

Masterarbeit



Brandenburgische
Technische Universität
Cottbus

Name: Linke, Marcel

Matrikel: 2702942

Ausgabe: 30.10.2012

Fakultät 3

Lehrstuhl Kommunikationstechnik

Avatare zur Darstellung von Gebärdensprache

Using avatars for sign language synthesis

Prof. Dr.-Ing. habil. Matthias Wolff

Lehrstuhlinhaber

T +49 (0)355 69 21 28

F +49 (0)355 69 21 50

E matthias.wolff@tu-cottbus.de

www.tu-cottbus.de/

kommunikationstechnik

Die zwischenmenschliche Kommunikation ist eine elementare Funktion in unserem Leben. Dabei stellt die gesprochene Sprache die bedeutendste und am meisten verwendete Kommunikationsform dar. Gehörlose Personen können Sprache jedoch akustisch nicht wahrnehmen. Die Muttersprache vieler Gehörloser ist die Gebärdensprache, bei der visuell wahrnehmbare Sprachelemente verwendet werden. Aufgrund der fehlenden auditiven Rückmeldung sind zudem oft auch die Lese- und Schreibfähigkeiten auf einem geringen Niveau. Für die Verbesserung der Kommunikationssituation sollten für gehörlose Personen die Informationsdarstellung entsprechend angepasst werden. Eine sinnvolle Möglichkeit ist animierte Darbietung von Gebärdensprache durch eine künstliche Figur (Avatar). Neben der Modellierung des virtuellen Menschen müssen aus gesprochener Sprache oder aus Texten Ansteuerparameter für den Avatar erzeugt werden, um eine computeranimierte Darstellung von Bewegungsabläufen zu realisieren.

Die Arbeit beinhaltet folgende Teilaufgaben:

- Kurze Darstellung der Phonologie und Grammatik der Deutschen Gebärdensprache
 - Ableitung der Anforderungen an die Animation des Avatars
 - Notationssysteme für Gebärdensprache (linguistisch und animationsorientiert)
- Avatarsysteme
 - Klassifikation nach Anwendungsgebieten und verwendeter Technik
 - Verarbeitungsschritte zur Umsetzung von Text in Gebärdensprachsynthese
 - Visualisierung der Avatare (Avatardesign)
- Maschinelle Übersetzung von Text in eine animationsorientierte Gebärdensprachnotation
 - Grundprinzipien
 - Stand der Technik
- Evaluationsmöglichkeiten
 - Maße für die Genauigkeit der Übersetzung
 - Möglichkeiten der Qualitätsbewertung von animierter Gebärdensprachdarstellung und Zusammenfassung bestehender Evaluationsergebnisse

Betreuer: M. Sc. Jens Lindemann

Prof. Dr.-Ing. habil. Matthias Wolff

Avatare zur Darstellung von Gebärdensprache

Using avatars for sign language synthesis

Masterarbeit

im Studiengang Informations- und Medientechnik

Brandenburgische Technische Universität Cottbus

Fakultät 3: Maschinenbau, Elektrotechnik und Wirtschaftsingenieurwesen

Lehrstuhl Kommunikationstechnik

Prof. Dr.-Ing. habil. Matthias Wolff

Verfasser:

Marcel Linke

Matrikelnummer 2702942

Betreuer:

M. Sc. Jens Lindemann

Cottbus, 30.04.2013

Kurzfassung

Diese Arbeit beschäftigt sich mit den Kommunikationsschwierigkeiten zwischen Gehörlosen und ihrer Umgebung. So sind Informationen der modernen digitalen Medien (z. B. Internet, Smartphone, Fernseher usw.) für die Gehörlosengemeinschaft meist nur schwer zu verstehen. Der Grund ist, dass der Informationsaustausch in der Lautsprache akustisch erfolgt und in der Gebärdensprache visuell. Des Weiteren fällt das Lese- und Schreibniveau der Gehörlosen meist sehr mangelhaft aus. Für die barrierefreie Kommunikation und Informationsdarstellung könnte eine animierte Figur, ein sogenannter Gebärdensprachavatar, hilfreich sein.

Diese Arbeit gliedert sich in vier Teile. Im ersten Teil erfolgt eine ausführliche Darstellung der Deutschen Gebärdensprache aus der sich die Anforderungen an einen Gebärdensprachavatar ableiten lassen. Der zweite Teil betrachtet Schriftsysteme, die nötig sind, um eine Gebärdensprache textuell darzustellen. Im dritten Teil werden die Strategien der Maschinellen Übersetzung und deren Bewertungsmöglichkeiten geschildert, die einen Text aus der Lautsprache in eine animationsorientierte Gebärdensprachnotation überführen. Abschließend folgt im vierten Teil die technische Betrachtung zur Erzeugung eines Avatar-Modells. Außerdem werden Anwendungsgebiete für Avatarsysteme aufgezeigt und vorhandene Evaluationsmöglichkeiten sowie Evaluationsergebnisse von Gebärdensprachavataren zusammengefasst.

Abstract

This work deals with the problems of communication between deaf and their environment. Usually the informations of modern digital media (such as Internet, Smartphone, television, etc.) are difficult to understand for the deaf community. The reason is that in the spoken language the exchange of information is acoustically and in sign language visually. Furthermore, the deaf have problems with writing and reading. For a better communication and presentation of information could help an animated character (called Sign language avatar).

This work is divided into four parts. The first part deals with the presentation of the german sign language (DGS) and includes the requirements for a sign language avatar. The second part shows writing systems for a sign language. After this, the third part gives an overview of machine translation strategies and evaluation methods. Moreover, it describes the transformation from a spoken language text into an animation-based notation. Finally, the fourth part shows techniques to generate an avatar model. In addition to this, it deals with the applications for avatar systems and shows you a summary of existing works and evaluation results of sign language avatars.

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	15
Abbildungsverzeichnis.....	19
Tabellenverzeichnis.....	23
1. Kapitel: Einleitung.....	25
1.1 Motivation.....	25
1.2 Aufgaben und Ziele der Arbeit.....	26
1.3 Aufbau der Arbeit.....	27
2. Kapitel: Deutsche Gebärdensprache (DGS).....	29
2.1 Einführung Gebärdensprache.....	29
2.2 Phonologie.....	32
2.2.1 Manuelle Komponenten/Ausdrucksmittel/Parameter.....	34
2.2.1.1 Handform.....	35
2.2.1.2 Handstellung.....	36
2.2.1.3 Ausführungsstelle.....	38
2.2.1.4 Bewegung.....	39
2.2.1.5 Zusammenfassung: manuelle Ausdrucksmittel.....	41
2.2.1.6 Zweihandgebärden.....	42
2.2.1.7 Simultaneität.....	43
2.2.2 Nicht-manuelle Komponenten/Ausdrucksmittel/Parameter.....	43
2.2.2.1 Gesichtsausdruck/Mimik.....	44
2.2.2.2 Augenausdruck/Blickrichtung.....	45
2.2.2.3 Kopf- und Oberkörperhaltung bzw. -bewegung.....	46
2.2.2.4 Mundartikulation.....	47
2.2.2.5 Zusammenfassung: nicht-manuelle Ausdrucksmittel.....	49
2.3 Grammatik und Syntax (Satzlehre).....	50
2.3.1 Satzglieder.....	50
2.3.2 Satzbausteine bzw. Satzbestandteile.....	53
2.3.3 Satztypen.....	57
2.3.4 Zusammenfassung zur Grammatik der DGS.....	60
2.4 Anforderung an die Animation des Avatars.....	62
3. Kapitel: Notationssysteme der Gebärdensprache.....	65
3.1 Glossen-Notation für menschliche Kommunikation.....	66
3.2 Linguistische Notationssysteme.....	68
3.2.1 Stokoe-Notation.....	68
3.2.2 Hamburger Notationssystem (HamNoSys).....	69

3.3 Animationsorientierte Notationssysteme.....	71
3.3.1 SIGML (Signing Gesture Markup Language).....	71
3.3.2 BML (Behavior Markup Language).....	72
3.3.3 EMBRScript (Embodied Agent Behavior Realizer Script).....	75
3.3.4 Fazit zu animationsorientierten Notationen für die Gebärdensprache.....	77
4. Kapitel: Maschinelle Übersetzung.....	79
4.1 Grundlegendes zur Maschinellen Übersetzung.....	79
4.2 Strategien der Maschinellen Übersetzung.....	81
4.2.1 Regel-basierte Verfahren – rule-based machine translation (RBMT).....	81
4.2.1.1 Direkte Übersetzung.....	81
4.2.1.2 Transfer Übersetzung.....	82
4.2.1.3 Interlingua.....	83
4.2.2 Daten-basierte Verfahren – data-based machine translation (DBMT).....	84
4.2.2.1 Statistik-basierte MÜ – statistic-based machine translation (SBMT).....	84
4.2.2.2 Beispiel-basierte MÜ – example-based machine translation (EBMT).....	90
4.2.3 Fazit: MÜ-Strategien.....	92
4.3 Evaluationsmöglichkeiten für die Maschinelle Übersetzung.....	93
4.3.1 WER (word error rate).....	94
4.3.2 PER (position-independent word error rate).....	94
4.3.3 TER (Translation Error Rate).....	95
4.3.4 SER (Sentence Error Rate).....	96
4.3.5 BLEU (Bilingual Evaluation Understudy).....	96
4.3.6 NIST (National Institute of Standards and Technology).....	98
4.3.7 METEOR (Metric for Evaluation of Translation with Explicit Ordering).....	98
4.4 Stand der Maschinellen Übersetzung im Bereich der Gebärdensprache.....	100
4.4.1 Regel-basierte MÜ-Systeme.....	102
4.4.2 Daten-basierte MÜ-Systeme.....	106
4.4.3 Hybride MÜ-Systeme.....	108
4.5 MÜ-Strategien für die Übersetzung von Texten in eine animationsorientierte Notation.....	111
4.5.1 RBMT-Workflow.....	111
4.5.2 EBMT-Workflow.....	112
4.5.3 SBMT-Workflow.....	113
5. Kapitel: Avatarsysteme.....	115
5.1 Design des Avatars.....	115
5.1.1 Modellierungsphase.....	116
5.1.1.1 Darstellungsarten der Geometrie.....	116
5.1.1.2 Modellierungsmethoden/-techniken.....	120

5.1.1.3 Techniken zur Anpassung/Verformung der Geometrie bzw. Gestalt des Avatars.....	123
5.1.2 Animationsphase.....	128
5.1.2.1 Keyframe Animation.....	129
5.1.2.2 Kinematische Techniken.....	130
5.1.2.3 Motion Capture (MOCAP).....	131
5.1.2.4 Prozedurale Animation.....	132
5.1.3 Rendern.....	133
5.1.4 Fazit: Gebärdensprachavatar.....	134
5.1.4.1 Design des Avatars.....	134
5.1.4.2 Erstellung des Avatars.....	136
5.2 Anwendungsbereiche Avatarsysteme.....	139
5.2.1 Übersicht: ausgewählte allgemeine Avatarsysteme.....	143
5.2.2 Übersicht: ausgewählte Gebärdensprach-Avatarsysteme.....	145
5.3 Evaluation von Gebärdensprachavataren.....	149
5.3.1 Methoden.....	149
5.3.2 Ergebnisse.....	150
5.3.3 Auswertungen und Erkenntnisse.....	153
6. Kapitel: Zusammenfassung der Ergebnisse und Ausblick.....	157
6.1 Zusammenfassung der Ergebnisse.....	157
6.2 Ausblick.....	160
Literaturverzeichnis.....	163
Anhang.....	175
A1: Symbole Stokoe-Notation.....	175
A2: Übersichtstabelle: Notationssysteme der Gebärdensprache.....	178
A2.1: Beispiel SWML.....	183
A2.2: Beispiel Zebedee-Notation.....	184
A2.3: Beispiel MURML.....	185
A3: Erklärung der LFG.....	186
A4: Beispielberechnung für BLEU-Evaluation.....	188
A5: Beispielberechnung für METEOR-Evaluation.....	190
A6: Zusammenfassung ausgewählter Evaluationen.....	192
A6.1: Huenerfauth'schen „multiple choice“ Test.....	192
A6.2: Delta-Evaluation (EMBR).....	195
A6.3: ViSiCAST Evaluation (speziell: Dialogsystem Poststelle).....	198
A6.4: Die 9 Evaluationen des Dicta-Sign Projektes.....	201
Erklärung.....	203

Abkürzungsverzeichnis

ACE	–	Articulated Communicator Engine
Adj	–	Adjektiv
AGB	–	Allgemeine Geschäftsbedingungen
ARP	–	Avatar Research Platform
Art	–	Artikel
ASCII	–	American Standard Code for Information Interchange
ASL	–	American Sign Language, Amerikanische Gebärdensprache
ATIS	–	Air Traffic Information System
AU	–	Action Unit
BGG	–	Behindertengleichstellungsgesetz
BLEU	–	Bilingual Evaluation Understudy
BMAS	–	Bundesministeriums für Arbeit und Soziales
BML	–	Behavior Markup Language
BP	–	brevity penalty
BSL	–	British Sign Language, Britische Gebärdensprache
BTS	–	Berkeley Transcription System
CAT	–	computer-aided translation
CCD	–	charge-coupled device
CL	–	Computerlinguistik
CMU	–	Carnegie Mellon University
CP	–	classifier predicates
CSE	–	Czech Sign Language, Tschechische Gebärdensprache
DARPA	–	Defense Advanced Research Projects Agency
DBMT	–	data-based machine translation, Daten-basierte Maschinelle Übersetzung
DCL	–	Doll Control Language
DCU	–	Dublin City University
DEF	–	Definitheit
DET	–	Determinierer
DFKI	–	Deutsches Forschungszentrum für Künstliche Intelligenz
DGS	–	Deutsche Gebärdensprache
DRS	–	Discourse Representation Structure
DSGS	–	Deutschschweizer Gebärdensprache
EBMT	–	example-based machine translation, Beispiel-basierte Maschinelle Übersetzung
EFFD	–	Erweiterte Freiformdeformation
ELAN	–	Eudico Linguistic Annotator

EM	–	Expectation-Maximization Algorithmus
EMBR	–	Embodied Agent Behavior Realizer
EMOTE	–	Expressive MOTion Engine
eSIGN	–	Essential Sign Language Information on Government Networks
EU	–	Europäische Union
FACS	–	Facial Action Coding System
FAP	–	Facial Animation Parameter
FAPU	–	Face Animation Parameter Unit
FDP	–	Face Definition Parameter
FFD	–	Freiformdeformation
FK	–	Forward Kinematics
FML	–	Function Markup Language
GIATI	–	Grammatical Inference and Alignments for Transducer Inference
GPL	–	General Public License
GR	–	Gebärdenraum
GS	–	Gebärdensprache
GSL	–	Greek Sign Language, Griechische Gebärdensprache
GuK	–	Gebärden-unterstützte-Kommunikation
HamNoSys	–	Hamburger Notationssystem
HMS	–	Human Modeling and Simulation
HPSG	–	Head-Driven Phrase Structure Grammar
IBM	–	International Business Machines Corporation
ICT	–	Information and Communications Technology
IK	–	Inverse Kinematics
ILES	–	Information, Language, writtEn and Signed Group
ISI	–	Information Sciences Institute
ISL	–	Indian Sign Language, Indische Gebärdensprache
ISL	–	Irish Sign Language, Irische Gebärdensprache
JSL	–	Japanese Sign Language, Japanische Gebärdensprache
KBMT	–	Knowledge-Based Machine Translation
KI	–	Künstliche Intelligenz
LBG	–	Lautsprachbegleitende Gebärden
LFG	–	Lexical functional grammar
LGPL	–	GNU Lesser General Public License
LIMSI	–	Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
LS	–	Lautsprache
LSE	–	Lengua de signos Española, Spanische Gebärdensprache
LSF	–	Langue des Signes Françaises, Französische Gebärdensprache
LUG	–	Lautsprachunterstützende Gebärden

LUT	–	Lookup Tabelle, Umsetzungstabelle
LV	–	Losson und Vannobel
MATREX	–	MACHine TRAnslation using EXamples
Max	–	the Multimodal Assembly eXpert
METEOR	–	Metric for Evaluation of Translation with Explicit Ordering
MLE	–	Maximum likelihood estimation
MOCAP	–	Motion Capture
MPEG	–	Moving Picture Experts Group
MRE	–	Mission Rehearsal Excercise
MT	–	machine translation
MURML	–	Multimodal Utterance Representation Markup Language
MÜ	–	Maschinelle Übersetzung
N	–	Nomen
NAACL	–	North American Chapter of the Association for Computational Linguistics
NGT	–	Nederlandse Gebarentaal, Niederländische Gebärdensprache
NIST	–	National Institute of Standards and Technology
NP	–	Nominalphrase
NPC	–	non-player character
NUM	–	Numerus
NURBS	–	Non-Uniform Rational B-Splines
ÖGS	–	Österreichische Gebärdensprache
PCFG	–	Probabilistic Context-Free Grammar
PDTS-SIGML	–	posture/ detention/ transition/ steady shift-Signing Gesture Markup Language
PER	–	position-independent word error rate
PERS	–	Person
PRED	–	Prädikat
Qs	–	Quellsprache
RBMT	–	rule-based machine translation, Regel-basierte Maschinelle Übersetzung
S	–	Satz
SAIBA	–	Situation Agent Intention Behavior Animation
SAMPA	–	Speech Assessment Methodes Phonetic Alphabet
SASL	–	South African Sign Language, Südafrikanische Gebärdensprache
SBB	–	Schweizerische Bundesbahn
SBMT	–	statistic-based machine translation, Statistik-basierte Maschinelle Übersetzung
SC	–	Signed Czech
SDO	–	Semantic Dimension Overlap

SER	–	Sentence Error Rate, String Error Rate
SFST	–	Stochastic Finite State Transducer
SIGML	–	Signing Gesture Markup Language
SL	–	source language
SNCF	–	Société nationale des chemins de fer français
STAG	–	Synchronous Tree Adjoining Grammar
STEP	–	Scripting Technology for Embodied Persona
SWML	–	SignWriting Markup Language
TAG	–	Tree-Adjoining Grammar
TEAM	–	Translation from English to ASL by Machine
TER	–	Translation Error Rate
TL	–	target language
TM	–	translation memory
UEA	–	University of East Anglia
USL	–	Urubu Sign Language, Urubu Gebärdensprache
V	–	Verb
VH	–	Virtual Human
VP	–	Verbalphrase
VP	–	Versuchspersonen
VRML	–	Virtual Reality Modeling Language
WER	–	word error rate
WN	–	word net
XML	–	Extensible Mark-up Language
Zs	–	Zielsprache

Abbildungsverzeichnis

Abb. 1: Hauptproblem der Arbeit.....	26
Abb. 2: DGS-Komponenten.....	34
Abb. 3: Manuelle Ausdrucksmittel.....	34
Abb. 4: Handform-Phoneme der DGS; verändert aus [Pap08].....	35
Abb. 5: Allophon Faust; aus [Pap08].....	36
Abb. 6: Ebenen im GR; verändert aus [Pap08].....	36
Abb. 7: Fingeransatzrichtung; aus [Pap08].....	37
Abb. 8: Handform und Handstellung der DGS.....	37
Abb. 9: Ausführungsstellen im GR; aus [Pap08].....	39
Abb. 10: Übersicht manuelle Komponenten der DGS.....	41
Abb. 11: Parallele, spiegelbildliche und alternierende Symmetrie; aus [Pri07], [Sie07].....	42
Abb. 12: Obere Hand: aktiv; untere Hand: passiv; aus [Sie07].....	43
Abb. 13: Nicht-manuelle Ausdrucksmittel.....	44
Abb. 14: Basisemotionen; aus [Pap08].....	44
Abb. 15: Unterscheidung durch Mundbild; aus [Sie07].....	48
Abb. 16: Übersicht nicht-manuelle Komponenten der DGS.....	49
Abb. 17: Zeitlinie in der GS; aus [Pap08].....	52
Abb. 18: links: Zeitlinie vor dem Körper; rechts: Zeitlinie quer vor dem Körper; aus [Pap08].....	52
Abb. 19: Vertikale Zeitlinie; aus [Boy95].....	53
Abb. 20: Substitutor-Verb „Gehen“; aus [Pap08].....	55
Abb. 21: Manipulator-Verben: Handformbeispiele; aus [Pap08].....	55
Abb. 22: Steigerung des Adjektivs „GUT“; aus [Pap08].....	57
Abb. 23: Ausgewählte Satztypen der DGS.....	57
Abb. 24: Entscheidungsfrage in der DGS; aus [Boy96].....	58
Abb. 25: von links nach rechts: Bitte-, Aufforderungs-, Befehls-Mimik; aus [Pap08].....	59
Abb. 26: Übersicht DGS-Grammatik.....	61
Abb. 27: Übersicht Notationssysteme für die Gebärdensprachen.....	65
Abb. 28: links: Glossen-Notation; aus [Wil10]; rechts: Glossen-Notation; aus [Pap08].....	67
Abb. 29: links: Stokoe-Notation für die Gebärde „Idee“; rechts: Zugehörige Gebärde; ergänzt aus [Hoh07].....	68
Abb. 30: links: Handformen; rechts: Handstellungen; aus [Han04].....	69
Abb. 31: HamNoSys-Gebärde „GEHEN-ZU“; ergänzt aus [Ken02].....	70
Abb. 32: SIGML-Gebärde „GEHEN-ZU“; verändert aus [Ell12].....	72
Abb. 33: SAIBA-Framework; aus [Vil07].....	73
Abb. 34: BML-Beispiel; verändert aus [Vil07].....	73
Abb. 35: BML-Synchronisationspunkte; aus [Kop06].....	74
Abb. 36: SAIBA-Framework mit EMBRScript; verändert aus [Hel10].....	75

Abb. 37: Konfiguration des Skeletts; aus [Hel09].....	76
Abb. 38: EMBRScript-Code für Augenzwinkern und Lächeln.....	76
Abb. 39: MÜ-Dreieck nach Vauquois.....	81
Abb. 40: Beispiel für eine Transfer Übersetzung.....	82
Abb. 41: Funktionsweise einer SBMT.....	85
Abb. 42: Beispiel für ein Wort-basiertes Übersetzungsmodell nach IBM-Modell 1; verändert aus [Koe13].....	87
Abb. 43: Phrasen-Alinierung; aus [Koe06].....	88
Abb. 44: links: Vorgehensweise EBMT; rechts: Beispiel in der EBMT.....	91
Abb. 45: Übersetzungsworkflow des ViSiCAST Projektes; aus [Saf01].....	102
Abb. 46: Architektur des Text-To-Indian Sign Systems; aus [Das07].....	104
Abb. 47: Zardoz Architektur; aus [Vea98].....	105
Abb. 48: links: Definition der CP's; rechts: Visualisierung der CP's; aus [Mat07].....	106
Abb. 49: DCU-Übersetzungssystem; verändert aus [Smi10].....	106
Abb. 50: Vorgehensweise für die Übersetzung.....	109
Abb. 51: Architektur des Driver License Systems; verändert aus [San09].....	110
Abb. 52: RBMT von einem Text in eine animationsorientierte GS-Notation.....	111
Abb. 53: EBMT von einem Text in eine animationsorientierte GS-Notation.....	112
Abb. 54: SBMT von einem Text in eine animationsorientierte GS-Notation.....	113
Abb. 55: Animation Pipeline; aus [Cun12].....	115
Abb. 56: Übersicht: Avatargenerierung.....	116
Abb. 57: links: Polygonnetz bestehend aus Dreiecke oder Vierecke; aus [Cun12] rechts: Gesicht mit zu wenig Polygone wirkt kantig; aus [Flü08].....	117
Abb. 58: Spline-Arten für die Modellierung; aus [Gia97].....	117
Abb. 59: links: Zunehmende Verfeinerung durch Unterteilungsflächen rechts: Unterteilungsflächen eines Kopfes; aus [Flü08].....	118
Abb. 60: links: Verschmelzung von zwei Metaballs; aus [Mae99] rechts: Polygonnetz als dehnbare Oberfläche; aus [Rat98].....	119
Abb. 61: links: Körper aus Metaballs; verändert aus [Plu02] rechts: Kopf aus Metaballs; verändert aus [Rat98].....	119
Abb. 62: Modellierungstechniken.....	120
Abb. 63: links: Box-Modelling; aus [Tüm07] mitte: Zweischichtiges Modell; aus [Mae99] rechts: Haptisches Eingabegerät Phantom Omni; aus [Tüm07].....	121
Abb. 64: 1-3D-Scanner, 2-Extrusion, 3-Lathing, 4-Sweeping, 5-Skinning; verändert aus [Gia97], [Tüm07].....	122
Abb. 65: Startbildschirm des Editors Makehuman.....	122
Abb. 66: Freiformdeformation; ergänzt aus [Cun12].....	123
Abb. 67: Deformierung bei falschen Skinning-Informationen; aus [Tüm07].....	124
Abb. 68: Verschiedene Morph-Targets; aus [Mor13].....	125

