



**Philosophische Fakultät  
Germanistische Sprachwissenschaft**

**Die Negation von *bar*-Adjektiven – eine korpuslinguistische  
Untersuchung mithilfe von Data-Mining-Methoden**

**Magisterarbeit**

**vorgelegt von Christiane Reiß**

**Chemnitz, den 6. März 2006**

**Gutachter:**

**Prof. Dr. Werner Holly**

**Prof. Dr. Werner Dilger**

## Inhaltsverzeichnis

	Seite
1 Zielsetzung	1
2 Vorgehensweise	3
3 Forschungsstand	4
3.1 Der Untersuchungsgegenstand: <i>bar</i> -Adjektive	4
3.2 Die Negation mittels <i>un</i> und <i>nicht</i> – Unterschiede in Bedeutung und Gebrauch	7
3.3 Bisherige Aussagen zur Verteilung von <i>un</i> und <i>nicht</i>	15
4 Untersuchungsmethoden	17
4.1 Die korpuslinguistische Analyse	17
4.2 Die benutzten Korpora	20
4.3 Das Data-Mining-Analyseverfahren	22
4.4 Probleme bei der Aufbereitung der Daten für die Analyse	28
5 Potenziell relevante Eigenschaften der Adjektive	31
5.1 Phonologische Struktur und Silbenstruktur von <i>bar</i> -Adjektiven	31
5.2 Morphologische Struktur von <i>bar</i> -Adjektiven	37
5.3 Syntaktische Merkmale von <i>bar</i> -Adjektiven	42
5.4 Semantische Merkmale von <i>bar</i> -Adjektiven	48
5.5 Weitere Merkmale von <i>bar</i> -Adjektiven	54
6 Die Data-Mining-Analyse	57
7 Auswertung der Analyseergebnisse	64
8 Zusammenfassung und Ausblick	66
Literatur	70
Anhänge	75
Anhang A: Liste der analysierten Adjektive in alphabetischer Reihenfolge	76
Anhang B: Die Attribute und ihre möglichen Werte	79
Anhang C: Beispieldatensätze	81
Selbstständigkeitserklärung	83

## 1 Zielsetzung

Die vorliegende Arbeit befasst sich mit der Negation von *bar*-Adjektiven. Obwohl die *bar*-Adjektive selbst gut erforscht sind, finden sich zu ihrer Negation in der Literatur nur vage, teils widersprüchliche Aussagen. Eine systematische Untersuchung zu diesem Thema fehlt bisher. Diese Arbeit versucht, mit einem Ansatz, der Korpusanalyse und Data Mining vereint, einen Beitrag zur Erforschung der Negation von *bar*-Adjektiven zu leisten.

*Bar*-Adjektive können auf zwei verschiedene Weisen negiert werden: durch Präfigierung mit *un* wie etwa in *undenkbar* und durch vorangestelltes *nicht* wie in *nicht essbar*.<sup>1</sup> Die Negation mit *nicht* ist immer möglich. Ausnahmen sind Bildungen, bei denen nur die mit *un* negierte Form, nicht aber das Basiswort existiert wie etwa *unentrinnbar*. Die Negation mit *un* unterliegt hingegen bestimmten Restriktionen; Formen wie *\*unfeststellbar*, *\*unessbar* und *\*unabschaltbar* klingen auffällig. Diese Restriktionen zu erforschen ist Gegenstand dieser Arbeit.

Die *bar*-Adjektive gehören zu den produktivsten Gebieten der deutschen Wortbildung. Jeder kompetente Sprecher des Deutschen kann neue *bar*-Adjektive bilden und nie vorher gehörte *bar*-Adjektive verstehen. Im Allgemeinen kann er auch – sowohl für lexikalisierte als auch für ad-hoc-Bildungen – entscheiden, ob das entsprechende *bar*-Adjektiv mit *un* negiert werden kann oder nicht. Es muss also bestimmte Regeln zur *un*-Negation geben, die implizit im Kopf eines Sprechers vorhanden sind.<sup>2</sup>

Wie kann man nun diese impliziten Regeln explizit machen? Ob ein *bar*-Adjektiv mit *un* negierbar ist oder nicht, kann sowohl von den Eigenschaften der Basis, also des *bar*-Adjektivs, als auch von den zu erwartenden Eigenschaften des mit *un* präfigierten *bar*-Adjektivs abhängen. Die Eigenschaften der Basis lassen sich ermitteln, zum Teil schlicht durch Betrachten der phonologischen und graphematischen Form des Adjektivs, zum Teil mithilfe einer Korpusanalyse. Anders verhält es sich mit den Eigenschaften des mit *un* präfigierten Adjektivs. Existiert das *un-bar*-Adjektiv nicht, lässt sich auch nicht direkt beschreiben, welche Eigenschaften

---

<sup>1</sup> Diese beiden Varianten sind nicht immer bedeutungsgleich; wir gehen darauf in Abschnitt 3.2 näher ein.

<sup>2</sup> Diese Regeln haben allerdings den Status von Präferenzen, denn in der Wortbildung gibt es viele Ausnahmen. Frans Plank formuliert das Problem folgendermaßen: „zur Wortbildungsmorphologie gehört, was mittels relativ irregulärer Regeln gebildet wird“ (Plank 1981, 14). In diesem Sinne sollten auch die Ergebnisse dieser Untersuchung betrachtet werden.

Ein weiteres Problem ist, dass nicht alle Sprecher einer Menge von Regeln die gleiche Hierarchie zuweisen. Das wirkt sich vor allem dann aus, wenn die Regeln untereinander in Konflikt stehen. Jean Aitchison hat festgestellt, dass verschiedene Sprecher in diesem Fall unterschiedlichen Regeln den Vorzug geben (vgl. Aitchison 1997, 216). Es kann also durchaus vorkommen, dass ein Sprecher ein mit *un* negiertes *bar*-Adjektiv als akzeptabel bewertet, ein anderer hingegen nicht.

dazu geführt haben, dass es nicht existiert. Glücklicherweise lassen sich viele seiner Eigenschaften auf das unpräfigierte Adjektiv zurückführen. Zwei Beispiele sollen das verdeutlichen: Angenommen, es existiert eine Regel, dass ein *bar*-Adjektiv höchstens drei Silben haben darf. Diese Regel lässt sich leicht auf das unpräfigierte Adjektiv zurückführen. Man kann dann sagen, dass das unpräfigierte *bar*-Adjektiv, damit eine Präfigierung mit *un* möglich ist, höchstens zwei Silben haben darf, denn *un* ist ja ebenfalls ein silbischer Bestandteil. Eine andere Regel könnte lauten, dass zwei betonte Silben am Anfang eines *bar*-Adjektivs vermieden werden, indem auf Negation mit *nicht* ausgewichen wird. Wenn man weiß, dass *un* im Allgemeinen betont ist, lässt sich die Regel folgendermaßen umformulieren: Das unpräfigierte *bar*-Adjektiv darf nicht auf der ersten Silbe betont sein.

Welchen praktischen Nutzen könnte eine Untersuchung wie diese haben? Möglicherweise ließen sich die Ergebnisse in der Lexikografie anwenden. Lexikografen stehen immer vor dem Problem zu entscheiden, welche Einträge sie ins Lexikon aufnehmen sollen und welche nicht. Alle Wörter einer Sprache aufzunehmen ist nicht möglich. Abgesehen davon, dass kein Lexikon dafür groß genug wäre, weiß auch niemand, wieviele Wörter eine Sprache zu einem bestimmten Zeitpunkt enthält. Die Lexikografen müssen also eine Auswahl treffen. Besonders problematisch ist das bei den Derivationen (vgl. etwa Haß-Zumkehr 2001, 34f.; Béjoint 2000, 192f.), zu denen auch die mit *un* präfigierten Adjektive gehören. Oft werden nur die Bildungen, die nicht transparent sind, ins Lexikon aufgenommen.<sup>3</sup> Das ist gerechtfertigt, denn wenn das Muster bekannt ist, nach dem ein Wort gebildet werden kann, ist ein separater Lexikon-eintrag nicht mehr nötig. Wenn man also weiß, unter welchen Voraussetzungen die Präfigierung mit *un* möglich ist, kann das Lexikon entlastet werden.

---

<sup>3</sup> Man nennt diese Bildungen auch *idiosynkratisch*. Zum Idiosynkrasiebegriff siehe Abschnitt 3.2.

## 2 Vorgehensweise

Um herauszufinden, welche Regeln für die *un*-Präfigierung gelten, sollen, wie im vorigen Kapitel bereits gesagt, möglichst viele Eigenschaften der *bar*-Adjektive betrachtet werden. Schließlich kann jede dieser Eigenschaften im Hinblick auf die Präfigierbarkeit mit *un* relevant sein. Interessant sind nicht die Eigenschaften, die für alle *bar*-Adjektive gleichermaßen gelten, denn sie können die *un*-Präfigierbarkeit nicht beeinflussen. Relevant für die Untersuchung sind nur die Eigenschaften, die sich von Adjektiv zu Adjektiv unterscheiden. Sie sind teils sehr allgemeiner Natur; ein Beispiel ist die Anzahl der Silben. Einige Eigenschaften sind jedoch sehr spezifisch für *bar*-Adjektive. Ein Beispiel dafür ist die Wortart der Basis, von der das *bar*-Adjektiv abgeleitet ist. Um möglichst viele dieser Eigenschaften zu finden, orientiere ich mich an der bisher publizierten Literatur zu *bar*-Adjektiven. Dabei erweist es sich als vorteilhaft, dass die *bar*-Adjektive gut erforscht sind. Danach werden für 263 *bar*-Adjektive alle diese Eigenschaften mithilfe einer korpuslinguistischen Untersuchung ermittelt.<sup>4</sup> Im Anschluss daran ermittle ich in einer Data-Mining-Analyse, welche der Eigenschaften die *un*-Präfigierbarkeit beeinflussen.

Die Gliederung der Arbeit ist an die Chronologie dieses Analyseprozesses angelehnt. In Kapitel 3 werden zunächst einige Aussagen getroffen zu *bar*-Adjektiven im Allgemeinen, zu Unterschieden zwischen der Negation mit *un* und der Negation mit *nicht* sowie zum Forschungsstand der Negation von *bar*-Adjektiven. Im nächsten Kapitel werden dann die Untersuchungsmethoden, die ich für die Aufgabenstellung einsetze, erläutert. Daran schließt sich der empirische Teil an. Hier werden alle möglicherweise relevanten Eigenschaften erläutert und in eine für die Data-Mining-Algorithmen verwendbare Form gebracht. In Kapitel 6 werden dann die wichtigsten Ergebnisse der Data-Mining-Analyse vorgestellt. Daran schließt sich Kapitel 7 an, in dem die Analyseergebnisse ausgewertet und interpretiert werden. In Kapitel 8 folgen schließlich eine kurze Zusammenfassung und ein Ausblick auf künftige Aufgaben.

---

<sup>4</sup> Es wird dafür das *public*-Korpus des Instituts für deutsche Sprache benutzt, das in Abschnitt 4.2 vorgestellt wird.

### 3 Forschungsstand

Im Folgenden soll ein Überblick über die Forschung zu *bar*-Adjektiven und zur Negation gegeben werden. In Abschnitt 3.1 geht es zunächst um den Forschungsstand zu *bar*-Adjektiven. In Abschnitt 3.2 wird die Negation mit *un* der Negation mit *nicht* gegenübergestellt, um Unterschiede und Gemeinsamkeiten herauszuarbeiten. In Abschnitt 3.3 werden schließlich die Aussagen in der Literatur zur Verteilung von *un* und *nicht* bei der Negation von Adjektiven, speziell *bar*-Adjektiven, referiert und diskutiert.

#### 3.1 Der Untersuchungsgegenstand: *bar*-Adjektive

Dieser Abschnitt gibt einen Überblick über die Forschung zu *bar*-Adjektiven. Dabei werden auch gegensätzliche Positionen diskutiert. Die meisten der hier behandelten Gesichtspunkte betreffen alle *bar*-Adjektive gleichermaßen; sie sind also für die nachfolgende Analyse, die ja die Unterschiede innerhalb der Klasse der *bar*-Adjektive betrachtet, nicht vordergründig relevant. Als Hintergrundinformationen sind sie dennoch wichtig.

Laut Henning Brinkmann beurteilen *bar*-Adjektive die „Welt auf die Eignung für Tätigkeiten des Menschen hin“ (Brinkmann 1962, 139). Sie werden daher oft als *Eignungswörter* oder *Eignungsadjektive* bezeichnet (vgl. Hentschel/Weydt 2003, 224; Erben 2000, 107). Prinzipiell lassen sich alle denkbaren Arten von Entitäten mithilfe von *bar*-Adjektiven charakterisieren: Gegenstände (*ein drehbarer Stuhl*), Lebewesen (*ein reizbarer Mensch*) und auch Abstrakta (*absehbare Folgen*). Allerdings müssen *bar*-Adjektive pragmatisch sinnvolle Eigenschaften beschreiben; das heißt, diese Eigenschaften dürfen nicht zum geteilten Grundwissen über die bezeichnete Entität gehören. Daher sind Beispiele wie *\*putzbare Nasen* inakzeptabel.<sup>5</sup> Wolfgang Motsch bezeichnet dies als *Prinzip des sinnvollen Wortes* (vgl. Motsch 2004, 26). Man kann die Bedeutung der meisten *bar*-Adjektive mit einer Konstruktion wie *kann geVERBt werden* paraphrasieren, wobei *VERB* hier für den Stamm des Basisverbs steht. „Adjektive auf *bar* bezeichnen Dispositionen“ (Eisenberg 2004, 279).

- (1) *unabdingbar, unentrinnbar, unverrückbar, unsinkbar*

---

<sup>5</sup> In sehr speziellen Kontexten sind jedoch auch putzbare Nasen vorstellbar, etwa in einer Werbung für ein neuartiges Nasenspray: *Dieses Spray macht Ihre Nase besser putzbar.*

Einige *bar*-Adjektive gibt es nur in der mit *un* negierten Form, etwa die in (1). Bei diesen Bildungen existiert das Basiswort nicht oder nicht mehr.<sup>6</sup> Einige dieser Basiswörter existierten in älteren Sprachstufen, einige andere sind nicht überliefert. Das heißt aber nicht, dass es sie nicht gegeben hat. Möglicherweise wurden sie nur nicht schriftlich fixiert oder die schriftlichen Belege sind verloren gegangen. Barbara Lenz betrachtet die *bar*-Adjektive ohne Basiswort synchron als Zusammenbildungen, da hier zwei Derivationsprozesse gleichzeitig ablaufen: *bar*-Suffigierung und *un*-Präfigierung (vgl. Lenz 1995, 103). Diese Adjektive sind für die vorliegende Arbeit nicht interessant, da diese Untersuchung ja gerade herausfinden will, welche Adjektive mit *un* präfigierbar sind und dabei von der nicht-negierten Form ausgeht.

Die Bildung von Adjektiven auf *bar* zählt zu den produktivsten Gebieten der Wortbildung. *Bar*-Adjektive können synchron von transitiven Verben abgeleitet werden.<sup>7</sup> Ob es noch andere produktive Muster zur Bildung von *bar*-Adjektiven gibt, wird in der Forschung kontrovers diskutiert. Susanne Riehemann hat eine Untersuchung zu *bar*-Adjektiven durchgeführt und ist zu dem Ergebnis gekommen, dass *bar*-Adjektive synchron auch von intransitiven Verben abgeleitet werden können (vgl. Riehemann 1993, 13ff.). Susann Siebert widerspricht dieser Auffassung und kritisiert, dass Riehemann lediglich eine Korpusanalyse und keine Befragung durchgeführt hat. Sie behauptet, ihrerseits eine Befragung durchgeführt zu haben, bei der die Befragten Riehemanns Beispiele – bis auf einige Ausnahmen, die allerdings ihrer Meinung nach von transitiven Verben abgeleitet sind – als ungrammatisch bewertet hätten (vgl. Siebert 1999, 68). Siebert gibt aber weder an, wie viele Informanten sie befragt hat, noch erscheinen ihre Erklärungen dafür, dass es sich bei den Ausnahmen um transitive Verben handelt, plausibel. Auch ihre Kritik an der Korpusanalyse ist nicht einzusehen. Ist eine Befragung in jedem Fall aussagekräftiger als eine bestimmte Anzahl von Belegen zu einem Wort in einem Korpus? Die Texte in einem Korpus wurden schließlich auch von Menschen geschrieben; ein Beleg in einem Korpus entspricht daher ebenso einem Grammatikalitätsurteil wie die positive Bewertung eines Wortes durch einen Informanten. Andrzej Kałny hält das Modell der *bar*-Ableitung von intransitiven Verben zwar nicht für produktiv, aber für aktiv (vgl. Kałny 1997, 83). Jindřich Toman bezeichnet diese *bar*-Adjektive als Minoritätstypen, die im Verschwin-

---

<sup>6</sup> Aufsehen erregte vor einigen Jahren die Bildung *unkaputtbar*. Auch hier existiert das Basiswort nicht (*\*kaputtbar*). Coca-Cola bewarb mit *unkaputtbar* 1990 seine neuen bruchsicheren Getränkeflaschen. Dieses Adjektiv ist eigentlich ungrammatisch, weil es nicht nach dem produktiven Muster gebildet, sondern von dem Adjektiv *kaputt* abgeleitet ist. Laut Barbara Lenz wird diese Bildung dennoch „heute so oft verwendet, dass sie mittlerweile schon als Bestandteil des Alltagswortschatzes betrachtet werden kann“ (Lenz 1998, 10). Tatsächlich ist *unkaputtbar* im *public*-Korpus mit 20 Belegen vertreten (zum *public*-Korpus siehe Abschnitt 4.2). Als Begründung dafür, dass sich diese ad-hoc-Bildung durchgesetzt hat, führt Lenz an, dass *unkaputtbar* trotz seiner Regelwidrigkeit verständlich ist. Außerdem gab es im Deutschen bisher kein Wort mit der Bedeutung von *unkaputtbar*, so dass eine kommunikative Notwendigkeit für ein solches Wort besteht.

<sup>7</sup> Peter Eisenberg nennt als weitere Einschränkung, dass *bar*-Adjektive von Verben mit *werden*-Passiv abgeleitet werden (vgl. Eisenberg 2004, 279).

den begriffen oder zumindest auf bestimmte Kontexte beschränkt sind (vgl. Toman 1987, 69). Peter Eisenberg lässt offen, ob *bar*-Adjektive synchron von intransitiven Verben abgeleitet werden können (vgl. Eisenberg 2004, 279). Ich schließe mich dieser Position an.

(2) \**redbar*, \**weinbar*, \**lachbar*, \**huldigbar*, \**grüelbar*

Elke Donalies ist der Ansicht, dass „im Prinzip [...] alle Verben mit dem Suffix *-bar* kombinierbar“ (Donalies 2002, 112) sind. Diese Auffassung ist zweifellos nicht richtig, wie die Bildungen in (2) belegen. Keine von ihnen ist im *public*-Korpus<sup>8</sup> vertreten, was für ihre Ungrammatikalität spricht.

(3) *brennbar*, *haltbar*, *haftbar*, *verzichtbar*

Sei das Muster produktiv oder nicht – es existiert eine ganze Reihe von *bar*-Bildungen, die von intransitiven Verben abgeleitet sind, etwa die in (3). Diesen *bar*-Adjektiven wird von vielen Wissenschaftlern aktivische Bedeutung attestiert, während solche, die von transitiven Verben abgeleitet sind, mit passivischer Bedeutung assoziiert werden (vgl. etwa Motsch 2004, 296ff.; Donalies 2002, 113; Fleischer/Barz 1995, 252f.).

- (4) (a) *ein Bedienelement, das man abnehmen kann*  
*ein abnehmbares Bedienelement*  
(b) *Das Bedienelement ist nicht abzunehmen.*  
*Das Bedienelement ist nicht abnehmbar.*  
(c) *ein abzunehmendes Bedienelement*  
*ein abnehmbares Bedienelement*

Robert Flury nennt drei verschiedene syntaktische Leistungen der *bar*-Adjektive (vgl. Flury 1964, 108ff.):

1. Mithilfe von *bar*-Adjektiven lassen sich komplexe Sachverhalte kompakt darstellen. So ist ein *bar*-Adjektiv in der Lage, verschiedene syntaktische Fügungen zu ersetzen. Dazu gehören Relativsätze ((4)(a)), Infinitive mit *zu* ((4)(b))<sup>9</sup> und Partizipien I mit *zu* ((4)(c)).

---

<sup>8</sup> Zum *public*-Korpus siehe Abschnitt 4.2.

<sup>9</sup> Konstruktionen wie die im ersten Satz von (4)(b) werden oft als *modale Infinitive* bezeichnet (vgl. dazu Brinkmann 1962, 139; Gelhaus 1977).



2. *Bar*-Adjektive können verbale Aussagen nominal wiedergeben. Dadurch treten sie in Textsorten, für die der Nominalstil charakteristisch ist, gehäuft auf.

3. *Bar*-Adjektive sind eine passivische Ausdrucksmöglichkeit; Konstruktionen mit *bar*-Adjektiven sind in der Lage, das Agens auszublenden. Das ist von Vorteil in Situationen, in denen der Handelnde nicht bekannt ist, nicht genannt werden soll oder nicht von Relevanz ist. Auch wenn das Agens wegfällt, bleibt der Bezug auf den Menschen implizit erhalten: Bei einer Äußerung wie in (4) wird mitverstanden, dass das Bedienelement nur von einem Menschen abgenommen werden kann (vgl. Erben 2000, 107).

### 3.2 Die Negation mittels *un* und *nicht* – Unterschiede in Bedeutung und Gebrauch

Die Negation mit *un* und die Negation mit *nicht* sind keine beliebig gegeneinander austauschbaren Varianten. In diesem Abschnitt sollen die Gemeinsamkeiten und Unterschiede erläutert werden, die zwischen den beiden Negationsarten bestehen. Es wird allerdings kein allgemeiner Überblick über den Stand der Negationsforschung gegeben. Das wäre im Rahmen dieser Arbeit auch gar nicht möglich.<sup>10</sup> Vielmehr sollen die für die nachfolgende Untersuchung relevanten Positionen in der Wissenschaft referiert werden; das heißt, das Thema Negation wird auf Adjektive und, wo möglich, speziell auf *bar*-Adjektive bezogen.

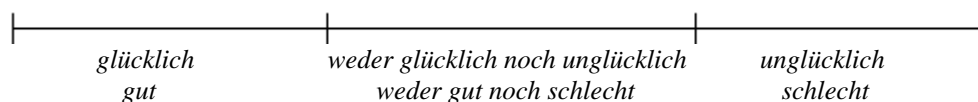
„Negation formt Inhalte in jeweils entgegengesetzte Inhalte um“ (Jacobs 1991, 560). So beschreibt Joachim Jacobs die Aufgabe der Negation. Sie ist in der Lage, drei Grundbedürfnissen des Menschen Rechnung zu tragen: dem Bedürfnis, Widerspruch oder Ablehnung zu signalisieren, dem Bedürfnis, die Welt in Gleiches und Gegensätzliches einzuteilen und dem Bedürfnis, kognitiv zu erfassen, was nicht der Fall ist (vgl. Brütsch/Nussbaumer/Sitta 1990, 2f.). *Un* und *nicht* bezeichnet Jacobs als semantisch einfache Negationsträger, weil sie nur die Negationsbedeutung tragen. Im Gegensatz dazu haben semantisch komplexe Negationsträger eine Nebenbedeutung. Im Falle von *nie* ist das etwa die temporale Existenzquantifikation (vgl. Jacobs 1991, 561f.).

Lenz unterscheidet zwischen kontradiktorischen und konträren Antonymen. Sie gibt folgende Definitionen: „Konträre Antonyme können nicht zusammen wahr, aber zusammen falsch sein. [...] Kontradiktorische Antonyme können weder zusammen wahr noch zusammen falsch sein“ (Lenz 1995, 34). Als Beispiele für konträre Antonyme nennt sie *glücklich* und *unglücklich*, als Beispiel für kontradiktorische *verheiratet* und *unverheiratet*.

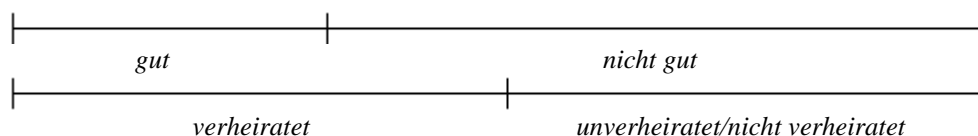
---

<sup>10</sup> Eine gute Bibliografie zum Thema Negation ist Brütsch/Nussbaumer/Sitta 1990. Dort findet der interessierte Leser alle bis 1990 erschienenen relevanten Werke sowie eine Einführung in die Thematik. Der einzige Nachteil dieser Bibliografie ist ihre mangelnde Aktualität.

## (1) (a) konträre Antonyme



## (b) kontradiktorische Antonyme



Der Unterschied zwischen konträren und kontradiktorischen Antonymen lässt sich am besten anhand von Skalen wie in (1) verdeutlichen. Konträre Antonyme stellen Bereiche an entgegengesetzten Endpunkten einer Skala dar. Dazwischen ist noch ein dritter Bereich möglich (*weder glücklich noch unglücklich*, (1)(a)). Bei kontradiktorischen Antonymen existiert dieser Zwischenbereich nicht ((1)(b)). Jacobs benutzt andere Begriffe: kontradiktorisch heißt bei ihm komplementär und konträr nennt er nicht-komplementär (vgl. Jacobs 1991, 561). Diese Unterscheidung ist nicht unproblematisch. Laut Jacobs existiert Komplementarität eigentlich nicht, da es immer Entitäten gibt, von denen man nicht sinnvoll sagen kann, sie seien etwa verheiratet oder unverheiratet – zum Beispiel Toaster oder Fußbälle (vgl. Jacobs 1991, 560f.). Streng genommen sind also alle kontradiktorischen Antonyme konträr.

Konträre Negation ist stärker als kontradiktorische. Das lässt sich gut an den jeweiligen Skalen verdeutlichen. Bei kontradiktorischer Negation ist die Skala in nur zwei Bereiche geteilt, bei konträrer Negation in mindestens drei (prinzipiell sind weitere Abstufungen denkbar). Daher ist der Bereich am negativen Ende der Skala in (1)(a) kleiner als der negative Bereich auf der Skala in (1)(b). Der Teilbereich *schlecht* in (1)(a) entspricht einem Teilbereich von *nicht gut* in (1)(b). Man kann also sagen, dass *nicht gut* aus *schlecht* folgt, nicht aber umgekehrt (vgl. Jacobs 1991, 579).

*Un* negiert Adjektive, die kein primäres Antonym haben. Außerdem bildet es die Antonyme zu abgeleiteten und partizipialen Adjektiven (vgl. Lenz 1995, 29f.). Dazu gehören auch die *bar*-Adjektive, bei denen die *un*-Präfigierung stark aktiv ist. Die Negation mit *un* ist allerdings nicht in jedem Fall möglich. Die Beschränkungen zu erforschen ist, wie bereits gesagt, Gegenstand dieser Arbeit.

Laut Lenz bildet *un* konträre und kontradiktorische Antonyme, wobei kontradiktorische Antonyme häufiger sind. Welche Art von Antonym vorliegt, hängt vom Basiswort ab (vgl. Lenz 1995, 35ff.). Eine andere Auffassung vertritt Wolfgang Motsch: „Das [mit *un*] negierte Adjektivkonzept charakterisiert in allen Fällen eine komplementäre Bedeutung“ (Motsch 2004,

288).<sup>11</sup> Dass diese Annahme nicht korrekt ist, hat bereits das Beispiel *glücklich – unglücklich* gezeigt.

Anders als *un* bildet *nicht* nur kontradiktorische Antonyme (vgl. Lenz 1995, 135; Lenz 1991, 12). Bildet *un* ein zum Grundwort kontradiktorisches Adjektiv, so entspricht seine Bedeutung der von *nicht*. Das sieht man in (1)(b) – *unverheiratet* hat die gleiche Bedeutung wie *nicht verheiratet*.

Jacobs unterscheidet zwischen *Begriffs-* und *Sachverhaltsnegation*. Begriffsnegation liegt bei morphologischen Negationsträgern wie *un* vor. Sie sind Teile von komplexen Wörtern und modifizieren die Bedeutung dieser Wörter (vgl. Jacobs 1991, 583). Motsch benutzt dafür den Ausdruck *wortinterne Negation* (vgl. Motsch 2004, 287). Syntaktische Negationsträger wie *nicht* modifizieren dagegen die Bedeutungen von Sachverhalten, also den „semantischen Korrelaten von Sätzen“ (Jacobs 1991, 568) oder „Teilen des Restsatzes“<sup>12</sup> (Jacobs 1991, 570; Hervorhebung im Original). Jacobs nennt dies *Sachverhaltsnegation* (Jacobs 1991, 569). Welche Teile des Restsatzes durch die syntaktischen Negationsträger modifiziert werden, lässt sich laut Jacobs an der Oberflächenreihenfolge im Satz ablesen (vgl. Jacobs 1991, 570f.). Oft steht ein Negationsträger wie *nicht* vor oder hinter dem Teil des Satzes, den er negiert.

- (2) (a) *Da ein Referenzzustand für eine Wirtschaft ohne Verkehr nicht fassbar und denkbar ist, kann seine Bedeutung nicht isoliert werden.*  
(b) *Man hat Monika Weber vorgeworfen, sie sei nicht fassbar und wankelmütig, eine politische Windfahne.*

Allerdings ist der Skopus von *nicht* an der Oberflächenreihenfolge nicht in jedem Fall klar erkennbar. Das lässt sich an den Beispielen in (2) erläutern. Beide stammen aus dem *public-*Korpus. *Nicht* negiert in (2)(a) *fassbar und denkbar*, in (2)(b) dagegen nur *fassbar*. In Abschnitt 5.5 wird noch deutlich werden, dass das besonders bei der Korpusanalyse zu Problemen führt.

- (3) (a) *In seinem Vortrag skizzierte Karl nicht undenkbbare Entwicklungen.*  
(b) *Karls Beruf ist mit seinem Privatleben unvereinbar.*

---

<sup>11</sup> Motsch verwendet, ebenso wie Jacobs, *komplementär* im Sinne von *kontradiktorisch*.

<sup>12</sup> Lenz ist der Auffassung, *nicht* würde stets den gesamten Restsatz negieren (vgl. Lenz 1991, 11). Dass diese Annahme nicht korrekt ist, zeigt (2).

