

Friedrich-Schiller-Universität Jena

Philosophische Fakultät

Institut für Germanistische Sprachwissenschaft

Die Skalierung epistemisch modaler Ausdrücke

Eine empirische Untersuchung mit Fokus auf biomedizinischen Texten

Masterarbeit zur Erlangung des akademischen Grades
Master of Arts (M.A.)

vorgelegt von Christine Engelmann
Matrikelnummer: 93803

geboren am 27.06.1987 in Dresden

Erstgutachter: Prof. Dr. Udo Hahn
Zweitgutachter: Prof. Dr. Holger Diessel

Jena, den 15.10.2013

Inhaltsverzeichnis

0	Einleitung	7
I	Theoretischer Teil	9
1	Relevante Konzepte	10
1.1	Epistemische Modalität	10
1.2	Evidentialität	12
1.3	Subjektivität	13
1.4	Faktualität	14
1.5	Hedges	15
2	Relevante sprachliche Mittel	16
2.1	Überblick zu den Wortarten	16
2.2	Einbezug des sprachlichen Kontextes	18
3	Epistemische Modalität in Wissenschaftstexten	20
3.1	Hedges als Strategie im Wissenschaftsdiskurs	20
3.2	Hedge-Arten in biologischen Texten	23
3.3	Hedge-Kombination in biologischen Texten	26
4	Die theoretische Skalierung	27
4.1	Die epistemische Skala als Kontinuum	27
4.2	Die epistemische Skala als diskrete Skala	29
4.3	Fazit zur theoretischen Skalierung	32
5	Vorhandene Annotations- und Klassifizierungsarbeiten	34
5.1	Arbeiten außerhalb des biomedizinischen Bereichs	34
5.2	Arbeiten im biomedizinischen Bereich	37
5.2.1	Frühe Arbeiten	37
5.2.2	BioScope-Korpus	39
5.2.3	Mehrdimensionalität	40
5.2.4	Meta-Knowledge-GENIA-Event-Korpus	41
5.2.5	Kombination von Sicherheit und Polarität	43
5.3	Fazit zu vorhandenen Arbeiten	44
II	Empirischer Teil	45
6	Fragestellung	46
6.1	Bedeutungsaspekt der epistemischen Skala	46
6.2	Empirische Fundierung	47

6.3	Ungenauigkeit der Ausdrücke	48
6.4	Systematische Bewerterunterschiede	49
6.5	Rolle des sprachlichen Kontextes	49
6.6	Fragen und vorläufige Hypothesen	50
7	Methodik	51
7.1	Auswahl der epistemisch modalen Ausdrücke	51
7.2	Konstruktion der Befragungssitems	53
7.3	Variablen und statistische Hypothesen	55
7.4	Durchführung	59
7.5	Probandenrekrutierung	61
8	Ergebnisse	63
8.1	Explorative Datenanalyse	63
8.1.1	Kritische Punkte	63
8.1.2	Auswahl gültiger Fälle	64
8.2	Überprüfung systematischer Bewerterunterschiede	69
8.3	Positionierung und Gruppierung der Ausdrücke	71
8.3.1	Darstellung der Wahrscheinlichkeitsbereiche	71
8.3.2	Genauere Beschreibung der Wahrscheinlichkeitsbereiche	71
8.3.3	Gruppierung der Ausdrücke	74
8.4	Überprüfung der Ungenauigkeits-Unterschiede	75
8.5	Überprüfung der Unterschiede im epistemischen Grad	79
8.6	Nicht epistemische Vorkommen	82
9	Diskussion	83
9.1	Angemessenheit der Skalenkonzeption	83
9.2	Fazit zu Bewerterunterschieden	86
9.3	Festlegung eines Kategoriensystems	89
9.4	Fazit zum sprachlichen Kontext	95
10	Schlussbetrachtung	98
A	Anhang	101
A.1	Liste der Befragungssitems	101
A.2	Fragebogen	106
	Literaturverzeichnis	109
	Eigenständigkeitserklärung	115

Abkürzungsverzeichnis

ANOVA	Analysis of Variance/Varianzanalyse
AV	abhängige Variable
D	D-Statistik (Kolmogorov-Smirnov-Test)
F	F-Statistik (Varianzanalyse)
F_1	F-Score (Relevanzmaß)
h	relative Häufigkeit
H_0	Nullhypothese
H_1	Alternativhypothese
k	Cohen's Kappa (Übereinstimmungsmaß)
M	Mittelwert
Max	Maximum
Md	Median
Min	Minimum
n	absolute Häufigkeit
p	Aussage
p	p-Wert/Signifikanz
$partial \eta^2$	partielles Eta-Quadrat (Effektstärke)
r	Pearson's r (Effektstärke)
$Range$	Spannweite
SD	Standardabweichung
U	U-Statistik (Mann-Whitney-U-Test)
UV	unabhängige Variable
α	Signifikanzniveau
ϵ	Greenhouse-Geisser-Epsilon (Korrekturfaktor)
χ^2	Chi-Quadrat-Statistik
#	pragmatische Unangemessenheit

Anmerkung zur Darstellung von Zahlen: In Anlehnung an die Publikationsvorgaben der American Psychological Association (APA) entfällt bei der Präsentation von Zahlen die Null vor dem Dezimalzeichen, wenn der statistische Kennwert nicht größer als 1 sein kann (z. B. $r = -.40$, $p = .001$). (vgl. American Psychological Association, 2010, S.113f)

Abbildungsverzeichnis

1	Parameter der Evidentialität bei Chafe (1986, S.263)	12
2	Klassifikation der Hedge-Verwendung bei Hyland (1996, S.438)	21
3	Epistemische Skala bei Holmes (1982, S.13)	31
4	Epistemische Skala bei Rubin (2010, S.536)	34
5	Eventannotation im GENIA-Event-Korpus	39
6	Negations- und Spekulationsannotation im BioScope-Korpus	40
7	Annotationsschema bei Thompson et al. (2011, S.398)	42
8	Sicherheitsannotation im Meta-Knowledge-GENIA-Event-Korpus	42
9	Epistemische Skala bei Velupillai (2012, S.34)	43
10	Histogramm zur Werteverteilung bei <i>unlikely</i>	66
11	Histogramm zur Werteverteilung bei <i>no evidence</i>	67
12	Histogramm zur Werteverteilung bei <i>cannot</i>	67
13	Balkendiagramm zur Bewertung nach Fachbereich	70
14	Balkendiagramm zur Bewertung nach englischer Sprachkompetenz	70
15	Hoch-Tief-Schluss-Diagramm zu den Wahrscheinlichkeitsbereichen	71
16	Dendrogramm zur hierarchischen Clusteranalyse	75
17	Boxplots zur Ungenauigkeit	76
18	Hoch-Tief-Schluss-Diagramm zu 99.8%-Konfidenzintervallen der Ungenauigkeits- Mittelwerte	78
19	Boxplots zum epistemischen Grad	79
20	Hoch-Tief-Schluss-Diagramm zu 99.8%-Konfidenzintervallen der Mittelwer- te des epistemischen Grades	81
21	Histogramm zur Werteverteilung bei <i>unknown</i> (kein Wahrheitswert)	82
22	Histogramm zur Werteverteilung bei <i>unclear</i> (kein Wahrheitswert)	82
23	Einleitungsteil des Fragebogens mit Aufgabenerklärung	106
24	Mittelteil des Fragebogens mit Items und Wahrscheinlichkeitsskala	107
25	Schlusssteil des Fragebogens mit Angaben zur Person	108

Tabellenverzeichnis

1	Ausgewählte Ausdrücke	52
2	Ausgewählte kritische Ausdrücke	52
3	Ausgewählte Vergleiche der sprachlichen Gesamtkonstruktion	53
4	Ausgewählte Vergleiche bei Kombination	53
5	Ausgewählter Vergleich bei Modifizierung	53
6	Beschreibung der Stichprobe hinsichtlich Geschlecht, Alter und Vertrautheit mit Wissenschaftstexten	62
7	Beschreibung der Stichprobe hinsichtlich Muttersprache und englischer Sprachkompetenz	62
8	Häufigkeit des Antwortverhaltens beim Testitem	63
9	Bewertung der kritischen Ausdrücke	64
10	Häufigkeit des Antwortverhaltens beim Testitem (Fachbereich Biomedizin) .	65
11	Häufigkeit des Antwortverhaltens bei kritischen Ausdrücken	68
12	Kennwerte zu den Variablen der Wahrscheinlichkeitsbereiche	72
13	Kennwerte zu den Variablen der Skalenposition	73
14	Präsentation der Drei-Cluster-Lösung	74
15	Signifikante Ungenauigkeits-Unterschiede zwischen Ausdrücken	77
16	Signifikante Unterschiede im epistemischen Grad zwischen Ausdrücken . . .	80
17	Vorgeschlagene diskrete epistemische Skala	94
18	Auflistung der konstruierten Items (3 Sätze pro Ausdruck)	101

0 Einleitung

Das alte Wissenschaftsideal, das absolut gesicherte Wissen [episteme], hat sich als ein Idol erwiesen. Die Forderung der wissenschaftlichen Objektivität führt dazu, daß jeder wissenschaftliche Satz *vorläufig* ist. Er kann sich wohl bewähren – aber jede Bewährung ist relativ, eine Beziehung, eine Relation zu anderen, gleichfalls vorläufig festgesetzten Sätzen. Nur in unseren subjektiven Überzeugungserlebnissen, in unserem Glauben können wir „absolut sicher“ sein.

Popper, 1969, S.225

Die Biomedizin, als Forschungsschnittstelle zwischen Biologie und Medizin, verzeichnet einen exponentiellen Zuwachs an Publikationen und wissenschaftlicher Literatur. Um dieser Informationsflut Herr zu werden und um Forscher mit effizienten Mitteln zur Lokalisierung der Information zu versorgen, ist die Automatisierung im Umgang mit den publizierten Texten erforderlich. Zu solchen automatisierten Methoden gehört die Informationsextraktion, d. h. die Extraktion von Fakten aus unstrukturierten Textdaten bezüglich zuvor festgelegter Inhalte. Die domänenspezifische Anpassung der dabei zum Einsatz kommenden, auf die Semantik ausgerichteten computerlinguistischen Systeme erfolgt mithilfe maschineller Lernverfahren, die auf der Grundlage bereits annotierter Dokumente trainiert werden. (vgl. Thompson et al., 2011, S.393f)

Ein nicht zu unterschätzendes Problem ist hierbei allerdings die Tatsache, dass in Wissenschaftstexten oft keine reinen und unumstrittenen Fakten präsentiert werden. Tatsächlich verwenden die Autoren eine Vielzahl an sprachlichen Mitteln, um kenntlich zu machen, dass für eine Aussage kein uneingeschränkter Geltungsanspruch erhoben wird. Bei der Informationsextraktion ist es daher notwendig, auch den sprachlichen Kontext einzubeziehen, in den zentrale Sachverhalte eingebettet sind. Die Äußerung in (1) macht dies deutlich.

- (1) *These results **suggest** that the narL gene product **might** be activated by the nitrate reductase operon.* (Thompson et al., 2011, S.394)

Der in (1) beschriebene biologische Prozess wird nicht als vollkommen sicher oder hundertprozentig wahrscheinlich dargestellt. Stattdessen gibt der Autor explizit an, dass die zentrale Aussage das Resultat eines Schlussfolgerungsprozesses ist. Die Grundlage für diese Schlussfolgerung sind Ergebnisse aus vorherigen Untersuchungen. Diese sind ein Hinweis für die Richtigkeit der Aussage, aber wahrscheinlich kein absoluter Beweis.

In der Linguistik werden Wörter wie *suggest* und *might* zum Bereich der epistemischen Modalität gezählt. Dabei steht u. a. die Frage im Raum, inwieweit solche

Wörter eine Skalierung erlauben. Ist es demnach möglich, aufgrund des sprachlichen Kontextes festzulegen, dass eine Aussage eher einem Fakt entspricht als eine andere? Eine derartige inhaltliche Klassifizierung wäre auch eine wertvolle Erweiterung der computerlinguistischen Systeme, da es Forschern so ermöglicht würde, definitive von leicht spekulativen oder stark spekulativen Inhalten zu unterscheiden. Eine Reihe bereits vorhandener Annotationsarbeiten widmet sich dieser Problematik, indem entsprechende Kategorien zur Klassifizierung wissenschaftlicher Inhalte vorgeschlagen werden. Diese Kategoriensysteme sind aber meist sehr subjektiv, da sie oft auf Einzelurteilen basieren oder in ihrer Bedeutung nicht klar genug definiert sind. Ein solches Vorgehen ist besonders fragwürdig, wenn man bedenkt, dass Wörter wie *suggest* und *might* semantisch eher vage und entsprechend variabel in ihrer Interpretation sind.

In der vorliegenden Masterarbeit soll auf solche kritischen Punkte näher eingegangen werden. Das Ziel ist es, eine epistemische Skala zu konstruieren, die für die Klassifizierung und Annotation wissenschaftlicher Inhalte in biomedizinischen Texten angemessen ist. Dabei sollen sowohl theoretische Überlegungen als auch empirische Befunde Beachtung finden.

Im theoretischen Teil der Arbeit (Teil I) werden die relevanten sprachlichen Mittel zunächst aus linguistischer Perspektive näher beleuchtet. Hierzu gehören die Betrachtung verschiedener linguistischer Konzepte (Kapitel 1) sowie ein Überblick zu den unterschiedlichen Ausdrucksmöglichkeiten (Kapitel 2). Kapitel 3 überträgt diese allgemeine Darstellung auf den speziellen Bereich der Wissenschaftstexte, da die Verwendung der Ausdrücke hier einer besonderen Motivation unterliegt. Daraufhin wird auf die theoretische (Kapitel 4) sowie in den Annotationsarbeiten bereits vorgenommene Skalierung (Kapitel 5) eingegangen. Die Erkenntnisse aus dem ersten Teil der Arbeit sind der Ausgangspunkt für eine eigene Studie, die im empirischen Teil der Arbeit (Teil II) beschrieben wird. Die Annahme ist hier, dass die Befragung einer Menge von Personen ein repräsentativeres und differenziertes Bild zu den Möglichkeiten der Skalierung sowie zu den dabei zu beachtenden Punkten liefern kann. Kapitel 6 präsentiert die zentrale Fragestellung. Anschließend werden Methodik (Kapitel 7) und Ergebnisse (Kapitel 8) der empirischen Untersuchung vorgestellt. Die Diskussion in Kapitel 9 soll die Resultate zu einer möglichen Skalierung zusammenfassen sowie Implikationen bezüglich der Annotation biomedizinischer Texte und der Klassifizierung der darin enthaltenen wissenschaftlichen Inhalte aufzeigen.

Teil I

Theoretischer Teil

1 Relevante Konzepte

Aus linguistischer Perspektive gibt es eine ganze Reihe von Konzepten, die mit den hier betrachteten Ausdrücken zur Einschränkung des Geltungsanspruchs einer Aussage in Verbindung stehen. Dieses Kapitel soll die relevanten Konzepte kurz umreißen, wichtige Begriffe definieren und den Untersuchungsgegenstand so in der linguistischen Forschungslandschaft positionieren.

1.1 Epistemische Modalität

Das Konzept der Modalität hat seinen Ursprung im Bereich der Logik. Die sogenannte Modallogik arbeitet mit den Begriffen der Möglichkeit und Notwendigkeit und stellt die Idee der möglichen Welten ins Zentrum. Ein Sachverhalt gilt als möglich, wenn er in mindestens einer der möglichen Welten vorliegt, und als notwendig, wenn das in allen möglichen Welten der Fall ist. In der formalen Notation wird dies durch bestimmte Modaloperatoren gekennzeichnet, die eine Aussage p wie in (2) modifizieren. (vgl. Perkins, 1983, S.1-12)

- (2) $\Diamond p$ - „Es ist möglich, dass p .“
 $\Box p$ - „Es ist notwendig, dass p .“

In der Linguistik hingegen gibt es für die Modalität keine wirklich einheitliche Definition, denn die Betrachtungen sind zahlreich und in ihrer Sichtweise sehr heterogen. (vgl. Nurmi, 2007) So orientiert sich bspw. die formale Semantik stark an der Modallogik. In der sprachtypologischen Forschung steht dagegen die grammatische Realisierung im Vordergrund, während unter pragmatisch-kognitiver Perspektive v. a. die Motivation für die Verwendung verschiedener modaler Ausdrücke beleuchtet wird. Zentral scheint allerdings stets die Feststellung zu sein, dass ein Sprecher mithilfe der Modalität seine Einstellung gegenüber einem Sachverhalt sprachlich kenntlich machen kann. (vgl. Palmer, 2001, S.1) Dies wird auch beim folgenden Lexikoneintrag deutlich:

[Die Modalität ist eine] semantisch-pragmatische Beschreibungsperspektive, welche sich im weiteren Sinne auf die Art und Weise der Stellungnahme des Sprechers zur Geltung des durch eine Äußerung ausgedrückten Sachverhaltes in der aktuellen Welt bezieht. Modalitätsunterschiede in der Bedeutung von Sätzen beziehen sich auf alternative Möglichkeiten zu den in der aktuellen Welt bestehenden Gegebenheiten. (Fries, 2010, S.433)

Tatsächlich kann ein Sprecher unterschiedliche Einstellungen gegenüber einem Sachverhalt manifest machen. Vorrangig Modalverben, die ein mögliches Ausdrucksmittel der Modalität darstellen, sind in dieser Hinsicht ambig.

- (3) a. *She **must** be at work.*
 b. *She **can** play cards.*

Für (3a) bieten sich zwei Interpretationen an. Man könnte einerseits auf Grundlage entsprechender Hinweise gezwungen sein, darauf zu schließen, dass die Person bei der Arbeit ist. Es kann sich aber genauso gut um eine Anweisung handeln („Ich verlange, dass sie bei der Arbeit ist.“). Im ersten Fall spricht man von epistemischer, im zweiten Fall von deontischer Modalität. Die deontische Modalität bezieht sich im Kern auf Verpflichtung und Erlaubnis. Auch (3b) hat zwei Lesarten. Eine davon ist deontisch („Sie darf Karten spielen.“), die andere hat mit der individuellen Fähigkeit der Person zu tun („Sie ist in der Lage, Karten zu spielen.“) und wird als dynamische Modalität bezeichnet, die sich im Kern auf die Fähigkeit und den Willen bezieht. (vgl. Huddleston, 2010, S.177ff) Deontische und dynamische werden in Abgrenzung zur epistemischen Modalität oft als Root-Modalität (im Sinne von „grundlegender Modalität“) zusammengefasst. (vgl. Simon-Vandenberg & Aijmer, 2007, S.3)

Über die Anzahl der verschiedenen Modalitätsarten herrscht keine Einigkeit. Eine Festlegung hierzu ist für das weitere Vorgehen allerdings auch nicht essenziell. An dieser Stelle soll primär festgehalten werden, dass die in der vorliegenden Arbeit betrachteten Ausdrücke der epistemischen Modalität zuzuordnen sind, da diese prototypisch den eingeschränkten Geltungsanspruch für eine Aussage aufseiten des Sprechers anzeigt. (vgl. Coates, 1987, S.112) Bei der Definition wird insbesondere auf den linguistischen Begriff der Proposition zurückgegriffen. Eine Proposition ist das, was über die Welt geäußert wird, entspricht also dem Äußerungsgehalt. Sie kann wahr oder falsch sein. (vgl. Linke et al., 2004, S.202) Ist sie wahr, handelt es sich um einem Sachverhalt. (vgl. Lehmann, 2012) Epistemisch modale Ausdrücke können eine Proposition zusätzlich modifizieren. Der Sprecher schränkt damit seine Festlegung gegenüber dem Wahrheitswert der Proposition ein:

Any utterance in which the speaker explicitly qualifies his commitment to the truth of the proposition expressed by the sentence he utters [...], is an epistemically modal, or modalized, utterance. (Lyons, 1977, S.797)

So entspricht bspw. die Proposition, die durch den Satz in (4a) geäußert wird, einer kategorischen Aussage. Bei (4b) hingegen handelt es sich um eine modalisierte Äußerung. Dieselbe Proposition (dass die Person bei der Arbeit ist) wird durch das modale Adverb *probably* modifiziert, d. h. der Sprecher erachtet die Proposition nicht uneingeschränkt für wahr.

- (4) a. *She is at work.*
 b. *She is **probably** at work.*

1.2 Evidentialität

Die epistemische Modalität ist nicht klar von bestimmten anderen Konzepten zu trennen. Hierzu zählt die Evidentialität. Die sprachlichen Mittel beider Phänomene markieren die Festlegung des Sprechers bezüglich des Wahrheitswerts der Proposition. Während die epistemische Modalität allerdings die Einschätzung selbst indiziert, zeigt die Evidentialität die Informationsquelle für diese Einschätzung an. (vgl. Simon-Vandenberg & Aijmer, 2007, S.24) In der Sprachtypologie wird die Evidentialität mitunter als eigenständige grammatische Kategorie mit rein morphosyntaktischer Realisierung definiert. (vgl. Aikhenvald, 2004, S.6) Es finden sich aber auch Betrachtungen, die lexikalische Mittel einbeziehen.

Palmer (2001) bspw. unterscheidet eine evidentielle Modalität von der epistemischen Modalität, fasst beide allerdings als propositionale Modalität zusammen. Dazwischen verzeichnet er Überschneidungen. Dies betrifft v. a. den Bereich der Schlussfolgerungen, also der auf Evidenz basierenden Einschätzungen. Die Interpretation bestimmter Modalverben macht dies deutlich.

(5) *She **must** be at work.*

In (5) gibt der Sprecher eine sichere Einschätzung darüber ab, dass die Person bei der Arbeit ist. Eine solche Einschätzung ist nach Palmer immer auch evidenzbasiert (z. B. könnte das Licht im Büro an sein), auch wenn die Einschätzungsgrundlage nicht explizit genannt wird. (vgl. Palmer, 2001, S.8f,24f)

Chafe (1986) präsentiert eine weite Auffassung der Evidentialität, die sich nicht ausschließlich auf Evidenz, sondern insgesamt auf Einstellungen gegenüber Wissen bezieht, wobei das Wissen hier mit der Proposition gleichgesetzt werden kann. Die Evidentialität umfasst nach Chafe die in Abbildung 1 aufgeführten Parameter.

source of knowledge	mode of knowing	knowledge matched against
		reliable
		k
		n
???	---> belief	---> o
evidence	---> induction	---> w ---> verbal resources
language	---> hearsay	---> l ---> expectations
hypothesis	---> deduction	---> e
		d
		g
		e
		unreliable

Abbildung 1: Parameter der Evidentialität bei Chafe (1986, S.263)

Im Zentrum der Skizze steht das Wissen. Der Status dieses Wissens kann durch Evidentialitätsmarker modifiziert werden. Diese zeigen verschiedene Arten des Wissenserwerbs an, die wiederum unterschiedliche Wissensquellen als Grundlage haben. Dabei wird erstens zwischen dem nicht ausschließlich auf Evidenz basierenden Glauben (s. (6a) für entsprechende Ausdrücke), zweitens der auf direkter oder

indirekter Evidenz basierenden Induktion (s. (6b)), drittens dem auf sprachlicher Information basierenden Hörensagen (s. (6c)) und viertens der Deduktion unterschieden, deren Grundlage Hypothesen sind (s. (6d)). Viele diesen Parameterwerten zugeordnete Ausdrücke können kontextabhängig in ihrer Einstufung variieren.

- (6) a. *think, guess, suppose*
- b. *must, obvious, seem, evidently, look like, sound like, feel*
- c. *have been said to*¹
- d. *should, presumably, can, could, would*
- e. *of course, oddly enough, in fact, actually*

Das Wissen selbst kann schließlich mit vorherigen Erwartungen abgeglichen werden. Ausdrücke, die diesen Abgleich kenntlich machen, sind unter (6e) aufgeführt.

Epistemische Modalität und Evidentialität sind offensichtlich nur schwer separat zu behandeln. Im weiteren Verlauf der Arbeit sollen deshalb beide Konzepte gemeinsam betrachtet werden. Zwar gibt es semantische Unterschiede; die zahlreichen Überschneidungen legen aber nahe, dass bei der Interpretation epistemisch modaler Ausdrücke sowohl die Einschätzung an sich als auch der evidentielle Status eine Rolle spielen. So kann die Festlegung zum Wahrheitswert einer Proposition durchaus durch die Art der Evidenz eingeschränkt werden. (vgl. Palmer, 1986, S.54)

1.3 Subjektivität

Auch das Konzept der Subjektivität, welches sich auf Perspektive und Standpunkt eines Sprechers bezieht, spielt bei epistemisch modalen Ausdrücken eine Rolle. (vgl. Simon-Vandenberg & Aijmer, 2007, S.33) Subjektivität kann einerseits als übergeordnete Dimension betrachtet werden, die auch die epistemische Modalität umfasst. Dies ist bspw. in den anwendungsbezogenen Studien von Wiebe et al. (2004)² der Fall. Subjektivität bezieht sich hier auf die Verwendung von Sprache zum Ausdruck von Meinungen, Emotionen, Bewertungen und Vermutungen und steht im Kontrast zur objektiven Präsentation faktischer Information. Die epistemische Modalität ist dabei dem Bereich der Vermutungen bzw. Spekulationen zuzuordnen. Diese Herangehensweise deckt sich z. T. mit Standpunkten aus der Linguistik. So gibt z. B. Palmer (1986) an, dass die Subjektivität ein obligatorisches Merkmal der epistemischen Modalität ist:

Modality could [...] be defined as the grammaticalization of a speaker's (subjective) attitudes and opinions. (Palmer, 1986, S.16)

¹In Wissenschaftstexten zählen hierzu v. a. Zitate. (vgl. Chafe, 1986, S.268f)

²Hierzu gehören verschiedene Annotationsprojekte. Die Texte (z. B. aus dem Wallstreet Journal Treebank Korpus) sind auf unterschiedlichen Ebenen (Dokument-, Satz-, Wortebene) bezüglich Subjektivität annotiert. Das Ergebnis ist das MPQA Opinion Korpus, welches zusammen mit anderen Ressourcen zur Markierung und Erkennung von Subjektivität unter folgender Adresse zu finden ist: <http://mpqa.cs.pitt.edu/>

Andererseits wird Subjektivität oft als Parameter innerhalb der epistemischen Modalität angesehen. Lyons (1977) bspw. führt den Kontrast subjektiv-objektiv auf die fachspezifische Perspektive zurück. Während sich die Linguistik mit der Sprechereinstellung auseinandersetzt, beschäftigt sich die Logik mit der Evidenz, die die Möglichkeit oder Notwendigkeit einer Aussage bestimmt. Lyons greift diese Unterscheidung auf, indem er zwischen einer subjektiven und einer objektiven epistemischen Modalität differenziert. Dies lässt sich an folgendem Beispiel illustrieren.

(7) *Alfred **may** be unmarried.*

Die Äußerung in (7) kann einerseits so interpretiert werden, dass der Sprecher seine Festlegung subjektiv modifiziert, d. h. Unsicherheit ausdrückt. Unter bestimmten Umständen (In einer Menge von 90 Leuten z.B., zu denen Alfred gehört, sind 30 verheiratet, aber auf wen genau dies zutrifft, ist unbekannt.) kann die gesamte Äußerung, also das Vorliegen einer Möglichkeit, aber auch als objektiver Fakt gewertet werden. Es liegt dann eine quantifizierbare Wahrscheinlichkeit auf der Grundlage mathematisch fundierter Evidenz vor, dass die Proposition wahr ist. (vgl. Lyons, 1977, S.797f) Eine derartige Unterscheidung wirkt auf den ersten Blick ziemlich konstruiert.³ Trotzdem sind solche Überlegungen für die vorliegende Arbeit interessant. So scheint es bei der epistemischen Modalität zu einem gewissen Grad möglich zu sein, eine inhaltliche Trennung zwischen der (subjektiven) Festlegung eines Sprechers zum Wahrheitswert der Proposition und der (objektiven) Angabe zur Wahrscheinlichkeit des beschriebenen Sachverhalts vornehmen zu können. Dies deutet darauf hin, dass die epistemische Modalität unterschiedliche Bedeutungsaspekte aufweist.

1.4 Faktualität

Das Konzept der Faktualität ist ebenfalls im Zusammenhang mit der epistemischen Modalität zu erwähnen. Lyons (1977) bezeichnet Äußerungen, die epistemisch modale Ausdrücke beinhalten und in denen der Sprecher weder Wahrheit noch Falschheit der Proposition annimmt, als nicht-faktiv. (vgl. Lyons, 1977, S.795f) Auch in der anwendungsbezogenen Literatur außerhalb der traditionellen Linguistik spielt die Faktualität eine Rolle. Bei Saurí (2008)⁴ ist darunter z. B. eine Informationsebene zu verstehen, die sich auf den faktischen Status von Ereignissen im Text bezieht. Die epistemische Modalität wird hier als grundlegendes Konzept angesehen. Dieses wird allerdings erweitert. Ein Ereignis kann nämlich auch anderweitig als nicht-faktisch gekennzeichnet werden. So liegt z. B. bei der Verwendung der Negationspartikel in (8a) oder durch das Verb *fail* in (8b) ein kontrafaktisches Ereignis vor. Bei anderen

³Lyons gibt dies selbst zu Beginn seiner Ausführungen zu: „This is not a distinction that can be drawn sharply in the everyday use of language; and its epistemological justification is, to say the least, uncertain“ (Lyons, 1977, S.797).

⁴Ergebnis dieser und verwandter Arbeiten ist u. a. FactBank, ein nach Faktualität annotiertes Korpus. (Saurí & Pustejovsky, 2009)

Verben wie *want* in (8c) kann die Faktualität eines Ereignisses auch unterspezifiziert bleiben. (vgl. Saurí, 2008, S.viii,44-57)

- (8) a. *The size of the contingent was **not** disclosed.* (Saurí, 2008, S.30)
- b. *She **failed** to follow the rules.* (Saurí, 2008, S.247)
- c. *The GOP just **wants** to use the tax issue as an electoral weapon.* (Saurí, 2008, S.49)

Festzuhalten bleibt: Epistemisch modale Ausdrücke kennzeichnen, dass eine Äußerung keinen Fakt beinhaltet. Andererseits gibt es viele weitere sprachliche Mittel, die dies ebenfalls markieren können. Diese werden in der vorliegenden Arbeit nicht näher betrachtet. Insbesondere betrifft das die Negation, die ein ebenso großes Untersuchungsfeld – v. a. hinsichtlich der Informationsextraktion – darstellt.

1.5 Hedges

Als letztes Konzept sollen hier die Hedges – die sogenannten Heckenausdrücke – angesprochen werden. Die Bezeichnung wurde von Lakoff (1972) geprägt. In der ursprünglichen Konzeption sind hierunter Ausdrücke zu verstehen, die Ungenauigkeiten bei der Verwendung von „Prädizierungen begrifflicher Art“ (Graefen, 2000) anzeigen, wobei diese Ungenauigkeiten durch mehrwertige Logiken⁵ beschrieben werden. Hedges wie *sort of* oder *rather* werden demnach eingesetzt, wenn das mit einem Wort Gemeinte nicht dem Ideal entspricht und der Sprecher diesen Zweifel verbalisieren möchte. Das Konzept wurde allerdings im späteren Verlauf erweitert, um auch die Modifizierung ganzer Propositionen einzuschließen. Kern der Hedge-Forschung sind mittlerweile die Motive für den Gebrauch. Man geht hier davon aus, dass ihre Rolle die eines verbalen Schutzschildes ist und sie dazu verwendet werden, das Gewicht einer Proposition abzuschwächen. Der Sprecher entzieht sich der Verantwortung für den Wahrheitsgehalt seiner Aussage und schützt sich damit vor Kritik und Ablehnung. (vgl. Hübler, 1983) Solche Motive werden v. a. in Hinblick auf Wissenschaftstexte näher untersucht. Epistemisch modale Ausdrücke können in diesem Zusammenhang als die wichtigsten Hedge-Realisatoren angesehen werden. (vgl. Clemen, 1998, S.29) Das Konzept ist allerdings noch weiter gefasst und vereint eine ganze Bandbreite an Strategien, mittels derer wissenschaftliche Aussagen einem Fachpublikum präsentiert werden können, ohne dass sich der Autor gegenüber diesen Aussagen eindeutig festlegt. Bevor auf solche Besonderheiten bei der Hedge-Verwendung in Wissenschaftstexten genauer eingegangen wird, soll im folgenden Kapitel vorerst ein allgemeiner Überblick zu den verschiedenen sprachlichen Mitteln der epistemischen Modalität gegeben werden.

⁵Hiermit ist die Fuzzy-Set-Theorie nach Zadeh (1965) gemeint, nach der ein Element zu einem bestimmten Grad einer Menge angehören kann. Es gibt bei solchen Mengen ein Kontinuum der Klassifikation. (vgl. Clemen, 1997, S.236)

2 Relevante sprachliche Mittel

Die Sprache bietet viele Möglichkeiten, um auszudrücken, dass gegenüber dem Wahrheitswert einer Proposition keine eindeutige Festlegung erfolgt. Es ist deshalb sinnvoll, die verschiedenen sprachlichen Mittel in formaler Hinsicht zu kategorisieren.

2.1 Überblick zu den Wortarten

Linguistische Arbeiten zur Modalität beschränken sich meist auf die Analyse von Modalverben, da diese (zumindest im Englischen) die einzige formal kohärente Gruppe modaler Ausdrücke darstellen. Damit sind sie leicht als syntaktische Klasse zu isolieren und können semantisch beschrieben werden. Andere sprachliche Mittel tauchen oft nur als stilistische Paraphrasierungen der Modalverbbedeutung auf, wobei bspw. Gruppierungen wie in (9) vorgenommen werden.

(9) *may – perhaps, it is possible that, there is a possibility that*

In der Sprachtypologie, die die Modalität vorrangig als grammatische Kategorie (ähnlich Tempus und Aspekt) definiert, ist darüber hinaus die sprachliche Realisierung in Form des Modus zentral. So weist das Deutsche z. B. die drei Modi Indikativ, Konjunktiv und Imperativ auf, wobei der Konjunktiv wie in (10) epistemisch gebraucht werden kann.

(10) *Sie **würde** nach Hause kommen.*

Im Englischen hingegen ist der Subjunktiv als Modus fast vollständig verschwunden. (vgl. Palmer, 2001, S.1-4,19) Solche engen Auffassungen zur Modalität sind für die vorliegende Arbeit eher ungeeignet. Eine angebrachtere ausgeweitete Betrachtung liefert hier Perkins (1983). Die folgende Darstellung konzentriert sich auf die von ihm beschriebenen sprachlichen Mittel zum Ausdruck von epistemischer Modalität.

An erster Stelle führt auch Perkins die Klasse der Modalverben an, wobei er zwischen primären (s. (11a)) und sekundären Modalverben (s. (11b)) unterscheidet. Letztere weisen ein zusätzliches (kontextabhängiges) semantisches Merkmal auf. So können Äußerungen, die sekundäre Modalverben enthalten, z. B. als hypothetischer, formaler, höflicher, vorsichtiger oder indirekter aufgefasst bzw. in temporaler Hinsicht eher in der Vergangenheit verortet werden.

(11) a. *can, may, must, will, shall*
b. *could, might, ought to, would, should*

Für *can* ist anzumerken, dass eine epistemische Verwendung durchaus umstritten ist. Dieses Modalverb wird meist eher als Ausdrucksmittel für deontische und dynamische, d. h. für Root-Modalität, angesehen, da „CAN [...] is not directly concerned

with the truth of propositions but rather with the disposition of circumstances vis-à-vis the occurrence of some events“ (Perkins, 1983, S.38).

Weitere modale Ausdrücke listet Perkins vorrangig als komplexe Konstruktionen auf. Hierzu gehören Konstruktionen mit nicht-deverbalen (s. (12a)) und deverbalen Adjektiven (s. (12b)) sowie auf Handlungen (s. (12c)) und mentale Zustände referierenden Partizipien (s. (12d)).

- (12) a. *(be) likely, possible, probable, apparent, evident (to/that)*
b. *(be) arguable, conceivable, debatable, doubtful, questionable (that)*
c. *(be) estimated, hypothesized, inferred, proposed, reported (to/that)*
d. *(be) assumed, believed, doubted, suspected, thought (to/that)*

Die in den Konstruktionen enthaltenen Adjektive können neben der prädikativen Verwendung auch wie in (13) attributiv gebraucht werden.

- (13) *likely outcomes*

Weiterhin gibt es Konstruktionen, die modale Nomen enthalten. Perkins betrachtet Nomen wie in (14), die in der Konstruktion *there is a ... to/that* vorkommen.

- (14) *belief, claim, doubt, hypothesis, observation*

Neben den beschriebenen Konstruktionen stellen modale Adverbien wie in (15) eine weitere Kategorie dar.

- (15) *apparently, clearly, maybe, perhaps, possibly*

Schließlich existieren modale lexikalische Verben wie die in (16) aufgeführten.

- (16) *estimate, believe, feel, suspect, think*

Mit den hier angegebenen Ausdrücken sind die Möglichkeiten zur Verbalisierung von epistemischer Modalität nach Perkins aber noch nicht erschöpft. So kann auch das Tempus als Modalität angesehen werden, insofern ein nicht aktueller Zustand als Zustand in einer der möglichen Welten (aus der Modallogik) definiert wird. Eine nicht spezifisch temporale Lesart liegt bspw. in (17a) und (17b) vor.

- (17) a. *If I **said** that, you would hit me.* (Perkins, 1983, S.107)
b. *Pretend there **was** a ladder on top [...]* (Perkins, 1983, S.107)

Der Kontext zeigt hier an, dass die Imperfekt-Formen nicht auf einen Sachverhalt in der Vergangenheit, sondern auf einen hypothetischen Sachverhalt verweisen. In (17a) ist laut Perkins auch die Konjunktion *if* als epistemisch modaler Ausdruck zu verstehen, da sie die Festlegung des Sprechers zum Wahrheitswert der Proposition einschränkt und Unsicherheit über die Existenz einer Bedingung anzeigt. (vgl. Perkins, 1983, S.61,66-99,102,106-115)

Die Übersicht zeigt, dass die epistemische Modalität mithilfe unterschiedlicher Wortarten ausgedrückt werden kann. Tatsächlich scheint es für einen Sprecher hinsichtlich der verwendeten sprachlichen Mittel aber sogar unendlich viele Möglichkeiten

zu geben.⁶ (vgl. Høye, 2009, S.117) Als Beispiel werden in (18) verschiedene semantische Entsprechungen für die modale lexikalische Verbkonstruktion *I believe that* angeführt. (vgl. Halliday & Matthiessen, 2004, S.616)

- (18) *It stands to reason that [...]* (Halliday & Matthiessen, 2004, S.616)
Nobody tries to deny that [...] (Halliday & Matthiessen, 2004, S.616)
[...] it is particularly difficult to avoid the conclusion that [...] (Halliday & Matthiessen, 2004, S.616)

An dieser Stelle soll deshalb festgehalten werden, dass sich die Betrachtungen in der vorliegenden Arbeit vorrangig auf klar umrissene lexikalische Mittel beziehen. Dies sind die in den westeuropäischen Sprachen vorkommenden Haupttypen des Ausdrucks von Modalität: Modalverben, (epistemisch) modale Adjektive, (epistemisch) modale Adverbien und (epistemisch) modale lexikalische Verben. Im Gegensatz zu den ersten drei bildet die letzte Gruppe dabei eine offene Klasse, sie ist also entsprechend umfangreich und erweiterbar. (vgl. Nuyts, 2001, S.xif,29,108) Darüber hinaus werden in Anlehnung an Perkins (1983) auch (epistemisch) modale Nomen einbezogen. Auf eine Betrachtung des Modus kann hingegen aufgrund der mittlerweile seltenen Verwendung des Subjunktivs im Englischen vorerst verzichtet werden. Andere Realisierungsmöglichkeiten (z. B. Tempus) haben keinen primär epistemisch modalen Status und sind damit ebenfalls nicht zentral. Ein Aspekt, der allerdings beachtet werden sollte, ist die Rolle des sprachlichen Kontextes, in dem diese Wortklassen auftreten.

