zum Teil; glücklicherweise sieht man insbesondere für die Pareto-Verteilung heute ein wenig (aber wirklich nur ein wenig) klarer. Inwieweit gerade diese Verteilungen für soziale Netzwerke, und damit für Modelle, mit denen sprachliche Innovationen untersucht werden können, anwendbar sind, darauf gehe ich weiter unten noch genauer ein.

Der erste, der sich mit der Verteilung von Graden in dem uns interessierenden Zusammenhang befaßt hat, war der Mathematiker Anatol Rapoport in [Rap57].

7.4 Andere statistische Kenngrößen für Netzwerke

7.4.1 Clustering

Bei dem *Clustering* geht es um die Frage, ob die Kanten innerhalb des Netzes gleichmäßig (wenn auch entsprechend einer Verteilung) verteilt sind oder ob es beispielsweise Cliquen von Knoten, sogenannte *Cluster*, gibt, die dichter untereinander vernetzt sind. Daß ein mathematisches Maß für diese Erscheinung in unserem Zusammenhang eine bedeutende Rolle spielt, das ist offensichtlich. Hier sollen einige der in der Literatur am häufigsten verwandten Maße für das Clustering vorgestellt werden. Ihnen liegen zwei Überlegungen zugrunde:

- Clustering ist eine lokale Eigenschaft. Ein Cluster entsteht, wenn aus einer Verbindung zwischen zwei Knoten a und b und einer zwischen b und c gefolgert werden kann, daß auch eine Verbindung zwischen a und c existiert. Das Clustering mißt, wie oft diese Folgerung gilt und wie oft nicht. Damit ist das Clustering ein Maß für die Transitivität der Verbindungsrelation⁴².
- Als eine Art Standardmaß bei der Messung des Clustering dient der Vollständige Graph V_n (siehe 6.2)⁴³, der am dichtesten geclusterte Graph überhaupt.

Die Lokalität aus der ersten Bedingung geht dadurch in die Definition ein, daß zuerst lokale Nachbarschaften betrachtet werden.

Definition: Seien $G = (V, E)$ ein Graph und $v \in V$ ein Knoten in G . Die *lokale Nachbarschaft* von v ist der Teilgraph $N(v) = (N, K)$ von G mit den Eigenschaften:

⁴² Ich erinnere daran, daß der Begriff der Transitivität bei der Definition der Halbordnung in Unterkapitel 4.1 eingeführt worden ist. Auch für Transitivität gilt, was ich oben schon über Relationen ganz allgemein geschrieben habe, nämlich daß sie fast überall vorkommt, ein in der Mathematik gut untersuchter einfacher Begriff ist und interessante Eigenschaften hat.

⁴³ Es läßt sich leicht ein Satz beweisen: Ein zusammenhängender, symmetrischer und reflexiver Graph, in dem die Verbindungsrelation transitiv ist, ist schon der Graph V_n .

- $N(v)$ besteht aus allen Nachbarn von v ⁴⁴, also
$$w \in N \Leftrightarrow (v,w) \in E$$
- Die Verbindungsrelation K in $N(v)$ ist genau die Einschränkung der Verbindungsrelation E von G auf $N(v)$, also: Seien
$$x \in N \wedge y \in N$$
Dann ist
$$(x,y) \in K \Leftrightarrow (x,y) \in E$$

Beispiel: In Abbildung 24 bestehe der gegebene Graph aus allen Objekten, welcher Farbe auch immer. Sei der blaue Knoten als v bezeichnet. Seine lokale Nachbarschaft besteht dann einmal aus den gelb gezeichneten Knoten, also seinen Nachbarn im Graphen. Die Kanten der lokalen Nachbarschaft sind die Kanten, die zwischen den gelben Knoten im Graphen vorhanden sind (ebenfalls gelb bezeichnet). Ausdrücklich nicht zur lokalen Nachbarschaft gehören v selber und die Kanten von v aus (blau bezeichnet). Das wäre nur dann anders, wenn es im Graphen eine Kante von v nach v gäbe. Das ist bei meiner Konstruktion nicht verboten, wird im Beispiel aber nicht thematisiert. Im Beispiel sieht man auch noch etwas anderes: Die lokale Nachbarschaft (alle gelben Objekte) muß nicht zusammenhängend sein.

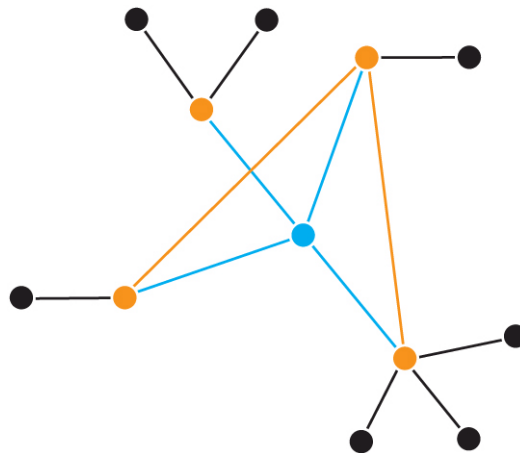


Abbildung 24: Lokale Nachbarschaft

Wir nehmen für einen Augenblick an, $N(v)$ wäre ein vollständiger Graph mit n Knoten. Da dann bei ihm jeder Knoten mit jedem anderen Knoten genau einmal verbunden wäre, überlegt man sich leicht, daß er dann genau $n(n-1)/2$ Kanten enthielte. Dieser Graph ist das oben angesprochene Standardmaß für das Clustering. Sei jetzt $N(v)$ wieder der tatsächlich existierende Teilgraph von G mit n Knoten und m Kanten. Das zahlenmäßige Verhältnis von tatsächlich vorhandenen zu potentiell vorhandenen Kanten ergibt für $N(v)$ den lokalen Clustering-Koeffizienten, hier $C(v)$ genannt. Dieser läßt sich allgemein für Graphen definieren:

⁴⁴ Man beachte, daß damit im allgemeinen ein Knoten v nicht in seiner Nachbarschaft enthalten ist, es sei denn, es gibt bei v eine Schleife: $(v,v) \in E$.

Definition: Zu einem Graphen G mit n Knoten⁴⁵ und m Kanten sei der *lokale Clustering-Koeffizient* C definiert durch

$$C = 2m / n(n-1).$$

Beispiel: Wäre $N(v)$ vollständig, so gälte also $C = 1$. Für den Fall $N(v)=V_4$, entstanden als lokale Nachbarschaft eines Knotens in V_5 , wird das durch das linke Bild in Abbildung 25 illustriert, wo die Farben dieselbe Bedeutung wie in Abbildung 24 haben. Wäre andererseits $N(v)$ im ursprünglichen Graphen nur durch v zusammengehalten worden und bestünde $N(v)$ folglich jetzt lediglich aus unverbunden Knoten, so wäre jetzt $C = 0$. Der letzte Fall tritt in Bäumen auf, wo sogar *alle* lokalen Clustering-Koeffizienten den Wert 0 haben. Dieser Fall wird das durch das rechte Bild in Abbildung 25 illustriert, wo die Farben ebenfalls dieselbe Bedeutung wie in Abbildung 24 haben.

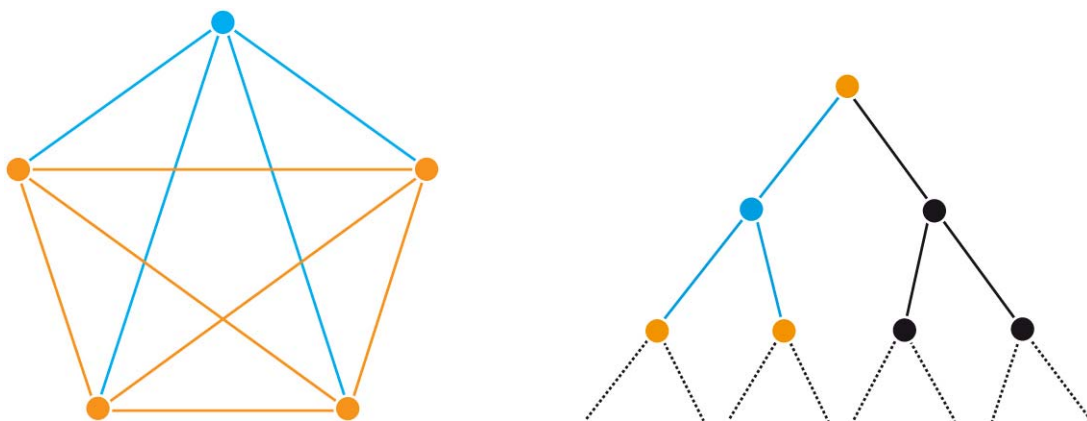


Abbildung 25: Lokale Nachbarschaften in speziellen Graphen

Wir erhalten also die nicht sehr überraschende Erkenntnis, daß es in Bäumen weniger Cliques als in vollständigen Graphen gibt.

Dieses ist wohl die am häufigsten benutzte aber keineswegs die einzige in der Literatur bekannte Definition eines Clustering-Koeffizienten. Insbesondere ist sie nur schlecht anwendbar, wenn man auch mehrfache Kanten zwischen zwei Knoten zuläßt. Eine alternative Definition macht sich den oben schon angesprochenen Begriff der Transitivität in der folgenden Weise zu Nutze: Die Größe der in einem Netzwerk vorhandenen Cliques wächst mit der Wahrscheinlichkeit, daß zu zwei gegebenen Kanten von einem Knoten A zu einem Knoten B sowie vom Knoten B zu einem weiteren Knoten C auch die Kante von A nach C existiert. Diese Überlegung führt zu der folgenden alternativen

Definition: Zu einem Graphen G sei der *lokale Clustering-Koeffizient* C definiert durch

$$C = 6 \cdot \text{Anzahl der Dreiecke in } G / \text{Anzahl der Pfade der Länge 2}.$$

⁴⁵ Ich werde weder hier noch im folgenden auf den Fall eingehen, daß $n=0$ oder $n=1$ ist.

Daß diese Definition genau unsere Intuition widerspiegelt, das soll das folgende Beispiel illustrieren:

Beispiel: Sei $G = V_3$ mit den Knoten A, B, C . Dann besteht G aus genau einem Dreieck. Dieses Dreieck kann aber auf sechs Arten durchlaufen werden:

ABC, ACB, BAC, CBA, BCA, CAB

Diese Tatsache erklärt die 6 in der Formel. Jeder dieser Arten des Durchlaufs entspricht genau ein Weg der Länge 2, blau in der Abbildung. Damit gilt auch bei dieser Definition für den lokalen Clustering-Koeffizienten in dem vorliegenden vollständigen Graphen der Größe 3 die Gleichung $C = 1$.

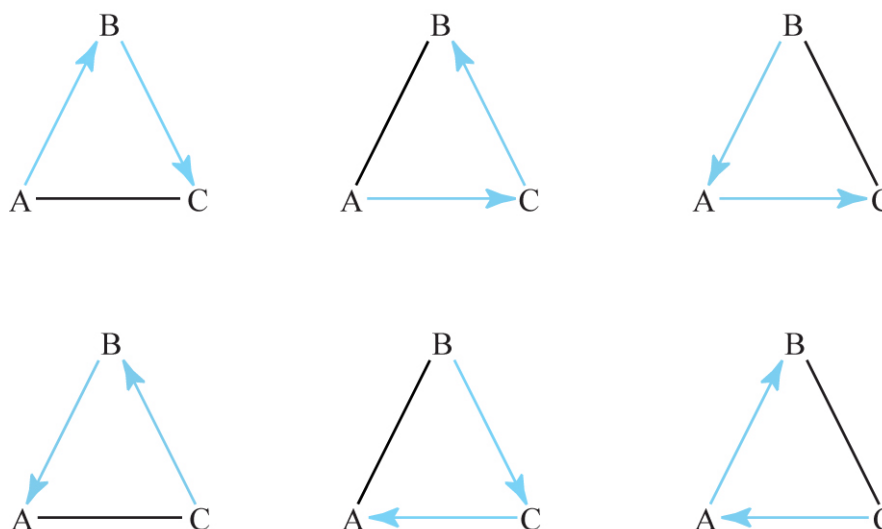


Abbildung 26: Durchläufe im gleichseitigen Dreieck

Wie man sich leicht überlegen kann, gilt auch bei dieser Definition für alle lokalen Clustering-Koeffizienten C eines beliebig großen vollständigen Graphen V_n wieder $C = 1$. Für einen Baum mit n Knoten und demzufolge $n-1$ Kanten gilt erneut immer $C = 0$, da es in Bäumen keine Dreiecke gibt.

Was noch nicht definiert worden ist, das ist der allgemeine (also nicht der lokale) Clustering-Koeffizient eines Graphen. Dieser wird normalerweise durch Mittelung der lokalen Clustering-Koeffizienten bestimmt:

Definition: Seien $G = (V, E)$ ein Graph und zu jedem Knoten $v \in V$ der lokale Clustering-Koeffizient mit $C(v)$ bezeichnet. Dann wird der Clustering-Koeffizient C von G berechnet durch

$$C = (1/|V|) * \sum_{v \in V} C(v)$$

Beispiel: Da sich die Aussagen über lokale Clustering-Koeffizienten von vollständigen Graphen und Bäumen auf alle Knoten beziehen, gelten auch für den allgemeinen Clustering-Koeffizienten nach beiden Definitionen die Werte $C=1$ und $C=0$. Es ist sogar so, daß vollständige Graphen und Bäume durch diese Werte gekennzeichnet sind.

In der Literatur finden sich aber auch für dieses Thema noch weitere Versuche zur Definition eines Maßes für das Clustering, die jedoch außerhalb des Bereiches dieser Arbeit liegen.

7.4.2 Hierarchie und Gradkorrelationen

Eng mit dem Clustering ist die Frage der *Hierarchie* verbunden: In vielen tatsächlich vorhandenen Netzwerken gibt es eine modulare Struktur. Man kann Gruppen von Knoten finden, die sehr stark untereinander vernetzt sind aber eine wesentlich geringere Anzahl von Verbindungen nach außen haben. Diese Erscheinung kann sich natürlich über mehrere Abstraktionsebenen hinweg fortsetzen. Solche Netze werden in [RB03] als „Hierarchische Netzwerke“ bezeichnet. Beispiele von dort sind das world wide web und, in unserem Zusammenhang besonders wichtig, soziale Netze.

Untersuchungen über die *Gradkorrelation* haben sehr disparate Ergebnisse gebracht. Bei der Gradkorrelation handelt es sich um ein Maß dafür, inwieweit der Grad eines Knotens mit den Graden seiner Nachbarknoten korreliert ist, platt ausgedrückt also dafür, ob Knoten mit hohem Grad vorrangig mit Knoten interagieren, die ihrerseits einen hohen Grad aufweisen, oder ob das nicht der Fall ist.

Beispiel:

In einem Netz aus Flughäfen und den zwischen ihnen bestehenden Verbindungen sind die Grade zwischen Nachbarn korreliert: Eine Verbindung von Frankfurt, also einem Verkehrsgrößknotenpunkt, wird eher zu einem anderen Verkehrsgrößknotenpunkt wie London führen und nicht zu einem kaum ausgebauten und benutzten Landeplatz irgendwo in der Provinz.

In einem Netz aus Städten und Autobahnverbindungen sind die Grade zwischen Nachbarn nicht korreliert: Die nächste Ortschaft neben dem hochverbundenen Hamburg ist nicht beispielsweise Berlin oder Bremen, beide ebenfalls hochverbunden, sondern etwa Stapelfeld oder Ramelsloh, die ihrerseits nicht viele Verbindungen haben.⁴⁶

Mit der Gradkorrelation haben wir einen der Fälle vor uns, bei dem sich soziale und andere Netzwerke grundlegend unterscheiden. Werden Netze bewußt entworfen, wie etwa technische Netze, so zeigt sich fast immer eine negative Gradkorrelation (Beispiel: [PVV01] für das Internet), während die Grade bei sozialen Netzen eindeutig positiv korreliert sind. Als ein Hauptgrund für dieses Phänomen wird in [NP03] die Tatsache identifiziert, daß Bekanntschaft dazu tendiert transitiv zu sein. Daher weisen soziale Netze, wie Newman und Park nachweisen, auch einen besonders hohen Clustering-Koeffizienten auf. Als tieferen Grund geben sie eine Erscheinung an, die sie „*community structure*“ nennen. Auf die Details werde ich noch näher eingehen, wenn ich Algorithmen vorstelle, durch die Netze mit

⁴⁶ Auf die Gründe für die Strukturunterschiede soll hier nicht weiter eingegangen werden. Die Netze werden hier lediglich so, wie sie vorhanden sind, zur Illustration einer Netzeigenschaft herangezogen.

bestimmten vorgegebenen Eigenschaften, in unserem Fall mit hohem Clustering-Koeffizienten und positiver Gradkorrelation, erzeugt werden können.

7.4.3 Der „small-world“-Effekt

In den Gesellschaftswissenschaften, insbesondere der Soziologie, ist schon Ende der zwanziger Jahre ein Effekt in großen Mengen von Individuen beobachtet worden: Wählt man sich zwei beliebige Individuen, etwa zwei Angehörige eines Volkes, und untersucht die Ketten von Bekanntschaften durch die diese miteinander verbunden sind, so sind diese Ketten erstaunlich kurz. Für die kürzestmögliche Verbindung wird seit der Arbeit [Mil67] der Wert 6 immer wieder herangezogen. Es gibt also vermutlich eine Kette höchstens der Länge 6 von wechselseitigen Bekanntschaften, durch die jeder Mensch mit dem Papst, G.W.Bush oder dem Mafiapaten Giorgio Bandito in Palermo verbunden ist.

Etwas genauer gesagt, wird in small-world-Netzwerken vorausgesetzt, daß ihr Durchmesser in irgendeinem Sinne klein ist⁴⁷, damit gesagt werden kann, solche Netzwerke zeigen den *small-world-Effekt*. Bei technischen Netzwerken wird aber auch von small-world-Netzen mit höheren Zahlen gesprochen. So wird für das world wide web in [AJB99] der Wert 19 angegeben⁴⁸. Erste gezielte wissenschaftliche Untersuchungen über den small-world-Effekt wurden schon in [Mil67] vorgenommen. Heute gibt es eine unübersehbare Literatur über Netze mit dem small-world-Effekt. Als Startpunkt für die analytische Untersuchung von small-world-Netzwerken gilt der kurze Artikel [WS98]. Darin werden diese Netzwerke als Kompromiß zwischen zwei Extremen erkannt und untersucht:

- regulären Netzwerken (also durch feste Regeln erzeugt),
- Zufallsnetzwerken (also durch Zufallsverfahren erzeugt).

In [MG00] werden small-world-Effekte durch eine Optimierung bezüglich zweier Ziele erklärt:

- einer möglichst engen (und daher besonders schnellen) lokalen Verkettung,
- einer möglichst sparsamen (und so vom Aufwand her optimierten) Verbindung ganz allgemein.

Auch fast alle uns interessierenden Netze zeigen den small-world-Effekt. Das Dynamische Sprachnetz wird ihn ebenfalls zeigen. Ein offensichtliches Beispiel für Netze ohne diesen Effekt sind hingegen die isotropen Gittergeometrien.

Ich möchte nur einen Punkt deutlich machen: Es ist nur dann sinnvoll, von einem small-world-Effekt zu sprechen, wenn es nur relativ wenige Verbindungen gibt, die Wege aber trotzdem kurz sind. Wenn wir etwa einen vollständigen Graphen vor uns haben, so ist es überhaupt nicht verwunderlich, wenn er einen small-world-Effekt zeigt. Man kommt eben mit

⁴⁷ Oft wird einfach 6 gewählt; streng genommen ist die Definition jedoch so, daß der Durchmesser mit dem Logarithmus der Knotenzahl wächst.

⁴⁸ Streng genommen würde es sich dann bei dem www nicht mehr um ein small-world-Netzwerk handeln, da der Wert 19 für einen Logarithmus zu hoch ist. Aber verschiedene Autoren verwenden verschiedene Definitionen. So wird auch statt des Durchmessers eines Netzwerkes N , also der Länge des längsten geodäsischen Pfades in N , manchmal die Länge eines durchschnittlichen geodäsischen Pfades in N genommen.

einem Schritt von jedem Knoten zu jedem Knoten. Bemerkenswert ist aber, daß die Netze, mit denen wir uns in dieser Arbeit beschäftigen, insbesondere die sprachlichen Netze, diesen Effekt zeigen, obwohl sie in ihren Verbindungen nur sparsam sind.

7.5 Die zentrale verbundene Komponente

Im allgemeinen muß ein Graph, wie oben schon erwähnt, nicht verbunden sein. In vielen Fällen gibt es aber, auch falls das nicht der Fall sein sollte, eine besonders große verbundene Komponente, die *zentrale verbundene Komponente* oder *giant connected component (gcc)*. Die größte Komponente eines Netzwerkes gilt als gcc, wenn sie sich sowohl in x - als auch in y -Richtung über das gesamte Netzwerk erstreckt. Wann bei einem Zufallsnetzwerk eine gcc entsteht und wie groß diese dann ist, das ist genauso ein im allgemeinen interessanter Forschungsgegenstand wie die Frage nach der Verteilung der Größen der einzelnen Zusammenhangskomponenten. Bei den hier beabsichtigten Untersuchungen spielen Zusammenhangskomponenten aber keine Rolle, da wir getrost annehmen können, daß es keine isolierten Individuen oder Sprechergruppen gibt. Entsprechend muß ein Modell gewählt werden, daß von Anfang an den Zusammenhang des Netzwerkes sicherstellt.

7.6 Weitere statistische Kenngrößen

In der Literatur gibt es eine Fülle weiterer statistischer Kenngrößen, beispielsweise die oben in 6.2 genannten Distanzmaße wie der Durchmesser. Eine ausführliche Erörterung kann in [BE05], Kap. 11, p.293 ff., gefunden werden.

7.7 Charakteristika von Netzwerken

Der Übersichtlichkeit halber gebe ich hier noch einmal die wichtigsten Kenngrößen für Netzwerke in Form einer kurzen Liste an:

- die Anzahl N der Knoten, oft *Größe* des Netzwerkes genannt,
- die Anzahl der Kanten,
- den totalen Grad des Netzwerkes K , die Summe aller Knotengrade,
- den mittleren Grad des Netzwerkes, also K/N ,
- die Verteilung der Grade,

- der Gradkorrelationskoeffizient,
- die durchschnittliche Länge von geodäsischen Pfaden,
- die Länge des längsten geodäsischen Pfades,
- der Clustering-Koeffizient,
- die Größe der zentralen verbundenen Komponente,
- die Verteilung der Größen der Zusammenhangskomponenten.

Um noch einmal herauszuheben, wie sehr die Forschung über Netzwerke noch in den Kinderschuhen steckt, möchte ich noch einmal darauf hinweisen, daß bisher kein theoretischer Gesamtrahmen dafür existiert. Man schaut an einige bestimmte Stellen, weil man dort etwas aussagen kann, ohne genau zu wissen, ob es überhaupt die richtigen Stellen sind. Mehr kann dem Schlußwort in [New03] entnommen werden.

8 Zufallsnetzwerke

Dieses und das nächste Kapitel stellen vor allem die für diese Arbeit benötigten mathematischen Ergebnisse anderer Autoren aus den letzten zehn Jahren vor. Sie mögen deshalb aus der Sicht der Sprachwissenschaft etwas technisch sein oder scheinbar vom Thema wegführen. Da es sich aber um so neue Erkenntnisse handelt, ist es überhaupt nicht verwunderlich, daß es noch nicht viele Verbindungen zur Linguistik gibt. Es ist ja gerade das Anliegen meiner Arbeit, zu zeigen, daß solche Ansätze mit Gewinn bei der Untersuchung von Sprache verwandt werden können. Insofern bin ich froh, wenigstens einige „Kronzeugen“ zu haben, die schon ähnliche Aspekte beleuchtet haben. Gegenwärtig gibt es Querverbindungen zwischen der Linguistik und der Theorie der Netzwerke, die zeigen, daß mein Ansatz zumindest erfolgversprechend ist; und auf solche Verbindungen werde ich auch hinweisen, sobald es die Logik der Arbeit nahelegt. Aber beim gegenwärtigen Stand der *mathematischen* Forschung noch viel mehr solcher Anwendungen zu erwarten, wie ich sie vorstellen werde, das wäre vermessen.

Will man statistische Untersuchungen in komplexen Netzwerken anstellen, so hat man, wie schon erwähnt wurde, zwei Möglichkeiten:

- Man kann versuchen, auf analytischem Wege Ergebnisse zu erzielen,
- oder man kann versuchen, interessierende Kenngrößen vorher festzulegen und dann Graphen von hinreichender Größe in Simulationen zu erzeugen, die die Anforderungen erfüllen.

Auch im letzten Fall ist natürlich ein Beweis erforderlich, daß die Anforderungen tatsächlich erfüllt sind. Der erste Ansatz ist zwar befriedigender, aber trotz aller in den letzten fünf bis zehn Jahren gemachten Fortschritte ist die Menge der Probleme, die analytisch gelöst werden können, noch enttäuschend klein und wird es sicher bleiben⁴⁹. Auch im zweiten Fall sind noch zu wenige Beweise für Aussagen bekannt, aber mathematisch abgesicherte Verfahren, die den Computer benutzen, sind doch in unvergleichlich größerem Maße bekannt und anwendbar.

Die Situation ist ähnlich wie bei der Theorie der partiellen Differentialgleichungen, wo sich das Herz jedes Mathematikers über analytische Lösungen freut; leider gibt es so wenige. Über mathematisch gesicherte Verfahren mit Computer freuen sich besonders die Physiker. Und sie erhalten bei ihren etwas bescheideneren Ansprüchen eine reiche Fülle von Lösungen für ihre Fragen.

Bei der Arbeit an linguistischen Anwendungen mit dem dynamischen Sprachnetz werde ich den zweiten Weg gehen und sogenannte *Zufallsnetzwerke* untersuchen, also Netzwerke, die zufällig erzeugt werden, aber mit mathematischer Sicherheit Eigenschaften aufweisen, wie sie in Netzen wünschenswert sind, bei uns in sozialen Netzen.

⁴⁹ Diese Aussage ist nicht exakt quantifizierend zu verstehen. Gemeint ist ein allgemeines Problem der Modellierung: Die reine Mathematik ist nicht weit genug entwickelt, als daß komplexe Systeme aus der Realität im allgemeinen tief genug analysiert werden könnten; grob gesagt: Man hat sehr oft bei analytischen Ansätzen die Wahl, mit einem vereinfachten aber recht unrealistischen Modell zu arbeiten oder mit einem realistischeren, das dann so komplex ist, daß schließlich zum Zwecke der Analyse noch Zusatzannahmen eingeführt werden müssen, die die Allgemeingültigkeit in Frage stellen.

Erste Untersuchungen in diese Richtung waren die bereits erwähnte Arbeit von Rapoport [Rap57] sowie die in [ER59], [ER60] präsentierten Modelle.

Für eine ausführliche Lektüre über Netzwerkmodelle verweise ich auf das schon mehrfach genannte Buch von Dorogovtsev und Mendes [DM03] und das Kapitel 13 in [BE05], pp.341 ff. . Ich möchte an dieser Stelle lediglich einige Verfahren vorstellen, mit deren Hilfe für uns wichtige Verteilungen erzeugt werden können.

8.1 *Das configuration-Modell („Igel-Modell“)*

Dieses seit den siebziger Jahren von vielen Autoren untersuchte Modell kann für jede vorgegebene Verteilung V ein Zufallsnetzwerk erzeugen, bei dem die Gradverteilung gleich V ist. Das Netzwerk wird so erzeugt:

Es werden für eine gegebene Zahl N von Knoten die Knotengrade k_i festgelegt ($i = 1, \dots, N$), so daß die k_i hinreichend genau der Verteilung V entsprechen. Auf die Frage, wie N Zufallszahlen so erzeugt werden können, daß sie genau einer gegebenen Verteilung entsprechen, will ich an dieser Stelle nicht eingehen, da es sich um ein eigenes Forschungsgebiet mit einer langen Tradition handelt. Interessierte Leser mögen sich etwa in [Zie72] näher darüber kundig machen. Auf jeden Fall kann man sich an dieser Stelle den Graphen wie eine Herde von N Igel vorstellen, wobei Igel Nr. i genau k_i Stacheln hat, die die Enden der Kanten sind. Dann werden Knotenpaare des Graphen (Paare von Igelstacheln) zufällig ausgewählt und miteinander verbunden.

Hierfür gibt es ausgefeilte Verfahren, die auf Monte Carlo Methoden (siehe Kapitel 12) oder auf genetischen Algorithmen basieren. Die Betrachtung der letzteren würde hier aber ebenfalls zu weit wegführen. Ich verweise für eine genauere Lektüre auf [MKI+04].

8.2 *Poisson-Verteilung*

Der „klassische“ Algorithmus zur Erzeugung dieser Verteilung ist schon seit den Arbeiten von Erdős und Rényi [ER59], [ER60] bekannt. Die einfachste seiner Varianten geht wie folgt:

- Es werden N Knoten festgelegt.
- Zwischen je zwei zufällig gewählten Knoten werden Kanten erzeugt, bis die vorgegebene Kantenzahl M erzeugt ist.

Die solcherart erzeugten Zufallsgraphen gehorchen der Poisson-Verteilung. Sie reproduzieren den small-world-Effekt, aber bezüglich fast aller anderer Charakteristika spiegeln sie das Verhalten tatsächlich existierender Netze nur schlecht wider (vgl. [New03]).

8.3 Exponential-Verteilung

Bei dieser Verteilung sind im Gegensatz zum letzten geschilderten Verfahren die Knoten nicht von Anfang festgelegt, sondern man startet mit einem einzigen Knoten. Dann werden die beiden folgenden Schritte solange durchgeführt, bis die vorher festgelegte Zahl von Knoten und Kanten erreicht ist:

- Füge eine zufällig gewählte Kante zwischen zwei schon existierenden Knoten ein (nicht notwendigerweise verschiedene Knoten).
- Füge dem Graphen einen weiteren Knoten hinzu.

Bei dieser Verteilung werden offensichtlich früher erzeugte Knoten gegenüber später erzeugten bevorzugt. Das Ergebnis ist eine Exponentialverteilung (vgl. [DM03]).

8.4 Wer hat, dem wird gegeben (Pareto-Verteilungen)

Die erste Untersuchung eines Netzes mit einer Pareto-Gradverteilung wurde in [Pri65] beschrieben. Es handelt sich um eine Untersuchung über das wissenschaftliche Zitieren. Die Knoten stellen wissenschaftliche Artikel dar. Zwei Knoten k_1 und k_2 sind verbunden, wenn k_1 in k_2 zitiert wird. An dieser Stelle soll nicht darauf eingegangen werden, daß hier ein gerichteter Graph vor uns liegt, da die Zitier-Relation nicht symmetrisch ist, was zu mathematisch wichtigen, für uns aber marginalen Komplikationen führt. In [Pri76] hat Price eine noch heute allgemein akzeptierte theoretische Erklärung gegeben: den sogenannten „Matthäus-Effekt“ („Denn wer hat, dem wird gegeben ...“, Matthäus 25:29 in [Bib]).

Was damit gemeint ist, das läßt sich am Beispiel des Ziternetzwerkes gut erläutern: Im Zweifelsfall würde man lieber einen berühmten Autor zitieren als einen unbekanntes; und deshalb bekommt ein berühmter, also ein schon oft zitierter Autor leichter neue Zitate hinzu als einer, den sowieso niemand kennt.

8.5 Die Anwendung des Matthäus-Effekts

Damit das Neue der Situation klar wird und weil wir hier einen zentralen Punkt der Beschäftigung mit Zufallsnetzwerken vor uns haben, vergleiche ich die Situation des Ziternetzwerkes mit den Netzen der Unterkapitel 8.2 und 8.3: Dort war die Wahrscheinlichkeit, daß ein fester Knoten k eine zusätzliche Verbindung bekommt, unabhängig von irgendwelchen Eigenschaften von k . Diese Wahrscheinlichkeit sei mit $p(k)$ bezeichnet. Dann ist, falls die Gesamtzahl der Knoten mit N bezeichnet wird,

$$p(k) = 1 / N$$

oder in der etwas handlicheren Beschreibung durch Proportionalität

$$p(k) \propto 1. \quad ^{50}$$

Was für einen dem eben skizzierten Ansatz als Zitiernetzwerk analogen Ansatz also nötig ist, das ist eine Art Vorrangfunktion $pref$, die die Wahrscheinlichkeit, daß ein Knoten k eine neue Kante bekommt, also $p(k)$, abhängig macht von dem Grad p_k von k . Deshalb wird diese Form der Erzeugung von Zufallsnetzwerken nach [BA99] auch „preferential linking“ oder „preferential attachment“ genannt. Eine offensichtlich mögliche Wahl einer Vorrangfunktion wäre die Gradfunktion selber. In der Beschreibung durch Proportionalität lautet der Ansatz dann

$$p(k) \propto p_k$$

Der Proportionalitätsfaktor S bestimmt sich dann durch

$$S = 1 / \sum_k p_k$$

Damit hinge die Wahrscheinlichkeit, eine neue Kante zu bekommen, also direkt von der Zahl der schon vorhandenen Kanten ab. Dieser einfache Ansatz hat einen Nachteil: Wenn die Wahrscheinlichkeit, eine neue Kante zu erhalten, proportional zu den schon vorhandenen Kanten ist, kann ein Knoten, der gar keine Kante besitzt, auch nie eine zusätzliche Kante bekommen. Die Wahrscheinlichkeit, eine neue Kante zu erhalten, wäre in diesem Falle ja gerade 0. Deshalb wird eine Konstante $A > 0$ gewählt, so daß die Wahrscheinlichkeit für eine neue Kante eines Knoten k proportional zu $p_k + A$ ist, also

$$p(k) \propto p_k + A$$

Damit ist $p(k)$ immer noch affin linear⁵¹ von dem bisherigen Grad abhängig. Für die Kantenverteilung bei Ende des Verfahrens gilt dann (was nicht völlig trivial zu beweisen ist)

$$p_k \propto k^{-(2+A/2)}$$

In [DM03], S.28f., kann ein Beweis⁵² dafür nachgelesen werden, daß alle Verfahren, bei dem die Wahrscheinlichkeit für eine neue Kante eines Knotens (affin) linear von dem bisherigen Grad dieses Knoten abhängt, zu Pareto-Verteilungen führen.

Ein Verfahren, das zu verhältnismäßig realistischen Verteilungen im Vergleich mit tatsächlich existierenden Netzwerken führt, wird in [BA99], [AJB00] vorgestellt:

Das Modell beginnt mit m_0 Knoten. Zu jedem Zeitpunkt t wird ein neuer Knoten erzeugt und mit m schon vorhandenen Knoten verbunden. Die Wahrscheinlichkeit p_i , daß der neue Knoten mit dem schon vorhandenen Knoten i verbunden wird, hängt von dessen Grad k_i ab. Es gilt

⁵⁰ In dieser Form kann in der Mathematik eine konstante Funktion ausgedrückt werden. Die Gleichung darüber ist insofern präziser, als dort auch gleich der Proportionalitätsfaktor $1/N$ sichtbar ist.

⁵¹ Eine Funktion f heißt affin linear, wenn sie durch eine Gleichung der Form

$$f(x) = ax + b$$

beschrieben wird.

⁵² Der dort angegebene Beweis enthält an zwei Stellen leider Fehler. Die lassen sich jedoch mit etwas Aufwand reparieren.

$$p(i) = k_i / \sum_j k_j$$

Für hinreichend große t folgt die Verteilung der Grade einem Paretogesetz der Form

$$p_k = 2m^2 / k^3.$$

Ein offensichtlicher Nachteil dieses Modells ist es, daß dieses Modell nur Netzwerke erzeugt, deren Gradverteilungen mit k^{-3} fallen. Genauere Untersuchungen ([ACL01]) haben gezeigt, daß die so erzeugten Netze außerdem zu baumähnlich sind im Vergleich zu realistischen Netzen. Deshalb wird dort eine Verallgemeinerung vorgeschlagen:

Es wird immer noch in jedem Schritt genau ein Knoten neu erzeugt, aber es werden nicht genau m Kanten zu schon vorhandenen Knoten konstruiert. Vielmehr werden für den gesamten Konstruktionsprozeß drei Verteilungen V_1, V_2, V_3 fest gewählt, gemäß denen bei jedem Schritt drei Zahlen m_1, m_2, m_3 zufällig bestimmt werden. Nach der Erzeugung eines neuen Knotens werden bei jedem Schritt m_1 Kanten innerhalb des alten Netzes, m_2 Kanten vom neuen Knoten zu Knoten im alten Netz mit preferential linking nach den Regeln in [BA99], [AJB00] und m_3 Kanten des neuen Knoten zu sich selbst konstruiert.

Die in diesem sogenannten „ACL-Modell“ entstehenden Netze sind nach vielen der oben aufgeführten Kriterien realistischen Netzwerken sehr ähnlich, ferner kann der Zusammenhang des Netzes erzwungen werden und auch viele mathematische Eigenschaften bleiben erhalten. Allerdings klappt das Verfahren nur bei wachsenden Netzen. Inwieweit alle Charakteristika auch erhalten bleiben, wenn die Streichung von Knoten erlaubt wird, das ist gegenwärtig eine offene Frage. Einige Ergebnisse, die in anderem Zusammenhang in [KKR99] erzielt werden konnten, deuten aber darauf hin.

Ich möchte den Abschnitt über das preferential linking nicht beenden, ohne kurz eine Erklärung des Zipfschen Gesetzes zu zitieren, die auf den Nobelpreisträger für Wirtschaftswissenschaften von 1978 Herbert A. Simon zurückgeht. Schon in [Sim55] stellt er ein Modell zur Erzeugung von Netzwerken vor:

Zu jedem Zeitpunkt wird ein neuer Knoten hinzugefügt.

- Mit Wahrscheinlichkeit p etabliert dieses neue Individuum eine neue Familie;
- Im komplementären Fall wird zufällig eines der alten Individuen gewählt, und das neue Individuum schließt sich dessen Familie an.

Jede Familie entspricht einem Wort einer Sprache und jeder Knoten einem Auftreten eines Wortes. Das Erstellen eines Textes entspricht der Konstruktion eines Netzwerkes. Hier haben wir offenbar einen Fall von preferential linking vor uns: Ein Wort, das wir in der Vergangenheit häufiger verwandt haben, werden wir auch in Zukunft häufiger verwenden. Es kann dann gezeigt werden (für einen eleganten Beweis siehe [DM03], p.213), daß die Gradverteilung einem Paretogesetz gehorcht:

$$P_k \propto k^{-\gamma} \quad \text{mit} \\ \gamma = 1 + 1/(1-p)$$

Da wir im allgemeinen immer wieder dieselben Wörter benutzen, ist p , die Wahrscheinlichkeit ein völlig neues Wort zu verwenden, sehr klein. Damit ist ziemlich genau $\gamma=2$. Das stimmt auch mit Zählungen zum Zipfschen Gesetz überein.

8.6 Noch einmal das Piotrovskij-Gesetz

Die übergroße Mehrheit der natürlich vorkommenden Netzwerke gehorcht einer Paretoverteilung. Laut Barabasi und Albert, mit deren Studien, etwa [BA99], [BAJ99], die Untersuchung von Netzen mit solchen Verteilungen begann, liegt der Grund dafür in den beiden oben beschriebenen Mechanismen:

- Die Anzahl der Knoten ist nicht von Anfang an festgelegt, sondern das Netzwerk wird nach und nach erzeugt.
- Die sich entwickelnden Kanten gehorchen dem preferential linking.

Wenn dem Wachstum eine lineare Vorrangfunktion zu Grunde liegt, ist, wie man sich sehr leicht überlegen kann, der durchschnittliche Knotengrad unabhängig von der Größe des Netzwerkes.

Wegen deren großer Bedeutung seien an dieser Stelle einige Bemerkungen über Netzwerke mit einer Paretoverteilung der Grade gemacht:

Mit unseren Betrachtungen über Zufallsnetzwerke sind wir jetzt in der Lage, den Bogen von Netzwerken zum Piotrovskij-Gesetz zu schlagen: Dieses Gesetz besagt ja, daß zumindest initial das Wachstum einer bestimmten sprachlichen Innovation exponentiell vor sich geht. Wir haben herausgearbeitet, daß ein solches Wachstum in isotropen zellularen Automaten nicht möglich ist (Unterkapitel 6.1), wohl aber in Bäumen (Unterkapitel 6.4). Eine Eigenschaft von Zufallsnetzwerken, siehe [DM03], S.17, führt uns jetzt weiter:

Wenn ein Zufallsnetzwerk mit Pareto-Verteilung hinreichend groß ist, hat es für gewöhnlich lokal eine baumähnliche Struktur.

Dabei handelt es sich um eine statistische Aussage, die in ihrer strengen, hier von mir unterschlagenen Form die Begriffe „hinreichend groß“, „für gewöhnlich“ und „baumähnlich“ genauer quantifiziert. Ein Beweis für diese seit 2002 bekannte Aussage kann für eine Klasse von Sonderfällen in [DM03], S.161ff., gefunden werden.

Wenn eine Innovation sich anfangs lokal ausbreitet, kann sie sich mithin zunächst, wie in Bäumen üblich, exponentiell schnell fortpflanzen, bevor in größerer Entfernung die baumartige Struktur verloren geht. Damit kann die Innovation später auf Grund ihres eigenen Erfolges auf schon markierte Knoten treffen, was die Geschwindigkeit der weiteren Ausbreitung behindert. Natürlich kann diese Behinderung in Bäumen nicht vorkommen, und so herrscht dort ungebremstes exponentielles Wachstum. Diese Überlegungen zeigen also, daß für irreversiblen Sprachwandel auch bei netzwerkbasierter Modellen ein S-förmiger

Verlauf zu erwarten ist, wie er bei den Arbeiten von Altmann durch die logistische Differentialgleichung erzwungen wird.

8.7 Zusammenhang der Sprechergemeinschaft (*Community structure*)

Die oben schon erwähnten Untersuchungen in [New03a], [NP03] haben für soziale Netze im Vergleich zu technischen Netzen ungewöhnlich hohe Clustering-Koeffizienten und eine positive Gradkorrelation nachgewiesen. Da wir uns in unserer Untersuchung mit sozialen Netzen beschäftigen, müssen wir uns mit den Gründen für dieses sogenannte „*Herdenphänomen*“ ebenso beschäftigen wie mit der Frage, wie diese Gründe in ein Modell für sprachliche Innovation eingearbeitet werden können.

Eine informelle Begründung ihrer Ergebnisse liefern die obigen Autoren dadurch, daß Individuen sozialer Netze Angehörige von im allgemeinen mehreren Gruppen sind, wobei die Angehörigen einer Gruppe potentiell vollständig miteinander fähig zur Kommunikation sind. Das Netzwerk, das dann entsteht, indem zwei Knoten genau dann miteinander verbunden werden, wenn es eine Gruppe gibt, der sie gemeinsam angehören, heißt „*community structure*“. Es zeigt alle potentiellen Verbindungen an.

Beispiel: Das folgende Beispiel aus [New03a] soll die Konstruktion erläutern: In Abbildung 27 wird die allgemeine Situation in Form eines *bipartiten Graphen*⁵³ geschildert. Die mit Zahlen bezeichneten Knoten stellen die vorhandenen Gruppen dar. Mit Buchstaben werden die Individuen bezeichnet. Eine Kante zeigt an, daß ein Individuum zu einer Gruppe gehört.

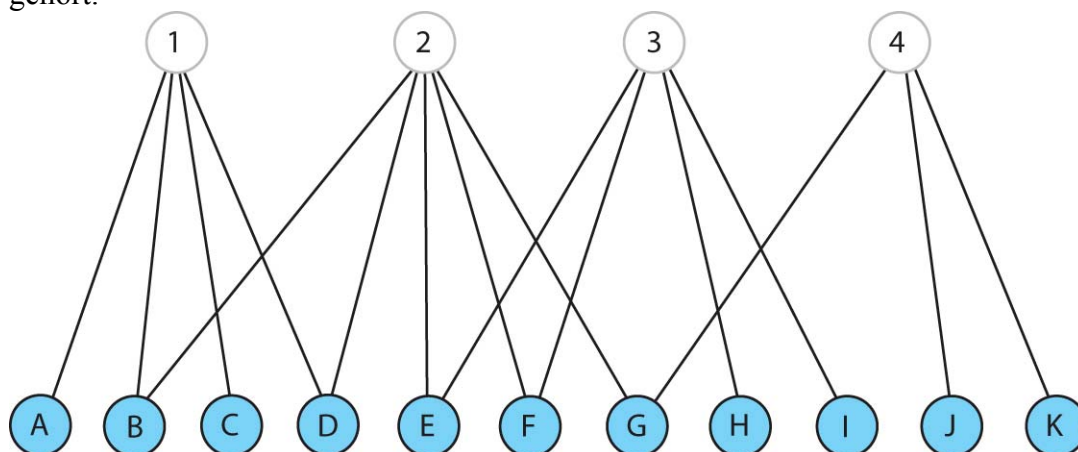


Abbildung 27: Die Community-Structure als bipartiter Graph

Innerhalb jeder Gruppe soll jedes Individuum mit jedem kommunizieren können. Dadurch entsteht aus der Struktur von Abbildung 27 das in Abbildung 28 dargestellte Netz potentieller Kommunikationen innerhalb der gesamten Gemeinschaft.

⁵³ Ein bipartiter Graph zeichnet sich dadurch aus, daß seine Knotenmenge aus zwei disjunkten Teilmengen besteht, wobei nur Kanten zwischen Elementen der verschiedenen Teilmengen erlaubt sind.

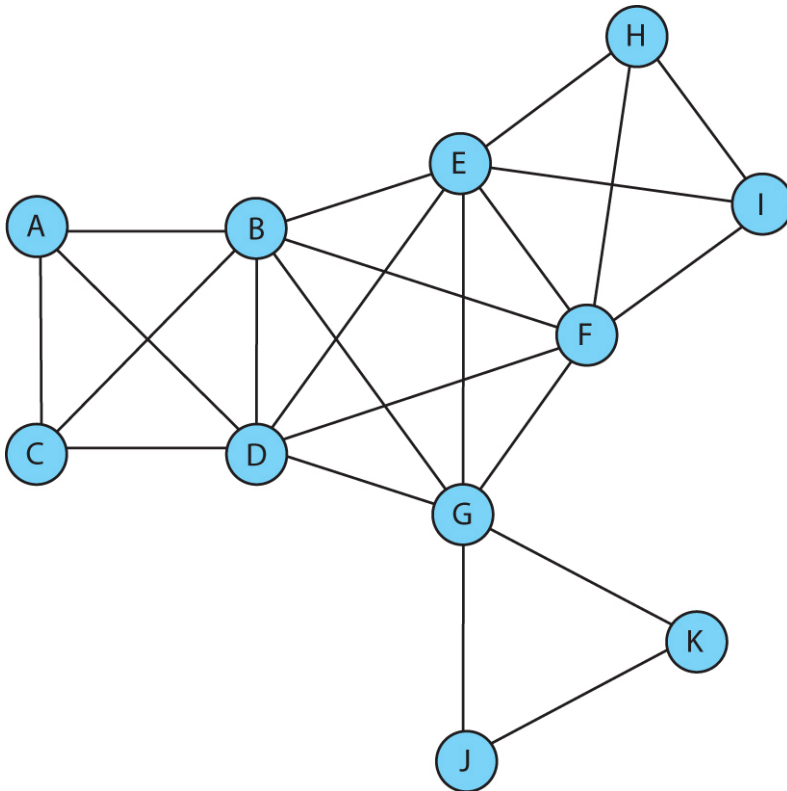


Abbildung 28: Potentielle Verbindungen in einer Community-Structure

Aus den potentiell vorhandenen Verbindungen müssen nun im Modell tatsächlich vorhandene Verbindungen durch das Streichen einiger der nur potentiell (aber nicht real) vorhandenen Verbindungen gemacht werden. Ein mögliches Ergebnis eines solchen Verfahrens zeigt Abbildung 29.

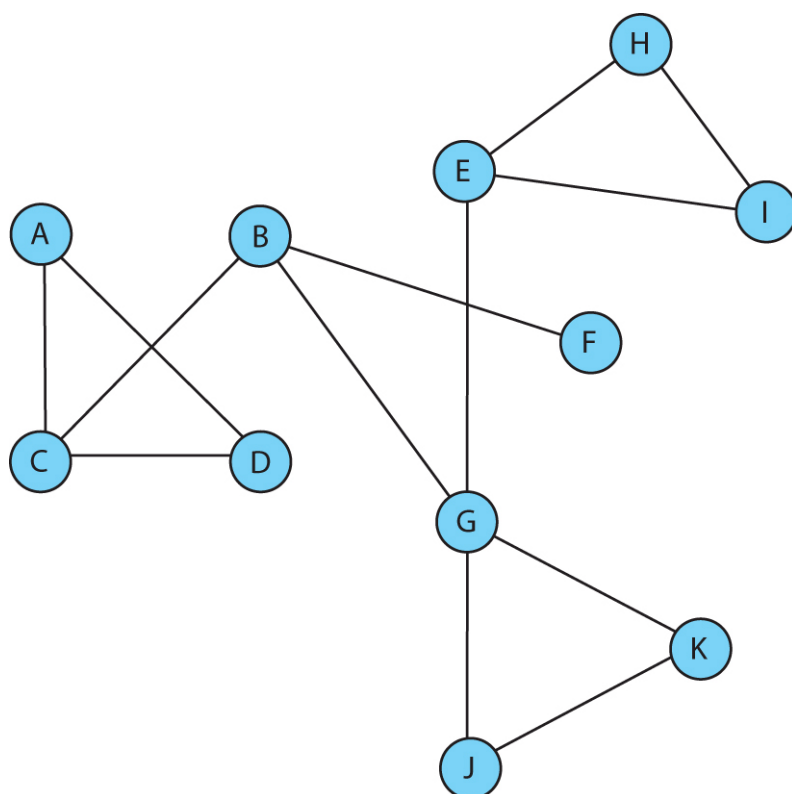


Abbildung 29: Realisierte Verbindungen in einer Community-Structure

Die Autoren in [NP03] schlagen für diesen Schritt vor, einen universellen Parameter p einzuführen und aus der potentiellen Verbindung innerhalb der community structure mit Wahrscheinlichkeit p eine echte Kommunikation im Sinne des Modells zu machen, also tatsächlich eine Kante zwischen den entsprechenden Knoten einzufügen. Dies hat meiner Meinung nach einen bedeutenden Nachteil: Das Ergebnis wird für jede Gruppe wieder eine Poisson-Verteilung liefern. Das ist aber, wie schon öfter festgestellt wurde, unrealistisch. Wir könnten community structure und preferential linking miteinander verbinden, indem neue Kanten gemäß den Prinzipien des preferential linking eingefügt werden, jedoch nur dann, wenn es die community structure erlaubt. Innerhalb jeder Gruppe ist damit die Paretoverteilung gesichert. Die experimentellen Befunde in [NP03] deuten darauf hin, daß eine solche Verteilung auch für das ganze Netzwerk gilt.