Die Unterscheidung zwischen Begriffs- und Sachverhaltsnegation ist nicht unproblematisch, denn sie legt den Schluss nahe, dass *un* immer Wörter, *nicht* dagegen immer größere Einheiten, also Satzglieder oder Sätze, negiert. Das ist jedoch nicht der Fall. Helbig/Albrecht haben darauf hingewiesen, dass *nicht* sehr wohl auch einzelne Wörter negieren kann, etwa in (3)(a). Hier wird nur das Wort *undenkbaren* negiert, nicht aber das Satzglied *undenkbare Entwicklungen* (vgl. Helbig/Albrecht 1993, 23). Ein weiteres Beispiel ist (2)(b). Wie oben bereits besprochen, wird in (2)(b) nur *fassbar* negiert. Umgekehrt ist die morphologische Negation mit *un* nicht auf Wörter beschränkt, sondern kann auch Satzglieder modifizieren. Ein Beispiel ist (3)(b): Semantisch modifiziert das Affix *un* die gesamte Phrase *mit seinem Privatleben vereinbar*. Morphosyntaktisch verbindet es sich aber nur mit einem Wort.<sup>13</sup> Edwin Williams bezeichnet solche Phänomene als *relatedness paradoxes* (vgl. Williams 1981, 258ff.).<sup>14</sup> Aus diesem Grund unterscheiden Helbig/Albrecht nicht zwischen Wortnegation auf der einen und Satzglied- beziehungsweise Satznegation auf der anderen Seite wie Jacobs. Sie ziehen die Grenze zwischen Satznegation auf der einen und Satzglied- beziehungsweise Wortnegation auf der anderen Seite. Letzteres nennen sie *Sondernegation* (vgl. Helbig/Albrecht 1993, 27f.). *Nicht* kann Satz- oder Sondernegation bewirken, *un* nur Sondernegation (vgl. Helbig/Albrecht 1993, 39). Bei Negation mit *nicht* ist allerdings nicht immer entscheidbar, ob es in einem konkreten Beispiel Satz- oder Sondernegation bewirkt, etwa in Sätzen mit Kopulaverben wie *Er ist nicht reizbar* (vgl. Helbig/Albrecht 1993, 33).

Eine andere Unterscheidung ist die zwischen *präsupponierender* und *nicht-präsupponierender Negation*. Dies soll an Beispielen aus Blau 1978 erläutert werden:

- (4) (a) *Der Marsmond ist nicht bewohnt.*  
(b) *Der Marsmond ist unbewohnt.*  
(c) *Es ist nicht der Fall, dass der Marsmond bewohnt ist.*  
(d) *\*Der Marsmond ist unbewohnt, es gibt nämlich gar keinen.*  
(e) *Es ist nicht der Fall, dass der Marsmond bewohnt ist, es gibt nämlich gar keinen.*

Präsupponierende Negation nennt Ulrich Blau auch *behauptende* oder *starke Negation*. Sie setzt voraus, dass das Bezeichnete tatsächlich existiert. (4)(b) ist ein Beispiel für starke Negation: Es setzt voraus, dass genau ein Marsmond existiert. Deshalb ist (4)(d) ungrammatisch.

---

<sup>13</sup> Die Konstruktion *mit ... unvereinbar* tritt keineswegs selten auf. Im *public*-Korpus ist sie mit 297 Belegen vertreten.

<sup>14</sup> Williams beschreibt *relatedness paradoxes* für Derivationen und Kompositionen. Die Terminologie lässt sich aber auf Phrasen übertragen.

Nicht-präsupponierende Negation, von Blau auch *schwache* oder *bestreitende Negation* genannt, setzt dagegen nicht voraus, dass das Bezeichnete existiert. So trifft (4)(c) zwar eine Aussage über den Marsmond, setzt aber nicht voraus, dass es wirklich genau einen Marsmond gibt. Deshalb ist (4)(e) grammatisch. Laut Blau ist Negation mit *un* präsupponierend, während Negation mit *nicht* präsupponierend oder nicht-präsupponierend sein kann (vgl. Blau 1978, 49f.).<sup>15</sup> (4)(a) kann deshalb im Sinne von (4)(b), aber auch im Sinne von (4)(c) gelesen werden.<sup>16</sup>

- (5) (a) ?*Helga ist viel unglücklicher als Renate, aber sie ist trotzdem glücklich.*  
(b) *Helga ist nicht viel glücklicher als Renate, aber sie ist trotzdem glücklich.*

Ähnliches gilt laut Lenz für Komparativkonstruktionen wie in (5): (5)(a) präsupponiert, dass Helga nicht glücklich ist, deshalb klingt (5)(a) zumindest auffällig.<sup>17</sup> (5)(b) präsupponiert dagegen nicht, dass Helga nicht glücklich ist und ist daher grammatisch (vgl. Lenz 1995, 125). Jacobs unterscheidet außerdem zwischen *replaziver* und *nicht-replaziver Negation*. Bei replaziver Negation wird immer ein Teil der Satzbedeutung durch einen anderen Ausdruck ersetzt. Deshalb muss sich im Deutschen an replazive Negation immer eine *sondern*-Phrase oder etwas Äquivalentes anschließen (vgl. (6)(a)).

- (6) (a) *Für taube Menschen ist Musik nicht hörbar, sondern fühlbar.*  
(b) \**Für taube Menschen ist Musik unhörbar, sondern fühlbar.*

Die *sondern*-Phrase kann auch fehlen, wenn sie aus dem Kontext zu erschließen ist. Wenn sie fehlt, wird sie mitverstanden. Bei nicht-replaziver Negation darf dagegen keine *sondern*-Phrase stehen.

Replazive Negation muss in gesprochener Sprache durch Betonung gekennzeichnet werden. Würde man (6)(a) aussprechen, wären *hör* und *fühl* betont. Replazive Negation ist immer syntaktisch; die Negation mit *nicht* kann daher replaziv sein, die Negation mit *un* nicht. Deshalb ist (6)(b) ungrammatisch (vgl. Jacobs 1991, 585f.). Laut Helbig/Albrecht kann replazive Ne-

---

<sup>15</sup> Es ist fraglich, ob Negation mit *un* in jedem Fall präsupponierend ist. Schon (4)(d) ist meines Erachtens nicht eindeutig ungrammatisch. Würde man eine Informantenbefragung durchführen, würden sicherlich einige der Befragten (4)(d) akzeptieren.

<sup>16</sup> Barbara Lenz ist der Ansicht, *nicht* sei grundsätzlich ein schwacher Negationsträger (vgl. Lenz 1995, 123), also nicht-präsupponierend. Wie wir oben gesehen haben, kann *nicht* aber ebenso als starker Negationsträger gebraucht werden.

<sup>17</sup> Ob (5)(a) tatsächlich von der Mehrheit der Muttersprachler als ungrammatisch bewertet wird, ließe sich wiederum nur durch eine Informantenbefragung feststellen. Zumindest für die gesprochene Sprache erscheint es mir nicht völlig inakzeptabel.

gation immer dort auftreten, wo Satz- und Sondernegation positionell zusammenfallen. Die Intonation entscheidet dann darüber, ob Sondernegation vorliegt oder nicht. Falls Sondernegation beabsichtigt ist, wird sie durch Kontrastbetonung gekennzeichnet (vgl. Helbig/Albrecht 1993, 30).

Lenz behandelt neben *un* und *nicht* noch den morphologischen Negationsträger *nicht*, der als *nicht-* wie in *nicht-metallisch* oder als *nicht* wie in *nichtmetallisch* auftreten kann. Dieses gebundene *nicht* bildet stets ein kontradiktorisches Antonym zum Basisadjektiv (vgl. Lenz 1995, 36). Es entspricht also bedeutungsmäßig dem freien *nicht*. Laut Lenz hat es allerdings den gleichen Skopus wie *un*, da es sich morphologisch mit dem zu negierenden Begriff verbindet, also Begriffsnegation realisiert (vgl. Lenz 1995, 135). Oben wurden bereits die Probleme im Zusammenhang mit der Redeweise von Begriffs- und Sachverhaltsnegation erörtert. Die Negation mittels gebundenem *nicht* habe ich in Bezug auf die *bar*-Adjektive nicht systematisch untersucht, denn es trat im *public*-Korpus praktisch nie mit hinreichend vielen Belegen auf. Im *public*-Korpus sind nur 476 Belege für *bar*-Adjektive mit vorangestelltem gebundenem *nicht* vorhanden. Zum Vergleich: Für *bar*-Adjektive mit vorangestelltem *un* existieren im *public*-Korpus 196597 Belege. Es lassen sich deshalb keine fundierten Aussagen über das Vorkommen des gebundenen *nicht* treffen. Es fällt allerdings auf, dass die Belege für das gebundene *nicht* in der Mehrzahl Adjektive betreffen, die sich nicht mit *un*, sondern lediglich mit dem freien *nicht* negieren lassen. Diese Vermutung wird durch die folgende Aussage von Lenz gestützt: „Andererseits verbindet sich *nicht-* nur selten mit Lexemen, die sich mit *un-* verbinden.“ (Lenz 1995, 134). Es ist zu vermuten, dass das gebundene *nicht* – zumindest in bestimmten Kontexten – mehr oder minder eine Variante des freien *nicht* darstellt. Zu diesem Punkt sind allerdings weitere Untersuchungen nötig.

Lenz diskutiert verschiedene Arten von *Idiosynkrasien*, die mit *un* präfigierte Wörter aufweisen können (vgl. Lenz 1995, 92ff.). Idiosynkrasien sind Eigenschaften, die durch das produktive Wortbildungsmuster nicht vorhersagbar sind (vgl. Motsch 2004, 3; Motsch 1988, 2), also Unregelmäßigkeiten. Es gibt drei verschiedene Arten von idiosynkratischen Eigenschaften, die bei den mit *un* präfigierten *bar*-Adjektiven auftreten können. Bei einer Gruppe von Adjektiven existiert das Basiswort nicht mehr. Beispiele dafür sind *unentrinnbar* (\**entrinnbar*) und *unbeirrbar* (\**beirrbar*). Diese Gruppe haben wir bereits in Abschnitt 3.1 besprochen. Wie bereits gesagt, sind diese Adjektive für die Untersuchung in dieser Arbeit nicht von Interesse. Als weitere mögliche Idiosynkrasie können Bedeutungsunterschiede zwischen dem Basiswort und dem mit *un* negierten Adjektiv auftreten. Wo diese Bedeutungsunterschiede so signifikant sind, dass das *un*-Adjektiv nicht als Negation des Basiswortes angesehen werden kann, etwa

bei *scheinbar* – *unscheinbar*, sollen die Adjektive wie zwei getrennte Wörter behandelt werden. Das bedeutet, es wird beispielsweise für *scheinbar* angenommen, dass es nicht mit *un*, sondern nur mit *nicht* negiert werden kann. Wie Lenz bemerkt, treten bei Negation mit *nicht* keine solchen Bedeutungsunterschiede zwischen dem negierten und dem nicht-negierten Adjektiv auf (vgl. Lenz 1991, 25).

Die dritte mögliche Idiosynkrasie ist ein Unterschied in der Argumentstruktur zwischen Basiswort und *un*-Wort:

- (7) (a) *für jeden Hinweis dankbar*  
(b) *\*für jeden Hinweis undankbar*  
(c) *nicht für jeden Hinweis dankbar*  
(d) *\*Karl ist mit seinem Bruder unvergleichbar.*  
(e) *\*Karl ist mit seinem Bruder unverwechselbar.*  
(f) *\*Diese schwere Arbeit ist Karl unzumutbar.*

Tritt ein *bar*-Adjektiv mit einer von ihm regierten Präpositionalgruppe auf wie in (7)(a), so kann es im Allgemeinen nicht mit *un* negiert werden; (7)(b) ist ungrammatisch.<sup>18</sup> Die einzig mögliche Negation ist (7)(c). Lenz spricht in diesem Zusammenhang von *Argumentblockierung* und weist darauf hin, dass diese Reduzierung der Argumentstruktur lediglich bei Negation mit *un*, nicht aber bei Negation mit *nicht* auftritt (vgl. Lenz 1995, 4). Argumentblockierung tritt immer dann auf, wenn die Ergänzung direkt vom Basisverb an das *bar*-Adjektiv vererbt wird.<sup>19</sup> Lenz' Auffassung, dass Argumentblockierung nur bei einem einzigen *bar*-Adjektiv – nämlich *dankbar* – vorkommt (vgl. Lenz 1995, 77), muss ich widersprechen. Auch in (7)(d) und (7)(e) tritt Argumentblockierung bei Vererbung einer Präpositionalgruppe auf. (7)(f) belegt, dass dieses Phänomen nicht auf Präpositionalgruppen beschränkt ist, sondern auch bei anderen Ergänzungen – etwa solchen im Dativ – auftritt.

- (8) *Karl ist ein undankbares Kind.*

Trotz der Argumentblockierung ist *dankbar* prinzipiell mit *un* negierbar, zum Beispiel in Kontexten wie dem in (8). Die Adjektive sollen also in diesem Fall nicht wie zwei getrennte Wörter behandelt werden. Das mit *un* präfigierte Wort wird als Negation des Basisworts betrachtet.

---

<sup>18</sup> Eine Ausnahme bilden die relatedness paradoxes, die wir oben besprochen haben.

<sup>19</sup> Zur Argumentvererbung ausführlicher Abschnitt 5.3.

Wie wir in diesem Abschnitt gesehen haben, sind die Negation mit *un* und die Negation mit *nicht* nicht immer bedeutungsgleich und daher auch nicht in jedem Fall gegeneinander austauschbar. Nachfolgend werden die wichtigsten Aussagen dieses Abschnitts zu den beiden Negationsarten gegenübergestellt.

- *Un* bildet konträre und kontradiktorische Antonyme.
- *Un* ist ein morphologischer Negationsträger und realisiert Begriffsnegation.
- *Un* kann Sondernegation bewirken.
- *Un* ist präsupponierend, setzt also voraus, dass das Bezeichnete tatsächlich existiert.
- *Un* kann keine replazive Negation realisieren.
- Bei Negation mit *un* unterscheidet sich die Bedeutung des negierten Adjektivs manchmal von der Bedeutung des nicht-negierten Adjektivs.
- Bei Negation mit *un* kann Argumentblockierung auftreten.
- *Nicht* bildet nur kontradiktorische Antonyme.
- *Nicht* ist ein syntaktischer Negationsträger und realisiert Sachverhaltsnegation. An der Oberflächenreihenfolge im Satz ist der Skopus von *nicht* nicht immer klar erkennbar.
- *Nicht* kann Satz- und Sondernegation bewirken.
- *Nicht* kann präsupponierend und nicht-präsupponierend verwendet werden.
- Bei Verwendung von *nicht* ist replazive Negation möglich.
- Bei Negation mit *nicht* treten keine Bedeutungsunterschiede zwischen dem negierten und dem nicht-negierten Adjektiv auf.
- Bei Negation mit *nicht* tritt keine Argumentblockierung auf.

Die meisten der hier erörterten Gesichtspunkte beziehen sich auf konkrete Verwendungszusammenhänge. Ziel dieser Arbeit ist jedoch nicht, herauszufinden, in welchen Kontexten *un* und *nicht* gegeneinander austauschbar sind. Hier soll untersucht werden, für welche Adjektive *un*-Negation prinzipiell möglich ist. Daher sind die meisten der in diesem Abschnitt besprochenen Gesichtspunkte für diese Arbeit nicht vordergründig relevant. Lediglich zwei Aspekte werden uns noch beschäftigen: die Probleme im Zusammenhang mit dem Skopus von *nicht*



und die Bedeutungsunterschiede, die bei Negation mit *un* auftreten können. Wie bereits erläutert, sollen zwei getrennte Adjektive angenommen werden, wenn sich die Bedeutung des negierten Adjektivs signifikant von der Bedeutung des nicht-negierten Adjektivs unterscheidet.

### 3.3 Bisherige Aussagen zur Verteilung von *un* und *nicht*

Der Forschungsstand zur *un*-Präfigierbarkeit von *bar*-Adjektiven ist, wie bereits in Kapitel 1 gesagt, keineswegs zufriedenstellend. Bisher ist nicht systematisch untersucht worden, für welche *bar*-Adjektive *un*-Negation möglich ist. In der Literatur zur Wortbildung und speziell zu *bar*-Adjektiven finden sich zu diesem Thema zumeist vage, teils unzutreffende Aussagen. Einige Aussagen zur Verteilung von *un* und *nicht* wurden schon im vorigen Abschnitt im Zusammenhang mit der Erläuterung der Unterschiede zwischen *un*-Negation und Negation mit *nicht* getroffen. Hier sollen nun noch weitere Positionen diskutiert werden, die in der Literatur zu *bar*-Adjektiven oder zu Adjektiven im Allgemeinen zu finden sind.

Lenz ist der Auffassung, dass fast alle *bar*-Adjektive mit *un* präfigiert werden können (vgl. Lenz 1995, 21). Ähnlich äußert sich Eisenberg: „Wenn wir wissen, dass ein *bar*-Adjektiv nach dem produktiven Muster von einem Verb abgeleitet ist, dann kann es im allgemeinen auch mit *un* präfigiert werden“ (Eisenberg 2004, 249). Aussagen wie diese sind nicht zutreffend; sie lassen sich durch meine Beispiele widerlegen. Ich habe insgesamt 263 *bar*-Adjektive untersucht; der weitaus größte Teil von ihnen ist nach dem produktiven Muster gebildet. Allerdings lassen sich nur 131 von ihnen, also 49,8 %, mit *un* negieren.

Lenz führt einige Beschränkungen der *un*-Präfigierung an, die ihrer Meinung nach für alle Adjektive gelten. Zum einen sei *un*-Präfigierung nicht möglich bei Adjektiven, die nur attributiv oder prädikativ verwendbar sind. Auch Adjektive, die nicht komparierbar sind, könnten in der Regel nicht mit *un* negiert werden. Könnte ein Adjektiv dagegen mit *un* präfigiert werden, ließe es sich im Allgemeinen auch adverbial verwenden (vgl. Lenz 1995, 28). Mit Fleischer/Barz werden *bar*-Adjektive, die von intransitiven Verben abgeleitet sind, also nicht nach dem produktiven Muster gebildet wurden, meistens mit *un* negiert (vgl. Fleischer/Barz 1995, 253). Auch hier bringt meine Untersuchung ein anderes Ergebnis. Insgesamt 33 der von mir analysierten Beispiele sind von intransitiven Verben abgeleitet – aber nur zwölf von ihnen sind mit *un* präfigierbar.

Wolfgang Motsch trifft die folgende Aussage zur *un*-Präfigierbarkeit von *bar*-Adjektiven: „Die Bildung komplementärer Adjektivkonzepte mit dem Präfix *un*- ist nicht beliebig. Sehr häufig kommt sie mit adjektivischen Partizipien 2 und mit *-bar*-Derivationen, die nicht gra-

dierbar sind, vor“ (Motsch 2004, 288). Diese Aussage ist, wie noch deutlich werden wird, nicht zutreffend. Es sind gerade die gradierbaren *bar*-Adjektive, die die Negation mit *un* bevorzugen. Die nicht gradierbaren *bar*-Adjektive werden dagegen zumeist mit *nicht* negiert. An dieser Stelle soll auch nochmals betont werden, dass ich, entgegen Motschs Auffassung, nicht der Meinung bin, *un* würde nur zur Bildung komplementärer Adjektivkonzepte genutzt. Dieser Punkt wurde bereits im vorigen Abschnitt erörtert.

Laut Motsch müssen *un*-Adjektive pragmatisch sinnvolle Eigenschaften beschreiben. Er erklärt das am Beispiel *\*unzusammenklappbar*. Die Auffälligkeit von Bildungen wie *\*unzusammenklappbarer Tisch* begründet Motsch damit, dass Möbel normalerweise nicht zusammenklappbar sind. Nichtzusammenklappbarkeit ist also die unmarkierte Eigenschaft von Möbeln. Diese Eigenschaft lässt sich nicht durch ein mit *un* präfigiertes Adjektiv ausdrücken, da *un* immer die markierte Eigenschaft anzeigt (vgl. Motsch 2004, 288f.). Hier sind Zweifel berechtigt. Ist *\*unzusammenklappbar* wirklich aus genau diesem Grund inakzeptabel? Wenn das so wäre, dann müssten theoretisch auch Bildungen wie *unüberhörbares Alarmsignal* auffällig sein, denn es ist die unmarkierte Eigenschaft eines Alarmsignals, dass es unüberhörbar ist. Für *\*unzusammenklappbar* sind daher andere Restriktionen wahrscheinlicher. Welche das genau sind, werden wir noch sehen.

Riehemann bemerkt, dass *un* scheinbar alle offenen Argumentstellen schließt. Soll ein Adjektiv, das ein Argument besitzt, negiert werden, so muss das mittels *nicht* geschehen (vgl. Riehemann 1993, 19f.). Es wurde bereits im vorigen Abschnitt im Zusammenhang mit *relatedness paradoxes* ein Gegenbeispiel angeführt: *unvereinbar mit*. Auf dieses Beispiel geht auch Riehemann ein. Sie merkt an, dass *unvereinbar* nicht dem produktiven Muster folgt. Es müsste eigentlich *\*unvereinbarbar* heißen. Diese Form ist aber phonologisch auffällig, weshalb die letzte Silbe getilgt wird. Außerdem bezweifelt Riehemann, dass in diesem Fall echte Argumentvererbung vorliegt. Diese Begründung scheint allerdings kaum gerechtfertigt. Ob *unvereinbar mit* lediglich eine Ausnahme darstellt, bedarf einer genaueren Untersuchung.

Riehemann führt noch zwei weitere Restriktionen für die *un*-Präfigierbarkeit von *bar*-Adjektiven an: Tritt ein Adjektiv stets in einem bestimmten Kontext auf wie zum Beispiel *abbaubar* (*biologisch abbaubar*), so ist es möglicherweise nicht mit *un* negierbar. Diese Vermutung bedürfe allerdings einer eingehenderen Überprüfung. Außerdem könnte die Länge oder die Komplexität eines *bar*-Adjektivs eine Restriktion darstellen. Als Beispiel nennt sie *\*unwiederverwertbar* (vgl. Riehemann 1993, 23).

## 4 Untersuchungsmethoden

In diesem Kapitel werden die „Werkzeuge“ für die nachfolgende Untersuchung vorgestellt: Korpusanalyse und Data Mining. Die Gliederung dieses Kapitels ist angelehnt an die Chronologie des Analyseprozesses: Abschnitt 4.1 gibt zunächst einen Überblick über die Entwicklung und Bedeutung der Korpuslinguistik und motiviert den Einsatz der Korpusanalyse für die Aufgabenstellung in dieser Arbeit. In Abschnitt 4.2 wird das *public*-Korpus des Instituts für deutsche Sprache, das ich für die Analyse verwende, mit seinen Teilkorpora vorgestellt. Mithilfe der Analyse des *public*-Korpus können die Daten, die für die Data-Mining-Analyse benötigt werden, erhoben werden. In Abschnitt 4.3 wird erläutert, wie Data Mining funktioniert, welche Data-Mining-Techniken für die Aufgabenstellung in dieser Arbeit von Interesse sind und wie die Ergebnisse der Korpusanalyse in der Data-Mining-Analyse weiterverarbeitet werden. In Abschnitt 4.4 werden schließlich Probleme bei der Aufbereitung der Daten für die Data-Mining-Analyse skizziert und Lösungsansätze erörtert.

### 4.1 Die korpuslinguistische Analyse

Eine Untersuchung wie die vorliegende wäre vor 30 Jahren undenkbar gewesen, und noch vor zehn Jahren hätte sie einen ungleich höheren Aufwand an Zeit, Speicherplatz und Rechenleistung erfordert. Erst die rapide Entwicklung der Korpuslinguistik in den letzten Jahren hat eine solche Untersuchung in dem gegebenen Zeitrahmen durchführbar gemacht. In diesem Abschnitt sollen Entwicklung und Bedeutung der Korpuslinguistik skizziert und verschiedene korpuslinguistische Verfahren vorgestellt werden.<sup>20</sup>

Die Korpuslinguistik ist eine sprachwissenschaftliche Disziplin, die sich mit der Erstellung, Verwaltung und Auswertung großer Korpora gesprochener und geschriebener Sprache beschäftigt. Ein Korpus ist eine Sammlung von Texten, die sich nach verschiedenen Gesichtspunkten durchsuchen lässt. Das Ziel der Korpuslinguistik ist, aus den großen Datenmengen, die die Korpora darstellen, linguistische Informationen zu extrahieren. Die Korpuslinguistik ist allerdings kein Teilgebiet der Linguistik wie Syntax oder Semantik; vielmehr stellt sie eine Methode zur Erforschung dieser Teilgebiete dar.

---

<sup>20</sup> Die Darstellung in diesem Abschnitt kann einen Überblick geben. Eine Auflistung der relevanten Literatur zu diesem Gebiet findet der interessierte Leser in der Bibliografie zur Korpuslinguistik von Susanne Lenz (Lenz 2000).

Graeme Kennedy unterscheidet vier verschiedene korpuslinguistische Aktivitäten: die Erstellung von Korpora, die Entwicklung von Software zur Analyse der Korpora, die Analyse der Korpora und die Entwicklung von Anwendungen, die auf Ergebnissen der korpuslinguistischen Analyse beruhen, zum Beispiel für die natürliche Sprachverarbeitung. Die vorliegende Untersuchung bedient sich der Analyse von Korpora. Darauf soll kurz näher eingegangen werden. Die Korpusanalyse untersucht große Korpora von gesprochener oder geschriebener Sprache qualitativ und quantitativ. Sie will nicht nur herausfinden, welche Ausdrücke in einem Korpus auftreten, sondern auch, wie oft diese Ausdrücke im Korpus vorkommen. Darüber hinaus stellt das Korpus auch Informationen bereit über den Kontext, in dem ein bestimmter Ausdruck auftritt (vgl. Kennedy 1998, 8f.). Dieser Kontext wird auch *Konkordanz* genannt. Mithilfe der Konkordanz lässt sich etwa herausfinden, in welchen syntaktischen Konstruktionen ein bestimmter Ausdruck bevorzugt auftritt und welche Ausdrücke oft gemeinsam verwendet werden. Das erleichtert beispielsweise Lexikografen die Arbeit enorm, wenn sie ermitteln wollen, welche idiomatischen Wendungen es in einer Sprache gibt. Ohne Korpora müssten sie auf den Zufall hoffen, um Belege für die idiomatischen Wendungen zu finden (vgl. Teubert 2004, 90f.). Die Lexikografie gehört zu den Bereichen der Sprachwissenschaft, die sich durch die Korpuslinguistik am meisten verändert haben.

Susanne Lenz unterscheidet zwei grundsätzlich verschiedene Ansätze der Arbeit mit Korpora: Zum einen lassen sich mithilfe korpuslinguistischer Methoden bestehende Theorien testen und empirisch fundieren. Zum anderen kann man die Korpora auch nutzen, um neue Theorien aufzustellen. In diesem Fall wird die Theorie erst aus den Daten entwickelt, die man durch die Korpusanalyse gewinnt. Auch die vorliegende Untersuchung verfährt auf diese Weise. Wie Lenz bemerkt, rückt diese Herangehensweise die Linguistik in die Nähe der Naturwissenschaften. Dort ist diese Methodik seit langem üblich (vgl. Lenz 2000, 10ff.).

Die Korpuslinguistik in ihrer heutigen Form ist eine relativ junge Disziplin und stammt aus der englischsprachigen Linguistik. In die deutschsprachige Linguistik hat sie erst vor einigen Jahren Eingang gefunden. Wer heute von Korpora spricht, meint damit fast immer elektronische Textsammlungen, die maschinell gelesen und ausgewertet werden können. Das war nicht immer so. Früher wurden auch gedruckte Textsammlungen als Korpora bezeichnet.

Durch die Entwicklung des Computers hat die Korpuslinguistik einen enormen Entwicklungsschub erfahren. Es gab allerdings auch schon vor der Erfindung des Computers korpuslinguistische Analysen, die aber mit ungleich höherem Aufwand verbunden waren als die heutigen Analysen. Edward Lee Thorndike hat beispielsweise 1921 die häufigsten Wörter des Englischen ermittelt. Dazu wertete er ein Korpus von 4,5 Millionen Wörtern aus, das aus Werken

der klassischen Literatur und Kinderbüchern bestand. Natürlich konnte Thorndike die Texte noch nicht elektronisch speichern und auswerten. Per Hand zählte er aus, wie oft jedes Wort in seinem Korpus vorkam (vgl. Kennedy 1998, 93). Zum Vergleich: Das für die vorliegende Analyse verwendete *public*-Korpus besteht aus 1,2 Milliarden Wörtern. Die Häufigkeit eines bestimmten Wortes zu ermitteln ist heute in wenigen Sekunden möglich.

Der Computer hat nicht nur die Geschwindigkeit der Auswertung von Korpora verändert, sondern auch den Umgang mit ihnen: Riesige Korpora können sehr einfach digital gespeichert und mithilfe von Programmen wie *COSMAS II*<sup>21</sup> durchsucht werden. Die Arbeit mit Karteikarten gehört damit der Vergangenheit an. Das Internet macht die Korpora zudem von jedem Ort aus problemlos zugänglich. Viele Korpora des Instituts für deutsche Sprache lassen sich beispielsweise unter <http://www.ids-mannheim.de/cosmas2> online durchsuchen. Dazu ist keine spezielle Software nötig; ein Webbrowser genügt. Da die Korpora zentral gespeichert sind, belegen sie auf dem Rechner des Nutzers auch keinen Speicherplatz.

Erst die Korpuslinguistik ebnet den Weg zu einer empirisch fundierten Linguistik. Linguisten müssen sich nun nicht mehr auf Introspektion und Intuition verlassen, sondern können direkt an Texten belegen, welche Wortformen wie oft in Texten vorkommen (vgl. Kennedy 1998, 7) – und zwar an Texten, die uns im Alltag begegnen. Muttersprachler erwerben ihre Sprachkompetenz schließlich nicht aus Lehr- und Wörterbüchern, sondern zum Beispiel durch die Lektüre von Zeitungen und das Anschauen von Filmen. Das ist genau das Material, das die Korpuslinguistik für ihre Analysen benutzt (vgl. Teubert 2004, 96). Damit ist sie ein wichtiger Beitrag zur Entwicklung weg von einer normativen und hin zu einer wirklich deskriptiven Linguistik.