2.2 Einbezug des sprachlichen Kontextes

Zur Rolle des sprachlichen Kontextes gibt Perkins (1983) an, dass dieser sich auf die semantische Interpretation epistemisch modaler Ausdrücke auswirken kann. So können z. B. epistemische Adjektive (s. (19a)) und Nomen (s. (19b)) weiter modifiziert werden. (vgl. Perkins, 1983, S.102)

- (19) a. *It is **quite possible** that he will go.* (Perkins, 1983, S.102)
b. *There is a **remote possibility** that he will go.* (Perkins, 1983, S.102)

Bei der Betrachtung von Nomen dürfte darüber hinaus schnell deutlich werden, dass der Sprecher unterschiedliche Einschränkungen vornimmt, wenn er bspw. (20a) oder (20b) äußert.

- (20) a. *There is a **claim** that ...*
b. *The **claim** that ... is not warranted.*

⁶Hier sollte erwähnt werden, dass die Menge der relevanten Ausdrucksmittel auch noch um paralinguistische (z. B. Intonation, Lautstärke) und non-verbale Merkmale (z. B. Mimik, Gestik) erweitert werden kann. (vgl. Holmes, 1982, S.11) Diese spielen allerdings bei der Betrachtung von Texten keine Rolle.

Weiterhin ist es möglich, dass mehrere Ausdrücke kombiniert auftreten. Zu den in dieser Hinsicht bisher untersuchten Kombinationen zählen v. a. Modalverb-Modaladverb-Kollokationen wie in (21). (vgl. Hoye, 1997, S.242f)

(21) *could maybe, will probably, must certainly*

Lyons (1977) unterscheidet hier zwischen modal harmonischen (s. (22a)) und modal nicht-harmonischen Kollokationen (s. (22b)). (vgl. Lyons, 1977, S.807f)

(22) a. *may perhaps*
b. *may certainly*

In harmonischen Kombinationen sind die Ausdrücke semantisch äquivalent und verstärken sich gegenseitig. Bei nicht-harmonischen Kombinationen befindet sich ein Ausdruck im Skopus des anderen. (22b) könnte daher paraphrasiert werden als *It is certainly the case that ... may ...* Die semantische Interpretation ist also abhängig vom Zusammenspiel der kombinierten Ausdrücke. (vgl. Halliday, 1970, S.331)

Aber auch die konkrete Wortform kann sich auf die epistemische Modifizierung auswirken. Nuyts (2001) unterscheidet bspw. zwischen Performativität und Deskriptivität. Mithilfe performativer Ausdrücke wie in (23) legt sich der Sprecher explizit selbst zum Sprechzeitpunkt fest.

(23) *I **think** they have run out of fuel.* (Nuyts, 2001, S.39)

Deskriptive Ausdrücke wie in (24) hingegen haben eine epistemische Funktion, ohne die Sprecherfestlegung zum Sprechzeitpunkt einzubeziehen. (vgl. Nuyts, 2001, S.39)

(24) *John **thinks** they have run out of fuel.* (Nuyts, 2001, S.39)
*I **thought** they have run out of fuel.* (Nuyts, 2001, S.39)

Darüber hinaus ist die Verwendung bestimmter Wortformen (z. B. unpersönliche Konstruktionen im Passiv) auch diskursstrategisch determiniert. Der Faktor der Diskursstrategie reflektiert v. a. die interpersonale Beziehung zwischen Sprecher und Hörer. Ein Grund für eine solche Strategie ist z. B. der Ausdruck von Höflichkeit gegenüber dem Gesprächspartner. (vgl. Nuyts, 2001, S.39,44f)

In Hinblick auf die mögliche Skalierung epistemisch modaler Ausdrücke ist die Feststellung, dass der sprachliche Kontext einen Einfluss auf die semantische Interpretation hat, nicht zu unterschätzen. Die Beobachtung legt nahe, dass die Annotation und Klassifizierung sich wahrscheinlich nicht auf Einzelexeme beschränken können, sondern dass die sprachliche Gesamtkonstruktion (inklusive der konkreten Wortform) sowie die Modifizierung und Kombination von Ausdrücken Beachtung finden müssen.

3 Epistemische Modalität in Wissenschaftstexten

Die Forschung zur epistemischen Modalität beschäftigt sich zu einem nicht geringen Teil mit dem Vorkommen von Hedges in Wissenschaftstexten. Hierunter fallen neben den bisher betrachteten sprachlichen Mitteln noch weitere Formulierungsmöglichkeiten, die sich nicht eindeutig lexikalisch oder grammatisch beschreiben lassen, da sie eher unter pragmatischer Perspektive zu betrachten sind. Entsprechend stehen immer auch Verwendungsgrund bzw. -motivation aufseiten des Autors im Fokus.

3.1 Hedges als Strategie im Wissenschaftsdiskurs

Wissenschaftstexte sind aus pragmatischer Sicht immer als in den wissenschaftlichen Diskurs eingebettet zu verstehen. Dieser gilt zwar als rational und neutral, verwendet im Grunde aber die gleichen Mechanismen wie die Alltagskommunikation, was durch die Verwendung eines speziellen Codes verdeckt wird. So sind wissenschaftliche Artikel nicht nur inhaltsbezogen und informativ, sondern haben auch das Ziel, die akademische Leserschaft zu beeinflussen und zu überzeugen. Hierzu gehört, dass der Autor die Zuverlässigkeit seiner Aussagen so weit wie möglich betont. (vgl. Markkanen & Schröder, 1997, S.9f) Zu den wichtigsten Verwendungsmotiven von Hedges zählt die Höflichkeit. Darunter ist zu verstehen, dass eine wissenschaftliche Aussage (z. B. eine Schlussfolgerung auf der Basis von Beobachtungen und Daten) nur als vorläufig markiert wird, da sie von der wissenschaftlichen Gemeinschaft noch nicht vollständig akzeptiert ist. „The hedging reflects [...] the appropriate attitude for offering a claim to the community“ (Myers, 1989, S.12f) und dient in diesem Sinne dazu, die Beziehung zwischen Autor und Leser vor einer Verletzung zu bewahren. Brown und Levinson (1978) sprechen hier von einer Gesichtsbedrohung, die minimiert werden kann, indem ein Sachverhalt in abgeschwächter Form präsentiert wird. (vgl. Brown & Levinson, 1978) Neben dem Motiv der Höflichkeit, welches sich eher an der Figur des Rezipienten orientiert, spielt aber ebenso das Motiv der Unsicherheit, welches sich vorrangig auf den Produzenten von Wissenschaftstexten bezieht, eine Rolle. Hedges erfüllen demnach auch ihre klassisch epistemisch modale Funktion, indem sie Mangel an Wissen aufseiten des Autors aufzeigen. Insgesamt gibt es also eine gewisse Bandbreite an Gründen – von der Vermeidung kategorischer Aussagen bis zur Interaktion mit der Leserschaft – für die Verwendung von Hedges in Wissenschaftstexten. (vgl. Behnam et al., 2012, S.20ff)

Ein umfassendes pragmatisches Framework zur Einordnung der Hedge-Motivation, welches sich an sozialen und institutionellen Kontexten des wissenschaftlichen

Diskurses orientiert und das solche verschiedenen Aspekte berücksichtigt, wird von Hyland (1996) vorgelegt. Für die vorliegende Arbeit ist es insbesondere relevant, da sich die Betrachtungen auf den Bereich der Biologie beziehen. Hyland betont, dass unterschiedliche Grade von Hedging notwendig sind, wenn Unsicherheit über den evidentialen Status von Ergebnissen bezüglich einer Hypothese besteht. Die Aufgabe des Lesers liegt dann darin, eine entsprechende Schlussfolgerung als akzeptabel einzuschätzen oder sie zurückzuweisen. Der Autor ist sich dieser aktiven Rolle der Leserschaft bei der Ratifizierung von Aussagen bewusst und zeigt deshalb durch Hedges an, dass er auch Widerspruch zum von ihm geäußerten Sachverhalt erwartet. Dabei können sich die Hedges sowohl auf den Inhalt selbst als auch auf den Leser beziehen. Leserbezogene Hedges orientieren sich an interpersonalen Faktoren und betreffen die Akzeptanz einer Proposition. Oft tritt der Autor dabei wie in (25) persönlich in Erscheinung, um das Zutreffen eines Sachverhalts als eine alternative Sichtweise und nicht als absolute Wahrheit zu präsentieren.

(25) *From our investigations we **conclude** that [...]* (Hyland, 1996, S.436)

Inhaltsbezogene Hedges reflektieren hingegen die Angemessenheit einer Proposition. Hier gibt es einerseits auf die Genauigkeit der Proposition bezogene Hedges, die zum einen die Intensität des Sachverhalts (Attribute Hedges wie in (26a)) und zum anderen die Zuversicht des Autors (Reliability Hedges wie in (26b)) anzeigen.

- (26) a. *Staining was **generally** confined to the vascular tissues.* (Hyland, 1996, S.437)
 b. *The photoreceptor involved is **somehow** related to the photosynthetic apparatus itself.* (Hyland, 1996, S.437)

Andererseits können einige Hedges als auf den Autor bezogen interpretiert werden. Der Autor antizipiert hier Kritik aufseiten der Leserschaft, indem er einen gefolgerten Sachverhalt wie in (27) als unabhängig von ihm selbst darstellt. Er übernimmt also keine Verantwortung für eine Behauptung.

(27) *The figures **suggest** that the determining factor for the protein response was whether [...]* (Hyland, 1996, S.437)

Das pragmatische Framework zeigt Abbildung 2. (vgl. Hyland, 1996, S.433-438)

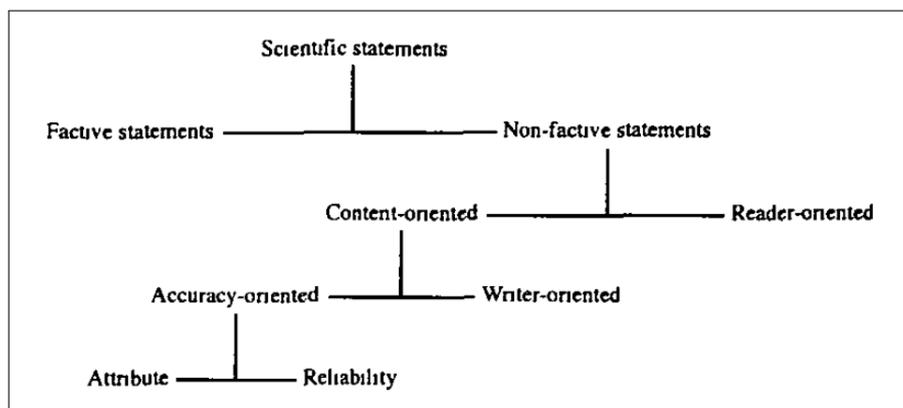


Abbildung 2: Klassifikation der Hedge-Verwendung bei Hyland (1996, S.438)

Die Ausführungen machen deutlich, dass Hedges in Wissenschaftstexten zwar insgesamt zur Einschränkung der Autorfestlegung beitragen, dies aber aufgrund vielfältiger Motive erfolgt. In ihrer interpersonalen Vermittlungsrolle weichen Hedges z. B. von einer rein epistemisch modalen Funktion ab. So steht nicht durchgängig der Wissensstatus des Autors im Zentrum (Hylands Reliability Hedges), sondern eben auch die für den Leser akzeptable Präsentation eines Sachverhalts (Hylands leserbezogene Hedges) oder der Schutz des Autors vor Kritik (Hylands autorbezogene Hedges). Hedges unterliegen also in Wissenschaftstexten einer strategischen Verwendung. Solche Strategien sollten beim Versuch, wissenschaftliche Inhalte zu klassifizieren, generell berücksichtigt werden. Darüber hinaus werfen diese Strategien (d. h. der reflektierte Einsatz relevanter sprachlicher Mittel) die Frage auf, ob der Umgang mit Hedges systematisch von der die Strategie umsetzenden Person abhängig ist. Zumindest für den produktiven Bereich, d. h. für die Verwendung von Hedges im Schreibprozess, gibt es Hinweise auf solche Personenunterschiede.

Dies betrifft zum einen die Vertrautheit im Umgang mit Wissenschaftstexten allgemein. Kommt – wie oben angegeben – ein bestimmter Code zum Einsatz, so muss dieser Code zunächst einmal erlernt werden. Hedging im Wissenschaftsdiskurs ist als eine notwendige rhetorische Fertigkeit zu verstehen, die erworben werden muss. Hierzu gehören die konkreten sprachlichen Mittel wie auch die Fähigkeit, mit ihrer Hilfe wissenschaftliche Gedanken nuanciert zu präsentieren. Verfasser von Wissenschaftstexten müssen lernen, diese Gedanken so kenntlich zu machen, dass eindeutig zwischen beobachteten Fakten und interpretativen Erklärungen unterschieden werden kann. (vgl. Salager-Meyer, 1994, S.165) Hyland (1998) spricht hier von „conventionalisation within the discourse community as a result of readers’ constant exposure to [the linguistic forms]“ (Hyland, 1998, S.148). Die Ausdrücke werden also zunehmend formelhaft verwendet.

Ein weiterer Faktor ist der kulturelle Hintergrund. So gelten für einzelne Sprachgemeinschaften spezifische Normen, was die Verwendung von Hedges anbelangt. Verschiedene Studien zeigen, dass ein Transfer solcher Normen von der Muttersprache in die Zielsprache erfolgen kann. (vgl. Markkanen & Schröder, 1997, S.13) Da das Englische als Lingua Franca des wissenschaftlichen Diskurses gilt und viele Autoren durch ihre internationale Publikationstätigkeit Texte nicht in der Muttersprache verfassen, kann ein solcher Transfer beim Umgang mit Hedges zu systematischen Unterschieden führen. Dabei spielt aber nicht nur die Art der Muttersprache eine Rolle, sondern auch die Kompetenz im englischen Sprachgebrauch, denn über die strategische Verwendung hinaus müssen die relevanten Ausdrücke sowie ihre Bedeutung ausreichend vertraut sein.

Schließlich ist zu bedenken, dass sich die Verwendung von Hedges auch am konkreten Fachgebiet orientieren kann. So weisen verschiedene Disziplinen unterschiedliche Terminologien und rhetorische Strategien auf. Während sich der Faktor

Sprache auf die Frequenz von Hedges insgesamt auszuwirken scheint (z. B. gibt es in englischen und norwegischen Wissenschaftstexten mehr epistemisch modale Ausdrücke als in französischen), sind es in Bezug auf die Disziplin meist einzelne Lexeme, die je nach Fachbereich präferiert werden. (vgl. Vold, 2006, S.61-64,83f) Außerdem variieren die damit verbundenen Funktionen. Dies wird mit den unterschiedlichen Arten von Daten in Zusammenhang gebracht. Bei abstrakten und unpräzisen Daten wird durch Hedges vorrangig die Sicherheit des Autors in den Vordergrund gestellt. Bei numerischen Daten, die akkurate Messungen darstellen und leichter verifiziert werden können, steht eher die Höflichkeitsfunktion bei der Präsentation im Zentrum. (vgl. Parladé & Diner, 2008, S.185ff) Die Biologie scheint in dieser Hinsicht eine Zwischenposition einzunehmen. (vgl. Rizomilioti, 2006, S.66)

Obwohl die hier angesprochenen Personencharakteristika (Vertrautheit im Umgang mit Wissenschaftstexten, Muttersprache, englische Sprachkompetenz, Fachbereich) v. a. für den Prozess des Schreibens untersucht worden sind, stellt sich die Frage, ob sich derartige systematische Unterschiede auch bei der Rezeption der Texte ergeben. Solche Unterschiede wären hinsichtlich der Klassifizierung wissenschaftlicher Inhalte durchaus relevant, weil sie eventuell zur Variabilität in der Interpretation von Hedges beitragen.

3.2 Hedge-Arten in biologischen Texten

In Einklang mit der Beobachtung, dass Hedges in Wissenschaftstexten strategisch verwendet werden, gibt es auch Besonderheiten bezüglich der sprachlichen Mittel, die in diesen Texten vorkommen. Eine detaillierte Übersicht, die das Ergebnis einer umfassenden Korpusanalyse ist,⁷ ist bei Hyland (1998) zu finden. Er unterscheidet zwischen lexikalischen Mitteln, die nach Wortart kategorisiert werden können, und diskursbasierten Mitteln, die eine hohe Variabilität aufweisen und bei denen daher keine explizite Einordnung möglich ist.

Aus der ersten Kategorie werden am häufigsten (27.4% aller in der Korpusstudie verzeichneten Hedge-Vorkommen) epistemische lexikalische Verben verwendet. Hyland führt einerseits Verben der Einschätzung an, die sich auf den faktischen Status eines Events beziehen und in ihrer Funktion spekulativ (s. (28a), im Sinne von subjektiver Meinung) oder schlussfolgernd (s. (28b)) sind.

- (28) a. *propose, believe, speculate, suspect*
b. *calculate, infer, conclude, deduce, imply, indicate*

Dabei können einige Lexeme je nach sprachlichem Kontext durchaus beide Funktionen erfüllen. Im Gegensatz zum spekulativen Vorkommen von *suggest* in (29a)

⁷Dabei wurden 26 Artikel aus dem Bereich Molekulare und Zellbiologie sowie die Inhalte von drei weiteren Korpora aus englischen Wissenschaftstexten mit über 2000 Textseiten analysiert. (vgl. Hyland, 1998, S.11)

distanziert sich der Autor in (29b) bspw. von der modifizierten Proposition, da er die erhobenen Daten selbst als Quelle der epistemischen Einschätzung präsentiert. Der Leser wird dadurch in die Position versetzt, eigenständig zur gleichen Schlussfolgerung zu gelangen.

- (29) a. *I **suggest** here that the photosynthetic apparatus [...]* (Hyland, 1998, S.121)
 b. *Taxonomic evidence **suggests**, therefore, that a species [...]* (Hyland, 1998, S.123)

Distanzierung kann ebenfalls durch unpersönliche Konstruktionen wie das Passiv in (30) erreicht werden, wodurch der Autor objektiver in Erscheinung tritt.

- (30) *It has been **calculated** that the concentration [...]* (Hyland, 1998, S.122)

Neben den Einschätzungsverben gibt es evidentielle Verben. Hierzu gehören Verben der indirekten Rede, die auf vorhandene Forschungsliteratur verweisen und formal größtenteils mit den spekulativen und schlussfolgernden Verben übereinstimmen. So haben *suggest* und *deduce* in (31a) und (31b) z. B. Zitatfunktion.

- (31) a. *Sandal et al. (1987) and Metz et al. (1988) have **suggested** that [...]* (Hyland, 1998, S.124)
 b. *Jofuku et al. (1989) **deduced** that [...]* (Hyland, 1998, S.124)

Weiterhin gehören zu den evidentialen Verben solche wie in (32), die auf den Akt der Wahrnehmung referieren.

- (32) *appear, seem*

Letztendlich zählt Hyland hier Verben hinzu, die das Ziel eines Forschungsvorhabens in den Vordergrund rücken und dieses in gewisser Weise mit den erhaltenen Ergebnissen kontrastieren. Verben wie in (33a) und (33b) werden als metadiskursiv betitelt.

- (33) a. *We **sought to investigate** this by [...]* (Hyland, 1998, S.125)
 b. *In these FTIR studies we **attempt to gain insight** into the [...]* (Hyland, 1998, S.126)

Auf die epistemischen lexikalischen Verben folgen als zweithäufigste Gruppe (24.7%) die epistemischen Adverbien. Hyland unterscheidet hier zwischen Downtoners, die die Stärke des durch sie modifizierten Verbs abschwächen (s. (34a)), und Disjuncts, die syntaktisch vom Rest des Satzes abgetrennt sind⁸ und entweder stilistische Funktion haben (s. (34b)) oder Grade von Sicherheit (s. (34c)) ausdrücken.

- (34) a. *quite, partially, rarely, virtually, entirely*
 b. *generally, broadly*
 c. *presumably, possibly, probably, apparently, evidently, intuitively, potentially, essentially, seemingly, approximately*

⁸D. h., sie können verschiedene Satzpositionen einnehmen, ohne dass die Bedeutungsrelation zwischen Satz und Adverb verändert wird. (vgl. Hyland, 1998, S.134)

Mit 22.1% stellen die epistemischen Adjektive die dritthäufigste Gruppe bei Hyland dar. Dazu zählen u. a. die Ausdrücke in (35).

(35) *(un)likely, possible, consistent with, somehow, proposed, potential, probable*

Im Falle von *possible* betont Hyland, dass nur wenige Vorkommen dieses Adjektivs tatsächlich epistemisch modal sind, da – wie beim Modalverb *can* – oft die deontische oder dynamische Bedeutung im Vordergrund steht. Während z. B. in (36a) tatsächlich die Einschätzung des Autors explizit gemacht wird, geht es in (36b) eher um Machbarkeit.

- (36) a. [...] *the possible involvement of phytochrome C, D or E in an interactive or synergistic manner with phytochrome A [...]* (Hyland, 1998, S.133)
b. *Further refinements of the proposed structure are possible in these regions.* (Hyland, 1998, S.133)

(36a) ist außerdem ein Beispiel für die attributive Verwendung epistemischer Adjektive als Teil komplexer Nominalphrasen. Nach Hyland referieren sie in diesem Fall auf abstrakte Propositionen, die existieren können oder nicht. Sie beziehen sich also auf die Existenz der Prozesse (in (36a) *involvement*), die in den Nominalphrasen beschrieben sind.

Die in der Linguistik besonders wichtige geschlossene Klasse der Modalverben nimmt mit 19.4% nur den vierten Platz in der Häufigkeit ein. Auch Hyland betont die Ambiguität dieser Ausdrücke. So sind bei seiner Korpusanalyse nur 39% der Modalverbvorkommen epistemisch modal, wobei am meisten *would* und *may* diese Bedeutung aufweisen. *Can* hat nach Hyland nur im negierten Kontext eine epistemisch modale Funktion und auch bei *could* liegt oft eine dynamische Bedeutung wie in (37) vor. Hier steht hauptsächlich das Vorhandensein externer Bedingungen im Vordergrund.

(37) *Long ago it was shown that the components of the photosystems could be partially separated [...]* (Hyland, 1998, S.110)

An letzter Stelle (6.3%) führt Hyland die epistemischen Nomen an, wobei er als Prototyp das Nomen *possibility* nennt. Ein Beispiel ist die Äußerung in (38).

(38) *One cannot exclude a possibility that the activity [...]* (Hyland, 1998, S.132)

Die zweite Kategorie von Hedges, die diskursbasierten Mittel, bezeichnet Hyland als nicht-lexikalisch. Zu dieser Art von Hedges zählen Angaben zu mangelndem Wissen und zur Unsicherheit im Zusammenhang mit fehlender Evidenz. Das Verb *know* taucht dabei besonders häufig auf. Beispiele hierfür sind (39a) und (39b).

- (39) a. *Nothing is known about the chemical constitution [...]* (Hyland, 1998, S.142)
b. *We do not know whether the increase in intensity [...]* (Hyland, 1998, S.142)

Weiterhin führt Hyland Beispiele an, bei denen es um die Defizite von Modellen, Theorien und Methoden geht. Der Autor kann auch auf diese Weise ausdrücken, dass er sich nicht vollständig festlegt, und deutlich machen, dass er bezüglich des Untersuchungskontextes Kritik antizipiert. Hier ist eine Vielzahl an Formulierungen – wie (40a), (40b) und (40c) – möglich.

- (40) a. *In the context of the proposed model the observation that [...]* (Hyland, 1998, S.144)
b. *In this scenario, breeding plants [...]* (Hyland, 1998, S.145)
c. *In spite of its shortcomings the method has been widely employed to evidence this type of [...]* (Hyland, 1998, S.144)

3.3 Hedge-Kombination in biologischen Texten

Das gemeinsame Auftreten einzelner Hedges ist in Wissenschaftstexten ein gängiges Phänomen. Hyland (1998) berichtet, dass bei ca. der Hälfte aller Vorkommen der entsprechende Satz einen weiteren Hedge-Ausdruck enthält. V. a. lexikalische Verben und Modalverben tauchen in Kombination auf. Beispiele für dieses und andere typische Muster sind in (41) gegeben. (vgl. Hyland, 1998, S.151)

- (41) *It has been **suggested** that the affects of light on PEPc **may** be mediated by a Calvin cycle product [...]* (Hyland, 1998, S.151)
*It now **seems possible** that the oxygen carrier function [...]* (Hyland, 1998, S.151)
*[...] this modification **could possibly** play a role in [...]* (Hyland, 1998, S.151)

Mit Bezug auf Lyons (1977) und Halliday (1970) gibt Hyland an, dass die meisten dieser Kombinationen harmonischer Art sind, dass also die Ausdrücke den gleichen epistemischen Grad (z. B. Möglichkeit, Wahrscheinlichkeit oder Sicherheit) kennzeichnen. Er sieht diese Kombination nach Werteähnlichkeit auch als einen Hinweis darauf, dass es möglich ist, epistemische Skalen zu konstruieren. Hinweise zur relativen Position der Ausdrücke auf einer solchen Skala könnte seiner Meinung nach die Häufigkeit ihrer Kookkurrenzen liefern. (vgl. Hyland, 1998, S.150-153) Dass es in der Linguistik bereits zahlreiche Ansätze zu einer Skalierung epistemisch modaler Ausdrücke gibt, soll im folgenden Abschnitt deutlich gemacht werden.

4 Die theoretische Skalierung

Theoretische Betrachtungen innerhalb der Linguistik zur Skalierung epistemisch modaler Ausdrücke unterscheiden sich primär in der Antwort auf die Frage, ob eine epistemische Skala kontinuierlicher Natur ist oder ob sie die natürliche Gruppierung der Ausdrücke in differenzierbare Kategorien reflektiert. Die folgenden beiden Abschnitte geben einen Überblick zu den verschiedenen Herangehensweisen. Dabei soll insbesondere darauf eingegangen werden, welche Bedeutungsaspekte der epistemischen Modalität die vorgestellten Skalen repräsentieren.

4.1 Die epistemische Skala als Kontinuum

Hyland (1998) charakterisiert die epistemische Skala als Kontinuum der Sprecherfestlegung bezüglich des Wahrheitswerts einer Proposition. Dieses Kontinuum reicht von „uncertain possibility to confident necessity“ (Hyland, 1998, S.2). Die Bezeichnungen *possibility* und *necessity* zeigen die Nähe zur Modallogik an. Die Verwendung von *uncertain* und *confident* hingegen verdeutlicht, dass es sich im Kern um eine Skala der Sicherheit handelt. Der Sprecher kann sich hinsichtlich der Wahrheit oder Falschheit einer Proposition unsicher bis sicher sein. Zu beachten ist, dass eine solche Definition zunächst keinen spezifischen Wahrheitswert zu implizieren scheint.

Auch bei Chafe (1986) kann das im Zentrum stehende Wissen (s. Abschnitt 1.2) als mehr oder weniger valide bzw. reliabel aufgefasst werden und bewegt sich daher auf einem Kontinuum der Reliabilität. Der Grad der Reliabilität kann zum einen direkt kenntlich gemacht werden. So drückt das epistemische Adverb *probably* (s. (42b)) z. B. einen höheren Reliabilitätsgrad als das Modalverb *might* (s. (42a)) aus.

- (42) a. *It **might** be easy.*
b. *It's **probably** easy.*

Zum anderen gibt es Ausdrücke, die eher indirekt auf den Grad der Reliabilität verweisen, da sie sich auf andere Parameter der Evidentialität beziehen. So zeigen bspw. sowohl das Modalverb *could* in (43a) als auch das epistemische Adverb *presumably* in (43b) den Prozess der Deduktion an. Letzterer Ausdruck impliziert allerdings einen höheren Reliabilitätsgrad als ersterer. (vgl. Chafe, 1986, S.262-269)

- (43) a. *No normal phonological rules **could** account for the loss of this h.* (Chafe, 1986, S.269)
b. *Adults **presumably** are capable of purely logical thought.* (Chafe, 1986, S.269)

Zur Konzeption der epistemischen Skala äußert sich Chafe folgendermaßen:

People are aware, though not necessarily consciously aware, that some things they know are surer bets for being true than others, that not all knowledge is

equally reliable. Thus, one way in which knowledge may be qualified is with an expression indicating the speaker's assessment of its degree of reliability, the likelihood of its being a fact. (Chafe, 1986, S.264)

Die Bezeichnung *likelihood* weist darauf hin, dass die epistemische Skala bei Chafe im Kern als Wahrscheinlichkeitsskala konzipiert ist. Wissen, welches nicht valide ist, ist demnach unwahrscheinlich, valides Wissen hingegen wahrscheinlich. Die Wahrscheinlichkeit bezieht sich dabei explizit darauf, dass der Sachverhalt eine Tatsache darstellt, d. h., dass die Proposition wahr ist. Im Gegensatz zur Sicherheitsskala scheint daher ein spezifischer Wahrheitswert impliziert zu sein.

Die von Nuyts (2001) eingenommene kognitiv-pragmatische Perspektive führt ihn ebenfalls zu der Annahme, dass die epistemische Modalität auf einer kontinuierlichen Skala der Wahrscheinlichkeit abgetragen werden kann. Einschätzungen zur Wahrscheinlichkeit, dass ein bestimmter Sachverhalt zutrifft, sieht er als wesentlichen Bestandteil der menschlichen Wahrnehmung und des menschlichen Handelns an. Die vorgeschlagene Skala hat dabei die folgende Form:

[T]his estimation of likelihood is situated on a scale (henceforth called the 'epistemic scale') going from certainty that the state of affairs applies, via a neutral or agnostic stance toward its occurrence, to certainty that it does not apply, with intermediary positions on the positive and the negative sides of the scale. (Nuyts, 2001, S.22)

Nuyts verbindet also die Bedeutungsaspekte der Sicherheit und der Wahrscheinlichkeit, indem er die Skala in eine negative und eine positive Seite unterteilt. Die Skalenmitte repräsentiert absolute Unsicherheit bzw. einen Mangel an Wissen. Die Proposition kann hier genauso gut wahr wie auch falsch sein. Die Pole der Skala repräsentieren jeweils absolute Sicherheit. Sicherheit bezüglich der Falschheit einer Proposition entspricht einer Position am negativen Ende der Skala. Sicherheit bezüglich der Wahrheit einer Proposition ist hingegen am positiven Ende der Skala zu verorten. Hier wird erneut deutlich, dass sich die Sicherheit auf verschiedene Wahrheitswerte beziehen kann, eine Wahrscheinlichkeitsskala bei solch einer Konzeption hingegen ausschließlich die Wahrheit einer Proposition betrifft. Die Annahme einer diskreten Skala weist Nuyts auf der Grundlage seiner Überlegungen zur Organisation des kognitiven Systems zurück. So gibt er an, dass die englische Sprache zwar über eine Grundterminologie (die Ausdrücke *impossibility*, *improbability*, *possibility*, *probability* und *certainty*) zur groben Kategorisierung verfügt, macht aber deutlich, dass durch die Hinzufügung verschiedener Modifikatoren wie in (44) im Grunde eine unendlich feine Differenzierung möglich ist.

(44) *very probable*

Eine kategoriale Perspektive kann solche sprachlichen Phänomene schlecht erklären. (vgl. Nuyts, 2001, S.21ff)

4.2 Die epistemische Skala als diskrete Skala

Die Annahme einer diskreten epistemischen Skala, d. h. ihre (natürliche) Unterteilung in Kategorien, scheint sich stark an der Analyse von Modalverben zu orientieren. Dies hängt sicherlich damit zusammen, dass Modalverben das Hauptuntersuchungsfeld der Linguistik in diesem Bereich darstellen.

Die von Palmer (2001) beschriebene Überschneidung von epistemischer und evidentialer Modalität (s. Abschnitt 1.2) spiegelt sich in der von ihm vorgeschlagenen Skala wider. Demnach gibt es drei Arten von Einschätzungen zum Wahrheitswert einer Proposition. Es ist möglich, Unsicherheit auszudrücken, Schlussfolgerungen auf der Grundlage von Allgemeinwissen oder Erfahrungen kenntlich zu machen oder anzuzeigen, dass Schlussfolgerungen aus beobachtbarer Evidenz gezogen werden. Palmer setzt diese Einschätzungsarten mit den typologischen Kategorien ‘Speculative’, ‘Assumptive’ und ‘Deductive’ gleich. Das Modalsystem des Englischen weist mit den Modalverben *may*, *will* und *must* Marker für alle drei Kategorien auf (s. (45)).

- (45) *She **may** be at work.* - „mögliche Schlussfolgerung“ (‘Speculative’)
*She **will** be at work.* - „vernünftige Schlussfolgerung“ (‘Assumptive’)
*She **must** be at work.* - „einzig mögliche Schlussfolgerung“ (‘Deductive’)

Diese grobe Dreiteilung lässt allerdings weitere Modifikationen zu. Auch innerhalb der Kategorien kann der Sprecher indizieren, dass seine Einschätzungen stärker (s. (46b)) oder schwächer (s. (46a)) sind. Im Englischen dient hierzu die Imperfektform der Modalverben. Dies deckt sich mit der bei Perkins (1983) vorgenommenen Unterscheidung zwischen primären und sekundären Modalverben (s. Abschnitt 2.1).

- (46) a. *She **might** be at work.*
b. *She **may** be at work.*

Auch auf die Rolle anderer Wortarten geht Palmer ein. Zentral ist hier die Unterscheidung zwischen Zuversicht des Sprechers und Stärke der Schlussfolgerung, also zwischen Konfidenz und Inferenz. Während Modalverben immer auch die Schlussfolgerungsstärke übermitteln, können durch epistemische Adverbien wie in (47) nur unterschiedliche Grade der Zuversicht angezeigt werden. (vgl. Palmer, 2001, S.24f,31f)

- (47) ***Perhaps** she is at work.*
*She is **probably** at work.*
*She is **certainly** at work.*

Es gibt allerdings auch Arbeiten, in denen keine derartige Unterscheidung vorgenommen wird. Bei Perkins (1983) z. B. ist das zentrale Kriterium allein die Schlussfolgerungsstärke, also die Beziehung zwischen vorhandener Evidenz und beschriebenem Sachverhalt. Beispiele hierfür gibt (48). (vgl. Perkins, 1983, S.37,59,64,77)

- (48) *be possible to/that* - „Evidenz schließt Sachverhalt nicht aus“
be likely to/that - „Evidenz ist günstig für Sachverhalt“
be certain to/that - „Evidenz bedingt Sachverhalt“

Solche Überlegungen machen deutlich, dass es neben Sicherheit und Wahrscheinlichkeit weitere Bedeutungsaspekte gibt, die der epistemischen Skala zugrunde liegen können. Während die Zuversicht des Sprechers dem der Sicherheit entspricht, bezieht sich die Stärke der Schlussfolgerung auf den evidentialen Status. Die verschiedenen Aspekte tragen alle zur Einschränkung der Sprecherfestlegung bei.

Die von Palmer (2001) und Perkins (1983) skizzierten diskreten Skalen weisen eine Dreiteilung auf, die für linguistische Betrachtungen typisch ist. Die einzelnen Kategorien sind dabei meist durch bestimmte Ausdrücke charakterisiert. Bezüglich der Frage, ob eine solche Kategorisierung direkt die natürlichen sprachlichen Gegebenheiten reflektiert, liegen allerdings unterschiedliche Sichtweisen vor.

Halliday und Matthiessen (2004), die in ihrem umfassenden Einführungswerk zur Funktionalen Grammatik⁹ auch das System der Modalität betrachten, begründen die Dreiteilung der epistemischen Skala mit Bezug auf das Verhalten der Negation. Diese kann einerseits Teil der Proposition sein wie in (49).

- (49) *That's possibly not true.*
That's probably not true.
That's certainly not true.

Das negative Merkmal kann aber auch wie in (50) auf die Modalität selbst übertragen werden.

- (50) *That's not possibly true.*
That's not probably true.
That's not certainly true.

Hierbei ist für die Autoren auffällig, dass in der mittleren Kategorie bei Verschiebung der Negation kein Bedeutungswechsel stattfindet, bei den beiden äußeren Kategorien allerdings der Wert der Modalität umgekehrt werden muss, um die Bedeutung zu erhalten (*That's possibly not true.* ist bspw. äquivalent zu *That's not certainly true.*). Die Konzeption der Skala wird also auf der Grundlage solcher linguistischer Besonderheiten erklärt. (vgl. Halliday & Matthiessen, 2004, S.147-149)

Die Ausführungen von Holmes (1982) sind hingegen ein Beispiel für Arbeiten, in denen eine Dreiteilung der epistemischen Skala nicht als natürlich gegeben, sondern eher als für die Zwecke der linguistischen Beschreibung günstig angesehen wird:

Although [discrete points on the scale] must be to some extent arbitrary, they can nevertheless provide useful points of reference for descriptive purposes.
(Holmes, 1982, S.12)

⁹Die Grammatik einer Sprache wird dabei in funktionaler Hinsicht untersucht, d. h. dahingehend, wie sie Bedeutung erzeugt und diese ausdrückt. (vgl. Halliday & Matthiessen, 2004, S.19)

P_j outranks P_i on a given scale iff a statement containing an instance of the former unilaterally entails the corresponding statement containing the latter.
(Horn, 2001, S.231)

Die Beziehungen zwischen den Prädikaten manifestieren sich in bestimmten syntaktischen Rahmen, in denen ein Sprecher explizit angibt, dass ein höherer Wert auf der relevanten Skala erzielt wird bzw. erzielt werden kann. Ein solcher syntaktischer Rahmen ist z. B. ... *if not* ... in (54).

- (54) *(it is) possible if not likely*
(it is) likely if not certain
(it is) unlikely if not impossible
(it is) uncertain if not unlikely

Die Trennung in eine positive und eine negative Skala begründet Horn damit, dass Elemente aus unterschiedlichen Ordnungen nicht in den relevanten syntaktischen Rahmen kombiniert werden können, da sich hier sprachliche (genauer: pragmatische) Anomalien ergeben. Dies ist z. B. in (55) der Fall. (vgl. Horn, 2001, S.231-236)

- (55) *#(it is) unlikely if not likely*

Eine ähnliche Skalenkonzeption findet sich bei Hoye (1997), der sechs Kategorien auf zwei getrennten Skalen annimmt (s. (56)). (vgl. Hoye, 1997, S.239-243)

- (56) ‘possibility’, ‘probability’, ‘certainty’
‘negative possibility’, ‘negative probability (unlikely)’, ‘negative certainty (impossibility)’

4.3 Fazit zur theoretischen Skalierung

Im Folgenden sollen einige Punkte festgehalten werden, die für das weitere Vorgehen relevant sind. Wichtig scheint die Beobachtung zu sein, dass eine epistemische Skala unterschiedliche Bedeutungsaspekte der epistemischen Modalität repräsentieren kann. (vgl. Westney, 1986, S.311) Beim Aspekt der Sicherheit geht es um den Status des Wissens, auf dessen Grundlage eine Aussage getroffen wird. Die Sicherheit ist damit am ehesten als Urteilssicherheit zu verstehen. Diese kann sich auf beide Wahrheitswerte einer Proposition beziehen. Geht es um die Wahrheit einer Proposition (wie bei *It's probable that it's going to rain.*), liegt positive Sicherheit vor. Bei Falschheit einer Proposition (wie bei *It's improbable that it's going to rain.*) kann man hingegen von negativer Sicherheit sprechen. Auf einer Sicherheitsskala kann folglich nur in Kombination mit dem konkreten Wahrheitswert eindeutig zwischen solchen konträren Bedeutungen unterschieden werden. Ähnlich der Sicherheitsskala ist auch eine Skala aufzufassen, die die Stärke der gezogenen Schlussfolgerung reflektiert. Dieses Phänomen ist v. a. dadurch begründet, dass die epistemische Modalität nur schwer von der Evidentialität zu trennen ist. Unter der Annahme, dass nicht

alle relevanten Ausdrücke einen evidentialen Charakter haben, alle evidentialen Ausdrücke aber zu einem gewissen Grad immer auch die Sicherheit des Sprechers implizieren, wäre eine Sicherheitsskala für den Einbezug der Fülle an Ausdrücken allerdings einer Skala der Schlussfolgerungsstärke vorzuziehen. Schließlich ist der Aspekt der Wahrscheinlichkeit zu nennen. Die Wahrscheinlichkeit bezieht sich entsprechend der obigen Überlegungen immer auf die Wahrheit einer Proposition bzw. das Zutreffen eines Sachverhalts und kann daher als Sachverhaltswahrscheinlichkeit umschrieben werden. Auf einer solchen Skala, deren Mitte Mangel an Wissen und deren Pole absolute Sicherheit (negativ und positiv) repräsentieren, lassen sich Unterschiede wie zwischen *probable* und *improbable* eindeutig abbilden.