Ein neueres Modell, das zu soziologischen Zwecken entwickelt [HK02] und mathematisch genauer in [BLT05] untersucht wurde und das zusätzlich seinerseits einige Anforderungen erfüllt, die an soziologische Modelle gestellt werden, insbesondere Paretoverteilung und positive Gradkorrelation, ist die folgende Variante des ACL-Modells:

Zusätzlich zu allen Verfahren von ACL gibt es eine kleine Wahrscheinlichkeit θ , mit der, nachdem ein neuer Knoten k_1 erzeugt worden und mit einem alten Knoten k_2 verbunden worden ist, die nächste neu zu schaffende Verbindung von k_1 vorrangig mit einem zufällig zu wählenden Nachbarn von k_2 erfolgt.

Durch diese zusätzliche Regel führt also zusätzliche Transitivität in die Kommunikationsbeziehungen ein, erhöht damit den Clustering-Koeffizienten und befördert so das Herdenphänomen. Alle anderen für uns wesentlichen Eigenschaften bleiben nach [BLT05] erhalten.

9 Prozesse auf Netzwerken

Wenn wir noch einmal auf den Anfang der Untersuchungen zurückkommen, die mit Überlegungen zu zellularen Automaten begannen, so war der Ansatz ja der, daß ich von zellularen Automaten zu komplexeren Graphen übergegangen bin, weil die mögliche Ausbreitungsgeschwindigkeit von Innovationen in isotropen Netzen zu gering ist. Darum habe ich die Gittergeometrien geändert. Aus dem in Abschnitt 8.6 gemäß [DM03] zitierten Satz und den in Abschnitt 6.4 durchgeführten Abschätzungen für Bäume habe ich dann auf Seite 84 die Konsequenz gezogen, daß bei Gittergeometrien, die ein Zufallsnetzwerk mit Pareto-Verteilung darstellen, strukturell eine hinreichende Ausbreitungsgeschwindigkeit erzielt werden kann. Damit habe ich den zentralen Bestandteil von Zellularen Automaten verändert. Woran wir aber eigentlich interessiert sind, das sind nicht nur die Netzwerke selber, sondern besonders Prozesse, die auf den Netzwerken ablaufen. Beispiele für allgemeine derartige prozeßbasierte Untersuchungsgegenstände sind Fragen, wie die, auf welche Art die Netzwerktopologie Suchalgorithmen im Internet oder Strategien zur Bekämpfung von Epidemien beeinflussen kann. Leider ist der Stand der Forschung bei solchen Fragen noch beklagenswerter als bei den Untersuchungen zur statischen Netzwerktopologie und deren dynamischer Veränderung ([New03]). Der Grund liegt einfach darin, daß zum Verständnis dynamischer Prozesse auf Netzwerken zunächst das Verständnis für die statische und die dynamische Struktur der Netzwerke selbst entwickelt werden muß, und auch die letztere Theorie ist ja noch sehr neu und lückenhaft.

Es ist sicher aufgefallen, daß bei der Erzeugung von Netzwerken mit bestimmten Gradverteilungen (Exponential- und Pareto-, nicht aber Poisson-Verteilung) schon mit der Dynamik des Entstehungsprozesses argumentiert wurde. Wir haben es insgesamt also mit zwei Formen des dynamischen Geschehens zu tun:

- der Dynamik durch das Entstehen (und später in dieser Arbeit auch das Verschwinden) von Knoten,
- der Dynamik durch Zustandsänderungen in den Knoten.

Hier sehen wir, daß die eher von der Informatik geprägte Fragestellung: „Wie können wir Netzwerke mit bestimmten Gradverteilungen erzeugen?“ insofern die ursprünglichere ist im Vergleich zu den in der Wissenschaft älteren mathematischen Fragestellungen: „Welche Verteilungen gibt es? Was kann über Netzwerke mit bestimmten Verteilungen ausgesagt werden?“. Die erste Frage, die auch den dynamischen Charakter unterstreicht, beleuchtet eben auch den Gesichtspunkt der Entstehung solcher Netzwerke. Als wichtige Faktoren bei der Entstehung von Netzwerken mit bestimmten Verteilungen gelten im allgemeinen unter anderem

- das **preferential linking**, zu dem ich ja schon einiges gesagt habe,
- Alterungsprozesse, die in der einschlägigen Literatur mit **aging** bezeichnet werden. Sie werden weiter unten in 9.7 beschrieben werden.
- die **Kosten**, die zum Einfügen einer neuen Kante (in irgendeiner Form) anfallen,

- die allgemeine Kapazität eines Knotens.

Es gibt eine Reihe von (uns hier nicht interessierenden) Untersuchungen, die beschreiben, wie sich die genannten Faktoren auf die Konstruktion eines real existierenden Netzwerkes auswirken und wie man sich solche Ideen bei der Computermodellierung zunutze machen kann. Für unsere Zwecke sind Ansätze, in denen Kosten oder Kapazitäten eine Rolle spielen, viel zu komplex. Ich verzichte deshalb auf die Definitionen der Begriffe.

Für die Modellierung erhebt sich nun die Frage, ob die beiden oben angesprochenen Prozesse nacheinander ablaufen sollen, was mathematisch leichter zu behandeln wäre, oder simultan. Eine Faustregel liefern [AB02]: Eine Modellierung auf der Grundlage einer festen Topologie ist dann angemessen, wenn die Größenordnungen der Zeiten, die die Netzwerktopologie und der Zeiten die die Dynamik auf dem Netzwerk beschreiben, sich deutlich unterscheiden. In unserem Zusammenhang sind wir sogar in einer besonderen Situation: Unter Umständen laufen die Veränderungen der Topologie und die Dynamik auf dem Netzwerk nicht nur auf in etwa gleichen Zeitskalen ab. Es kann sich sogar die Topologie schneller ändern als die Netzdynamik. Das besagt, daß in unserem Zusammenhang eine simultane Modellierung beider Dynamiken unumgänglich sein wird.

Einige bisher erzielte Ergebnisse für die Dynamik auf Netzen möchte ich in diesem Kapitel noch kurz vorstellen.

9.1 Aufrechterhaltung der Kommunikation

Ein praxisrelevantes Ergebnis der Ausfalltoleranz ist etwa das von Albert und Barabási in [AB00] und [AJB00], daß Netzwerke mit einer Pareto-Verteilung der Grade sehr widerstandsfähig gegen den zufälligen Ausfall von Knoten sind. Das heißt, auch wenn relativ viele Knoten ausfallen, zerfällt das Netz noch nicht in Teile, und eine Kommunikation zwischen allen Knoten des Netzes ist immer noch möglich, wenn auch im allgemeinen umständlicher. Die Situation ändert sich jedoch, wenn gezielt die Knoten mit hohem Kantengrad lahmgelegt werden. Etwas genauer können die Ergebnisse von Albert und Barabási so formuliert werden:

Es sei ein Graph G gegeben. Es sei f eine Funktion, die einer Zahl n , die die Anzahl zufällig ausfallender Knoten in G darstellt, den Durchmesser des Graphen ohne diese ausgefallenen Knoten zuordnet. Dann mißt f also in etwa, wie sich die Wege (genauer: der längstmögliche Weg) durch den Graphen mit der Anzahl zufällig ausfallender Knoten verlängert. Es ist offenkundig, daß eine solche Funktion nur statistisch definiert werden kann, da wir ja nicht über die ausgefallenen Knoten selber reden, sondern nur von ihrer Anzahl und dabei dann außerdem die Topologie der Graphen eine erhebliche Rolle spielt. Unter diesen Voraussetzungen können die Autoren zeigen, daß für die ursprünglichen Poisson-verteilten Graphen nach Erdős und Rényi (also mit zufällig erzeugten Kanten bei vorgegebenen Knoten) die Funktion f ein lineares Wachstum im Sinne der oben beschriebenen Komplexitätstheorie zeigt⁵⁴, daß also der Durchmesser des Graphen ungefähr genauso steigt wie die Anzahl der ausgefallenen Knoten. Bei Graphen mit Pareto-Verteilung ändert sich die Lage:

⁵⁴ Genauer gilt: $f \in o(n) \cap \Omega(n)$. Dies ist genau die Menge der Funktionen, die sich im Limes wie lineare Funktionen verhalten.

- Fallen die Knoten nur zufällig aus, bleiben f und damit der Graphdurchmesser annähernd konstant.
- Werden gezielt Knoten mit einem hohen Verknüpfungsgrad lahmgelegt, so wächst f sogar noch schneller als linear.

Dieses Ergebnis kann dadurch erklärt werden, daß bei diesen Verteilungen verhältnismäßig wenige Knoten mit dafür umso höherem Grad existieren. Bei zufälligem Ausfall werden daher mit hoher Wahrscheinlichkeit lediglich Knoten mit kleinem Grad betroffen sein, was dann recht geringen Schaden anrichtet. Bei einer gezielten Attacke können jetzt aber genau Knoten mit besonders hohem Grad als lohnendes Ziel anvisiert werden.

Es ist aber überhaupt eine allgemeine Beobachtung, daß komplexe Systeme ein überraschend hohes Maß an Widerstandskraft gegen Fehler an den Tag legen ([AJB00], [AB02]). Das mag teilweise an Redundanzen bei der Verkopplung in den Netzwerken liegen. Daß die Netzwerktopologie auch darüber hinaus eine Rolle spielt, das ist heute schon klar. Einzelheiten sind zur Zeit jedoch noch offen.

Modelle, die extrem dicht an unseren Forschungsgegenstand sind, sind aus der Epidemiologie bekannt, wobei man (innerhalb eines Modells) davon ausgehen kann, daß sich sprachliche Neuerungen in der Sprachgemeinschaft ähnlich wie Epidemien ausbreiten. Die Idee, etwa die Ausbreitung von Lehnwörtern als Epidemie zu untersuchen, kann schon bei Altmann nachgelesen werden.

9.2 Sprachliche Neuerungen als epidemiologische Prozesse

Der Vergleich einer sprachlichen Neuerung mit einer Epidemie, die die Sprache überschwemmt, liegt nahe und ist vielfach gezogen worden, siehe etwa [BK83], [Köh86]. Auch das Piotrovskij-Gesetz kann als epidemiologischer Ansatz interpretiert werden, auch wenn es ursprünglich nicht als solcher konzipiert worden ist. Ein Ansatz, bei dem epidemiologische Überlegungen auf genau den Untersuchungsgegenstand übertragen werden, mit dessen Hilfe ich mein noch genauer zu entwickelndes Sprachwandelmodell durch reale Befunde zu stützen gedenke, die Übernahme deutscher Lehnwörter ins Polnische, ist [Hen95]. Jedoch bestehen bedeutende Unterschiede zwischen meiner Auffassung dessen, was ich als Epidemie ansehe, und dem, was dort untersucht wird. Bei Hentschel (und andernorts auch bei anderen Autoren) wird die *Gesamtmenge* der deutschen Wörter als „Seuche“ aufgefaßt (natürlich nur modellhaft, nicht im ideologischen Sinne). Jedes neue Lehnwort ist – in dieser Metaphorik – ein neuer Krankheitsfall. Dagegen stellt in meinem Modell *jedes einzelne Wort* eine sich ausbreitende Änderung dar. Jede neue Anwendung eines Lehnworts ist ein neuer Krankheitsfall. Der Hentschelsche Ansatz würde sich bei mir also eine tausendfache Epidemie darstellen. Es könnte ein lohnender Versuch sein, die beiden Ansätze miteinander zu verheiraten, also eine mehrfache Infektion im Dynamischen Sprachnetz zu untersuchen. Die wissenschaftliche Untersuchung mehrfacher Epidemien würde aber sicher sehr tief in bisher unerforschte Gefilde führen. Zumindest würden sicher umfangreiche Erweiterungen nötig. Ich gebe nur beispielhaft einen Punkt zu bedenken: Wenn untersucht werden soll, wie die Zu- und Abnahme der Anzahl übernommener fremder Wörter modelliert werden kann, so könnten sich für ein solches umfassenderes Modell außer statistischen

Phänomenen auch Ergebnisse der Sprachkontaktforschung oder soziolinguistische Faktoren bezüglich der Interaktion verschiedener Lehnwörter als relevant erweisen, die bisher noch gar nicht modelliert werden können.

Im allgemeinen wird in der Epidemiologie zwischen *epidemischen* Krankheiten und *endemischen* Krankheiten unterschieden. Die ersteren sind solche, die sich schnell ausbreiten, für einen kurzen Zeitraum einen beträchtlichen Teil der Bevölkerung befallen und dann wieder verlöschen, zum Beispiel die Grippe. Eine endemische Krankheit setzt sich in der Bevölkerung fest und infiziert dauerhaft einen Teil der Bevölkerung, zum Beispiel die Masern. Wir werden uns im folgenden mit der Frage beschäftigen, ob sprachliche Neuerungen eher dem epidemischen oder dem endemischen Typ angehören, und mit der Frage, wie sich die beiden Typen hinsichtlich ihrer Modellierung in Netzen unterscheiden.

Das traditionelle Modell zur Beschreibung von Epidemien ist das *SIR-Modell*:

Definition: Das *SIR-Modell* ist ein Netzwerkmodell, bei dem jeder Knoten zu genau einer von drei Klassen gehört, bei dem also die Bevölkerung in drei Klassen eingeteilt wird:

- S (*susceptible*) ist jemand, der die Krankheit nicht hat, der sie aber bekommen kann.
- I (*infected*) ist jemand, der die Krankheit hat und sie weiter verbreiten kann.
- R (*recovered* oder *removed*) ist jemand, der die Krankheit überwunden hat, immun ist und sie auch nicht weiter verbreiten kann. Der Zustand kann auch bedeuten, daß das entsprechende Individuum gestorben ist. Ob diese zweite Interpretation anwendbar ist, das hängt vom der Art der Untersuchung ab. Die Idee ist in beiden Fällen, daß ein Individuum im Zustand R die Epidemie nicht weiter verbreiten kann.

Frühere auf analytischer Mathematik beruhende Überlegungen können durch offensichtliche Änderungen auf Netzwerke übertragen werden, indem man jetzt Zellen mit einer dreielementigen Zustandsmenge benutzt. Eine Rechtfertigung für solche Ansätze, in denen mit Zufallsgraphen gearbeitet wird, liefern [NSW01] etwa dadurch, daß soziale Kontakte eben zufällig geschehen.

Als mögliche Gründe für den Schwund eines Lehnwortes könnten ein Verschwinden des Denotats oder eine Verdrängung durch andere Lehn- oder Eigenwörter in Frage kommen. Eine genauere Klassifikation von Varianten eines solchen Ausscheidens kann in [HM01] gefunden werden.

Ein Beispiel für die praktische Anwendung der hier sehr knapp skizzierten Überlegungen sei hier gegeben: Eine Konsequenz des oben zitierten Ergebnisses von Barabási und Albert über den Ausfall von Knoten in Netzwerken ergibt beispielsweise eine Impfstrategie: Es ist effektiver, Menschen mit einer hohen Zahl von Kontakten gezielt zu impfen als den Impfschutz zufällig zu verbreiten.

Auch in unserem Zusammenhang kann eine testbare Hypothese abgeleitet werden: Da Innovationen als Störungen aufgefaßt werden können und unsere Netzwerke immer einer Pareto-Verteilung der Grade gehorchen, sollte gelten:

*Durchgängig erfolgreiche sprachliche Innovationen beginnen in der Regel bei den Individuen, die eine hohe Vernetzung (also besonders vielfältige Kommunikation) aufweisen.*⁵⁵

Eine Innovation, die nicht so beginnt, verhält sich ja eher wie eine zufällige Störung. Eine zufällige Störung kann sich zwar möglicherweise ebenfalls durchsetzen, es wird aber wesentlich schwieriger werden. Im Rahmen der vorliegenden Untersuchung soll die eben geäußerte Hypothese aber nicht weiter verfolgt werden. Es gibt aber erste Hinweise, die tatsächlich in diese Richtung deuten ([Lab01]).

Das SIR-Modell wird vorwiegend bei epidemisch verlaufenden Krankheiten angewandt. Für endemische Krankheiten hat sich ein einfacheres Modell bewährt, das *SIS-Modell*, bei dem der R-Zustand fehlt.

Definition: Das *SIS-Modell* ist ein Netzwerkmodell, bei dem jeder Knoten zu genau einer von zwei Klassen gehört, bei dem also die Bevölkerung in zwei Klassen eingeteilt wird:

- S (*susceptible*) ist jemand, der die Krankheit nicht hat, der sie aber bekommen kann.
- I (*infected*) ist jemand, der die Krankheit hat und sie weiter verbreiten kann.

Diese Modellierung ist natürlich auch bei den Krankheiten naheliegend, die nicht zu einer Immunisierung derjenigen ehemaligen Träger führen, die die Krankheit überwunden haben. In diesem Zusammenhang sind besonders die Untersuchungen von Pastor-Satorras und Vespignani (siehe Literaturliste) zu nennen.

9.3 Momente und erzeugende Funktion

Bevor ich näher auf die Details der für uns nützlichen Ergebnisse der Epidemiologie eingehe, schildere ich noch einige altbekannte Hilfsmittel aus der Statistik:

Bei *Momenten* handelt es sich um Durchschnittswerte für Verteilungen, die wie folgt definiert sind: Das *n-te Moment* einer diskreten Verteilung p_0, p_1, \dots ist

$$\langle k^n \rangle = \sum_k k^n p_k$$

Dabei ist im Falle der Netze p_k die Wahrscheinlichkeit, daß ein Knoten genau k Kanten hat. Es sind also $\langle k \rangle$ die durchschnittliche Kantenzahl im Netzwerke und $\langle k^2 \rangle$ die durchschnittliche Zahl der Kantenquadrate usw. .

Alle Momente können durch einfache mathematische Operationen aus der „erzeugenden Funktion“ gewonnen werden:

⁵⁵ Diese Aussage gilt ausdrücklich nur für durchgängig erfolgreiche Neuerungen. Wird dagegen etwa eine Neuerung in einer Fachsprache betrachtet, so wird diese sich nicht von den global aktivsten Sprechern aus ausbreiten. Aber es könnten die Sprecher dieser speziellen Sprache als isolierter Graph modelliert werden. Dort sollte die oben geäußerte Vermutung wieder gelten.

Definition: Sei p_0, p_1, \dots eine diskrete Verteilung. Dann sei die *erzeugende Funktion* definiert durch

$$G_0(x) = \sum_k p_k x^k.$$

Aus der erzeugenden Funktion lassen sich viele wichtige Größen und Gleichungen ableiten:

$$G_0(1) = \sum_k p_k = 1$$

(Die Summe aller Wahrscheinlichkeiten ist 1.), außerdem

$$p_k = 1/k! \left. d^k G_0 / dx^k \right|_{x=0}$$

(k -fache Differentiation von G_0 , Division durch $k!$ und Auswertung an der Stelle 0), ferner

$$\langle k^n \rangle = \left((x d/dx)^n G_0(x) \right) \Big|_{x=1}$$

(n -fache Differentiation und Multiplikation mit x sowie Auswertung an der Stelle 1) für das n -te Moment.

Ferner sei

$$G_1(x) = G_0'(x) / G_0'(1).$$

9.4 Die Epidemische Schwelle oder: Wieso gewinnt nicht jede sprachliche Neuerung?

Der Begriff der „*Epidemischen Schwelle*“ („*epidemic threshold*“) ist grundlegend in der klassischen Epidemiologie. Dabei geht es um die Frage, wann, in Abhängigkeit von gewissen Parametern, eine Krankheit sich ausbreitet, bis sie endemisch wird und wann sie von selber zurückgeht und schließlich ausstirbt.⁵⁶ Beim einfachsten klassischen SIS-Modell können folgende Parameter angenommen werden:

Jeder gesunde Knoten (Zustand S) wechselt seinen Zustand mit der Rate ν , wenn er mindestens einen infizierten Nachbarn hat.

Jeder infizierte Knoten (Zustand I) wird mit der Rate δ geheilt.

⁵⁶ Aus systemtheoretischer Sicht handelt es sich hier um die Suche nach „*stationären Zuständen*“, also nach Zuständen, die die Eigenschaft haben, daß sie durch die Vorgänge, die im nächsten Zeitschritt ablaufen, nicht mehr verändert werden. Mathematisch werden diese Zustände „*Fixpunkte*“ genannt. Offensichtliche stationäre Zustände sind die, bei denen entweder keine Individuen infiziert sind oder alle. Eine wichtige Frage ist dann, ob es weitere stationäre Zustände geben kann. Das könnte zum Beispiel der Fall sein, wenn (in etwa) gleich viele Menschen krank werden wie gesunden.

Eine andere Frage ist die, wohin sich das System schließlich entwickeln wird. Dem entspricht mathematisch eine Limesbildung für $t \rightarrow \infty$. Der Übergang von einer Parameterkombination, die im Limes einen stationären Zustand z_1 erreicht, zu einer Parameterkombination, die im Limes einen anderen stationären Zustand z_2 erreicht, wird auch als „*Phasenübergang*“ bezeichnet, siehe Unterkapitel 2.3. Die Frage, wann ein Phasenübergang stattfindet, ist von höchster Relevanz.

Die effektive Ausbreitungsrate ist dann $\lambda = \nu / \delta$. In der klassischen nur auf globale Effekte ausgerichteten Epidemiologie ist das eine recht simple Ableitung mit Hilfe von Differentialgleichungen. Seien t der Zeitparameter und ρ der Anteil der infizierten Individuen. Für $t \rightarrow \infty$ gilt dann

$$\begin{aligned} \rho &\rightarrow 0 \text{ für } \lambda < 1 \text{ und} \\ \rho &\rightarrow 1 \text{ für } \lambda > 1. \end{aligned}$$

Dieses λ wird dann als die Epidemische Schwelle bezeichnet. Dort findet also ein ganz typischer Phasenübergang statt. Diese mit klassischer Mathematik erzielten Überlegungen und Ergebnisse lassen sich auch auf Netzwerke mit isotroper Gittergeometrie übertragen – immerhin sollte es sich bei den Netzwerk Betrachtungen ja auch möglichst um eine konservative Erweiterung der Theorie handeln. Für eine Darstellung siehe [DM03].

Bei allgemeinen Netzwerken verändert sich die Lage. Für unkorrelierte Netze konnte in [PV00], [PV01] die Epidemische Schwelle bestimmt werden durch

$$\lambda = \langle k \rangle / \langle k^2 \rangle,$$

also den Quotienten aus dem ersten und dem zweiten Moment. Da man sich hier besonders für große Netzwerke interessiert, ist für diese Untersuchungen das Verhalten für $k \rightarrow \infty$ wichtig. Zur Erinnerung: Wir beschäftigen uns hier mit Netzwerken, bei denen den p_k eine Pareto-Verteilung der Form

$$p_k \propto k^{-\gamma}$$

unterliegt. Bekanntermaßen divergiert $\langle k^2 \rangle = c \sum_k k^2 k^{-\gamma}$ für alle $\gamma \leq 3$. (Dabei ist c eine mathematische Konstante.) Für unkorrelierte Netze heißt das in diesem Falle:

Bei Netzen mit einer Pareto-Verteilung und einem Exponenten $\gamma \leq 3$ gibt es keine Epidemische Schwelle.

Mit Hilfe von Methoden der Matrixalgebra konnte dieses Ergebnis in [BPV03] auch auf korrelierte Netze ausgedehnt werden. Wie ich schon in Unterkapitel 7.3 bemerkt habe, liegt γ in fast allen natürlich vorkommenden Netzen genau in diesem Bereich. Ferner konnten die Autoren in [BPV03] die Abwesenheit einer Epidemischen Schwelle auch dann für die meisten Netzwerke nachweisen, wenn das SIR-Modell benutzt wird. Dieser insgesamt völlig überraschende Befund führt zu der spannenden Frage:

„Wieso sind wir nicht alle tot?“

Die Abwesenheit einer Epidemischen Schwelle müßte doch eigentlich dazu führen, daß jede Infektion sofort endemisch wird, und im Zusammenhang mit sprachlichen Innovationen wäre die Konsequenz, daß jede Innovation sich wegen des small-world Effekts buschfeuerartig ausbreitet und dann permanent wird. Eine langfristige sprachliche Kommunikation wäre dann überhaupt nicht mehr möglich.

Laut [DM03] ist der entscheidende Punkt der, daß es unstatthaft ist, $\langle k^2 \rangle$ (und verwandte Größen) auf Divergenz zu untersuchen, da man immer zwar potentiell unendliche aber praktisch immer nur endliche Netzwerke vor sich hat, und da ist die Konvergenz der

Parametersummen trivial. Tendenziell bleiben die Beobachtungen etwa über eine dramatisch sinkende Epidemische Schwelle oder die Störungsresistenz / -anfälligkeit aber richtig, insbesondere, wenn wir sehr große Netzwerke vor uns haben.

9.5 Analytische Ergebnisse

Ich habe ja schon wiederholt erwähnt, daß analytische Ergebnisse nur schwer und zumeist unter unrealistischen Vereinfachungen erreichbar sind. Trotzdem möchte ich einige wesentliche analytische Resultate der Epidemiologie, zumeist nach [New02], hier vorstellen:

Die Grundannahme, die Newman macht, ist die, daß ein SIR-Modell vorliegt. Es seien Familien unabhängiger gleichverteilter Zufallsvariablen gegeben

r_{ij} Dieses gebe die Rate an, mit der in einem Zeitschritt eine Krankheit vom Individuum i auf das Individuum j übertragen wird.

t_i Dieses gebe die Zeit an, für die das Individuum i infiziert bleibt, sobald es infiziert wurde.

Mit Hilfe von Gegenwahrscheinlichkeiten läßt sich dann leicht die Wahrscheinlichkeit einer Infektion des Individuums j durch das Individuum i berechnen:

$$T_{ij} = 1 - (1 - r_{ij})^{t_i}$$

Man beachte, daß im einfachen SIR-Modell nach dem Überstehen der Infektion Immunität herrscht, und daher nach t_i Zeitschritten eine weitere Übertragung nicht mehr möglich ist. Der Durchschnitt über alle i und j ergibt die durchschnittliche Infektionswahrscheinlichkeit T (engl. „*transmissibility*“). Die genaue Form der Berechnungsformel spielt für uns hier aber keine Rolle. Es sei nur daran erinnert, daß, wie es für alle Wahrscheinlichkeiten der Fall ist, auch für T gilt:

$$0 \leq T \leq 1$$

Newman definierte ferner zwei Hilfsfunktionen, deren genaue Bedeutung hier aber ebenso keine Rolle spielt:

$$\begin{aligned} G_0(x;T) &= G_0(1+(x-1)T), \\ G_1(x;T) &= G_1(1+(x-1)T). \end{aligned}$$

Nach diesen Vorüberlegungen können einige Ergebnisse erzielt werden:

Sei $P_s(T)$ die Verteilung der Größen s von Krankheitsausbrüchen auf dem Netzwerk. Für diese Verteilung ist die erzeugende Funktion definiert durch

$$H_0(x;T) = \sum_s P_s(T) x^s.$$

Mit deren Hilfe kann die durchschnittliche Größe $\langle s \rangle$ eines Krankheitsausbruches errechnet werden zu

$$\langle s \rangle = 1 + G_0'(1;T)/(1 - G_1'(1;T)) = 1 + T G_0'(1)/(1 - T G_1'(1))$$

Dieser Ausdruck geht genau dann gegen unendlich, wenn $T G_1'(1) = 1$ gilt. Damit kann die „critical transmissibility“ T_c , also die Größe, bei der sich eine Epidemie über das ganze Netz ausbreitet, berechnet werden zu

$$T_c = 1 / G_1'(1) = G_0'(1) / G_0''(1) = \sum_k k p_k / \sum_k k(k-1) p_k$$

Die Bedeutung dieser Gleichung liegt darin, daß es unter den gegebenen Voraussetzungen einen Ausdruck gibt, mit dessen Hilfe sich bestimmen läßt, wann eine Epidemie sich überall hin ausbreitet – ein Ergebnis, das sich sicher auch auf die Ausbreitung sprachlicher Innovationen übertragen läßt, das aber vorwiegend nur von theoretischem Interesse sein dürfte.

Die vorhergehenden Ausführungen sollen nur beispielhaft zeigen, daß man mit analytischer Mathematik erstaunliche Ergebnisse erzielen kann, daß aber solche Methoden jetzt und sicher auch in absehbarer Zeit nur in Ausnahmefällen reichen werden, um Fragestellungen einer realistischen Komplexität mit der für diese Mathematik charakteristischen Exaktheit zu beantworten. Da es sich bei der vorliegenden Abhandlung um eine linguistische und nicht um eine mathematische Arbeit handelt, sehe ich von der Entwicklung neuartiger analytischer Ansätze, die uns ohnehin für die eigentliche Frage höchstwahrscheinlich nicht weit genug geführt hätten, ab und tue das, was die Mathematiker normalerweise in einer solchen Situation ebenfalls tun: Ich werde weniger exakte Lösungen anstreben, die aber für das eigentliche Problem ausreichend sind.

9.6 Änderungsregeln

Wann und wie soll sich der Zustand eines Knotens in Abhängigkeit von seinen Nachbarn ändern? Das war bei zellularen Automaten der zentrale Punkt, und er bleibt auch hier äußerst relevant. Auch in anderen Zusammenhängen als dem uns hier für sprachliche Veränderungen interessierenden sind schon früher einige Änderungsregeln benutzt worden. Ich führe zwei davon hier auf, besonders um deutlich zu machen, welches Erweiterungspotential noch in dem hier entwickelten Ansatz liegt:

9.6.1 Die Allgemeine einfache Mehrheitsregel

(AEMR, im Original: *General Simple Majority Rule*).

Diese Regel verallgemeinert den intuitiven Ansatz, den wir schon bei zellularen Automaten kennengelernt haben, nämlich daß der Anteil der Nachbarn, die sich in einem bestimmten Zustand befinden, den Nachfolgezustand des gerade betrachteten Knotens festlegen. So werde auch ich verfahren.

Die ursprüngliche deterministische Regel aus [WW95] wurde von Ferrer y Cancho dahingehend verändert, daß der Übergang stochastisch erfolgt. Genauer gesagt: Es gibt eine

mit einem Parameter β versehene Menge von Übergangsregeln. Seien k die Gesamtzahl der Nachbarn eines Knoten n und k die Anzahl der Nachbarn von n im komplementären Zustand. Dann erfolgt ein Zustandswechsel mit der Wahrscheinlichkeit

$$f_{\beta}(k) = 1 / (1 + e^{2\beta(2-k/k-1)})$$

Ein so kompliziertes Modell könnte für spätere Anwendungen vielleicht auch in der Linguistik nützlich sein. Aber für diese Arbeit ist die Regel doch zu detailliert.

9.6.2 Die Regel des höchsten Lohnes

(RHL, im Original: *Highest Current Reward Rule*)

Diese Regel stammt aus der Spieltheorie (beispielsweise aus [Ten96], [ST97]). Insbesondere in der zweiten Quelle wird eine sehr ausgefeilte Strategie mit mehreren Parametern und einem leistungsfähigen Gedächtnis für soziale Prozesse und für die KI-Forschung nutzbar gemacht. In einer einfachsten Variante der Regel gehen die Autoren insbesondere von einem Gedächtnis aus, das nur einen Schritt rückwärts wirkt. Das Szenario vereinfacht sich dann folgendermaßen:

Zu jedem Zeitpunkt werden Paare von Knoten ausgewählt, jeweils der erste Knoten zufällig und der zweite zufällig aus dessen Nachbarn. Diese spielen ein Spiel gegeneinander, bei dem sie je zwei mögliche Aktionen durchführen können und eine Auszahlung von $+1$ oder -1 in Abhängigkeit von ihren Aktionen erhalten, etwa $+1$, wenn sie identische Aktionen wählen und -1 im dualen Fall. Danach wird für den nächsten Zug die Aktion vorgemerkt, die in der Vergangenheit ein besseres Ergebnis gezeigt hat. Schon die Idee dieses Beispiels ist sehr komplex, und es ist kaum zu erwarten, daß für unsere speziellen Zwecke sich der Aufwand für die Realisierung einer solchen Regel und deren Ertrag in einem erträglichen Verhältnis befinden.

9.7 Auswirkung von Alterungsprozessen auf Veränderungen der Sprache

Verfahren, mit denen Alterungsprozesse in mathematischen Modellen nachgebildet werden, sind nicht neu. Sie kommen in vielen Bereichen der Wissenschaft von Simulation bis zur Praktischen Informatik vor.

Im Rahmen der Theorie der dynamischen Netze kann durch die Alterung sowohl die Topologie des Netzes beeinflusst werden als auch die Dynamik auf den Knoten des Netzwerkes.

Eine Änderung der Topologie kann dadurch geschehen, daß alte Knoten verschwinden (Beispiel: Tod eines Individuums) oder weniger Kanten binden können (Beispiel: Alte Artikel werden seltener zitiert als neue.). Dieser Ansatz wird in [DM00] für Alterungsfunktionen untersucht, die die Form $t^{-\alpha}$ haben. Das Ergebnis ist, daß die Form des Netzes extrem subtil von dem Parameter α abhängt. Für $\alpha > 1$ etwa verliert das Netz seine Potenz-Verteilung. Schon allein dieses Ergebnis zeigt, wie vorsichtig bei der Modellierung vorgegangen werden muß.

Eine Änderung der Dynamik könnte andererseits dadurch geschehen, daß die Zustandsmenge verändert und beispielsweise ein besonderer Alterungszustand angenommen wird.

Da es sich bei meiner Arbeit um einen der ersten Versuche handelt, dynamische Netze für die Linguistik nutzbar zu machen, kann das Modell noch nicht so elaboriert sein wie etwa die Modelle der Physiker. Deshalb nutze ich keine derartig komplexen Verfahren sondern wende das einfache in Unterkapitel 13.8 erläuterte Verfahren zur Modellierung von Alterungsprozessen im Dynamischen Sprachnetz an.

10 Anwendungen von Netzwerken in der Linguistik

Nachdem jetzt alle Begriffe zur Verfügung gestellt worden sind und bevor ich im nächsten Kapitel mit der Konstruktion des Dynamischen Sprachnetzes beginne, erscheint es angebracht, noch einen kurzen Überblick darüber zu geben, wie Netzwerke in den letzten Jahren zur Untersuchung von Sprachen benutzt worden sind. Da es sich bei der Theorie der Netzwerke ganz allgemein um einen sehr jungen Gegenstand handelt, gibt es natürlich noch keine große Menge von bereits speziell im Bereich der Sprachwissenschaft durchgeführten Untersuchungen und demzufolge zur Zeit noch vergleichsweise wenige konkrete Ergebnisse für die Linguistik ganz allgemein. Der hauptsächliche Fokus liegt noch im wesentlichen, aber nicht ausschließlich darauf, die genauen Eigenschaften von linguistischen Netzwerken zu verstehen. Ehe ich mich den eigentlichen Anwendungen von Netzwerken in der Linguistik zuwende, möchte ich einige hauptsächliche Eigenschaften von aus der Sprache abgeleiteten Netzwerken nennen, die schon früher in dieser Arbeit betrachtet worden sind und auch in den in diesem Kapitel genannten Publikationen immer wieder festgestellt werden:

- kurze Wege im Graphen, also kleiner Graphdurchmesser und eng damit zusammenhängend
- small-world-Eigenschaft,
- hoher Clusteringkoeffizient,
- Skaleninvarianz (siehe Unterkapitel 6.5),
- Pareto-Verteilung der Knotengrade.

Aus zwei Gründen scheinen solche Eigenschaften wichtig: Sprachliche Netze haben erstens die gleichen signifikanten statistischen Eigenschaften, die gleichen Universalien wie Netzwerke, die in anderen Bereichen zu finden sind. Sprachen sind also der gleichen Dynamik unterworfen wie andere natürliche und soziale Systeme. Diese statistischen Beobachtungen sind zweitens Folgen universeller Gesetze, die die Entwicklung von Netzwerken steuern und, wie noch viel deutlicher darzulegen ist, sie sind Ursache für viele sprachliche Phänomene. Es geht bei der Nutzung von Netzwerken in der Sprachwissenschaft also nicht darum, zusätzlich zu bisherigen linguistischen Forschungen einen weiteren vom Standpunkt der Linguistik aus abwegigen Fragenkomplex neu zu erfinden und diesen dann in sich abgeschlossen zu behandeln; es soll viel mehr versucht werden, auf neuen Wegen alte Fragen zu beantworten.

10.1 Sprachliche Universalien

Die Ursachen und die Evolution von Sprache seien trotz vieler Theorien noch weit davon entfernt, geklärt zu sein, meinen die Autoren von [SMVS05], einem Artikel, den ich in diesem Kapitel immer wieder zitieren werde. Sprachen haben eine Reihe von Universalien, beispielsweise Phoneminventare (etwa die Einteilung in Konsonanten und Vokale), die Eigenschaft, Inhalt und Ausdruck zu besitzen, das Zipfsche Gesetz und anderes. Als eine tiefergehende Eigenschaft sei etwa zu nennen, daß eine Sprache mit flektierten Hilfsverben vor dem Verb normalerweise auch in ihrem Bestand Präpositionen und keine Postpositionen

aufweist. Das Ziel der meisten in diesem Kapitel vorgestellten Untersuchungen ist in den Worten von [Fer05] das Auffinden neuer möglicher Wege, wie universelle Eigenschaften von Sprachen verstanden werden können. Woher aber stammen aber diese Universalien? In [SMVS05] werden zunächst drei Faktoren für deren Entstehung genannt:

- der menschliche Körper und die Hirnstruktur

So ist die Unterscheidung zwischen Vokalen und Konsonanten auf das menschliche Artikulationssystem zurückzuführen. Ebenso sei Sprache Bedingungen unterworfen, die auf das kognitive System zurückgehen, [Cho02], auf Berechnungsprinzipien [PU04], also generelle Gesetze der allgemeinen Algorithmik, oder auf Grenzen des Gedächtnisses.

- die Aufgaben, für die die Sprache benutzt wird

Da Sprache ein Kommunikationswerkzeug ist, sollen die Kommunikation und die Ausdrucksfähigkeit optimiert werden, während die kognitive und physische Anstrengung zu minimieren ist. Nach [FS03] und [Ste05] sind solche Optimierungstendenzen die treibende Kraft hinter grammatischen Erscheinungen wie etwa der Prädikat-Argument-Struktur. Auch hier spielt die Erlernbarkeit eine wichtige Rolle.

- Familienverwandtschaften zwischen Sprachen und Kontakte

Viele sprachliche Eigenheiten gehen nach [SMVS05] darauf zurück, daß Sprachen gleiche Ursprünge haben und es durch räumliche Nähe immer wieder zu Sprachkontakten und damit zu sprachlicher Beeinflussung komme. So können gemeinsame Eigenschaften auch von nicht verwandten Sprachen entstehen, die sich einer tieferen Erklärung entzögen. Es soll hier nicht erörtert werden, ob diese dritte Klasse der in [SMVS05] genannten Faktoren wirklich als „Universalien“ bezeichnet werden sollte.

Dann wird in [SMVS05] aber besonders eine weitere Hypothese untersucht: *Die Gesetze, die die Organisation und Evolution von Netzwerken steuern, sind ein bedeutender vierter Faktor für die Natur und die Evolution der Sprache ganz allgemein.* Erscheinungen, die in allen Sprachen feststellbar sind, könnten Indikator für allgemeine, unausweichliche Mechanismen auf der Netzebene sein, unabhängig davon, welchen Weg die Evolution nun genau genommen hat. Damit könnten sprachliche Universalien sich als Instanzierungen allgemeinerer Netzgesetze erweisen. Ähnliches ist schon früher in [MMLD02] vermutet worden.

Dabei ist der Begriff der Evolution in diesem Zusammenhang, immer noch nach [SMVS05], auf (mindestens) zwei Ebenen anzusiedeln: Erstens ist die Emergenz innerhalb menschlicher Populationen zu nennen, zweitens die Ebene des Spracherwerbs von Individuen⁵⁷. Sprachentwicklung ist eng an die kognitiven Potentiale von Individuen gekoppelt aber auch an die Einflüsse, denen es durch die Sprachnutzung anderer Individuen ausgesetzt ist. Sprachliche Erscheinungen sollten im allgemeinen als kollektive Phänomene mit kontinuierlicher Anpassung von Sprechern und Hörern an die Sprachkonventionen und die Sicht auf außersprachliche Realien angesehen werden. Die Frage, wie diese Sicht der Dinge sich auf die sprachlichen Netze und damit auf die Sprache selbst auswirkt, ist zur Zeit

⁵⁷ Es handelt sich also um nichts weiter als um den bekannten Unterschied zwischen Phylogenese und Ontogenese.

noch ziemlich offen, wiederum weil die Theorie der Netze selber noch zu jung ist. Ansätze bieten [ST05] und [DM01].

Es gibt weitere Ansätze, mit deren Hilfe Universalien gefunden werden sollen, die für sprachliche Netze charakteristisch sind, die sozusagen deren „*Alleinstellungsmerkmale*“ gegenüber anderen Netzen bilden. So werden in [MKI+04] und [MIK+04] systematisch die Ähnlichkeiten in lokalen Unterstrukturen von allgemeinen Netzwerken untersucht. Ein Beispiel aus einer etwas früheren Arbeit, [MSI+02], bilden die sogenannten „*motifs*“; das sind lokale Verknüpfungsmuster im Netz. Es zeigt sich: Wird ein motif in hoher Konzentration in einer Sprache gefunden, findet es sich mit hoher Wahrscheinlichkeit auch in hoher Konzentration bei einer anderen Sprache, auch wenn die genauen statistischen Werte von Sprache zu Sprache schwanken. Nach Meinung der Autoren reflektiert die Neigung zu gewissen motifs die Art, wie Wörter in Sprachen benutzt werden. Netzwerke aus anderen Bereichen, etwa biologische Netzwerke, bevorzugen andere motifs. Dadurch lassen sich „*Signifikanzprofile*“ für Netzklassen definieren. Durch die Ähnlichkeit solcher Signifikanzprofile entstehen „*Superfamilien*“. Eine derartige Superfamilie bilden die weiter unten erwähnten word co-occurrence networks. Andere Superfamilien werden unter anderem durch gewisse genetische Netze, Proteinstrukturen oder neuronale Verknüpfungen gebildet.

In [RB03] wird ein anderes Alleinstellungsmerkmal sprachlicher Netze vorgestellt: Es wird gezeigt, daß in sprachlichen Netzen das lokale Clustering eines Knotens in merkwürdiger, bei anderen Netzen selten anzutreffender Weise abhängig vom Knotengrad ist: Je höher der Grad, desto geringer der lokale Clustering-Koeffizient. Es gehören also gering verbundene Knoten zu dicht verbundenen Clustern, und umgekehrt gehören die gut verbundenen Knoten, die hubs, nicht zu dicht verbundenen Clustern. Daraus kann gefolgert werden, daß die Organisationsstruktur solcher Netze sich auf verschiedenen Ebenen wiederholt; es liegt also auch hier ein Hinweis auf Skaleninvarianz vor.

Um was für Netze geht es nun? Es lassen sich in der linguistischen Forschung der letzten Jahre zwei Sorten von potentiellen Anwendungen von Netzen unterscheiden:

- Netzwerke in einem existierenden Sprachsystem
- Netze von Sprachnutzern

Diese Netzarten werden in den folgenden Unterkapiteln noch genauer vorgestellt werden.

Neben den für die Linguistik sicher interessanten, wenn auch vielleicht etwas esoterischen und noch in der Zukunft liegenden Überlegungen zu Universalien gibt es auch schon heute die eine oder andere für die reine Linguistik spannende Konsequenz, und die meisten der im folgenden aufgeführten Autoren erwähnen große Anwendungsmöglichkeiten dafür.

10.2 Netzwerke in einem existierenden Sprachsystem

Darunter sind solche Netzwerke zu verstehen, deren Knoten eine sprachinhärente Interpretation besitzen, etwa Wörter oder formal-semantische Inhalte. Die Interpretation der Interaktion zwischen den Knoten ist sehr uneinheitlich; die Forschung ist da noch in einem experimentellen Stadium.

In dem Bereich der Netzwerke, die sich mit Beziehungen innerhalb der Sprache beschäftigen, sind die Arbeiten von Steyvers und Tenenbaum, [ST01] und [ST05], grundlegend. Sie sollen schon vor der Jahrtausendwende bekannt geworden sein und bedeutenden Einfluß gehabt haben, bevor eine Kurzfassung, eben [ST01], sehr prominent wurde, der zweite genannte Artikel ist die Langform. Laut [Fer05] sind [ST01] und [ST05] die einzigen Arbeiten, die sich gründlich mit mehreren deutlich verschiedenen Formen der sprachlichen Netzwerkmodellierung befassen.

Ich möchte im folgenden einige herausragende Beispiele für linguistische Netzwerke kurz schildern:

10.3 *Das word web*

Zu diesem Bereich gehört eine Reihe von Untersuchungen über Wortnähen, bei denen die Knoten von Wörtern einer Sprache, im allgemeinen der englischen, gebildet werden und wo Kanten dann gegeben sind, wenn die beiden zu verbindenden Wörter in irgendeiner Form miteinander interagieren. Die Bedeutung der Kanten ist in der Literatur nicht einheitlich festgelegt. Ein Vorschlag ist es, eine Kante zwischen zwei Wörtern zu konstruieren, wenn sie in einem Satz eines gegebenen Corpus benachbart oder nur durch ein weiteres Wort getrennt sind [FS01]. Es gibt aber auch andere Vorschläge, etwa den, es solle eine Kante existieren, wenn es einen Satz gibt, in dem sie gemeinsam vorkommen. Ebenso spielt in manchen Ansätzen die Wortreihenfolge eine Rolle⁵⁸. Die so entstehenden Netzwerke werden allesamt als „word web“ oder auch als „word co-occurrence networks“ bezeichnet. Sie besitzen die small-world-Eigenschaft und sind durch ein in [MIK+04] entwickeltes Signifikanzprofil gekennzeichnet.

Untersuchungen über die Dynamik solcher Netzwerke haben noch einmal den bekannten Befund bestätigt, daß die Funktionswörter (wie Artikel, Hilfsverben, Präpositionen) der stabilste Teil der Sprache sind. Immerhin haben sie die höchste Konnektivität, [SFMV03], sind also die hubs im Netz. Da der Verknüpfungsgrad mit der Häufigkeit eines Wortes korreliert ist und diese wieder mit dem Alter eines Wortes, haben die Autoren dort auch gefolgert, daß in der Dynamik des word web ein preferential linking vorliegt.

Auch von Dorogovtsev und Mendes (besonders in [DM01] und [DM03]) sind Forschungen über das word web publiziert worden, besonders über statistische Eigenschaften bezüglich der Stabilität des Wortschatzes, so eine Untersuchung über die Dynamik der Sprachentwicklung, die ebenfalls auf dem word web aufbaut und von zwei Grundannahmen ausgeht [DM01]:

- Sprache ist ein evolvierendes Netzwerk interagierender Wörter.
- Während ihrer Evolution organisiert die Sprache sich selber zu einer komplexen Struktur.

⁵⁸ In diesem Falle entstehen gerichtete Graphen als Modelle.

Auf der Grundlage eines extrem einfachen Modells werden sehr realistische Konsequenzen beispielsweise über die Veränderung des Kernwortschatzes einer Sprache gezogen.

In [Jin04] wird der Spracherwerb von Kindern mit solchen Netzen analysiert. Es wird von einer Spannung zwischen Lexikongröße und Flexibilität berichtet: Kindersprache mit kleinerem Lexikon wird durch ein Netz mit höherer Konnektivität reflektiert. Auch die hubs verändern sich, wenn sich die Sprache eines Kindes entwickelt: weg von Inhaltswörtern wie „mama“, hin zu Funktionswörtern.

Ein Ansatz, bei dem die untersuchten Einheiten nicht die Wörter sondern die Silben sind, ist [MCL05]. Dort wird ein co-occurrence Netzwerk für die Silben der portugiesischen Sprache untersucht. Auch auf der Ebene der Silben sind vergleichbare Netzeigenschaften wie für Netze auf der Grundlage von Wörtern zu finden.

10.4 Semantische Netzwerke

Der Begriff „Semantisches Netzwerk“ ist seit den sechziger Jahren des letzten Jahrhunderts im Zusammenhang mit Wissensrepräsentationstechniken benutzt worden. Auch dabei gilt es, Beziehungen zwischen Entitäten mit Hilfe von Graphen zu modellieren. Hier gilt es zu unterscheiden zwischen einerseits den „klassischen“ Netzen, die bei der reinen Wissensrepräsentation verwandt werden, wo Graphentheorie lediglich zur bequemeren Darstellung dient, wo aber keine ernsthaften graphentheoretischen Methoden benutzt werden, und andererseits einfacheren jüngeren Darstellungen, die semantische Grundlagen benutzen, wo die Netze aber konzeptionell so einfach gestaltet sind, daß graphentheoretische Methoden zu ihrer Analyse anwendbar sind.

Der erste Typ von Netzen zeichnet sich dadurch aus, daß er sehr komplex ist. Es soll ja, zumindest im Prinzip, *das Weltwissen* repräsentiert werden. Da kommt man mit so einfachen Graphen, wie sie in dieser Arbeit betrachtet werden, nicht sehr weit. In der Regel gibt es viele verschiedene Typen von Kanten. In dem Modell „MultiNet“, [Hel01], einer der derzeit führenden Methodiken, sind nicht nur wie üblich Kanten zwischen Knoten sondern auch Kanten zwischen Kanten und Knoten zugelassen. Vergleichbare Erweiterungen kommen in den meisten vergleichbaren Formalismen vor. Zu ihrer Behandlung bedürfte es Logik höherer Ordnung⁵⁹. Beim jetzigen Stand der Wissenschaft sind keine tieferen Erkenntnisse durch mathematische Analysen zu erwarten. In der Computerlinguistik, als prominentestes Beispiel diene der Bereich der automatischen Übersetzung, ist eine semantische Komponente unerlässlich; dazu können solche Netze dienen.

Zu den Netzen des ersten, des mathematisch komplexen Typs können auch die Netzwerke gezählt werden, die im Rahmen der „Word Grammar“ konstruiert worden sind. In dieser Theorie sind Wörter, Wortformen und Laute die drei Arten von linguistischen Einheiten, nicht aber syntaktische Einheiten. Insbesondere wird die Phrasenstrukturgrammatik verworfen. Zusätzlich gibt es eine reichhaltige Hierarchie von Beziehungen zwischen linguistischen Einheiten und auch zwischen linguistischen und nicht-linguistischen Einheiten. Auch das auf diese Art entstehende Netzwerk kann als Teil des Weltwissens betrachtet

⁵⁹ Es spielt für unsere Untersuchungen keine Rolle, wie sich die Logik höherer Ordnung von der klassischen Logik unterscheidet. Eine angemessene Behandlung ist in jener auf jeden Fall schwieriger.

werden. Zur dieser Theorie ganz allgemein kann in [Hud84] mehr erfahren werden, zu den in diesem Umfeld konstruierten Netzen in [Hud06].