Die Korpuslinguistik stellt aber nicht nur eine Möglichkeit bereit, die linguistische Beschreibung einer Sprache zu beschleunigen und zu fundieren; sie ist auch imstande, Fakten zutage zu fördern, die mit herkömmlichen Methoden nicht zu entdecken gewesen wären. Bengt Altenberg hat beispielsweise 1991 herausgefunden, dass sich *absolutely* stärker mit negativen Wörtern wie *not* verbindet als andere Fokuspartikeln. Ohne die Hilfe der Korpuslinguistik hätte er das kaum belegen können (vgl. Kennedy 1998, 9f.). Die vorliegende Untersuchung ist ähnlich gelagert – methodisch geht sie allerdings ein Stück weiter. Ich werte nicht nur die Korpora maschinell aus, sondern auch die Ergebnisse der Korpusanalyse. Damit lassen sich Strukturen und Regeln erkennen, die dem „bloßen Auge“ verborgen bleiben. Die Verbindung von Korpusanalyse und Data Mining ist also nicht nur in der Lage herauszufinden, welche

---

<sup>21</sup> Nähere Erläuterungen zu *COSMAS II* folgen im nächsten Abschnitt.

Ausdrücke wie häufig auftreten, sondern vermag auch automatisch Schlüsse daraus zu ziehen – nämlich wovon das Auftreten einer bestimmten Form abhängt.

## 4.2 Die benutzten Korpora

Für die Analyse benutze ich, wie bereits gesagt, das *public*-Korpus des Instituts für deutsche Sprache, das alle öffentlich zugänglichen Korpora des Instituts für deutsche Sprache enthält. Diese Korpusammlung umfasst neben Texten aus deutschen, österreichischen und schweizerischen Tageszeitungen auch literarische Texte und Fachtexte. Das *public*-Korpus ist ein Monitorkorpus; das heißt, seine Zusammensetzung ändert sich mit der Zeit, weil ständig neue Texte dazukommen und bestehende Texte überarbeitet werden. Zum gegenwärtigen Zeitpunkt hat es einen Umfang von 1,2 Milliarden Wörtern und enthält im Einzelnen die folgenden Korpora:<sup>22</sup>

- *Berliner Morgenpost* der Jahrgänge 1997 bis 1999
- Bonner Zeitungskorpus – Artikel aus den Tageszeitungen *Neues Deutschland* (DDR) und *Die Welt* (Bundesrepublik Deutschland) aus den Jahren 1949, 1954, 1959, 1964, 1969 und 1974
- *COMPUTER ZEITUNG* der Jahrgänge 1993 bis 1998
- *Die Presse* der Jahrgänge 1991 bis 2000
- *Frankfurter Rundschau* der Jahrgänge 1997 bis 1999
- *Goethe*-Korpus – literarische Texte sowie autobiographische, ästhetische und naturwissenschaftliche Schriften; insgesamt 29 Texte
- Grammatik-Korpus – acht literarische Texte aus den Jahren 1971, 1977 und 1990
- *Grimm*-Korpus – 796 Sagen, Märchen und Legenden vom Anfang des 19. Jahrhunderts
- Handbuchkorpora – Zeitungs- und Zeitschriftenartikel der Jahrgänge 1985 bis 1988; hauptsächlich aus *Mannheimer Morgen*, *Die Zeit*, *stern*
- *Kleine Zeitung* der Jahrgänge 1996 bis 2000
- *LIMAS*-Korpus – 500 Textstücke mit je 2000 Textwörtern aus verschiedenen Publikationen der Jahre 1970 und 1971
- Korpus *Kartei der Gesellschaft für deutsche Sprache in Wiesbaden* – Wörter und Wortgruppen der Sprachdokumentation der Gesellschaft für deutsche Sprache vom Anfang der 70er Jahre bis Ende 1991

---

<sup>22</sup> Vgl. <http://www.ids-mannheim.de/kt/projekte/korpora/archiv.html> (03.03.2006).

- Korpus *Magazin Lufthansa Bordbuch* aus den Jahren 1995 bis 1997
- Mannheimer Korpora – Zeitungs- und Zeitschriftenartikel, wissenschaftliche und populärwissenschaftliche Texte, Handbücher, Gesetzestexte aus den Jahren 1949 bis 1974
- *Mannheimer Morgen* der Jahrgänge 1989, 1991 sowie 1994 bis 2004
- *Marx-Engels-Korpora* – 1151 Texte aus der Mitte des 19. Jahrhunderts
- *Neue Kronen-Zeitung* der Jahrgänge 1994 bis 2000
- *Oberösterreichische Nachrichten* der Jahrgänge 1996 bis 2000
- *Salzburger Nachrichten* der Jahrgänge 1991 bis 2000
- *St. Galler Tagblatt* der Jahrgänge 1997 bis 2001
- *Tiroler Tageszeitung* der Jahrgänge 1996 bis 2000
- *Vorarlberger Nachrichten* der Jahrgänge 1997 bis 2000
- Wendekorpus – Texte aus Ost- und Westdeutschland aus dem Zeitraum Mitte 1989 bis Ende 1990
- *Züricher Tagesanzeiger* der Jahrgänge 1996 bis 2000

Diese Textsammlung ist zweifelsfrei umfangreich, aber dennoch nicht unproblematisch. Die größte Gefahr dürfte darin bestehen, dass nicht alle Textsorten und Themengebiete in der gleichen Zusammensetzung repräsentiert sind, wie sie den Sprechern des Deutschen im Alltag begegnen. Fachtexte sind beispielsweise unterrepräsentiert, da die Fachsprachenkorpora des Instituts für deutsche Sprache aus urheberrechtlichen Gründen nicht öffentlich zugänglich sind. Ein weiterer Nachteil ist, dass das *public*-Korpus lediglich geschriebene Sprache enthält. Es stellt dennoch ein brauchbares Instrument für die nachfolgende Analyse dar, da es die weltweit größte Korpusammlung dieser Art ist. Auch wurden alle enthaltenen Korpora für die Untersuchung verwendet, denn es wäre schwierig gewesen zu bestimmen, welche Korpora ausgeklammert werden sollten, um die Analyse repräsentativer zu gestalten.

Das *public*-Korpus lässt sich mit dem System *COSMAS II*, das vom Institut für deutsche Sprache entwickelt wurde, sehr komfortabel durchsuchen.<sup>23</sup> *COSMAS II* zählt die einzelnen Belege für einen bestimmten Suchausdruck und bietet die Möglichkeit, die Belege auch einzeln anzuschauen. Dabei wird der Kontext eines Belegs angezeigt, so dass man analysieren kann, in welchen Zusammenhängen ein Ausdruck gebraucht wird. Das ist für die Analyse in dieser Arbeit von großer Bedeutung, wenn etwa festgestellt werden soll, in welchen syntaktischen Kontexten ein *bar*-Adjektiv auftreten kann.

---

<sup>23</sup> Weiterführende Informationen gibt das Institut für deutsche Sprache unter <http://www.ids-mannheim.de/cosmas2> (03.03.2006).

- (1) (a) *weil die Reform für mich nicht denkbar ist ohne Gesetzesänderung*  
(b) *Eine Trennung ist für beide nicht denkbar.*  
(c) *Ein Wasserexport sei für ihn nicht denkbar.*

Mithilfe von COSMAS II lässt sich das *public*-Korpus nach Wörtern oder ganzen Phrasen durchsuchen. Der Suchanfrageausdruck *denkbar\** findet beispielsweise alle Belege für *denkbar* in seinen verschiedenen Flexionsformen. Neben *denkbar* werden also auch *denkbare*, *denkbares*, *denkbarem*, *denkbaren* und *denkbarer* gefunden. Phrasen lassen sich durch Suchanfrageausdrücke wie *für /+w3 denkbar\** finden. Dieser Ausdruck liefert alle Belege, bei denen *denkbar* auf *für* folgt, wobei zwischen *für* und *denkbar* bis zu zwei weitere Wörter stehen dürfen. Einige der Belege, die dieser Ausdruck im *public*-Korpus findet, sind in (1) aufgelistet.

### 4.3 Das Data-Mining-Analyseverfahren

Mithilfe von Data Mining lassen sich – vereinfacht gesagt – Strukturen und Regeln aus großen Datenmengen extrahieren. Diese Strukturen und Regeln sind mit bloßem Auge oft nicht zu erkennen. Manchmal sind verschiedene Merkmale für sich allein auch bedeutungslos, in Kombination mit anderen jedoch wichtig. Solche Merkmalskombinationen lassen sich oft ohne algorithmische Unterstützung kaum finden. Data Mining ist für solche Aufgaben jedoch gut geeignet. Im Folgenden soll kurz erklärt werden, wie Data Mining funktioniert. Danach werden die für die nachfolgende Analyse relevanten Data-Mining-Techniken vorgestellt und erläutert.<sup>24</sup>

Ziel einer Data-Mining-Analyse ist die Erstellung eines oder mehrerer *Klassifikatoren*. Ein Klassifikator kann für jedes Beispiel vorhersagen, welcher Klasse es angehört – in diesem konkreten Fall soll der Klassifikator entscheiden, ob ein *bar*-Adjektiv mit *un* oder nur mit *nicht* negiert werden kann. Zur Erstellung von Klassifikatoren gibt es verschiedene Algorithmen; wir gehen darauf später detailliert ein. Alle Algorithmen, die ich für die Data-Mining-Analyse benutze, können mit den gleichen Eingabedaten arbeiten. Zunächst soll nun erklärt werden, wie man die Eingabedaten für die Algorithmen vorbereitet.

Zuerst müssen alle Merkmale, die für die Klassifikation relevant sein könnten, bestimmt werden. Für die *bar*-Adjektive sind das beispielsweise die Anzahl der Silben, die Art der Derivation (desubstantivisch, deadjektivisch oder deverbale) und die Frage, ob das *bar*-Adjektiv gra-

---

<sup>24</sup> Die Darstellung muss sich auf das Wesentlichste beschränken und kann daher nur einführenden Charakter haben. Einen guten Überblick über die Thematik bietet Witten/Frank 2005.



dierbar ist oder nicht. Diese Merkmale nennt man *Attribute*. Die einzelnen Attribute, die für diese Untersuchung verwendet werden, werden in Kapitel 5 detailliert beschrieben. In dieser Analyse sind das insgesamt 52 Attribute. Bevor der Klassifikator erstellt werden kann, muss für jedes *bar*-Adjektiv jeder einzelne Attributwert bestimmt werden. Es entstehen also für 263 *bar*-Adjektive je 52 Attributwerte. Einige dieser Attributwerte lassen sich an der phonologischen oder der graphematischen Form des *bar*-Adjektivs ablesen, zum Beispiel die Anzahl der Silben. Die meisten der übrigen Attributwerte, etwa die Frage, ob das *bar*-Adjektiv gradierbar ist oder nicht, werden mithilfe der Korpusanalyse ermittelt.<sup>25</sup> Wir haben im vorigen Abschnitt gesehen, wie man ein Korpus gezielt durchsuchen kann. *Hörbar* hat beispielsweise folgende Attributwerte:<sup>26</sup>

(1)

Attribut	Attributwert
Anzahl Silben	2
Derivation	deverbal
Gradierbarkeit	ja
...	...
class	un

Insgesamt sind es 52 Attributwerte; zur Verdeutlichung sind in (1) nur vier davon aufgelistet. Eines der Attribute, *class*, ist das Zielmerkmal und kann die Werte *un* und *nicht* annehmen. Die 52 Attributwerte ergeben zusammen einen *Merkmalsvektor*<sup>27</sup>. Jedes *bar*-Adjektiv wird durch einen Merkmalsvektor repräsentiert. Einen Merkmalsvektor kann man sich vorstellen wie die rechte Spalte der Tabelle in (1). Der Merkmalsvektor ist also eine Art Visitenkarte für das zugehörige *bar*-Adjektiv. Die Merkmalsvektoren der 263 untersuchten Adjektive dienen nun als Eingabe für einen Data-Mining-Algorithmus. Der Algorithmus arbeitet also im Prinzip mit einer riesigen Tabelle, die aus 263 Zeilen und 52 Spalten besteht.<sup>28</sup> Mithilfe der Eingaben kann der Algorithmus nun einen Klassifikator erstellen. Das kann man sich – stark ver-

<sup>25</sup> Es existiert noch eine kleine Gruppe von Attributen, die die Semantik der *bar*-Adjektive beschreiben und deren Werte sich nicht direkt durch die Korpusanalyse oder durch Betrachten der Formseite des Adjektivs, sondern nur durch Introspektion ermitteln lassen. Dazu gehört beispielsweise das Attribut *dauerhaft* – wir gehen darauf in Abschnitt 5.4 näher ein.

<sup>26</sup> In Anhang C befindet sich eine Tabelle, die beispielhaft für drei *bar*-Adjektive alle 52 Attributwerte auflistet.

<sup>27</sup> Eine andere Bezeichnung für *Merkmalsvektor* ist *Attributvektor*. Oft spricht man auch einfach von *Datensätzen*.

<sup>28</sup> Die Tabelle mit den 263 Datensätzen befindet sich als *.xls*-Datei und als *.pdf*-Datei auf der CD, die dieser Arbeit beiliegt. Die dazugehörigen 263 Adjektive sind auch in Anhang A aufgelistet.

einfacht – folgendermaßen vorstellen: Der Algorithmus versucht herauszufinden, welche der 51 ersten Spalten sich auf die letzte Spalte, also das Zielmerkmal *class*, auswirken.

Zur Erstellung eines Klassifikators muss die Menge der Datensätze aufgeteilt werden. Grundsätzlich gibt es dafür zwei Möglichkeiten. Bei der ersten Methode teilt man die Menge der Datensätze in Trainings- und Testdaten. Man könnte hier zum Beispiel 200 Datensätze für das Training und 63 Datensätze zum Testen verwenden. Mit den Trainingsdaten kann der Klassifikator trainiert werden; die Testdatensätze dienen zur Evaluation eines erstellten Klassifikators. Die Erstellung eines Klassifikators verläuft in zwei Phasen: In der Trainingsphase wird anhand der Trainingsdaten ein Klassifikator erstellt. Dieser wird in der Testphase ausgewertet; das heißt, es wird bestimmt, wie viele Testdatensätze der Klassifikator korrekt klassifiziert. Dazu durchläuft jeder Testdatensatz den Klassifikator. Danach wird die Vorhersage, die der Klassifikator für das Zielmerkmal dieses Datensatzes trifft, mit dem tatsächlichen Wert des Zielmerkmals verglichen. Stimmen beide Werte überein, hat der Klassifikator das Beispiel korrekt klassifiziert. Für das Beispiel *hörbar* heißt das: Sagt der Klassifikator voraus, dass *hörbar* mit *un* negiert werden kann, dann hat er *hörbar* korrekt klassifiziert, denn *class* hat bei diesem Beispiel den Wert *un*.

Die Unterteilung in Trainings- und Testdatenmenge ist geeignet, wenn genügend Datensätze zur Verfügung stehen. Das ist hier aber nicht der Fall. Ich benutze daher eine andere Methode, die eine bessere Ausnutzung der Datenmenge gewährleistet. Dafür werden die Datensätze nicht in Trainings- und Testdaten unterteilt, sondern mithilfe der so genannten *Kreuzvalidierung* evaluiert. Für die Kreuzvalidierung wird die Menge der Datensätze in  $n$  etwa gleich große Teilmengen, so genannte *Folds*, geteilt. Bewährt hat sich  $n = 10$ , auch bekannt als *tenfold cross-validation*. Nun führt man die Erstellung und die Evaluation eines Klassifikators  $n$ -mal durch, wobei jeweils  $n-1$  Folds als Trainingsdaten und ein Fold als Testdaten dienen. Jeder Fold dient somit genau einmal als Testdatenmenge. Der beste Klassifikator wird ausgegeben; die Fehlerrate wird über alle  $n$  Wiederholungen gemittelt. Um ein möglichst gutes Ergebnis zu bekommen und Abweichungen zu minimieren, kann man dafür sorgen, dass die Klassenzugehörigkeiten in den Folds proportional vertreten sind. Wenn beispielsweise 70 % der Beispiele zur Klasse *un* gehören und 30 % zur Klasse *nicht*, dann sollten die Folds so gewählt werden, dass jeder Fold ebenfalls 70 % *un*-Beispiele und 30 % *nicht*-Beispiele enthält. Das nennt man *stratifizierte Kreuzvalidierung*. In meiner Analyse arbeite ich mit stratifizierter Kreuzvalidierung und zehn Folds.

Im Folgenden sollen einige Data-Mining-Techniken, die für die nachfolgende Analyse relevant sind, kurz vorgestellt werden. Es existieren verschiedene Klassen von Data-Mining-Algorithmen. Einige Beispiele:

- Entscheidungsbaumalgorithmen
- Bayessche Netze
- Neuronale Netze
- Regellerner
- Supportvektormaschinen
- Clustering-Algorithmen

Für die Analyse in dieser Arbeit kommen nur Entscheidungsbaumalgorithmen und Regellerner infrage. Der Grund liegt darin, dass „menschenslesbare“ Klassifikatoren erzeugt werden sollen. Das heißt, es muss erkennbar sein, wie der Algorithmus die Merkmale gewichtet hat und welche Merkmalskombinationen von Bedeutung sind. Das ist bei Bayesschen Netzen, Neuronalen Netzen und Supportvektormaschinen nicht der Fall. Diese Algorithmen erzeugen zum Teil sehr genaue, aber nicht interpretierbare Klassifikatoren. Auch Clustering-Algorithmen sind für die Analyse in dieser Arbeit nicht brauchbar: Sie arbeiten ohne Klassenzugehörigkeiten und teilen die Datenmenge aufgrund von Ähnlichkeiten in Gruppen, die man *Cluster* nennt. Sie würden also keine Aussagen darüber treffen, ob ein *bar*-Adjektiv mit *un* negiert werden kann oder nicht.

Ich benutze für meine Analyse das frei verfügbare *WEKA*-System<sup>29</sup>, das alle für die geplante Untersuchung relevanten Entscheidungsbaumalgorithmen und Regellerner enthält. Nachfolgend soll kurz erläutert werden, wie Entscheidungsbaumalgorithmen und Regellerner arbeiten.

Ein Entscheidungsbaumalgorithmus versucht, möglichst exakte und kompakte Entscheidungsbäume zu generieren. Er wählt dafür aus der Datenmenge, in diesem konkreten Fall also aus 51 Attributen, das Kriterium, das die Datenmenge am besten in Teilmengen teilt. Es sind nur 51 Attribute, weil das Zielmerkmal dafür nicht verwendet werden kann. Das ermittelte Kriterium setzt der Algorithmus an die Wurzel des Entscheidungsbaums. Anhand dieses Kriteriums verzweigt der Baum. Wenn also beispielsweise das Attribut *Gradierbarkeit*, das die möglichen Werte *ja* und *nein* hat, als erstes Kriterium ausgewählt wird, entstehen auf der

---

<sup>29</sup> *WEKA* ist ein Akronym für *Waikato Environment for Knowledge Analysis* und wurde an der Universität von Waikato von Ian H. Witten und Eibe Frank entwickelt. Mehr Informationen dazu gibt Witten/Frank 2005.

zweiten Ebene zwei neue Knoten. Dem linken Knoten werden nun alle Merkmalsvektoren zugeordnet, bei denen das Attribut *Gradierbarkeit* den Wert *ja* annimmt. Dem rechten Knoten werden entsprechend die Merkmalsvektoren zugeordnet, bei denen das Attribut *Gradierbarkeit* den Wert *nein* annimmt. Auf der zweiten Ebene wählt der Algorithmus das nächste Kriterium aus, generiert neue Knoten in der dritten Ebene und so weiter. Der Algorithmus stoppt, wenn sich nur noch Merkmalsvektoren in einem Knoten befinden, bei denen das Zielmerkmal den gleichen Wert annimmt – in diesem konkreten Fall also *un* oder *nicht*. Um den Baum möglichst klein zu halten, muss dieses Stoppkriterium allerdings nicht immer exakt eingehalten werden. In der Praxis wird oft eine Majoritätsentscheidung als Stoppkriterium benutzt; das heißt, der Algorithmus terminiert, wenn der größte Teil der Beispiele in einem Knoten einer bestimmten Klasse angehört. Diesem Knoten wird dann die entsprechende Klasse – also *un* oder *nicht* – zugewiesen. Der Vorteil einer solchen Verfahrensweise ist, dass der Algorithmus mit relativ unscharfen und verrauschten Daten arbeiten kann. Das ist für die Aufgabenstellung in dieser Arbeit sehr wichtig, denn linguistische Daten sind immer unscharf. Für meine Analyse benutze ich die beiden Entscheidungsbaumalgorithmen *J4.8*<sup>30</sup> und *ADTree*<sup>31</sup>. Die von ADTree erzeugten Entscheidungsbäume unterscheiden sich in der Benutzung von den Bäumen, die J4.8 erstellt. Dieser Unterschied wird in Kapitel 6 noch näher erläutert.

Im Folgenden wird das Ergebnis der oben beschriebenen Verfahrensweise, ein Entscheidungsbaum, vorgestellt und erläutert, wie man mit seiner Hilfe ein *bar*-Adjektiv hinsichtlich der Negation mit *un* oder *nicht* klassifizieren kann.

- (1) Anzahl\_Silben  $\leq 3$   
| Gradierbarkeit = nein: nicht (30.0/9.0)  
| Gradierbarkeit = ja: un (182.0/64.0)  
Anzahl\_Silben  $> 3$ : nicht (34.0/10.0)

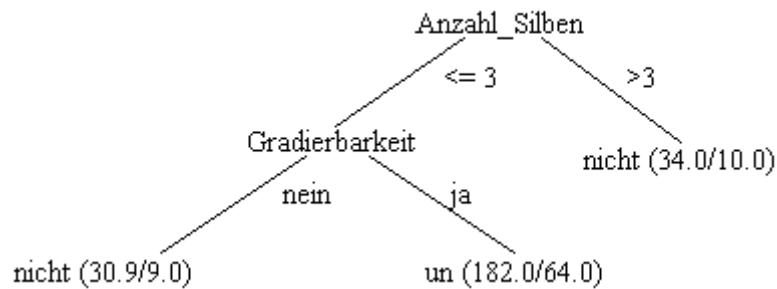
Der Entscheidungsbaum in (1) entspricht nicht den tatsächlichen Untersuchungsergebnissen; diese werden in Kapitel 6 dargestellt. Dieser Baum wurde lediglich zu Demonstrationszwecken erstellt. Der Form nach entspricht er allerdings den Bäumen, die von den Entscheidungsbaumalgorithmen im *WEKA*-System ausgegeben werden. Die Form dieses Baums mag Lesern, die mit Entscheidungsbäumen vertraut sind, ungewöhnlich erscheinen. Der Baum in (1) ließe sich auch folgendermaßen darstellen:

---

<sup>30</sup> J4.8 ist eine Implementation des *C4.5*-Algorithmus von Ross Quinlan (Quinlan 1993).

<sup>31</sup> ADTree ist eine Implementation des Algorithmus *Alternating Decision Tree* von Yoav Freund und Llew Mason (Freund/Mason 1999).

(2)



Diese Art der Darstellung ist die für Entscheidungsbäume übliche. Sie ist allerdings nur für kleine Entscheidungsbäume praktikabel. Data-Mining-Algorithmen erzeugen in der Praxis aber zum Teil wesentlich komplexere Bäume. Diese wären in der Form wie in (2) nicht mehr darstellbar. Die Ausgaben der Entscheidungsbaumalgorithmen werden daher im Folgenden stets in der Form wie in (1) dargestellt.

Die Bäume in (1) und (2) sind folgendermaßen zu lesen: Hat ein *bar*-Adjektiv höchstens drei Silben, so hängt die Klassifizierung vom Merkmal *Gradierbarkeit* ab. Ist das Adjektiv nicht gradierbar, wird es eher mit *nicht* negiert, andernfalls eher mit *un*. Hat das *bar*-Adjektiv mehr als drei Silben, so ist irrelevant, ob es gradierbar ist oder nicht. Man kann sofort sagen, dass es eher mit *nicht* negiert wird. Die Regeln in den Blättern eines Entscheidungsbaums enthalten stets das Zielmerkmal. Man fängt also oben an und bewegt sich Ebene für Ebene abwärts, bis man an einem Blatt ankommt. In (1) ist jede neue Ebene durch / gekennzeichnet. Die Zahlen in Klammern sind ein Maß dafür, wie exakt die Klassifikation an einem Blatt ist. Die erste Zahl bezeichnet die Anzahl an Beispielen, die an diesem Blatt „ankommen“, also dort klassifiziert werden. Die zweite Zahl gibt an, wie viele falsch klassifizierte Beispiele sich darunter befinden.

Zur Verdeutlichung wird nun eine Klassifikation am Beispiel *hörbar* durchgeführt: *Hörbar* hat zwei Silben, also nehmen wir den Zweig *Anzahl\_Silben*  $\leq 3$ . Als nächstes müssen wir nun das Merkmal *Gradierbarkeit* betrachten. Das Adjektiv ist gradierbar, deshalb wählen wir die Alternative *Gradierbarkeit* = *ja*. Nun stellen wir fest, dass wir an einem Blatt angekommen sind und schauen uns die Klassenzugehörigkeit an. Der Baum sagt vorher, dass *hörbar* mit *un* negiert werden kann.

Regellerner arbeiten ähnlich wie Entscheidungsbaumalgorithmen. Sie erstellen allerdings keine Bäume, sondern Regeln, anhand derer Adjektive klassifiziert werden können. Oft erzeugen Regellerner intern einen Entscheidungsbaum und „übersetzen“ ihn in Regeln. Für das Beispiel *hörbar* könnte ein Regellerner etwa die folgenden Regeln erzeugen:

- (3) Anzahl\_Silben  $\leq 3$  AND Gradierbarkeit = nein: nicht (30.0/9.0)  
Anzahl\_Silben  $\leq 3$  AND Gradierbarkeit = ja: un (182.0/64.0)  
: nicht (34.0/10.0)

Die Regeln in (3) sind hierarchisch geordnet. Um ein Beispiel zu klassifizieren, müssen sie also von oben nach unten abgearbeitet werden. Nicht alle Regellerner erzeugen solche hierarchischen Regeln. Einer der beiden Algorithmen, die für die Untersuchung in dieser Arbeit verwendet werden, *OneR*<sup>32</sup>, erzeugt nur eine einzige Regel. Der andere Algorithmus, *PART*<sup>33</sup>, erzeugt Hierarchien wie die in (3). Diese Hierarchie bewirkt, dass 30 Beispiele durch die erste Regel klassifiziert werden. Sie ist folgendermaßen zu lesen:

- (4) Wenn die Anzahl der Silben kleiner oder gleich drei ist und das Adjektiv nicht gradierbar ist, dann gehört es der Klasse *nicht* an.

Beispiele, die diese Bedingungen nicht erfüllen, werden an die zweite Regel weitergereicht, die wiederum 182 Beispiele klassifiziert. Sie kann folgendermaßen paraphrasiert werden:

- (5) Wenn die Anzahl der Silben kleiner oder gleich drei ist und das Adjektiv gradierbar ist, dann gehört es der Klasse *un* an.

Die übrigen Beispiele werden nun durch die dritte Regel klassifiziert. Diese Regel hat einen leeren Bedingungsteil. Dadurch werden alle Beispiele, die nicht durch die ersten beiden Regeln klassifiziert wurden, automatisch durch die dritte Regel klassifiziert.

### 4.4 Probleme bei der Aufbereitung der Daten für die Analyse

Bei einer Untersuchung wie dieser tritt eine Reihe von Schwierigkeiten auf. Manche von ihnen liegen in der Methodik begründet; andere hängen mit dem verwendeten Datenmaterial zusammen. Hier sollen nun die Probleme umrissen und mögliche Lösungsansätze für einige von ihnen diskutiert werden.

Eines der wichtigsten Probleme besteht darin, die Kriterien festzulegen, nach denen die zu untersuchenden Adjektive ausgewählt werden sollen. Dabei gibt es zwei konkurrierende An-

---

<sup>32</sup> OneR ist die Implementation eines sehr einfachen Regellerners von Robert C. Holte (Holte 1993).

<sup>33</sup> PART basiert auf dem Algorithmus *J4.8* und generiert Regeln aus einem zuvor erstellten Entscheidungsbaum. PART wurde von Ian H. Witten und Eibe Frank (Witten/Frank 1998) entwickelt.

forderungen: Zum einen sollen die Adjektive, die im Sprachgebrauch der deutschen Sprachgemeinschaft am häufigsten auftreten, in der Analyse vertreten sein – je häufiger ein Adjektiv vorkommt, desto größer muss die Wahrscheinlichkeit sein, dass es erfasst wird. Dieses Kriterium soll *Proportionalitätskriterium* heißen. Es werden also diejenigen *bar*-Adjektive untersucht, die mit den meisten Belegen im *public*-Korpus vertreten sind. Diese Vorgehensweise garantiert, dass die wichtigsten *bar*-Adjektive analysiert werden. Zudem lassen sich umso genauere Daten erheben, je mehr Belege zu einem Adjektiv existieren. Zum anderen sollen aber die Werte aller Attribute, die für die Data-Mining-Analyse gewählt werden, in ausreichendem Maße repräsentiert sein. Ein Beispiel ist das Attribut *un\_lich* – wir kommen später darauf zurück, hier sei nur so viel gesagt: Dieses Attribut hat die möglichen Werte *ja* und *nein*, nimmt allerdings bei fast allen *bar*-Adjektiven den Wert *nein* an. Um das Attribut *un\_lich* trotzdem in geeigneter Weise in die Analyse einbeziehen zu können, muss man gezielt nach Adjektiven suchen, bei denen *un\_lich* den Wert *ja* annimmt – und verletzt damit zwangsläufig das Proportionalitätskriterium. Das Gleiche gilt für die älteren desubstantivischen und deadjektivischen Bildungen. Um sie angemessen zu repräsentieren, habe ich gezielt Adjektive aus Flurys Untersuchung (Flury 1964) für die Analyse ausgewählt, sofern sie mit hinreichend vielen Belegen im *public*-Korpus vertreten sind – unabhängig davon, ob sie zu den häufigsten *bar*-Adjektiven gehören oder nicht. Die Untergrenze liegt bei 50 Belegen. Adjektive, die mit weniger als 50 Belegen im *public*-Korpus vertreten sind, können in der Analyse nicht berücksichtigt werden, da sich über sie kaum verlässliche Aussagen treffen lassen.

In einigen Fällen ist es auch nicht möglich, alle möglichen Attributwerte zu erfassen. Das ist beispielsweise der Fall bei den Präfixen und Verbpartikeln, die bei den *bar*-Adjektiven vorkommen können. Da die Zahl der möglichen Präfixe und Verbpartikeln potenziell unendlich ist, dürfte klar sein, dass nicht alle von ihnen in der Analyse erfasst werden können.

