



MIKO

(Mitschreiben in Vorlesungen: Ein multimodales Lehr-Lernkorpus) Handbuch

Leonore Spiegel, Maria Parker, Tim Feldmüller, Lisa Lenort, Katrin Wisniewski

Version 1.0

Stand: 7. Dezember 2020

https://home.uni-leipzig.de/sprastu/korpora/miko

Korpusteam

Dr. Katrin Wisniewski: Projektleitung

Leonore Spiegel, M. A.: Datenerhebung, Korpusaufbereitung, Transkription und Annotation, Qualitätskontrolle

Maria Parker, M. Sc.: Datenerhebung, Korpusaufbereitung, Qualitätskontrolle

Tim Feldmüller, M. A.: Korpusaufbereitung, Transkription und Annotation, Qualitätskontrolle, Vorlesungssegmentierung

Lisa Lenort, B. A.: Mitschriftenaufbereitung und -bewertung, Qualitätskontrolle

Elisabeth Muntschick, M. A.: korpustechnische Fragen

Anna Grimmer, B. A.: Qualitätskontrolle

Lydia Thumser, B. A.: Datenaufbereitung

Inhalt

Τ	abeller	n- un	d Abbildungsverzeichnis	i
Е	inleitu	ng		1
Ι	Nu	ıtzunş	g, Daten und Erstellung	3
1	Nu	ıtzunş	g des Korpus	4
	1.1	Da	tenschutz und Urheberrecht	4
	1.2	Ko	rpuslizenz	4
	1.3	Zu	gang	4
	1.3	.1	Datenbank für Gesprochenes Deutsch	4
	1.3	.2	Langzeitarchiv des IDS-Repositoriums	5
	1.3	.3	Archiv für Gesprochenes Deutsch	6
	1.3	.4	SpraStu-Homepage	6
	1.4	Ziti	ieren des Korpus	7
2	Da	tenül	persicht	8
	2.1	.1	Primärdaten: Audio- und Videodateien	8
	2.1	.2	Sekundärdaten: Transkripte	10
	2.1	.3	Zusatzmaterialien: Mitschriften und Metadaten	11
	2.1	.4	Sprechereignisübersicht	12
3	Me	tadat	en	13
	3.1	Spr	echereignisse	13
	3.2	Mit	schriften	18
4	Vo	rlesu	ngen	28
	4.1	Da	tenerhebung	28
	4.2	Da	tenaufbereitung	29
	4.2	.1	Schneiden der Videos	29
	4.2	.2	Anonymisierung der Videos	29
	4.3	Tra	nskriptions- und Annotationsprozess und Qualitätssicherung	30
	4.3	.1	Phase 1: Entwicklung eines Goldstandards	30
	4.3	.2	Phase 2: Transkriptions- und Annotationsphase	31
	4.3	.3	Phase 3: Prüfung der Transkripte durch die Transkribentin/den Transkribente	en31
	4.3	.4	Phase 4: Reviewphase	31
	4.3	.5	Reliabilität der Transkription und Annotation	32
	4.3	.6	Automatische Annotationen	32

5	Mit	schri	ften	33
	5.1	Dat	enerhebung	33
	5.1.	1	Einsammeln der Vorlesungsmitschriften	33
	5.1.	2	Fragebogen zu den Mitschriften	33
	5.2	Dat	renaufbereitung	33
	5.3	Ana	alysemethoden Mitschriften	34
	5.3.	1	Erfassung von Kenncharakteristika der Mitschriften	34
	5.3.	2	Segmentierung der Vorlesungen in Wissenselemente	37
	5.3.	3	Bewertung der Mitschriften	42
II	Trans	kript	ion und Annotation	47
6	Übo	ersicl	nt Spurstruktur	48
7	Tra	nskri	ptionssystem HIAT	50
	7.1	Beg	ründung der Auswahl des Transkriptionssystems	50
	7.2	Lite	erarische Umschrift und Normalisierung	51
	7.3	Seg	mentierungseinheiten	52
	7.3.	1	Äußerungen	52
	7.3.2		Wörter	52
	7.3.	3	Events	52
8	Ver	balsp	our Dozent/-in X[v]	54
	8.1	Äuſ	Berungsmodus und Äußerungsendzeichen	54
	8.1.	1	Assertion	54
	8.1.	2	Frage	54
	8.1.	3	Aufforderung, Ausruf	55
	8.1.	4	Abbruch	56
	8.1.	5	Äußerung ohne Modus	57
	8.2	Pau	isen	57
	8.3	Bes	ondere Wörter	58
	8.3.	1	Verzögerungspartikeln	58
	8.3.	2	Interjektionen	59
	8.3.	3	Abkürzungen	59
	8.3.	4	Buchstabieren	59
	8.3.	5	Zahlen	60
	8.3.	6	Fremdwörter und fremdsprachige Äußerungen	60

8.3	.7 Teilwörter	61
8.3	.8 Morpheme	61
8.3	.9 Eigennamen	61
8.4	Unverständliches und Zweifelsfälle	61
8.5	Intrasegmentale Phänomene	63
8.5	.1 Wiederholungen (Reihung und Repetition)	63
8.5	.2 Sprechhandlungsaugmente	63
8.5	.3 Herausstellungen und Ausklammerungen	64
8.5	.4 Parenthesen	64
8.5	.5 Ankündigungen	65
8.5	.6 Reparaturen	66
8.6	Uneigentliches Sprechen	68
8.6	.1 Zitierende Handlungen	69
8.6	.2 Metasprachliche Verwendung von Begriffen	70
8.6	.3 Überschriften, Titel, Themen- und Kategorienbenennungen	72
8.7	Nicht-phonologische akustische Phänomene	73
9 Bes	sondere Betonungen X[akz]	74
10 An	notationen auf Wortebene X[wort]	75
10.1	Versprecher	75
10.2	Pausen im Wort	76
10.3	Vorschläge bei schwer Verständlichem	76
11 Spr	rechweise X[sprech]	77
12 Me	etasprachliche Erläuterungen $X[meta]$	78
13 No	onverbale Handlungen $X[nv]$ und $X[nv2]$	80
13.1	((zeigt auf))	81
13.2	((fasst an))	81
13.3	((demonstriert))	81
13.4	((veranschaulicht))	82
13.5	Schreiben und Zeichnen	82
13.6	Korrekturen und Unterstreichungen	82
14 Tex	xt-Bild-Bezüge X[bild]	84
14.1	Abk	87
14.2	Feh	88

14.3	Abw	88
14.4	Alt	90
15 Ass	sistent $Assist[v]$	92
15.1	Verbale Äußerungen	92
15.2	Nonverbale Handlungen	92
16 Au	ditorium $NN[v]$	93
16.1	Verbale Äußerungen	93
16.2	Nicht-phonologische akustische Phänomene	94
17 Ko	mmentare zum Transkript $X[komm]$	95
Literatu	rverzeichnis	96

Tabellen- und Abbildungsverzeichnis

Tabelle 1: MIKO-Daten in der DGD	5
Tabelle 2: MIKO-Daten im LZA	6
Tabelle 3: Technische Daten zu den Audiodateien	8
Tabelle 4: Technische Daten zu den Videodateien	9
Tabelle 5: Technische Daten zu den Transkripten	10
Tabelle 6: Technische Daten zu den Mitschriften	11
Tabelle 7: Technische Daten zu den Metadaten	11
Tabelle 8: Sprechereignisübersicht	12
Tabelle 9: Metadatenvariablen der Sprechereignisse	17
Tabelle 10: Metadaten der Mitschriften	27
Tabelle 11: Deskriptoren für die Bewertung der Korrektheit	45
Tabelle 12: Deskriptoren für die Bewertung der Organisation und Nachvollziehbarkeit	45
Tabelle 13: Transkriptionsspuren	48
Tabelle 14: Manuelle Annotationsspuren	49
Tabelle 15: Automatische Transkriptions- und Annotationsspuren	49
Tabelle 16: Mathematische Phänomene, die mit Text-Bild-Bezug getaggt werden	86
Abbildung 1: Videoausschnitt MIKO_E_00007, 60:18	29
Abbildung 2: Erfasste Kenncharakteristika der Mitschriften	34
Abbildung 3: Codierfenster in MAXQDA, MIKO_E_00007	40
Abbildung 4: Liste der Codings in MAXQDA, MIKO_E_00007	41
Abbildung 5: Eventgrenzen bei wortgenauer Annotation; MIKO E 00001, 23:14	53

Einleitung

Multimodale Korpora sind linguistische Text- oder Diskurssammlungen geschriebener oder gesprochener Sprache, die empirische Daten verschiedener "sensory modalit[ies]" (Allwood, 2008, S. 208) enthalten. Aufgrund der (korpus-)technischen Entwicklungen der letzten zwei Jahrzehnte nehmen multimodale Korpora eine zunehmend wichtige Rolle in der Korpuslinguistik ein und werden für Forschende immer leichter zugänglich. In der Regel handelt es sich um "Verschriftlichungen gesprochener Sprache [in Verbindung] mit anderen Modalitäten" (Sasaki & Witt, 2016, S. 195), die die Untersuchung vielfältiger linguistischer Fragestellungen ermöglichen.

MIKO (Mitschreiben in Vorlesungen: Ein multimodales Lehr-Lernkorpus) ist ein multimodales, wissenschaftssprachliches Vorlesungskorpus und beinhaltet sowohl gesprochene (Vorlesungen mit Audio- und Videoaufnahmen) als auch geschriebene Sprache (studentische Mitschriften). Neben dem hier vorgestellten Korpus sind weitere mulitmodale Korpora gesprochener Wissenschaftssprache zum Beispiel euroWiss (http://hdl.handle.net/11022/0000-0001-7DBA-2) und GeWiss (https://gewiss.uni-leipzig.de).

MIKO entstand im BMBF-geförderten Forschungsprojekt "Sprache und Studienerfolg bei Bildungsausländer/-innen" (SpraStu; Laufzeit 2017-2020; FKZ: 01PX16020A; Universitäten Leipzig (Gesamtleitung) und Würzburg; https://home.uni-leipzig.de/sprastu/projekt/). SpraStu untersuchte sprachliche (und andere) Einflussfaktoren auf den Studienerfolg und widmete sich unter anderem der Erstellung dieses Korpus zur Untersuchung des Mitschreibens in Vorlesungen in der Studieneingangsphase. Für weitere Informationen zu den Projektbausteinen, Zielen, Ergebnissen und Analysen von SpraStu sei auf die Publikation Sprache und Studienerfolg bei Bildungsausländer/-innen (Wisniewski, Lenhard et al., in Vorbereitung) verwiesen.

MIKO enthält korpuslinguistisch aufbereitete Vorlesungsaufnahmen, Mitschriften von Studierenden mit Deutsch als L1 und L2 sowie umfangreiche Metadaten, die im Rahmen des weitgefächerten Studiendesigns erhoben wurden, darunter Sprachkompetenzmessungen sowie Ergebnisse von Fragebögen mit den Studierenden.

Das Mitschreiben in Vorlesungen gilt als äußerst komplex, da simultan sowohl rezeptive als auch produktive Fertigkeiten beansprucht werden (Steets, 2003). Insbesondere Studierende mit Deutsch als L2 empfinden das Mitschreiben zu Studienbeginn als extrem schwierig (Bärenfänger et al., 2015; Marks, 2015). Mitschriften sind empirisch-linguistisch zudem herausfordernde Untersuchungsobjekte (Wisniewski, 2019), da sie nie für sich alleine stehen, sondern sich immer eng auf visuellen und auditiven Input beziehen. Aus diesem Grund ist der Zugang zu einem multimodalen Korpus für ihre weitere Erforschung vielversprechend (Wisniewski, Spiegel, Parker et al., in Vorbereitung).

Das vorliegende Handbuch gliedert sich in zwei Teile. Teil I stellt das Korpus zunächst hinsichtlich seiner Nutzbarkeit (Kapitel 1) sowie der verfügbaren Daten (Kapitel 2) und Metadaten (Kapitel 3) vor. Außerdem wird die Korpuserstellung, unterteilt nach Vorlesungen (Kapitel 4) und Mitschriften (Kapitel 5) dokumentiert. Teil II beinhaltet die Transkriptions- und Annotationsrichtlinien, nach denen MIKO erstellt wurde. Es werden ein Überblick über die in MIKO vorhandenen Spuren gegeben (Kapitel 6), die Grundlagen des Transkriptionssystems Halbinterpretative Arbeitsinterpretation (HIAT) (Kapitel 7) vorgestellt und anschließend alle

Transkriptions- und Annotationsrichtlinien von MIKO, unterteilt nach den einzelnen Spuren, erläutert (Kapitel 8 bis 17).

I Nutzung, Daten und Erstellung

1 Nutzung des Korpus

1.1 Datenschutz und Urheberrecht

Der Umgang mit dem Datenschutzrecht, allgemeinen Persönlichkeitsrecht und Urheberrecht orientiert sich an den Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) zu rechtlichen Aspekten bei der Erstellung von Sprachkorpora (DFG, 2013) und steht im Einklang mit der europäischen Datenschutzgrundverordnung (https://dsgvo-gesetz.de/).

Die Dozierenden und ein Assistent der Vorlesungen haben in schriftlichen Einverständniserklärungen der Verarbeitung und Veröffentlichung personenbezogener Daten (Audio- und Videoaufnahmen sowie deren Transkription) in einem Online-Korpus zugestimmt. Jede/-r Dozierende konnte entscheiden, ob eine Pseudonymisierung des Namens erfolgen soll oder nicht. Eine Dozentin stimmte der Veröffentlichung der Videoaufnahmen nicht zu, sodass diese nicht verfügbar sind.

Alle personenbezogenen und personenbeziehbaren Daten (visuell oder auditiv) der an den Vorlesungssitzungen teilnehmenden Studierenden wurden durch technische Nachbearbeitung anonymisiert (Kapitel 4.2.2). In den Transkripten sind Redebeiträge Studierender ebenfalls anonymisiert (Kapitel 16).

Die Studierenden stimmten in schriftlichen Einverständniserklärungen der Verarbeitung ihrer Mitschriften zu. Die anonymisierten Mitschriften können beim Archiv für Gesprochenes Deutsch (AGD) unter der E-Mail-Adresse agd@ids-mannheim.de unter Angabe des Forschungsvorhabens angefragt werden (Kapitel 1.3.3).

1.2 Korpuslizenz

Ein durch Authentifizierung eingeschränkter Personenkreis erhält Zugriffs- und Nutzungsrechte für das Korpus zur wissenschaftlichen, nicht kommerziellen Nutzung. Die in dem Korpus bereitgestellten Daten dürfen heruntergeladen, verändert und ausgedruckt werden. Die Daten dürfen nur für Forschung, Lehre und Studium genutzt werden. Für andere Zwecke ist eine Nutzung nicht gestattet. Insbesondere darf das Korpus nicht an Dritte weitergegeben, publiziert oder in anderer Form der Öffentlichkeit zugänglich gemacht werden. Die angebotenen Daten können im Rahmen des Zitatrechtes mit Angabe der Quelle in eigenen selbstständigen Werken wiedergegeben werden (Kapitel 1.4).

1.3 Zugang

1.3.1 Datenbank für Gesprochenes Deutsch

MIKO ist ab 2021 in der Datenbank für Gesprochenes Deutsch (DGD) des Leibniz-Instituts für Deutsche Sprache (IDS) in Mannheim verfügbar (https://dgd.ids-mannheim.de) und für

registrierte Nutzer in dieser durchsuchbar. Die DGD als Korpusmanagementsystem ist eine browserbasierte Plattform mit umfangreichen Möglichkeiten des Abfragens und Durchsuchens von Transkripten und Metadaten für korpuslinguistische Analysen speziell gesprochener Sprache.

In der DGD werden folgende MIKO-Daten zur Verfügung gestellt:

Daten	Dateiformat	Anzahl
Vorlesungstranskripte (z. B. MIKO_E_00001_SE_01_T_01)	exb	8
Audioaufnahmen (z. B. <i>MIKO_E_00005_SE_01_A_01_DF_01</i>)	wav	12
Videoaufnahmen (z. B. <i>MIKO_E_00012_SE_01_V_01_DF_01</i>)	mp4	9
MIKO-Handbuch (MIKO_Z_01_Handbuch)	PDF/A	1
Erläuterung zum Mitschriftenzugriff (MIKO_Z_02_Mitschriftenzugriff)	txt	1
Metadaten zu den studentischen Mitschriften (MIKO_Z_03_Metadaten_Mitschriften)	CSV	1
Übersicht über die Metadatenvariablen zu den studentischen Mitschriften (MIKO_Z_04_Metadaten_Mitschriften_Variablen)	PDF/A	1

Tabelle 1: MIKO-Daten in der DGD

1.3.2 Langzeitarchiv des IDS-Repositoriums

MIKO ist ab 2021 im Langzeitarchiv (LZA) des IDS-Repositoriums verfügbar (PID: https://hdl.handle.net/10932/00-0534-6426-9660-0101-7). Der Zugang ist kostenfrei über die Authentifizierungs- und Autorisierungsinfrastruktur (AAI) möglich. Interessierte Nutzer/-innen, die keiner Forschungsinstitution angehören, die Teil der AAI ist, können sich unter https://idm.clarin.eu/user/home registrieren und sich so beim LZA für die Nutzung identifizieren. Das IDS ist Teil der CLARIN-Infrastruktur, sodass das Korpus über einige Rahmenmetadaten auch im Virtual Language Observatory auffindbar ist (https://www.clarin.eu/content/virtual-language-observatory-vlo).

Folgende MIKO-Daten sind im LZA verfügbar:

Daten	Dateiformat	Anzahl
untokenisierte Vorlesungstranskripte (z. B. <i>MIKO_E_00004_SE_01_T_01</i>)	exb, txt	8
tokenisierte Vorlesungstranskripte mit Lemmatisierung und POS-Tagging (z. B. MIKO_E_00004_SE_01_T_02)	exb	8
Audioaufnahmen (z. B. <i>MIKO_E_00005_SE_01_A_01_DF_01</i>)	wav	12
Videoaufnahmen (z. B. <i>MIKO_E_00012_SE_01_V_01_DF_01</i>)	mp4	9
MIKO-Handbuch (MIKO_Z_01_Handbuch)	PDF/A	1
Erläuterung zum Mitschriftenzugriff (MIKO_Z_02_Mitschriftenzugriff)	txt	1
Metadaten zu den studentischen Mitschriften (MIKO_Z_03_Metadaten_Mitschriften)	CSV	1
Übersicht über die Metadatenvariablen zu den studentischen Mitschriften (MIKO_Z_04_Metadaten_Mitschriften_Variablen)	PDF/A	1
Metadaten zu den Sprechereignissen (MIKO_Z_05_Metadaten_Vorlesungen)	CSV	1
Übersicht über die Metadatenvariablen zu den Sprechereignissen	PDF/A	1
(MIKO_Z_06_Metadaten_Mitschriften_Variablen)		

Tabelle 2: MIKO-Daten im LZA

1.3.3 Archiv für Gesprochenes Deutsch

Studentische Mitschriften aus den zwölf Vorlesungssitzungen (N=146, PDF/A-Format) können über das AGD unter der E-Mail-Adresse <u>agd@ids-mannheim.de</u> unter Angabe des Forschungsvorhabens angefragt werden (Kapitel 1.1).

1.3.4 SpraStu-Homepage

Das MIKO-Handbuch sowie ausgewählte Metadaten sind außerdem auf der Homepage des Forschungsprojekts SpraStu verfügbar: https://home.uni-leipzig.de/sprastu/korpora/miko/

1.4 Zitieren des Korpus

Die Daten von MIKO sollen im Rahmen des Zitatrechtes wiedergegeben werden. Um auf MIKO als Ganzes zu verweisen, zitieren Sie bitte folgenden Beitrag:

Wisniewski, K., Spiegel, L., Parker, M., Feldmüller, T. & Lenort, L. (in Vorbereitung). Mitschreiben in Vorlesungen in der Studieneingangsphase: Das multimodale Lehr-Lernkorpus MIKO. In K. Wisniewski, W. Lenhard, J. Möhring & L. Spiegel (Hg.), *Sprache und Studienerfolg bei Bildungsausländer/-innen*. Waxmann.

Für das Zitieren von spezifischen Daten aus dem Korpus geben Sie folgenden PID an: https://hdl.handle.net/10932/00-0534-6426-9660-0101-7.

Wenn Ihre Zitieranforderungen dies verlangen, können Sie Wisniewski, K., Spiegel, L., Parker, M., Feldmüller, T., Lenort, L. als Herausgeber/-innen des Korpus zitieren.

Für das Zitieren eines Transkripts geben Sie die Kennung des Sprechereignisses an, z. B. MIKO_E_00012.

Für das Zitieren konkreter Ausschnitte aus einem Transkript ist zusätzlich die jeweilige Startzeit der Aufnahme, gerundet auf ganze Sekunden, anzugeben, z. B. MIKO_E_00004, 71:15.

Erfolgt der Zugriff auf das Korpus über die DGD, gelten die dort festgelegten Zitierhinweise.

2 Datenübersicht

MIKO besteht aus korpuslinguistisch aufbereiteten Vorlesungssitzungen (N=8) aus prüfungsrelevanten Pflichtvorlesungen der Fächer Medizin (Funktionelle Anatomie I, Physik für Mediziner), Deutsch als Fremdsprache (Grundzüge der Lexikologie der deutschen Gegenwartssprache) und Wirtschaftswissenschaften (Bürgerliches Recht für Wirtschaftswissenschaftler), die alle für das erste Fachsemester vorgesehen sind und im Wintersemester 2017/18 an der Universität Leipzig mit Audio und Video aufgezeichnet wurden.

Weitere gefilmte, jedoch nicht korpuslinguistisch aufbereitete Vorlesungssitzungen (N=4, je zwei der bereits zitierten Vorlesungen aus Deutsch als Fremdsprache und Wirtschaftswissenschaften) stehen ebenfalls zur Verfügung.

In den insgesamt zwölf Vorlesungssitzungen erhobene, anonymisierte Mitschriften (N=146) werden in archivierter Form, jedoch nicht maschinenlesbar aufbereitet, zur Verfügung gestellt.

Das Forschungsdatenmanagement und die Dateiformate orientieren sich an den Empfehlungen der DFG zu datentechnischen Standards bei der Erhebung von Sprachkorpora (DFG, 2019). Im Folgenden werden die verschiedenen Datentypen des Korpus dargestellt.

2.1.1 Primärdaten: Audio- und Videodateien

Daten zu den Audiodateien	Details
Anzahl	12
Aufnahmetechnik	Funkmikrofon sowie fest installiertes Mikrofon des E- Learning-Service der Universität Leipzig
Dateiformat	wav (unkomprimiert, lineares PCM-Format)
Codec	PCM S16 LE
Kanäle	Mono
Samplingrate	48 000 Hz
Dauer der Aufnahmen gesamt	938 Min, entspricht 15:38 h
Dateigröße gesamt	5136 MB

Tabelle 3: Technische Daten zu den Audiodateien

Daten zu den Videodateien	Details
Anzahl	9
Aufnahmetechnik	Handkameras auf Tisch- bzw. Standstativen sowie fest installierte Aufnahmetechnik des E-Learning-Service der Universität Leipzig
Dateiformat	mp4 (Video File)
Auflösung	1920px x 1080px
Codec	H264/ MPEG-4 AVC
Framerate	25 fps (Bilder pro Sekunde)
Audio	Kanäle: Mono
	Samplingrate: 48 000 Hz
Dauer der Aufnahmen gesamt	770 Min, entspricht 12:50 h
Dateigröße gesamt	15,38 GB

Tabelle 4: Technische Daten zu den Videodateien

Detaillierte Ausführungen zur Aufzeichnung der Vorlesungen sind Kapitel 4.1 zu entnehmen.

2.1.2 Sekundärdaten: Transkripte

Daten zu den Transkripten	Details
Anzahl	8 (liegen jeweils mit und ohne Tokenisierung und Part-of- speech-Tagging (POS-Tagging) vor)
Tool	EXMARaLDA Partitur-Editor (Schmidt & Wörner, 2014)
Auszeichnungssprache	UTF-8
Dateiformat	exb
	txt
Transkription	Dozierende/-r, Assistent, Auditorium (anonymisiert)
Annotation	automatisch: Lemmatisierung und POS-Tagging mithilfe des TreeTaggers (Schmid, 1994), der das Stuttgart-Tübingen- TagSet (STTS) (Schiller et al., 1999) verwendet
	manuell: Betonungen, Annotationen auf Wortebene, Sprechweise, metasprachliche Erläuterungen, nonverbale Handlungen, Text-Bild-Bezüge, Kommentare
Transkriptionssystem	modifizierte HIAT-Richtlinien (Rehbein et al., 2004)
Token¹ gesamt	82075
Types ² gesamt	6076

Tabelle 5: Technische Daten zu den Transkripten

Für die Korpuserstellung wurde das Editor-Tool EXMARaLDA (Schmidt & Wörner, 2014; https://exmaralda.org/de/partitur-editor-de/) gewählt, da dieses besonders für die Arbeit mit mündlichen Daten geeignet ist und die Verknüpfung mit Audio- und Videodateien ermöglicht. Die Transkripte und Audio-/Videoaufnahmen sind durch Zeitmarken (Alignment) verknüpft. Neben der flexiblen Transkription und Annotation im Partitur-Editor bietet das Tool außerdem die Möglichkeit, Metadaten umfassend im Corpus-Manager (COMA) zu organisieren.

Von den insgesamt zwölf gefilmten Vorlesungssitzungen wurden acht für die Transkription ausgewählt, weil zu ihnen die meisten studentischen Mitschriften vorliegen.

¹ Die Tokenzahl der Transkripte lässt sich automatisch über den EXMARaLDA Partitur-Editor und die Funktion *Transcription* > *Count Segments* ermitteln. *HLAT:w* gibt die Anzahl der fortlaufenden Wortvorkommnisse ohne Satzzeichen in den Verbalspuren an. Eine zusätzliche Analyse der Tokenzahl ergab eine Abweichung von drei Token. Der hier angegebene Wert entspricht der korrigierten Tokenzahl.

² Die Typezahl des Korpus wurde über einen automatischen Export aller Lemmata ermittelt. Als <*UNKNOWN>* bzw. ambig (|) getaggte Lemmata wurden bei der Analyse manuell korrigiert. In den Transkripten wurden jedoch keine Korrekturen vorgenommen. In den Metadaten stehen zusätzlich zur Gesamttypezahl des Korpus auch die Typezahlen der einzelnen Transkripte zur Verfügung (Kapitel 3.1)

Die vier weiteren, nicht transkribierten Vorlesungssitzungen werden mit Audio- und Videoaufnahmen archiviert zur Verfügung gestellt, sodass Transkriptionen auf deren Grundlage nachträglich angefertigt werden können.

Die Transkripte liegen im LZA jeweils in einer untokenisierten und einer tokenisierten Version vor. Es sei darauf hingewiesen, dass die tokenisierten Transkripte aufgrund ihrer Größe einige Minuten zum Öffnen benötigen. Auch ist in ihnen keine Satzzeichentokenisierung enthalten, sodass zu den Satzzeichen kein POS-Tagging vorliegt.

2.1.3 Zusatzmaterialien: Mitschriften und Metadaten

Daten zu den Mitschriften	Details
Anzahl	146 (Deutsch als L2: 88; Deutsch als L1: 58)
Dateiformat	PDF/A
Dateigröße gesamt	413 MB

Tabelle 6: Technische Daten zu den Mitschriften

Detaillierte Ausführungen zur Erhebung, Aufbereitung und Auswertung der Mitschriften sind Kapitel 5 zu entnehmen.

Daten zu den Metadaten	Details
Anzahl der Dokumente	2 (Sprechereignisse; Mitschriften)
Dateiformat	csv
Auszeichnungssprache	UTF-8

Tabelle 7: Technische Daten zu den Metadaten

Weitere Zusatzmaterialien stehen in der DGD bzw. im LZA zur Verfügung. Diese sind in Kapitel 1.3 gelistet.