In unserem Zusammenhang sind alle diese eben genannten komplexen Netze nicht anwendbar. Ich erwähne sie nur der Vollständigkeit wegen.

Netzwerke des zweiten Typs sind wesentlich simpler, zumindest auf der Ebene der Knoten und Kanten. Dem steht eine verblüffende Komplexität des Gesamtnetzes gegenüber. Zu den Netzklassen, die auf semantischer Grundlage erstellt werden, die aber keine klassischen Netze der Wissensrepräsentation sind, gehören die in den beiden folgenden Unterkapiteln geschilderten Netzwerkklassen. Einen Überblick über semantische Netzwerke vorwiegend des zweiten Typs bietet [BJ06].

In [ST05] wird vorgeschlagen, die Topologien von semantischen Netzwerken zu verschiedenen Sprachen zu untersuchen, damit dann die Kontextabhängigkeit von Bedeutungen sowie der Zusammenhang zwischen mentalen Konzepten und sprachlichen Realisierungen in Abhängigkeit von der jeweiligen Sprache genauer erforscht werden können⁶⁰.

Neben theoretischen Ergebnissen gibt es eine Reihe praktischer Anwendungen semantischer Netzwerke des zweiten Typs: automatische Wortdisambiguierung, [Bor03], [Ver04], Beantwortung natürlichsprachlicher Fragen an automatische Sprachsysteme, [BDM03]. In [Bor03] gibt es auch Ansätze zu psycholinguistischen Forschungen über den Spracherwerb.

10.5 Thesaurus-Netzwerke

Thesauri werden benutzt als Liste von Einträgen, wobei diese Einträge aus einem Wurzelwort und einer Liste von Wörtern bestehen, die sich in einer gewissen semantischen Nähe zum Wurzelwort befinden (unter Umständen in Synonymie). Dieser Ansatz legt eine Modellierung durch einen Graphen sofort nahe: Die Knoten sind die Wörter; die Liste der semantisch nahen Wörter definiert die Liste der Nachbarn⁶¹.

Soweit es Thesaurus-Netzwerke betrifft, stützen sich die Artikel anderer Autoren stark auf die Arbeit von Steyvers und Tenenbaum. Als zugrunde liegende Thesauri werden *Roget's Thesaurus* (in [ST01], [ST05], [MMLD02], [New03]) oder der *Thesaurus von Merriam-Webster* (in [AB02]) verwandt. In [JTK+03] wird das auf dem *Moby Thesaurus II*, [War02], dem größten frei im Weltnetz verfügbaren Thesaurus in Englisch, beruhende Netzwerk auf viele der Charakteristika hin untersucht, wie sie in dieser Arbeit schon vorgestellt worden sind.

Ich möchte nur ein Beispiel anführen, das aus [MMLD02] stammt: Die small-world-Eigenschaft dieser Netze beruht darauf, daß es Wörter gibt, die mit mehreren sehr verschiedenen Konzepten korrespondieren. So läßt sich eine Kette von Begriffen konstruieren:

⁶⁰ Viele der dort benutzten Netzwerke sind bipartite Graphen (siehe Seite 84). Die beiden laut der dort angegebenen Definition vorhandenen Gruppen von Knoten sind durch die mentalen Konzepte und die möglichen sprachlichen Realisierungen gegeben.

⁶¹ Genauso, also mit Hilfe von Nachbarschaftslisten, habe übrigens auch ich die Graphen in meinem Modell implementiert.

universe - nature - character - letter

Dabei ist jedes Wort inhaltlich verwandt mit seinen Nachbarn. Solche Ketten bilden genau die Abkürzungen, die bei small-world-Netzen nötig sind. Es ist spekuliert worden, [MMLD02], daß die Existenz solcher Ketten mit einer assoziativen Speicherung im Gehirn zusammenhängen könnte. Diese stelle unter dem Gesichtspunkt des Wiederauffindens gespeicherter Information ein ideales Hilfsmittel dar, das die Suche deutlich beschleunige. Zu ähnlichen Schlüssen kommen die am Ende des nächsten Unterkapitels zitierten Autoren.

Ein häufig zitierter, an der Physik orientierter Ansatz bei der allgemeinen Untersuchung von Thesaurus-Netzwerken ist [KML+02].

10.6 Das WordNet

Das WordNet, zu finden im Internet unter [Word], nicht zu verwechseln mit dem word web, beruht auf psycholinguistischen Grundlagen. Es werden nicht nur Wort-Bedeutungs-Assoziationen sondern auch Bedeutung-Bedeutungs-Assoziationen betrachtet. Dadurch wird erstens die Nachbarschaftsbeziehung vage und nicht mehr klar mathematisch faßbar sowie zweitens die Netze selber sehr komplex. Andererseits gelingt es auf diese Weise, viele Beziehungen mathematisch präzise zu fassen, für die das in einfacheren Modellen nicht möglich ist. In WordNet werden nur Verben, Nomina, Adverbien und Adjektive studiert. Auch auf diesem Gebiet sind [ST01] und [ST05] grundlegend.

Daneben ist [SC02] zu nennen. In dieser Arbeit werden wie üblich Netzcharakteristika bestimmt, aber wieder in einem besonderen Netz: Darin wird mit Hilfe semantischer Relationen besonders Polysemie untersucht. Zwei Konzepte a und b , die Knoten des Graphen, sind durch eine Kante verbunden, wenn sie solchen semantischen Relationen genügen, etwa „ a ist ein b “, „ a ist ein Teil von b “ oder „ a ist das Gegenteil von b “. Die hubs sind in solchen Netzen die polysemen Wörter. Nach Meinung der Autoren dient Polysemie dazu, das Netz in kompakter und kanonischer Weise zu organisieren. Diese Tatsache könne das durchgängige Vorhandensein von Polysemie in den verschiedenen Sprachen erklären helfen. Polysemie sei demzufolge kein „Betriebsunfall“ sondern eine notwendige Komponente aller Sprachen.

Nach [ST05] bewirkt die skalen- oder größeninvariante Topologie semantischer Netze Einschränkungen für die Art, wie diese Netze und folglich auch die durch sie repräsentierten Sprachen im Gehirn verarbeitet werden können. Auch das hohe Clustering und die small-world Eigenschaft haben große Auswirkungen auf die Suchverfahren, die angewandt werden. So werden assoziative Speicherverfahren von den Autoren aus mathematischen Gründen für nützlicher gehalten als strenge Verfahren, wie sie im Gegensatz dazu in den meisten Computern heute üblich sind.

10.7 Wort-Assoziations-Netzwerke

Diese Netzwerke fußen, neben ebenfalls auf [ST01] und [ST05] beruhenden Grundlagen, auf psychologischen Experimenten. So werden beispielsweise die Kanten

zwischen den durch Wörter gegebenen Knoten dadurch etabliert, daß auf einen Stimulus, ein vorgegebenes Wort, mit einem assoziierten Wort geantwortet werden muß. Mathematisch führt dieser Ansatz zu gerichteten Graphen (vom Stimulus zur Antwort). [CSCC05]

10.8 Netzwerke syntaktischer Abhängigkeiten

In Anlehnung an die Dependenzgrammatik kann die Struktur eines Satzes durch einen Abhängigkeitsgraphen dargestellt werden. Die Wurzel bildet normalerweise das Prädikat. Deren Kinder bilden die Aktanten, die wieder in Bäume entfaltet werden. Es geht jeweils eine Kante vom Kopf zum Modifikator. Es gibt aber auch Ansätze mit Hilfe ungerichteter Graphen. Untersucht werden Beziehungen zwischen der Position von Wörtern in einem Satz und der Struktur dieses Satzes.

Grundlegend ist ein Ökonomieprinzip mit einer Kostenfunktion, die darauf fußt, daß die Entfernungen durch die syntaktische Struktur verbundener Komponenten ein Maß für die Kosten liefern, die durch deren Verarbeitung im Gehirn entstehen, [Fer05]. Die begrenzten Ressourcen im menschlichen Hirn führen zu Einschränkungen der vorhandenen Distanzen innerhalb von Sätzen. Von den Untersuchungen im Word Web und anderen word co-occurrence Netzwerken unterscheiden sich diese Untersuchungen also durch ein elaborierteres Distanzmaß: Es werden zur Bestimmung der Distanz zwischen zwei Wörtern w_1 und w_2 nicht einfach die trennenden Wörter gezählt, sondern die Satzstruktur wird ausgenutzt, indem die Entfernung zwischen w_1 und w_2 im Abhängigkeitsgraph gemessen wird. Genauere Untersuchungen zeigen nach [Fer05], daß gewisse statistische Eigenschaften der Netze in diesen beiden Ansätzen unterschiedlich sind.

Auf diese Weise können dann ähnlich wie im Word Web verschiedenen Sprachgemeinschaften auf relativ kanonische Weise Netze zugeordnet werden. Ein Beispiel, bei dem Charakteristika solcher Netze für verschiedene Sprachen untersucht werden (Rumänisch, Tschechisch und Deutsch), ist [FSK04]. Die Ergebnisse dieser vergleichenden Untersuchung für mehrere Sprachen wie auch einer Untersuchung für das Chinesische, [ZCC03], legen es nahe, daß sich solche Charakteristika auch als sprachliche Universalien entpuppen könnten. Die in [FSK04] berichteten Eigenschaften der Netze (und damit der Syntaxorganisation) seien keine trivialen Konsequenzen der Satzstruktur sondern emergente Eigenschaften der Netze. Die Autoren erinnern ferner daran, daß es für viele allgemeine sprachliche Universalien, zum Beispiel für die Fähigkeit, Sätze aus Wörtern zu kombinieren, also Syntax zu bilden, noch keine Erklärung gebe.

In [Fer05] wird berichtet, die Struktur des globalen syntaktischen Abhängigkeitsnetzwerkes reflektiere die Hirnstruktur (als Netzwerk von durch Synapsen verbundenen Neuronen) in vielerlei Weisen, die über bloße physikalische Ähnlichkeit weit hinausgehe. Von dort gesehen böte sich nach Meinung von Ferrer y Cancho eine Erweiterung des klassischen Phrasenstrukturmodells von Chomsky [Cho57] mit Hilfe von auf Netzwerken syntaktischer Abhängigkeiten fußenden Formalismen an. Netzwerkprozesse modellierten Vorgänge im Gehirn eben besser als einfache Ersetzungsregeln, wie sie von Chomsky und anderen postuliert werden.

In dem Übersichtsartikel [Ke07], auf den ich mich im nächsten Unterkapitel noch häufiger beziehen werde, wird hervorgehoben, daß es sich bei den meisten bisher in diesem Kapitel behandelten Anwendungen von Netzwerken in einem existierenden Sprachsystem

zumeist um synchrone Studien handele. Insbesondere seien die Kräfte, die bei der Entwicklung der Netzeigenschaften wirken, noch viel zu wenig verstanden, so daß verstärkt ein evolutionärer Standpunkt eingenommen werden müsse; so sehen das auch die Autoren von [SMVS05]. Es sei zu hoffen, daß durch die Vermittlung von Netzwerken künftig auch komplexere Aspekte der Sprache, vielleicht mit Hilfe von Ökonomieprinzipien, aus anderen Bereichen abgeleitet werden können, etwas aus kognitiven oder sozialen Modellen.

Ein Beispiel dafür ist der Ansatz in [KY06]. Dort wird die Entwicklung von in Longitudinalstudien untersuchter Kindersprache (und auch der Sprache der sie betreuenden Personen) dargestellt als eine Folge von Netzwerken. Nach Meinung der Autoren sei eine Einteilung etwa in frühe und späte Sprecher bei weitem zu grob. Die Entwicklung erfolge in vielen Dimensionen, beispielsweise können die Größe des Vokabulars, die MLU („mean length of utterance“) oder viele andere Parameter als Maßeinheit für die Entwicklung herangezogen werden. Solche Meßgrößen entsprechen oft charakteristischen Netzgrößen. So entspreche die Größe des Vokabulars der Netzgröße und die MLU passe bei richtiger Interpretation zum durchschnittlichen Knotengrad. In dem genannten Artikel werden auch Untersuchungen zur Entwicklung der Morphologie und zur Benutzung von bestimmtem, unbestimmtem oder gar keinem Artikel bei Kleinkindern geschildert (letzteres mit Hilfe unter anderem von hubs). Durch den Ansatz in [KY06] wird den Netzen zwar eine Dynamik hinzugefügt. Dies geschieht aber auf einer rein phänomenologischen Ebene. Es wird etwas beschrieben, nämlich die globalen Muster eines solchen Entwicklungspfades für die Sprache eines Kindes durch einen vieldimensionalen Raum, aber nicht auf mathematisch-theoretischer Grundlage erklärt. Eine Beschreibung, die sich auf Netze stützt, sei besonders für syntaktische Strukturen geeignet. Sicher seien viel genauere Ergebnisse möglich, sobald an den Kantenrelationen noch gefeilt werde⁶². Beispielsweise könne der unterschiedlichen Intensität der Interaktion zwischen zwei Wörtern (oder anderen Einheiten) durch mehrfache Kanten oder durch die Einführung von Gewichten an den Kanten Rechnung getragen werden.

Es werden in [Ke07] aber auch weitere interessante Möglichkeiten für die Zukunft prognostiziert, insbesondere psycholinguistische Ansätze in Gebieten wie Zungenbrecher, Koordination bei Synonymie, statistische Neigung zu verschiedenen Formen bei Synonymen oder statistische Verteilung von Zugangszeiten zum mentalen Lexikon, ferner Verbesserungen bei Untersuchungen zur Aphasie. So könne der Ausfall von Knoten in einem Netz Wege dramatisch verändern; darauf wurde auch von mir ja schon in Unterkapitel 9.1 hingewiesen. Das könne Krankheitsbilder näher verständlich machen, die ihren Ausgangspunkt bei einem Verlust der Funktionswörter, also der hubs, haben.

Natürlich haben sprachliche Netze ihre Grenzen, es kann nicht alles aus ihnen abgeleitet werden.

- Es gibt universelle Entwicklungen und Einschränkungen, die einfach keine Verbindung zu statistischen Eigenschaften von Netzwerken haben: der menschliche Körperbau (Gehirn, Artikulationsorgane), die Zwecke des Sprechens, geschichtliche Zufälligkeiten.
- Gerade die Tatsache, daß viele statistische Gesetze universell über die Linguistik hinaus gültig sind, zeigt, daß die Untersuchung von Netzen noch nicht hinreichend genau auf die Untersuchung von Sprachen hin spezialisiert ist. Die Einmaligkeiten von Sprachen werden noch nicht gut genug erfaßt.

⁶² Hier handelt es sich um einen erneuten Hinweis darauf, daß der Kantenrelation gegenwärtig noch etwas Willkürliches anhaftet, wie ich es ja bei der Schilderung des word webs schon erwähnt habe.

10.9 Netze von Sprachnutzern

Von den beiden Netzklassen, die, wie ich oben in diesem Kapitel schon erwähnt habe, in der linguistischen Forschung der letzten Jahre eine Rolle gespielt haben, sind die Netzwerke in einem existierenden Sprachsystem bisher die bei weitem bedeutendere Klasse. Zur Darstellung der Forschungsstandes bei Netzen von Sprachnutzern werde ich immer wieder den Überblicksartikel [Ke07] zitieren. Die Forschung mit solchen Methoden habe sich nach Meinung von Ke vor allem auf synchrone Variationen in sozialen Netzwerken konzentriert, besonders in kleinen lokalen Gemeinschaften. Zur Zeit existierende Modelle seien nicht adäquat:

„As a complementary approach, computational modelling studies can offer a way to study the long-term effect of language change in speech communities of any sizes and structures. Existing models of language change, however, either do not consider the actual population structure, or assume regular or random networks as the population structure.“
[Ke07]

Ke nennt die Situation paradox: Intuitiv gesehen seien soziale Netzwerke ein wichtiger Faktor beim Sprachwandel. Es gebe jedoch sehr wenige empirische Studien, die sich quantitativ über längere Zeiträume mit der Wirkung von sozialen Netzen beim Sprachwandel beschäftigen.

Es hat im theoretischen Bereich allerdings in neuerer Zeit auch Versuche gegeben, mit Hilfe sogenannter „Sprachspiele“⁶³ eine evolutionäre Dynamik in die Netze zu inkorporieren, so daß Sprache als ein selbstorganisierendes System zum optimalen Informationsaustausch beschrieben werden kann, etwa in [DBBL06], [TLW+07]. So ist das „Naming Game“ ein Modell für die selbstorganisierte Emergenz von linguistischen Konventionen in einem Kommunikationssystem paarweise interagierender Partner. Bei diesem minimalen Modell beobachten Agenten ein Objekt, wobei jedem Agenten eine Liste mit möglichen Namen für das Objekt zugeordnet ist⁶⁴. Durch Kommunikation ändert sich das Wissen der Agenten um den nützlichsten Namen. In [DBBL06] kann gezeigt werden, daß in solchen Kommunikationsnetzen, außer in einigen Netzen mit ganz spezieller Struktur, immer ein globaler Konsens, also eine linguistische Konvention, erzielt werden kann.

In [Ke07] hebt der Autor insbesondere die großen Potentiale von computergestützten Untersuchungen hervor:

„Computer simulations provide us with a convenient platform to study the effect of the social network systematically, as we can manipulate the various parameters, such as size, structure of the network, etc. We can simulate the evolution of the network under control conditions, and address questions of language change at different time scales. ...

⁶³ Damit sind natürlich keine Kinderspiele gemeint, sondern Spiele im Sinne der Spieltheorie: Es handelt sich um einfache Modelle interagierender Agenten, die eine kollektive Einigung auf eine Abbildung zwischen Wörtern und Objekten oder Bedeutungen entwickeln. Dabei geht es um die Entstehung eines Systems sprachlicher Konventionen ohne allgemeine Aufsicht oder gemeinsames Wissen a priori.

⁶⁴ Der Zusammenhang mit den Netzwerken ist dadurch gegeben, daß die Agenten die Knoten bilden. Der Zustand des Agenten ist die Liste mit dem Agenten bekannten möglichen Namen für das Objekt. Dieser Zustand ändert sich durch Kommunikation, damit auch der Sprachgebrauch.

Computational models are particularly useful in addressing questions at a larger time scale.“ [Ke07]

Damit fordert Ke genau das, was in dieser Arbeit getan worden ist: Computersimulationen zur Untersuchung im Rahmen größerer Zeiträume. Ke beklagt, daß es zur Zeit noch zu wenige Daten für eine exakte Modellierung sozialer Netzwerke gebe⁶⁵.

„So far there have been few data of large scale social networks with linguistic behavioral data. ... Though the models presented so far are still simple and highly idealized, they can be used to address some empirical questions in historical linguistics as well as sociolinguistics.“ [Ke07]

Auch hier deckt sich die Beschreibung von Ke mit meiner Arbeit. Natürlich habe ich versucht, das Modell so einfach wie möglich zu halten⁶⁶ und möglichst viel zu idealisieren. Auch der Zweck ist der von Ke angegebene, die Analyse empirischer Fragen in der historischen Linguistik.

Ein weiterer Grund für die Beschäftigung mit sozialen Netzwerken sei auch ihre Bedeutung für die Frage, wie die Sprache an sich überhaupt entstanden sei, so die Autoren von [GTK+04]. Auch für eher soziale Aspekte, insbesondere den Gegensatz von langue und parole, könnten soziale Netzwerke zukünftig wichtig werden:

„With the development of these techniques⁶⁷, it is expected that more complex linguistic networks can be constructed and analyzed, and provide a new paradigm to study the organisation and evolution of language, both for language in the individual speakers and language in the speech communities.“ [Ke07]

Insbesondere können soziale Netzwerke (beispielsweise mit Hilfe von Erkenntnissen der Diffusionsdynamik) zu neuen Erkenntnissen über den Sprachwandel führen.

„The models allow systematic studies on how different network structures affect the diffusion dynamics and language change in the long run, and provide a computational framework to address some theoretical questions in historical linguistics, ...“ [Ke07]

Ke kommt zum gleichen Fazit wie auch andere in diesem Kapitel zitierte Autoren:

„All in all, network research will bring a new perspective to linguistics, provide a new methodology to carry out quantitative analyses and suggest new questions and insights; ...“ [Ke07]

⁶⁵ Ich habe den Mangel an sozialen Daten durch die umfangreiche Datenbank ausgleichen können.

⁶⁶ Es ist ja nicht so daß die Güte eines Modells mit seiner Komplexität steigt. Die Güte eines Modells mißt sich eher an seiner Aussagekraft und an seiner Handhabbarkeit. Diese ist eher in negativer Korrelation zur Komplexität.

⁶⁷ Gemeint sind neuere Techniken der Netzwerkanalyse.

10.10 Das Zipfsche Gesetz

Daß das Zipfsche Gesetz eine Verbindung mit den für Netzwerke wichtigen Pareto-Verteilungen hat, habe ich ja schon in Unterkapitel 7.3 erwähnt. In den letzten Jahren haben sich im Themenbereich „Zipfsches Gesetz und Netzwerke“ einige neuere Erkenntnisse ergeben, die es rechtfertigen, jetzt die Zusammenhänge zwischen diesem in der Linguistik wohlbekannten Gesetz und der Theorie der Netzwerke noch in einem eigenen Unterkapitel genauer darzustellen.

Ob dem Zipfschen Gesetz ein Ökonomieprinzip zugrunde liegt, das ist auch heute noch umstritten. Etwa in [Her66] wird nicht nur bestritten, daß durch Zipfs Beobachtung ein ökonomischer Charakter der Sprache offenbar werde; es wird sogar die Wissenschaftlichkeit des Gesetzes an sich bestritten. Hier komme nur ein rein statistisches Phänomen zum Ausdruck. In der Tat scheinen Beobachtungen, zu deren Erklärung Ökonomie keine Rolle spielt, wo sich aber trotzdem das Zipfsche Gesetz als gültig erweist, die Behauptung zu bestätigen, daß das Gesetz eine andere Ursache als sparsames Verhalten haben müsse. Nehmen wir beispielsweise die berühmten Affen, die auf einer Schreibmaschine „Texte“ aus Zufallszeichen erzeugen. Dann kann gezeigt werden, daß die durchschnittliche Länge von „Wörtern“, also Zeichenketten zwischen zwei Leerzeichen, gemäß dem Zipfschen Gesetz, besser: einer Pareto-Verteilung, verteilt ist, siehe [Fer06]. Ebenso ist die Anzahl der Intervalle in der Klaviermusik von F.Chopin annähernd danach verteilt, siehe [Her66]. Nach der in [Fer06] vertretenen Meinung seien solche Phänomene in der Tat rein statistisch zu erklären; bei Sprachen könne aber gerade nicht rein statistisch argumentiert werden, da Wörter nicht rein zufällig auftreten. Dann würde man die Korrelationen zwischen ihnen ignorieren. In [FS03] zeigen die Autoren, daß man zumindest unter der Annahme eines Ökonomieprinzips mit Hilfe des word webs das Zipfsche Gesetz ableiten kann. Seitdem gibt es einige auf Netzwerken basierende Untersuchungen, die nachzuweisen versuchen, daß das Auftreten der Pareto-Verteilung im Zusammenhang mit Sprache alles andere als trivial ist. Im allgemeinen wird davon ausgegangen, daß die Sprecher zwei normalerweise entgegengesetzte Ziele haben: zuverlässige Kommunikation und Energieersparnis bei deren Aufrechterhaltung. Als Maß für die Kosten, also meistens die aufgewandte Energie, wird der durchschnittliche Informationsgehalt einer Einheit (Zeichens, Wortes) benutzt, das heißt, entsprechend der klassischen Definition, die Entropie. Das wird mit dem höheren Wiedererkennungswert häufig benutzter Wörter begründet. Die Begriffe

große Häufigkeit – hoher Wiedererkennungswert –
niedriger Informationsgehalt – geringe Kosten

entsprechen einander. Seien E_s die Entropie und E_h die Ambiguität für den Hörer. Dann gilt es, eine Kombination dieser Werte

$$\Omega(\lambda) = \lambda E_s + (1-\lambda) E_h$$

mit einem Parameter $0 \leq \lambda \leq 1$ zu minimieren. Solche Ansätze werden in [Fer06] und [SMVS05] gewählt. Als besonders erfolgreich haben sie sich im Zusammenhang mit Netzen erwiesen, die auf der Grundlage von Wort-Bedeutungs-Assoziationen aufgebaut sind, bei Netzen, wo Polysemie eine große Rolle spielt, etwa dem WordNet. Es bedeuten $\lambda = 0$, daß nur die Kosten des Hörers berechnet werden, und $\lambda = 1$, daß nur die Kommunikationslast des Sprechers zählt. Im ersten Fall sollte es möglichst zu jeder Bedeutung ein Wort geben und

umgekehrt. Beim anderen Extrem vermindert Polysemie die Kosten eines Sprechers. Minimale Kosten liegen in diesem Fall vor, wenn es in der Sprache nur ein einziges Wort gibt. Zu beiden Extremen gehören Netze mit bestimmten Charakteristica. Wird der Wert von λ kontinuierlich von 0 nach 1 verändert, so ergibt sich nach [SMVS05] eine aufregende Entdeckung: Die zu den Extremen gehörenden Netzklassen gehen nicht langsam in einander über, sondern es gibt eine scharfe Änderung bei einem kritischen Wert λ_c zwischen 0 und 1. Genau an diesem Punkt, der nach den Autoren genau den Phasenübergängen in der Physik entspreche, erscheint bei den Knotengraden plötzlich eine Pareto-Verteilung; genau dort wird also das Zipfsche Gesetz gültig. Nach [FBR05] und [Sol05] folgen bei Netzen, die Charakteristika aufweisen, wie sie genau beim Wert λ_c vorliegen, aus dem Zipfschen Gesetz starke Einschränkungen für die Syntax. Hier werden also mit Hilfe von Netzwerken aus einem Optimalitätskriterium Konsequenzen für die Syntax gezogen. In [Fer06] zieht der Autor auch noch wahrscheinlichkeitstheoretische Überlegungen heran, um nachzuweisen, daß das Zipfsche Gesetz kein statistischer Selbstgänger ist, sondern ein Indikator für eine sehr subtile Balance von Sprachbedürfnissen, die in einem Netz ausgedrückt werden.

10.11 Sprache und Netzwerke: Weitere Aspekte

Die Autoren von [SMVS05] sehen die Notwendigkeit die vielen verschiedenen Typen von Netzen zu systematisieren, insbesondere die Art, wie sich die wechselseitigen Einflüsse der verschiedenen Typen von Netzen, wie auch ich sie hier recht detailliert geschildert habe, untereinander bemerkbar machen. Dort wird daher ein erstes Netzwerk von Netzwerken vorgeschlagen, das diese Beziehungen strukturiert.

Auch die Autoren von [SMVS05] sehen Gründe, warum die Wissenschaft von sprachlichen Netzwerken noch in den Kinderschuhen steckt:

- Es muß sich eine dynamische Sicht auf die sprachlichen Netzwerke breit machen. Dem habe für viele Jahrzehnte eine strukturalistische, auf synchrone Beschreibungen von Sprache fixierte Sicht entgegengestanden.
- Es müssen treffendere Modelle für die kognitiven Prozesse entwickelt werden, die den sprachlichen Netzwerken zugrunde liegen. Diese können dann komplexe Aspekte der Sprache, etwa die Grammatik, besser erklären.
- Die abgeleiteten Universalien sollten nicht in erster Linie als Begrenzungen aufgefaßt werden, sondern als etwas Produktives. Die universellen Kräfte in den Netzen könnten zur Emergenz von geordneten komplexen Strukturen in den Netzen und damit folglich auch in den Sprachen führen. Solche Zusammenhänge zwischen Universalien, Strukturen in den Netzen und Strukturen in den Sprachen gelte es zu entdecken.

Der erste der eben genannten Punkte bezieht sich vorwiegend auf die sozialen Netze aus Unterkapitel 10.9, der zweite eher auf die Netzwerke in einem existierenden Sprachsystem aus den Unterkapiteln 10.2 bis 10.8. Zum Abschluß möchte ich noch eine Liste von Fragen danach erwähnen, wie die Forschung zu sprachlichen Netzwerken künftig nach Meinung der Autoren von [SMVS05] aussehen könnte:

- Wie wirkt sich der Spracherwerb auf sprachliche Netzwerke aus, insbesondere auf das Wachstum?
- Gibt es statistische Unterschiede zwischen den Netzwerken für verschiedene Sprachen?
- Gibt es eine Typologie von Sprachen, so daß die genealogischen Beziehungen durch die Netzeigenschaften widerspiegelt werden?
- Welche Prinzipien der statistischen Physik⁶⁸ können für die Untersuchung von sprachlichen Netzwerken und Sprachen nutzbar gemacht werden?
- Wie werden sprachliche Netzwerke durch Alterung oder Hirnschädigungen verändert?
- Welche Verbindungen gibt es zwischen dem physischen Netz im Gehirn, insbesondere den für die Sprache zuständigen Teilen, und den untersuchten sprachlichen Netzwerken?

⁶⁸ Dies ist der Bereich, wo die Theorie der Netzwerke am weitesten entwickelt ist und am meisten benutzt wird. Dort ist folglich das größte Potential für Anstöße in der Linguistik.

11 Das „Dynamische Sprachnetz“

Die erste Frage, die bei der Modellierung gestellt werden muß, ist die, welche Aspekte der Wirklichkeit überhaupt modelliert werden sollen und wovon abstrahiert werden sollte. An dieser Stelle will ich gar nicht mehr detailliert begründen, wieso dazu ein dynamisches Netz geeignet sein könnte. Dieser Ansatz ist ja geradezu konstitutiv für diese Arbeit und er wird außerdem durch die schon genannten epidemiologischen und die im letzten Kapitel kurz geschilderten linguistischen Arbeiten gerechtfertigt. Wenn man sich einmal für Netze entschieden hat, bleibt die Frage, was in den Knoten, Kanten und den übrigen Bestandteilen dargestellt wird. In dem zur Zeit wohl wichtigsten die Sprachwissenschaft betreffenden Modell, welches auf dynamischen Netzen basiert, dem Word Web, werden, wie oben beschrieben, in den Knoten die Wörter beschrieben. Dagegen ist schon die Bedeutung der Kanten unklar und wird von verschiedenen Autoren willkürlich verschieden interpretiert. Auch bei den meisten anderen im letzten Kapitel, 10.2 bis 10.8, beschriebenen Modellen stehen innersprachliche Objekte als Netzwerkknoten im Mittelpunkt. Auch bei ihnen haftet der Auswahl der Kanten eine gewisse Willkür an. Für meine Zwecke taugen aus den in 10.9 geschilderten Gründen und in Übereinstimmung mit [Ke07] das Word Web oder verwandte Ansätze nur schlecht, so daß ich ein neues Modell mit dem Namen „*Dynamisches Sprachnetz*“ entwickelt habe, dessen Einzelheiten in diesem Kapitel begründet werden sollen.

Zunächst möchte ich begründen, warum ich nicht die Netzwerke in einem existierenden Sprachsystem im Sinne der Klassifikation aus Unterkapitel 10.1 berücksichtigt sondern ein eigenes Modell, ein Netz von Sprachnutzern, entwickelt habe:

- Eine gewisse Willkür in der Abgrenzung zwischen dem, was explizit modelliert werden soll, und dem, von dem abstrahiert wird, ist bei jeder Modellierung unvermeidlich. Mir scheint aber die Frage, was es etwa beim Word Web heißt, daß ein Wort ein anderes beeinflusst, aus sprachwissenschaftlicher Sicht so schwierig zu sein, daß dort jedes einigermaßen einfache Modell in seiner Aussagekraft stark beschränkt ist. Diese Ansicht wird meines Erachtens auch dadurch gestützt, daß es eben in der Literatur über das Word Web keinen Konsens über die Interpretation der Kanten gibt. Ich möchte aber noch einmal unterstreichen, daß die Forschungen zu innersprachlichen Netzen trotzdem zu gewichtigen Erklärungen sprachlicher Phänomene kommt. Daß zum Beispiel das Word Web trotz anfechtbarer Grundannahmen zu erstaunlich genauen Ergebnissen kommt (siehe [DM03]), das zeigt noch einmal, wie leistungsfähig der Netzansatz ist.
- Der Begriff der Dynamik ist bei den innersprachlichen Netzen anders als er von mir benötigt wird. Dort findet eine Dynamik, soweit sie ein Wort betrifft, nur statt, indem ein Wort plötzlich da ist. Weder die Gesetze der Übernahme noch die weitere Entwicklung, wie dieses Wort genutzt wird, werden betrachtet. Dies ist auch der Kern des von mir in Unterkapitel 10.9 zitierten Argumentes von Ke, nach dem die innersprachlichen Netze bisher vor allem für synchrone Betrachtungen geeignet seien.
- Insbesondere verschwindet in diesen Modellen ein einmal eingedrungenes Wort nie wieder. Aber gerade der reversible Sprachwandel soll ja in dieser Arbeit untersucht werden.

- In epidemiologischen Anwendungen dynamischer Netze gibt es vom Grundgedanken, der Interpretation von Knoten und Kanten, her verwandte, in der Wissenschaft akzeptierte Modelle, die die eben angesprochene Dynamik in den Knotenzuständen während der Ausbreitung und der weiteren Entwicklung besser abbilden. Auch die epidemiologischen Modelle können natürlich nicht einfach nur übernommen werden. Es gilt vielmehr, sie an unsere Zwecke anzupassen.

Insbesondere die drei letzten Argumente sind für mich entscheidend. Ich habe ja auch schon, als ich den Inhalt von [Ke07] referiert habe, einige Male darauf hingewiesen, daß die Ausführungen dort über Netze, die sich mehr mit der Darstellung menschlicher Kommunikation beschäftigen, und die Gründe für die Anwendung solcher Netze ebensogut auf das Dynamische Sprachnetz hätten gemünzt worden sein können.

In Übereinstimmung mit epidemiologischen und soziologischen Modellen werde ich künftig Knoten als Abstraktion von Menschen betrachten, deren Kommunikation über die Kanten stattfindet. Anstelle einzelner Sprecher könnte man sich unter den Knoten auch Cluster von Sprechern oder andere Handelnde, etwa Massenmedien, vorstellen. Mit solchen Überlegungen ließe sich beispielsweise auch der Einfluß von Gruppenzwängen oder Massenmedien (durch Einfügen von Knoten mit extrem hohem Grad) modellieren.

In den Kanten kann natürlich nicht jeder einzelne Sprechakt berücksichtigt werden. Eine Verbindung bedeutet nur, daß eine gewisse, zumindest sporadische Kommunikation stattfindet. Ebenso bedeutet der Zustand eines Knotens nicht den Sprachgebrauch während eines einzelnen Sprechaktes sondern kennzeichnet nur den überwiegenden Sprachgebrauch also eine generelle Neigung zu einem bestimmten Sprachgebrauch. Insbesondere vernachlässige ich für den Fall, daß die betrachteten Änderungen lexikalischer Art sind, einen Aspekt, der bei Lehnwörtern eine bedeutende Rolle spielt: den des semantischen Wandels. Lehnwörter werden aus semantischer Sicht so gut wie nie 1:1 aus der Ursprungssprache übernommen. So werden von polysemen Lexemen in der Gebersprache häufig nur monoseme Verbindungen von Ausdruck und Inhalten übernommen, beziehungsweise nur eine kleinere Zahl von Inhalten als in der Ursprungssprache. Und auch später geht ein Lehnwort im Laufe seiner Nutzung bedeutungsmäßig in der Regel eigene Wege. Darüber findet sich mehr in Unterkapitel 15.1. Das ist aber nicht nur ein notwendiges Übel meines Ansatzes sondern angemessen, da auch bei meinen späteren Auswertungen der Lehnwortdatenbank nur auf das Vorhandensein eines Wortes, eine Disposition zu seiner Verwendung, Bezug genommen wird und nicht auf semantischen Wandel. Insofern passen die entwickelte Theorie und der experimentelle Befund vom methodischen Ansatz her zusammen.

Natürlich gilt auch für das jetzt angepeilte Dynamische Sprachnetz wie für alle anderen in dieser Arbeit erwähnten Modelle, daß sie keine genauen quantitativen Aussagen liefern können. Die Ergebnisse solcher Untersuchungen können nur idealtypischer Art sein.

Alle diese Aspekte müssen dann auch später bei der experimentellen Überprüfung des Modells anhand der Lehnwörterdatenbank berücksichtigt werden.

11.1 Designentscheidungen für das Dynamische Sprachnetz

Nach der Frage, was überhaupt modelliert werden soll, muß geklärt werden, wie dies geschehen soll. Durch die bisherigen Untersuchungen ist es so weit, daß zwei wichtige Fragen bezüglich der technischen Details beantwortet werden können:

- Welche Eigenschaften muß ein auf Netzwerken beruhendes Modell zur Untersuchung einer sprachlichen Neuerung (nach dem gegenwärtigen Stand der Wissenschaft) haben? Dieses Modell soll außerdem möglichst einfach sein.
- Wie kann ein solches Modell realisiert werden, damit wir überhaupt eine Chance haben, wirklichkeitsnahe Ergebnisse zu bekommen?

Die nächste naheliegende Frage, nämlich die der experimentellen Überprüfung, also ob die im Modell erzielten Ergebnisse tatsächlich der Wirklichkeit entsprechen, wird später in Kapitel 16 beantwortet.

Wie in [PV02] zu ersehen ist, führen korrelierte Netzwerke bei der Untersuchung epidemiologischer Vorgänge auf ihren Knoten (in jenem Falle im SIS-Modell) leicht zu paradoxen Phänomenen, wenn die Systemparameter nicht sehr sorgfältig gewählt werden, gerade bei dem uns interessierenden Wert, der Frage nach der Dauer einer Epidemie, also den Fragen:

- Wird der Ausbruch endemisch?
- Wie lange dauert es im Schnitt, bis die Epidemie wieder verschwindet?

Deshalb liste ich hier noch einmal die Desiderata für ein angemessenes Modell auf. In den Klammern habe ich dabei die Art der Anforderung vermerkt, ob es um die Erzeugung des Netzes an sich geht, um Eigenschaften des Ergebnisses dieses Erzeugungsprozesses oder um die Dynamik auf dem Netz.

- preferential linking (Anforderung an die Erzeugung des Netzes) mit der Folge eines Potenzverteilungsgesetzes,
- Zusammenhang des gesamten entstehenden Netzes (Anforderung an die Erzeugung des Netzes), also die Abwesenheit von isolierten Individuen oder Gruppen von Individuen,
- Hohe Gradkorrelation (Anforderung an das Netz) mit der Folge eines hohen Clusteringkoeffizienten,
- aging (dynamische Anforderung an das Netz und an den Prozeß),
- gleichzeitige Veränderung von Netz und epidemiologischem Prozeß,
- SIS- oder SIR-Modell (dynamische Anforderung nur an den Prozeß).

Alle diese Forderungen, die im bisherigen Verlauf herausgearbeitet wurden, sind in der Literatur einzeln und auch in Teilkombinationen schon betrachtet worden, bisher aber, soweit mir bekannt ist, noch nie in einer vollständigen Kombination aller hier als notwendig erkannten Eigenschaften.

Wie ich schon in Unterkapitel 8.5 verdeutlicht habe, hat die dort geschilderte und in [BLT05] mathematisch untersuchte Variante des ACL-Modells die Eigenschaft, daß sie Netze erzeugt, die in vielen Parametern den hier zu untersuchenden sozialen Netzen nahe kommen. Alle anderen vergleichbaren Modelle zeigen nicht die richtige Struktur des Netzwerkes oder sind zu umständlich zu implementieren. Das gilt insbesondere für die oben genannten Modelle AEMR und RHL. Ich werde stattdessen mit Hilfe einer Kombination aus mathematischen, sprachlichen und schlicht anschaulich-logischen Argumenten einfachere Übergangsregeln ableiten. Das Ergebnis steht in manchem dem AEMR nahe.

Nach der Erörterung der Netzwerkstruktur möchte ich noch auf einige Einzelheiten der Dynamik auf dem Netz eingehen:

- Obwohl es sich bei der Ausbreitung sprachlicher Neuerungen um ein tendenziell endemisches Phänomen handelt und obwohl ich oben angedeutet habe, daß solche Phänomene oft durch das SIS-Modell untersucht werden, möchte ich zumindest im Grundsatz vom SIR-Modell ausgehen. Dabei werde ich einen Trick benutzen, der sich in vergleichbaren Fällen schon bewährt hat und auf den auch die Doppelinterpretation des R in SIR als „recovered“ und als „removed“ schon hindeutet: In diesem Zustand sollen sowohl verstorbene Individuen zusammengefaßt werden als auch solche, die eine Neuerung zeitweilig verwandt, diese dann aber wieder abgelegt haben. Durch diesen Ansatz wird also beispielsweise das nur temporäre Vorhandensein von Modewörtern mit erfaßt. Das Ergebnis erfordert im Dynamischen Sprachnetz an einer Stelle delikate Implementierungsdetails, siehe Abschnitt 14.2, und kommt dann insgesamt de facto dem SIS-Modell nahe.
- Bei Lehnwörtern (und auch anderen Neuerungen) kann unter Umständen keine *alte* Form angenommen werden, etwa wenn mit dem Lehnwort eine neue mit ursprünglichen Wörtern nicht zu benennende (oder nicht benannte) Realie bezeichnet wird. Trotzdem kann die entlehnte Bildung später durch eine Bildung verdrängt werden, die der entlehnenden Sprache entstammt. Für das Ausscheiden des Lehnwortes ist es im uns interessierenden Zusammenhang unerheblich, ob diese Verdrängung auf die letzte geschilderte Weise geschieht oder durch ein Wiedererstarken einer alten Form. Deshalb ist die Einführung eines zusätzlichen Zustandes, beispielsweise „verdrängt durch einen Neologismus“ oder etwas Ähnliches, überflüssig.
- Damit die Dauer, während der ein Individuum sich einer Neuerung bedient, nachgebildet werden kann, gibt es verschiedene Möglichkeiten. Man kann unter einigen etablierten Methoden wählen und
 1. eine feste Konstante vereinbaren, die ganz universell angibt, welche Zeit zwischen dem Übergang in den Zustand I und dem Übergang in den Zustand R verstreicht.

2. eine Wahrscheinlichkeit p universell vereinbaren, mit der bei jedem Schritt ein Übergang aus dem Zustand I in den Zustand R erfolgt. Die Verteilung der Aufenthaltsdauern im Zustand I gehorcht dann der „geometrischen Verteilung“. Dies ist leider nicht sehr realistisch.⁶⁹
3. eine realistischere Verteilung der Aufenthaltsdauern im Zustand I erzwingen. Diese könnte beispielsweise der Verteilung der Lebenserwartung entsprechen. Ein solches Vorgehen ist unnötig aufwendig, und es wird außer bei sehr detaillierten epidemiologischen Untersuchungen in der Literatur vermieden.

In Übereinstimmung mit den Ratschlägen in [BF99], wo der Leser auch Näheres zu diskreten Übergangswahrscheinlichkeiten erfahren kann, werde ich den zweiten Ansatz wählen. Mit ihm werden wir hinreichend signifikante Ergebnisse bei vertretbarem Aufwand erzielen. Mit dieser Übergangswahrscheinlichkeit finden wir einen eminent wichtigen Systemparameter. Die Übergänge von I in R werden damit die Methode sein, mit der das aging nachgebildet wird. Ein Knoten im Zustand R soll bei der Bestimmung der Übergänge seiner Nachbarn keine Rolle mehr spielen und auch selber nicht mehr aus diesem Zustand herauskommen. Das gewählte Verfahren vereinfacht das Modell, und es hat sowohl aus sprachlicher Sicht als auch auf der Grundlage der epidemiologischen Erfahrungen einen Sinn: Werden die Änderungen als Epidemien aufgefaßt, entspricht das genannte Verfahren einer Immunisierung. Sprachlich kann ein solcher Knoten als ein Individuum aufgefaßt werden, das die Änderung eine Zeit lang, dem Reiz des Neuen huldigend, benutzt hat, das dann aber endgültig zu seinem ursprünglichen Sprachgebrauch zurückgekehrt ist.

- Auch bei dem Übergang aus dem Zustand S in den Zustand I haben wir prinzipiell die gleichen Möglichkeiten. Hier muß eben nur zusätzlich bedacht werden, daß dieser Übergang für jedes Paar verbundener Knoten betrachtet werden muß. Auch für diesen Fall ist es sinnvoll, eine Übergangswahrscheinlichkeit ν anzunehmen, die den Übergang beschreibt, wenn genau *ein* Nachbar im Zustand I ist. Damit liegt hier ein weiterer wichtiger Systemparameter vor. Zu bemerken ist noch, daß das ν hier nicht genau das aus 9.4 ist. Jenes ν beschreibt die allgemeine Infektionswahrscheinlichkeit, während hier nur die Infektionswahrscheinlichkeit von Individuum zu Individuum gemeint ist.

Schließlich sollen, nur damit ich mich gegen noch komplexere Modelle abgrenzen kann, einige Punkte genannt werden, die hin und wieder in der Literatur zu finden sind, deren Zweck aber zumeist eine tiefgehende Detailuntersuchung ist, teilweise für praktische Anwendungen, teilweise weil mit Hilfe solcher zusätzlicher Parameter interessante mathematische Fragestellungen in allerdings anderweitig vereinfachten Modellen untersucht werden können. Für unsere Ziele wäre ein Modell, das auch noch die folgenden Parameter inkorporiert, unnötig kompliziert. Aus einer Fülle von Beispielen dafür seien nur wenige herausgegriffen:

⁶⁹ Statt einer Definition gebe ich hier nur ein Beispiel: Sei beispielsweise $p = 1/2$, und sei $p(n)$ die Wahrscheinlichkeit, daß ein Individuum n Zeitschritte im Zustand I verharrt, so besagt die geometrische Verteilung, daß gilt:

$$p(n) = 1/2^n$$

- *weitere Zustände neben SIR*

Hier handelt es sich um ein in der Epidemiologie übliches Verfahren, mit dessen Hilfe realistischere Ergebnisse erzielt werden sollen. Als wichtigstes Beispiel einer solchen Zustandserweiterung sei eine Zerlegung des I-Zustandes genannt in einen Zustand, bei dem ein Individuum bereits infiziert ist und die Epidemie weitergeben kann, wobei die Krankheit aber noch nicht ausgebrochen ist, und in einen Zustand, bei dem die Krankheit offen zutage tritt.

- *Einführung eines Gedächtnisses in den Knoten*

Ein Gedächtnis haben wir schon bei der Regel des höchsten Lohnes in 9.6.2 und beim Naming Game in 10.9 kennengelernt. Mit Hilfe eines solchen Gedächtnisses könnten beispielsweise Anforderungen modelliert werden wie: „Die Krankheit dauert mindestens drei Zeitschritte.“ Weiter unten werden wir sehen, daß solche Anforderungen im Dynamischen Sprachnetz eine untergeordnete Rolle spielen. In der Theorie der formalen Sprachen, und darunter könnten wir unsere Überlegungen bei großzügiger Auslegung des Begriffs auch subsumieren, spielt der Unterschied zwischen gedächtnislosen und mit Gedächtnis behafteten Formalismen eine bedeutende Rolle. Auf diese Weise läßt sich zum Beispiel der Unterschied zwischen regulären und kontextfreien Sprachen erklären. Daß sich hier eine große Spielwiese für Mathematiker ergibt, das ist offensichtlich. Als eins von vielen Beispielen sei [KE01] genannt. Aber auch hier gilt das schon mehrfach gesagte: Der zusätzliche große Aufwand würde in unserem Fall nicht durch einen entsprechenden Informationsgewinn gerechtfertigt werden.

11.2 Anforderungen an Statische Zustandsüberführungsregeln

Es ist inzwischen sicher deutlich geworden, wieso ich oben, in Abschnitt 3.4, geschrieben habe, daß die Zustandsüberführungsregeln das Herzstück des Modells bilden. Deswegen ist es unbedingt erforderlich, sich an dieser Stelle bei der Auswahl des dann tatsächlich verwandten Modells Gedanken darüber zu machen. In der oben versprochenen Kombination von mathematischen, sprachlichen und logisch-anschaulichen Argumenten, werde ich mögliche Antworten auf einige Fragen zu finden versuchen: Welche Zustandsüberführungsregeln sind in einem mathematischen Modell überhaupt sinnvoll und möglich? Solche Regeln sollen

- das sprachliche Geschehen widerspiegeln,
- mathematisch angemessen, insbesondere auch möglichst einfach⁷⁰ sein,
- möglichst einfach zu implementieren sein,
- den in Kapitel 4 entwickelten Metaregeln Monotonie, Angemessenheit und Symmetrie⁷¹ genügen,

⁷⁰ Ich erinnere noch einmal an die Bemerkungen, die ich in Unterkapitel 2.5 über die Einfachheit in einem guten wissenschaftlichen Modell gemacht habe.

- für alle Knoten gleich sein,
- die Zustände S und I gleich behandeln.

Alle diese Punkte sind evident oder wurden schon erörtert. Nur die beiden letzten bedürfen möglicherweise einer Erklärung:

Zur Gleichheit der Knoten: Im Dynamischen Sprachnetz unterscheiden sich die Individuen nur durch ihre Zustände S, I oder R, sowie die Zahl ihrer sozialen Kontakte, nicht aber durch die Form der Zustandsüberführungsregeln. Ich habe insbesondere darauf verzichtet, zwischen mehr oder weniger konservativen Sprechern zu unterscheiden. Auch eine solche Unterscheidung taucht nur gelegentlich in der Literatur auf. Hier könnte durchaus ein Weg für weitere Forschung liegen, etwa was den Einfluß bewußter Sprachpflege angeht.

Zur Gleichbehandlung der Zustände: Ich gehe davon aus, daß ein Lehnwort, sofern es einmal übernommen worden ist, von allen Sprechern akzeptiert wird wie ein eigenes. Oft wissen viele Menschen sehr bald nicht einmal mehr, daß das Wort einst fremd war. Darauf, daß beispielsweise die Erkennbarkeit eines deutschen Lehnwortes keinen Einfluß auf die Häufigkeit seiner Ersetzung durch ein polnisches Eigenwort hat, wird in [HM01] hingewiesen⁷². Dort untersuchen die Autoren auch den Begriff der „Erkennbarkeit“ näher und zeigen die Willkür in den Ersetzungsvorgängen, soweit die Erkennbarkeit betroffen ist, auf. Das alles bedeutet, daß es im Modell keinen wie auch immer gearteten Bonus für Eigenwörter gibt. Mathematisch folgt aus dieser Festlegung, daß die Übergangsfunktion invariant gegen eine Vertauschung der Zustände sein muß.⁷³

Da die Metaregeln bisher nur für isotrope Netze formuliert worden sind, mache ich noch einige Bemerkung zu der Form, die sie im vorliegenden Kontext annehmen: Eine wichtige Bemerkung ist der folgende

Satz: Eine Übergangsregel f in einem System mit nur zwei Zuständen ist genau dann symmetrisch, wenn sie semitotalistisch ist.