Ein weiterer Punkt ist die Frage, ob es sich bei der epistemischen Modalität um ein Kontinuum oder eher um eine diskrete Skala handelt. Eine diskrete Skala kann zur Klassifizierung wissenschaftlicher Inhalte sehr gut übernommen werden. Die in der Linguistik gängige und relativ klare Dreiteilung ('possible', 'probable', 'certain') ist allerdings der Tatsache geschuldet, dass der Hauptuntersuchungsgegenstand die Modalverben sind und die Grundterminologie der (englischen) Sprache darüber hinaus bestimmte epistemische Kategorien nahelegt. Problematisch ist hier, dass das System der Modalverben eine geschlossene Klasse darstellt und die Ausdrucksmöglichkeiten natürlich auch weit über die Grundterminologie hinausgehen. Unter kognitiv-pragmatischer Perspektive und in Hinblick auf den Einfluss des sprachlichen Kontextes scheint die Konzeption der epistemischen Skala als Kontinuum nahe liegend. Dies würde allerdings bedeuten, dass eine Kategorisierung zum Zwecke der Klassifizierung zu einem gewissen Grad immer willkürlich erfolgen muss.

Ein letzter Punkt, der an dieser Stelle angesprochen werden soll, ist die von Holmes gemachte Beobachtung, dass epistemisch modale Ausdrücke kaum eindeutig auf einer epistemischen Skala zu positionieren sind. Sie scheinen also in gewissem Sinne unpräzise zu sein. Nuyts (2001) spricht hier von der Spezifität der Referenz zur epistemischen Skala. Er gibt an, dass diese Spezifität zu semantischen Unterschieden zwischen den verschiedenen sprachlichen Mitteln führt. So scheinen sich epistemische Adverbien auf recht spezifische Positionen auf der Skala zu beziehen, Modalverben in dieser Hinsicht allerdings sehr vage zu sein. (vgl. Nuyts, 2001, S.228) Bei der Skalierung epistemisch modaler Ausdrücke sollten solche Überlegungen berücksichtigt werden. Einschränkungen ergeben sich dabei v. a. bezüglich der möglichen Detailliertheit eines epistemischen Kategoriensystems. Zu viele Kategorien werden der Tatsache, dass sich die Ausdrücke teilweise unpräzise auf der Skala bewegen, nicht gerecht.

Im Folgenden soll nun darauf eingegangen werden, inwieweit bereits durchgeführte Annotations- und Klassifizierungsarbeiten auf solche theoretischen Überlegungen zur Skalierung epistemisch modaler Ausdrücke zurückgreifen und ob sich die hier beschriebenen Problematiken wiederfinden.

5 Vorhandene Annotations- und Klassifizierungsarbeiten

Es gibt einige anwendungsbezogene Arbeiten, in denen die Skalierung epistemisch modaler Ausdrücke eine Rolle spielt. In den nächsten Abschnitten werden die epistemischen Kategoriensysteme der entsprechenden Annotationsrichtlinien sowie die dafür berichteten Übereinstimmungswerte bei der Annotation kritisch beleuchtet.

5.1 Arbeiten außerhalb des biomedizinischen Bereichs

Das Annotationsschema von Rubin (2006, 2007, 2010) zur Annotation von Nachrichtentexten weist als zentrales Konzept die Sicherheit auf, welche verschiedene Dimensionen umfasst. Hierzu gehört die Sicherheitsebene. (vgl. Rubin et al., 2006, S.66ff) Rubin bezieht sich explizit auf die typische Dreiteilung ('possible', 'probable', 'certain') in der Linguistik und erweitert diese zur fünfstufigen Skala in Abbildung 4.

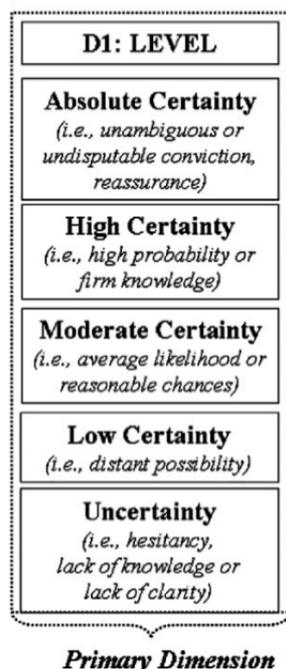


Abbildung 4: Epistemische Skala bei Rubin (2010, S.536)

Auffällig ist, dass die Sicherheit über den Aspekt der Wahrscheinlichkeit definiert wird, d. h. hohe Sicherheit geht mit hoher Wahrscheinlichkeit, geringe Sicherheit mit geringer Wahrscheinlichkeit einher. Diese Gleichsetzung scheint nach den vorhergehenden Überlegungen problematisch. Deutlich wird das am Beispiel in (57).

(57) *It's improbable that she's at work.*

Nach der vorgeschlagenen Skalierung ist eine eindeutige Klassifizierung hier schwierig. Der Autor drückt einerseits aus, dass er über ausreichendes Wissen (nach Rubin (2010) *firm knowledge*) zur Urteilsfindung verfügt, was der Kategorie ‘High Certainty’ entspricht. Andererseits impliziert der Ausdruck *improbable* eine geringe Wahrscheinlichkeit des beschriebenen Sachverhalts. Dies entspricht der Kategorie ‘Low Certainty’. Ebenso scheint die Erklärung von *average likelihood* für die Kategorie ‘Moderate Certainty’ problematisch. Wird dem beschriebenen Sachverhalt eine mittlere Wahrscheinlichkeit zugewiesen, muss man von fehlendem Wissen aufseiten des Autors ausgehen, da er sich weder bezüglich der Wahrheit noch bezüglich der Falschheit der Proposition sicher sein kann. Das wäre die Kategorie ‘Uncertainty’. Ein Beispiel hierfür ist der Satz in (58).

(58) *I don't know whether she's at home.*

Die Verbindung der Aspekte der Sicherheit und Wahrscheinlichkeit wie bei Rubin führt also zu Kategorisierungsproblemen. Eine Alternative wäre die Skala von Nuyts (2001) mit hoher Sicherheit an beiden Polen der Wahrscheinlichkeitsskala.

Zur Übereinstimmung zwischen vier Annotatoren in einem Annotationsexperiment (10 Nachrichtenartikel im Gesamtumfang von 272 Sätzen) berichtet Rubin einen durchschnittlichen Kappa-Wert¹⁰ von $\kappa = .15$. (vgl. Rubin, 2007, S.142f) Das wirft die Frage auf, ob eine fünfstufige Skala für die Klassifizierung vielleicht bereits zu detailliert ist, da sie der fehlenden Präzision der Ausdrücke bezüglich der Skala bzw. der Interpretationsvarianz zwischen bewertenden Personen nicht gerecht wird:

[I]ndependent coders' perceptions of the boundaries between shades of certainty in epistemically modalized statements are highly subjective and present difficulties for manual annotation and consequent automation [...] (Rubin, 2007, S.141)

Relativ geringe Übereinstimmungswerte werden auch bei einer vierstufigen Sicherheitsskala wie der von Henriksson und Velupillai (2010) in (59) berichtet.

(59) ‘very certain’
‘quite certain’
‘quite uncertain’
‘very uncertain’

Die paarweise Annotation von zehn Dokumenten der World Bank liefert einen durchschnittlichen F-Score¹¹ von $F_1 = .28$. (vgl. Henriksson & Velupillai, 2010, S.41ff)

¹⁰Cohen's Kappa ist ein Maß der Interrater-Reliabilität zwischen zwei Annotatoren. Es findet bei Nominalskalen Verwendung und bezieht im Gegensatz zur rein prozentualen Übereinstimmung die Tatsache mit ein, dass die Bewerter auch rein zufällig dieselbe Kategorie zuordnen können. Von einer guten Übereinstimmung spricht man ab einem Kappa-Wert von $\kappa = .60$. (vgl. Bortz & Döring, 2006, S.276f)

¹¹Der F-Score ist ein besonders für das Gebiet des Information Retrieval, d. h. der Informationsrückgewinnung, relevantes Maß. Es vereinigt die dort verwendeten Standardrelevanzmaße Pre-

Beide bisher vorgestellten Sicherheitsskalen lassen den konkreten Wahrheitswert der Proposition offen. Damit ist entsprechend der vorherigen Überlegungen keine Unterscheidung zwischen teilweise gegensätzlichen Inhalten (wie bei *It's probable that it's going to rain.* und *It's improbable that it's going to rain.*) möglich, da die Ausdrücke ein ähnliches Maß an Urteilssicherheit implizieren. Es ist aber davon auszugehen, dass eine solche fehlende Unterscheidbarkeit für die Annotation und Klassifizierung problematisch ist, da sie einen nicht unerheblichen Informationsverlust mit sich bringt. Der Sachverhalt (dass es regnen wird) ist in den beiden obigen Beispielsätzen epistemisch unterschiedlich modifiziert und sollte entsprechend eingestuft werden. Hier zeigen sich Differenzen zwischen einer theoretischen Betrachtung der epistemischen Modalität in der Linguistik, die sich primär mit dem ausgedrückten Wissensstatus des Sprechers auseinandersetzt, und der praktischen Anwendbarkeit einer solchen Skala. Es scheinen sich nun zwei Möglichkeiten der Handhabung anzubieten. Zum einen könnte der Aspekt der Sicherheit bei der Klassifizierung auf einen konkreten Wahrheitswert bezogen werden. Zum anderen wäre es möglich, stattdessen eine Wahrscheinlichkeitsskala zu verwenden.

Die erste Herangehensweise liegt bspw. USAS, dem UCREL Semantic Analysis System, zugrunde. Hierbei handelt es sich um eine Software zur automatischen semantischen Analyse englischer Texte. Das umfangreiche Tagset umfasst auch den Bereich der Modalität. (vgl. Rayson et al., 2004, S.1) Zu den dort vermerkten prototypischen Beispielen gehören die in (60) aufgeführten.¹²

- (60) ACHIEVABLE (+), ARGUABLE (-), CAN (+), CERTAINTY (+), CIRCUMSTANTIAL (-), CONCEIVABLE, CONVINCING, CONCLUSIVE (+++), CONTENTIOUS (-), DEFINITE (+++), DON'T STAND AN EARTHLY (—), BY ALL MEANS (+), DITHER ABOUT (-), GOOD CHANCE (+++), GREY AREA (-), HIT AND MISS (-), IN LIMBO (-), MADE CERTAIN (+), MAY AS WELL (+), NO UNCERTAIN TERMS (+)

Die Plus- und Minus-Zeichen indizieren den spezifischen Wahrheitswert. Die Anzahl der Zeichen ist ein Hinweis auf die Stärke der epistemischen Einschätzung.

Die zweite mögliche Herangehensweise findet sich z. B. bei Nirenburg und McShane (2008). Im Fokus des sogenannten OntoSem-Projekts (Ontological Semantics) (vgl. McShane et al., 2005) steht ebenfalls die automatische semantische Textannotation. Innerhalb der semantischen Repräsentation wird dabei auch die Modalität modelliert. Für die verschiedenen Modalitätsarten wird jeweils eine Skala von 0 bis 1 mit dezimalen Zwischenstufen vorgegeben. Der Wert 0 impliziert, dass etwas definitiv nicht der Fall ist. Die Wahrscheinlichkeit, dass die Proposition wahr ist,

recision – Anzahl der relevanten an den gefundenen Ergebnissen – und Recall – Anzahl der gefundenen an den relevanten Ergebnissen – und kann Werte zwischen 0 und 1 annehmen. (vgl. Powers, 2011, S.38) Zur Bestimmung der Interrater-Reliabilität müssen dabei die Annotationen eines Bewerbers als Goldstandard definiert werden, an denen die anderen Annotationen gemessen werden. (vgl. Henriksson & Velupillai, 2010, S.43)

¹²<http://ucrel.lancs.ac.uk/usas/>

entspricht also 0%. Der Wert 1 hingegen gibt an, dass etwas definitiv der Fall ist. Die Wahrscheinlichkeit, dass die Proposition wahr ist, entspricht also 100%. (vgl. Nirenburg & McShane, 2008, S.1)

5.2 Arbeiten im biomedizinischen Bereich

5.2.1 Frühe Arbeiten

Eine frühe Arbeit im biomedizinischen Bereich, die ein epistemisches Kategoriensystem vorschlägt, ist Friedman et al. (1994). Im Zentrum steht die Übersetzung der in Radiologieberichten enthaltenen klinischen Information in die Terme eines kontrollierten Vokabulars. Das formale Modell zur strukturierten Repräsentation dieser Information umfasst auch die Sicherheitsskala in (61), die sich im Kern an der bekannten Dreiteilung orientiert. (vgl. Friedman et al., 1994, S.161-171)

- (61) ‘no’
 - ‘low certainty’
 - ‘moderate certainty’
 - ‘high certainty’
 - ‘cannot evaluate’

Ein ähnliches Kategoriensystem (s. (62)) wird bei Light et al. (2004) vorgestellt.

- (62) ‘highly speculative’
 - ‘low speculative’
 - ‘definite’

Hier steht die automatische Klassifikation von Sätzen bezüglich darin enthaltener Spekulation im Vordergrund. Als Grundlage dient die Annotation von Spekulation auf der Satzebene. Zu den Annotatoren gehören die drei Autoren (aus der Linguistik) sowie ein Fachexperte aus dem Bereich der Biowissenschaften. Die Übereinstimmung zwischen den Annotatoren bei der Annotation von 163 Abstracts (747 Sätze) wird mit Kappa-Werten zwischen $\kappa = .03$ (für ‘highly speculative’ vs. ‘low speculative’) und $\kappa = .68$ (für allgemein ‘speculative’ vs. ‘definite’) angegeben. Die Autoren schließen daraus, dass eine Unterscheidung zwischen spekulativen und nicht-spekulativen Sätzen einigermaßen reliabel vorgenommen werden kann, nicht aber eine Unterscheidung zwischen verschiedenen Stufen der Spekulation. (vgl. Light et al., 2004, S.17-24) Eine Frage, die sich an dieser Stelle auch stellt, ist, ob die geringe Übereinstimmung zwischen den Annotatoren vielleicht teilweise dem unterschiedlichen Fachbereich der Annotatoren geschuldet ist. Hinweise darauf liefern die Überlegungen zur fachspezifischen strategischen Verwendung von Hedges.

Eine ähnliche Vorgehensweise der Satzannotation liegt bei Medlock und Briscoe (2007) vor. Das von den Autoren zum Zwecke des Trainings eines statistischen Modells annotierte Korpus besteht aus fünf biologischen Volltext-Artikeln. Die Anno-

tation erfolgte durch einen der Autoren sowie durch einen Domänenexperten. Als spekulative Instanzen von Sätzen gelten u. a. das Vorkommen von Hedges (s. (63a)), Angaben zu mangelndem Wissen (s. (63b)) oder zu Hypothesen (s. (63c)).

- (63) a. *Dl and Ser **have been proposed** to act redundantly in the sensory bristle lineage.* (Medlock & Briscoe, 2007, S.993)
- b. *How endocytosis of Dl leads to the activation of N **remains to be elucidated.*** (Medlock & Briscoe, 2007, S.994)
- c. ***To test whether** the reported sea urchin sequences represent a true RAG1-like match, we repeated the BLASTP search against all GenBank proteins.* (Medlock & Briscoe, 2007, S.994)

Die Übereinstimmung zwischen den Annotatoren ist mit einem Kappa-Wert von $\kappa = .93$ sehr hoch. Allerdings umfasst das Annotationsmodell nur die Kategorien ‘speculative’ und ‘non-speculative’ und keine Abstufungen der Spekulationsstärke. (vgl. Medlock & Briscoe, 2007, S.992-999) Gleichzeitig wird offensichtlich, dass die Annotation hier über die bisherigen Instanzen modalisierter Äußerungen hinausgeht. So geben die Autoren zwar an, dass sich die betrachteten Ausdrücke auf den Wahrheitswert von Propositionen beziehen, geben mit (63b) aber ein Beispiel, bei dem es sich nicht um einen epistemisch modalisierten Satz handelt. Das komplexe Prädikat *remains to be elucidated* bezieht sich nicht auf den Wahrheitswert der Proposition im Subjektsatz *how endocytosis of DI leads to the activation of N*. Stattdessen geht es um die genaue Art der Mechanismen, die zum bereits feststehenden Sachverhalt führen, wie durch das verwendete Adverb *how* deutlich wird. Die bisherigen Definitionen von epistemischer Modalität schließen solche Angaben aus. Die Einordnung des Satzes in die Kategorie ‘speculative’ ist deshalb schwierig, denn der tatsächliche Informationsgehalt des Satzes (die positive Regulierung) ist keineswegs spekulativ. Weiterhin beziehen die Autoren mit (63c) Äußerungen ein, in denen auf noch zu prüfende Hypothesen verwiesen wird. Solche Ausdrücke (z. B. *test whether*) wurden bisher nicht als epistemisch modal definiert. Einen Hinweis darauf, dass ein solcher Einbezug allerdings möglich ist, liefert Perkins (1983), der in seiner Übersicht zu den sprachlichen Mitteln auch die Konjunktionen *if* und *whether* anführt. Auch auf der epistemischen Skala von Nuyts (2001) würden solche Ausdrücke positioniert werden können, da die Skalenmitte den Bereich des vollkommenen Mangels an Wissen repräsentiert, was bei Hypothesen der Fall ist.

Im Zusammenhang mit der Satzannotation ist auch das GENIA-Event-Korpus (Kim et al., 2008) anzuführen. Dieses ist eine umfangreiche Ressource für das Textmining im biomedizinischen Bereich. Es umfasst einen Teil (1000 MEDLINE¹³ Abstracts mit insgesamt 36858 Events) des GENIA-Korpus (Kim et al., 2003).

¹³MEDLINE, Medical Literature Analysis and Retrieval System Online, ist eine bibliografische Datenbank der U.S. National Library of Medicine (NLM) mit über 19 Millionen Referenzen auf Journalartikel aus dem Gebiet der Lebenswissenschaften mit Schwerpunkt auf der Biomedizin. (<http://www.nlm.nih.gov/pubs/factsheets/medline.html>)

Dieses wurde im Zuge des GENIA-Projekts¹⁴ zusammengestellt und auf mehreren linguistischen Ebenen annotiert. Das GENIA-Event-Korpus stellt dabei eine Erweiterung um die Annotation von biomedizinischen Events dar. Zu einer solchen Event-Annotation gehört typischerweise die Markierung von Event-Clues (Wörter oder Phrasen, die das Eintreten eines Events indizieren), Event-Typen (Art der Information, die durch das Event ausgedrückt wird) und Event-Partizipanten (Entitäten oder andere Events, die Teil der Event-Repräsentation sind und die anhand ihrer semantischen Rollen – z. B. THEME und CAUSE – charakterisiert werden). Beispiel für die strukturierte Repräsentation (im XML-Format) ist Abbildung 5.¹⁵

```

-<event id="E1">
  <type class="Positive_regulation"/>
  <theme idref="E4"/>
  <cause idref="T1"/>
-<clue>
  Leukotriene B4
  <clueType>stimulates</clueType>
  c-fos and c-jun gene transcription and AP-1 binding activity in human monocytes.
</clue>
</event>

```

Abbildung 5: Eventannotation im GENIA-Event-Korpus

Das Annotationsschema schließt hier auch den Aspekt der Unsicherheit ein. Die Eventbeschreibung umfasst jeweils einen der Sicherheitsgrade in (64).

- (64) ‘certain’ (Default-Wert)
 ‘probable’
 ‘doubtful’

Genauere Hinweise zu diesen Kategorien liefert die Annotationserklärung nicht. (vgl. Ohta et al., 2007, S.8ff) Eine Analyse der einzelnen Annotationen lässt allerdings darauf schließen, dass die Kategorie ‘probable’ epistemisch modalisierte Äußerungen im eigentlichen Sinne markiert. Die Kategorie ‘doubtful’ scheint sich dagegen wie bei Medlock und Briscoe (2007) auf Hypothesen zu beziehen (s. (66a)).

- (65) *We have **examined** the effect of leukotriene B₄ (LTB₄), a potent lipid proinflammatory mediator, the expression of the proto-oncogenes c-jun and c-fos.*

5.2.2 BioScope-Korpus

Das von Vincze et al. (2008) erstellte BioScope-Korpus, welches mehr als 20000 Sätze aus drei verschiedenen Textsorten (Radiologieberichte, biologische Volltexte, biologische Abstracts) umfasst, weicht von der Satz- oder Eventannotation ab. Die Markierungen beziehen sich stattdessen auf Schlüsselwörter der Negation und Spekulation sowie deren Skopus. Dabei verfolgen die Autoren eine sogenannte Min-Max-Strategie. Bezüglich der Schlüsselwörter wird die minimale Einheit annotiert,

¹⁴<http://www.nactem.ac.uk/genia/>

¹⁵<http://www.nactem.ac.uk/genia/genia-corpus/event-corpus>

die Spekulation oder Negation ausdrückt. Als Skopus gilt hingegen die größtmögliche syntaktische Einheit, auf die sich ein Schlüsselwort bezieht. Eine Beispielannotation findet sich in Abbildung 6.¹⁶ (vgl. Vincze et al., 2008, S.9-17)

```

-<sentence id="S13.5">
  However, these genes are
  -<xcope id="X13.5.2">
    <cue type="negation" ref="X13.5.2">not</cue>
    constitutively active in B lymphocytes
  </xcope>
  ,
  -<xcope id="X13.5.1">
    <cue type="speculation" ref="X13.5.1">suggesting</cue>
    that other regulatory mechanisms must play a role in determining the patterns of expression
  </xcope>
  .
</sentence>

```

Abbildung 6: Negations- und Spekulationsannotation im BioScope-Korpus

Obwohl auch hier keine Abstufung der Spekulation erfolgt, ist das BioScope-Korpus für die vorliegende Arbeit interessant. Aufgrund der expliziten Annotation auf der Wortebene gibt es einen Überblick zu den in biomedizinischen Texten vorkommenden epistemisch modalen Ausdrücken. Problematisch ist lediglich, dass die eingesetzte Min-Max-Strategie nicht unbedingt der Rolle des sprachlichen Kontextes gerecht wird. Dies zeigt z. B. der Unterschied in der möglichen epistemischen Interpretation zwischen den beiden Vorkommen des Nomens *hypothesis* in (66a) und (66b).

- (66) a. *Specific staining as well as electron microscopic examinations revealed the accumulation of metal gold in the cells, **supporting our previous** <cue>hypothesis</cue> that gold ions could block NF-kappaB-DNA binding by a redox mechanism.*
- b. *The current study **investigates the** <cue>hypothesis</cue> that CD40 expressed on lung fibroblasts is an activation structure and mechanism for interaction with hemopoietic cells.*

Aus annotationstechnischer Sicht und in Hinblick darauf, dass der Gebrauch von Hedges in Wissenschaftstexten eine zu erlernende rhetorische Strategie darstellt, ist schließlich interessant, dass zu den Annotatoren des BioScope-Korpus sowohl Studenten als auch Wissenschaftler gehören.

5.2.3 Mehrdimensionalität

Bei einigen biomedizinischen Korpora sind neben der epistemischen Modalität noch weitere für die Fragestellungen relevante Modifizierungsebenen (z. B. Art der Information oder Art der Evidenz) annotiert. So umfasst das Annotationsschema von Wilbur et al. (2006) bspw. mehrere Dimensionen von Metawissen, mithilfe derer Inhalte klassifiziert werden können. Der Aspekt der Sicherheit wird dabei durch die vierstufige Skala in (67) repräsentiert. (vgl. Wilbur et al., 2006, S.7ff)

¹⁶<http://www.inf.u-szeged.hu/rgai/bioscope>

- (67) ‘0’ - „vollständige Unsicherheit bzw. Mangel an Wissen bezüglich eines Phänomens“
 ‘1’ - „geringe Sicherheit“
 ‘2’ - „hohe Wahrscheinlichkeit ohne vollständige Sicherheit“
 ‘3’ - „vollständige Sicherheit, Fakt“

Die Mehrdimensionalität umfasst auch das Konzept der Polarität und erlaubt es theoretisch, den jeweiligen Sicherheitswert auf einen konkreten Wahrheitswert zu beziehen. Dass dies allerdings nicht konsistent gehandhabt wird, zeigen Beispielsätze aus den Annotationsrichtlinien. So gibt es Fälle wie in (68), in denen hohe Sicherheitswerte auf negative Polarität bezogen werden. (vgl. Shatkay et al., 2005)

- (68) *Our blood studies of 61 children with autism do **not show evidence** of BDV infection.* - ‘negativ 2’ (Shatkay et al., 2005, S.17)

Andererseits wird *unlikely* in (69a) wie *may* in (69b) der Kategorie ‘1’ zugeordnet.

- (69) a. *Another potential factor in promotion of apoptosis, inducible NO synthase is limited in distribution to perivascular infiltrates at the peak of inflammation and is **unlikely** to contribute to widespread neuronal loss.* - ‘positiv 1’ (Shatkay et al., 2005, S.3)
 b. *Rather, the increased somal [Ca²⁺]_i and lethality of proximal neurite injuries **may** be due to the decreased path length/increased diameter for Ca²⁺ entering the transection site to reach the soma.* - ‘positiv 1’ (Shatkay et al., 2005, S.2)

Außerdem gibt es Parallelen zur Arbeit von Medlock und Briscoe (2007), wenn sich Ausdrücke nicht auf den Wahrheitswert einer Proposition beziehen (s.(70a)) bzw. wenn Angaben bezüglich zu testender Hypothesen gemacht werden (s. (70b)).

- (70) a. *Future structural and functional studies will be necessary to **understand precisely how** She2p binds ASH1 mRNA and how interactions with She3p influence the formation of a functional localization complex.* - ‘positiv 0’ (Shatkay et al., 2005, S.7)
 b. *Since She2p forms a stable dimer in solution, we sought to **establish whether or not** She2p dimerization is required for RNA binding.* - ‘positiv 0’ (Shatkay et al., 2005, S.6)

Die Übereinstimmungen zwischen den Annotatoren ist lediglich prozentual für eine Auswahl an Satzfragmenten angegeben. Die relativ hohen Werte zwischen .787 und .861 sind damit nicht besonders informativ bezüglich der Handhabbarkeit der hier vorgestellten Sicherheitsskala.

5.2.4 Meta-Knowledge-GENIA-Event-Korpus

Eine Erweiterung des GENIA-Event-Korpus ist das Meta-Knowledge-GENIA-Event-Korpus (vgl. Thompson et al., 2011). Das Annotationsschema orientiert sich hier an der Mehrdimensionalität des Metawissens bei Wilbur et al. (2006). Eine Übersicht zu den festgelegten Annotationsdimensionen liefert Abbildung 7.

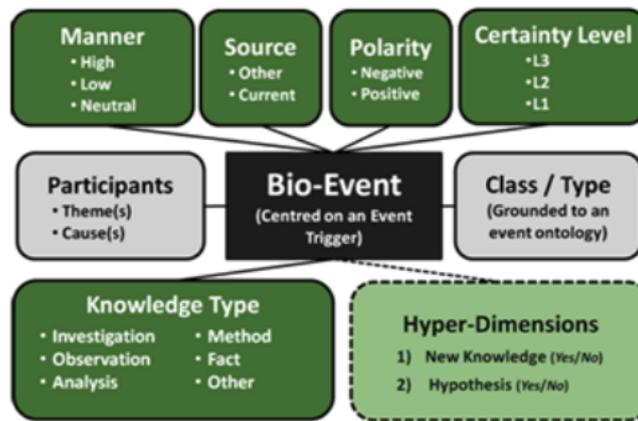


Abbildung 7: Annotationsschema bei Thompson et al. (2011, S.398)

Von zentralem Interesse ist wiederum der Aspekt der Sicherheit, wobei die Dimension drei Stufen umfasst. ‘L3’ ist Default-Kategorie für alle Events, die nicht explizit bezüglich Unsicherheit charakterisiert sind. ‘L2’ impliziert hohes, aber nicht vollständiges Vertrauen in das Eintreten eines Events. Ausdrücke, die nach Angabe der Autoren in diese Kategorie fallen, sind u. a. diejenigen in (71).

(71) *can, suggest, indicate, ability, appear, able, likely*

Bei ‘L1’ hingegen ist ein Event sehr spekulativ, was typischerweise durch die Ausdrücke in (72) angezeigt wird. (vgl. Thompson et al., 2011, S.398f)

(72) *may, might, could, possible, potential, possibility, possibly, potentially, perhaps, propose*

Die sprachlichen Mittel sind als Schlüsselwörter im Korpus explizit annotiert (s. Abbildung 8).¹⁷

```

- <event CL="L2" KT="Analysis" id="E28">
  <type class="Positive_regulation"/>
  <theme idref="E27"/>
  <cause idref="A7"/>
- <clue>
  This threshold
  <clueCL>likely</clueCL>
  <clueKT>reflects</clueKT>
  the NF-AT concentration-
  <clueType>dependent</clueType>
  assembly of transcription complexes at the promoter.
</clue>
</event>

```

Abbildung 8: Sicherheitsannotation im Meta-Knowledge-GENIA-Event-Korpus

Auch im Meta-Knowledge-GENIA-Event-Korpus erfolgt keine Kombination der Sicherheit mit der Polarität, um mit Ausdrücken wie bspw. *unlikely* umzugehen. So wird z. B. in einer Vorarbeit zum Thema *unlikely* eine geringe Sicherheit zugeschrieben. In der gleichen Kategorie befindet sich dort auch der Ausdruck *unknown*. (vgl. Thompson et al., 2008, S.34) Eine solche Zusammenfassung scheint

¹⁷<http://www.nactem.ac.uk/meta-knowledge/download.php>

allerdings unglücklich gewählt, denn es macht durchaus einen Unterschied, ob über ein Event kein Wissen vorliegt oder ob dieses als unwahrscheinlich eingestuft wird. Im Korpus selbst ist *unlikely* hingegen Marker für negative Polarität. Auch dieses Vorgehen ist schwierig, da es sich im Grunde um keine eindeutige Negation handelt.

Die Anwendbarkeit des Schemas wird bei 104 zufällig ausgewählten Abstracts aus dem Korpus überprüft. Die Übereinstimmung zwischen den Annotatoren (ein Linguist und ein Biologe; Muttersprache nicht Englisch) ist dabei relativ hoch, bezüglich der Dimension ‘Certainty’ wird ein Kappa-Wert von $\kappa = 0.864$ erzielt. (vgl. Thompson, 2011, S.401,407)

5.2.5 Kombination von Sicherheit und Polarität

Eine tatsächliche Kombination von Sicherheit und Polarität findet sich schließlich in der Arbeit von Velupillai (2012). Im Zentrum steht die Annotation diagnostischer Aussagen in Patientenakten. Die verwendete Skala orientiert sich an Saurí und Pustejovsky (2009), die mit dem FactBank-Korpus eine auf Event-Faktualität annotierte Sammlung von Nachrichtentexten vorlegen. Hier findet sich die aus der Linguistik stammende Dreiteilung in die Kategorien ‘certain’, ‘probable’ und ‘possible’ wieder. Darüber hinaus werden aber die beiden Polaritäten ‘positive’ und ‘negative’ angenommen, sodass das Modell in Abbildung 9 mit sechs Annotationsklassen entsteht, welches als Kontinuum zu verstehen ist. (vgl. Velupillai, 2012, S.1,34ff)

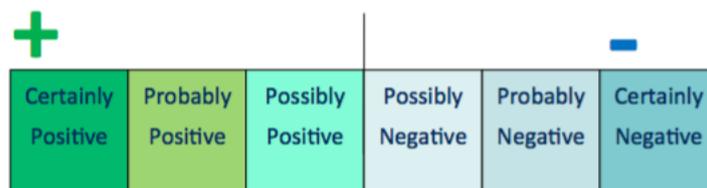


Abbildung 9: Epistemische Skala bei Velupillai (2012, S.34)

Eine solche Skala ist mit der von Nuyts (2001) vorgeschlagenen Wahrscheinlichkeitsskala vergleichbar (s. Abschnitt 4.1). Die Sicherheit bezieht sich immer explizit auf einen Polaritätswert und füllt damit beide Seiten der Skala aus. Problematisch scheint allerdings die Einordnung von Ausdrücken, die einen Mangel an Wissen indizieren. So ist nicht klar, welcher Kategorie eine Äußerung mit *It is unknown whether* zuzuordnen wäre, da die Polarität hier undefiniert bleibt und die Skala keine mittlere Kategorie anbietet.

Die Annotation wurde ausschließlich durch Domänenexperten, d. h. von Ärzten, vorgenommen. Der Übereinstimmungswert der differenzierten Skala mit sechs Kategorien liegt bei einem Kappa von $\kappa = .60$. Werden nur die Sicherheitsebenen unter Ausschluss der damit verbundenen Polarität betrachtet, liegt die Übereinstimmung sogar bei einem Kappa von $\kappa = .82$. (vgl. Velupillai, 2012, S.50)

5.3 Fazit zu vorhandenen Arbeiten

Bei den hier vorgestellten anwendungsbezogenen Skalen zur Klassifizierung von Äußerungen¹⁸ handelt es sich vorrangig um von den Autoren der entsprechenden Arbeiten festgelegte Einteilungen als Teil instruierender Annotationsschemata. Die Skalen unterscheiden sich dabei in ihrer Gesamtkonzeption (Sicherheit vs. Sicherheit kombiniert mit Polarität vs. Wahrscheinlichkeit; diskrete Skala vs. Kontinuum). Problematisch scheinen in diesem Zusammenhang Skalen zu sein, bei denen die Aspekte der Sicherheit und Wahrscheinlichkeit gleichgesetzt werden. Reine Sicherheitsskalen wiederum führen zu Informationsverlust bei bestimmten Unterscheidungen (z. B. bei Äußerungen mit *likely* und *unlikely*), während bei der Kombination von Sicherheit und Polarität Angaben zu fehlendem Wissen (z. B. *unknown whether*) nicht gut eingeordnet werden können. Die eindeutigste Kategorisierung scheint nach diesen Überlegungen bei einer Wahrscheinlichkeitsskala möglich zu sein. Diese findet sich z. B. in Arbeiten außerhalb des biomedizinischen Bereichs.

Unterschiede bestehen außerdem auch in der Anzahl der vorgeschlagenen Kategorien. Mitunter wird dabei explizit auf die in der Linguistik vorgeschlagene Dreiteilung der epistemischen Skala verwiesen. Bezüglich der Übereinstimmung zwischen mehreren Annotatoren werden sehr heterogene Werte berichtet, wobei die Übereinstimmung mit zunehmender Detailliertheit des Kategoriensystems abnimmt. Hier kann auf die Annahme verwiesen werden, dass einige sprachliche Mittel bezüglich der Skala relativ unpräzise sind. Dies spricht gegen ein detailliertes Kategoriensystem. Darüber hinaus stellt sich die Frage, ob die fehlende Übereinstimmung teilweise auch systematischen Unterschieden zwischen den Annotatoren (z. B. hinsichtlich des Fachbereichs, der Vertrautheit mit Wissenschaftstexten, der Muttersprache und englischen Sprachkompetenz) geschuldet ist. Bei den Arbeiten sind derartige Unterschiede auffällig. Solche Überlegungen beziehen sich auf die strategische Verwendung von Hedges in Wissenschaftstexten. Faktoren, die sich auf den Einsatz von Hedges beim Schreiben beziehen, könnten demnach auch Einfluss auf die Interpretation haben.

Schließlich wird deutlich, dass in den konkreten Projekten Ausdrücke markiert sind, die den bisher betrachteten Rahmen der epistemischen Modalität ausweiten. Hierzu gehören einerseits Angaben zu Hypothesen und Untersuchungsvorhaben. Andererseits sind Ausdrücke markiert, die in bestimmten Kontexten keine epistemisch modale Funktion erfüllen, da sie sich nicht auf den Wahrheitswert einer Proposition, sondern auf bestimmte Konzepte und Mechanismen beziehen.

¹⁸Für weitere Beispiele zur Satzannotation siehe Desclés et al. (2011). Zur Schlüsselwortannotation siehe Bongelli et al. (2012). Einen umfassenden Überblick zu Studien in diesem Bereich geben Morante und Sporleder (2012).

Teil II

Empirischer Teil

6 Fragestellung

Nach den theoretischen Betrachtungen im ersten Teil der vorliegenden Arbeit soll nun im zweiten empirischen Teil eine eigene Studie vorgestellt werden. Diese Studie dient als Grundlage für die Konstruktion einer epistemischen Skala, die für die Klassifizierung wissenschaftlicher Inhalte in biomedizinischen Texten geeignet ist. Beim zentralen Bedeutungsaspekt, den die Skala repräsentiert, soll die Konstruktion theoretisch geleitet sein. Erkenntnisse darüber, wie die einzelnen Ausdrücke auf der Skala bewertet werden und wie detailliert ein epistemisches Kategoriensystem sein kann, liefern hingegen die Ergebnisse einer in der Studie durchgeführten Probandenbefragung. Weiterhin sollen auf dieser Grundlage Aussagen dazu getroffen werden, ob es bei der Klassifizierung systematische Unterschiede zwischen den bewertenden Personen gibt. Schließlich werden einzelne Beobachtungen bezüglich der bereits vorhandenen Projekte einbezogen.

6.1 Bedeutungsaspekt der epistemischen Skala

Nicht jede Skalenkonzeption eignet sich für die Klassifizierung und Annotation. So ist die v. a. in der Linguistik gängige und in anwendungsbezogenen Arbeiten meist übernommene Sicherheitsskala problematisch, da eine Bewertung hier, insofern sie nicht explizit auf einen konkreten Wahrheitswert der Proposition bezogen ist, zu einem Informationsverlust führt (s. Abschnitt 5.3). Entsprechend soll die hier konstruierte Skala nicht die Urteilsicherheit des Autors, sondern die Sachverhaltswahrscheinlichkeit repräsentieren. Dabei geht es um die Wahrscheinlichkeit, dass die Proposition wahr ist, also einen Sachverhalt darstellt. Die Skalenkonzeption folgt den Ausführungen von Nuyts (2001) (s. Abschnitt 4.1). Im unteren Bereich der Skala sind theoretisch Ausdrücke wie in (73) positioniert, die eine geringe Wahrscheinlichkeit bezüglich der Wahrheit der Proposition implizieren.¹⁹

(73) *It is **improbable** that x interacts with y .*

Im oberen Bereich liegen theoretisch Ausdrücke wie in (74), die eine hohe Wahrscheinlichkeit bezüglich der Wahrheit der Proposition anzeigen.²⁰

(74) *It is **probable** that x interacts with y .*

Ausdrücke wie in (75), mit denen der Sprecher absolute Unsicherheit und Mangel an Wissen explizit macht, nehmen wiederum eine Position im Bereich der Skalenmitte ein. Hier gibt es keinerlei Kenntnisse zum beschriebenen Sachverhalt.

(75) *It is **unknown whether** x interacts with y .*

¹⁹Dies ist gleichzusetzen mit hoher Sicherheit, dass die Proposition falsch ist.

²⁰Dies ist gleichzusetzen mit hoher Sicherheit, dass die Proposition wahr ist.

In diesem Bereich könnten auch die über die epistemische Modalität hinausgehenden Angaben zu Hypothesen und Untersuchungen (s. (76)) positioniert werden.

(76) *We tested whether x interacts with y.*

Bei der Konzeption der Skala soll zunächst auf eine (grobe) Kategorisierung verzichtet werden. Die in der Linguistik gängige Dreiteilung ist zu stark an der Einzelanalyse der geschlossenen Klasse der Modalverben orientiert und die tatsächliche Fülle an sprachlichen Mitteln legt eher ein Kontinuum der epistemischen Einschätzungen nahe. Die Wahrscheinlichkeitsskala ist deshalb zunächst auch als Kontinuum zu verstehen. Ausgehend von einer solchen Analogskala soll herausgefunden werden, ob sich bei der empirischen Untersuchung eindeutige Gruppierungen von Ausdrücken hinsichtlich ihrer Position ergeben oder ob eine Einteilung in epistemische Kategorien eher willkürlich erfolgen muss. In Anlehnung an Hyland (1998) (s. Abschnitt 3.3) wird die Skalenposition dabei im weiteren Verlauf als *epistemischer Grad* (im Sinne der implizierten Wahrscheinlichkeit) eines Ausdrucks bezeichnet.