2.1.4 Sprechereignisübersicht

Sprechereignis ID	Vorlesungssitzung	Transkript (exb, txt)	Audio (wav)	Video (mp4)	Anzahl Mitschriften (PDF/A)
MIKO_E_00001	Funktionelle Anatomie 1	ja	ja	nein	16
MIKO_E_00002	Funktionelle Anatomie 2	ja	ja	nein	18
MIKO_E_00003	Funktionelle Anatomie 3	ja	ja	nein	19
MIKO_E_00004	Physik für Mediziner 1	ja	ja	ja	18
MIKO_E_00005	Physik für Mediziner 2	ja	ja	ja	16
MIKO_E_00006	Physik für Mediziner 3	ja	ja	ja	17
MIKO_E_00007	Bürgerliches Recht für Wirtschaftswissenschaftler 1	ja	ja	ja	9
MIKO_E_00008	Bürgerliches Recht für Wirtschaftswissenschaftler 2	nein	ja	ja	3
MIKO_E_00009	Bürgerliches Recht für Wirtschaftswissenschaftler 3	nein	ja	ja	5
MIKO_E_00010	Grundzüge der Lexikologie der deutschen Gegenwartssprache 1	nein	ja	ja	8
MIKO_E_00011	Grundzüge der Lexikologie der deutschen Gegenwartssprache 2	nein	ja	ja	7
MIKO_E_00012	Grundzüge der Lexikologie der deutschen Gegenwartssprache 3	ja	ja	ja	10
Talvanzahl assamt	. 02 075				

Tokenzahl gesamt: 82 075

Typezahl gesamt: 6076

Dauer der Audioaufnahmen: 15:38 h / 938 Min (davon 10:00 h / 600 Min transkribiert und annotiert)

Dauer der Videoaufnahmen: 12:50 h / 770 Min (davon 7:12 h / 432 Min transkribiert und annotiert)

Tabelle 8: Sprechereignisübersicht

3 Metadaten

Es liegen umfangreiche Metadaten zu den Sprechereignissen und den Mitschriften vor. Für die Veröffentlichung in der DGD wurden die für die Plattform relevanten Metadaten zu den Sprechereignissen im **EXMARaLDA** Corpus Manager (COMA) verwaltet. Die Mitschriftenmetadaten stehen in der DGD zusätzlich als csv-Tabelle zur Verfügung. Die vollständigen Metadaten zu den Sprechereignissen und den Mitschriften liegen im LZA als csv-Tabellen vor. Als Variablenbenennungskonvention wurde eine Kombination aus snake_case und camelCase gewählt, d. h. Bezeichnungen, die verschiedene Variablenbenennungsebenen ansprechen, werden durch Unterstriche abgetrennt; Bezeichnungen, die Variablengruppen veranschaulichen, durch Großschreibung.

3.1 Sprechereignisse

Die korpus-, vorlesungs- und transkriptbezogenen Metadaten orientieren sich an Granger & Paquot (2017) und wurden um projekt-, korpus- und DGD-spezifische Variablen ergänzt.

Folgende Metadaten zu den Sprechereignissen liegen vor:

- Erstellungsprojekt (Forschungsprojekt "Sprache und Studienerfolg", in dem das Korpus entstanden ist),
- Korpus (Administratives, Korpusdesign),
- Sprechereignisse (Vorlesungssitzungen, verwendete Medien, Themenliste, zugehörige Audio- und Videodateien, Auditorium),
- Transkripte (Token, Types, Annotationen, Transkribent/-in, Reviewer/-in) und
- Sprecher/-innen (Dozierende/-r, Assistent).

Ebene	Variable	Erläuterung
Erstellungsprojekt	project_name	Name des Erstellungsprojekts
Erstellungsprojekt	project_place	Ort des Erstellungsprojekts
Erstellungsprojekt	project_country	Land des Erstellungsprojekts
Erstellungsprojekt	project_institution	Einrichtung des Erstellungsprojekts
Erstellungsprojekt	project_type	Art des Erstellungsprojekts
Erstellungsprojekt	project_head	Leiter/-in des Erstellungsprojekts
Erstellungsprojekt	project_contact	Kontakt des Erstellungsprojekts
Erstellungsprojekt	project_duration	Laufzeit des Erstellungsprojekts
Erstellungsprojekt	project_publication	Publikation des Erstellungsprojekts
Korpus (Administrativ)	corpus_title	Titel des Korpus
Korpus (Administrativ)	corpus_acronym	Akronym des Korpus
Korpus (Administrativ)	corpus_head	Verantwortliche des Korpus
Korpus (Administrativ)	corpus_contributors	Mitwirkende am Korpus
Korpus (Administrativ)	corpus_contact	Kontakt des Korpus
Korpus (Administrativ)	corpus_distributor	Einrichtung des Korpus
Korpus (Administrativ)	corpus_availability	Zugang zum Korpus
Korpus (Administrativ)	corpus_license	Lizenz des Korpus
Korpus (Administrativ)	corpus_characterEncoding	Zeichenkodierung
Korpus (Administrativ)	corpus_markupLanguage	Auszeichnungssprache

Ebene	Variable	Erläuterung
Korpus (Design)	corpus_language	Sprache des Korpus
Korpus (Design)	corpus_sizeTranscripts	Anzahl der Transkripte
Korpus (Design)	corpus_sizeToken	Anzahl der Token
Korpus (Design)	corpus_sizeTypes	Anzahl der Types
Korpus (Design)	corpus_mode	Modalität des Korpus
Korpus (Design)	corpus_textType	Textart
Korpus (Design)	corpus_transcriptionGuidelines	Transkriptionsrichtlinien
Korpus (Design)	corpus_transcriptionConvention	Transkriptionssystem
Korpus (Design)	corpus_longitudinal	Längsschnitt
Korpus (Design)	corpus_field	Sprachliche Domäne
Korpus (Design)	corpus_compData	Erhebung weiterer Daten als Kontrollgruppe
Korpus (Design)	corpus_annotation	Wurden die Texte annotiert?
Korpus (Design)	corpus_annotationComplete	Ist das gesamte Korpus annotiert?
Korpus (Design)	corpus_posTagged	POS-Tagging des Korpus
Korpus (Design)	corpus_posTagset	POS-Tagset des Korpus
Sprechereignis	communication_id	ID des Sprechereignisses
Sprechereignis	communication_interactionType	Diskursart
Sprechereignis	communication_soundID	Name der Audiodatei
Sprechereignis	communication_soundDuration	Länge der Audiodatei

Ebene	Variable	Erläuterung
Sprechereignis	communication_soundProperties	Format und Eigenschaften der Audiodatei
Sprechereignis	communication_videoID	Name der Videodatei
Sprechereignis	communication_videoDuration	Länge der Videodatei
Sprechereignis	communication_videoProperties	Format und Eigenschaften der Videodatei
Sprechereignis	communication_subject	Studienfach
Sprechereignis	communication_lecture	Name der Vorlesung
Sprechereignis	communication_media	verwendete Medien (PPT, Overhead etc.)
Sprechereignis	communication_semester	Fachsemester der Vorlesung nach Studienverlaufsplan
Sprechereignis	communication_audience	Anzahl der anwesenden Studierenden
Sprechereignis	communication_topicKeywords	Themenliste der Sitzung
Sprechereignis	communication_date	Datum der Erhebung
Sprechereignis	communication_place	Ort der Erhebung
Sprechereignis	communication_institution	Einrichtung der Erhebung
Sprechereignis	communication_sizeNotes	Anzahl der Mitschriften
Transkript	transcript_id	ID des Transkripts
Transkript	transcript_transcriberID	ID der Transkribentin/des Transkribenten
Transkript	transcript_reviewerID	ID der Reviewerin/des Reviewers
Transkript	transcript_token	Anzahl der Token des Transkripts
Transkript	transcript_types	Anzahl der Types des Transkripts

Ebene	Variable	Erläuterung
Transkript	transcript_posTagged	POS-Tagging des Transkripts
Transkript	transcript_posTagset	POS-Tagset des Transkripts
Transkript	transcript_annotationOther	Vorhandensein weiterer Annotationen
Sprecher/-in	speaker1_id	ID der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker1_pseudonym	Pseudonym der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker1_acronym	Akronym der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker1_sex	Geschlecht der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker1_age	Alter der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker1_profession	Beruf der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker1_l1	L1 der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker1_place	Herkunftsregion der ersten Sprecherin/des ersten Sprechers
Sprecher/-in	speaker2_id	ID der zweiten Sprecherin/des zweiten Sprechers
Sprecher/-in	speaker2_pseudonym	Pseudonym der zweiten Sprecherin/des zweiten Sprechers
Sprecher/-in	speaker2_acronym	Akronym der zweiten Sprecherin/des zweiten Sprechers
Sprecher/-in	speaker2_sex	Geschlecht der zweiten Sprecherin/des zweiten Sprechers
Sprecher/-in	speaker2_age	Alter der zweiten Sprecherin/des zweiten Sprechers
Sprecher/-in	speaker2_profession	Beruf der zweiten Sprecherin/des zweiten Sprechers
Sprecher/-in	speaker2_l1	L1 der zweiten Sprecherin/des zweiten Sprechers
Sprecher/-in	speaker2_place	Herkunftsregion der zweiten Sprecherin/des zweiten Sprechers

Tabelle 9: Metadatenvariablen der Sprechereignisse

3.2 Mitschriften

Mitschriftenmetadaten Variablen Die beinhalten aus verschiedenen Quellen: Sprechereignismetadaten, Fragebögen, die die Studierenden zu den jeweiligen Vorlesungssitzungen ausgefüllt haben, Mitschriftenerfassungen und -bewertungen (Kapitel 5.3) sowie ein Fragebogen zu Vorbedingungen des Studiums. Im Forschungsprojekt SpraStu wurden darüber hinaus umfangreiche Sprachkompetenzerhebungen der Proband/-innen mit einer Reihe standardisierter, auf den Gemeinsamen europäischen Referenzrahmen für Sprachen (Europarat, 2001) bezogenen Verfahren (onSET: g.a.s.t., 2017; TestDaF-Schreibaufgabe: https://www.testdaf.de/; ACTFL-Lesetest: ACTFL, 2013; ACTFL-Hörtest: ACTFL, 2014) sowie einem produktiven und einem rezeptiven Wortschatztest (Institut für Testforschung und Testentwicklung e. V.; https://ittleipzig.de/) erhoben. Die Ergebnisse dieser Sprachtests zum Zeitpunkt der Vorlesungs- und Mitschriftenerhebung (Oktober bis Dezember 2017) stehen ebenfalls in den Metadaten zur Verfügung.

Folgende Metadaten zu den Mitschriften liegen vor:

- Vorlesungssitzung (Sprechereignis, empfundene sprachliche und inhaltliche Schwierigkeit, wichtige Themen aus Sicht der Studierenden³),
- Mitschriften (Deskriptiva, Bewertungen, Nützlichkeit der Mitschrift, Mitschreibabsichten) und
- Autor/-innen (Studium, Sprachlernbiografie, Sprachtestergebnisse).

_

³ Freitextangaben der Studierenden wurden nicht orthografisch korrigiert.

Ebene	Variable	Erläuterung
Vorlesungssitzung	communication_id	ID des Sprechereignisses
Vorlesungssitzung	communication_lecture	Name der Vorlesung
Vorlesungssitzung	communication_lecture_number	Nummer der Vorlesungssitzung
Vorlesungssitzung	communication_subject	Studienfach
Vorlesungssitzung	communication_media	verwendete Medien (PPT, Overhead etc.)
Vorlesungssitzung	communication_semester	Fachsemester
Vorlesungssitzung	communication_audience	Anzahl der anwesenden Studierenden
Vorlesungssitzung	communication_topicKeywords	Themenliste der Sitzung
Vorlesungssitzung	communication_date	Datum der Erhebung
Vorlesungssitzung	communication_place	Ort der Erhebung
Vorlesungssitzung	communication_institution	Einrichtung der Erhebung
Vorlesungssitzung	communication_sizeNotes	Anzahl der Mitschriften
Vorlesungssitzung	communication_speed	sprachliche Schwierigkeit: Geschwindigkeit (4-stufige Skala: leicht, eher leicht, eher schwer, schwer)
Vorlesungssitzung	communication_pronunciation	sprachliche Schwierigkeit: Aussprache/Akzent (4-stufige Skala: leicht, eher leicht, eher schwer, schwer)
Vorlesungssitzung	communication_vocabulary	sprachliche Schwierigkeit: Wörter (4-stufige Skala: leicht, eher leicht, eher schwer, schwer)
Vorlesungssitzung	communication_grammar	sprachliche Schwierigkeit: Grammatik (4-stufige Skala: leicht, eher leicht, eher schwer, schwer)

Ebene	Variable	Erläuterung
Vorlesungssitzung	communication_handwriting	sprachliche Schwierigkeit: Handschrift des Dozierenden (4-stufige Skala: leicht, eher leicht, eher schwer, schwer)
Vorlesungssitzung	communication_content_difficulty	inhaltliche Schwierigkeit insgesamt (4-stufige Skala: leicht, eher leicht, eher schwer, schwer)
Vorlesungssitzung	communication_content_previousKnowledge	Vorwissen zum Thema der Vorlesungssitzung (4-stufige Skala: sehr viel, viel, fast nichts, nichts)
Vorlesungssitzung	communication_content_comprehension	Verständnis des Inhalts der Vorlesungssitzung (4-stufige Skala: gut, eher gut, eher nicht gut, überhaupt nicht gut)
Vorlesungssitzung	communication_topics_notUnderstood	Themen, die überhaupt nicht verstanden wurden
Vorlesungssitzung	communication_topic1	wichtige Themen aus Sicht der Studierenden: Thema 1
Vorlesungssitzung	communication_topic1_comprehension	Verständnis von Thema 1 (4-stufige Skala: gut verstanden, eher gut verstanden, eher nicht gut verstanden, überhaupt nicht gut verstanden)
Vorlesungssitzung	communication_topic2	wichtige Themen aus Sicht der Studierenden: Thema 2
Vorlesungssitzung	communication_topic2_comprehension	Verständnis von Thema 2 (4-stufige Skala: gut verstanden, eher gut verstanden, eher nicht gut verstanden, überhaupt nicht gut verstanden)
Vorlesungssitzung	communication_topic3	wichtige Themen aus Sicht der Studierenden: Thema 3
Vorlesungssitzung	communication_topic3_comprehension	Verständnis von Thema 3 (4-stufige Skala: gut verstanden, eher gut verstanden, eher nicht gut verstanden, überhaupt nicht gut verstanden)
Mitschrift	text_id	ID der Mitschrift
Mitschrift	text_token	Tokenzahl: lexikalische Wörter, Teilwörter, Abkürzungen sowie genormte (alphabetische) Kürzel und einzelne Buchstaben

Ebene	Variable	Erläuterung
Mitschrift	text_token_notDeu	Tokenzahl nicht Deutsch (und nicht fachsprachlich)
Mitschrift	text_groupOfCharacters	Anzahl größerer Symbolgruppen mit mindestens drei Elementen mindestens zweier verschiedener Zeichenklassen
Mitschrift	text_numbers	Anzahl von Zahlzeichen, Ziffern, auch römische Zahlen außerhalb von Zeichengruppen
Mitschrift	text_characters	Anzahl typografischer Zeichen, mathematischer Zeichen und Sonderzeichen außerhalb von Zeichengruppen oder Grafiken
Mitschrift	text_structuralElements	Anzahl von Aufzählungszeichen, Nummerierungen (numerisch, alphabetisch) und andere Zeichen, die grafisch eine unter- oder nebengeordnete Beziehung zwischen zwei Elementen ausdrücken
Mitschrift	text_ownTables	über die Medien der/des Dozierenden hinaus erstellte eigene Tabellen
Mitschrift	text_ownPictures	über die Medien der/des Dozierenden hinaus gezeichnete eigene Bilder
Mitschrift	text_otherGraphics	über die Medien der/des Dozierenden hinaus erstellte sonstige Grafiken
Mitschrift	text_medium	Medium, auf dem die Mitschrift verfasst wurde (Skript, Zettel)
Mitschrift	text_modus	Modus der Mitschrift (handschriftlich, computergestützt)
Mitschrift	text_completeness_total_max_abs	inhaltliche Vollständigkeit: max. Punkte gesamt (absolut)
Mitschrift	text_completeness_total_achieved_abs	inhaltliche Vollständigkeit: erreichte Punkte gesamt (absolut)
Mitschrift	text_completeness_total_achieved_rel	inhaltliche Vollständigkeit: erreichte Punkte gesamt (relativ)

Ebene	Variable	Erläuterung
Mitschrift	text_completeness_total_notRatable_abs	inhaltliche Vollständigkeit: Punkte nicht bewertbar gesamt (absolut)
Mitschrift	text_completeness_total_notRatable_rel	inhaltliche Vollständigkeit: Punkte nicht bewertbar gesamt (relativ)
Mitschrift	text_completeness_level1_max_abs	inhaltliche Vollständigkeit: Ebene 1 max. Punkte gesamt (absolut)
Mitschrift	text_completeness_level1_achieved_abs	inhaltliche Vollständigkeit: Ebene 1 erreichte Punkte gesamt (absolut)
Mitschrift	text_completeness_level1_achieved_rel	inhaltliche Vollständigkeit: Ebene 1 erreichte Punkte gesamt (relativ)
Mitschrift	text_completeness_level1_notRatable_abs	inhaltliche Vollständigkeit: Ebene 1 Punkte nicht bewertbar gesamt (absolut)
Mitschrift	text_completeness_level1_notRatable_rel	inhaltliche Vollständigkeit: Ebene 1 Punkte nicht bewertbar gesamt (relativ)
Mitschrift	text_completeness_level2_max_abs	inhaltliche Vollständigkeit: Ebene 2 max. Punkte gesamt (absolut)
Mitschrift	text_completeness_level2_achieved_abs	inhaltliche Vollständigkeit: Ebene 2 erreichte Punkte gesamt (absolut)
Mitschrift	text_completeness_level2_achieved_rel	inhaltliche Vollständigkeit: Ebene 2 erreichte Punkte gesamt (relativ)
Mitschrift	text_completeness_level2_notRatable_abs	inhaltliche Vollständigkeit: Ebene 2 Punkte nicht bewertbar gesamt (absolut)
Mitschrift	text_completeness_level2_notRatable_rel	inhaltliche Vollständigkeit: Ebene 2 Punkte nicht bewertbar gesamt (relativ)
Mitschrift	text_completeness_level3_max_abs	inhaltliche Vollständigkeit: Ebene 3 max. Punkte gesamt (absolut)
Mitschrift	text_completeness_level3_achieved_abs	inhaltliche Vollständigkeit: Ebene 3 erreichte Punkte gesamt (absolut)
Mitschrift	text_completeness_level3_achieved_rel	inhaltliche Vollständigkeit: Ebene 3 erreichte Punkte gesamt (relativ)
Mitschrift	text_completeness_level3_notRatable_abs	inhaltliche Vollständigkeit: Ebene 3 Punkte nicht bewertbar gesamt (absolut)

Ebene	Variable	Erläuterung
Mitschrift	text_completeness_level3_notRatable_rel	inhaltliche Vollständigkeit: Ebene 3 Punkte nicht bewertbar gesamt (relativ)
Mitschrift	text_correctness	Bewertung der Korrektheit des Mitgeschriebenen
Mitschrift	text_structure	Bewertung der Organisation und Nachvollziehbarkeit des Mitgeschriebenen
Mitschrift	text_usefulness_memory	Nützlichkeit der Mitschrift: um sich später besser an die Vorlesung zu erinnern (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_usefulness_examPreparation	Nützlichkeit der Mitschrift: um sich später auf die Klausur vorzubereiten (4- stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_usefulness_concentration	Nützlichkeit der Mitschrift: um sich in der Vorlesung weiter zu konzentrieren (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_usefulness_boredom	Nützlichkeit der Mitschrift: um sich nicht zu langweilen (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_usefulness_other	Nützlichkeit der Mitschrift: Sonstiges
Mitschrift	text_purpose_unknown	Mitschreibabsicht: was man noch nicht wusste (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_known	Mitschreibabsicht: was man schon wusste (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_importance	Mitschreibabsicht: was der/dem Studierenden persönlich am wichtigsten erschien (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)

Ebene	Variable	Erläuterung
Mitschrift	text_purpose_exam	Mitschreibabsicht: was für die Klausur wichtig ist (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_maximum	Mitschreibabsicht: so viel wie die/der Studierende konnte (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_lecturersOpinion	Mitschreibabsicht: Meinungen der/des Dozierenden (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_ownOpinion	Mitschreibabsicht: eigene Meinung (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_facts	Mitschreibabsicht: nur Fakten (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_media	Mitschreibabsicht: was auf den Folien oder auf der Tafel stand (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_noMedia	Mitschreibabsicht: was nicht auf den Folien oder der Tafel stand (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_emphasis	Mitschreibabsicht: was die/der Dozierende besonders betont hat (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_comprehension	Mitschreibabsicht: was nicht verstanden wurde (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)

Ebene	Variable	Erläuterung
Mitschrift	text_purpose_bodyLanguage	Mitschreibabsicht: was laut Körpersprache der/des Dozierenden wichtig sein musste (4-stufige Skala: trifft voll zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu)
Mitschrift	text_purpose_other	Mitschreibabsicht: Sonstiges
Autor/-in	author_ID	ID der/des Studierenden
Autor/-in	author_deu_l1_l2	Deutsch als L1 oder L2
Autor/-in	author_uniQualificationPlace_group	Land der Hochschulzulassungsberechtigung nach Herkunftsregionen (ISCED; http://uis.unesco.org/en/isced-mappings)
Autor/-in	author_uniEntranceLanguageTest	Hochschulzulassungssprachtest
Autor/-in	author_majorSubject	Hauptfach
Autor/-in	author_semester	Fachsemester des Hauptfaches
Autor/-in	author_age	Alter
Autor/-in	author_sex	Geschlecht
Autor/-in	author_l1_1	erste L1
Autor/-in	author_l1_2	zweite L1
Autor/-in	author_11_3	dritte L1
Autor/-in	author_l1_group	L1 nach Sprachfamilien
Autor/-in	author_deu_stayMonths	alle Aufenthalte in Deutschland vor dem Studium (in Monaten)
Autor/-in	author_degree	Höchster Bildungsabschluss
Autor/-in	author_studyExperience	Studienerfahrung

Ebene	Variable	Erläuterung
Autor/-in	author_deu_studyExperience	Studienerfahrung in Deutschland
Autor/-in	author_Studienkolleg	Besuch des Studienkollegs
Autor/-in	author_lCount	Anzahl gesprochener Sprachen inklusive der L1
Autor/-in	author_deu_selfAssessment	Selbsteinschätzung der Deutschkenntnisse
Autor/-in	author_deu_ageOfOnset_group	Erwerbsbeginn Deutsch gruppiert
Autor/-in	author_deu_duration	Anzahl der Jahre des Deutschunterrichts
Autor/-in	author_irregDefinition	Beschreibung der Irregularität des Studienverlaufs der/des Studierenden
Autor/-in	author_onSET_level	onSET-Stufe
Autor/-in	author_TestDaF_level	Bewertung des TestDaF: gesamt
Autor/-in	author_TestDaF_Lesefluss	Bewertung des TestDaF: Lesefluss
Autor/-in	author_TestDaF_Gedankengang	Bewertung des TestDaF: Gedankengang
Autor/-in	author_TestDaF_Textstruktur	Bewertung des TestDaF: Textstruktur
Autor/-in	author_TestDaF_PunkteAufgabenstellung	Bewertung des TestDaF: Aufgabenstellung
Autor/-in	author_TestDaF_BeschreibungGrafik	Bewertung des TestDaF: Grafik
Autor/-in	author_TestDaF_Argumentation	Bewertung des TestDaF: Argumentation
Autor/-in	author_TestDaF_SprachlicheMittel	Bewertung des TestDaF: Sprachliche Mittel
Autor/-in	author_TestDaF_Wortschatz	Bewertung des TestDaF: Wortschatz
Autor/-in	author_TestDaF_Korrektheit	Bewertung des TestDaF: Korrektheit

Ebene	Variable	Erläuterung
Autor/-in	author_LPT_level	Bewertung des Listening Proficiency Test (GER-Niveau)
Autor/-in	author_LPT_sumscore	Bewertung des Listening Proficiency Test (Summenscore)
Autor/-in	author_RPT_level	Bewertung des Reading Proficiency Test (GER-Niveau)
Autor/-in	author_RPT_sumscore	Bewertung des Reading Proficiency Test (Summenscore)
Autor/-in	author_vocabRec_percent	Bewertung des rezeptiven Wortschatztests (in Prozent)
Autor/-in	author_vocabProd_percent	Bewertung des produktiven Wortschatztests (in Prozent)

Tabelle 10: Metadaten der Mitschriften

4 Vorlesungen

4.1 Datenerhebung

In den Kernfächern des Projekts SpraStu wurden Vorlesungen zur Aufnahme ausgewählt, die möglichst sowohl eine Pflichtvorlesung für alle Studierenden des ersten Fachsemesters sind als auch mit einer Klausur abgeschlossen werden. Somit wurden – nach Absprache mit den zuständigen Dozierenden – folgende Vorlesungen aufgenommen:

- Medizin: Funktionelle Anatomie I (Ana), Physik für Mediziner (Phy)
- Wirtschaftswissenschaften: Bürgerliches Recht für Wirtschaftswissenschaftler (BGB)
- Deutsch als Fremd- und Zweitsprache: Grundzüge der Lexikologie der deutschen Gegenwartssprache (Lex)

In Anlehnung an das euroWiss-Projekt (Heller et al., 2013) wurden von jeder Vorlesung möglichst drei aufeinanderfolgende Vorlesungssitzungen aufgezeichnet, um ein repräsentatives Bild der Vorlesungen zu erhalten. Die Nummerierung der Vorlesungssitzungen ergibt sich aus der Reihenfolge der Aufnahme.

Jede Vorlesung wurde mit zwei Kameras aufgezeichnet, sodass parallel die Projektion bzw. das Tafelbild und die/der Dozierende gefilmt werden konnten. Die BGB-Vorlesungen fanden in einem Hörsaal statt, in dem Aufnahmetechnik des E-Learning-Service der Universität Leipzig fest installiert war, die das Projekt für die Aufnahme der Projektion bzw. des Tafelbildes und die Audioaufnahme (Mikrofon der/des Dozierenden) nutzen konnte. Zusätzlich wurde mit einer Handkamera auf Stativ die/der Dozierende gefilmt. In den drei anderen Vorlesungen kamen jeweils zwei Handkameras zum Einsatz; eine davon war mit einem Empfänger für ein Funkmikrofon ausgestattet, das der/dem Dozierenden angeheftet wurde und dessen Tonaufnahme für das endgültige Video verwendet wurde.

Da es sich bei den Vorlesungsaufzeichnungen um "Aufnahmen unter nicht vollständig kontrollierten Bedingungen" (Schmidt et al., 2013) handelt, weisen die Daten Qualitätsunterschiede auf. Eine technische Herausforderung bei der Vorlesungsaufzeichnung bestand unter anderem in der Enge der Hörsäle in der Medizin, wegen der für eine Kamera ein Tisch- statt eines Standstativs eingesetzt wurde. Dies hatte jedoch beim Schwenken der Kamera ein Ruckeln in der Aufnahme der Projektion bzw. des Tafelbildes zur Folge. Außerdem resultierte das fehlerhafte Anheften des Funkmikros am Dozenten in der ersten Physikvorlesungssitzung in knackenden Störgeräuschen bei der Audioaufnahme. Der Assistent in den Physikvorlesungssitzungen erhielt kein eigenes Mikrofon, sodass dessen kurze Redebeiträge nur über das Mikrofon des Dozenten erfasst wurden.

Die vom E-Learning-Service aufgenommenen BGB-Vorlesungen mussten hinsichtlich der Qualität der Tonaufnahmen stark nachbearbeitet werden; zwar war die Sprache sehr gut zu verstehen, trat allerdings eine Pause ein, war ein intensives Rauschen zu vernehmen. Außerdem nahm der Service in der ersten BGB-Vorlesung aufgrund ablauftechnischer Probleme die ersten 11 Minuten und 10 Sekunden nicht auf. Aus diesem Grund stehen für diesen Vorlesungsabschnitt

nur die Videoaufnahmen des Dozenten und der Projektion zur Verfügung. Das Transkript wurde anhand der Audiospur der Handkamera angefertigt und liegt daher vollständig vor.