Ein ganzer Problemkomplex betrifft die Attribute, mit deren Hilfe die *bar*-Adjektive beschrieben werden sollen. Welche Eigenschaften von *bar*-Adjektiven sind für die geplante Analyse relevant? Ich bediene mich in diesem Punkt einer recht simplen Vorgehensweise: Potenziell relevant sind alle möglichen Eigenschaften, denn die Analyse will ja gerade herausfinden, welche Eigenschaften von Bedeutung sind und welche nicht. Ich kann keine Vorauswahl treffen; das ist auch nicht notwendig. Also versuche ich, für jedes *bar*-Adjektiv möglichst alle Eigenschaften, die die Literatur zu *bar*-Adjektiven und zu Adjektiven im Allgemeinen auflistet, zu bestimmen. Das ist mit einem gewissen Aufwand verbunden, verringert aber die Wahrscheinlichkeit, dass ein wichtiges Attribut ausgelassen wird. Ein weiteres Problem ist eher beschreibungsspezifischer Natur. Manche Attribute sind einander sehr ähnlich oder

„überlappen“ sich. Beispiele dafür sind die Anzahl der Silben und die Betonungswerte für jede einzelne Silbe. Wenn ein *bar*-Adjektiv vier Silben hat, nehmen die Betonungswerte für die fünfte und sechste Silbe stets den Wert - an, da es ja keine fünfte und sechste Silbe gibt. Manche Merkmale sind für einen Teil der *bar*-Adjektive auch überhaupt nicht relevant, ja sogar unpassend. Viele Attribute sind beispielsweise nur für deverbale *bar*-Adjektive von Belang.<sup>34</sup> Ein Beispiel ist das Attribut *deverbal\_stark\_schwach*, das erfasst, ob ein *bar*-Adjektiv von einem starken, schwachen oder unregelmäßigen Verb abgeleitet wurde. Dieses Attribut klammert die deadjektivischen und desubstantivischen *bar*-Adjektive völlig aus und muss für diese Adjektive den Wert - annehmen. Diese Unzulänglichkeiten sind nicht zu vermeiden, denn der Fokus dieser Arbeit liegt nicht darauf, die *bar*-Adjektive so klar und redundanzarm wie möglich zu beschreiben, sondern so, dass eine Data-Mining-Analyse die bestmöglichen Ergebnisse bringt.

Ein Problem stellen auch die Textwörter dar. Da die *bar*-Adjektive eine offene Klasse sind, dürfte ein signifikanter Anteil an *bar*-Adjektiven in nicht lexikalisierten ad-hoc-Bildungen bestehen. Diese sind allerdings kaum zu erfassen. Das liegt zum einen daran, dass hier nur geschriebene Sprache betrachtet wird, wo Textwörter seltener vorkommen als in gesprochener Sprache. Zum anderen ist es ja gerade das Charakteristikum von Textwörtern, dass sie für einen konkreten Kontext gebildet werden. Daher treten sie im Allgemeinen nicht in so vielen Belegen auf, dass es gerechtfertigt oder überhaupt nur möglich wäre, sie in der Analyse zu berücksichtigen.

Wichtig ist auch, hier nochmals festzuhalten, dass lediglich untersucht werden soll, ob ein *bar*-Adjektiv prinzipiell mit *un* negiert werden kann oder nicht. Wie wir in Abschnitt 3.2 gesehen haben, ist es allerdings von verschiedenen Faktoren abhängig, ob ein prinzipiell mögliches *un-bar*-Adjektiv in einem bestimmten Zusammenhang auch tatsächlich verwendet werden kann oder nicht.

---

<sup>34</sup> Diese Attribute sind trotzdem sehr aussagekräftig, denn 89 % der in dieser Arbeit untersuchten Beispiele sind deverbale *bar*-Adjektive.



## 5 Potenziell relevante Eigenschaften der Adjektive

Im Folgenden wird dargestellt, welche Eigenschaften der *bar*-Adjektive Einfluss auf die Negation mittels *un* oder *nicht* haben könnten. Zudem wird erläutert, wie diese Eigenschaften für jedes Adjektiv ermittelt werden und wie sie als Attribute für die Data-Mining-Analyse umgesetzt werden.<sup>35</sup>

Die fraglichen Eigenschaften beziehen sich stets auf das gesamte Wortparadigma, nicht auf die konkrete Wortform. Das heißt, dass beispielsweise *essbare*, *essbarer*, *essbares*, *essbarem* und *essbaren* nicht getrennt betrachtet werden, sondern als Paradigma *essbar*. Das ist ohne Probleme möglich, denn Restriktionen hinsichtlich der Präfigierbarkeit mit *un* gelten aller Wahrscheinlichkeit nach stets für das gesamte Paradigma, nicht nur für einzelne Formen. Die korpuslinguistische Analyse legt diesen Schluss jedenfalls nahe. Wenn die Grundform mit keinem oder nur sehr wenigen Belegen im *public*-Korpus vorhanden ist, gilt das Gleiche für die Flexionsformen und umgekehrt.<sup>36</sup> Für das Beispiel *essbar* heißt das: *\*Unessbar* ist nicht im *public*-Korpus vorhanden und auch *\*unessbare*, *\*unessbarer*, *\*unessbares*, *\*unessbarem* und *\*unessbaren* nicht.

Die potenziell relevanten Eigenschaften sind in vier Gruppen gegliedert, die in den folgenden vier Abschnitten besprochen werden: phonologische, morphologische, syntaktische und semantische Eigenschaften. Daran schließt sich noch eine Restgruppe mit weiteren Eigenschaften an.

### 5.1 Phonologische Struktur und Silbenstruktur von *bar*-Adjektiven

Zur phonologischen Struktur und zur Silbenstruktur gehören alle Merkmale, die die Formseite eines gesprochenen Wortes charakterisieren. Hier offenbart sich bereits eine Schwierigkeit: Es werden *bar*-Adjektive in ihrer graphematischen Form untersucht, dieser Abschnitt bezieht sich aber auf ihre phonologische Form, denn möglicherweise hängt die *un*-Präfigierbarkeit von Kriterien ab, die das gesprochene Wort betreffen. Dies bereitet einige Probleme. Wir kommen an gegebener Stelle darauf zurück.

---

<sup>35</sup> Eine vollständige Liste aller Attribute mit ihren möglichen Werten befindet sich in Anhang B.

<sup>36</sup> Eine Ausnahme bilden theoretisch Adjektive, die nur attributiv verwendbar sind. In diesem Fall wäre zu erwarten, dass die Flexionsformen im *public*-Korpus vorhanden sind, nicht aber die Grundform. Man könnte dann nicht mit Sicherheit sagen, ob die Restriktionen für die Flexionsformen auch für die Grundform gelten. Für diese Untersuchung spielen solche Erwägungen aber keine Rolle, da sich unter den untersuchten Adjektiven keines befindet, das nur attributiv verwendbar ist. Ob es überhaupt *bar*-Adjektive gibt, die nur attributiv verwendbar sind, kann nicht mit Sicherheit gesagt werden.

Im Folgenden werden die segmentalen und suprasegmentalen Merkmale der *bar*-Adjektive besprochen, die für die Analyse relevant sind.

Ein wichtiges Kriterium könnte die Anzahl der Silben sein. Damit ist hier die Anzahl der Silben vor einer eventuellen Präfigierung mit *un* gemeint. Das entsprechende Attribut für die Analyse heißt *Anzahl\_Silben* und kann die Werte 2 bis 6 annehmen. Ein *bar*-Adjektiv kann nicht weniger als zwei Silben haben, denn dem Suffix *bar* muss mindestens ein silbischer Bestandteil vorangehen. *Bar*-Adjektive mit mehr als sechs Silben sind vorstellbar, traten aber in der Korpusanalyse nicht auf. Denkbar wäre, dass für längere *bar*-Adjektive Komplexitätsrestriktionen greifen und diese eher mit *nicht* negiert werden. Wolfgang Motsch nennt diese Beschränkung *das Prinzip der Knappheit von Wortbildungen*: Demnach darf die phonologische Form von Wortbildungen nicht beliebig umfangreich sein. Der Komplexität von Mehrfachkomposita sind also Grenzen gesetzt (vgl. Motsch 2004, 26f.).

Auch die Betonung eines Adjektivs könnte Einfluss auf seine Präfigierbarkeit mit *un* haben. *Un* trägt normalerweise den Wortakzent (vgl. Kohler 1995, 188; Wiese 1996, 294). Falls das *bar*-Adjektiv auf der ersten Silbe betont ist, kann es daher bei *un*-Präfigierung zu einem Akzentzusammenstoß kommen.<sup>37</sup> In diesem Fall wird möglicherweise die Negation mit *nicht* bevorzugt, denn dadurch wird der Akzentzusammenstoß vermieden. Es muss allerdings angemerkt werden, dass die Betonung neu zugewiesen wird, wenn das Wort komplexer wird (vgl. Féry 2000, 192). Die Betonung der Basis bleibt also nach der *un*-Präfigierung nicht in jedem Fall erhalten. Das heißt, dass es nicht unbedingt zum Akzentzusammenstoß kommen muss. Ein oft angeführtes Beispiel ist *sichtbar* – bei Präfigierung mit *un* gibt *sicht* seinen Akzent auf (*únsichtbar*; vgl. etwa Kiparsky 1966, 94).<sup>38</sup> Wilbur Benware weist zudem darauf hin, dass manchmal auch Betonungsvarianten möglich sind, bei denen *un* keinen Akzent trägt – etwa bei Ableitungen mit der Struktur *un* + Präfix + Lexem + Suffix: Diese „weisen aus syntaktischen und auch aus rhythmischen Gründen Alternation auf“ (Benware 1988, 275). Benware nennt als Beispiele unter anderem *unabweisbar* (*únabweisbar* vs. *unabweísbar*) und *unantastbar* (*únantastbar* vs. *unantástbar*). Solche Varianten sind aber offenbar selten. Es scheint eine Tendenz zu geben, bei abgeleiteten Wörtern die Akzentstruktur der Basis beizubehalten. Darauf weist Jean Aitchison hin (vgl. Aitchison 1997, 214f.). Muss die Akzentstruktur verändert werden, neigen Probanden eher dazu, einen Akzent am Wortende zu verändern als am Wortanfang. Aitchisons Aussagen beziehen sich zwar auf das Englische, aber die glei-

---

<sup>37</sup> Es ist oft festgestellt worden, dass bei vielen Adjektiven auf *isch* die vorletzte Silbe betont ist (vgl. etwa Kohler 1995, 186). Möglicherweise gilt das auch für *bar*-Adjektive. Das würde bedeuten, dass bei zweisilbigen *bar*-Adjektiven oft die erste Silbe betont ist.

<sup>38</sup> Diese Akzentverschiebung scheint aber nicht in jedem Fall möglich zu sein. Das könnte ein Grund dafür sein, warum *\*unessbar* ungrammatisch ist.

chen Präferenzen könnten auch für das Deutsche gelten. Die Betonung des Basiswortes kann also durchaus aufschlussreich sein im Hinblick auf mögliche phonotaktische Restriktionen bei der *un*-Präfigierung, beispielsweise den erwähnten Akzentzusammenstoß. Daher ist es notwendig, die Betonung in die Untersuchung einzubeziehen. Um zu beschreiben, welche Silben eines Wortes betont und welche unbetont sind, bieten sich drei verschiedene Möglichkeiten. Peter Eisenberg unterscheidet zwischen den Termini *Akzent-*, *Betonbarkeits-* und *Betonungsstruktur* (vgl. Eisenberg 1991, 46).

- (1) *höhenverstellbar*
- (a) Betonbarkeitsstruktur: { + , - , - , + , + }
  - (b) Betonungsstruktur: { <sup>-</sup> , \_ , \_ , <sup>-</sup> , \_ }
  - (c) Akzentstruktur: { τ , \_ , \_ , <sup>-</sup> , \_ }

Die Betonbarkeitsstruktur gibt an, welche Silben prinzipiell betonbar sind und welche nicht. In dem Beispiel *höhenverstellbar* sind die Silben *hö*, *stell* und *bar* betonbar; nicht betonbar sind die Schwasilben *hen* und *ver* ((1)(a)). Damit wird aber keine Aussage darüber getroffen, welche Silben wirklich betont sind, denn „eine betonbare Silbe wird nur dann tatsächlich betont, wenn damit die Wortform vollständig in wohlgeformte Füße gegliedert werden kann“ (Eisenberg 1991, 56). Die Betonbarkeitsstruktur ist daher zur Formalisierung der Betonungsverhältnisse bei *bar*-Adjektiven nicht die erste Wahl. Besser geeignet ist dafür die Betonungsstruktur, die angibt, welche Silben betont und welche unbetont sind (siehe (1)(b)). Sie unterscheidet allerdings bei den betonten Silben nicht zwischen Haupt- und Nebenakzent wie die Akzentstruktur. Sowohl (1)(b) als auch (1)(c) gibt an, welche Silben betont sind, nämlich *hö* und *stell*.<sup>39</sup> Aus (1)(c) kann man zusätzlich entnehmen, dass der Wortakzent auf *hö* liegt. Die Entscheidung fällt also zugunsten der Akzentstruktur, da sich mit ihrer Hilfe die Betonung der *bar*-Adjektive genauer beschreiben lässt als mittels der Betonungsstruktur. Wir führen also für die Akzentstruktur insgesamt sechs Attribute mit den Namen *Akzentbelegung\_Silbe\_1* bis *Akzentbelegung\_Silbe\_6* ein. Jedes dieser Attribute kann die Werte *H* (Hauptbetonung beziehungsweise Wortakzent), *N* (Nebenbetonung) oder *U* (unbetont) annehmen, ab *Akzentbelegung\_Silbe\_3* auch den Wert - (Silbe nicht vorhanden).

---

<sup>39</sup> Einige Autoren nehmen an, dass *bar* einen Nebenakzent trägt, andere Derivationsuffixe wie *lich*, *ig* und *isch* hingegen nicht (vgl. etwa Kiparsky 1966, 72f.). Wiese kritisiert diese Ansicht. Er ist der Auffassung, dass Muttersprachler sich nicht bewusst sind, dass *bar* etwa in *sichtbar* einen (wenn auch schwachen) Akzent trägt, *lich* in *sichtlich* aber nicht (vgl. Wiese 1996, 274f.). Ich schließe mich Wieses Auffassung an und weise *bar* nur dann einen Nebenakzent zu, wenn es für die Gliederung der Wortform in wohlgeformte Füße notwendig ist, etwa in *ánrechenbàr* und *vernàchlàssigbàr*.

Zur Beschreibung der Silbenstruktur eines Wortes existiert in der Forschung eine Fülle unterschiedlicher Begriffe und Modelle. Einige von ihnen sollen hier kurz umrissen werden. Im *Kern*<sup>40</sup> einer Silbe befindet sich meist ein Vokal oder Diphthong. Ein Vokal kann *gespannt* wie in *Wal* oder *ungespannt* wie in *Wall* sein. Vor dem Kern liegt der *Anfangsrand*<sup>41</sup>, nach dem Kern der *Endrand*<sup>42</sup>. Anfangsrand und Endrand können aus null, einem oder mehreren Konsonanten bestehen. Dem entsprechen die Begriffe *leerer*, *einfacher* und *komplexer Anfangs-* beziehungsweise *Endrand*. Ist der Anfangsrand leer, enthält er also keinen Konsonanten, so bezeichnet man die Silbe als *nackt*, andernfalls als *bedeckt*. Silben mit leerem Endrand heißen *offen*, Silben mit nicht-leerem Endrand dagegen *geschlossen* (vgl. Eisenberg 2004, 101). Schwierigkeiten bereitet in diesem Zusammenhang der glottale Plosiv [ʔ], zumindest wenn man ihn als Phonem akzeptieren will. Dieser Konsonant besetzt immer dann den Anfangsrand einer Silbe, wenn sich dort kein anderer Konsonant befindet.<sup>43</sup> Demnach dürfte es streng genommen keine nackten Silben geben, da im Anfangsrand ja [ʔ] als Default-Konsonant steht. Ich schließe mich dieser Sichtweise nicht an, denn der Status von [ʔ] ist nicht hinreichend geklärt. Von einigen Wissenschaftlern, beispielsweise von Richard Wiese, wird er nicht als Phonem angesehen. Wiese führt zwei Argumente für seine Auffassung an: Das Vorkommen von [ʔ] ist vollständig vorhersagbar – er kann am Anfang von Silben stehen, die andernfalls mit einem Vokal beginnen würden. Zudem ist die Artikulation von [ʔ] optional (vgl. Wiese 1996, 58). Ihm entspricht auch kein Graphem. Daher ist die Beschäftigung mit [ʔ] bestenfalls von theoretischem Interesse. Wir werden der Einfachheit halber von leeren Anfangsrändern sprechen, wo [ʔ] stehen müsste.

Peter Eisenberg und Theo Vennemann unterscheiden zwischen *leichten* und *schweren* Silben. Bei Eisenberg sind leichte Silben solche mit ungespanntem Vokal und einfachem oder leerem Endrand. Als schwer bezeichnet er dagegen Silben, deren Kern aus einem gespannten Vokal oder einem Diphthong besteht oder die einen komplexen Endrand haben (vgl. Eisenberg 1991, 41). Vennemann gebraucht die Begriffe etwas anders. Um dies näher zu erläutern, müssen noch einige weitere Termini eingeführt werden. Vennemann unterscheidet nicht zwischen gespannten und ungespannten Vokalen, sondern zwischen Silben mit *scharfem* und Silben mit

---

<sup>40</sup> Eine andere Bezeichnung für *Kern* ist *Nukleus*.

<sup>41</sup> Der Anfangsrand wird auch *Onset* genannt.

<sup>42</sup> Oft verwendet man für den Endrand auch die Bezeichnung *Koda*.

<sup>43</sup> Eine Ausnahme bilden zwei aufeinander folgende Vokale, zwischen denen eine Silbengrenze verläuft und bei denen der zweite Vokal unbetont ist wie in *re|alisierbar*. Hier steht [ʔ] nicht vor *a*. Dazu ausführlicher Wiese 1996, 58f.

*sanftem Schnitt*. Vollsilben<sup>44</sup> haben einen Energieverlauf aus Crescendo und Decrescendo. Scharf geschnitten sind alle Silben, bei denen das Decrescendo nach dem Kern beginnt wie in *Bett*. Bei sanft geschnittenen Silben wie *Beet* beginnt das Decrescendo hingegen schon im Kern. Das ist bei allen Silben mit gespanntem Vokal oder Diphthong der Fall. Als *leicht* bezeichnet Vennemann nun alle Silben, die „offen, monophthongisch und sanft geschnitten sind“ (Vennemann 1991, 97). Alle anderen Silben bezeichnet er als *schwer* (vgl. Vennemann 1991, 87ff.).

Ein anderer Ansatz benutzt zur Beschreibung des Silbengewichts den Begriff der *Morigkeit*. Caroline Féry benutzt dieses Konzept folgendermaßen: Die morigen Bestandteile der Silbe sind Kern und Endrand; sie tragen Gewicht zur Silbe bei. Es gibt nicht-morige, einmorige, zweimorige und dreimorige Silben. Nicht-morig sind beispielsweise Schwasilben. Einmorig sind offene Silben mit ungespanntem Vokal wie *pri* in *Paprika*. Zweimorige Silben haben entweder einen gespannten Vokal im Kern und einen leeren Endrand wie *ma* in *Komma* oder einen ungespannten Vokal im Kern und einen einfachen Endrand wie *Kom* in *Komma*. Dreimorige Silben haben einen gespannten Vokal im Kern und einen einfachen Endrand wie *tom* in *Symptom* oder einen ungespannten Vokal im Kern und einen komplexen Endrand wie *Symp* in *Symptom* (vgl. Féry 2000, 163f.).

Welcher dieser Ansätze ist nun am besten geeignet, die Silbenstruktur der *bar*-Adjektive zu formalisieren? Die Theorien ähneln sich zwar, legen aber den Fokus auf verschiedene Aspekte. Es besteht die Gefahr, dass ein wichtiger Gesichtspunkt ausgeklammert wird, wenn man sich für einen Ansatz – und damit gegen die anderen – entscheidet. Aus diesem Grund soll versucht werden, einen gemeinsamen Nenner für alle drei Theorien zu finden. Das geht am besten, wenn man von den phonologischen Grundbegriffen ausgeht. Wir setzen daher für jede Silbe drei Attribute an und behandeln Anfangsrand, Kern und Endrand separat. Somit ergeben sich insgesamt 18 Attribute für das Konzept der Silbenstruktur, denn es werden ja *bar*-Adjektive mit bis zu sechs Silben betrachtet. Die Attribute tragen die Namen *Anfangsrand\_Silbe\_1*, *Kern\_Silbe\_1* und *Endrand\_Silbe\_1*; entsprechend natürlich für die Silben zwei bis sechs. *Anfangsrand\_Silbe\_1* und *Endrand\_Silbe\_1* können die Werte *einfach*, *komplex* und *leer* annehmen, *Kern\_Silbe\_1* die Werte *gespannt*, *ungespannt*, *Diphthong*, *Schwa*. Für die dritte bis sechste Silbe können alle drei Attribute auch den Wert *-* annehmen, da diese Silben ja nicht in jedem Fall vorhanden sind.

---

<sup>44</sup> Der Gegenbegriff zu *Vollsilbe* ist *Reduktionssilbe*. Reduktionssilben sind alle Silben, deren Kern aus Schwa, einem Liquid oder einem Nasal besteht.

(2)(a) *Anfangsrand\_Silbe\_2 = einfach*

*Kern\_Silbe\_2 = ungespannt*

*Endrand\_Silbe\_2 = einfach*

(2)(b) *Anfangsrand\_Silbe\_2 = einfach*

*Kern\_Silbe\_2 = gespannt*

*Endrand\_Silbe\_2 = leer*

Auf diese Weise sind alle drei der oben vorgestellten Ansätze integriert. Ist die zweite Silbe eines Wortes leicht in Eisenbergs Verwendungsweise des Begriffes, so könnten die Attribute zum Beispiel die Werte in (2)(a) annehmen. Eine leichte Silbe im Sinne von Vennemann wäre beispielsweise die in (2)(b). Beide Beispiele entsprechen einer zweimorigen Silbe.

Ein Problem ist, wie bereits gesagt, dass sich die Werte der in diesem Abschnitt eingeführten Attribute nicht aus dem geschriebenen Wort ablesen lassen, denn dieses besteht ja aus Graphemen statt Phonemen und enthält auch keine Informationen zu Silbengrenzen. Diesem Teil der Untersuchung liegt daher die gesprochene Form der Adjektive zugrunde. Es kann dennoch darauf verzichtet werden, für jedes *bar*-Adjektiv die phonetische Repräsentation anzugeben, denn die Attributwerte lassen sich in der Regel leicht bestimmen, indem man das Adjektiv laut ausspricht.

(3) (a) /fəʀ/

*Anfangsrand\_Silbe\_2 = einfach*

*Kern\_Silbe\_2 = Schwa*

*Endrand\_Silbe\_2 = einfach*

(b) /fɛ/

*Anfangsrand\_Silbe\_2 = einfach*

*Kern\_Silbe\_2 = unbetont*

*Endrand\_Silbe\_2 = leer*

Schwierigkeiten bereiten auch Silben mit Schwa im Kern. Soll man beispielsweise für *ver* in *vereinbar* die phonetische Repräsentation /fəʀ/ oder eher /fɛ/ zugrunde legen? Diese Frage beeinflusst die Attributwerte: Für /fəʀ/ ergibt sich (3)(a), für /fɛ/ (3)(b). Beide Varianten sind

nicht unproblematisch.<sup>45</sup> In dieser Arbeit lege ich die Lautung zugrunde, die Eisenberg *Explizitlautung* nennt (vgl. Eisenberg 2004, 12). Für das Beispiel *ver* ist das die Variante (3)(a).

## 5.2 Morphologische Struktur von *bar*-Adjektiven

Zur morphologischen Struktur der *bar*-Adjektive gehören alle Merkmale, die den Aufbau der Adjektive aus einzelnen Morphemen charakterisieren oder die mit der Morphologie des zugrunde liegenden Basiswortes zusammenhängen. Diese morphologischen Merkmale sollen im Folgenden eingeführt werden. Außerdem wird wieder erläutert, wie diese Merkmale in Attribute für die Data-Mining-Analyse umgesetzt werden. Die Werte aller in diesem Abschnitt eingeführten Attribute lassen sich durch Betrachten der graphematischen Form des Adjektivs ermitteln.

*Bar*-Adjektive können von einfachen und von komplexen Stämmen abgeleitet sein. Unter einfachen Stämmen sollen solche verstanden werden, die aus einem Morphem bestehen; unter komplexen Stämmen entsprechend solche, die aus mehreren Morphemen bestehen.<sup>46</sup> Das sind zum Beispiel präfigierte Verben (*erzählen*) und Verben mit Derivationsuffix (*flektieren*). Wir führen dafür das Attribut *Komplexitaet\_Basis* mit den Werten *einfach* und *komplex* ein. Pseudosuffixe, die zur nominalen Basis gehören wie *el* in *verhandelbar* und *er* in *entzifferbar* sollen nicht als eigene Morpheme angesehen werden. Das ist nicht ganz unproblematisch, da Substantive wie *Handel* und *Ziffer* möglicherweise früher einmal morphologisch komplex waren. Synchron sind die Pseudosuffixe jedoch ohne eigenständige Bedeutung. Deshalb sollen Derivationsbasen, die nur aus dem Stamm mit Pseudosuffix bestehen, als morphologisch einfach betrachtet werden.<sup>47</sup>

Eine Vielzahl von *bar*-Adjektiven ist von Verben mit Präfix oder Verbpartikel abgeleitet.<sup>48</sup> Diese Verben zu systematisieren bringt einige Schwierigkeiten mit sich. Ich stelle hier die Ansätze von Fleischer/Barz (1995) und Eisenberg (2004) gegenüber und entwickle daraus die für die Analyse benötigten Attribute.

Fleischer/Barz gliedern die Präfixverben in zwei Gruppen: *trennbare* und *untrennbare* Verben. Bei trennbaren Verben liegt der Akzent auf dem Präfix. Außerdem trennt beim Partizip II

---

<sup>45</sup> Zu den Schwierigkeiten ausführlich Eisenberg 2004, 98f.

<sup>46</sup> Riehemann, die ebenfalls eine korpuslinguistische Studie zu *bar*-Adjektiven durchgeführt hat, stellt fest, dass nur ein Bruchteil der *bar*-Adjektive von einfachen Stämmen abgeleitet ist (vgl. Riehemann 1993, 24). Zum gleichen Ergebnis kommt Toman. Er bemerkt, dass nur 10 % der Neubildungen des 19. Jahrhunderts und 10 % der Neubildungen des 20. Jahrhunderts einfache Basen haben (vgl. Toman 1987, 75f.). Meine Untersuchung bestätigt dies.

<sup>47</sup> Zu dieser Problematik ausführlicher Eisenberg 2004, 211.

<sup>48</sup> 59 % der von mir untersuchten Beispiele sind von Präfix- oder Partikelverben abgeleitet.

ge das Präfix vom Stamm (*abgeholt*). Beim *zu*-Infinitiv steht entsprechend *zu* zwischen Präfix und Verbstamm (*abzuholen*). Eine Ausnahme bilden Verben mit dem Suffix *ieren* und doppelt präfigierte Verben: Sie bilden das Partizip II ohne *ge* (*ausradiert*, *umverteilt*). Untrennbare Verben sind dagegen stammbetont, bilden das Partizip II ohne *ge* (*ergangen*) und beim *zu*-Infinitiv steht *zu* durch ein Leerzeichen getrennt vor der Verbform wie bei *zu ergehen* (vgl. Fleischer/Barz 1995, 342). Einige der Verbpräfixe kommen nicht frei vor. Dazu zählen *ge*, *be*, *er*, *miss*, *ver*, *ent* und *zer*. Sie bilden nur untrennbare Verben. Die Verbpräfixe, die auch frei vorkommen, sind jünger und bilden – von wenigen Ausnahmen abgesehen – nur trennbare Verben. Zu dieser Gruppe gehören *auf*, *bei*, *ab*, *nach*, *an*, *vor*, *aus*, *dar*, *los*, *zu* und *ein*. Neben trennbaren und untrennbaren Verben gibt es auch doppelförmige Verben, die trennbar oder untrennbar sein können (*umfahren*). Welche Variante gemeint ist, muss durch die Betonung kenntlich gemacht werden: Die untrennbaren Verben sind stammbetont, die trennbaren haben die Betonung auf dem Präfix. Doppelförmige Verben werden mithilfe von *wider*, *über*, *ob*, *durch*, *um* und *unter* gebildet.<sup>49</sup> Fleischer/Barz zählen *hinter*, *voll* und *wieder* „aus semantischen Gründen“ (Fleischer/Barz 1995, 294) nicht zu den Präfixen, sondern zu den Kompositionsgliedern<sup>50</sup> (vgl. Fleischer/Barz 1995, 294).

Eisenberg klassifiziert anders. Präfixverben sind bei ihm alle Verben, die das Partizip II ohne *ge* und den *zu*-Infinitiv mit vorangestelltem *zu* bilden. Dagegen grenzt er die so genannten *Partikelverben* ab. Deren Erstglieder nennt er nicht *Präfixe*, sondern *Verbpartikeln*. Bei den Partikelverben steht *ge* im Partizip II zwischen Verbpartikel und Stamm; beim *zu*-Infinitiv trennt *zu* die Verbpartikel vom Stamm. Außerdem tritt das flektierte Verb oft in zweigeteilter Form auf wie etwa in *ich höre zu* (vgl. Eisenberg 2004, 255). Man könnte nun leicht den Eindruck bekommen, Eisenbergs Präfixverben entsprächen den untrennbaren Verben bei Fleischer/Barz, während Fleischer/Barz' trennbare Verben dasselbe bezeichnen wie Eisenbergs Partikelverben. Dem ist nicht so. Die Klassen besitzen große Schnittmengen, sind aber nicht äquivalent. Für Eisenberg gehören Formen wie *hintergehen*, *wiederholen* und *vollziehen*, die Fleischer/Barz als Kompositionen bezeichnen, zu den Präfixverben, wenngleich er sie als „Restgruppe“ (Eisenberg 2004, 256) bezeichnet.

- (1) (a) *kopfrechnen*, *worthalten*, *eislaufen*
- (b) *krankschreiben*, *fertigmachen*, *totlachen*
- (c) *spazierengehen*, *steckenlassen*, *kennenlernen*

---

<sup>49</sup> Das heißt aber nicht, dass mit diesen Präfixen ausschließlich doppelförmige Verben gebildet werden können. Es gibt auch trennbare (*durchstehen*) und untrennbare (*umkreisen*) Verben mit diesen Präfixen.

<sup>50</sup> Nicht ganz schlüssig ist, warum *vor* als Präfix, *hinter* aber Kompositionsglied klassifiziert wird.