Abb. 69: Umsetzung von AU2: Heben der Augenbraue; ergänzt aus [Jac06].....	126
Abb. 70: 1-Feature Points legen das Drahtgittermodell fest, 2-FAP-Gruppen, 3-Positionierung der FAPUs; zusammengestellt aus [Par02], [Jac06].....	127
Abb. 71: 1 ohne Überschleifen, 2 mit Überschleifen, 3 ohne Überschleifen; verändert aus [Ber11].....	130
Abb. 72: MOCAP-Arten zur Datenerfassung: von links nach rechts: optisch, magnetisch, mechanisch und mit Gyroskope; aus [Tüm07].....	132
Abb. 73: Laufanimation durch prozedurale Animation; aus [Tüm07].....	132
Abb. 74: von links nach rechts: Video „The Forest“ von Vcom3D, Avatar „Max“ von Televirtual/UEA, ein spanischer Geschichtenerzähler, spanischer Junge, PR-Video der deafworld 2000 und ein Avatar für mobile Endgeräte von Vcom3D; aus [Kip11].....	135
Abb. 75: Effekt des Uncanny Valley (unheimliches Tal); aus [Poh09].....	135
Abb. 76: Generierung eines Gebärdensprachavatars.....	137
Abb. 77: Mögliche Anwendungsbereiche für Avatare.....	139
Abb. 78: Virtuelle Konferenzen; aus [Tüm07].....	141
Abb. 79: Avatarmodell im Bereich der Medizin; aus [Cun12].....	141
Abb. 80: Ergonomieuntersuchung Fahrzeuginnenraum; aus [Tüm07].....	141
Abb. 81: Avatare aus Shrek 2; aus [Cun12].....	141
Abb. 82: NPCs in die SIMS 2; aus [Tüm07].....	141
Abb. 83: Interaktiver Hilfeassistent von Microsoft Office.....	141
Abb. 84: Max als Museumsführer; aus [Tüm07].....	142
Abb. 85: Robert von T-Online; aus [Dus11].....	142
Abb. 86: Avatar aus Projekt MRE; aus [Len02].....	144
Abb. 87: Interaktiver Museumsführer Max; aus [Wac09].....	144
Abb. 88: Avatar Greta; von der Webseite.....	144
Abb. 89: Demonstrationsvideo Elckerlyc; von der Webseite.....	144
Abb. 90: Virtual Guido aus dem eSIGN Projekt; aus [Kip11].....	148
Abb. 91: Avatar Max; von der BMAS Internetseite.....	148
Abb. 92: Vcom3D-Avatar aus dem Werbevideo: „The Forest“ by Jason Stewart; aus [Kip11].....	148
Abb. 93: LIMSI Avatar für Bahndurchsagen; aus [Seg08].....	148
Abb. 94: Avatarsystem EMBR aus dem Java Applet.....	148
Abb. 95: Problemdarstellung der Arbeit.....	159
Abb. 96: Verarbeitungsschritte zur Umsetzung von einem Text in eine Gebärdensprachavatar-Darstellung.....	160
Abb. 97: SWML-Gebärde für das Wort „Brasilien“; aus [Cos01].....	183
Abb. 98: Zebedee-Gebärde für das Wort „BALL“; aus [Fil09].....	184
Abb. 99: MURML-Beispiel; aus [Kop07].....	185
Abb. 100: Beispiel einer LFG; verändert aus [Sch13].....	186

Abb. 101: Auswahlmöglichkeiten A, B und C für die Versuchsperson; aus [Hue07].....	193
Abb. 102: Balkendiagramm für Huenerfauth'schen „multiple choice“ Test; aus [Hue07].....	194
Abb. 103: Ablauf der Delta-Evaluation; aus [Ngu11].....	196
Abb. 104: Ergebnisse der Delta-Evaluation zu EMBR; aus [Kip11].....	197
Abb. 105: Ergebnisse der Evaluierung des Poststellen Systems; zusammengestellt aus [Cox02].....	199

Tabellenverzeichnis

Tab. 1: Entwicklung der GS in Deutschland.....	31
Tab. 2: Vor- und Nachteile der Glossen-Notation.....	67
Tab. 3: Vor- und Nachteile der Stokoe-Notation.....	68
Tab. 4: Vor- und Nachteile HamNoSys.....	71
Tab. 5: Vor- und Nachteile SIGML-Notation.....	72
Tab. 6: Vor- und Nachteile BML.....	74
Tab. 7: Vor- und Nachteile EMBRScript-Notation.....	77
Tab. 8: Übersichtstabelle der Arbeiten.....	101
Tab. 9: Übersicht Animationstechniken.....	128
Tab. 10: Ausgewählte allgemeine Avatarsysteme.....	143
Tab. 11: Ausgewählte Gebärdensprach-Avatarsysteme.....	145
Tab. 12: Ergebnisse des europäischen ViSiCAST Projektes; Werte aus [She03], Evaluation Report 3.....	151
Tab. 13: Ausführungsstellen nach Stokoe; aus [Sto12].....	175
Tab. 14: Handformen nach Stokoe; aus [Sto12].....	176
Tab. 15: Bewegungen nach Stokoe; aus [Sto12].....	177
Tab. 16: Anmerkungen zu Notationen; aus [Boy95], [Mar03], [Han04], [San06], [Kat08], [Kip11], [Dot12].....	182
Tab. 17: Datenbasis für das Experiment; aus [Hue07].....	192
Tab. 18: Datenbasis der Evaluation zu dem Avatar EMBR; aus [Hel11].....	195
Tab. 19: Zusammenfassung der Evaluationen aus dem Dicta-Sign Projekt; aus [Web11].....	202

1. Kapitel: Einleitung

1.1 Motivation

Die Kommunikation zwischen Gehörlosen und hörenden Menschen kann schwierig ausfallen. Der Hörende nutzt die Lautsprache mit ihren akustischen Äußerungen um sich zu verständigen und der Gehörlose artikuliert seine Aussagen visuell mit einer Gebärdensprache. Die Allgemeinheit glaubt, dass die Gebärdensprache die visuelle Äußerung der Lautsprache eines Landes ist. Diese Annahme ist falsch. Die Lautsprache und die Gebärdensprache innerhalb eines Landes sind grundsätzlich verschiedene Sprachen mit jeweils eigener Grammatik und eigener Sprachkultur. In vielen Schulen können zwar die Gehörlosen die Lautsprache des Heimatlandes erlernen, aber das Lese- und Schreibniveau reicht meist nur für eine oberflächliche Kommunikation mit ihren hörenden Mitmenschen aus. Einer der Hauptgründe dafür ist, dass gehörlose Kinder als Muttersprache zuerst die Gebärdensprache vermittelt bekommen. Ihr geprägtes Sprachwissen unterscheidet sich somit grundlegend von dem Sprachwissen eines hörenden Menschen, der sich als Muttersprache frühzeitig die Lautsprache angeeignet hat. Somit müssen gehörlose Kinder mit einem anderen Sprachsystem das Schriftsystem der Lautsprache in der Schule erlernen und haben damit größere Probleme als ihre hörenden Schulkameraden. Das Resultat ist ein schwaches Lese- und Schreibniveau der Gehörlosen. Ein Überblick über wissenschaftliche Arbeiten im Bereich des Lese- und Schreibniveaus gehörloser Menschen gibt die Quelle [Kra01]. So befindet sich unter anderem in ihr eine Studie von Beate Krausmann (1997), die den Titel trägt „Schriftsprachkompetenz Gehörloser - eine Untersuchung anhand der Faxkommunikation“. Ihr Ergebnis lautet im Bezug auf die Gehörlosen in Deutschland: „Nur 4 % der gehörlosen Schulabgänger erreichen ein altersangemessenes Leseniveau. 40 % bewegen sich auf dem Niveau von Zweit- bis Drittklässlern und mehr als 50 % gelten als Nicht-Leser [Kra01].“ Andere Studien in [Kra01] zeigen ähnliche Ergebniswerte.

Somit kommen für die zwischenmenschliche Kommunikation zwischen hörenden und gehörlosen Menschen meist Dolmetscher zum Einsatz, um die Sprachbarriere zu überbrücken. So helfen sie Gehörlosen bei persönlichen Gesprächen (z. B. Arztbesuch), vermitteln die Information von Reden oder Nachrichtensendungen in Fernsehbeiträgen und können in Gebärdensprachvideos Informationen auf Internetseiten präsentieren. Der letzte Punkt findet in einem überschaubaren Rahmen statt. So werden meist nur Seiten von staatlichen Institutionen von einem Dolmetscher dargestellt. Andere Internetseiten kann ein Gehörloser nur nutzen, wenn er ein gutes Lese- und Schreibniveau besitzt. Dies ist auch die Voraussetzung in Diskussionsforen. So können sich Gehörlose nur selten an Diskussionen oder sogenannten sozialen Netzwerken wie z. B. Facebook beteiligen.

Um die barrierefreie Informationsdarstellung in den modernen Medien für Gehörlose zu erleichtern, kann als Alternative ein sogenannter Gebärdensprachavatar zur Anwendung kommen. Ein Avatar ist eine künstlich erstellte animierte Figur. Ihr Aussehen und ihr Bewegungsablauf lässt sich an einem Computer erstellen. So kann diese Figur dem Menschen ähnlich sehen. Das Ziel eines Gebärdensprachavatars ist es, die Information durch visuelle Darstellung von Gebärden den Gehörlosen vereinfacht bereit zu stellen und die Kommunikation mit den hörenden Mitmenschen zu erleichtern. Der Vorteil des Avatarsystems ist, dass es eine kostengünstige Alternative zum Dolmetscher darstellt. So lassen sich Avatare flexibler gestalten und in verschiedenen Bereichen anwenden. Des Weiteren erlaubt ein Avatar den gehörlosen Internetnutzer sich anonymisiert in Internetforen oder -beiträgen zu beteiligen. Ein Avatarsystem kann dem Nutzer auch interaktive Möglichkeiten bieten. So können Aussehen, Kamerawinkel oder der Gebärdenstil personalisiert werden. Die Arbeit soll die Verarbeitungsschritte in einem Avatarsystem aufzeigen und mögliche Anwendungsgebiete ermitteln.

1.2 Aufgaben und Ziele der Arbeit

Die Abbildung 1 zeigt die Problemdarstellung der Arbeit. Es stellt sich die Frage, wie ein Text aus der Lautsprache automatisch in eine Gebärdensprache übersetzt und von einer animierten Figur gebärdet wird. Das Ziel der Arbeit ist es, diese Fragestellung zu lösen.

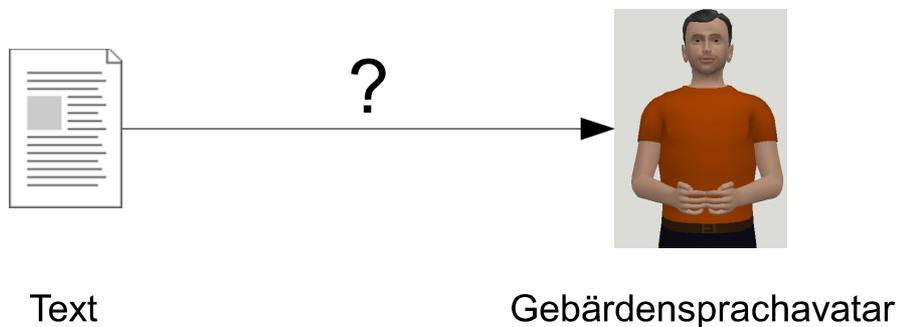


Abb. 1: Hauptproblem der Arbeit

Dazu muss die Phonologie und die Grammatik der Deutschen Gebärdensprache recherchiert werden. Aus diesen Kenntnissen lassen sich dann die Anforderungen für einen Gebärdensprachavatar ableiten. Ein weiterer Aufgabenpunkt beschäftigt sich mit den Notationssystemen einer Gebärdensprache. Die Frage, die sich stellt ist, wie lässt sich eine visuelle Sprache schriftlich abbilden? Das Ziel ist, linguistische und animationsorientierte Notationen zu finden und zu beschreiben.

Des Weiteren soll untersucht werden, mit welchen Techniken sich Avatare visualisieren bzw. erstellen lassen und in welchen Bereichen Avatarsysteme einsetzbar sind.

Für den Generierungsprozess von einem Text in die Gebärdensprachsynthese muss der Punkt der Maschinellen Übersetzung betrachtet werden. Ziel ist es, Grundprinzipien der Maschinellen Übersetzung zu recherchieren, den aktuellen Stand im Bereich der Gebärdensprachavatare zu ermitteln und mit den gewonnenen Kenntnissen einen Workflow von einem Text in eine animationsorientierte Gebärdensprachnotation zu erstellen.

Abschließend soll die Frage geklärt werden, wie sich der Übersetzungsprozess und ein Avatarsystem evaluieren lassen. Außerdem sollen vorhandene Ergebnisse von animierten Gebärdensprachdarstellungen recherchiert und zusammengestellt werden.

1.3 Aufbau der Arbeit

Um die Aufgabenstellung zu lösen, wird zuerst die Zielsprache des Avatarsystems betrachtet. Als Vertreter für die Gebärdensprache untersucht das 2. Kapitel die Deutsche Gebärdensprache. Hierbei werden die Bestandteile dieser visuellen Sprache, die sogenannten manuellen und nicht-manuellen Komponenten, vorgestellt und erläutert. Außerdem gibt das Kapitel einen Einblick in die grammatikalische Funktionsweise der Deutschen Gebärdensprache und bildet aus den gewonnenen Kenntnissen die Anforderungen eines Gebärdensprachavatars ab.

Das 3. Kapitel zeigt wie eine bildliche Sprache, so wie es die Gebärdensprache ist, textuell notiert werden kann. Dazu klassifiziert eine Übersicht vorhandene Gebärdensprach-Schriftsysteme in die Bereiche „menschliche Kommunikation“, „linguistische Notation“ und „animationsorientierte Notation“. Ausgewählte Vertreter aus diesen Bereichen werden genauer betrachtet und geben somit einen Einblick in den Aufbau und die Funktionsweise solcher Gebärdensprach-Notationen.

Um von einem Text aus der Lautsprache in die animationsorientierte Gebärdensprachnotation für den Avatar zu kommen, benötigt es einen automatischen Ablauf. Das 4. Kapitel beschäftigt sich mit den Maschinellen Übersetzungsstrategien. Es zeigt, welche Strategien schon für Gebärdensprachavatare zur Anwendung gekommen sind und wie diese sich mit Evaluationsmethoden vergleichen lassen.

Im 5. Kapitel geht es um das Design und die Erzeugung eines Avatar-Modells. Des Weiteren stellt es Anwendungsbereiche für Avatare vor und zeigt ausgewählte Avatarsysteme im allgemeinen Bereich und speziell für den Bereich der Gebärdensprache. Abschließend beinhaltet das 5. Kapitel Evaluationsmöglichkeiten, die bei der Bewertung von Gebärdensprachavataren vorkommen und stellt vorhandene Evaluationsergebnisse vor.

Das 6. Kapitel fasst die Ergebnisse der Masterarbeit zusammen und gibt einen Ausblick für weitere Arbeiten im Bereich des Gebärdensprachavatars.

2. Kapitel: Deutsche Gebärdensprache (DGS)

Dieses Kapitel widmet sich dem Aufbau und Ablauf der Deutschen Gebärdensprache (kurz: DGS). Die DGS ist die Muttersprache der Gehörlosen bzw. der Gehörgeschädigten. Sie unterscheidet sich wesentlich von der deutschen Lautsprache und wird somit als eine separate Sprache anerkannt, die nicht nur in ihrer Bildung, sondern auch in der Grammatik eigene Merkmale setzt. Die Kommunikation erfolgt hier nicht durch die Bildung von akustischen Ausdrucksmitteln, sondern durch visuell manuellen und nicht-manuellen Komponenten, die der Sprecher mittels seines Körpers umsetzt.

2.1 Einführung Gebärdensprache

Allgemeines

Die Gebärdensprache (GS) wurde aufgrund ihrer ikonischen bzw. bildhaften Darstellung häufig als Pantomime oder eine Aneinanderreihung von Gesten verstanden. So glaubten Sprachwissenschaftler, dass diese Form der Kommunikation ähnlich dem Verhalten von hörenden Menschen ist, die sich unter anderem mit „Hand und Fuß“ unterhalten, wenn sie aus verschiedenen Ländern kommen und die gegenseitige Sprache nicht beherrschen. Sie sahen keine Anzeichen einer eigenen Sprache, da es in einer GS keine Artikel (der, die, das) sowie Kopulae (z. B. sein, haben) gibt und Pronomen auf den ersten Blick nicht erkenntlich waren. Nach mehreren Studien und Untersuchungen stellte sich heraus, dass die Gebärdensprache eine eigene Grammatik, Syntax und Semantik besitzt.

Es gibt nicht nur „die Gebärdensprache“. Jedes Land besitzt aufgrund seiner kulturellen Entwicklung eine eigene (z. B. BSL (*British Sign Language*), ASL (*American Sign Language*), LSF (*Langue des Signes Françaises*), LSE (*Lengua de signos Española*) oder eben die DGS) [Blu06].

In der Regel sind ca. 0,1 % der Bevölkerung eines Landes gehörlos und nutzen dementsprechend eine Gebärdensprache. Somit kann man annehmen, dass es in Deutschland ca. 82.000 gehörlose Menschen gibt, die sich mit Hilfe der DGS unterhalten. Insgesamt umfasst die Sprachgemeinschaft ca. 200.000 Menschen, zu denen auch Hörgeschädigte und Dolmetscher zählen.

Die GS eines Landes kann sich sogar von Region zu Region unterscheiden, was vergleichbar ist mit den verschiedenen Dialekten einer Lautsprache (LS), wobei die Grammatik bei den verschiedenen regionalen Varianten keine Unterschiede aufweist, allerdings Abweichungen in der Ausführung, Bewegung und Darstellung der Gebärdenzeichen [Sie07]. So lassen sich beispielsweise Erkenntnisse der Sprachforschung aus der ASL in die DGS übertragen (z. B. die Verbarten) [Blu06]. Die deutsche Lautsprache ist zum Beispiel in Deutschland, Österreich und in der Schweiz eine offizielle Amtssprache. Gebärdende aus diesen Ländern hätten jedoch Probleme, sich vollständig zu verständigen, da diese Länder eine separate und eigenständige

Gebärdensprache besitzen (Deutsche Gebärdensprache (DGS), Österreichische Gebärdensprache (ÖGS), Deutschschweizer Gebärdensprache (DSGS)). Gebärdende aus verschiedenen Ländern können sich trotzdem nach einer kurzen Eingewöhnungsphase unterhalten, da bestimmte Komponenten z. B. die Handform bei der Ausführung ähnlich sind.

Geschichtliche Entwicklung

Die Tabelle 1 soll kurz und prägnant die geschichtliche Entwicklung der GS – im speziellen die markanten deutschen Ereignisse – darstellen.

Jahr	Ereignis
1760	Erste Gehörlosenschule der Welt in Paris; Gründer Abbé Charles Michel de l'Épée (1712-1789)
1778	Erste deutsche Gehörlosenschule in Leipzig; Gründer Samuel Heinicke (1727-1790)
1838	Erster Gehörlosenverein der Welt von Ferdinand Berthier (1803-1886) in Paris gegründet
1848	Erster Gehörlosenverein mit dem Namen „ <i>Allgemeiner Taubstummer-Unterstützungsverein von Groß-Berlin</i> “ in Deutschland; Gründer Eduard Fürstenberg (1827-1885) → Startpunkt der deutschen Gehörlosenbewegung
ab 1850	Negative Haltung der Gehörlosenpädagogen gegenüber den Gebärdensprachen stieg an, da GS als eine niedrige Kommunikationsform angesehen wurde
1880	Kongress der Gehörlosenlehrer in Mailand; Ergebnis: LS wird anstatt der GS gelehrt → Einführung der „ <i>oralen Methode</i> “; Grund: LS wurde einen höheren geistlichen Mehrwert als GS zugesagt
ab 1900	„ <i>orale Methode</i> “ wurde durch verschiedene Fachbücher kritisiert → andere Lernmethoden wurden eingeführt
1911	Einführung der gesetzlichen Schulpflicht für alle gehörlosen Kinder
1960	William Stokoe untersucht erstmalig linguistisch die ASL (<i>American Sign Language</i>); Titel der Arbeit „ <i>Sign Language Structure</i> “
1979	Edward Klima und Ursula Bellugi stärkten die Untersuchung der GS; Sie betrachteten die Produktionsgeschwindigkeit von Gebärden; Titel der Arbeit: „ <i>The Signs of Language</i> “ → Anerkennung der GS stieg in der linguistischen Forschungswelt an, da diese Untersuchungen den Aussagen von 1880 widersprachen → GS und LS wurden auf eine Stufe gestellt
1986	Veröffentlichung der deutschen Zeitschrift „ <i>Das Zeichen</i> “; Herausgeber „ <i>Gesellschaft für Gebärdensprache und Kommunikation Gehörloser e.V.</i> “ → Plattform zur Abbildung von sprachwissenschaftlichen Arbeiten
1987	Siegfried Prillwitz (Vorreiter bei der Untersuchung der DGS) gründet das Zentrum

	zur „ <i>Erforschung der deutschen Gebärdensprache und Kommunikation Gehörloser</i> “ in Hamburg
1988	Forderung des EU-Parlaments an die europäischen Länder zur Anerkennung der GS
1998	Institut für „ <i>Deutsche Gebärdensprache und Kommunikation Gehörloser</i> “ wurde der Universität Hamburg angegliedert → entwickelten die HamNoSys-Notation und diverse Lehrbücher und -hefte
2002	Rechtliche Anerkennung der DGS mit dem „ <i>Behindertengleichstellungsgesetz (§ 6 BGG)</i> “

Tab. 1: Entwicklung der GS in Deutschland

Abgrenzung der DGS zu anderen visuellen Sprachsystemen

Die Deutsche Gebärdensprache ist kein gebärdetes Deutsch oder ein lautsprachbegleitendes Gebärden, sondern besitzt eine eigene Grammatik und ein spezifisches Lexikon. Sie wird jedoch von anderen manuell-visuellen Kodierungssysteme der deutschen Sprache verwechselt. Neben der Gebärdensprache gibt es das „*Fingeralphabet*“, „*Lautsprachbegleitende Gebärden (LBG)*“, „*Lautsprachunterstützende Gebärden (LUG)*“ und „*Gebärden-unterstützte Kommunikation (GuK)*“.

Beim Fingeralphabet wird ein Wort, für das es in der GS keine Gebärde gibt, mit Handzeichen buchstabiert. Dieser Fall tritt üblich bei Personennamen, Eigennamen (Organisationen, Unternehmen, Städte usw.) und Fachwörter auf. Es findet bevorzugt Anwendung in den Vereinigten Staaten und wird mit nur einer Hand gebärdet. In der DGS wird es aus kulturellen Gründen dem Lippenlesen nachgestellt.

Lautsprachbegleitende Gebärden (LBG) oder „gebärdetes Deutsch“ orientieren sich an die Lautsprache des Landes. So werden Sätze aus der LS eins zu eins in der gleichen Reihenfolge in die GS übersetzt. Die Darstellung der Wörter erfolgt zwar mit Gebärden, aber die Grammatik orientiert sich an die Lautsprache. „Wörter wie Konjunktionen oder Morpheme wie Wortendungen (z. B. -en, -t), für die es keine passende Gebärde gibt, werden üblicherweise durch erfundene Gebärden ersetzt [Wil10].“ Dadurch gibt es in der LBG wesentlich mehr Gebärden, mit denen man in Kombination mit dem Fingeralphabet unter anderem Artikel und Wortflexionen, äußern kann. Dieses manuell-visuelle Kodierungssystem wird für Lehrzwecke genutzt. So können gehörlose Menschen, deren Muttersprache die DGS ist, die visuelle Darstellung und Grammatik der deutschen Lautsprache erlernen.