Bei der Aufzeichnung der Vorlesungen waren jeweils mindestens zwei Mitarbeiter-/innen anwesend, die die Kameras bedienten. Die technische Ausstattung wurde durch das Zentrum für Medientechnik und Kommunikation (ZMK) der Universität Leipzig (Kamera, Stativ, Funkmikro) und durch das Herder-Institut (Kamera) bereitgestellt.

4.2 Datenaufbereitung

4.2.1 Schneiden der Videos

Als Ausgangsmaterial standen Video- und Audioaufnahmen der/des Dozierenden sowie Video- und Audioaufnahmen der Projektion bzw. des Tafelbildes zur Verfügung. Die beiden Videos wurden jeweils so geschnitten, dass sie die bestmögliche Audioqualität haben und die Aufnahme der Projektion bzw. des Tafelbildes und die Nahaufnahme der/des Dozierenden nebeneinander zu sehen sind. Somit sind auf den Videos alle nonverbalen Handlungen der/des Dozierenden, einschließlich Zeigegesten auf Projektion und Tafelbild, sichtbar.



Abbildung 1: Videoausschnitt MIKO_E_00007, 60:18

Das ZMK der Universität Leipzig unterstützte maßgeblich bei Videoschnitt und -bearbeitung mit Adobe Premiere Pro.

4.2.2 Anonymisierung der Videos

Vor der Veröffentlichung im Korpus wurden umfangreiche Nachbearbeitungen der Vorlesungsvideos vorgenommen, um die Persönlichkeitsrechte von Studierenden und dritten, in

den Vorlesungen erwähnten Personen zu gewährleisten und durch Anonymisierung eine nachträgliche Identifizierung zu verhindern bzw. erheblich zu erschweren.

Die Nachbearbeitung erfolgte ebenfalls mit Adobe Premiere Pro, hier wurden Weichzeichnungsmasken angewendet und Audioabschnitte durch Rauschen ersetzt. Studierende, bei denen das Gesicht komplett/im Profil oder einzelne Teile des Kopfes (Nase, Stirn, Ohr, Hals) sichtbar sind, wurden verpixelt. Auch ausgefallene Haarschnitte, Frisuren, Kleidung oder Accessoires, die einen Rückschluss auf die Person zulassen könnten, sowie Hände, Beine, Füße und Schuhe wurden weichgezeichnet. In Ausnahmefällen wurde die Videoaufzeichnung der/des Dozierenden mit einer schwarzen Farbfläche überblendet, wenn entweder zu viele Studierende eindeutig im Bild zu sehen waren (z. B. bei MIKO_E_00007) oder urheberrechtlich geschütztes Material gezeigt wurde (z. B. bei MIKO_E_00006).

Darüber hinaus wurden verbale Äußerungen einzelner Studierender aus dem Auditorium durch 1/f-Rauschen (rosa Rauschen) in der Audiospur unkenntlich gemacht. Wenn ein/-e Studierende/-r identifizierbar gleichzeitig mit der/dem Dozierenden spricht, wurde je nach Lautstärke des Redebeitrags entweder die Verbalspur der/des Dozierenden komplett verrauscht oder lediglich ein Rauschen darüber gelegt. Husten, Niesen oder allgemeine Geräuschkulissen aus dem Plenum wurden nicht nachbearbeitet. Personennamen von Diskursbeteiligten sowie sensible Informationen wurden ebenfalls anonymisiert (Kapitel 8.3.9).

4.3 Transkriptions- und Annotationsprozess und Qualitätssicherung

Nach der technischen Einarbeitung in EXMARaLDA (Schmidt & Wörner, 2014) sowie dem Besuch einer EXMARaLDA-Schulung am Hamburger Zentrum für Sprachkorpora wurden die Transkriptions- und Annotationsrichtlinien ausgehend von den HIAT-Richtlinien (Rehbein et al., 2004) festgelegt. Diese wurden im Laufe des Transkriptionsprozesses stetig angepasst und liegen in ihrer finalen Fassung in diesem Handbuch vor (ab Kapitel 8).

Der Transkriptions- und Reviewprozess (2018 bis 2020) gliederte sich in mehrere Phasen, die im Folgenden genauer beschrieben werden. Während des gesamten Prozesses waren stets dieselben zwei Transkribent/-innen beteiligt.

4.3.1 Phase 1: Entwicklung eines Goldstandards

Die erste Phase hatte die Etablierung eines Goldstandards zum Ziel. Ausgehend vom ersten Entwurf der Transkriptions- und Annotationsrichtlinien wurden 20 Minuten der Vorlesungssitzung Funktionelle Anatomie 1 unabhängig von den beiden Transkribent/-innen transkribiert und annotiert. Die Transkripte wurden verglichen, Abweichungen und Zweifelsfälle dokumentiert, diskutiert und Entscheidungen getroffen. Das daraus entstandene Transkript und die angepassten Richtlinien dienten der Festlegung eines Standardverständnisses für die weitere Transkriptions- und Annotationsarbeit. Anschließend wurde ein zyklisches Vorgehen verfolgt.

4.3.2 Phase 2: Transkriptions- und Annotationsphase

Je ein/-e Transkribent/-in war hauptverantwortlich für die Transkription und Annotation einer Vorlesungssitzung zuständig und führte diese eigenständig durch. Fach- und vorlesungsspezifische Phänomene, wie z.B. Formeln und die Durchführung von Experimenten in den Physikvorlesungen sowie der indirekte Umgang mit PowerPoint-Folien in der Lexikologie- und BGB-Vorlesung, wurden prozessbegleitend in den Richtlinien ergänzt. Die Annotationsspur *Text-Bild-Bezüge* wurde erst zu einem späteren Zeitpunkt entwickelt und daher zeitversetzt unter stetiger Qualitätskontrolle in die Transkripte eingefügt.

4.3.3 Phase 3: Prüfung der Transkripte durch die Transkribentin/den Transkribenten

Im Anschluss an die Transkription und Annotation einer Vorlesungssitzung führte die/der Transkribent/-in zunächst eine Überprüfung des eigenen Transkripts auf verschiedenen Ebenen durch:

- Prüfen der Transkription (Verbalspur inkl. Pausen),
- Erneutes Anhören von unverständlichen Sequenzen und
- Suche nach Struktur- und Segmentierungsfehlern (im Partitur-Editor über *Transcription* > *Structure Errors* bzw. *Segmentation Errors*) und Behebung derselben.

4.3.4 Phase 4: Reviewphase

Die/der jeweils andere Transkribent/-in übernahm das Review. Für den Reviewprozess wurde mit zwei Hilfsspuren gearbeitet, die in den finalen Transkripten nicht mehr vorhanden sind. Korrekturen, die ohne Rücksprache mit der Transkribentin/dem Transkribenten im Transkript vorgenommen werden konnten, sind:

- Korrektur des Alignments in allen Spuren,
- Ökonomisieren von Eventgrenzen,
- Rechtschreibung und Kommasetzung,
- Pausen,
- Betonungen,
- Sprechweise,
- Zeitangaben (Nonverbales, Auditoriumsbeiträge) und
- fehlende/fehlerhafte Text-Bild-Bezüge.

Korrekturvorschläge, die einer Rücksprache bedurften, sind:

- Fehlen ganzer Wörter,
- Äußerungsendzeichen (Äußerungsmodi),
- Änderungen an der Syntax,
- Anmerkungen der Transkribentin/des Transkribenten zu Unsicherheiten im Transkriptionsprozess und
- Zweifelsfälle in allen Spuren.

In wöchentlichen Besprechungen wurden das Review und alle Änderungsvorschläge diskutiert, Zweifelsfälle geklärt und getroffene Entscheidungen dokumentiert. Bei bestehender Uneinigkeit wurde eine dritte Person hinzugezogen.

4.3.5 Reliabilität der Transkription und Annotation

Für sechs der acht Transkripte wurden während des Reviews die Abweichungen nach Spuren tabellarisch erfasst. Der daraus ermittelte durchschnittliche Fehlerquotient liegt bei 0,99% der Gesamttoken. Die niedrigste Reliabilität wurde bei der Annotation der Betonungen erreicht. 20% der erfassten Abweichungen wurden in dieser Spur ermittelt, was darauf zurückzuführen ist, dass die nur auditiv gestützte Annotation sehr subjektiv ist.

4.3.6 Automatische Annotationen

Neben den manuellen Transkriptions- und Annotationsspuren wurden nach Abschluss der Reviewphase morphosyntaktische Annotationen auf Wortebene automatisch hinzugefügt. Dafür wurden die Verbalspuren zunächst im COMA automatisch segmentiert und anschließend mithilfe des TreeTaggers (https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/; Schmid, 1994), der das Stuttgart-Tübingen-TagSet (STTS) (Schiller et al., 1999) verwendet, automatisch lemmatisiert und mit POS-Tagging annotiert. Die Tokenisierung, Lemmatisierung und das POS-Tagging wurden nicht manuell überprüft. Die Transkripte liegen im LZA jeweils in tokenisierter (mit automatischen Annotationen) und untokenisierter Form (ohne automatische Annotationen) vor (Kapitel 1.3.2).

5 Mitschriften

Zusätzlich zu den Vorlesungstranskripten stehen in MIKO studentische Mitschriften zur Verfügung (der Zugang zu den Mitschriften ist in Kapitel 1.3.3 dokumentiert). Im Folgenden wird erläutert, wie diese erhoben, aufbereitet und ausgewertet wurden.

5.1 Datenerhebung

5.1.1 Einsammeln der Vorlesungsmitschriften

Die Mitschriften der SpraStu-Teilnehmenden in den Vorlesungen wurden in unmittelbarem Anschluss an das Vorlesungsende von den anwesenden Mitarbeiter-/innen mit Digitalkameras abfotografiert, wobei Karten mit den SpraStu-IDs der Teilnehmenden auf die Mitschriften gelegt wurden, um eine spätere eindeutige Zuordnung zu gewährleisten. Jegliche personenbezogenen Hinweise auf die Studierenden (einschließlich der SpraStu-IDs) wurden bei der weiteren Verarbeitung geschwärzt und durch MIKO-IDs ersetzt (Kapitel 5.2).

Die Mitschriften wurden teilweise auf uneinheitliche Weise fotografiert, sodass sich Unterschiede in der Qualität der Aufnahmen ergeben. Ein weiteres Problem stellte der große Zeitdruck dar, unter dem die Mitschriften fotografiert werden mussten, da die Studierenden zum Teil nur wenige Minuten zum Wechseln des Hörsaals zur Verfügung hatten und eine unmittelbare Qualitätskontrolle und gegebenenfalls -korrektur der Aufnahmen daher nicht möglich war.

5.1.2 Fragebogen zu den Mitschriften

Die Studierenden füllten zu jeder Vorlesungssitzung, in der Mitschriften eingesammelt wurden, einen kurzen Fragebogen aus. Dieser konnte entweder in Papierform direkt im Anschluss an die Vorlesungssitzung oder online ausgefüllt werden. Der Link zum Fragebogen wurde an demselben oder am Folgetag an die SpraStu-Proband/-innen gesandt, die in der Vorlesung anwesend waren. Ergebnisse des Fragebogens stehen in den Mitschriftenmetadaten zur Verfügung (Kapitel 3.2).

5.2 Datenaufbereitung

Um eine bestmögliche Bildqualität zu erreichen, wurden die abfotografierten Mitschriften hinsichtlich Kontrast und Helligkeit nachbearbeitet. Alle nicht zur jeweiligen Vorlesungssitzung gehörigen Inhalte wurden entfernt. Außerdem wurden personenbezogene Daten wie der Name, die projektinterne SpraStu-ID oder auf dem Foto sichtbare Finger der/des Studierenden geschwärzt und die Mitschriften mit einer MIKO-ID versehen. Diese setzt sich wie folgt zusammen: Kürzel der Vorlesung_Sitzungsnummer_MIKO fortlaufende Nummerierung pro Person_L1/L2, z. B. Ana_1_MIKO01_L2.

Mitschriften, die von Studierenden angefertigt wurden, die zum Zeitpunkt der Erhebung nicht im ersten Fachsemester studierten, tragen die Endung _irreg. In den Mitschriften-Metadaten ist die Irregularität in der Variable author_irregDefinition definiert (Kapitel 3.2).

5.3 Analysemethoden Mitschriften

Mitschreiben ist eine stark kontextualisierte Handlung, die von zahlreichen internen (z. B. Motivation, Vorwissen, Sprachkompetenz) und externen Faktoren (z. B. sprachliche oder fachliche Schwierigkeit der Vorlesungssprache) beeinflusst wird. Sie bedarf also strategischen, sprachlichen und fachlichen Wissens und ist zudem abhängig von individuellen, einzelsituationsabhängigen Motivationen und Absichten (mehr Hintergründe in Wisniewski, 2019). Daher verwundert es nicht, dass auch die resultierenden Produkte in ihrer Ausgestaltung stark variieren.

Von den zwölf im Korpus enthaltenen Vorlesungssitzungen liegen Transkripte zu den acht Sitzungen, in denen die meisten Mitschriften erhoben wurden, vor. Für die in diesen acht Sitzungen erhobenen Mitschriften (N=123; davon Deutsch als L2: N=78, Deutsch als L1: N=45) wurden Deskriptiva (Kenncharakteristika; Kapitel 5.3.1) erfasst und Qualitätsbeurteilungen vorgenommen (diese liegen in den Metadaten vor, Kapitel 3.2). Die Bewertungen wurden nach einem Bewertungsraster auf drei Ebenen vorgenommen: Inhaltliche Vollständigkeit (Kapitel 5.3.3.1), Korrektheit und Gesamteindruck (Kapitel 5.3.3.2). Der Bewertung der inhaltlichen Vollständigkeit ging eine Sequenzierung der Vorlesungstranskripte in Wissenselemente voraus (Kapitel 5.3.2). Ergebnisse der Mitschriftenanalyse sind nachzulesen in Wisniewski, Spiegel, Lenort & Feldmüller (in Vorbereitung).

5.3.1 Erfassung von Kenncharakteristika der Mitschriften

Die erhobenen Mitschriften wurden zunächst deskriptiv hinsichtlich allgemeiner, linguistischer und strukturierender Merkmale erfasst. Folgende Kategorien wurden entwickelt:

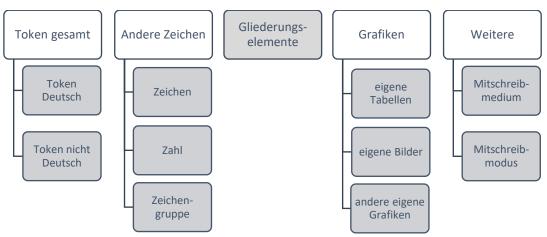


Abbildung 2: Erfasste Kenncharakteristika der Mitschriften

5.3.1.1 Token Deutsch text_token

Als Token wurden lexikalische Wörter, Teilwörter, Abkürzungen sowie genormte (alphabetische) Kürzel und einzelne Buchstaben gezählt. Interpunktionszeichen, die eine satzgliedernde Funktion erfüllen, wurden nicht erfasst. So gelten beispielsweise *Sehnen, u.* (als Abkürzung für *und*), *DD* (als fachspezifische Abkürzung für *Differentialdiagnose*), *nm* (als physikalische Einheit in *10nm*) und *a*) (als gliederndes Aufzählungselement) als Token.

5.3.1.2 Token nicht Deutsch text_token_notDeu

In dieser Kategorie wurden Wörter, Teilwörter und Abkürzungen erfasst, die nicht auf Deutsch mitgeschrieben wurden. Hiervon ausgenommen sind fachsprachliche Termini, z. B. lateinischgriechische im Fach Medizin: *Articuatio sacoiliaca* oder *Foramen transversarium*.

5.3.1.3 Andere Zeichen

Nicht bzw. nicht nur alphabetische Zeichen (vgl. *Nichtwörter* im STTS; Schiller et al., 1999) wurden je einer der drei folgenden Kategorien zugeordnet:

Zahl text numbers

Als Zahlen wurden arabische und römische Ziffern und Ziffernfolgen gewertet, die als Ordinalund Kardinalzahlen verwendet werden und nicht Teil einer Zeichengruppe sind, z. B. 1977 (Jahreszahl, 1 Zahl), 1., 6., 7. (Ordinalzahlen, 3 Zahlen), 2.4.2 (Nummerierung, 1 Zahl), 1,5 (Dezimalzahl, 1 Zahl), I, II, III (Nummerierungen, 3 Zahlen).

Zeichen text characters

In die Kategorie *Zeichen* fallen typografische Zeichen, mathematische Zeichen und Sonderzeichen außerhalb von Zeichengruppen oder Grafiken, z. B. € (1 Zeichen), + (außerhalb einer Formel; 1 Zeichen), +++ (3 Zeichen). Pfeile und als solche verwendete andere Zeichen, wie z. B. lang gezogene Schweifklammern, wurden grundsätzlich als Zeichen gewertet.

Zeichengruppe text_groupOfCharacters

Eine Zeichengruppe ist eine Kombination aus Token (oft einzelne Buchstaben), Zahlen und/oder Zeichen, die eine sinnhafte Einheit darstellt. Zeichengruppen bestehen aus mindestens drei Elementen mindestens zweier verschiedener Zeichenklassen (Buchstaben, Sonderzeichen, Zahlen, mathematische Zeichen, nicht: Interpunktion), wobei Zeichenfolgen einer Klasse als ein Element gewertet wurden. Die Grenze einer Zeichengruppe wird graphisch bestimmt (Absatz, Leerraum, Gliederungselement). Beispiele:

- ∫ 426a (1) → 1 Gruppe (Elemente: Sonderzeichen, Ziffernfolge, Buchstabe, Ziffer; Klammern werden nicht berücksichtigt, da sie Interpunktionszeichen sind)
- ∫∫ 611ff. → 1 Gruppe (Elemente: Sonderzeichenfolge, Ziffernfolge, Buchstabenfolge)
- $F = QE \rightarrow 1$ Gruppe (Elemente: Buchstabe, mathematisches Zeichen, Buchstabenfolge)

In Kombinationen von zwei Elementen aus zwei verschiedenen Kategorien wurden die einzelnen Elemente in den entsprechenden Kategorien und nicht als Zeichengruppe gezählt, z. B. *1-achsig* (1 Zahl, 1 Token).

5.3.1.4 Gliederungselemente text_structuralElements

Die Kategorie *Gliederungselement* bezieht sich auf die Funktion von Zeichen auf wissensorganisatorischer Ebene. Die Erfassung von Gliederungselementen erfolgte unabhängig von und zusätzlich zu den oben genannten Kategorien *Token*, *Zahl*, *Zeichen* und *Zeichengruppe*, sodass solche Elemente in zwei Kategorien gezählt wurden, wenn sie eine gliedernde Funktion ausüben.

Als Gliederungselemente zählen (rein grafische) Aufzählungszeichen, Nummerierungen (numerisch, alphabetisch) und andere Zeichen, die grafisch ausdrücken, dass eine hierarchisch unter- oder nebengeordnete Beziehung zwischen zwei Elementen besteht. U. U. zählen dazu auch Interpunktionszeichen, die in den oben erläuterten Kategorien ausgeschlossen wurden, wenn sie als gliederndes Element verwendet werden. Als Gliederungselemente zählen beispielsweise 3., a) und - (Spiegelstrich).

5.3.1.5 Eigene Tabellen *text_ownTables*, eigene Bilder *text_ownPictures*, andere eigene Grafiken *text_otherGraphics*

Innerhalb dieser Kategorie wurde erfasst, ob die Studierenden über das Abschreiben von präsentierten sichtbaren Medien (Tafelanschrift, Folie etc.) hinaus eigene Tabellen, Bilder oder Zeichnungen notiert haben. Dies betrifft vor allem Versuchsaufbauten, Präparate oder Demonstrationen am Körper.

5.3.1.6 Mitschreibmedium text_medium und -modus text_modus

Außerdem wurde festgehalten, auf welchem Medium (Zettel, Skript, Folie, digitales Dokument) und auf welche Weise (mit der Hand, auf dem Computer) die Studierenden mitgeschrieben haben.

5.3.1.7 Interrater-Reliabilität der Erfassung der Kenncharakteristika

Zur Überprüfung der Reliabilität der Auszählung wurden die soeben beschriebenen Kategorien Token, Andere Zeichen und Gliederungselemente bei N=22 Mitschriften aus allen Vorlesungssitzungen von einer anderen Bewerterin blind zweiterfasst. Innerhalb der Kategorie Token Deutsch⁴ liegt die durchschnittliche prozentuale Übereinstimmung bei 97,3%, die mittlere Abweichung betrug 5 Token. In der Kategorie Andere Zeichen (Zeichen, Zahl, Zeichengruppe) stimmten die erfassten Werte durchschnittlich in 87,7% der Elemente überein, die mittlere Abweichung betrug 2,5 Elemente

_

⁴ Nur vier der zweiterfassten Mitschriften enthielten nicht-deutsche Token, sodass für diese Kategorie keine Reliabilität angegeben wird.

(Zeichen: 4,8; Zahl: 1,6; Zeichengruppe: 1,3). Auf der Ebene Gliederungselemente gab es eine Übereinstimmung von 89,4%, die durchschnittliche Abweichung betrug 3,2 Gliederungselemente.

Die Kategorien Eigene Tabellen, Eigene Bilder, Andere eigene Grafiken, Mitschreibmedium und Mitschreibmodus wurden nicht zweiterfasst, da sie eine geringe Fehleranfälligkeit aufweisen.

5.3.2 Segmentierung der Vorlesungen in Wissenselemente

Vorlesungen sind wissensvermittelnde Diskurse. Sie "bestehen aus Ensembles von Sprechhandlungen, die zur Wissenstradierung unter bildungsinstitutionellen Bedingungen geeignet sind." (Redder, 2016). Studentische Mitschriften wiederum werden von Redder (2016) als "Fenster in die wissensverarbeitenden Prozesse" bezeichnet. In zahlreichen empirischen Studien zur Lernförderlichkeit von Mitschriften hat sich herausgestellt, dass die Behaltens- und Lernleistung insbesondere von solchen Mitschriften profitiert, die möglichst viele Wissenseinheiten (sogenannte *idea units*) auf verschiedenen Relevanzebenen des zu vermittelnden Wissens enthalten (für einen Überblick vgl. Wisniewski, 2019).

Deshalb wurden die acht transkribierten Vorlesungssitzungen im Rahmen einer sequenziellen Strukturanalyse in Wissenselemente segmentiert mit dem Ziel, zu einer evidenzbasierten, differenzierten Einschätzung der Mitschriftenqualität, insbesondere der inhaltlichen Vollständigkeit (Kapitel 5.3.3.1), zu gelangen. Gleichzeitig ist das gewählte Vorgehen deutlich weniger granular als in der Kognitionspsychologie oder der kognitiven Semantik, wo in der Regel kurze Sequenzen meist nicht authentischer Diskurse segmentiert werden.

Als **Wissenselemente** zählen bei der sequenziellen Strukturanalyse in MIKO grundsätzlich alle Vorlesungsinhalte, in denen Wissen vermittelt wird.

Jedes identifizierte Segment enthält einen identifizierbaren Wissensinhalt. Die Vorlesungen wurden dabei nicht nach formalen Kriterien (etwa grammatischen, intonatorischen) segmentiert. Vielmehr wurden Sequenzen mit möglichst spezifischen, voneinander abgrenzbaren Inhalten, die zu vermittelndes Wissen enthalten, segmentiert. Unterstützend wurde dabei auf lexikalische Hinweise geachtet, etwa auf Inhaltswörter und Schlüsselwörter sowie Segmentgrenzen. Dennoch ist das Vorgehen nicht frei von Interpretation der/des Segmentierenden. Schwierigkeiten lagen beispielsweise in der reliablen Einschätzung von Segmentumfängen bzw. der Granularität der Identifikation von Wissensstrukturen.

5.3.2.1 Relevanzebenen

Jeder herausgearbeitete Inhalt wurde zusätzlich hinsichtlich der Ebene des vermittelten Wissens kategorisiert. Diese nach Bedeutsamkeit hierarchisierende Unterteilung folgt Song (2012), die fünf Ebenen unterscheidet: "main idea" (Level 1), "major topical ideas" (Level 2), "subtopic ideas" (Level 3), "supporting details" (Level 4) und "minor details" (Level 5). Bei der vorliegenden Kategorisierung der Vorlesungssitzungen wurden drei Hierarchieebenen von Wissenselementen unterschieden:

- Ebene 1: Hauptthemen
- Ebene 2: Zugehörige wesentliche Themen
- Ebene 3: Nebenthemen, Details, Beispiele

Darüber hinaus wurden Sequenzen kodiert, die sich zwar auf das Thema der Vorlesung beziehen, jedoch kein zu vermittelndes Wissen beinhalten. Sequenzen der Ebene 4 gelten nicht als Wissenselemente im Sinne oben genannter Definition, sodass nicht erwartet werden kann, dass sie mitgeschrieben werden:

- Ebene 4: Marginalia, z. B. Vor- und Rückverweise, Anekdoten

Sequenzen, in denen es gar nicht um das Thema der Vorlesung geht, erhielten keine Codierung und keine Zuordnung zu einer Ebene. Dabei handelt es sich z. B. um Bemerkungen zu Störungen der Vorlesung oder zum Forschungsprojekt SpraStu.

Im Folgenden werden Beispiele für die Segmentierungen und Codierungen auf den vier Ebenen gegeben, obwohl deren vollständige Nachvollziehbarkeit ohne die Kenntnis des Vorlesungskontextes vermutlich nicht möglich ist.

Ebene 1: Hauptthemen

Die Einleitung eines Hauptthemas bzw. der Wechsel von einem Thema zum nächsten wird von der/dem Dozierenden i. d. R. explizit verbalisiert:

BS[v]

Und unser großes Thema • • • soll heute sein: ((1,1s)) "Wie ist denn ein Durchschnittsknochen sozusagen aufgebaut?". Wir wollen uns mal so einen ganz klassischen Röhrenknochen anschauen. • • • Und generell: "Wie verhalten sich denn Knochen • • • bei Belastung?"

Funktionelle Anatomie 1 (MIKO_E_00001, 00:48)

→ Code⁵: Aufbau eines Röhrenknochens & Verhalten bei Belastung <1>

DH[v]

Und wir kommen zum nächsten Punkt, zwei Punkt vier Punkt drei Punkt. Habe ich irgendein Experiment vergessen, oder … Nö, ne, ist alles okay, ne? • • • Genau. • • "Stromleitung im Elektrolyten". ((17,6s)) Stromleitung im Elektrolyten

Physik für Mediziner 2 (MIKO_E_00005, 60:47)

→ Code: Stromleitung im Elektrolyten <1>

Ebene 2: Zugehörige wesentliche Themen

Wissenselemente mit dem Code <2> sind relevant für das Verständnis des Hauptthemas, dem sie zugeordnet sind. Mitunter wird die Relevanz des Elements durch die/den Dozierende/-n explizit hervorgehoben.

38

⁵ Die Codes wurden in MAXQDA gesetzt, siehe Kapitel 5.3.2.2.

BS[v] Der hat keine Blutgefäße, das ist ganz wichtig, das brauchen wir gleich noch mal.

Funktionelle Anatomie 2 (MIKO E 00002, 27:15)

→ Code: Gelenkknorpel hat keine Blutgefäße <2>

Ganz wichtig, was Sie sich bitte hier merken: ((1,1s)) So ein Dienstvertrag • • • beinhaltet • • • als Hauptleistungspflicht desjenigen, der die Dienste ausübt, ((2,2s)) die Verpflichtung zur Bemühung, ((1,6s)) nicht zu einem Erfolg. ((3,0s))

Bürgerliches Recht für Wirtschaftswissenschaftler 1 (MIKO_E_00007, 31:21)

→ Code: Hauptleistungspflicht d. Ausübenden: Bemühung <2>

Ebene 3: Nebenthemen, Details, Beispiele

In mit <3> codierten Wissenselementen werden weiterführende Informationen zu den Codes der Ebene 2 gegeben. Es handelt sich häufig um Beispiele, Merkhilfen, Formelzeichen (Physik), Übersetzungen, Paragrafen (BGB) oder Randinformationen.