(d) *dabeibleiben, hinaufgehen, dazukommen*

(e) *zugutehalten, emporblicken, anheimstellen*

Auch zu Eisenbergs Partikelverben gehören einige, die Fleischer/Barz zu den Komposita zählen. Das sind solche mit substantivischer ((1)(a)), adjektivischer ((1)(b)), verbaler ((1)(c)) oder adverbialer Verbpartikel ((1)(d)). Auch einige Verben, deren Partikel nicht frei vorkommt ((1)(e)), zählen dazu (vgl. Eisenberg 2004, 267f.).<sup>51</sup> Wie Eisenberg die Verben mit Stämmen auf *ier* und die Doppelpräfigierungen einordnet, ist unklar. Sie tragen Merkmale sowohl der Präfixverben (das Partizip II wird ohne *ge* gebildet) als auch der Partikelverben (*zu* tritt beim *zu*-Infinitiv zwischen Verbpartikel und Stamm; zweigeteilte Form etwa bei *ich radiere aus*). Vermutlich würde er sie allerdings eher zu den Partikelverben zählen, denn *ge* fehlt im Partizip II wahrscheinlich lediglich aus prosodischen Gründen.

Ob Eisenbergs Einteilung oder die von Fleischer/Barz nützlicher für die Analyse ist, ist kaum zu sagen. Daher sollen beide Ansätze als Attribute realisiert werden. Um die Struktur des Basisverbs zu beschreiben, führen wir das Attribut *Aufbau\_Basis* ein. Die möglichen Werte für dieses Attribut orientieren sich an Fleischer/Barz' Systematik: *trennbar* für trennbare Verben, *untrennbar* für untrennbare Verben, *Kompositum* für Komposita, *infV* für alle anderen Verben (das können auch Verben mit einem Derivationssuffix wie *ier* sein) und - für den Fall, dass das *bar*-Adjektiv nicht von einem Verb abgeleitet ist. Die doppelförmigen Verben sollen nicht gesondert behandelt werden; das wäre auch nicht sinnvoll, denn die Kategorie der doppelförmigen Verben ist eine Oberkategorie. Sollte ein doppelförmiges Verb als Derivationsbasis auftreten, so soll es wie ein trennbares beziehungsweise untrennbares Verb behandelt werden. Es lässt sich im Allgemeinen ohne Probleme sagen, ob das *bar*-Adjektiv von der trennbaren oder der untrennbaren Variante des Verbs abgeleitet ist. Zusätzlich führen wir das Attribut *Basis\_Partikelverb* mit den Werten *ja*, *nein* und - ein, um Eisenbergs Partikelverben zu beschreiben. - steht wiederum, wenn das *bar*-Adjektiv nicht von einem Verb abgeleitet ist. Mit den Attributen *Aufbau\_Basis* und *Basis\_Partikelverb* sind beide Beschreibungsmodelle gut erfasst.

Weiterhin führen wir das Attribut *Praefix\_Basis* ein. Dieses Attribut soll alle Präfixe in Fleischer/Barz' und Eisenbergs Verwendungsweise des Begriffs aufnehmen, außerdem Eisenbergs Verbpartikeln. Daraus ergibt sich, dass auch Formen wie *voraus*, *fest*, *wahr*, *hinter*, *voll* und *wieder*, die Fleischer/Barz als Kompositionsglieder bezeichnen, in diesem Attribut repräsentiert werden sollen. Es werden hier also auch Formen erfasst, die weder Eisenberg noch

---

<sup>51</sup> Die Beispiele in (1) sind aus Eisenberg 2004.

Fleischer/Barz zu den Präfixen zählen. Daher ist der Name *Praefix\_Basis* etwas irreführend. Diese Ungenauigkeit kann man aber im Interesse einer besseren Lesbarkeit in Kauf nehmen.<sup>52</sup>

Der Wertebereich des Attributs *Praefix\_Basis* ist potenziell unendlich, da die *bar*-Adjektive eine offene Klasse sind und somit zumindest von Partikelverben ständig neue *bar*-Adjektive abgeleitet werden können.

Zusätzlich zu den Bestandteilen, die dem einfachen Stamm vorausgehen, sollen auch solche, die dem Stamm folgen, in der Analyse berücksichtigt werden. Darunter sind Derivationsuffixe wie *ier* in *definierbar* zu verstehen, aber auch Formen, die keine eigenen Morpheme darstellen wie die bereits besprochenen Pseudosuffixe (etwa *el* in *handelbar*). Einige dieser Suffixe und Pseudosuffixe beeinflussen die Betonungsstruktur – *el* ist nicht betonbar, *ier* trägt dagegen stets einen Akzent, wenn auch nicht immer den Wortakzent (vgl. Wiese 1996, 287). Daher könnten sie durchaus die Präfigierbarkeit des Adjektivs mit *un* beeinflussen. Wir führen dafür das Attribut *Stammende* ein. Dieses Attribut kann Werte wie *ier*, *el*, *en*, *er*, *ig* und *lich* annehmen. Es sind sicher noch weitere Werte möglich. Ob der Wertebereich endlich ist, kann hier nicht mit letzter Gewissheit geklärt werden. Es ist aber für diese Untersuchung auch nicht von Interesse. Dass hier Suffixe und Pseudosuffixe gemeinsam repräsentiert werden, mag zunächst methodisch unsauber erscheinen. Die Pseudosuffixe verhalten sich jedoch in vielerlei Hinsicht wie Suffixe. Das wird deutlich, wenn man Homonyme wie *er* in *wunderbar* (Pseudosuffix) und *er* in *verlängerbar* (Derivationsuffix) betrachtet. Diese Homonyme unterscheiden sich lediglich morphologisch, haben aber die gleichen prosodischen Eigenschaften – beide sind nicht betonbar. Es ist daher gerechtfertigt, sie gemeinsam in einem Attribut zu repräsentieren.

In einigen Fällen wird die Stammendung der Basis nicht an das *bar*-Adjektiv weitergegeben, sondern fällt bei der Derivation weg. Das kommt zum Beispiel bei *ig* vor (*entschuldigen*, aber *entschuldbar*; vgl. Riehemann 1993, 10). Man muss aber festhalten, dass *ig* in manchen Fällen auch erhalten bleibt, etwa bei *vernachlässigbar*. Ein anderes Beispiel ist *n*, das bei Derivation von *berechnen* zu *berechenbar* getilgt wird (vgl. Fleischer/Barz 1995, 252). Um solche Fälle zu beschreiben, führen wir das Attribut *fallengelassene\_Bestandteile* ein. Es kann beispielsweise die Werte *ig* und *n* annehmen. Auch hier ist der Wertebereich möglicherweise unendlich. Dazu sollen in dieser Arbeit aber keine weiteren Betrachtungen angestellt werden. *Bar*-Adjektive können von Verben, Substantiven und Adjektiven abgeleitet sein. Die Wortart der Basis stellt ein wesentliches Kriterium zur Charakterisierung der *bar*-Adjektive dar, denn

---

<sup>52</sup> Bei Mehrfachpräfigierungen wird *Praefix\_Basis* stets das erste Präfix zugewiesen. Bei *wiederverwendbar* wäre das beispielsweise *wieder*. Ideal wäre natürlich, alle Präfixe zu berücksichtigen. Für die Analyse ist das aber schwer umsetzbar, da man in diesem Fall für jedes neue Präfix ein separates Attribut bräuchte.

die deverbalen *bar*-Adjektive unterscheiden sich in vielerlei Hinsicht von den älteren desubstantivischen und deadjektivischen Bildungen. Produktiv ist heute lediglich die Ableitung von Verben, während die Verhältnisse in mittelhochdeutscher Zeit umgekehrt lagen: Es gab im Mittelhochdeutschen 118 desubstantivische, aber nur 20 deverbale Neubildungen.<sup>53</sup> Erst nach dem 13. Jahrhundert sank die Produktivität der Substantivableitungen, während die Produktivität der Verbalableitungen stieg (vgl. Flury 1964, 30ff.). Die Gründe für diese Veränderungen sind vielfältig und können hier nur auszugsweise wiedergegeben werden.<sup>54</sup>

Das Suffix *bâri* entfernte sich schon in althochdeutscher Zeit von seiner ursprünglichen Bedeutung *fähig hervorzubringen*. Diese Bedeutungsabschwächung führte zu reger Produktivität im Mittelhochdeutschen. Das mittelhochdeutsche Suffix *baere* zeigt nur noch an, dass es sich bei dem Wort um ein Adjektiv handelt, hat darüber hinaus aber kaum noch eine eigenständige Bedeutung. Es war daher – was Dichter beim Reimen zu schätzen wussten – leicht austauschbar gegen die Suffixe *lich*, *ig*, *haft* und *sam* (vgl. Flury 1964, 35). Besonders oft wurde *bar* mit *lich* vertauscht. Nach dem 13. Jahrhundert begann *lich*, sich auf Substantivableitungen festzulegen, während *bar* sich zunehmend auf Verbalableitungen beschränkte.<sup>55</sup> Zudem gab es mehr und mehr Neubildungen auf *bar* mit passivischer Bedeutung. Im Neuhochdeutschen überwiegen die passivischen Neubildungen deutlich. Die älteren Substantivableitungen tragen zum Teil heute noch die Bedeutungskomponente *fähig hervorzubringen* in sich (*dankbar*, *fruchtbar*). Substantiv- und Adjektivableitungen weichen bedeutungsmäßig in der Regel von den deverbalen Bildungen ab. Gerade die älteren deadjektivischen und desubstantivischen *bar*-Adjektive gehören allerdings zu den am häufigsten verwendeten. Flury gibt dafür eine plausible Erklärung: Diese Adjektive werden, weil sie nicht dem produktiven Muster entsprechen, nicht mehr als Derivationen wahrgenommen, sondern als einfache Adjektive (vgl. Flury 1964, 95). Wir führen für die Wortart der Basis das Attribut *Derivation* mit den Werten *deverbal*, *desubstantivisch*, *deadjektivisch*, *desubst\_deverb*, *?* und *Komposition* ein. Dies bedarf einiger Erläuterungen. *Desubst\_deverb* beschreibt Bildungen, die weder der einen noch der anderen Derivationsart zugeordnet werden können. Sie „weisen im Nhd. in ihrer Bedeutung Strukturen auf, wie sie sowohl den Substantiv- als auch den Verbalableitungen eigen sind“ (Flury 1964, 104). Beispiele sind *dankbar* und *wunderbar*. *?* beschreibt Derivationen, bei denen die Wortart des Basiswortes nicht mehr zu ermitteln ist. Das betrifft nur zwei der unter-

---

<sup>53</sup> Deadjektivische Bildungen waren damals schon im Rückgang begriffen; das Mittelhochdeutsche verzeichnet lediglich 12 deadjektivische Neubildungen. Ein heute noch existierendes Relikt aus dem Althochdeutschen ist *offenbar*.

<sup>54</sup> Eine ausführlichere Darstellung der Entwicklung der *bar*-Adjektive bietet Flury 1964.

<sup>55</sup> Die Konkurrenz zwischen *bar*, *lich*, *ig* und *sam* besteht zum Teil auch heute noch. Diese Arbeit beschäftigt sich allerdings nicht näher damit. Der interessierte Leser sei an Flury 1964 (insbesondere 15ff.), Fleischer/Barz 1995 (insbesondere 253ff.) und Eschenlohr 1996 verwiesen.

suchten Beispiele: *strafbar* und *urbar*. *Strafbar* könnte von *bestrafen* oder von *Strafe* abgeleitet sein; *urbar* ist wahrscheinlich eine Umdeutung aus dem älteren *urbor* (vgl. Flury 1964, 58). In Zweifelsfällen gibt Flury 1964 Auskunft über die Wortart der Basis. *Komposition* beschreibt Formen wie *höhenverstellbar* und *maschinenlesbar*. Es handelt sich hierbei um Kompositionen mit *bar*-Adjektiv als Zweitglied. Dagegen lässt sich einwenden, dass es methodisch unsauber ist, Kompositionen unter dem Merkmal *Derivation* einzuordnen. Dies kann man aber im Interesse der Lesbarkeit in Kauf nehmen, da es für die Analyse keine Nachteile mit sich bringt und Kompositionen unter den untersuchten Beispielen die absolute Ausnahme darstellen.

Wie Fleischer/Barz bemerken, können deverbale *bar*-Adjektive von starken und schwachen Verben abgeleitet sein (vgl. Fleischer/Barz 1995, 252). Davon sind zusätzlich die unregelmäßigen Verben abzugrenzen.<sup>56</sup> Hentschel/Weydt unterscheiden zwei Gruppen von unregelmäßigen Verben: solche, die das Präteritum mit Ablaut und *te* bilden (*nennen* – *nannte* – *genannt*) und solche, die zusätzlich zum Ablaut eine Veränderung im Konsonantismus aufweisen (*gehen* – *ging* – *gegangen*; vgl. Hentschel/Weydt 2003, 52). Wir führen dafür das Attribut *deverbal\_stark\_schwach* mit den Werten *stark*, *schwach*, *unregelmässig* und *- ein.* *-* steht wiederum bei Adjektiven, die nicht von Verben abgeleitet sind.

### 5.3 Syntaktische Merkmale von *bar*-Adjektiven

Zu den syntaktischen Merkmalen von *bar*-Adjektiven gehören alle Eigenschaften, die mit der Verwendung der Adjektive in syntaktischen Konstruktionen zusammenhängen. Das sind beispielsweise die Positionen im Satz, die das Adjektiv einnehmen kann. Ist das *bar*-Adjektiv von einem Verb abgeleitet, soll auch betrachtet werden, ob und wie die vom Basisverb regierten Ergänzungen an das Adjektiv weitergegeben werden. Nachfolgend werden die für die Data-Mining-Analyse verwendeten syntaktischen Attribute vorgestellt. Die Werte fast aller in diesem Abschnitt eingeführten Attribute werden mithilfe des *public*-Korpus ermittelt. Auf welche Weise das *public*-Korpus dafür durchsucht wird, wird in diesem Abschnitt für jedes Attribut erläutert.

- (1) (a) *die messbare Strahlung*
- (b) *Die Strahlung, messbar, stellt kaum eine Gefährdung dar.*
- (c) *Die Strahlung ist messbar.*

---

<sup>56</sup> Die unregelmäßigen Verben werden auch oft als *gemischte Verben* bezeichnet.

(d) *Die Strahlungsintensität hat sich messbar verringert.*

(e) *Die Strahlung ist messbar gering.*

Wolfgang Motsch unterscheidet zwischen *attributiver*, *appositiver*, *prädikativer*, *adjunktiver* und *adjektivmodifizierender* Verwendungsweise von Adjektiven. Wird ein Adjektiv attributiv verwendet, so ist es Teil einer Nominalphrase wie in (1)(a). In appositiver Verwendungsweise ist es dem Nomen nachgestellt ((1)(b)). Ist das Adjektiv als Prädikatsnomen Teil des Prädikats, so spricht Motsch von prädikativer Verwendungsweise ((1)(c)). Adjunktiv wird ein Adjektiv dann verwendet, wenn es Teil der Verbalphrase ist ((1)(d)).<sup>57</sup> Adjektivmodifizierende Verwendungsweise liegt vor, wenn das Adjektiv die Bedeutung eines anderen Adjektivs oder Adverbs verändert wie in (1)(e) (vgl. Motsch 2004, 158f.). Die fünf verschiedenen Verwendungsmöglichkeiten sollen als Attribute in die Analyse integriert werden. Dabei stellt sich das Problem, dass entschieden werden muss, wie viele Belege für jede Verwendungsweise notwendig sind, um diese Verwendungsweise als möglich zu akzeptieren. Diese Frage kann nicht eindeutig beantwortet werden. Man kann nur versuchen Grenzen festzulegen, die mit dem intuitiven Verständnis eines Muttersprachlers übereinstimmen. Wir führen für die syntaktischen Verwendungsweisen das folgende Kriterium ein: Das Adjektiv muss im Korpus mindestens dreimal in der entsprechenden syntaktischen Position auftreten, damit diese syntaktische Position als möglich akzeptiert wird.<sup>58</sup> Bei weniger als drei Belegen ist die Wahrscheinlichkeit groß, dass diese Belege lediglich Abweichungen sind. Dieses Kriterium allein reicht aber nicht aus, da es im *public*-Korpus Adjektive mit 200 Belegen neben solchen mit nahezu 100000 Belegen gibt. Es ist ein großer Unterschied, ob ein Adjektiv in drei von 200 Belegen adjektivmodifizierend verwendet wird oder in drei Belegen von 100000. Wir führen deshalb ein zweites Akzeptabilitätskriterium ein: Das Adjektiv muss mindestens einmal pro 100 Belege in der entsprechenden Verwendungsweise vorkommen. Das wirft ein Problem auf: In der Korpusanalyse hat sich gezeigt, dass *bar*-Adjektive in appositiver Verwendungsweise praktisch nie einmal pro 100 Belege oder öfter auftreten und somit auch nie das oben genannte Kriterium erfüllen. Die appositive Verwendungsweise scheint überaus selten und auf spezielle Kontexte beschränkt zu sein. Diese Vermutung wird bestätigt durch die Ergebnisse von Hermann Gelhaus, der ebenfalls eine Korpusanalyse zu *bar*-Adjektiven durchgeführt hat. Gelhaus

---

<sup>57</sup> Das heißt aber nicht, dass ein adjunktiv verwendetes Adjektiv in jedem Fall das Prädikat modifiziert. Es kann sich auch auf das Subjekt oder andere Satzglieder wie Objekte beziehen. In anderen Grammatiken wird dies oft als *adverbiale Verwendungsweise* bezeichnet.

<sup>58</sup> Ist ein Adjektiv mit weniger als 200 Belegen im Korpus vertreten, können Ausnahmen von dieser Regel gemacht werden. Darüber muss im Einzelfall entschieden werden. Generell sind solche Ausnahmen jedoch problematisch, da bei nur zwei Belegen unklar ist, ob diese Belege nicht lediglich Abweichungen sind.

hat insgesamt 1205 Belege untersucht. Lediglich 23 von ihnen sind *bar*-Adjektive in appositiver Verwendungsweise (vgl. Gelhaus 1997, 283). Es lassen sich also keine genauen Aussagen treffen, welche *bar*-Adjektive appositiv verwendbar sind. Daher kann die appositive Verwendungsweise in der Analyse nicht berücksichtigt werden. Wir führen also die Attribute *praedikativ*, *attributiv*, *adjektivmodifizierend* und *adjunktiv* ein. Alle vier Attribute können die Werte *ja* und *nein* annehmen. Um herauszufinden, ob ein *bar*-Adjektiv prädikativ, attributiv, adjektivmodifizierend oder adjunktiv verwendbar ist, muss man im *public*-Korpus gezielt nach Belegen für das Adjektiv in den einzelnen syntaktischen Verwendungsweisen suchen. Für die attributive Verwendungsweise wird dabei der Suchanfrageausdruck *ADJEKTIV* benutzt, für die anderen Verwendungsweisen der Ausdruck *ADJEKTIV\**.<sup>59</sup> *ADJEKTIV* findet alle Belege für die undeklinierte Form des Adjektivs, etwa *denkbar*. *ADJEKTIV\** findet auch Belege für die deklinierten Formen, also *denkbare*, *denkbarer*, *denkbares*, *denkbaren* und *denkbarem*. Um zu ermitteln, welche syntaktische Verwendungsweise in einem bestimmten Beleg vorliegt, muss man den Kontext des Adjektivs anschauen. Für jedes Adjektiv zählt man, wie oft es in prädikativer, attributiver, adjektivmodifizierender und adjunktiver Verwendungsweise auftritt. Bei Adjektiven mit mehreren tausend Belegen wäre es natürlich ziemlich zeitaufwändig, alle Belege einzeln anzuschauen. Es genügt allerdings, 200 bis 300 Belege zu betrachten und die ermittelten Zahlen auf die Gesamtanzahl der Belege hochzurechnen.

Schwieriger gestaltet sich die Beschreibung der Weitergabe von Argumenten des Basisverbs an das derivierte Adjektiv. Oft wird in diesem Zusammenhang von *Vererbung* gesprochen. Toman unterscheidet grundsätzlich drei Arten von Vererbung (vgl. Toman 1987, 77ff.).

(2) (a) *Man kann den Jungen leicht mit seinem Bruder verwechseln.*

*Der Junge ist leicht mit seinem Bruder verwechselbar.*

(b) *Derartige Grausamkeiten können wir uns kaum vorstellen.*

*Derartige Grausamkeiten sind für uns kaum vorstellbar.*

In (2)(a) wird die Präpositionalgruppe *mit seinem Bruder* ohne Veränderung ihrer Formseite vom Verb an das *bar*-Adjektiv vererbt. Auch ihre syntaktische Funktion als präpositionale Ergänzung bleibt erhalten, sie wird lediglich vom Verbkomplement zum Komplement des abgeleiteten Adjektivs. Die Nominalgruppe *den Jungen* wird dagegen vom direkten Objekt zum Subjekt des Satzes mit dem *bar*-Adjektiv, also von der Akkusativergänzung zur Nominativergänzung. Toman spricht in beiden Fällen von *Argumentvererbung*. In (2)(b) liegt dage-

---

<sup>59</sup> *ADJEKTIV* ist hier nur ein Platzhalter und wird im konkreten Fall durch das Adjektiv ersetzt, für das man Belege finden will. *ADJEKTIV* könnte zum Beispiel *denkbar* sein, *ADJEKTIV\** etwa *denkbar\**.

gen keine Argumentvererbung vor. Hier wird die Nominativergänzung *wir* zur präpositionalen Ergänzung *für uns*. Toman nennt dies *periphrastische Vererbung* (vgl. Toman 1987, 64) und erklärt es folgendermaßen: Es gibt zwei mögliche Präpositionen für die Präpositionalgruppe: *für* und *von*.<sup>60</sup> Welche von beiden gewählt wird, hängt vom Basisverb ab. Ist das Basisverb stark intentional, wird die Präpositionalgruppe mit *von* gebildet ((3)(c)), bei schwach intentionalen Verben mit *für* ((3)(b)).<sup>61</sup> Da die Semantik des Basisverbs für die Wahl der Präposition verantwortlich ist, liegt keine Argumentvererbung vor, denn diese „betrifft nur [...] semantisch neutrale Positionen“ (Toman 1987, 80). Für die Data-Mining-Analyse ist dieser Unterschied aber nicht relevant. Daher sollen beide Vererbungsarten gemeinsam in einem Attribut beschrieben werden. Wir führen die Attribute *Valenz\_NomE*, *Valenz\_AkkE*, *Valenz\_DatE* und *Valenz\_praepE* ein.<sup>62</sup> Diese Attribute beschreiben, wie die Ergänzungen des Basisverbs an das *bar*-Adjektiv weitergegeben werden, wie sich also die Valenz des Verbs auf die Valenz des abgeleiteten Adjektivs auswirkt.

- (3) (a) *Das Material brennt.*  
*Das Material ist brennbar.*
- (b) *Die Menschen können das Ausmaß der Katastrophe kaum fassen.*  
*Das Ausmaß der Katastrophe ist für die Menschen kaum fassbar.*
- (c) *Jeder kann dieses Urteil anfechten.*  
*Dieses Urteil ist von jedem anfechtbar.*
- (d) *Die Gründe versteht wohl jeder.*  
*Die Gründe sind wohl für jeden verstehbar.*  
*Die Gründe sind wohl von jedem verstehbar.*
- (e) *Zum Telefonieren und SMS-Schreiben kann der Benutzer das Handy aufklappen.*

---

<sup>60</sup> Susann Siebert ist der Auffassung, dass nur *für* als Präposition möglich ist, nicht aber *von* (vgl. Siebert 1999, 75). Meine Untersuchung widerlegt diese Position.

<sup>61</sup> Die Untersuchung in dieser Arbeit hat ergeben, dass diese Abhängigkeit eher den Charakter einer Präferenz als den einer Regel besitzt. Zur Unterscheidung von *stark* und *schwach intentional* siehe Abschnitt 5.4.

<sup>62</sup> Diese Attribute sind angelehnt an die Ergänzungsklassen in VALBU (2004, vgl. 29ff.). Es sind allerdings nicht alle dieser Ergänzungsklassen als Attribute repräsentiert. Eigentlich müssten wir zusätzlich das Attribut *Valenz\_GenE* einführen. Darauf kann aber verzichtet werden, da Verben mit Genitivobjekten äußerst selten sind. Das bestätigt auch Toman (1987, vgl. 79). Auch die Adverbativergänzungen (AdvE) können vernachlässigt werden, da ihre Abgrenzung von den präpositionalen Ergänzungen ohnehin problematisch ist. Ebenso fehlen die Prädikativergänzungen (PrädE) und die Verbativergänzungen (VerbE) in der vorliegenden Analyse. Prädikativergänzungen sind Prädikatsnomen; und Verben mit Prädikatsnomen bilden höchstwahrscheinlich keine *bar*-Adjektive. Verbativergänzungen sind selten und können als satzförmige Akkusativergänzungen interpretiert werden, so dass sie unter dem Attribut *Valenz\_AkkE* subsumiert werden sollen.

- \* *Zum Telefonieren und SMS-Schreiben ist das Handy für den Benutzer aufklappbar.*
- \* *Zum Telefonieren und SMS-Schreiben ist das Handy von dem Benutzer aufklappbar.*

*Valenz\_NomE* kann die Werte *vererbbar*, *vererbbar\_fuer*, *vererbbar\_von*, *vererbbar\_fuer\_von*, *nicht\_vererbbar*, *?* und *-* annehmen. *Vererbbar* bedeutet, dass die Ergänzung im Nominativ vom Basisverb direkt an das *bar*-Adjektiv weitergegeben wird. Das Subjekt im Satz mit dem zugrunde liegenden Verb ist also auch Subjekt im Satz mit dem derivierten Adjektiv ((3)(a)). Der Wert *vererbbar\_fuer* wird einem Adjektiv zugewiesen, wenn das Subjekt im Satz mit dem Basisverb im Satz mit dem Adjektiv als Präpositionalgruppe mit *für* auftreten kann ((3)(b)).<sup>63</sup> Es ist wichtig festzuhalten, dass eine solche Präpositionalgruppe nicht obligatorisch ist; es handelt sich hierbei also um eine Angabe. Das Gleiche gilt für *vererbbar\_von* ((3)(c)). (3)(d) ist ein Beispiel für *vererbbar\_fuer\_von*. Hier kann die Nominativ-Ergänzung des Verbs sowohl als *für*- als auch als *von*-Phrase an das *bar*-Adjektiv weitergegeben werden. *Vererbbar\_fuer*, *vererbbar\_von* und *vererbbar\_fuer\_von* repräsentieren die periphrastische Vererbung. Ist das Subjekt nicht vererbbar, nimmt *Valenz\_NomE* den Wert *nicht\_vererbbar* an. Ein Beispiel dafür ist (3)(e). *-* bedeutet, dass die Ergänzung im Nominativ nicht vorhanden ist und folglich auch nicht vererbt werden kann. Das ist beispielsweise bei deadjektivischen und desubstantivischen *bar*-Adjektiven der Fall. In welchen Fällen *Valenz\_NomE* der Wert *?* zugewiesen wird, werden wir gleich noch sehen.

*Valenz\_AkkE*, *Valenz\_DatE* und *Valenz\_praepE* können jeweils die Werte *vererbbar*, *vererbbar\_NomE*, *nicht\_vererbbar* und *-* annehmen. *Vererbbar*, *nicht\_vererbbar* und *-* haben die gleiche Bedeutung wie oben. Einer näheren Erläuterung bedarf lediglich *vererbbar\_NomE*. Dieser Wert wird einem Attribut zugewiesen, wenn die entsprechende Ergänzung des Basisverbs im Satz mit dem Adjektiv zum Subjekt wird. Meistens erhält das Attribut *Valenz\_AkkE* diesen Wert, da die meisten *bar*-Adjektive von transitiven Verben abgeleitet sind und daher in der Regel das Akkusativ-Objekt des Verbs zum Subjekt im Satz mit dem Adjektiv wird.

Auch hier stellt sich wieder die Frage, wie man feststellen kann, ob die entsprechenden Vererbungsweisen für ein bestimmtes *bar*-Adjektiv existieren. Bei der Argumentvererbung ist

---

<sup>63</sup> Wie Riehemann bemerkt, kann die *für*-Phrase auch angeben, aus welcher Perspektive betrachtet die bezeichnete Eigenschaft zutrifft. Als Beispiel nennt sie *für den Präsidenten unantastbar*. Die Bedeutung der *für*-Phrase kann hier der in *für den Präsidenten heilig* entsprechen (vgl. Riehemann 1993, 21f.). Die *für*-Phrase drückt also nicht in jedem Fall das Subjekt im Satz mit dem Basisverb aus. Meistens sind die fraglichen Strukturen ambig: Die *für*-Phrase in *für den Präsidenten unantastbar* kann sowohl als Realisierung des Agens als auch als Perspektivierung gelesen werden. Wir legen daher fest: Wenn die *für*-Phrase auch oder nur als Agens gelesen werden kann, wird dem Attribut *Valenz\_NomE* der Wert *vererbbar\_fuer* zugewiesen.



das relativ einfach zu entscheiden: Man betrachtet die Valenz des Basisverbs. Bei transitiven Verben wird, wie schon gesagt, das direkte Objekt durch die Derivation zum Subjekt. Wir können in diesem Fall also *Valenz\_AkkE* den Wert *vererbbar\_NomE* zuweisen. Andere Objekte werden meist direkt vererbt; das entsprechende Attribut erhält also den Wert *vererbbar*. Ob ein Objekt vererbt werden kann, lässt sich im Allgemeinen zuverlässig feststellen: Wenn ein Verb außer dem Akkusativ-Objekt noch weitere Objekte hat, sind sie in der Regel auch vererbbar.<sup>64</sup> In den äußerst seltenen Zweifelsfällen muss im *public*-Korpus nach Belegen für die Vererbung gesucht werden. Manchmal existiert ein Verb auch in mehreren Varianten mit verschiedenen Valenzen und es ist nicht eindeutig, von welcher Variante das dazugehörige *bar*-Adjektiv abgeleitet ist. Dann weiß man natürlich auch nicht, welche Argumente für die Vererbung infrage kommen. In solchen Fällen bringt VALBU 2004 Klarheit. Dort ist verzeichnet, welche Verben mit welchen Valenzen *bar*-Adjektive bilden können.