Die Lautsprachunterstützenden Gebärden (LUG) nutzen Kinder mit Entwicklungsverzögerungen, Behinderungen oder Hörschädigungen. Ihre Muttersprache ist die Lautsprache, die sie jedoch nur in eingeschränkten Maßen nutzen können. Um die Kommunikation mit anderen Menschen zu erleichtern, werden inhaltstragende Wörter (sog. Signalwörter) gebärdet.

Der Unterschied zur LBG ist, dass nicht der komplette Satz mit Gebärdenzeichen vollzogen wird, sondern nur einzelne Wörter. Die Reihenfolge der Wörter des lautsprachlichen Satzes orientiert sich dabei an die der Lautsprache. Eine eigene Grammatik besitzt sie somit nicht und wird dementsprechend nicht als eigenständige Sprache anerkannt. Spätertaubte und mittelgradig schwerhörige Menschen sind ebenfalls eine Nutzergruppe dieser Kommunikationsart.

Die Gebärden-unterstützte-Kommunikation (GuK) erfüllt die gleiche Funktion wie die LUG. Sie findet häufig Einsatz bei Kindern, die an dem Down-Syndrom, welches die kognitive Fähigkeiten stark beeinträchtigt, leiden. Die GuK wurde von Frau Prof. Etta Wilken entwickelt und soll Kindern mit einer verzögerten Sprachentwicklung unter zu Hilfenahme von Gebärden das Erlernen der Lautsprache erleichtern. Wie bei der LUG werden diese Zeichengebärden bei den wesentlichen aussagekräftigen Wörtern eines Satzes zur Unterstützung der Informationsübermittlung genutzt. Somit fördert die GuK den Spracherwerb der Lautsprache. Der Unterschied zu den Lautsprachunterstützenden Gebärden ist, dass die GuK keine Gebärden aus der DGS nutzt, sondern eigene konzipierte vereinfachte Gebärdenzeichen.

Zusammenfassend lässt sich sagen, dass die Deutsche Gebärdensprache aufgrund ihres eigenen Wortschatzes, der eigenständigen, abweichenden, komplexen Grammatik zu der Lautsprache sowie ihrer kulturellen Entwicklung in der Sprachwissenschaft akzeptiert und sich als eigener Forschungszweig in der Linguistik etabliert hat. Das folgende Unterkapitel beschäftigt sich mit dem Aufbau und der Funktionsweise der Deutschen Gebärdensprache.

2.2 Phonologie

Die „Phonologie“ ist ein Teilbereich der Linguistik. In der Sprachwissenschaft bezeichnet dieser Begriff die Lehre der Lautsprache und beschäftigt sich mit der Funktion der Laute. Sie kann folgendermaßen definiert werden.

„Die Phonologie als Teilgebiet der Linguistik bezeichnet die Analyse der Rollen und Funktionen von Lauten sowie der Anwendung von Lautunterscheidungen innerhalb eines Sprachsystems [Wil10].“

William Stokoe empfand die Bezeichnung Phonologie für die Gebärdensprache nicht passend, da diese nicht auf Laute sondern auf lautlose visuell sprachliche Gebärden basiert. Er definierte den Begriff „*cherology*“, das aus dem griechischen Wort „*cheros*“ stammt und übersetzt „die Hand“ bedeutet.

Jedoch konnte sich diese Bezeichnung nicht in der modernen Gebärdensprachforschung durchsetzen, weil nicht nur über Handzeichen in der GS kommuniziert wird, sondern auch über nicht-manuelle Komponenten (z. B. Mimik). Somit nutzt man heute weiterhin den Begriff der Phonologie. Dieser wird jedoch etwas allgemeiner definiert, um die gesprochenen und gebärdeten Elemente abzudecken [Boy95], [Dot01], [Pap08].

„Phonologie [ist] die „Lehre von den (Sprech-) Lauten“ und beschäftigt sich mit der Beschreibung der kleinsten bedeutungsunterscheidenden Elemente einer Sprache, den so genannten Phonemen [Pap08].“

Auf den ersten Blick ähnelt diese Definition der vorherigen. Der Unterschied liegt darin, dass in der ersten Definition der Schwerpunkt auf die Analyse von *Lauten* liegt. Bei der zweiten geht es um die Analyse bzw. Beschreibung von bedeutungsunterscheidenden *Elementen*, zu denen Gebärdenzeichen, nonverbale Komponenten aber auch Laute zählen können.

In anderen Fachbüchern wird anstatt Phonologie die Bezeichnung *„sublexikalische Ebene“* verwendet. Eine lexikalische Einheit ist das Wort, das aus den sublexikalischen Einheiten, den Buchstaben, besteht.

Ein weiterer Begriff, den man im Zusammenhang mit der Phonologie setzt, ist die *„Morphologie“*. Dieser Bereich in der Linguistik befasst sich mit den kleinsten bedeutungstragenden Einheiten – den sogenannten Morphemen. In der Lautsprache sind Morpheme die einzelnen Bausteine eines Wortes, die Träger einer Bedeutung sind. „Das Wort *betrachtete* zum Beispiel setzt sich aus drei Morphemen zusammen: dem Stamm (*-tracht-*), der Vorsilbe (*be-*) und der Endung (*-ete*). Ferner können die einzelnen Bausteine der Morpheme, die sog. Phoneme, analysiert werden. Bei den Phonemen handelt es sich um kleinste bedeutungsunterscheidende, nicht aber selbst bedeutungstragende sprachliche Einheiten (hier: *b-e-t-r-a-c-h-t-e-t-e*). Ein Satz der gesprochenen Sprache kann demnach als eine zeitliche Abfolge von Segmenten betrachtet werden, als eine Folge von Wörtern, die ihrerseits aus Folgen von Morphemen, die wiederum aus Folgen von Phonemen bestehen [Boy95].“ In der Gebärdensprache gibt es Gebärden, die nur aus einem Morphem bestehen können und andere die mehrere besitzen. Bei der Gebärde *„ARBEITEN-lustlos“* ist ein Morphem das Gebärdenzeichen für *„ARBEITEN“*, ein weiteres für die Ausführung der langsamen Bewegung, welches für die Lustlosigkeit steht und ein drittes könnte die Mimik sein, die ebenfalls Lustlosigkeit suggeriert und somit die Aussage bekräftigt [Pap08].

Im Laufe des Kapitels stehen komplett groß geschriebene Wörter im Text oder in einer Abbildung für ein Beispiel aus der Gebärdensprache in der Glossen-Notation. Weitere Informationen über dieses Notationssystem enthält das 3. Kapitel, Abschnitt 3.1. In diesem Fall enthält die Glosse *„ARBEITEN-lustlos“* die angehängte kleingeschriebene Emotion *„lustlos“*.

Die phonologischen Sprachinstrumente der Gebärdensprache können in zwei artikulatorische Gesichtspunkte aufgeteilt werden. Diese sind die „**manuellen**“ und „**nicht-manuellen Komponenten**“, welche in den folgenden Abschnitten näher gebracht werden sollen. Eine Übersicht, die als eine Art Mindmap zu verstehen ist und sich in diesem Abschnitt weiterentwickelt, soll die Zusammenhänge der phonologischen Aspekte erleichtern und grundlegende Bewegungsinformationen für die Entwicklung eines Gebärdensprachavatars aufzeigen. Die Begriffe sowie die Unterteilung der einzelnen Elemente sind an den Kenntnissen der DGS von [Pap08] angelehnt und teilweise abgeändert bzw. erweitert. Nach dem aktuellen Verständnisstandpunkt lässt sich die Grafik in folgender Abbildung 2 darstellen.

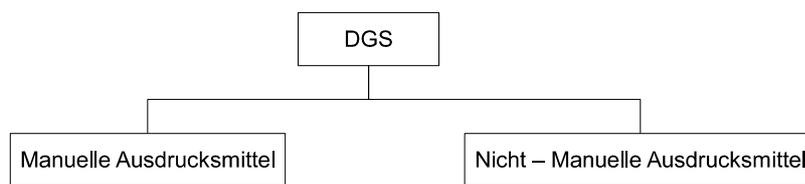


Abb. 2: DGS-Komponenten

2.2.1 Manuelle Komponenten/Ausdrucksmittel/Parameter

Am Anfang der wissenschaftlichen Untersuchungen zu den Gebärdensprachen wurde das Augenmerk eher auf die manuellen Komponenten gelegt. Zu diesen zählt man alle Bewegungen die der Gebärdende per Hand oder Arm ausführt.

Stokoe unterteilte bei der Analyse der ASL 1960 in seiner Arbeit „Sign Language Structure“ diesen Bereich in *Tabula* („**Ausführungsstelle**“), *Designator* („**Handform**“) und *Signation* („**Bewegung**“). Zu diesen grundlegenden Erkenntnissen entwickelten sich spätere Modelle, die zu den Aspekten die „**Handstellung**“ hinzufügten. Die drei Parameter Handstellung, Handform und Ausführungsstelle lassen sich in jedem Gebärdenzeichen wiederfinden. Der Parameter der Bewegung muss nicht immer vorkommen, so zum Beispiel bei der Gebärde für Brille, bei der man die Hände ähnlich wie beim Halten eines Fernglases positioniert. Hierbei spricht man von „*Nullbewegung*“ [Pap08].

Diese vier manuellen Grundkomponenten (vgl. Abb. 3) sind essentiell für die Bildung eines Gebärdenzeichens und werden auf den folgenden Seiten detaillierter erläutert.

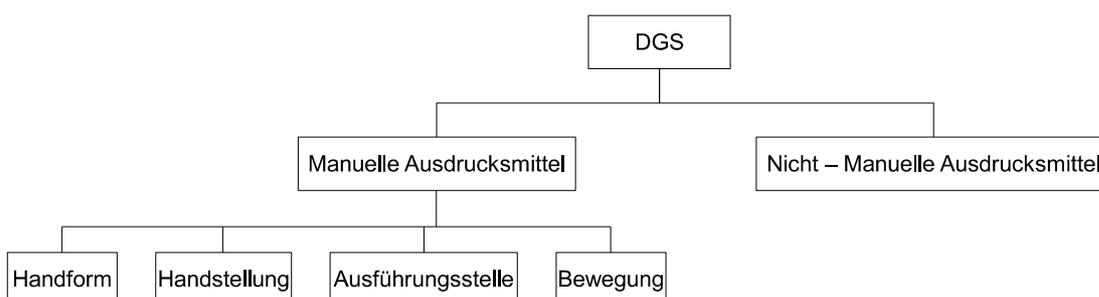


Abb. 3: Manuelle Ausdrucksmittel

2.2.1.1 Handform

In der Gebärdensprache versteht man unter der Handform die äußere Gestalt der Hand. Diese kann in den unterschiedlichen visuellen Sprachen vielseitig und grundlegend anders sein. Zu den bis heute untersuchten GS kristallisierten sich 6 Handformen heraus, die in allen Sprachen („*international*“) auftraten. Hierzu zählt die Faust (1), die Flachhand (2), die „O“-Hand (3), die „C“-Hand (4), der gestreckte Zeigefinger (5) und die Spreizhand (6), welche in der nachfolgenden Abbildung 4 in den roten Kreisen dargestellt sind.

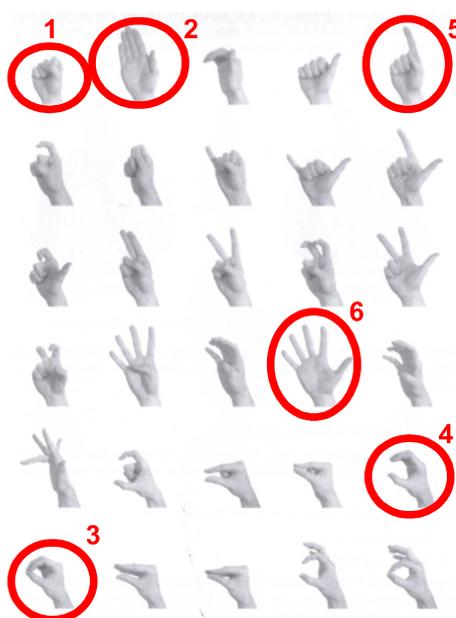


Abb. 4: Handform-Phoneme der DGS;
verändert aus [Pap08]

[Dot01] beziffert die Anzahl der Handformen auf 20 bis 30 in einer GS. Nach [Pri07] und [Pap08] gibt es in der Deutschen Gebärdensprache ca. 30 Handform-Phoneme („*national*“), wie sie in Abbildung 4 zu sehen sind. [Leu04] fand für diese 32 und [Boy95] zählt für die Schweizerdeutsche Gebärdensprache 36 Handformen auf. Eine genaue Anzahl von Handformen ist wahrscheinlich schwer zu ermitteln, da sich die GS ebenfalls wie die Lautsprache ständig weiterentwickelt. So können neue Wörter, Dialekte oder einfach der eigene Sprach-/Gebärdensstil Abweichungen erzeugen.

Zur Bestimmung dieser Phoneme – den kleinsten bedeutungsunterscheidenden Elemente – wird die „*Minimalpaarmethode*“ angewandt [Boy95]. Hierbei wird geschaut, ob kleinste Veränderungen in der Gebärde eine neue Bedeutung erzeugt. In der Lautsprache wäre ein solches „*Minimalpaar*“ „Baum, Raum“ oder „Tisch, Fisch“. Durch die Änderung eines Lautes bekommt das Wort eine neue Bedeutung. Für ein Beispiel in der GS betrachte man die letzte Zeile der Abbildung 4. Die letzte und vorletzte Zeichengebärde ähneln sich sehr stark in ihrem Aussehen. Der Unterschied liegt lediglich in der Berührung zwischen Daumen und Zeigefinger.

Ein weiteres Problem bei der genauen Unterscheidung von Phonemen sind die sogenannten „*Allophone*“. Ein Phonem kann in unterschiedlichen Varianten gebärdet werden, z. B. bei der Handform „Faust“ (vgl. Abb. 5) kann der Gebärdende seinen Daumen kreuzweise über die anderen Finger legen oder er legt diesen seitlich an.



Abb. 5: Allophon Faust; aus [Pap08]

Beide Varianten sind möglich, ändern aber nicht die Bedeutung der Gebärde. Somit beschreibt ein Allophon die Vielseitigkeit der Ausführung eines Phonems ohne Einfluss auf dessen Semantik zu haben. Ein Beispiel für die Analogie dieses Sachverhaltes in der Lautsprache wäre der Laut „r“, der sich je nach Region (in Sachsen wird das „r“ „gerollt“) unterscheiden kann.

2.2.1.2 Handstellung

Die Handstellung beschreibt die Stellung bzw. Lage der Handfläche und der ausgestreckten Finger im „*Gebärdenraum*“. Als Gebärdenraum (kurz GR) bezeichnet man den Wirkungskreis, den ein Gebärdender für die Darstellung der Gebärden nutzt. Dieser dreidimensionale Spielraum lässt sich in den drei Ebenen „*frontal*“, „*waagerecht*“ und „*senkrecht*“ unterteilen, welche in Darstellung 6 ersichtlich sind.

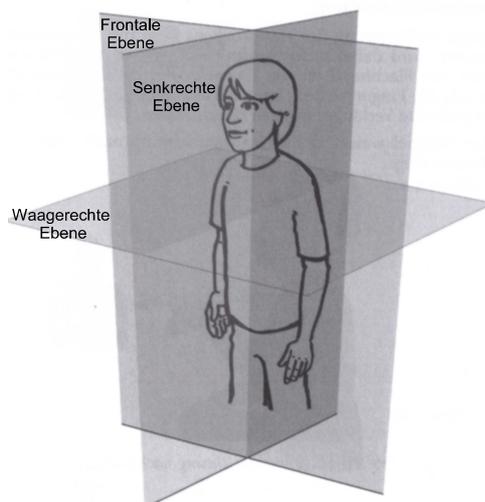


Abb. 6: Ebenen im GR; verändert aus [Pap08]

[Pap08] unterscheidet das Phonem Handstellung in die Subparameter „*Handflächenorientierung*“ und „*Fingeransatzrichtung*“ (bei [Leu04] als „Handrichtung“ bezeichnet).

Die Handflächenorientierung zeigt aus der Perspektive des Gebärdenden die Richtung an, zu der sich die Handinnenseite orientiert. Hierbei gibt es acht häufig auftretende Neigungen, die sich grob unterscheiden lassen nach: rechts oben, rechts, rechts unten, unten, links unten, links, links oben und oben. Weitere können in Richtung des Körpers oder vom Körper weg zeigen.

Bei der Fingeransatzrichtung geht es nicht um die Richtung, in der der Finger zeigt, sondern in welche Richtung – ebenfalls aus der Sicht des Gebärdenden – der Handrücken bei einer geradlinigen Verlängerung nach vorne zeigt (siehe Abbildung 7). Somit kann z. B. bei der Handform „Flachhand“ der Handrücken und die Finger in dieselbe Richtung zeigen, jedoch bei der Handform „Faust“ zeigen die Finger auf einen anderen Punkt als die gedachte Linie über den Handrücken [Boy95], [Pap08].



Abb. 7: Fingeransatzrichtung; aus [Pap08]

Durch die Beweglichkeit der Fingergelenke und des Handgelenks lassen sich die Fingeransatzrichtung und die Handflächenorientierung in allen drei Ebenen des Gebärdenraums zuordnen. Somit können diese in jede Richtung des dreidimensionalen Raumes zeigen.

Die betrachteten Aspekte werden in der Darstellung 8 mit aufgenommen. Diese sieht nun wie folgt aus:

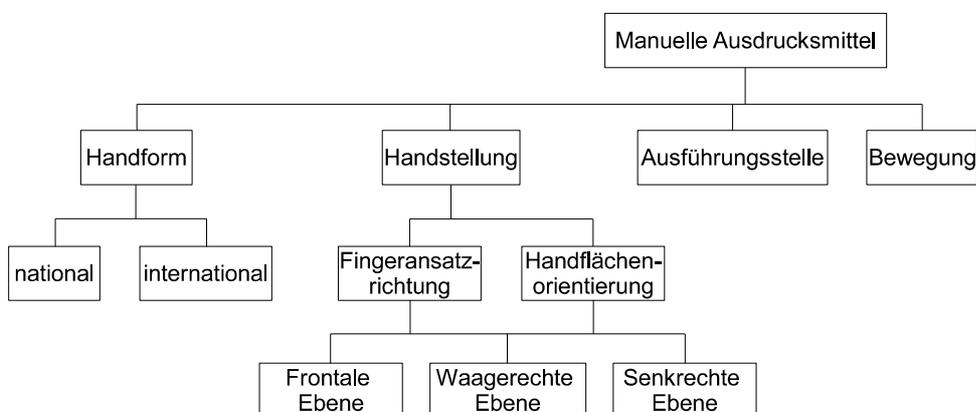


Abb. 8: Handform und Handstellung der DGS

2.2.1.3 Ausführungsstelle

Die Ausführungsstelle definiert den Bereich in dem eine Gebärde vollzogen wird. Es wird unterschieden zwischen körpergebundenen Gebärden und welche, die sich im GR ausführen lassen. Zu den „*körpergebundenen Gesten*“ zählen die Regionen, die sich nah am Körper befinden. Dazu zählen Kopf, Gesicht, Rumpf, Arm und Hand. [Dot01] zählt generell etwa 15 bis 20 distinktive Körperstellen für Gebärdensprachen auf. Nach Liddell & Johnson gibt es in der ASL 20 Gebiete. [Pap08] definiert insgesamt 30 Ausführungsstellen für die DGS, die er folgendermaßen untergliedert [Pap08]:

- Kopf: vor dem Kopf, über dem Kopf, hinter dem Kopf, Schläfe, Ohr, Hals,
- Gesicht: Stirn, Auge, Nase, Wange, Mund, Kinn,
- Rumpf: Schulter, Brust, Bauch, Unterleib, Hüfte,
- Arm/Hand: Oberarm, Ellenbogen, Armbeuge, Unterarm, Handgelenk, Puls, Handrücken, Handinnenfläche, Handkante, Fingerkante, Fingerbeere bzw. Fingerkuppe, Fingerspitze, Fingernagel

Es könnte der Eindruck entstehen, dass die DGS mehr Ausführungsstellen als die ASL hat. Jedoch muss dabei beachtet werden, dass es keine eindeutigen Raumdefinitionen für die einzelnen Körperstellen gibt. So bezeichnen z. B. Liddel & Johnson eine Stelle als „Oberkörper“. Zu dieser könnte man natürlich nun die Bereiche „Bauch, Unterleib, Hüfte, Brust“ usw. aus [Pap08] zählen. Auf der anderen Seite wird bei Liddel & Johnson ein Teil des Gesichtes in „Mund, Lippe, Unterkiefer, Kinn“ aufgeschlüsselt, was detaillierter ist als bei [Pap08]. Es wäre wünschenswert eine eindeutige Begriffsverwaltung zu verwenden, da so sich die unterschiedlichen Gebärdensprachen besser vergleichen lassen könnten.

Wie aus der Untergliederung zu erkennen ist, werden die Gebärden aus Komfortabilität über der Gürtellinie gebärdet. Dabei muss nicht die Körperstelle direkt berührt werden, sondern kann auch kontaktlos erfolgen.

Die „*Ausführungsstellen im Gebärdenraum*“ befinden sich in einem dreidimensionalen Raum und werden dementsprechend in 3 Ebenen gegliedert, zu denen die frontale, waagerechte und senkrechte Ebene zählt. Diese können in weitere grobe Zonen untergliedert werden (siehe Abbildung 9).

Frontale Ebene:

- in Kopfhöhe
- in Oberkörperhöhe
- in Bauchhöhe

Waagerechte Ebene:

- nah am Körper
- vor dem Körper (mitte)
- weit vor dem Körper (weit)

Senkrechte Ebene:

- auf der linken Seite
- auf der rechten Seite
- Mitte vor dem Körper
(„*neutrale Zone*“)

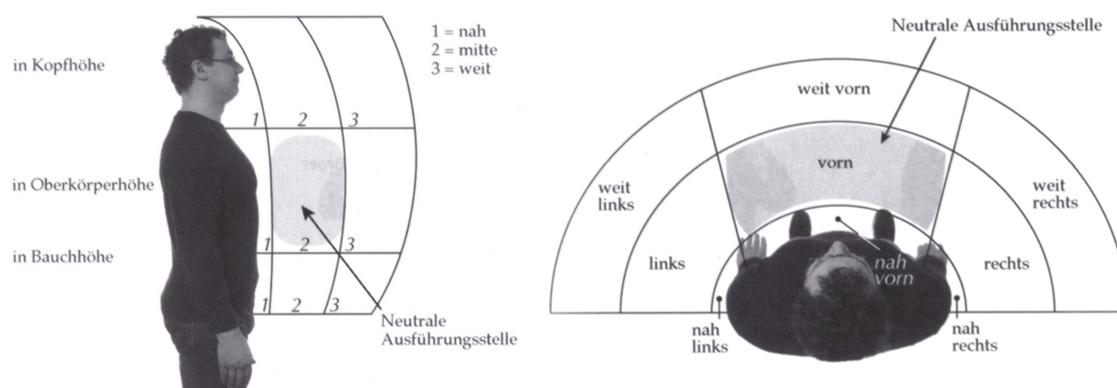


Abb. 9: Ausführungsstellen im GR; aus [Pap08]

Die meisten Gebärden werden auf der Höhe der Brust kurz vor dem Körper produziert. Dieser Bereich wird als neutrale Ausführungsstelle bezeichnet [Pap08]. Will sich der Gebärdende z. B. an einem Esstisch mit dem Sitznachbar ungestört unterhalten, führt er die Gesten im Raum etwas kleiner aus, reduziert die Bewegung und setzt die Ausführungsstelle etwas tiefer an als üblich. In der Lautsprache würden sich die Kommunikationspartner durch senken der Lautstärke in ihrer Stimme mittels „flüstern“ unterhalten. Sitzt nun der Gesprächspartner auf der anderen Seite des Esstisches, kann die Geste im Gebärdenraum größer als üblich ausfallen, was in der Lautsprache mit erhöhen der Lautstärke („zurufen“) vergleichbar ist [Boy95].