DH[v] So, und jetzt haben wir hier ((1,2s)) den induktiven Widerstand. • • • Die Induktivität • • • ähm wird mit dem Formelzeichen "L" bezeichnet, ne? • • • Also hier an den Widerstand schreibe ich das Formelzeichen "R" ran, • • • an die Spule schreibe ich das Formelzeichen "L" dran.

Physik für Mediziner 2 (MIKO_E_00005, 30:05)

→ Code: FZ Induktivität = L <3>

Und es ist tatsächlich so, wenn Sie sich mal einen Kran anschauen: • • • Das ist ja auch kein Riesenklumpen oder so ein Betonfels, ja, • • • obwohl der natürlich enorme • • • Belastungen • • • auszuhalten hat, • • • sondern eigentlich sehen die ja relativ filigran aus.

Funktionelle Anatomie 2 (MIKO_E_00002, 02:56)

→ Code: Bsp: Kran <3>

Ebene 4: Marginalia, z. B. Vor- und Rückverweise, Anekdoten

Sequenzen, die als <4> codiert werden, sind keine Wissenselemente im oben genannten Sinn, da sie kein zu vermittelndes Wissen enthalten. Es handelt sich um Vor- und Rückverweise auf bereits Besprochenes bzw. noch zu Besprechendes sowie Nebenbemerkungen zum Thema, die nicht für das Verständnis des Themas relevant sind. Wiederholungen von Inhalten vorangegangener Vorlesungssitzungen zählen etwa dazu. Auch für Sequenzen, in denen Experimente durchgeführt werden (Physikvorlesungen), wurde der Code <4> vergeben, wenn das Experiment lediglich zur Veranschaulichung von zuvor oder im Anschluss vermitteltem Wissen dient. Es wurde bei Elementen dieser Relevanzebene kein zusätzlicher inhaltlicher Code vergeben.

DH[v]

Sie erinnern sich an Epsilon. Epsilon, die Dielektrizitätszahl. • • • Die war spezifisch für bestimmte Materialien. Wir hatten ja gesagt, dass wir die Kraft zwischen • • • äh Ladungen • • • durch die/ durch das Medium, in dem sich die Ladungen befinden, • • • versp/ • • verstärken oder abschwächen können.

Physik für Mediziner 3 (MIKO_E_00006, 39:28)

→ Code: <4>

SL[v]

Und fünf sechsunddreißig regelt dann zum Beispiel Mietminderung. \bullet \bullet Braucht Sie nicht weiter zu interessieren, Sie sollen einfach nur wissen, dass es das gibt, ((1,5s)) ja? Wir machen das hier nicht so detailliert, dass Sie jetzt alles Mögliche hier können müssen aus dem Mietrecht.

Bürgerliches Recht für Wirtschaftswissenschaftler 1 (MIKO E 00007, 22:58)

→ Code: <4>

Zweifelsfälle

Einige Wissenselemente der Ebenen 2 und 3 konnten keinem Hauptthema zugeordnet werden, z. B. wenn der BGB-Vorlesung vorangestellt allgemeine Erklärungen zum Aufbau des Bürgerlichen Gesetzbuches gegeben wurden.

In vielen Vorlesungssitzungen werden zu Beginn Inhalte der vorherigen Vorlesung wiederholt oder Hausaufgaben besprochen, was mit <4> codiert wurde. Mitunter war es uneindeutig, wann der Übergang zum neuen Wissen und somit dem ersten Hauptthema oder dem ersten zu codierenden Wissenselement der Vorlesungssitzung erfolgt. Unter der Annahme, dass erst das neue Wissen mitgeschrieben wird, wurde in solchen Fällen auf die Mitschriften zurückgegriffen, um zu rekonstruieren, wo dieser Zeitpunkt liegt.

5.3.2.2 Vorgehen

Die Segmentierung der Transkripte erfolgte mit der Software MAXQDA, da das Programm umfangreiche Möglichkeiten zur qualitativen Datenanalyse bietet. Die Sequenzierung wurde von einem der Transkribenten der Vorlesungssitzungen durchgeführt, wobei sich die bereits sehr gute Kenntnis der Vorlesungen als Vorteil erwiesen hat.

Die Transkripte wurden zunächst in Wissenselemente segmentiert und jede Sequenz mit einer Codierung versehen, die das Wissenselement inhaltlich benennt. Anschließend wurde die Ebene im Code kategorisiert (z. B. mit dem Kürzel <2> für Ebene 2) und der Code dem hierarchisch nächst höheren und inhaltlich zugehörigen Code als Subcode zugeordnet (Beispiele in Kapitel 5.3.2.1).

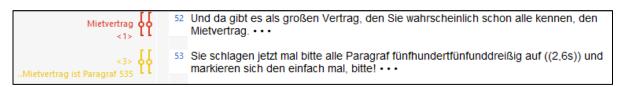


Abbildung 3: Codierfenster in MAXQDA, MIKO_E_00007

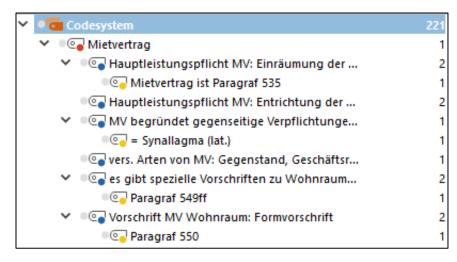


Abbildung 4: Liste der Codings in MAXQDA, MIKO_E_00007

Die minimale Segmentiereinheit war ein Satz. Wurden mehrere Wissenselemente in einem Satz vermittelt, wurde dieser Satz mehrfach codiert. Abschnitte, die der Ebene 4 zugeordnet sind, erhielten keine inhaltliche Codierung, da sie bei der Auswertung der Mitschriften nicht berücksichtigt wurden.

Aus der Segmentierung wurde pro Vorlesungssitzung eine Liste mit den verschieden gewichteten Wissenselementcodierungen der Ebenen 1 bis 3 angelegt, die Grundlage für die Analyse der inhaltlichen Vollständigkeit war (Kapitel 5.3.3.1).

5.3.2.3 Reliabilität der Segmentierung

Die Identifikation von Wissenselementen ist in hohem Maße subjektiv und interpretativ. Die Segmentierungen wurden von einem Transkribenten der Vorlesungssitzungen durchgeführt. Ein Review der Segmentierungen erfolgte durch die zweite Transkribentin. Für sechs der acht Vorlesungssegmentierungen wurden die Abweichungen systematisch erfasst. Marginale Abweichungen wurden dabei nicht gezählt, z. B.:

- doppelte Vergabe von Codes, also das zusätzliche Codieren einer Sequenz mit einem bereits existierenden Code,
- Umbenennung von Codes ohne inhaltlich prägnante Abänderung und
- Abweichungen, die die Ebene 4 betreffen, außer es wurde statt der Ebene 4 ein Code der Ebene 3, 2 oder 1 vergeben.

Wichtige Abweichungen hingegen wurden gezählt:

- Änderung der Ebene,
- Teilen und Zusammenführen von Codes und
- Umbenennung von Codes mit inhaltlich prägnanter Abänderung.

Der Segmentierende konnte einen Code für Unsicherheiten vergeben. Diese Codes wurden im Rahmen des Reviews besprochen und gemeinsam entschieden und daher nicht als Abweichung erfasst. Die so ermittelten Abweichungen betragen durchschnittlich 6,2% der Sequenzen, die mit Ebene 1, 2 oder 3 codiert wurden.

5.3.3 Bewertung der Mitschriften

Die Beurteilung der Qualität der Mitschriften ist subjektiv und wurde unabhängig von der Funktion, welche die Studierenden selbst der jeweiligen Mitschrift zuschreiben, vorgenommen (Metadaten zur Mitschreibabsicht und Nützlichkeit, Kapitel 3.2). Die Mitschriften sind nicht für weitere Adressat/-innen geschrieben und müssen auch nicht prinzipiell für andere nachvollziehbar sein. Trotzdem ist eine Beurteilung nötig, um die Mitschriften auszuwerten und z. B. mit Sprachkompetenzen in Zusammenhang bringen zu können.

Die Mitschriften wurden auf drei Ebenen hinsichtlich ihrer Qualität bewertet. Dem Kriterium Inhaltliche Vollständigkeit liegt die soeben beschriebene Sequenzierung der Vorlesungssitzungen (Kapitel 5.3.2) zugrunde, die Kriterien Korrektheit und Gesamteindruck wurden holistisch anhand von Deskriptoren bewertet.

5.3.3.1 Inhaltliche Vollständigkeit

Das Kriterium *Inhaltliche Vollständigkeit* dient der quantitativen Einschätzung der Mitschreibleistung in Bezug auf die maximal mitschreibbaren Wissenselemente der Vorlesungssitzung (betrifft also **nur Ebene 1 bis 3**, Kapitel 5.3.2.1). Die aus der Sequenzierung gewonnenen hierarchisch gegliederten Wissenselemente wurden zur Bewertung in eine je Vorlesungssitzung aus MAXQDA extrahierte Exceltabelle überführt, welche pro Mitschrift ausgefüllt wurde.

Bei der Bewertung der inhaltlichen Vollständigkeit wurde jedes Wissenselement dahingehend bewertet, in welchem Umfang es mitgeschrieben wurde. Ein Wissenselement konnte vollständig (2 Punkte), teilweise/indirekt (1 Punkt) oder nicht mitgeschrieben worden sein (0 Punkte).

2 Punkte (klar, eindeutig, zufriedenstellend, genau, ausführlich, detailliert oder vollständig mitgeschrieben)

Als vollständig mitgeschrieben gelten Wissenselemente, wenn sie eindeutig notiert wurden. Dazu zählen auch Formulierungen, die als synonyme Ausdrucksweise der zu vermittelnden Inhalte gelten können. Außerdem wurde festgelegt, dass bei der Nennung mehrerer Beispiele in der Vorlesung (i. d. R. als Ebene 3 codiert) das Mitschreiben eines Beispiels ausreicht, um zwei Punkte für das Wissenselement zu erhalten.

1 Punkt (unvollständig/indirekt mitgeschrieben)

Nicht direkt oder nur unvollständig in den Mitschriften abgebildete Wissenselemente wurden mit einem Punkt bewertet. Zur indirekten Realisierung zählen einerseits textuell-implizit transferierte Wissenselemente (uneindeutige Formulierungen, einzelne/nicht zugeordnete Stichworte) und andererseits auch Zeichnungen oder das grafische Hervorheben von Bestandteilen in Zeichnungen.

Beispiel für **textuell-implizit** mitgeschriebene Wissenselemente:

```
Das heißt also, ((3,6s)) die Leistung, • • • die Sie haben, • • • hängt mit einem Mal von T ab, von der Zeit. ((2,3s)) Zu jeden Zeitpunkt • • • haben Sie möglicherweise eine unterschiedliche Leistung.
```

Physik für Mediziner 2 (MIKO_E_00005, 51:04)

→ Code: Energie und Leistung im Wechselstromkreis <1> \ Formel: P = U0*cos(Omega t + Delta kl Phi)*I0*cos Omega t <2> \ Leistung hängt von der Zeit ab <3>

Bei diesem Wissenselement gilt z. B. als indirekt mitgeschrieben, wenn neben der entsprechenden Formel der Begriff *Momentanleistung* notiert wurde, da dieser implizit einen Hinweis darauf enthält, dass die Leistung von einem zeitlichen Verlauf abhängig ist.

Beispiel für indirekte Realisierung von Wissenselementen in Zeichnungen:

```
BS[v]
Das glaube ich nicht! ((lacht)) Selbst die zarteste Person kann das so nicht machen. Warum denn nicht? • • • Weil wir ((1,0s)) an dieser Stelle • • • eine Weichteilhemmung haben. Also mein Bizeps ist jetzt auch nicht so dolle. • • • Aber trotzdem hat jeder von uns dort eine Weichteilhemmung. (Anmerkung: Dozierende veranschaulicht Weichteilhemmung an eigenem Ellenbogengelenk)
```

Funktionelle Anatomie 3 (MIKO E 00003, 03:26)

→ Code: Nachtrag: Gelenke <1> \ weiterer Einfluss auf Beweglichkeit: Weichteilhemmung <2> \ Bsp. Ellenbogengelenk <3>

Wird zu diesem Wissenselement das Stichwort Weichteilhemmung mitgeschrieben und daneben eine kleine Zeichnung mit einem Arm angefertigt, bei dem das Ellenbogengelenk gebeugt ist, dann gilt das Wissenselement als indirekt abgebildet (wegen ausschließlich bildlicher Darstellung des Ellenbogengelenks).

Unvollständig mitgeschrieben sind Wissenselemente beispielsweise dann, wenn nur einzelne Teile mitgeschrieben wurden, Begriffserklärungen fehlen oder die enthaltenen Begriffe ohne sinnvolle Relation zueinander notiert wurden.

Beispiel für **unvollständig** mitgeschriebene Wissenselemente:

```
SL[v] Wir haben erst mal ein Dreipersonenverhältnis, • • • und zwar das Verhältnis zwischen Lieferant, Leasinggeber und Leasingnehmer. ((2,4s))
```

Bürgerliches Recht für Wirtschaftswissenschaftler 1 (MIKO_E_00007, 63:58)

→ Code: Leasingvertrag <1> \ Dreipersonenverhältnis: Lieferant, Leasinggeber, Leasingnehmer <2>

Wenn bei diesem Wissenselement nur die Begriffe Lieferant, Leasinggeber und Leasingnehmer mitgeschrieben wurden, gilt es als unvollständig mitgeschrieben, da es sich bei Dreipersonenverhältnis um einen wesentlichen Teil des Wissenselements handelt.

<u>0 Punkte (nicht mitgeschrieben)</u>

Wurde ein Wissenselement gar nicht mitgeschrieben, wurden null Punkte vergeben. Auch inhaltlich nicht korrekt mitgeschriebene Wissenselemente wurden als nicht mitgeschrieben gewertet.

99 (nicht bewertbar)

Hindernisse für die Beurteilung der inhaltlichen Vollständigkeit einzelner Wissenselemente können auf unterschiedlichen Ebenen vorkommen.

War aufgrund von Mängeln bei der Fotoqualität nicht eindeutig erkennbar, ob ein Wissenselement realisiert wurde oder nicht, wurde die Bewertung 99 vergeben, z.B. bei abgeschnittenen oder unscharfen Stellen der Mitschrift.

Auch Wissenselemente, die nicht auf Deutsch oder in mehr als einer Sprache mitgeschrieben wurden, konnten nicht miteinbezogen werden und erhielten daher die Bewertung 99. Ausgenommen davon sind Fachtermini (z. B. lateinisch-griechische in der Medizin), die innerhalb eines nicht-deutschen Mitschriftenabschnitts durch ihre Verortung innerhalb der Mitschrift oder durch sonstige Kennzeichen zuzuordnen waren und daher (vermutlich) richtig mitgeschrieben wurden.

Wenn in einer Mitschrift ein eindeutiger Hinweis enthalten war, dass Wissenselemente nicht mitgeschrieben wurden, weil die/der Studierende sie in Skripten, Lehrbüchern, Mitschriften anderer Lehrveranstaltungen o. Ä. findet oder bereits mitgeschrieben hat, wurden die entsprechenden Wissenselemente ebenfalls mit 99 bewertet. Daher zählen bei Mitschriften, die auf einem Skript angefertigt wurden, alle bereits im Skript enthaltenen Wissenselemente ebenfalls als nicht bewertbar, es sei denn, die/der Mitschreibende hat durch Hervorhebungen oder Erweiterungen die im Skript enthaltenen Informationen ergänzt oder modifiziert. Jene Inhalte wurden wie oben beschrieben mit den regulären Bewertungen versehen.

Interrater-Reliabilität der inhaltlichen Vollständigkeit

Bei jeder Vorlesungssitzung, zu der die Mitschriften angefertigt wurden, wurde zu Beginn des Bewertungsprozesses über eine doppelblinde Bewertung ein Goldstandard etabliert. Dazu wurden pro Vorlesungssitzung zwei bis vier Mitschriften ausgewählt, die auffällig erschienen (z. B. durch Umfang, Unübersichtlichkeit, Mitschreibmedium, viele Zeichnungen, Lücken o. Ä.). Insgesamt wurden 17% (N=21) aller zu bewertenden Mitschriften doppelt bewertet. Sämtliche Abweichungen wurden besprochen und jeweils vorlesungssitzungsspezifische Einigungen getroffen und dokumentiert. Die Übereinstimmung bei der Bewertung lag bei durchschnittlich 87% der Wissenselemente (minimal: 77%, maximal: 93%). Im Anschluss an diese Phase wurden die Bewertungen von einer Person vorgenommen und Zweifelsfälle regelmäßig und fortlaufend im Bewerterinnenteam besprochen.

5.3.3.2 Korrektheit und Gesamteindruck

Die Kriterien Korrektheit und Gesamteindruck (Organisation und Nachvollziehbarkeit) sind als holistische Bewertungen zu verstehen und beruhen auf der qualitativen Beurteilung durch zwei Bewerterinnen. Alle Mitschriften (N=123) wurden doppelblind bewertet. Anschließend wurde in einem diskursiven Verfahren ein Konsens ermittelt.

⁶ Korrektheit bezieht sich in der vorliegenden Analyse ausschließlich auf inhaltliche und nicht auf sprachliche Aspekte.

Der qualitativen Bewertung liegen folgende Deskriptoren zugrunde:

Korrektheit des Mitgeschriebenen		
3	Es wurde durchgängig korrekt mitgeschrieben.	
2	Die mitgeschriebenen Inhalte enthalten einige wenige Ungenauigkeiten oder kleinere Unklarheiten, sind aber weitgehend korrekt.	
1	Die mitgeschriebenen Inhalte sind durch viele Fehler geprägt; es gibt nicht nur Lücken, sondern auch klar falsch Festgehaltenes.	
0	Die mitgeschriebenen Inhalte sind nahezu durchgängig nicht korrekt.	
n.b.	Die Korrektheit ist aus verschiedenen Gründen nicht beurteilbar.	

Tahelle 11: Deskriptoren für die Bewertung der Korrektheit

Gesamtei	ndruck: Organisation und Nachvollziehbarkeit
3	Die Mitschrift ist durchgängig klar aufgebaut und gegliedert. Es werden nicht nur einzelne Inhaltswörter mitgeschrieben, sondern Inhalte und Themen sowie deren Beziehungen untereinander werden angemessen und nachvollziehbar sprachlich, grafisch oder typografisch dargestellt. Die Mitschrift ist für jemanden, der den Bezugstext kennt, durchgängig nachvollziehbar. Um inhaltliche Kohärenz zu erreichen, weicht die Mitschrift evtl. stellenweise von der chronologischen Abfolge der Vorlesung ab.
2	Die Mitschrift ist überwiegend klar aufgebaut und gegliedert. Inhalte und Themen sowie ihre Beziehungen untereinander werden meist angemessen und nachvollziehbar sprachlich, grafisch oder typografisch gegliedert, auch wenn es einzelne Unklarheiten bezüglich einzelner Themen/Inhalte gibt. Die Mitschrift ist für jemanden, der den Bezugstext kennt, meist nachvollziehbar, auch wenn einige unklare Stellen auftreten können.
1	Die Mitschrift ist nur ansatzweise gegliedert. Inhalte und Themen werden nur teilweise gegliedert dargestellt, und Beziehungen der Inhalte und Themen untereinander werden selten oder nicht überzeugend markiert. Die Mitschrift wirkt insgesamt unorganisiert und unübersichtlich. Einzelne Inhalte der Vorlesung werden festgehalten, ohne dass deren Bedeutung und Zusammenhang immer klar würden. Es wird insgesamt nicht klar, was wichtige und unwichtige Informationen sind. Die Mitschrift ist somit auch für jemanden, der die Vorlesung kennt, schlecht nachvollziehbar. Für jemanden, der den Bezugstext nicht kennt, ist die Mitschrift nur schwer nachvollziehbar.
0	Es ist keine Gliederung erkennbar, die Mitschrift ist durchgängig nicht nachvollziehbar.
n.b.	Die Organisation und Nachvollziehbarkeit ist aus verschiedenen Gründen nicht beurteilbar.

Die Bewertung der Mitschriften ergab insgesamt ein relativ homogen-positives Bild. Hinsichtlich der Korrektheit wurde keine Mitschrift mit weniger als zwei Punkten bewertet. Nur zwei Mitschriften erhielten in Bezug auf den Gesamteindruck einen Punkt. Eine Mitschrift konnte aufgrund ihrer Kürze (acht Token, davon zwei fremdsprachlich, keine weiteren Elemente) in diesen beiden Kategorien nicht bewertet werden.

Interrater-Reliabilität der Korrektheit und des Gesamteindrucks

Bei der doppelblinden Bewertung hat sich eine Bewerterin jeweils intensiv mit einer einzelnen Mitschrift in allen zu erfassenden Kategorien beschäftigt und die zweite Bewerterin pro Vorlesungssitzung nacheinander mit allen Mitschriften hinsichtlich einer Bewertungskategorie befasst. Auf diese Weise fußt die Beurteilung einerseits auf der detaillierten Kenntnis der einzelnen Mitschriften (nützlich für die Beurteilung der Korrektheit) und andererseits auf einem vergleichenden Blick auf alle Mitschriften innerhalb einer Vorlesungssitzung (nützlich für den Gesamteindruck).

Bei der Bewertung der Korrektheit betrug die Übereinstimmung 94,3% und beim Gesamteindruck 90,2%. Die Abweichungen betrugen maximal einen Punkt.

II Transkription und Annotation

6 Übersicht Spurstruktur

Im Folgenden wird ein Überblick über die einzelnen Transkriptions- und Annotationsspuren der Vorlesungstranskripte gegeben. Eine detaillierte Beschreibung der Transkriptions- und Annotationsrichtlinien der einzelnen Spuren und Phänomene erfolgt ab Kapitel 8. Das Kürzel X wird in allen Spuren durch die Sprechersigle der/des jeweiligen Dozierenden ersetzt (Metadatenvariable $speaker_acronym$).

Transkriptionsspuren⁷

Kürzel	Beschreibung
X[v]	Transkription der verbalen Äußerungen und Pausen der/des Dozierenden
Assist[v]	Transkription verbaler Äußerungen und Pausen des Assistenten sowie ausgewählter relevanter außersprachlicher Handlungen
NN[v]	Anonymisierte Transkription verbaler Äußerungen oder außersprachlicher Handlungen aus dem Auditorium

Tabelle 13: Transkriptionsspuren

Ein besonderer Fokus von MIKO liegt auf der Multimodalität der Vorlesung, mit der die Studierenden simultan während des Mitschreibens umgehen müssen (modality interdependencies, Reimer et al., 2015). Die Auswahl der manuellen Annotationsspuren orientiert sich an dem Ziel, diese Multimodalität abzubilden, z. B. Betonungen, die Sprechweise oder nonverbale Handlungen. Einen Schwerpunkt bei der Annotation bildet zudem die Beziehung zwischen gesprochener und auf Vorlesungsmedien geschriebener Sprache, die in der Spur Text-Bild-Bezüge X[bild] erfasst wurde (Kapitel 14).

Manuelle Annotationsspuren

Kürzel	Beschreibung
X[akz]	Annotation besonderer Betonungen
X[wort]	Annotationen auf Wortebene (Versprecher, Pausen und Verzögerungssignale innerhalb von Wörtern, Vorschläge bei schwer verständlichen Wörtern oder Passagen)
X[sprech]	Annotation besonderer Sprechweisen
X[meta]	Annotation sprachlicher Erläuterungen
X[nv]	Annotation nonverbaler Handlungen
X[nv2]	Annotation parallel stattfindender nonverbaler Handlungen

⁷ Es wurden je Vorlesungstranskript nur die Spuren verwendet, in denen mindestens ein Phänomen transkribiert bzw. annotiert werden konnte. Daher enthält nicht jedes Transkript alle Spuren.

Kürzel	Beschreibung
X[bild]	Annotation von Text-Bild-Bezügen, d. h. zeitliche Bezüge zwischen Verbalem und auf Medien sichtbaren Begriffen
X[komm]	Anmerkungen zum Transkript, freie Beschreibung von nicht festgelegten Phänomenen

Tabelle 14: Manuelle Annotationsspuren

Die tokenisierten Versionen der Vorlesungstranskripte, die im LZA zur Verfügung stehen, enthalten darüber hinaus automatische Transkriptions- und Annotationsspuren.

Automatische Transkriptions- und Annotationsspuren

Kürzel	Beschreibung
X[tok]	Tokenisierung der Verbalspur der/des Dozierenden $X[v]$ ohne Satzzeichentokenisierung
X[pos]	POS-Tagging
X[p-pos]	Angabe der Wahrscheinlichkeit für die Korrektheit des POS-Taggings
X[lemma]	Lemmatisierung
Assist[tok]	Tokenisierung der Verbalspur des Assistenten Assist[v] ohne Satzzeichentokenisierung
Assist[pos]	POS-Tagging
Assist[p-pos]	Angabe der Wahrscheinlichkeit für die Korrektheit des POS-Taggings
Assist[lemma]	Lemmatisierung
NN[tok]	Tokenisierung der Auditoriumsspur $NN[v]$ ohne Satzzeichentokenisierung

Tabelle 15: Automatische Transkriptions- und Annotationsspuren

Da die Spur NN[v] des Auditoriums keine eigentlichen Verbaläußerungen enthält, stehen die Spuren pos, p-pos und lemma zu dieser Spur nicht zur Verfügung.

7 Transkriptionssystem HIAT

7.1 Begründung der Auswahl des Transkriptionssystems

Bei MIKO wurden die Vorlesungssitzungen nach leicht modifizierten HIAT-Richtlinien (Halbinterpretative Arbeitstranskription, Rehbein et al., 2004) transkribiert. Das etablierte Transkriptionssystem HIAT wurde einerseits um projektspezifische Richtlinien erweitert, andererseits wurden ausgewählte Konventionen verändert. Die entsprechende Dokumentation erfolgt ab Kapitel 7.2. Sofern nicht anders angegeben, ist das HIAT-Handbuch stets die Grundlage für die hier erläuterten Richtlinien und wird aus diesem Grund nur bei projektspezifischen Abweichungen oder Ergänzungen zitiert.

In Erwägung gezogen wurden außerdem die relativ verbreitete Transkription nach GAT 2 (Gesprächsanalytisches Transkriptionssystem 2, Selting et al., 2009) sowie nach CHAT (Codes for the Human Analysis of Transcripts, MacWhinney, 2000). Während auch diese Transkriptionskonventionen Vorteile bieten, schienen sie für den gegebenen Kontext nicht in gleicher Weise geeignet zu sein. Im Forschungsprojekt SpraStu standen keine im engeren Sinne gesprochensprachlichen Aspekte der transkribierten Vorlesungen im Forschungsinteresse: SpraStu untersuchte mit MIKO vor allem lexikalische und grammatische sowie ausgewählte pragmatische Fragestellungen der Vorlesungen, insbesondere im Zusammenhang mit von Studierenden angefertigten Mitschriften. Deshalb wurde eine ökonomische, teilnormalisierte Transkription gewählt.

HIAT orientiert sich stark an der Orthographie der Schriftsprache (*literarische Umschrift*, Kapitel 7.2). Viele Regeln, z. B. das Markieren von Äußerungsmodi oder die Anwendung von Groß- und Kleinschreibung, können von der Schriftsprache übernommen werden. Dadurch gestaltet sich nicht nur die Transkription als intuitiver und damit weniger fehleranfällig und schneller, sondern auch das Lesen der Transkripte wird beschleunigt. Außerdem werden automatisierte Verarbeitungsschritte der Transkripte wie etwa Tokenisierung, Lemmatisierung, POS-Tagging und automatisierte Komplexitätsanalysen durch eine möglichst der Standardsprache angenäherte, zumindest nahezu normalisierte Transkription erheblich erleichtert.

Im Gegensatz dazu berücksichtigt die deutlich aussprachenähere GAT 2-Konvention (Selting et al., 2009) viel stärker gesprochensprachliche Aspekte, wie etwa gefüllte Pausen und Verzögerungsphänomene, und ist klarer von der Schriftsprache zu unterscheiden. Das hätte zu einem erheblichen Mehraufwand bei der Transkription geführt. Deshalb kam GAT 2 für MIKO auch in modifizierter Form nicht in Frage, auch wenn es sich dabei um eine im deutschsprachigen Raum verbreitete und die beispielsweise im GeWiss-Korpus (Fandrych & Schmidlin, 2019) verwendete Konvention handelt.