Beweis: In Unterkapitel 4.4 habe ich ja schon für isotrope Netze bemerkt, daß semitotalistische Regeln symmetrisch sind. Die dort aufgeführten Argumente gelten auch im Fall beliebiger Netzwerke. Für die nun folgende Analyse ist jedoch die Umkehrung wichtiger, daß nämlich symmetrische Regeln semitotalistisch sind. Daher müssen wir uns nur noch mit solchen Regeln beschäftigen. Sei also f eine symmetrische Übergangsregel. Sei ferner eine feste Zelle z gegeben. Nach Definition gilt für alle Zustandsvektoren v in der Nachbarschaft

⁷¹ Ich möchte darauf hinweisen, daß das vorliegende Modell sich insbesondere durch die Metaregel der Symmetrie von epidemiologischen Modellen in der Medizin unterscheidet, wo sie offensichtlich unangebracht ist. Ich werde Symmetrien in der Zukunft aber ausführlich benutzen.

⁷² „Alte deutsche Lehnwörter, die im Polnischen des 19. Jhs. oder des frühen 20. Jhs. außer Gebrauch geraten sind, weisen also unabhängig von der Frage, ob sie noch als deutsch (fremd) zu erkennen sind oder nicht, ähnliche Phänomene des lexikalischen Wandels auf wie nicht-entlehntes Material.“ [HM01]

⁷³ Es muß also beispielsweise gelten

$$f(0,1,1) = 1-f(1,0,0)$$

Da zum Zustand z der inverse Zustand sich durch $1-z$ ausrechnet, heißt das anschaulich: Wenn bei allen Nachbarn die Zustände S und I (oder 0 und 1) getauscht werden, muß auch der Folgezustand getauscht werden; daher steht auf der rechten Seite $1-f(1,0,0)$ und nicht $f(1,0,0)$.

von z und für alle Permutationen $perm$ der Nachbarschaft die Gleichung $f(v) = f(perm(v))$. Seien jetzt v ein beliebig gewählter Zustandsvektor in der Nachbarschaft von z und n die Anzahl der Nachbarn von z im Zustand I , also die Anzahl der Komponenten von v im Zustand I . Sei v' ein weiter Zustandsvektor in der Nachbarschaft von z mit n Komponenten im Zustand I . Da es nur die beiden Zustände 0 und 1 gibt und die Länge der beiden Vektoren gleich ist, unterscheiden sich v und v' nur durch die Stellen, an denen die Nullen und Einsen stehen aber nicht durch die Anzahl der Nullen und Einsen. Das heißt, es gibt eine Permutation $perm$ der Nachbarschaft mit $v' = perm(v)$. Also gilt wegen der Symmetrie

$$f(v) = f(perm(v)) = f(v')$$

Da v und v' beliebig gewählt waren mit der Eigenschaft dieselbe Anzahl von Komponenten im Zustand I zu besitzen, heißt dieses, daß f semitotalistisch ist.

Qed

11.3 Detaillierte Analyse der Zustandsüberführungsregeln

Bevor die eben gewonnenen Erkenntnisse über metaregelgerechte Zustandsüberföhrungsfunktionen angewandt werden können, ist noch eine weitere kleine Vorbemerkung nötig: Weiter oben, bei isotropen Graphen, war die Lage die, daß sich der Zustand der Zelle z zu einem Zeitpunkt t , mit $z(t)$ bezeichnet, in Abhängigkeit von den Zuständen der Nachbarn von z , mit z_1, \dots, z_n bezeichnet, ändert. Dann wird die Zustandsänderung durch die Funktion f beschrieben mit

$$z(t+1) = f(z(t), z_1(t), \dots, z_n(t))$$

Es ist dort also insbesondere so, daß für alle Zellen die gleiche Übergangsfunktion f gilt. Das sollte auch in unseren sich dynamisch ändernden Netzen der Fall sein. Dabei bekommen wir ein mathematisches Problem:

Sei die Menge der möglichen Zustände eines Knotens mit Z bezeichnet. Dann hat f die Funktionalität

$$f: Z^{n+1} \rightarrow Z$$

Dabei sei n die Anzahl der Nachbarn. Es ist das Wesen isotroper Graphen, daß alle Knoten dieselbe Anzahl von Nachbarn haben. Damit ist n wohlbestimmt. In dem jetzt behandelten Modell beliebiger Netzwerke entfällt diese Voraussetzung. Scheinbar benötigen wir mehrere Funktionen f_n der Form

$$f_n: Z^{n+1} \rightarrow Z$$

Dabei ist wieder n die diesmal variable Anzahl der Nachbarn. Was sollte dann die Forderung nach einer einheitlichen Übergangsfunktion, also letztlich nach einer Gleichheit der Knoten, bedeuten?

Eine erste Vereinfachung des Problems kann durch die Forderung nach Gleichbehandlung aller Zustände erzielt werden. Das zeigt das folgende Argument:

Angenommen, die Übergangsfunktion sei schon definiert für Knoten im Zustand 0 . Das heißt für alle Nachbarn mit Zuständen z_1, \dots, z_n sei der Funktionswert schon bekannt. Es gebe also ein x mit

$$f(0, z_1, \dots, z_n) = x$$

Sind alle diese Werte bekannt, dann sind auch alle Werte für Knoten im Zustand 1 bekannt, denn wegen der Invarianz gegen Vertauschung der Zustände gilt ja

$$f(1, 1-z_1, \dots, 1-z_n) = 1-x = 1-f(0, z_1, \dots, z_n)$$

und wenn (z_1, \dots, z_n) alle möglichen Kombinationen von 0 und 1 durchläuft, tut dies auch $(1-z_1, \dots, 1-z_n)$. Das bedeutet, daß, wenn alle Werte für einen Zustand bekannt sind, auch die Werte für den anderen Zustand festliegen. Daher kann unsere Definition der Übergangsfunktion unabhängig vom gegenwärtigen Zustand des Knotens festgelegt werden. Deshalb werde ich künftig, auch wenn ich es nicht ausdrücklich schreibe, davon ausgehen, daß der Ursprungszustand 0 ist, und das erste Argument in f einfach unterdrücken. Die Tatsache, daß an Zustandswechseln Knoten im Zustand R nicht beteiligt sind, und die Forderung nach Symmetrie weisen nun den eigentlichen Ausweg aus dem Problem mit der Stelligkeit von f . Es kommt ja gar nicht auf die genauen Zustände der einzelnen Nachbarn an. Wesentlich zur Berechnung des Nachfolgezustandes eines gegebenen Knotens i sind damit zunächst nur noch zwei Parameter:

- die Anzahl der Nachbarn von i , was oben in der Stelligkeit der Funktionen f_n ausgedrückt worden ist,
- wie viele der Nachbarn von i sich im komplementären Zustand befinden.

Angenommen, eine Zelle z habe zum Zeitpunkt t genau n Nachbarn, von denen k im Zustand 1 sind, so kann f ausgedrückt werden in der Form

$$z_{t+1} = f(n, k)$$

Dabei ist natürlich z_{t+1} der Nachfolgezustand von z zum Zeitpunkt $t+1$.

Nach dieser Einsicht ist die Behandlung der Monotonie bei metaregelgerechten Zustandsüberföhrungsfunktionen einfach.

Bemerkung: Eine semitotalistische Übergangsregel f in einem System mit nur zwei Zuständen, genügt der Metaregel der Monotonie genau dann, wenn für alle Zustände z und alle n_1, n_2 gilt:

$$n_1 \leq n_2 \Rightarrow f(z, n_1) \leq f(z, n_2)$$

Diese Bemerkung ist unmittelbar einsichtig; auf einen Beweis wird verzichtet.

Die Metaregel der Angemessenheit hat folgende mathematische Interpretation:

Sind ein Knoten i und alle seine Nachbarn zu einem Zeitpunkt t im gleichen Zustand z , so ist i auch zum Zeitpunkt $t+1$ im Zustand z .

Diese Interpretation ist auch aus sprachlicher Sicht sinnvoll, da Veränderungen im allgemeinen nur durch Kontakt weitergegeben werden können; und wenn sich alle Kontaktpersonen gleich verhalten, wird normalerweise keine Änderung im Sprachverhalten auftreten.

In der Situation einer metaregelgerechten Zustandsüberföhrungsfunktion kann ich in 11.4 eine bedeutsame Folgerung ziehen. Diese werde ich nur unter der Zusatzannahme erklären, daß der betrachtete Knoten i wieder im Zustand 0 ist. Der Fall des Zustandes 1 wird dann wegen der Symmetrie der Zustände analog behandelt. Also bildet meine Zusatzannahme über den Zustand keine Einschränkung.

Es habe also eine Zelle n Nachbarn, von denen k im Zustand 1 sind, so daß f in der obigen Form mit zwei Argumenten ausgedrückt werden kann.

Wegen der Angemessenheit gilt $f(n,0) = 0$. Jetzt gibt es offensichtlich zwei Fälle:

- Es gibt eine Stelle mit $f(n,k) = 0$ und $f(n,k+1) = 1$, wobei gilt $k < n$.
- Es gilt durchgängig $f(n,k) = 0$ für alle k mit $k \leq n$.

Der zweite Fall bedeutet, daß (für dieses und sinnvollerweise dann für alle n) der ursprüngliche Zustand unabhängig vom Zustand der Nachbarn beibehalten wird. Solche Übergangsregeln sind uninteressant; dabei kann es überhaupt keine Veränderungen und somit keine Epidemieausbreitungen geben. Jetzt soll der erste Fall genauer untersucht werden.

11.4 Die Umschaltfunktion

Wegen der Monotonie kann es keine Stelle mit $f(n,k) = 1$ und $f(n,k+1) = 0$ sowie $k < n$ geben. Das bedeutet, daß jedem n genau ein Wert $u(n)$ zugeordnet werden kann, den ich „Umschaltwert“ nennen möchte, an dem nämlich gerade die Definition dieses Falles gilt:

$$f(n,u(n)) = 0, \quad f(n,u(n)+1) = 1$$

Auf diese Weise wird eine Funktion u , die „Umschaltfunktion“ definiert⁷⁴. Einige ihrer Eigenschaften sollen jetzt hergeleitet werden:

- Der Definitionsbereich von u sind die natürlichen Zahlen, der Wertebereich ist eine Teilmenge der natürlichen Zahlen.

Die Behauptung folgt für den Definitionsbereich aus der Tatsache, daß wir es hier mit Graphen zu tun haben, die im Prinzip beliebig groß werden können. Daher gibt es auch keine Grenzen für den Grad eines Knotens, also die Anzahl seiner Nachbarn. Die Behauptung für den Wertebereich ist offensichtlich.

⁷⁴ Ich halte mich hier an die auch schon früher benutzte mathematische Konvention, daß die Funktion mit u und der Wert von u an der Stelle n mit $u(n)$ bezeichnet wird. So bedeutet beispielsweise $u = n$, daß $u(x) = n$ an allen Stellen x des Definitionsbereiches von u gilt.

- Für alle n gilt $u(n) < n$. Für alle obigen k gilt $k < n$.

Da $u(n)$ eins dieser k ist, folgt die Behauptung.

- Die Funktion u ist monoton.

Wäre u nicht monoton, so gäbe es laut der Definition der Monotonie eine Stelle mit $u(n) > u(n+1)$. Das hieße anschaulich: Damit ein Zustandswechsel von 0 nach 1 herbeigeführt werden kann, sind für einen Knoten mit n Nachbarn mehr Knoten im Zustand 1 nötig als für einen Zustandswechsel bei einem Knoten mit $n+1$ Nachbarn. Das ist unlogisch.

- Die Funktion u sollte (im wesentlichen) konstant oder linear sein.

Diese Aussage ist keine logisch zwingende Folgerung. Sie läßt sich jedoch mit Praktikabilitätsargumenten aus den vorherigen Punkten ableiten. Ich bitte die Leser zu bedenken, daß wir ja gar keine vollständige Untersuchung anstreben, welche Funktionen in Frage kommen. Wir wollen nur zwei Dinge: Eine Menge von sinnvollen, einfachen, handhabbaren, implementierbaren Funktionen und die mathematisch gesicherte Überzeugung, daß keine wesentliche alternative Wahl einer solchen Funktion übersehen worden ist. Das liefert die folgende Argumentation:

Wegen $u(n) < n$ ist die Funktion u in der Komplexitätsklasse $o(n)$ ⁷⁵. In dieser Klasse liegen einige bekannte Funktionen, etwa die Wurzel-Funktion oder die Logarithmus-Funktion. Die konstanten und die linearen Funktion bezeichnen jedoch die beiden Extreme in $o(n)$; sie sind die am schnellsten (nämlich linear) und die am langsamsten (nämlich konstant, also überhaupt nicht) wachsenden monotonen Funktionen innerhalb dieser Komplexitätsklasse. Zur Begründung, warum ich nur die beiden extremen Funktionenklassen betrachte gilt es, folgende Punkte zu beachten:

- Die Implementierung von Funktionen, deren Wachstumsverhalten zwischen den beiden Extremen liegt, zeigt keine wesentlichen Unterschiede des Systemverhaltens im Vergleich zu den beiden genannten Klassen.
- Es gibt keinen mathematischen oder sprachlichen Grund, einen anderen Verlauf von u anzunehmen. Wodurch sollte sprachlich ein sich wild veränderndes Verhalten entstehen? Aus mathematischer Sicht läßt sich argumentieren, daß dieses keine chaostheoretische⁷⁶ Untersuchung ist, wir uns also nicht in Parameterbereichen bewegen, bei denen kleine Änderungen dramatische Konsequenzen des Systemverhaltens bewirken.

⁷⁵ Ich erinnere an die Definition von Komplexitätsklassen in Unterkapitel 4.6.

⁷⁶ Ein System zeigt *chaotisches Verhalten*, wenn (unter anderem) kleine Änderungen der Parameter zu großen Änderungen des Systemverhaltens führen. In der Chaostheorie werden genau solche Parameterveränderungen untersucht. Hier möchte ich gar nicht untersuchen, ob die Sprache ein chaotisches System ist, obwohl manches darauf hindeutet. Insbesondere bin ich nicht an Variationen der Parameter interessiert, die zu chaotischem Verhalten führen. Das könnte zwar eine wissenschaftlich akzeptable Fragestellung sein. Ich weiß aber nicht, wie ich die Relevanz eines solchen chaotischen Verhaltens an gewissen Parametern beurteilen sollte.

- Eine andere Implementierung von u würde unnötigerweise der oben geforderten Einfachheit der Zustandsüberföhrungsfunktion widersprechen.

Trotzdem handelt es sich hier nicht um einen strengen mathematischen Beweis sondern um eine Plausibilitätsbetrachtung.

- Es sollte Sonderregeln für kleine n geben.

Dieser Punkt in unserer Liste der Eigenschaften der Funktion u präzisiert die relativierende Klammer im letzten Punkt, daß nämlich die Funktion u nur *im wesentlichen* konstant oder linear sein sollte. Er bedarf einer etwas genaueren Analyse: Zunächst einmal widerspricht eine solche Sonderregelung nicht der obigen prinzipiellen Argumentation für Konstanz oder Linearität, da ich meine Schlüsse nur aus der o -Notation gezogen habe. Deren Wesen ist es aber gerade, daß sie das Funktionsverhalten für kleine Werte ignoriert. Für die o -Notation interessiert besonders das Verhalten bei wachsendem n .

Nach dieser Klarstellung gehen wir der Einfachheit halber zunächst von einer konstanten Funktion u aus. Wieder sei der ursprüngliche Zustand 0 . Ist $u = 0$, so gilt auf Grund der Definition von u :

$$f(n,0) = 0, \quad f(n,1) = 1 \quad \text{für alle } n.$$

Das bedeutet, daß ein einziger infizierter Nachbar schon einen Zustandswechsel bewirkt. Damit haben wir hier zunächst eine scheinbar durchaus akzeptable Regel.⁷⁷ Auch der Fall $u = 1$ ist auf den ersten Blick akzeptabel. Er hat jedoch eine Macke, die für den Fall $u = 2$ noch deutlicher wird. Es gilt dann nämlich

$$f(n,2) = 0, \quad f(n,3) = 1 \quad \text{für alle } n.$$

Das bedeutet anschaulich, daß ein Zustandswechsel nur durch das Zusammenwirken von mindestens drei infizierten Nachbarn bewirkt werden kann. Die normalerweise unerwünschte Konsequenz dessen ist, daß Knoten mit einem Grad kleiner als drei überhaupt keinen Zustandswechsel mehr zulassen; in der sprachlichen Interpretation: Individuen mit geringer Kommunikation ändern ihr sprachliches Verhalten überhaupt nicht. Nun ist das Phänomen von Einsiedlern oder isolierten Sprachgemeinschaften, die, zumindestens bezüglich des untersuchten Aspekts, einen alten Sprachstand konserviert haben, obwohl er sonst überall ausgestorben ist, wohlbekannt. Aus zwei Gründen möchte ich solches Verhalten in die Modelle jedoch nicht aufnehmen:

- Das angesprochene Phänomen ist grundsätzlicher Art. Für *alle* Individuen mit nur zwei Verbindungen gilt für *alle* Zeiten, daß sie immun gegen Veränderungen sind. Solche Immunitäten kommen zwar auch vor, etwa in oberitalienische Gemeinden, die altertümliches Deutsch sprechen. Aber ein solches Verhalten stellt doch eher eine große Ausnahme dar. Von der modellmäßigen Ausnahmslosigkeit, die sich ergäbe, kann keine Rede sein.

⁷⁷ Wegen der Gleichberechtigung der Zustände führt eine solche Regel fast notwendigerweise zu in jedem Schritt alternierenden Zuständen. Dieses eindeutig unerwünschte Verhalten behandle ich später, wenn ich mich mit Latenzen beschäftige.

- Die Zahl solcher durch Isolation immun gewordener Gruppen ist vermutlich so gering, daß sie die Ausbreitung sprachlicher Veränderungen in der Realität nicht wesentlich beeinflussen dürften. Man würde sicher davon sprechen, daß sich die Gruppe von der allgemeinen Entwicklung der Hochsprache abgekoppelt hat. Das Ziel meiner Untersuchungen ist aber nicht das Studium solcher Randphänomene sondern das Studium der Hauptströmung der Hochsprache.

Im übrigen sind meine Implementierungen so entwickelt, daß eine nachträgliche Aufnahme strikt konstanter (ebenso wie strikt linearer) Funktionen leicht, nämlich durch eine Änderung nur an einer einzigen Stelle in den Programmen, bewirkt werden kann.

Was ich hier für Konstanten erläutert habe, gilt prinzipiell ähnlich auch für lineare Funktionen. Wir werden weiter unten jedoch sehen, daß aus mathematischen Gründen in diesem Fall keine Sonderregeln für kleine Werte nötig sind. Aus der genannten Motivation habe ich bei der Funktionenklasse der konstanten Funktionen für kleine n eine Absenkung der Zahl der für einen Zustandswechsel benötigten infizierten Nachbarn ins Modell eingebaut. Diese wird noch erläutert werden.

Typische Verläufe sinnvoller Umschaltfunktionen werden in Abbildung 32 bis Abbildung 34 angegeben. Man kann daraus ersehen, daß wir damit zwar bei weitem nicht alle möglichen Umschaltfunktionen erfaßt haben aber einen angemessenen Teil davon.

Damit ist unsere Liste mit Eigenschaften der Umschaltfunktionen abgeschlossen. Wir können demnach Übergangsfunktionen f der Form

$$f(x, u(x)) \quad \text{mit} \quad u(x) = c \quad \text{oder} \quad u(x) = dx$$

(zuzüglich Sonderregeln für kleine Werte) als sinnvolle Auswahl aus der Menge der möglichen Funktionen zu identifizieren. Welche Werte für c, d sind sinnvoll? Ich verweise erneut darauf, daß dieses keine chaostheoretische Untersuchung ist. Deshalb beschränke ich mich auf diskrete Werte⁷⁸. Diese Einschränkung ist nicht wesentlich, sie macht aber die Programmierung und die Auswertung der Simulationen einfacher.

11.4.1 Konstante Umschaltfunktionen

Wie ich oben schon bemerkt habe, bedeutet der Fall $u = n$ anschaulich, daß ein Zustandswechsel nur durch das Zusammenwirken von mindestens $n+1$ infizierten Nachbarn bewirkt werden kann. Es ist zu erwarten, daß bei einem zu großen n eine Infektion sich nicht mehr ausbreiten kann. Testsimulationen zeigen, daß dieser Effekt in der Tat spätestens bei $n = 5$ eintritt.

⁷⁸ Deshalb werde ich in Übereinstimmung mit den Konventionen in der Informatik die Parameter jetzt nicht mehr mit c sondern mit n bezeichnen.

11.4.2 Lineare Umschaltfunktionen

Aus der obigen Ungleichung $u(n) < n$ folgt, daß die Steigung (das d von oben) von f höchstens 1 betragen kann. Aus Gründen der Einfachheit wähle ich die Werte $1/2, 1/3, \dots$. Bekanntlich nähern sich lineare Funktionen, deren Steigung immer kleiner wird, immer mehr einer Konstanten an. In unserem Fall treten schon ab dem Wert $1/5$ in den Simulationen vergleichbare Effekte wie bei Konstanten ein. Deshalb sind solche Parameter für den weiteren Verlauf der Untersuchung irrelevant und werden daher künftig ausgeblendet.

11.5 Spezielle Zustandsüberföhrungsfunktionen

Aus den genannten Gründen habe ich mich im eigentlichen Modell auf einige spezielle Zustandsüberföhrungsfunktionen beschränken können, ohne wesentliche Effekte zu verpassen.

Seien ein Knoten k und ein Zeitpunkt t fest gewählt. Dieser Knoten habe m Nachbarn, von denen v in einem anderen Zustand sind als k . Sei ferner n eine natürliche Zahl mit $1 \leq n \leq 4$. In den beiden hergeleiteten Fällen gelten dann für den Folgezustand z_{t+1} von k , und damit für die Zustandsüberföhrungsfunktion die folgenden Verläufe:

- Verlauf 1: konstantes u

In diesem Falle gilt

$$\begin{aligned} z_{t+1} &= 1, & \text{falls } v \geq n \text{ ist, und} \\ z_{t+1} &= 0, & \text{sonst.} \end{aligned}$$

Diese einfachste aller Situationen wird in Abbildung 30 idealisiert als kontinuierliche Funktion dargestellt: In x-Richtung sind die insgesamt vorhandenen Nachbarn eines Knotens eingetragen, in y-Richtung die Anzahl der zu einem Zustandswechsel benötigten Nachbarn. Wie man sieht, ist diese Zahl unabhängig von der Zahl der tatsächlich vorhandenen Nachbarn.

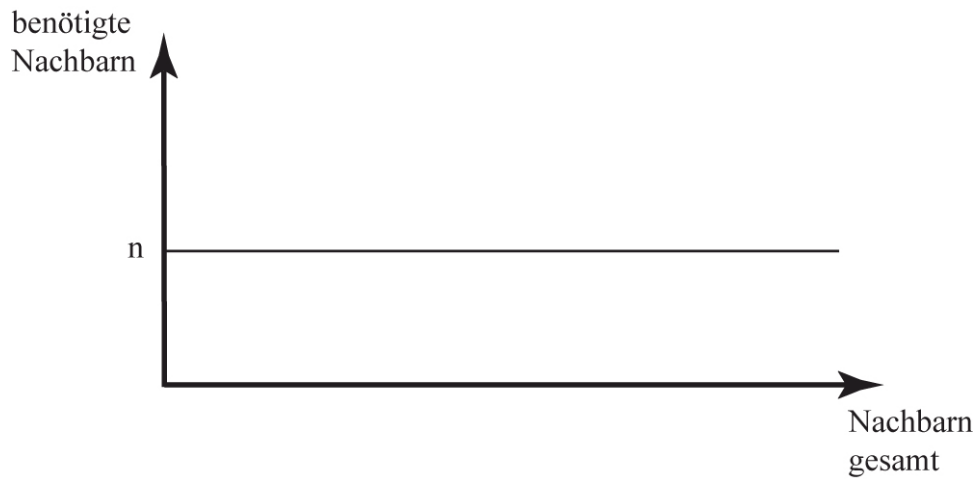


Abbildung 30: Konstante Umschaltfunktion

- Verlauf 2: lineares u

In diesem Falle gilt

$$\begin{aligned} z_{t+1} &= 1, && \text{falls } v \geq \text{ceil}(m / (1+n)) \text{ ist, und} \\ z_{t+1} &= 0, && \text{sonst.} \end{aligned}$$

Der Verlauf für die Steigung $\frac{1}{2}$ ist in Abbildung 31 zu sehen. Die tatsächlichen Umschaltwerte werden durch die Punkte dargestellt. Die durchgezogene Linie illustriert die idealisierte Gerade, die daraus resultiert.

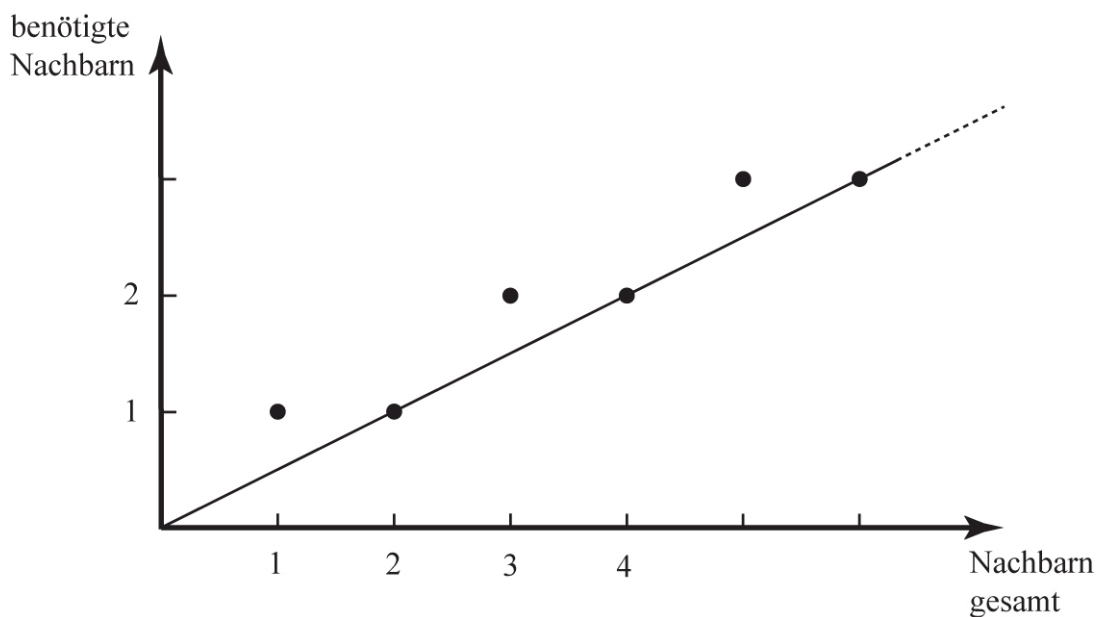


Abbildung 31: Lineare Umschaltfunktion

Dabei ist *ceil* die *ceiling*-Funktion, die jeder reellen Zahl r den Wert $ceil(r)$ zuordnet, die kleinste natürliche Zahl, die größer als r ist (oder gleich r). Der Ausdruck innerhalb der *ceiling*-Funktion folgt aus elementarer Geometrie: Wegen $1 \leq n \leq 4$ liegt der Nenner im Bereich zwischen 2 und 5, also im Bereich derjenigen Nenner, die ich als sinnvoll für lineares u identifiziert habe. Die Werte $1/2, \dots, 1/5$ geben die Steigung von Geraden an, auf denen die Umschaltwerte liegen. Der Umschaltwert für den aktuell betrachteten Knoten wird dann an der Stelle m , der Zahl seiner Nachbarn insgesamt, abgelesen. Die *ceiling*-Funktion kommt dadurch ins Spiel, daß wir uns hier ja nur mit ganzzahligen Werten beschäftigen.

11.6 Sonderregeln für kleine n

Für die Sonderregeln für kleine n gelten die folgenden Überlegungen: Die Absenkung der Schwelle soll natürlich in Abhängigkeit von dem Knotengrad erfolgen. Gesucht ist also eine Zustandsüberföhrungsfunktion, die durch eine Fallunterscheidung (kleine vs. große n) definiert wird. Für den Normalfall (große n) soll die Funktion konstant (oder linear) sein. Ich verdeutliche die Überlegungen wieder anhand eines Beispiels:

Die Funktion u habe wie oben bei meiner Motivation der Sonderregeln die Form $u(n) = 2$ für alle n . Es seien also drei infizierte Nachbarn nötig, damit ein Zustandswechsel erzwungen werden kann. Als unerwünschte Konsequenz dessen habe ich oben schon genannt, daß bei Knoten mit einem Grad kleiner als drei überhaupt keine Zustandswechsel mehr zugelassen werden, und auch bei einem höheren Grad wird ein Zustandswechsel erschwert. Damit dieses Phänomen kompensiert werden kann, soll nun also der Umschaltwert für kleine n herabgesetzt werden. Dies muß nun wieder mit einer möglichst einfachen Funktion geschehen. Wie wir oben gesehen haben, sind die geeigneten Kandidaten für möglichst einfache Funktionen die linearen Funktionen, da die noch einfacheren konstanten Funktionen hier offensichtlich nicht als Modifikation benutzt werden können, also:

$$u(n) = c * n$$

Wie groß sollte dann c sein? Auf Grund der Forderung $u(n) < n$ muß $c < 1$ gelten, und wie oben könnte ich mich auf Werte der Form $1/2, 1/3, \dots$ zurückziehen. Ich werde mich jedoch noch weiter beschränken, nämlich auf $c = 1/2$. Auch an dieser Stelle ist das implementierte Modell so beschaffen, daß sich ein zusätzlicher Parameter leicht hinzufügen ließe. Um die Beschränkung auf $c = 1/2$ zu verstehen, schauen wir uns u an: Es ist

$$\begin{array}{lll} u(n) = ceil(1/2 * n) & \text{für} & 0 \leq n \leq 4 \\ u(n) = 2 & \text{für} & 4 < n \end{array}$$

Das ist in der folgenden Abbildung 32 zu sehen:

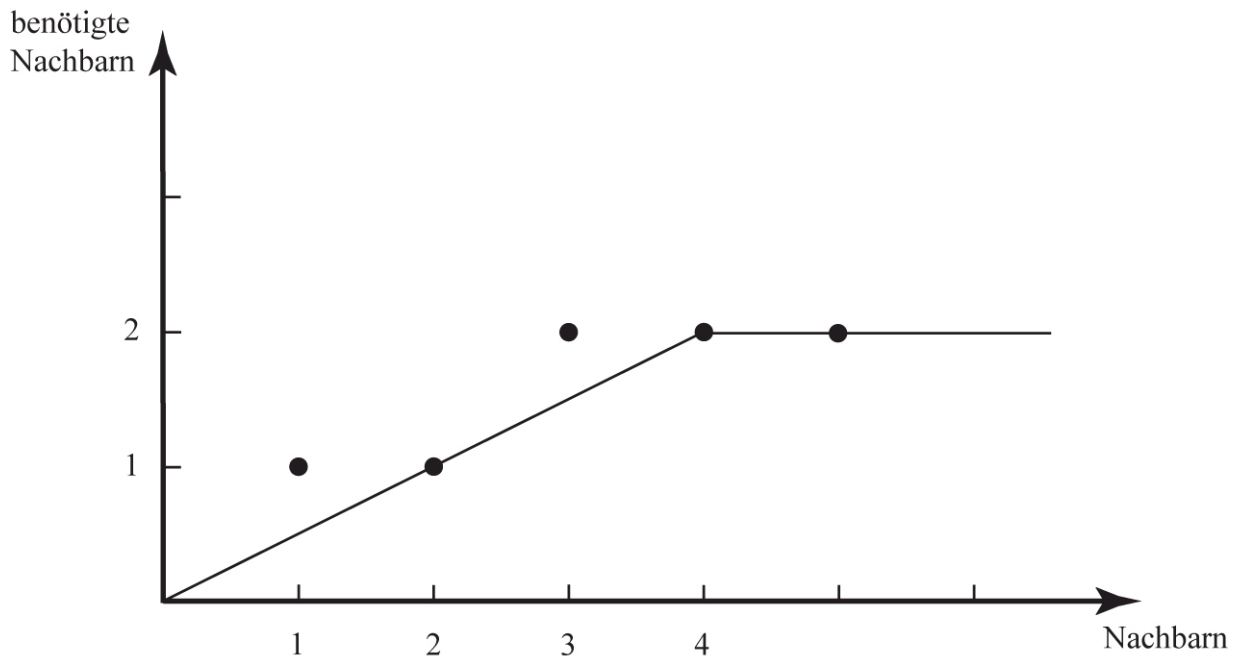


Abbildung 32: Konstante Umschaltfunktion mit linearer Absenkung ($c=1/2$)

Das allgemeine Bildungsgesetz dürfte klar sein. Es handelt sich um eine Mischfunktion, die bis zum Wert $2/c$ linear und danach konstant ist. Alle möglichen Umschaltfunktionen u liegen zwischen den Extremen, die durch $c = 1/2$ einerseits und durch lineare Funktionen andererseits gegeben sind. Diese werden in meiner Implementierung untersucht. Auch eine Einbeziehung anderer als linearer Funktionen hat nur einen geringen Effekt. Es können ja für $c = 1/2$ überhaupt nur die u -Werte im Bereich $0 \leq n \leq 4$ abgeändert werden. Berücksichtigt man zusätzlich noch die obigen Metaregeln, gibt es für die Wahl alternativer Funktionsverläufe von u nur eine minimale Auswahl, und diese hat, da wir uns, wie schon erwähnt, hier ja nicht um chaostheoretische Effekte kümmern wollen, auch nur minimale Wirkung.

Die Zustandsüberföhrungsfunktion für konstantes u für diesen Fall in der obigen Notation die Form

$$\begin{aligned} z_{t+1} &= 1, & \text{falls } v \geq \min(\text{ceil}(m/2), n) \text{ ist, und} \\ z_{t+1} &= 0, & \text{sonst.} \end{aligned}$$

Die Absenkung wird somit durch die Funktion \min , also die Wahl eines Minimums, bewirkt, bei dem das erste Argument als linearer Anteil für kleine und das zweite Argument als konstanter Anteil für größere Werte angesehen werden können.

Für $c = 1/3$ gölte:

$$\begin{aligned} u(n) &= \text{ceil}(1/3 * n) & \text{für } 0 \leq n \leq 6 \\ u(n) &= 2 & \text{für } 6 < n \end{aligned}$$

Damit ergäbe sich der in Abbildung 33 gezeigte Verlauf.

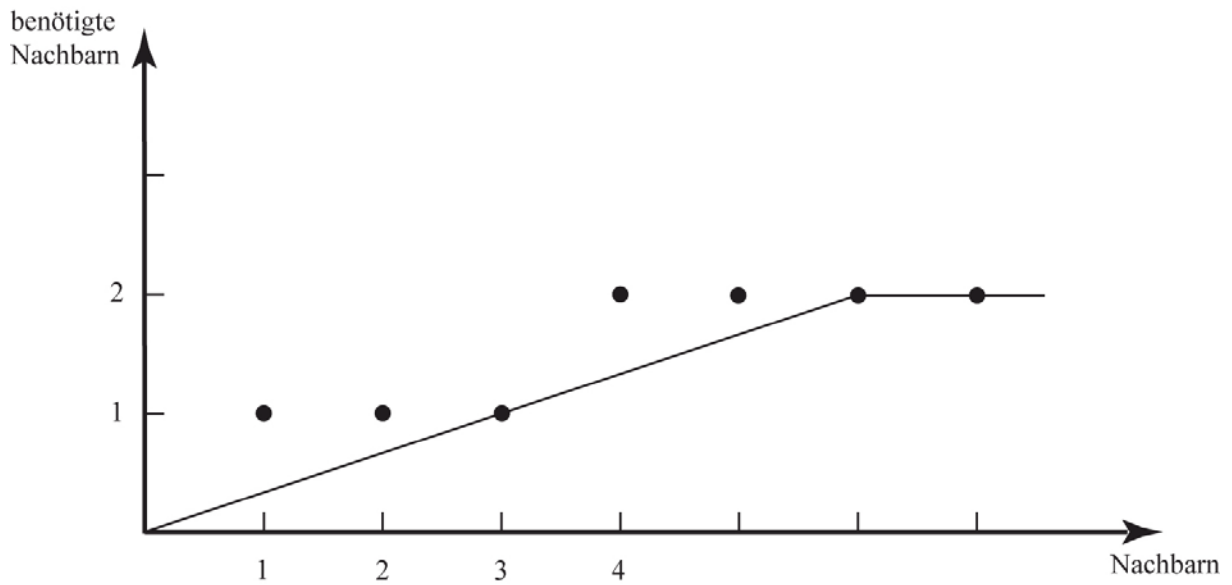


Abbildung 33: Konstante Umschaltfunktion mit linearer Absenkung ($c=1/3$)

Einen beispielhaften Verlauf für einen höheren konstanten Wert und eine lineare Absenkung $c = 1/2$ illustriert Abbildung 34:

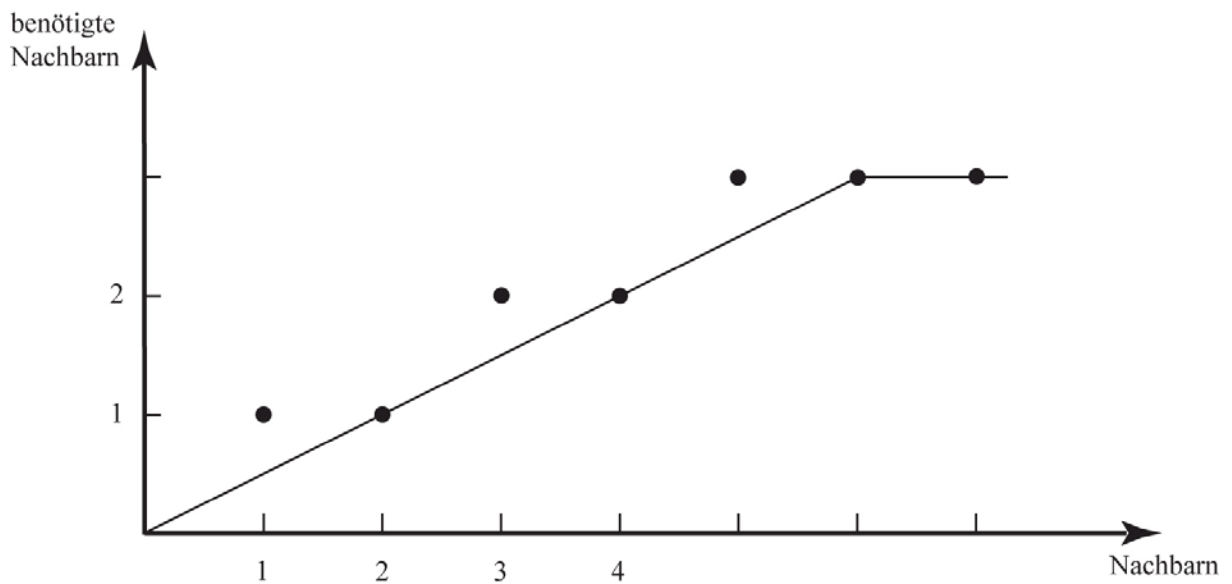


Abbildung 34: Hohe konstante Umschaltfunktion mit linearer Absenkung ($c=1/2$)

Nachdem wir uns davon überzeugt haben, daß die Absenkung einer konstanten Umschaltfunktion u für kleine n am sinnvollsten durch lineare Funktionen bewerkstelligt wird, wird auch deutlich, warum wir für die andere mögliche Funktionenklasse, das sind ja gerade die linearen Funktionen, keine weitere Absenkung brauchen. Sie sind streng genommen selber bereits die Extremform einer Absenkungsfunktion.

11.7 Dynamische Zustandsüberführungsregeln

Bisher bin ich in der gesamten Untersuchung davon ausgegangen, daß sich die Zustandsüberführungsregeln selber mit der Zeit nicht ändern. Das hat seinen Grund: In dem überwiegenden Teil der Literatur wird das genau auf diese Weise gemacht; aber es gibt auch Ausnahmen. In [BoT97], [DM99] sind solche Ansätze bei isotropen Netzwerken vorgestellt. Der Grund für die Zurückhaltung bei der Wahl zeitlich variabler Zustandsüberführungsregeln liegt wieder einmal in dem Wunsch, diese Regeln so einfach wie möglich zu halten. Natürlich werde ich mich an meine eigenen Metaregeln halten. Trotzdem muß ich zwei möglichst klein gehaltene Einflüsse zeitlicher Art in die Regeln einfügen. Wie üblich werde ich zunächst begründen, wozu diese Zusätze nötig sind, und dann erläutern, wie die Probleme technisch gelöst werden können.

Wird die Zustandsüberführungsfunktion in allgemeiner Form von der Zeit abhängig gemacht, bekommt sie die Zeit als ein neues zusätzliches Argument. Angenommen, ein Knoten i habe wie oben n Nachbarn, von denen k im Zustand l sind, wobei wieder ohne Beschränkung der Allgemeinheit für den Knotenzustand z_t zum Zeitpunkt t der Wert 0 vorliegt, so kann f also in dieser Situation als semitotalistische Funktion mit drei Argumenten ausgedrückt werden in der Form

$$z_{t+1} = f(t, n, k)$$

Dann werden wir künftig zusätzlich zwei Formen der zeitlichen Abhängigkeit benötigen – eine Sonderregel für die Anfangsphase der Ausbreitung und einen Schutz gegen ständigen Zustandswechsel, der als „*Latenz*“ bezeichnet werden wird.

11.8 Anfangsphase der Ausbreitung

Hier beziehe ich mich auf epidemiologische Überlegungen: Jede Infektion hat einen Herd. Dieser erste infizierte Knoten ist natürlich von lauter nicht infizierten Knoten umgeben. Man kann jetzt rein logisch zwei Typen von Übertragungsregeln unterscheiden:

(*) Schon ein infizierter Nachbar reicht zur Übertragung der Infektion aus.

(**) Es bedarf mehr als eines einzigen infizierten Nachbarn zur Übertragung.

Im zweiten Fall kommt es offensichtlich nicht zu weiteren Infektionen und die Neuerung stirbt aus. Eine solche Regel taugt in dieser einfachen Form natürlich nicht zur Modellierung. Im ersten Fall würde sich bei strenger Durchführung die Neuerung rasend schnell ausbreiten und im allgemeinen entweder sehr schnell das ganze Netz infizieren oder ein völlig unrealistisches alternierendes Verhalten zeigen, auf das ich im Zusammenhang mit der Latenz noch eingehen werde.

Wir wollen jedoch auch andere Ausbreitungsmuster als Ergebnis einer Simulation des Modells erhalten können:

- Die Neuerung verdrängt die ältere Form völlig. Das kann die Regel (*) liefern. Aber auch hier hat die strenge Anwendung einen Schönheitsfehler: Manchmal geht in der Wirklichkeit eine solche Verdrängung sehr langsam vor sich, während (*) zu einer buschfeuerartigen Ausbreitung führt. Auch das langsame Verhalten sollte jedoch im Modell nachgebildet werden können.
- Die Neuerung verschwindet nach einiger Zeit wieder. Auch hier sollte die Zeit ihres Vorhandenseins variabel sein (nicht nur ein Schritt wie im Falle der Regel (**)).
- Es soll möglich sein, daß ein Zustand einer Koexistenz der beiden Formen entsteht. Das ist bisher völlig ausgeschlossen.

Eine allzu rigide Anwendung der Regeln (*) und (**) würde also zu einer Wahl zwischen Pest und Cholera führen, dem schnellen und vollständigen Sieg einer der Formen. Natürlich kann es in der Wirklichkeit tatsächlich Ausbreitungsmuster geben, wie sie in den beiden Fällen nahegelegt werden, aber diese sehen eben nicht ausschließlich so aus.

Die Analyse der statischen Zustandsüberführungsregeln hat gezeigt, daß der Fall (**) der interessantere ist, auch wenn ich in meinen Simulationen den Fall (*) nicht völlig ausschließen werde. Um einer simulierten Neuerung überhaupt eine Chance auf Ausbreitung zu geben, werde ich Neuerungen in einer Anfangsphase grundsätzlich die Ausbreitung nach Regel (*) gestatten. Nach dem Ende dieser Phase findet die Ausbreitung gemäß den oben entwickelten allgemeinen statischen Regeln mit zufällig ermittelten Parametern statt. Wir haben hier also einen neuen Systemparameter erkannt: die Dauer der Anfangsphase. Es erhebt sich die Frage nach der Größe dieses Parameters. Diese Frage ist leicht zu beantworten, wenn man bedenkt, daß unsere Netze die small world Eigenschaft haben. Schon bei sehr kleinen Werten wäre das ganze Netz infiziert. In real existierenden Netzen würde im allgemeinen der in 7.4.3 genannte Wert 6 reichen⁷⁹. Auf Grund der Kleinheit der hier simulierten Netze kann leicht berechnet werden, daß dort schon der Wert 4 ausreicht, um das ganze Netz zu infizieren. In den Simulationen werden deshalb Werte von 0 bis 3 (einschließlich) zugelassen.

⁷⁹ Genauer gesagt: Es würde der Durchmesser des Netzes reichen, der ja gerade angibt, wie viele Schritte höchstens nötig sind, damit man von jedem beliebigen Knoten zu jedem beliebigen Knoten kommt. Oben habe ich ja schon bemerkt, daß tatsächlich bestimmte Werte unter Umständen auch größer als 6 sein können.

11.9 Latenz

Es ist leicht zu bemerken, daß dynamische Zustandsüberführungsregeln bisher so benutzt worden sind, daß sie bestimmte Ausbreitungsmuster ermöglichen, daß ich aber noch nichts über die Unterdrückung unerwünschter Muster geschrieben habe. Daß es so etwas gibt, habe ich aber oben im Zusammenhang mit Regel (*) schon erwähnt. Ich zeige hier ein besonders extremes

Beispiel: Nehmen wir an, wir haben die Anfangsphase durchlaufen und eine Situation erreicht, in der jeder Knoten im Zustand S nur mit Knoten im Zustand I verbunden ist und umgekehrt⁸⁰. Jetzt verfahren wir streng nach der Regel (*). Dann zeigt das Netz ein ziemlich bizarres Verhalten: Alle Knoten wechseln in jedem Zeitschritt ihren Zustand. In Abbildung 35 ist dieses Verhalten durch rote und blaue Knoten dargestellt. Offensichtlich erzwingt die Regel (*) ein solches Verhalten.

Da Zustände nicht einmalige Sprechhandlungen charakterisieren sondern langfristige Neigungen, wird das Modell der Wirklichkeit nicht gerecht. Auch die Tatsache, daß unsere Netze wachsen, ändert nichts Prinzipielles daran, daß solch ein Verhalten unrealistisch ist.

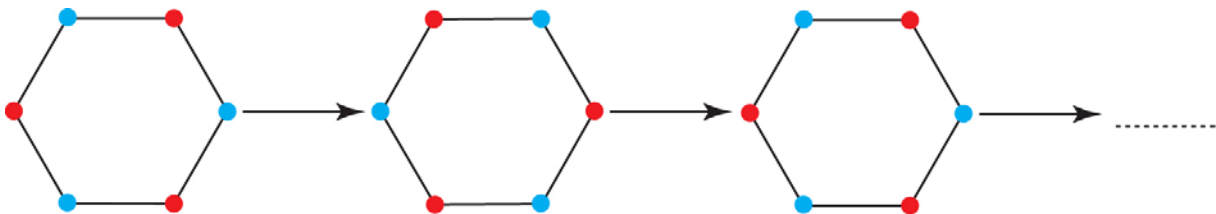


Abbildung 35: Alternierendes Verhalten in einem einfachen Netz

Man könnte sich auf den Standpunkt stellen, daß die Untersuchungen ja ohnehin mit statistischen Methoden durchgeführt werden, solch abartiges Verhalten unsignifikant ist und daher in der Meßungenauigkeit verschwindet. Das ist wahr. Aber vergleichbares Verhalten kommt, wenn auch nicht in der eben vorgeführten extremen Form, überraschenderweise eben doch manchmal vor. Und da ziehe ich es vor, eine leichte Änderung ins Modell einzubauen, die zwar bei der Untersuchung dynamischer Netze nicht üblich ist, die aber ihren Zweck erfüllt und die mir aus eigenen Forschungen in anderem Zusammenhang [Sche97] wohlvertraut ist. Dabei handelt es sich um die *Latenz*.

Dieser Begriff ist sehr vielfältig. Auch in der Informatik gibt es mehrere Bedeutungen, zum Beispiel die Zeitspanne, bis nach einer Anfrage an ein Massenspeichermedium die ersten Daten übertragen werden. Ich beziehe mich hier auf eine Definition aus der Spezifikation von Realzeitsystemen. Dort bedeutet Latenz die Zeitdauer, die nach einem Ereignis mindestens zum Zwecke interner, also latenter Berechnungen verstrichen sein muß, bevor ein neues Ereignis akzeptiert werden kann. In unserem Zusammenhang möchte ich mit Latenz die Zeit bezeichnen, die mindestens nach einem Zustandswechsel eines Knotens vor einem erneuten

⁸⁰ Selbst dieser extreme Fall ist mit leichten Variationen wiederholt in meinen Simulationen aufgetreten, als ich ihn noch nicht unterdrückt hatte. Die allgemeine Situation in den Simulationen war aber natürlich wesentlich komplexer als im hier angegebenen Beispiel.

Zustandswechsel dieses Knotens vergehen muß. Mit der Latenz haben wir einen neuen Systemparameter erkannt.

Wieder erhebt sich die Frage nach einer sinnvollen Größe dieses Parameters für die Simulationen. Diese Frage ist rein mathematisch nur schwer zu beantworten. Auf einer rein linguistischen Ebene ist die Interpretation so, daß eine hohe Latenz ein konservatives Sprachverhalten begünstigt. Dadurch wird das Entstehen einer neuen sprachlichen Form nicht unmöglich gemacht; aber dann kommt es nicht wieder zu einer Verdrängung der Neuerung. Insbesondere wird dann die Koexistenz von alter und neuer Sprachform erschwert. In den Simulationen hat es sich herausgestellt, daß die Zahl der Simulationen mit schließlicher Koexistenz in Abhängigkeit von der Latenz etwa Pareto-verteilt ist. Ich zeige hier nur einen beispielhaften Simulationsverlauf.

Simuliert wurden insgesamt 300 Fälle der Ausbreitung einer Neuerung. Es sollte der Einfluß der Latenz auf deren langfristige Entwicklung untersucht werden. Als Parameterwerte für die Ausgangsdaten wurden Zufallszahlen innerhalb der in der bisherigen Analyse als sinnvoll erkannten Grenzen erzeugt. Lediglich für die Latenz wurden Werte gleichverteilt (hier zwischen 21 und 38 mal) im Bereich von 0 bis 9, also einem verhältnismäßig großen Intervall, erzeugt. Die Ergebnisse wurden in drei Kategorien eingeteilt: 0 und 2 zeigen den Sieg der alten und den der neuen Form an; das Ergebnis 1 signalisiert die Koexistenz. Es zeigt sich deutlich, daß eine Koexistenz bei höheren Latenzen immer mehr erschwert wird. Das Wesentliche ist in der mittleren Zeile ersichtlich.