Bei der periphrastischen Vererbung ist es schwieriger, ein Akzeptabilitätskriterium festzulegen. Das Hauptproblem ist, dass *für*- beziehungsweise *von*-Phrasen relativ selten sind. Für *bar*-Adjektive mit wenigen Belegen lässt sich daher kaum eine verlässliche Aussage treffen, ob periphrastische Vererbung möglich ist oder nicht. Daher erhält bei Adjektiven mit weniger als 200 Belegen das Attribut *Valenz\_NomE* den Wert ?. Für Adjektive mit mindestens 200 Belegen führen wir zwei Kriterien ein: Es müssen mindestens drei Belege für *für*- beziehungsweise *von*-Phrasen existieren;<sup>65</sup> außerdem muss die entsprechende Phrase mindestens einmal pro 2000 Belege auftreten. Um herauszufinden, wie oft ein *bar*-Adjektiv im *public*-Korpus mit einer *für*- oder *von*-Phrase auftritt, werden die Suchanfrageausdrücke *für* /+w3 ADJEKTIV\* und *von* /+w3 ADJEKTIV\* verwendet. Wie schon in Abschnitt 4.2 gesagt, bedeutet /+w3, dass zwischen der Präposition und dem Adjektiv null bis zwei Wörter stehen können. Setzt man für ADJEKTIV *bezahlbar* ein, finden diese Ausdrücke Belege wie *für jedermann bezahlbar*, *zu für alle bezahlbaren Preisen* und *von allen bezahlbar*.

Mit den Attributen *Valenz\_NomE*, *Valenz\_AkkE*, *Valenz\_DatE* und *Valenz\_praepE* sind indirekt auch andere Fragestellungen erfasst – beispielsweise ob das Adjektiv von einem transitiven oder von einem intransitiven Verb abgeleitet ist: Hat das Attribut *Valenz\_AkkE* den Wert *vererbbar\_NomE*, so ist das Adjektiv von einem transitiven Verb abgeleitet und gehört damit dem produktiven Muster an. Andernfalls ist es von einem intransitiven Verb abgeleitet, sofern zusätzlich das Attribut *Derivation* den Wert *deverbal* hat. Es gehören allerdings auch einige Adjektive, bei denen das Attribut *Derivation* den Wert *desubst\_deverb* hat, in diese Gruppe. Diese Adjektive verhalten sich in vielen Fällen wie deverbale Adjektive.

---

<sup>64</sup> Für Dativobjekte und präpositionale Objekte bestätigt dies auch Toman (1987, vgl. 79).

<sup>65</sup> Ausnahmen sind auch hier bei Adjektiven mit wenigen Belegen möglich.

#### 5.4 Semantische Merkmale von *bar*-Adjektiven

Die semantischen Merkmale lassen sich von allen Eigenschaften der *bar*-Adjektive am schwersten systematisieren und in Attribute fassen. Ganz gleich, welche Kategorisierungen man ansetzt – es wird immer Adjektive geben, die sich mithilfe dieser Kategorisierungen besser beschreiben lassen als andere. Um die Bedeutungen der einzelnen *bar*-Adjektive so gut wie möglich zu erfassen, muss man daher versuchen, möglichst viele unterschiedliche Beschreibungsmöglichkeiten für die Semantik der *bar*-Adjektive zu finden. Es kommt dabei vor, dass sich die Kategorisierungen ähneln oder überschneiden. Das ist aber bestenfalls ein theoretisches Problem und stört bei der Analyse nicht. Wie schon in Abschnitt 4.3 skizziert, kommt bei einigen semantischen Merkmalen noch ein weiteres Problem dazu: Die Werte einiger Attribute lassen sich nicht direkt durch Betrachten der Formseite des Adjektivs oder durch die Korpusanalyse ermitteln.<sup>66</sup> Man ist hier ganz auf Introspektion angewiesen. Diese Methode birgt natürlich die Gefahr mangelnder Objektivität. Dem kann man begegnen, indem man möglichst viele Belege für Verwendungsweisen des Adjektivs betrachtet und so versucht herauszufinden, ob das fragliche Attribut eher den einen oder den anderen Wert annimmt. Diese Verfahrensweise ist nicht unproblematisch, aber sie ist die einzige Alternative, denn diese Attribute bezeichnen keine „harten“, messbaren Eigenschaften.

Im Folgenden sollen die für die Data-Mining-Analyse verwendeten semantischen Attribute eingeführt werden. Einige der Attribute beziehen sich auch hier wieder auf die Derivationsbasis, andere auf das Derivat – das *bar*-Adjektiv.

Wolfgang Motsch beschreibt Wortbildungsprozesse mithilfe von *Wortbildungsmustern*. Ein Wortbildungsmuster besteht aus einem *semantischen Muster* und einer konkreten phonologisch-morphologischen Realisierung dieses Musters (vgl. Motsch 2004, 1). Ein semantisches Muster kann auf verschiedene Weisen realisiert werden. Das lässt sich am besten am konkreten Beispiel illustrieren: Das semantische Muster *die Fähigkeit, das durch V<sup>67</sup> bezeichnete Geschehen zu vollziehen, ist eine Eigenschaft von x*<sup>68</sup> kann durch zwei verschiedene Wortbildungsmuster realisiert werden. Die erste Möglichkeit besteht darin, an einen Verbstamm das Suffix *bar* anzuhängen (*brennbar*). Die zweite Möglichkeit ist die Komposition eines verbalen Erstglieds mit dem adjektivischen Zweitglied *fähig* (*lernfähig*). Die entstehenden Wort-

---

<sup>66</sup> Das betrifft die Attribute *Intentionalitaet*, *idiosynkratisch*, *Bereich\_Herkunft*, *deverbal\_Verbart* und *modaler Nebensinn*, die alle auf den folgenden Seiten eingeführt werden.

<sup>67</sup> V bedeutet *Verb*.

<sup>68</sup> Die Variable *x* bezeichnet die Argumentstelle für das Bezugswort – etwa *Verbindung* in *explodierbare Verbindung*.

bildungsprodukte haben aktivische Bedeutung – ein *brennbarer Stoff* ist ein *Stoff, der brennen kann*.

*Bar*-Derivate haben außerdem Anteil an einem weiteren semantischen Muster: *von einem Geschehen V betroffen sein zu können, ist eine Eigenschaft von x*. Nach diesem Muster erzeugte Bildungen haben passivische Bedeutung – *hörbarer Lärm* ist *Lärm, der gehört werden kann* (vgl. Motsch 2004, 296ff.).

Motschs Einteilung bestätigt also, was wir schon in Abschnitt 3.1 festgestellt haben: *Bar*-Adjektive können sowohl aktivische als auch passivische Bedeutung haben. Dies repräsentieren wir durch das Attribut *Motsch* mit den möglichen Werten *faehig\_agens*, *faehig\_patiens* und *-*.<sup>69</sup> - steht, falls das Adjektiv weder aktivische noch passivische Bedeutung hat, also beispielsweise bei desubstantivischen *bar*-Adjektiven. Die Attributwerte lassen sich ermitteln, indem man die Derivationsbasis, also das zugrundeliegende Verb, betrachtet.

Kühnhold, Putzer und Wellmann kategorisieren anders. Sie betrachten die Morphologie und Syntax der Derivationsbasis sowie die Semantik des entstehenden *bar*-Adjektivs. Es ergeben sich sieben Klassen von *bar*-Adjektiven – drei deverbale und vier desubstantivische. Die siebte Klasse stellt eine Restklasse dar. Deadjektivische *bar*-Adjektive berücksichtigen sie nicht.

(1)

Klasse	Beispiel	Umschreibung	Basiswortart
<i>-bar</i> <sup>1</sup>	<i>heilbare Krankheit</i>	<i>Krankheit, die geheilt werden kann</i>	Verb
<i>-bar</i> <sup>2</sup>	<i>brennbares Material</i>	<i>Material, das brennen kann</i>	Verb
<i>-bar</i> <sup>3</sup>	<i>ehrbare Frau</i>	<i>Frau, die Ehre hat</i>	Substantiv
<i>-bar</i> <sup>4</sup>	<i>unentrinnbare Verlockung</i>	<i>Verlockung, der man nicht entinnen kann</i>	Verb
<i>-bar</i> <sup>5</sup>	<i>fruchtbarer Acker</i>	<i>Acker, der Frucht bringt</i>	Substantiv
<i>-bar</i> <sup>6</sup>	<i>jagdbares Tier</i>	<i>Tier, das gejagt werden kann</i>	Substantiv
<i>-bar</i> <sup>r</sup>			Substantiv

(vgl. Kühnhold/Putzer/Wellmann 1978, 106)

Die einzelnen Klassen sind in (1) dargestellt. Wir führen dafür das Attribut *Innsbruck*<sup>70</sup> mit den Werten *1, 2, 3, 4, 5, 6, R* und *- ein*.<sup>71</sup> Die Ziffern entsprechen den Klassen bei Kühnhold, Putzer und Wellmann; *-* steht bei Adjektiven, die in keine der Klassen passen. Das ist bei-

<sup>69</sup> Bereits in Abschnitt 3.1 wurde festgestellt, dass *bar*-Adjektive mit aktivischer Bedeutung von intransitiven Verben abgeleitet sind. Für die Attributwerte heißt das: *Motsch* = *faehig\_agens* korreliert bei deverbale *bar*-Adjektiven mit *Valenz\_AkkE* = *-*.

<sup>70</sup> Das Attribut heißt *Innsbruck*, weil Kühnhold, Putzer und Wellmann zur Innsbrucker Forschungsstelle des Instituts für deutsche Sprache gehören.

<sup>71</sup> Die vierte Klasse spielt für die Analyse allerdings keine Rolle, da sie nur wenige Adjektive beinhaltet, die – mit Ausnahme von *erinnerbar* – sämtlich *un*-Bildungen sind.

spielsweise bei deadjektivischen *bar*-Adjektiven der Fall. Welcher Klasse ein *bar*-Adjektiv angehört, lässt sich durch Nachschlagen in Kühnhold/Putzer/Wellmann 1978 herausfinden.

- (2) (a) *Lies dieses Buch!*  
(b) *\*Verstehe diese Aufgabenstellung!*  
(c) *Versuche, diese Aufgabenstellung zu verstehen!*  
(d) *\*Erwarte diesen Ausgang der Geschichte!*  
(e) *\*Versuche, diesen Ausgang der Geschichte zu erwarten!*  
(f) *Dieser Ausgang der Geschichte war erwartbar.*

Ein weiteres wichtiges Kriterium ist das der *Intentionalität*. Jindřich Toman unterscheidet *stark intentionale* Verben von *schwach intentionalen* und *nicht-intentionalen* (vgl. Toman 1987, 71ff.). Stark intentional ist ein Verb, wenn man den Imperativ bilden kann.<sup>72</sup> In (2)(a) ist das möglich, in (2)(b) nicht. *Lesen* ist also – im Gegensatz zu *verstehen* – ein stark intentionales Verb. Schwach intentionale Verben haben keinen Imperativ, können aber in so genannten „Anstrengungskontexten“ (Toman 1987, 72) auftreten, zum Beispiel in einem Satz mit *versuchen* ((2)(c)). *Verstehen* ist also schwach intentional. Verben, die weder einen Imperativ haben, noch in Anstrengungskontexten auftreten können, sind nicht-intentional. Ein Beispiel dafür ist *erwarten* ((2)(d), (2)(e)). Laut Toman sind *bar*-Adjektive von stark und schwach intentionalen, nicht aber von nicht-intentionalen Verben ableitbar (vgl. Toman 1987, 73). Ich schließe mich dieser Auffassung nicht an. *Erwartbar* ((2)(f)) ist ein Gegenbeispiel zu Tomans Theorie, und es ist kein Einzelfall. In dieser Untersuchung treten noch weitere *bar*-Adjektive auf, die von nicht-intentionalen Verben abgeleitet sind (beispielsweise *verwechselbar* und *erahnbare*). Wir führen daher das Attribut *Intentionalität* mit den Werten *stark*, *schwach*, *nicht\_intentional* und - ein. - steht wieder bei nicht von Verben abgeleiteten *bar*-Adjektiven.

Mit dem Intentionalitätskriterium eng verwandt ist die Frage, ob das *bar*-Adjektiv von einem Vorgangs-, Zustands- oder Handlungsverb abgeleitet ist. Flury bringt Vorgangs- und Zustandsverben mit aktivischen, Handlungsverben mit passivischen *bar*-Ableitungen in Verbindung (vgl. Flury 1964, 101ff.). Motsch ist der Ansicht, dass für das passivische Wortbildungsmuster „Verben mit Akkusativobjekt, die keine Handlungen bezeichnen, [...] als Basiswörter ausgeschlossen“ (Motsch 2004, 300) sind. Auch diese Auffassung kann durch Beispi-

---

<sup>72</sup> Toman nennt als Eigenschaft der stark intentionalen Verben außerdem die Verwendbarkeit mit *absichtlich* (Toman 1987, 71). Dieses Kriterium scheint aber kaum geeignet, da Sätze wie *?Karl hat das Buch absichtlich gelesen* fragwürdig sind.

le wie *brauchbar* und *erwartbar* widerlegt werden. Wir führen das Attribut *deverbal\_Verbart* mit den Werten *Vorgang*, *Zustand*, *Handlung* und - ein.

Die Attribute *Intentionalitaet* und *deverbal\_Verbart* ähneln sich zwar, bezeichnen aber nicht dasselbe. Der Wert *stark\_intentional* entspricht fast 1:1 dem Wert *Handlung*; bei den anderen Werten existiert allerdings keine solche Entsprechung. Schwach intentionale Verben können Vorgangs-, Zustands- und Handlungsverben entsprechen; nicht-intentionale Verben entsprechen Zustands- oder Vorgangsverben. Man muss jedoch festhalten, dass die Zuordnung zu diesen Kategorien in manchen Fällen sehr schwierig ist. Es sind keine „harten“ Kategorien. Das trifft, wie bereits gesagt, auf die meisten semantischen Merkmale zu.

- (3) (a) *Das ESP ist in diesem Auto zuschaltbar.*
- (b) *Karl ist beeinflussbar.*
- (c) *\*Das ESP ist in diesem Auto zuschaltbarer als in anderen Autos.*
- (d) *Karl ist beeinflussbarer als Renate.*
- (e) *Karl ist leicht beeinflussbar.*

Eine weitere semantische Subklassifizierung ist die Unterteilung in *relative* und *absolute* Adjektive. Diese Begriffe sollen hier so verstanden werden wie sie etwa Hentschel/Weydt (2003, vgl. 201) oder Motsch (2004, vgl. 165) verwenden. Absolute Adjektive bezeichnen demnach Eigenschaften, die nicht auf einem Vergleich mit den Eigenschaften anderer Entitäten beruhen. Ein Beispiel ist (3)(a). Relative Adjektive bezeichnen dagegen Eigenschaften, die Entitäten im Vergleich zu anderen Entitäten zugeschrieben werden ((3)(b)). Sie sind daher – im Gegensatz zu den absoluten Adjektiven – gradierbar: Jemand kann beeinflussbarer sein als jemand anderes, aber etwas kann nicht zuschaltbarer sein als etwas anderes. Deshalb ist (3)(d) grammatisch, (3)(c) jedoch nicht. Wolfgang Motsch nennt zwei Möglichkeiten der Gradierung für die relativen Adjektive. Zum einen bilden sie einen Komparativ<sup>73</sup> ((3)(d)) und einen Superlativ. Zum anderen können sie nach Fokuspartikeln stehen wie in (3)(e) (vgl. Motsch 2004, 165).<sup>74</sup> Wir führen dafür das Attribut *Gradierbarkeit* mit den Werten *ja* und *nein* ein. Um zu testen, ob ein *bar*-Adjektiv gradierbar ist, muss man nach Belegen suchen, in denen das Adjektiv im Komparativ oder mit den Fokuspartikeln *sehr*, *leicht*, *schwer* oder *kaum* auf-

---

<sup>73</sup> Susann Siebert behauptet, *bar*-Adjektive seien „im Gegensatz zu den adjektivischen Simplizia nicht komparierbar“ (Siebert 1999, 81). Dies widerspricht der gängigen Auffassung in der Forschung (vgl. dazu Šved 1977, 276; Lenz 1995, 21, 28). Auch meine Untersuchung zeigt, dass Sieberts Behauptung nicht zu halten ist.

<sup>74</sup> Laut Siebert können *bar*-Adjektive nicht mit den Gradpartikeln *sehr* und *ziemlich* auftreten (vgl. Siebert 1999, 78f.). Dem widerspricht die korpuslinguistische Analyse: *Sehr achtbar* ist beispielsweise im *public*-Korpus mit 51 Belegen vertreten, *sehr dehnbar* mit 28 Belegen und *sehr belastbar* mit 41 Belegen.

tritt. Wir führen dafür die folgenden Kriterien ein: Es müssen mindestens drei Belege mit einer der genannten Gradierungsmöglichkeiten existieren, damit das Adjektiv als gradierbar gilt. Außerdem muss mindestens ein Beleg mit Gradierung pro 100 Belege vorkommen. Ist ein *bar*-Adjektiv mit 3000 Belegen im *public*-Korpus vertreten, muss es also in 30 Belegen mit Gradierung auftreten. Wie viele Belege davon jeweils auf den Komparativ und auf Vorkommen mit Gradpartikeln entfallen, ist unerheblich. Diese Festlegungen stimmen in etwa mit der intuitiven Vorstellung von Gradierbarkeit überein.

Um die Belege zu ermitteln, soll für den Komparativ der Suchanfrageausdruck *ADJEKTIVer\** benutzt werden. Dieser Ausdruck findet etwa *angreifbare*, *angreifbarer*<sup>75</sup>, *angreifbarer*, *angreifbarer*, *angreifbarer* und *angreifbarer*. Für die Vorkommen mit Gradpartikeln werden die Ausdrücke *sehr* /+w1 *ADJEKTIV\**, *leicht* /+w1 *ADJEKTIV\**, *schwer* /+w1 *ADJEKTIV\** und *kaum* /+w1 *ADJEKTIV\** benutzt. Das *bar*-Adjektiv muss also direkt auf die Gradpartikel folgen.

Nicht alle *bar*-Adjektive sind transparent. Wie alle lexikalisierten Wortbildungsprodukte können auch lexikalisierte *bar*-Adjektive idiosynkratische Eigenschaften aufweisen.<sup>76</sup> Das betrifft vor allem lexikalisierte *bar*-Adjektive, die nicht nach dem produktiven Muster gebildet sind<sup>77</sup>; es bedeutet aber nicht, dass diese Adjektive völlig unregelmäßig sind. Diese Bildungen folgen durchaus Regeln, die aber zum Teil „im Dunkeln liegen“ (Motsch 1988, 2) und „ihren Musterstatus verloren [haben], d.h., entweder der semantisch-syntaktische Teil oder der phonologisch-morphologische Teil oder beide Teile des Musters sind defekt“ (Motsch 2004, 4). Im Unterschied zu Motsch soll der Idiosynkrasiebegriff hier stärker auf die Semantik bezogen werden. Es werden alle *bar*-Adjektive als idiosynkratisch bezeichnet, deren Bedeutung von der Grundbedeutung des produktiven Musters abweicht. Diese Grundbedeutung lässt sich etwa folgendermaßen beschreiben: *Es ist möglich, an der vom Bezugswort bezeichneten Entität die vom Basisverb bezeichnete Tätigkeit*<sup>78</sup> *zu vollziehen*. Dazu einige Beispiele: Ein *überprüfbarer Vorgang* ist ein *Vorgang, an dem man die Tätigkeit des Überprüfens vollziehen kann*; ein *schwenkbarer Roboterarm* ist ein *Roboterarm, an dem man die Tätigkeit des Schwenkens vollziehen kann*. Wir führen dafür das Attribut *idiosynkratisch* mit den Werten *ja* und *nein* ein. Als idiosynkratisch sollen also alle *bar*-Adjektive bezeichnet werden, die nicht spontan von einem kompetenten Sprecher des Deutschen produziert und rezipiert werden können. Aufgrund ihrer vom Grundmuster abweichenden Eigenschaften müssen sie im Ge-

---

<sup>75</sup> *Angreifbarer* kann auch die Genitivform im Positiv sein. Um sie vom Komparativ abzugrenzen, muss man den Kontext des Adjektivs betrachten.

<sup>76</sup> Eine Definition für den Ausdruck *idiosynkratische Eigenschaften* findet sich in Abschnitt 3.2.

<sup>77</sup> Wie bereits gesagt, zählen dazu beispielsweise desubstantivische *bar*-Adjektive wie *furchtbar* und *fruchtbar*.

<sup>78</sup> Je nach Art des Basisverbs kommen neben Tätigkeiten auch Zustände und Vorgänge vor.

dächtnis des Sprechers gespeichert und ins Lexikon der Sprache aufgenommen werden.<sup>79</sup> Dazu gehören auch *bar*-Adjektive, die neben der oben genannten Grundbedeutung zusätzliche Bedeutungskomponenten haben. Diese Adjektive können nach dem produktiven Muster gebildet worden sein, entscheidend ist die Nebenbedeutung. Ein Beispiel ist *essbar*. Wie Riehemann bemerkt, impliziert *essbar*, dass das betreffende Objekt gefahrlos gegessen werden kann (vgl. Riehemann 1993, 10). Diese Bedeutung geht über die Grundbedeutung hinaus. Flury unterscheidet in diesem Zusammenhang verschiedene modale Nebenbedeutungen: „was leicht (getan) werden kann“, „was gut (getan) werden kann“, „was wert, würdig ist (getan) zu werden“, „was (getan) werden muss“, „was [...] ohne Bedenken (getan) werden kann“, „was (getan) werden darf“, „was mit Recht (getan) werden kann“ (Flury 1964, 102f.). Wir führen für diese modalen Nebenbedeutungen das zusätzliche Attribut *modaler\_Nebensinn* ein. Die möglichen Werte orientieren sich an Flurys Systematik: *darf*, *kann*, *kann\_leicht*, *muss* und *soll*. *Kann* ist die modale Bedeutung, die schon in der Grundbedeutung vorhanden ist. Alle anderen Werte bezeichnen Idiosynkrasien.

*Bar*-Adjektive können dauerhafte und nicht dauerhafte Eigenschaften bezeichnen; das heißt, diese Eigenschaften können „einem Gegenstand oder einer Klasse von Gegenständen in allen oder in kontextuell begrenzten Situationen zukommen“ (Motsch 2004, 300).<sup>80</sup> Wir bestimmen drei Kriterien für Dauerhaftigkeit: 1. Die Eigenschaft muss der betreffenden Entität über eine längere Zeitspanne, also über eine Einzelsituation hinaus zukommen. 2. Die Eigenschaft muss Teil der Charakterisierung der Entität sein. 3. Die Zuschreibung der Eigenschaft darf nicht stark subjektiv sein, also in hohem Maße von der Einschätzung des Betrachters abhängen. Dazu zwei Beispiele: *Brauchbar* bezeichnet keine dauerhafte Eigenschaft, denn es hängt stark vom einzelnen Betrachter ab, ob etwas als brauchbar eingeschätzt wird oder nicht. *Abnehmbar* bezeichnet dagegen eine dauerhafte Eigenschaft, denn ein Gegenstand – etwa eine Anhängerkupplung – ist entweder immer abnehmbar oder nie. Wir führen dafür das Attribut *dauerhaft* mit den möglichen Werten *ja* und *nein* ein.

Das letzte semantische Kriterium, das betrachtet werden soll, beschreibt die Verwendungsmöglichkeiten eines Adjektivs in bestimmten Zusammenhängen. Die verschiedenen Verwendungsmöglichkeiten sind jedoch schwer zu systematisieren. Welche Kontexte soll man ansetzen? Soll man beispielsweise danach kategorisieren, auf welche Arten von Entitäten man mit einem Adjektiv Bezug nehmen kann? Einen solchen Ansatz verfolgt Motsch. Er schlägt folgende Kategorien vor: *physikalische Gegenstände*, *Personen*, *Artefakte*, *Geschehen* und *Ei-*

---

<sup>79</sup> Diese Wörter werden von DiSciullo/Williams als *listemes* bezeichnet (vgl. DiSciullo/Williams 1987, 3).

<sup>80</sup> Riehemann ist der Auffassung, dass alle Adjektive, also auch *bar*-Adjektive, dauerhafte Eigenschaften bezeichnen (vgl. Riehemann 1993, 7). Dieser Position schließe ich mich nicht an.

*enschaften* (vgl. Motsch 2004, 162f.). Dabei ergibt sich das Problem, dass viele Adjektive auf mehrere dieser Kategorien beziehbar sind. Es dürfte wohl schwer fallen, für jedes Adjektiv festzustellen, welche das im Einzelnen sind. Man könnte auch versuchen, danach zu kategorisieren, in welchen Bereichen des Lebens uns ein Adjektiv begegnet. Mögliche Kategorien wären dann *Politik, Justiz, Sport...* Dabei ergeben sich ähnliche Schwierigkeiten wie bei Motschs Kategorisierung. Außerdem dürfte es schwierig sein, die Kategorien erschöpfend und überschneidungsfrei festzulegen. Wir beschränken uns daher auf die Unterscheidung zwischen Alltagssprache und Fachsprache und führen dafür das Attribut *Bereich\_Herkunft* mit den Werten *Alltagssprache* und *Fachsprache* ein. Dabei soll ein Adjektiv als fachsprachlich klassifiziert werden, wenn es überwiegend in einem bestimmten Kontext auftritt oder wenn es aus einer Fachsprache stammt und die Fachsprachlichkeit noch in einer signifikanten Anzahl von Belegen erkennbar ist. Andernfalls wird es der Alltagssprache zugeordnet. Zwei Beispiele: *Einklagbar* ist ein Begriff aus der Justiz und tritt stets in diesem speziellen Kontext auf. *Urbar* stammt aus der Landwirtschaft. Heute wird es zwar auch in anderen Zusammenhängen gebraucht, aber die fachsprachliche Herkunft ist noch deutlich zu erkennen.

### 5.5 Weitere Merkmale von *bar*-Adjektiven

In diese Kategorie fallen alle Merkmale, die sich in keine der anderen Kategorien einordnen lassen. Das sind lediglich zwei Merkmale: eines, das weder phonologischer, noch morphologischer, noch syntaktischer oder semantischer Natur ist, weil es sich nicht direkt auf ein *bar*-Adjektiv bezieht, sowie das Merkmal, das angibt, ob ein *bar*-Adjektiv mit *un* negiert werden kann oder nicht. Diese beiden Eigenschaften werden im Folgenden näher erläutert.

Lenz hat festgestellt, dass es eine Reihe von Adjektiven auf *lich* gibt, die mit *un* negiert sind, zu denen aber kein Basiswort ohne *un* existiert. Interessant ist, dass zu diesen *un-lich*-Adjektiven parallele *bar*-Bildungen und parallele *un-bar*-Bildungen existieren. Beispiele sind *unauflöslich* – *\*auflöslich* – *unauflösbar* – *auflösbar* und *undurchdringlich* – *\*durchdringlich* – *undurchdringbar* – *durchdringbar*. Diese parallelen Bildungen sind wahrscheinlich Relikte aus älteren Sprachstufen. Sie stammen aus der Zeit, als *bar* und *lich* noch weitgehend synonym waren.<sup>81</sup> Insgesamt listet Lenz 42 dieser Formen auf (vgl. Lenz 1995, 100f.). Es liegt die Vermutung nahe, dass *bar*-Adjektive, zu denen es eine parallele *un-lich*-Bildung, aber keine parallele *lich*-Bildung gibt, generell mit *un* negiert werden können. Dieser Gesichtspunkt scheint also für die Analyse von Interesse zu sein. Wir führen das Attribut *un\_lich* ein,

---

<sup>81</sup> Diese Vermutung wird gestützt durch die Beobachtung, dass die fraglichen *un-lich*-Adjektive allesamt von Verben abgeleitet sind, während sich *lich* synchron nur mit Substantiven verbindet.



das den Wert *ja* annimmt, falls zu dem entsprechenden *bar*-Adjektiv eine parallele *un-lich*-Bildung, aber keine parallele *lich*-Bildung existiert. Andernfalls nimmt es den Wert *nein* an. Wie entscheidet man, ob *un\_lich* den Wert *ja* annimmt? Wir führen dafür die folgenden Kriterien ein: Das entsprechende *un-lich*-Adjektiv muss mindestens zehnmal im *public*-Korpus vorhanden sein und die Zahl der dazugehörigen *lich*-Bildungen darf nicht mehr als fünfzehn Prozent der Anzahl der *un-lich*-Bildungen betragen. Ein Beispiel: *Unvergleichlich* tritt im *public*-Korpus in 3512 Belegen auf, *vergleichlich* dagegen nur in acht; *un\_lich* nimmt daher den Wert *ja* an. Um die Belege für die *lich*-Bildungen im *public*-Korpus zu ermitteln, wird der Suchanfrageausdruck *ADJEKTIV\** (etwa *denklich\**), für die *un-lich*-Bildungen *unADJEKTIV\** (etwa *undenklich\**) benutzt.

Das letzte Attribut ist das Zielattribut für die Data-Mining-Analyse. Es gibt an, ob ein *bar*-Adjektiv mit *un* negiert werden kann oder nicht. Wir führen dafür das Attribut *class*<sup>82</sup> mit den Werten *un* und *nicht* ein.<sup>83</sup> Auch für dieses Attribut muss entschieden werden, wie viele Belege nötig sind, um *class* den Wert *un* zuzuweisen. Wir führen dafür die folgenden Kriterien ein: Das mit *un* negierte *bar*-Adjektiv muss in mindestens drei Belegen auftreten. Außerdem muss die Anzahl der Belege mindestens 15 % der entsprechenden Belege mit *nicht* betragen. *Undenkbar* ist zum Beispiel im *public*-Korpus mit 5038 Belegen vertreten, *nicht denkbar* mit 1832. Da 5038 mehr als 15 % von 1832 ist, nimmt *class* für *denkbar* den Wert *un* an.

Zur Ermittlung der Belege für die Negation mit *un* wird der Suchanfrageausdruck *unADJEKTIV\** benutzt. Dieser Ausdruck findet Belege für die mit *un* negierten *bar*-Adjektive in allen deklinierten Formen, also neben *undenkbar* auch *undenkbare*, *undenkbaren*, *undenkbarer*, *undenkbarem* und *undenkbares*. Um die Belege für die Negation mit *nicht* im *public*-Korpus zu ermitteln, wird der Ausdruck *nicht /+w3 ADJEKTIV\** benutzt. Das bedeutet: Zwischen *nicht* und dem entsprechenden Adjektiv dürfen null bis zwei Wörter stehen; außerdem sind alle deklinierten Formen des Adjektivs erlaubt. Dieser Ausdruck ist nicht unproblematisch. Die meisten Belege, die er im *public*-Korpus findet, entsprechen zwar dem Muster *nicht ADJEKTIV* (also etwa *nicht denkbar*); vereinzelt findet er aber auch Belege wie *Mittelfristig ist auch ein Einstieg in nicht deutschsprachige Märkte denkbar*. In diesem Beleg negiert *nicht* das ihm direkt folgende Adjektiv *deutschsprachige*, nicht das *bar*-Adjektiv. Wir haben die Probleme im Zusammenhang mit dem Skopus von *nicht* bereits in Abschnitt 3.2 besprochen. Man könnte nun argumentieren, dass für die Suchanfrage vielleicht der Ausdruck *nicht /+w1*

---

<sup>82</sup> Der Name des Zielattributs ist von dem Data-Mining-System *WEKA* vorgegeben.