Die CHAT-Konventionen (MacWhinney, 2000) lassen sich ebenfalls kontextangemessen anpassen, sind an frei verfügbare Analysetools geknüpft und zudem auch international weit verbreitet. Da jedoch das .cha-Format ganz besondere Spezifika aufweist, die die Daten eventuell nicht für eine gleichermaßen breite Nutzung prädestiniert hätten, wurde CHAT nicht berücksichtigt.

Auch aus technischer Sicht bietet sich HIAT als Grundlage an, denn der EXMARaLDA Partitur-Editor sowie die zugehörigen Tools sind hervorragend darauf abgestimmt (Kapitel 2.1.2).

Transkriptionen und Annotationen sollten und können in Anbetracht der hohen Kosten stets nur so feinkörnig sein, wie es dem jeweiligen Forschungskontext angemessen ist (Hovy & Lavid, 2010). Gleichzeitig sollte MIKO von Anfang an in einem öffentlich zugänglichen Korpus verfügbar gemacht werden. Hierdurch entstand ein gewisses Spannungsfeld, denn zweifelsohne wären die Daten für viele weitere, auch im engeren Sinne gesprochensprachliche Untersuchungsfragen nutzbar, wenn sie granularer transkribiert wären. Bei MIKO wurde deshalb ein Partiturformat (multi-layer standoff, Lüdeling et al., 2005) gewählt, das die spätere Ergänzung weiterer Transkriptions- und Annotationsspuren problemlos ermöglicht.

7.2 Literarische Umschrift und Normalisierung

Die Transkription ist gemäß des Transkriptionsverfahrens HIAT (Rehbein et al., 2004, S. 11) in literarischer Umschrift verfasst. Für die Transkription ist die deutsche Standardorthografie nach Duden (Dudenredaktion, o. J.) maßgeblich. Bei mehreren orthografischen Varianten wird die von Duden empfohlene Schreibung gewählt. Umgangssprachliche und dialektale Varianten, die laut Duden akzeptiert sind, werden beibehalten, z. B. heut statt heute, nu statt nun, bissel statt bisschen, nix oder nischt statt nichts, gerne statt gern, vorne statt vorn, zwohundertzwanzig statt zweihundertzwanzig. Auch verdoppelte Pronominaladverbien wie dadrin, dadrauf, dadraus fallen unter diese Dudenregelung. Ooch statt auch hingegen ist beispielsweise laut Duden nicht akzeptiert und wird daher normalisiert. Dialektal bedingte Kasusabweichungen werden nicht normalisiert, wenn sie auditiv eindeutig feststellbar sind, z. B. mit einen Stift.

Reduktionen des Schwa-Lautes in den Endsilben -en, -em und -el werden nicht erfasst:

Sprecher/-in	Transkription
wir laufn	wir laufen
Atm	Atem
Tafl	Tafel

Elisionen/Tilgungen einzelner oder mehrer Laute sowie Änderungen von Flexionsmerkmalen werden abweichend von HIAT (Rehbein et al., 2004, S. 12) nicht erfasst:

Sprecher/-in	Transkription
nich	nicht
un	und
ne	eine
ich hab	ich habe

Eine Ausnahme bilden Tilgungen, die laut Standardorthografie des Duden als umgangssprachlich akzeptiert sind, z. B. grad oder gerad statt gerade, drum statt darum oder drauf statt darauf.

Klitisierungen werden ebenfalls abweichend von HIAT (Rehbein et al., 2004, S. 12) nicht erfasst, sondern standardorthografisch transkribiert:

Transkription
ob sich es
ob es sich
gab es
kannst du
stimmt es

7.3 Segmentierungseinheiten

7.3.1 Äußerungen

Mündliches sprachliches Handeln lässt sich in Äußerungen segmentieren. Rehbein et al. (2004, S. 19) definieren Äußerungen als "Einheiten des Diskurses, die von einem Aktanten aus der Sprecherrolle heraus an einen ko-präsenten Hörer adressiert werden und der Umsetzung illokutionärer und propositionaler Akte dienen." Verschiedene Kriterien wie beispielsweise der kommunikative Kontext, Morphosyntax, Pausen und intonatorische Merkmale können zwar herangezogen werden, um Äußerungsgrenzen und -modi zu bestimmen, ein gewisser Interpretationsspielraum verbleibt jedoch, dem durch das Review Rechnung getragen wird (Kapitel 4.3.4). Die weitere Untergliederung von Äußerungen in Teiläußerungen (durch Kommata, Parenthesen etc.) wird grundsätzlich so gehandhabt, wie der Duden (Dudenredaktion, 2016) Teilsätze und Einschübe definiert, sofern in diesen Richtlinien nicht anders festgelegt.

7.3.2 Wörter

Die Segmentierung des Lautstroms in Wörter orientiert sich an der Standardorthografie nach Duden unter Berücksichtigung der Groß- und Kleinschreibung. Wortgrenzen werden durch Leerzeichen bzw. Interpunktion gekennzeichnet.

7.3.3 Events

Im EXMARaLDA Partitur-Editor wird in Events (Ereignissen) transkribiert. Events teilen den Redestrom in einzelne Segmente, die dem jeweiligen Audioabschnitt zugeteilt werden (Alignment). Diese Untergliederung hat keine analytische linguistische Bedeutung. Dennoch wurden verschiedene Richtlinien zum Setzen von Eventgrenzen befolgt:

- Events sollen in der Regel nicht länger als 5 Sekunden sein,
- Eventgrenzen werden nach dem Abschluss einer Äußerung (markiert durch ein Äußerungsendzeichen) sowie bei längeren Äußerungen, die aus einer Reihung von Teiläußerungen bestehen, nach Teiläußerungsgrenzen (z. B. Komma, Doppelpunkt, bei Parenthesen etc.) gesetzt,
- nach sehr kurzen Äußerungen (z. B. Einwortäußerungen) kann, muss aber keine Eventgrenze gesetzt werden,
- werden Annotationen zu Abschnitten verbaler Handlungen hinzugefügt, wird der betreffende verbale Abschnitt wortgenau in einem separaten Event transkribiert. Interpunktionszeichen werden nicht extra abgetrennt,

843 [23:14.5]	844 [23:15.]	845 [23:15.8]
Das ist also nicht	zufällig,	diese Anordnung.
	-	

Abbildung 5: Eventgrenzen bei wortgenauer Annotation; MIKO_E_00001, 23:14

- Pausen über einer Sekunde werden in ein eigenes Event geschrieben,
- Wörter in der Verbalspur werden nicht in mehrere Events untergliedert und
- in Verbalspuren endet jedes Event mit einem Leerzeichen.

8 Verbalspur Dozent/-in X/v

In der Spur X[v] werden alle verbalen Äußerungen der/des Dozierenden, nicht-phonologische akustische Phänomene wie Lachen sowie Pausen transkribiert. Im folgenden Kapitel werden die Richtlinien für Äußerungsendzeichen (Kapitel 8.1), Pausen (Kapitel 8.2), besondere Wörter (Kapitel 8.3), Unverständliches und Zweifelsfälle (Kapitel 8.4), intrasegmentale Phänomene (Kapitel 8.5), uneigentliches Sprechen (Kapitel 8.6) sowie nicht-phonologische akustische Phänomene (Kapitel 8.7) erläutert.

8.1 Äußerungsmodus und Äußerungsendzeichen

8.1.1 Assertion

Notation: . (Punkt)

Assertionen sind "Aussagen über einen Sachverhalt" (Rehbein et al., 2004, S. 20) und werden mit einem Punkt markiert. Sie sind oft durch terminale Intonation am Äußerungsende und Verbzweitstellung gekennzeichnet.

MIKO E 00003, 18:50

Als Antworten auf Fragen können Assertionen auch Einwortäußerungen bestehend aus nur einer Interjektion sein.

DH[v]	Wir hatten/ • • • zwanzig, zwanzig Milliampere hatten wir ungefähr gesagt, ne?	
Assist[v]		Ja.

MIKO_E_00006, 86:09

8.1.2 Frage

Notation: ? (Fragezeichen)

Äußerungen mit mindestens einem Frageindikator (z. B. Fragewort, Wortstellung, steigende Intonation) werden als Frage interpretiert und erhalten ein Fragezeichen. Von Bedeutung ist dabei nicht, ob die/der Sprecher/-in eine hörerseitige Antwort auf die Frage erwartet. Es kann sich also auch um rhetorische Fragen handeln.

MIKO_E_00012, 30:04

BS[v]	Können Sie	sich erinnern,	als wir	über d	en Begriff	"serra",	"die Säge"	
ps[v]	gesprochen	haben?						

MIKO_E_00003, 05:47

Wird eine Assertion mit einem Augment (*ja, oder, ne* etc.) und steigender Intonation geschlossen, gilt das Gesagte als unsicher und wird daher mit einem Fragezeichen als Äußerungsendzeichen versehen.

DHTV	1	Eisen	ist	ja	•	•	• einer	der	wenigen	Stoffe,	die	wirklich	ferromagnetisch	
וחטון	J	sind,	ne?											

MIKO_E_00006, 33:06

CE [v]	Ähm • • • ähm ((2,0s)) ich habe als Fahrgast viele Wünsche, ja? • • • Aber äh
CF[V]	die Straßenbahn bietet mir genau einen an, nämlich einen Knopf, • • • ja?

MIKO E 00012, 77:09

BS[v] • • • Man könnte doch das auch miteinander verwachsen lassen, oder?

MIKO_E_00003, 07:43

Äußerungen, die durch ihre Wortstellung als Assertion gekennzeichnet sind, aufgrund ihrer steigenden Intonation am Äußerungsende und dem Kontext jedoch als Frage interpretiert werden können, erhalten ein Fragezeichen.

CF[v] Sie haben schon nachgeguck	t?
----------------------------------	----

MIKO_E_00012, 90:30

Teiläußerungen, die durch ihre Wortstellung als Frage gekennzeichnet sind, deren Kontext jedoch deutlich macht, dass es sich um einen Relativsatz mit Verbzweitstellung handelt, erhalten kein Fragezeichen.

BS[v]	Diese ganze Physengeschichte beschäftigt sich einfach mit den Teilen in der
D2[v]	Mitte und am Ende, und was ist dazwischen.

MIKO_E_00001, 24:26

8.1.3 Aufforderung, Ausruf

Notation: ! (Ausrufezeichen)

Aufforderungen mit eindeutig imperativem Modus und Ausrufe, die vor allem durch Lautheit und Emotionalität (z. B. Überraschtheit, Ärger, Freude) gekennzeichnet sind, werden mit einem Ausrufezeichen notiert.

Aufforderungen:

CE[v]	• • • Einzige Hausaufgabe: Achten Sie auf die Nachrichten und äh informieren Sie
CF[V]	• • • Einzige Hausaufgabe: Achten Sie auf die Nachrichten und an informieren Sie sich über das Unwort des Jahres!

MIKO_E_00012, 90:22

BS[v]	Sagen	Sie	es	mal	laut!	
-------	-------	-----	----	-----	-------	--

MIKO_E_00003, 13:49

Ausrufe:

BS[v] • • • Danke für den Hinweis!	
------------------------------------	--

MIKO E 00003, 39:47

DH[v]	So!		•	Damit	sind	wir	dann	•	•	•	für	heute	am	Ende.	
-------	-----	--	---	-------	------	-----	------	---	---	---	-----	-------	----	-------	--

MIKO_E_00004, 86:06

8.1.4 Abbruch

Notation: ... (drei Punkte)

Bleibt die Äußerung einer Sprecherin/eines Sprechers unvollendet und beginnt sie/er oder ein/-e andere/-r Sprecher/-in anschließend eine neue Äußerung, wird diese Äußerung als Abbruch gewertet und erhält drei Punkte als Äußerungsendzeichen. Als neu gilt die nachfolgende Äußerung, wenn die erste abgebrochene Äußerung nicht fortgeführt wird. Abbrüche sind äußerungsexterne Phänomene und grenzen sich dadurch zu äußerungsinternen Reparaturen ab (Rehbein et al., 2004, S. 22).

Bei einem Abbruch kann der Turn an eine/-n andere/-n Sprecher/-in übergehen. Führt ein/-e Sprecher/-in die eigene Äußerung bewusst nicht zu Ende, woraufhin ein/-e andere/-r Sprecher/-in den Turn übernimmt, handelt es sich um einen Turn-Verzicht. Wird ein/-e Sprecher/-in unterbrochen und schließt die Äußerung deshalb nicht ab, verliert sie/er den Turn (Rehbein et al., 2004, S. 23).

Aufgrund der Monologizität der Vorlesungssituation und der geringen Zahl an Gesprächsinteraktionen sind turninterne Abbrüche vorherrschend, bei denen der Turn beibehalten wird.

Eine Mischform aus den genannten Szenarien stellt folgendes Beispiel dar:

CF[v]	Noch andere,	die typisch s	• • • Wie bitte?	
NN[v]		<pre>((Student/-in antwortet auf Frage))?</pre>		<pre>((Student/-in wiederholt Antwort))?</pre>

MIKO E 00012, 11:54

CF bricht seine Äußerung ab, weil ein/-e Studierende/-r aus dem Plenum zeitgleich eine Frage stellt. Daraufhin macht CF eine Pause, in der der Turn von der/dem Studierenden übernommen werden könnte. Da die/der Studierende die Frage bereits vollständig formuliert hat, übernimmt sie/er den Turn jedoch nicht, sodass CF die/den Studierenden bittet, die Frage zu wiederholen. Der vermeintliche Turn-Verlust von CF durch die Unterbrechung wird demnach zu einer Turn-Übergabe durch CF.

Ähnlich verhält es sich bei Äußerungen, die abgeschlossen sind, dann jedoch durch ein Augment mit schwebender Intonation erneut geöffnet und dann nicht fortgeführt werden. Bei Rehbein et al. (2004, S. 24) wird dieses Phänomen zwar beschrieben, es gibt aber keine Transkriptionsvorschläge. Um diesen Fall von einem Augment mit steigender Intonation abzugrenzen, werden solche Äußerungen in den vorliegenden Transkripten als Abbruch notiert.

I (F I V I	Ab wann, ja, ab wann haben Sie sich so als Nichtraucher gefühlt, oder	
NN[v]		((Student/-in antwortet auf Frage)).

CF[v]	Noch gar nicht? • • • Also Sie haben jetzt erst am ersten Januar aufgehört, oder		Okay.
NN[v]		<pre>((Student/-in antwortet auf Frage,5s)).</pre>	

MIKO E 00012, 32:27

8.1.5 Äußerung ohne Modus

Notation: (hochgestellter Punkt)

Hat eine Äußerung keinen der zuvor beschriebenen Modi, wird der hochgestellte Punkt als Äußerungsendzeichen vergeben. Dazu zählen:

- Hörersignale⁸ ohne Anspruch auf Turnübernahme,
- Äußerungen, die so unverständlich sind, dass ihre Illokution nicht eindeutig bestimmt werden kann und
- relevante nicht-phonologische akustische Phänomene (z. B. Lachen, Kapitel 8.7).

CF[v]		Laut,		Ja˙		Okay '		Okay '		O kay		Ja˙
NN[v]	((Stu	((Student/-in antwortet auf Frage,23s)).										

MIKO_E_00012, 32:46

8.2 Pausen

Notation: • • • (drei Pausenpunkte) bzw. ((1,2s))

Als Pausen gelten Sequenzen, innerhalb derer ein/-e Sprecher/-in während eines Turns keine verbale Handlung vollzieht oder andere relevante nicht-phonologische akustische Phänomene wie Lachen auftreten. Pausen werden demnach in Abweichung von den HIAT-Transkriptionsrichtlinien (Rehbein et al., 2004, S. 34) auch bei parallel stattfindenden nonverbalen Handlungen der Sprecherin/des Sprechers oder anderer Sprecher/-innen transkribiert.

BS[v]	((19,8s))	So, ich gebe Ihnen mal was vor,	((1,0s))	und Sie sagen	mir mal,	was das soll.
BS[nv]	<pre>((schreibt auf Folie,20s))</pre>	((zeichnet auf Folie	e,4s))			

MIKO_E_00002, 01:20

_

⁸ Dabei handelt es sich um hörerseitige Äußerungen "ohne Anspruch auf Turnübernahme" (Rehbein et al. (2004), S. 19) und ohne propositionalen Gehalt, wie z. B. *mhm*, *hm* oder *ja*.

Pausen innerhalb von oder zwischen zwei Äußerungen werden ab einer Länge von 0,25s berücksichtigt. Dabei gibt es abweichend von HIAT (Rehbein et al., 2004, S. 35) eine zweistufige Transkription:

- Pausen mit einer Länge zwischen 0,25s und 1s: drei Pausenpunkte • •
- Pausen ab 1s werden gemessen, auf die nächste Zehntelsekunde gerundet und in einem eigenen Event transkribiert: ((3,2s))

```
CF[v] • • • Was ein typisches • • • ähm • • • Auto ist, zum Beispiel, hat sich in den letzten dreißig Jahren • • • sicherlich verändert.
```

MIKO E 00012, 03:30

Pausen zwei Äußerungen werden jeweils der zweiten Äußerung bzw. der/dem nächsten Sprecher/-in zugeteilt. Akustische Phänomene wie hörbares Atmen, Räuspern, Husten und Knarren werden ebenfalls als Pausen transkribiert.

Wenn vor der ersten Äußerung der Vorlesung keine relevante nonverbale Handlung stattfindet, wird das erste Wort als Beginn der kommunikativen Handlung angesehen und keine einleitende Pause notiert. Die Transkription beginnt mit dem ersten Wort der/des Dozierenden.

8.3 Besondere Wörter

8.3.1 Verzögerungspartikeln

Verzögerungspartikeln (*Planungsindikatoren*) wie *ähm*, *äh*, *hm* deuten auf einen Sprechplanungsprozess der Sprecherin/des Sprechers hin (Rehbein et al., 2004, S. 51). Sie werden daher immer dem nächsten Sprechereignis zugeordnet (neue Äußerung, Nebensätze, nach Abbrüchen oder Reparaturen, in Parenthesen etc.). Am Äußerungsanfang werden die Partikeln beginnend mit einer Majuskel und innerhalb von Äußerungen immer klein geschrieben.

Die Verzögerungspartikeln werden grundsätzlich nicht durch Kommata von der restlichen Äußerung getrennt.

```
BS[v] ••• Äh••• genau, Oberflächenvergrößerung, ja?
```

MIKO_E_00002, 41:44

Wenn die Syntax jedoch ein Komma erfordert, beispielsweise um Nebensätze abzugrenzen, werden der Syntax folgend Kommata gesetzt, wie die folgende Sequenz zeigt:

$CF[v]$ \bullet \bullet Die Werbung bombardiert uns mit \bullet \bullet neuen Wörtern, ähm weil \circ	sie…
--	------

MIKO_E_00012, 75:48

Bei Wiederholung von Verzögerungspartikeln werden diese (ergänzend zu Rehbein et al., 2004, S. 51) nicht durch Komma getrennt.

DH[v]

• • Das ist das Riesenproblem, weswegen das mit Island nicht funktioniert mit der/ • • • mit den, mit den billigen • • • äh Strom, den man dort **äh äh** produzieren kann, um dort große Computernetzwerke aufzubauen.

MIKO E 00005, 60:12

8.3.2 Interjektionen

Interjektionen (Empfindungswörter) werden bei MIKO reduziert transkribiert. Tonbewegungen und unterschiedliche Vokalqualitäten zur Abbildung von Bedeutungsdifferenzen werden im Gegensatz zu Rehbein et al. (2004, S. 31f.) nicht berücksichtigt. Oh, ah, nee, ach, na ja usw. werden durch Komma von der restlichen Äußerung getrennt.

DH[v] • • • Ah, das ist doch mal beeindruckend, he? Na ja, gut.

MIKO E 00004, 82:07

8.3.3 Abkürzungen

Zu den Abkürzungen zählen Kurzwörter im engeren Sinn sowie Schreibabkürzungen. Wenn sie in ihrer kurzen Form gesprochen werden, werden sie der Standardorthografie folgend ohne Leerzeichen notiert.

CF[v] • • • Aber weiß jemand, was es hieß, zu DDR-Zeiten?

MIKO_E_00012, 82:08

Vollständig gesprochene Schreibabkürzungen werden standardorthografisch ausgeschrieben. Eine Mischform stellt folgendes Beispiel dar:

DH[v] • • • Spulendurchmesser wäre dann ein Zentimeter, hundert Windungen, Eisenkern et cetera pp.

MIKO_E_00006, 52:26

Et cetera als ausgesprochene Form der Schreibabkürzung etc. wird ausgeschrieben. Das nachfolgende pp hingegen wird als Kurzform gesprochen, sodass hier die erste Regelung Anwendung findet.

8.3.4 Buchstabieren

Werden Wörter oder Silben buchstabiert bzw. einzelne Buchstaben verbalisiert und handelt es sich nicht um eine Abkürzung, so werden die Grapheme in Abweichung zu HIAT (Rehbein et al., 2004, S. 32) als Majuskeln und durch Leerzeichen getrennt ohne Komma notiert.

CF[v] • • • "Ist ein Fugenelement, das **E N** in 'Sprachenpolitik'."

 $\mathsf{BS}[\mathsf{v}] \qquad \bullet \quad \bullet \quad \mathsf{Und} \ \mathsf{in} \ \mathsf{Sachsen} \ \mathsf{ist} \ \mathsf{das} \ \mathsf{gef\"{a}hrlich} \ \mathsf{mit} \ \mathsf{dem} \ \mathsf{harten} \ \mathsf{und} \ \mathsf{dem} \ \mathsf{weichen} \ \mathbf{B}, \ \bullet \quad \bullet \quad \mathsf{ja}?$

MIKO E 00001, 56:58

Buchstaben in Formeln werden ebenfalls immer großgeschrieben, unabhängig von ihrer üblichen Schreibung in dem Fach.

DH[v] P ist gleich Q • • • mal • • • D.

MIKO E 00004, 27:48

Buchstaben aus anderen Alphabeten als dem lateinischen werden in ihrer laut Duden geltenden Schreibung notiert, z. B. Omega, Phi, Pi.

8.3.5 Zahlen

Kardinal- und Ordinalzahlen werden grundsätzlich ausgeschrieben.

BS[v] **Eins, zwei, drei, vier** Köpfe.

MIKO_E_00001, 12:58

CF[v]

• • • Kommt aus dem neunzehnten Jahrhundert, aus der/ ähm den Ursprüngen sozusagen der • • • äh äh germanistischen • • • äh Sprachbeschreibung.

MIKO_E_00012, 07:00

8.3.6 Fremdwörter und fremdsprachige Äußerungen

Wechselt ein/-e Dozierende in eine andere Sprache als Deutsch, so richtet sich die Schreibung der fremdsprachigen Wörter nach der Orthografie der jeweiligen Sprache. Dies betrifft auch Groß- und Kleinschreibung sowie Erweiterungen des lateinischen Alphabets. Es erfolgt abweichend zu Rehbein et al., 2004, S. 57ff.) keine Übersetzung der nicht-deutschen Sequenzen.

```
BS[v] Und - \cdot \cdot \cdot last, but not least - \cdot \cdot \cdot wir können natürlich auch rotieren, \cdot \cdot \cdot ja?
```

MIKO_E_00003, 12:43

CF[v] • • • Auch diese Lehnbedeutung stammt aus dem Englischen/ äh von dem englischen "to fire".

MIKO_E_00012, 70:22

Fachbegriffe aus anderen Sprachen wie beispielsweise Latein oder Griechisch werden so geschrieben, wie sie im Fachdiskurs üblicherweise verwendet werden, auch wenn die Groß- und Kleinschreibung u. U. nicht der Orthografie der Herkunftssprache folgt:

BS[v] Das ist mal unser kleiner Processus coracoideus und das ist die Facies glenoidalis.

MIKO E 00002, 13:06

8.3.7 Teilwörter

Bei Teilwörtern werden alle Wortteile mit Bindestrich verbunden, z. B. E-Mails, Café-und-Kneipen-Kultur, Druck- und Spannungslinien, Minimum-Maximum-Prinzip, Multiple-Choice-Aufgabe, Nord-Süd-Richtung.

8.3.8 Morpheme

Wenn über Morpheme gesprochen wird, z. B. Präfixe oder Suffixe, um Wissen über Wörter zu vermitteln, werden diese klein und ohne Bindestrich, jedoch mit Anführungszeichen notiert.

```
Ich wiederhole noch mal: • • • "Abduktion", • • • "Adduktion", also bitte • • • "BS[v] "ab" • • • und "ad" nicht verwechseln an der Stelle, das kommt nicht gut, weil es unterschiedliche Bedeutungen hat.
```

MIKO E 00002, 59:27

Am Äußerungsanfang beginnen sie jedoch mit einer Majuskel.

BS[v]	"Itis" ist	ja die	Endung für	"Entzündung".
-------	------------	--------	------------	---------------

MIKO_E_00002, 45:30

8.3.9 Eigennamen

Notation: NameFunktion

Die Identität Diskursbeteiligter soll durch Anonymisierung geschützt werden. Hierzu wird der Name durch die Kodierung Name+Funktion ersetzt, z. B. NameDozentIn, NameAssistentIn, NameForscherIn, um das Verständnis des Erwähnungszusammenhanges zu sichern. Zusätzlich werden die entsprechenden Sequenzen mit Rauschen unterlegt.

Als diskursbeteiligt gelten Personen und Institutionen, deren Identität nicht öffentlich bekannt ist. Namen historischer Fachvertreter/-innen, Namen von Personen des öffentlichen Lebens, Autor/-innen, Institutionen, Firmen usw., deren Identität allgemein bekannt ist bzw. über deren Identität Informationen öffentlich zugänglich sind, werden nicht anonymisiert.

8.4 Unverständliches und Zweifelsfälle

Sequenzen, die akustisch nur schwer, unsicher oder überhaupt nicht verständlich sind, werden wie folgt transkribiert:

Unverständliche Sequenzen bis 0,25s:

Notation: ((unv.))

BS[v] Sagen Sie es laut, ich ((unv.)) es nicht ganz…
--------------------------------	------------------------

Unverständliche Sequenzen ab 0,25s:

Notation: ((unverständlich)), ((unverständlich,2s))

Ab einer Länge von 1s wird die Zeitangabe auf ganze Sekunden gerundet mit Komma und ohne Leerzeichen notiert.

```
Wenn Sie mal auf einem Autofriedhof waren, • • • dann sehen Sie genau solche Magneten, • • • die dann benutzt werden, um die Leute • • • ähm ((unverständlich,3s)) derer entledigen, • • • die eben • • • ein bisschen lästig für Sie sind, ja?
```

MIKO E 00006, 49:20

Unverständliche Sequenzen am Äußerungsanfang:

Notation: ((Unv.)), ((Unverständlich))

BS[v]	• • • ((Unv.)) • • • Also wie das aussieht, ist ja egal, es geht ja nicht um Schönheit.
[-]	Schönheit.

MIKO_E_00002, 47:07

Unsicher verstandene Sequenzen (Zweifelsfälle), bei denen die/der Transkribent/-in eine Vermutung hat:

In der Verbalspur wird die zweifelhafte Sequenz in einfachen runden Klammern notiert.

CF[v]	(Wie	würde	das	klingen)?	
-------	------	-------	-----	-----------	--

MIKO_E_00012, 85:58

Unsicher verstandene Sequenzen (Zweifelsfälle), bei denen die/der Transkribent/-in alternative Vermutungen hat:

In der Verbalspur wird die am wahrscheinlichsten identifizierte Vermutung transkribiert und die zweifelhafte Sequenz in einfachen runden Klammern notiert. In der Spur X[wort] (Kapitel 10.3) werden die Vermutungen dann in einfachen runden Klammern und mit Schrägstrich getrennt notiert.