2.2.1.4 Bewegung

Der Parameter der Bewegung sagt aus, wie die Bewegung innerhalb eines Gebärdenzeichens vollzogen wird. Am Anfang wurde dieser Teil von den Sprachwissenschaftlern nahezu gemieden. Die anderen manuellen Ausdrucksmittel waren leichter wissenschaftlich zu analysieren. Bis eine Studie von Poizner, Bellugi und Lutes-Driscoll (1981) diesen Aspekt in den Vordergrund rückte. Bei der Untersuchung bekamen Versuchspersonen an bestimmten Körperstellen Lichtquellen angebracht. Diese Testpersonen haben in vollständiger Dunkelheit Gesten gebärdet, welche von anderen Gehörlosen per Video erkannt werden mussten. Das Ergebnis der Studie zeigte, dass anhand der Bewegung 86 % der Zuschauer die Gebärden erkennen konnten und somit wurde das Ausdrucksmittel der Bewegung für die Linguisten relevanter [Boy95].

Heute unterscheidet [Pap08] diesen Parameter in „**externe Handbewegung**“, „**interne Handbewegung**“ und „**Bewegungsarten**“.

Handexterne Bewegungen können im Gebärdenraum oder am Körper vollzogen werden. Die Änderung der Position innerhalb des Gebärdenzeichens und somit die Bewegung zwischen dem Anfang und dem Ende der Gebärde bezeichnet man als „**Bewegungsspur**“, welche man sich bildlich vorstellt. Ein Beispiel hierfür wäre, wenn ein Auto eine schlangenförmige Linie auf der Straße fährt, welches durch eine kurvenartige Bewegung mit der Hand nachgestellt

wird. Wichtig ist in diesem Beispiel auch die Richtung. Fährt das Auto rückwärts oder vorwärts? Diese Erkenntnis der „**Bewegungsrichtung**“ gewinnt der Kommunikationspartner durch den Ablauf der Gebärde, d. h. durch den Start- und Zielpunkt.

Handinterne Bewegungen sind Bewegungen, bei denen die Finger bzw. die ganze Hand eine Rolle spielen. Die Finger können in die verschiedensten Stellungen gebracht werden. Sie können sich aneinander reiben, krümmen, knicken, spreizen, öffnen, schließen, auffächern und alternierend auf- und abbewegen (Fingerspiel). Dieser handinterne Bewegungsablauf wird dementsprechend als „**Fingerbewegung**“ bezeichnet. Ein weiterer Untergliederungspunkt neben der Fingerbewegung ist die „**Bewegung mit Handgelenksbeteiligung**“. Dazu nutzt der Gebärdende nicht nur sein Handgelenk, er kann auch den Unterarm und das Ellenbogengelenk zur Unterstützung des Gebärdenzeichens nehmen. Übliche Bewegungen sind Beugen, Drehen und Schütteln des Handgelenks. Die handinternen und -externen Bewegungen können miteinander kombiniert werden.

Der dritte Subparameter des manuellen Ausdrucksmittel Bewegung ist die Bewegungsart. Ein Gebärdenzeichen kann unbestimmt wiederholt werden. Diese Art der „**Bewegungshäufigkeit bzw. -wiederholung**“ findet zum Beispiel bei der Verneinung eines Verbs Anwendung. Wird das Verb „ARBEITEN“ zweimal wiederholt, wobei bei der zweiten Wiederholung eine leichte Versetzung der Gebärde im Gebärdensraum erfolgt, heißt das in der Semantik „nicht arbeiten“. Eine weitere Funktion, die durch die Wiederholung erreicht wird, ist eine Aussage zu bekräftigen. Eine andere Bewegungsart ist die „**Bewegungsgeschwindigkeit**“, die ebenfalls unterschiedliche Funktionen erfüllen kann. Ein Beispiel wurde schon gebracht. Indem man mit seinem Partner „leise“ am Tisch kommuniziert. Die Bewegungsgeschwindigkeit kann auch einen semantischen Nutzen haben und somit die Aussage bekräftigen. Sie dient unter anderem als Stilmittel in Geschichten, bei der zum Beispiel ein hüpfender Hase mit der Hand in Zeitlupe dargestellt wird. Um eine Aussage zu verstärken, kann auch der „**Bewegungsumfang**“ eine enorme Rolle spielen. Der Umfang einer Gebärde kann bei jedem sehr unterschiedlich sein. Zum Einen kann das der eigene Gebärdensstil sein, zum Anderen kann es von der Situation abhängen (z. B. jemandem etwas „zurufen“). Eine andere Anwendung dieser Bewegungsform ist das Beschreiben von Objekten durch Maßangaben [Pap08].

Auf der nachfolgenden Seite sind die grundlegenden manuellen Komponenten in einer zusammenfassenden Übersichtsgrafik (vgl. Abb. 10) abgebildet. Sie soll als Überblick zu den bereits erläuterten Aspekten dienen und schafft die Grundlage zur Ableitung der Anforderungen eines Avatars. Die nächsten beiden Abschnitte behandeln die Zweihandgebärden und den Punkt der Simultaneität, die ebenfalls bei der Darstellung der Gebärdensprache mit einer animierten Figur einfließen sollten. Nach diesem kleinen Einschub folgt die detaillierte Erläuterung zu den nicht-manuellen Komponenten im Abschnitt 2.2.2 der Deutschen Gebärdensprache.

2.2.1.5 Zusammenfassung: manuelle Ausdrucksmittel

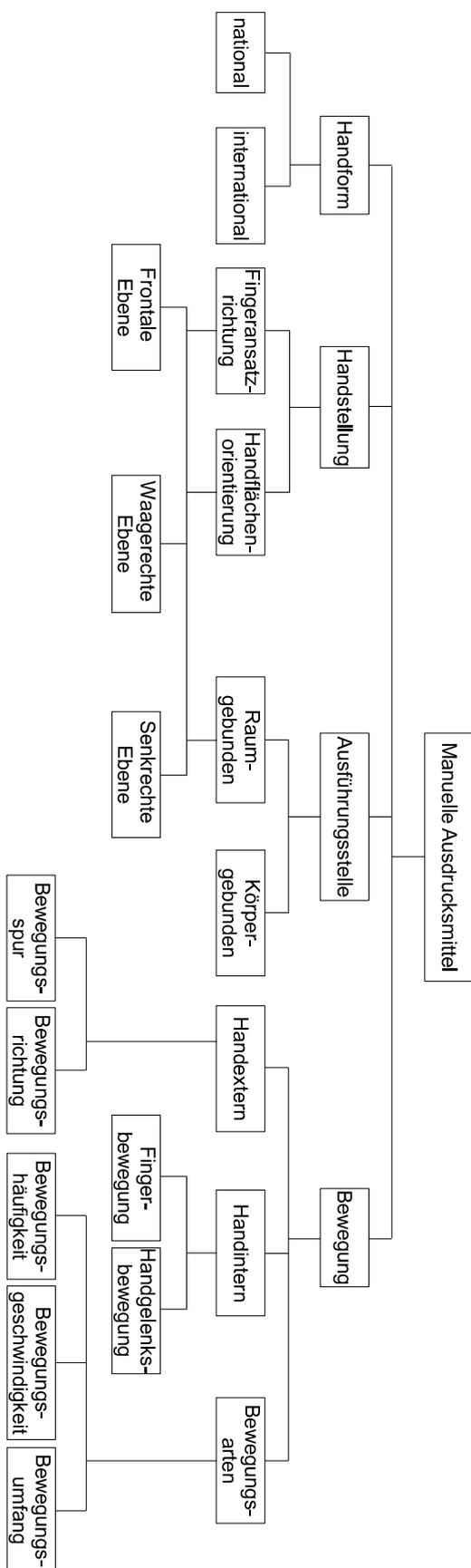


Abb. 10: Übersicht manuelle Komponenten der DGS

2.2.1.6 Zweihandgebärden

Die bis hierher erläuterten manuellen Komponenten können alle mit einer Hand gebärdet werden. Die bevorzugte Hand, in der Regel die Schreibhand, für die Darstellung der Gebärden wird als starke bzw. „*dominante Hand*“ bezeichnet und findet ihren Einsatz vorrangig bei Einhandgebärden. Die schwächere Hand, bei Rechtshänder die linke Hand, erhält die Bezeichnung der „*nichtdominanten Hand*“. Nutzt der Gebärdende diese beiden Artikulatoren, muss er bestimmte Regeln beachten.

◆ Regel der Symmetrie

Bei der Ausführung der Gebärden mit zwei Händen müssen diese sich gleichzeitig bewegen und dieselbe Handform aufweisen. Die manuellen Komponenten Handform, Ausführungsstelle und Bewegung fallen gleich aus. Die dominante und nichtdominante Hand können dabei synchrone oder alternierende Bewegungen vorführen, je nachdem wie der Parameter der Handstellung gebärdet wird. Die synchronen symmetrischen Bewegungen können parallel oder spiegelbildlich verlaufen. Die Gewichtung der Hände (dominant und nichtdominant) spielt keine entscheidende Rolle. Die Darstellung 11 zeigt Beispiele für die parallele, spiegelbildliche und alternierende Symmetrie.

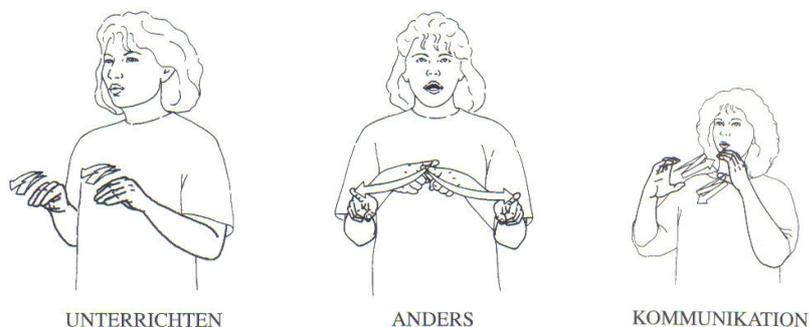


Abb. 11: Parallele, spiegelbildliche und alternierende Symmetrie; aus [Pri07], [Sie07]

◆ Regel der Dominanz

Besitzen die Hände in der DGS bei einer Zweihandgebärde unterschiedliche Handformen wird die dominante Hand aktiv bewegt und die nicht-dominante Hand bleibt in einer passiven Stellung und kann gelegentlich als Ausführungsstelle genutzt werden (vgl. Abbildung 12). „Darüber hinaus muß die nichtdominante Hand eine der 6 Grundhandformen annehmen. Die Grundhandformen können dabei jede beliebige Orientierung einnehmen [Gro99].“ Die Aussage des Zitats findet man in verschiedenen Fachbüchern und kann eher als Hypothese aufgefasst werden. Dieser Aspekt muss sich erst noch bei der Untersuchung der DGS herausstellen, um die Aussage zu unterstützen [Leu04], [Pri07], [Sie07].

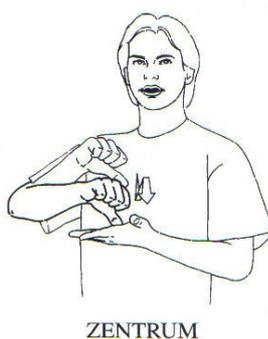


Abb. 12: Obere Hand: aktiv; untere Hand: passiv; aus [Sie07]

2.2.1.7 Simultaneität

Einer der wesentlichen Punkte bei der Unterscheidung zwischen Lautsprache und Gebärdensprache ist der der „*Simultaneität*“ oder auch Gleichzeitigkeit. Die Lautsprache nutzt den akustischen Kanal zur Informationsübermittlung, bei dem die Laute vom menschlichen Ohr nur sequentiell verarbeitet werden können. In der Gebärdensprache gibt es ebenfalls einen sequentiellen Anteil, bei dem die Informationen in einer konkreten festgelegten zeitlichen Abfolge durch die Satzglieder übermittelt werden. Anstatt des akustischen Übertragungskanal nutzt die GS den visuellen Kanal. Dieser erlaubt Informationen parallel bzw. simultan neben den sequentiell dargestellten manuellen Gebärdenzeichen, mittels nicht-manuelle Komponenten, in denen weitere Informationen codiert werden, zu tätigen.

Diese Simultaneität erlaubt den Gebärdenden, trotz zeitlich langsameren Zeichengebärden, die im Gegensatz zu den schnellen akustischen Lauten der LS stehen, den Informationsgehalt in ähnlicher Zeit den Gesprächspartner zu vermitteln. In den Quellen [Kar09] und [Wil10] können weitere Informationen zur Produktionsgeschwindigkeit von Gebärdensprache und Lautsprachen entnommen werden.

2.2.2 Nicht-manuelle Komponenten/Ausdrucksmittel/Parameter

Die nicht-manuellen Komponenten spielten früher in der Forschung eine geringe Rolle und wurden bei Untersuchungen der Gebärdensprache vernachlässigt. Später wurde entdeckt, dass diese visuellen Kommunikationsmittel, zu denen der „*Gesichtsausdruck/Mimik*“, die „*Blickrichtung*“, die „*Kopf- sowie Oberkörperhaltung*“ und die „*Mundartikulation*“ gehören (vgl. Abbildung 13), eine wesentliche Bedeutung in der Gebärdensprache haben. So können sie die Ausdrucksweise der manuellen Komponenten erheblich beeinflussen und ermöglichen Informationen simultan über den visuellen Informationskanal zu den Kommunikationspartnern zu übermitteln.

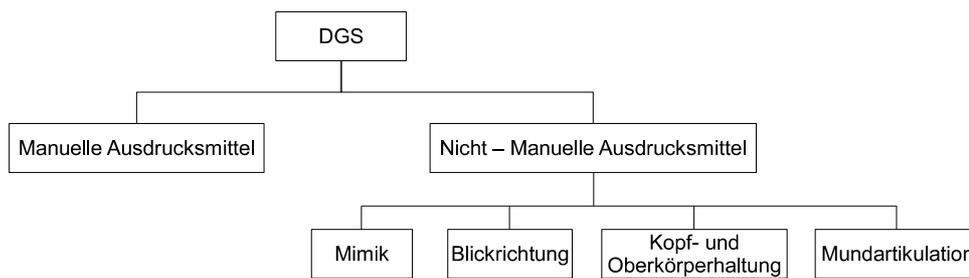


Abb. 13: Nicht-manuelle Ausdrucksmittel

2.2.2.1 Gesichtsausdruck/Mimik

Das menschliche Gesicht besitzt 26 Gesichtsmuskeln von denen im wesentlichen 11 Muskeln für die Mimik, die die sichtbaren Bewegungen der Gesichtsoberfläche anzeigt, verwendet werden. Die wichtigsten Bewegungen der Mimik sind im Bereich des Mundes und der Augen zu finden.

[Pap08] unterscheidet die Mimik in „*nichtsprachliche Mimik*“ und „*sprachliche DGS-Mimik*“. Zu dem nichtsprachlichen Gesichtsausdruck zählen die Grundemotionen (vgl. Darstellung 14) eines Menschen (bezeichnet als „*universelle Mimik*“) und die „*kulturelle Mimik*“. Ein Beispiel für kultureller Gesichtsausdruck ist, dass in Deutschland Trauer in der Öffentlichkeit akzeptiert wird und Leute dementsprechend hilfstellend reagieren können. In Japan werden negative Gefühle wie Zorn, Trauer oder Enttäuschung nur den engsten Freunden oder Verwandten angezeigt, aber in der Öffentlichkeit kann über ein trauriges Geschehen mit einem Lächeln im Gesicht erzählt werden. Dieses Lächeln dient dazu, den Schmerz zu kaschieren und dem Gesprächspartner eine gewisse Verpflichtung zur Hilfestellung zu ersparen.



Abb. 14: Basisemotionen; aus [Pap08]

Zu dem Bereich der sprachlichen DGS-Mimik zählen Gesichtsausdrücke, die eine Funktion erfüllen. Diese sind unter anderem das Darstellen von Adverbien, die Kennzeichnung von Hinweispronomen, Satztypen (z. B. Relativ- und Konditionalsätze, Fragen, Verneinungen), Satzmodifizierungen, Satzverbindungen und der Rollenübernahme [Pap08], [Dot12].

2.2.2.2 Augenausdruck/Blickrichtung

Die Blickrichtung spielt nicht nur bei der Gebärdensprache eine Rolle, sondern auch bei der nonverbalen Kommunikation in der Lautsprache. Dieser Aspekt der zwischenmenschlichen Kommunikation kann bewusst oder unbewusst ausgeführt werden. So können die Augen den Gemütszustand (z. B. gelangweilt, ängstlich) oder sogar Charakterzüge (z. B. selbstbewusst) des Gesprächspartners abbilden. In der Gebärdensprache gibt es vier wesentliche Bereiche, die bewusst mit den Augen anvisiert werden können. Dazu zählt der „**Blick zum Gesprächspartner**“, der „**Blick aus der Sicht einer Rolle**“, der „**Blick in den Gebärdenraum**“ und die „**Fokussierung auf die gebärdende Hand**“.

In einem Gespräch gilt es als höflich und respektvoll, der gegenüberliegenden Person in zufälligen Zeitabständen ins Gesicht (z. B. auf den Mund) oder direkt in die Augen zu schauen. Dadurch erhält der aktive Gesprächspartner Informationen, ob der Kommunikationspartner den Inhalt des Gesprochenen versteht und folgen kann. Der gegenüberliegende Part kann durch den Blick zum Beispiel symbolisieren, ob er eine Aussage zustimmt, ob er ihn unterbrechen will, da er vielleicht eine andere Meinung hat oder durch einen interessierten Blick weiter zu hört.

Will der Gebärdende eine Geschichte oder ein Gespräch zwischen zwei Menschen „nacherzählen“, nutzt er die Methode der Rollenübernahme. Wird zum Beispiel eine Konversation zwischen einem Mann und einer Frau wiedergegeben, schlüpft der Erzähler in die Rollen dieser Personen. Dabei stellt er auch den Augenausdruck und die Kopf- sowie Körperhaltung der zu spielenden Personen dar. Ist die Frau in diesem Gesprächsverlauf traurig, nutzt der Gebärdende diese Emotion und schaut traurig in Richtung des Mannes.

Der Blick in den Gebärdenraum wird genutzt, um Objekte oder Personen, die vorher durch Indizierung einen Standpunkt im Gebärdenraum zugewiesen bekommen haben, dem Gesprächspartner anzuzeigen. Mit Hilfe dieses Blickes wird die Verständlichkeit in der Kommunikation erhöht, da der Gesprächsteilnehmer weiß, über welches Objekt oder über welche Person das Gespräch gerade gehalten wird. Aus grammatikalischer Sicht nutzt der Sprecher hier die Funktion von Pronomen (Personalpronomen (z. B. ich, du, er/sie), Possessivpronomen (z. B. mein, dein, sein/ihr), Hinweispronomen (z. B. dieses, jenes, dort)). Wichtig ist, dass der Erzähler eine klare Abgrenzung zwischen den indizierten Gegenständen bzw. Menschen, z. B. durch einen gewissen Raumabstand bei der Indizierung, lässt und nicht zu viele Indizes in den Gebärdenraum setzt, da sonst die Verständlichkeit und Übersicht beim Gegenüberliegendem verloren gehen und Missverständnisse entstehen können.

Der Blick auf die Hand ist dann von Nöten, wenn der Erzähler die Aufmerksamkeit des Gesprächspartners führen will. So kann zuerst auf einen Bezugspunkt im Raum geschaut bzw. gedeutet werden, der zum Beispiel für ein Auto steht. Dieser Punkt könnte im linken Bereich des Gebärdenraums indiziert sein. Ein zweiter Punkt im rechten Teil des Gebärdenraums könnte für ein Krankenhaus stehen. Durch den folgenden Blick auf die Hand und deren Bewegung zum Krankenhaus-Bezugspunkt kann dargestellt werden, wie sich das Auto bewegt (z. B. schlängelnd, rasend). Andere Anwendungsbereiche sind skizzenhafte oder maßanzeigende Gebärden, bei denen die Hand die Größe und Form/Gestalt eines Objektes nachzeichnet [Pap08].

2.2.2.3 Kopf- und Oberkörperhaltung bzw. -bewegung

Ein weiterer Subparameter der nicht-manuellen Ausdrucksmittel stellt die Kopf- bzw. Körperhaltung und deren Bewegungen dar. Die Positionierung der Beine haben in der Gebärdensprache keine Funktion. Dies lässt sich dadurch begründen, dass es für den Gebärdenden umständlich und zeitaufwendig sowie praktisch eher hinderlich wäre. In der Lautsprache gibt es verschiedene Sprachstile, welche sich z. B. durch zu schnelles, undeutliches, langsames, stockendes Sprechen äußern können. In der Gebärdensprache treten ebenfalls verschiedene Gebärdenstile auf, bei der die Oberkörperhaltung eine Rolle spielt. So kann diese sehr individuell von bewegungsarm bis bewegungsreich ausfallen. Jedoch gibt es feste Bereiche, wo die Kopf- und Oberkörperbewegung nicht fehlen dürfen. Wie in dem Abschnitt 2.2.2.2 schon angedeutet, ist einer dieser Bereiche die Rollenübernahme. Wenn aus der Sicht einer Rolle gebärdet wird, positioniert sich der Kopf und Körper entsprechend der Position im Gebärdenraum. Zum Beispiel wird eine alte Dame mit Buckel dementsprechend nachgestellt. Diese gestischen Informationen zeigen dem Gesprächspartner an, wer zu wem gerade spricht. Somit kann auch die grammatikalische Funktion der direkten und indirekten Rede dargestellt werden.

Direkte und indirekte Rede

Ein Beispiel für die indirekte Rede ist: „Paul sagte, dass er ein kluger Kerl sei!“

Für diese Aussage zeigt der Gebärdende auf eine Person (im Beispiel Paul), die im Gebärdenraum indiziert wurde. Die gebärdende Person bleibt dabei mit dem Körper zum Kommunikationspartner gewandt. Der Gesichtsausdruck ist neutral und der Blick wird zum Zuhörer gehalten.

Bei der direkten Rede fällt der Satz wie folgt aus: Paul sagte: „Du bist ein kluger Kerl!“

In diesem Fall nimmt der Gebärdende durch Drehung des Oberkörpers die Rolle von Paul ein, der vorher mittels Indizierung im Gebärdenraum festgelegt wurde. Die Blickrichtung und Mimik von Paul wird übernommen und der Gebärdende gebärdet die Aussage „Du bist ein kluger Kerl!“ [Boy95].

Die Kopfhaltung kann auch grammatikalische Aufgaben erfüllen. So wird bei Fragestellungen, die mit Ja/Nein beantwortet werden können oder Entscheidungsfragen, der Kopf leicht nach vorne in Richtung des Gesprächspartners geneigt. Konditionalsätze, also Sätze, bei denen eine Aussage mit einer Bedingung verknüpft ist, werden ebenfalls mit leicht geneigtem Kopf nach vorne angezeigt. Zieht der Gebärdende den Kopf zurück, stellt er eine „W-Frage“. Zustimmung oder Verneinungen werden wie in der Lautsprache mittels Kopfnicken bzw. Kopfschütteln verdeutlicht [Boy95].

2.2.2.4 Mundartikulation

Der Mund ist nicht nur ein wesentlicher Bestandteil der Sprachartikulation in der Lautsprache, sondern er hat auch eine relevante Funktion in der Gebärdensprache. Man unterscheidet das „*Mundbild*“ und die „*Mundgestik*“.

Das Mundbild oder auch als Wortbild bezeichnet, ist die vollständige oder teilweise stimmlose Artikulation eines Wortes aus der Lautsprache, das simultan zu einer Gebärde „nachgesprochen“ wird. Dabei ähneln sich die Lippenbewegungen der Artikulation von Wörtern der Gebärdensprache denen der Lautsprache. So wird zum Beispiel bei der Gebärde für Wetter das Wort „Wetter“ mit dem Mund artikuliert. Dabei kann der Gebärdende das Wort vollständig oder nur teilweise ausdrücken. In der Regel wird hier der markant visuell gestische Teil des Wortes auf den Lippen sichtbar dargestellt. Der Zweck von Mundbildern ist es, Wörter zu artikulieren, für die es keine Gebärdenzeichen gibt. So können Namen, Fachwörter oder Eigenbezeichnungen wie Städte mit Hilfe des Lippenablesens dem gegenüberliegenden Gesprächspartner vermittelt werden.