<sup>83</sup> Der Wert *un* bedeutet nicht, dass das *bar*-Adjektiv *a u s s c h l i e ß l i c h* mit *un* negiert werden kann, sondern lediglich, dass es *a u c h* mit *un* negiert werden kann. Die Negation mit *nicht* ist, wie bereits gesagt, immer möglich.

*ADJEKTIV\** angemessener wäre, denn dann würden nur Belege gefunden, in denen das *bar*-Adjektiv direkt auf *nicht* folgt. Auch diese Variante bringt jedoch zwei Probleme mit sich. Zum einen wird auch hier nicht in jedem Fall das *bar*-Adjektiv negiert, beispielsweise wenn es adjektivmodifizierend gebraucht wird wie in *wenn die Arbeitslosigkeit nicht spürbar sinkt*. Zum anderen werden auf diese Weise Belege wie *nicht mehr denkbar* und *nicht fassbar und denkbar* ausgeschlossen, in denen *denkbar* negiert wird, obwohl es nicht direkt auf den Negationsträger folgt. Man muss also die Belege, in denen durch den Ausdruck *nicht /+w3 ADJEKTIV\** das „falsche“ Satzglied negiert wird, in Kauf nehmen. In der Praxis sind diese Belege allerdings kaum relevant für die Klassifizierung eines Adjektivs als *un* oder *nicht*. Der Ausdruck *nicht /+w3 ADJEKTIV\** findet im Allgemeinen nicht signifikant mehr Belege als der Ausdruck *nicht /+w1 ADJEKTIV\**. Der Unterschied zwischen den beiden Ausdrücken ist also sehr gering. Die 15%-Schwelle für die *un-bar*-Adjektive ist indessen sehr großzügig gewählt. Wenn das *un-bar*-Adjektiv nicht existiert, ist es meist gar nicht oder nur in sehr wenigen Belegen im *public*-Korpus vorhanden, so dass diese Belege nicht annähernd die 15 % erreichen.

## 6 Die Data-Mining-Analyse

In Vorbereitung auf die Data-Mining-Analyse wurden insgesamt 244 *bar*-Adjektive mithilfe des *public*-Korpus untersucht. Beim Betrachten der einzelnen Belege für die Adjektive stellte sich heraus, dass viele *bar*-Adjektive polysem sind. Bei einigen gehen die semantischen Unterschiede einher mit signifikanten Unterschieden in der Syntax.

- (1) (a) *Dieser Politiker ist für die Partei kaum noch tragbar.*  
(b) *ein tragbarer Fernseher*

Ein Beispiel dafür ist *tragbar*. In (1)(a) ist seine Bedeutung abstrakt, in (1)(b) konkret. Im Folgenden soll die abstrakte Verwendungsweise mit *tragbar\_abstrakt* und die konkrete mit *tragbar\_konkret* bezeichnet werden.<sup>84</sup>

*Tragbar\_abstrakt* und *tragbar\_konkret* unterscheiden sich in einigen wesentlichen Eigenschaften. Die korpuslinguistische Analyse hat ergeben, dass das direkte Objekt von *tragen* bei Derivation zu *tragbar\_abstrakt* als *für*- oder *von*-Phrase vererbt werden kann, bei Derivation zu *tragbar\_konkret* hingegen nur als *von*-Phrase. Darüber hinaus ist die Derivationsbasis *tragen* bei *tragbar\_abstrakt* nicht intentional, bei *tragbar\_konkret* dagegen stark intentional. Es erscheint sinnvoll, von *tragen\_abstrakt* und *tragen\_konkret* zu sprechen, denn die Unterschiede zwischen *tragbar\_abstrakt* und *tragbar\_konkret* gehen offensichtlich auf Bedeutungsunterschiede zwischen den Derivationsbasen zurück.

Es existieren noch einige weitere Unterschiede zwischen *tragbar\_abstrakt* und *tragbar\_konkret*. Der wichtigste Unterschied dürfte darin liegen, dass *tragbar\_abstrakt* mit *un* negiert werden kann, *tragbar\_konkret* hingegen nicht. In Fällen, wo sich die Verwendungsweisen derart stark voneinander unterscheiden, sollen zwei getrennte *bar*-Adjektive angenommen werden, obwohl es sich streng genommen jeweils um ein einziges handelt. Man kann aber in solchen Fällen nicht sagen, ob sich das Adjektiv als solches mit *un* negieren lässt

---

<sup>84</sup> Wolfgang Motsch unterscheidet zwischen *primärer* und *metaphorischer* Bedeutung von Adjektiven, wobei metaphorische Bedeutungen durch Umdeutung von primären Bedeutungen entstehen (vgl. Motsch 2004, 163f.). Ich schließe mich dieser Sichtweise nicht an, denn sie ist bestenfalls für primäre Adjektive zutreffend. Abgeleitete Adjektive wie die *bar*-Adjektive können zahlreiche Bedeutungsvarianten haben, und keine davon muss primär im Sinne von Motschs Terminologie sein. *Absetzbar* existiert zum Beispiel in drei Varianten: im Kontext von Steuern, Personen und Waren. Primär ist keine der drei Bedeutungen, denn *absetzbar* in dem hier gebrauchten Sinn ist von dem Verb *absetzen* in seiner abstrakten Bedeutung abgeleitet. Wir benennen diese Bedeutungsvarianten mit *absetzbar\_Steuer*, *absetzbar\_Person* und *absetzbar\_Ware*. *Absetzbar* mit konkreter Bedeutung existiert vermutlich nicht, weil es keine kommunikative Notwendigkeit für eine solche Bildung gibt. Es würde eine triviale Eigenschaft beispielsweise von Mützen bezeichnen und damit das „Prinzip des sinnvollen Wortes“ (Motsch 2004, 26) verletzen.

oder nicht, sondern muss diese Aussage getrennt für die verschiedenen Bedeutungsvarianten treffen. Ob für ein Adjektiv mehrere Bedeutungsvarianten existieren und welche das sind, muss von Fall zu Fall durch die Analyse der Belege für jedes Adjektiv festgestellt werden. Auf diese Weise sind 263 Testdatensätze für die Data-Mining-Analyse entstanden, obwohl ich nur 244 *bar*-Adjektive untersucht habe.

Die Untersuchungsergebnisse der Korpusanalyse wurden in Merkmalsvektoren für die Data-Mining-Analyse umgewandelt. Jedem Adjektiv entspricht ein Merkmalsvektor, der wiederum 51 Attribute sowie das Zielmerkmal *class* enthält. Diese Merkmalsvektoren dienen als Eingabe für die verschiedenen Data-Mining-Algorithmen. Wir haben den Ablauf der Data-Mining-Analyse bereits in Abschnitt 4.3 besprochen.

Nicht alle Algorithmen liefern gute Ergebnisse, wenn man sie mit allen Attributen laufen lässt. Manchmal ist es sinnvoll, bestimmte Attribute von der Analyse auszuschließen, wenn die Vermutung besteht, dass sie nicht relevant sind oder wenn gezielt bestimmte Aspekte untersucht werden sollen. Dies wird im Folgenden noch deutlich werden.

Die verschiedenen Algorithmen liefern zum Teil recht ähnliche Ergebnisse. Der Prozentsatz an korrekt klassifizierten Beispielen kann von Testlauf zu Testlauf variieren, da im Allgemeinen zehnfache Kreuzvalidierung eingesetzt wird und die Auswahl, welche Merkmalsvektoren in welchen Fold kommen, zufällig ist.<sup>85</sup> Daher kommt es vor, dass derselbe Klassifikator bei verschiedenen Testläufen eine unterschiedliche Anzahl von Beispielen korrekt klassifiziert. Zu bemerken ist, dass keiner der erzeugten Klassifikatoren alle Testdatensätze gut klassifiziert. Es wird oft ein bestimmter Teil der Datenmenge sehr gut, andere Datensätze dagegen schlechter klassifiziert. Daher lohnt es sich, in den Klassifikatoren nach solchen Regeln zu suchen, die besonders viele Beispiele richtig klassifizieren. Nachfolgend sollen die Data-Mining-Ergebnisse diskutiert werden, die am aufschlussreichsten sind. Die Klassifikatoren, also die Entscheidungsbäume und Regeln, werden dabei stets in der Form wiedergegeben, wie sie vom Algorithmus erzeugt wurden. In Abschnitt 4.3 wurde erläutert, wie diese Klassifikatoren zu lesen sind.

Der Algorithmus *J4.8* liefert den folgenden Entscheidungsbaum, wenn man ihn mit allen 51 Attributen startet:

---

<sup>85</sup> Wie zehnfache Kreuzvalidierung funktioniert, wurde bereits in Abschnitt 4.3 erläutert.

- (2) un\_lich = ja: un (10.0)
  - un\_lich = nein
    - | Gradierbarkeit = ja
      - | | Akzentbelegung\_Silbe\_5 = H: un (1.0)
      - | | Akzentbelegung\_Silbe\_5 = N: un (0.0)
      - | | Akzentbelegung\_Silbe\_5 = U: nicht (6.0/1.0)
      - | | Akzentbelegung\_Silbe\_5 = -
        - | | | adjektivmodifizierend = ja: nicht (21.0/6.0)
        - | | | adjektivmodifizierend = nein: un (155.0/53.0)
    - | Gradierbarkeit = nein
      - | | dauerhaft = ja: nicht (54.0/5.0)
      - | | dauerhaft = nein
        - | | | Endrand\_Silbe\_2 = einfach
          - | | | | idiosynkratisch = ja: nicht (6.0)
          - | | | | idiosynkratisch = nein
            - | | | | | Akzentbelegung\_Silbe\_1 = H: un (4.0)
            - | | | | | Akzentbelegung\_Silbe\_1 = N: un (0.0)
            - | | | | | Akzentbelegung\_Silbe\_1 = U: nicht (4.0)
          - | | | | Endrand\_Silbe\_2 = komplex: nicht (0.0)
          - | | | | Endrand\_Silbe\_2 = leer: un (2.0)

Dieser Baum erzielt insgesamt eine Klassifikationsgenauigkeit von 70 %. Auch wenn dies – wie wir noch sehen werden – nicht das beste Ergebnis ist, sind hier einige Teilergebnisse von Bedeutung. Die Regel *un\_lich = ja: un* klassifiziert alle zehn Beispiele, die dieses Blatt erreichen, korrekt. Diese Regel wiederholt sich in mehreren anderen Klassifikatoren. Scheinbar werden *bar*-Adjektive, zu denen es parallele *un-lich*-Bildungen, aber keine *lich*-Bildungen gibt, mit *un* negiert. Allerdings ist zu bemerken, dass diese Erkenntnis aufgrund der dünnen Datenbasis nicht als gesichert gelten kann. Das zweite wichtige Teilergebnis aus diesem Baum ist das folgende:

- (2)(a) | Gradierbarkeit = nein
  - | | dauerhaft = ja: nicht (54.0/5.0)

Diese Regel klassifiziert 49 von 54 Beispielen (90,7 %) korrekt. Offenbar ist Nichtgradierbarkeit in Verbindung mit Dauerhaftigkeit ein sehr starkes Indiz dafür, dass ein *bar*-Adjektiv nicht mit *un* negiert werden kann.

Der Algorithmus *J4.8* klassifiziert insgesamt die größte Zahl von Beispielen korrekt, wenn man einige Attribute<sup>86</sup> weglässt. Er erzeugt dann den folgenden Entscheidungsbaum:

- (3) Gradierbarkeit = ja
  - | adjektivmodifizierend = ja: nicht (23.0/8.0)
  - | adjektivmodifizierend = nein
    - | | Bereich = Alltagssprache: un (151.0/46.0)
    - | | Bereich = Fachsprache: nicht (19.0/7.0)
- Gradierbarkeit = nein: nicht (70.0/11.0)

Dieser Entscheidungsbaum erreicht – je nach Testlauf – eine Genauigkeit von bis zu 72,6 %. Für Adjektive, die mit *un* negierbar sind, erzielt der Baum eine höhere Genauigkeit – in dieser Gruppe liegt der Anteil der korrekt klassifizierten Beispiele bei 76,6 %, während er in der Gruppe der nicht mit *un* negierbaren Adjektive bei 67,9 % liegt.

Einen ähnlichen Entscheidungsbaum liefert der Algorithmus *ADTree*:

- (4) : 0.004
  - | (1)Gradierbarkeit = ja: -0.25
  - | (1)Gradierbarkeit = nein: 0.801
  - | (2)Bereich = Alltagssprache: -0.093
    - | | (3)adjektivmodifizierend = ja: 0.648
    - | | (3)adjektivmodifizierend = nein: -0.103
  - | (2)Bereich = Fachsprache: 0.599
- Legend: -ve = un, +ve = nicht

Dieser Baum erreicht 70,7 % Genauigkeit und ist ein wenig anders zu lesen als die Bäume, die *J4.8* erzeugt: Um ein bestimmtes *bar*-Adjektiv zu klassifizieren, geht man alle für dieses Adjektiv möglichen Wege von der Wurzel zu den Blättern und summiert die Zahlen entlang der Wege auf. Ist das Ergebnis negativ, wird das Adjektiv als *un* klassifiziert, bei positivem Ergebnis als *nicht*. Der Betrag des Ergebnisses ist ein Maß dafür, wie eindeutig die Klassifikation ist: je höher der Betrag, desto eindeutiger. Für ein gradierbares *bar*-Adjektiv aus der Alltagssprache, das nicht adjektivmodifizierend verwendbar ist, müsste man folgende Zahlen aufsummieren:  $0,004 + (-0,25) + (-0,093) + (-0,103) = -0,442$ . Der Betrag des Ergebnisses ist

---

<sup>86</sup> Folgende Attribute wurden verwendet: *Praefix\_Basis*, *Komplexitaet\_Basis*, *fallengelassene\_Bestandteile*, *Stammende*, *Derivation*, *deverbal\_stark\_schwach*, *Aufbau\_Basis*, *Basis\_Partikelverb*, *praedikativ*, *attributiv*, *adjektivmodifizierend*, *adjunktiv*, *Valenz\_NomE*, *Valenz\_AkkE*, *Valenz\_DatE*, *Valenz\_praepE*, *Motsch*, *Innsbruck*, *Intentionalitaet*, *idiosynkratisch*, *Gradierbarkeit*, *Bereich*, *deverbal\_Verbart*, *modaler\_Nebensinn*.

0,442. Das bedeutet, dass die Klassifikation relativ eindeutig ist. Der Betrag der Zahl in einem Knoten des Entscheidungsbaums gibt an, wie relevant der Wert in diesem Knoten für die Klassifikation ist. Demnach ist *Gradierbarkeit = nein* mit 0,801 am relevantesten, gefolgt von *adjektivmodifizierend = ja* (0.648) und *Bereich = Fachsprache* (0.599).

Ein ebenso gutes Ergebnis (70,7 % Genauigkeit) liefert J4.8, wenn man ihn lediglich mit den beiden Attributen *Gradierbarkeit* und *adjektivmodifizierend* laufen lässt:

- (5) Gradierbarkeit = ja
  - | adjektivmodifizierend = ja: nicht (23.0/8.0)
  - | adjektivmodifizierend = nein: un (170.0/58.0)
- Gradierbarkeit = nein: nicht (70.0/11.0)

Bemerkenswert ist, dass bereits der einfache Regellerner *OneR*, der nur Regeln aus einem einzigen Merkmal generiert, 68,1 % Genauigkeit erzielt. Er erzeugt die folgenden Regeln:

- (6) Gradierbarkeit:
  - ja -> un
  - nein -> nicht

Dieses Ergebnis bestätigt, dass *Gradierbarkeit* das wichtigste Kriterium ist.

Lässt man J4.8 nur mit den Attributen *Komplexitaet\_Basis* und *Anzahl\_Silben* laufen, so erzeugt er den folgenden Entscheidungsbaum:

- (7) Komplexitaet\_Basis = einfach: nicht (74.0/33.0)
- Komplexitaet\_Basis = komplex
  - | Anzahl\_Silben <= 4: un (177.0/81.0)
  - | Anzahl\_Silben > 4: nicht (12.0/2.0)

Auffällig ist hier, dass die Merkmale *Anzahl\_Silben* und *Komplexitaet\_Basis* bei *bar*-Adjektiven, die nicht mehr als vier Silben haben, praktisch keinen Informationsgehalt zu besitzen scheinen – beinahe die Hälfte der Beispiele wird falsch klassifiziert (81 von 177). Bei *bar*-Adjektiven mit fünf oder sechs Silben hingegen werden zehn von zwölf Beispielen korrekt als *nicht* klassifiziert. Es lässt sich daher vermuten, dass mit der Länge der *bar*-Adjektive die Wahrscheinlichkeit für *nicht* steigt. Allerdings lassen sich hier keine sicheren Aussagen treffen, da *bar*-Adjektive mit mehr als vier Silben selten sind.

Startet man J4.8 nur mit den phonologischen und morphologischen Merkmalen<sup>87</sup>, erzeugt er folgenden Baum:

- (8) Akzentbelegung\_Silbe\_5 = H: un (1.0)
  - Akzentbelegung\_Silbe\_5 = N: nicht (1.0)
  - Akzentbelegung\_Silbe\_5 = U: nicht (10.0/1.0)
  - Akzentbelegung\_Silbe\_5 = -
    - | Akzentbelegung\_Silbe\_2 = H: un (86.0/31.0)
    - | Akzentbelegung\_Silbe\_2 = N: nicht (1.0)
    - | Akzentbelegung\_Silbe\_2 = U: nicht (164.0/74.0)

Die Regel *Akzentbelegung\_Silbe\_5 = U: nicht* ist lediglich eine andere Formulierung für die Regularität, die oben bereits festgestellt wurde: *Bar*-Adjektive mit mehr als vier Silben werden in der Regel nicht mit *un* negiert. Wenn die fünfte Silbe unbetont ist, bedeutet das in erster Linie, dass eine fünfte Silbe überhaupt existiert. Interessanter ist der folgende Teilbaum:

- (8)(a) Akzentbelegung\_Silbe\_5 = -
  - | Akzentbelegung\_Silbe\_2 = H: un (86.0/31.0)
  - | Akzentbelegung\_Silbe\_2 = N: nicht (1.0)
  - | Akzentbelegung\_Silbe\_2 = U: nicht (164.0/74.0)

Offensichtlich werden *bar*-Adjektive mit höchstens vier Silben eher mit *un* negiert, wenn der Wortakzent auf der zweiten Silbe liegt, und eher mit *nicht*, wenn die zweite Silbe unbetont ist. Dies ist jedoch nur eine relativ schwache Präferenz, wie man an der großen Zahl falsch klassifizierter Beispiele sieht.

Es wurde weiterhin eine Testreihe mit dem Algorithmus *J4.8* und dem *Resampling*-Filter durchgeführt. Resampling funktioniert folgendermaßen: Einige Merkmalsvektoren werden vor der Analyse zufällig ausgewählt, gelöscht und durch zufällig ausgewählte Duplikate anderer Merkmalsvektoren ersetzt. Die Datenmenge erhält dadurch eine Zufallskomponente, was möglicherweise zu neuen Erkenntnissen führt. Die Analyseergebnisse sind in diesem Fall allerdings mit Vorsicht zu betrachten, da einige Merkmale stärker repräsentiert sind als in der

---

<sup>87</sup> Es wurden die folgenden Attribute benutzt: *Anzahl\_Silben*, *Akzentbelegung\_Silbe\_1*, *Akzentbelegung\_Silbe\_2*, *Akzentbelegung\_Silbe\_3*, *Akzentbelegung\_Silbe\_4*, *Akzentbelegung\_Silbe\_5*, *Akzentbelegung\_Silbe\_6*, *Anfangsrand\_Silbe\_1*, *Kern\_Silbe\_1*, *Endrand\_Silbe\_1*, *Anfangsrand\_Silbe\_2*, *Kern\_Silbe\_2*, *Endrand\_Silbe\_2*, *Anfangsrand\_Silbe\_3*, *Kern\_Silbe\_3*, *Endrand\_Silbe\_3*, *Anfangsrand\_Silbe\_4*, *Kern\_Silbe\_4*, *Endrand\_Silbe\_4*, *Anfangsrand\_Silbe\_5*, *Kern\_Silbe\_5*, *Endrand\_Silbe\_5*, *Anfangsrand\_Silbe\_6*, *Kern\_Silbe\_6*, *Endrand\_Silbe\_6*, *Praefix\_Basis*, *Komplexitaet\_Basis*, *fallengelassene\_Bestandteile*, *Stammende*, *Derivation*, *deverbal\_stark\_schwach*, *Aufbau\_Basis*, *Basis\_Partikelverb*.



Ausgangsdatenmenge; andere Merkmale entsprechend schwächer. Deshalb wurde eine Testreihe mit zehn Versuchen durchgeführt. Dadurch dürften sich eventuell auftretende Schwankungen ausgleichen. Die Klassifikatoren, die in dieser Testreihe entstanden, weisen durchweg eine hohe Klassifikationsgenauigkeit auf; in einigen Fällen über 81 %. Dieses Ergebnis ist allerdings nicht sehr aussagekräftig, da die Kreuzvalidierung auf den gesampelten Daten durchgeführt wird. Diese enthalten insgesamt weniger Informationen als die Ausgangsdatenmenge, da ja Duplikate vorkommen.

Die einzelnen Entscheidungsbäume aus der Resampling-Testreihe sollen hier nicht aufgeführt werden, denn es sind zu viele und sie sind auch zu groß. Interessant ist nur ein Ergebnis dieser Testreihe: In acht der zehn Versuche findet sich das Merkmal *Praefix\_Basis* im generierten Entscheidungsbaum, in den meisten Fällen tiefer als die Regeln *Gradierbarkeit = ja* und *Akzentbelegung\_Silbe\_5 = -* beziehungsweise *Anfangsrand\_Silbe\_5 = -*.<sup>88</sup> Daraus lässt sich schlussfolgern, dass bei *bar*-Adjektiven, die gradierbar sind und nicht mehr als vier Silben haben, ein eventuell vorhandenes Präfix Einfluss auf die Klassifikation hat. Auffällig ist außerdem, dass die zwei Entscheidungsbäume, die das Merkmal *Praefix\_Basis* nicht enthalten, weniger Beispiele korrekt klassifizieren (73 % beziehungsweise 75,6 %) als die Bäume, in denen das Merkmal vorkommt (zwischen 77 % und 82 %). Dies kann ebenfalls als Indiz für die Relevanz des Merkmals *Praefix\_Basis* gewertet werden.

Einige Präfixe sind scheinbar besonders deutliche Indizien für eine Negation mit *un* oder *nicht*: *ab* für *nicht*; *ver*, *über*, *be* und *an* für *un*. Die Entscheidungsbäume aus mindestens sieben der zehn Versuche klassifizieren die Mehrheit der Adjektive mit diesen Präfixen korrekt. Dieses Ergebnis wird – zumindest teilweise – durch einen anderen Versuch gestützt. Lässt man den Algorithmus *PART* nur mit den beiden Merkmalen *Praefix\_Basis* und *Gradierbarkeit* laufen, so erzeugt er die folgenden Regeln:

- (9) *Praefix\_Basis = be AND Gradierbarkeit = ja: un* (24.0/4.0)  
*Gradierbarkeit = ja: un* (169.0/69.0)  
*: nicht* (70.0/11.0)

Die Regel *Praefix\_Basis = be AND Gradierbarkeit = ja: un* klassifiziert 20 von 24 Beispielen korrekt – ein weiteres Anzeichen dafür, dass bei gradierbaren *bar*-Adjektiven das Präfix Einfluss auf die Klassifikation haben kann.

---

<sup>88</sup> Die letzten beiden Regeln kann man zu einer einzigen zusammenfassen: *Das Wort darf nicht mehr als vier Silben haben*.

## 7 Auswertung der Analyseergebnisse

Die Data-Mining-Analyse hat ergeben, dass vor allem die semantischen Merkmale eines *bar*-Adjektivs darüber entscheiden, ob das Adjektiv mit *un* negierbar ist oder nicht.

Das stärkste Kriterium ist die Gradierbarkeit beziehungsweise Nichtgradierbarkeit. Ist ein *bar*-Adjektiv nicht gradierbar, so kann es mit großer Wahrscheinlichkeit nicht mit *un* negiert werden.<sup>89</sup> Folgende Interpretation dieses Ergebnisses liegt nahe: Nicht gradierbare Adjektive bilden meist kontradiktorische Antonyme. Wie wir in Abschnitt 3.2 gesehen haben, entspricht kontradiktorisches *un* bedeutungsmäßig *nicht*. Die fehlende *un*-Negation bewirkt daher keine semantische Lücke, denn das Gleiche kann auch mit *nicht* ausgedrückt werden. Das *un*-Adjektiv könnte also aus Gründen der Sprachökonomie fehlen.

Ist das Adjektiv nicht gradierbar und beschreibt zusätzlich eine dauerhafte Eigenschaft, ist die Wahrscheinlichkeit, dass es nicht mit *un* negiert werden kann, noch höher.<sup>90</sup> Auch Fachsprachlichkeit ist ein Indikator für fehlende *un*-Negation.

Bei den gradierbaren Adjektiven spielen noch andere Merkmale eine Rolle. Ein syntaktisches Kriterium ist offenbar von großer Bedeutung: Ist das *bar*-Adjektiv adjektivmodifizierend verwendbar, wird es eher mit *nicht* negiert, andernfalls eher mit *un*. Bei kürzeren gradierbaren *bar*-Adjektiven wird die Negation unter Umständen von einem Präfix beeinflusst – *ver*, *über* und *be* sind Indizien für *un*-Negation.<sup>91</sup> Über die Gründe kann man spekulieren: *Ver*, und *be* sind nicht betonbar; *über* kann betont sein oder auch nicht. Somit ergibt sich bei Präfigierung mit *un* am Wortanfang die prosodisch vorteilhafte Betonungsfolge *betont* – *unbetont*, denn *un* ist normalerweise betont, wie wir bereits in Abschnitt 5.1 festgestellt haben. Bei den Präfixen *an* und *ab* ist die Sache komplizierter. Es lassen sich kaum Vermutungen anstellen, warum die Analyse ergeben hat, dass *an* auf *un*-Negation, *ab* dagegen auf *nicht*-Negation hindeutet, denn die Datenbasis ist hier besonders dünn: Es wurden nur vier gradierbare *bar*-Adjektive mit dem Präfix *ab* und sechs gradierbare *bar*-Adjektive mit dem Präfix *an* untersucht.

Zum Teil sind noch weitere prosodische Präferenzen von Belang. Kürzere *bar*-Adjektive, die den Wortakzent auf der zweiten Silbe haben, neigen zur Negation mit *un*, während solche mit

---

<sup>89</sup> Wie schon in Abschnitt 3.3 gesagt, vertritt Lenz eine andere Ansicht: Sie ist der Auffassung, dass Adjektive, die nicht komparierbar sind, nicht mit *un* präfigiert werden können (vgl. Lenz 1995, 28). Aufgrund meiner Beobachtungen während der Korpusanalyse komme ich allerdings zu einem anderen Schluss. Es gibt eine Reihe von *bar*-Adjektiven – etwa *vermeidbar*, *bezwingbar* und *überwindbar* – die nicht komparierbar sind, aber mit Fokuspartikeln modifiziert und mit *un* negiert werden können. Die *un*-Präfigierbarkeit hängt also scheinbar – zumindest bei den *bar*-Adjektiven – eher von Gradierbarkeit im Allgemeinen als von Komparierbarkeit ab.

<sup>90</sup> Dieses Ergebnis könnte eine Erklärung dafür sein, warum das Beispiel *zusammenklappbar* aus Abschnitt 3.3 nicht mit *un* negiert werden kann.

<sup>91</sup> Diese Aussage lässt sich allerdings nur für *be* mit Sicherheit treffen.

unbetonter zweiter Silbe oft mit *nicht* negiert werden. Das könnte daran liegen, dass bei der Mehrheit der *bar*-Adjektive mit unbetonter zweiter Silbe die erste Silbe den Wortakzent trägt. Da *un* gern den Wortakzent auf sich zieht, käme es in diesem Fall zu einem Akzentzusammenstoß, der – falls das *bar*-Adjektiv seinen Akzent nicht aufgibt – durch eine Negation mit *nicht* vermieden wird. Längere *bar*-Adjektive bevorzugen offenbar ebenfalls die Negation mit *nicht* – wahrscheinlich, weil die Negation mit *un* das Adjektiv noch verlängern würde. Wolfgang Motschs *Prinzip der Knappheit von Wortbildungen* (vgl. Motsch 2004, 26f.) scheint hier also zu greifen.

Alle untersuchten *bar*-Adjektive, zu denen es eine parallele *un-lich*-Bildung, aber keine parallele *lich*-Bildung gibt, werden mit *un* negiert. Es lässt sich allerdings nicht mit Sicherheit sagen, ob diese Feststellung verallgemeinert werden kann, da ich im *public*-Korpus nur zehn *bar*-Adjektive finden konnte, auf die die genannten Kriterien zutreffen und die in mehr als 50 Belegen vorkommen. Deshalb konnten auch nur diese zehn Adjektive in die Analyse einbezogen werden. Lenz' Untersuchung stützt jedoch die obige Vermutung (vgl. Lenz 1995, 100f.).

## 8 Zusammenfassung und Ausblick

In dieser Arbeit wurde die Negation von *bar*-Adjektiven untersucht. Prinzipiell ist die Negation auf zwei Arten möglich: durch Präfigierung mit *un* und durch den syntaktischen Negationsträger *nicht*. Ersteres ist restringiert, letzteres ist bei jedem *bar*-Adjektiv möglich. Ob die Negation mit *un* möglich ist oder nicht, wird durch die Eigenschaften des *bar*-Adjektivs bestimmt. Dafür kommen potenziell alle Eigenschaften des *bar*-Adjektivs infrage. Ziel dieser Arbeit war es, herauszufinden, welche Eigenschaften tatsächlich relevant für die Negation sind. Diese Eigenschaften wurden in einem integrierten Ansatz aus einer Korpusanalyse und einer Data-Mining-Untersuchung ermittelt und ausgewertet. Zunächst habe ich so viele Eigenschaften wie möglich für 263 Beispiele ermittelt, einige mithilfe einer Analyse des *public*-Korpus des Instituts für deutsche Sprache. Diese Eigenschaften wurden dann in Merkmalsvektoren umgewandelt und mit Data-Mining-Algorithmen analysiert. Es stellte sich heraus, dass vor allem die semantischen Eigenschaften eines *bar*-Adjektivs beeinflussen, ob es mit *un* negiert werden kann oder lediglich mit *nicht*. Darüber hinaus existiert noch eine Reihe phonologischer, syntaktischer und morphologischer Beschränkungen.