BS[v]	• • • So, das bedeutet, es ist		sinnvoll, • • • diese Muskelbäuche • • • in Sehnen auslaufen zu lassen, und die Sehnen dienen dann der Kraftübertragung.
BS[wort]		(jetzt ja/ja sehr)	

MIKO E 00001, 58:07

Endet eine Äußerung mit etwas Unverständlichem oder einem Zweifelsfall, ergibt sich auch hier die Notwendigkeit ein Äußerungsendzeichen zu vergeben. Ist der Modus der Äußerung beispielsweise aufgrund der Intonation deutlich zu erkennen, erhält die Äußerung das entsprechende Äußerungsendzeichen. Ist der Modus nicht eindeutig identifizierbar, schließt die Äußerung mit einem hochgestellten Punkt für Äußerungen ohne Modus.

8.5 Intrasegmentale Phänomene

Im Folgenden werden äußerungsinterne Phänomene beschrieben, die die Struktur der Äußerung betreffen.

8.5.1 Wiederholungen (Reihung und Repetition)

Notation:, (Komma)

Werden ein oder mehrere Wörter oder Äußerungsteile identisch wiederholt, kann dies sprecherseitig intendiert zur "Gewichtung, Intensitätssteigerung oder Aufzählung" (Rehbein et al., 2004, S. 46) (Reihung) oder als Ausdruck des Sprechplanungsprozesses (Repetition) geschehen. Die wiederholten Wörter oder Äußerungsteile werden in beiden Fällen mit Komma getrennt.

	CETVI	Ähm • • • äh und • • • äh auch Tiere, die einem, die einem nicht begegnen, • • • fallen einem nicht so schnell ein.
ļ	CF[V]	fallen einem nicht so schnell ein.

MIKO_E_00012, 11:02

Das Komma zwischen Wiederholungen entfällt, wenn ein Verzögerungspartikel zwischen den Wiederholungen realisiert wird.

DH[v]	Auch wenn der Effekt minimal ist, minimal klein ist, • • • ist das ein ganz
ויחונין	entscheidender Effekt, der es uns erlaubt, • • • Funktionen im Gehirn • • • über

		DH[v]	MRT	•	•	•	ähm	•	•	•	über	MRT	zu	((1,6s))	detektieren.
--	--	-------	-----	---	---	---	-----	---	---	---	------	-----	----	----------	--------------

MIKO_E_00006, 89:19

Wie bereits in Kapitel 8.3.1 erläutert, wird bei der Wiederholung von Verzögerungspartikeln kein Komma notiert.

```
Und • • • wir hören das ja immer wieder, dass eine ganze Menge • • • Unfälle/ dass DH[v] da Patienten zu Tode kommen, • • • weil Zivilisten • • • Angst haben, sich schnell genug ähm ähm zu, zu äh zu beteiligen, den Menschen zu helfen.
```

MIKO_E_00006, 12:17

8.5.2 Sprechhandlungsaugmente

Nachgelagerte Augmente wie *ja*, *oder*, *ne*, die das Gesagte als fraglich oder unsicher kennzeichnen, werden mit Komma von der restlichen Äußerung getrennt.

CF[v]	Die meinten Sie, ne?	((2,5s))	Und nur dreiundzwanzig Prozent waren dezidiert
CF[v]	anderer Meinung, • • •	ja?	

MIKO_E_00012, 47:53

Sie können auch innerhalb von Äußerungen vorkommen.

BS[v] Hier ist es auch viel glatter, **ja**, an der Stelle, als der Rest.

MIKO_E_00001, 14:26

8.5.3 Herausstellungen und Ausklammerungen

Herausstellungen sind der Äußerung nach vorn oder hinten ausgelagerte Äußerungsteile, die außerhalb des Satzrahmens ein Element des Satzrahmens erneut aufgreifen. Sie werden mit Komma abgetrennt.

BS[v] Das ist also nicht zufällig, diese Anordnung.

MIKO E 00001, 23:14

SL[v]

• • • Die Bank selbst hat über den Kaufvertrag • • • gegenüber dem Lieferanten • • Mängelgewährleistungsrechte, • • • eben für dieses Auto.

MIKO E 00007, 67:14

Ausklammerungen von Satzgliedern ohne Aufgreifen eines Elements des Satzrahmens werden hingegen ohne Komma notiert.

CF[v] Fallen Ihnen welche ein • • • in der Linguistik?

MIKO_E_00012, 06:01

8.5.4 Parenthesen

Notation: - (m-dash, Gedankenstrich)

Wird eine Äußerung durch einen Einschub (Parenthese) unterbrochen und der Matrixsatz danach weitergeführt, wird der Einschub mit Gedankenstrichen und Leerzeichen vor und nach den Strichen von der übrigen Äußerung getrennt. In der Parenthese sind keine Äußerungsendzeichen möglich, auch wenn die Parenthese eine vollständige Äußerung mit einem identifizierbaren Modus enthält. Die Interpunktion des Matrixsatzes wird abweichend von HIAT (Rehbein et al., 2004, S. 44) standardorthografisch nach der Parenthese notiert (Kommata bei Nebensätzen, Doppelpunkt bei Ankündigungen). Pausen vor Parenthesen werden in der Parenthese notiert, Pausen danach werden dem Matrixsatz zugeordnet.

DH[v]	Warum kann ich die • • • Stärke • • • des Lichtes oder eben des Stromes, • • • der				
	Warum kann ich die • • • Stärke • • • des Lichtes oder eben des Stromes, • • • der durch diese Spule – • • • ach, jetzt ist hier wieder aus – • • • fließt,				

DH[v] ((6,2s)) steuern?

MIKO_E_00006, 90:15

Werden Wörter oder Äußerungsteile des Matrixsatzes nach der Parenthese wiederholt, entfällt das Komma für Wiederholungen (analog zu den Verzögerungspartikeln, Kapitel 8.3.1).

```
DH[v]

• • Denn X Quadrat ist ja – und dann steht hier noch das Minus – denn X Quadrat
• • • ist ja sozusagen der Abstand im Quadrat.
```

MIKO_E_00004, 45:23

Auch bei nicht völlig identischen Wiederholungen von Wörtern oder Teilen des Matrixsatzes nach der Parenthese wird kein Komma gesetzt. Diese Fälle gelten nicht als Reparatur, da die Wiederholung als Wiederaufgreifen des Matrixsatzes nach einer zeitlichen Verzögerung durch die Parenthese gewertet wird und keine tatsächliche Korrektur stattfindet.

```
BS[v] Und jetzt schreibe ich – das mache ich selten, weil ich immer viel zu faul bin – aber jetzt schreibe ich einmal einen Merksatz auf, weil das wichtig ist.
```

MIKO E 00003, 19:27

Es kommt vor, dass Anführungszeichen verbalisiert werden, um das Gesagte als ironisch, metaphorisch o. Ä. zu kennzeichnen. In diesem Fall wird die Verbalisierung der Anführungszeichen als Parenthese notiert.

```
BS[v] • • • Und dieser Strick — in Anführungsstrichen — ist natürlich kein Strick, sondern ein Nerv, • • • den Sie eigentlich schon alle kennen.
```

MIKO E 00001, 02:40

8.5.5 Ankündigungen

Notation: : (Doppelpunkt)

Ankündigungen innerhalb einer Äußerung werden mit Doppelpunkt notiert. Angekündigt werden können unter anderem wiederzugebende Äußerungen (direkte Rede, Zitate etc.), Aufzählungen, Zusammenfassungen, Schlussfolgerungen, Überschriften oder Themen.

```
SL[v] • • • Frage eins zu diesem Fall: Was kann Luisa tun, um den bereits überwiesenen Kaufpreis erstattet zu bekommen?
```

MIKO_E_00007, 12:51

```
DH[v] • • • Einheit: Newton pro Coulomb, • • • ja?
```

MIKO E 00004, 09:02

Pausen vor dem Angekündigten werden nach dem Doppelpunkt notiert.

BS[v]	Und dann • • • sind hier die Begriffe: • • • "Substantia	((1,2s))	compacta"

MIKO E 00001, 22:04

Handelt es sich bei dem Angekündigten um in Anführungszeichen gesetzte uneigentliche Rede, werden die Pausenpunkte nach dem Doppelpunkt und vor den Anführungszeichen notiert.

```
SL[v] Der Kommilitone hat jetzt gesagt: • • • "Es kommt immer dadrauf an, was die Parteien eigentlich wollen.", • • • ja?
```

MIKO_E_00007, 06:08

Nachgestellte Erläuterungen, die mit das heißt, das ist, das bedeutet etc. eingeleitet werden, werden mit Komma und nicht mit Doppelpunkt notiert.

MIKO_E_00001, 21:46

Ein Grenzfall zwischen Ankündigung und Nebensatz zeigt sich in folgendem Beispiel:

DH[v] Sie wissen, Leistung ist Spannung mal Stromstärke.

MIKO E 00005, 05:22

Sie wissen kann als Ankündigung im Sinne von Sie wissen Folgendes: verstanden werden, es kann aber auch die Einleitung eines Nebensatzes Sie wissen, dass sein. Im Beispiel folgt jedoch ein Nebensatz mit Verbzweitstellung. Um einheitliche Transkription zu gewährleisten, wurde folgende Regel aufgestellt: Macht die/der Sprecher/-in nach Sie sehen, Sie merken, Sie wissen, Sie erinnern sich eine Pause von mindestens 0,25s, wird ein Doppelpunkt notiert. Wird die Äußerung unmittelbar fortgeführt, wird ein Komma notiert.

8.5.6 Reparaturen

Notation: /(slash, Schrägstrich)

Es liegt eine Reparatur vor, wenn die/der Sprecher/-in eine äußerungsinterne Korrektur, also sprachliche Verbesserung innerhalb einer vollständigen Äußerung, vornimmt. Kennzeichnend dafür ist der Beginn einer Äußerung, das anschließende Unterbrechen an einer reparaturbedürftigen Stelle und die abschließende Korrektur durch Fortführung der begonnenen Äußerung. Die Reparatur kann, muss aber nicht durch Pausen oder Verzögerungspartikeln gekennzeichnet sein.

Das zu reparierende Element (Reparandum) muss nicht falsch sein, es kann sich auch um eine Präzisierung des Geäußerten handeln. Ausschlaggebend für die Notation einer Reparatur ist hierbei die Reparaturabsicht der Sprecherin/des Sprechers (Rehbein et al., 2004, S. 53), die sich beispielsweise auch durch die Intonation ausdrücken kann. Auch Fehlstarts sind eine Form der Reparatur.

Hinter dem Reparandum wird ein Schrägstrich notiert, danach ein Leerzeichen. Dann folgt das reparierte Wort oder die reparierte Äußerung mit Kleinschreibung, da es sich nicht um den Beginn einer neuen Äußerung handelt. Dies gilt auch, wenn ein Satzanfang repariert wird.

DH[v] • • • Im el/ wir können das elektrische Feld nachweisen durch eine Kraftwirkung auf Probekörper.

MIKO_E_00004, 12:53

Verzögerungspartikeln können nicht repariert werden. Geht die Reparatur eines Wortes mit der Realisation eines Verzögerungspartikels einher, wird der Schrägstrich hinter dem Reparandum notiert. Es wird also niemals *ähm*/ transkribiert.

DH[v] Dann kommen wir heute **zum/ ähm** zu unserer Vorlesung.

CE[v]	Das	ist	eben	das,	was/	•	•	wann/	äh	was	besonders	häufig	und	besonders	relevant	
CF[v]	ist.															

MIKO_E_00012, 04:31

Reparaturen innerhalb eines Wortes

Beginnt die/der Sprecher/-in die Realisation eines Wortes, unterbricht innerhalb des Wortes und korrigiert dann nicht das vollständige Wort, sondern nur den reparierten Wortbestandteil, so wird das Wort zunächst in der Verbalspur normalisiert transkribiert. Anschließend erfolgt die Notation der Reparatur in der Spur X/wort/ (Kapitel 10.1) wie folgt:

CF[v]	Das ist keine • • • eindeutige • • • und leicht abgrenzbare/ • • äh leicht	abgrenzbares	Konzept, ja?
CF[wort]		abgrenzga/ -bares	

MIKO_E_00012, 36:11

Anders verhält es sich, wenn ein/-e Sprecher/-in an einer defizitären Stelle im Wort unterbricht und dann das ganze Wort wiederholt. Hier wird die Reparatur in der Verbalspur notiert.

DH[v]	• • • Die Einheit der Indiktiv/ Induktivität ist Voltsekunde • • • pro Ampere.
-------	--

MIKO E 00006, 47:09

Reparaturen und zulässige Alternativen

Wird eine Reparatur mit *oder* eingeleitet, ist inhaltlich und sprachlich nicht immer eindeutig, ob es sich um eine Reparaturabsicht oder eine Alternative handelt, die die/der Sprecher/-in als ebenso korrekt empfindet wie das vermeintliche Reparandum. In diesen Fällen wird die Intonation (z. B. erhöhtes Sprechtempo, Zögern, Pausen) als ausschlaggebendes Kriterium herangezogen.

	Wir, wir, wir äh könnten auch noch • • • weniger typische • • • Elemente/ oder
CF[v]	Vertreter, die • • • nur einige dieser Elemente/ oder Merkmale erfüllen, • • • noch
	dieser Kategorie zuordnen.

MIKO_E_00012, 08:30

((haspelt))

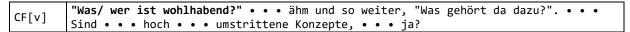
Die Notation ((haspelt)) erfolgt, wenn mehrere Laut- oder Silbenreparaturen zu schnell und zu undeutlich sind, um konkrete Laute, Lautfolgen oder Wörter zu identifizieren.

DC [v]	Ja, • • • nein, wissen.	((haspelt))	dann könner	Sie es	vielleicht	wirklich noch	nicht
ps[v]	wissen.						

MIKO_E_00001, 46:48

Reparaturen bei uneigentlichem Sprechen

Tritt eine Reparatur innerhalb einer Sequenz uneigentlichen Sprechens auf, folgt die Notation den festgelegten Richtlinien.



MIKO_E_00012, 16:06

Fällt die Reparatur an einer Wortgrenze mit den Anführungszeichen zusammen, ist zu entscheiden, ob die Reparatur auf Matrixsatzebene erfolgt oder Teil der uneigentlichen Rede ist. Im folgenden

Beispiel erfolgt die Reparatur auf Ebene des Matrixsatzes, sodass erst die schließenden Anführungzeichen und dann der Schrägstrich notiert werden.

```
CF[v] | • • Ähm • • • aber • • ich habe zum Beispiel in Leipzig/ als ich nach Leipzig gezogen bin, gelernt, • • • was "Fahrgatz"/ was "Fahrgastwunsch" eigentlich heißt, • • • ja?
```

MIKO E 00012, 76:58

Identifikation des Reparandums

Sprachlich und inhaltlich ist in Ausnahmefällen nicht ersichtlich, welches Wort das Reparandum ist. In diesen Fällen ist die Intonation (z. B. erhöhter Sprechdruck) ausschlaggebend. Im folgenden Beispiel könnte sowohl ein als auch präzise das Reparandum sein. Der erhöhte Sprechdruck bei präzise lässt aber vermuten, dass das Adjektiv die Reparatur ist.

```
• • • Na ja, jetzt müssen wir eigentlich uns wieder ein/ präzise überlegen, dass wir hier • • • F einen Vektor haben, V einen Vektor • • • und B einen Vektor.
```

MIKO_E_00006, 60:51

Orthografie des Reparandums

Bei Reparaturen nach nur wenigen Lauten unterliegt die Orthografie des transkribierten Reparandums einer gewissen Spekulation darüber, was die/der Sprecher/-in eigentlich sagen wollte und welche Wortart das unterbrochene Wort haben könnte. Anhand des Kontextes und der im Anschluss korrigiert realisierten Wortform muss die wahrscheinlichste Option identifiziert werden.

	• • • Ähm • • • Jugendsprachen oder jugendsprachliche Register — es gibt ja nicht
CF[v]	"die Jugendsprache", sondern viele • • • Register • • • der Jugendsprache – dient
	genau dazu, • • • sich • • • von den Erwachsenen und von der etablierten

CF[v]	Gesellschaft erst mal zu	((1,0s))	unterscheiden, sich abzugrenzen, s/ eine eigene • • • S/ ein eigenes Selbstwertgefühl und eine eigene Gruppenidentität aufzubauen.
-------	-----------------------------	----------	--

MIKO_E_00012, 71:49

Bei der ersten Reparatur sich abzugrenzen, s/ eine eigene ist es am wahrscheinlichsten, dass die Aufzählung reflexiver Verben fortgeführt werden sollte, weshalb das Reparandum sich angenommen wird und Kleinschreibung erforderlich ist. Die zweite Reparatur eine eigene • • • \$ / ein eigenes Selbstwertgefühl lässt darauf schließen, dass Selbstwertgefühl das Reparandum ist, die Kasuskongruenz jedoch nicht gegeben war, weshalb CF eine Reparatur einleitet. In diesem Fall wird also ein Nomen angenommen, welches Großschreibung erfordert.

8.6 Uneigentliches Sprechen

Notation: "" (Anführungszeichen)

Zu den Formen uneigentlichen Sprechens gehören laut HIAT (Rehbein et al., 2004, S. 47) "zitierende sprachliche Handlungen" (Kapitel 8.6.1) sowie "metasprachliche Äußerungen" (Kapitel 8.6.2). Beide werden mit Anführungszeichen notiert. Ergänzend zu HIAT werden bei MIKO auch Überschriften, Titel von Büchern oder Aufsätzen, Themen- und Kategorienbenennungen in Anführungszeichen gesetzt (Kapitel 8.6.3).

8.6.1 Zitierende Handlungen

Zitierende Handlungen sind aus dem Gedächtnis wiedergegebene bzw. vorgelesene Zitate, die (zum Teil auch hypothetische) Wiedergabe von direkter Rede oder Sprichwörter, die eindeutig als solche von der/dem Sprecher/-in gekennzeichnet sind. Ausschlaggebend ist die Zitierabsicht der Sprecherin/des Sprechers, die Korrektheit des Wortlautes ist dabei nicht von Bedeutung und wird "als authentisch qualifiziert" (Rehbein et al., 2004, S. 47).

Eingrenzung der als Zitat behandelten Phänomene

Da es sich bei Vorlesungen i. d. R. um textgestützte und oft skriptbasierte Vortragssituationen handelt und die Bezugnahme des Sprechers/der Sprecherin auf Folien oder Notizen nicht eindeutig erkennbar ist, wird das Vorlesen abweichend von HIAT (Rehbein et al., 2004) nur dann in Anführungszeichen gesetzt, wenn es sich um zitierte Äußerungen im engeren Sinne handelt. Text, der lediglich von einem Präsentationsmedium vorgelesen wird, wird bei MIKO nicht als Zitat behandelt, es sei denn, der Text ist auf dem Medium als Zitat einer anderen Quelle gekennzeichnet.

```
• • • Ähm • • • Thomas Alva Edison s/ spricht/ ((haspelt)) sagt Folgendes: • • • DH[v]

"Ein einheitlicher Druck • • • ist entscheidend für eine gleichmäßige Beleuchtung.".
```

MIKO E 00005, 06:20

Zuweilen kann nicht eindeutig bestimmt werden, ob es sich um eine wörtliche (direkte) oder sinngemäße (indirekte) Wiedergabe einer Äußerung handelt. Das ist besonders dann der Fall, wenn die Äußerung in einen Matrixsatz eingebettet ist, aber in Hauptsatzstellung wiedergegeben wird, da in gesprochenem Deutsch auch Nebensätze Verbzweitstellung aufweisen können, wobei bei dass-Sätzen die Konjunktion wegfällt. Dadurch ist eine Unterscheidung zwischen einem Haupt- und einem Nebensatz und somit zwischen einer wörtlichen und einer sinngemäßen Redewiedergabe nicht immer möglich. In diesen Fällen wird ein wörtliches Zitat nur dann angenommen, wenn die/der Sprecher/-in es durch andere (z. B. prosodische oder nonverbale) Mittel als solches kennzeichnet.

CF[v]	Gut, und äh ich hatte letztes Mal auch schon gesagt, ähm Prototypen sind
CF[V]	anschaulich.

MIKO E 00012, 04:38

Ein spezieller Fall zitierender Handlung entsteht, wenn die/der Dozierende in Frage-Antwort-Interaktion mit einzelnen Studierenden tritt. Zuweilen werden die Redebeiträge der Studierenden paraphrasiert oder wortgenau von der/dem Dozierenden wiederholt. Diese Äußerungen, Teiläußerungen oder Einzelwörter werden nicht als Zitate notiert.

SL[v]		<pre>Genau, Sie würden als allererstes schauen, ist es ein • • • anfänglicher oder ein nachträglicher Mangel. Zu welcher • • • Lösung würden Sie kommen?</pre>
NN[v]	• • • ((Student/-in antwortet auf Frage,4s))?	

SL[v]		Genau, ein nachträglicher Mangel. Warum?	
NN[v]	<pre>((Student/-in antwortet auf Frage)).</pre>		((Student/-in antwortet auf Frage,4s)).

Lediglich in solchen Fällen, in denen die Zitiertheit wie bei direkter Redewiedergabe expliziert wird, wird die Äußerung als zitierende Handlung notiert.

CL [v]	• • • Der Kommilitone hat das richtig gesagt: "'Fernabsatzverträge' kennzeichnet
SL[v]	sich dadurch, • • • dass man irgendwelche Fernkommunikationsmittel benutzt.".

MIKO_E_00007, 01:20

Notation zitierender Handlungen

Sind zitierte Äußerungen vollständige Äußerungen, so erfordern sie Großschreibung am Äußerungsbeginn und ein Äußerungsendzeichen innerhalb der Anführungszeichen. Ist die zitierte Äußerung in einen Matrixsatz eingebettet, erhalten sowohl zitierte Äußerung als auch Matrixsatz ein Äußerungsendzeichen.

		((1,0s))	typischerweise kommt, wenn ich sage: "Ich habe meine Familie eingeladen, morgen ist, ist mein dreißigster
	eigentlich		Geburtstag, ich habe meine Familie eingeladen.".

MIKO_E_00012, 38:34

Steht die zitierte Äußerung eigenständig ohne Einbettung in einem Matrixsatz, erhält sie nur ein Äußerungsendzeichen innerhalb der Anführungszeichen.

```
CF[v] "Der Einfluss anderer Sprachen auf die deutsche Sprache nimmt stark zu."
```

MIKO_E_00012, 49:40

Pausen bzw. Verzögerungspartikeln am Anfang von zitierten Äußerungen werden nicht dem Zitat zugeordnet und daher vor den eröffnenden Anführungszeichen notiert.

```
• • • Ähm äh wenn ich, wenn ich lese: Ähm "Südafrika ist ein bisschen Afrika, Europa und Kalifornien.", • • • dann muss ich mir daraus irgendwas zusammenbauen.
```

MIKO E 00012, 26:00

Pausen bzw. Verzögerungspartikeln sowie Reparaturen innerhalb von zitierten Äußerungen sind zulässig.

```
| • • • Auf jeden Fall stand das da: • • • "Auf den/ in, in den Sch/ in den Straßen und den Straßencafés • • • ähm herrscht ein italienisches Flair.", • • • ja?
```

MIKO_E_00012, 24:52

Die Orthografie vorgelesener Zitate wird an die neue deutsche Rechtschreibung laut Duden angepasst, z. B. Rückschluss statt auf der PowerPoint-Folie sichtbar Rückschluß (MIKO_E_00012, 54:20).

8.6.2 Metasprachliche Verwendung von Begriffen

Rehbein et al. (2004, S. 48) unterscheiden zwischen Sach- und Ausdrucksebene eines Begriffes. Die Sachebene bezieht sich auf die inhaltliche Seite des Wortes, die Ausdrucksebene auf die sprachliche Form des Wortes. Bei MIKO gilt die metasprachliche Verwendung eines Begriffes als dann gegeben, wenn primär über den Begriff als *signifiant* gesprochen wird und nicht oder nur sekundär über das Referierte (*signifiè*).

DC [v/]	• • • Und die Grube, die hier also oberhalb liegt — • • • A nur, weil "Fossa" eben auch weiblich ist — • • • äh heißt "supraspinata".
pS[v]	eben auch weiblich ist — • • • äh heißt "supraspinata".

MIKO_E_00001, 09:28

Auch Übersetzungen geäußerter Begriffe in andere Sprachen gelten als metasprachliche Verwendung von Begriffen.

```
BS[v] Also "Compacta" ist diese äußere, kräftige • • • Schicht • • • der Diaphyse, des "Schafts", wie Ihre Kollegin sagte.
```

MIKO_E_00001, 22:17

Werden Begriffe innerhalb einer zitierten Äußerung metasprachlich/uneigentlich verwendet, werden diese Begriffe ergänzend zu HIAT (Rehbein et al., 2004) in einfache Anführungszeichen⁹ gesetzt.

```
CF[v] Ich habe gesagt: • • • "Gibt es nicht, ja? • • • Ist ein Fugenelement, das E N in 'Sprachenpolitik'.".
```

MIKO_E_00012, 17:16

Bestimmung der metasprachlichen und nicht-metasprachlichen Verwendung von Begriffen

Wenn (Fach-)Begriffe und bestimmte Phänomene einander zugeordnet werden, hängt es vom Kontext und der Intention der Sprecherin/des Sprechers ab, ob es sich um eine metasprachliche Verwendung des Begriffs handelt oder nicht. Grundsätzlich lassen sich (1) die Zuordnung eines Begriffs zu einem (konkreten) Phänomen und (2) die Zuordnung eines (konkreten) Phänomens zu einem Konzept unterscheiden. Im ersten Fall wird der Begriff metasprachlich verwendet, im zweiten Fall nicht.

(1) Zuordnung eines Begriffs zu einem (konkreten) Phänomen

Die/der Sprecher/-in benennt ein konkretes Phänomen mit der Absicht, einen Fachbegriff erstmals einzuführen oder ihn zu wiederholen. Dem Phänomen wird also ein Begriff zugeordnet, die implizite Bedeutung der Aussage lautet: "Dieses Phänomen heißt/wird bezeichnet mit dem Wort X". Damit handelt es sich um eine metasprachliche Verwendung des Begriffs X, welcher somit durch Anführungszeichen gekennzeichnet wird.

DC [v]	Wie, wie hieß nochmal der Fortsatz? Jetzt habe ich den wieder vergessen.
PS[A]	((1,2s)) "Rabenschnabelfortsatz", genau, "Processus coracoideus".

MIKO_E_00001, 04:31

DH[v] Wir haben eine Ladung, • • • die bezeichnen wir mit "groß Q".

MIKO_E_00004, 41:31

(2) Zuordnung eines (konkreten) Phänomens zu einem Konzept

Die/der Sprecher/-in nennt einen Begriff mit der Absicht, ein Beispiel für das vom Begriff bezeichnete Konzept – die Abstraktion eines konkreten Phänomens – an einem Objekt zu identifizieren bzw. identifizieren zu lassen. Dem Konzept – versprachlicht durch seine

⁹ Da einfache Anführungszeichen im HIAT-Transkriptionssystem nicht vorgesehen sind, führen diese zu technischen Schwierigkeiten bei der automatischen Verarbeitung in EXMARaLDA, etwa bei der Lemmatisierung. Dies ist bei der Arbeit mit den tokenisierten Transkripten zu beachten.

Bezeichnung, d. h. durch den Begriff – wird also ein konkretes Phänomen zugeordnet. Die implizite Bedeutung der Aussage lautet: "Dieses konkrete Phänomen ist ein Beispiel für das Konzept X". Die Zuordnung des Begriffs zum Konzept wird hier, anders als bei der metasprachlichen Verwendung des Begriffs, vorausgesetzt.

Die Dozentin liest im folgenden Beispiel einen lateinischen Begriff von der Overhead-Folie vor, den sie sodann ins Deutsche übersetzt. Durch den Kontext wird allerdings deutlich, dass es hier nicht primär um die Übersetzung des lateinischen Begriffs geht, sondern darum, Beispiele für bestimmte Phänomene zu finden, in diesem Fall für einen *Sulcus*. Damit wird *Sulcus* hier nicht metasprachlich verwendet.

```
BS[v] • • • Vielleicht nur ein Beispiel noch: Schauen Sie mal: Sulcus, • • • "die Furche".
```

MIKO_E_00001, 01:54

In manchen Fällen kann auch unter Berücksichtigung des Kontexts nicht eindeutig bestimmt werden, ob es sich um eine metasprachliche Verwendung eines Begriffs handelt oder nicht, d. h. es ist unklar, ob die/der Sprecher/-in einem konkreten Phänomen einen Begriff zuordnet oder einem abstrakten Konzept ein konkretes Beispiel. Eine metasprachliche Verwendung wird dann angenommen, wenn der Begriff in der aktuellen Vorlesungssitzung noch nicht vorkam, sodass nicht vorausgesetzt werden kann, dass den Rezipient/-innen die Zuordnung eines Begriffs zu einem Phänomen geläufig ist, was jedoch eine Voraussetzung für die nicht-metasprachliche Verwendung des Begriffs wäre.