Latenz:	0	1	2	3	4	5	6	7	8	9	gesamt
Ergebnis											
0	19	13	19	16	15	16	18	14	8	20	158
1	15	13	2	2	1	2	1	0	0	1	37
2	0	2	6	7	11	18	15	16	13	17	105
gesamt	34	28	27	25	27	36	34	30	21	38	300

Vergleichbare Verläufe zeigen sich durchgängig in allen durchgeführten Simulationen. Welche Werte für die Latenz sinnvoll sind hängt oft von den konkreten Anwendungen ab.

11.10 Formale Definition eines Dynamischen Sprachnetzes

An dieser Stelle soll eine formale Definition eines Dynamischen Sprachnetzes gegeben werden: Zunächst definiere ich einen dynamischen Graphen, bei dem noch keine Dynamik auf dem Netz zu beobachten ist. Es gibt nur topologische Veränderungen:

Definition: Ein Quadrupel $G = (V, E, R1, R2)$ heißt *DT-Graph*, wenn gilt:

- (V, E) ist ein endlicher Graph.
- $R1$ ist eine Regel zur Erzeugung neuer Knoten und ihrer Anbindung.
- $R2$ ist eine Regel für das Ausscheiden von Knoten aus dem Graphen.

Der Name „DT-Graph“ steht für die dynamische Topologie. Das Paar (V, E) beschreibt die Anfangssituation. In vielen Fällen ist dies ein sehr kleiner Graph (leer oder ein Knoten).

Es sollen anfangs aber auch größere Graphen zugelassen werden. Darüber kann mehr in Abschnitt 13.2.2 erfahren werden. Bei der Formulierung der Regeln habe ich auf eine genaue Bestimmung verzichtet. Dadurch wird es möglich auch andere Netzstrukturen zu erzeugen als die in dieser Arbeit als zielführend erkannten Netze mit konstanter oder linearer Umschaltfunktion und Pareto-Verteilung. Es könnten jetzt etwa auch die klassischen Modelle aus den in 8.2 und 8.3 vorgestellten Arbeiten von Erdős und Rényi [ER59], [ER60] benutzt werden.

Die eigentliche Definition eines Dynamischen Sprachnetzes schließt dann auch die Dynamik auf dem Netz ein:

Definition: Ein Sechs-Tupel $D = (G, Z, f, init, lat, infek)$, heißt *Dynamisches Sprachnetz* wenn gilt:

- $G = (V, E, R1, R2)$ ist ein DT-Graph.
- Z ist eine endliche Menge von Zuständen.
- Es ist f ist eine monotone, semitotalistische, angemessene Überföhrungsfunktion mit der Funktionalität $Z^* \rightarrow Z$.⁸¹
- Schließlich seien $init$, lat , $infekt$ natürliche Zahlen, die die Dauer der Anfangsphase, die Latenz und den Infektionszeitpunkt bezeichnen.

Auch hier habe ich die allgemeine Definition etwas großzügiger, als es bisher abgeleitet wurde, gewählt, um in Zukunft bei Bedarf auch allgemeinere Verfahren zulassen zu können.

⁸¹ Die Menge Z^* bezeichnet üblicherweise die Menge aller endlichen Folgen mit Elementen aus Z . Dadurch wird hier der Tatsache Rechnung getragen, daß die Überföhrungsfunktion für beliebig große Nachbarschaften eines Knotens definiert sein muß.

12 Die Monte Carlo Methode

Nach den theoretischen Vorüberlegungen zum Dynamischen Sprachnetz stelle ich jetzt eine Methode vor, die für die Durchführung der Simulationen in dem Modell und ihre Auswertung hilfreich sein wird. Ich beginne mit einer Motivierung, die unabhängig von unserem eigentlichen Thema ist.

Gegeben sei eine Landkarte, die vor allem ein Land zeigt; sagen wir, es heiße Germanistan, und man könne sehen, wie groß der Maßstab der Karte ist. Dann kann man durch Ausmessen der Kartengröße leicht berechnen, wie groß die Fläche des ausgemessenen Ausschnittes in der Wirklichkeit ist. Vermutlich interessiert das in den meisten Fällen recht wenig, aber sehr relevant ist die Frage:

„Wie groß ist die Fläche von Germanistan?“

Ein Mathematiker käme vielleicht auf die Idee, die Grenze dieses Landes durch eine mathematische Funktion zu beschreiben. Mit etwas Glück und viel Schweiß könnte er möglicherweise dann die Fläche durch Integration bestimmen. Damit erhielte er das, was alle Mathematiker lieben: eine exakte Lösung. Bei einem solchen Ansatz tritt unter anderem ein Problem auf: Wo bekommt man eine Funktion zur Beschreibung der Grenze her? In vielen Anwendungen ist es zweckmäßig, die Ansprüche zu mäßigen. Unter Umständen soll die Fläche von Germanistan ja gar nicht bis auf den Durchmesser eines Atoms genau bestimmt werden, wie es der integrative Ansatz liefern kann. Dann kann der Zufall bei der Lösung der Aufgabe helfen, genau wie er auch der Spielbank von Monte Carlo bei ihrer Aufgabe hilft, viel Geld zu verdienen. Deswegen heißt die Monte Carlo Methode auch „*Monte Carlo Methode*“. Man geht folgendermaßen vor:

- Sei A die Fläche die von der ganzen Karte abgebildet wird.
- Es werde zufällig eine Anzahl M von Punkten auf der Karte erzeugt.
- Es werde die Anzahl N von Punkten, die in Germanistan liegen, bestimmt.
- Der Wert AN/M sollte eine recht gute Annäherung für die gesuchte Fläche sein.

Einige Fakten sind klar:

- Natürlich gibt es keine Garantie, daß das Ergebnis stimmt. Beim Zufall ist eben nichts sicher, aber:
- Je größer M ist, desto genauer sollte das Ergebnis sein.
- Die Punkte sollten wirklich zufällig sein. Wenn die Auswahl einen Drall nach links unten hat, landen vielleicht alle Punkte im Nachbarland Frankistan und man erhält den Wert 0 als Größe von Germanistan.

12.1 Die Anwendung der Monte Carlo Methode

Im Falle von Germanistan war die Zahl der Dimensionen unserer Karte gleich zwei. Natürlich könnte dasselbe Verfahren auch zur Bestimmung des Volumens von Körpern, also von dreidimensionalen Objekten, benutzt werden. Es ist nur ein quantitativer Unterschied aber qualitativ nichts Neues, wenn man zu noch höheren Dimensionen übergeht. Als Dimension stellt man sich in den Anwendungen der Monte Carlo Methode die Anzahl der frei wählbaren Parameter vor. Die Rolle von Germanistan wird dann gespielt von der Menge aller Parameterkombinationen, die gewissen Eigenschaften genügen. Das wird auch in dieser Arbeit so sein.

Eine nächste Erweiterung der Monte Carlo Methode erhält man, wenn man sie nicht zur Bestimmung von Inhalten oder Volumina sondern zur Bestimmung von Durchschnittswerten benutzt.

Nehmen wir an, die Karte von Germanistan sei so organisiert, daß zu jedem Punkt seine Höhe h über dem Meeresspiegel abgelesen werden kann (durch Höhenlinien oder als geographisches Informationssystem). Bei der Beantwortung der Frage

„Wie groß ist die durchschnittliche Höhe von Germanistan?“

verfährt man wie bei der einleitenden Anwendung; aber jetzt werden nicht einfach die Punkte, die innerhalb von Germanistan zu liegen kommen, gezählt. Vielmehr wird, sobald ein Punkt x_i erzeugt worden ist, gleich $h(x_i)$ bestimmt. Daraus kann dann, sobald alle M Punkte erzeugt wurden, der Durchschnitt errechnet werden.⁸²

Ich werde zu diesem Zweck $x_i \in G$ schreiben, falls ein zufällig erzeugter Punkt x_i innerhalb von Germanistan zu liegen kommt. Germanistans durchschnittliche Höhe sei mit H bezeichnet. Dann lautet die Formel folgendermaßen:

$$H = \sum_{x_i \in G} h(x_i) / N$$

Mathematisch dient die Monte Carlo Methode also zur Vermeidung einer Integration zur Flächenbestimmung. Die entsprechenden Integrale zu berechnen ist ohnehin in den meisten Fällen unmöglich, und man müßte zu den weniger exakten numerischen Methoden übergehen.

12.2 Dimensionsreduktion

Die Dimensionsreduktion ist ein bedeutendes Problem der Statistik. Um erläutern zu können, worum es dabei geht und wie es sich in unserem Zusammenhang äußert, nutze ich eine offensichtliche Beobachtung. Um H ausrechnen zu können, müssen M Punkte erzeugt

⁸² Mathematisch ist die ursprüngliche Anwendung nur ein Spezialfall der zweiten. Bei der Flächenbestimmung wird nichts weiter gemacht als eine Durchschnittsbildung für die „Charakteristische Funktion“ X_G , für die gilt:

$$\begin{array}{ll} X_G(x) = 1 & \text{falls } x \in G \\ X_G(x) = 0 & \text{falls } x \notin G \end{array}$$

werden, von denen aber nur N , nämlich die, die in Germanistan landen, in die Berechnung eingehen. Der Rest wird einfach weggeworfen.

Nehmen wir jetzt der Einfachheit halber an, die Landkarte sei quadratisch und Germanistan sei ebenfalls quadratisch, wobei der Abstand des Landes zum Kartenrand jeweils 10 % der gesamten Breite der Karte ausmache. Die Größe der Karte sei zu 1 normiert. In diesem Falle beträgt der flächenmäßige Anteil von Germanistan an der gesamten Kartenfläche genau $0,8 * 0,8 = 0,64$. Das läßt sich leicht im rechten Teil von Abbildung 36 ausmessen.

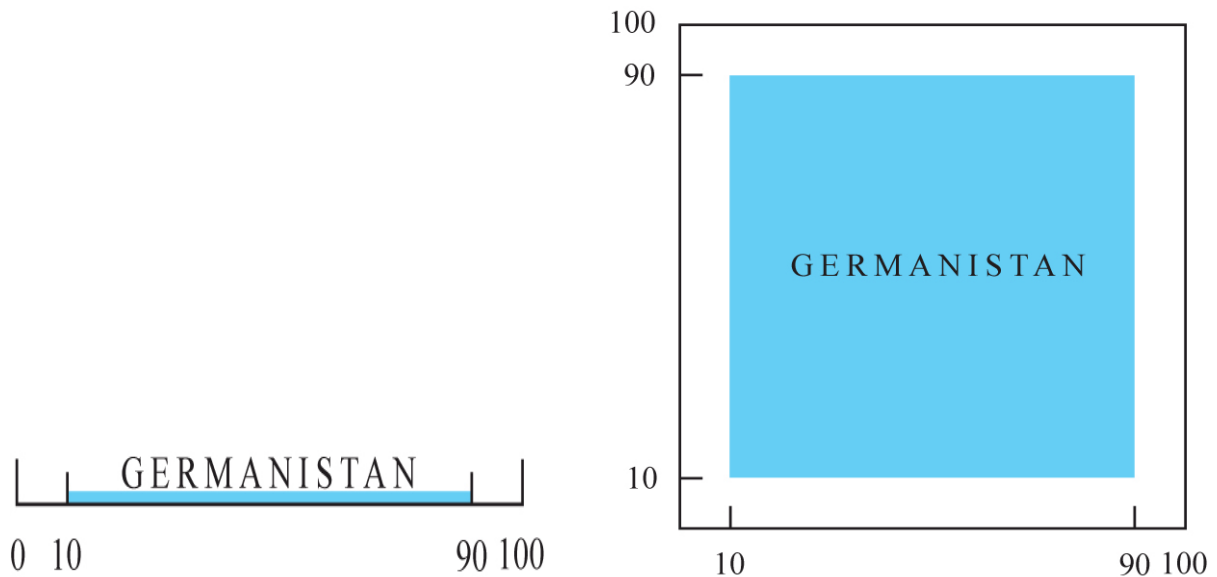


Abbildung 36: Germanistan (eindimensional, zweidimensional)

Wäre die Karte eindimensional, so wäre der flächenmäßige Anteil von Germanistan an der gesamten Kartenfläche genau $0,8$; im Falle einer dreidimensionalen Karte wäre der entsprechende Anteil $0,8^3$ und ganz analog wäre er im Falle von n Dimensionen gleich $0,8^n$. Dabei wird wieder davon ausgegangen, daß ein Punkt genau dann nicht zu Germanistan gehört, wenn er in einer Richtung weniger als 10% vom Rand entfernt liegt. Diese Werte sind zugleich der Anteil der für eine Berechnung der durchschnittlichen Höhe von Germanistan nutzbaren Zufallswerte. Dieser Anteil sinkt etwa für $n = 6$ auf $0,8^6 = 0,262144$. In dieser Situation heißt das, daß nur noch gerade ein Viertel aller Meßwerte überhaupt brauchbar ist. Dabei ist die Ausgangssituation außerordentlich günstig. Wäre der Anteil in jede Richtung nicht bei im allgemeinen unrealistisch günstigen $0,8$, sondern bei eher realistischen $0,1$, so sänke der Anteil der brauchbaren Meßwerte auf $0,1^6 = 0,000001$, auf ein Millionstel. Dabei ist auch eine Dimensionalität von 6 noch gering.

Beim Übergang zu höheren Dimensionen haben wir also das Problem, daß die Größe des uns interessierenden Wertebereiches mit wachsender Dimensionszahl n im Vergleich zur Größe des gesamten möglichen Wertebereiches sehr schnell (wieder einmal exponentiell) gegen 0 gehen wird, wodurch fast alle Messungen wertlos werden.

Da die Dimensionszahl der Anzahl der freien Parameter in den Messungen entspricht, ist die Konsequenz dieser bedauerlichen Tatsache, daß, falls es hinreichend viele Parameter gibt, die meisten erzeugten Zufallspunkte außerhalb des interessanten Bereichs liegen und

man sehr, sehr viele Parametermessungen durchführen muß, um einigermaßen verlässliche Ergebnisse zu bekommen. Man spricht auch vom „curse of dimensionality“, dem „Fluch der hohen Dimensionen“. Daher ist man heute in der Statistik ganz allgemein an solchen Reduktionen der Dimension sehr interessiert.

Dimensionsreduktionen werden auch in dieser Arbeit benutzt werden. Ich werde dabei keine ausgefeilten modernen Techniken nutzen, sondern einen ganz einfachen Weg gehen: Zunächst wird ein Simulationslauf mit allen Parametern aber verhältnismäßig wenigen Meßwerten gemacht. Dann wird mit Hilfe des Korrelationskoeffizienten⁸³ bestimmt, welche Parameter überhaupt einen Einfluß auf die Zielgröße (im Falle von Germanistan also die Durchschnittshöhe, im Falle unseres Modells die Aufenthaltsdauer eines Worts oder eine davon abgeleitete Größe) haben. Durch Signifikanztests, etwa den in Unterkapitel 14.4 erläuterten und benutzten χ^2 -Unabhängigkeitstest, kann festgestellt werden, ob festgestellte Unabhängigkeiten zufällig entstanden sind, oder ob sie auch bei Simulationen mit mehr Meßwerten erwartet werden können.

Als Faustregel kann allgemein angenommen werden, daß die Zahl der benötigten Meßwerte exponentiell, also sehr schnell, mit der Zahl der Parameter wächst. Jeder durch solche Tests als unabhängig erkannte Parameter reduziert die Dimension des Raumes der möglichen Messungen um 1 und erhöht dadurch die Wahrscheinlichkeit, in den eigentlichen Simulationsläufen brauchbare Werte zu erhalten, ganz beträchtlich.

⁸³ Der *Korrelationskoeffizient* ist eine Maßeinheit für die Abhängigkeit von zwei Meßreihen gleicher Länge, wo jeder Punkt der einen Reihe einem Punkt der anderen Reihe zugeordnet ist (und diesen möglicherweise beeinflusst). Normalerweise besteht die eine Reihe aus Eingabewerten, die andere aus Ergebnissen. Eine genaue Definition mit Erklärungen kann in [Sac97], Abschnitt 51, gefunden werden.

13 Die Implementierungen des Modells

Ganz allgemein ist es notwendig, Programme zu kommentieren. Dazu gehören mindestens die folgenden Teile:

- eine Erklärung der vorhandenen Schnittstellen, einschließlich deren Bedeutung,
- eine Erklärung der benutzten Datenstrukturen, einschließlich deren Bedeutung,
- eine Erklärung der Algorithmen, die auf den Datenstrukturen arbeiten,
- eine Erklärung ansonsten schwer durchschaubarer Programmdetails.

In diesem Kapitel möchte ich die beigefügten Implementierungen kommentieren⁸⁴, im folgenden Unterkapitel zunächst, die in 4.5. schon erwähnten Programme, mit deren Hilfe die Situation für eindimensionale Gittergeometrien vollständig geklärt wird.

Der Rest des Kapitels ist der ausgiebigen Kommentierung der Implementierung des in Kapitel 11 entwickelten Modells des Dynamischen Sprachnetzes gewidmet. Alle Implementierungen sind von mir in C++, also mit Hilfe des objektorientierten Programmierparadigmas⁸⁵, durchgeführt worden.

13.1 Programme für Eindimensionale Geometrien

Durch die jetzt zu besprechenden Programme werden alle möglichen Regeln einer eindimensionalen Gittergeometrie mit drei Zuständen klassifiziert, die den Metaregeln genügen. Zunächst führe ich eine einheitliche Beschreibung der Regeln ein. Das geschieht durch das folgende

Beispiel: In der nach [Wol02] für unseren Fall adaptierten Schreibweise

```
00 || 01 || 02 || 10 || 11 || 12 || 20 || 21 || 22 ||  
0 || 0 || 0 || 0 || 1 || 1 || 0 || 1 || 2 ||
```

bedeutet die obere Zeile den alten Zustand und die zweite Zeile den neuen. Die neun Spalten entsprechen den neun möglichen Fällen für den alten Zustand :

00 der Sprecher und seine Nachbarn benutzen die Neuerung nie,

⁸⁴ In den Programmen selber befindet sich eine sehr detaillierte Kommentierung.

⁸⁵ Das herausragende Kennzeichen objektorientierter Programmierung ist, wie schon der Name sagt, die Fixierung auf Objekte und deren Eigenschaften. Auf Objekte kann beispielsweise nicht beliebig zugegriffen werden, sondern nur mittels wohldefinierter Zugriffsmethoden, die von den Objekten angeboten werden. Die Formate solcher Schnittstellen werden in sogenannten header-Dateien beschrieben. Diese sind keine Kommentare sondern integraler Bestandteil der Programmierung. Sie sind an der Endung .h im Namen zu erkennen.

- 01 der Sprecher benutzt die Neuerung nie, seine Nachbarn gelegentlich,
... ..
22 der Sprecher und seine Nachbarn benutzen die Neuerung immer.

Ich bitte zu beachten, daß eine der Metaregeln, die Symmetrie, schon dadurch in der Schreibweise der Regeln eingeht, daß das Verhalten der Nachbarn nur einmal angegeben und daher nicht näher ausdifferenziert wird. Eine Regel ist dadurch charakterisiert, daß jedem der neun Fälle ein möglicher Folgezustand, also ein mögliches künftiges Sprechverhalten, aus $\{0,1,2\}$ zugeordnet wird. Damit bestimmt sich die Gesamtzahl der Regeln zu

$$3^9 = 19683$$

Denn jede der neun Möglichkeiten ist zunächst einmal frei mit jeder anderen kombinierbar, überall kann einer von drei Werten angenommen werden. Die Frage ist die, welche davon den Metaregeln genügen.⁸⁶ Die im Beispiel gegebene Darstellung einer Regel läßt sich etwas weniger redundant gestalten. Die Regel wird durch den Vektor der möglichen Nachfolgezustände gekennzeichnet:

000011012

Es entspricht also jeder derartige neundimensionale Vektor einer Regel und umgekehrt. Ich mache mir dann einen Trick zunutze, den ich wieder bei Wolfram abgeschaut und auf unsere Situation umgeschrieben habe: Ich betrachte den Vektor als Darstellung im 3-adischen Zahlensystem.⁸⁷ Das bedeutet folgendes:

Wenn man die Zahlen von 0 bis $3^9 - 1 = 19682$ durchläuft, durchlaufen die zugehörigen 3-adischen Darstellungen genau die möglichen Darstellungen von Regeln. Genau das tut das kleine Programm

regeln.cpp

In einer Schleife von 0 bis 19682 werden mit Hilfe einer Konversionsroutine die 3-adische Darstellungen aller möglichen Regeln erzeugt. Währenddessen werden die metaregelkonformen Regeln bestimmt und in ihrer 3-adische Darstellung in die Datei

regeln.txt

geschrieben. Diese Datei dient dann als Eingabe für das ebenfalls kleine Programm

⁸⁶ Die Regel aus dem Beispiel genügt allen Metaregeln. Sie ist etwas Besonderes insofern, als sie unter allen metaregelkonformen Regeln diejenige ist, die die wenigsten Nachkommen produziert, sozusagen die „konservativste“ aller derartigen Regeln.

⁸⁷ Unser herkömmliches Zahlensystem beruht auf dem dezimalen (oder 10-adischen) System. Das heißt, jede natürliche Zahl n läßt sich eindeutig darstellen durch die Form

$$n = \sum_i a_i 10^i$$

Das läßt sich auch auf der Basis 3 (oder jeder anderer natürlicher Zahl, die größer als 1 ist,) statt der Basis 10 realisieren. Beispielsweise ist

$$139 = 1*10^2 + 3*10^1 + 9*10^0 = 1*3^4 + 2*3^3 + 0*3^2 + 1*3^1 + 1*3^0$$

Entsprechend hat 139 die 3-adische Darstellung 12011.

regelausfuehren.cpp

Solange in *regeln.txt* noch Regeln vorhanden sind, werden diese eingelesen und *max*-mal angewandt. Dabei ist *max* eine willkürlich gewählte Konstante. In dem Beispiel, das in 4.5 knapp und auf der CD ausführlich dokumentiert ist, habe ich *max* auf 20 gesetzt. Das reicht bei weitem, um zu erkennen, was vor sich geht. Der Startzustand für die Regelarbeitung ist immer

20...0

Das bedeutet, daß das eindimensionale Feld mit lauter nicht infizierten Knoten besetzt ist außer einem, bei dem ständiger Gebrauch der Neuerung vorherrscht. Wegen der Symmetrie der Regeln habe ich ohne Beschränkung der Allgemeinheit festsetzen können, daß die Änderung zuerst links außen auftritt. Die Ergebnisse dieser Regelanwendungen werden in

ergebnisse.txt

abgespeichert. Darin ist zu ersehen, daß alle möglichen Regeln einer eindimensionalen Gittergeometrie mit drei Zuständen zu den in 4.5 aufgezählten Ausbreitungsmustern führen.

13.2 Grobdesign des Dynamischen Sprachnetzes

13.2.1 Objekte

Die wesentlichen Objekte im Modell des Dynamischen Sprachnetzes sind

- Listen,
- Zellen.

Listen sind in der Informatik übliche Datenstrukturen. Die Arbeit mit Listen ist ein Grundpfeiler jeder Informatikausbildung. Bei mir sind alle Listen (außer den Listen für den in 14.4 benutzten χ^2 -Test) geordnet – je kleiner ein Listenelement ist, desto weiter vorne befindet es sich also in der Liste. Dazu gehören verschiedene übliche Zugriffsmethoden: Einfügen eines Elementes (sowohl mit möglichem mehrfachen Auftreten als auch ohne dieses), Herausnehmen eines Elementes (unter Umständen in gewünschter Vielfalt), Bestimmen der Listenlänge und weitere Methoden, die in *listen.cpp* implementiert (sowie gründlich kommentiert) und deren Schnittstellen in der header-Datei *zellenlisten.h* beschrieben werden.

Mehr spezifisch für den uns interessierenden Zusammenhang sind die Zellen. Sie stehen für die Knoten im Graphen und haben zwei wesentliche Attribute

- den Zustand, in dem sich der Knoten gerade befindet (Variable *state*),
- eine Liste der Nachbarn des Knotens (Zeiger auf Listen mit Namen *nachbarn*).

Aus Implementierungsgründen gibt es noch zwei weitere Hilfsattribute *newstate*, *schutzstate*, die nur kurzfristig benötigt werden. Die Schnittstellen der Zellen werden ebenfalls in der header-Datei *zellenlisten.h* beschrieben; die Methoden werden in *zellen.cpp* implementiert sowie gleichfalls gründlich kommentiert.

13.2.2 Struktur des Hauptprogramms

Alle der im folgenden erläuterten Parameter sind so gestaltet, daß sie durch eine Änderung nur an einer Stelle im Programm einen neuen Wert erhalten können.

Das Hauptprogramm hat den Namen *simulation.cpp*. Es steuert die Simulationen. Deren Anzahl wird durch *anzahlsimul* bestimmt. Die jeweils aktuelle Zahl bereits durchgeführter Simulationen wird in der Laufvariablen⁸⁸ *laufnummer* festgehalten.

In jeder Simulation wird eine durch die Grenze in *gesamtgenerationen* bestimmte und durch die Laufvariable *generation* gezählte Anzahl von Generationen⁸⁹ durchgerechnet. Alle Aussagen statistischer Art, die im bisherigen Text gemacht wurden, gelten in jedem Simulationslauf von Anfang an. Aber Statistik behandelt Massenphänome. Deshalb kann es in einem Teil der Simulationsläufe zu Verzerrungen kommen, wenn der dogmatische Weg beschritten wird: Aufbau des Netzes aus einem einzigen Knoten und dann abwechselnde Erzeugung von Knoten und Kanten. Das würde sich auf Dauer zwar statistisch herausmitteln, ich habe aber ein übliches Verfahren gewählt, mit dessen Hilfe solch pathologisches Verhalten von Anfang an erheblich erschwert wird. Es handelt sich im wesentlichen um die Idee aus [BA99], [AJB00], die schon in Unterkapitel 8.5 zitiert wurde:

Die Simulationen starten dann nicht mit einem einzigen Knoten sondern mit einer durch *initgenerationen* einstellbaren Anzahl von Knoten, oben m_0 genannt. In den folgenden Generationen wird es dann eine gewisse kurze Zeit dauern, bis sich das System entsprechend der Theorie verhält, bis sich etwa eine Pareto-Verteilung einstellt. Die Zahl *initgenerationen* sollte also hinreichend groß sein, damit sie ihren Zweck erfüllen kann, aber auch nicht zu groß, so daß der Einschwingprozeß nicht zu lange dauert. Experimente haben gezeigt, daß die Zahl 10 ein passender Wert ist.

Die Hauptstruktur des Modells ist also eine Doppelschleife: Außen werden Simulationen, innen werden Generationen innerhalb der gerade aktuellen Simulation gezählt.

⁸⁸ In der vorliegenden Implementierung (und in der Programmierung ganz allgemein) wird eine dynamisch veränderbare Größe oft durch zwei Variablen gekennzeichnet:

- eine Variable die immer den aktuellen Wert der betrachteten Größe enthält; wird diese während der Abarbeitung einer Schleife immer um 1 erhöht, wird sie auch als Laufvariable bezeichnet,
- eine Variable die den Wert für die Grenze der Laufvariablen enthält; diese ist oft ein Systemparameter, oder sie wird für jeden Simulationslauf neu bestimmt und gilt dann für den ganzen Simulationslauf.

⁸⁹ Der Name „Generation“ stammt aus der Theorie der zellularen Automaten. Gemeint ist einfach ein Zeitschritt auf der imaginären diskreten Zeitachse der Simulation.

13.3 Dynamik beim Aufbau des Netzes

Nach der Initialisierung werden in jeder Generation *knoterz* Knoten erzeugt, danach *einfinintern* interne Kanten zwischen schon vorher vorhandenen Knoten und schließlich wird zwischen jedem der neu erzeugten Knoten und *einextern* zufällig gewählten alten Knoten eine Verbindung geschaffen. Die Nachbarn eines Knotens *i* werden in einer Liste *nachbarliste[i]* verwaltet. Diese Liste enthält also alle Nachbarn von *i* und zwar in geordneter Reihenfolge.

Die Knoten, zwischen denen neue Verbindungen entstehen, werden mit Hilfe des preferential linking erzeugt. Um das zu ermöglichen, habe ich eine Liste *preflist* eingeführt. In dieser geordneten Liste werden so viele Kopien eines Knotens gehalten, wie seinem augenblicklichen Vorrang entspricht. Dann wird einfach zufällig ein Knoten in *preflist* als neuer Endpunkt einer Verbindung bestimmt. Die Wahrscheinlichkeit, daß ein bestimmter Knoten *i* an einer neuen Verbindung beteiligt ist, errechnet sich damit durch

$$\text{anz}(i, \text{preflist}) / \text{getlaenge}(\text{preflist})$$

Dabei bedeutet $\text{anz}(i, \text{preflist})$ die Anzahl der Vorkommen von *i* in *preflist*; die Methode *getlaenge* bestimmt ganz allgemein die Länge einer gegebenen Liste.

Der Aufbau von *preflist* geschieht folgendermaßen: Wie in den theoretischen Ausführungen schon dargelegt wurde, soll es eine Funktion *f* geben, die die Wahrscheinlichkeit $p(i)$, daß ein bestimmter Knoten *i* an einer neuen Verbindung beteiligt ist, von seinem Grad $k(i)$ abhängig macht. Zur Erzeugung einer Pareto-Verteilung muß dieses *f* affin linear, also von der Form

$$p(i) = f(k(i)) = c * k(i) + d$$

sein, wie noch aus 8.5 bekannt ist. Im vorliegenden System nehmen die Variablen *prefmult* und *prefadd* die Rolle von *c* und *d* ein.

13.4 Die Dynamik auf dem Netz

Die zweite Art der Dynamik in dem in dieser Arbeit entwickelten Modell neben der Dynamik des Netzaufbaus selber ist die der Veränderung der Knotenzustände, unter anderem also insbesondere die Ausbreitung von Neuerungen. In klassischer Weise werden die möglichen Zustände von Individuen durch Buchstaben dargestellt, die durch Zahlen als Programmkonstanten implementiert sind:

S	0	susceptible
I	1	infected
R	2	recovered / removed

Für jedes Individuum sind diese Werte im Feld *state* gespeichert.

Wie breitet sich eine Neuerung aus? Sie beginnt, wie schon gesagt, mit einem Individuum und breitet sich in ihrer Anfangsphase geschätzt aus. Für den Zeitpunkt des ersten Auftretens gibt es, wie für alle zufälligen äußeren Ereignisse, die das Systemverhalten beeinflussen⁹⁰, im Prinzip zwei Möglichkeiten, [LK91], je nach den Voraussetzungen und dem Ziel der Untersuchung. Ich bin bei den verifizierenden Experimenten beide Wege gegangen.

13.4.1 rein zufällige Werte

Dieser Weg wird bei eher grundlegend-theoretischen Untersuchungen eingeschlagen. Grundlage ist dann eine gut untersuchte theoretische Verteilung, die das Verhalten der Störung nachbildet. Auf diesen Weg wollte ich nicht verzichten, da mich die Theorie der Netze auch unabhängig von der in dieser Arbeit hauptsächlich beabsichtigten konkreten Anwendung auf die Verbreitung von Lehnwörtern interessiert. Da ich mich zunächst nur an möglichst wenigen Stellen ausdrücklich auf Lehnwörter bezogen habe, sollte das Dynamische Sprachnetz auch auf andere Phänomene des Sprachwandels anwendbar sein. Außerdem zeigen die Überlegungen zur Dimensionsreduktion ja gerade, daß die prinzipiellen Mechanismen grundsätzlicher Art sind, also unabhängig vom Zeitpunkt der Übernahme. Die Grenzen für die Auswahl der Zeitpunkte der Übernahme werden dann in erster Linie nach mathematischen Kriterien festgelegt.

Der Zeitpunkt des ersten Auftretens wird am Anfang jeder Simulation in einer Variablen *anfgeninfekt* festgehalten. Der genaue Zeitpunkt läßt sich etwas eingrenzen:

- Am Anfang sollte eine Infektion noch nicht möglich sein, solange das Netz noch zu klein ist, als daß es sinnvoll wäre, es mit statistischen Mitteln zu behandeln. Die Anfangsbedingungen sind zunächst noch zu stark⁹¹.
- Auch eine zu späte Infektion erscheint mir zumindest bei dem Ansatz mit den zufälligen Werten nicht sinnvoll. Sonst hätte sie ja gar keine Zeit mehr, sich auszubreiten.

Ich habe den Zeitraum, zu dem anfangs eine Infektion ausgeschlossen ist, sinnvollerweise vom durchschnittlich erreichbaren Alter der Individuen abhängig gemacht und ihn willkürlich, aber durch meine Simulationsergebnisse gedeckt, auf das fünffache des Alters festgesetzt, das ein Individuum im Durchschnitt erreicht. Die Implementierung dieses Durchschnittswertes geschieht mittels eines Systemparameters *alter*, der in 13.8 näher erläutert wird. Ebenso habe ich eine Infektion am Ende eines Simulationslaufes ausgeschlossen; auch dieser Zeitraum wird auf das fünffache des Alters festgesetzt, das ein Individuum durchschnittlich erreicht. Die Variable *anfgeninfekt*, die den Anfangszeitpunkt der Infektion beschreibt, wird folglich durch die Programmzeile

$$\text{anfgeninfekt} = 5 * \text{alter} + \text{rand}() \% (\text{gesamtgenerationen} - 10 * \text{alter}); \quad ^{92}$$

⁹⁰ In der Systemtheorie werden solche äußeren Ereignisse oft *Störungen* genannt.

⁹¹ In systemtheoretischer Sprache heißt das, daß das System sich erst einmal *einschwingen* muß.

⁹² Der Term $\text{rand}() \% (m)$ erzeugt mit gleicher Wahrscheinlichkeit Zufallszahlen im Bereich $0, \dots, m-1$. Durch $5 * \text{alter} + \text{rand}() \% (\text{gesamtgenerationen} - 10 * \text{alter})$ wird also eine Zufallszahl im Bereich zwischen $5 * \text{alter}$ und $\text{gesamtgenerationen} - 5 * \text{alter} - 1$ erzeugt.

initialisiert. Sie wird für die eigentliche Simulation nur an einer Stelle im Programm, die auch durch einen Kommentar gekennzeichnet ist, abgefragt. Aber dieser Wert ist natürlich auch für die statistische Auswertung von Bedeutung.

13.4.2 empirische Daten

Dieser Weg wird begangen, wenn schon umfangreiche Erfahrungen mit den untersuchten Phänomenen vorliegen, insbesondere dann, wenn die Störungen keiner theoretischen Verteilung gehorchen.

Im uns interessierenden Zusammenhang bietet sich dieser Weg offensichtlich an: Wie in Kapitel 15 noch dargestellt werden wird, wird das Ausmaß der Aufnahme deutscher Wörter ins Polnische in hohem Maße durch die geschichtlich-politische Situation, also durch kontingente Ereignisse, bestimmt. Es ist daher überhaupt nicht zu erwarten, daß sich diese Aufnahme mit Hilfe theoretischer Methoden, also rein mathematischer Verteilungen, beschreiben läßt. Stattdessen habe ich die Zeiten, zu denen die Übernahme eines Lehnwortes stattgefunden hat, direkt aus der Datenbank extrahiert und als Infektionszeitpunkt interpretiert. Dann muß verifiziert werden, daß die Zeitpunkte, zu denen die Wörter wieder verschwinden, sofern dies geschieht, vernünftig verteilt sind. Die genauen Experimente, mit denen die Verifikation gelungen ist, werden in Kapitel 16 beschrieben.

13.5 Ausbreitung der sprachlichen Neuerung

Wie wird nun, nachdem ich den Anfang der Infektion beschrieben habe, deren Ausbreitung modelliert? Der Nachfolgezustand für ein Individuum i berechnet sich in Abhängigkeit vom alten Zustand von i und den alten Zuständen der Nachbarn: In der Prozedur *changestate* wird die Fallunterscheidung nach dem alten Zustand von i durchgeführt. Der eigentliche Zustandswechsel wird in der von dort aufgerufenen Prozedur *wechsel* vollzogen. Die dabei benutzten Regeln werden durch weiter unten erklärte Parameter gesteuert.

Zwei Implementierungsdetails sind wichtig:

- Da der alte Zustand eines Knotens i mehrfach benutzt wird, nämlich bei der Berechnung der neuen Zustände aller Nachbarn von i , kann er durch *changestate* nicht einfach überschrieben werden, sobald der neue Zustand berechnet worden ist. Daher ist jedem Individuum ein weiteres Feld *newstate* zugeordnet, das diesen neuen Wert aufnimmt. Die Übertragung von *newstate* als neuem aktuellem Zustand nach *state* erfolgt im Hauptprogramm an der durch einen Kommentar angegebenen Stelle.
- Bei dieser Übertragung muß kurzzeitig unterschieden werden, ob ein gerade neu berechnetes Individuum im Zustand R diesen Zustand schon länger angenommen hat, oder ob er gerade neu erreicht wurde. In diesem Falle sind noch einige Bereinigungen im Hauptprogramm nötig. Deshalb gibt es für das Feld *newstate* die Möglichkeit, kurzfristig neben den Werten S, I und R auch den Wert 3 anzunehmen. Damit wird angezeigt, daß diese Bereinigungen noch nicht

durchgeführt worden sind. Im Hauptprogramm wird dann *R* als neuer Zustand vermerkt.

13.6 Die Anfangsphase der Ausbreitung

Damit überhaupt festgestellt werden kann, ob sich eine Infektion in ihrem Anfangsstadium befindet, gibt es die Variable *infektionsdauer*, die am Anfang eines Simulationslaufes mit 0 initialisiert wird. Solange noch keine Infektion stattgefunden hat, wird das durch 0 signalisiert. Mit Infektionsbeginn wird die Variable zu einer Laufvariablen und in jedem Schritt um 1 hochgezählt. Die Dauer der Anfangsphase der Infektion wird in der Variablen *anfinfekt* festgehalten. Wie wir oben gesehen haben, ist dies ein Systemparameter, und zwar sogar einer, der das Systemverhalten stark beeinflusst. Die Tatsache, daß eine Infektion sich noch am Beginn ihrer Ausbreitung befindet, wird durch eine Boolesche Variable⁹³ ausgedrückt. Diese wird mit Hilfe eines Makros⁹⁴ definiert.

```
#define anfangsstatus (infektionsdauer>0 && infektionsdauer<=anfinfekt)
```

Die Variable *anfangsstatus* ist also genau dann wahr, wenn die Infektion schon begonnen hat (*infektionsdauer*>0) und wenn die Laufvariable *infektionsdauer* noch nicht oft hochgezählt worden ist (*infektionsdauer*<=anfinfekt). Im Programm wird die Variable *anfangsstatus* nur an einer Stelle abgetestet. Sie ist einer der Aufrufparameter der Prozedur *changestate*, die dann den eigentlichen Zustandsübergang bewirkt. Gilt noch der Anfangsstatus für die Infektion, so schreibt *changestate* unabhängig von den übrigen Parametern grundsätzlich eine ungebremste Ausbreitung vor; jeder Kontakt führt also im nächsten Schritt zu einer Infektion.

13.7 Übertragungsregeln

Es gilt nun, die Implementierung der Übertragungsregeln, die ich in Kapitel 11 motiviert und beschrieben habe, zu erläutern. Die Implementierung der Regeln erfolgt in der Prozedur *wechsel*. Deren Aufruf, insbesondere die Auswahl der Werte für die dabei benutzten Parameter, wird durch *changestate* gesteuert. Die Prozedur *changestate* sorgt ferner nicht nur dafür, daß die Anfangsphase berücksichtigt wird; sie steuert auch die Alterung (siehe unten).

⁹³ Eine *Boolesche Variable* ist eine Variable, die sich dadurch auszeichnet, daß sie nur zwei Werte annehmen kann. Diese werden je nach Kontext als *wahr / falsch* oder *1 / 0* bezeichnet. In C++ gibt es eigentlich keine Boolesche Variablen; *anfangsstatus* ist, streng genommen, ganzzahlig. Es ist aber üblich, Variablen, deren Semantik nur die zwei Werte 0 und 1 zuläßt, auch in C++ als Boolesch zu bezeichnen.

⁹⁴ *Makros* sind Zeichenketten, die beim Übersetzungsvorgang in das Programm eingesetzt werden. Dazu wird ohne jegliche Konsistenzprüfung an allen Stellen ihres Auftretens die linke Seite (*anfangsstatus*) durch die rechte (den Booleschen Ausdruck) ersetzt.

Da die Ideen der Übertragungsregeln schon ausgiebig entwickelt worden sind, braucht an dieser Stelle nur aufgezählt zu werden, wie sich die theoretisch hergeleiteten Systemparameter in der Implementierung wiederfinden. Es bedeuten

<i>regelkonstante</i>	die Konstante der Umschaltfunktion u , falls diese konstant ist,
<i>regelsteigung</i>	den Steigungsgrad der Umschaltfunktion u , falls diese linear ist,
<i>latenz</i>	die Latenz, die Dauer des Schutzes für den Zustand vor einer erneuten Änderung.

Wie ich es oben schon erwähnt habe, ist es schwierig, eine mathematische Argumentation für einen bestimmten Wert von *latenz* zu finden. Um bequem zwischen den möglichen Maximalwerten wechseln zu können, habe ich einen Systemparameter *maxlat* eingeführt, mit dem der Bereich der möglichen Zufallszahlen für die Latenz festgelegt wird.

13.8 Alterung im Dynamischen Sprachnetz

Ich weise der Deutlichkeit wegen noch einmal darauf hin, daß die Individuen, die Knoten, hier Sprecher(gruppen) bedeuten. Da die Simulationsläufe einen sehr langen Zeitraum überdecken, muß der Tod eines Individuums ebenfalls modelliert werden. Eine wichtige Rolle bei der Dynamik auf dem Netz spielt daher die Alterung. Das Alter eines jeden Individuums wird nicht explizit gespeichert. Es gibt stattdessen den schon erwähnten Systemparameter *alter*, der angibt, wie viele Generationen ein Individuum im Durchschnitt lebt. Dafür sorgt eine Codesequenz in der Prozedur *changestate*. Diese Sequenz lautet:

```
int i = rand()%(alter);  
if (i==0) {setnewstate(3);return;};
```

In der ersten Zeile wird also eine Zufallszahl zwischen 0 und *alter-1* bestimmt. Mit einer Wahrscheinlichkeit von $1/alter$ ist diese gleich 0, und es findet ein Zustandswechsel des Individuums in den Zustand *R* statt⁹⁵; die weitere Berechnung eines Folgezustands unterbleibt und die Prozedur *changestate* wird verlassen. Der so implementierte Mechanismus ist nicht völlig realistisch, da hier die Sterberaten in absoluten Zahlen unabhängig vom gerade erreichten Alter der Individuen sind. Die Ungenauigkeit ist jedoch von vernachlässigbarer Größe. In Unterkapitel 11.1 habe ich diesen Punkt ja schon angesprochen und mögliche Alternativen erläutert. Dort habe ich auch schon erläutert, daß die Verteilung der Alterswerte von Individuen durch die geometrische Verteilung beschrieben wird. Hier ist insbesondere wichtig, daß sich das durchschnittliche Lebensalter von Individuen als Konsequenz der geometrischen Verteilung in der gegebenen Situation zu *alter* berechnet.

⁹⁵ Warum ich hier *newstate* auf 3 setze, das habe ich ja schon in 13.5 erläutert.

13.9 Sprachliche Erneuerung oder alter Bestand?

Für Auswertungszwecke werden die Knoten in Abhängigkeit von ihrem Zustand in einer der Listen *gesundliste*, *infzliste*, *remliste* verwaltet. Diese Listen müssen natürlich in jedem Simulationslauf neu angelegt werden.

Um bestimmen zu können, wie die drei möglichen Verteilungen der Zustände am Ende der Simulation (Sieg einer der beiden Formen oder Koexistenz) sich langfristig verhalten, habe ich ein array *anzlaeufe* geschaffen, in dem einfach für jeden Satz von Simulationsläufen gezählt wird, wie oft welche Endkonstellation am Ende eines Laufes vorgelegen hat. Es bedeuten also

anzlaeufe[0] die Anzahl der Läufe mit dem Sieg der alten Form⁹⁶,
anzlaeufe[1] die Anzahl der Läufe mit einer Koexistenz beider Formen,
anzlaeufe[2] die Anzahl der Läufe mit dem Sieg der neuen Form,

jeweils am Ende des gesamten Satzes von Simulationsläufen.

Außerdem habe ich die Möglichkeit implementiert, sich bei einzelnen Läufen die Gradverteilung anzusehen. Nach den unten erläuterten Programmverfeinerungen lieferten die Simulationen ausschließlich Pareto-Verteilungen. Ein solches Bild ist Abbildung 39. Die Pareto-Verteilung kann sehr gut gesehen werden.

⁹⁶ Ich erinnere noch einmal an die am Anfang von Kapitel 11 gemachte Bemerkung, daß ich hierunter auch die Wiederverdrängung des Lehnwortes durch einen Neologismus und Vergleichbares subsumiere.

14 Verfeinerungen des Modells

Die bisher beschriebenen Eigenschaften des Modells haben einige Konsequenzen, die zusätzliche Verfeinerungen nötig machen.

14.1 Zerfall des Netzes

Man kann sich bildhaft die Struktur des Netzes so vorstellen, als gebe es einen Kern von Knoten im Zustand R. Um diesen Kern gruppieren sich die lebendigen Knoten (Zustände S,I) herum. Da kein Knoten aus dem Zustand R wieder herauskommt⁹⁷ und somit dieser Kern beständig wächst, während wegen der Alterung die Anzahl der lebendigen Knoten (größenordnungsmäßig) gleich bleibt, ist zu erwarten, daß der Zusammenhang zwischen den lebendigen Knoten verloren geht. Diese bilden dann kein geschlossenes Teilnetz mehr sondern ein Teilnetz mit mehreren nur über R-Knoten mit einander verbundenen und ansonsten isolierten Komponenten. Diesem Teilnetz der Knoten in einem anderen Zustand als R möchte ich in diesem Kapitel ad hoc den Namen *L-Netz* geben. Da sich eine Infektion über R-Knoten nicht ausbreiten kann, bleibt der Gültigkeitsbereich der Neuerung hauptsächlich auf die Zusammenhangskomponente des L-Netzes beschränkt, in der sie erstmals aufgetreten ist. Eine Ausnahme böte lediglich der Fall, daß zwei getrennte Zusammenhangskomponenten durch eine neu eingefügte interne Kante zufällig verbunden würden. Im allgemeinen wird es tendenziell viele kleine derartige Komponenten geben und mit fortschreitender Zeit werden die Zusammenhangskomponenten immer mehr zersplittern. Das hat zur Folge, daß eine Komponente, in der eine Erneuerung dabei ist, sich auszubreiten, eher selten durch neu eingefügte Kanten vergrößert wird. Da die Alterung aber recht schnell zur Schrumpfung der Komponente führt, im allgemeinen sogar zu ihrem vollständigen Verschwinden, ist auch ein allmähliches Verschwinden der auf diese Komponente beschränkten Änderung zu erwarten. Ein typischer durch das so konstruierte Modell prognostizierter Verlauf wäre also folgendermaßen:

- steiler S-förmiger Anstieg auf ein Niveau, das weit weg ist von einer vollständigen Sättigung,
- sehr langsamer Abstieg bis zum völligen Verschwinden.

Experimentelle Befunde haben diese Überlegung bestätigt.

Schuld an diesem unrealistischen Verlauf ist der ebenfalls unrealistische⁹⁸ Zerfall des L-Netzes. Wie also läßt sich dieser verhindern? Es muß eine Verbreitung der Änderung auch weiterhin über das ganze L-Netz möglich sein. Das heißt, dessen Zusammenhang muß zu jedem Zeitpunkt erhalten bleiben. Ein erster scheinbar zielführender Ansatz ist ganz einfach:

⁹⁷ In Übereinstimmung mit der vorherrschenden Interpretation in der Epidemiologie habe ich nicht zwischen gegen die Neuerung immunen und toten Individuen unterschieden. Auf lange Sicht sollte aber der Zustand R einfach auf Grund der Sterblichkeit als „removed“ („tot“) angesehen werden.

⁹⁸ Dem Zerfall entspräche ja ein Zerfall der Sprachgemeinschaft in Gruppen, die nicht mehr miteinander reden.

Bei jedem Ausscheiden eines Knotens k aus dem L-Netz werden alle lebendigen Nachbarn von k , also in der Bezeichnung von Abschnitt 7.4.1 die gesamte lokale Nachbarschaft, miteinander verbunden. Das zeigt das folgende

Beispiel: Knoten mit 3 oder 4 Nachbarn.

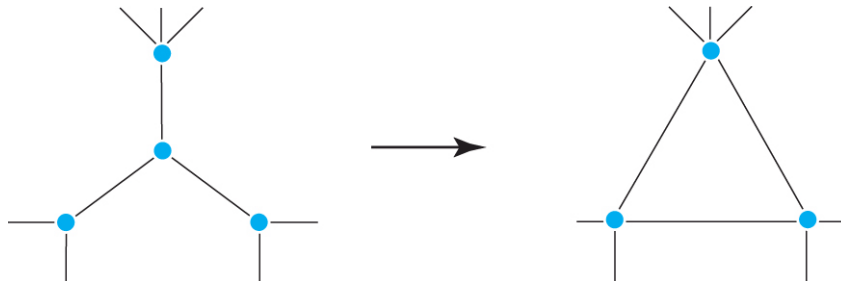


Abbildung 37: Ausfügen eines Knotens mit drei Nachbarn

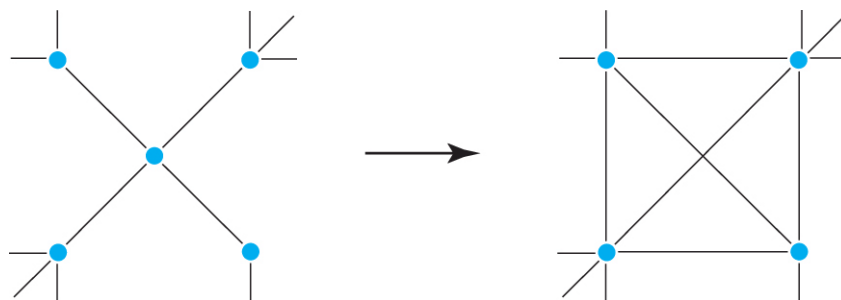


Abbildung 38: Ausfügen eines Knotens mit vier Nachbarn

Die Situation in den Abbildungen ist insofern idealisiert, als es im alten Graphen keine Verbindungen zwischen den Nachbarn des auszufügenden Knotens gibt. Da ich aber Mehrfachverbindungen zwischen Knoten im Modell zulasse, ergibt sich auch im Falle der Anwesenheit solcher Verbindungen im alten Graphen keine Notwendigkeit zu einer prinzipiellen Änderung des Verfahrens.