[Boy95] schildert zwei Varianten für das Mundbild. In der ersten Variante werden die Wörter vollständig stimmlos ausgesprochen, was dem Flüstern von Wörtern und Sätzen in einer Lautsprache sehr nah kommt. Das Mundbild muss nicht parallel zu einer Gebärde produziert werden und kann über längere Phasen im Gespräch vorkommen. Diese Form von Mundartikulation nutzen vermehrt Menschen, deren Muttersprache eine Lautsprache ist und die Gebärdensprache zusätzlich gelernt haben (z. B. Dolmetscher). Eine Anwendung dieser Variante des Mundbildes findet man in den lautsprachbegleitenden Gebärdungen LBG (vgl. Abschnitt 2.1).

Bei der zweiten Variante ahmt der Gebärdende Wörter aus der gesprochenen Sprache mit seinen Lippen nach. Diese Wörter stammen ursprünglich aus der Lautsprache. Somit wird diese Form der nicht-manuellen Komponenten nicht über große Teile des Gesprächs genutzt, sondern nur in Verbindung mit manuellen Komponenten. In der Regel begleitet das Mundbild in einem Satz 2 bis 3 manuelle Gebärden. Diese Variante wird in der DGS genutzt.

Eine weitere Eigenschaft des Mundbildes ist, dass es über ein Wort hinaus gedehnt werden kann. Zum Beispiel bei der Gebärde „GEHÖRLOS DU“ erfolgt das Mundbild für das Wort „gehörlos“ über die komplette Aussage. Des Weiteren wird das Wort bei der Darstellung nicht flektiert, also gebeugt, z. B. wenn man sagt „ich aß“ wird beim Mundbild „essen“ artikuliert,

bei Verben der Infinitiv oder eine Stammform und beim Substantiv stellt man das Wort im Singular Nominativ dar. Im Groben hat das Mundbild die Funktion, Informationen zur Verfügung zu stellen, die eine Gebärde eindeutig erklärt. So wird zum Beispiel „BRUDER“ und „SCHWESTER“ mit der gleichen Gebärde vorgeführt, wie Abbildung 15 zeigt.

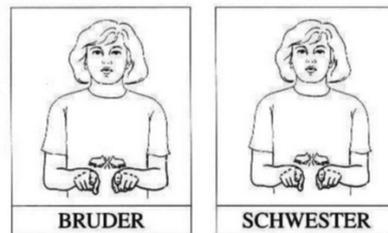


Abb. 15: Unterscheidung durch Mundbild; aus [Sie07]

Der Kommunikationspartner erkennt den Unterschied nur an Hand des Mundbildes, welches „Bruder“ oder „Schwester“ artikuliert. Des Weiteren verbessert es den Kontext einer Aussage. So wird das Wort „Hamburger“ bei der Gebärde für „FLEISCH“ per Mund ausgesprochen. Bei Namen von Menschen oder Orte dient das Mundbild als alleiniger Bedeutungsträger und muss nicht in Verbindung mit einer manuellen Komponente geäußert werden.

Mundbild international

In Amerika findet das Mundbild in der ASL selten Anwendung. Für die Artikulation von Fachwörtern oder Eigennamen wird das Fingeralphabet genutzt, bei dem man per Fingerzeichen die Wörter buchstabiert. Dies hat kulturelle Hintergründe, da das Gebärden ohne Nutzung des Mundbildes als ästhetischer und wohl erzogener galt als das übertriebene Sprechverhalten, was eher verspottet wurde. In der heutigen Zeit liegt eine ansteigende Akzeptanz gegenüber der Nutzung des Mundbildes in Amerika vor. Wird diese Art der Mundartikulation genutzt, wird jedoch versucht die Darstellung nicht zu sehr zu übertreiben. In den europäischen Ländern, wie zum Beispiel in der Schweiz und Deutschland, findet das Mundbild ein höheres Ansehen und wird dem Fingeralphabet bevorzugt verwendet.

Mundgestiken, häufig auch als Mundmimik bezeichnet, haben die Eigenschaft, Gebärden mit Informationen zu unterstützen bzw. zu bekräftigen. So kann bei der Gebärde für „WIND“ der Gebärdende parallel mit dem Mund Luft herausblasen. Eine Reihe von Gebärden, darunter viele Adverbien, nutzen diese Form der Artikulation. Der Unterschied zum Mundbild ist, dass die Mundmimik kein Bezug auf die Artikulationsbewegung, wie sie bei der lautsprachlichen Aussprache des Wortes auftreten würde, hat und somit sich das Wort nicht durch Lippenlesen verdeutlichen lässt. „[...] sie tritt immer mit der entsprechenden Gebärde gemeinsam auf und darf nicht weggelassen werden [Sie07].“ So muss bei der Gebärde für „ZUSAMMENSTOßEN“ die Mundgestik ein „bam“ darstellen. Die Mundgestik kann neutral, gespreizt oder offen sein. Demgemäß können auch die Zungenspitze und die Zähne beim Verständnis der Geste hilfreich sein [Boy95], [Dot01], [Sie07], [Pap08], [Kar09].

2.2.2.5 Zusammenfassung: nicht-manuelle Ausdrucksmittel

Die Abb. 16 stellt die Gesamtübersicht für die nicht-manuellen Komponenten dar. Zusammen mit der Darstellung 10 aus 2.2.1.5 deckt sie den Bereich der Phonologie für die Deutsche Gebärdensprache ab. Der nachfolgende Abschnitt 2.3 stellt die grammatikalischen Aspekte der DGS vor.

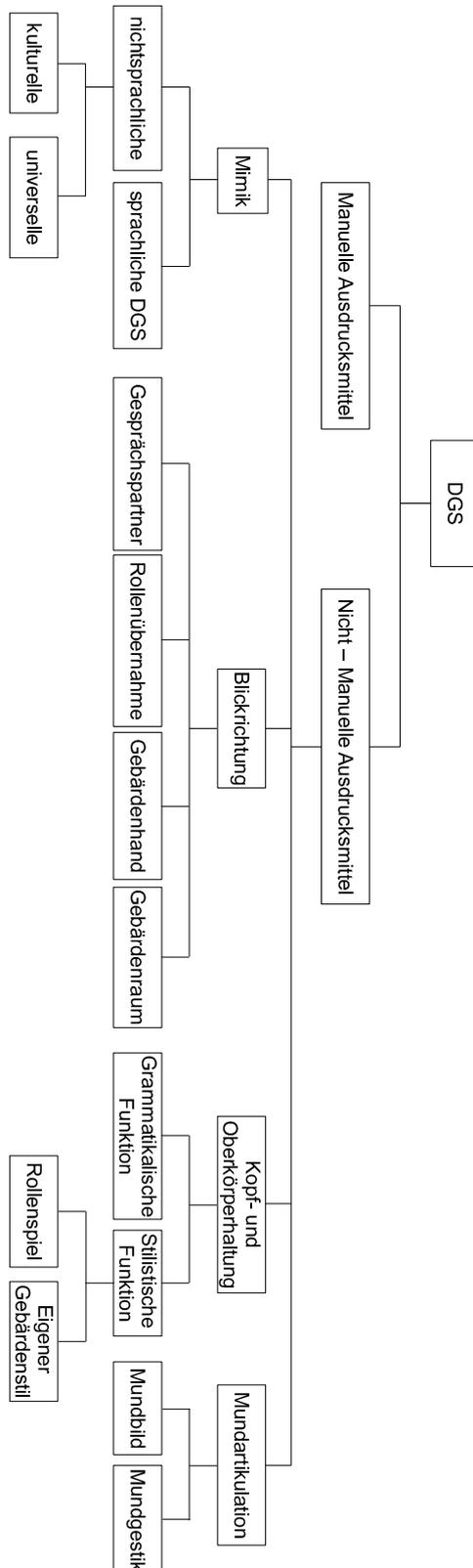


Abb. 16: Übersicht nicht-manuelle Komponenten der DGS

2.3 Grammatik und Syntax (Satzlehre)

In diesem Teilkapitel werden grundlegende Aspekte der Grammatik der Deutschen Gebärdensprache näher gebracht. Diese kann man nicht mit der Grammatik der deutschen Sprache vergleichen, sondern hat speziell ihre eigenen Merkmale. So gibt es in der DGS keine Artikel (der, die, das) oder Kopulae (ist, sein, haben). Am Ende des Kapitels verdeutlicht eine Übersicht (vgl. Abschnitt 2.3.4) die wesentlichen Bestandteile der DGS-Grammatik.

2.3.1 Satzglieder

„Um einen Sachverhalt sprachlich auszudrücken, werden Satzglieder benötigt. Sie bilden die grammatischen Einheiten, welche die Aussage, die Bezugsgegenstände und ergänzende Informationen über den Sachverhalt ausdrücken [Pap08].“ Man kann sie als die Bausteine bezeichnen, die für eine grammatikalisch korrekte Bildung eines Satzes benötigt werden. In der deutschen Lautsprache folgt nach einem Subjekt das Prädikat und nach diesem das Objekt. In der Deutschen Gebärdensprache unterscheidet sich die Reihenfolge durch das Vertauschen von Objekt und Prädikat. Das hat zur Folge, dass ein Satz in der DGS folgenden Ablauf der Satzglieder erhält:

Subjekt → Objekt → Prädikat

„[...] DGS-Sätze[-] erlaub[en] keine Umstellung der Satzglieder und das Verb steht stets am Satzende. Dies gilt sowohl für Haupt- als auch für Nebensätze und Verben der DGS werden nicht nach Tempus flektiert [Leu04].“

Subjekt

In der Sprachwissenschaft bezeichnet das Subjekt eine Person oder ein Sachverhalt in einem Satz. Es kann nicht alleine in einem Satz stehen und bildet zusammen mit dem Prädikat die Hauptaussage eines Satzes. In der Deutschen Gebärdensprache wird dieses Satzglied durch ein Nomen oder ein Pronomen (ich, du, er/sie, usw.) bestimmt. Wurde das Subjekt in einer Aussage zuvor im Gebärdenraum indiziert, kann dieses durch den Kontext ebenfalls für den Kommunikationspartner erkannt werden. Eine vierte Möglichkeit das Subjekt darzustellen, ist das Nutzen von Transfer-Verben, bei denen die Bewegungsrichtung entscheidet, wer oder was als Subjekt oder als Objekt in einem Satz steht. So wird beispielsweise bei der Aussage „Ich gebe dir“ die Gebärde für „GEBEN“ ausgeführt, wobei die Bewegung vom Gebärdenden (das Subjekt) in Richtung des Gesprächspartners (das Objekt) verläuft [Pap08].

Objekt

Aus grammatikalischer Sicht ist das Objekt eine Satzergänzung, auf der sich das Prädikat bezieht. Die Handlung eines Satzes ist entweder direkt auf dieses Satzglied bezogen oder kann neben dem Subjekt an einer anderen Tätigkeit beteiligt sein. Im Gegensatz zum Subjekt muss das Objekt nicht nur aus ein Wort bestehen, es kann sich über mehrere Wörter erstrecken, was abhängig ist vom Verbgebrauch im Satz. „Objekte können ebenso wie das Subjekt in das Prädikat inkorporiert werden. Dies geschieht durch die Bewegungsrichtung, die Ausführungsstelle im Gebärdenraum oder auch die Handform des Prädikats. Dafür muss das Objekt in einer vorangegangenen Äußerung eingeführt und ggf. verortet worden sein [Pap08].“ [Pap08] unterscheidet dieses Satzglied in „Ziel-Objekt“ und „Sach-Objekt“. Beim Beispiel „Ich finde dich.“ ist „dich“ das Ziel-Objekt, auf dem sich die Aussage bezieht. In der Regel deckt dieses Objekt den Bereich der Lebewesen und Gegenstände ab. Bei einem belebten oder unbelebten Gegenstand hilft das Sach-Objekt dem Subjekt bei der Ausführung der Handlung [Pap08].

Prädikat

Das Prädikat ist das bestimmende Satzteil für die Struktur des Satzes und hat die Funktion, eine Aussage über das Subjekt zu treffen. In der Deutschen Gebärdensprache ist das Prädikat meistens eine Verbform. Es kann aber auch in der Form eines Nomens erscheinen, wobei es dann bei der Verortung von Gegenständen oder Lebewesen im Gebärdenraum in Erscheinung tritt. Ein Beispiel hierfür wäre, wenn zwei Autos nebeneinander parken, dann wird das linksstehende Auto im GR auf der linken Seite festgelegt und das rechtsstehende, rechts daneben. Weiterhin benutzt man es als Nomen, wenn Lebewesen oder Gegenstände bestimmte Eigenschaften zugewiesen bekommen, z. B. „HAUS INDEX-li VILLA“. Die Glossen-Notation sagt aus, dass im Gebärdenraum links („INDEX-li“) die Gebärde „HAUS“ gesetzt wurde, die jetzt die Eigenschaft „VILLA“ erhält. Des Weiteren können Adjektive ebenfalls in der Rolle des Prädikats auftreten (z. B. „JUNGE KRANK“). Eine weitere Besonderheit ist, dass ein Satz nur alleine aus dem Prädikat bestehen kann, ohne dass ein Objekt oder Subjekt auftaucht. Dies ist aber nur der Fall, wenn der Zusammenhang im Gespräch ersichtlich ist. Wird beispielsweise nach dem Wetter gefragt, reicht es für die Antwort aus, die Gebärde für „SCHEINEN“ aufzuführen, was für den Satz stehen kann „Die Sonne scheint“ [Pap08].

Zeitangaben

In der deutschen Lautsprache werden Verben mittels Konjugation (ich sehe, er sah usw.) an dem zeitlichen Ereignis angepasst. Anstatt in der Gebärdensprache die Verben ebenfalls zu konjugieren, nutzt man den visuellen Gebärdenraum und lässt die Verben in ihrer Ausgangsform. Der Gebärdende kann unter Verwendung der sogenannten Zeitlinie im Gebärdenraum, die Vergangenheit, Gegenwart und Zukunft ausdrücken. Diese gedachte Linie zeigt die Abbildung 17.

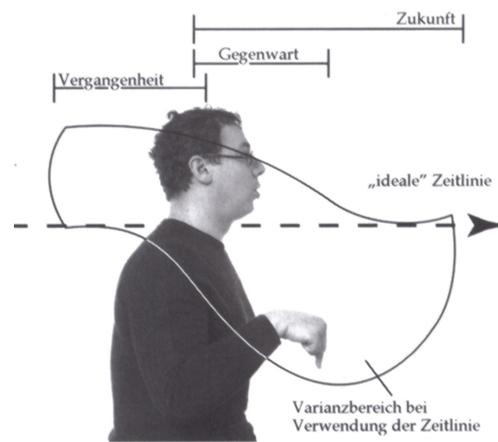


Abb. 17: Zeitlinie in der GS; aus [Pap08]

Gebärden, die mit etwas Zukünftigem zu tun haben, werden weiter vor dem Körper ausgeführt. Bei zurückliegenden Sachverhalten orientieren sich die Gebärden vom Körper ausgesehen rückwärts und bei der Gegenwart liegen diese nah vor dem Körper. Jedoch ist es auch möglich diese Zeitlinie im Gebärdenraum zu verschieben. Sollen mehrere zeitlich aufeinander folgende Ereignisse, z. B. der Tages- oder Wochenablauf, dem Gesprächspartner näher gebracht werden, kann der Gebärdende die Zeitlinie mit einer Handbewegung von der Schulter abnehmen und diese vor dem Körper legen (vgl. Bild 18 (links)). Mit dem Zeigefinger wird dann ein Referenzpunkt im Gebärdenraum festgelegt, der für die Gegenwart steht. Ereignisse, die in der Vergangenheit vorkamen, beziehen sich dann auf Punkte vor diesem Referenzpunkt, also in Richtung des Körpers. Nachfolgende oder zukünftige Ereignisse stehen nach diesem Bezugspunkt in Richtung des Gebärdenraums.



Abb. 18: links: Zeitlinie vor dem Körper; rechts: Zeitlinie quer vor dem Körper; aus [Pap08]

Eine Alternative ist es, diese Zeitlinie um 90 Grad zu drehen und quer vor dem Körper zu präsentieren (vgl. Abb. 18 (rechts)). Ebenfalls wird hier mit dem Zeigefinger der Bezugspunkt zur Gegenwart gesetzt. Geschehnisse, die zu der Vergangenheit zählen, sind aus der Sicht des Erzählers links vom Bezugspunkt und zukünftige Ereignisse rechts. Die horizontale Zeitlinie, die quer vor dem Körper liegt, wird genutzt, um Kalender-Einheiten (letzte Woche, vorher, nachher usw.), eine zeitliche Reihenfolge oder eine Dauer auszudrücken.

Eine dritte Methode, die [Boy95] schildert, ist eine Zeitlinie in vertikaler Richtung, welche von der Hüfte aus bis zum Kopf gehen kann (vgl. Darstellung 19). Sie wird in der Westschweizerischen Gebärdensprache genutzt, um z. B. Ereignisse wie das Aufwachsen eines Kindes zu schildern.



Abb. 19: Vertikale Zeitlinie; aus [Boy95]

Solange Zeitangaben nicht aus dem Kontext zu erkennen sind, müssen diese am Anfang eines Satzes stehen. Dieser gesetzte Zeitindex kann für längere Phasen des Gespräches gelten und wird erst beim Erzählen eines anderen temporalen Geschehnis neu gesetzt. Für die Vergangenheitsangaben können spezifische Zeitgebärden wie zum Beispiel „VERGANGENHEIT, GESTERN, VORGESTERN, FRÜHER“ usw. auftreten. Bei der Gegenwart gibt es die Gebärden für „JETZT“ oder „HEUTE“ und für die Zukunft lassen sich die Gebärden „ZUKUNFT, MORGEN, ÜBERMORGEN, DEMNÄCHST“ usw. nutzen [Boy95], [Blu06], [Pap08].

2.3.2 Satzbausteine bzw. Satzbestandteile

Die Begriffe Subjekt, Objekt und Prädikat können in der Linguistik auch als Satzbausteine betitelt sein. In dieser Arbeit wird dieser Begriff für die Wortarten Nomen, Verben, Adjektive verwendet. Neben diesen drei Satzbausteinen, die nach meiner Ansicht am Wichtigsten sind, da diese sehr häufig im praktischen Gebärdengebrauch Anwendung finden, gibt es natürlich einige andere, wie Adverbien, Modalverben, Pronomen, Attribute usw. Für weitere Informationen über diese Wortarten können die Quellen [Erl00], [Leu04] und [Pap08] nachgelesen werden.

Nomen

Mit Hilfe des Nomens können Lebewesen, Gegenstände, Zustände oder Vorgänge ausgedrückt werden. Grob lassen sich diese untergliedern in Nomen für „Eigennamen“, welche bei Namen, Orte oder Markennamen Anwendung finden und für „Sachbezeichnungen“, zu denen Gegenstände, Lebewesen und Vorgänge gezählt werden. In der Gebärdensprache sind Nomen häufig Wörter, die für sich stehen und dementsprechend mit einer Gebärde vollzogen werden. Häufig begleitet eine Mundartikulation das Nomen. Wörter, die aus mehreren Nomen bestehen, z. B. „Fußballmannschaft“, wurden durch den Einfluss der deutschen Lautsprache

allmählich mit in die DGS aufgenommen. Bei solchen zusammengesetzten Wörtern, gebärdet man die einzelnen Wörter separat, in diesem Fall „FUßBALL“ und „MANNSCHAFT“. Bei der Mehrzahl wird das Nomen an einer leicht versetzten Ausführungsstelle wiederholt. Eine weitere Möglichkeit ist das Nomen im Singular darzustellen und davor eine Mengenangabe wie z. B. „FÜNF AUTO“ zu gebärden. Seltener genutzt, kann auch das Nomen mehrmals im Gebärdenraum indiziert sein und so die Anzahl durch das Anzeigen mit dem Finger bestimmt werden [Pap08].

Verben

Die Aufgabe des Verbs im Satz ist eine Tätigkeit, ein Geschehen oder ein Zustand auszudrücken. Die Einteilung der Verben in der Gebärdensprache fällt in der Literatur sehr vielfältig aus. „Trotz unterschiedlicher Forschungsansätze, hat sich die Verbklasseneinteilung von Padden (1988) weitläufig etabliert. Sie untersuchte die Verben der ASL gemäß ihrem morphosyntaktischen Verhalten. Laut Keller sind diese Erkenntnisse ohne weiteres auf die DGS übertragbar, da sich Verben sowohl in der ASL als auch in der DGS gleichsam verhalten. Paddens Analyse wurde von vielen Forschern übernommen und teilweise auch erweitert [Sch09].“ Somit lassen sich Verben grob in „nichtflektierende einfache Verben“, „flektierende Orts- und Bewegungsverben (Raumverben)“ und „flektierende Richtungsverben“ einteilen [Sch09]. [Dot01] klassifiziert sie in „einfache Verben“, „Lokationsverben“, die den Raumverben entsprechen und „Kongruenz-Verben“ (Richtungsverben). [Boy95] unterteilt die Verben in die Klassen „Raumverben“ und „Übereinstimmungsverben“ (Kongruenz-Verben), die auch umgekehrt auftauchen können, bezeichnet als „umgekehrtes Übereinstimmungsverb“. Ebenfalls werden bei [Pap08] die Verben in den zwei Klassen „*einfache Verben*“ und „*komplexe Verben*“ grob unterteilt. Da diese Quelle die aktuellste von den genutzten Quellen im Bereich DGS-Grammatik ist und von gehörlosen Fachleuten über ein dreijähriges Projekt konzipiert wurde, werden diese Kenntnisse in dieser Arbeit als Grundlage genommen.

Die einfachen Verben können nicht flektiert werden. Sie lassen sich unterteilen in „*körpergebundene Verben*“ und „*nicht-körpergebundene Verben*“. Wie die Bezeichnung schon vermuten lässt, zählen zu den körpergebundenen Verben Verbgebärden, deren Ausführungsstelle in Nähe des Körpers liegen (z. B. bei „SCHLAFEN, KENNEN“), wobei hingegen zu den nicht-körpergebundenen Verben Gebärden zählen, deren Ausführungsstelle im Gebärdenraum (z. B. bei „STERBEN, KOCHEN“) – meistens im Bereich der neutralen Stelle – auszuführen sind. Durch eine wiederholende Darstellung der Gebärde an einer leicht versetzten Ausführungsstelle können einfache Verben in den Plural überführt werden [Pap08].

Die komplexen Verben lassen sich in einem Satz flektieren und unterteilen sich in „*Substitutor-Verben*“, „*Manipulator-Verben*“ und „*Transfer-Verben*“.

Substitutor-Verben unterstützen eine Aussage über einen Gegenstand oder einem Lebewesen, in dem die Hand als Stellvertreter die Bewegung oder Form des Gegenstandes bzw. des

Lebewesens nachstellt. So kann z. B. die Art und Weise einer Fortbewegung einer Person gebärdet werden. Die Abbildung 20 zeigt die Gebärde „GEHEN“, bei der die Hände die Füße des Menschen darstellen und sich dementsprechend bewegen.



Abb. 20: Substitutor-Verb „Gehen“; aus [Pap08]

In der Regel wird erst das Nomen gebärdet, z. B. Flugzeug. Anschließend kann eine Hand die Flugbewegung des Flugzeuges andeuten, wobei die andere Hand vorbeiziehende Objekte, wie z. B. Wolken darstellt. Ein weiteres Anwendungsgebiet für Substitutor-Verben ist die Bildung des Plurals. Die bildhafte Darstellung mit den Händen erlaubt es, den Gegenstand oder die Personen mehrmals aufzuführen. So kann die linke Hand und die rechte Hand jeweils einen Läufer abbilden oder ein Apfel wird mehrmals leicht versetzt im Gebärdenraum wiederholt skizziert.