6.2 Empirische Fundierung

Hinweise zum epistemischen Grad einzelner Ausdrücke sollen die Ergebnisse einer Befragung liefern. Die Idee hinter dieser Herangehensweise ist, dass die Interpretation der Ausdrücke nicht eindeutig ist bzw. sein kann und eine Festlegung auf Grundlage von Einzelurteilen deshalb immer Gefahr läuft, zu subjektiv auszufallen:

[W]hen it comes to such a vague concept as uncertainty, there is always a thin line between having guidelines capturing the general perception of uncertainty in the language and capturing a definition of uncertainty that is specific to the writers of the guidelines. (Dalianis & Skeppstedt, 2010, S.9)

Die Zusammenfassung mehrerer Einschätzungen kann hier ein repräsentativeres Bild zu den sprachlichen Gegebenheiten liefern. Das Vorgehen ist als eine Art Konsensfindung zu verstehen. Es stellt eine Alternative zur Validierung subjektiv konstruierter Kategoriensysteme durch Berechnung der Bewerterübereinstimmung dar. (vgl. Dalianis & Skeppstedt, 2010, S.9) Studien dieser Art, bei denen von den individuellen Differenzen in der Interpretation abstrahiert wird, finden sich v. a. in der psychologischen Forschung. So berichten Clark (1990) sowie Teigen und Brun (2003) von einer Vielzahl an Arbeiten zur numerischen Übersetzung sogenannter Verbal Probabilities, also sprachlicher Mittel, die Wahrscheinlichkeiten anzeigen. Diese Verbal Probabilities sind im Grunde nichts anderes als epistemisch modale Ausdrücke. Die psychologischen Arbeiten haben allerdings eine relativ enge Perspektive, was die betrachteten sprachlichen Mittel anbelangt. Sie beschränken sich überwiegend auf epistemische Adjektive wie in (77), da diese relativ gut isoliert präsentiert werden können, ohne dass ein konkreter Sachverhalt angeführt werden muss.

(77) *doubtful, possible, certain*

Kern solcher Studien ist die Zuweisung von Wahrscheinlichkeitswerten (auf einer Skala von 0 bis 1 oder von 0 bis 100) zu den Ausdrücken. Metaanalysen zeigen dabei eine hohe Konsistenz der Ergebnisse, d. h., im Mittel stimmen die Einschätzungen zwischen den Studien überein. Gleichzeitig wird eine hohe Variabilität innerhalb der Probandengruppen deutlich. Die einzelnen Einschätzungen bezüglich eines Ausdrucks unterscheiden sich also mitunter sehr stark. (vgl. Teigen & Brun, 2003, S.125f) Eine Befragung kann daher einerseits reliable und andererseits im Sinne der Zusammenfassung von Einzelurteilen repräsentative Ergebnisse zum epistemischen Grad der Ausdrücke liefern. Beispiel für eine Skalenkonstruktion auf Grundlage einer umfangreichen Bewertungserhebung ist Rohrman (1978). Das Ziel seiner Arbeit ist die Erstellung von Antwortskalen für psychologisch relevante, aber subjektive Beurteilungsdimensionen (Rating Scales für z. B. Häufigkeit, Intensität, Wahrscheinlichkeit). Bei diesen soll gewährleistet sein, dass zwischen den durch spezifische Ausdrücke (bspw. *schwerlich, eventuell, zweifellos*) charakterisierten Skalenstufen quantitativ gleiche Abstände vorliegen, damit subsequente Berechnungen auf Intervallskalenniveau gerechtfertigt sind. Da dies nach Rohrman bei einer willkürlichen Verbalisierung der Stufen nicht der Fall ist, ist eine empirische Fundierung notwendig. (vgl. Rohrman, 1978, S.222-225) Analog dazu soll die Skalierung in der vorliegenden Arbeit ebenfalls empirisch fundiert sein.

6.3 Ungenauigkeit der Ausdrücke

Die in den anwendungsbezogenen Arbeiten (s. Kapitel 5) berichteten Übereinstimmungswerte sind Hinweis dafür, dass allzu detaillierte epistemische Kategoriensysteme kaum eine einheitliche Klassifizierung ermöglichen. Ausgehend von den theoretischen Betrachtungen ist anzunehmen, dass einer der Hauptgründe hierfür die Ungenauigkeit bezüglich der Position auf der epistemischen Skala ist (s. Abschnitt 4.3). So scheinen sich die Ausdrücke jeweils in bestimmten Bereichen auf der Skala zu bewegen, was eine eindeutige Kategorisierung schwer möglich macht. Dies betrifft nach Nuyts (2001) einige Wortarten (z. B. Modalverben) mehr als andere (z. B. epistemische Adverbien). Darüber hinaus lassen sich auf der Skala anscheinend auch keine Grenzen festlegen, nach denen ein Ausdruck klar zutrifft oder nicht.²¹ (vgl. Holmes, 1982, S.12) Außerdem sind viele Ausdrücke ambig. So drückt das Modalverb *can* vorrangig dynamische und nur zweitrangig epistemische Modalität aus. Um festzustellen, wie detailliert ein epistemisches Kategoriensystem also sein kann, ist es sinnvoll, sich ein Bild von der Ungenauigkeit epistemisch modaler Ausdrücke hinsichtlich ihres epistemischen Grades zu machen. Werden den

²¹In diesem Zusammenhang wird bspw. die Konstruktion von Zugehörigkeitsfunktionen über die Wahrscheinlichkeitsskala vorgeschlagen, wobei sich dieses Vorgehen an der Fuzzy-Logik orientiert. (vgl. Wallsten et al., 1986)

Ausdrücken bspw. große Bereiche auf der epistemischen Skala zugewiesen, spricht dies gegen eine hohe Zahl an epistemischen Kategorien. Solche Einschätzungen zur Ungenauigkeit sollen ebenfalls empirisch fundiert sein.

6.4 Systematische Bewerterunterschiede

Die Feststellung, dass epistemisch modale Ausdrücke (bzw. Hedges) in Wissenschaftstexten vorrangig strategisch verwendet werden (s. Kapitel 3), wirft weiterhin die Frage auf, ob die oben angesprochene fehlende Übereinstimmung teilweise von systematischen Personenunterschieden bei der Kategorisierung herrührt. So ist den anwendungsbezogenen Arbeiten zu entnehmen, dass die beteiligten Annotatoren mitunter unterschiedlichen Fachgebieten (Linguistik und Biologie) angehören, verschieden umfangreiche Erfahrungen mit Wissenschaftstexten gesammelt haben und nicht durchgängig Muttersprachler des Englischen sind. Solche Faktoren stellen potenzielle Quellen der systematischen Variabilität in der Bewertung dar. Grundlage hierfür ist die Beobachtung, dass es beim Einsatz von Hedges im Schreibprozess Unterschiede hinsichtlich dieser Faktoren gibt. Entsprechend könnten sie auch auf die Verwendung der Wahrscheinlichkeitsskala Einfluss haben. Die Betonung muss dabei auf der Systematik dieser Unterschiede liegen, denn natürlich ist davon auszugehen, dass sich Personen in ihren Einschätzungen rein subjektiv unterscheiden. Ein weiteres Ziel der Arbeit ist es deshalb, Aussagen darüber zu treffen, ob sich systematische Interpretationsunterschiede nachweisen lassen. Ein tatsächlicher Einfluss hätte Konsequenzen dafür, wie die Auswahl der Annotatoren erfolgen kann, wenn solche Bewertungen vorgenommen werden sollen.

6.5 Rolle des sprachlichen Kontextes

Die epistemisch modalen Ausdrücke können nur bedingt isoliert interpretiert werden. Stattdessen ist oft entscheidend, in welchen sprachlichen Kontext sie eingebettet sind (s. Abschnitt 2.2). In der vorliegenden Arbeit sollen stichprobenartig Aussagen dazu gemacht werden, welchen Einfluss die sprachliche Gesamtkonstruktion, die Modifizierung sowie die Kombination mit anderen Ausdrücken auf den epistemischen Grad eines Ausdrucks haben. Solche Betrachtungen können einen Hinweis darauf liefern, wie bei der Klassifizierung und Annotation vorgegangen werden muss. In diesem Zusammenhang ist auch auf das Vorkommen von Ausdrücken einzugehen, die in bestimmten Kontexten keine epistemisch modale Funktion erfüllen, da sie sich nicht auf die Wahrheit der Proposition beziehen. Das Antwortverhalten bei derartigen Äußerungen kann Auskunft dazu geben, inwieweit ein Einbezug solcher Phänomene überhaupt möglich ist.

6.6 Fragen und vorläufige Hypothesen

Die oben angegebenen Fragen sollen sowohl mittels beschreibender als auch schließender statistischer Verfahren beantwortet werden. Zur Entscheidung darüber, welche Werte (d. h. die Bewertungen welcher Personen) als Grundlage für die Konstruktion einer epistemischen Skala im biomedizinischen Bereich verwendet werden können, soll zuerst folgende Hypothese getestet werden, die sich darauf bezieht, inwieweit sich Personen bezüglich bestimmter Charakteristika in ihrer Einschätzung unterscheiden:

- Hypothese 1: Der epistemische Grad eines Ausdrucks ist abhängig vom Fachbereich, von der Vertrautheit mit Wissenschaftstexten und von der englischen Sprachkompetenz der Person, die hierzu eine Einschätzung abgibt.

Die Positionierung und mögliche Gruppierung der Ausdrücke auf der Skala soll dann zunächst deskriptiv erfolgen. Zur Festlegung einer sinnvollen Anzahl an epistemischen Kategorien wird daraufhin die durch Probanden geschätzte Ungenauigkeit der Ausdrücke einbezogen. Dabei ist folgende Hypothese zu überprüfen, die sich darauf bezieht, ob die verschiedenen sprachlichen Mittel vergleichbar präzise sind:

- Hypothese 2: Die Ausdrücke unterscheiden sich in ihrer Ungenauigkeit bezüglich der epistemischen Skala.

Ausgehend davon soll getestet werden, ob die durch die Gruppierung festgelegten Kategoriengrenzen empirisch nachweisbar sind und ob die Zuweisung zu den Kategorien entsprechend eindeutig ist:

- Hypothese 3a: Ausdrücke, die an einer deskriptiv festgelegten Kategoriengrenze liegen, unterscheiden sich in ihrem epistemischen Grad.

In Hinblick auf den Unterschied im epistemischen Grad werden weitere Hypothesen formuliert, die stichprobenartig den Einfluss des sprachlichen Kontextes auf die Bewertung betreffen:

- Hypothese 3b: Der epistemische Grad eines Ausdrucks unterscheidet sich je nach sprachlicher Gesamtkonstruktion, in der dieser Ausdruck vorkommt.
- Hypothese 3c: Der epistemische Grad einer als harmonisch geltenden Kombination von Ausdrücken unterscheidet sich nicht vom epistemischen Grad der einzelnen Ausdrücke. Auch die einzelnen Ausdrücke unterscheiden sich nicht voneinander.
- Hypothese 3d: Der epistemische Grad eines Ausdrucks ändert sich, wenn der Ausdruck selbst modifiziert wird.

Die Frage nach dem Umgang mit Ausdrücken, die sich nicht auf die Wahrheit einer Proposition beziehen, soll schließlich erneut deskriptiv auf der Grundlage des Antwortverhaltens der Probanden beantwortet werden.

7 Methodik

Die empirische Grundlage für die Skalenkonstruktion sind die Ergebnisse einer Fragebogenstudie. Im Folgenden soll die Methodik der Untersuchung, d. h. die Fragebogenerstellung sowie die Durchführung der Onlinebefragung, beschrieben werden.

7.1 Auswahl der epistemisch modalen Ausdrücke

Zunächst musste eine für die Befragung angemessene und möglichst charakteristische Menge an sprachlichen Mitteln, die in biomedizinischen Texten eine epistemisch modale Funktion übernehmen können, ausgewählt werden. Hyland (1998) gibt an, dass einige wenige Ausdrücke (die ca. 100 Wörter mit der höchsten Frequenz) fast die Hälfte aller Hedge-Vorkommen in Wissenschaftstexten ausmachen und die restlichen Wörter meist nur einmalig auftauchen. (vgl. Hyland, 1998, S.149) Als Auswahlkriterium wurde daher die Vorkommenshäufigkeit festgelegt. Als Grundlage dienten das BioScope-Korpus (s. Abschnitt 5.2.2) und das Meta-Knowledge-GENIA-Event-Korpus (s. Abschnitt 5.2.4). Diese Korpora sind einerseits umfangreich genug für ein repräsentatives Bild zu den in der Biomedizin verwendeten Ausdrücken. Andererseits sind die relevanten Wörter (bei BioScope als *keywords*, bei Meta-Knowledge-GENIA-Event als *clues* bezeichnet) im Gegensatz zur Satzannotation direkt markiert und können gut extrahiert werden. Zur Vergleichbarkeit der beiden Textgrundlagen wurden von den drei Textsorten des BioScope-Korpus nur die Abstracts genutzt, da auch das Meta-Knowledge-GENIA-Event-Korpus ausschließlich Abstracts umfasst.²² Hierbei ist zu beachten, dass es zwischen den beiden Korpora Überschneidungen gibt, denn das BioScope-Korpus umfasst auch Teile des allgemeinen GENIA-Korpus. Da die Annotationen der Textgrundlagen unter Verwendung verschiedener Schemata allerdings unabhängig sind, wurden diese in der vorliegenden Arbeit auch als eigenständig erachtet. Aus den im XML-Format vorliegenden Korpora wurden mithilfe einer selbst geschriebenen Java-Klasse alle durch die entsprechenden Tags (bei BioScope `<cue type="speculation"/>`, bei Meta-Knowledge-GENIA-Event `<clueCL/>`) markierten Ausdrücke zusammen mit den Sätzen, in denen sie vorkamen (bei BioScope 2649, bei Meta-Knowledge-GENIA-Event 1897 Sätze), extrahiert. Dabei wurde jeweils die absolute Häufigkeit der Ausdrücke notiert. Der zweite Schritt bestand darin, die durch ihr textuelles Vorkommen determinierten Wortformen (z. B. *suggest*, *suggesting*) zu entsprechenden Lexemen

²²Die so getroffene Auswahl an Ausdrücken ist daher relativ spezifisch für biomedizinische Abstracts. Tatsächlich scheint das Vorkommen von Hedges von der konkreten Textsorte bzw. von der jeweiligen Passage in Wissenschaftsartikeln abhängig zu sein. (vgl. Hyland, 1998, S.153ff) Da sich die meisten Anwendungen der Informationsextraktion im biomedizinischen Bereich allerdings – bedingt durch die meist öffentliche Zugänglichkeit – auf Abstracts beziehen, ist diese Entscheidung zu einem gewissen Grad gerechtfertigt. (vgl. Vincze et al., 2008, S.14)

(z. B. *suggest*) zusammenzufassen und unterschiedliche Annotationsstrategien (z. B. *indicate, indicate that*) auszugleichen. Aufgrund des unterschiedlichen Umfangs der Korpora wurde dann aus der absoluten die relative Häufigkeit berechnet. Dann wurden die 19 Wörter mit der höchsten relativen Häufigkeit (*h*) über beide Korpora hinweg bestimmt.²³ Diese sind in Tabelle 1 zu finden.

Ausdruck	<i>h</i>	Ausdruck	<i>h</i>
<i>suggest</i>	.231	<i>putative</i>	.015
<i>may</i>	.185	<i>propose</i>	.014
<i>can</i>	.131	<i>think</i>	.013
<i>indicate</i>	.113	<i>seem</i>	.012
<i>appear</i>	.058	<i>unknown</i>	.010
<i>might</i>	.026	<i>possibly</i>	.009
<i>could</i>	.024	<i>imply</i>	.008
<i>likely</i>	.023	<i>potentially</i>	.008
<i>possible</i>	.020	<i>hypothesis</i>	.008
<i>potential</i>	.017		

Tabelle 1: Ausgewählte Ausdrücke

Ausgeschlossen blieben nicht epistemisch modale Ausdrücke wie die Konjunktionen *or* und *either*, die auf Alternativen verweisen, die Konjunktion *whether*, die eher als sprachlicher Kontext einzubeziehen ist (z. B. bei *unknown whether*), sowie die dynamisch modalen Ausdrücke *ability, able, capable* und *capacity*. Der Ausdruck *implicate* wurde ausgelassen, da er stets in der Form *has/have been implicated* vorkommt, wobei die Konstruktion Teil der Proposition ist und diese nicht modifiziert (s. (78)).

(78) *Redox regulation of NF-kappa B **has been implicated** in the activation of the human immuno-deficiency virus (HIV).*

Ausdrücke, die eine geringe Wahrscheinlichkeit anzeigen, sind in beiden Korpora selten. Ein Grund könnte sein, dass sie mitunter als Negation annotiert sind. Da die theoretische Skalenkonzeption aber v. a. auch die Bedeutung solcher Wörter eindeutig erfassen soll, wurden sie trotzdem in die Betrachtung einbezogen. Um die besondere Rolle zu betonen, die den drei dabei ausgewählten Ausdrücken (s. Tabelle 2) zukommt, werden sie im weiteren Verlauf der Arbeit als *kritisch* bezeichnet.

Ausdruck	<i>h</i>
<i>no evidence</i>	.0008
<i>unlikely</i>	.0008
<i>cannot</i>	.0004

Tabelle 2: Ausgewählte kritische Ausdrücke

²³Wurde ein Ausdruck aufgrund der relativen Häufigkeit in einem der Korpora bereits ausgewählt, wurde er im anderen Korpus entsprechend ignoriert.

Die Bestimmung verschiedener sprachlicher Kontexte erfolgte auf Basis von Auffälligkeiten in den Korpora.²⁴ So tritt das epistemische lexikalische Verb *suggest* häufig in der Passivkonstruktion auf. Das epistemische Nomen *hypothesis* wiederum kann Argument verschiedener Verben sein. Entsprechend wurden bezüglich der sprachlichen Gesamtkonstruktion die Kontraste in Tabelle 3 festgelegt.

<i>suggest</i>	<i>has/have been suggested</i>
<i>investigate + hypothesis</i>	<i>support + hypothesis</i>

Tabelle 3: Ausgewählte Vergleiche der sprachlichen Gesamtkonstruktion

Bei der Kombination von Ausdrücken ist v. a. das gemeinsame Auftreten von epistemischen lexikalischen Verben und Modalverben auffällig. Besonders häufig (in ca. 20% der Fälle des Vorkommens von *suggest*) ist hierbei die Kombination von *suggest* und *may*, die nach Hyland (1998) als harmonisch zu werten ist (s. Abschnitt 3.3). Entsprechend wurden die Kontraste in Tabelle 4 festgelegt.

<i>suggest</i>	<i>suggest + may</i>
<i>may</i>	<i>suggest + may</i>
<i>suggest</i>	<i>may</i>

Tabelle 4: Ausgewählte Vergleiche bei Kombination

Bezüglich der Modifizierung von Ausdrücken entspricht der in Tabelle 5 aufgeführte Kontrast schließlich der Kombination mit dem intensivierenden Adverb *strongly*.

<i>suggest</i>	<i>strongly suggest</i>
----------------	-------------------------

Tabelle 5: Ausgewählter Vergleich bei Modifizierung

Um Ausdrücke einzubeziehen, mit denen Angaben zu Hypothesen gemacht werden, wurde die Verbkonstruktion *examine whether* ausgewählt. Die Anzahl der für die Untersuchung vorgesehenen Ausdrücke wurde damit auf insgesamt 27 erweitert.²⁵

7.2 Konstruktion der Befragungsisems

Anders als bei den in psychologischen Arbeiten betrachteten epistemischen Adjektiven (s. Abschnitt 6.2) können die meisten Ausdrücke (z. B. *may*, *has/have been suggested*) nicht isoliert und ohne Bezug zum jeweiligen Sachverhalt zur Bewertung vorgelegt werden. Daher wurden die Items für die Befragung so konstruiert, dass sie immer einen Ausdruck im Zusammenhang mit einer modifizierten

²⁴Die Untersuchung ist diesbezüglich auch nicht als umfassend zu verstehen, sondern soll lediglich erste Hinweise zum Einfluss des sprachlichen Kontextes auf den epistemischen Grad der Ausdrücke liefern.

²⁵Um lange Beschreibungen zu vermeiden, fallen im weiteren Verlauf auch die oben angeführten sprachlichen Kontexte und Konstruktionen unter die Bezeichnung *Ausdruck*.

Proposition enthalten. Hierfür wurde auf die Originalsätze in den beiden Korpora zurückgegriffen. Beachtet werden muss allerdings, dass bei Präsentation solcher Items nicht auszuschließen ist, dass die Wahrscheinlichkeit des beschriebenen Sachverhalts selbst (z. B. aufgrund bekannter wissenschaftlicher Vorarbeiten oder aufgrund von Eigenschaften der Entitäten) einen Einfluss auf die Interpretation der Ausdrücke hat. Auf eine entsprechende Abhängigkeit der Bewertungen vom außersprachlichen Kontext weisen Brun und Teigen (1988) hin.²⁶ Um den möglichen Einfluss dieser Grundwahrscheinlichkeit zumindest auszugleichen, wurden für jeden Ausdruck drei verschiedene Originalsätze ausgewählt. Die Auswahl erfolgte primär auf Grundlage der Satzlänge, da kurze Sätze die Bearbeitung erleichtern sollten. Eine zweite Anforderung war, dass die Sätze wenig anaphorische Bezüge (wie z. B. durch das Demonstrativpronomen *these* in *These results suggest*) enthalten, um isoliert präsentiert werden zu können. Außerdem sollte der Ausdruck – um Einheitlichkeit zu gewährleisten – in den drei Sätzen in der gleichen Wortform auftauchen. Konnten für einen Ausdruck nicht genügend Sätze gefunden werden, die diese drei Kriterien erfüllen, wurden die Originalsätze angepasst. Ein Beispiel ist (79).

- (79) *Here, we show by gel retardation analysis that TCF-1 specifically recognizes the T beta 5 element of the TCR beta enhancer and the T delta 7 element of the TCR delta enhancer. Comparison of the sequences of all elements recognized by TCF-1 defines a consensus motif A/T A/T C A A/G A G. These observations **imply** that TCF-1 is involved in the control of several T cell-specific genes and might thus play an important role in the establishment and maintenance of the mature T cell phenotype.*

Die beiden Sätze in (79) wurden für die Untersuchung zu (80) zusammengefasst.

- (80) *TCF-1 specifically recognizes the T beta 5 element of the TCR beta enhancer and the T delta 7 element of the TCR delta enhancer, **implying** that TCF-1 is involved in the control of several T cell-specific genes.*

Weitere Anpassungen waren in Fällen nötig, in denen ein Originalsatz mehrere Ausdrücke enthält. Um die Bewertung z. B. tatsächlich dem epistemischen Adjektiv *possible* zuordnen zu können, wurde der Satz in (81a) zu (81b) geändert.

- (81) a. *These results indicate that G(Anh)MTetra induces IL-1 beta and IL-6 expression in human monocytes **suggesting** a **possible** role for G(Anh)MTetra in the release of cytokines during sepsis.*
 b. *We report a **possible** role for G(Anh)MTetra in the release of cytokines during sepsis.*

Auch für die Erstellung von Items mit kritischen Ausdrücken mussten Abänderungen vorgenommen werden, da diese in den Korpora selten sind und die vorhandenen

²⁶Ein alltagssprachliches Beispiel ist die Proposition im Satz *Es wird womöglich nachher regnen*. Der Sachverhalt an sich hat hier eine unterschiedliche (außersprachliche) Wahrscheinlichkeit, wenn zum Zeitpunkt der Äußerung die Sonne scheint oder wenn bereits dicke Wolken aufziehen.

Sätze sich nicht unbedingt für eine Befragung eignen. Um solche Sätze inhaltlich wenig zu verändern, wurden Originalsätze gewählt, in denen eine Proposition negiert wird (s. (82a)). Die Negation wurde dann durch die kritischen Ausdrücke ersetzt (s. (82b)).

- (82) a. *These results suggest that, while lipopolysaccharide-induced expression of inflammatory mediators requires tyrosine kinase activity, tyrosine kinase activity is **not** obligatory for lipopolysaccharide signal transduction.*
- b. *While lipopolysaccharide-induced expression of inflammatory mediators requires tyrosine kinase activity, tyrosine kinase activity **cannot** be obligatory for lipopolysaccharide signal transduction.*

Um herauszufinden, wie die Probanden Ausdrücke bewerten, die sich nicht auf die Wahrheit der Proposition beziehen, wurden schließlich Sätze ausgewählt, in denen die beiden Adjektive *unknown* und *unclear* wie in (83a) und (83b) keine epistemisch modale Funktion haben. Die Anzahl der Befragungssitems lag somit bei 29.

- (83) a. *The mechanisms by which GATA-1 controls cell survival are **unknown**.*
- b. *The mechanisms by which Tax activates NF-kappa B remain **unclear**, and findings presented in the literature are often controversial.*

Eine Liste der Befragungssitems steht im Anhang (s. Abschnitt A.1) in Tabelle 18.

7.3 Variablen und statistische Hypothesen

Nach der Beschreibung der ausgewählten Ausdrücke und der konstruierten Befragungssitems soll nun die konkrete Operationalisierung der theoretischen Konstrukte erläutert werden. Die verwendete Wahrscheinlichkeitsskala wurde als Rationalskala von 0% bis 100% konzipiert. Der von den Probanden auf dieser Skala zugewiesene Wert ist generell als durch die Ausdrücke implizierte Wahrscheinlichkeit bezüglich der Wahrheit einer Proposition zu verstehen. Die Skala wurde in Dezimalschritte unterteilt, um den kognitiven Aufwand aufseiten der Probanden zu verringern. Ein ähnliches Vorgehen findet sich bei Scott et al. (2012). Die Autoren messen hier auf einer elfstufigen Skala die Wahrscheinlichkeit, die durch Hedges für Krankheiten, welche in Patientenakten beschrieben werden, angezeigt wird.

In Hinblick auf die Fragestellung war es notwendig, von den Probanden sowohl eine Einschätzung zum epistemischen Grad wie auch zur Ungenauigkeit eines Ausdrucks bezüglich der Skala zu erheben. Als Ungenauigkeit wurde der Umfang eines von den Probanden festgelegten Bereichs auf der Wahrscheinlichkeitsskala definiert. Dieser Bereich war durch einen unteren sowie einen oberen Wahrscheinlichkeitswert zu begrenzen. So konnten Ausdrücke als mehr (z. B. bei einem Umfang von 10% zwischen den Grenzwerten 30% und 40%) oder weniger präzise (z. B. bei einem Umfang von

60% zwischen den Grenzwerten 40% und 100%) eingeschätzt werden. Diese Herangehensweise orientiert sich an einer psychologischen Arbeit von Bocklisch (2011). Um die Vagheit epistemisch modaler Ausdrücke zu untersuchen, werden hier von der Autorin für jeden Ausdruck eine untere sowie eine obere Grenze auf einer Wahrscheinlichkeitsskala erhoben. (vgl. Bocklisch, 2011, S.25) Als epistemischer Grad wurde in der vorliegenden Arbeit aus dem Wahrscheinlichkeitsbereich ein repräsentativer Einzelwert berechnet. Dieser entspricht der Mitte des Wahrscheinlichkeitsbereichs, also dem Mittelwert aus unterer und oberer Grenze. Die Gültigkeit dieses Vorgehens wurde in einem kleinen Pretest überprüft. Die Probanden bestätigten hier, dass sie, wenn sie einem Ausdruck einen Einzelwert zuzuweisen hätten, die Mitte der von ihnen angegebenen Wahrscheinlichkeitsspanne wählen würden. Die gemessenen abhängigen Variablen (AV) sind damit die folgenden:

AV₁: UNTERE GRENZE (0% - 100%) – rationalskaliert

AV₂: OBERE GRENZE (0% - 100%) – rationalskaliert

AV₃: EPISTEMISCHER GRAD (0% - 100%) – rationalskaliert

AV₄: UNGENAUIGKEIT (0% - 100%) – rationalskaliert

Um auf eventuelle systematische Bewertungsunterschiede zwischen Probandengruppen zu testen, erfolgte eine Einteilung nach verschiedenen Charakteristika. Zu den Annotatoren in den vorhandenen Projekten gehören sowohl Biologen als auch Linguisten. Deshalb sollte in der vorliegenden Studie bezüglich des Fachbereichs ein ähnlicher Vergleich erfolgen. Der Biomedizin zugehörig wurden Probanden definiert, die zum Zeitpunkt der Erhebung auf einem verwandten Gebiet (z. B. Genetik, Bioinformatik) tätig waren bzw. ein entsprechendes Hauptfach studierten. Probanden, die auf sprachwissenschaftlichen Feldern aktiv waren, wurden dem Fachbereich Linguistik zugeordnet. Die Vertrautheit mit Wissenschaftstexten wurde über den formalen Bildungsabschluss definiert, da davon auszugehen ist, dass Personen mit höherer wissenschaftlicher Qualifikation hier mehr Erfahrung aufweisen. Schließlich erfolgte eine Einteilung bezüglich der englischen Sprachkompetenz. Diese wurde auf einer fünfstufigen Likert-Skala erfasst. Die Sprachkompetenz wurde dabei auf den Bereich des Lesens und Schreibens bezogen, da es beim Umgang mit Wissenschaftstexten hauptsächlich um den schriftlichen Sprachgebrauch geht. Die in der Untersuchung festgelegten unabhängigen Variablen (UV) sind damit die folgenden:

UV₁: FACHBEREICH ('Biomedizin' – 'Linguistik') – nominalskaliert

UV₂: VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN ('Hochschulzugangsberechtigung' – 'Undergraduate Degree' – 'Postgraduate Degree' – 'Promotion')²⁷ – nominalskaliert

²⁷Die Bezeichnungen *Undergraduate Degree* und *Postgraduate Degree* sind an die im englischsprachigen Raum verbreiteten Terme angelehnt. Sie lassen sich im Sinne eines ersten berufsquali-

UV₃: ENGLISCHE SPRACHKOMPETENZ (‘Grundkenntnisse’ – ‘ausreichend’ – ‘gut’ – ‘sehr gut’ – ‘ausgezeichnet (muttersprachliches/bilinguales Niveau)’ – ordinalskaliert

An dieser Stelle ist darauf hinzuweisen, dass bei der Einteilung nach den Faktorstufen der unabhängigen Variablen keine Randomisierung, d. h. keine zufällige Zuordnung der Probanden, erfolgte. Da die Variablen natürlich variiert sind (die Charakteristika der Personen sind bereits vorgegeben) und nicht künstlich manipuliert werden, handelt es sich bei der Überprüfung entsprechender Einflüsse auch um ein Quasiexperiment. Diese Untersuchungsvariante hat eine eingeschränkte interne Validität, d. h., die Ergebnisse sind nur bedingt kausal interpretierbar, da eventuelle Drittvariablen nur schlecht kontrolliert werden können. (vgl. Bortz & Döring, 2006, S.53-57)

Um zu testen, ob es zwischen den Bewertungen verschiedener sprachlicher Mittel Unterschiede gibt, wurde als weitere unabhängige Variable die Art des Ausdrucks festgelegt. Die Faktorstufen ergeben sich durch die Auswahl der Befragungssitems:

UV₄: AUSDRUCK (27 verschiedene Ausdrücke)²⁸ – nominalskaliert

Entsprechend der Operationalisierung wurden für die vorläufigen Hypothesen (s. Abschnitt 6.6) die statistischen Null- und Alternativhypothesen (H_0 und H_1) formuliert. Die Überprüfung des Einflusses der Personenmerkmale sollte für jeden der 27 epistemisch modalen Ausdrücke mittels einer mehrfaktoriellen Varianzanalyse (entsprechend eines Between-Subject-Designs – jeder Proband wurde entsprechend seiner Angaben eindeutig einer Gruppe zugewiesen) erfolgen. Dabei waren neben den Haupteffekten auch ausgewählte Interaktionseffekte zu testen. Diese beziehen sich auf die Frage, ob der Einfluss der VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN oder der ENGLISCHEN SPRACHKOMPETENZ vom FACHBEREICH der bewertenden Person abhängig ist:

- Hypothese 1:

$H_{0.1}$: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich des FACHBEREICHS einer Person. ($F = 1$)

$H_{1.1}$: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich des FACHBEREICHS einer Person. ($F > 1$)

$H_{0.2}$: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich der VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN einer Person. ($F = 1$)

$H_{1.2}$: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich der VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN einer Person. ($F > 1$)

fizierenden Studienabschlusses (z. B. Bachelor) und weiterführenden Studienabschlusses (z. B. Master) definieren. Zu Letzterem sind auch die alten Bildungsabschlüsse Magister und Diplom zu zählen, worauf die Probanden durch Vorgabe entsprechender Beispiele hingewiesen wurden.

²⁸Das nicht epistemisch modale Vorkommen der Adjektive *unknown* und *unclear* zählt nicht dazu.

- H_{0.3}: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich der ENGLISCHEN SPRACHKOMPETENZ einer Person. ($F = 1$)
- H_{1.3}: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich der ENGLISCHEN SPRACHKOMPETENZ einer Person. ($F > 1$)
- H_{0.4}: Der Effekt der ENGLISCHEN SPRACHKOMPETENZ auf den geschätzten EPISTEMISCHEN GRAD eines Ausdrucks ist nicht abhängig vom FACHBEREICH einer Person. ($F = 1$)
- H_{1.4}: Der Effekt der ENGLISCHEN SPRACHKOMPETENZ auf den geschätzten EPISTEMISCHEN GRAD eines Ausdrucks ist abhängig vom FACHBEREICH einer Person. ($F > 1$)
- H_{0.5}: Der Effekt der VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN auf den geschätzten EPISTEMISCHEN GRAD eines Ausdrucks ist nicht abhängig vom FACHBEREICH einer Person. ($F = 1$)
- H_{1.5}: Der Effekt der VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN auf den geschätzten EPISTEMISCHEN GRAD eines Ausdrucks ist abhängig vom FACHBEREICH einer Person. ($F > 1$)

Die Überprüfung der UNGENAUIGKEITS-Unterschiede zwischen den Ausdrücken erfolgte mittels einer Varianzanalyse mit Messwiederholung (entsprechend eines Within-Subject-Designs – jeder Proband hatte alle Ausdrücke zu bewerten):

- Hypothese 2:

H₀: Es gibt keinen Unterschied in der geschätzten UNGENAUIGKEIT bezüglich des konkreten AUSDRUCKS. ($F = 1$)

H₁: Es gibt einen Unterschied in der geschätzten UNGENAUIGKEIT bezüglich des konkreten AUSDRUCKS. ($F > 1$)

Beim EPISTEMISCHEN GRAD wurde eine allgemeine Hypothese formuliert. Als Testverfahren war auch hier eine Varianzanalyse mit Messwiederholung vorgesehen:

- Hypothese 3:

H₀: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD bezüglich des konkreten AUSDRUCKS. ($F = 1$)

H₁: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD bezüglich des konkreten AUSDRUCKS. ($F > 1$)

Durch Post-Hoc-Analysen sollten die folgenden Unterhypothesen überprüft werden:

- Hypothese 3a:

H₀: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen den an einer deskriptiv festgelegten Kategoriengrenze liegenden Ausdrücken. ($F = 1$)

H₁: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen den an einer deskriptiv festgelegten Kategoriengrenze liegenden Ausdrücken. ($F > 1$)

- Hypothese 3b:

H_{0.1}: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *has/have been suggested*. ($F = 1$)

H_{1.1}: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *has/have been suggested*. ($F > 1$)

H_{0.2}: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *investigate + hypothesis* und *support + hypothesis*. ($F = 1$)

H_{1.2}: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *investigate + hypothesis* und *support + hypothesis*. ($F > 1$)

- Hypothese 3c:

H_{0.1}: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *suggest + may*. ($F = 1$)

H_{1.1}: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *suggest + may*. ($F > 1$)

H_{0.2}: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *may* und *suggest + may*. ($F = 1$)

H_{1.2}: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *may* und *suggest + may*. ($F > 1$)

H_{0.3}: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *may*. ($F = 1$)

H_{1.3}: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *may*. ($F > 1$)

- Hypothese 3d:

H₀: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *strongly suggest*. ($F = 1$)

H₁: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD zwischen *suggest* und *strongly suggest*. ($F > 1$)

7.4 Durchführung

Im Rahmen der Studie wurde eine Onlinebefragung durchgeführt. Der Fragebogen wurde mithilfe des Softwarepaketes SoSci Survey²⁹ erstellt und über den entsprechenden Server bereitgestellt. Die Probanden wurden gezielt per E-Mail,

²⁹<http://www.soscisurvey.de/>

Mailverteiler bzw. Newsletter kontaktiert. Die Nachricht enthielt den Link zum Fragebogen sowie die Bitte, an der empirischen Untersuchung teilzunehmen. Der Link führte zur ersten Seite des Fragebogens mit einer genauen Instruktion inklusive der Erläuterung von Beispielen. Aufgabe der Probanden war es, einzuschätzen, für wie wahrscheinlich der Autor eines Wissenschaftstextes eine Beobachtung oder ein Ergebnis (die gewählte Formulierung für die fachspezifische Bezeichnung *Proposition*) hält, wenn er diese im Kontext eines speziellen Signalwortes (die gewählte Formulierung für die fachspezifische Bezeichnung *epistemisch modaler Ausdruck*) präsentiert. Zur besseren Orientierung wurden die Ausdrücke und der sprachliche Kontext, in den sie eingebettet sind (z. B. *it is improbable that*), rot, der eigentliche Sachverhalt, der dadurch modifiziert wird, hingegen blau markiert.³⁰

Außerdem wurde den modalisierten Beispieläußerungen ein Satz ohne epistemisch modalem Ausdruck kontrastiv gegenübergestellt. Ein solcher Satz wurde auch zu den Befragungssitems hinzugefügt (Insgesamt umfasste der Fragebogen damit 30 Items.). Dieser Satz wurde als Testitem konzipiert, um sicherzustellen, dass sich die Bewertungen der Probanden tatsächlich auf die markierten Ausdrücke beziehen. Sollte sich bei der Auswertung zeigen, dass einige Probanden der unmodalisierten Äußerung eine Wahrscheinlichkeit von weniger als 100% zugewiesen hatten, wäre dies Indiz dafür, dass sich diese Probanden eher an der Grundwahrscheinlichkeit des beschriebenen Sachverhalts orientiert hatten, und ein Kriterium für den Ausschluss ihrer Bewertungen.

Die Probanden wurden schließlich darauf hingewiesen, ihre Einschätzung mithilfe zweier Schieberegler vorzunehmen, um auf der Skala denjenigen Wahrscheinlichkeitsbereich anzugeben, der ihrer Meinung nach durch die verwendeten Signalwörter impliziert wird. Sie hatten dabei aber auch die Möglichkeit, beide Regler auf einen einzelnen Wert zu schieben. Auf den anschließenden drei Seiten wurden jeweils zehn Befragungssitems zusammen mit der Wahrscheinlichkeitsskala präsentiert. Für jeden Probanden erfolgte dabei eine willkürliche Auswahl aus den drei möglichen Sätzen. Die Präsentationsreihenfolge der Befragungssitems war darüber hinaus bei jedem Fragebogensaufruf randomisiert. Am Ende wurden die relevanten personenbezogenen Daten abgefragt, durch die eine Zuweisung der Probanden zu den jeweiligen Faktorstufen der unabhängigen Variablen erfolgte. Darüber hinaus konnten Angaben zum Alter, Geschlecht und zur Muttersprache gemacht sowie Kommentare bezüglich der Fragebogenbearbeitung abgegeben werden. Ausschnitte aus dem Fragebogen befinden sich in Abschnitt A.2 im Anhang.

³⁰Um zu verhindern, dass Ausdrücke überlesen werden, wurde diese Markierung auch für die Sätze übernommen, in denen die Adjektive *unknown* und *unclear* nicht auf den Wahrheitswert der Proposition referieren, obwohl die Satzstruktur hier eine andere ist.