Der beste Klassifikator, der in der Data-Mining-Analyse erzeugt wurde, klassifiziert 72,6 % der Beispiele korrekt. Dass der Prozentsatz nicht höher ist, liegt sicher zum Teil an der Art der Daten. Wie bereits gesagt, sind linguistische Daten per se relativ unscharf. Zum Teil mag es aber auch an der Wahl der Attribute und ihrer möglichen Werte liegen. Sinnvoll wäre sicherlich, die Analyse auszubauen und zu präzisieren, um genauere Ergebnisse zu erhalten. Im Folgenden sollen einige Anregungen gegeben werden, wie die Arbeit zu diesem Thema fortgesetzt werden könnte.

Denkbar wäre beispielsweise, nominale Werte für diejenigen Attribute einzuführen, denen in dieser Untersuchung Wahrheitswerte zugeordnet wurden, etwa für das Zielattribut *class* oder das Attribut *Gradierbarkeit*. Es würde sich anbieten, die Anzahl der Belege für die einzelnen Attributwerte aus dem *public*-Korpus dafür zu verwenden. *Unhörbar* ist im *public*-Korpus mit 301 Belegen vertreten, *nicht hörbar* mit 168. Es werden also 64 % der Negationen von *hörbar* mit *un* und 36 % mit *nicht* gebildet. Als Zielattribut würde sich beispielsweise der Prozentsatz der Negationen mit *un* anbieten. Für das Beispiel *hörbar* wäre das 64. Die Klassifikatoren würden dann natürlich auch nicht mehr *un* oder *nicht*, sondern einen Prozentsatz vorhersagen. Dies könnte unter Umständen bessere Ergebnisse bringen. Sinnvoll wäre sicherlich auch, das Attribut *Gradierbarkeit* aufzuteilen in mehrere Attribute. Man könnte zum Beispiel eines ha-

ben für jede Art der Gradierung, also für die Verwendung eines *bar*-Adjektivs im Komparativ, mit *sehr*, *kaum*, *leicht* und *schwer*. All das war mir aus Zeitgründen nicht möglich. Möglicherweise ließen sich auch mehr Regeln finden, wenn man insgesamt mehr Beispiele sowie für jeden Attributwert hinreichend viele Beispiele untersuchen würde.

Denkbar wäre auch, die Ergebnisse der Analyse in einen optimalitätstheoretischen Rahmen einzubetten. Die Optimalitätstheorie stammt aus der generativen Grammatik und wurde von Alan Prince und Paul Smolensky entwickelt.<sup>92</sup> Sie hat ihren Ursprung in der Phonologie, lässt sich aber auf andere Bereiche der Grammatik übertragen. Auch für die Thematik der Negation von *bar*-Adjektiven lässt sie sich – stark vereinfacht – adaptieren. Ihr Vorteil ist, dass sie als Formalismus dienen kann, aber auch algorithmisch umsetzbar ist.

Der Grundbegriff der Optimalitätstheorie ist die Markiertheit. Markierte Strukturen sind solche, die vom Prototypischen, vom Idealen abweichen. Markierte Oberflächenformen sind nicht von vorn herein ungrammatisch, sondern nur dann, wenn es eine bessere, „optimalere“ Oberflächenform gibt. Das bedeutet umgekehrt, dass jede tatsächlich in der Sprache vorkommende Oberflächenform optimal ist. Optimal bedeutet in diesem Zusammenhang eben nicht, dass diese Form keine grammatischen Regeln verletzt, sondern nur, dass sie weniger grammatische Regeln verletzt als jede andere mögliche Oberflächenform. Das Ziel ist nun, diese optimale Oberflächenform zu finden. Dazu definiert man eine Menge von Regeln<sup>93</sup>, die verletzt werden dürfen, und ordnet sie hierarchisch an – von der wichtigsten zur unwichtigsten. Dann bestimmt man für jede mögliche Ausgabe zu einer bestimmten Eingabe, welche Constraints diese Ausgabe verletzt. Die Ausgabe, die die wenigsten wichtigen Constraints verletzt, „gewinnt“ und wird zur Oberflächenform. Am besten lässt sich das an zwei der untersuchten Beispiele verdeutlichen:

(1) *hörbar*

Input: NEG ( <i>hörbar</i> ) <sup>94</sup>	1	2	3	4	5	6
<i>nicht hörbar</i>				*	*	*
→ <i>unhörbar</i>						

<sup>92</sup> Siehe Prince/Smolensky 2004; eine lesbare Einführung in das Gebiet bietet auch Kager 1999.

<sup>93</sup> Man nennt diese Regeln auch *Constraints*.

<sup>94</sup> Zu lesen als: Negation von *hörbar*.

(2) *erkennbar*

Input: NEG ( <i>erkennbar</i> )	1	2	3	4	5	6
→ <i>nicht erkennbar</i>				*		*
<i>unerkennbar</i>		*				

Den Ziffern in den Tabellenköpfen entsprechen folgende Regeln:<sup>95</sup>

1 *Gradierbarkeit* = *nein*: *nicht*

2 *adjektivmodifizierend* = *ja*: *nicht*

3 *Bereich* = *Fachsprache*: *nicht*

4 *Gradierbarkeit* = *ja*: *un*

5 *adjektivmodifizierend* = *nein*: *un*

6 *Bereich* = *Alltagssprache*: *un*

Die Sternchen markieren die Regelverstöße; ein Pfeil vor einem Output bedeutet, dass dieser Output als optimaler Output ausgewählt wurde. *Unhörbar* ist der ideale Output zum Input *NEG* (*hörbar*): Er verletzt keinerlei Regeln; *nicht hörbar* verletzt dagegen drei Regeln. *Nicht erkennbar* ist nicht ideal, aber trotzdem der optimale Output zum Input *NEG* (*erkennbar*), obwohl er gegen zwei Constraints verstößt. *Unerkennbar* verstößt zwar lediglich gegen einen Constraint; dieser steht allerdings in der Hierarchie über den beiden Constraints, die *nicht erkennbar* verletzt.

Als Formalismus zur Veranschaulichung der Untersuchungsergebnisse ist der optimalitätstheoretische Ansatz gut geeignet. Es wäre nun allerdings notwendig, die Leistungsfähigkeit dieses Verfahrens auch für die Klassifikation zu testen, indem man es als Programm implementiert und ermittelt, wie viele Merkmalsvektoren korrekt klassifiziert werden. Das sollte mit etwas Zeitaufwand möglich sein. Man könnte dann auch feststellen, ob die Constraints aus nur einer Bedingung – wie sie in den obigen Beispielen verwendet wurden – gute Ergebnisse liefern oder ob man besser Konjunktionen von mehreren Constraints wie etwa *Gradierbarkeit* = *nein* AND *Bereich* = *Fachsprache*: *nicht* verwenden sollte.

Die Untersuchungsergebnisse legen außerdem die Vermutung nahe, dass die Restriktionen für die *un*-Präfigierung nicht für *bar*-Adjektive allein gelten, denn die meisten relevanten Kriterien – etwa *Gradierbarkeit*, *Fachsprachlichkeit* und die Möglichkeit der adjektivmodifizieren-

<sup>95</sup> Diese Regeln stammen aus dem von ADTree erzeugten Entscheidungsbaum aus Kapitel 6. Sie wurden nach den Beträgen der Zahlen in den Knoten gewichtet. Man hätte natürlich auch einen anderen Klassifikator verwenden können.

den Verwendung – lassen sich auf andere Adjektivklassen übertragen. Interessant wäre daher, die vorliegende Untersuchung auf andere derivierte Adjektive, etwa solche auf *lich*, *sam* oder *ig* auszuweiten.

## Literatur

Aitchison, Jean (1997): Wörter im Kopf. Eine Einführung in das mentale Lexikon. Tübingen: Niemeyer (= Konzepte der Sprach- und Literaturwissenschaft 56).

Béjoint, Henri (2000): Modern Lexicography. An Introduction. Oxford: University Press.

Benware, Wilbur (1988): Die Betonung im Wörterbucheintrag. Ein Plädoyer für Präzisierung. In: Wirkendes Wort 2(1988). S. 274–278.

Blau, Ulrich (1978): Die dreiwertige Logik der Sprache. Ihre Syntax, Semantik und Anwendung in der Sprachanalyse. Berlin, New York: de Gruyter.

Brinkmann, Henning (1962): Die deutsche Sprache. Gestalt und Leistung. Düsseldorf: Schwann.

Brütsch, Edgar/Markus Nussbaumer/Horst Sitta (1990): Negation. Heidelberg: Groos (= Studienbibliographien zur Sprachwissenschaft 1).

DiSciullo, Anna Maria/Edwin Williams (1987): On the Definition of Word. Cambridge, London: MIT Press (= Linguistic inquiry monographs 14).

Donalies, Elke (2002): Die Wortbildung des Deutschen. Ein Überblick. Tübingen: Narr (= Studien zur Deutschen Sprache 27).

Eisenberg, Peter (1991): Syllabische Struktur und Wortakzent. Prinzipien der Prosodik deutscher Wörter. In: Zeitschrift für Sprachwissenschaft 10(1991). S. 37–64.

Eisenberg, Peter (2004): Grundriß der deutschen Grammatik. Band 1: Das Wort. 2., überarbeitete und aktualisierte Auflage. Stuttgart, Weimar: Metzler.

Erben, Johannes (2000): Einführung in die deutsche Wortbildungslehre. 4., aktualisierte und ergänzte Auflage. Berlin: Schmidt (= Grundlagen der Germanistik 17).



Eschenlohr, Stefanie (1996): Derivational morphology and the system of word classes in German. In: *Acta Linguistica Hungarica* 43(1995). S. 93–110.

Féry, Caroline (2000): *Phonologie des Deutschen. Eine optimalitätstheoretische Einführung*. Potsdam: Publikationsstelle der Universitätsbibliothek (= Linguistics in Potsdam 7).

Fleischer, Wolfgang/Irmhild Barz (1995): *Wortbildung der deutschen Gegenwartssprache*. 2., durchgesehene und ergänzte Auflage. Tübingen: Niemeyer.

Flury, Robert (1964): *Struktur- und Bedeutungsgeschichte des Adjektivsuffixes -bar*. Winterthur: Keller.

Freund, Yoav/Llew Mason (1999): The Alternating Decision Tree Learning Algorithm. In: *Proceedings of the 16th International Conference on Machine Learning*. San Francisco: Morgan Kaufmann. S. 124–133.

Gelhaus, Hermann (1977): *Der modale Infinitiv*. Tübingen: Narr (= Forschungsberichte des Instituts für deutsche Sprache 35).

Haß-Zumkehr, Ulrike (2001): *Deutsche Wörterbücher – Brennpunkt von Sprach- und Kulturgeschichte*. Berlin, New York: de Gruyter.

Helbig, Gerhard/Helga Albrecht (1993): *Die Negation. Zur Theorie und Praxis des Deutschunterrichts für Ausländer*. 6. Auflage. Leipzig u.a.: Langenscheidt/Verlag Enzyklopädie.

Hentschel, Elke/Harald Weydt (2003): *Handbuch der deutschen Grammatik*. 3., völlig neu bearbeitete Auflage. Berlin, New York: de Gruyter.

Holte, Robert C. (1993): Very simple classification rules perform well on most commonly used datasets. In: *Machine Learning* 11(1993). S. 63–91.

Jacobs, Joachim (1991): Negation. In: Arnim von Stechow/Dieter Wunderlich (Hrsg.): *Semantik. Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin, New York: de Gruyter. S. 560–596 (= Handbücher zur Sprach- und Kommunikationswissenschaft 6).

Kager, René (1999): *Optimality Theory*. Cambridge: University Press.

Kątny, Andrzej (1997): Die deverbalen „-bar“-Adjektive im Deutschen und deren Entsprechungen im Polnischen. In: *Studia Germanica Posnaniensia* 23(1997). S. 83–95.

Kennedy, Graeme (1998): *Introduction to Corpus Linguistics*. London, New York: Longman.

Kiparsky, Paul (1966): Über den deutschen Akzent. In: *Studia Grammatica VII. Untersuchungen über Akzent und Intonation im Deutschen*. Berlin: Akademie. S. 69–98.

Kohler, Klaus J. (1995): *Einführung in die Phonetik des Deutschen*. 2., neubearbeitete Auflage. Berlin: Schmidt (= *Grundlagen der Germanistik* 20).

Kühnhold, Ingeburg/Oskar Putzer/Hans Wellmann (1978): *Deutsche Wortbildung: Typen und Tendenzen der Gegenwartssprache*. Hauptteil 3. Das Adjektiv. Düsseldorf: Schwann (= *Sprache der Gegenwart* 43).

Lenz, Barbara (1991): *un-Affigierung im Lexikon*. Universität Wuppertal. (= *Arbeiten des Sonderforschungsbereiches* 282 „Theorie des Lexikons“ 9)

Lenz, Barbara (1995): *un-Affigierung. unrealisierbare Argumente, unausweichliche Fragen, nicht unplausible Antworten*. Tübingen: Narr (= *Studien zur deutschen Grammatik* 50).

Lenz, Barbara (1998): *Unkaputtbar ist unkaputtbar. Zur Etablierung einer Ad-hoc-Bildung*. In: *Sprachdienst* 42(1998). S. 10–12.

Lenz, Susanne (2000): *Korpuslinguistik*. Tübingen: Groos Brigitte Narr (= *Studienbibliographien zur Sprachwissenschaft* 32).

Motsch, Wolfgang (1988): On Inactivity, Productivity and Analogy in Derivational Processes. In: *Linguistische Studien des ZISW* 179(1988). S. 1–30.

Motsch, Wolfgang (2004): *Deutsche Wortbildung in Grundzügen*. 2., überarbeitete Auflage. Berlin, New York: de Gruyter (= *Schriften des Instituts für deutsche Sprache* 8).

Prince, Alan/Paul Smolensky (2004): *Optimality Theory. Constraint Interaction in Generative Grammar*. Malden, Oxford, Carlton: Blackwell.

Quinlan, J. Ross (1993): *C4.5. Programs for Machine Learning*. Los Altos: Morgan Kaufmann.

Riehemann, Susanne (1993): *Word Formation in Lexical Type Hierarchies. A Case Study of *bar*-Adjectives in German*. Sfs-Report-02-93. Seminar für Sprachwissenschaft, Eberhard-Karls-Universität Tübingen (<http://doors.stanford.edu/~sr/sfsreport.ps>, 03.03.2006).

Siebert, Susann (1999): *Wortbildung und Grammatik. Syntaktische Restriktionen in der Struktur komplexer Wörter*. Tübingen: Niemeyer (= Linguistische Arbeiten 408).

Šved, V. I. (1977): *Vergleichende Analyse von deverbativen *-bar*-Adjektiven und Partizipien der deutschen Gegenwartssprache*. In: *Deutsch als Fremdsprache* 14(1977). S. 276–279.

Teubert, Wolfgang (2004): *Language and corpus linguistics*. In: Halliday, Michael A. K. u.a. (Hrsg.): *Lexicology and Corpus Linguistics. An Introduction*. London, New York: Continuum. S. 73–112.

Toman, Jindřich (1987): *Wortsyntax. Eine Diskussion ausgewählter Probleme deutscher Wortbildung*. 2., erweiterte Auflage. Tübingen: Niemeyer (= Linguistische Arbeiten 137).

VALBU – Valenzwörterbuch deutscher Verben (2004). Bearbeitet von Helmut Schumacher u.a. Tübingen: Narr (= Studien zur deutschen Sprache 31).

Vennemann, Theo (1991): *Skizze der deutschen Wortprosodie*. In: *Zeitschrift für Sprachwissenschaft* 10(1991). S. 86–111.

Wiese, Richard (1996): *The Phonology of German*. Oxford: University Press.

Williams, Edwin (1981): *On the Notions “Lexically Related” and “Head of a Word”*. In: *Linguistic Inquiry* 12(1981). S. 245–274.

Witten, Ian H./Eibe Frank (1998): Generating accurate rule sets without global optimization. In: Proceedings of the 15th International Conference on Machine Learning. San Francisco: Morgan Kaufmann. S. 144–151.

Witten, Ian H./Eibe Frank (2005): Data Mining. Practical machine learning tools and techniques. 2. Auflage. San Francisco: Morgan Kaufmann.

## Anhänge

Die folgenden Seiten enthalten Datenmaterial, das aus Gründen der Übersichtlichkeit nicht in den Text integriert werden konnte.

In Anhang A sind alle untersuchten Adjektive in alphabetischer Reihenfolge aufgelistet. Die Darstellung ist so gewählt, dass erkennbar ist, welche Adjektive gemeinsam und welche getrennt analysiert wurden: Varianten in neuer und alter Rechtschreibung (etwa *essbar/eßbar*) gelten beispielsweise als ein einziges Adjektiv, ebenso regionale Varianten (*benutzbar/benützbar*). Tritt ein Adjektiv dagegen auf mehreren Zeilen auf, handelt es sich um Polyseme, die getrennt analysiert wurden. Hier ist jeweils in Klammern vermerkt, um welche Bedeutungsvariante es sich handelt. Aus Gründen der Lesbarkeit wurde auf Kursivierung verzichtet.

Anhang B ordnet alle Attribute ihren möglichen Werten zu. Dieser Teil führt also in Tabellenform die in Kapitel 5 eingeführten Merkmale nochmals auf.

In Anhang C sind drei Beispieldatensätze in Tabellenform aufgelistet. Es handelt sich um die Datensätze zu den Adjektiven *einklagbar*, *fruchtbar* und *wahrnehmbar*.

Alle 263 für die Data-Mining-Analyse benutzten Datensätze befinden sich auf der CD, die dieser Arbeit beiliegt. Die Datensätze liegen als Tabellen in zwei unterschiedlichen Dateiformaten vor: als *.pdf*-Datei und als *.xls*-Datei.<sup>96</sup> *.pdf*-Dateien können zum Beispiel mit dem frei erhältlichen *Acrobat Reader* betrachtet werden. In der *.pdf*-Tabelle stehen die Adjektive in der ersten Spalte, danach folgen die 52 Attribute. In der letzten Spalte steht das Zielmerkmal *class*. *.xls*-Dateien können mit gängigen Tabellenkalkulationsprogrammen wie *OpenOffice Calc* oder *Microsoft Excel* angeschaut und bearbeitet werden. Die Darstellung der *.xls*-Tabelle ist in einer Form belassen, die eine problemlose Konvertierung der Daten in das vom WEKA-System benötigte *.arff*-Format zulässt. Daher steht das zu einem Datensatz gehörige Adjektiv in der letzten Spalte als Kommentar hinter dem Zielmerkmal.<sup>97</sup> Auf diese Weise ist gewährleistet, dass die Daten leicht weiterverarbeitet werden können.

---

<sup>96</sup> Die Tabellen konnten dem Anhang nicht beigelegt werden, da sie zu groß sind.

<sup>97</sup> Kommentare werden im *.arff*-Format vom %-Zeichen eingeleitet.

## Anhang A: Liste der analysierten Adjektive in alphabetischer Reihenfolge

abbaubar	beschreibbar (wörtlich)
ableitbar	besiegbar
ablesbar	beispielbar
abnehmbar	bestimmbar
abrufbar	bewältigbar
abschaltbar	beweisbar
abschätzbar	bewohnbar
abschließbar (übertragen)	bezahlbar
abschließbar (wörtlich)	bezifferbar
absehbar	bezwingbar
absetzbar (Person)	brauchbar
absetzbar (Steuer)	brennbar
absetzbar (Ware)	buchbar
abwendbar	dankbar
achtbar	darstellbar
akzeptierbar	definierbar (Alltag)
amortisierbar	definierbar (Fachsprache)
anfechtbar	dehnbar
angreifbar	demontierbar
annehmbar	denkbar
anrechenbar	dienstbar
ansprechbar	drehbar
antastbar	durchführbar
anwendbar	durchhörbar
aufblasbar	durchschaubar
auffindbar	durchsetzbar
aufklappbar	einholbar
auflösbar	einklagbar
aufschiebbar	einlösbar
ausbaubar	einschätzbar
ausfahrbar	einsehbar
ausführbar	einsetzbar
austauschbar	einstellbar
auswechselbar	empfangbar
ausziehbar	entflammbar
bebaubar	entschuldbar
beeinflussbar/beeinflußbar	entzifferbar
befahrbar	entzündbar
begehbar	erahnbar
begreifbar	erfahrbar
begründbar	erfassbar/erfaßbar
behandelbar	erfüllbar
beheizbar	erkennbar
beherrschbar	erklärbar
belastbar	erlebbar
belegbar	erlernbar
bemerkbar	erneuerbar
benutzbar/benützbar	erpressbar/erpreßbar
beobachtbar	erreichbar
berechenbar (Alltag)	ersetzbar
berechenbar (Mathematik)	eruierbar
beschreibbar (übertragen)	erwartbar

erweiterbar	nachvollziehbar
erziehbar	nachweisbar
erzielbar	nutzbar
essbar/eßbar	offenbar
fahrbar	passierbar
fassbar	patentierbar
fehlbar	planbar
feststellbar	programmierbar
finanzierbar	quantifizierbar
förderbar	realisierbar
formbar	recyclierbar/rezyklierbar
fruchtbar	regulierbar
fühlbar	reizbar
furchtbar	reparierbar
gangbar	reproduzierbar
genießbar	ruchbar
gestaltbar	rückzahlbar
gewinnbar	schandbar
greifbar	schätzbar
haftbar	scheinbar
haltbar	schiffbar
handelbar	schlagbar
handhabbar	schwenkbar
heilbar	sichtbar
herausnehmbar	sonderbar
herstellbar	spaltbar
herzeigbar	spielbar
höhenverstellbar	spürbar
hörbar	stapelbar
identifizierbar	steuerbar
integrierbar (Alltag)	steuerbar (Österreich/Schweiz)
integrierbar (Mathematik)	strafbar
interpretierbar	streitbar
jagdbar	tanzbar
kalkulierbar	teilbar (Alltag)
klagbar	teilbar (Mathematik)
kombinierbar	therapierbar
kompostierbar	tolerierbar
konsumierbar	tragbar (abstrakt)
kontrollierbar	tragbar (konkret)
kopierbar	trennbar
korrigierbar	trinkbar
kostbar	überbietbar
kündbar	überblickbar
leistbar	überhörbar
lenkbar	überprüfbar
lernbar	überschaubar
lesbar	übersehbar
lieferbar	übertragbar
lösbar	überwindbar
machbar	umkehrbar
manipulierbar	umklappbar
maschinenlesbar	umlegbar
messbar	umsetzbar
mittelbar	unterscheidbar
nachprüfbar	unterteilbar

urbar	verwendbar
veränderbar	verwertbar
verantwortbar	verwirklichbar
vereinbar	verwundbar
verfügbar	verzichtbar
verführbar	vollziehbar
vergleichbar	voraussehbar
verhandelbar	vorhersagbar
verkraftbar	vorhersehbar
verlängerbar	vorstellbar
verletzbar	vorzeigbar
vermarktbar	wählbar
vermeidbar	wahrnehmbar
vermietbar	wandelbar
vermittelbar	widerlegbar
vernachlässigbar	wiederholbar
vernehmbar	wiederverwendbar
verrottbar	wiederverwertbar
verschiebbar	wunderbar
versicherbar	wünschbar
verstehbar	zählbar
verstellbar	zumutbar
vertretbar	zuschaltbar
verwechselbar	



**Anhang B: Die Attribute und ihre möglichen Werte**

Attribute	Mögliche Werte
Anzahl_Silben	2, 3, 4, 5, 6
Akzentbelegung_Silbe_1	H, N, U
Akzentbelegung_Silbe_2	H, N, U
Akzentbelegung_Silbe_3	H, N, U, -
Akzentbelegung_Silbe_4	H, N, U, -
Akzentbelegung_Silbe_5	H, N, U, -
Akzentbelegung_Silbe_6	H, N, U, -
Anfangsrand_Silbe_1	einfach, komplex, leer
Kern_Silbe_1	gespannt, ungespannt, Diphthong, Schwa
Endrand_Silbe_1	einfach, komplex, leer
Anfangsrand_Silbe_2	einfach, komplex, leer
Kern_Silbe_2	gespannt, ungespannt, Diphthong, Schwa
Endrand_Silbe_2	einfach, komplex, leer
Anfangsrand_Silbe_3	einfach, komplex, leer, -
Kern_Silbe_3	gespannt, ungespannt, Diphthong, Schwa, -
Endrand_Silbe_3	einfach, komplex, leer, -
Anfangsrand_Silbe_4	einfach, komplex, leer, -
Kern_Silbe_4	gespannt, ungespannt, Diphthong, Schwa, -
Endrand_Silbe_4	einfach, komplex, leer, -
Anfangsrand_Silbe_5	einfach, komplex, leer, -
Kern_Silbe_5	gespannt, ungespannt, Diphthong, Schwa, -
Endrand_Silbe_5	einfach, komplex, leer, -
Anfangsrand_Silbe_6	einfach, komplex, leer, -
Kern_Silbe_6	gespannt, ungespannt, Diphthong, Schwa, -
Endrand_Silbe_6	einfach, komplex, leer, -
Praefix_Basis	ab, an, auf, aus, be, dar, de, durch, ein, ent, er, fest, ge, -, ...
Komplexitaet_Basis	einfach, komplex
fallengelassene_Bestandteile	bar, ig, n, zu, -, ...
Stammende	el, en, er, ier, ig, lich, -, ...
Derivation	deverbal, desubstantivisch, deadjektivisch, desubst_deverb, Komposition, ?
deverbal_stark_schwach	stark, schwach, unregelmässig, -

Anhang B: Die Attribute und ihre möglichen Werte

Attribute	Mögliche Werte
Aufbau_Basis	einfV, trennbar, untrennbar, Kompositum, -
Basis_Partikelverb	ja, nein, -
praedikativ	ja, nein
attributiv	ja, nein
adjektivmodifizierend	ja, nein
adjunktiv	ja, nein
Valenz_NomE	vererbbar, nicht_vererbbar, vererbbar_fuer, vererbbar_von, vererb- bar_fuer_von, ?, -
Valenz_AkkE	vererbbar, vererbbar_NomE, nicht_vererbbar, -
Valenz_DatE	vererbbar, vererbbar_NomE, nicht_vererbbar, -
Valenz_praepE	vererbbar, vererbbar_NomE, nicht_vererbbar, -
Motsch	faehig_agens, faehig_patiens, -
Innsbruck	1, 2, 3, 4, 5, 6, R, -
Intentionalitaet	stark, schwach, nicht_intentional, -
idiosynkratisch	ja, nein
Gradierbarkeit	ja, nein
dauerhaft	ja, nein
Bereich_Herkunft	Alltagssprache, Fachsprache
deverbal_Verbart	Handlung, Vorgang, Zustand, -
modaler_Nebensinn	darf, kann, kann_leicht, muss, soll, -
un_lich	ja, nein
class	un, nicht

**Anhang C: Beispieldatensätze**

Attribute	Beispiele		
	<i>einklagbar</i>	<i>fruchtbar</i>	<i>wahrnehmbar</i>
Anzahl_Silben	3	2	3
Akzentbelegung_Silbe_1	H	H	H
Akzentbelegung_Silbe_2	U	U	U
Akzentbelegung_Silbe_3	U	-	U
Akzentbelegung_Silbe_4	-	-	-
Akzentbelegung_Silbe_5	-	-	-
Akzentbelegung_Silbe_6	-	-	-
Anfangsrand_Silbe_1	leer	komplex	einfach
Kern_Silbe_1	Diphthong	ungespannt	gespannt
Endrand_Silbe_1	einfach	komplex	einfach
Anfangsrand_Silbe_2	komplex	einfach	einfach
Kern_Silbe_2	gespannt	gespannt	gespannt
Endrand_Silbe_2	einfach	einfach	einfach
Anfangsrand_Silbe_3	einfach	-	einfach
Kern_Silbe_3	gespannt	-	gespannt
Endrand_Silbe_3	einfach	-	einfach
Anfangsrand_Silbe_4	-	-	-
Kern_Silbe_4	-	-	-
Endrand_Silbe_4	-	-	-
Anfangsrand_Silbe_5	-	-	-
Kern_Silbe_5	-	-	-
Endrand_Silbe_5	-	-	-
Anfangsrand_Silbe_6	-	-	-
Kern_Silbe_6	-	-	-
Endrand_Silbe_6	-	-	-
Praefix_Basis	ein	-	wahr
Komplexitaet_Basis	komplex	einfach	komplex
fallengelassene_Bestandteile	-	-	-
Stammende	-	-	-
Derivation	deverbal	desubstantivisch	deverbal
deverbal_stark_schwach	schwach	-	stark

Attribute	Beispiele		
	<i>einklagbar</i>	<i>fruchtbar</i>	<i>wahrnehmbar</i>
Aufbau_Basis	trennbar	-	Kompositum
Basis_Partikelverb	ja	-	ja
praedikativ	ja	ja	ja
attributiv	ja	ja	ja
adjektivmodifizierend	nein	nein	nein
adjunktiv	ja	ja	ja
Valenz_NomE	vererbbar_fuer_von	-	vererbbar_fuer_von
Valenz_AkkE	vererbbar_NomE	-	vererbbar_NomE
Valenz_DatE	-	-	-
Valenz_praepE	-	-	-
Motsch	faehig_patiens	-	faehig_patiens
Innsbruck	1	5	1
Intentionalitaet	stark	-	schwach
idiosynkratisch	nein	ja	nein
Gradierbarkeit	nein	ja	ja
dauerhaft	ja	ja	nein
Bereich_Herkunft	Fachsprache	Alltagssprache	Alltagssprache
deverbal_Verbart	Handlung	-	Vorgang
modaler_Nebensinn	kann	-	kann
un_lich	nein	nein	nein
class	nicht	un	nicht

## **Selbstständigkeitserklärung**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt, nicht anderweitig zu Prüfungszwecken vorgelegt und keine anderen als die angegebenen Hilfsmittel verwendet habe. Sämtliche wissentlich verwendeten Textauschnitte, Zitate oder Inhalte anderer Verfasser wurden ausdrücklich als solche gekennzeichnet.

Chemnitz, den 6. März 2006

*Christiane Reiß*

Christiane Reiß