Bei den Manipulator-Verben wird die Hand genutzt, um darzustellen wie Gegenstände gehalten oder benutzt werden. So kann alleine schon die Handform anzeigen wie groß (vgl. Abb. 21) oder welche Form ein Gegenstand hat, vorausgesetzt dieser wurde vorher schon gebärdet. Die „F“-Handform deutet kleine, dünne Dinge an (z. B. eine Nadel oder ein Blatt Papier), die „C“-Handform runde Objekte (z. B. ein Rohr), die gekrümmte Flachhand größere rundliche Gegenstände (z. B. eine Vase) und die letzte Handform könnte für die Dicke eines Buches stehen.



Abb. 21: Manipulator-Verben: Handformbeispiele; aus [Pap08]

Anwendung finden diese Art von komplexen Verben bei der Rollenübernahme, wo der Gebärdende eine Tätigkeit nachspielt, z. B. das Essen eines Apfels. Andere Verben können „HALTEN, ZIEHEN, GEBEN“ usw. sein [Pap08].

Zusammenfassend verkörpern die Substitutor-Verben einen/ein Gegenstand/Lebewesen und die Manipulator-Verben bilden die Handhabung eines Gegenstandes ab.

Die Transfer-Verben (bei [Dot01] die Kongruenzverben, [Sch09] die Richtungsverben und [Boy95] die Übereinstimmungsverben) bezeichnen den Übergang eines Objektes von einem Startpunkt zu einem Zielpunkt. Bei einem Objekt kann es sich um eine zwischenmenschliche Botschaft (Verben: fragen, anrufen) oder einen Gegenstand (Verben: senden, abholen) handeln. Ein Beispiel wäre: „Ich frage dich“. Das Transfer-Verb „FRAGEN“ wird vom Gebärdenden in Richtung des Gesprächspartners vollzogen, also vom Subjekt zum Objekt. Stimmt der Anfangspunkt nicht mit dem Subjekt, sondern mit dem Objekt überein, bezeichnet man dies als „umgekehrtes Übereinstimmungsverb“, z. B. bei „Ich lade ihn ein.“, wo die Gebärde „EINLADEN“ von dem Gast zum Gastgeber sich bewegt [Boy95]. Sollen mehrere Gäste eingeladen werden, wird die Mehrzahl durch eine zickzack-Bewegung oder einer Bogen-Bewegung angezeigt. Transfer-Verben können auch unter Nutzung des Gebärdensraums Anwendung finden. Hierbei erfolgt die Indizierung des Subjektes und Objektes im Raum und können mit dieser Verbart in Verbindung gesetzt werden [Pap08]. Transfer-Verben können sich noch weiter unterteilen lassen. Für weitere Details zu den Verbklassen sei an dieser Stelle auf [Pap08] verwiesen. Für eine sehr linguistische Sichtweise empfiehlt sich [Hän04].

Adjektive

Die Hauptfunktion der Adjektive besteht darin, Lebewesen, Gegenständen, Handlungen oder Zuständen eine Eigenschaft oder bestimmte Merkmale zuzuschreiben. Adjektive werden auch Eigenschaftswörter genannt. Einige Beispiele sind Farben (blau, grün, usw.), Maßangaben (groß, klein, usw.) oder auch Zustände (heiß, kalt, usw.). Sie lassen sich steigern (z. B. kleiner-kleiner-am kleinsten) und flektieren. In der Gebärdensprache steht das Adjektiv überwiegend nach dem Nomen (z. B. „MANN GEHÖRLOS“). In andern Fällen kann es aber auch davor stehen, wie zum Beispiel bei „HÜBSCHE FRAU“. Werden mehrere Adjektive genutzt, dann müssen sie hinter den Nomen gebärdet werden: „ICH MÖGEN KAFFEE HEISS STARK SÜSS“.

[Pap08] unterteilt die Adjektive in „Einfache“, „Skizze“ und „Maß“-Adjektive. „Unter einfachen Adjektiven verstehen wir konventionelle Gebärden, die eine feste Grundform haben, z. B. die Gebärde „GESUND“ [Pap08].“ Sie stehen entweder vor dem Nomen oder danach. Zu den Skizze-Adjektiven zählen Gebärden, die eine hohe visuelle Ausdrucksform haben. Die Hände können dabei die Form durch nachzeichnen des Umrisses eines Gegenstandes oder die Art der Bewegung (z. B. zickzackförmig) einer Person oder Objektes aufführen. Diese Art der Adjektive stehen hinter dem Nomen. Maß-Adjektive finden Anwendung bei Angaben über die Größe eines Gegenstandes bzw. einer Person. Der Maßstab für die Größenangabe wird durch die Perspektive des Sprechers bestimmt.

Die Steigerung bzw. stufenweise Intensivierung von Adjektiven lässt sich durch Vergrößerung, Beschleunigung oder Verlangsamung der Gebärde darstellen. Die Mimik unterstützt die Steigerung (vgl. Abbildung 22), indem sie von minimal bis überspitzt gestikuliert wird [Pap08].



Abb. 22: Steigerung des Adjektivs „GUT“; aus [Pap08]

2.3.3 Satztypen

In diesem Abschnitt werden einige Möglichkeiten von Sätzen dargestellt, die der Gebärdende nutzen kann, um eine Aussage, Frage oder Aufforderung zu tätigen. Die Grafik 23 zeigt ein Überblick von Satztypen, welche nachfolgend im einzelnen erläutert werden. Am Ende des Abschnittes folgt eine Aufzählung von Satzverbindungen, die es erlauben, verschachtelte Sätze in der Gebärdensprache zu bilden.

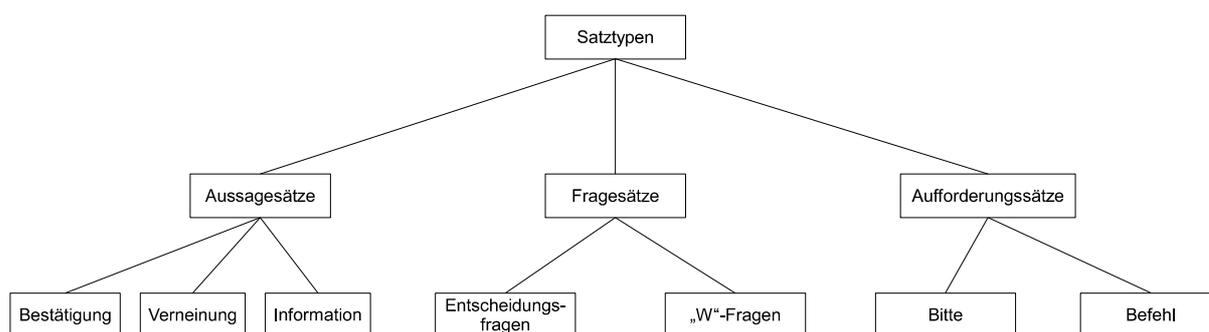


Abb. 23: Ausgewählte Satztypen der DGS

Aussagesätze

In einer Sprache treten am Häufigsten die Aussagesätze auf. Unter Verwendung dieses Satztyps kann der Sprecher die wesentliche Information, z. B. eine Behauptung, eine eigene Meinung oder einen Sachverhalt in einer Aussage formulieren. In der Gebärdensprache funktioniert das durch das Anzeigen einer neutralen Mimik. Eine Aussage kann aber auch in einer Bestätigung oder Verneinung auftreten, wobei hier die Mimik nicht neutral ausfällt. Bei der Zustimmung kann man eine „JA“-Gebärde mit einem unterstützenden positiven Blick tätigen oder die Aussage des Kommunikationspartners wird wiederholt und zusätzlich mit einem Kopfnicken bestätigt.

Analog zur Bestätigung gibt es bei der Verneinung eine Gebärde für „NEIN“, die durch ein seitliches Kopfschütteln begleitet wird. Weitere Möglichkeiten zur Verneinung können durch alleiniges seitliches Kopfschütteln, Stirnrunzeln, oft mit zusätzlichem Nasenrumpfen, dem Gesprächspartner mitgeteilt werden [Boy95], [Pap08], [Dot12].

Fragesätze

Ein weiterer häufig verwendeter Satztyp in einer Sprache sind die Fragen. Grob lassen sich diese in „Entscheidungsfragen“ und den „W-Fragen“ einteilen. Bei einer Entscheidungsfrage kann der Gesprächspartner diese Frage mit Ja oder Nein beantworten. Sie wird in der Gebärdensprache mit einem leicht nach vorn gebeugten Kopf sowie einem großen offenen Augenausdruck und mit dem Anheben der Augenbrauen geäußert. Die Abbildung 24 zeigt, wie aus der Gebärde „GEHÖRLOS“ eine Entscheidungsfrage „Bist du gehörlos?“ formuliert wird.

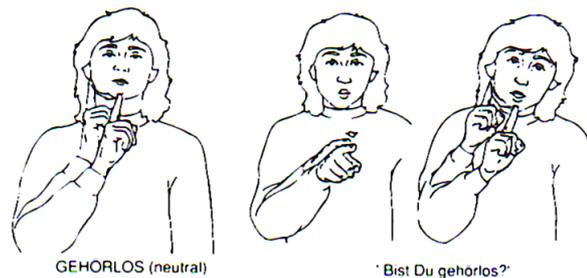


Abb. 24: Entscheidungsfrage in der DGS; aus [Boy96]

Zu den „W-Fragen“ lassen sich die Fragewörter: wie, wo, wer, was, wann, warum usw. zuordnen und erwarten keine Ja/Nein, sondern eine ausführliche informative Antwort. [Pap08] beschreibt die Mimik als verschlossene Fragemimik, bei der der Kopf leicht interessiert vorgestreckt ist und die Augenbrauen leicht skeptisch gerunzelt wirken. [Boy95] hingegen schildert den ähnlichen Blick mit den zusammenziehenden Augenbrauen in Kombination mit einer nach hinten geneigten Kopfhaltung. Jedoch ist aus dem Kontext nicht erkenntlich, ob diese Kopfhaltung sich generell auf Gebärdensprachen bezieht oder ob diese speziell für die DGS gedacht ist [Boy95], [Sie07], [Pap08].

Aufforderungssätze

Mit Hilfe von Aufforderungssätzen können Befehle und Bitten ausgedrückt werden. Bei einer freundlichen Aufforderung nutzt der Sprecher die Gebärde „BITTE“, in der der Erzähler mit dem Handrücken an seine Wange vorbei streicht (vgl. Abbildung 25). Je nach Situation oder Gesprächspartner kann diese Bitte eher weich oder verstärkt ausfallen. Soll ein Kind sein Zimmer aufräumen, kann die Mutter die Mimik zur Gebärde etwas streng und auffordernd halten. Fragt die Mutter das Kind nach Hilfe für den Haushalt, fällt die Emotion und Gesichtsmimik etwas weicher, liebevoller, unterwürfiger aus. Bei sachlichen Dingen, z. B. wenn ein Kollege nach einen Kugelschreiber bittet, kann die Mimik auch neutral ausfallen. Die Gebärde „BITTE“ kann auch bei einem Befehl auftreten, wobei diese etwas energischer als bei der freundliche Bitte ausfällt. Wird die Anweisung bzw. die Aufforderung ohne Bitte geäußert, stellt der Gebärdende diese mit einer sachlichen, energischen, leicht strengen Mimik dar [Pap08].



Abb. 25: von links nach rechts: Bitte-, Aufforderungs-, Befehls-Mimik; aus [Pap08]

Satzverbindungen

In der deutschen Lautsprache gibt es Hauptsätze und Nebensätze, um mehrere Informationen in einer Aussage zu vermitteln. Diese Satzteile können mit den verschiedensten Satzverbindungen, wie zum Beispiel „und“, „oder“, „aber“, „dann“ usw. miteinander verbunden werden. Dazu nutzt die Deutsche Gebärdensprache nicht nur manuelle Ausdrucksmittel, sondern auch die nicht-manuellen Ausdrucksmittel. So kann alleine schon eine leichte Veränderung in der Oberkörperhaltung, eine seitliche Neigung des Kopfes oder ein Wechsel in der Blickrichtung, ein Anzeichen sein, dass der Sprecher gerade zwei Sachverhalte gegenüberstellt. Ein Beispiel sind die Konditionalsätze, bei denen der erste Satzteil eine Bedingung ausdrückt – in der deutschen Lautsprache meistens mit „falls“ oder „wenn“ eingeleitet – und der zweite die Konsequenz dieser Bedingung („dann“) ist. Die Einleitung des Satzes wird mit einer leicht vorgestreckten Kopfhaltung und erhobenen Augenbrauen durchgeführt. Nach einer kurzen Pause folgt dann die Aussage mit der Konsequenz, bei der die Kopfhaltung sowie eine Änderung des Oberkörpers je nach Situation und Aussage abgeändert wird.

Um zwei gleichrangige Satzteile zu verknüpfen, benutzt man in der Lautsprache das Wort „und“. Dieses Wort existiert nicht in der Gebärdensprache. Der Gebärdende nutzt auch hier den Gebärdenraum aus, indem er zwei Sachverhalte im Raum an unterschiedlichen Orten indiziert und beim Wiedergeben zwischen diesen, entweder eine kurze Pause einlegt oder den Blick bzw. den Oberkörper auf diesen Index dreht. Zum Beispiel bei der Aussage „Papa schaut fern (links im GR indiziert) und Mutti wäscht ab“ (rechts im GR indiziert).

Ebenfalls kann diese Methode beim Wort „oder“ genutzt werden, wobei hier der Kopf seitlich geneigt ist und die Augenbrauen angehoben sind. Alternativ gibt es auch ein Gebärdenzeichen für diese Satzverbindung.

Ein „aber“ lässt sich entweder durch eine manuelle Gebärde darstellen oder mittels nicht-manuellen Komponenten, bei der die Mimik eine überraschende Emotion anzeigt meist verknüpft mit weit geöffneten Augen. Das letzte Beispiel an dieser Stelle ist das Satzverbindungsglied „dann“, für das es ebenso ein Gebärdenzeichen gibt, welches in Kombination mit dem Mundbild „und dann“ geäußert wird [Pap08].

2.3.4 Zusammenfassung zur Grammatik der DGS

Ähnlich wie beim Thema Phonologie gibt es auf der nächsten Seite eine zusammenfassende Grafik (vgl. Grafik 26) der gerade behandelten grammatischen Ansichten. Diese Abbildung soll nicht den kompletten Bereich der DGS abdecken, sondern fasst nur die behandelten Punkte zusammen, die in diesem Kapitel betrachtet wurden. Eine ausführlichere Darstellung der DGS-Grammatik ist den Sprachwissenschaftlern vorbehalten.

Der Zweck der Grafik ist es, zum Beispiel Personen/Laien, die vielleicht aus dem Animationsbereich kommen, einen Ablauf der Satzglieder näher zu bringen. So könnte ein Computer-Animator bei dem Satz der deutschen Lautsprache „Ich kaufe ein Haus“ mit Hilfe der Grafik erkennen, dass in der Deutschen Gebärdensprache der Artikel „ein“ wegfällt und die Abfolge der Gebärden für ein Avatar: „ICH“, „HAUS“ und „KAUFEN“ ist. Sie soll somit ein Verständnis dafür schaffen, dass die GS eine andere Grammatik als die der deutschen Lautsprache aufweist.

Die oberste Ebene zeigt die Position der Satzglieder in einem Satz an. Eine Zeitangabe muss am Satzanfang gesetzt werden (siehe Pfeil von oben). Nachfolgende Sätze können dann beim Subjekt oder Prädikat beginnen, welches ebenfalls durch den Pfeil von oben symbolisiert wurde. Dies hängt vom Kontext des Gespräches ab. Bei einem neuen zeitlichen Geschehen muss der Satz wieder mit der Zeitangabe beginnen.

Die Gebärde für „FERTIG“ zeigt dem Gesprächspartner an, dass der Gebärdende seine Ausführungen beendet hat. Diese Gebärde kann optional gesetzt werden, da auch eine längere Pause oder ein nicht-manuelles Verhalten (z. B. Augenkontakt) in einer Konversation den Abschluss einer Handlung anzeigen kann.

Die darunter liegenden Ebenen legen keine feste Reihenfolge fest, sondern stellen nur optionale Möglichkeiten dar, aus welchen Komponenten das überliegende Satzglied/Satzbaustein bestehen kann. Für detailliertes Wissen können dann die zugehörigen Abschnitte in der Arbeit gelesen werden.

Die blauen Kästchen beschreiben die Möglichkeit der Pluralbildung. Bei dem Kästchen „Zahlen- oder Mengenangaben“ bedeutet „Zahlenangabe“ das der Gebärdende eine Ziffer vor dem Nomen gebärdet, z. B. „FÜNF AUTOS“. Der Begriff „Mengenangabe“ bedeutet, dass der Sprecher 5 Autos im Gebärdenraum festgelegt hat und die Mehrzahl mit den daraufzeigenden Finger darstellt.

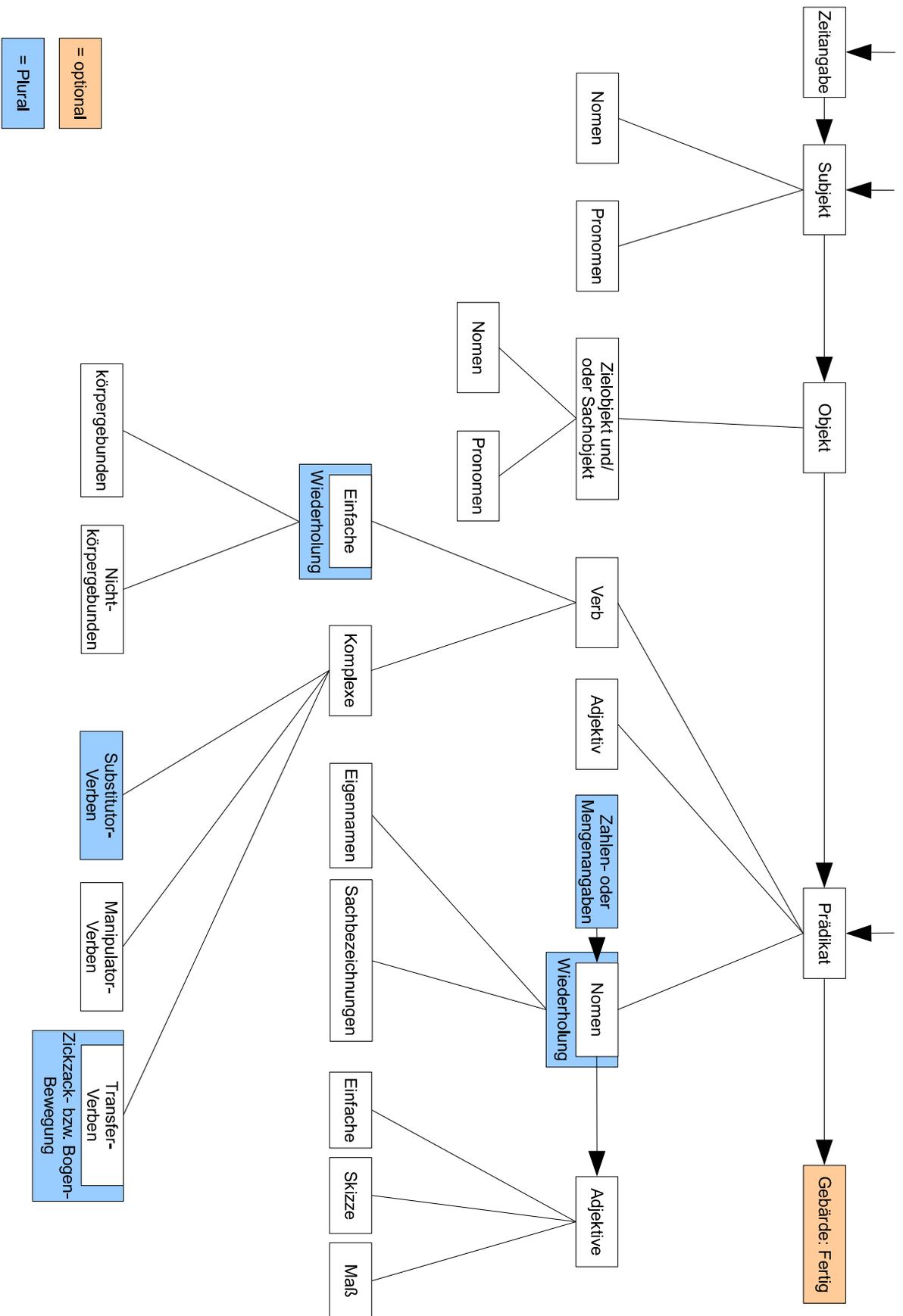


Abb. 26: Übersicht DGS-Grammatik

2.4 Anforderung an die Animation des Avatars

Nach der ausführlichen Betrachtung der Deutschen Gebärdensprache in diesem Kapitel lassen sich die Anforderungen für eine Animation eines Gebärdensprachavatars ableiten. Die gewonnenen Kenntnisse sind in den Aufzählungen „manuelle Komponenten“, „nicht-manuelle Komponenten“ und „Generelles“ enthalten.

Manuelle Komponenten

Handform:

- Separate Fingerbewegung und Daumenbewegung
- Fingergelenk/-glied Bewegung/Krümmung
- Spreiz- und Schließbewegung der Finger
- Schließen und Öffnen der Hand
- *Krümmung der Handfläche*

Handstellung:

- Handgelenk mit (*unterstützender Unterarmbewegung*) 360° Drehung

Ausführungsstellen (Körper/Gebärdenraum):

- Kopf, Gesicht, Rumpf, Arm, Hand in frontaler, waagerechter und senkrechter Ebene bewegbar

Bewegungen:

- Einhandgebärden
- Zweihandgebärden (Regel der Symmetrie und Dominanz)
- Geschwindigkeitsregulierung der Ausführungen
- Regulierung des Bewegungsumfanges (z. B. „flüstern“, „zurufen“)
- *Wiederholungsmöglichkeit der Gebärde*
- *Wiederholungsmöglichkeit der Aussage bei Problemen der Verständlichkeit*
- *Bewegungsspur sichtbar*
- *Inkorporation*

Nicht-manuelle Komponenten

Mimik:

- Bewegung der Augenbrauen
- Naserümpfen bzw. Bewegung der Nasenflügel
- Stirnrunzeln
- Kinnrunzeln
- Wangenbewegungen (z. B. aufblasen, pusten)

Blickrichtung:

- Pupillenbewegung/Richtungssehen
- Regulierung der Augenöffnung und -schließung
- *Augenlidbewegung (z. B. natürliches Augenzwinkern)*

Kopfhaltung:

- vertikale Kopfbewegung (z. B. heben, senken)
- horizontale Kopfbewegung (z. B. neigen, drehen)
- frontale Kopfbewegung (z. B. vorstrecken, zurückziehen)

Oberkörperhaltung:

- vertikale Körperbewegung (z. B. strecken, krümmen)
- horizontale Körperbewegung (z. B. neigen, drehen)
- frontale Körperbewegung (z. B. vorbeugen, zurückbeugen)
- *Atembewegung vom Brustkorb*

Mundartikulation:

- Lippenbewegung
- Mundbewegung (z. B. öffnen, schließen)
- Darstellung der Zähne und Zunge

Generelles

- Kollisionserkennung
- flüssige und weiche Übergänge zwischen den Animationen
- *Möglichkeiten zur Kamerapositionierung bzw. Blickwinkelwahl*
- *natürliche Bewegungen/Animationen*
- *Anpassungen an Umgebung und Aussehen des Avatars*
- *Auswahl zwischen verschiedenen Avatar-Modellen*
- *Auswahl zwischen verschiedenen Hintergrundumgebungen*

Für eine korrekte Ausführung eines Gebärdenzeichens bzw. einer Informationsaussage muss der Avatar die nicht kursiv geschriebenen Punkte beherrschen. Anforderungen, die kursiv hervorgehoben sind, sollen als zusätzliche Optionen angesehen werden. So können diese den Computer-Animator bei der Erleichterung der Arbeit helfen („Wiederholungsmöglichkeit der Gebärde“ für Mehrzahlbildung), die Akzeptanz der animierten Figur durch ein natürliches Verhalten steigern („Krümmung der Handfläche“, „unterstützende Unterarmbewegung“, „Augenlidbewegung“, „Atembewegung vom Brustkorb“, „natürliche Bewegungen/ Animationen“) oder eine Gebärde in ihrer Aussage bekräftigen („Bewegungsspur sichtbar“, „Inkorporation“, „Möglichkeiten zur Kamerapositionierung bzw. Blickwinkelwahl“).