7.5 Probandenrekrutierung

Hinsichtlich der Rekrutierung der Probanden ist es notwendig, die Grundgesamtheit, d. h. die Population, über die Aussagen getroffen werden sollen, zu definieren. Dies gestaltete sich schwierig. Die Grundgesamtheit von Personen, die Einschätzungen zu epistemisch modalen Ausdrücken in biomedizinischen Texten abgeben können, setzt sich im Grunde aus denjenigen Personen zusammen, die mit diesen Texten umgehen. Da es sich um ein interdisziplinäres Fachgebiet handelt, gehören hierzu Personen aus der Medizin und der Biologie. Aber auch Personen, die auf angrenzenden Feldern (z. B. Bioinformatik) tätig sind, können hinzugezählt werden. Zur Diversität der Fachbereiche kommen die unterschiedlichen Nationalitäten und Muttersprachen hinzu. Ausgehend von dieser Problematik ist es kaum möglich, zu einer repräsentativen Stichprobe zu gelangen. Die Rekrutierung der Probanden erfolgte entsprechend auch nicht über die Ziehung einer Zufallsstichprobe aus der Grundgesamtheit, die eine solche Repräsentativität gewährleisten kann. Stattdessen wurde eine Ad-Hoc-Stichprobenziehung vorgenommen. (vgl. Albert & Marx, 2010, S.34) Der Nachteil solcher nicht-probabilistischen anfallenden Stichproben (Nicht jedes Element der Grundgesamtheit hat die gleiche oder wenigstens bekannte Chance, in die Stichprobe zu gelangen.) ist, dass sie nur schwer Schlussfolgerungen bezüglich der Population erlauben, d. h., nur unter Vorbehalt verallgemeinerbar sind. (vgl. Bortz & Döring, 2006, S.402) Da die vorliegende Arbeit allerdings eine erste Annäherung an eine empirisch fundierte Skalierung epistemisch modaler Ausdrücke darstellt, ist diese Art der Stichprobengewinnung vorerst als ausreichend anzusehen.

Für den Vergleich bezüglich der oben operationalisierten unabhängigen Variablen musste die Stichprobe Probanden mit entsprechenden Charakteristika umfassen. So wurden einerseits Personen aus dem Projektumfeld des Lehrstuhls für Computerlinguistik an der Friedrich-Schiller-Universität Jena kontaktiert. Diese waren in biomedizinischen Gebieten tätig. Weiterhin wurden verschiedene sprachwissenschaftliche Institute angeschrieben. Um sicherzustellen, dass die Probanden aus der Linguistik ähnliche Erfahrungen im Umgang mit englischsprachigen Wissenschaftstexten hatten, lag der Fokus dabei auf Instituten der Anglistik. Der gezielte Einbezug ausreichender Probanden mit verschiedenen Muttersprachen war nicht möglich. Hierin besteht allerdings ein nicht unerhebliches Problem, denn bei der oben beschriebenen Ad-Hoc-Stichprobe ist anzunehmen, dass deutsche Muttersprachler stark überrepräsentiert sind. Daher wurde beschlossen, zumindest Probanden mit englischer Muttersprache einzubeziehen. Englische Muttersprachler konnten über die unabhängige Variable ENGLISCHE SPRACHKOMPETENZ kategoriell eindeutig erfasst werden, was einen (eingeschränkten) Vergleich zwischen deutscher und englischer Muttersprache ermöglichen sollte. So wurden weiterhin zehn willkürlich ausgewählte biomedizinische Institute im britischen Raum kontak-

tiert. Eine ähnliche Auswahl der Probanden erfolgte bei der Vergleichsgruppe der Linguisten mit englischsprachigem Hintergrund. Die Datenerhebung selbst erfolgte innerhalb eines Zeitraums von ca. zwei Monaten im Sommersemester 2013. Alle darauf aufbauenden statistischen Berechnungen wurden mithilfe der Software IBM SPSS Statistics (Version 21)³¹ durchgeführt.

Insgesamt lieferte der Onlinefragebogen 97 Datensätze, von denen 87 auf einer vollständigen Bearbeitung beruhten. Die verbleibenden 10 Befragungen waren abgebrochen worden. Zwei Probanden hatten dabei gar keine Bewertungen vorgenommen. Bei den restlichen 8 Datensätzen lagen zwar Bewertungen für einzelne Items, allerdings keine persönlichen Angaben vor. Für die Untersuchung der faktoriellen Einflüsse waren diese Datensätze nicht verwendbar und wurden aussortiert. Ein Proband hatte kommentiert, die Aufgabenstellung falsch interpretiert zu haben, da der beschriebene Sachverhalt und nicht die durch die markierten Wörter angezeigte Wahrscheinlichkeit bewertet worden war. Auf diesen Datensatz musste deshalb ebenfalls verzichtet werden. Die Tabellen 6 und 7 beschreiben die $n = 86$ Fälle umfassende Gesamtstichprobe.³² Bezüglich des Alters sind Mittelwerte (M) und Standardabweichungen (SD) angegeben. Alle anderen Werte stellen Häufigkeiten der jeweiligen Ausprägungen bzw. Ausprägungskombinationen der Variablen dar.

Fachbereich	Geschlecht		Alter		Vertrautheit m. Wiss.-Texten			
	w	m	M	SD	HZB	UGD	GD	P
Biomedizin	22	33	34.63	11.63	1	3	27	26
Linguistik	21	8	25.93	5.18	12	7	7	3
Gesamt	43	41	31.70	10.73	13	10	34	29

Tabelle 6: Beschreibung der Stichprobe hinsichtlich Geschlecht, Alter und Vertrautheit mit Wissenschaftstexten³³

Fachbereich	Muttersprache			Englische Sprachkompetenz				
	De.	Eng.	And.	Gr.	ausr.	gut	s. gut	ausg.
Biomedizin	41	12	4	1	2	21	20	12
Linguistik	24	4	1	0	0	6	15	8
Gesamt	65	16	5	1	2	27	35	20

Tabelle 7: Beschreibung der Stichprobe hinsichtlich Muttersprache und englischer Sprachkompetenz³⁴

³¹<http://www-01.ibm.com/software/de/analytics/spss/products/statistics/>

³²Zu beachten ist, dass bei zwei Fällen keine Angabe zum Geschlecht, bei einem weiteren Fall keine Angabe zur Sprachkompetenz vorlag.

³³Zu den verwendeten Abkürzungen: w = weiblich, m = männlich, HZB = Hochschulzugangsberechtigung, UGD = Undergraduate Degree, GD = Graduate Degree, P = Promotion

³⁴Zu den verwendeten Abkürzungen: De. = Deutsch, Eng. = Englisch, And. = Andere, Gr. = Grundkenntnisse, ausr. = ausreichend, s. gut = sehr gut, ausg. = ausgezeichnet

8 Ergebnisse

Im Folgenden sollen die mittels der Onlinebefragung erhobenen Daten beschrieben und analysiert sowie die Ergebnisse der Studie präsentiert werden.

8.1 Explorative Datenanalyse

Um Auffälligkeiten in der Datenstruktur ausfindig zu machen, die die Ergebnisse verzerren könnten, wurde zuerst eine explorative Datenanalyse bezüglich der Variablen EPISTEMISCHER GRAD durchgeführt. Dabei wurden kritische Punkte entdeckt, die vor einer weiteren statistischen Analyse zu klären waren.

8.1.1 Kritische Punkte

a) Ausreißer: Zwei Fälle waren bezüglich der Bewertung mehrerer Ausdrücke als Extrem- oder Ausreißerwerte markiert. Die genaue Analyse der Datensätze zeigte insgesamt unplausible Einschätzungen dieser beiden Probanden hinsichtlich der Aufgabenstellung. So hatte eine Person für mehr als zwei Drittel der Ausdrücke eine Wahrscheinlichkeit von 0% festgelegt. Im zweiten Fall standen Einschätzungen von 100% (z. B. bei *suggest, may, appear, possible*) Bewertungen von 0% (z. B. bei *propose, can*) gegenüber. Die entsprechenden Datensätze wurden aussortiert.

b) Wahrscheinlichkeitsbereich: Nicht alle Probanden hatten jeweils einen unteren und oberen Wert zur Begrenzung eines Wahrscheinlichkeitsbereichs angegeben. So war von 27 Personen nur einer der beiden Schieberegler bewegt worden. Zur Handhabung solcher Fälle wurde erneut auf Erfahrungen aus dem Pretest zurückgegriffen. War im Pretest nur ein Schieberegler bedient worden, hatten die Probanden auf Nachfrage angegeben, die Aufgabenstellung nicht gründlich genug gelesen und jeweils nur einen einzelnen Wahrscheinlichkeitswert festgelegt zu haben, der der Interpretation eines Ausdrucks am ehesten entspricht. Bei der Datenanalyse wurde dieser Einzelwert daher als Ausprägung der Variable EPISTEMISCHER GRAD gewertet. Bei der Betrachtung von UNTERER GRENZE, OBERER GRENZE und UNGENAUIGKEIT wurden die betreffenden Datensätze hingegen ausgeschlossen.

c) Testitem: Beim Testitem wurden durchgängige Bewertungen von 100% erwartet. Die Daten wichen von dieser Annahme ab, da beinahe 40% der Probanden nicht erwartungsgemäß geantwortet hatten, wie Tabelle 8 zu entnehmen ist.

	Epistemischer Grad		
	100%	<100%	keine Angabe
<i>n</i>	53	19	14

Tabelle 8: Häufigkeit des Antwortverhaltens beim Testitem

Eine Möglichkeit der Handhabung lag darin, die fragwürdigen Fälle auszuschließen. Eventuell war hier bei allen Items die Grundwahrscheinlichkeit der beschriebenen Sachverhalte, nicht die durch die Ausdrücke implizierte Wahrscheinlichkeit bewertet worden (s. Abschnitt 7.4). Da so ein beträchtlicher Teil der Datensätze wegfallen würde, musste allerdings vorab geklärt werden, ob die Aufgabenstellung vielleicht für das Testitem zu ungenau war. Da sie darin bestand, die in den Sätzen markierten Signalwörter zu bewerten, konnte ein Satz, der keine Markierung enthält, durchaus zu Verwirrung führen. Die Annahme wurde durch Kommentare einzelner Probanden gestützt, die eine Bewertung des Testitems aufgrund der fehlenden Markierung nicht für möglich hielten bzw. eher intuitiv vorgegangen waren (d. h., ein eventuelles Signalwort war erraten worden). Es wurde also vermutet, dass das Antwortverhalten beim Testitem keinen guten Hinweis auf das allgemeine Vorgehen bot.

d) Kritische Ausdrücke: Bei Items mit kritischen Ausdrücken (Sätze mit *unlikely*, *no evidence* und *cannot*) waren hohe Maxima (*Max*) und Standardabweichungen auffällig. Einige Probanden hatten denjenigen Ausdrücken, die nach der Skalenkonzeption eigentlich geringe Wahrscheinlichkeiten implizieren, also hohe Werte zugewiesen. Tabelle 9³⁵ zeigt, dass mitunter Bewertungen von 90% bzw. 100% vorlagen. Ca. zwei Drittel der Bewertungen bezüglich *no evidence* lagen zwischen 0.00% und 70.60%, bezüglich *cannot* sogar zwischen 2.41% und 87.83%.

Ausdruck	Epistemischer Grad					
	<i>M</i>	<i>Md</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Range</i>
<i>unlikely</i>	29.70	20.00	26.19	0.00	90.00	90.00
<i>no evidence</i>	35.17	20.00	35.43	0.00	100.00	100.00
<i>cannot</i>	45.12	40.00	42.71	0.00	100.00	100.00

Tabelle 9: Bewertung der kritischen Ausdrücke

Diese Werte könnten damit erklärt werden, dass einige Probanden die Skala als Sicherheits- und nicht als Wahrscheinlichkeitsskala aufgefasst hatten. Es musste daher untersucht werden, ob das Antwortverhalten dieser Personen auch bei den anderen Ausdrücken abweichend und die Werte damit insgesamt unbrauchbar waren.

8.1.2 Auswahl gültiger Fälle

Um die kritischen Punkte zu klären und eine Entscheidung darüber zu treffen, welche Datensätze in die eigentliche Untersuchung eingehen, wurden Probandengruppen, die sich in ihrem Antwortverhalten bei bestimmten Items unterschieden, hinsichtlich ihrer Bewertung der restlichen Items verglichen. Ein signifikanter Unterschied hätte impliziert, dass die problematischen Datensätze ausgeschlossen werden müssen.

³⁵Die dort aufgeführten statistischen Kennwerte sind neben Mittelwert, Standardabweichung und Maximum außerdem Median (*Md*), Minimum (*Min*) und Spannweite (*Range*) der erhobenen Werte.

Zuerst wurden Probanden verglichen, die unterschiedlich auf das Testitem geantwortet hatten. Die Annahme war, dass die Bewertungen bei Personen, die dem Testitem nicht 100% Wahrscheinlichkeit zugewiesen hatten, generell von der Art des beschriebenen Sachverhalts abhängig waren, bei den anderen Personen hingegen nicht. Die Probanden der erstgenannten Gruppe hätten demnach die Grundwahrscheinlichkeit des jeweiligen Sachverhalts bewertet, letztere Probanden wie erwartet die Wahrscheinlichkeit, die explizit durch die Ausdrücke impliziert wird. Es war also zu überprüfen, ob es einen Interaktionseffekt zwischen dem Antwortverhalten beim Testitem und dem konkreten Sachverhalt (d. h. dem für jeden Probanden zufällig aus drei Möglichkeiten ausgewählten Satz) auf die Bewertung der Ausdrücke gab. Einem solchen Test stand jedoch entgegen, dass die Art des Sachverhalts keine bewusst manipulierte Variable darstellte. Tatsächlich lagen keinerlei Informationen zur Grundwahrscheinlichkeit der Sachverhalte vor (Es ist nicht einmal klar, wie eine solche Grundwahrscheinlichkeit überhaupt gemessen werden kann.). Daher wurde nur ein eventueller Haupteffekt des Antwortverhaltens beim Testitem auf die Bewertung der einzelnen Ausdrücke überprüft. Die abhängige Variable war der EPISTEMISCHE GRAD des jeweiligen Ausdrucks. Die unabhängige Variable wurde wie folgt operationalisiert:

UV₅: ANTWORTVERHALTEN BEIM TESTITEM ('100%' – '<100%' – 'keine Angabe')
– nominalskaliert

Die zu testenden Null- und Alternativhypothesen waren:

- Hypothese 4:

H₀: Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich des ANTWORTVERHALTENS BEIM TESTITEM.
($F = 1$)

H₁: Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich des ANTWORTVERHALTENS BEIM TESTITEM.
($F > 1$)

Der Test wurde ausschließlich bezüglich der Probanden aus dem Fachbereich Biomedizin durchgeführt. Diese hatten wie in Tabelle 10 aufgeführt geantwortet.

	Epistemischer Grad		
	100%	<100%	keine Angabe
<i>n</i>	38	11	8

Tabelle 10: Häufigkeit des Antwortverhaltens beim Testitem (Fachbereich Biomedizin)

Bei Personen aus dem Fachbereich Linguistik war davon auszugehen, dass der beschriebene Sachverhalt aufgrund der fehlenden Fachkenntnisse generell keine Rolle bei der Bewertung spielt. Die Signifikanztests erfolgten einzeln bezüglich der 27

relevanten Ausdrücke.³⁶ Bei 11 Ausdrücken wurde eine einfaktorielle ANOVA (Varianzanalyse) über die drei Stufen der UV durchgeführt, da hier die Voraussetzung der Normalverteilung der Residuen erfüllt war. Bei den restlichen 16 Ausdrücken wurde aufgrund des Nichtvorliegens einer solchen Normalverteilung und wegen der unterschiedlichen Gruppengrößen als nicht-parametrischer Test ein H-Test nach Kruskal und Wallis durchgeführt. Alle Tests erfolgten auf einem Signifikanzniveau von $\alpha = .002$, da aufgrund der vielen Einzeltests eine Bonferroni-Korrektur (Das Ausgangssignifikanzniveau von $\alpha = .05$ wurde durch die Anzahl der unabhängigen Tests geteilt.) notwendig war, um die Gefahr eines α -Fehlers, d. h. die falsche Annahme der Alternativhypothese, zu minimieren. (vgl. Nachtigall & Wirtz, 2006, S.190ff) Bei keinem der 27 Ausdrücke lag ein signifikanter Haupteffekt des ANTWORTVERHALTENS BEIM TESTITEM auf den EPISTEMISCHEN GRAD des Ausdrucks vor. Die Nullhypothese konnte also in keinem Fall verworfen werden. Probanden aus dem Fachbereich Biomedizin, die das Testitem nicht erwartungsgemäß mit 100% bewertet hatten, unterschieden sich in ihrer Bewertung der einzelnen Ausdrücke nicht von den anderen Probanden. Auf Grundlage dieser Ergebnisse schien eine Zusammenfassung der verschiedenen Datensätze gerechtfertigt.

Weiterhin wurden Probanden verglichen, die die kritischen Ausdrücke unterschiedlich bewertet hatten. Die Probandengruppierung basierte auf der Werteverteilung bei diesen Ausdrücken, wie in den Abbildungen 10, 11 und 12 wiedergegeben.

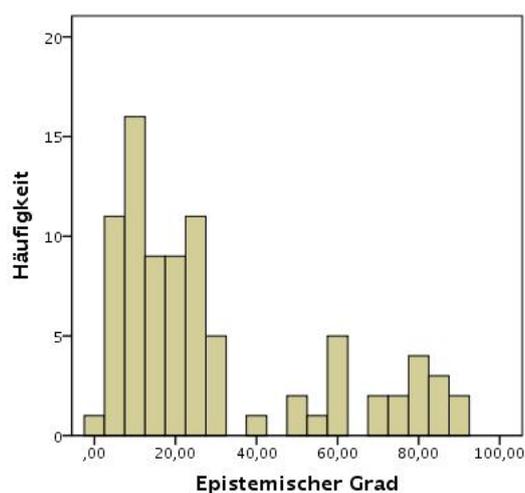


Abbildung 10: Histogramm zur Werteverteilung bei *unlikely*

³⁶Das nicht epistemisch modale Vorkommen der Adjektive *unknown* und *unclear* sowie das Testitem selbst zählten entsprechend nicht dazu.

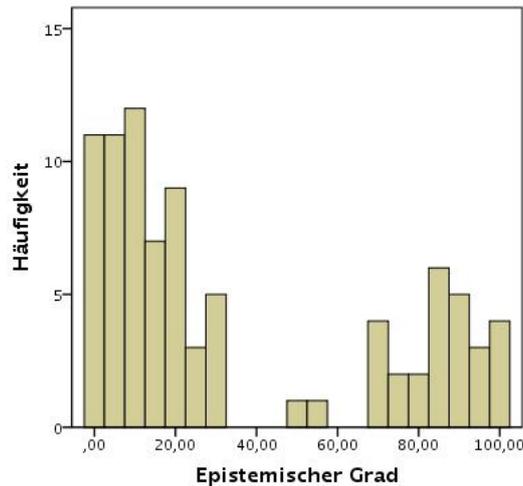


Abbildung 11: Histogramm zur Werteverteilung bei *no evidence*

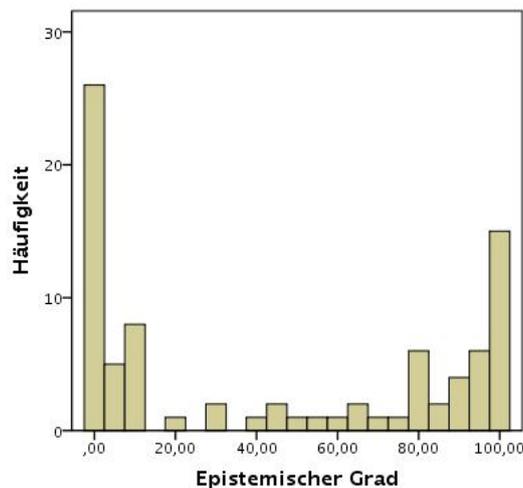


Abbildung 12: Histogramm zur Werteverteilung bei *cannot*

Ein Kolmogorov-Smirnov-Test zeigte, dass bei keinem der Ausdrücke eine Normalverteilung der Wahrscheinlichkeitsbewertungen vorlag (*unlikely*: $D(84) = 0.25$, $p = .000$; *no evidence*: $D(86) = 0.25$, $p = .000$; *cannot*: $D(85) = 0.25$, $p = .000$). Die visuelle Inspektion legte nahe, dass es sich stattdessen um bimodale bzw. binormale Verteilungen handelte, wobei zumindest bei *unlikely* und *no evidence* relativ eindeutig eine Trennung zwischen Personen, die den Ausdrücken weniger als 50% Wahrscheinlichkeit zugewiesen hatten, und Personen, deren Wahrscheinlichkeitsbewertung 50% oder mehr betrug, möglich war. Diese 50%-Grenze wurde daher als Gruppierungskriterium herangezogen. Darüber hinaus floss in die Kategorisierung ein, ob die Bewertungen über die drei Ausdrücke hinweg konsistent oder inkonsistent waren. Die unabhängige Variable wurde diesmal wie folgt operationalisiert:

UV₆: ANTWORTVERHALTEN BEI KRITISCHEN AUSDRÜCKEN

(‘durchgehend <50%’ – ‘gemischt’ – ‘durchgehend ≥50%’) – nominalskaliert

Der Umfang der Probandengruppen sah damit wie in Tabelle 11 aus.

	Epistemischer Grad		
	durchgehend < 50%	gemischt	durchgehend ≥ 50%
<i>n</i>	39	32	15

Tabelle 11: Häufigkeit des Antwortverhaltens bei kritischen Ausdrücken

Die zu testenden Null- und Alternativhypothesen waren:

- Hypothese 5:

H_0 : Es gibt keinen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich des ANTWORTVERHALTENS BEI KRITISCHEN AUSDRÜCKEN. ($F = 1$)

H_1 : Es gibt einen Unterschied im geschätzten EPISTEMISCHEN GRAD eines Ausdrucks bezüglich des ANTWORTVERHALTENS BEI KRITISCHEN AUSDRÜCKEN. ($F > 1$)

Die Signifikanztests erfolgten einzeln bezüglich der 24 relevanten Ausdrücke³⁷, analog zu oben entsprechend der Erfüllung der Normalverteilungsvoraussetzung mittels einer einfaktoriellen ANOVA (10 Ausdrücke) oder einem H-Test nach Kruskal und Wallis (14 Ausdrücke) sowie auf einem Bonferroni-korrigierten Signifikanzniveau von $\alpha = .002$. Bei 23 Ausdrücken lag kein signifikanter Haupteffekt des ANTWORTVERHALTENS BEI KRITISCHEN AUSDRÜCKEN auf den EPISTEMISCHEN GRAD des Ausdrucks vor. Die Nullhypothese konnte hier also nicht verworfen werden. Probanden, die die kritischen Ausdrücke nicht durchgängig erwartungsgemäß (d. h. mit einer geringen Wahrscheinlichkeit) bewertet hatten, unterschieden sich in ihrer Bewertung der einzelnen Ausdrücke nicht von den anderen Probanden. Bei *unknown whether* zeigte sich allerdings ein signifikanter Haupteffekt ($F(2, 78) = 10.942, p = .000, partial \eta^2 = .22$). Die Nullhypothese musste in diesem Fall also verworfen werden. Einzelvergleiche zwischen den Gruppen mittels Mann-Whitney-U-Tests zeigten, dass Probanden, die den kritischen Ausdrücken durchweg Wahrscheinlichkeiten größer oder gleich 50% zugewiesen hatten ($Md = 60.00$), sich signifikant von Probanden unterschieden, deren Wahrscheinlichkeitsbewertungen konsistent unter 50% lagen ($Md = 30.00$) ($U = 88.50, p = .000, r = -.53$), sowie von Probanden, deren Antworten gemischt waren ($Md = 40.00$) ($U = 111.50, p = .006, r = -.41$). Die letzten beiden Gruppen hatten dabei niedrigere Wahrscheinlichkeitsbewertungen abgegeben. Auf Grundlage dieser Ergebnisse wurden bei der weiteren statistischen Analyse bezüglich der kritischen Ausdrücke *unlikely*, *no evidence* und *cannot* alle Fälle aussortiert, bei denen die Bewertung gleich oder größer als 50% war. Bei der Analyse von *unknown whether* wurden alle Fälle ausgelassen, bei denen die Bewertung der kritischen Ausdrücke

³⁷Neben dem nicht epistemisch modalen Vorkommen der Adjektive *unknown* und *unclear* sowie dem Testitem zählten die kritischen Ausdrücke selbst nicht dazu.

konsistent bei oder über 50% lag. Hinsichtlich aller anderen Ausdrücke wurden die verschiedenen Datensätze zusammengefasst.

8.2 Überprüfung systematischer Bewerterunterschiede

Um für Hypothese 1 zu überprüfen, ob der FACHBEREICH, die VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN sowie die ENGLISCHE SPRACHKOMPETENZ einer Person einen Einfluss auf den geschätzten EPISTEMISCHEN GRAD eines Ausdrucks haben, wurde für jeden der relevanten 27 Ausdrücke eine 2 x 2 x 3 ANOVA gerechnet. Aufgrund mitunter kleiner und heterogener Gruppengrößen mussten einige Datensätze aus der Analyse ausgeschlossen bzw. eine Rekategorisierung von Variablen vorgenommen werden. So hatten sich nur drei Personen bei ihrer ENGLISCHEN SPRACHKOMPETENZ den Kategorien ‘Grundkenntnisse’ oder ‘ausreichend’ zugeordnet. Diese drei Fälle und damit auch die beiden Kategorien wurden nicht berücksichtigt. Außerdem wurden bei der VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN die Kategorien ‘Hochschulzugangsberechtigung’ und ‘Undergraduate Degree’ zur Kategorie ‘Student’ sowie die Kategorien ‘Graduate Degree’ und ‘Promotion’ zu ‘Wissenschaftler’ zusammengefasst. Auf Signifikanz überprüft wurden die für Hypothese 1 spezifizierten Haupt- und Interaktionseffekte. Die 27 Einzeltests erfolgten auf einem Bonferroni-korrigierten Signifikanzniveau von $\alpha = .002$. Bei Nichtvorliegen einer Normalverteilung der Residuen wurden die Werte der abhängigen Variablen zuvor durch das Ziehen der Quadratwurzel transformiert. (vgl. Field, 2009, S.202f) Bei keinem der 27 Ausdrücke liegen signifikante Haupt- oder Interaktionseffekte des FACHBEREICHS, der VERTRAUTHEIT MIT WISSENSCHAFTSTEXTEN oder der ENGLISCHEN SPRACHKOMPETENZ einer Person auf den geschätzten EPISTEMISCHEN GRAD des Ausdrucks vor. Die Nullhypothesen können also nicht verworfen werden.

Für den Ausdruck *can* sind allerdings zwei Beobachtungen anzuführen. Zwischen den beiden Faktorstufen der unabhängigen Variablen FACHBEREICH liegt ein Mittelwertsunterschied von mehr als 13 Prozentpunkten vor. Probanden aus der Biomedizin hatten dem Ausdruck im Mittel eine Wahrscheinlichkeit von 77.79% zugewiesen, Probanden aus der Linguistik hingegen 64.66%. Ähnlich ist es bei der unabhängigen Variablen ENGLISCHE SPRACHKOMPETENZ. Ausgezeichnete Sprecher hatten den Ausdruck mit durchschnittlich 85.50% Wahrscheinlichkeit bewertet, sehr gute Sprecher hingegen mit 66.62% und gute Sprecher mit 72.04%. Obwohl nach der Bonferroni-Korrektur keine signifikanten Haupteffekte des FACHBEREICHS ($F(1, 73) = 7.31, p = .009$) oder der ENGLISCHEN SPRACHKOMPETENZ ($F(2, 73) = 6.32, p = .003$) auf den EPISTEMISCHEN GRAD von

can vorliegen, sind solche Unterschiede, verglichen mit den anderen Ausdrücken, auffällig, wie die Abbildungen 13 und 14 verdeutlichen.

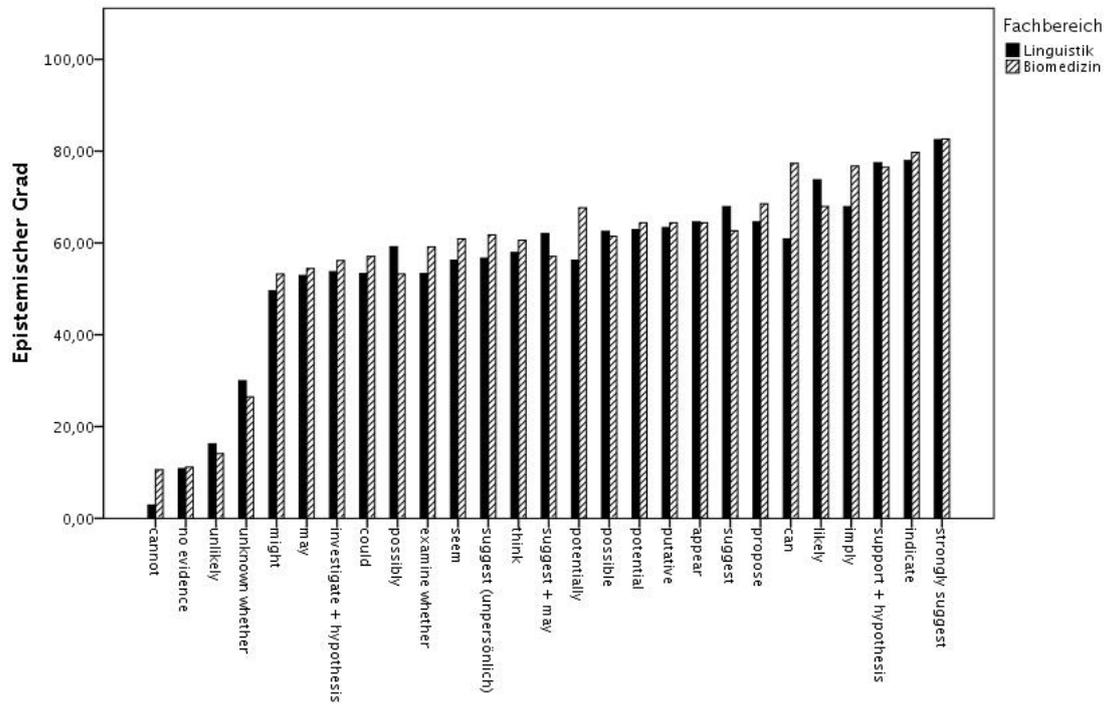


Abbildung 13: Balkendiagramm zur Bewertung nach Fachbereich

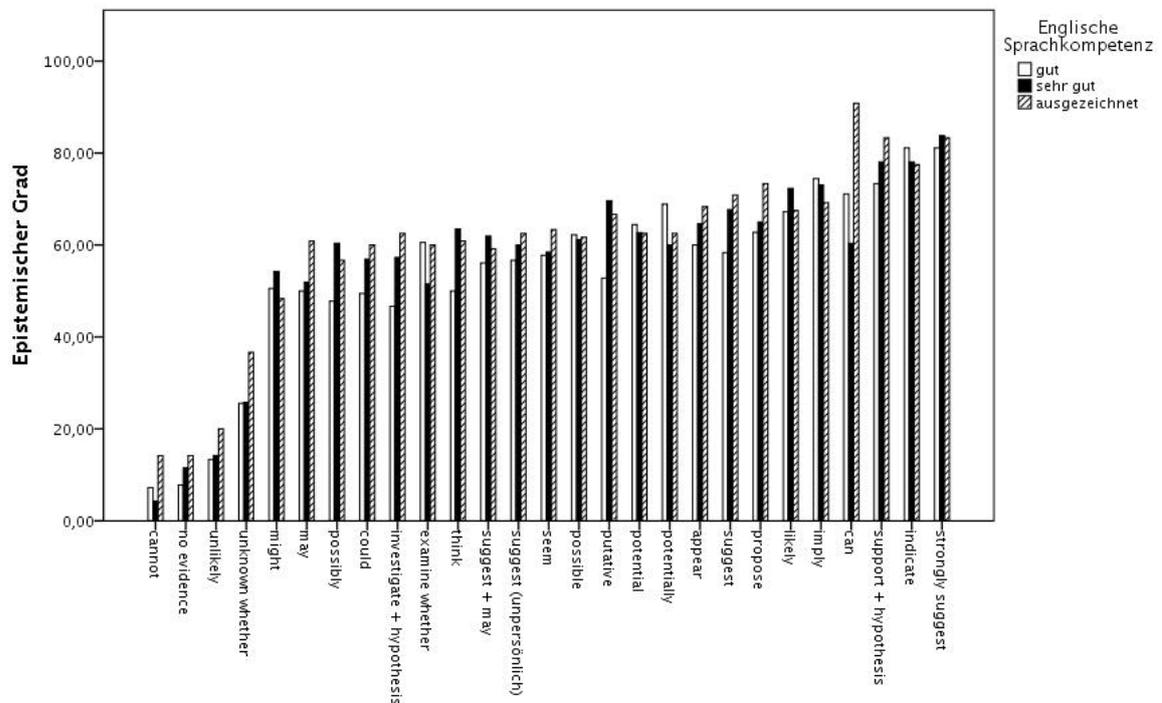


Abbildung 14: Balkendiagramm zur Bewertung nach englischer Sprachkompetenz

Aufgrund der generellen Vergleichbarkeit der Wahrscheinlichkeitsbewertungen erfolgte keine weitere Filterung der Datensätze. Die Antworten aller Probanden gingen – unabhängig vom jeweiligen Fachbereich – in die folgenden Analysen ein.

8.3 Positionierung und Gruppierung der Ausdrücke

8.3.1 Darstellung der Wahrscheinlichkeitsbereiche

Abbildung 15 präsentiert die den Ausdrücken zugewiesenen Wahrscheinlichkeitsbereiche auf der epistemischen Skala. Diese ergeben sich aus der durchschnittlichen UNTEREN GRENZE und der durchschnittlichen OBEREN GRENZE (untere und obere Querstriche im Diagramm). Außerdem ist für jeden Ausdruck der durchschnittliche EPISTEMISCHE GRAD angegeben (Kreise im Diagramm). Die Grafik zeigt, dass sich zwei hinsichtlich ihrer Positionierung auf der Skala benachbarte Ausdrücke stets überschneiden.³⁸

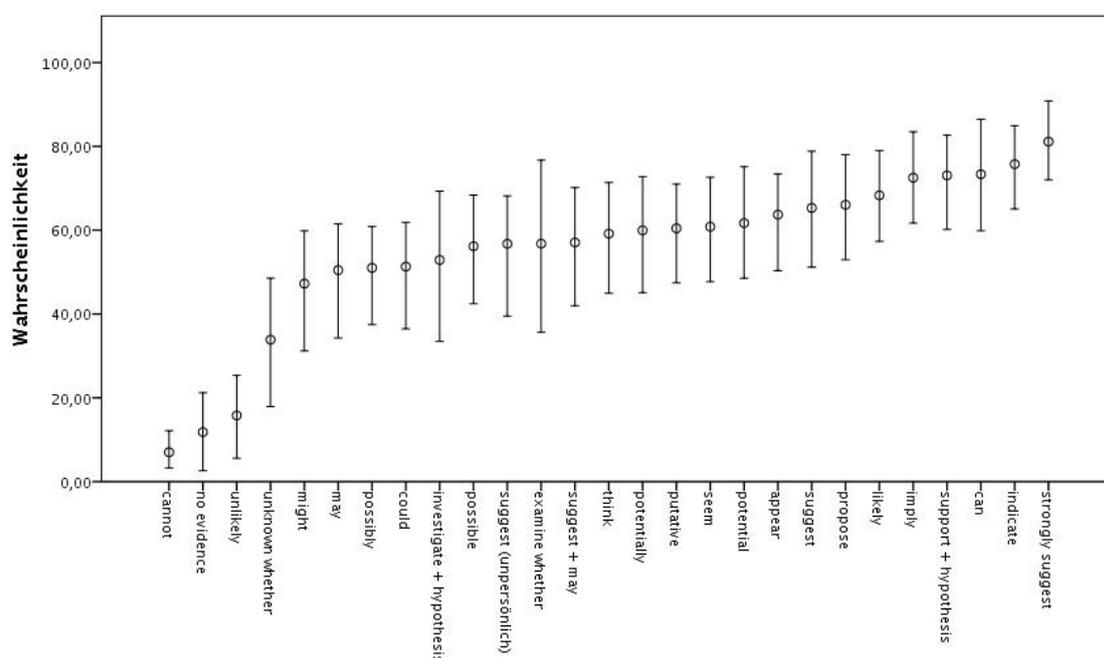


Abbildung 15: Hoch-Tief-Schluss-Diagramm zu den Wahrscheinlichkeitsbereichen

8.3.2 Genaue Beschreibung der Wahrscheinlichkeitsbereiche

Die die Wahrscheinlichkeitsbereiche kennzeichnenden statistischen Werte sind in Tabelle 12 genauer angegeben. Die relevanten Variablen sind die UNTERE GRENZE und OBERE GRENZE sowie die dadurch determinierte UNGENAUIGKEIT der Ausdrücke. Als statistische Kennwerte sind der Mittelwert und Median sowie die Standardabweichung ausgewählt. Diese Kennwerte beziehen sich auf die jeweils

³⁸Zu beachten ist, dass die Werte für den EPISTEMISCHEN GRAD nicht genau der Mitte der hier angegebenen Wahrscheinlichkeitsbereiche entsprechen, da für die Berechnung der UNTEREN GRENZE und OBEREN GRENZE nur Datensätze verwendet werden konnten, bei denen die Probanden entsprechend zwei Werte angegeben hatten (s. Abschnitt 8.1.1). Beim EPISTEMISCHEN GRAD wurden hingegen auch Fälle mit Einzelwertangaben einbezogen.

als gültig festgelegten Fälle³⁹.

Ausdruck	Ungenauigkeit			unt. Grenze			ob. Grenze			n
	M	Md	SD	M	Md	SD	M	Md	SD	
<i>cannot</i>	8.93	10.00	10.31	3.21	0.00	10.56	12.14	10.00	15.72	28
<i>no evidence</i>	18.54	10.00	17.75	2.62	0.00	5.44	21.22	20.00	16.15	41
<i>str. suggest</i>	18.83	20.00	8.46	72.00	70.00	12.73	90.83	90.00	9.62	60
<i>unlikely</i>	19.78	20.00	11.96	5.56	0.00	7.56	25.33	20.00	13.25	45
<i>indicate</i>	19.84	20.00	13.23	65.08	70.00	17.67	84.92	90.00	16.49	61
<i>likely</i>	21.67	20.00	11.81	57.33	60.00	16.76	79.00	80.00	14.48	60
<i>imply</i>	21.83	20.00	13.47	61.67	60.00	20.10	83.50	90.00	14.94	60
<i>supp. + hyp.</i>	22.50	20.00	12.97	60.17	60.00	17.80	82.67	90.00	14.48	60
<i>appear</i>	23.11	20.00	12.72	50.33	50.00	17.22	73.44	80.00	18.25	61
<i>possibly</i>	23.39	20.00	11.83	37.46	40.00	17.18	60.85	60.00	19.50	59
<i>putative</i>	23.62	20.00	14.59	47.41	50.00	24.17	71.03	75.00	21.74	58
<i>seem</i>	24.92	20.00	15.01	47.70	50.00	19.01	72.62	70.00	19.05	61
<i>propose</i>	25.08	20.00	16.29	52.95	50.00	20.60	78.03	80.00	18.60	61
<i>could</i>	25.42	20.00	16.85	36.44	40.00	16.48	61.86	60.00	17.47	59
<i>possible</i>	25.90	20.00	17.45	42.46	40.00	20.47	68.36	70.00	20.59	61
<i>think</i>	26.44	20.00	15.06	44.92	40.00	19.33	71.36	70.00	17.67	59
<i>can</i>	26.67	20.00	26.98	59.83	65.00	29.72	86.50	90.00	17.16	60
<i>potential</i>	26.67	20.00	17.63	48.50	50.00	20.90	75.17	70.00	16.42	60
<i>may</i>	27.21	20.00	16.24	34.26	40.00	18.67	61.48	60.00	19.31	61
<i>potentially</i>	27.70	20.00	17.46	45.08	50.00	20.46	72.69	80.00	16.45	61
<i>suggest</i>	27.70	20.00	14.99	51.15	50.00	18.13	78.85	80.00	13.91	61
<i>sugg. may</i>	28.20	30.00	14.89	41.97	50.00	17.68	70.16	70.00	16.88	61
<i>suggest (u.)</i>	28.67	20.00	17.61	39.50	40.00	18.81	68.17	70.00	19.00	60
<i>might</i>	28.67	20.00	21.59	31.17	30.00	17.38	59.83	60.00	19.00	60
<i>unknown wh.</i>	30.63	20.00	29.92	17.92	10.00	21.13	48.54	50.00	29.10	48
<i>inv. + hyp.</i>	35.86	20.00	36.66	33.44	30.00	30.92	69.31	70.00	28.58	58
<i>examine wh.</i>	41.07	20.00	40.81	35.46	40.00	32.76	76.73	90.00	26.53	55

Tabelle 12: Kennwerte zu den Variablen der Wahrscheinlichkeitsbereiche

Einen besonders geringen Umfang des Wahrscheinlichkeitsbereichs auf der epistemischen Skala weist der Ausdruck *cannot* auf. Die UNGENAUIGKEIT beträgt hier 8.93%. Ein besonders großer Umfang von 41.07% liegt hingegen beim Ausdruck *examine whether* vor. Während die Mittelwerte der UNGENAUIGKEIT der Ausdrücke zwischen diesen beiden Extremwerten variieren, ist der gegenüber Ausreißern

³⁹Bei den kritischen Ausdrücken *cannot*, *no evidence* und *unlikely* sind dies nur Bewertungen unterhalb der 50%-Grenze. Bei *unknown whether* wurden ausschließlich Fälle einbezogen, bei denen diese kritischen Ausdrücke nicht durchgängig mit Wahrscheinlichkeiten gleich oder über 50% bewertet worden waren. (s. Abschnitt 8.1.2)

robuste Median relativ stabil (mit Ausnahme von drei Ausdrücken beträgt der Median 20.00%). Die durchschnittliche UNGENAUIGKEIT über alle Ausdrücke hinweg (d. h. der Mittelwert der Mittelwerte) kann mit 25.14% angegeben werden.