Das im Beispiel illustrierte Verfahren löst offensichtlich unser Problem: Der Zusammenhang des L-Netzes bleibt gewahrt, und die Änderung, die sich vorher etwa in unserem Beispiel über den zentralen Knoten ausbreiten konnte, kann jetzt direkt über Nachbarschaftsbeziehungen weitergegeben werden. Diese einfache Lösung leidet aber unter einem entscheidenden Nachteil, den ich durch eine kurze Reflektion verdeutlichen möchte. Wir steigen noch einmal in das Beispiel ein.

Beispiel (Fortsetzung): Wieviele Kanten im L-Netz werden gestrichen und wieviele Kanten werden neu eingefügt? Bei dem Knoten mit 3 Nachbarn werden 3 Kanten gestrichen und 3 Kanten eingefügt. Bei dem Knoten mit 4 Nachbarn werden 4 Kanten

gestrichen und 6 Kanten eingefügt. Allgemein gilt: Sei n die Länge der Nachbarschaftsliste des zu streichenden Knotens. Dann werden n Kanten gestrichen und potentiell $n(n-1)/2$ neue Kanten hinzugefügt. Da im allgemeinen n deutlich größer als 3 oder 4 ist, ist die Zahl der neuen Kanten normalerweise *erheblich* größer als die Zahl der gestrichenen⁹⁹. Die Folge davon ist, daß das L-Netz nach einer gewissen Zeit im wesentlichen die Struktur eines riesigen zentralen Blocks hat, bei dem jeder Knoten mit jedem verbunden ist. Darum herum gibt es einige relativ junge Einzelknoten, die mit Knoten im zentralen Block verbunden sind (wegen des preferential linking mit an Sicherheit grenzender Wahrscheinlichkeit). Für einen solchen Knoten m außerhalb gibt es nur zwei Möglichkeiten:

- Er stirbt vor allen seinen Nachbarn. Dann spielt er für die weitere Entwicklung des Netzwerkes keine Rolle mehr.
- Es gibt einen Nachbarn n innerhalb des zentralen Blockes, der vor m stirbt. In diesem Augenblick werden gemäß Konstruktion alle Nachbarn von n , also der gesamte zentrale Block, zu Nachbarn von m . Damit wird auch m zu einem Mitglied des zentralen Blockes.

Diese Überlegung zeigt, daß recht schnell – Experimente haben gezeigt, daß etwa 30 Generationen ausreichen, - das L-Netz zu einem vollständig zusammenhängenden Graph wird¹⁰⁰. Anschaulich bedeutet eine solche totale Verbindung, daß jedes Individuum mit jedem kommuniziert und dieses beeinflusst. In einem so strukturierten Graph breiten sich Neuerungen augenblicklich aus. Das ist offensichtlich zu schnell, wie man etwa am von Piotrovskij und dem von Altmann in seinen mathematischen Begründungen postulierten S-förmigen Anstieg sieht, der gerade keine augenblickliche Ausbreitung zuläßt.

14.2 Umstrukturierung des Netzes

Damit wir zu einem sinnvollen Ansatz zur Umstrukturierung des Netzes unter Wahrung des Zusammenhalts des L-Netzes kommen, ohne daß im Übermaß neue Kanten eingefügt werden, ist eine kleine Anforderungsanalyse nötig:

Es sollten (größenordnungsmäßig) beim Ausscheiden eines Knotens k etwa gleich viele Kanten gestrichen werden wie neu zwischen ehemaligen Nachbarn von k konstruiert werden. Und deren Anzahl ist die der zuvor von k ausgehenden Kanten.

Sei also $d(k)$ der Grad des Knotens k . Dann sollte für Knoten k_1, k_2 gelten:

$$p(k_1, k_2) \propto d(k_1) * d(k_2)$$

Dabei soll $p(k_1, k_2)$ die Wahrscheinlichkeit bedeuten, daß eine Kante zwischen k_1 und k_2 errichtet wird. Die Formel sagt dann, daß diese Wahrscheinlichkeit, genau

⁹⁹ Die Zahl der gestrichenen Kanten wächst *linear*, die Zahl der neu eingefügten Kanten wächst *quadratisch*.

¹⁰⁰ Nach wievielen Generationen dieses passiert, das hängt auch von den Parametern einfintern und einfextern ab. Der Effekt tritt aber in jedem Fall auf. Der angegebene Wert 30 gilt für einfintern = einfextern = 1.

wie es das lineare preferential linking verlangt, proportional zu den Graden der beiden Knoten sein soll.

Sei derjenige Knoten, der beseitigt wird, mit x bezeichnet, und seien k_1, \dots, k_n die Nachbarn von x . Es seien zwei Hilfsgrößen definiert:

$$r = \sum_i d(k_i), \quad R = \sum_i d(k_i)^2$$

Dann ist

$$\sum_{i < j} d(k_i) * d(k_j) = (r^2 - R) / 2$$

Beweis: Es ist nach Definition

$$r^2 = \sum_{i,j} d(k_i) * d(k_j)$$

Werden davon die Summanden abgezogen, bei denen die beiden Indizes gleich sind, deren Summe ja gerade gleich R sind, erhalten wir $r^2 - R$. Für die Hälfte der Summanden gilt $i < j$. Wird über alle diese summiert, ergibt sich als Ergebnis $(r^2 - R) / 2$.

qed

Daher könnte durch die Programmzeile

$$k = \mathbf{rand}() \% ((r^2 - R) / 2);$$

zunächst eine Zufallszahl mit $0 \leq k < \sum_{i < j} d(k_i) * d(k_j)$ erzeugt werden. Würde zur Erzeugung einer Kante von k_i nach k_j dieser Wert mit $d(k_i) * d(k_j)$ verglichen, etwa durch einen Pseudocode der Form

if ($k < d(k_i) * d(k_j)$) erzeuge eine Kante von k_i nach k_j ;

so wäre die Wahrscheinlichkeit, eine Kante von k_i nach k_j zu erzeugen, genau proportional zu $d(k_i) * d(k_j)$. Würde ferner die Summe über alle verschiedenen i, j mit $i < j$ gebildet, so wäre in etwa der Wert 1 zu erwarten, also würde im Mittel etwa eine Kante erzeugt. Damit im Mittel ungefähr n (, was ja gerade gleich der Anzahl der alten Nachbarn des zu beseitigenden Knotens ist,) Kanten erzeugt werden, wird das obige Programmfragment modifiziert¹⁰¹:

if ($k < d(k_i) * d(k_j) * n$) erzeuge eine Kante von k_i nach k_j ;

Wird der Knoten x gestrichen, so entfallen auch seine Kanten zu k_1, \dots, k_n ; insgesamt sind das also n Kanten. Genauso viele Kanten werden im Mittel durch das obige Programmfragment, das in der Prozedur namens „umstrukturieren“ benutzt wird, in der Nachbarschaft von x neu erzeugt. Auch die Forderung nach preferential linking wird auf diese Weise gewahrt.

¹⁰¹ Strikt gesehen, müßte an dieser Stelle etwas komplexer, nämlich mit Gegenwahrscheinlichkeiten, gearbeitet werden. Das würde aber zu einer unnötigen Komplexität bei einem vergleichsweise unbedeutenden Detail des Programms führen.

14.3 Änderung der Zielgröße: Verteilung auf Kategorien

Um zunächst feststellen zu können, ob einer der wichtigsten Ansprüche erfüllt ist, den ich an das Dynamische Sprachnetz gestellt habe, daß nämlich sowohl der komplette Sieg der ursprünglichen Form als auch ein kompletter Sieg der Neuerung als auch eine Koexistenz beider Formen möglich sein soll, habe ich zunächst die Simulationsergebnisse nicht für die eigentliche Zielgröße, die Verweildauer der Neuerung, ausgewertet sondern nur einfach gezählt, wie viele Neuerungen in welche der drei Kategorien fallen. Alle drei Kategorien sind in den Läufen gut vertreten. Das zeigt die beigefügte Tabelle in der Datei ergebnisvordimred.txt. Damit ist vorerst zumindest kein offensichtlich unbrauchbares Ergebnis entstanden.

14.4 Dimensionsreduktion

Nachdem in Unterkapitel 12.2 die Idee der Dimensionsreduktion allgemein geschildert worden ist, möchte ich jetzt beschreiben, wie ich sie für das konkrete Problem durchgeführt habe:

Gemäß der Konstruktion ist im Programm eine Reihe von unabhängig veränderbaren Systemparametern, also Dimensionen im Untersuchungsraum, vorhanden. Die folgenden habe ich zunächst einzeln untersucht:

- einfintern, einfextern

Diese Parameter bezeichnen die Anzahl der in jeder Generation eingefügten internen und externen Kanten. Nach der Theorie sind die Form des Netzes und die Dynamik nicht von den genauen Werten abhängig. Das haben meine Simulationen im großen und ganzen bestätigt. Es entstehen bei verschiedenen Parameterwerten jeweils eine Pareto-Verteilung und vergleichbare Verteilungen bei der Zahl von möglichen Benutzern der beiden Sprachformen. Jedoch zeigt sich, daß mit steigendem Parameterwert das Programm immer langsamer und etwa ab dem Wert 4 sogar extrem gebremst wird. Da diese Werte schon von der Theorie her nicht signifikant sind und zusätzlich unter Umständen praktisch zu Schwierigkeiten führen, habe sie einfach für alle wichtigen Durchführungen auf den Wert 1 gesetzt. Selbstverständlich lassen sich jedoch auch andere Werte durch eine Änderung an nur einer Stelle im Programm abtesten.

- initgenerationen

Dieser Parameter ist in Unterkapitel 13.2 eingeführt und motiviert worden. Dort habe ich bereits darauf hingewiesen, daß es sich hier um einen aus pragmatischen Gründen eingeführten Wert handelt, der nur einigen mathematischen Unregelmäßigkeiten vorbeugen soll. Die weiterhin dort angegebenen Argumente rufen geradezu nach einer Dimensionsreduktion in der angedeuteten Art: Der Parameter sollte also hinreichend groß sein, damit er seinen Zweck erfüllen kann, aber auch nicht zu groß, so daß der

Einschwingprozeß nicht zu lange dauert. Experimente haben gezeigt, daß die Zahl 10 ein passender Wert ist. Wieder lassen sich jedoch auch andere Werte durch eine Änderung an nur einer Stelle im Programm abtesten.

- alter, gesamtgenerationen

Diese beiden Parameter stehen in einem engen Verhältnis zueinander: Meine Verwendung des Begriffs „Generation“ steht in Übereinstimmung mit dem Sprachgebrauch bei Zellularen Automaten / Dynamischen Netzen. Es handelt sich einfach um die diskreten Zeitpunkte, zu denen das System beobachtet wird. Damit ist gesamtgenerationen ein Wert, der zunächst keinen Bezug in der realen Welt besitzt. Man könnte eine einzelne Simulation auch so interpretieren: Es wird einfach eine gewisse Anzahl von Lebensaltern simuliert, nämlich so viele wie der Quotient gesamtgenerationen / alter angibt. Dann bezeichnet alter nichts weiter als die Granularität der Messungen, nämlich wie oft während eines Lebensalters hingeguckt und gemessen wird. Die Dauer einer Simulation sollte im vorliegenden Fall keinen wesentlichen Einfluß auf das Ergebnis haben¹⁰². Da es sich nicht um chaotische Systeme handelt, sollte auch die Granularität der Untersuchung keinen bedeutenden Einfluß haben, vorausgesetzt, das Zeitraster wird nicht zu grob. Diese Vorüberlegungen haben sich in einer Reihe von Simulationsläufen tatsächlich bewahrheitet. Ich habe die beiden Parameter deshalb auf eine mittlere Größe festgesetzt, die die Länge der Simulationen nicht in nicht mehr zu bewältigende Größen steigert andererseits aber noch hinreichend umfangreiche und hinreichend fein granulare Messungen zuläßt.

- anfgeningekt

Dieser Parameter bestimmt den zufälligen Zeitpunkt der Infektion. Es versteht sich von selbst, daß dieser Zeitpunkt im Modell keinen Einfluß auf das prinzipielle Wachstumsverhalten der Infektion haben sollte. Dabei muß allerdings auf die in 13.4.1 erwähnten Sonderbedingungen am Anfang und am Ende der Simulation geachtet werden. Aus den dort genannten Gründen, insbesondere dem der allgemeineren Verwendbarkeit, habe ich die zufallsgesteuerte Wahl des Infektionszeitpunktes beibehalten, jedoch dafür gesorgt, daß eine Zufallszahl im Bereich zwischen $5 \cdot \text{alter}$ und $\text{gesamtgenerationen} - 5 \cdot \text{alter} - 1$ erzeugt wird. Damit ist gewährleistet, daß die bei initgenerationen erwähnte Einschwingphase beendet ist und sich andererseits das Fortschreiten der Neuerung noch für einige Zeit beobachten läßt. Den Parameter anfgeningekt habe ich dann natürlich in den unten näher erläuterten Test übernommen, eben um die Unabhängigkeit der Messungen davon zu zeigen. Die Durchführung der Tests hat dann gezeigt, daß der Parameter in der Tat keinen Einfluß auf die Zielgröße hat.

Nach diesen Schilderungen und den obigen theoretischen Vorüberlegungen werden zusätzlich zu anfgeningekt noch sechs unabhängige Systemparameter identifiziert, die nicht durch unsere Anschauung erfaßbar oder durch mathematische Überlegungen vorhersagbar sind. Dabei bedeuten

¹⁰² Eigentlich müßte hier zwischen zeitlich begrenzten Phänomenen wie etwa der initialen Ausbreitungsphase einer einzelnen Neuerung und globalen Phänomenen des Systems unterschieden werden. Näheres kann bei [Wol02] nachgelesen werden. Im vorliegenden Fall sind beliebig lange Simulationen ohnehin sinnlos und die Simulationsdauern willkürlich gewählt.

regelkonst	regelsteig	anfinf	latenz	prefadd	prefmult
Siehe unten	Siehe unten	Die Dauer des Anfangs der Infektion	Die Dauer der Latenz	Die additive Konstante beim Vorrang	Die multiplikat. Konstante beim Vorrang

Hier muß genauer untersucht werden, wie sich ihr Einfluß manifestiert. In 500 Simulationsläufen werden diese Parameter unabhängig von einander variiert. Das Ergebnis ist einschließlich statistischer Auswertungen in der Datei ergebnisvordimred.txt zu ersehen.

Hier folgen einige Erläuterungen zu den Ergebnissen: Es ist jeweils genau einer der Werte regelkonstante oder regelsteigung gleich 0, je nachdem, ob wir eine konstante oder eine lineare Umschaltfunktion u in der Zustandsüberföhrungsfunktion vor uns haben, also Funktionen der Form

$$u(n) = c \quad \text{oder} \quad u(n) = d \cdot n$$

Zum einen wird in dem Programm durch die Regelkonstante der Wert der Funktionskonstante c bezeichnet zum anderen durch die Regelsteigung der Wert d der konstanten Steigung von u .

Für die Zielgröße habe ich die naheliegenden drei Kategorien gebildet. Dabei bedeuten

- 0 den kompletten Sieg der ursprünglichen Form,
- 1 eine Koexistenz beider Formen,
- 2 einen kompletten Sieg der Neuerung,

gemessen jeweils nach 1000 Zeitschritten. Es muß also für jeden der insgesamt sieben Parameter untersucht werden, ob die Zielgröße, die Verteilung auf die Kategorien, von diesem Parameter abhängig ist oder nicht. Wie sind diese Parameter in der sprachlichen Wirklichkeit zu interpretieren? Welche Einflüsse sind bei naiver Interpretation zu erwarten?

- Die Parameter regelkonst und regelsteig geben an, wie viele Nachbarn eines jeden Individuums an einer Änderung seines Sprachverhaltens beteiligt sein müssen. Sie charakterisieren also den „Konservatismus“ der Sprecher. Der dürfte intuitiv einen großen Einfluß auf das Schicksal einer Änderung haben.
- Der Parameter anfinf gibt an, wie lange wir uns das Anfangsstadium der Infektion vorstellen. Da dabei die Infektion ungebremst verläuft, könnte dieser Parameter als „anfängliche Intensität“ interpretiert werden. Auch hier sollte ein großer Einfluß zu merken sein.
- Der Parameter latenz wurde schon bei seiner Einführung in Unterkapitel 11.9 diskutiert. Dort hatte ich mit Latenz die Zeit bezeichnet, die mindestens nach einem Zustandswechsel eines Knotens vor einem erneuten Zustandswechsel dieses Knotens vergehen muß. Es wurde auch hervorgehoben, daß eine hohe Latenz ein

konservatives Sprachverhalten begünstige. Dadurch werde das Entstehen einer neuen sprachlichen Form nicht unmöglich gemacht; aber dann komme es nur schwer wieder zu einer Verdrängung der Neuerung.

- Die Parameter *prefadd* und *prefmult* beeinflussen die Form des Kommunikationsnetzes. Je höher sie sind, desto mehr kommunikative Individuen, also solche mit vielen Kommunikationspartnern, sind zu erwarten. Inwieweit sich die Existenz von kommunikativen Partnern auf die Ausbreitung von Neuerungen auswirkt, ist durch eine naive Interpretation nicht zu entscheiden¹⁰³.
- Die Generation, in der die Infektion stattfindet, ist nichts weiter als der modellmäßig erfaßte Zeitpunkt des ersten Auftretens der Neuerung. Da wir von universellen, insbesondere also zu jedem Zeitpunkt gültigen allgemeinen Gesetzen ausgehen, dürfte der überhaupt keinen regelmäßigen Einfluß haben.

Um zunächst einen groben Überblick über die Lage zu kriegen, so daß festgestellt werden kann, ob die naiven Erwartungen den Ergebnissen des Modells intuitiv entsprechen, habe ich zunächst den Korrelationskoeffizienten zwischen der Zielgröße und je einem der Parameter berechnet¹⁰⁴. Hier seien zunächst die Korrelationskoeffizienten zwischen der Zielgröße und dem jeweiligen Parameter aus der Datei *ergebnisvordimred.txt* aufgeführt:

<i>regelkonst</i>	<i>regelsteig</i>	<i>anfinf</i>	<i>latenz</i>	<i>prefadd</i>	<i>prefmult</i>	Gener.d.Infekt.
-0.367673	0.214084	0.527373	0.159921	-0.08221	0.058168	0.023711

Schon auf den ersten Blick¹⁰⁵ bestätigen die Ergebnisse das, was zu erwarten war: die Unabhängigkeit (also relativ kleine Korrelationskoeffizienten) der Zielgröße von den Parametern Präferenz (multiplikativ und additiv) sowie Zeitpunkt der Infektion. Genauere Untersuchungen haben gezeigt, daß diese Unabhängigkeiten sowohl für konstante als auch für lineare Umschaltfunktionen gelten; der einzige Unterschied ist der, daß die Unabhängigkeit für lineare Umschaltfunktionen noch ausgeprägter ist. Diese Nachweise sind jederzeit reproduzierbar. Deshalb, und weil ich meine Leser nicht mit zuviel Statistik erschlagen möchte, gehe ich hier nicht näher auf die entsprechenden Untersuchungen ein.

Um nachweisen zu können, daß die genannten Unabhängigkeiten kein Zufall in diesem Testlauf sind, habe ich den χ^2 -Unabhängigkeitstest (vgl. [Sac97])¹⁰⁶ verwandt. Dieser Test soll jetzt etwas genauer erklärt werden:

¹⁰³ Die in Unterkapitel 9.1 zitierten Ergebnisse lassen einige Vermutungen zu, die ich im Ausblick (Unterkapitel 17.2) noch etwas erläutern werde.

¹⁰⁴ Um zusätzlich zu unterstreichen, wie sinnvoll das gewählte Vorgehen ist, habe ich in gesonderten Simulationsläufen die Zielgröße präzisiert: Dabei habe ich für jeden Simulationslauf den Anteil der infizierten Knoten an der Gesamtzahl der Knoten am Ende des Laufes, nach 1000 Zeitschritten berechnet und diesen Quotienten als Zielgröße anstelle der reinen Kategorieinteilung benutzt. Da der Unterschied zwischen den beiden Vorgehensweisen sich in jedem Falle als minimal herausgestellt hat, werde ich diesen Ansatz hier nicht weiter verfolgen oder dokumentieren.

¹⁰⁵ Da es sich hier nur um einen groben ersten Überblick handelt, verzichte ich auf eine formale Untersuchung bei den Korrelationskoeffizienten. Eine formale Analyse wird dann mit dem gleich erklärten χ^2 -Unabhängigkeitstest nachgeliefert werden. Diesen ersten Schritt habe ich besonders deshalb durchgeführt, weil eine Übersicht mit einem Korrelationskoeffizienten für die meisten Menschen, die damit zu tun haben, erfahrungsgemäß anschaulicher ist als der χ^2 -Test.

¹⁰⁶ Im Grunde sind alle χ^2 -Techniken Unabhängigkeitstests, so auch etwa der χ^2 -Anpassungstest, der auf die Frage antwortet, ob zwei gegebene Merkmale unabhängig oder die entsprechenden Verteilungen gut aneinander

Ich wähle dazu als Beispiel den Parameter anfinf, die Dauer der Anfangsphase einer Infektion. Ich habe in Unterkapitel 11.8 ja schon dahingehend argumentiert, daß dann nur die Betrachtung von Parameterwerten 0,1,2,3 sinnvoll ist. Da die Zielgröße in die drei Kategorien 0,1,2 eingeteilt ist, kann die Analyse einer *Zweiwegtafel*¹⁰⁷ des Typs 3x4 angewandt werden¹⁰⁸. Für den allgemeinen Fall muß eine Tabelle mit r Zeilen und c Spalten betrachtet werden. Dazu seien

- n der Stichprobenumfang,
- n_{ij} die Besetzungszahl in der i -ten Zeile und der j -ten Spalte,
- n_{i*} die Summe der Besetzungszahlen in der i -ten Zeile,
- n_{*j} die Summe der Besetzungszahlen in der j -ten Spalte.

Dann ist eine Prüfgröße zu bestimmen:

$$\chi^2 = n (\sum_i \sum_j n_{ij}^2 / n_{i*} n_{*j} - 1)$$

Bei Gültigkeit der Nullhypothese, in unserem Fall der Hypothese der Unabhängigkeit, und bei hinreichend großem n muß die so berechnete Prüfgröße mit dem in einer Tabelle vorliegenden Wert für χ^2 mit $(r-1)(c-1)$ Freiheitsgraden verglichen werden. Bei der Überprüfung von anfinf liegen also $(3-1)(4-1) = 6$ Freiheitsgrade vor. Das n gilt als hinreichend groß, wenn alle Erwartungshäufigkeiten mindestens den Wert 1 haben. Das ist bei uns erfüllt. Die Auswertung im konkreten Beispiel für alle Parameter ergibt folgende χ^2 -Werte (obere Zahlenreihe):

regelkonst	regelsteig	anfinf	latenz	prefadd	prefmult	Gener.d.Infekt.
85.855423	40.271626	167.833023	127.565742	7.971048	10.604679	12.771309
9.49	9.49	12.59	12.59	12.59	12.59	28.87

Die untere Zahlenreihe bezeichnet den einer Tabelle entnommenen Wert, der über Annahme oder Ablehnung der Nullhypothese von der Unabhängigkeit auf dem 5%-

angepaßt sind. Bei uns werden die Merkmale mehrfach gestuft sein, das heißt, sie können Werte in mehreren Auswertungsbereichen annehmen.

¹⁰⁷ Eine Zweiwegtafel beschreibt die Zusammenhänge zwischen Merkmalen aus zwei gestuften Kategorien. Der Typ 3x4 bedeutet, daß ein Merkmal in drei und das andere in vier Wertebereiche gestuft ist. Der bekannteste Test mit einer sehr einfachen Zweiwegtafel ist der χ^2 -Vierfeldertest. Dabei sind beide Merkmale in zwei Wertebereiche gestuft.

¹⁰⁸ Diese Analyse kann im allgemeinen nur verwandt werden, wenn die Observable nur diskrete Werte annehmen kann. Sonst müssen die Parameterwerte in Kategorien eingeteilt werden. Bei der vorliegenden Arbeit ist das beispielsweise bei angeninfekt der Fall.

In den durchgeführten Untersuchungen waren in jedem Simulationslauf die gesamte Anzahl der Generationen gleich 1000 und das Alter gleich 50. Damit besteht der oben angesprochene Wertebereich für angeninfekt zwischen $5 \cdot \text{alter}$ und $\text{gesamtgenerationen} - 5 \cdot \text{alter} - 1$ aus dem Zahlenintervall [250,749]. Das habe ich zum Zwecke des χ^2 -Unabhängigkeitstest in zehn gleich gleichlange Teilintervalle zerlegt. Dadurch konnte eine Zweiwegtafel des Typs 3x10 aufgestellt werden.

Signifikanzniveau entscheidet¹⁰⁹. Für das Beispiel des Parameters anfinf ergibt sich dann eine Entscheidung:

$$167.833023 > 12.59 = \chi^2_{6;0,05}$$

Damit ist die Nullhypothese abgelehnt.

Insgesamt bestätigt der χ^2 -Test die Interpretation der Ergebnisse der Korrelationskoeffizienten, also „das, was ohnehin schon zu sehen ist“¹¹⁰: Die Simulationsergebnisse sind deutlich beeinflusst durch die Parameter regelkonst , regelsteig , anfinf und latenz aber weitgehend unabhängig von prefadd und prefmult sowie der Generation der Infektion. Diese Ergebnisse sind bei näherem Hinsehen auch überhaupt nicht überraschend. Daß etwa eine höhere Konstante Änderungen erschwert ist genauso anschaulich wie die Tatsache, daß der Eintrittszeitpunkt die Dynamik der Änderung nicht beeinflussen dürfte.

Für die eigentlichen Simulationen habe ich dann folgende Entscheidungen getroffen:

- Latenz

Angesichts der schon in Unterkapitel 11.9 festgestellten Tatsache, daß eine hohe Latenz extreme Ergebnisse fördert, habe ich einen Parameter maxlat für den Maximalwert eingeführt und diesen auf 4 gesetzt, was also Latenzen von 0 bis 3 zuläßt. Wie üblich braucht nur dieser Parameter geändert werden, damit eine andere Bandbreite von Werten möglich wird.

- Anfang der Infektion

Die Variable anfinfdauer gibt an, wie lange eine Infektion sich zu Beginn mit ungebremster Dynamik ausbreitet: Kontakte führen zu sofortiger Ansteckung. Da in den betrachteten Netzen die Pareto-Verteilung gilt und der small-world-Effekt auftritt, sind hohe Werte sinnlos. Ich habe auch hier einen Parameter maxininf für den Maximalwert eingeführt und ihn auf 4 gesetzt, da unsere Netze relativ klein sind¹¹¹.

- prefadd , prefmult

Diese sowohl gemäß der Theorie als auch nach dem obigen χ^2 -Test irrelevanten Parameter habe ich, um eine gewisse Bandbreite zu haben, auf den Bereich $\{1,2,3,4\}$ festgesetzt.

- regelkonstante , regelsteigung

Auch hier ist es so, daß sowohl die Anschauung als auch die durchgeführten Simulationen es nahelegen, daß zu hohe Werte die Simulation sinnlos machen. Sie wirken hier prohibitiv. Ich habe die Werte zunächst auf den Bereich $\{1,2,3\}$ festgesetzt, wie ich es in den Abschnitten 11.4.1 und 11.4.2 schon begründet habe. Hier ist die einzige Stelle, wo ich später bei den durchgeführten Experimenten ein Problem bezüglich der Übereinstimmung mit den realen Daten bekommen habe. Darauf werde ich bei der Schilderung der Experimente noch genauer eingehen.

¹⁰⁹ Ein einfacher Blick in eine χ^2 -Tabelle zeigt, daß sich auf dem 1%-Signifikanzniveau dieselben Ergebnisse ergeben hätten.

¹¹⁰ Da tatsächlich vorhandene Verteilungen nie genau zu den theoretischen Verteilungen passen, kommt es bei sehr großen n zu dem Phänomen, daß bei einem χ^2 -Test alles signifikant sein kann. Allerdings sind in dem vorliegenden Fall alle Werte für n so klein, daß dieses Problem hier vernachlässigt werden kann.

¹¹¹ Ich erinnere daran, daß ich in Abschnitt 7.4.3 den Wert 6 als kennzeichnend für wesentlich größere Netze, etwa die gesamte Menschheit, zitiert habe.

14.5 Beispiel eines Sprachnetzes

Zum Abschluß der Überlegungen zu Netzen zeige ich hier noch beispielhaft ein Netz, das mit Hilfe des Werkzeuges Viator [HTK+02] auf der Grundlage des Dynamischen Sprachnetzes gezeichnet wurde. Es handelt sich dabei um den ersten der von `simulation.cpp` erzeugten und in der beigefügten Datei `gesamtprognoszuf.txt` (nähere Erläuterungen dazu in Unterkapitel 16.5) dokumentierten Läufe.

Die Parameter dieses zufällig erzeugten Simulationslaufes sind
regelkonstante = 0,
regelsteigung = 3,
Dauer des Anfangs der Infektion: `anfinfoaduer` = 0,
latenz = 2,
prefadd = 1,
prefmult = 3,
Generation der Infektion: `infektstart` = 419.

Das folgende Bild zeigt die Situation in der Sprechergemeinschaft nach 1000 Generationen:

Nach 1000 Simulationsschritten sind insgesamt noch 55 Knoten, also Individuen(gruppen), vorhanden; alle übrigen sind in den Zustand *R* gewechselt und werden nicht mehr angezeigt.

Von den noch vorhandenen Knoten sind 28 gesund (Zustand *S*), in der Abbildung blau gezeichnet; die 27 infizierten Individuen (Zustand *I*) sind in rot gehalten. Der besseren Identifizierbarkeit wegen habe ich den Knoten (nur für diese Zeichnung) einen eigenen Namen gegeben. Er besteht aus der zweimaligen Zustandsbezeichnung und der Nummer des Ranges seiner Erzeugung. So besagt etwa der Namen `SS945`, daß der Knoten im Zustand *S* ist und als 945. Knoten im Verlauf der Simulation erzeugt wurde. Eine Kante zwischen zwei Individuen zeigt an, daß sie miteinander kommunizieren. Die Topologie des Netzes zeigt eine klare Potenzverteilung: 25 Individuen haben einen Nachbarn, 12 haben zwei Nachbarn. Andererseits können die kommunikativsten Individuen `II874`, `II865` und `SS952` mit 24, 11 bzw. 9 Nachbarn als hubs angesehen werden.

Mehr über diesen Beispiellauf ist in den Dateien `potenzverteilung.txt` und `potenzvertgenerationen.txt` dokumentiert (siehe auch 16.5).

15 Unsicherheiten die Übernahme von Lehnwörtern betreffend

Wir sind jetzt an der Stelle angekommen, an der im Rahmen der bisherigen Untersuchungen die beiden ersten Ebenen aus Abbildung 1 realisiert sind. Die Theorie ist im Kern ausgearbeitet, wenn auch noch an vielen Stellen Möglichkeiten zur Verfeinerung bestehen; und die Computersimulation läuft offenbar fehlerfrei. Wir müssen jetzt den Schritt zur realen Welt wagen. Bevor aber Experimente, also Auswertungen der Oldenburger Datenbank, dazu herangezogen werden können, die bis hierher entwickelten Gedankengänge zu verifizieren (oder im Popperschen Sinne: eine Falsifikation zu ermöglichen), sind noch einige Bemerkungen zu dem Untersuchungsgegenstand nötig, von denen einige auf die Arbeiten [Hen95], [HM01], [Hen01] zurückgehen.

15.1 Der Begriff „Lehnwort“

Zunächst ist der Begriff des *Lehnwortes* selber zu präzisieren: Im Rahmen des experimentellen Teils meiner Untersuchungen ist ein Lehnwort schlicht und einfach das, was in der Datenbank verzeichnet ist¹¹². Insbesondere fließen nur Wörter in die Untersuchungen ein, die deutschen (germanischen) Ursprungs sind und ohne Vermittlung über dritte Sprachen (etwa das Tschechische) ins Polnische übernommen wurden. Dies ist somit ein Idealtyp des "deutschen Lehnwortes im Polnischen"¹¹³. Ferner beschränkt sich der Bestand auf die Hochsprache. Dabei wird ein dezidiert diachroner Standpunkt eingenommen: Von den Wörtern, die diese Kriterien erfüllen, sind, und das ist für mich natürlich besonders wichtig, alle, auch die inzwischen wieder ausgeschiedenen, verzeichnet.

Es wird für die uns interessierenden Daten der Erstbeleg verzeichnet, sofern dieser vorhanden ist. Alternativ, allerdings nur bei einer Minderheit der Wörter, ist hingegen das Datum der Erstbuchung angegeben. Der Unterschied zwischen beiden Begriffen besteht darin, daß eine *Buchung* das Auftreten des Wortes in einem zeitgenössischen Wörterbuch bezeichnet, also nur einen indirekten Nachweis liefert, und ein *Beleg* das Auftreten des Wortes in einem nichtlexikographischen Text, [Hen95]¹¹⁴. Ich werde künftig zwischen diesen

¹¹² Dieses Verfahren ist selbstverständlich legitim, da die Oldenburger Datenbank nach wissenschaftlichen Kriterien erstellt worden ist, die ich damit implizit übernehme.

¹¹³ Weitere Typen mit Abweichungen vom Ideal wären

- Wörter deutschen (germanischen) Ursprungs, die über vermittelnde Sprachen ins Polnische eingegangen sind,
- Wörter mit anderer Herkunft, die aus dem Deutschen direkt ins Polnische übernommen worden sind,
- Wörter mit anderer Herkunft, die über das Deutsche und anschließend eine vermittelnde Sprache (vor allem Tschechisch) ins Polnische übernommen worden sind.

In der umfangreichen philologischen Literatur der letzten Jahre sind Elemente als deutsche Lehnwörter im Polnischen beschrieben worden, die einem dieser vier Typen entsprechen, in der Regel ohne diese auf Entlehnungswegen beruhenden Typen zu definieren. In der Oldenburger Datenbank und somit auch hier geht es nur um den Idealtypus. Bei nicht ganz eindeutiger Lage wurde bei der Erstellung der Datenbank vom Idealtypus ausgegangen.

¹¹⁴ Eine Buchung sollte eigentlich nur dann genommen werden, wenn der Erstbuchung ein Beleg folgt. Nur gebuchte Lehnwörter sind (faktisch, aus Gründen der ausgewerteten Wörterbücher) zweifelhaft hinsichtlich ihres hochsprachlichen Status und insbesondere hinsichtlich ihrer Verbreitung.

beiden Begriffen nicht unterscheiden und allgemein auch dann von Erstbelegen reden, wenn Erstbuchungen gemeint sind¹¹⁵.

Das erste Auftreten eines Lehnwortes in einer Sprache liegt in der Regel früher als der erste schriftliche Nachweis, Belege in der Regel früher als Buchungen. Da in Zeiten mit geringerer Verschriftlichung die Differenz zwischen diesen beiden Zeitpunkten größer sein dürfte als in Zeiten, für die viele Quellen erhalten geblieben sind, ist zu erwarten, daß im Schnitt ein Wort, je älter es ist, desto länger schon vor dem ersten schriftlichen Auftreten in der mündlichen Sprache verwandt wurde.

Lehnwörter werden aus semantischer Sicht so gut wie nie 1:1 aus der Ursprungssprache übernommen. So werden von polysemen Lexemen in der Gebersprache zumeist nur monoseme Verbindungen von Ausdruck und Inhalten übernommen, beziehungsweise nur eine kleinere Zahl von Inhalten als in der Ursprungssprache. Und auch später geht ein Lehnwort bedeutungsmäßig in der Regel eigene Wege. Bei polysemen Lehnwörtern liegen in der Datenbank prinzipiell Erstbelege für jede Verbindung einer (Teil-)Bedeutung mit dem entsprechenden entlehnten Ausdruck vor, wobei hinzukommt, daß eine Untermenge der Teilbedeutungen unter Umständen eine semantische Eigenentwicklung der entnehmenden Sprache ist. So muß man sich entscheiden, wie man sich bezüglich Ausdrucks- und Inhaltsorientierung verhalten will.

In dieser Arbeit gehe ich weder rein inhalts- noch rein ausdrucksorientiert vor. Reines inhaltsorientiertes Arbeiten läßt sich im allgemeinen deshalb schwer durchhalten, weil eine formale theoretische Darstellung von Inhalten nicht möglich ist. Deshalb muß man sich mit Paraphrasierungen behelfen; und genau das ist in der Oldenburger Datenbank geschehen. Deshalb ist sie – wie alle herkömmlichen Wörterbücher – für eine Verifikation nach einem rein inhaltsorientierten Vorgehen gar nicht geeignet¹¹⁶. Für eine reine Ausdrucksorientierung wäre das dann zu entwickelnde Modell extrem kompliziert: In diesem Falle müßte jede Ausdrucksvariation als etwas anderes, als neues Lehnwort, interpretieren werden, denn der "vereinigende Aspekt des Inhalts" entfielen ja. Dann müßte das Modell so etwas wie eine Aufspaltung von Knoten enthalten. Bisher fehlen alle Ansätze zu einer geeigneten mathematischen Theorie für solche Modelle. Andererseits ergibt sich hier durchaus ein Ansatz zu zukünftigen Forschungen mit tiefergehendem mathematischem Schwerpunkt als in dieser Arbeit.

Das in der Arbeit gewählte Vorgehen könnte man als "zeichenorientiert" bezeichnen, wobei "Zeichen" wie üblich als "Ausdruck und Inhalt" zu verstehen ist¹¹⁷. Eine Inhaltsorientierung liegt insofern vor, als ich monoseme Einheiten aus der Datenbank übernommen habe. Der jeweilige Ausdruck wird dann sozusagen als "Etikett" für solche monosemen Einheiten benutzt.

An dieser Stelle ist eine gewisse empirische Unsauberkeit nicht zu vermeiden: Ist das erste Auftreten eines Lehnwortes auf jeden Fall eine echte Entlehnung, so kann das schon beim Erscheinen eines zweiten Paares mit dem gleichen Etikett aber einem unterschiedlichen Inhalt schon nicht mehr immer sicher gesagt werden. Die Frage, ob es sich um eine zweite

¹¹⁵ Erstbuchungen werden in der Oldenburger Datenbank kompensatorisch für „zweifelhafte“ Erstbelege verwandt.

¹¹⁶ Eine reine Inhaltsorientierung wäre nicht zuletzt in den Fällen unsinnig, wo Lehnwörter offenbar als Modeerscheinungen für alte genuine Wörter übernommen werden. Hier ist die "interlinguale Synonymie" Voraussetzung für die Verwendung des entlehnten Ausdrucks.

¹¹⁷ Da Lexeme immer Einheiten aus Ausdruck- und Inhalt sind, ist eigentlich nur die Option der Zeichenorientierung vernünftig. Im Grunde genommen diskutiere ich die Optionen der Ausdrucks- oder Inhaltsorientierung hier nur, um sie zu verwerfen.

Entlehnung desselben Wortes handelt oder um eben die zweite Verwendung eines schon etablierten monosemen Zeichens in der aufnehmenden Sprache, und damit, streng genommen, nicht um das Ergebnis einer echten Entlehnung, ist oft nur schwer zu entscheiden. Außerdem ist der Unterschied zwischen echter Entlehnung und einer Neubildung ja nicht einmal eine reine Alternative. Es können beide Prozesse parallel ablaufen und einander beeinflussen.

Mitunter ergeben sich spezifische Probleme: Der genaue Grad, inwieweit eine inhaltliche Adaption vorliegt, läßt sich eben recht schwer bestimmen.

Ich möchte einige Beispiele nennen:

- Beim polnischen Wort „*herb*“ zum Beispiel, aus dem deutschen Wort „*Erbe*“ (durch Vermittlung über das Tschechische) entlehnt, kann die ursprünglich im Deutschen nicht angelegte heutige Bedeutung 'Wappen' nur durch inhaltliche Anpassung im Slawischen erklärt werden.
- Ein Wort kann andererseits mehrfach entlehnt werden. Eine solche Interpretation liegt für das deutsche Wort „*Kittel*“ nahe. Es ist im 17. Jahrhundert in der Bedeutung 'Arztkittel' ins Polnische entlehnt worden. Zuvor hatte schon eine Entlehnung desselben Wortes in der Bedeutung 'Soldatenmantel' stattgefunden¹¹⁸.
- Schwerer zu fassen ist auch die Tatsache, daß ein Wort unter Umständen eine sehr hohe Ausdrucksvariation hat. Ein Beispiel dafür ist das polnische Wort „*blacha*“ mit vielen Varianten, das aus dem deutschen Wort „*Blech*“ stammt. In der Oldenburger Datenbank ist für jede Variante, soweit sie nicht rein orthographisch ist, der Erstbeleg vermerkt. Für diese Untersuchungen werden jedoch nur die ersten Belege des Inhalts mit einer beliebigen von gegebenenfalls mehreren Ausdrucksvarianten (zum Beispiel *blach*, *blech*, *blecha*, *plach*, ...) verwendet, und so gibt es mit diesem Phänomen natürlich in unserem Modell keine Schwierigkeiten.

Ein weiterer Grund, warum ich mich für ein zeichenorientiertes Vorgehen entschieden habe, bei dem eine mehrfache Übernahme eines von der Form her identischen Wortes auch mehrfach gezählt wird, ist auch der, daß ein solches zeichenorientiertes Vorgehen kein singulärer Ansatz dieser Arbeit sondern auch in vielen linguistischen Untersuchungen auf ähnlicher Datenbasis üblich ist. Im Grunde bedeutet der von mir gewählte Weg, daß monoseme Einheiten aus Ausdruck und Inhalt angenommen werden. Diese werden dann im Grunde als „unabhängige Homonyme“ behandelt.

Eine andere Idealisierung betrifft das Ausscheiden eines Wortes aus einer Sprache. Auch dieses muß nicht endgültig sein. Es gibt Beispiele für die Wiederauferstehung eines bereits toten Wortes, selbst bei Lehnwörtern [Hen01]. Solche Vorgänge sind relativ selten und können ebenfalls bei uns als „statistisches Rauschen“ ignoriert werden.

Es gibt ferner kontingente Ereignisse, die in einer systematischen Untersuchung wie der vorliegenden nicht berücksichtigt werden können. Unter der Voraussetzung, daß kein systematischer Fehler vorliegt, sind die meisten derartigen Einflüsse wieder wegen des

¹¹⁸ Eigentlich handelt es sich hier um einen Zwitter zwischen Polysemie und Homonymie. Es soll an dieser Stelle aber nicht untersucht werden, um wieviele Entlehnungen es sich bei den Abkömmlingen von „*Kittel*“ nun tatsächlich handelt. Das ist für mich auch absolut kein Problem, da ich ohnehin mit monosemen Einheiten arbeite, womit eine mitunter mögliche Unterscheidung von Polysemie und Homonymie unerheblich wird. Im allgemeinen geht man bei Homonymie trotz gleichem Ausdruck von zwei Wörtern aus. Diese werden im Falle von Lehnwörtern unabhängig entlehnt. Polyseme Lehnwörter sollten mit allen Semen, die aus der Urbildsprache überhaupt übernommen werden, „gleichzeitig“ entlehnt werden. Das ist natürlich eine Idealisierung.

statistischen Charakters der Modelle unerheblich, da die Effekte kleiner als die statistischen Schwankungen sind.

15.2 Die benutzten Daten

In meiner Untersuchung habe ich alle *hapax legomena*¹¹⁹ sowie alle nur gebuchten Lehnwörter ausgeschlossen. Auch hier schließe ich mich einem in der Linguistik üblichen, auch von Hentschel, so in [Hen01], praktizierten Verfahren an, mit dem den Folgen einer unsicheren Datenlage vorgebaut und eine Unterscheidung zwischen „echten“ und „gelegentlichen“ Lehnwörtern (etwa Zitatwörtern oder nicht hochsprachlichen Elementen) getroffen werden soll. Der Ausschluß von *hapax legomena* und nur gebuchten Lehnwörtern ist auch ein Filter zur Absicherung des hochsprachlichen Status des jeweiligen Elementes.

Gleichfalls den Folgen einer unsicheren Datenlage vorbauen soll eine andere Konstruktion, die ebenfalls bereits in der Datenbank benutzt wird: Das *Indexjahr* ist entweder das „exakte“ Jahr des Erstbelegs, sofern dieses bekannt ist, oder es wurde berechnet. Ist eine Quelle, die den ersten Beleg eines Lehnwortes enthält, nicht genau datierbar, das heißt nur in einem Intervall zu datieren, so werden „mittelwertorientierte“ Indexjahre willkürlich, jedoch nach festen Regeln fixiert. So würde etwa die Angabe „Erste Hälfte des 15. Jahrhunderts“ zu dem Indexjahr 1425 führen. Bei Angaben der Form „vor 1730“ wäre das Indexjahr jedoch 1730, da – wie gesagt – schriftliche Belege in der Regel nach der Übernahme erfolgen.

In der Oldenburger Datenbank werden prinzipiell nur solche Elemente mit deutscher Herkunft als Lehnwörter anerkannt, die in der Quelle keinen Hinweis auf die Fremdheit haben. Wenn in einer polnischen Quelle beispielsweise ein deutsches Element in Anführungszeichen steht und keine Anzeichen einer lautlichen, morphologischen oder orthographischen Adaption zeigt, so wird es als Fremd- oder Zitatwort eingestuft und nicht berücksichtigt.

Die Angaben zu den Letztbelegen sind so zu lesen, daß das Wort aus der Sprache ausgeschieden ist, falls ein solches Datum vorliegt. Beim Fehlen eines Letztbeleges gehört das Wort noch zum Sprachbestand¹²⁰.

Um die Unsicherheiten, die mit der genauen Erfassung von Wörtern in den letzten Jahren vor der Erstellung eines Wörterbuches immer einhergehen, zu vermeiden, habe ich bei der Auswertung das Ende willkürlich auf das Jahr 1950 festgesetzt.

15.3 Der deutsche Einfluß auf das Polnische

Bevor wir uns aber der Verifikation der Modelle zuwenden, will ich noch auf einige spezifisch polnische Faktoren aus der Sprachgeschichte eingehen, deren Einfluß doch außerhalb von Zufallsschwankungen liegen dürfte. Diese Faktoren hängen selbstverständlich

¹¹⁹ (griech.: einmal gelesen): Wörter, die nur durch eine Quelle überliefert sind, auch dann wenn sie in dieser Quelle mehrfach benutzt werden.

¹²⁰ Den Begriff „*Letztbeleg*“ zu definieren ist recht schwierig. In [Hen01] wird das Fehlen eines Wortes in bestimmten einschlägigen Wörterbüchern (z.B. Doroszewski,[Dor58]) als starker Indikator für das Ausscheiden des Wortes angesehen, zumindest aus der Hochsprache, um die es hier geht.

eng mit den allgemeineschichtlichen Vorgängen zusammen. Das Wichtige sind für uns aber nicht die historischen Ereignisse sondern in erster Linie die aus ihnen resultierenden Folgen für die Sprache.

Die folgende Einteilung habe ich den Arbeiten [Hen95], [HM01], [Hen01] entnommen, wo viele weitere Details zu finden sind.

- 13./14. Jahrhundert

Zu dieser Zeit kam es auf Grund der deutschen Siedlungsbewegung nach Osten zu starken Sprachkontakten und zu relativ vielen Übernahmen deutscher Wörter. Andererseits sind die verfügbaren polnischen Quellen bis zur Mitte des 14. Jahrhunderts äußerst spärlich, so daß es wenige Belege für diese Zeit gibt.

- 15./16. Jahrhundert

In dieser Zeit scheint in beträchtlichem Maße ein gleichbleibender Zustrom deutscher Wörter stattgefunden zu haben; jedoch ist die Quellenlage sehr unterschiedlich. In der zweiten Hälfte des 16. Jahrhunderts, während des polnischen „*Goldenen Zeitalters*“ (der Renaissance) ist, bedingt durch das Schaffen bedeutender Schriftsteller wie *Mikolaj Rej* und *Jan Kochanowski* sowie durch die Erfindung des Buchdrucks, das polnische Schrifttum ganz allgemein besser belegt als zuvor. Das gilt dann natürlich auch für die überlieferten, belegten deutschen Lehnwörter.

- 17./18. Jahrhundert

Dieser Zeitraum ist gekennzeichnet durch das Ende einer nennenswerten deutschen Einwanderung und, auch auf Grund der *Gegenreformation*, eine deutliche Schwächung des deutschen Einflusses. In diese Zeit fällt auch die Literatur des Barock. Das Ende dieser Epoche ist eindeutig durch die *Polnischen Teilungen* (1772, 1793, 1795) gegeben.

- 19./20. Jahrhundert

Da durch die Polnischen Teilungen große Gebiete Polens unter die Herrschaft zweier deutschsprachiger Staaten gerieten, wuchsen die Extensität und Intensität der Sprachkontakte zwischen deutschen und polnischen Bevölkerungsteilen wieder stark an und daher auch der Einfluß des Deutschen. Ein gewisser Rückgang der Neuentlehnungen seit der erneuten polnischen Unabhängigkeit im Kontext des Ersten Weltkrieges ist nicht zu übersehen. Die Verstärkung dieser Tendenz nach dem Zweiten Weltkrieg fällt wegen meiner Zeitbeschränkungen für diese Arbeit nicht mehr ins Gewicht.

Hentschel faßt die beiden ersten Epochen dann, etwa in [Hen01], zu einer einzigen Epoche zusammen.

16 Das Modell spiegelt die Wirklichkeit wider

Zur Überprüfung, ob das bisher entwickelte Modell tatsächlich die Wirklichkeit widerspiegelt, habe ich zwei Arten von Experimenten durchgeführt. Diese haben im allgemeinen sehr gute Übereinstimmungen zwischen den Ergebnissen des Modells und den Werten der Datenbank gezeigt. Wir wollen uns aber zunächst überlegen, was „sehr gute Übereinstimmung“ überhaupt bedeutet. Was können wir in den Experimenten erwarten und was nicht?

Als erstes sollte bemerkt werden, daß in dem Lehnwörterbuch (in der Datenbank dann in entsprechender Größenordnung) etwa 2500 Artikel vorhanden sein werden, davon etwa 600 über wieder ausgeschiedene Wörter. In diesen werden etwa 4600 monoseme Einheiten aus Ausdruck und Inhalt abgehandelt, von denen inzwischen 800 wieder aus dem lexikalischen Bestand des Polnischen ausgeschieden sind. Damit ist der Datenbestand hinreichend groß für valide statistische Aussagen, und die von mir in den Simulationen produzierten Datenmengen haben eine angemessene Größe.