Der Begriff „Kollisionserkennung“ oder „Kollisionskontrolle“ bedeutet, dass Körperteile oder Segmente des Avatars nicht ineinander verlaufen können (z. B. die Finger an der Hand dürfen nicht bei der Ausführung einer Handform durch die Handfläche verlaufen). Sind Gebärden nicht auf den ersten Blick verständlich, könnte eine Bedienungsoption die Kameraperspektive verändern. Die manuelle Auswahl einer Kameraperspektive in der Animation hätte den Vorteil, dass der Nutzer schwer verständliche Handformen oder andere Details der Ausführung des Gebärdensprachavatars genauer betrachten könnte. Ebenso kann durch manuelles Eingreifen des Benutzers in die Geschwindigkeitsdarstellung der Gebärden das gewünschte Tempo eingestellt werden, was die Verständlichkeit der Ausführungen des Avatars steigern könnte. Eine weitere Informationshilfedarstellung ist die Option der sichtbaren Bewegungsspur. So kann eine Art Schliereneffekt, welcher zum Beispiel bei Substitutor-Verben Anwendung finden könnte, den Bewegungsverlauf der Gebärde sichtbar darstellen und somit die Aussage optisch unterstützen.

Inkorporation erlaubt den Gebärdenden, Informationen schneller zu vermitteln, indem zwei Gebärden ineinander geführt werden. So kann die Aussage „fünf Autos“ mit zwei Gebärdenzeichen erfolgen oder die Zahl „fünf“ wird mit in der Gebärde „AUTO“ dargestellt. Für den Informationsgehalt ändert sich hierbei nichts, dass somit der Punkt als Option anzusehen ist. Für die Anerkennung bzw. Sympathie der animierten Figur würde ein natürlicher Bewegungsablauf nützlich sein. So sollte sich zum Beispiel bei einer Handgelenksdrehung der Unterarm bzw. die Oberflächentextur des Avatars natürlich mitbewegen. Da ein sympathisches ansprechendes Aussehen in einer Unterhaltung für die Akzeptanz des Gesprächspartners von Bedeutung sein kann, sollte die Szene bzw. Umgebung und der Avatar Gestaltungsmöglichkeiten erlauben. So könnte zum Beispiel der Nutzer entscheiden, in welcher Räumlichkeit die Animation stattfindet oder ob der Avatar weiblich oder männlich ist. Denkbar ist auch das Aussehen (z. B. Kleidung, Größe, Hautfarbe, Frisur, usw.) der animierten Figur abzuändern. Alternativ könnten verschiedene Umgebungen oder Avatar-Modelle zur Auswahl stehen. Diese Punkte sind in erster Linie nur als Option anzusehen, können aber die Akzeptanz und somit indirekt die Verständlichkeit des Avatars steigern.

Eine Hauptanforderung an den Avatar ist, dass sich sein Bewegungsablauf stark an den Bewegungen des Menschen orientiert. So sollten im wesentlichen die Hände und das Gesicht detailliert modelliert werden. Je nach Darstellung des Avatars auf dem Bildschirm können die Beine modelliert sein, besitzen aber für die Aussagen in der Gebärdensprache keine Funktion und müssen somit nicht animiert werden. Außerdem sollte die Darstellung der Gebärden möglichst kontinuierlich und nicht stockend oder ruckartig ablaufen.

Werden alle Punkte beachtet, kann ein Avatar Aussagen in der Gebärdensprache in einer guten Qualität dem Kommunikationspartner vermitteln. Nachdem die Anforderungen eines Gebärdensprachavatars bekannt sind, muss für die Übersetzung einer Lautsprache in die Gebärdensprache eine schriftliche Form der Gebärdensprache gefunden werden. Das nächste Kapitel stellt Notationssysteme vor, die eine bildliche Sprache wie die Gebärdensprache in eine textuelle Form abbildet.

3. Kapitel: Notationssysteme der Gebärdensprache

Ein Notationssystem ist ein formales System, das eine Sprache in ihrer komplexen Struktur schriftlich abbildet. Um dieses System nutzen zu können, muss der Anwender dessen Codierungsvorschrift kennen. In der deutschen Lautsprache benutzt man das Schriftsystem, das als Codierungsform das lateinische Alphabet nutzt. Eine weitere Codierungsmöglichkeit wäre die Lautschrift.

In der Gebärdensprache gibt es nach aktuellem Stand kein einheitliches Notationssystem. So gibt es Transkriptionen (= schriftliche Verfassung von Sprache bzw. Gebärden), die für die Erstellung von Wörterbüchern geeignet sind, die für die Kommunikation zwischen Gehörlosen genutzt werden oder die zur Entwicklung und Ansteuerung von Gebärdensprachavataren Anwendung finden. In der folgenden Abbildung 27 sind ausgewählte Notationssysteme zu den Bereichen „*menschliche Kommunikation*“, „*linguistische Notation*“ und „*animationsorientierte Notation*“ aufgeführt.

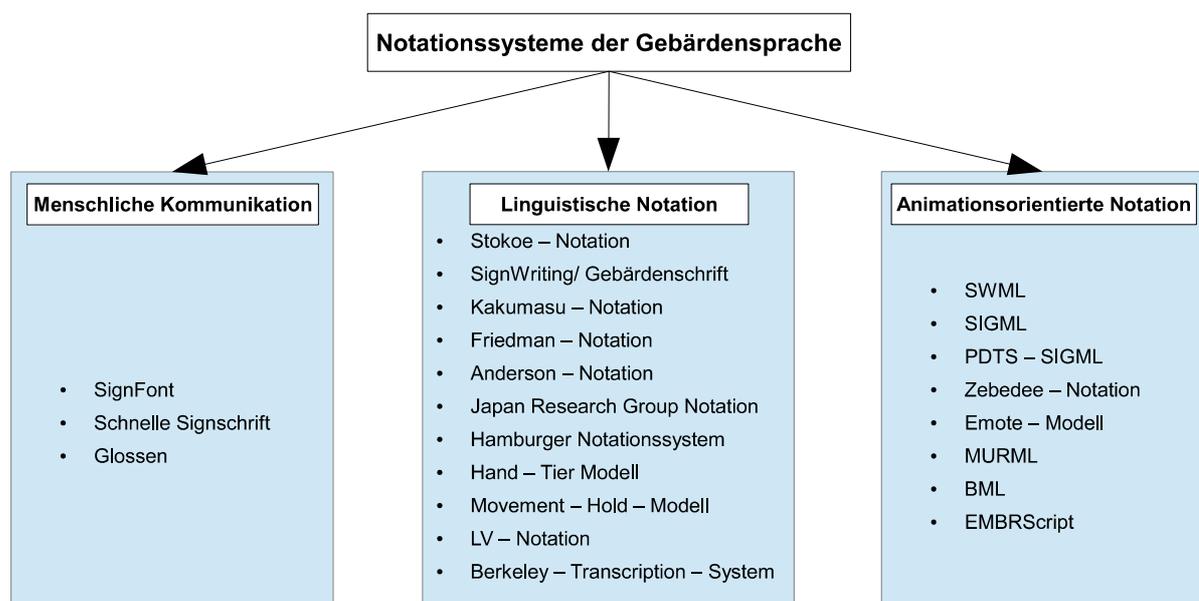


Abb. 27: Übersicht Notationssysteme für die Gebärdensprachen

Zu den Bereich der menschlichen Kommunikation zählen Verschriftungssysteme, die Gehörlose erlauben, miteinander in schriftlicher Form zu kommunizieren. Diese Schreibsysteme bilden nicht alle Details einer Gebärdensprache ab. So stehen sie lediglich für die Semantik der Aussage und zeigen nicht aus welchen manuellen und nicht-manuellen Komponenten eine Gebärde besteht. Somit lassen sie sich für eine unkomplizierte und schnelle Kommunikation verwenden (z. B. kurze Notizen, Merkhilfen für Vokabeln usw.).

Die linguistischen Notationen haben die Zielsetzung eine Gebärdensprache ausführlich im Detail widerzuspiegeln. Sie beschreiben die manuellen und nicht-manuellen Komponenten einer Gebärde und versuchen die grammatikalischen Strukturen der GS zu erfassen. Ihr Nutzen liegt momentan in der Sprachforschung, soll aber im günstigsten Fall zu einem der ersten allgemein anerkannten Schriftsysteme für Gebärdensprachen, die die Gehörlosengemeinschaft auch zur Kommunikation verwendet, etabliert werden. So gibt es erste Ansätze von linguistischen Notationen in Form von Wörterbüchern, um die Akzeptanz für diese Schreibsysteme unter den Gehörlosen zu verbreitern und zu steigern.

Animationsorientierte Notationen haben den Zweck, die Sprachbarriere zwischen Gehörlose und Hörende zu verringern sowie der Gehörlosengemeinschaft Informationen aus den modernen medialen Techniken wie TV, Internet, Smartphone usw. vereinfacht bereit zu stellen. Sie basieren teilweise auf den Erkenntnissen der linguistischen Notationen und ermöglichen diese Parameter zur Ansteuerung von Avataren, die Gebärdensprache ausüben sollen, zu nutzen.

Eine ausführlichere Darstellung dieser verschiedenen Notationssysteme ist im Anhang A2 zu finden, die kurz die Merkmale und ein Beispiel für jede Notation präsentiert. Im folgenden Verlauf des Kapitels wird die Glossen-Notation aus dem Bereich der menschlichen Kommunikation sowie die Stokoe-Notation und das Hamburger Notationssystem als Vertreter der linguistischen Notationen genauer erläutert. Die SIGML-Notation, die auf dem Hamburger Notationssystem basiert, dient als Übergang zu den animationsorientierten Notationen und wird neben BML und EMBRScript detaillierter beschrieben.

3.1 Glossen-Notation für menschliche Kommunikation

Eine der häufigsten verwendeten Notation stellt die Glossen-Notation dar. Es gibt keine einheitliche Codierungsvorschrift. So findet man in Büchern eine Vielzahl von verschiedenen Glossenumschriften. Lediglich gleich ist, dass Glossen aus lateinischen Buchstaben bestehen und die daraus resultierenden Wörter vollständig groß geschrieben werden. Sie erfüllt den Zweck, eine Aussage hinsichtlich ihrer Gebärden in die inhaltlich grammatikalisch richtige Reihenfolge schriftlich darzustellen. Die Wörter sind an die jeweilige Lautsprache des Landes angelehnt und stellen keine Übersetzung der Gebärde in Lautsprache dar, sondern geben ausschließlich die Bedeutung der Gebärde wieder. So gibt es nicht für jedes Wort aus der Lautsprache eine zugehörige Glosse. Sie dienen als eine Gedankenhilfe, mit denen sich die Abfolge von Einzelgebärden notieren lässt. Glossen können zusätzliche Zeichen/Symbole haben, die Gesichtsausdrücke, Emotionen und grammatikalische Funktionen, wie das Anzeigen von Indizierungen, Rollenübernahmen, Wiederholungen, Adjektive/Adverbien, Transfer-Verben usw. abbildet. Auf eine umfassende Beschreibung der manuellen und nicht-manuellen Komponenten wurde jedoch verzichtet, da sie für eine schnelle einfache Gebärden-Glossie-

ung gedacht ist, die auch zur Beschreibung von Gebärden-Beispielen in Wörterbüchern bzw. Büchern Anwendung findet [Pap08].

Beispiel: *Glossen-Notation*

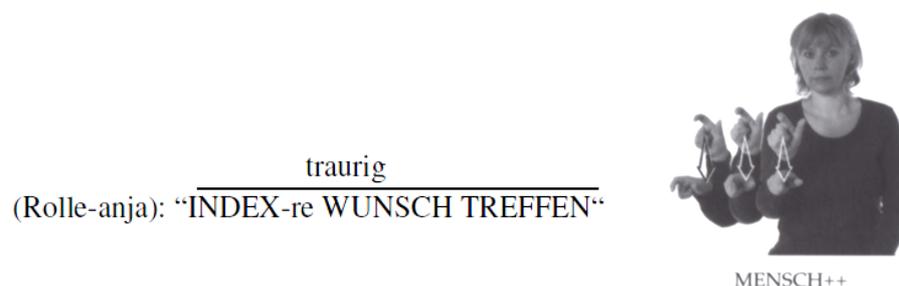


Abb. 28: links: *Glossen-Notation*; aus [Wil10]; rechts: *Glossen-Notation*; aus [Pap08]

Das linke Beispiel in der Abbildung 28 steht für die Aussage „Möchtest du dich mit mir treffen?“, fragte meine Freundin Anja traurig.“. Der Index sagt aus, dass der Gebärdende die Person Anja vorher rechts im Gebärdenraum indizierte und bei der direkten Aussage in ihre Rolle schlüpft und somit ihre Emotion in dem Fall „traurig“ wiedergibt. Eine andere Form der Glossenverschriftung zeigt das rechte Bild. Die beiden Pluszeichen in der Glosse sagen aus, dass die Gebärde MENSCH nach dem ersten Ausführen zweimal wiederholt wird mit der Bedeutung, dass der Gebärdende den Plural (Menschen) gebärdet.

Alleine aus der Glossen-Notation kann der Leser nicht erkennen, wie die speziellen Parameter einer unbekanntes Gebärde ablaufen. Somit kann der Inhalt der Glosse verstanden werden, aber diese Gebärde selber auszuführen, wäre nicht möglich.

Vor- und Nachteile der Glossen-Notation zeigt die nachstehende Tabelle 2.

Vorteile	Nachteile
<ul style="list-style-type: none"> + angelehnt an die Lautsprache des Landes + einfache Codierungsvorschrift + geeignet für eine schnelle Kommunikation + leicht verständlich + einfach zu erlernen + erweiterbar bzw. anpassbar 	<ul style="list-style-type: none"> – keine einheitliche Codierungsvorschrift – beschreibt nicht die Ausführung der Gebärden mittels manuellen und nicht-manuellen Parametern

Tab. 2: *Vor- und Nachteile der Glossen-Notation*

3.2 Linguistische Notationssysteme

3.2.1 Stokoe-Notation

Der US-amerikanische Linguist William Stokoe (1919 – 2000) setzte mit seiner Untersuchung der ASL (*American Sign Language*) 1960 den Beginn der modernen Gebärdensprachforschung. In diesem Jahr veröffentlichte er in seinem Werk „A Dictionary of American Sign Language on Linguistic Principles“ die Stokoe-Notation. Es war die erste linguistische Schreibweise für eine Gebärdensprache und beeinflusste viele folgende Notationen, die sich an Stokoe's System orientierten. Bei der Untersuchung der ASL fand Stokoe die Parameter Ausführungsstelle (*tab*), Handform (*dez*) und Bewegung (*sig*), für die er 12, 19 und 24 Symbole entwarf. Der Anhang A1 zeigt diese definierten Symbole und deren Bedeutung. Diese insgesamt 55 Glyphen basieren auf lateinischen Buchstaben, Zahlen und einfach gehaltene bildliche Symbole. Die Notation ist von links nach rechts zu lesen und besitzt für eine Gebärde die Reihenfolge „tab – dez – sig“. Der Zeichensatz vom Bewegungsparameter beinhaltet Symbole zur Drehung und Richtungsänderung, was bei neueren Notationen getrennt wurde und als Handstellung ausführlicher aufgelistet wird. Die Notation war für akademische Zwecke gedacht. Sie sollte zeigen, dass eine Gebärdensprache eine feste Struktur und Grammatik aufweist [Mar03], [Sto12].

Beispiel: Stokoe-Notation

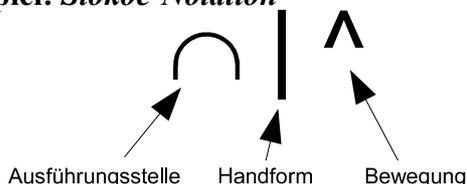


Abb. 29: links: Stokoe-Notation für die Gebärde „Idee“; rechts: Zugehörige Gebärde; ergänzt aus [Hoh07]

Die Darstellung 29 bildet ein Beispiel für die Stokoe-Notation ab. Im linken Bereich ist die Gebärde für das Wort „Idee“ notiert, rechts daneben ist die Ausführung zu sehen. Der Gebärdende führt die Hand von der Ausführungsstelle „Stirn“ mit der Handform „pinkie“ (abgespreizten kleinen Finger) in einer Bewegung nach oben. Die nachfolgende Tabelle 3 zeigt Vorteile und Nachteile der Notation.

Vorteile	Nachteile
<ul style="list-style-type: none"> + leicht zu erlernen + für Wörterbücher geeignet + erweiterbar + für PC geeignet (Unicode) 	<ul style="list-style-type: none"> – nur für die ASL geeignet – Parameter Handstellung nicht separat – unvollständig → es gibt mehr Handformen, Ausführungsstellen und Bewegungen in der ASL – nicht-manuelle Komponenten fehlen

Tab. 3: Vor- und Nachteile der Stokoe-Notation

3.2.2 Hamburger Notationssystem (HamNoSys)

Das Hamburger Notationssystem wurde 1984 vom Institut für „Deutsche Gebärdensprache und Kommunikation Gehörloser“ der Universität Hamburg entwickelt und im Jahre 1987 erstmalig veröffentlicht. Es basiert auf den Grundideen der Stokoe-Notation. Seitdem wird das phonetische Transkriptionssystem unter Berücksichtigung neuer Erkenntnisse aus der Linguistik stetig erweitert und befindet sich momentan in der Version 4.0. Es findet Anwendungen in der linguistischen Praxis von Gebärdensprachvideos (z. B. videobasiertes Fachgebärdenlexika) bis hin zur Animation von Avataren, bei der es in einer abgewandelten Form (SIGML) in den Projekten ViSiCAST und eSIGN zum Einsatz gekommen ist.

Das HamNoSys basiert auf bildlichen Darstellungen, den sogenannten Piktogrammen bzw. Glyphen und ist für mehrere Gebärdensprachen geeignet. Innerhalb eines Piktogramms sind keine parallelen Informationen abgebildet und somit weist dieses Notationssystem eine lineare logische formale Struktur auf. Die Notation wird von links nach rechts gelesen und besteht aus (Symmetrie-Operator), (nicht-manuelle Komponenten), Handform, Handstellung, (Ausführungsstelle) und Bewegung. In dieser Reihenfolge treten die in Klammern gesetzten Begriffe als Option auf und müssen dementsprechend nicht im Hamburger Notationssystem für eine Gebärde abgebildet sein. Durch die Bildhaftigkeit der verwendeten Symbole ist es für den Nutzer leicht, diese zu erlernen und sich zu merken. Mit Hilfe verschiedener PC-Tools (HamNoSys gibt es als Unicode) liefert es eine gute Integrationsfähigkeit und kann somit z. B. für Datenbankanwendungen genutzt werden.

Zusätzlich zu den Parametern (Handform, Ausführungsstelle und Bewegung) aus der Stokoe-Notation besitzt das HamNoSys den Parameter der Handstellung, der in Fingeransatzrichtung und Handflächenorientierung unterteilt wird. Die linke Seite der Abbildung 30 zeigt im oberen Bereich ausgewählte mögliche Handformen und die darunter liegenden zugehörigen Glyphen im HamNoSys. Die rechte Seite stellt mögliche Handstellungen dar. Bei der Orientierung im Gebärdenraum sind die verschiedenen Posen jeweils im 45° Winkel definiert, was die Notation zwar limitiert, da nicht alle Richtungen dargestellt werden, jedoch für die Übersichtlichkeit und Verständlichkeit als Kompromiss zu sehen ist. Ebenfalls wie in der linken Abbildung bestehen die Glyphen nur aus den Richtungspfeilen und Linien. Die Person, der Arm und die Hände in der Abbildung 30 dienen nur zur Verdeutlichung der Symbole.

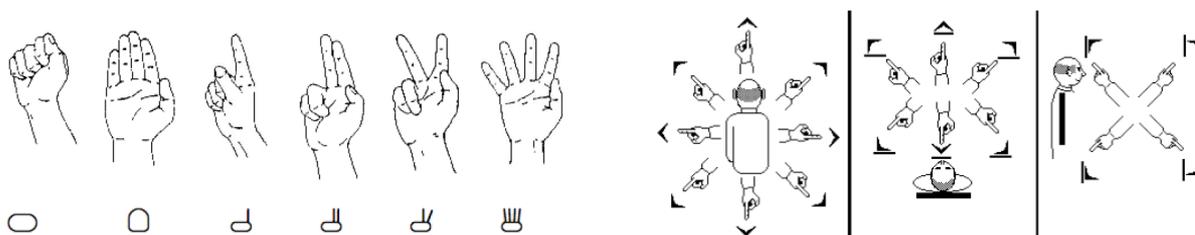


Abb. 30: links: Handformen; rechts: Handstellungen; aus [Han04]

Die hier gezeigten Symbole sind nur eine kleine Auswahl. Insgesamt gibt es ca. 200 Symbole, wobei die meisten für eine Handform stehen. Durch die Kombinierbarkeit dieser sollen alle bislang untersuchten Handformen in Gebärdensprachen abgedeckt sein. Die Quelle [Bak09] stellt einen großen Teil der HamNoSys-Glyphen dar.

Neben diesen manuellen Komponenten unterstützt die Notation ebenfalls die nicht-manuellen Komponenten, welche aber im Moment noch stark eingeschränkt sind. Kopf- und Oberkörperbewegungen sowie Schulterbewegungen sind jedoch möglich. Andere nicht-manuelle Ausdrucksmittel sollten eigentlich auch durch Piktogramme abgebildet werden. Jedoch gab es bei der Entwicklung praktische Grenzen, da es eine Vielzahl von Piktogrammen benötigt, um diesen Bereich abzudecken. Somit führte man Code-Schemen für Gesichtsausdruck, Mundbilder, Augen, Augenbrauen, Nase und Wangen ein. So steht zum Beispiel in dem Code-Schema für das Mundbild der Codeausdruck „C01“ für das „Aufblasen der Wangen“. Des Weiteren gibt es Symbole für die Darstellung von Zweihandgebärden unter Beachtung der dominanten und nicht-dominanten Hand [Ken02], [Han04], [Kip11], [Ham12].

Beispiel: *HamNoSys*

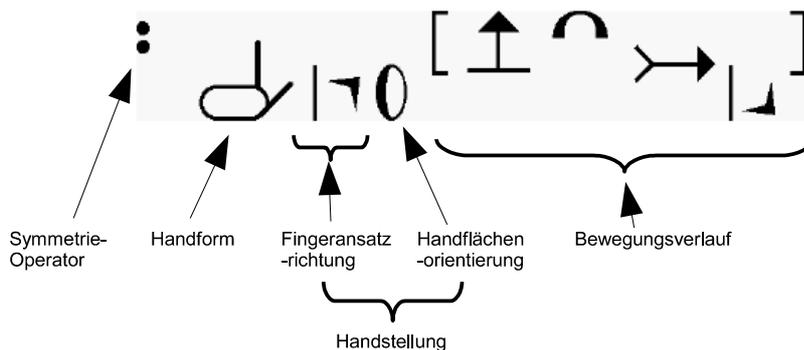


Abb. 31: *HamNoSys*-Gebärde „GEHEN-ZU“; ergänzt aus [Ken02]

Zur Verständlichkeit der Notation soll ein Beispiel in der Abbildung 31 dienen. Die beiden Punkte am Anfang zeigen, dass beide Hände spiegelsymmetrisch benutzt werden. Das darauffolgende Symbol steht für die Handform und zeigt wie Zeigefinger und Daumen gehalten werden. Die Handstellung bestehend aus der Fingeransatzrichtung und der Handflächenorientierung zeigt die Richtung der Handform im Gebärdenraum an. Diese bedeuten, dass die Hand vor dem Körper nach oben gerichtet ist (Fingeransatzrichtung) und die schwarz ausgemalte Seite des Kreises, der für die Handinnenfläche steht, die rechte Hand symbolisiert. In den eckigen Klammern wird der Bewegungsablauf ausgedrückt. Dieser besagt, dass die Hände nach oben, bogenförmig vorwärts in der senkrechten Ebene bewegt werden und dabei die Handstellung mit der Fingeransatzorientierung am Ende der Aktion nach schräg unten zeigt. In diesem Beispiel steht kein Symbol für die Ausführungsstelle der Gebärde, was bedeutet, dass diese in der „neutralen Ausführungsstelle“ auszuführen ist.