Die die Positionierung im engeren Sinne kennzeichnenden Werte sind in Tabelle 13 analog zu oben genauer angegeben. Die relevante Variable ist der EPISTEMISCHE GRAD der Ausdrücke.

Ausdruck	Epistemischer Grad			n
	M	Md	SD	
<i>cannot</i>	7.00	0.00	12.22	45
<i>no evidence</i>	11.81	10.00	9.30	58
<i>unlikely</i>	15.79	15.00	8.76	63
<i>unknown wh.</i>	33.86	37.50	19.92	66
<i>might</i>	47.24	50.00	15.11	85
<i>may</i>	50.47	50.00	17.37	86
<i>possibly</i>	51.01	50.00	16.95	84
<i>could</i>	51.31	50.00	15.49	84
<i>inv. + hyp.</i>	52.90	50.00	22.87	81
<i>possible</i>	56.16	52.50	14.43	86
<i>suggest (u.)</i>	56.76	60.00	17.81	85
<i>examine wh.</i>	56.78	50.00	19.94	79
<i>sugg. + may</i>	57.06	60.00	15.55	85
<i>think</i>	59.16	60.00	17.65	83
<i>potentially</i>	59.94	60.00	15.65	85
<i>putative</i>	60.43	60.00	21.86	82
<i>seem</i>	60.82	60.00	16.22	85
<i>potential</i>	61.71	60.00	16.19	85
<i>appear</i>	63.71	65.00	16.59	85
<i>suggest</i>	65.30	65.00	13.85	84
<i>propose</i>	66.06	70.00	17.48	85
<i>likely</i>	68.31	70.00	14.38	83
<i>imply</i>	72.53	70.00	15.54	85
<i>supp. + hyp.</i>	73.06	75.00	14.48	85
<i>can</i>	73.35	80.00	20.46	85
<i>indicate</i>	75.76	80.00	14.98	86
<i>str. suggest</i>	81.14	80.00	10.22	83

Tabelle 13: Kennwerte zu den Variablen der Skalenposition

Eine besonders geringe Wahrscheinlichkeit wird durch den Ausdruck *cannot* impliziert. Der EPISTEMISCHE GRAD beträgt hier 7.00%. Eine besonders hohe Wahrscheinlichkeit zeigt hingegen der Ausdruck *strongly suggest* mit einem EPISTEMISCHEN GRAD von 81.14% an.

8.3.3 Gruppierung der Ausdrücke

Um herauszufinden, inwieweit eine Gruppierung der Ausdrücke möglich ist, wurde eine hierarchische Clusteranalyse durchgeführt. Die Clusteranalyse stellt ein deskriptivstatistisches Klassifizierungsverfahren dar, mit dem eine nicht weiter differenzierte Menge von Objekten hinsichtlich einer Anzahl von Merkmalen in homogene Gruppen unterteilt werden kann. (vgl. Rasch & Kubinger, 2006, S.385f) Die Elemente innerhalb der Gruppen sollen dabei in ihren Variablenausprägungen möglichst ähnlich und zwischen den Gruppen möglichst verschieden sein. (vgl. Gries, 2009, S.307) Bei der hierarchischen Clusteranalyse bilden alle Objekte der Grundmenge ein eigenes Cluster. Nacheinander werden daraufhin jeweils die beiden ähnlichsten Cluster auf der Grundlage spezifischer Kriterien zusammengefügt, sodass am Ende ein Cluster übrig bleibt, welches alle Objekte umfasst. Die verschiedenen Stufen der Clusterbildung können näher analysiert werden, um Cluster mit ähnlichen Objekten ausfindig zu machen. (vgl. Field, 2009, S.3)

Die zu gruppierenden Objekte waren bei der hier durchgeführten Clusteranalyse die epistemisch modalen Ausdrücke. Als Variablen, hinsichtlich derer eine Gruppierung erfolgt, wurden die UNTERE GRENZE und OBERE GRENZE festgelegt. Als Distanzmaß wurde die quadrierte euklidische Distanz gewählt. Der verwendete Cluster-Algorithmus war das Average-Linkage-Verfahren innerhalb der Gruppen. Die Distanz zwischen zwei Clustern entspricht hier dem Mittel der Distanzen aller Objektpaare, die sich bezüglich der beiden Cluster bilden lassen, einschließlich Paaren aus demselben Cluster. (vgl. Brosius, 2011, 734-739)

Die sukzessiven Clusterstufen sowie die relativen Distanzen zwischen den jeweils gebildeten Clustern sind in Abbildung 16 präsentiert. Dem Dendrogramm lässt sich entnehmen, dass die ersten 19 Stufen der Clusterbildung eine Acht-Cluster-Lösung ergeben. Die Ausdrücke *unknown whether*, *strongly suggest* und *cannot* liegen dabei noch als singuläre Cluster vor. Bei einer Drei-Cluster-Lösung sind alle Ausdrücke zu komplexeren Clustern hinzugefügt, wobei die Distanz zwischen den zusammengeführten Clustern auf einer Skala von 0 bis 25 bei maximal 5 liegt. Die Gruppierungen dieser Drei-Cluster-Lösung sehen dabei wie in Tabelle 14 angegeben aus.

Cluster	Ausdrücke
Cluster 1	<i>unlikely, no evidence, cannot</i>
Cluster 2	<i>imply, support + hypothesis, can, indicate, suggest, propose, likely, strongly suggest</i>
Cluster 3	<i>could, possibly, may, might, investigate + hypothesis, suggest + may, possible, has/have been suggested, think, potentially, putative, seem, appear, potential, examine whether, unknown whether</i>

Tabelle 14: Präsentation der Drei-Cluster-Lösung

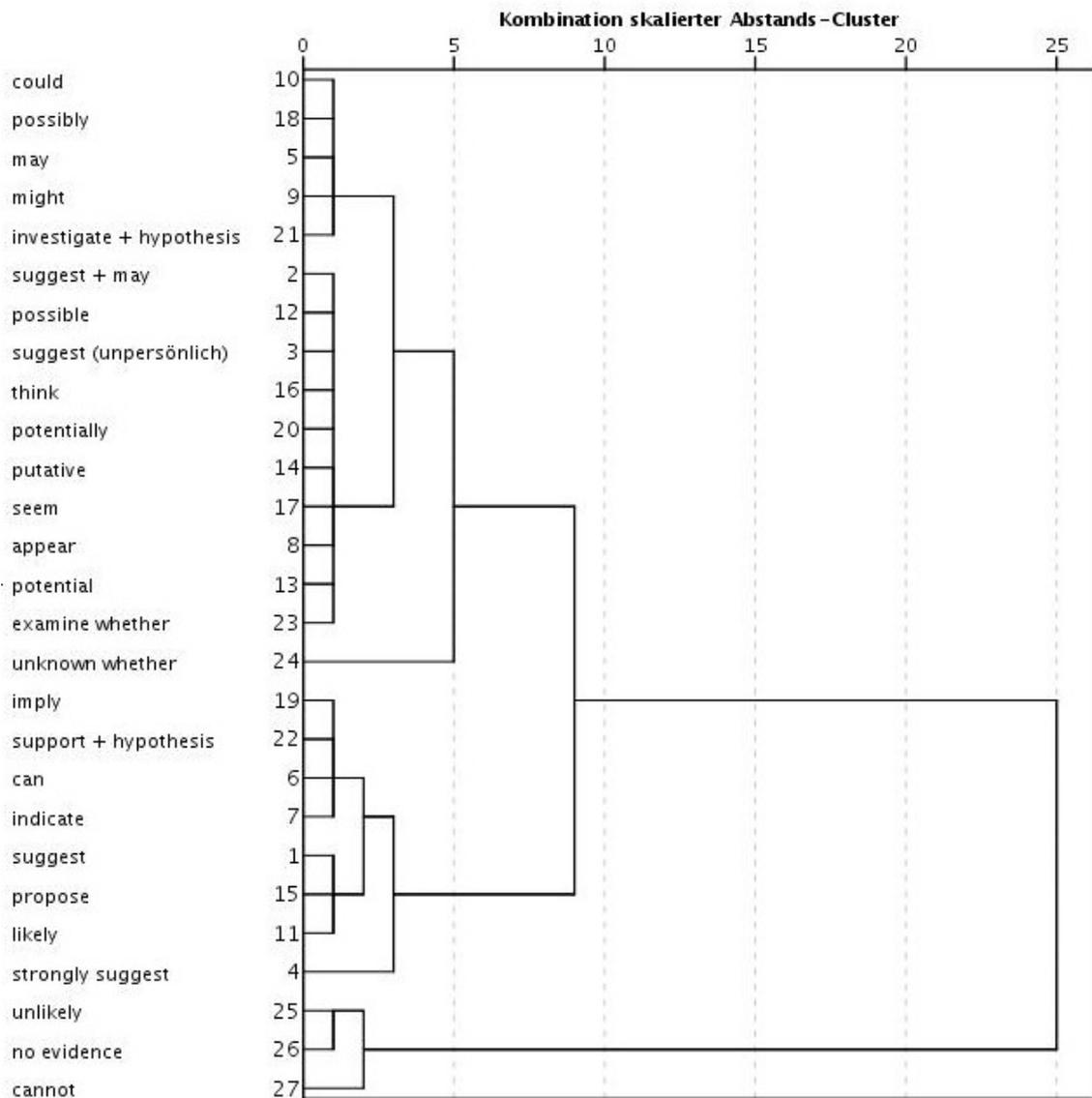


Abbildung 16: Dendrogramm zur hierarchischen Clusteranalyse

8.4 Überprüfung der Ungenauigkeits-Unterschiede

Abbildung 17 gibt einen Überblick über die Verteilung der von den einzelnen Probanden geschätzten UNGENAUIGKEITS-Werte bezüglich der verschiedenen Ausdrücke. Entsprechend der Angaben in Tabelle 12 ist die zentrale Lage der Werte (d. h. der Median) über die Ausdrücke hinweg relativ ähnlich. Bezüglich der Streuung der Werte weisen die Ausdrücke *examine whether*, *investigate + hypothesis* und *unknown whether* die größten Interquartilsabstände (und in Tabelle 12 die höchsten Standardabweichungen) auf.

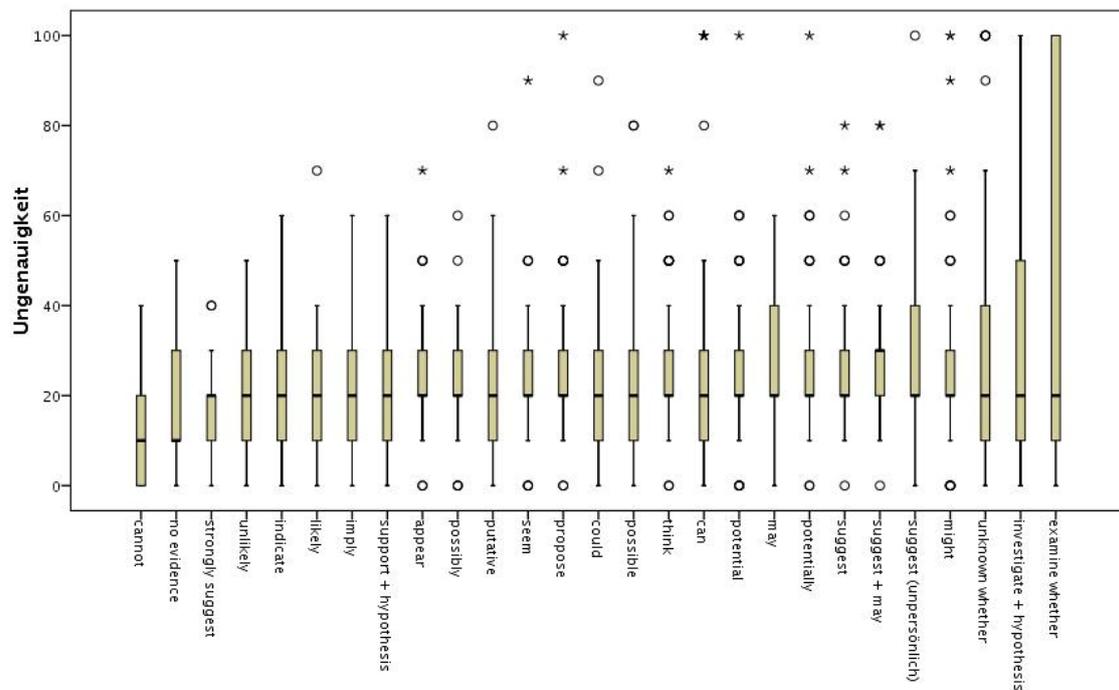


Abbildung 17: Boxplots zur Ungenauigkeit

Um für Hypothese 2 zu überprüfen, ob die und welche Ausdrücke sich in ihrer UNGENAUIGKEIT unterscheiden, wurde eine einfaktorische ANOVA mit Messwiederholung durchgeführt. In den Datensätzen wurden einzelne fehlende Werte zuvor durch Mittelwertimputation ersetzt, da für eine Varianzanalyse mit Messwiederholung vollständige Datensätze vorliegen müssen, einige Probanden aber nicht alle Ausdrücke bewertet hatten. Die unabhängige Variable als Innersubjektfaktor war der AUSDRUCK (23 Faktorstufen). Die drei kritischen Ausdrücke *unlikely*, *no evidence* und *cannot* sowie der Ausdruck *unknown whether* wurden, da hier aufgrund der Festlegung gültiger Datensätze für zahlreiche Fälle keine Bewertung vorlag (s. Abschnitt 8.1.2), gesondert betrachtet (s. u.). Die abhängige Variable war die UNGENAUIGKEIT. Die Ergebnisse des Mauchly-Tests zeigten, dass die Annahme der Sphärizität verletzt war ($\chi^2(252) = 870.79, p = .000$). Die Freiheitsgrade wurden deshalb mithilfe der Greenhouse-Geisser-Schätzung korrigiert ($\epsilon = .29$). Es liegt ein signifikanter Effekt des AUSDRUCKS auf die UNGENAUIGKEIT vor ($F(6.28, 376.63) = 6.28, p = .000, \text{partial } \eta^2 = .10$). Um herauszufinden, welche Ausdrücke sich im Speziellen in ihrer UNGENAUIGKEIT unterscheiden, wurden Post-Hoc-Vergleiche nach Bonferroni durchgeführt. Tabelle 15 gibt die Ergebnisse an. Eine leere Zelle entspricht der Nicht-Signifikanz des Unterschieds zwischen zwei Ausdrücken. Ein signifikanter Unterschied in der UNGENAUIGKEIT zwischen zwei Ausdrücken ist hingegen durch die Angabe des entsprechenden p-Werts gekennzeichnet. Die Ausdrücke *strongly suggest*, *indicate* und *likely* unterscheiden sich in ihrer UNGENAUIGKEIT signifikant von einigen anderen Ausdrücken. Zwischen den restlichen Ausdrücken liegen keine signifikanten Unterschiede vor.

Die vier Ausdrücke *unlikely*, *no evidence*, *cannot* und *unknown whether* konnten nicht als Faktorstufen in die obige Analyse einbezogen werden. Um trotzdem einen Vergleich hinsichtlich der UNGENAUGKEIT vorzunehmen, wurden die Konfidenzintervalle für die Mittelwerte bei diesen vier Ausdrücken mit den Konfidenzintervallen bei den restlichen 23 Ausdrücken verglichen. Hierfür wurden im Sinne einer Bonferroni-Korrektur jeweils 99.8%-Konfidenzintervalle berechnet (Das Ausgangssignifikanzniveau von $\alpha = .05$ wurde durch die Anzahl der unabhängigen Intervallberechnungen geteilt und von 1 – entsprechend 100% – subtrahiert.). Ein 99.8%-Konfidenzintervall stellt denjenigen Wertebereich dar, in den der wahre Populationsmittelwert bei 99.8% der gezogenen Stichproben fällt. Auf dieser Grundlage sind Gruppenvergleiche möglich. Es ist dann von einem signifikanten Unterschied zwischen verschiedenen Gruppen auszugehen, wenn sich die entsprechenden Konfidenzintervalle nicht überlappen. (vgl. Field, 2009, S.48) Abbildung 18 gibt einen Überblick zu den 99.8%-Konfidenzintervallen der UNGENAUGKEITS-Mittelwerte bei den verschiedenen Ausdrücken (Die Intervalle werden durch die Begrenzungen angegeben; die Kreise entsprechend den Mittelwerten.).

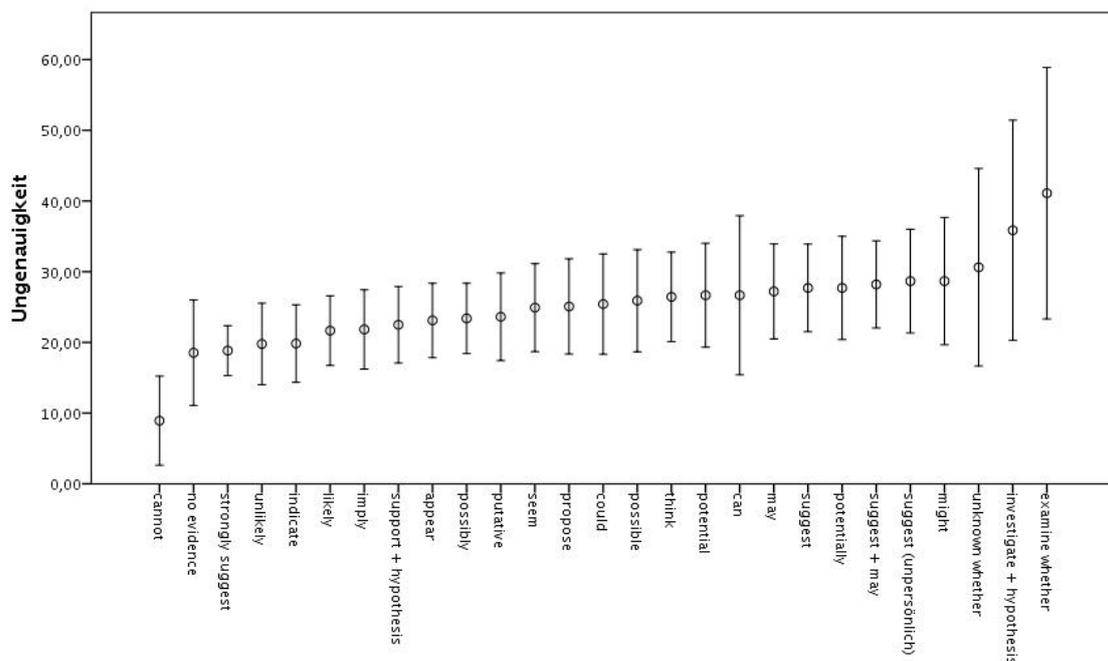


Abbildung 18: Hoch-Tief-Schluss-Diagramm zu 99.8%-Konfidenzintervallen der Ungenauigkeits-Mittelwerte

Es liegt ein signifikanter Unterschied zwischen *cannot* und allen Ausdrücken außer *no evidence*, *strongly suggest*, *unlikely*, *indicate*, *can* und *unknown whether* vor. Alle anderen Unterschiede zwischen den kritischen Ausdrücken und den restlichen 23 Ausdrücken sind nicht signifikant.

8.5 Überprüfung der Unterschiede im epistemischen Grad

Abbildung 19 gibt einen Überblick über die Verteilung der von den einzelnen Probanden geschätzten Werte des EPISTEMISCHEN GRADS bezüglich der verschiedenen Ausdrücke. Entsprechend der Angaben in Tabelle 13 sind die Ausdrücke relativ gleichmäßig über die epistemische Skala hinweg verteilt und decken diese ab. Bezüglich der Streuung der Werte weisen die Ausdrücke *unknown whether*, *putative* und *can* die größten Interquartilsabstände auf.

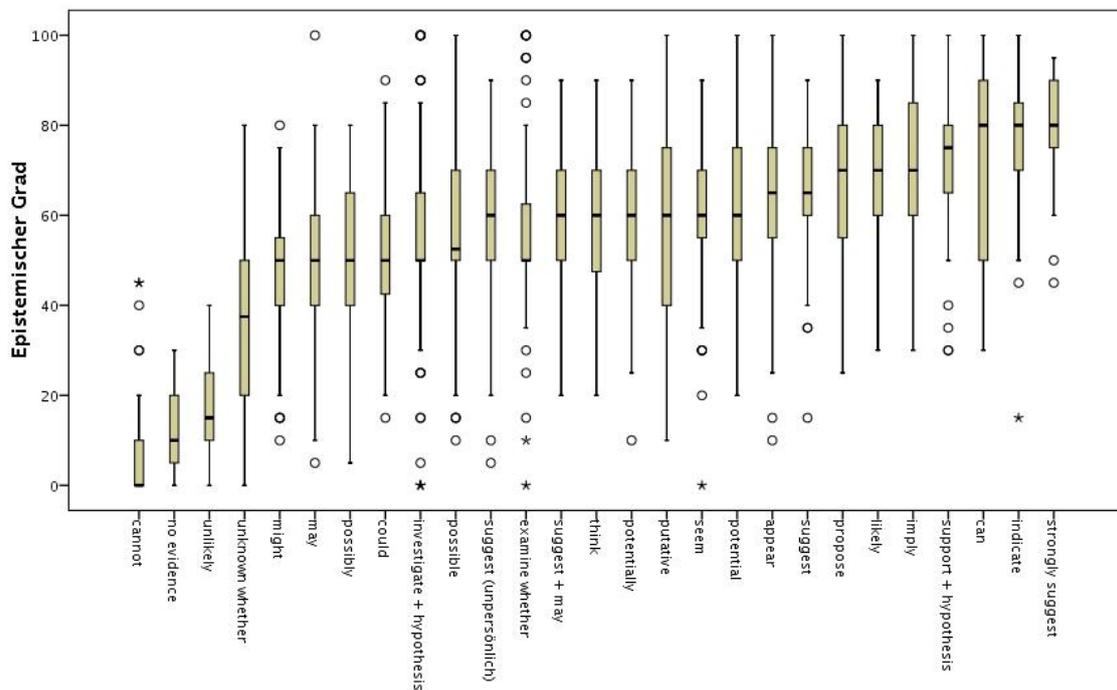


Abbildung 19: Boxplots zum epistemischen Grad

Um für Hypothese 3 zu überprüfen, ob die und welche Ausdrücke sich in ihrem EPISTEMISCHEN GRAD unterscheiden, wurde erneut eine einfaktorielle ANOVA mit Messwiederholung durchgeführt. Die unabhängige Variable als Innersubjektfaktor war der AUSDRUCK (23 Faktorstufen). Die drei kritischen Ausdrücke *unlikely*, *no evidence* und *cannot* sowie der Ausdruck *unknown whether* wurden gesondert betrachtet (s. u.). Die abhängige Variable war der EPISTEMISCHE GRAD. Die Ergebnisse des Mauchly-Tests zeigten, dass die Annahme der Sphärizität verletzt war ($\chi^2(252) = 448.68, p = .000$). Die Freiheitsgrade wurden deshalb mithilfe der Greenhouse-Geisser-Schätzung korrigiert ($\epsilon = .63$). Es liegt ein signifikanter Effekt des AUSDRUCKS auf den EPISTEMISCHEN GRAD vor ($F(13.80, 1172.98) = 33.26, p = .000, partial \eta^2 = .28$). Für die Unterschiede im Speziellen wurden Post-Hoc-Vergleiche nach Bonferroni durchgeführt. Tabelle 16 zeigt die entsprechenden Ergebnisse analog zu oben an. Ein signifikanter Unterschied im EPISTEMISCHEN GRAD zwischen zwei Ausdrücken ist durch Angabe des p-Werts gekennzeichnet.

Die Ergebnisse zeigen (da die Ausdrücke nach EPISTEMISCHEM GRAD geordnet aufgeführt sind), dass zwei hinsichtlich ihrer Positionierung auf der epistemischen Skala benachbarte Ausdrücke sich jeweils nicht signifikant voneinander unterscheiden. Für die vier Ausdrücke *unlikely*, *no evidence*, *cannot* und *unknown whether* erfolgte erneut ein Vergleich der 99.8%-Konfidenzintervalle der Mittelwerte. Abbildung 20 gibt einen Überblick zu diesen Konfidenzintervallen der Mittelwerte des EPISTEMISCHEN GRADS bei den verschiedenen Ausdrücken.

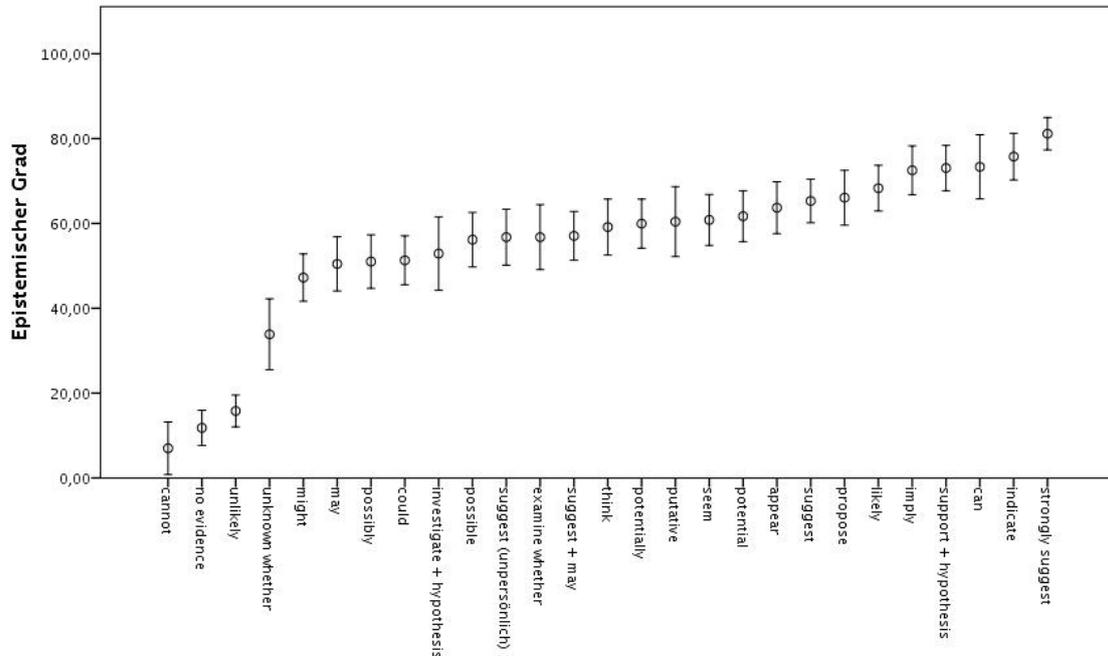


Abbildung 20: Hoch-Tief-Schluss-Diagramm zu 99.8%-Konfidenzintervallen der Mittelwerte des epistemischen Grades

Es liegt ein signifikanter Unterschied zwischen *unknown whether* und den drei kritischen Ausdrücken vor. *Cannot*, *no evidence* und *unlikely* sind nicht signifikant verschieden voneinander. Jeder der vier Ausdrücke unterscheidet sich jedoch signifikant von den restlichen 23 Ausdrücken.

Die Ergebnisse der Post-Hoc-Vergleiche bilden die Grundlage zur Überprüfung der Unterhypothesen 3b, 3c und 3d. Bezüglich Hypothese 3b liegt ein signifikanter Unterschied im EPISTEMISCHEN GRAD zwischen *suggest* und *has/have been suggested* ($p = .004$) sowie zwischen *investigate + hypothesis* und *support + hypothesis* ($p = .000$) vor. Die Nullhypothese kann also in beiden Fällen verworfen werden. Bezüglich Hypothese 3c gibt es einen signifikanten Unterschied zwischen *suggest* und *suggest + may* ($p = .009$), zwischen *may* und *suggest + may* ($p = .016$) sowie zwischen *suggest* und *may* ($p = .000$). Auch hier muss die Nullhypothese jeweils verworfen werden. Bezüglich Hypothese 3d ist schließlich ebenfalls ein signifikanter Unterschied zwischen *suggest* und *strongly suggest* ($p = .000$) zu verzeichnen, weshalb auch hier die Nullhypothese verworfen werden kann.

8.6 Nicht epistemische Vorkommen

Die hinsichtlich ihrer epistemisch modalen Funktion ambiguen Adjektive *unclear* und *unknown* wurden sehr heterogen bewertet, wenn sie sich nicht auf die Wahrheit der Proposition beziehen. Dies zeigen die Abbildungen 21 und 22.

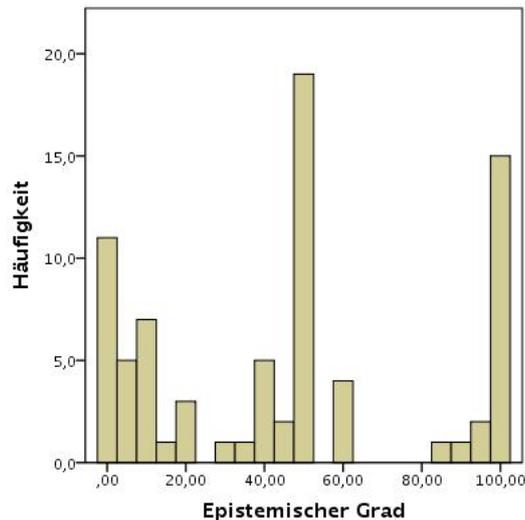


Abbildung 21: Histogramm zur Werteverteilung bei *unknown* (kein Wahrheitswert)

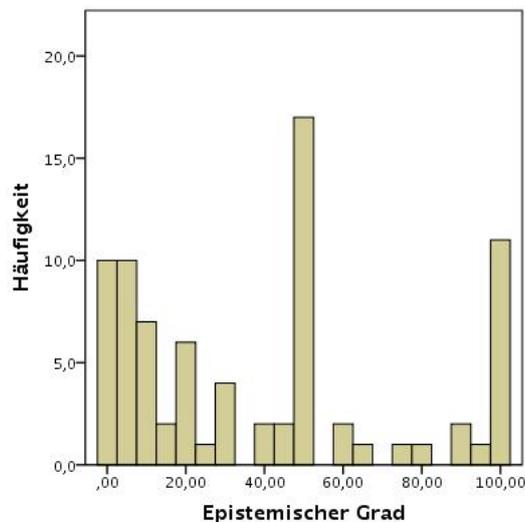


Abbildung 22: Histogramm zur Werteverteilung bei *unclear* (kein Wahrheitswert)

Der Modus der Verteilungen liegt bei 50%. Viele Probanden hatten aber auch sehr niedrige oder sehr hohe Wahrscheinlichkeitswerte angegeben. In sechs Fällen war das Item mit *unclear* nicht bewertet worden, acht Personen hatten die Bewertung von *unknown* ausgelassen. Fünf Probanden hatten die Kommentarfunktion genutzt und angegeben, dass bei beiden Items aufgrund der abweichenden Struktur der Äußerung die Ausdrücke eher nicht als Einschränkung der Wahrscheinlichkeit einer wissenschaftlichen Beobachtung bzw. eines Ergebnisses interpretiert werden können.

9 Diskussion

Im Folgenden sollen die Ergebnisse der Studie hinsichtlich der Fragestellung und unter Beachtung der theoretischen Überlegungen in Teil I näher diskutiert werden.

9.1 Angemessenheit der Skalenkonzeption

Gerade hinsichtlich der Befunde der explorativen Datenanalyse und der Auswahl gültiger Fälle (s. Abschnitt 8.1) ist zu klären, inwieweit die Wahrscheinlichkeitsskala tatsächlich angemessener als die bisher verwendeten Sicherheitsskalen ist. Als Bedeutungsaspekt, den die epistemische Skala repräsentiert, wurde die Sachverhaltswahrscheinlichkeit verwendet, weil hinsichtlich der kritischen Ausdrücke (d. h. der Ausdrücke, die zwar hohe Urteilssicherheit des Autors, aber geringe Wahrscheinlichkeit des Sachverhalts implizieren) bestimmte semantische Kontraste (z. B. zwischen *likely* und *unlikely*) eindeutiger darstellbar sind und kein Informationsverlust eintritt. Es scheint deshalb sinnvoll, auf das Antwortverhalten bei den kritischen Ausdrücken näher einzugehen. Die bimodalen bzw. binormalen Werteverteilungen legen hier nahe, dass es u. U. zwei verschiedene Vorgehensweisen bei der Interpretation dieser Ausdrücke gibt. So lässt sich eine Probandengruppe identifizieren, die den Ausdrücken erwartungsgemäß geringe Wahrscheinlichkeitswerte (unterhalb von 50%) zugewiesen hatte. Bei einer zweiten differenzierbaren Gruppe hingegen liegen die Werte bei 50% oder mehr. Das Verhalten der zweiten Gruppe wirft die Frage auf, wie die epistemische Skala hier verwendet worden war und ob eine eventuell abweichende Interpretation auf alle oder nur auf die kritischen Items zutrifft. Auffällig ist, dass die Bewertungen nicht immer konsistent sind. Knapp ein Fünftel der Probanden hatte den kritischen Ausdrücken durchgängig hohe, d. h. abweichende, Wahrscheinlichkeitswerte zugewiesen. Bei zirka einem Drittel sind die Antworten hingegen gemischt. Dies deutet darauf hin, dass eine nicht erwartungsgemäße Interpretation sich nur auf bestimmte Ausdrücke bezieht.

Zum einen scheint es generell einen Kontrast zwischen den kritischen und anderen sprachlichen Mitteln zu geben. Probanden, die sich in ihrem Antwortverhalten bei den kritischen Ausdrücken unterscheiden, zeigen keine systematischen Unterschiede bei den restlichen Bewertungen. So ist lediglich bei *unknown whether* ein Einfluss des Antwortverhaltens festzustellen. Ansonsten sind die Bewertungen über die verschiedenen Probandengruppen hinweg vergleichbar. Es bieten sich nun verschiedene Erklärungen dazu an, warum es speziell bei den kritischen Ausdrücken Unterschiede in der Interpretation geben könnte. Zentral ist dabei die Beobachtung, dass hier die Negation eine wichtige Rolle spielt. In Anlehnung an Halliday und Matthiessen (2004) (s. Abschnitt 4.2) kann das negative Merkmal einerseits auf die Modalität (s.

(84a) und (84c)) und andererseits auch auf die Proposition übertragen werden (s. (84b) und (84d)).

- (84)
- a. *It is **unlikely** that de novo induction of GATA-1 is required for globin gene activation following cell fusion.*
 - b. *It is **likely** that de novo induction of GATA-1 is **not** required for globin gene activation following cell fusion.*
 - c. *There is **no evidence** that TK activity is obligatory for CD14-mediated signal transduction to occur in response to LPS.*
 - d. *There is **evidence** that TK activity is **not** obligatory for CD14-mediated signal transduction to occur in response to LPS.*

Denkbar wäre, dass bei der Befragung die Äußerungen mit kritischen Ausdrücken mitunter entsprechend semantisch paraphrasiert wurden. Möglicherweise konnte das negative Element nicht immer eindeutig der epistemischen Modalität zugewiesen werden. Stattdessen wurde die negierte Proposition als wissenschaftliche Beobachtung oder als Ergebnis, d. h. als Sachverhalt, gedeutet. Die Wahrscheinlichkeitsbewertung bezieht sich dann darauf, dass der in den Äußerungen beschriebene Zusammenhang zwischen den genannten biologischen Prozessen nicht besteht, was die mitunter hohen Werte erklären könnte. Eine weitere Möglichkeit besteht darin, dass die Bewertung der kritischen Ausdrücke entgegen der Aufgabenstellung eher auf einer Sicherheitsskala erfolgt. Eine grundlegende Überlegung ist dabei, dass die verschiedenen Bedeutungsaspekte der epistemischen Modalität bei der Interpretation stets präsent, aber unterschiedlich dominant sein könnten. Eine Einschätzung zur Sachverhaltswahrscheinlichkeit ist bei den meisten Ausdrücken vermutlich gut zu treffen. Bei den kritischen Ausdrücken spielt hingegen das negative Element eine besondere Rolle. Hier wäre es möglich, dass deswegen überwiegend die Falschheit der Proposition in den Vordergrund gestellt wird. Anzunehmen wäre, dass bei diesen sprachlichen Mitteln der Aspekt der Urteilssicherheit, der sich auf beide Wahrheitswerte beziehen kann, entsprechend dominanter ist. Die Sicherheit des Autors, dass die Proposition falsch ist, ist dann hoch. Dabei scheint es unerheblich, dass die Aufgabenstellung eigentlich die Bezeichnung *likelihood* verwendete, denn es ist durchaus möglich, auch den Aspekt der Sicherheit durch Wahrscheinlichkeitsformulierungen auszudrücken. Dies zeigt z. B. die Äußerung *Ich bin mir zu 80% sicher*.

Dass die abweichende epistemische Interpretation zum anderen auch von strukturellen Merkmalen abhängig sein könnte, legt das Antwortverhalten bei *cannot* nahe. Im Gegensatz zu *no evidence* und *unlikely*, bei denen nur von einem Viertel der Probanden abweichende Bewertungen vorliegen, sind hier in fast der Hälfte der Fälle die Wahrscheinlichkeitswerte hoch. Ein Grund für diese Differenz könnte die syntaktische Struktur sein. Bei den für die Befragung konstruierten Items sind sowohl *unlikely* als auch *no evidence* Teil des Matrixsatzes, während die dadurch modifizierte Proposition im Komplementsatz (in (85a) und (85b) unterstrichen) angegeben ist.

- (85) a. *It is **unlikely** that de novo induction of GATA-1 is required for globin gene activation following cell fusion.*
 b. *There is **no evidence** that TK activity is obligatory for CD14-mediated signal transduction to occur in response to LPS.*

Das Modalverb *cannot* hingegen ist Teil eines komplexen Prädikats. Während dieser Teil als epistemisch modale Angabe zu werten ist, gehört der andere Teil, der Infinitiv des Vollverbs, bereits zur Proposition (s. (86a)). Möchte man die obige Struktur erhalten, muss man *cannot* semantisch paraphrasieren (s. (86b)).

- (86) a. [...] *tyrosine kinase activity **cannot** be obligatory for lipopolysaccharide signal transduction.*
 b. *It is **impossible** that tyrosine kinase activity is obligatory for lipopolysaccharide signal transduction.*

Bei den ersten beiden Ausdrücken ist die epistemische Modalität also strukturell eindeutiger von der modifizierten Proposition zu trennen als beim dritten Ausdruck. Obwohl die Trennung von Modifizierer und Modifiziertem durch die Farben Rot und Blau im Onlinefragebogen verdeutlicht wurde, könnte es sein, dass diese Unterscheidung bei der Bewertung nicht immer klar war. Eine Folge davon wäre, dass das negative Element entsprechend noch stärker an die Proposition gebunden bzw. die Falschheit der Proposition noch betonter ist. In Anlehnung an bisherige Überlegungen sind dann hohe Wahrscheinlichkeitsbewertungen (d. h., dass der beschriebene biologische Zusammenhang nicht besteht) bzw. hohe Sicherheitsbewertungen (Der Autor ist sich zu 100% sicher, dass die Proposition falsch ist.) denkbar.