DC [1/]	Das ist doch - ja, ich merke schon, Sie wissen, worauf ich hinauswill - • • • so
po[v]	ein unangenehmes Gefühl, ja? • • • Und das ist der "Nervus ulnaris".

MIKO_E_00001, 02:51

8.6.3 Überschriften, Titel, Themen- und Kategorienbenennungen

Überschriften (z. B. von Kapiteln), Publikationstitel sowie Benennungen von Themen, Kategorien und Konzepten werden ergänzend zu HIAT (Rehbein et al., 2004) als uneigentliche Äußerungen aufgefasst und mit doppelten Anführungszeichen gekennzeichnet. Die Notation erfolgt nach den Regeln der Notation zitierender Handlungen (Kapitel 8.6.1).

```
CF[v] Gut, das ware ahm • • • das Thema • • • "Prototypensemantik".
```

MIKO E 00012, 40:29

CE[v]	• • • Ähm aber • • • wir sehen, die Kategorie "Tier" ist noch relativ, • • • relativ abstrakt.
CF[V]	relativ abstrakt.

MIKO E 00012, 11:09

8.7 Nicht-phonologische akustische Phänomene

Notation: ((lacht)), ((lacht))

Nicht redebegleitende nicht-phonologische akustische Phänomene der Sprecherin/des Sprechers wie Lachen werden in doppelten runden Klammern und in der dritten Person notiert. Ab einer Dauer von drei Sekunden wird die Zeit auf ganze Sekunden gerundet in den Klammern vermerkt, getrennt mit einem Komma, aber ohne Leerzeichen, z. B. ((lacht,4s)). Außerhalb einer verbalen Äußerung erhält das Phänomen das Äußerungsendzeichen für Äußerungen ohne Modus (hochgestellter Punkt). Die Kleinschreibung wird in diesem Fall beibehalten, da es sich nicht um die Transkription einer verbalen Äußerung handelt.

BS[v]	Das glaube ich nicht!	((lacht))	Selbst di	e zarteste	Person	kann	das	so	nicht	
D2[V]	machen.									

MIKO E 00003, 03:26

Gefüllte Pausen, hörbares Atmen, Knarren, Husten, Räuspern, Schreibgeräusche etc. der Sprecherin/des Sprechers werden nicht berücksichtigt und als Pausen transkribiert (Kapitel 8.2).

9 Besondere Betonungen *X*[akz]

Notation: - (Bindestrich)

In der Betonungsspur werden besondere Betonungen annotiert, die über die reguläre Satzakzentuierung der Sprecherin/des Sprechers hinausgehen. Diese werden auditiv ermittelt, wobei erhöhte Lautstärke, stark abweichende Tonhöhe und auffällige Lautdehnung als Kriterien dienen.

Die Annotation der Betonung erfolgt auf Wortebene, auch wenn nur eine Silbe des Wortes die Betonung trägt. Das besonders betonte Wort wird in einem eigenen Event abgegrenzt, wobei das zeitliche Alignment an den Eventgrenzen berücksichtigt wird. Die Notation erfolgt mit einem einfachen Bindestrich in dem Event.

BS[v]	Das hängt miteinander zusammen, • • • aber es ist nicht das	Gleiche.
BS[akz]		-

MIKO_E_00001, 01:18

Mehrere betonte Wörter hintereinander werden durchgehend annotiert.

BS[v]	• • • Es gibt also	immer nur	• • • magnetische • • • Dipole, es gibt	keine
BS[akz]		-		-

	BS[v]	magnetischen	Monopole.
--	-------	--------------	-----------

MIKO_E_00006, 30:54

Pausen zwischen betonten Wörtern werden durchgehend mit der Annotation versehen, Pausen vor oder nach betonten Abschnitten werden jedoch nicht mit annotiert.

10 Annotationen auf Wortebene X/wort]

In der Spur X[wort] werden Annotationen vorgenommen, die die Wortebene betreffen. Dazu zählen Versprecher, Pausen innerhalb von Wörtern sowie Vorschläge bei auditiv nicht eindeutig identifizierbaren Wörtern. Diese Spur übernimmt in begrenztem Umfang die Funktion der bei Rehbein et al. (2004) vorgesehenen sprecherspezifischen Kommentarspur.

10.1 Versprecher

Auditiv eindeutig identifizierbarer Versprecher, der von der/dem Sprecher/-in nicht korrigiert wird:

Das Wort wird in der Verbalspur normalisiert und in der Spur X[wort] in seiner abweichenden Realisation notiert. Dafür wird das Wort in einem eigenen Event abgegrenzt.

BS[v]	Ich	zeichne	die	also	von	frontal.
BS[wort]					vron	

MIKO E 00002, 07:06

Auditiv nicht eindeutig identifizierbarer Versprecher, der von der/dem Sprecher/-in nicht korrigiert wird:

Das undeutlich artikulierte Wort wird in der Verbalspur normalisiert, es erfolgt keine weitere Annotation.

Versprecher, der von der/dem Sprecher/-in korrigiert wird, wobei im Rahmen der Reparatur das ganze Wort korrekt realisiert wird (Kapitel 8.5.6):

Dieser Fall wird als Reparatur behandelt und entsprechend in der Verbalspur notiert.

Versprecher, der von der/dem Sprecher/-in korrigiert wird, wobei im Rahmen der Reparatur nur der korrigierte Wortteil, nicht jedoch das ganze Wort korrekt realisiert wird (Kapitel 8.5.6):

Das Wort wird in der Verbalspur normalisiert transkribiert. In der Spur X[wort] wird anschließend die Reparatur mit Schrägstrich, Leerzeichen und Bindestrich für Teilwörter notiert.

DH[v]	Wenn wir keinen	Permanentmagneten	haben, • • • dann können wir • • • Magnetfelder durch Stromflüsse erzeugen.
DH[wort]		Permanentmagneter/ -magneten	

MIKO_E_00006, 41:09

10.2 Pausen im Wort

Unterbrechungen innerhalb von Wörtern werden ab 0,25s notiert. Das Wort wird in der Verbalspur normalisiert und in der Spur X[wort] mit Teilwortstrichen und Pausenzeichen annotiert.

Pausen zwischen 0,25s und 1s:

BS[v]	• • • Also heißt das Ding "Sulcus • • •	intertubercularis".
BS[wort]		inter- • • • -tubercularis

MIKO_E_00001, 07:18

Pausen ab 1s:

DH[v]	Dielektrizitätszahl.
DH[wort]	Dielek- ((1,8s)) -trizi- ((3,0s)) -tätszahl

MIKO E 00004, 21:42

Genauso verhält es sich bei Verzögerungspartikeln innerhalb von Wörtern.

CF[v]	Also beispielsweise in literarischen Texten oder wenn ich ein	Wortfindungsproblem	habe.
CF[wort]		Wortfindungs- äh -problem	

MIKO_E_00012, 65:01

Pausen oder Verzögerungspartikeln bei Auslassungen (z. B. Fach- und Wissenschaftssprache, vier- und fünfundzwanzig) werden in der Verbalspur vermerkt.

DH[v] Und dieses Magnetfeld ist natürlich wieder ((2,5s)) durch einen • • • magnetischen Nord- • • • und einen • • • magnetischen Südpol • • • gekennzeichnet.
--

MIKO_E_00006, 60:02

10.3 Vorschläge bei schwer Verständlichem

Wie in Kapitel 8.4 erläutert, werden akustisch nicht eindeutig identifizierbare Passagen (mindestens ein Wort), bei denen die/der Transkribent/-in mehrere Vorschläge angeben möchte, um welches Wort oder welche Wortgruppe es sich handeln könnte, wie folgt notiert: In der Verbalspur wird die am wahrscheinlichsten identifizierte Vermutung transkribiert und die zweifelhafte Sequenz in einfachen runden Klammern notiert. In der Spur X[wort] werden die Vermutungen dann in einfachen runden Klammern und mit Schrägstrich getrennt notiert.

BS[v]	• • • Sie müssen das dann	(nur)	wirklich draußen in der Bewegung machen.
BS[wort]		(nur/nun/nu auch)	

MIKO_E_00001, 27:46

Bei Abschnitten, die so unverständlich sind, dass der Wortlaut nur erraten werden kann, wird auf Vermutungen verzichtet und stattdessen ((unverständlich)) transkribiert (Kapitel 8.4).

11 Sprechweise *X*/*sprech*/

Besondere Intonation, die vom normalen Stil und Stimmklang der Sprecherin/des Sprechers abweicht und der besonderen Hervorhebung dient, bedeutungsverändernd wirkt oder das Verständnis der Vorlesung beeinflusst, wird in der Spur X[sprech] in der Form des Partizip I bzw. als Adjektiv in doppelten runden Klammern notiert. Hierzu zählen ausgewählte prosodische Phänomene wie die Veränderung der Lautstärke ((leise)), ((laut)) oder des Stimmklangs ((lachend)), aber auch die übermäßige Dehnung von Lauten.

BS[sprech]		((lachend))	
IRSIVI	• • • Also wer noch nie eine Säge gesehen hat, darf sich vertrauensvoll	an meinen Kollegen wenden.	• • • Der zeigt Ihnen eine.

MIKO E 00001, 34:20

Ab einer Dauer von drei Sekunden wird die Zeit auf ganze Sekunden gerundet in den Klammern vermerkt (mit Komma, ohne Leerzeichen).

((Stimme verstellend))

Unter diesem Tag sind auffällige Sprechweisen zusammengefasst, die mit einer stimmlichen Entfremdung einhergehen.

	Was würden Sie denn machen,	wenn Sie so eine kleine Gelenkfläche nur hätten, die wirklich ein bisschen mickrig ist?
BS[sprech]		((Stimme verstellend,7s))

MIKO E 00001, 51:51

((silbifizierend))

Als silbifizierend wird das Sprechen beim Schreiben (z. B. an die Tafel) annotiert, das sich durch eine Verlangsamung, Silbifizierung und einen leicht skandierenden Rhythmus auszeichnet.

BS[v]	Also ich schreibe mal die Einzahl jetzt auf:	"Meniscus".
BS[wort]		Menis- • • -cus
BS[sprech]		((silbifizierend))
BS[nv]		((schreibt auf Folie))

MIKO_E_00002, 39:13

12 Metasprachliche Erläuterungen X/meta]

Zusätzlich zur Notation der metasprachlichen Verwendung von Begriffen durch Anführungszeichen in der Verbalspur (Kapitel 8.6.2) werden bei MIKO in der Spur X[meta] Sequenzen annotiert, in denen sprachliche Erläuterungen über Begriffe gegeben werden. Dazu zählen Hinweise zur Übersetzung in andere Sprachen, zur Etymologie, zum Genus, zur Morphologie, Morphosyntax, Orthografie oder Aussprache. Damit die Sequenzen in der Analyse ohne Kontext Sinn ergeben, umfasst die Annotation in der Regel ganze Äußerungen, bzw. Teiläußerungen bei längeren Äußerungen. Sie werden mit ((Meta)) getaggt.

Bei mehreren Sequenzen metasprachlicher Erläuterungen hintereinander wird das Tag durchgehend vergeben, da eine Untergliederung in einzelne Erläuterungseinheiten aufgrund der nicht-linearen Wissensstrukturierung nicht immer möglich ist.

Die metasprachlichen Erläuterungen müssen nicht als solche von der/dem Sprecher/-in angekündigt oder benannt werden. Entscheidend ist die Absicht der Sprecherin/des Sprechers, Informationen über Sprache zu vermitteln.

Direkte metasprachliche Erläuterung

BS[v]	Die Menisci, genau!	((4,8s))	Also ich schreibe mal die Einzahl jetzt auf: "Meniscus".
BS[meta]	((Meta))		

MIKO E 00002, 39:07

Die Sequenz lässt sich in folgende Erläuterungseinheiten unterteilen:

- Kennzeichnung des Wortes Menisci als Plural durch anschließende Benennung des Singular Meniscus und
- 2. Benennung des Wortes Meniscus als Singular.

Die erste metasprachliche Erläuterung erfolgt indirekt, die zweite wird direkt als solche benannt: Also ich schreibe mal die Einzahl jetzt auf.

Indirekte metasprachliche Erläuterung

BS[v]	• • • Und die zweite Geschichte • • • sind die "Zwischenwirbelscheiben",	((5,7s))	also die "Disci"
BS[meta]	((Meta)) ¹⁰		

BS[v]	"Diskuswerfer", ja, das Wort "Scheibe" kennen Sie ja.	((1,3s))	"Discus", • • • also • • "Disci	((1,1s))	intervertebrales".
BS[meta]	((Meta))				

MIKO_E_00003, 42:29

_

¹⁰ Für den gesamten abgebildeten Transkriptausschnitt ist nur ein durchgehendes Tag ((Meta)) vergeben. Im Gegensatz zum Transkript muss das Handbuch den Ausschnitt durch Zeilenumbrüche trennen. Um darzustellen, dass auch noch die dritte Zeile zur metasprachlichen Erläuterung gehört, wird das Tag im Handbuch in jeder Zeile wiederholt.

Die Sequenz lässt sich in folgende Erläuterungseinheiten unterteilen:

- 1. Übersetzung des Wortes Zwischenwirbelscheiben ins Lateinische: Disci intervertrebrales,
- 2. Herleitung der Übersetzung des Wortes Discus vom Wort Diskuswerfer,
- 3. Übersetzung des Wortes Discus ins Deutsche: Scheibe und
- 4. Ableitung des Plural Disci aus dem Singular Discus.

Keine der vier metasprachlichen Erläuterungen werden von BS direkt benannt.

Synonyme innerhalb des Deutschen werden nur getaggt, wenn sie als solche kenntlich gemacht werden. Im folgenden Beispiel wird der Äußerungsteil zu gut Deutsch als direkte Benennung gewertet.

BS[v]	• • • "Plan", • • • also einfach "platt" • • • zu gut Deutsch, • • • ja?
BS[meta]	((Meta))

MIKO_E_00003, 14:41

Einen Grenzfall zur metasprachlichen Erläuterung stellt die Zuordnung von Formel- oder Einheitszeichen zur Einheit oder zum Fachbegriff in der Physik dar:

DH[v]	L ist d	ie Länge	der Spule.
-------	---------	----------	------------

MIKO_E_00006, 46:04

Im Beispiel wird das Zeichen definiert, wobei es einen scheinbaren sprachlichen Bezug zwischen L und Länge gibt. Da der Dozent jedoch keine sprachliche Erläuterung vornimmt, wird die inhaltliche Definition nicht als sprachliche Übersetzung gewertet. Ebenso in folgendem Beispiel in der Anatomievorlesung:

BS[v]	Also "DD" soll immer "Differenzialdiagnose" heißen, also • • • Unterschied
[א] כם	zur Arthrose, darum geht es jetzt hier.

MIKO_E_00002, 47:31

13 Nonverbale Handlungen X[nv] und X[nv2]

Nonverbale Handlungen sind außersprachliche Phänomene der Kommunikation, die gestisch, proxemisch oder mimisch an den Hörer übermittelt werden und dadurch die Kommunikationssituation beeinflussen. In den Vorlesungstranskripten sind ausschließlich gestische nonverbale Handlungen annotiert.

Nonverbales wird in doppelten runden Klammern und in der dritten Person notiert. Dabei sind die Tags so kurz wie möglich gestaltet, z. B. durch das Weglassen von Artikeln: ((hält Schädelknochen vor Kamera)). Ab einer Dauer von drei Sekunden wird die Zeit auf ganze Sekunden gerundet in den Klammern vermerkt (mit Komma, ohne Leerzeichen).

In der Interaktion mit Gegenständen sollen diese so genau wie möglich benannt werden, um keine fachlichen Fehler zu riskieren und um das Verständnis auch ohne Vorliegen des Videos zu sichern.

Nicht berücksichtigt wurden folgende Phänomene:

- Gesten, die nicht der Präsentation von Vorlesungsinhalten diesen, z. B. das Kratzen am Kopf oder Tippen auf einen Tisch,
- Hilfsmittel, mit denen auf Medien oder Gegenstände gezeigt wird, z. B. Laserpointer, Zeigestock,
- das Nehmen und Hinlegen von Gegenständen, sondern lediglich die mit ihnen stattfindende Interaktion,
- das erste Auflegen oder letzte Wegnehmen von Folien auf dem Overheadprojektor,
- Vor- und Nachbereiten von Experimenten,
- das Hoch- und Herunterschieben von Tafeln und
- das Wechseln zwischen PowerPoint-Folien.

Das Wechseln von Folien auf dem Overheadprojektor hingegen ist mit größerem Aufwand verbunden und wird daher annotiert.

Bei der Durchführung von Experimenten werden diese nicht genauer beschrieben, sondern lediglich als Experiment bezeichnet: ((zeigt auf Experiment)), ((führt Experiment durch)).

Die Spur X[nv2] wird verwendet, wenn parallele nonverbale Handlungen stattfinden. Die Richtlinien sind identisch mit denen der Spur X[nv].

Neben der freien Beschreibung von nonverbalen Handlungen gibt es verschiedene wiederkehrende Tags, die im Folgenden vorgestellt werden.

13.1 ((zeigt auf ...))

Das Tag beschreibt das Zeigen auf Medien (Folie, Tafel), Gegenstände oder Personen ohne direkten Körperkontakt. Das Zeigen kann mit Hilfsmitteln oder mit der Hand erfolgen.

BS[v]	Das zeige ich di/ Ihnen	dann aber unter der Videokamera, da können Sie das besser sehen.
BS[nv]	((zeigt auf Knochen))	

MIKO E 00001, 22:49

Zeigt die/der Sprecher/-in auf ein Experiment, das per Durchlicht oder Overheadprojektor an die Wand projiziert wird, wird das Tag ((zeigt auf Experiment)) anstelle von ((zeigt auf Folie)) verwendet.

13.2 ((fasst an ...))

Das Tag wird verwendet, wenn beim Zeigen direkter Körperkontakt mit einem Gegenstand, Experiment oder sich selbst besteht. Beim Anfassen von Medien (Overheadprojektor-Folien, Tafel) wird jedoch das Tag ((zeigt auf ...)) gewählt, wenn es um das Zeigen auf Inhalte geht.

DH[v]	Wir haben also hier	unsere beiden Kondensatorplatten.
DH[nv]		((fasst an Kondensatorplatten))

MIKO_E_00004, 76:38

13.3 ((demonstriert ...))

Das Tag bezeichnet den unmittelbaren Bezug zwischen einem oder mehreren Wörtern und einer Geste, d. h. wenn ein Wort gleichzeitig oder in unmittelbarer zeitlicher Nähe durch eine (unbewusste) illustrierende Geste unterstützt wird. Die Annotationen beruhen selbstverständlich auf Interpretationen und bleiben subjektiv.

Das Wort oder die Wortgruppe wird wortgenau aus der Verbalspur übernommen und in die runden Doppelklammern geschrieben.

CF[v]	• • • Und ähm die Wortbedeutungen sind	lgarada e e a an inran/ e e e	also äh gerade an den Rändern…
CF[nv]		((demonstriert Rändern))	

MIKO E 00012, 01:25

Handelt es sich bei dem demonstrierten Wort um ein trennbares Verb, das aufgrund der Syntax getrennt wird, wird die Trennung im Tag aufgehoben.

DH[v]	• • • Wenn Sie so eine normale	Glühbirne	haben, fassen die an:
DH[nv]		((demonstriert Glühbirne))	((demonstriert anfassen))

DH[v]	Die	ist	total	heiß,	ja?

MIKO_E_00005, 08:49

13.4 ((veranschaulicht ...))

Das Tag beschreibt die gestische Veranschaulichung inhaltlich-fachlicher Zusammenhänge oder komplexer Bewegungsabläufe durch den Körper oder andere Hilfsmittel ohne konkreten Bezug zu einzelnen Wörtern. Es gibt keine standardisierte Vorgabe für die Benennung, die Prämisse ist jedoch eine möglichst einfache Beschreibung.

BS[v]	Das, was ich Ihnen vorhin vorgemacht habe, • • • zum Beispiel die Beugung	und den/ die Streckung, geht unterschiedlich gut • • • in den einzelnen Abschnitten.
BS[nv]	((veranschaulicht Beugung und Streckung mit ihrem Oberkörper,3s))	

MIKO_E_00003, 20:53

13.5 Schreiben und Zeichnen

Als *Schreiben* gilt das Verschriftlichen von Wörtern, Buchstaben, Formeln, Formelzeichen und Zahlen auf Medien. Das Medium wird im Tag benannt, jedoch nicht das Geschriebene.

DH[v]	Also das ist jetzt hier • • • der • • •	kapazitive • • • Widerstand	in Gelb.
DH[nv]		((schreibt an Tafel))	

MIKO_E_00005, 37:42

Als Zeichnen gilt das Festhalten von Formen, Pfeilen, Diagrammen und Bildern auf Medien. Das Medium wird auch hier benannt.

BS[v]	Und jetzt male ich mal hier was hin, was Sie auch schon kennen.	• • • Da sind zwei Begriffe möglich.
BS[nv]	((zeichnet auf Folie,6s))	

MIKO_E_00001, 17:46

Mischen sich Schreiben und Zeichnen, wird nur ein Tag mit der überwiegenden Handlung vergeben.

13.6 Korrekturen und Unterstreichungen

Nimmt die/der Sprecher/-in Korrekturen des Geschriebenen oder typografische Ergänzungen wie Unterstreichungen oder Einklammerungen vor, wird das Tag entsprechend benannt und das betreffende Wort, Teilwort oder der Buchstabe orthografisch von der Folie/Tafelanschrift übernommen.

DH[v]	• • • Und	für • • • F/	für, sagen wir mal, für	T ist gleich/
DH[nv]		((schreibt an Tafel))	((korrigiert f))	((schreibt an Tafel))

DH[v]	• • • äh nee, für klein	T ist gleich groß T, ne?
DH[nv]	((korrigiert T))	((schreibt an Tafel))

MIKO_E_00005, 19:09

BS[v]	• • • Und das ist ja das, was wir	hier als	"Apophysen" bezeichnet haben.
BS[nv]		((unterstreicht Apophysen))	

MIKO_E_00001, 26:59

14 Text-Bild-Bezüge *X/bild/*

Notation: ((Medium Tafel))

In der Interaktion der/des Dozierenden mit verschiedenen Medien (PowerPoint, Overheadprojektor, Tafel etc.) werden Bezüge zwischen Gesprochenem (Text) und Geschriebenem (Bild) hergestellt. Diese werden in der Spur für Text-Bild-Bezüge X[bild] mit dem Tag ((Medium)) annotiert. Diese projektspezifische Annotationsspur stellt ergänzend zu HIAT (Rehbein et al., 2004) einen besonderen Fokus von MIKO dar. So lassen sich beispielsweise modality interdependencies (Reimer et al., 2015) korpuslinguistisch untersuchen.

Getaggt werden Wörter, die von der/dem Dozierenden gesprochen und gleichzeitig oder in unmittelbarer zeitlicher Folge visuell durch ein Medium unterstützt werden. Das bedeutet konkret:

- bei der ersten Erwähnung eines Wortes, das auf einem Medium sichtbar ist und
- bei darauffolgenden Erwähnungen des Wortes nur noch dann, wenn von der/dem Sprecher/-in ein direkter Bezug zum Medium hergestellt wird, z.B. durch Zeigen, Schreiben oder Blickkontakt mit dem Medium.

Wenn eine neue Folie erscheint, dasselbe Wort wie auf der vorigen Folie zu sehen ist (z. B. in der Überschrift) und das Wort seit Folienwechsel erneut das erste Mal genannt wird, unterliegt dieses wiederum der Ersterwähnungsregel, auch wenn kein direkter Bezug durch die/den Dozierenden hergestellt wird. Dies gilt auch, wenn die/der Dozierende bewusst auf eine frühere Folie zurückgeht.

Das Tag wird je nach Medium spezifiziert und hat folgenden Aufbau:

- Tafel: ((Medium_Tafel))
- Overheadprojektor oder Durchlicht: ((Medium_OHP))
- PowerPoint: ((Medium_PPT))

Die Notation erfolgt in runden Doppelklammern und ohne Zeitangabe. Liegt bei mehreren Wörtern hintereinander ein Text-Bild-Bezug vor, wird das Tag durchgehend annotiert. Dabei ist unerheblich, ob die Wörter auch auf dem Medium unmittelbar hintereinander folgen oder ob die Wortstellung in der Verbalisierung geändert wird. Entscheidend ist, dass alle gesprochenen Wörter auf dem Medium zu sehen sind und ein direkter Bezug besteht. Im folgenden Beispiel steht auf der Folie: heute weniger gelesen wird als früher, die Umformulierung kann also durchgehend annotiert werden.

CF[v]	Das am häufigsten genannte • • • Merkmal war	"Heute wird weniger gelesen als früher.",	• • • ja?
CF[bild]		((Medium_PPT))	

MIKO E 00012, 49:29

Dies gilt auch bei trennbaren Verben. Auf der Folie steht: der Einfluss anderer Sprachen auf die deutsche Sprache stark zunimmt.

CF[v]	"Der Einfluss anderer Sprachen auf die deutsche Sprache nimmt stark zu."
CF[bild]	((Medium PPT))

Bei Pausen oder Verzögerungspartikeln zwischen Wörtern, die mit Text-Bild-Bezug annotiert werden, wird das Tag durchgehend vergeben und nicht unterbrochen.

IRSIVI	• • • Und jetzt drehen wir die ganze Geschichte rum • • • Spina • • • scapulae.
BS[bild]	((Medium_PPT))

MIKO E 00001, 08:25

Ebenso verhält es sich mit Reparaturen.

DH[v]	Also das	sauerstoffbeladene • • • Hämoglobin	• • • verstärkt das	Magnetfeld,
DH[bild]		((Medium_Tafel_Abk: oxyHb))		<pre>((Medium_Tafel_Abw: diamagnetisch))</pre>

DH[v]	• • • während das • • •	nichtsauerstoffgebundene Magn/ äh Hämoglobin	das
DH[bild]		((Medium_Tafel_Abk: deoxyHb))	

DH[v]	Magnetfeld	etwas abschwächt.
DH[bild]	<pre>((Medium_Tafel_Abw: paramgnetisch))</pre>	

MIKO E 00006, 89:06

Bei Komposita mit Bindestrich können die einzelnen Teile des Kompositums getrennt getaggt werden, wenn sie einzeln verbalisiert werden, im folgenden Beispiel das geschriebene Wort *DDR-Neologismen*.

CF[v]	Neologismen,	die in der	DDR	entstanden sind
CF[bild]	((Medium_PPT))		((Medium_PPT))	

MIKO_E_00012, 81:18

Die Interpunktion, Syntax sowie Groß- und Kleinschreibung der Transkription bei Wörtern mit Text-Bild-Bezug ordnet sich stets der Transkription der verbalisierten Äußerung unter, nicht umgekehrt. Im folgenden Beispiel steht auf der Folie: Basiskategorie v.a. für Konkreta leicht zu ermitteln, aber bei Abstrakta wesentlich aufwendiger / weniger eindeutig. Bei der Verbalisierung ergeben sich dabei zwei Äußerungen:

CF[v]	• • • Ähm • • • wir haben • • • äh auch hier das	Problem,	dass es	für Konkreta
CF[bild]		((Medium_PPT_Abw: Probleme))		((Medium_PPT))

CF[v]	• • • relativ	leicht	ist, solche • • •	Basiskategorien
CF[bild]		((Medium_PPT))		((Medium_PPT_Abw: Basiskategorie))

CF[v]	zu • • • äh ermitteln.	((1,3s))	Aber bei	den • • •	Abstrakta	• • • ähm ist es
CF[bild]	((Medium_PPT))				((Medium_PPT))	

CF[v]	wesentlich	schwieriger,	ja?
CF[bild]	((Medium_PPT))		

MIKO_E_00012, 14:10

Der Fokus der Annotation liegt grundsätzlich auf Inhaltswörtern (Nomen, Verben, Adjektive), nicht auf Funktionswörtern (Konjunktionen, Pronomen etc.). Diese werden nur dann getaggt, wenn sie auf einem Medium sichtbar sind und ein direkter Bezug zu einem getaggten Inhaltswort erkennbar ist. Wenn also auf einer Folie *der* steht und die/der Dozierende *der* in einem davon

unabhängigen Kontext sagt, während ein Bezug zur Folie hergestellt wird, wird dies nicht annotiert. Anders verhält es sich beispielsweise beim Vorlesen eines Stichpunktes wie im folgenden Beispiel. Auf der Folie steht: wäre bei einem Stromfluss von 20 A in etwa 3 s leer. Einen, von und in werden hier getaggt, da der unmittelbare Bezug erkennbar ist.