In der Datenbank sind zum zweiten (und wesentlich wichtiger) systematisch nur Erst- und, sofern vorhanden, Letztbelege angegeben¹²¹. Es kann damit also keine Analyse durchgeführt werden, ob der Ausbreitungsvorgang für ein einzelnes Wort so modelliert worden ist, wie er typischerweise auch in der Realität stattfindet. Was ich für den Fall eines einzelnen Wortes tun konnte und getan habe, das ist lediglich, das Programm gegen die Theorie zu testen, also die mittlere Ebene in Abbildung 1 gegen die oberste, nicht aber die unterste gegen die oberste. Mit Hilfe des Programmes wird für den Fall, daß nur ein Simulationslauf für ein einziges Wort durchgeführt wird, eine Datei mit Kenndaten für das entstandene Netz angelegt. Es zeigt sich in Abbildung 39, daß tatsächlich eine Pareto-Verteilung entstanden ist. Das beweise ich gar nicht erst. Man sieht es sofort! Ein Indikator ist die Existenz von hubs. Auf den einfach zu führenden mathematischen Nachweis (etwa durch den unten beschriebenen Kolmogoroff-Smirnoff-Test) habe ich an dieser Stelle verzichtet, da die Existenz einer Pareto-Verteilung für den Vergleich mit der Datenbank zwar eine für die Theorie notwendige Bedingung ist aber an sich keine relevante Erkenntnis für die Sprachwissenschaft. Deshalb ist das Vorhandensein der richtigen Verteilung für das eigentliche Ziel, die formale Untersuchung der Sprachdynamik, ohne Belang.

Was allerdings getan werden kann, das ist, daß man große Wortmengen betrachten und diese mit statistischen Methoden untersuchen kann. Da für uns nur die Erst- und Letztbelege wichtig sind, ist zu untersuchen, ob Eigenschaften der Mengen dieser Daten auch durch das Modell reproduziert werden können. Wieder gibt es eine ganz natürliche Grenze der Aussagefähigkeit des Modells: Wir haben keine Chance, irgendwelche Eigenschaften der Menge der Erstbelege vorherzusagen. Dazu sind diese Vorgänge viel zu sehr an zufällige reale Vorgänge gebunden, die keinesfalls in einem Modell erfaßt werden können. Es ist aber eine der Grundannahmen dieser Arbeit, daß es, sobald ein Wort, sozusagen als Angebot, eingedrungen ist, *objektive, allgemeingültige Gesetzmäßigkeiten* gibt, die in den lokalen Kommunikationen verborgen sind und die den weiteren Verlauf seiner Existenz steuern. Folglich müßten die Aufenthaltsdauern von realen Wörtern sich in irgendeiner Form in den Ergebnissen des Modells wiederfinden lassen. Und die statistischen Eigenschaften der Aufenthaltsdauern sind genau das, was ich zur Untermauerung meines Modells heranziehen werde. Natürlich überlagern die eben angesprochenen objektiven

¹²¹ Es liegen vielfach auch weitere Belege vor, manchmal auch weitere Zahlenangaben. Diese können aber nicht als repräsentativ für die Verwendung von Lehnwörtern angesehen werden und erlauben nicht die Ableitung statistisch gesicherter allgemeiner Gesetze zum Sprachwandel.

Gesetzmäßigkeiten sich mit zufälligen Ereignissen (zum Beispiel Moden, während derer verstärkt Lehnwörter aufgenommen werden¹²², oder Wellen von Purismus, in denen zahlreiche Lehnwörter wieder verdrängt werden können), aber die Hoffnung ist, daß sich solche Ereignisse bei einem Massenphänomen wie der Betrachtung sehr vieler Wörter über sehr lange Zeiträume herausmitteln und im „statistischen Rauschen“ untergehen. Umgekehrt gilt: Wenn trotz prinzipiell nicht vorhersehbarer Störungen wesentliche Charakteristika der Aufenthaltsdauern sich im Modell reproduzieren lassen, ist das ein klares Indiz für die Angemessenheit des Modells (und mehr als Indizien kann es ja im Sinne Poppers nicht geben). *Schließlich ist es äußerst unwahrscheinlich, daß das Modell fehlerhaft ist aber die Wirklichkeit durch eine zufällige Störung gerade so verändert wird, daß sie zum Modell paßt.*

Wovon man zur Auswertung eines konkreten Experiments unter diesen Prämissen also ausgehen muß, ist die Existenz zweier Zahlenmengen. In diesem Kapitel werde ich zunächst schildern, wie ich sie erhalten habe, und dann beschreiben, wie ich sie ausgewertet habe. Das schließt die Beschreibung und die Begründung des Kriteriums für den Begriff der „Ähnlichkeit“ der beiden Mengen ein.

16.1 Die Experimente

Für jeden Eintrag in der Datenbank, der Erst- und Letztbeleg enthält, wird die Differenz der beiden Werte herangezogen. Ist kein Letztbeleg vorhanden, kann angenommen werden, daß das Wort noch vorhanden ist. Dann wird das Jahr 1950, das Ende meiner Betrachtungen, dafür eingesetzt. Dieses Verfahren führt deswegen zu keinem systematischen Fehler, weil, wie unten noch geschildert wird, auch die Simulationen nach einer gewissen Zeit einfach enden.

Das in Unterkapitel 15.1 angesprochene Problem der semantischen Differenzierung habe ich ignoriert, indem ich von möglicherweise mehrfacher Übernahme monosemer Ausdrucks-Inhaltsverbindungen ausgegangen bin. Auch in diesem Fall sorgt das statistische Rauschen dafür, daß man sich darüber keine Sorgen zu machen braucht.

Bei den Simulationen habe ich der Entwicklung eines jeden Wortes 1000 Schritte (Generationen) Zeit gegeben. Dieser Wert ist frei gewählt, aber nicht völlig willkürlich. Eine Transformation bei der Auswertung wird dann besonders einfach. Bei dieser Transformation handelt es sich um das „*skaling*“. Dieser Begriff muß jetzt erläutert werden, da ich ihn öfter benutzen werde.

Die Simulation muß in die Wirklichkeit abgebildet werden. Dabei erhebt sich die Frage, wieviel Zeit in der Wirklichkeit entspricht einer Generation des Modells? Es geht also wieder um die Frage der Granularität der Simulation. Stelle ich mich beispielsweise auf den Standpunkt, ich rechne eine Generation des Modells gleich einem Jahr in der Wirklichkeit, so muß ich das bei der Auswertung berücksichtigen. Soll dagegen eine Generation einem halben Jahr entsprechen, so brauche ich nicht nur doppelt so viele Schritte (und erhalte eine im allgemeinen genauere Simulation), ich muß das auch bei der Auswertung berücksichtigen. Die beiden Zahlenmengen für die Aufenthaltsdauern dürfen also nicht einfach verglichen

¹²² Es sollte beachtet werden, daß Moden in meiner Modellierung nicht Interpretation der oben erwähnten S-Kurven sind: Jedes einzelne Wort folgt der S-Kurve, aber über das gehäufte Auftreten von Wörtern mache ich keine Aussagen. Durch meinen Trick, mit Aufenthaltsdauern zu arbeiten statt mit absoluten Werten, habe ich die Wirkung solcher zufälliger Ereignisse neutralisiert. Das bestätigen sowohl die Theorie als auch die noch zu schildernden Experimente.

werden. Man muß die simulierten Ergebnisse noch mit einem entsprechenden Skalierungsfaktor multiplizieren¹²³. Meine Simulationen sind auf das Skalierungsverhältnis von 1:1 ausgelegt.

In der beigegeführten Beispielsimulation habe ich dann die Ausbreitung von 500 Wörtern mit zufällig erzeugten Parametern, die in den schon begründeten Bereichen liegen, durchgespielt. Das ist bei weitem ausreichend für statistisch signifikante Resultate¹²⁴. Da bei meiner Modellbildung hapax legomena nicht vorkommen, habe ich als Ersatz alle Wörter mit einer kleinen Aufenthaltsdauer (bis zu 10 Generationen) aus den Simulationsergebnissen gestrichen¹²⁵. Durch dieses Experiment erhält man zwei Mengen von Daten, eine mit Daten aus der Datenbank, eine mit Daten aus der Simulation. Die Frage ist, ob die beiden Mengen aus einer gemeinsamen Grundgesamtheit stammen können.

Bei dem eben beschriebenen Experiment habe ich nur die Wörter betrachtet, die wieder ausgeschieden sind. Einmal um auch einen etwas anderen Ansatz heranzuziehen, andererseits auch um alle Wörter, nicht nur die wieder verschwundenen, in die Untersuchungen einbeziehen zu können, habe ich noch ein zweites Experiment durchgeführt: Wurden bei dem eben geschilderten Verfahren Wörter rein zufällig, wie in Abschnitt 13.4.1 beschrieben, erzeugt, so habe ich diesmal, wie in Abschnitt 13.4.2 motiviert, nicht zufällig erzeugte Eintrittsdaten verwandt sondern diese (wieder im Skalierungsverhältnis von 1:1) aus der Datenbank entnommen und die Simulation im Jahre 1000 gestartet. Dadurch sind Eingaben im Bereich von 213 (für das früheste nachweisbare Wort im Jahre 1213) bis 950 (wegen des Simulationsendes im Jahre 1950) entstanden. Die wieder in der Simulation ausgeschiedenen Wörter habe ich wie oben behandelt. Für die in der Simulation noch vorhandenen Wörter, seien es solche, die die alte Form wieder vollständigen verdrängt haben,

¹²³ Es mag hier willkürlich erscheinen, noch einen Faktor einzuführen. Daß das nicht so ist, zeigt die folgende Überlegung: Das früheste belegte Wort ist 1213 ins Polnische eingedrungen. Wir betrachten also einen Zeitraum von 1950 – 1213 = 737 Jahren. Andererseits sind von den 1000 Schritten der Simulation etwa die ersten 250 noch nicht brauchbar, da das System noch nicht „eingeschwungen“ ist, zum Beispiel durch die anfangs noch nicht korrekten Altersverteilungen. Deshalb ist die Simulation so programmiert, daß Infektionen erst nach 250 Schritten möglich sind. Die mögliche Ausbreitungsdauer von 750 Schritten legt dann ein Skalierungsverhältnis von 1:1 nahe. Hätte ich 2000 Schritte simuliert, müßte das Skalierungsverhältnis eher 1:2 sein.

Es könnte sich das Bedenken einschleichen, der zusätzliche Faktor könne das Ergebnis verfälschen. Aber wir wollen ja nur die charakteristischen Eigenschaften, insbesondere also die Verteilungen der beiden vorliegenden Mengen von Aufenthaltsdauern untersuchen, genauer: Wir wollen nachweisen, daß die Verteilungen gleich sind. Lügen nun zwei verschiedene Verteilungen vor, könnten sie auch durch Multiplikation mit einem Faktor nicht ineinander überführt werden. Das zeigt ein kleines Beispiel. Wenn die eine Menge etwa gleichverteilt ist (also zwischen den Extremwerten überall ungefähr gleich viele Werte liegen) und bei der anderen Menge sich etwa eine deutliche Häufung bei den niedrigen Werten für die Aufenthaltsdauern ergibt, so kann man herummultiplizieren, so viel man will, die Gleichverteilung bei der einen Menge und die Häufung bei der anderen kriegt man nicht weg; die eine Menge kann also nicht als Simulationsergebnis für die andere angesehen werden, und der unten beschriebene Kolmogoroff-Smirnoff-Test würde dieses Ergebnis auch tatsächlich liefern.

¹²⁴ Daß 500 ein angemessener Wert ist, das kann sowohl der beigegeführten Datei auswertung.xls entnommen als auch an Abbildung 40 und Abbildung 41 gesehen werden. Nehmen wir das in Abbildung 40 dargestellte erste Experiment. Dabei werden nur wieder ausgetretene Wörter betrachtet. Es zeigt sich, daß von allen Wörtern der Datenbank 64 wieder ausgeschieden und damit hier verwendbar sind. Von den 500 Simulationsläufen liefern 169 ein Beispiel für reversiblen Sprachwandel, also ein erneutes Ausscheiden. Es hätten also eher weniger als mehr Simulationsläufe herangezogen werden sollen. Die Größenordnungen stimmen aber noch. Beim zweiten Experiment lauten die entsprechenden Zahlen: Es wurden 442 Wörter aus der Datenbank und 488 angewendete Simulationen herangezogen.

¹²⁵ Sporadische Untersuchungen mit Einschluß der hapax legomena und der Wörter mit einer kleinen Aufenthaltsdauer lassen die Vermutung entstehen, daß die Ergebnisse der Simulationen dann eher besser als schlechter werden. Zumindest sind die Ergebnisse der statistischen Auswertungen mit Hilfe des noch zu erläuternden Kolmogoroff-Smirnoff-Tests noch besser.

oder solche, die in Koexistenz mit der alten Form existieren¹²⁶, habe ich eine entsprechende Zahl von Wörtern ohne Letztbeleg aus der Datenbank berücksichtigt. Das heißt, es wurde eine entsprechende Zahl von Werten der Form *1950-Erstbeleg*, also die bis zum Ende des betrachteten Zeitraumes gültige Mindestaufenthaltsdauer, aus der Datenbank zu der Stichprobe mit den realen Werten für den reversiblen Sprachwandel hinzugefügt. Wieder erhält man zwei Mengen von Daten und wieder erhebt sich die Frage, ob die beiden Mengen aus einer gemeinsamen Grundgesamtheit stammen können.

16.2 Die Ergebnisse

Beim ersten Augenschein schien zunächst keine gute Annäherung zwischen den Werten der Datenbank und der Simulation zu bestehen. Es waren bei der Simulation zu viele Wörter mit kurzer Aufenthaltsdauer vorhanden. Eine nähere Betrachtung der Simulationsergebnisse ergab, daß diese Wörter zu einem äußerst hohen Anteil auf die Parameterkombination

regelkonstante = 0, regelsteigung = 1

zurückzuführen war¹²⁷. Diese Kombination führte fast immer zu sehr kurzen Aufenthaltsdauern, so stark, daß das Gesamtergebnis stark verfälscht wurde. Der Befund läßt sich im nachhinein dadurch erklären, daß besonders bei hochverbundenen Knoten genau diese Kombination diejenige ist, die einen Zustandswechsel am schwierigsten macht. Sie bedeutet ja, daß für einen Zustandswechsel mindestens die Hälfte der Nachbarn im komplementären Zustand sein muß. Das bedeutet für eine sich ausbreitende Neuerung nach der Anfangsphase eine unüberwindliche Hürde. In meinen Simulationen habe ich die kritische Konstellation durch die folgende Programmsequenz ausgeschlossen. Der Befund sagt dann allgemein, daß das Dynamische Sprachnetz das Verhalten der deutschen Lehnwörter im Polnischen hervorragend modelliert, sofern keine extreme Hürden für die Ausbreitung errichtet werden.

Den Ausschluß der kritischen Parameterkombination habe ich in der folgenden Programmsequenz bewerkstelligt:

```
int steigung = rand()%2;
regelwert = steigung+1+rand()%(3);
regelsteigung = steigung*regelwert;
regelkonstante = (1-steigung)*regelwert;
```

¹²⁶ Die Unterscheidung zwischen diesen beiden Wortklassen kann der Datenbank ja nicht entnommen werden. Andererseits ist er aus theoretischer Sicht bedeutsam. Ein Modell, das nicht beides vorhersagen kann, taugt nichts.

¹²⁷ Die verschiedenen Simulationsläufe unterscheiden sich ja dadurch, daß anfangs verschiedene Parameter zufällig gesetzt werden, beispielsweise die Latenz oder die Generation der Infektion. Zwei dieser Parameter sind eben regelkonstante und regelsteigung. Die obige Parameterkombination sagt also nur, daß, unabhängig von den anderen Parametern, dem Parameter regelkonstante der Zufallswert 0 und dem Parameter regelsteigung der Zufallswert 1 zugewiesen wurden. Diese Kombination ist natürlich unter allen möglichen Parameterkombinationen ein eher seltenes Ereignis.

Durch die erste Zeile wird festgelegt, ob die Form der Umschaltfunktion konstant oder linear ist¹²⁸. Danach müssen die konkreten Werte festgelegt werden, sei es die Höhe der nötigen Konstante (zur Erinnerung: Das ist die Anzahl der für einen Zustandswechsel benötigten Nachbarn.) oder sei es die Größe der Steigung (Auch das bestimmt die Anzahl der für einen Zustandswechsel benötigten Nachbarn, aber auf eine etwas kompliziertere Weise.) Das geschieht in der zweiten Zeile. Die Variable regelwert nimmt dann folgende Bereiche an:

konstanter Fall (steigung hat den Wert 0): regelwert $\in \{1,2,3\}$
linearer Fall (steigung hat den Wert 1): regelwert $\in \{2,3,4\}$

Der Trick ist der, daß zuerst durch den Term $1+\text{rand}()\%3$ Zufallszahlen im Bereich $\{1,2,3\}$ erzeugt werden und dann im Falle einer linearen Umschaltfunktion noch 1 addiert wird. Dadurch wird die kritische Konstellation vermieden: Eine lineare Funktion mit dem Parameterwert 1 ist nicht mehr möglich. In den letzten beiden Zeilen werden die tatsächlichen Parameterwerte zugewiesen:

- konstante Umschaltfunktion:

Dann hat steigung hat den Wert 0 und damit auch die Variable regelsteigung; regelkonstante ist gleich dem Wert der eben berechneten Variablen regelwert.

- lineare Umschaltfunktion:

Dann hat steigung hat den Wert 1 und damit die Variable regelkonstante den Wert 0 (In diesem Falle ist ja (1-steigung) = 0.); regelsteigung ist gleich dem Wert der eben berechneten Variablen regelwert.

16.3 Der Kolmogoroff-Smirnoff-Test

Nach der Implementierung des funktionalen Modells ist zu untersuchen, ob für jedes der beiden Experimente tatsächlich eine Übereinstimmung der Ergebnisdaten vorliegt. Dies geschieht üblicherweise durch den *Kolmogoroff-Smirnoff-Test*. Wie dieser Test zu bewerten ist, das läßt sich am besten in den Worten von [Sac97] (Abschnitt 393) ausdrücken (alle Hervorhebungen im Original, Auslassungen von mir an der mit „...“ bezeichneten Stelle):

„Sind zwei unabhängige Proben von Meßwerten (oder von Häufigkeitsdaten) hinsichtlich der Frage zu vergleichen, ob sie aus derselben Grundgesamtheit stammen, dann gilt der Test von *Kolmogoroff* (1933) und *Smirnoff* (1939) als SCHÄRFSTER HOMOGENITÄTSTEST. **Er erfaßt Unterschiede der Verteilungsform aller Art:** Insbesondere Unterschiede der zentralen Tendenz (Mittelwert, Median), der Streuung, der Schiefe und des Exzesses, d.h. Unterschiede der Verteilungsfunktion (vgl. ...), d.h. $H_0: F_1=F_2$ gegen $H_A: F_1 \neq F_2$.“¹²⁹

¹²⁸ Ein Ausdruck der Form $\text{rand}()\%n$ erzeugt, wie schon erwähnt, ganzzahlige Zufallszahlen im Bereich $\{0, \dots, n-1\}$. Die erste Zeile bestimmt also eine Zufallszahl, die 0 oder 1 sein kann. Damit wird festgelegt, ob die Umschaltfunktion konstant (steigung = 0) oder linear (steigung = 1) sein soll.

¹²⁹ Wie üblich stehen bei Sachs H_0 und H_A für Null- und Alternativhypothese.

Zur Durchführung des Testes wird für jede der beiden Meßreihen, im statistischen Kontext meist „Stichproben“ genannt, die zugehörige empirische Verteilung (in dem Sinne, wie es in Abschnitt 7.2.1 beschrieben ist,) ermittelt. Zur Erklärung soll das folgende Beispiel dienen. Es ist insofern etwas vereinfacht, als der Kolmogoroff-Smirnoff eigentlich nur für stetige Verteilungen angewandt werden kann; aber es illustriert die Punkte, auf die es ankommt, und es ist gut verständlich. In den Anwendungen dieser Arbeit liegen dann hinreichend viele Meßwerte vor, so daß man die Verteilungen als stetig annehmen und den Test anwenden kann.

Beispiel: Gegeben sei eine Messung mit den Werten 1,2,2,3,5. Man kann sich etwa die nach der Größe geordneten Würfe mit einem Würfel denken. Diese Meßreihe kann als endliche Stichprobe aus einer Grundgesamtheit unbekannter Größe interpretiert werden, der Menge aller möglichen Ergebnisse von Würfeln mit diesem Würfel. Die empirische Verteilungsfunktion f gibt für jede reelle Zahl x den Anteil der Meßwerte an, die kleiner sind als x (oder gleich). Für $x=2,5$ sind drei Meßwerte kleiner als x von insgesamt fünf, und daher ist $f(2,5) = 3/5$.

Die so entstandene Funktion hat alle Eigenschaften einer (empirischen) Verteilungsfunktion. Beim Vergleich zweier Stichproben werden zunächst die beiden Verteilungsfunktionen bestimmt und dann mit Hilfe des Kolmogoroff-Smirnoff-Tests das Maximum der Differenzen zwischen ihnen gesucht. Das heißt mathematisch: Seien zu zwei Stichproben die Verteilungsfunktionen f_1 und f_2 gegeben. Dann berechnet sich das Maximum D der Differenzen durch

$$D = \max_x |f_1(x) - f_2(x)|$$

Beispiel: Sei eine weitere Messung mit den Werten 1,2,3,4,5,6 gegeben. So ergeben sich die Differenzen $D(x)$ der Verteilungsfunktionen an relevanten Stellen x zu

$$D(0)=0, D(1)=1/5-1/6, D(2)=3/5-2/6, D(3)=4/5-3/6, D(4)=4/5-4/6, D(5)=1-5/6, D(6)=0.$$

Das Maximum dieser Zahlen liegt also bei $D(3)=4/5-3/6=3/10$.

Das Ergebnis dieser Auswertung muß mit einer Prüfgröße D_α verglichen werden, die von dem sogenannten Signifikanzniveau α abhängt. Das *Signifikanzniveau* bestimmt für den Test die Irrtumswahrscheinlichkeit. Erreicht oder übersteigt der aus den Stichproben ermittelte Wert D die Prüfgröße D_α , so liegt auf dem verwendeten Niveau ein statistisch signifikanter Unterschied vor. Die Nullhypothese ist dann abzulehnen. Für die Auswertung meiner Simulationsexperimente habe ich mich auf das 1%-Niveau bezogen. Es gibt also eine einprozentige Irrtumswahrscheinlichkeit, ein ziemlich strenges Kriterium. Für kleine Werte liegen die Prüfgrößen tabelliert (siehe etwa [Sac97]) vor.

Für unser kleines **Beispiel** ergibt sich beim 1%- Signifikanzniveau für D_α bei den Stichprobengrößen 5 und 6 der Wert 0,838. Da der größer ist als $3/10$, stammen damit beide Stichproben mit hoher Wahrscheinlichkeit aus der gleichen Grundgesamtheit¹³⁰.

¹³⁰ Eine der beiden Stichproben kann auch theoretisch sein. So könnte die zweite Reihe auch allgemein als die theoretisch zu erwartenden Werte für einen Würfel angesehen werden. Unser Beispiel sagt dann auch, daß ein Ergebnis wie in der ersten Stichprobe und das theoretische zweite Ergebnis mit hoher Wahrscheinlichkeit aus

Für große Datenmengen, das heißt für $n_1+n_2>35$, wobei n_1, n_2 die Stichprobengrößen sind, kann die Prüfgröße D_α durch eine Formel berechnet werden:

$$D_\alpha = K_\alpha ((n_1+n_2)/(n_1*n_2))^{1/2}$$

Dabei ist K_α eine nur von der Irrtumswahrscheinlichkeit α abhängige wohlbekannte Konstante, die ebenfalls [Sac97] entnommen werden kann.

16.4 Die Auswertung der Ergebnisse

Bei den oben geschilderten Experimenten ergaben sich die in Abbildung 40 und Abbildung 41 zu sehenden Verteilungsfunktionen f_1 und f_2 (in den Bezeichnungen aus der Erläuterung des Kolmogoroff-Smirnoff-Tests).

Die genauen Resultate können den Dateien `gesamtprognoszuf.txt` und `gesamtprognosdbwerte.txt` entnommen werden.

Für das erste Experiment mit zufälliger Erzeugung von Übernahmezeiten und ausschließlicher Betrachtung des reversiblen Sprachwandels zeigt in Abbildung 40 die violette Farbe den Verlauf der Simulation, die gelbe Farbe den Verlauf der Daten aus der Datenbank. In beiden Fällen sind bei den Verteilungsfunktionen nur die Sprungstellen markiert. Ich möchte hier die Diagramme noch einmal erläutern, obwohl ihre Bedeutung schon im Zusammenhang mit dem Kolmogoroff-Smirnoff-Test schon ausgiebig thematisiert worden ist. Jetzt ist die Erläuterung einfach plastischer.

Die Abbildungen sind so zu lesen, daß die Werte an der x-Achse die Verweildauer in Jahren bedeuten. Ein Wert $y = f(x)$ bedeutet bei Abbildung 40 den Anteil der vor dem Zeitpunkt x wieder verschwundenen Wörter an der Gesamtzahl der überhaupt wieder verschwundenen Wörter. Die Werte an der y-Achse bezeichnen diesen Anteil. Entsprechendes gilt für die Achsenbezeichnungen bei Abbildung 41, nur daß dort nicht nur die wieder verschwundenen Wörter betrachtet werden. Wie aus der Datenbank zu ersehen ist, sind (im Rahmen meiner Auswertungen ohne hapax legomena und mit zeichenorientierter Interpretation) 64 Ausdrucks-Inhaltsverbindungen wieder ausgeschieden. Also macht die Sprungfunktion Sprünge in Vielfachen von $1/64 = 0,015625$. Wir betrachten zum Beispiel den gelben Punkt, der am zweitweitesten rechts steht. Der Wert der x-Koordinate ist 358. Das bedeutet, daß es ein Wort gibt, welches nach 358 Jahre wieder aus dem Polnischen ausgeschieden ist. Nach 357 Jahren sind noch $62/64$ aller derartigen Wörter vorhanden, nach 359 Jahren noch $63/64$. Folglich macht die Kurve bei 358 einen Sprung um $1/64$ von $0,96875$ auf $0,984375$.

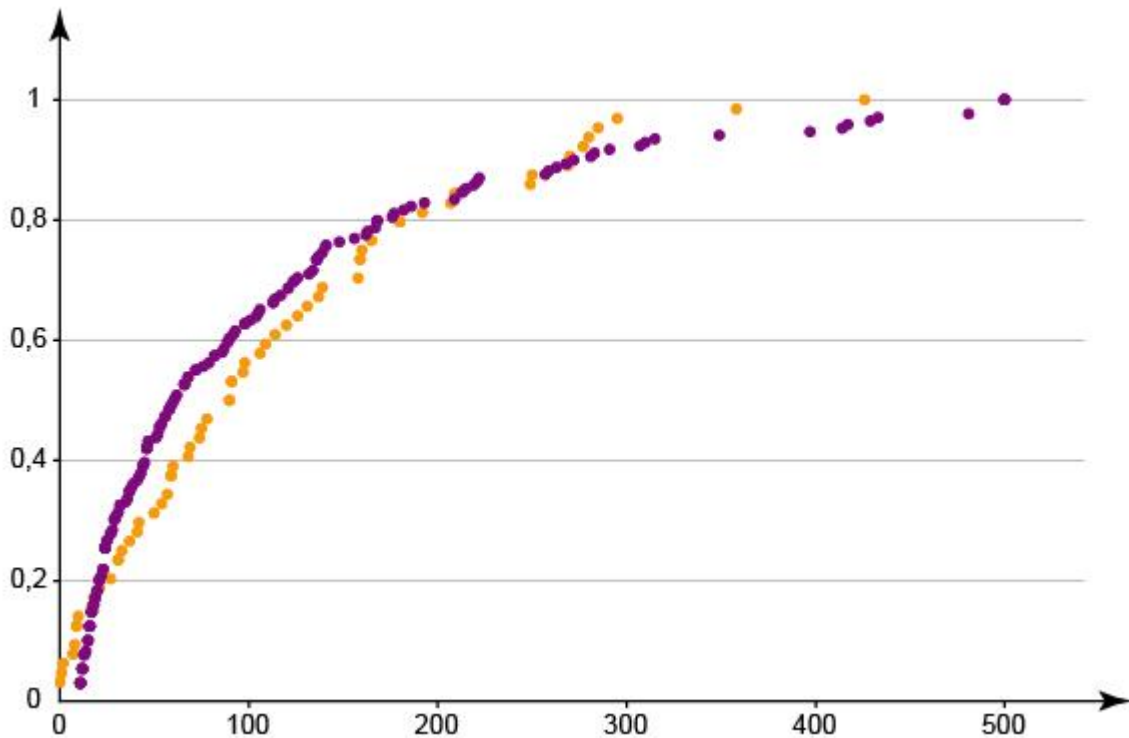


Abbildung 40: Verteilung beim reversiblen Sprachwandel

Für die Kolmogoroff-Smirnoff-Auswertung dieses ersten Experiments mit zufälliger Erzeugung von Übernahmezeiten und ausschließlicher Betrachtung des reversiblen Sprachwandels gilt

Prüfgröße: 0.239239 Gemessene maximale Abweichung: 0.145248

Das Maximum der Differenzen zwischen den beiden Funktionen liegt bei 56. Der Wert aus der Datenbank beträgt 0,328125, der Wert aus den Simulationen 0,47337278.

Für das zweite Experiment mit Übernahmezeiten aus der Datenbank und zusätzlicher Betrachtung des Sprachverhaltens bis zum Ende der Simulationen und bis zum Jahre 1950 zeigt in Abbildung 41 die blaßblaue Farbe den Verlauf der Simulation, die kräftige blaue Farbe den Verlauf der Daten aus der Datenbank. In beiden Fällen sind bei den Verteilungsfunktionen wieder nur die Sprungstellen markiert. Die Diagramme brauchen nicht noch einmal erläutert zu werden.

Für die Kolmogoroff-Smirnoff-Auswertung dieses zweiten Experiments mit Entnahme der Übernahmezeiten aus der Datenbank und zusätzlicher Betrachtung des irreversiblen Sprachwandels gilt

Prüfgröße: 0.107094 Gemessene maximale Abweichung: 0.097984

Die Zahlen sind hier absolut niedriger als beim ersten Experiment, da hier deutlich mehr Werte für die Auswertung relevant sind. Es sind ja auch deutlich mehr deutsche Wörter noch im Polnischen vorhanden als wieder ausgeschieden¹³¹.

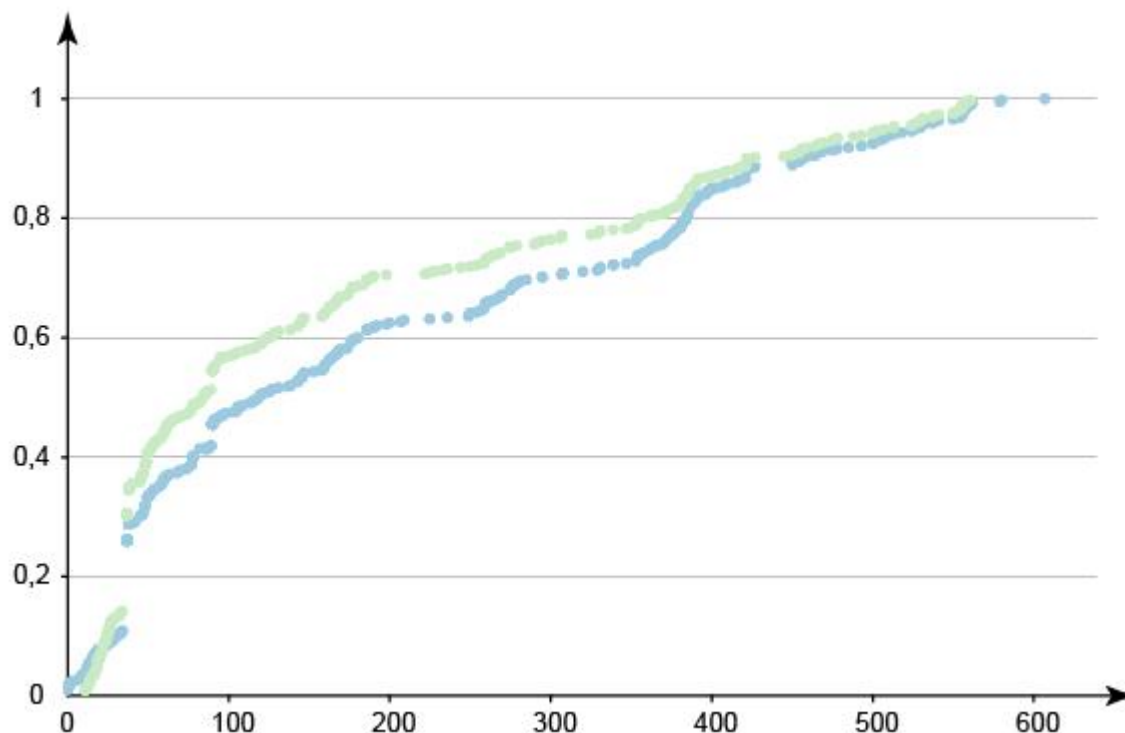


Abbildung 41: Verteilung beim Sprachwandel allgemein

In beiden Fällen muß die Nullhypothese nicht abgelehnt werden, und es darf angenommen werden, daß in beiden Fällen die beiden Stichproben aus derselben Grundgesamtheit stammen, mit anderen Worten:

Im Rahmen der statistischen Genauigkeit spiegelt das entwickelte Modell, also das Dynamische Sprachnetz, das Verhalten sprachlicher Neuerungen für den speziellen Fall der deutschen Lehnwörter im Polnischen angemessen wider.

16.5 Beigefügte Dateien und Programme

Hier soll jetzt kurz dokumentiert werden, welche Dateien und Programme der Arbeit beigefügt sind. Neben den in Unterkapitel 13.1 schon beschriebenen sind dies:

¹³¹ Die Formel für D_α in der Beschreibung des Kolmogoroff-Smirnoff-Tests zeigt deutlich, daß die Prüfgröße für Tests mit größeren Stichprobenumfängen sinkt. Das ist auch hier der Fall.

- Für die Eingabe in die Simulationen:

Datenbankauszug.txt	Der ursprüngliche Auszug aus der Datenbank, der für alle dort verzeichneten Wörter vier Angaben enthält: INHALTKURZ, HAPAX, INDEXNEU, LETZT,
uebernahmezeiten.txt	Datei mit Übernahmezeiten aus der DB,
dauernallewoerter.txt	Datei mit Aufenthaltsdauern aller Wörter aus der DB,
ausgeschieden.txt	Datei mit Aufenthaltsdauern der wieder verschwundenen Wörter aus der DB,
nochda.txt	Datei mit Aufenthaltsdauern der noch vorhandenen Wörter aus der DB.

Die Datei Datenbankauszug.txt habe ich direkt der Lehnwort-Datenbank (Dateien LWAUSDRU.DBF und LWINHALT.DBF) entnommen. Daraus wurden dann die übrigen vier Dateien herausgefiltert. Die Bedeutung der Angaben ist:

INHALTKURZ	Kurzbeschreibung der Wortbedeutung,
HAPAX	Indikator, ob ein hapax legomena vorliegt,
INDEXNEU	Angabe des Indexjahres,
LETZT	Angabe des Jahres des letzten Beleges.

Die Datei dauernallewoerter.txt ist die Vereinigung der Dateien ausgeschieden.txt und nochda.txt . Beim Übergang von Datenbankauszug.txt zu den anderen vier Dateien habe ich auch einige Inkonsistenzen aus den Werten bereinigt¹³². So wurden zum Beispiel Wörter gestrichen, die offensichtlich falsche Eingangsdaten hatten (Werte kleiner als 100) oder solche, bei denen als Erstbeleg ein Wert angegeben worden war, der über dem Letztbeleg lag.

- die eigentliche Simulation:

Das auf der CD vorhandene Programm ist in dem Sinne „multifunktional“, daß es neben den in Kapitel 13 beschriebenen Komponenten der eigentlichen Simulation auch alle zusätzlichen Statistik-Routinen¹³³ (Berechnung des Korrelationskoeffizienten, χ^2 - Test, Kolmogoroff-Smirnoff-Test) enthält, obwohl diese logisch nicht zusammengehören, weil die Dimensionsreduktion (mit dem χ^2 - Test) zuerst und erst danach unabhängig davon die eigentlichen Simulationen (mit dem Kolmogoroff-Smirnoff-Test) durchgeführt wurden. Aber ich wollte meine Leser nicht mit zu vielen verschiedenen Versionen des im wesentlichen gleichen Programmes verwirren.

Es befinden sich also auf der CD die in Unterkapitel 13.2 erläuterten Dateien

simulation.cpp	das eigentliche Simulationsprogramm,
zellen.cpp	Zugriffsprozeduren für Zellen,
listen.cpp	Zugriffsprozeduren für Listen,

¹³² Natürlich habe ich nicht viele gefunden.

¹³³ Ich hätte auch versuchen können, die Ausgaben meiner Simulationen an ein herkömmliches kommerzielles Statistik-Programm anzuschließen. Aber erstens hätte ich mir ein solches zunächst beschaffen müssen, und vor allem hätte ich zweitens die Ausgaben der Simulationen in ein Format konvertieren müssen, das von dem Statistik-Programm gelesen werden kann. Angesichts des dafür erforderlichen Aufwandes schien es mir trotz eines gewissen Aufwandes einfacher, die ohnehin bekannten statistischen Verfahren selber zu implementieren.

zellenlisten.h header-Datei mit Schnittstellen für Zellen und Listen.

Damit die Simulation korrekt statistisch ausgewertet werden kann, müssen die Dateien uebernahmezeiten.txt, ausgeschieden.txt, nochda.txt mit den relevanten Vergleichsdaten aus der Datenbank sich in demselben Verzeichnis befinden.

- für die Ergebnisse der Simulationen

auswertung.xls	EXCEL-Datei zur Vorbereitung des Kolmogoroff-Smirnoff-Tests. Hier werden die Werte erstellt, die in Abbildung 40 und Abbildung 41 zu sehen sind.
gesamtprognoszuf.txt	Datei mit allen Resultaten und Parameterwerten der Simulationen (Experiment 1) + Ergebnis des Kolmogoroff-Smirnoff-Tests,
gesamtprognosdbwerte.txt	Datei mit allen Resultaten und Parameterwerten der Simulationen (Experiment 2) + Ergebnis des Kolmogoroff-Smirnoff-Tests,
ergebprognoszuf.txt	Datei mit Aufenthaltsdauern aller Wörter (im Experiment 1 errechnet),
ergebprognosdbwerte.txt	Datei mit Aufenthaltsdauern aller Wörter (im Experiment 2 errechnet).

Die beiden letzten Dateien dienen als Eingabe für den Kolmogoroff-Smirnoff-Test.

- Zusätzliche Dokumentation vor der Dimensionsreduktion

Um die in Unterkapitel 14.4 aufgestellten Behauptungen über die Zeit vor der Dimensionreduktion zu dokumentieren, habe ich noch folgende Dateien eingefügt:

simulationvordimred.cpp	das eigentliche Simulationsprogramm (Version vor Dimensionsreduktion, ohne Kolmogoroff-Smirnoff-Test),
zellen.cpp	Zugriffsprozeduren für Zellen (dazu passend),
listen.cpp	Zugriffsprozeduren für Listen,
zellenlisten.h	header-Datei mit Schnittstellen für Zellen und Listen,
ergebnisvordimred.txt	Dokumentation aller Ergebnisse (einschließlich Korrelationskoeffizienten und χ^2 -Test).

- Dokumentation eines Beispiellaufes

Zur Dokumentation der Vorgänge bei einem einzelnen Beispiellauf (dem ersten, der von simulation.cpp mit zufälligen Übernahmezeiten geliefert wird,) dienen die Dateien

potenzvertgenerationen.txt	dokumentiert die Anzahl aller Knoten in den verschiedenen Zuständen Generation für Generation,
potenzverteilung.txt	dokumentiert die Grade aller Knoten am Ende der Simulation, führt daraus den Nachweis der Potenzverteilung,

graphzeichnung.txt Eingabe für das Zeichenprogramm nach [HTK+02].

Diese Dateien werden automatisch erzeugt, wenn `anzahlsimul=1`; gesetzt wird. Die Situation am Ende des Simulationslaufs wurde in Abbildung 39 illustriert.

17 Fazit und Ausblick

17.1 Was wurde erreicht?

In der Arbeit wurde ein Modell, das Dynamische Sprachnetz, entwickelt, mit dessen Hilfe eine noch recht neue Theorie, die der dynamischen Netze, für die Sprachwissenschaft nutzbar gemacht wird. Es ist gelungen, mit Hilfe dieses Modells das Verhalten einer Sprache oder besser gesagt bestimmter Elemente einer Sprache in einer ganz konkreten Fragestellung erstaunlich präzise zu reproduzieren. Ich möchte noch einmal auf die auch für mich überraschend genauen Ergebnisse verweisen, die in Abbildung 40 und Abbildung 41 dargestellt sind. Damit hat das Modell seine „Prognosefähigkeit“ zumindest in einem Bereich bewiesen und ist über das Stadium einer Spielwiese für Theoretiker hinausgewachsen¹³⁴.

Die Schritte der Arbeit von den wissenschaftstheoretischen Grundlagen und dem Literaturstudium über die Theorie und die Konstruktion des Modells bis hin zur Durchführung der Simulationen sind leicht nachvollziehbar; die Experimente sind jederzeit reproduzierbar. Genau das muß von einer akzeptablen wissenschaftlichen Methode erwartet werden.

Ich möchte auch noch einem Mißverständnis vorbeugen, das sich möglicherweise einschleichen könnte: Ich habe mein Modell mit Daten über die Übernahme von Lehnwörtern überprüft. Es handelt sich aber nicht um ein Modell für die Ausbreitung von Lehnwörtern. Insofern ginge ein Einwand mit dem Tenor, es gebe auch andere Formen von Entlehnungen, etwa in der Phraseologie oder morphologische Lehnprägungen, fehl. Solche anderen Formen sind in dem vorliegenden Modell für die allgemeine Ausbreitung sprachlicher Neuerungen implizit mit enthalten und könnten ebenfalls mit seiner Hilfe untersucht werden. Meine experimentelle Untersuchung hat sich aber eben auf einen anderen Aspekt bezogen.

Der propagierte Ansatz wurzelt tief in der Tradition der Wissenschaft, die man als „reduktionistisch“ bezeichnen könnte. Es geht bei der vorgeschlagenen Methode jedoch nicht darum, natürliche Sprache in ein mathematisch beschreibbares Prokustes-Bett zu pressen oder etwa wissenschaftlich abseitige Wege zu beschreiten; es geht auch nicht darum, zu behaupten, Sprache sei „nur“ das, was die Formeln und Simulationen liefern. Sprache ist ja wirklich mehr als das. Es geht lediglich darum, gewisse Facetten des Forschungsgegenstandes in genauerem Detail zu untersuchen, als es ohne die benutzten Formalismen möglich gewesen wäre, auch wenn solche Methoden im vorliegenden Kontext noch nicht allgemein akzeptiert sind. In diesem Sinne sehe ich meine Methode als angemessen an. Das Ziel ist in meinen Augen die Mithilfe beim Verständnis der Sprache als etwas Ganzem.

Eines der herausragenden Kennzeichen der vorliegenden Arbeit ist ihre Interdisziplinarität, die sich besonders im Einsatz von Mitteln und Methoden aus Mathematik, Physik, Informatik für Ziele der Sprachwissenschaft zeigt; und die in Kapitel 10 zitierte Literatur zeigt, daß sich auch schon mancher andere Autor mit interdisziplinären Ansätzen um die Sprachwissenschaft verdient gemacht hat. Man mag es als zweifelhaft ansehen, ob eine Arbeit unbedingt den Modeströmungen der Zeit folgen sollte; auf jeden Fall komme ich damit

¹³⁴ Wie bei jeder Theorie muß natürlich noch durch weitere Experimente abgegrenzt werden, wie weit diese Theorie tatsächlich trägt.

aber einer immer wieder an die Wissenschaft gestellten Forderung nach solchen Ansätzen nach. Die Forderung nach interdisziplinären Ansätzen heie ich bei dem hier behandelten Forschungsgegenstand auch selber gut, obwohl sie vom Zeitgeist forciert wird (und ich zeitgeistigen Strmungen reserviert gegenberstehe). Die Sprachwandelforschung, zumindest so wie ich sie dargestellt und verwendet habe, erfordert einfach, da Methoden aus vielen Ursprngen benutzt werden.

17.2 Was ist fr die Zukunft zu erwarten?

Zwei Fragenkomplexe liegen immer, wenn sich ein Modell in demselben Stadium befindet wie das Dynamische Sprachnetz, auf der Hand:

- Wie lt sich das Modell, so wie es schon da ist, noch weiter ausnutzen? Welche neuen Prognosen kann es noch zustzlich liefern, zustzlich zu bereits abgeleiteten und mglicherweise schon experimentell untermauerten Folgerungen?¹³⁵ Kann es alte Aussagen aus frheren Modellen verifizieren?
- Wie lt sich das Modell weiter verfeinern? Welche Ergnzungen knnten ntig werden? Welcher Preis mu dafr bezahlt werden?

17.2.1 Ausnutzung des Modells

Bezglich des ersten Fragenkomplexes liegt es auf der Hand, da weitere experimentelle Fragestellungen behandelt werden sollten. Fr zustzliche sttzende Befunde mssen natrlich dann auch entsprechende Datenmengen nutzbar gemacht werden. Weiterhin bin ich mir sicher, da die Theorie der dynamischen Graphen noch viele einschlgige theoretische Erkenntnisse und testbare Prognosen liefern wird. Dort gibt es schon heute viele mglicherweise anwendbare Ergebnisse, und das Forschungsfeld der Netzwerke ist ja auch immer noch in strmischer Entwicklung begriffen. Insbesondere drfte mit der vorliegenden Arbeit auch die Grundlage fr eine bertragung wohlausgearbeiteter mathematischer Ideen auch auf andere sprachwissenschaftliche Phnomene als das der Ausbreitung¹³⁶ von Lehnwrtern gelegt worden sein.

So scheinen mir die Resultate zum Zerfall des Netzes oder zur epidemischen Schwelle heie Anwrter fr Lieferanten interessanter, neuer sprachlicher Erkenntnisse, etwa das in Unterkapitel 9.1 zitierte Ergebnis zur Ausfalltoleranz, da Netzwerke mit der im Modell vorliegenden Pareto-Verteilung der Grade sehr widerstandsfhig gegen den zuflligen Ausfall von Knoten sind.

In Unterkapitel 9.2 habe ich ja auch schon eine testbare Hypothese formuliert:

¹³⁵ Eine konkrete Frage in unserem Fall knnte lauten: Kann das Modell etwas ber deutsche Lehnwrter im Polnischen in der Zukunft sagen?

¹³⁶ Man sollte hier zwei Dinge nicht verwechseln: Das vorgeschlagene Modell beschreibt ganz allgemein die bernahme und die Fortentwicklung von sprachlichen Neuerungen. Bei der in der Arbeit vorgeschlagenen experimentellen berprfung hingegen geht es nur um die Verweildauer von Wrtern. Genauere Analysemethoden lieen sich aus den Daten der Oldenburger Datenbank nicht ableiten.

Durchgängig erfolgreiche sprachliche Innovationen beginnen in der Regel bei den Individuen, die eine hohe Vernetzung (also besonders vielfältige Kommunikation) aufweisen.

Es sollen noch kurz beispielhaft einige Prognosemöglichkeiten vorgestellt werden, die auf der Verwendung von in der Arbeit vorgestellten theoretischen Erkenntnissen und deren mögliche Übertragung in die Linguistik beruhen. Zur Überprüfung müßten aber sicher in erheblichem Maße bisher noch unbekannte empirische Resultate herangezogen werden.

- Mit der gerade genannten Prognose eng verwandt ist eine Prognose über die Effekte von Sprachenpolitik: Falls eine solche gewünscht sein sollte, wird sie ohne den Einfluß von Multiplikatoren, besonders den der Massenmedien, also der Netzwerk-hubs, nicht machbar sein. Ein solcher Einfluß auf die Sprachentwicklung wird durch die Analyse von Netzwerken sicher gründlich analysierbar sein.
- Prognosen, die weniger auf der Hand liegen, können zum Beispiel einigen Anwendungen der Perkolationstheorie für Netze entnommen werden: Nach [AB02] ist die Größe der unabhängigen, nicht zusammenhängenden Komponenten eines Zufallsnetzes exponentiell (siehe Abschnitt 7.2.2.2) verteilt. Die entsprechende linguistische Übertragung läßt bei dem Fall der Existenz von mehreren verschiedenen Varianten vermuten, daß die Anzahl der Personen, die (überwiegend) diese Varianten benutzen, ebenfalls exponentiell verteilt ist¹³⁷.
- In der Untersuchung [KETS02] wird festgestellt, daß Änderungen in Systemen desto schneller vor sich gehen, je ähnlicher die handelnden Einheiten einander sind. Auch bei sprachlichen Änderungen müßte die Ähnlichkeit der Sprecher die Ausbreitung von Neuerungen befördern. Auch quantitative Aussagen werden in [KETS02] getroffen. Für eine Übertragung solcher Ergebnisse müßte natürlich der Begriff der "Ähnlichkeit" in einer der Sprachwissenschaft angemessenen Weise definiert werden. In der Originalarbeit ist er so umfassend, daß es auch möglich sein könnte, zu untersuchen, wie sich verschiedene Formen von Ähnlichkeit auswirken.
- Ein verwandter Effekt kann in [MV02] nachgewiesen werden. Ich habe in Unterkapitel 9.4 schon darauf hingewiesen, daß die Epidemische Schwelle, also ein Maß für die Widerstandskraft gegen die Ausbreitung von Krankheiten, bei Netzwerkanalysen praktisch sehr niedrig ist. Nach [MV02] gilt die Faustregel: Je homogener die Netzknoten desto niedriger die Epidemische Schwelle. Das müßte sich auch bei der Übertragung auf sprachliche Verhältnisse beobachten lassen.
- In [MB05] wird auf die Möglichkeit des Auftretens von Diskontinuitäten, hier zu verstehen als radikale Änderung des Systemverhaltens, hingewiesen¹³⁸. Es werden Bedingungen angegeben, wann solche Sprünge zu erwarten sind. Auch beim Verhalten des Systems "Sprache" könnten diese Sprünge unter analogen Bedingungen erwartet werden.

¹³⁷ Dieses Ergebnis wäre hoch signifikant, da die Exponentialverteilung in unserem Zusammenhang nur sehr selten auftaucht.

¹³⁸ Das bemerkenswerteste der von den Autoren untersuchten Beispiele ist das plötzliche Abbrechen von Beifall, simultan und ohne ersichtlichen Grund beim gesamten Publikum.