Die Ergebnisse zeigen damit einerseits, dass die Verwendung der epistemischen Skala bei den kritischen Ausdrücken nicht völlig eindeutig ist bzw. dass hier Abweichungen von der vorgegebenen Skalenkonzeption möglich sind. Bei der Klassifizierung und Annotation sollte diese Problematik beachtet werden. So sollte (z. B. beim Aufbau entsprechender Annotationsschemata) klar definiert sein, welcher Bedeutungsaspekt der epistemischen Modalität durch die Skala tatsächlich repräsentiert wird und wie epistemisch modale Ausdrücke syntaktisch in einen Satz eingebettet sein können bzw. wie eine Unterscheidung zwischen beschriebenem Sachverhalt und modifizierenden Elementen getroffen werden kann. Andererseits muss der Einfluss der Negation, deren spezielle Betrachtung in der vorliegenden Arbeit nicht erfolgt, näher untersucht und modelliert werden. Dies betrifft sowohl die Interpretation der kritischen Ausdrücke als auch die generelle Frage, wie bei der Annotation mit negierten Propositionen umgegangen wird. So impliziert *likely* in (87) z. B. eine hohe Wahrscheinlichkeit, dass die Proposition wahr ist. Diese Proposition umfasst aber zusätzlich das Negationselement. Der Zusammenhang zwischen den biologischen Prozessen wird also sowohl durch die epistemische Modalität als auch durch die Negation modifiziert. Beides muss für die Zwecke der Klassifizierung modelliert werden.

- (87) *It is **likely** that de novo induction of GATA-1 is **not** required for globin gene activation following cell fusion.*

Unter Beachtung der Tatsache, dass bei einem Großteil der Probanden die Bewertungen der kritischen Items erwartungsgemäß sind, kann schließlich aber auch festgehalten werden, dass es die hier vorgestellte Wahrscheinlichkeitsskala im Gegensatz zur Sicherheitsskala durchaus erlaubt, für die Klassifizierung und Annotation relevante Unterscheidungen (z. B. zwischen (88a) und (88b)) darzustellen.

- (88) a. *It is **unlikely** that de novo induction of GATA-1 is required for globin gene activation following cell fusion.*
- b. *It is **likely** that de novo induction of GATA-1 is required for globin gene activation following cell fusion.*

9.2 Fazit zu Bewerterunterschieden

Hypothese 1 bezieht sich auf die Frage, ob die Positionierung eines Ausdrucks auf der epistemischen Skala von bestimmten Charakteristika der bewertenden Person abhängt. Hintergrund ist die Beobachtung, dass Hedges in Wissenschaftstexten zum Großteil strategisch verwendet werden. Die Vertrautheit einer Person mit dieser Strategie könnte sich auf die Interpretation der Ausdrücke auswirken. So könnten Merkmale wie der Fachbereich, die Vertrautheit mit Wissenschaftstexten oder auch die englische Sprachkompetenz zu systematischen Unterschieden in der Bewertung führen, die möglicherweise ein Grund für die teilweise fehlende Übereinstimmung zwischen Annotatoren in den vorgestellten Klassifizierungsarbeiten sind. Die Ergebnisse zeigen, dass bei keinem der Ausdrücke ein derartiger Einfluss vorliegt. Für die epistemische Einschätzung scheint es also unerheblich, ob die bewertende Person im Bereich Biomedizin oder Linguistik tätig ist, ob sie noch studiert oder weitergehend forscht und daher entsprechende Erfahrungen im Umgang mit Wissenschaftstexten aufweist und ob sie ihre englische Sprachkompetenz als gut, sehr gut oder exzellent einstuft. Dies ist ein Hinweis darauf, dass zumindest diejenigen Differenzen zwischen Annotatoren, die bezüglich der vorgestellten Klassifizierungsprojekte auffällig sind, wahrscheinlich nicht zur Variabilität in den Bewertungen beitragen. Stattdessen ist davon auszugehen, dass die fehlende Übereinstimmung eher rein subjektiven Unterschieden zwischen den einzelnen Annotatoren geschuldet ist.

Eine Möglichkeit, die Ergebnisse zu deuten, besteht in der Annahme, dass die hier betrachteten Personenmerkmale eher keine relevanten Kontraste darstellen. Die fehlenden Unterschiede bezüglich der englischen Sprachkompetenz könnten z. B. daher rühren, dass die verglichenen Faktorstufen (gut, sehr gut, exzellent) relativ vergleichbar sind und gravierendere Kontraste (z. B. beim Vergleich mit grundlegender oder ausreichender Sprachkompetenz) aufgrund der Gegebenheiten der Studie

nicht betrachtet werden können. Da allerdings davon auszugehen ist, dass diejenigen Personen, die mit (englischen) biomedizinischen Wissenschaftstexten umgehen, über eine entsprechend hohe Sprachkompetenz verfügen, hätten solche Unterschiede wohl keine großartigen Konsequenzen für Annotation und Klassifizierung. Ein wichtiger Punkt, der im Rahmen dieser Arbeit allerdings nicht beleuchtet werden kann, ist der Einfluss verschiedener Muttersprachen auf den epistemischen Grad der Ausdrücke. Unter der Annahme, dass bei der Verwendung von Hedges ein Transfer sprachgemeinschaftlicher Normen erfolgen kann, wäre die Überprüfung eines solchen Transfers bei der Interpretation epistemisch modaler Ausdrücke durchaus interessant. Einen Hinweis darauf, dass sich zumindest deutsche und englische Muttersprachler in ihren Einschätzungen vermutlich nicht stark unterscheiden, liefern die Ergebnisse dadurch, dass die beiden Sprachen relativ eindeutig verschiedenen Kategorien der Sprachkompetenz zugeordnet werden können. So umfasst die Kategorie der exzellenten Sprachkompetenz alle englischen und kaum deutsche Muttersprachler. Ein fehlender Unterschied hinsichtlich der Sprachkompetenz könnte also (wenn auch nur tendenziell) auf eine solche Vergleichbarkeit hindeuten. Für genauere Aussagen wäre allerdings eine systematischere Untersuchung nötig, die darüber hinaus auch andere Sprachen einschließt.

Hinsichtlich der fehlenden Bewertungsunterschiede beim Fachbereich ist anzuführen, dass sich die Bereiche der Biomedizin und Linguistik aus wissenschaftstheoretischer Perspektive vielleicht weniger unterscheiden, als man annehmen könnte. Die Linguistik ist eine empirische Wissenschaft, bei der die Präsentation von (auch experimentellen) Daten eine große Rolle spielt. Die Art der Daten ist dabei vermutlich nicht wesentlich abstrakter oder unpräziser als auf dem Gebiet der Biologie. Nach den theoretischen Betrachtungen wäre die Verwendung (d. h. Art und Funktion) von Hedges entsprechend vergleichbar (s. Abschnitt 3.1). Diese Vergleichbarkeit könnte sich in der Interpretation der Ausdrücke zeigen. Um solche Überlegungen weiter zu verfolgen, könnte das Vorkommen von Hedges in biomedizinischen und linguistischen Texten formal wie auch pragmatisch näher beleuchtet werden.

Eine andere Möglichkeit der Erklärung der Ergebnisse besteht darin, den strategischen Hedge-Einsatz als eher auf die Textproduktion bezogen zu sehen. So könnte es sein, dass es beim Verfassen von Wissenschaftsartikeln zwar Unterschiede in der Hedge-Frequenz gibt, bestimmte Lexeme präferiert werden und die mit den sprachlichen Mitteln verbundenen Funktionen (z. B. Stärke der Festlegung des Autors oder Ausdruck von Höflichkeit) variieren, die Interpretation der Wörter aber von solchen Häufigkeits-, Terminologie- und Motivunterschieden relativ unabhängig ist. Einen Hinweis darauf liefert der Befund, dass die Einschätzungen bei unterschiedlicher Vertrautheit mit Wissenschaftstexten vergleichbar sind. Hyland (1998) geht hier bezüglich der Textproduktion von einer zunehmend formelhaften Verwendung von Hedges im Zusammenhang mit akademischer Erfahrung aus. Der fehlende Un-

terschied in der Interpretation bezüglich des formalen Bildungsabschlusses könnte dahingehend verstanden werden, dass eine entsprechend formelhafte Deutung bei der Textrezeption allerdings nicht der Fall ist.⁴⁰ Eine eventuelle Dissoziation zwischen Textproduktion und Textrezeption könnte bei der vorliegenden Untersuchung auch dadurch bedingt sein, dass die epistemische Skala hier den Bedeutungsaspekt der Wahrscheinlichkeit repräsentiert. Möglicherweise werden dadurch v. a. die Stärke der Autorfestlegung und weniger die rhetorischen Motive betont. Es wäre an dieser Stelle also insgesamt interessant, zu überprüfen, ob es tatsächlich entsprechende Unterschiede in der Interpretation epistemisch modaler Ausdrücke bei der aktiven Formulierung von Sätzen in Wissenschaftstexten und beim passiven Lesen entsprechender Äußerungen gibt. Solche Befunde wären hinsichtlich der Klassifizierung wissenschaftlicher Inhalte durchaus wichtig, denn es müsste womöglich eine Entscheidung darüber getroffen werden, welche Perspektive bei der Bewertung eingenommen wird bzw. woran sich die Positionierung auf der Skala orientiert.

Die vorgestellten Befunde haben Implikationen für das Vorgehen bei der Annotation epistemisch modaler Ausdrücke und der Klassifizierung wissenschaftlicher Inhalte im biomedizinischen Bereich. So scheint es zumindest für die hier untersuchten Personenmerkmale nicht unbedingt notwendig, bei der Auswahl der Annotatoren Einschränkungen zu treffen. Die Bewertung und Kategorisierung kann vermutlich sowohl von Personen aus der Biomedizin als auch aus der Linguistik, von Personen mit unterschiedlicher Sprachkompetenz (allerdings mit Beschränkung der Ergebnisse auf deutsche und englische Muttersprachler) und mit unterschiedlicher Erfahrung im Umgang mit wissenschaftlichen Texten ausreichend übereinstimmend vorgenommen werden. Zur sicherlich vorhandenen subjektiven Variabilität kommen hier also scheinbar keine systematischen Interpretationsunterschiede hinzu.

Beachtet werden sollten an dieser Stelle allerdings die tendenziellen Unterschiede in der Bewertung des Ausdrucks *can*. Probanden aus der Biomedizin hatten hier höhere Werte als Linguisten, Personen mit ausgezeichneter Sprachkompetenz höhere Werte als Personen mit guter oder sehr guter Sprachkompetenz vergeben. Obwohl diese Unterschiede sich nicht als signifikant herausstellen, sind sie gegenüber dem Antwortverhalten bei den restlichen Ausdrücken doch auffällig. Diese Auffälligkeit bei der Bewertung deckt sich damit, dass *can* bereits in der theoretischen Betrachtung als problematisch aufgefasst werden konnte (s. Abschnitt 2.1 und Abschnitt 3.2). So wurde mehrfach betont, dass das hinsichtlich der Modalitätsart ambigue Modalverb vorrangig als dynamisch (oder deontisch) modal zu interpretieren ist und nur selten (z. B. im negierten Kontext) eine epistemisch modale Funktion aufweist. Insbeson-

⁴⁰Natürlich ist auch nicht auszuschließen, dass die hier vorgenommene Operationalisierung des Konstrukts 'Vertrautheit mit Wissenschaftstexten' auf der Grundlage des formalen Bildungsabschlusses schlicht nicht angemessen sein könnte und die Variable entsprechend nicht erwartungsgemäß gemessen wird.

dere bei der Präsentation experimenteller Ergebnisse scheint dabei die Beobachtung von Perkins (1983) zu gelten, dass *can* weniger die Wahrheit der Proposition als das Vorliegen bestimmter Bedingungen betrifft. Ein Beispiel ist die Äußerung in (89).

- (89) *The p49 (100) DNA binding subunit, together with p65, can act in concert with Tat-I to stimulate the expression of HIV-CAT plasmid.*

Man kann annehmen, dass *can* hier weniger ein Indikator für Wahrscheinlichkeit als für Frequenz ist, wobei sich die Frequenz darauf bezieht, dass bestimmte Umstände zum Vorliegen des Sachverhalts führen, andere Umstände hingegen nicht. Auf Grundlage solcher Überlegungen ist es nicht abwegig, dass Personen aus der Biomedizin, die wahrscheinlich eher Erfahrungen im Laborbereich mitbringen, dem Ausdruck tendenziell höhere Wahrscheinlichkeitswerte zuweisen als Probanden aus der Linguistik. Die verschiedenen Modalitätsarten könnten bei der Interpretation unterschiedlich dominant sein. Die unterschiedliche Bewertungstendenz hinsichtlich der englischen Sprachkompetenz ist hingegen nicht ohne Weiteres zu erklären. Vielleicht ist hier weniger die Sprachkompetenz an sich als der generelle Unterschied zwischen deutscher und englischer Muttersprache bedeutsam, der durch diesen Faktor teilweise miterfasst wird. So könnte es auch hier Dominanzunterschiede in der Modalitätsart geben. Auf weitere Spekulationen soll an dieser Stelle allerdings verzichtet werden. Festgehalten werden soll, dass das Modalverb *can* in biomedizinischen Texten womöglich gar nicht als epistemisch modaler Ausdruck aufgefasst und bei der Annotation vielleicht auch eher ausgelassen werden sollte.

9.3 Festlegung eines Kategoriensystems

Die empirische Erhebung zeigt, dass die von den Probanden für die einzelnen Ausdrücke festgelegten Wahrscheinlichkeitsbereiche ungefähr eine Spanne von 3% (untere Grenze bei *cannot*) bis 91% (obere Grenze bei *strongly suggest*) und damit beinahe die gesamte Wahrscheinlichkeitsskala abdecken. Werden die Ausdrücke nach ihrem epistemischen Grad geordnet, wird deutlich, dass sich die Wahrscheinlichkeitsbereiche zweier auf der Skala benachbarter Ausdrücke stets überlappen. Es bietet sich daher auch nicht von vornherein eine eindeutige (und natürliche) Unterteilung der Wahrscheinlichkeitsskala im Sinne epistemischer Kategorien (wie das bei der Orientierung an der Grundterminologie der Sprache möglich ist) an. Stattdessen spiegeln die Bewertungen die Annahme von z. B. Nuyts (2001) wider, dass die epistemische Skala grundsätzlich eher als Kontinuum aufzufassen ist (s. Abschnitt 4.1). Tatsächlich scheint die Vielzahl an sprachlichen Mitteln (Hierunter fallen sowohl Einzellexeme als auch komplexere sprachliche Konstruktionen.) eine feine Differenzierung bei der Einschränkung der Festlegung zum Sachverhalt zu ermöglichen. Entsprechende Nuancen zeigen sich bei der Interpretation der Ausdrücke. Eine Gruppierung dieser Ausdrücke muss daher z. T. willkürlich erfolgen. Grundlage für eine

derartige Gruppierung sind in der vorliegenden Arbeit die Ergebnisse einer Clusteranalyse sowie die Befunde zur Ungenauigkeit der sprachlichen Mittel.

Hypothese 2 bezieht sich auf die Frage, ob es zwischen den Ausdrücken Unterschiede in der Ungenauigkeit gibt, d. h., ob die verschiedenen sprachlichen Mittel unterschiedlich große Bereiche auf der Wahrscheinlichkeitsskala einnehmen. Die Ergebnisse zeigen, dass sich die Ausdrücke *cannot*, *strongly suggest*, *indicate* und *likely* signifikant von einigen anderen Ausdrücken unterscheiden. Zwischen den restlichen sprachlichen Mitteln liegen keine signifikanten Unterschiede vor.

Dies zeigt einerseits, dass die von Nuyts (2001) angenommenen wortartabhängigen Unterschiede in der Spezifität der Referenz zur epistemischen Skala (s. Abschnitt 4.3) in der Untersuchung nicht nachgewiesen werden können. So ist nicht festzustellen, dass bspw. epistemische Adverbien wie *possibly* und *potentially* spezifischer, d. h. in der implizierten Wahrscheinlichkeit genauer sind als Modalverben wie *cannot*, *can*, *may* und *might*. Eine Tendenz, die sich bei den erhobenen Daten stattdessen abzeichnet, ist der Zusammenhang der Ungenauigkeit mit dem epistemischen Grad der Ausdrücke. Sprachliche Mittel, die eher extreme Wahrscheinlichkeitswerte anzeigen, sind mitunter auch präziser. So impliziert das negierte Modalverb *cannot* besonders geringe Wahrscheinlichkeitswerte, die epistemischen lexikalischen Verbkonstruktionen *strongly suggest* und *indicate* sowie das epistemische Adjektiv *likely* hingegen besonders hohe Wahrscheinlichkeitswerte. Diese vier Ausdrücke unterscheiden sich teilweise signifikant von einigen anderen Ausdrücken mit weniger extremen epistemischen Graden, indem sie weniger Ungenauigkeit aufweisen. Es scheint also, als ob bei starker epistemischer Modifizierung (in beide Richtungen auf der Wahrscheinlichkeitsskala) die Interpretation der Ausdrücke eindeutiger ist. Diese Ergebnisse decken sich mit Befunden aus der Psychologie. Hier sind meist hohe Streuungswerte für die Verteilung der von den Probanden angegebenen Wahrscheinlichkeitswerte auffällig, wenn die untersuchten Wörter eher eine mittlere Wahrscheinlichkeit implizieren. Bei sehr niedrigen und hohen Wahrscheinlichkeitswerten ist die Streuung hingegen gering. (vgl. Clark, 1990, S.205)

Andererseits zeigen die Ergebnisse an, dass die Ausdrücke unabhängig von dieser Tendenz in ihrer Ungenauigkeit relativ ähnlich sind. Der durchschnittliche Wahrscheinlichkeitsbereich, der auf Basis der Befragung festgelegt werden kann, beträgt über alle Ausdrücke hinweg ca. 25%. In der Fragestellung wurde die Ungenauigkeit als Kriterium für die Detailliertheit eines epistemischen Kategoriensystems definiert. Die grundlegende Annahme ist, dass die Anzahl der für Annotation und Klassifizierung vorgegebenen epistemischen Kategorien der Tatsache gerecht werden muss, dass die Ausdrücke hinsichtlich der implizierten Wahrscheinlichkeit relativ vage sind und in ihrer Interpretation variieren können. Entsprechend soll dieser Wert als Näherungswert für die durchschnittliche Kategoriengröße übernommen werden. Eine geschätzte Ungenauigkeit von 25% ist dann als Hinweis dafür zu sehen, dass eine

Wahrscheinlichkeitsskala von 0% bis 100% in maximal vier epistemische Kategorien unterteilt werden sollte, um der Interpretationsvariabilität gerecht zu werden.

In Hinblick auf die Ungenauigkeit gibt es noch eine weitere Auffälligkeit. So weisen *examine whether*, *investigate + hypothesis* und *unknown whether* gegenüber anderen sprachlichen Mitteln hohe Streuungswerte auf. Diese Ausdrücke haben die Gemeinsamkeit, dass sie vollständigen Mangel an Wissen anzeigen. *Unknown whether* ist dabei nach Hyland (1998) nur epistemisch modal im weiteren Sinne, da nach seiner Definition Angaben zu fehlendem Wissen zu den diskursbasierten, nicht-lexikalischen Hedges zählen (s. Abschnitt 3.2). Die anderen beiden Ausdrücke liegen ebenfalls außerhalb des Kernbereichs der epistemischen Modalität, da sie anzeigen, dass Hypothesen und damit Sachverhalte erst untersucht werden müssen. Sie weisen aber nicht auf eine Einschränkung der Festlegung an sich hin. Auch sind sie nicht unbedingt mit den bei Hyland (1998) aufgeführten metadiskursiven Verben (z. B. *sought to investigate*) zu vergleichen, da diese eher strategisch verwendet werden, um einen Kontrast zu den erhaltenen Ergebnissen darzustellen. Die fraglichen Ausdrücke sind also an sich bereits Grenzphänomene. Die hohen Streuungen lassen sich dadurch erklären, dass hier bei der Bewertung auf einer Wahrscheinlichkeitsskala verschiedene Vorgehensweisen möglich sind. So kann zum einen eine Wahrscheinlichkeit von ca. 50% angegeben werden, da es keinerlei Hinweise für das Zutreffen des beschriebenen Sachverhalts gibt. Zum anderen ist denkbar, dass die Ausdrücke einen Wahrscheinlichkeitsbereich von 0% bis 100% implizieren, da eine genaue Einschätzung zum Sachverhalt noch aussteht. Beide (aber nicht ausschließlich diese) Vorgehensweisen finden sich in den Datensätzen. Hier stellt sich die Frage, ob derartige Ausdrücke tatsächlich auf einer Wahrscheinlichkeitsskala, d. h. hinsichtlich einer epistemischen Einschränkung, bewertet werden sollten oder ob bei der Annotation eine separate Markierung sinnvoller ist. Thompson et al. (2011) gehen so vor (s. Abschnitt 5.2.4). Hier werden Angaben zu durchgeführten Untersuchungen unter der Kategorie ‘Investigation’ des Parameters ‘Knowledge Type’ aufgeführt, weil im Grunde keine Aussage über die Sicherheit des Autors getroffen werden kann. Da eine Kategorie wie ‘Investigation’ den Ausdruck *unknown whether* allerdings nicht einschließt, löst das die Problematik nicht vollständig. Für das weitere Vorgehen ist in der vorliegenden Arbeit deshalb wichtig, dass der Median der Ungenauigkeitswerte, auf den abweichende Bewertungen (wie die Angabe eines 100%-Wahrscheinlichkeitsbereichs) weniger einen Einfluss haben, zwischen den drei fraglichen und den restlichen Ausdrücken relativ vergleichbar ist. Da auch nach der theoretischen Skalenkonzeption von Nuyts (2001) Angaben zum Mangel an Wissen in der Mitte der Skala (d. h. bei ca. 50%) positioniert werden, ist davon auszugehen, dass ein Einbezug solcher sprachlichen Mittel mit entsprechender Wahrscheinlichkeitsbewertung möglich ist.

Die oben festgelegte maximale Kategorienanzahl kann als Kriterium für mögliche Gruppierungen der Ausdrücke bei der Clusteranalyse genutzt werden. So lässt

sich diejenige Stufe der Clusterbildung ausfindig machen, bei der unter Beachtung der Ungenauigkeit der sprachlichen Mittel eine angemessene Clusterzahl vorliegt. Die Clusteranalyse basiert auf den von den Probanden festgelegten Begrenzungen der Wahrscheinlichkeitsbereiche. Es können also solche Gruppierungen von Ausdrücken auf der Skala ausfindig gemacht werden, die bezüglich dieser Begrenzungen in sich möglichst ähnlich und untereinander möglichst verschieden sind. Die im Dendrogramm präsentierte Acht-Cluster-Lösung nach den ersten 19 Stufen der Agglomeration stellt hinsichtlich der Ungenauigkeit ein zu detailliertes epistemisches Kategoriensystem dar. Bei einer Vier-Cluster-Lösung liegt der Ausdruck *unknown whether* hingegen noch als singuläres Cluster vor. Oben wurde bereits auf die semantische Ähnlichkeit zwischen *unknown whether*, *examine whether* und *investigate + hypothesis* eingegangen, die alle einen Mangel an Wissen verdeutlichen. Es ist interessant, dass der Ausdruck *unknown whether* allerdings einen von den anderen beiden sprachlichen Mitteln abweichenden Wahrscheinlichkeitsbereich zu implizieren scheint und deshalb erst spät zu deren Cluster hinzugefügt wird. Der durch *unknown whether* ausgedrückte epistemische Grad ist dabei jeweils geringer. Eine Erklärung wäre, dass bei Angaben zu bevorstehenden Untersuchungen die zu prüfende Hypothese meist schon theoretisch begründet ist, es also oft bereits Hinweise zum Zutreffen eines Sachverhalts gibt. Die Ausdrücke *examine whether* und *investigate + hypothesis* könnten daher eine höhere Wahrscheinlichkeit als *unknown whether* implizieren, weil der Sachverhalt plausibler ist. Bei letzterem Ausdruck wäre davon auszugehen, dass die Wahrheit der Proposition noch nicht durch theoretische Vorüberlegungen gestützt wird. Trotz einer solchen Erklärung scheint die Vier-Cluster-Lösung aufgrund der semantischen Ähnlichkeit zwischen den drei Ausdrücken nicht unbedingt gerechtfertigt. In der vorliegenden Arbeit wird deshalb eine Drei-Cluster-Lösung angenommen. Hier sind alle Einzelausdrücke einem Cluster zugeordnet. Diese Cluster werden als epistemische Kategorien interpretiert.

Da die Clusteranalyse ein rein deskriptives Verfahren darstellt, dessen Ergebnisse sich vorrangig auf die erhobenen Daten beziehen, stellt sich die Frage, ob die festgelegten Kategorien auch inhaltlich beschrieben werden können. Hier kann auf die theoretischen Betrachtungen zurückgegriffen werden. Eine relativ eindeutige Gruppierung wird durch die kritischen Ausdrücke *cannot*, *no evidence* und *unlikely* gebildet. Diese implizieren eine geringe Wahrscheinlichkeit, dass die Proposition wahr ist, und liegen daher im unteren Bereich der Skala. Aus inhaltlicher Sicht ist diese erste Kategorie also gut beschreibbar. Eine weitere Gruppierung bilden Ausdrücke, die sich eher im oberen Bereich der Skala befinden und eine hohe Wahrscheinlichkeit implizieren, dass die Proposition wahr ist. Aus semantischer Perspektive umfasst diese Gruppierung v. a. Ausdrücke, die (in der konkreten Form ihres Vorkommens) eher induktiv evidential sind, d. h., die nach Chafe (1986) den Prozess der Schlussfolgerung auf der Grundlage direkter Evidenz anzeigen (s. Abschnitt 1.2).

Diese Evidenz – z. B. experimentelle Ergebnisse – ist dabei wie in (90) explizit.

- (90) *Pretreatment of PBMCs with pertussis toxin blocked the functions of C3a and C3a(desArg), **indicating** that the actions of these two molecules are mediated by a G protein-coupled pathway.*

Ausdrücke solcher Art sind *suggest, imply, support + hypothesis, indicate* und *strongly suggest*. Die Zugehörigkeit von *can* zum zweiten Cluster lässt sich durch den eher dynamischen Charakter dieses Modalverbs erklären. Zum modalen lexikalischen Verb *propose* gibt schließlich Hyland (1998) an, dass dieses in seiner Funktion v. a. performativ ist. (vgl. Hyland, 1998, S.120) Die hohe Bewertung wäre dann damit zu erklären, dass der Autor hier weniger eine Einschränkung der Wahrscheinlichkeit vornimmt als eher seine aktive Rolle bei der Präsentation des Sachverhalts betont. Insgesamt fällt bei der Betrachtung der so gebildeten Gruppierung auf, dass die zugehörigen Ausdrücke aus pragmatischer Sicht zu denjenigen sprachlichen Mitteln gehören, die in Wissenschaftstexten eher strategisch verwendet werden. Diese stellen nicht durchgängig den Wissensstatus des Autors ins Zentrum, sondern beziehen sich eher auf die akzeptable Präsentation eines Sachverhalts oder den Schutz des Autors vor Kritik. Die hohen Bewertungen wären dann damit zu erklären, dass eine Wahrscheinlichkeitsmodifizierung eher zweitrangig ist. Der einzige Ausdruck in der Gruppe, der tatsächlich direkt die Wahrscheinlichkeit kenntlich macht, ist das epistemische Adjektiv *likely*, dessen semantischer Gehalt bereits relativ hohe Werte impliziert. Ausgehend von dieser inhaltlichen Beschreibung der zweiten Gruppierung kann das dritte Cluster von Ausdrücken am besten dadurch charakterisiert werden, dass die sprachlichen Mittel hier tatsächlich dazu dienen, Mangel an Wissen auszudrücken. Zu den Ausdrücken gehören dabei einerseits das primäre Modalverb *may* sowie die sekundären Modalverben *might* und *could*. In Anlehnung an die Überlegungen von Perkins (1983) impliziert *might* dabei eine geringere Wahrscheinlichkeit als *may*, der Unterschied ist allerdings nicht signifikant. Weiterhin sind hier die epistemischen Adverbien *possibly* und *potentially* sowie die korrespondierenden epistemischen Adjektive *possible* und *potential* sowie das Adjektiv *putative* anzuführen. Die epistemischen lexikalischen Verben *seem* und *appear* lassen sich als sprachliche Mittel beschreiben, die nach Chafe (1986) den Prozess der Schlussfolgerung auf der Grundlage indirekter Evidenz anzeigen. Sie unterscheiden sich semantisch von den Verben aus der ersten Gruppierung dahingehend, dass die Evidenz nicht explizit angegeben ist und die Festlegung eher unspezifisch auf der perzeptiven Wahrnehmung gründet. Aus pragmatischer Sicht könnte angenommen werden, dass bei diesen Verben auch eher der Wissensstatus des Autors als die Art der Präsentation des Sachverhalts im Vordergrund steht. Ebenfalls zur zweiten Gruppierung gehören die Angaben zu Hypothesen und Untersuchungen (*investigate + hypothesis* und *examine whether*) sowie zu vollständigem Mangel an Wissen (*unknown whether*). Der etwas unklare Status dieser Ausdrücke wurde bereits be-

sprochen. Die übrigen sprachlichen Konstruktionen *has/have been suggested*, *suggest + may* und *think* sind vorerst als Sonderformen zu betrachten (*think* tritt in den Items in der Form *is thought to* auf), auf die weiter unten näher eingegangen wird.

Nach den Überlegungen zur Clusteranalyse scheint bei einer Unterteilung der Wahrscheinlichkeitsskala ein epistemisches Kategoriensystem aus drei Kategorien angemessen. Die Kategorien lassen sich hinsichtlich des Bedeutungsaspekts der Wahrscheinlichkeit beschreiben. Diese wird aber oft nur indirekt impliziert, da teilweise andere Aspekte (z. B. die eher strategische Verwendung, die Art der Evidenz) im Vordergrund stehen. Die Wahrscheinlichkeitsangaben können diese Aspekte zwar reflektieren, aber eben vermutlich nur indirekt. Für die Klassifizierung und Annotation wissenschaftlicher Inhalte in biomedizinischen Texten wäre nach diesen Betrachtungen das epistemische Kategoriensystem in Tabelle 17 vorzuschlagen.

geringe Wahrscheinlichkeit	mittlere Wahrscheinlichkeit	hohe Wahrscheinlichkeit
geringe Wahrscheinlichkeit, dass die Proposition wahr ist bzw. dass der Sachverhalt zutrifft; ausgedrückt durch z. B. <i>cannot</i> , <i>no evidence</i> , <i>unlikely</i>	mittlere Wahrscheinlichkeit, dass die Proposition wahr ist bzw. dass der Sachverhalt zutrifft; zeigt u. a. Mangel an Wissen aufseiten des Autors oder nur indirekte Evidenz an; ausgedrückt durch z. B. <i>unknown whether</i> , <i>may</i> , <i>possibly</i> , <i>seem</i> oder durch Angaben zu Hypothesen wie <i>examine whether</i>	hohe Wahrscheinlichkeit, dass die Proposition wahr ist bzw. dass der Sachverhalt zutrifft; zeigt u. a. explizit subjektive Sichtweise des Autors bei der Präsentation des Sachverhalts, direkte Evidenz oder hohe Wahrscheinlichkeit an; ausgedrückt durch z. B. <i>(these results) suggest</i> , <i>(we) propose</i> , <i>likely</i>

Tabelle 17: Vorgeschlagene diskrete epistemische Skala

Es bleibt schließlich die Frage zu klären, ob sich die hier festgelegten Kategoriengrenzen tatsächlich auch in den Bewertungen der Ausdrücke widerspiegeln. Dabei dient Hypothese 3a der Überprüfung, ob sich zwei Ausdrücke, die auf den verschiedenen Seiten einer Kategoriengrenze liegen, in ihrem epistemischen Grad unterscheiden. Die Post-Hoc-Vergleiche beim Testen der allgemeinen Hypothese 3 liefern die entsprechenden Werte. An der Kategoriengrenze zwischen ‘geringer Wahrscheinlichkeit’ und ‘mittlerer Wahrscheinlichkeit’ liegen die Ausdrücke *unlikely* und *unknown whether*. Zwischen diesen gibt es einen signifikanten Unterschied im epistemischen Grad. Die Annahme, dass die Ausdrücke verschiedenen Kategorien zugeordnet werden können, wird daher durch ihre unterschiedliche Position auf der Skala gestützt. An der Kategoriengrenze zwischen ‘mittlerer Wahrscheinlichkeit’ und ‘hoher Wahrscheinlichkeit’ liegen die Ausdrücke *appear* und *suggest*. Zwischen diesen

gibt es keinen signifikanten Unterschied im epistemischen Grad. Die Zugehörigkeit zu verschiedenen Kategorien kann also nicht durch unterschiedliche Positionen auf der Skala gerechtfertigt werden. Dies weist erneut darauf hin, dass die epistemische Skala eher als Kontinuum zu verstehen ist. Eine Kategorisierung muss teilweise willkürlich bzw. auf der Basis inhaltlicher und semantischer Überlegungen erfolgen.

Ein letzter interessanter Befund ist, dass die hier vorgestellte Kategorisierung auf Grundlage von Probandenurteilen eine z. T. von bisherigen Annotationsschemata abweichende Einordnung einzelner Ausdrücke mit sich bringt. So impliziert *propose* z. B. – ähnlich *suggest* und *indicate* – eine hohe Wahrscheinlichkeit des beschriebenen Sachverhalts. Bei Thompson et al. (2011) hingegen ist *propose* als Ausdruck definiert, der starke Spekulation anzeigt. *Appear* hingegen ist in der vorliegenden Kategorisierung (in Übereinstimmung mit dem evidentialen Status) spekulativer als z. B. *suggest* oder *indicate* zu werten, mit denen es bei Thompson et al. semantisch gleichgesetzt wird. Dies zeigt, dass eine empirische Untersuchung im Sinne einer Konsensfindung zu anderen (und vermutlich repräsentativeren) Festlegungen führen kann als solche, die das Ergebnis subjektiver Einzelurteile sind.

9.4 Fazit zum sprachlichen Kontext

Die Überprüfung der Hypothesen 3b, 3c und 3d zeigt stichprobenartig, dass der sprachliche Kontext, in den ein epistemisch modaler Ausdruck eingebettet ist, einen Einfluss auf dessen Interpretation haben kann. So liegt bei allen ausgewählten Kontrasten ein signifikanter Unterschied im epistemischen Grad vor.

Bezüglich der sprachlichen Gesamtkonstruktion wurden die Lexeme *suggest* und *hypothesis* betrachtet. Bezieht sich das Verb *suggest* wie in (91a) auf die explizite Darstellung von Erkenntnissen und Fakten, auf deren Grundlage Schlussfolgerungen gezogen werden, wird eine höhere Wahrscheinlichkeit des beschriebenen Sachverhalts impliziert, als beim Auftreten des Verbs in der Passivkonstruktion (s. (91b)).

- (91) a. *E3330 decreased activated NF-kappa B in nuclei, **suggesting** that E3330 inhibits NF-kappa B activation and/or translocation of the nuclei.*
- b. *Antioxidant molecules **have been suggested** to be of therapeutic value in the treatment of HIV-infected patients.*

Ein solcher Unterschied lässt sich durch den evidentialen Status erklären. Nach den Ausführungen von Chafe (1986) verweist die Konstruktion in (91a) auf den auf direkter Evidenz basierenden Prozess der Induktion, bei (91b) steht der Prozess des Hörensagens im Vordergrund. Der Sachverhalt in der ersten Äußerung wird wahrscheinlicher eingeschätzt, weil es genauere Hinweise dafür gibt. Die Evidenzgrundlage in der zweiten Äußerung wird hingegen nicht genau spezifiziert, da die Quelle der Information nicht angegeben ist. Hyland (1998) weist darauf hin, dass sich der Autor durch Verwendung des Passivs von einer Aussage distanzieren

kann. Diese Distanzierung in Zusammenhang mit nicht explizit angegebener Evidenz scheint sich also in einer geringen Wahrscheinlichkeitseinschätzung zu äußern. Tatsächlich können die beiden Ausdrücke nach den vorherigen Überlegungen sogar unterschiedlichen epistemischen Kategorien zugewiesen werden. Während die Unterschiede bei *suggest* v. a. die konkrete Wortform betreffen, ist in den ausgewählten Beispielen beim Nomen *hypothesis* insbesondere relevant, bei welchem Verb es als Argument auftritt. So ist die Konstruktion *investigate + hypothesis* (s. (92a)) semantisch mit *examine whether* gleichzusetzen, die nicht epistemisch modal im engeren Sinne ist. Dieser Konstruktion wird ein geringerer Wahrscheinlichkeitswert zugewiesen als dem Ausdruck *support + hypothesis*, der erneut einen stark evidentialen Charakter hat, da die Evidenzgrundlage spezifiziert ist (s. (92b)).

- (92) a. We **investigate the hypothesis** that anti-LPS Abs neutralize endotoxin by blocking cellular uptake through mCD14.
 b. Vpr can regulate NF-kappa B, which **supports the hypothesis** that some aspects of viral pathogenesis are the consequence of cell dysregulation by Vpr.

Auch diese kontrastierenden Ausdrücke können unterschiedlichen epistemischen Kategorien zugewiesen werden.

Bezüglich der Kombination von Ausdrücken wurden die Lexeme *suggest* und *may* untersucht. Bei dieser häufigen Kombination von lexikalischen Verben und Modalverben ist zuerst auffällig, dass man im Gegensatz zur Annahme von Hyland (1998) hier nicht unbedingt von einer harmonischen Kombination ausgehen kann (s. Abschnitt 3.3). Tatsächlich unterscheiden sich die beiden Ausdrücke signifikant im epistemischen Grad und werden unterschiedlichen Kategorien zugeordnet. Der Unterschied lässt sich auch hier wieder auf Grundlage des evidentialen Status erklären, der bei *suggest* (s. (91a)) spezifischer ist als bei *may* (s. (93a)). Interessant ist, dass die Kombination *suggest + may* (s. (93b)) in der Untersuchung auf der epistemischen Skala ziemlich genau zwischen den beiden Einzelexemen positioniert wird. Dies könnte ein Hinweis darauf sein, dass bei Kombination die Interpretation tatsächlich auch von beiden Ausdrücken abhängt. Zwar liegt explizite Evidenz vor, der Autor schränkt seine Festlegung aber noch zusätzlich ein.

- (93) a. The GM-CSF locus **may** represent one of several target genes that are dysregulated in acute myeloid leukemia.
 b. Increased expression of wild-type NFAT1 substantially increases IL-4 promoter activity in unprimed CD4 T cells, **suggesting** that NFAT1 **may** be limiting for IL-4 gene expression in this cell type.

Bezüglich der Modifizierung wurde schließlich die Kombination von *suggest* mit dem intensivierenden Adverb *strongly* untersucht (s. (94)).

- (94) Dexamethasone has no effect on TPA-induced activation of PKC, **strongly suggesting** that this glucocorticoid inhibits signals downstream or parallel to this enzyme.

Obwohl sich der evidentielle Status des Verbs nicht durch die Modifizierung ändert, wird die Festlegung des Autors bezüglich der Wahrheit der Proposition doch wesentlich stärker eingeschätzt und die durch den Ausdruck implizierte Wahrscheinlichkeit entsprechend hoch bewertet. Hinsichtlich der hier vorgeschlagenen Kategorisierung liegt allerdings kein Unterschied vor.

Die obigen Ausführungen sind ein Hinweis dafür, dass sich die Klassifizierung wissenschaftlicher Inhalte auf der Grundlage explizit einschränkender sprachlicher Mittel nicht nur am Vorkommen von Einzelexemen orientieren sollte. Mitunter führt der sprachliche Kontext zu unterschiedlichen Einschätzungen, wobei die Differenzen (z. B. bei *investigate + hypothesis* und *support + hypothesis*) mitunter relativ groß sind, sodass die Zuweisung zu unterschiedlichen epistemischen Kategorien vorgenommen werden muss. Darüber hinaus gibt es Implikationen für das Vorgehen bei der Annotation. Die bspw. im BioScope-Korpus vorgenommene Min-Max-Strategie (s. Abschnitt 5.2.2) wäre für die Kennzeichnung von Ausdrücken im Zusammenhang mit ihrer epistemischen Kategorisierung nicht unbedingt angebracht, da der minimale Ausdruck, der Spekulation anzeigt, mitunter keine ausreichenden Hinweise darauf gibt, wie der betreffende Sachverhalt zu klassifizieren ist.