DH[v]	Wenn Sie	Einen	Stromfluss von zwanzig	Ampere
DH[bild]		((Medium_PPT_Abw: einem))	((Medium_PPT))	((Medium_PPT_Abk: A))

DH[v]	erzeugen, dann wäre die	in	wenigen	Sekunden	leer.
DH[bild]		((Medium_PPT))		((Medium_PPT_Abk: s))	((Medium_PPT))

MIKO_E_00006, 53:05

Getaggt werden folgende besondere Phänomene:

- Verbalisierung von mathematischen Symbolen, während diese als Zeichen auf dem Medium zu sehen sind:

gesprochen	auf Medium geschrieben
plus	+ (Pluszeichen)
minus	– (Minuszeichen)
mal	· (Multiplikationszeichen)
(geteilt) durch	: oder /
Pi Halbe	$\pi/2$
ist gleich	= (Gleichheitszeichen)
ungefähr, rund	~ bzw. ≈ (Gerundet-Zeichen)
proportional zu/zum/zur	lpha (Zeichen für Proportionalität)
pro	/ (Slash in Formeln)
Kosinus(funktion)	cos (Kosinus)

Tabelle 16: Mathematische Phänomene, die mit Text-Bild-Bezug getaggt werden

- positiv oder negativ, wenn Plus- oder Minuszeichen auf dem Medium zu sehen sind
- Zahlen in Formeln
- Funktionen (z. B. *U von T*)

Nicht getaggt werden folgende Phänomene:

- Verbalisierung von Nummerierungen und Aufzählungszeichen (z. B. Kapitel zwei Punkt zwei)
- Verbalisierung von Integralrechnungen und Vektoren
- Verbalisierung geometrischer Formen, z. B. Kreis, Viereck, Linie, Pfeil

Es werden verschiedene Arten von Abweichungen zwischen dem Verbalisierten und dem Medium festgehalten. Die dafür vorgesehenen Tags werden im Folgenden erläutert.

14.1 Abk

Notation: ((Medium PPT Abk: ...))

Bei der vollständigen Verbalisierung von Abkürzungen, die als solche auf einem Medium zu sehen sind, erhält das Tag die Ergänzung *_Abk* und die Angabe der Abkürzung. Die Orthografie wird dabei inklusive etwaiger Interpunktion original vom Medium übernommen.

BS[v]	• • • Und zwar gibt es da ein	"laterales",	((1,3s))	ein
BS[bild]		((Medium_OHP_Abk: lat.))		

BS[v]	"intermediäres"	• • • und ein	"mediales"	Os cuneiforme.
BS[bild]	((Medium_OHP_Abk: interm.))		((Medium_OHP_Abk: med.))	((Medium_OHP))

MIKO_E_00002, 08:02

Zeichen für Einheiten sowie Formelzeichen gelten als Abkürzungen.

DH[v]	Also	Fläche	der Spule	geteilt durch	Länge
DH[bild]		((Medium_Tafel_Abk: A))		((Medium_Tafel))	((Medium_Tafel_Abk: 1))

DH[v]	der Spule, das ist die	Induktivität.
DH[bild]		((Medium_Tafel_Alt: L, Induktivität))

MIKO_E_00006, 46:59

Stehen mehrere Einheitszeichen in einem mathematischen Verhältnis und werden diese zusammen verbalisiert, wird ein durchgehendes Tag vergeben.

DH[v]	• • • Sie wissen, die	Kapazität	hat die Einhei/ äh hat die Einheit
DH[bild]		<pre>((Medium_Tafel_Abk: C))</pre>	

DH[v]	((2,4s))	Amperesekunde • • • pro Volt.
DH[bild]		((Medium_Tafel_Abk: As/V))

MIKO_E_00005, 36:37

Bei Auslassungen wird derjenige Wortteil mit Ergänzungsbindestrich dennoch als vollständiges Wort und nicht als Abkürzung getaggt. Im folgenden Beispiel erhält *Fach*- bei Bild-Bezug zum Wort *Fachsprache* auf der Folie daher das Tag ((Medium_PPT)).

CF[v]	Ähm • • • also • • • sehr häufig ist das Phänomen, dass	((1,0s))	Wörter aus der
-------	---	----------	----------------

CF[v]	Fach-	oder Wissenschaftssprache	in die • • • Allgemeinsprache	kommen.
CF[bild]	((Medium_PPT))		((Medium_PPT))	

MIKO_E_00012, 67:10

14.2 Feh

Notation: ((Medium_OHP_FehV: ... statt ...))

Wenn das Gesagte und das Geschriebene voneinander abweichen und dabei ein fachlicher oder orthografischer Fehler vorliegt, erhält das Tag die Ergänzung _Feh, zusätzlich das Kürzel V für Fehler im Verbalen bzw. M für Fehler auf dem Medium. Anschließend wird der Fehler innerhalb der runden Doppelklammern wie folgt notiert: inkorrekt statt korrekt.

BS[v]	• • • Und das ist also ganz besonders wichtig für die	Diaphyse.
BS[bild]		((Medium_OHP_FehV: Diaphyse statt Epiphyse))

MIKO_E_00002, 05:32

DH[v]	Ein	Schwingkreis.
DH[bild]		<pre>((Medium_Tafel_FehM: Schwinkreis statt Schwingkreis))</pre>

MIKO E 00005, 40:05

Numerale werden als Zahlen notiert.

DH[v]	Sagen wir mal so um die	achthundert	Ohm.
DH[bild]		((Medium_OHP_FehV: 800 statt 100))	((Medium_OHP))

MIKO E 00005, 02:03

Bei orthografischen Abweichungen zwischen dem Transkript und dem Medium, die laut Duden akzeptierte Alternativen sind, wird kein Fehler annotiert, z. B. grafische (Transkript) und graphische (Tafel).

Bei orthografischen Abweichungen zwischen dem Transkript und dem Medium aufgrund alter Rechtschreibung, die laut Duden unzulässig sind, wird ein Fehler annotiert, z. B. bei *Stromfluß* (Tafel) statt *Stromfluss* (Transkript).

14.3 Abw

Notation: ((Medium OHP Abw: ...))

Wenn das Gesagte und das Geschriebene morphologisch voneinander abweichen, erhält das Tag die Ergänzung *Abw* und die Angabe des auf dem Medium sichtbaren Wortes. Die Orthografie wird dabei original vom Medium übernommen. Folgende morphologische Veränderungen werden getaggt:

Flexion

a) Konjugation

DH[v]	Also er	Erzeugt	mit einem Mal
DH[bild]		((Medium_PPT_Abw: erzeugen))	

MIKO_E_00006, 50:11

b) Deklination

BS[v]	• • • Und dann sind die • • •	Epiphysen	((2,4s))	anders
BS[bild]		((Medium_OHP_Abw: Epiphyse))		

BS[v]	aufgebaut als die	Diaphysen.
BS[bild]		((Medium_OHP_Abw: Diaphyse))

MIKO_E_00001, 21:19

Derivation

CF[v]	Witz,	Humor, • • • Kreativität, ja?
CF[bild]	((Medium_PPT_Abw: witzig))	

MIKO_E_00012, 75:30

Komposition

a) Auf dem Medium ist ein Kompositum sichtbar, verbalisiert wird aber nur ein Teil des Kompositums.

CF[v]	Und damit auch ihre	Bedeutung.
CF[bild]		((Medium PPT Abw: Bedeutungswandel))

MIKO_E_00012, 67:04

b) Verbalisiert wird ein Kompositum, auf dem Medium ist aber nur ein Teil des Kompositums sichtbar.

DH[v]	Aber jetzt hier für die	Magnetfeldbeschreibungen	und insbesondere für
DH[bild]		((Medium_Tafel_Abw: Magnetfeld))	

DH[v]	das Verständnis von/ • • • ähm • • • vom MRT • • • ist das hier sehr wichtig.

MIKO_E_00006, 39:00

c) Verbalisiert wird ein Kompositum, dessen Wortbestandteile alle auf dem Medium sichtbar sind. Diese stehen jedoch nicht in Form eines Kompositums, sondern als Wortgruppe auf dem Medium.

DH[v]	Also die	Stromwirkungen	• • • haben zum
DH[bild]		((Medium Tafel Abw: Wirkung von Strömen))	

MIKO_E_00005, 85:02

d) Auf dem Medium ist ein Kompositum sichtbar, dessen Wortbestandteile alle verbalisiert werden. Diese werden jedoch nicht in Form eines Kompositums, sondern als Wortgruppe mit u. U. morphologischen Abweichungen verbalisiert.

DH[v]	Natürlich haben wir auch ein • • •	magnetisches Feld	unserer
DH[bild]		((Medium_Tafel_Abw: Magnetfeld))	

DH[v]	Erde.
DH[bild]	((Medium Tafel))

MIKO_E_00006, 34:33

e) Verbalisiert wird ein Kompositum, auf dem Medium ist auch ein Kompositum sichtbar, es stimmt jedoch nur ein Teil des Kompositums überein.

DH[v]	So, nu muss man noch mal abschätzen • • • das	Gefährdungspontenzial	• • • für ähm
DH[bild]		((Medium_Tafel_Abw: Gefährdungsstufen))	

DH[v]	• • • für uns
ואחומן	Menschen.

MIKO E 00006, 16:35

Bei mehr als einer Kategorie von Abweichungen (Abkürzung, Fehler, Abweichung) innerhalb eines Wortes wird immer das Tag *Abw* vergeben.

Im folgenden Beispiel steht das Formelzeichen t für Zeit, daher handelt es sich um eine Abkürzung. Zeitlich ist zusätzlich eine Derivation von Zeit und muss deshalb als Abweichung getaggt werden.

DH[v]	Es muss sich nur • • •	zeitlich	ändern, ja?
DH[bild]		<pre>((Medium_Tafel_Abw: t))</pre>	

MIKO_E_00005, 13:47

Im zweiten Beispiel schreibt der Dozent *Sekundar* anstelle von *Sekundär* an die Tafel, daher handelt es sich um einen Fehler (Kapitel 14.2). *Sekundärspule* ist zusätzlich ein Kompositum, von dem nur der erste Teil auf dem Medium zu sehen ist, sodass hier ebenfalls eine Abweichung getaggt wird.

DH[v]	Das heißt also, • • • in dieser	Sekundärspule	wird jetzt eine
DH[bild]		((Medium_Tafel_Abw: Sekundar))	

DH[v]	Spannung	induziert.
DH[bild]	((Medium_Tafel_Abk: U))	

MIKO E 00006, 82:06

14.4 Alt

Notation: ((Medium OHP Alt: ..., ...))

Wenn die/der Dozierende ein Wort verbalisiert, das mehrfach in unterschiedlichen Wortformen auf dem Medium sichtbar ist, jedoch nicht eindeutig ist, auf welche Wortform die/der Dozierende referiert, erhält das Tag die Ergänzung _Alt und die Angabe der auf dem Medium sichtbaren Alternativen des Wortes, getrennt durch Kommata. Die Orthografie wird dabei original vom Medium übernommen. Es wird nicht weiter differenziert, ob die Wörter eine der oben genannten Abweichungen (Abkürzung, Fehler, Abweichung) enthalten.

DH[v]	Das heißt aber, da das	Volumen	dieses Zylinders genauso groß sein muss wie das	Volumen
DH[bild]		<pre>((Medium_OHP_Alt: V, Volumen))</pre>		<pre>((Medium_OHP_Alt: V, Volumen))</pre>

DH[v]	dieses Zylinders, • • • funktioniert das nur, • • • wenn der	Durchmesser	von diesem Zylinder hier kleiner wird, ja?
DH[bild]		<pre>((Medium_OHP_Alt: d, Durchmesser))</pre>	

Wird ein Begriff erstmalig und ohne direkten Bezug zum Medium erwähnt und gibt es mehrere (abweichende) Wortformen zum gesagten Wort, werden auch bei Ersterwähnung alle Alternativen aufgeführt. Lässt sich bei Ersterwähnung ein lexikalisch und morphologisch nächster Begriff feststellen, erhält dieser hingegen ein reguläres *Medium-*Tag, ggf. mit Abweichungsannotation.

15 Assistent Assist/v/

In den Physikvorlesungen unterstützt ein Assistent den Dozenten aktiv durch die Bereitstellung von Materialien und das Auf- und Abbauen von Experimenten. Daher kommt es zu Interaktionen, die sich auf die Kommunikationssituation der Vorlesung auswirken.

15.1 Verbale Äußerungen

Für verbale Äußerungen gelten die Richtlinien der Verbalspur X[v] (Kapitel 8). Da der Assistent kein eigenes Mikrofon hat, sind manche Passagen nur schwer oder nicht verständlich.

DH[v]		Dazu den Löffel, bitte sehr.
DH[nv]		((gibt Assistent Löffel))
Assist[v]	Ich brauche aber den Löffel, bitte.	

MIKO_E_00004, 15:06

15.2 Nonverbale Handlungen

Nonverbale Handlungen des Assistenten werden ebenfalls in der Spur Assist[v] notiert. Es werden nur solche nonverbalen Handlungen berücksichtigt, die für das Vorlesungsgeschehen relevant sind, z. B. das Durchführen von Experimenten. Als nicht relevant gelten das Abwischen von Tafeln, das Hin- und Hertragen von Gegenständen sowie das Auf- und Abbauen von Experimenten.

Führt der Assistent eine für das Vorlesungsgeschehen relevante nonverbale Handlung aus, übernimmt oder hat aber nicht den Turn, wird in der Spur X/v eine Pause notiert.

DH[v]	Jetzt ziehen wir das Ding raus.	((5,4s))	Und Sie sehen:
Assist[v]		((führt Experiment durch,5s))	

DH[v]	• • • Der Strom • • • geht zwar zurück, • • • aber ist bei weitem noch nicht
חווען	null, • • • ne?

MIKO_E_00006, 85:08

16 Auditorium *NN/v*/

Dem Auditorium wird in Vorlesungen grundsätzlich die Hörerrolle zugeschrieben. Dennoch sind Vorlesungen nicht frei von Interaktion zwischen Dozierenden und Studierenden, so z. B. bei folgenden Szenarien:

- 1. Die/der Dozierende stellt eine Frage und erhält eine Antwort aus dem Plenum.
- 2. Ein/-e Studierende/-r meldet sich zu Wort und stellt eine Frage.
- 3. Die/der Dozierende kommentiert Handlungen des Auditoriums (z. B. Unruhe, Fehlverhalten, Zuspätkommen).
- 4. Studierende im Auditorium lachen über etwas, das die/der Dozierende sagt.
- 5. Das Auditorium applaudiert am Ende der Vorlesung oder bei besonderer Begeisterung.

In der Auditoriumsspur NN[v] werden die Inhalte verbaler Äußerungen einzelner Studierender und nicht-phonologische akustische Phänomene notiert.

16.1 Verbale Äußerungen

Notation: ((Beschreibung der verbalen Handlung))

Verbale Äußerungen einzelner Studierender wurden durch Rauschen in der Aufnahme ersetzt (Kapitel 4.2.2), um die Anonymität zu gewährleisten. Im Transkript werden die entsprechenden Sequenzen in doppelten runden Klammern notiert und die verbale Handlung beschrieben: ((Student/-in antwortet auf Frage)), ((Student/-in bejaht)). Sofern die verbale Äußerung akustisch nicht eindeutig zu identifizieren ist, wird die Sequenz wie folgt transkribiert: ((nicht zuzuordnender Beitrag aus Auditorium)) bzw. ((Murmeln)) für undeutliches Reden mehrerer Studierender zur selben Zeit. Ab einer Dauer von drei Sekunden wird die Zeit auf ganze Sekunden gerundet in den Klammern vermerkt (mit Komma, ohne Leerzeichen). Ist in der Aufnahme eine Illokution der Äußerung identifizierbar, wird das entsprechende Äußerungsendzeichen notiert.

NN[v]	'	• • • ((Student/-in antwortet auf Frage)).	es auch.
BS[v]	Was gibt es noch so für Knorpel?		Faserknorpel gibt

MIKO_E_00001, 52:22

Ist der Modus der Äußerung nicht eindeutig bzw. handelt es sich um ein Hörersignal ohne Turnübernahme, erhält die Äußerung einen hochgestellten Punkt.

16.2 Nicht-phonologische akustische Phänomene

Notation: ((Lachen))

Nicht-phonologische akustische Phänomene im Plenum wie Lachen, Klatschen oder Klopfen werden in doppelten runden Klammern und in nominalisierter Form notiert. Das Rascheln und Murmeln des Plenums am Anfang und Ende der Vorlesung wird nicht berücksichtigt. Ab einer Dauer von drei Sekunden wird die Zeit auf ganze Sekunden gerundet in den Klammern vermerkt (mit Komma, ohne Leerzeichen). Außerhalb einer verbalen Äußerung erhält das Phänomen das Äußerungsendzeichen für Äußerungen ohne Modus (hochgestellter Punkt).

BS[v]	• • • Bei mir ist da was, bei Ihnen sollte da auch was sein,	• • • ja?
NN[v]		((Lachen))

MIKO_E_00001, 02:35

Liegt der Turn bei der/dem Dozierenden und macht sie/er parallel zu den transkribierten akustischen Phänomenen des Plenums eine Pause innerhalb einer Äußerung oder zwischen zwei Äußerungen, werden die Pausen in der Verbalspur der/des Dozierenden vermerkt.

SL[v]	Das hier wird ja nicht bewertet, es wird nur auf Video aufgenommen.	((1,7s))
NN[v]		((Lachen))

SL [v]	Ist überhaupt gar kein Druck heute.	((2,3s))	Es hört Ihnen so gut wie niemand zu.
NN[v]		((Lachen))	

MIKO_E_00007, 47:26

Liegt der Turn bei einem Studierenden oder wird der Turn im Anschluss an ein Lachen von einer/einem Studierenden übernommen, wird parallel keine Pause in der Verbalspur der/des Dozierenden notiert, da der Turn als bei der/dem Studierenden gewertet wird.

CF[v]]		Si/ • • • wi	rd noch benutzt?			Ja˙
NN[v]]	((Student/-in widerspricht,5s)).		((Lachen))	((Student/-in widerspricht,5s))		

MIKO_E_00012, 60:59

Dies gilt auch, wenn kein Turn mehr auf das Phänomen folgt, d. h. am Ende der Vorlesung.

BS[v]	Ich bedanke mich, dass Sie so gut mitgemacht haben.	
NN[v]		((Klopfen,4s))

MIKO_E_00003, 46:12

17 Kommentare zum Transkript *X*[komm]

Kommentare zu Besonderheiten des Transkripts, die über die erläuterten Konventionen hinausgehen und von der Transkribentin/dem Transkribenten als relevant angesehen werden, werden in der Spur X[komm] notiert. Sie werden in runde Doppelklammern und ab einer Dauer von drei Sekunden auf ganze Sekunden gerundet in den Klammern vermerkt (mit Komma, ohne Leerzeichen). Bei MIKO werden verschiedene für sprecherspezifische Kommentarspuren bei HIAT (Rehbein et al., 2004) vorgesehene Phänomene bereits in anderen Annotationsspuren notiert, sodass die Kommentarspur in den vorliegenden Transkripten zwei Phänomene beinhaltet.

Schlechte Aufnahmequalität:

Ist die Tonaufnahme in einzelnen Sequenzen so gestört, dass das Verstehen beeinträchtigt oder verhindert wird, oder gibt es eine Aufnahmelücke, wird dies entsprechend notiert: ((schlechte Aufnahmequalität)).

DH[nv]	So wie es beispielsweise Elektronen und Ionen gibt, ne?			
DH[komm]	komm] ((schlechte Aufnahmequalität))			

MIKO E 00006, 42:13

Freie Beschreibung von nicht festgelegten Phänomenen

Im Folgenden gibt es ein Beispiel für ein von den Richtlinien nicht festgelegtes Phänomen, für dessen Beschreibung die Kommentarspur genutzt wurde. Es handelt sich um Auffälligkeiten in der Aussprache.

	DH[7/]	Und das ist das, was man im Allgemeinen als das Phänomen "Faradayscher Käfig"
	H[v]	bezeichnet, • • • nach

DH[v]	Michael/ Michael	Faraday.
DH[komm]	((Aussprache wird von Deutsch zu Englisch korrigiert))	

MIKO_E_00004, 35:35

Literaturverzeichnis

- ACTFL (Hg.). (2013). ACTFL Reading Proficiency Test (RPT): FAMILIARIZATION MANUAL & ACTFL PROFICIENCY GUIDELINES 2012 READING. White Plains. https://www.languagetesting.com/reading-proficiency-test
- ACTFL (Hg.). (2014). ACTFL Listening Proficiency Test (LPT): FAMILIARIZATION MANUAL & ACTFL PROFICIENCY GUIDELINES 2012 Listening. White Plains. https://www.languagetesting.com/listening-proficiency-test
- Allwood, J. (2008). Multimodal corpora. In A. Lüdeling & M. Kytö (Hg.), Handbücher zur Sprachund Kommunikationswissenschaft /HSK]: Bd. 29.1. Corpus linguistics: An international handbook (S. 207–225). De Gruyter.
- Bärenfänger, O., Lange, D. & Möhring, J. (2015). Sprache und Bildungserfolg: Sprachliche Anforderungen in der Studieneingangsphase. Research papers in assessment: Bd. 1. Institut für Testforschung und Testentwicklung e.V. http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-188820
- DFG. (2013). Informationen zu rechtlichen Aspekten bei der Handhabung von Sprachkorpora.
- DFG. (2019). Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora.
- Dudenredaktion. (o. J.). Duden online. https://www.duden.de/woerterbuch
- Dudenredaktion (Hg.). (2016). Der Duden in zwölf Bänden. Die Grammatik: Unentbehrlich für richtiges Deutsch (9. Aufl.). Dudenverlag.
- Europarat (Hg.). (2001). Gemeinsamer europäischer Referenzrahmen für Sprachen ((J. Quetz, Übers.)). Langenscheidt.
- Fandrych, C. & Schmidlin, R. (Hg.). (2019). Bulletin suisse de linguistique appliquée: Bd. 109. Wissenschaftssprache(n) kontrastiv.
- g.a.s.t. (Hg.). (2017). Über on SET. https://www.onset.de/home/ueber-onset/
- Granger, S. & Paquot, M. (2017). *Core metadate for learner corpora: draft 1.0.* Louvain-la-Neuve: Université catholique de Louvain.
- Heller, D., Hornung, A., Redder, A. & Thielmann, W. (2013). The euroWiss-Project: Linguistic Profiling of European Academic Education (Germany/Italy). *European Journal of Applied Linguistics*, 1(2), 317–320. https://doi.org/10.1515/eujal-2013-0018
- Hovy, E. & Lavid, J. (2010). Towards a 'science' of corpus annotation: A new methodological challenge for corpus linguistics. *International Journal of Translation Studies*, 22, 13–36.
- Lüdeling, A., Walter, M., Kroymann, E. & Adolphs, P. (2005). *Multi-Level Error Annotation in Learner Corpora*.
- MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk (third edition): Volume I: Transcription format and programs, Volume II: The database. *Computational Linguistics*, 26(4), 657. https://doi.org/10.1162/coli.2000.26.4.657

- Marks, D. (2015). Prüfen sprachlicher Kompetenzen internationaler Studienanfänger an deutschen Hochschulen Was leistet der TestDaF? Zeitschrift für Interkulturellen Fremdsprachenunterricht, 20(1), 21–39.
- Redder, A. (2016). Theoretische Grundlagen der Wissenskonstruktion im Diskurs. In J. Kilian, B. Brouer & D. Lüttenberg (Hg.), *Handbuch Sprache in der Bildung* (S. 297–318). De Gruyter.
- Rehbein, J., Schmidt, T., Meyer, B., Watzke, F. & Herkenrath, A. (2004). *Handbuch für das computergestützte Transkribieren nach HIAT: Version 1.0. Arbeiten zur Mehrsprachigkeit Folge B: Bd.* 56. Sonderforschungsbereich 538 (Mehrsprachigkeit), Universität Hamburg.
- Reimer, E., Trevisan, B., Eraßme, D., Schmidt, T. & Jakobs, E.-M. (2015). Annotating Modality Interdependencies. In B. Fisseni, B. Schröder & T. Zesch (Hg.), *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology* (S. 110–111). https://doi.org/10.13140/RG.2.1.2311.1126
- Sasaki, F. & Witt, A. (2016). *Linguistische Korpora* (H. Lobin & L. Lemnitzer, Hg.). Institut für Deutsche Sprache, Bibliothek.
- Schiller, A., Teufel, S., Stöckert, C. & Thielen, C. (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung; Universität Tübingen, Seminar für Sprachwissenschaft. http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf
- Schmid, H. (1994). Probabilistic part-of-speech tagging using Decision Trees. *Proceedings of the International Conference on New Methods in Language*. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf
- Schmidt, T. & Wörner, K. (2014). EXMARaLDA. In J. Durand, U. Gut & G. Kristoffersen (Hg.), *The Oxford Handbook of Corpus Phonology* (S. 402–419). Oxford University Press.
- Schmidt, T., Wörner, K., Hedeland, H. & Lehmberg, T. (2013). Leitfaden zur Beurteilung von Aufbereitungsaufwand und Nachnutzbarkeit von Korpora gesprochener Sprache. Hamburger Zentrum für Sprachkorpora, Universität Hamburg; Archiv für Gesprochenes Deutsch, IDS Mannheim.
- Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., Günthner, S., Hartung, M., Kern, F., Mertzlufft, C., Meyer, C., Morek, M., Oberzaucher, F., Peters, J., Quasthoff, U., Schütte, W., . . . Uhmann, S. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). Gesprächsforschung Online-Zeitschrift zur verbalen Interaktion, 2009(10), 353–402. http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf
- Song, M.-Y. (2012). Note-taking quality and performance on an L2 academic listening test. *Language Testing*, 29(1), 67–89.
- Steets, A. (2003). Die Mitschrift als universitäre Textart-schwieriger als gedacht, wichtiger als vermutet. In K. Ehlich & A. Steets (Hg.), *Wissenschaftlich schreiben-lehren und lernen* (S. 51–64). De Gruyter.
- Wisniewski, K. (2019). Mitschreiben in Vorlesungen. Ein interdisziplinärer Forschungsüberblick mit Fokus Deutsch als L2. In C. Fandrych & R. Schmidlin (Hg.), *Bulletin suisse de linguistique appliquée: Bd. 109. Wissenschaftssprache(n) kontrastiv.* (S. 153–170).

- Wisniewski, K., Lenhard, W., Möhring, J. & Spiegel, L. (Hg.). (in Vorbereitung). *Sprache und Studienerfolg bei Bildungsausländer/-innen*. Waxmann.
- Wisniewski, K., Spiegel, L., Lenort, L. & Feldmüller, T. (in Vorbereitung). Herausforderung Wissenschaftssprache I: Mitschreiben. In K. Wisniewski, W. Lenhard, J. Möhring & L. Spiegel (Hg.), Sprache und Studienerfolg bei Bildungsausländer/-innen. Waxmann.
- Wisniewski, K., Spiegel, L., Parker, M., Feldmüller, T. & Lenort, L. (in Vorbereitung). Mitschreiben in Vorlesungen in der Studieneingangsphase: Das multimodale Lehr-Lernkorpus MIKO. In K. Wisniewski, W. Lenhard, J. Möhring & L. Spiegel (Hg.), Sprache und Studienerfolg bei Bildungsausländer/-innen. Waxmann.