- Von einem ebenfalls Netze, allerdings komplexere als in der vorliegenden Arbeit, involvierenden Befund wird in [Swa01] berichtet: Der Autor betrachtet in diesem nur rudimentär Mathematik benutzenden Buch die Gesamtheit aller Sprachen als ein Netzwerk. Die Verbindung zwischen den Sprachen wird durch mehrsprachige Sprecher aufrechterhalten. Grundlage für eine auf einem Ökonomieprinzip beruhende Analyse der Verständigungsmöglichkeiten in verschiedenen multilingualen Gesellschaften (unter anderem Indien, dem postkolonialen Afrika und der EU) ist der *Q-Wert*, ein Maß für den Nutzen, den das Erlernen einer Fremdsprache mit sich bringt. Die Netzwerke, die de Swaan benötigt, sind komplexer als die in dieser Arbeit benutzten. Die Kanten müssen noch annotiert werden, indem ihnen die Sprache zugeordnet ist, in der die Verständigung erfolgt. Solche Annotationen sind in der Mathematik nicht unüblich; und es könnte durchaus Erweiterungen des Dynamischen Sprachnetzes geben, bei denen mit weiteren Methoden der Graphentheorie neue Ergebnisse in die von de Swaan vorgegebene Richtung erzielt werden können. Es gibt aber schon jetzt eine ganz starke Prognose, die auf dieser Verwendung von Netzwerken beruht: Wenn man die Definition¹³⁹ des Q-Wertes einer Sprache L betrachtet, so sieht man, daß dieser linear von der Zahl derjenigen Sprecher abhängt, die L schon als Fremdsprache gelernt haben. Die Zahl der neuen Schüler von L hängt wieder vom Q-Wert linear ab. Es liegt also ein klassischer Fall für den Matthäus-Effekt vor. Die Formeln aus Unterkapitel 8.5 sagen dann voraus, daß, ganz global gesehen, die Zahlen, von wie vielen Spechern eine Sprache als Fremdsprache gelernt wird, der Pareto-Verteilung gehorchen.

Ohne bisher noch tiefer in diese spezielle Materie eingestiegen zu sein, vermute ich insgesamt stark, daß die neueren Ereignisse der Theorie der dynamischen Netze auch weitere interessante Konsequenzen in der Linguistik zeigen dürften.

17.2.2 Erweiterungen des Modells

Was den zweiten Fragenkomplex angeht, so geht es um von den Methoden her mehr mathematisch ausgerichtete Fragen. Die dort erzielten Ergebnisse können unter Umständen auch in der Linguistik benutzt werden und sollten daher ebenfalls als Angebot an die Linguistik zur Erlangung neuer Erkenntnisse aufgefaßt werden.

Eine durch die Sprachwissenschaft motivierte Erweiterung, die sich aber vermutlich nicht ganz einfach ins Modell integrieren ließe, wäre die Modellierung, der in Unterkapitel 9.2 erwähnten verschiedenen Varianten des Ausscheidens eines Lehnwortes nach [HM01].

Eine mathematisch naheliegende Frage ist die nach weiteren Zuständen: Eine Möglichkeit ist aus der Epidemiologie wohlbekannt: ein Zustand, der dort Immunität nach einer Infektion bedeutet. Mit Hilfe eines jetzt anders zu modellierenden Immunitätszustandes, nicht nur nach einer Übernahme der Neuerungen, könnte der Einfluß eines Kerns von Verweigerern oder die Wirkung einer gezielten Sprachpflege untersucht werden. Da die Knoten Menschen oder Gruppen von Menschen symbolisieren, gibt es sicher viele Möglichkeiten, wie durch neue Zustände neue Eigenschaften dieser Menschen ins Spiel

¹³⁹ Deren genaue Form kann in [Swa01], p. 34, gefunden werden. An dieser Stelle spielt die exakte Formel aber keine Rolle.

gebracht werden könnten. Da sehe ich aber eher die Gefahr, daß die Modelle irgendwann zu kompliziert werden. Es ist aber eine wohlbekannte Tatsache, daß zu komplexe Modelle irgendwann zwar noch Phänomene beschreiben können, und insofern immer noch solide Wissenschaft bieten, aber ihre Erklärungsfähigkeit verlieren. Die Frage nach dem Warum kann dann nicht mehr seriös beantwortet werden.

Ein anderes bisheriges Dogma ist das der *Gleichheit der handelnden Individuen*: Im Dynamischen Sprachnetz unterscheiden sich die Individuen nur durch ihre Zustände S, I oder R, sowie die Zahl ihrer sozialen Kontakte, nicht aber durch die Form der Zustandsüberführungsregeln, also durch ihre Haltung zu sprachlichen Neuerungen. Ich habe insbesondere darauf verzichtet, zwischen mehr oder weniger konservativen Sprechern zu unterscheiden.

In Unterkapitel 11.2 habe ich schon darauf hingewiesen, daß ich davon ausgehe, daß ein Lehnwort, sofern es einmal eingedrungen ist, von allen Sprechern akzeptiert wird wie ein eigenes, daß es also keinen Bonus für Eigenwörter gibt. Diese *Gleichbehandlung der Zustände* könnte ebenfalls relativiert werden.

Eine andere Erweiterung des Modells ist ebenfalls aus der Graphentheorie lange bekannt: Man könnte die Kanten in den Graphen mit Gewichten versehen und dadurch die Intensität der Kommunikation widerspiegeln.

Wieder eine andere Möglichkeit ist die Einführung eines Gedächtnisses in den Knoten. Daß diese Veränderung aus der Sicht der theoretischen Informatik naheliegt, das habe ich ja schon in Unterkapitel 11.1 ausgeführt. Hier sehe ich aber keine unmittelbaren Anwendungsgebiete für die Sprachwissenschaft, wo ein Gedächtnis wirklich nötig sein sollte.

Ich möchte abschließend noch drei weiter entfernt liegende Ansätze erwähnen:

1. Zunächst soll noch einmal an die in Unterkapitel 9.2 erwähnten Überlegungen von Hentschel in [Hen95] erinnert werden. Der Autor kommt dort zu einem sehr überraschenden Befund: Obwohl es auf Grund der vielfältigen zufällig auftretenden Faktoren¹⁴⁰ in der Geschichte überhaupt nicht zu erwarten ist, daß die Übernahme von Lehnwörtern allein durch Methoden der mathematischen Epidemiologie zu erklären ist, so läßt sich doch feststellen, und das ist das eigentlich Überraschende, daß *der jeweilige Bestand* an deutschen Lehnwörtern im Polnischen in seinem zeitlichen Verlauf sich ziemlich genau durch die aus vergleichbaren mathematischen oder epidemiologischen Modellen, zum Beispiel dem Piotrovskij-Gesetz, bekannten S-förmigen Kurven beschreiben läßt. Für diesen Befund wird in [Hen95] allerdings keine Erklärung angeboten. Es könnte ein lohnender Versuch sein, den Hentschelschen und meinen Ansatz zu verbinden, also eine mehrfache Infektion im Dynamischen Sprachnetz zu untersuchen. Dann kann das beschriebene Phänomen möglicherweise durch wissenschaftliche Untersuchungen im Dynamischen Sprachnetz *erklärt* und nicht nur *beschrieben* werden (oder sich möglicherweise als Zufall herausstellen). Wie ich oben aber schon erwähnt habe, würde eine solche Studie zu mehrfachen Epidemien zumindest umfangreiche Erweiterungen meines Modells nötig machen und möglicherweise nicht nur sprachlich sondern auch mathematisch in Neuland führen. Eine solche Untersuchung stünde insbesondere vor dem Problem, meine Auffassung von vielfachen Epidemien durch einzelne Wörter mit der zweifachen Epidemie im Hentschelschen Sinne (bis 1550 und im 19. Jahrhundert), die man in einem verallgemeinerten Ansatz auf der Grundlage von Netzen vielleicht als „Metaepidemien“ bezeichnen könnte, unter einen Hut zu bringen.

¹⁴⁰ So führen die in Kapitel 15 kurz geschilderten Ereignisse zu einer doppelten S-Kurve bei der Verteilungsfunktion, die sich durch die Zahl der Worteingänge in Abhängigkeit von der Zeit ergibt. Der erste steile Anstieg liegt im Mittelalter und der Renaissance und der zweite zur Zeit der Polnischen Teilungen.

Einen Schritt jedoch in die Richtung, die in [Hen95] angemahnt worden ist, habe ich mit dieser Arbeit getan: Hentschel unterstreicht am Ende seines Artikels den noch beträchtlichen Forschungsbedarf für das Ausscheiden eines Lehnwortes, also den Teil des zeitlichen Verlaufs, wo die beschreibenden Kurven unter Umständen nicht mehr S-förmig sind. Mein Ansatz geht nicht nur in die dadurch angesprochene Richtung des reversiblen Sprachwandels sondern ermöglicht auch die Modellierung, auf längere Sicht also vielleicht eine Erklärung, des Phänomens, daß sich eine Neuerung nur unvollständig durchsetzt.

2. In Kapitel 5 habe ich auf die Tatsache hingewiesen, daß bei den Überlegungen zum Piotrovskij-Gesetz immer wieder aus rein mathematischen Gründen Konstanten auftreten, deren Wert und unter Umständen auch deren Bedeutung sich nicht aus sprachlicher Sicht erklären läßt. Das einfachste derartige Beispiel lieferte die dort mit (*) bezeichnete Gleichung, die in etwa besagt, daß (zumindest initial) die Ausbreitungsgeschwindigkeit eines Lehnwortes proportional zu dem schon erreichten Bestand ist. Doch wie der dadurch ins Spiel kommende Proportionalitätsfaktor nun tatsächlich zu bestimmen ist, wird nicht gesagt. Dabei ist diese Konstante im Vergleich zu den anderen in Kapitel 5 vorgestellten Konstanten gedanklich noch recht leicht nachzuvollziehen. Ich bin fest davon überzeugt, daß die Werte dieser Konstanten im Piotrovskij-Gesetz zum Teil durch lokale Einflüsse begründet sind, was die Möglichkeit nahelegt, daß hier die Struktur des Netzes eine wichtige Rolle spielt, aber ich habe keine Ahnung, wie eine solche Erklärung genau aussehen sollte. Immerhin habe ich ja in Unterkapitel 6.4 mehrere Beispiele dafür angegeben, wie die Ausbreitungsgeschwindigkeit durch die Struktur des Netzes beeinflußt wird. Vielleicht gibt es also Möglichkeiten, mit dem Dynamischen Sprachnetz oder Erweiterungen davon, weitere Schritte auf dem Weg zu einer Antwort auf diese für so manchen Sprachwissenschaftler interessante Frage nach Wert und Bedeutung der Konstanten zu gehen.

3. In Kapitel 15 habe ich schon erwähnt, daß im Modell nur die reine Ausbreitung des übernommenen Wortes untersucht wird. Dazu wurden die Wörter als monoseme Einheiten aus Ausdruck und Inhalt betrachtet, die als unabhängige Homonyme behandelt wurden. Insofern wurden Polysemie und Homonymie ignoriert, was wiederum heißt, daß auch nicht zwischen ihnen differenziert wurde. Eine Erweiterung könnte dann darin bestehen, daß das Wort, dort wo es notwendig ist, eben auch auch als polyseme Einheit behandelt würde, (und womöglich auch Homonymie modelliert würde). Das wäre gleichbedeutend damit, daß auch die Herkunft der Polysemie berücksichtigt würde: sowohl übernommene als auch eigenständig entwickelte. Sollte noch eine solche semantische Auffächerung eines einzelnen Wortes einbezogen, also inhaltsorientiert gearbeitet werden, so wäre das sicher aus linguistischer Sicht ein lohnender Forschungsgegenstand, aber auch hier würde das Modell extrem kompliziert. Bisher fehlen dazu alle Ansätze zu einer geeigneten mathematischen Theorie auf der Grundlage von Netzwerken. Eine solche Theorie dynamischer Graphen mit sich aufspaltenden Knoten wäre sicher nicht nur für die Mathematik und die Linguistik interessant sondern auch für andere Anwendungen.

Im Verlaufe dieser Arbeit habe ich viele Modelle besonders auch deshalb aufgeführt, um die Vielfalt der möglichen Verfeinerungen des Modells aufzuzeigen, beispielsweise AEMR und RHL in Unterkapitel 9.6 Aber mit solchen Ansätzen beginne ich damit, mich der Spekulation zu nähern. Diese Ansätze dienen bei mir nur dazu, den einen Punkt zu unterstreichen:

Da ist noch viel Raum für zukünftige Forschungen!

Literatur

[ACL01] W.Aiello, F.Chung, L.Lu: Random evolution in massive graphs. 42th Annual IEEE Symposium on Foundations of Computer Science, 510-519 (2001).

[AJB99] R.Albert, H.Jeong, A.-L. Barabási: Diameter of the World-Wide Web. Nature 401, 130-131 (1999).

[AJB00] R.Albert, H.Jeong, A.-L. Barabási: Error and attack tolerance of complex networks. Nature 406, 378-382 (2000).

[AB00] R.Albert, A.-L. Barabási: Topology of evolving networks: Local events and universality. Phys. Rev. Lett. 85 5234; cond-math / 0005085.

[AB02] R.Albert, A.-L. Barabási: Statistichal Mechanics of Complex Networks. Rev. Of Modern Physics 74, 47-101 (2002); cond-math / 0106096.

[ABRS83] G.Altmann, H.v.Buttlar, W.Rott, U.Strauß: A law of Change in Language. pp.104-115 in [Brai83].

[Alt83] G.Altmann: Das Piotrovski-Gesetz und seine Verallgemeinerungen. pp.59-90 in [BK83].

[Alt85] G.Altmann: On the Dynamic Approach to Language. pp.181-189 in [Bal85].

[Alt85a] G.Altmann: Die Entstehung diatopischer Varianten. Zeitschrift für Sprachwissenschaft. Vandenhoeck&Ruprecht, Göttingen 1985.

[Alt02] G.Altmann: Einführung in die quantitative Lexikologie. Göttinger linguistische Abhandlungen, Bd. 5, Peukert & Gutschmidt, Göttingen 2002.

[AH78] K.Appel, W.Haken: The Four-Color Problem. In: Mathematics Today, ed. L.A.Stehen, pp. 153-190. Springer 1978.

[AO94] K.R.Apt, E.-R.Olderog: Programmverifikation. Springer, Berlin, Heidelberg, New York 1994.

[AC83] M.V.Arapov, M.M.Cherc: Mathematische Methoden in der historischen Linguistik. Quantitative linguistics, Bd.17. Studienverlag Brockmeyer, Bochum 1983.

[Ara83] M.V.Arapov: Word Replacement Rates for Standard Russian. pp. 50-61 in [Brai83a].

[Bai73] Ch.J.N.Bailey: Variation and Linguistic Theory. Arlington, Center for Applied Linguistics, 1973.

[BJ06] M.E.Bales, S.B.Johnson: Graph theoretic modeling of large-scale semantic networks. *J. of Biomed. Inf.* 39(4), 451-464 (2006).

[Bal85] T.T.Ballmer (ed.): *Linguistic Dynamics*. W.deGruyter, Berlin, New York 1985.

[Ban94] R.B.Banks: *Growth and Diffusion Phenomena*. Texts in Applied Mathematics, Bd.14. Springer, Berlin, Heidelberg, New York+ 1994.

[BA99] A.-L.Barabási, R.Albert: Emergence of scaling in random networks. *Science* 286, 509-512 (1999).

[BAJ99] A.-L.Barabási, R.Albert, H.Jeong: Mean-field theory for scale-free random networks. *Physica A* 272, 173-187 (1999).

[Bar03] A.-L.Barabási: *Linked. How everything is Connected to Everything else and what it means for Business, Science, and for Everyday Life*. Penguin, New York, London, Camberwell+ 2003.

[BLT05] M.Bartolozzi, D.B.Leinweber, A.W.Thomas: Stochastic Opinion Formation in Scale-Free Networks. APS preprint; cond-mat / 0504168.

[BDM03] H.Berger, M.Dittenbach, D.Merkl: Activation on the move: Querying tourism information via spreading activation. pp. 474-483. In: V.Marik, W.Retschitzegger, O. Stepankova (eds.): *Proceedings of the 14th International Conference and Workshop on Database and Expert Systems Applications (DEXA 03)*. Prag, Springer 2003.

[BK83] K.-H.Best, J.Kohlhase (eds.): *Exakte Sprachwandelforschung*. edition herodot, Göttingen 1983.

[BA86] K.-H.Best, G.Altmann: Untersuchungen zur Gesetzmäßigkeit von Entlehnungsprozessen im Deutschen. *Folia Linguistica Historica* 7, 31-41.

[BBA90] K.-H.Best, E.Beöthy, G.Altmann: Ein methodischer Beitrag zum Piotrowski-Gesetz. pp. 115-124. In: Hammerl, Rolf (ed.): *Glottometrika* 12, Brockmeyer, Bochum 1990.

[Best03] K.-H.Best: *Quantitative Linguistik: eine Annäherung*. (2. Aufl.) Göttinger linguistische Abhandlungen, Bd. 3, Peukert & Gutschmidt, Göttingen 2003.

[Best03a] K.-H.Best: Spracherwerb, Sprachwandel und Wortschatzwachstum in Texten. Zur Reichweite des Piotrowski-Gesetzes. *Glottometrics* 6, 9-34.

[Bib] Die Bibel. Einheitsübersetzung 1980.

[BF99] N.Boccara, H.Fuks: Modeling diffusion of informations with probabilistic cellular automata. pp. 263-278 in [DM99].

[BPV03] M.Boguna, R.Pastor-Satorras, A.Vespignani: Epidemic spreading in complex networks with degree correlations. cond-mat / 0301149.

- [BoT97] C.Borchard-Tuch: Computer-Systeme – Ebenbilder der Natur? Vieweg, Braunschweig, Wiesbaden 1997.
- [Bor03] S.Bordag: Sentence co-occurrences as small-world graphs: A solution to automatic lexical disambiguation. pp. 329-332. In: A.Gelbukh (ed.): Computational Linguistics and Intelligent Text Processing. Proceedings. Springer, Mexico City 2003.
- [Bos94] H.Bossel: Modellbildung und Simulation. Vieweg, Braunschweig, Wiesbaden 1994.
- [Brai83a] B.Brainerd (ed.): Historical Linguistics. Quantitative linguistics, Bd.18. Studienverlag Brockmeyer, Bochum 1983.
- [Brai83b] B.Brainerd: A Stochastic Model for Language Change. pp. 25-49 in [Brai83a].
- [BE05] U.Brandes, Th.Erlebach (eds.): Network Analysis – Methodological Foundations. Springer Lecture Notes in Computer Science 3418. Springer, Berlin, Heidelberg 2005.
- [Bre01] Joan Bresnan: Lexical-Functional Syntax. Blackwell textbooks in linguistics, Bd.16. Blackwell, Malden (Mass.) 2001.
- [Buc03] M.Buchanan: Nexus: Small Worlds and the Groundbreaking Science of Networks. Verlag W.W.Norton, Neuauflage 2003.
- [CSCC05] A.Capocci, V.D.P.Servedio, G.Caldarelli, F.Colaioni: Detecting communities in large networks. Physica A352, 669-676, (2005).
- [Cho57] N.Chomsky: Syntactic structures. Mouton, New York 1957.
- [Cho02] N.Chomsky: Interview on minimalism. In: A.Belletti, L.Rizzi (eds.): On Nature and Language. Cambridge University Press, New York.
- [CP02] A.Congelosi, D.Parisi: Simulating the Evolution of Language. Springer, London, Berlin, Heidelberg 2002.
- [DBBL06] L.Dall'Asta, A.Baronchelli, A.Barrat, V.Loreto: Non-equilibrium dynamics of language games on complex networks. Phys. Rev. E 74: 036105.
- [Dal02] Mary Dalrymple: Lexical Functional Grammar [Nachdr.]. Syntax and Semantics, Bd.34. Academic Press, San Diego 2002.
- [DH86] P.J.Davis, R.Hersh: Erfahrung Mathematik. Birkhäuser, Basel, Boston, Stuttgart 1986.
- [Del99] M.Delorme: An introduction to cellular automata. pp. 5-50 in [DM99].
- [DM99] M.Delorme, J.Mazoyer: Cellular Automata – A Parallel Model. Kluwer, Dordrecht, Boston, London 1999.
- [Dij70] E.W.Dijkstra: Notes On Structured Programming, 1970, zu finden in <http://www.cs.utexas.edu/users/EWD/ewd02xx/EWD249.PDF>

- [DM00] S.N.Dorogovtsev, J.F.F.Mendes: Evolution of reference networks with aging. Phys. Rev. E 62, 1842; cond-mat / 0001419.
- [DM01] S.N.Dorogovtsev, J.F.F.Mendes: Language as an evolving Word Web. Proc. R. Soc. London B 268, 2603; cond-mat / 0105093.
- [DM03] S.N.Dorogovtsev, J.F.F.Mendes: Evolution of Networks. Oxford UP, Oxford 2003.
- [Dor58] W.Doroszewski: Słownik języka polskiego [Wörterbuch der polnischen Sprache], Warschau 1958-1969.
- [Emb86] S.M.Embleton: Statistics in historical linguistics. Quantitative linguistics, Bd.30. Studienverlag Brockmeyer, Bochum 1986.
- [ER59] P.Erdős, A.Rényi: On Random Graphs. Publ. Math. Debrecen 6, 290.
- [ER60] P.Erdős, A.Rényi: On the Evolution of Random Graphs. Publ. Math. Inst. Hung. Acad. Sci. 5, 17.
- [Fed89] J.Feder: Fractals. 4.Aufl. Plenum Press, New York, London (1989).
- [FS01] R.Ferrer y Cancho, R.V.Solé: The small world of human language. Proc. R. Soc. London B 268, 2261-2265 (2001).
- [FS03] R.Ferrer y Cancho, R.V.Solé: Least effort and the origins of scaling in human language. Proc. Nat. Ac. Sci. 100.3, 788-791 (2003).
- [Fer04] R.Ferrer y Cancho: Language: universals, principles and origins. Ph.D. Thesis Universitat Politecnica de Catalunya, Barcelona (2004).
- [FSK04] R.Ferrer y Cancho, R.V.Solé, R.Köhler: Patterns in syntactic dependency networks. Phys. Rev. E 69, 051915, 2004.
- [Fer05] R.Ferrer y Cancho: The structure of syntactic dependency networks: insights from recent advances in network theory. pp. 60-75 in: "The problems of quantitative linguistics", G.Altmann, V.Levickij, V.Perebyinis (eds.). Chernivtsi: Ruta 2005.
- [FBR05] R.Ferrer y Cancho, R.Bollobás, O.Riordan: The consequences of Zipf's law for syntax and symbolic reference. Proc. R. Soc. Lond. B 272, 561-565 (2005).
- [Fer06] R.Ferrer y Cancho: On the universality of Zipf's law for word frequencies. pp. 131-140 in: "Exact methods in the study of language and text. In honor of Gabriel Altmann", P.Grzybek, R.Köhler (eds.). de Gruyter, Berlin 2006.
- [Gar91] M.Garzon: Models of Massive Parallelism. Springer, Berlin, Heidelberg, New York 1991.
- [GN96] R.J.Gaylord, K.Nishidate: Modeling Nature. Berlin, Heidelberg, New York 1996.

- [GS95] M.Gerhardt, H.Schuster: Das digitale Universum. Vieweg, Braunschweig 1995.
- [GDM03] A.V.Goltsev, S.N.Dorogovtsev, J.F.F.Mendes: Critical Phenomena in networks. Phys. Rev. E 67, 026123; cond-mat / 0204596
- [GTK+04] Gong, Tao, Jinyun Ke, J.W.Minett, W.S.-Y.Wang: A computational framework to simulate the coevolution of language and social structure. Artificial Life 9, Boston (2004).
- [Gui82] H.Guiter (ed.): Studies on Zipf's law. Quantitative linguistics, Bd.16. Studienverlag Brockmeyer, Bochum 1982.
- [Hak81] H.Haken: Erfolgsgeheimnisse der Natur. Stuttgart 1981.
- [HR94] B.Hannon, M.Ruth: Dynamic Modeling. Springer, Berlin, Heidelberg, New York 1994.
- [Har06] K.Hartmann (ed.): Festschrift anlässlich der Emeritierung von Ivan Seder, ISBN 3-8322-5401-3, Aachen 2006.
- [Har95] A.C.Harvey: Zeitreihenmodelle (2.Aufl.). Oldenbourg, München, Wien 1995.
- [Hen95] G.Hentschel: Zur „Seuche“ des deutschen Lehnwortes im Polnischen und zu den „Selbstheilungskräften“ dagegen. In: Bochnakowa, A. & . St. Widłak (eds.), Munus Amicitiae. Studia linguistica in honorem Witoldi Mańczak Septuagenarii. Kraków, 69-78.
- [HM01] G.Hentschel, T.Menzel: Zum Ausscheiden alter deutscher Lehnwörter aus dem Wortschatz der polnischen Literatursprache im 19. und Anfang des 20. Jahrhunderts. In: M. Gierlak, M. Klentak-Zabłocka, L. Żyliński (eds.): Im Wechselspiel der Kulturen. Festschrift für Professor Karol Sauerland. Toruń, 185-208.
- [Hen01] G.Hentschel: Das deutsche Lehnwort in der Geschichte der polnischen Sprache: Quantitäten in chronologisch-qualitativer Perspektive. In: K. Sauerland (ed.): Kulturtransfer Polen-Deutschland. Wechselbeziehungen in Sprache, Kultur und Gesellschaft. B. 2. Bonn, 153-169.
- [Hel01] H.Helbig: Die semantische Struktur natürlicher Sprache. Springer, Berlin, Heidelberg, New York 2001.
- [Her66] G.Herdan: The Advanced Theory of Language as Choice and Chance. Springer, Berlin, Heidelberg, New York 1966.
- [HTK+02] S. Heymann, K. Tham, A. Kilian, G. Wegner, P. Rieger, D. Merkel and J. C. Freytag: Viator - A Tool Family for Graphical Networking and Data View Creation, Proc. 28th VLDB Conference, Hong Kong, China, 2002
- [Hof73] L.Hoffmann (ed.): Sprachstatistik. Fink, München, Salzburg 1973.
- [HK02] P.Holme, B.J.Kim: Phys. Rev. E 65, 026107 (2002).

- [HMU01] J.E.Hopcroft, R.Motwani, J.D.Ullman: Einführung in die Automatentheorie, Formale Sprachen, Komplexitätstheorie (2.Aufl.). Addison-Wesley, München 2001.
- [Hud84] R. Hudson: Word Grammar. Blackwell, Oxford 1984.
- [Hud06] R. Hudson: Word Grammar. In: K. Brown (ed.) Encyclopedia of Language and Linguistics, (2.Aufl.), ISBN: 0-08-044299-4, Elsevier.
- [JTK+03] A. de Jesus Holanda, I.Torres Pisa, O.Kinouchi, A.Souto Martinez, E.E.Seron Ruiz: Thesaurus as a complex network. cond-mat / 0312586v1, eingereicht für Physica A.
- [Jin04] K.Jinyun: Selforganization and language evolution: system, population and individual. Ph.D.Thesis, City University of Honkong 2004.
- [JM04] M.Jünger, P.Mutzel (eds.): Graph Drawing Software. Springer, Berlin, Heidelberg, New York 2004.
- [Jun91] D.Jungnickel: Graphs, Networks and Algorithms. Springer, Berlin, Heidelberg, New York 1991.
- [KW01] M.Kaufmann, D.Wagner (eds.): Drawing Graphs. LNCS2025. Springer, Berlin, Heidelberg, New York 2001.
- [Kam59] E.Kamke: Differentialgleichungen. Lösungen und Lösungsmethoden. Akad. Verl.ges. Leipzig 1959.
- [KY06] J.Ke, Y.Yao: Analysing language development from a network approach. <http://arxiv.org/ftp/cs/papers/0601/0601005.pdf> (2006)
- [Ke07] J.Ke: Complex networks and human language. arXiv:cs/0701135v1 (2007)
- [Kel85] R.Keller: Towards a Theory of linguistic Change. pp.211-237 in [Bal85].
- [KML+02] O.Kinouchi, A.S.Martinez, G.F.Lima, G.M.Lourenco, S.Risau-Gusman: Deterministic walks in random networks: an application to thesaurus graphs. Physica A 315, 665-676, 2002.
- [KKR99] J.Kleinberg, S.R.Kumar, P.Raghvan et.al.: The web as a graph: Measurements, models and methods. Proceedings of the International Conference on Combinatorics and Computing 1999.
- [KE01] K.Klemm, V.M.Eguíluz: Highly clustered scale-free networks. Phys. Rev. E 65, 036123; cond-mat / 0107606.
- [KETS02] K.Klemm, V.M.Eguíluz, R.Toral, M. San Miguel: Global culture: A noise induced transition in finite systems. cond-mat / 0205188.
- [Klü95] J.Klüver: Soziologie als Computerexperiment. Vieweg – Serie Wissenschaftstheorie, Wissenschaft und Philosophie, Bd.41, Braunschweig, Wiesbaden 1995.

- [Knu93] D.E.Knuth: The Stanford Graph Base: A Platform for Combinatorial Computing. Addison-Wesley, Reading 1993.
- [Köh86] R.Köhler: Zur linguistischen Synergetik. Struktur und Dynamik der Lexik. Quantitative linguistics, Bd.31. Studienverlag Brockmeyer, Bochum 1986.
- [KR91] R.Köhler, B.B.Rieger (eds.): Contributions to Quantitative Linguistics. Proceedings of the First International Conference on Quantitative Linguistics QUALICO, Trier 1991. Kluwer, Dordrecht, Boston, London.
- [Köh91] R.Köhler: Synergetic Linguistics. pp. 41-51 in [KR91].
- [Kör04] H.Körner: Zur Entwicklung des deutschen (Lehn)Wortschatzes. Glottometrics 7, 25-49.
- [KN99] U.Krause, T.Nesemann: Differenzengleichungen und diskrete dynamische Systeme. Teubner, Stuttgart, Leipzig 1999.
- [Lab94] W.Labov: Principles of Linguistic Change (I): Internal Factors. Blackwell, Oxford (1994).
- [Lab01] W.Labov: Principles of Linguistic Change (II): Social Factors. Blackwell, Oxford (2001).
- [Lan88] C.Langton (ed.): Artificial Life. Menlo Park (Calif.) 1988.
- [Lass80] R.Lass: On Explaining Language Change. Cambridge University Press, Cambridge, London, New York + 1980.
- [Lass97] R.Lass: Historical Linguistics and Language Change. Cambridge University Press, Cambridge 1997.
- [LK91] A.M.Law, W.D.Kelton: Simulation Modeling & Analysis. McGraw-Hill, New York, St. Louis + 1991.
- [Leo98] E.Leopold: Stochastische Modellierung lexikalischer Evolutionsprozesse. Verlag Dr. Kovac, Hamburg 1998.
- [Lot26] A.J.Lotka: The frequency distribution of scientific productivity. J. Washington Acad. Sci.. 16, 317.
- [Man67] B.B.Mandelbrot: How long is the coast of Britain? Science 155, 636-638.
- [MG00] N.Mathias, V.Gopal: Small-worlds: How and why? Cond-mat / 0002076.
- [MCL05] M.Medeiros Soares, G.Corso, S.L.Lucena: The network of syllables in Portuguese. Physica A355, 678-684, (2005).
- [MB05] Q.Michard, J.-P.Bouchaud: Theory of collective opinion shifts: from smooth trends to abrupt swings. New Scientist, 7.5.2005, 15-25; auch: Cond-mat / 0504079.

- [MSI+02] R.Milo, S.Shen-Orr, S.Itzkovitz, N.Kashtan, D.Chklovskij, U.Alon: Network motifs: Simple building blocks of complex networks. *Science* 298, 824-827.
- [MKI+04] R.Milo, N.Kashtan, S.Itzkovitz, M.E.J.Newman, U.Alon: On the uniform generation of random graphs with prescribed degree sequences. *Cond-mat / 0312028*, eingereicht für *Phys. Rev. E*.
- [MIK+04] R.Milo, S.Itzkovitz, N.Kashtan, R.Levitt, S.Shen-Orr, I.Ayzenshtat, M.Sheffer, U.Alon: Superfamilies of evolved and designed networks. *Science* 303, 1538-1542.
- [Mil67] S.Milgram: The small world problem. *Psychologie Today* 2, 60-67.
- [MV02] Y.Moreno, A.Vazquez: Disease Spreading in Structured Scale-Free Networks. *Cond-mat / 02120362 v1*, eingereicht für *Phys. Rev. E*.
- [MMLD02] A.E.Motter, A.P.S.de Moura, Y.-C.Lai, P.Dasgupta: Topology fo the conceptual network. *Phys. Rev. E* 65, 065162 (R), 2004.
- [Mun01] T.Munzner: H3Viewer. <http://grapuics.stanford.edu/~munzner/h3/>.
- [NSW01] M.E.J.Newman, S.H.Strogatz, D.J.Watts: Random graphs with arbitrary degree distribution and their applications. *Phys. Rev. E* 64, 026118 (2001); *cond-mat / 0107235*.
- [New02] M.E.J.Newman: The spread of epidemic disease on networks. *Phys. Rev. E* 66, 016128 (2002); *cond-mat / 0205009*.
- [New03] M.E.J.Newman: The Structure and Function of Complex Networks. *SIAM Review* 45, 167-256 (2003); *cond-mat / 0303516*.
- [New03a] M.E.J.Newman: Properties of highly clustered networks. *cond-mat / 0303183*.
- [NP03] M.E.J.Newman, J.Park: Why social nets are different from other types of networks. *Phys. Rev. E* 68, 036122 (2003); *cond-mat / 0305612*.
- [NBW06] M.E.J.Newman, A.-L.Barabasi, D.J.Watts: *The Structure and Dynamics of Networks*. B&T, Princeton Studies in Complexity, 2006.
- [OS65] Ch.E.Osgood, Th.Sebeok: *Psycholinguistics*. Bloomington, Indiana UP (1965).
- [PV00] R.Pastor-Satorras, A.Vespignani: Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200 (2000); *cond-mat / 0010317*.
- [PV01] R.Pastor-Satorras, A.Vespignani: Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* 63, 066117 (2001); *cond-mat / 0102028*.
- [PVV01] R.Pastor-Satorras, A.Vázquez, A.Vespignani: Dynamical and Correlation Properties of the Internet. *Phys. Rev. Lett.* 87, 258701 (2001).

- [PV02] R.Pastor-Satorras, A.Vespignani: Epidemics and immunization in scale-free networks. cond-mat / 0205260.
- [Par1897] V.Pareto: Cours d'Economie Politique. Macmillan, London 1897.
- [PR86] H.-O. Peitgen, P.-H. Richter: The Beauty of Fractals. Springer, Berlin 1986.
- [PU04] M.Piatelli-Palmarini, J.Uriagereka: The immune syntax: the evolution of the language virus. pp. 341-377 in Lyle Jenkins (ed.): Variation and Universals in Biolinguistics,. Elsevier, Oxford 2004.
- [Pin00] St. Pinker: Wörter und Regeln. Spektrum, Heidelberg, Berlin 2000.
- [PP74] A.A.Piotrovskaja, R.G. Piotrovskij: Matematičeskie Modeli diachronii i tekstobrazovanija, in: Statistika reci i avtomatičeskij analiz teksta. Leningrad, Nauka 1974, 361 – 400.
- [PBP85] R.G.Piotrovskij, K.B.Bektaev, A.A.Piotrovskaja: Mathematische Linguistik. Quantitative linguistics, Bd.27. Studienverlag Brockmeyer, Bochum 1985.
- [PLL90] R.G.Piotrovskij, M.M.Lesochin, K.F,Lukjanenkov: Introduction of elements of mathematics to linguistics. Quantitative linguistics, Bd.44. Studienverlag Brockmeyer, Bochum 1990.
- [Pri65] D.J. de S. Price: Networks of Scientific Papers. Science 149, 510-515 (1965).
- [Pri76] D.J. de S. Price: A general theory of bibliometric and other cumulative advantage processes. J. Amer. Soc. Inform. Sci. 27, 292-306 (1976).
- [PS81] I.Prigogine, I.Stengers: Dialog mit der Natur. Piper, München, Zürich 1981.
- [Rap57] A.Rapoport: Contribution to the theory of random and biased nets. Bulletin of the Mathematical Biophysics 19, 257-277 (1957).
- [RB03] E.Ravasz, A.-L.Barabási: Hierarchical organisation in complex networks. Phys. Rev. E 67, 026112; cond-mat / 0206130 (2003).
- [Roj93] R.Rojas: Theorie der neuronalen Netze – Eine systematische Einführung. Springer, Berlin, Heidelberg, New York 1993.
- [Rox97] E.O.Roxin: Control Theory and its Applications. Gordon and Breach, Amsterdam 1997.
- [Sac97] L.Sachs: Angewandte Statistik (8.Aufl.). Springer, Berlin, Heidelberg, New York 1997.
- [Sau99] Th.Sauerbier: Theorie und Praxis von Simulationssystemen. Vieweg, Braunschweig, Wiesbaden 1999.

- [Sche97] M.Schenke: Development of Correct Real-Time Systems by Refinement. Habilitationsschrift, Oldenburg 1997.
- [Sche06] M.Schenke: Wachstum in zellularen Automaten. pp. 55-76 in [Har06].
- [Schw97] H.R.Schwarz: Numerische Mathematik. Teubner, Stuttgart 1997.
- [ST97] Y.Shoham, M.Tennenholtz: On the emergence of social conventions: Modelling, Analysis and Simulations. Artificial Intelligence N.94, 139-166 (1997).
- [SC02] M.Sigman, G.A.Cecchi: Global organisation of the Worldnet lexicon, Proc. Nat. Acad. Sci. 99,3, 1742-1747 (2002).
- [Sim55] H.A.Simon: On a class of skew distribution functions. Biometrics 42, 425 (1955).
- [SFMV03] R.V.Solé, R.Ferrer y Cancho, J.M.Montoya, S.Valverde: Selection, Tinkering and Emergence in Complex Networks. Wiley Periodicals, Inc., vol.8, no.1.
- [Sol05] R.V.Solé: Syntax for free? Nature 434, 289 (2005).
- [SMVS05] R.V.Solé, B.C.Murtra, S.Valverde, L.Steels: Language Networks: Their structure, function and evolution.
<http://www.santafe.edu/research/publications/workingpapers/05-12-042.pdf>
- [Ste05] L.Steels: The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. Connection Sci. 17(3-4), 11-28 (2005).
- [ST01] M.Steyvers, J.P.Tenenbaum: The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. cond-mat/0110012 (2001).
- [ST05] M.Steyvers, J.P.Tenenbaum: The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. Cognitive Science 29, 41-78 (2005).
- [Stro01] S.H.Strogatz: Exploring complex networks. Nature 410, 268-276 (2001).
- [Swa01] A.de Swaan: Words of the World. Blackwell, Oxford 2001.
- [TLW+07] C.-L. Tang, B.-Y. Lin, W.-X. Wang, M.-B. Hu, B.-H. Wang: Role of connectivity-induced weighted words in language games. Phys. Rev. E 75, 027101 (2007)
- [Ten96] M.Tennenholtz: Convention Evolution in Organizations and Markets. Computational and Mathematical Organization Theory N.2, 261-283 (1996).
- [Tul95] J.Tuldava: Methods in Quantitative Linguistics. Quantitative linguistics, Bd.54. Wissenschaftlicher Verlag, Trier 1995.
- [Ver04] J.Veronis: HyperLex: lexical cartography for information retrieval. Comput. Speech Lang. 18(3), 223-252 (2004).

[WDLP] A. de Vincenz, G.Hentschel: Wörterbuch der deutschen Lehnwörter im Polnischen. In Vorbereitung.

[WW95] A.Walker, M.Wooldridge: Understanding the Emergence of Conventions in Multi-Agent Systems. Proceedings of the 1st International Conference on Multi-Agent Systems (V.Lesser, ed.), AAAI Press, 384-389 (1995).

[Wal98] D.Wallach: Komplexe Regelungsprozesse. Deutscher Universitäts Verlag, Wiesbaden 1998.

[War02] G.Ward: Moby Thesaurus II. Project Gutenberg Literary Archive Foundation (2002). <ftp://ibiblio.org/pub/docs/books/gutenberg/etext02/mthes10.zip>

[WS98] D.J.Watts, S.H.Strogatz: Collective dynamics of `small-world` networks. Nature 393, 440-442 (1998).

[Wat04] D.J.Watts: Six Degrees: The Science of a Connected Age. W W Norton & Co Ltd (Open Market Edition), 2004.

[WLH68] U.Weinreich, W.Labov, M.I.Herzog: Empirical Foundation for a Theory of Language Change. In: W.P.Lehmann, Y.Malkiel (eds.): Directions for Historical Linguistics. Austin, University of Texas Press, 95-188 (1968).

[Wil99] M.L.Wilkins: Computer Simulation of Dynamic Phenomena. Springer, Berlin, Heidelberg, New York 1999.

[Wol02] S.Wolfram: A new kind of Science. Wolfram Media 2002.

[Word] Wordnet: A lexical database for the english language.
<http://www.cogsci.princeton.edu/~wn>

[YSH96] E.K.Yeagers, R.W.Shonkwiler, J.V.Herod: An Introduction to the Mathematics of Biology. Birkhäuser, Boston, Basel, Berlin 1996.

[ZCC03] M.Zhu, Z.Cai, O.Cai: Automatic keywords extraction of Chinese document using small world structure. pp. 438-443 in International Conference in Natural Language Processing and Knowledge Engineering, Peking. IEEE, 2003.

[Zie72] R.Zielinski: Erzeugung von Zufallszahlen. Fachbuch-Verlag Leipzig 1972.

[Zip49] G.K.Zipf: Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley, Cambridge, MA, 1949.

Abbildungsverzeichnis

Abbildung 1: Drei Ebenen wissenschaftlicher Erkenntnis.....	15
Abbildung 2: Translation auf einem regulären Gitter	27
Abbildung 3: Rotation in einem regulären Gitter.....	27
Abbildung 4: Die von-Neumann-Nachbarschaft.....	29
Abbildung 5: Die Moore-Nachbarschaft.....	29
Abbildung 6: Belegungen im eindimensionalen Gitter.....	31
Abbildung 7: Spiel des Lebens – Der „Blinker“.....	33
Abbildung 8: Ausbreitungsmuster eindimensional, drei Zustände.....	40
Abbildung 9: Wachstum zweier Funktionen (kleine Werte)	41
Abbildung 10: Wachstum zweier Funktionen (größerer Bereich).....	42
Abbildung 11: Wachstum von Quadratzahlen	43
Abbildung 12: Die Stadt Königsberg und ihre Brücken	49
Abbildung 13: Die Brücken von Königsberg als Graph	50
Abbildung 14: Linie als Graph.....	51
Abbildung 15: Der Vollständige Graph V_6 der Größe 6.....	51
Abbildung 16: Der Vollständige Binäre Baum.....	52
Abbildung 17: Reguläre Gitter als Graphen.....	53
Abbildung 18: Alternierende Zustände	55
Abbildung 19: Ausbreitung im Vollständigen Binären Baum.....	56
Abbildung 20: Ein Binärer Baum als fraktale Struktur.....	57
Abbildung 21: Sierpinski-gasket.....	58
Abbildung 22: Pareto-Verteilungen und Poisson-Verteilungen.....	66
Abbildung 23: Pareto-Verteilungen und Poisson-Verteilungen (logarithmiert).....	67
Abbildung 24: Lokale Nachbarschaft	70
Abbildung 25: Lokale Nachbarschaften in speziellen Graphen.....	71
Abbildung 26: Durchläufe im gleichseitigen Dreieck.....	72
Abbildung 27: Die Community-Structure als bipartiter Graph.....	84
Abbildung 28: Potentielle Verbindungen in einer Community-Structure	85
Abbildung 29: Realisierte Verbindungen in einer Community-Structure	86
Abbildung 30: Konstante Umschaltfunktion.....	127
Abbildung 31: Lineare Umschaltfunktion.....	127
Abbildung 32: Konstante Umschaltfunktion mit linearer Absenkung ($c=1/2$).....	129
Abbildung 33: Konstante Umschaltfunktion mit linearer Absenkung ($c=1/3$).....	130
Abbildung 34: Hohe konstante Umschaltfunktion mit linearer Absenkung ($c=1/2$).....	130
Abbildung 35: Alternierendes Verhalten in einem einfachen Netz	133
Abbildung 36: Germanistan (eindimensional, zweidimensional).....	138
Abbildung 37: Ausfügen eines Knotens mit drei Nachbarn	151
Abbildung 38: Ausfügen eines Knotens mit vier Nachbarn	151
Abbildung 39: Netz nach 1000 Generationen.....	161
Abbildung 40: Verteilung beim reversiblen Sprachwandel	174
Abbildung 41: Verteilung beim Sprachwandel allgemein	175

Index

- $\Omega(g)$ 41
- χ^2 -Test 139, 157
- ζ -Funktion 65
- 3-adisches Zahlensystem 141
- ACL-Modell 82
- AEMR 96
- affin linear 81
- aging 88
- Allgemeine einfache Mehrheitsregel 96
- Alternativhypothese 171
- Anfangsphase 131
- Anna Karenina 62
- Antisymmetrie 36
- Aphasie 107
- Ausbreitungsmuster 40
- ausdrucksorientiert 163
- Barock 166
- Baum 51
- Baum, binärer 51
- Beleg 162
- Blinker 33
- Boolesche Variable 147
- Buchung 162
- ceiling 128
- chaotisches Verhalten 123
- Charakteristische Funktion 137
- Churchsche These 16
- Clustering 69
- Clustering-Koeffizient 72
- community structure 73, 84
- configuration-Modell 78
- Conway, John Horton 32
- critical transmissibility 96
- Dependenzgrammatik 106
- Descartes, René 9
- Dimensionsreduktion 137
- Distanzmaße 75
- DT-Graph 134
- Durchmesser eines Graphen 51
- Dynamisches Sprachnetz 8, 113
- Dynamisches Sprachnetz (formal) 135
- Einstein, Albert 9
- endemische Krankheit 91
- endlicher Automat 52
- Entropie 64
- epidemische Krankheit 91
- Epidemische Schwelle 93
- Erstbeleg 8
- erzeugende Funktion 93
- Euler, Leonhardt 49
- Exponentialverteilung 64
- exponentielles Wachstum 56
- fat-tail-distribution 65
- Fixpunkt 93
- Fraktal 56
- Funktionswörter 102
- Gegenreformation 166
- General Simple Majority Rule 96
- Generation 143
- geodäsischer Pfad 51
- geometrische Verteilung 117
- Gesetz der wissenschaftlichen Produktivität 62
- giant connected component (gcc) 75
- Gittergeometrie 26
- Goldenes Zeitalter (in Polen) 166
- Grad eines Knotens 63
- Gradkorrelation 73
- Gradkorrelationskoeffizient 76
- Graph 50
- Graph, bipartiter 84
- Graph, symmetrischer 50
- Graph, vollständiger 51
- Graph, zusammenhängender 51
- Graphentheorie 49
- Halbordnung 36
- hapax legomena 165
- Herdenphänomen 84
- Hierarchie 73
- Highest Current Reward Rule 97
- hub 65
- Igel-Modell 78
- Indexjahr 165
- infected (Zustand) 91, 92
- inhaltsorientiert 163
- Internet 62
- Isomorphismus 26
- isotroper Raum 26
- Kante 50
- Knoten 50
- Kochanowski, Jan 166
- Kolmogoroff-Smirnoff-Test 171
- Komponente eines Graphen 63

- Königsberg, Problem der sieben Brücken 49
- Korrelationskoeffizient 139
- Kreis im Graph 51
- Kumulative Verteilungsfunktion 63
- Latenz 133
- Laufvariable 143
- Lehnwort 162
- Letztbeleg 165
- L-Netz 150
- lokale Nachbarschaft 69
- lokaler Clustering-Koeffizient 71
- Lotka, Alfred 62
- Makro 147
- Makroverhalten 19
- Matthäus-Effekt 80
- mean length of utterance 107
- Merriam-Webster-Thesaurus 104
- Metaregel 35
- Micro-Macro-Link 20
- Mikroverhalten 19
- mittlerer Grad 75
- Moby Thesaurus 104
- Momente 92
- Monotonie, monoton 36
- Monte Carlo Methode 12, 136
- Moore-Nachbarschaft 29
- motif 101
- MultiNet 103
- Nachbarschaft 28
- Naming Game 108
- Netzwerk 48
- Neumann, John von 21
- neuronale Netze 62
- n -tes Moment 92
- Nullhypothese 171
- objektorientierte Programmierung 140
- o-Notation 40
- Pareto-Verteilung 17, 65
- Permutation 38
- Pfad durch den Graph 50
- Phasenübergang 18, 93
- Piotrovski-Gesetz 8, 45
- Poisson-Verteilung 64
- Polnische Teilungen 166
- polynomielle Geschwindigkeit 44
- polynomielles Wachstum 44
- Polysemie 105
- Popper, Karl 15
- Potenzverteilung 65
- preferential attachment 81
- preferential linking 81
- Q-Wert 182
- Randbedingungen 32
- Rapoport, Anatol 69
- recovered (Zustand) 91
- Redundanz 64
- Reflexivität 36
- Regel des höchsten Lohnes 97
- reguläre Gittergeometrie 26
- Rej, Mikolaj 166
- removed (Zustand) 91
- Renaissance 166
- RHL 97
- Riemannsche ζ -Funktion 65
- Roget's Thesaurus 104
- Rotation 27
- Selbstähnlichkeit 56
- semantisches Netzwerk 103
- semitotalistische Regel 31
- Sierpinski-gasket 58
- Signifikanzniveau 172
- Signifikanzprofile 101
- SIR-Modell 91
- SIS-Modell 92
- Skaleninvarianz 56
- skaling 168
- small-world-Effekt 74
- Spiel des Lebens 24, 32
- sprachliche Universalien 99
- Sprachspiele 108
- stationärer Zustand 93
- Superfamilien 101
- susceptible (Zustand) 91, 92
- Symmetrie, symmetrisch 38
- Thesaurus-Netzwerke 104
- totaler Grad 75
- totalistische Regel 31
- Transitivität 36
- Translation 26
- transmissibility 95
- Turing-Maschine 16
- Ulam, Stanislaw 21
- Umschaltfunktion 122
- Umschaltwert 122
- Verteilung 63
- Verteilung, diskrete 63
- Vierfarbenvermutung 17
- V_n 51
- vollständige Induktion 42

von-Neumann-Nachbarschaft 28
Vorrangfunktion 81
Watson-Prozeß 47
Weg im Graph 50
Weißsche Bereiche 18
word co-occurrence networks 102
Word Grammar 103
Word Net 105
word web 18, 102
world wide web 62
Wort-Assoziations-Netzwerke 105
zeichenorientiert 163
zellulärer Automat 9, 24, 33
zellulärer Graph 48
zentrale verbundene Komponente 75
Zipf, George Kingsley 17
Zipfsches Gesetz 17, 110
Zufallsnetzwerk 77
Zustandsüberföhrungsfunktion 30
Zustandsüberföhrungsregel 30
Zweiwegtafel 158