Beim sprachlichen Kontext wird weiterhin deutlich, dass einige Ausdrücke nicht durchgängig eine epistemisch modale Funktion erfüllen. So zeigt das Adjektiv *unknown* z. B. nur in Kombination mit der Konjunktion *whether* eine Einschränkung hinsichtlich der Proposition an. Ein Bezug zum Wahrheitswert liegt nur dann vor, wenn der Sachverhalt wie in (95a) als Subjektsatz (im Beispiel unterstrichen) realisiert ist, auf den sich *unknown* prädikativ bezieht. Ist das Subjekt wie in (95b) hingegen kein entsprechend eingeleiteter Nebensatz, wird durch das prädikative Adjektiv auch keine Proposition modifiziert.

- (95) a. *In contrast to the type I receptor, the type II IL-1R has a small cytoplasmic tail, and it is **unknown whether this receptor is a signal-transducing molecule.***
b. *The mechanisms by which GATA-1 controls cell survival are **unknown.***

Tatsächlich ist die Äußerung in (95b) überhaupt nicht spekulativ. Hyland (1998) fasst zwar beide Fälle als diskurs-basierte sprachliche Mittel zusammen. Es scheint aber durchaus einen Unterschied zu machen, ob sich mangelndes Wissen auf das Zutreffen eines Sachverhalts oder auf ein einzelnes Konzept (z. B. einen Mechanismus, der den beschriebenen Prozess ermöglicht) bezieht. Dass im zweiten Fall entsprechend keine Bewertung auf der epistemischen Skala vorgenommen werden kann, zeigen das unsichere Antwortverhalten (Antworthäufungen bei sowohl 0%, 50% als auch 100%) und die Kommentare der Probanden bezüglich solcher Items. Bei der Klassifizierung und Annotation sollten solche syntaktischen Konstruktionen daher ausgeschlossen und Adjektive wie *unknown* und *unclear* nicht generell und uneingeschränkt als epistemisch modal gewertet werden.

10 Schlussbetrachtung

Das Ziel der vorliegenden Masterarbeit war es, die Möglichkeiten der Skalierung epistemisch modaler Ausdrücke hinsichtlich der Klassifizierung wissenschaftlicher Inhalte in biomedizinischen Forschungsartikeln auszuloten. Mit epistemisch modalen Ausdrücken kann ein Autor kenntlich machen, dass für eine Aussage kein uneingeschränkter Geltungsanspruch erhoben wird. Als Ergebnis sollte eine epistemische Skala konstruiert werden, die für die Annotation von Wissenschaftstexten angemessen ist. Die Annotation stellt die Grundlage für die domänenspezifische Adaptation computerlinguistischer Informationssysteme dar.

Im ersten Teil, welcher theoretisch ausgerichtet war, wurde herausgearbeitet, dass aus linguistischer Perspektive verschiedene Konzepte (z. B. Subjektivität, Faktualität) dabei eine Rolle spielen. Besonders wichtig ist hier die Evidentialität, die die Art der Informationsquelle für eine Aussage anzeigt und die nicht klar von der epistemischen Modalität zu trennen ist. Ein Überblick zu den relevanten sprachlichen Mitteln zeigte weiterhin, dass die vielfältigen Ausdrucksmöglichkeiten sehr differenzierte Einschätzungen (wahrscheinlich auf einem Kontinuum) zum Zutreffen eines Sachverhalts ermöglichen. Gleichzeitig scheinen einige Ausdrücke in ihrer Bedeutung aber auch wenig präzise zu sein. Als nächstes wurden die Besonderheiten beim Gebrauch der Ausdrücke in Wissenschaftstexten betont. Den zentralen Untersuchungsgegenstand bilden hier die sogenannten Hedges, die zwar epistemisch modale Ausdrücke einschließen, aber in Hinblick auf die Realisierungsmöglichkeiten noch weiter gefasst sind. Neben der reinen Einschränkungsfunktion ist bei Hedges die strategische Verwendung zentral, mit der eine Akzeptanz von Erkenntnissen in der Wissenschaftsgemeinschaft erreicht werden soll. Die Betrachtung theoretischer und praktischer Skalierungsansätze führte schließlich zur Schlussfolgerung, dass von den verschiedenen Bedeutungsaspekten der epistemischen Modalität v. a. die Sachverhaltswahrscheinlichkeit für die Skalenkonzeption sinnvoll ist. Im Gegensatz zur Urteilssicherheit des Autors, die auf beide Wahrheitswerte bezogen werden kann, betrifft die Sachverhaltswahrscheinlichkeit nach der Definition nur die Wahrheit der Proposition, welche durch epistemisch modale Ausdrücke modifiziert wird. Damit wird ein Informationsverlust verhindert. In den vorhandenen Annotationsarbeiten scheinen solche unterschiedlichen Bedeutungsaspekte nicht sauber genug getrennt und definiert zu sein. Außerdem ist bei einigen Projekten eine geringe Bewerterübereinstimmung auffällig, die (neben den vermutlich zu detaillierten Kategoriensystemen) möglicherweise daher rührt, dass sich die Annotatoren hinsichtlich bestimmter Charakteristika systematisch unterscheiden.

Die empirische Untersuchung im zweiten Teil hatte zum Ziel, die häufigsten Ausdrücke, die in biomedizinischen Texten vorkommen, auf der so konzipierten

Wahrscheinlichkeitsskala mittels einer Probandenbefragung bewerten zu lassen. Dabei sollten auch einige auffällige Punkte aus dem theoretischen Teil Beachtung finden. Die Studie lieferte mehrere Erkenntnisse. Erstens scheint die Interpretation einzelner Ausdrücke (derjenigen, bei denen die Negation eine Rolle spielt) nicht immer eindeutig zu sein, was teilweise zu Bewertungen führen kann, die von der Skalenkonzeption abweichen, und impliziert, dass für die Annotation entsprechend klare Instruktionen (z. B. bezüglich Semantik oder syntaktischer Struktur) notwendig sind. Zweitens scheinen die näher betrachteten Charakteristika der bewertenden Personen (z. B. Fachbereich, Vertrautheit mit Wissenschaftstexten, englische Sprachkompetenz), die bei den vorhandenen Annotationsprojekten auffällig waren, zumindest im Rahmen der hier untersuchten Ausprägungen keinen Einfluss auf die Bewertung zu haben. Für die Annotation könnte es demnach unerheblich sein, ob die Annotatoren einen biologischen oder linguistischen Hintergrund haben, wieviel Erfahrung sie im Umgang mit Wissenschaftstexten haben und ob sie eine gute, sehr gute oder ausgezeichnete englische Sprachkompetenz aufweisen. Schließlich wurde die Skaleneinteilung im Sinne eines epistemischen Kategoriensystems angestrebt. Aufgrund der von den Probanden geschätzten Ungenauigkeit der Ausdrücke hinsichtlich der implizierten Wahrscheinlichkeit bietet sich eine Dreiteilung an. Die Kategorien lassen sich mit ‘geringe Wahrscheinlichkeit’, ‘mittlere Wahrscheinlichkeit’ und ‘hohe Wahrscheinlichkeit’ betiteln. Dabei ist zu beachten, dass die Ausdrücke diese Wahrscheinlichkeit mitunter nur indirekt implizieren. Insbesondere sprachliche Mittel, die nach Angabe der Probanden eine hohe Wahrscheinlichkeit anzeigen, lassen sich eher als strategische Verwendungen beschreiben, mit denen der Autor eine akzeptable Präsentation oder den Schutz vor Kritik anstrebt. Bei solchen Vorkommen ist oft die Evidenz als Schlussfolgerungsgrundlage explizit angegeben. Darüber hinaus muss betont werden, dass die hier vorgenommen Kategorisierung nicht unbedingt eindeutig durch die sprachlichen Gegebenheiten festgelegt werden kann. Stattdessen muss eine solche Kategorisierung in Anlehnung an die Annahme eines Kontinuums immer zu einem gewissen Teil willkürlich erfolgen. Die Ergebnisse der Studie verdeutlichen weiterhin, dass häufig nicht allein das Vorkommen von Einzelexemen, sondern auch die konkrete Wortform oder der sprachliche Kontext über die Zugehörigkeit zu einer Kategorie entscheiden. Auch sind einige Ausdrücke dahingehend ambig, dass sie in bestimmten syntaktischen Konstruktionen keine epistemisch modale Funktion erfüllen. Bei der Annotation und Klassifizierung wäre dies zu beachten.

Die hier vorgestellte Skalierung kann als erste Annäherung an eine sowohl theoretisch als auch empirisch fundierte Vorgabe für die Annotation als Grundlage von Informationsextraktionsverfahren im Bereich der Biomedizin verstanden werden. Auffällig ist, dass bei dieser empirischen Betrachtung teilweise andere Kategorienzuweisungen als in den bisher vorhandenen Annotationsvorgaben vorgenommen wer-

den konnten. Um die Validität zu überprüfen wäre es nun sinnvoll, ein Maß für die Bewerterübereinstimmung bei der Verwendung einer solchen Skala, die sich von den bisherigen Kategoriensystemen unterscheidet, zu erhalten. Gerade deshalb, weil sich die Konstruktion hier auf eine geringe Auswahl der möglichen Ausdrucksmittel bezieht, allerdings noch unzählige weitere Lexeme, Wortformen und sprachliche Kontexte zur epistemischen Modalisierung beitragen, wäre die Anwendbarkeit zu testen.

Ein wichtiger Punkt, der in der vorliegenden Arbeit nicht betrachtet werden konnte, ist die Negation. Diese muss im Rahmen eines umfangreichen Annotationschemas ebenfalls modelliert werden, denn neben der Frage, wie wahrscheinlich ein Sachverhalt ist, muss auch festgestellt werden, ob dieser Sachverhalt vielleicht darin besteht, dass ein biologischer Prozess verneint wird, d. h. gar nicht vorliegt. Neben der epistemischen Modalität ist also auch das Phänomen der Negation unbedingt näher zu untersuchen. Nur so können die relevanten Informationen aus Wissenschaftstexten korrekt extrahiert werden. In Einklang mit bisherigen Annotationsprojekten scheint dabei eine gemeinsame Betrachtung beider Phänomene im Rahmen eines Annotationschemas sinnvoll.

A Anhang

A.1 Liste der Befragungsitems

Tabelle 18: Auflistung der konstruierten Items (3 Sätze pro Ausdruck)

Ausdruck	Befragungsitem
<i>suggest</i>	Induction of fra-1 following engagement of CD40 did not require protein synthesis, suggesting that the PKC-dependent linkage between CD40 and fra-1 is direct.
	E3330 decreased activated NF-kappa B in nuclei, suggesting that E3330 inhibits NF-kappa B activation and/or translocation of the nuclei.
	IFN-alpha sensitized T cells to IL-2-induced proliferation, suggesting that IFN-alpha is involved in the regulation of T-cell mitogenesis.
<i>may</i>	The GM-CSF locus may represent one of several target genes that are dysregulated in acute myeloid leukemia.
	E3330 may suppress the production of active oxygen species serving as common messengers to activate NF-kappa B.
	Bcl-3 may aid activation of select NF-kappa B-regulated genes, including those of the human immunodeficiency virus.
<i>can</i>	The p49 (100) DNA binding subunit, together with p65, can act in concert with Tat-I to stimulate the expression of HIV-CAT plasmid.
	T cells representing different activation and/or differentiation stages can be differentially responsive to ROI-mediated signals.
	Antisense oligomers to p65 can be used to define the role of NF-kappa B in the activation pathways of neutrophils.
<i>indicate</i>	R24-stimulated increases in proliferation, cytotoxicity, and cell surface protein expression could be blocked by cyclosporin and staurosporin, indicating that cyclophilin/calcineurin and protein kinase C is involved in the R24 signaling pathway.
	Pretreatment of PBMCs with pertussis toxin blocked the functions of C3a and C3a(desArg), indicating that the actions of these two molecules are mediated by a G protein-coupled pathway.
	The addition of GM-CSF to the M-TAT/EPO cell culture decreased the amount of hemoglobin, even in the presence of EPO, indicating that the EPO signal for erythroid differentiation is suppressed by GM-CSF.
<i>appear</i>	Hemin induction appears to nonspecifically increase the expression of transiently transfected genes in K562 cells.
	NFAT1 appears to be the major NFAT family member responsible for the initial increased expression of IL-4 by primed CD4 T cells.
	The CMAT protein appears to be a novel inducible protein involved in the regulation of c-myb expression during T-cell activation.
<i>might</i>	Platelet-induced activation of the NF-kappaB system might contribute to early inflammatory events in atherogenesis.
	The immunosuppressive effect of oxLDLs might operate via a dysregulation of the T-lymphocyte activation mechanisms.
	C-Fos/c-Jun heterodimers might contribute to the repression of DRA gene expression.

Ausdruck	Befragungssitem
<i>could</i>	In monocytic cell lineage, HIV-1 could mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.
	NF-kappa B activation could constitute one of the mechanisms whereby the expression of kappa B-responsive genes is enhanced in phagocytosing neutrophils.
	Secretion of PGE2 by macrophages in response to infection or inflammatory activators could induce signaling events resulting in activation of proviral DNA present into T cells latently infected with HIV-1.
<i>likely</i>	HB24 is likely to have an important role in lymphocytes as well as in certain developing tissues.
	Tap is likely to be an important cellular mediator of Tip function in T cell transformation by herpesvirus saimiri.
	Egr-2 is likely to play a direct role in aberrant fasL up-regulation in lpr/lpr and gld/gld CD4(-)CD8(-) T cells.
<i>possible</i>	We report a possible role for G(Anh)MTetra in the release of cytokines during sepsis.
	We report the possible role of nuclear factor-kappa B activity in germline C epsilon transcription in a human Burkitt lymphoma B cell line.
	We report a possible signaling role for the Syk tyrosine kinase.
<i>potential</i>	The human G0/G1 switch gene, G0S2, has potential NFAT-binding sites in the 5' flank.
	Fos plays a potential role in the regulation of B cell proliferation and differentiation.
	Sequence analysis of the p40 promoter region identified two potential nuclear factor (NF)-kappaB binding sites conserved between mouse and human.
<i>putative</i>	Here we summarize the importance of the IgH 3'LCR and its putative functional role in IgH gene expression.
	Expression of PILOT, a putative transcription factor, requires two signals and is cyclosporin A sensitive in T cells.
	A putative function of BHRF1 in EBV-infected epithelial cells is to protect virus-infected cells from TNF- and/or anti-Fas- induced cell death.
<i>propose</i>	We propose that direct inhibition of K+ channels in T cells by progesterone contributes to progesterone-induced immunosuppression.
	We propose that the MNP protein(s) binding at the MNP site constitutes a novel transcription factor(s) expressed in hematopoietic cells.
	We propose that NFKB2 synthesis and processing allows continuous nuclear expression of an otherwise cytoplasmic protein.
<i>think</i>	Human T-cell leukemia virus type 1 (HTLV-1) Tax is thought to play a pivotal role in immortalization of T cells.
	Activation of NF-kappaB is thought to be required for cytokine release from LPS-responsive cells, a critical step for endotoxic effects.
	Bioactivation of sulphamethoxazole (SMX) to chemically-reactive metabolites and subsequent protein conjugation is thought to be involved in SMX hypersensitivity.
<i>seem</i>	A unique transcriptional mechanism distinct from those regulating the IL-2 or IL-4 genes seems to control the IL-5 gene.
	Enhanced IL-5 production by helper T cells seems to cause the eosinophilic inflammation of both atopic and nonatopic asthma.
	Tumor proliferative activity as evaluated by the monoclonal antibody Ki-67 seems to be an effective indicator of prognosis in breast cancer for DFS and OS.

Ausdruck	Befragungssitem
<i>unknown whether</i>	Whether virus-induced NF-kappa B activation is a mechanism that favors continuous viral replication in macrophages is unknown .
	In contrast to the type I receptor, the type II IL-1R has a small cytoplasmic tail, and it is unknown whether this receptor is a signal-transducing molecule.
	Several studies have established that Tat releases an elongation block to the transcription of HIV long terminal repeat (LTR); however, it is unknown whether this mechanism alone is sufficient to explain the block to HIV replication in human T cells when Tat is absent.
<i>possibly</i>	Xenogeneic human serum promotes leukocyte adhesion to porcine endothelium under flow conditions, possibly through the activation of the transcription factor NF-kappa B.
	Switching is hampered by CD30 coengagement, possibly through interference with the CD40-mediated NF-kappaB-dependent transcriptional activation of downstream C(H) genes.
	The ras oncogene blocks terminal differentiation to plasma cells, possibly by inhibiting the activity of lymphocyte-specific transcription factors.
<i>imply</i>	TCF-1 specifically recognizes the T beta 5 element of the TCR beta enhancer and the T delta 7 element of the TCR delta enhancer, implying that TCF-1 is involved in the control of several T cell-specific genes.
	The CD4 silencer is also active in CD4+ CD8int/- cells of the thymus, implying that an anti-silencer is required to resume CD4 expression in this cell population.
	Coexpression of a reciprocally mutant Myc protein capable of forming functional heterodimers with the mutant Max can compensate for the dominant negative activity and restore activation-induced apoptosis, implying that Myc promotes activation-induced apoptosis by obligatory heterodimerization with Max.
<i>potentially</i>	The monovalent gold compound AuTG is a potentially useful drug for the treatment of patients infected with HIV.
	Tyloxapol is potentially useful as a new antiinflammatory therapy for CF lung disease.
	Any event that enables T cells to become less responsive to IFN-gamma potentially alters immune responsiveness to Ag.
<i>no evidence</i>	There is no evidence that the molecular mechanism of HIV inhibition by ascorbate is mediated via NF-kappa B inhibition.
	There is no evidence that TK activity is obligatory for CD14-mediated signal transduction to occur in response to LPS.
	There is no evidence that the NFkB family members Rel-A, Rel-B and c-Rel are required for the transcription of IL-2 in the peripheral T lymphocytes of patients with breast cancer.
<i>unlikely</i>	It is unlikely that the NF-IL6 gene has an exclusive role as a combinatorial differentiation switch during myelopoiesis in human cells.
	It is unlikely that the limited activation of NF-kappa B by HIV protease has a significant effect on virus expression or T cell function.
	It is unlikely that de novo induction of GATA-1 is required for globin gene activation following cell fusion.

Ausdruck	Befragungssitem
<i>cannot</i>	While lipopolysaccharide-induced expression of inflammatory mediators requires tyrosine kinase activity, tyrosine kinase activity cannot be obligatory for lipopolysaccharide signal transduction.
	Activation of NF-kappa B plays an important role in the regulation of many proinflammatory cytokine genes, but cannot be the only mechanism, since several cytokine genes lack respective binding sites in their promoter regions.
	While ICN appeared primarily in the nucleus, DeltaE localized to cytoplasmic and nuclear membranes. Therefore, intranuclear localization cannot be essential for oncogenesis or transcriptional activation.
<i>has/have been suggested</i>	The lymphoid-specific Oct2B protein has been suggested to be involved in the function of the octamer motif in the context of the immunoglobulin heavy-chain (IgH) enhancer.
	Antioxidant molecules have been suggested to be of therapeutic value in the treatment of HIV-infected patients.
	Nuclear factor kappa B (NF-kappa B; heterodimer of p50 and p65) proteins have been suggested to play an important role in gene transcription of inflammatory mediators when monocytes are stimulated with lipopolysaccharide.
<i>investigate + hypothesis</i>	We investigate the hypothesis that low shear-induced activation of NF-kappaB is responsible for enhanced expression of vascular cell adhesion molecule (VCAM-1) resulting in augmented endothelial cell-monocyte (EC-Mn) adhesion.
	We investigate the hypothesis that cellular activation events occurring in T lymphocytes and monocytes and mediated through translocation of the transcription factor NF-kappa B are dependent upon the constitutive redox status of these cells.
	We investigate the hypothesis that anti-LPS Abs neutralize endotoxin by blocking cellular uptake through mCD14.
<i>support + hypothesis</i>	Specific staining as well as electron microscopic examinations revealed the accumulation of metal gold in the cells, which supports the hypothesis that gold ions block NF-kappaB-DNA binding by a redox mechanism.
	Vpr can regulate NF-kappa B, which supports the hypothesis that some aspects of viral pathogenesis are the consequence of cell dysregulation by Vpr.
	The tight correlation of HIV-1 transcriptional induction to T cell activation supports the hypothesis that the nuclear factor of activated T cells (NF-AT) influences HIV-1 replication.
<i>suggest + may</i>	All NK clones and cell lines studied were found to express hGATA-3-specific mRNA, suggesting that hGATA-3 may be involved in the regulation of the unrearranged TcR delta gene expression in NK cells.
	Increased expression of wild-type NFAT1 substantially increases IL-4 promoter activity in unprimed CD4 T cells, suggesting that NFAT1 may be limiting for IL-4 gene expression in this cell type.
	The sequence between -1.7 and -0.2 kb linked to a heterologous thymidine kinase promoter failed to respond to PRL stimulation, suggesting that the activity of upstream PRL response elements may require an interaction with promoter-proximal elements.

Ausdruck	Befragungsitem
<i>strongly suggest</i>	Even when LMP1 is unable to activate NF-kappaB, it is still capable of inducing certain characteristics of activated B cells, strongly suggesting that LMP1 also activates cells independently of NF-kappaB.
	The inhibitory effect of insulin on the GR binding power is both dose- and time-dependent, strongly suggesting that GR is tonically controlled by insulin concentration change under physiological conditions.
	Dexamethasone has no effect on TPA-induced activation of PKC, strongly suggesting that this glucocorticoid inhibits signals downstream or parallel to this enzyme.
<i>examine whether</i>	We examined whether NF(P), a nuclear factor specific for the P sequence, is related to NF-kappa B and NF-AT.
	We examined whether signal transducing components in T cells activate transcription of the GM-CSF gene.
	We examined whether BSAP plays a role in the transcriptional regulation of the epsilon germline promoter in human B cells.
<i>unknown (ohne Wahrheitswert)</i>	The mechanism by which early lymphoid cells are selectively transformed by v-Abl is unknown.
	The mechanisms by which GATA-1 controls cell survival are unknown.
	The molecular mechanism by which ATRA increases the nuclear abundance of a functional VDR is unknown.
<i>unclear (ohne Wahrheitswert)</i>	The mechanism by which TGF-beta inhibits Ig chain synthesis is unclear.
	The mechanisms by which Tax activates NF-kappa B are unclear.
	The exact mechanism by which HTLV-I contributes to leukemogenesis is unclear.
Testitem	Dexamethasone down-regulates TPA-induced transcription of the c-jun gene during monocytic differentiation by inhibiting activation of the AP-1 site.
	The human immunodeficiency virus type 1 LTR possesses functional redundancy which ensures virus replication in different T-cell types.
	EBNA-2 trans-activates cellular and viral genes like CD23, c-fgr, latent membrane protein 1 (LMP1) and terminal protein 1 (TP1).

A.2 Fragebogen



0% completed

Dear participant,

Thank you for taking part in our study! The following questionnaire should only take 10 - 15 minutes to complete.

The focus of our study is the way speculation can be verbalized in scientific articles. By using specific signalling words such as *improbable* or *suspect* (in the examples below in red) an author can indicate that a scientific observation or result (in the examples below in blue) is not an undisputable fact, but more or less speculative and thus has a reduced likelihood. The following *observation / result* is presented as a fact, i.e. it is definitely the case:

Lipid A myristoyl fatty acid is necessary for the induction of Mn SOD.

By adding the signalling word *improbable* the author attributes a low likelihood to the same *observation / result*:

It is improbable that lipid A myristoyl fatty acid is necessary for the induction of Mn SOD.

The signalling word *suspect*, on the other hand, indicates a moderate likelihood of the *observation / result*:

We suspect that lipid A myristoyl fatty acid is necessary for the induction of Mn SOD.

You will now be presented a number of sentences for which we would like you to rate the contained signalling words. Please specify on a scale from 0 - 100% (0% = 'definitely not the case', 100% = 'definitely the case') the likelihood you think the author attributes to a scientific observation / result by using these signalling words. In order to do this, please indicate a range of likelihood (e.g. between 40% and 60%) by dragging the two sliders on the left side to the desired positions on the scale:

is definitely not the case is definitely the case

0 10 20 30 40 50 60 70 80 90 100



Please make sure to use both sliders! If in some cases you would like to specify a single value instead of a range you can move the two sliders to the same position.

Of course all answers will remain confidential.

Abbildung 23: Einleitungsteil des Fragebogens mit Aufgabenerklärung

Please specify the likelihood you think the author attributes to a **scientific observation / result** by using **signalling words**.

NF-kappa B activation **could** constitute one of the mechanisms whereby the expression of kappa B-responsive genes is enhanced in phagocytosing neutrophils.

is definitely
not the case

0 10 20 30 40 50 60 70 80 90 100

is definitely
the case



The lymphoid-specific Oct2B protein **has been suggested to** be involved in the function of the octamer motif in the context of the immunoglobulin heavy-chain (IgH) enhancer.

is definitely
not the case

0 10 20 30 40 50 60 70 80 90 100

is definitely
the case



Specific staining as well as electron microscopic examinations revealed the accumulation of metal gold in the cells, **which supports the hypothesis that gold ions block NF-kappaB-DNA binding by a redox mechanism**.

is definitely
not the case

0 10 20 30 40 50 60 70 80 90 100

is definitely
the case



Abbildung 24: Mittelteil des Fragebogens mit Items und Wahrscheinlichkeitsskala

Finally, we would like to ask a few questions pertaining to your person.

1.) You are ... ▾

2.) How old are you? years old

3.) What is your native language?

4.) How would you rate your proficiency in reading and writing English?

basic fair good very good excellent (native/bilingual)

5.) What is your highest educational qualification?

University entrance degree

Undergraduate degree (e.g. Bachelor)

Graduate degree (e.g. Master, Diplom)

PhD

Other:

6.) What is your major field of study? What area are you currently working in?

Do you have any comments concerning this questionnaire?
If you noticed anything negative during your participation or if some of the questions were not clear enough – please inform us about it.

Abbildung 25: Schlussteil des Fragebogens mit Angaben zur Person

Literaturverzeichnis

- Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford (u.a.): Oxford University Press.
- Albert, R. & Marx, N. (2010). *Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht*. Tübingen: Narr.
- American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6. Aufl.). Washington, DC: Autor.
- Behnam, B., Naeimi, A. & Darvishzade, A. (2012). A Comparative Genre Analysis of Hedging Expressions in Research Articles: Is Fuzziness Forever Wicked? *English Language and Literature Studies*, 2 (2), 20-38.
- Bocklisch, F. (2011). *Die Analyse und Bewertung vager linguistischer Terme mittels fuzzy Methodik*. Unveröffentlichte Dissertation, Technische Universität Chemnitz. Zugriff auf <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-83410>
- Bongelli, R., Canestrari, C., Riccioni, I., Zuczkowski, A., Buldorini, C., Pietrobon, R., Lavelli, A. & Magnini, B. (2012). A Corpus of Scientific Biomedical Texts Spanning over 168 years annotated for Uncertainty. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (S. 2009-2014).
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Brosius, F. (2011). *SPSS 19*. Heidelberg (u.a.): mitp.
- Brown, P. & Levinson, S. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Hrsg.), *Questions and politeness. Strategies in social interaction* (S. 56-311). Cambridge (u.a.): Cambridge University Press.
- Brun, W. & Teigen, K. H. (1988). Verbal Probabilities: Ambiguous, Context-Dependent, or Both? *Organizational Behavior and Human Decision Processes*, 41, 390-404.
- Chafe, W. (1986). Evidentiality in English Conversation and Academic Writing. In W. Chafe & J. Nichols (Hrsg.), *Evidentiality: The Linguistic Coding of Epistemology* (S. 261-272). Norwood, New Jersey: Ablex.
- Clark, D. A. (1990). Verbal Uncertainty Expressions: A Critical Review of Two Decades of Research. *Current Psychology: Research and Reviews*, 9 (3), 203-235.
- Clemen, G. (1997). The Concept of Hedging: Origins, Approaches and Definitions. In R. Markkanen & H. Schröder (Hrsg.), *Hedging and Discourse. Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (S. 235-248). Berlin und New York: de Gruyter.

- Clemen, G. (1998). *Hecken in deutschen und englischen Texten der Wirtschaftskommunikation*. Unveröffentlichte Dissertation, Universität Siegen. Zugriff auf <http://dokumentix.ub.uni-siegen.de/opus/volltexte/2006/121/>
- Coates, J. (1987). Epistemic Modality and Spoken Discourse. *Transactions of the Philological Society*, 85 (1), 110-131.
- Dalianis, H. & Skeppstedt, M. (2010). Creating and Evaluating a Consensus for Negated and Speculative Words in a Swedish Clinical Corpus. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010)* (S. 5-13).
- Desclés, J., Alrahabi, M. & Desclés, J.-P. (2011). BioExcom: Detection and categorization of speculative sentences in biomedical literature. In Z. Vetulani (Hrsg.), *Human Language Technology. Challenges for Computer Science and Linguistics* (S. 478-489). Heidelberg (u.a.): Springer.
- Field, A. (2009). *Discovering Statistics Using SPSS* (3. Aufl.). London (u.a.): Sage.
- Friedman, C., Alderson, P. O., Austin, J. H. M., Cimino, J. J. & Johnson, S. B. (1994). A General Natural-language Text Processor for Clinical Radiology. *Journal of the American Medical Informatics Association*, 1 (2), 161-174.
- Fries, N. (2010). Modalität. In H. Glück (Hrsg.), *Metzler Lexikon Sprache* (S. 433). Stuttgart und Weimar: Metzler.
- Graefen, G. (2000). „Hedging“ als neue Kategorie? Ein Beitrag zur Diskussion. Zugriff auf http://www.kuwi.europa.uni.de/de/lehrstuhl/sw/sw2/forschung/hedging/hedging_als_neue_kategorie/index.html
- Gries, S. T. (2009). *Statistics for Linguistics with R. A Practical Introduction*. Berlin: de Gruyter.
- Halliday, M. A. K. (1970). Functional Diversity in Language as Seen from a Consideration of Modality and Mood in English. *Foundations of Language*, 6 (3), 322-361.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar* (3. Aufl.). London: Arnold.
- Henriksson, A. & Velupillai, S. (2010). Levels of Certainty in Knowledge-Intensive Corpora: An Initial Annotation Study. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010)* (S. 41-45).
- Holmes, J. (1982). Expressing Doubt and Certainty in English. *RELC Journal*, 13 (2), 9-28.
- Horn, L. R. (2001). *A Natural History of Negation*. Stanford, California: CSLI.
- Hoye, L. (1997). *Adverbs and Modality in English*. London und New York: Longman.
- Hoye, L. (2009). Modality in Discourse: The Pragmatics of Epistemic Modality. In A. Tsangalidis & R. Facchinetti (Hrsg.), *Studies on English Modality* (S. 99-132). Bern: Peter Lang.

- Huddleston, R. (2010). The verb. In R. Huddleston & K. Pullum Geoffrey (Hrsg.), *The Cambridge Grammar of the English Language* (S. 71-212). Cambridge (u.a.): Cambridge University Press.
- Hyland, K. (1996). Writing Without Conviction? Hedging in Science Research Articles. *Applied Linguistics*, 17 (4), 433-454.
- Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam und Philadelphia: John Benjamins.
- Kim, J.-D., Ohta, T., Tateisi, Y. & Tsujii, J. (2003). GENIA corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19 (Suppl 1), i180-i182.
- Kim, J.-D., Ohta, T. & Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9, 10-34.
- Lehmann, C. (2012). *Situation*. Linguistische Skripte; Universität Erfurt. Zugriff auf http://www.christianlehmann.eu/ling/lg_system/sem/index.html?http://www.christianlehmann.eu/ling/lg_system/sem/situation.php
- Light, M., Qiu, X. Y. & Srinivasan, P. (2004). The Language of Bioscience: Facts, Speculations, and Statements in Between. In *Proceedings of the Workshop on Linking Biological Literature, Ontologies and Databases (BioLink 2004)* (S. 17-24).
- Linke, A., Nussbaumer, M. & Portmann, P. R. (2004). *Studienbuch Linguistik* (5. Aufl.). Tübingen: Niemeyer.
- Lyons, J. (1977). *Semantics* (Bd. 2). Cambridge (u.a.): Cambridge University Press.
- Markkanen, R. & Schröder, H. (1997). Hedging: A Challenge for Pragmatics and Discourse Analysis. In R. Markkanen & H. Schröder (Hrsg.), *Hedging and Discourse. Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (S. 3-18). Berlin und New York: de Gruyter.
- McShane, M., Nirenburg, S., Beale, S. & O'Hara, T. (2005). Semantically Rich Human-Aided Machine Annotation. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky* (S. 68-75).
- Medlock, B. & Briscoe, T. (2007). Weakly Supervised Learning for Hedge Classification in Scientific Literature. In *Proceedings of 45th Annual Meeting of the Association of Computational Linguistics* (S. 992-999).
- Morante, R. & Sporleder, C. (2012). Modality and Negation: An Introduction to the Special Issue. *Computational Linguistics*, 38 (2), 223-260.
- Myers, G. (1989). The Pragmatics of Politeness in Scientific Articles. *Applied Linguistics*, 10 (1), 1-35.
- Nachtigall, C. & Wirtz, M. (2006). *Wahrscheinlichkeitsrechnung und Inferenzstatistik. Statistische Methoden für Psychologen Teil 2* (4. Aufl.). Weinheim und München: Juventa.

- Nirenburg, S. & McShane, M. (2008). *Annotating Modality. OntoSem Final Project Report* (Bericht). University of Maryland. Zugriff auf <http://intranet.umiacs.umd.edu/conferences/sapm05/Semantic-Annotation-Guidelines/UMBC-Report-Modality.doc>
- Nurmi, A. (2007). Employing and elaborating annotation for the study of modality. *Studies in Variation, Contacts and Change in English*, 1. Zugriff auf <http://www.helsinki.fi/varieng/journal/volumes/01/nurmi/>
- Nuyts, J. (2001). *Epistemic Modality, Language, and Conceptualization: A Cognitive-Pragmatic Perspective*. Amsterdam und Philadelphia: John Benjamins.
- Ohta, T., Kim, J.-D. & Tsujii, J. (2007). *Guidelines for event annotation*. Zugriff auf <http://docs.google.com/viewer?a=v&pid=sites&srcid=Z2VuaWFwcm9qZWNOIm9yZ3xtYWlufGd4OjI4M2RmNTVjNTg5ZjZkZTI>
- Palmer, F. R. (1986). *Mood and Modality*. Cambridge (u.a.): Cambridge University Press.
- Palmer, F. R. (2001). *Mood and Modality* (2. Aufl.). Cambridge (u.a.): Cambridge University Press.
- Parladé, I. V. & Giner, D. (2008). Beyond Mood and Modality: Epistemic Modality Markers as Hedges in Research Articles. A Cross-Disciplinary Study. *Revista Alicantina de Estudios Ingleses*, 21, 171-190.
- Perkins, M. R. (1983). *Modal Expressions in English*. London: Frances Pinter.
- Popper, K. R. (1969). *Logik der Forschung* (3. Aufl.). Tübingen: Mohr.
- Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2 (2), 37-63.
- Rasch, D. & Kubinger, K. D. (2006). *Statistik für das Psychologiestudium*. München: Spektrum Akademischer Verlag.
- Rayson, P., Archer, D., Piao, S. & McEnery, T. (2004). The UCREL Semantic Analysis System. In *Proceedings of the Beyond Named Entity Recognition Semantic Labeling for NLP Tasks Workshop* (S. 7-12).
- Rizomilioti, V. (2006). Exploring Epistemic Modality in Academic Discourse Using Corpora. In E. A. Macià, A. S. Cervera & C. R. Ramos (Hrsg.), *Information Technology in Languages for Specific Purposes. Issues and Prospects*. Berlin (u.a.): Springer.
- Rohrman, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222-245.
- Rubin, V. L. (2007). Stating with Certainty or Stating with Doubt: Intercoder Reliability Results for Manual Annotation of Epistemically Modalized Statements. In *Proceedings of NAACL HLT 2007, Companion Volume* (S. 141-144).

- Rubin, V. L. (2010). Epistemic modality: From uncertainty to certainty in the context of information seeking as interactions with texts. *Information Processing and Management*, 46, 533-540.
- Rubin, V. L., Liddy, E. D. & Kando, N. (2006). Certainty Identification in Texts: Categorization Model and Manual Tagging Results. In J. G. Shanahan, Y. Qu & J. Wiebe (Hrsg.), *Computing Attitude and Affect in Text: Theory and Applications* (S. 61-76). Dordrecht: Springer.
- Salager-Meyer, F. (1994). Hedges and Textual Communicative Function in Medical English Written Discourse. *English for Specific Purposes*, 13 (2), 149-170.
- Saurí, R. (2008). *A Factuality Profiler for Eventualities in Text*. Unveröffentlichte Dissertation, Brandeis University. Zugriff auf http://www.cs.brandeis.edu/~roser/pubs/sauriDiss_1.5.pdf
- Saurí, R. & Pustejovsky, J. (2009). FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43, 227-268.
- Scott, D., Barone, R. & Koeling, R. (2012). Corpus Annotation as a Scientific Task. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (S. 1481-1485).
- Shatkay, H., Wilbur, W. J. & Rzhetsky, A. (2005). *Appendices to the Guidelines for Annotation*. Zugriff auf <http://www.ncbi.nlm.nih.gov/CBBresearch/Wilbur/Appendices.pdf>
- Simon-Vandenberg, A.-M. & Aijmer, K. (2007). *The Semantic Field of Modal Certainty. A Corpus-Based Study of English Adverbs*. Berlin und New York: de Gruyter.
- Teigen, K. H. & Brun, W. (2003). Verbal Expressions of Uncertainty and Probability. In D. Hardman & L. Macchi (Hrsg.), *Thinking. Psychological perspectives on reasoning, judgment and decision making* (S. 125-145). Chichester (u.a.): Wiley.
- Thompson, P., Nawaz, R., McNaught, J. & Ananiadou, S. (2011). Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics*, 12, 393-411.
- Thompson, P., Venturi, G., McNaught, J., Montemagni, S. & Ananiadou, S. (2008). Categorising Modality in Biomedical Texts. In *Proceedings of the Workshop on Building and Evaluating Resources for Biomedical Text Mining (LREC 2008)* (S. 27-34).
- Velupillai, S. (2012). *Shades of Certainty. Annotation and Classification of Swedish Medical Records*. Unveröffentlichte Dissertation, Stockholms Universitet. Zugriff auf <http://www.diva-portal.org/smash/get/diva2:512263/FULLTEXT01.pdf>
- Vincze, V., Szarvas, G., Farkas, R., Móra, G. & Csirik, J. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes.

- BMC Bioinformatics*, 9 (Suppl 11), 9-17.
- Vold, E. T. (2006). Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics*, 16 (1), 61-87.
- Wallsten, T. S., Zwick, R. & Forsyth, B. (1986). Measuring the Vague Meanings of Probability Terms. *Journal of Experimental Psychology: General*, 115 (4), 348-365.
- Westney, P. (1986). How to be More or Less Certain in English: Scalarity in Epistemic Modality. *International Review of Applied Linguistics in Language Teaching*, 24 (4), 311-336.
- Wiebe, J., Wilson, T., Bruce, R., Bell, M. & Martin, M. (2004). Learning Subjective Language. *Computational Linguistics*, 30 (3), 277-308.
- Wilbur, W. J., Rzhetsky, A. & Shatkay, H. (2006). New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC Bioinformatics*, 7, 356-365.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8, 338-353.

Internetquellen (letzter Zugriff jeweils 04.10.13)

- BioScope-Korpus: <http://www.inf.u-szeged.hu/rgai/bioscope>
- GENIA-Event-Korpus: <http://www.nactem.ac.uk/genia/genia-corpus/event-corpus>
- GENIA-Projekt: <http://www.nactem.ac.uk/genia/>
- MEDLINE-Datenbank: <http://www.nlm.nih.gov/pubs/factsheets/medline.html>
- Meta-Knowledge-GENIA-Event-Korpus: <http://www.nactem.ac.uk/meta-knowledge/download.php>
- MPQA-Opinion-Korpus: <http://mpqa.cs.pitt.edu/>
- SocSci Survey: <http://www.soscisurvey.de/>
- SPSS: <http://www-01.ibm.com/software/de/analytics/spss/products/statistics/>
- USAS: <http://ucrel.lancs.ac.uk/usas/>

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit mit dem Titel

Die Skalierung epistemisch modaler Ausdrücke

Eine empirische Untersuchung mit Fokus auf biomedizinischen Texten

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, alle Ausführungen, die anderen Schriften wörtlich oder sinngemäß entnommen wurden, kenntlich gemacht sind und die Arbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung war.

Christine Engelmann