Zusammenfassend gliedert die Tabelle 4 Vor- und Nachteile dieser linguistischen Notation.

Vorteile	Nachteile
+ stetige Weiterentwicklung	– limitierte Handstellung im Gebärdenraum
+ gute Skalierbarkeit	– Transkription in HamNoSys ist sehr aufwendig → bei Glossen geht es schneller
+ geeignet für mehrere Gebärdensprachen → international anwendbar	– keine detaillierte vollständige Gebärden- beschreibung → z. B. Armhaltung fehlt
+ vielseitig anwendbar	→ keine Geschwindigkeitsregulierung
+ gute Integrationsfähigkeit	– nicht-manuelle Komponenten ausbaufähig
+ unterstützt manuelle und nicht-manuelle Komponenten	
+ weit verbreitet und anerkannt	
+ am PC nutzbar	
+ hohe Bildhaftigkeit → leicht merkbar und erlernbar	

Tab. 4: Vor- und Nachteile HamNoSys

3.3 Animationsorientierte Notationssysteme

3.3.1 SIGML (Signing Gesture Markup Language)

Die SIGML-Notation ist eine abgewandelte XML (*Extensible Mark-up Language*) Variante vom Hamburger Notationssystem und wurde speziell für die Ansteuerung von Gebärdensprachavataren in den Projekten ViSiCAST und eSIGN innerhalb von 3 Jahren entwickelt. Gegenüber dem HamNoSys beherrscht sie unter Verwendung von „SAMPa“ (*Speech Assessment Method Phonetic Alphabet*), welches ein auf ASCII basiertes phonetische Alphabet ist, Möglichkeiten zur genauen Einstellung des Mundbildes. Darüber hinaus gibt es Erweiterungen, die die nicht-manuellen Komponenten wie Mimik, Kopf, Blick, Schulter und Oberkörper enthalten. Zusätzlich können Parameter die Geschwindigkeit, die Dauer und die Kameraposition des Avatars festlegen. Der Anwendungsbereich ist auf die Ansteuerung von Avataren begrenzt. Für die Ausführung der Animation wird eine zusätzliche Software (z. B. SIGML Signing) benötigt. Der Vorteil liegt in der XML-Struktur, die erlaubt, die Notation für Anwendungen im Internet zu nutzen [Ken02], [Kip11].

Das Beispiel in der Abb. 32 bildet die gleiche Gebärde in SIGML ab, wie das Beispiel aus der Abb. 31 im HamNoSys.

Beispiel: SIGML

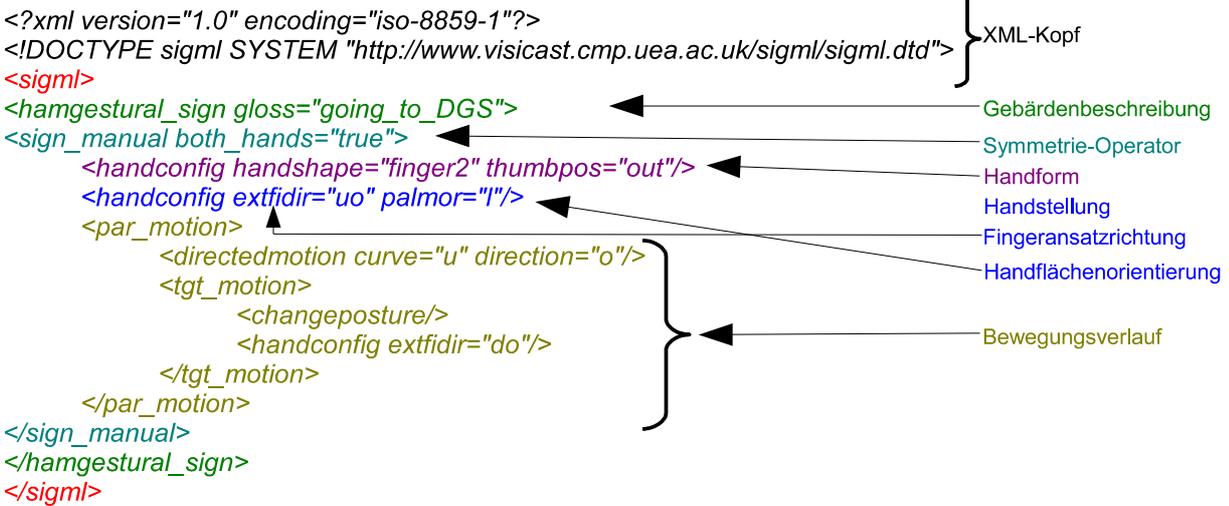


Abb. 32: SIGML-Gebärde „GEHEN-ZU“; verändert aus [Ell12]

Die farbliche Darstellung des XML-Codes soll die Struktur der Gebärde hervorheben. Die Reihenfolge entspricht hierbei der des Hamburger Notationssystem. Der Symmetrie-Operator sagt aus, dass beide Hände benutzt werden. Handform und Handstellung definieren die Haltung und Orientierung der Hand. Der Bewegungsverlauf, der im HamNoSys Beispiel schon beschrieben wurde, verläuft an dieser Stelle gleich ab. Vor- und Nachteile der Notation fasst die Tabelle 5 zusammen.

Vorteile	Nachteile
+ Mundbild kann spezifiziert werden	– abhängig vom HamNoSys
+ basiert auf HamNoSys	– XML-Kenntnisse sind nötig
+ vollständige Unterstützung der manuellen und nicht-manuellen Komponenten	– benötigt weitere Software für die Animation eines Avatars
+ Geschwindigkeitsregulierung der Gebärden möglich	
+ Kamera positionierbar	
+ geeignet für mehrere Gebärdensprachen	
+ Anwendbar im Internet (XML-Struktur)	

Tab. 5: Vor- und Nachteile SIGML-Notation

3.3.2 BML (Behavior Markup Language)

Im Jahr 2005 entwickelten Forscher aus verschiedenen Nationen in einem Workshop an der Universität Reykjavik das SAIBA-Framework. Das Ziel der Wissenschaftler war es ein Standard für Echtzeit multimodale Verhaltensgenerierung im Bereich der Kommunikationsavatare zu kreieren. Das Ergebnis zeigt Abbildung 33 mit den 3 Ebenen der Verhaltensgenerierung, die das Framework beschreibt.

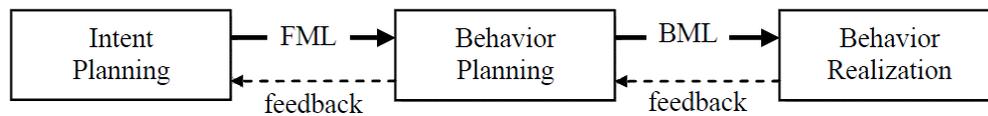


Abb. 33: SAIBA-Framework; aus [Vil07]

Die Abkürzung „SAIBA“ steht für *Situation* (die Umstände), *Agent* (Ziele, Emotionen, Planung), *Intention* (Was soll ausgedrückt werden?), *Behavior* (Wie soll es ausgedrückt werden?) und *Animation* (die Realisierung). Es ist unabhängig von anderen Anwendungen, Sound oder Grafik-Modellen (z. B. Avatare).

In der „Intent Planning“ Phase wird die kommunikative Absicht festgelegt, das heißt die Rahmenbedingung werden definiert. In welcher Situation befindet sich die zu animierende Figur (z. B. zwei Avatare führen ein Dialog) und wie sieht der Ablauf dieser Kommunikation aus. Diese Informationen (z. B. auch Emotionen) beschreibt das FML (*Function Markup Language*)-Skript. Anschließend werden in der „Behavior Planning“ Phase die Verhaltensgesten für diese geplante Situation mittels BML definiert. Zum Schluss realisiert ein Animationssystem in der „Behavior Realization“ Phase den Bewegungsablauf bzw. die Verhaltensweise des Avatars. Die Unterteilung in diese Ebenen gibt SAIBA eine gewisse Flexibilität.

In BML erfolgt die abstrakte Beschreibung der Aktionen. Es basiert auf XML und kann durch die Anweisung „<bml>“ aufgerufen werden.

Beispiel: BML

Behavior Block	{	<code><bml></code>	
		<code><speech id="s1" type="application/ssml+xml"></code>	Behaviors
		<code><text>This is an <mark name="wb3"> example</text></code>	
		<code></speech></code>	
		<code><head id="h1" type="NOD" stroke="s1:start"/></code>	Behavior-ID
		<code><gesture id="g1" stroke="s1:wb3" relax="s1:end" type="BEAT"></code>	
		<code><description level="1" type="MURML">...</code>	Synchronisationspunkt
		<code></description></code>	
		<code></gesture></code>	
		<code><gaze id="z1" target="PERSON1" stroke="g1:stroke-0.1"/></code>	Namespace Extensions
<code><body id="p1" posture="RELAXED" start="after(s1:end)"/></code>			
<code><cadia:operate target="SWITCH1" stroke="p1:ready"/></code>			
<code></bml></code>			

Abb. 34: BML-Beispiel; verändert aus [Vil07]

Das Beispiel in Darstellung 34 zeigt das BML Sprache, Gesten (Arm- und Handkoordination), Blickrichtung, Kopf- und Körperbewegung abbilden kann. Weitere mögliche Elemente sind Gesichtsmimiken (speziell Mund- und Augenausdruck/-bewegung), Beine, Wartehaltung, Raumbewegungen (Avatar bewegt sich in der Umgebung) und Lippen (Darstellung von Viseme). Neben diesen Kern-BML Elementen können eigene Erweiterungen (*Namespace Extensions*) neue Verhaltensdefinitionen hinzufügen.

Der *Behavior Block* beschreibt eine kommunikative Verhaltensweise und besteht aus mindestens einem Verhalten (*Behavior*). Jedes Verhalten erhält eine eindeutige Identifizierung (*Behavior-ID*), die dann einem *Synchronisationspunkt* zugeordnet werden kann. Insgesamt beinhaltet BML sieben dieser Synchronisationspunkte, die in der Abbildung 35 dargestellt sind.

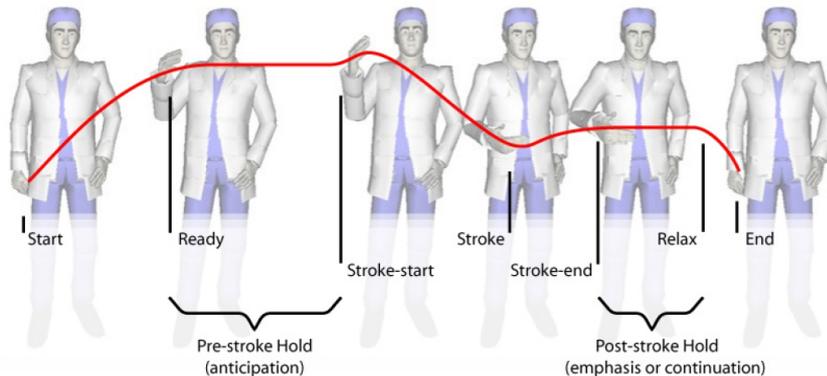


Abb. 35: BML-Synchronisationspunkte; aus [Kop06]

Der Synchronisationspunkt „Start“ legt die Ausgangshaltung des Avatars fest. Ist die Geste ausgeführt, bewegt sich der Avatar wieder in diese Ausgangshaltung und verharrt in diese („End“). Bei „Ready“ führt der Avatar seine Verhaltensweise zu der Ausführungsstelle der Geste bzw. Gebärde. Analog ist „Relax“ der Bewegungsverlauf zurück zur Endpose. Die eigentliche Aktion erfolgt zwischen den beiden Punkten Ready und Relax. Dabei beschreibt die „Stroke“ Phase den Teil der Aktion, bei dem die meiste Bewegung vollzogen wird. „Stroke-start“ und „Stroke-end“ charakterisieren lediglich den Anfang und das Ende dieser aussagekräftigen Bewegung. Ein Verhalten muss diese Stroke Phase nicht unbedingt haben. So kann bei einem Blick in den Raum nur die Ready (Start des Augenkontaktes) und Relax Phase (Abbrechen des Augenkontaktes) ablaufen. Die Synchronisationspunkte erlauben es, parallele Verhaltensweisen ablaufen zu lassen und erzeugen somit den Ablaufplan der Gesten. So kann z. B. während der Lippenbewegung auch Sprache synchron ausgegeben werden [Kop06], [Vil07].

Die Vor- und Nachteile der BML-Notation sind in der Tabelle 6 zusammengefasst.

Vorteile	Nachteile
<ul style="list-style-type: none"> + unterstützt manuelle und nicht-manuelle Komponenten + unabhängig und flexibel + erweiterbar + Ansatz eines Standards für animationsorientierte Avatarprojekte 	<ul style="list-style-type: none"> – eingeschränkte Verhaltensbeschreibungen → EMBRScript oder MURML nötig – XML-Kenntnisse sind nötig

Tab. 6: Vor- und Nachteile BML

Die Funktionen zur Beschreibung von Gesten sind in BML sehr eingeschränkt. Für eine höhere Funktionsweise und Präzision im Bewegungsablauf wurden weitere animationsorientierte Notationen wie zum Beispiel MURML oder EMBRScript entwickelt, die auf den BML-Standard und dem SAIBA-Framework aufsetzen. Für eine ausführlichere Darstellung des Behavior Markup Language Standard sei an dieser Stelle auf die Quelle [BML12] hingewiesen.

3.3.3 EMBRScript (Embodied Agent Behavior Realizer Script)

Die EMBRScript Notation bildet eine neue Ebene (*Animation Layer*) im SAIBA-Framework (vgl. Abb. 36) und liegt zwischen der Verhaltensbeschreibung BML und dem Animationssystem.



Abb. 36: SAIBA-Framework mit EMBRScript; verändert aus [Hel10]

Das SAIBA-Framework gliedert die Kontrolle einer animierten Figur in drei unabhängige Level ein. Das erlaubt den Nutzer, selber Animationen zu erstellen, da eine einheitliche Schnittstelle geschaffen wurde. Die zusätzliche Ebene EMBRScript ermöglicht es, die Parameter für die Animation des Avatars präziser und komplexer im Verhalten anzusprechen. So kann zum Beispiel in BML der Avatar die Anweisung bekommen „Schau auf den Punkt A“, in EMBRScript kann zusätzlich die Art der Bewegung, die Zeit und die genauen Koordinaten angesprochen werden. EMBRScript bietet den Vorteil, dass es unabhängig vom direkten Planungsprozess ist und somit nicht auf andere theoretische Modelle basiert.

Die Notation wird innerhalb eines BML-XML-Files aufgerufen. In EMBRScript wird jede Bewegungsanimation durch Schlüsselposen (*key poses*) definiert. Eine Schlüsselpose beschreibt den Status des Avatars (z. B. Augenlid schließen) zu einem bestimmten Zeitpunkt (*TIME_POINT*), welcher für eine gewisse Zeitdauer gehalten werden kann (*HOLD*). Durch die Aneinanderreihung von verschiedenen Posen, die auch parallel ablaufen können, ist jede manuelle und nicht-manuelle Komponente der Gebärdensprache in einer Animation realisierbar. Die Notation vermeidet tiefe verschachtelte Strukturen, was der Übersicht zu Gute kommt und dadurch die Benutzerfreundlichkeit steigert. Der Übergang zwischen benachbarten Posen basiert auf Interpolation für die es vier wählbare Methoden gibt. Diese sind: „Skelettkonfiguration“, „Zielverformung“ (*morph targets*), „Schattierer“ (*shader*) und „autonomes Verhalten“.

Bei der Skelettkonfiguration erhält die animierte Figur ein starres Skelett. Es gibt zwei Möglichkeiten zur Definition. Mit Forward Kinematics (FK) werden alle Winkel und Gelenke

unter Nutzung eines 3D-Animation-Tools konfiguriert. Bei der Inverse Kinematics (IK) kann ein Set von Beschränkungen (*constraints*) automatisch durch den sogenannten *IK-Solver* erkannt und erstellt werden. Diese Methode wird zum Beispiel benutzt, um einen Punkt im Raum zu erreichen, bei der sich der Charakter nach vorne beugen muss. Das Beispiel in der Abbildung 37 zeigt, wie die rechte Hand (Position und Orientierung) mittels kartesischen Koordinaten im Raum festgelegt wird.

Beispiel: Skelettkonfiguration

```
BEGIN POSITION_CONSTRAINT
  BODY_GROUP: rarm
  TARGET: 2.2; 0.0; -2
  JOINT: rhand
END
BEGIN ORIENTATION_CONSTRAINT
  BODY_GROUP: rarm
  NORMAL: 0.0; 0.0; -1.0
  JOINT: rhand
END
```

Abb. 37: Konfiguration des Skeletts; aus [Hel09]

Die Zielverformung und der Schattierer sind für die nicht-manuellen Komponenten gedacht. Mit Hilfe des *morph targets* kann die Gesichtsmimik (z. B. Augenzwinkern, Lächeln (vgl. Abbildung 38)) oder Emotionen dargestellt werden. Der *shader* kann dabei eine unterstützende Funktion liefern. Dieser erlaubt die Hautfarbe der animierten Figur anzupassen. So können Emotionen wie Nervosität, Angst oder Aufgeregtheit durch Erröten oder Erblassen gekennzeichnet sein.

Beispiel: Zielverformung

```
BEGIN K_POSE_SEQUENCE
CHARACTER:Amber
START:asap
FADE_IN:200
FADE_OUT:200
BEGIN K_POSE # +++ REST POSE +++
TIME_POINT:+500
HOLD:0
BEGIN MORPH_TARGET # +++ REST POSE +++
MORPH_KEY:ModBlinkRight
MORPH_VALUE:1.0
END
BEGIN MORPH_TARGET # +++ REST POSE +++
MORPH_KEY:ExpSmileClosed
MORPH_VALUE:1.0
END
END
BEGIN K_POSE # +++ REST POSE +++
TIME_POINT:+800
HOLD:0
BEGIN MORPH_TARGET # +++ REST POSE +++
MORPH_KEY:ModBlinkRight
MORPH_VALUE:0.0
END
BEGIN MORPH_TARGET # +++ REST POSE +++
MORPH_KEY:ExpSmileClosed
MORPH_VALUE:0.0
END
END
END
```



Abb. 38: EMBRScript-Code für Augenzwinkern und Lächeln

Das autonome Verhalten bildet die Grundlage der menschlichen Verhaltensweise ab und soll den Avatar natürlicher wirken lassen. Zu diesen teils unbewussten Handlungsweisen zählen Blickverfolgung, Augenbewegungen, Augenzwinkern, Gewichtsverlagerung, Balance Kontrolle und Atembewegung. Implementiert sind bisher jedoch nur das Augenzwinkern und die Atembewegung. Die Entwickler wollten die EMBRScript-Notation nicht zu restriktiv definieren und den Nutzern die Möglichkeit geben, eigene Ideen und Verhaltensweisen zu implementieren [Hel09], [Hel10].

Ein Überblick über die Vor- bzw. Nachteile der EMBRScript-Notation gibt die Tabelle 7.

Vorteile	Nachteile
+ Theorie unabhängig → basiert auf keiner anderen Notation + übersichtliche Struktur + jeder Bewegungsverlauf programmierbar → manuelle und nicht-manuelle Komponenten werden unterstützt + unterstützt natürliches Verhalten + Echtzeitanwendung + vielseitig anwendbar → GS-Avatar, Film, Computerspiele	– Bewegungen des Avatars wirken kantig – XML-Kenntnisse sind nötig

Tab. 7: Vor- und Nachteile EMBRScript-Notation

3.3.4 Fazit zu animationsorientierten Notationen für die Gebärdensprache

Das Kapitel zeigt, dass es mehrere Möglichkeiten gibt eine Gebärdensprache schriftlich abzubilden. Ein einheitlich anerkanntes Schriftsystem gibt es nicht. Um ein Avatar ansteuern zu können, benötigt man eine animationsorientierte Notation. Diese lassen sich grob in den zwei Klassen „Notations-basierte“ Schriftsysteme und „Verhaltens-basierte“ Schriftsysteme unterteilen.

Die „Notations-basierten“ Schriftsysteme bauen auf eine linguistische Notation auf. Beispiele hierfür sind die animationsorientierten Notationen SIGML und SWML, die auf die linguistischen Notationen HamNoSys und dem SignWriting basieren. Sie haben den Vorteil, dass sie sich aus Texten der zugehörigen linguistischen Schreibsysteme mit Hilfe von Computerprogrammen generieren lassen. Somit benötigt der Anwender kein Hintergrundwissen zum Aufbau und zur Funktionsweise dieser „Notation-basierten“ Schriftsysteme.

Zu den „Verhaltens-basierten“ Schriftsystemen lassen sich die Notationen BML, EMBRScript, MURML und Zebedee zählen. Sie basieren auf keiner anderen Notation und sind somit vollständig unabhängig. Das hat den Vorteil, dass sie flexibel anpassbar für eine Avataranwendung sind. Somit können diese Notationen nicht nur für die Gebärdensprache benutzt werden, sondern auch für andere Bereiche in denen Avatare zur Kommunikation zum Einsatz kommen sollen. Die Verhaltensweisen und Animationen für den Avatar muss der Nutzer selber programmieren. Das hat den Nachteil, dass bei einem Gebärdensprachavatar der Programmierer die Gebärdensprache beherrschen muss, um die Verhaltensgesten (Gebärden) in den Skriptsprachen richtig zu programmieren.

Nachdem einige Schriftsysteme für die Gebärdensprache vorgestellt wurden, behandelt das folgende 4. Kapitel Strategien und Bewertungsmöglichkeiten im Bereich der Maschinellen Übersetzung, um einen Text aus der Lautsprache in eine animationsorientierte Gebärdensprachnotation zu überführen.

4. Kapitel: Maschinelle Übersetzung

Das Kapitel 4 ist folgendermaßen untergliedert. Abschnitt 4.1 stellt kurz die Terminologie und Herausforderungen an die Maschinelle Übersetzung vor. Im zweiten Abschnitt 4.2 werden die grundlegenden Übersetzungsstrategien der Maschinellen Übersetzung erläutert. Anschließend folgt im dritten Abschnitt 4.3 eine Übersicht der Evaluierungsmöglichkeiten, die die Genauigkeit und Qualität eines MÜ-Systems bewerten können. Der vierte Abschnitt 4.4 enthält eine Zusammenfassung veröffentlichter Arbeiten, die eine Maschinelle Übersetzung für den Bereich der Gebärdensprache benutzen. Abschließend fasst Abschnitt 4.5 die Erkenntnisse zusammen und bildet einen Workflow zur Übersetzung von einem Text in eine animationsorientierte Notation ab.

4.1 Grundlegendes zur Maschinellen Übersetzung

Terminologie

Die „*Maschinelle Übersetzung (MÜ)*“ (*machine translation (MT)*) nutzt technische Möglichkeiten für die vollständige automatische Übersetzung einer textuellen oder gesprochenen natürlichen Sprache in eine andere natürliche Sprache. Die Motivation für MÜ-Systeme sind Sprachbarrieren zu überwinden (Erleichterung der Kommunikation), Übersetzungskosten zu verringern und hohe Übersetzungsgeschwindigkeiten zu erreichen.

Die Ausgangssprache für die Übersetzung bezeichnet man als „*Quellsprache (Qs)*“ (*source language (SL)*), die in die sogenannte „*Zielsprache (Zs)*“ (*target language (TL)*) überführt werden soll. Ein MÜ-System kann so definiert sein, dass es nur in eine Sprachrichtung (z. B. Deutsch nach Englisch) übersetzt. Dann spricht man von einer „*unidirektionalen*“ Übersetzung. Ist eine Übersetzung zwischen beiden Sprachrichtungen möglich, wird dies als „*bidirektionale*“ Übersetzung bezeichnet. Bei der Übersetzung zwischen einem Sprachpaar spricht man von einem „*bilingualen*“ Übersetzungssystem. Verwendet das System mehr als zwei Sprachen, gehört es zu den „*multilingualen*“ Übersetzungssystemen. Wird der Übersetzungsprozess vollautomatisch durchgeführt, ist es ein „*Batch-System*“. Kann oder muss der Nutzer beim Übersetzungsprozess interagieren, zählt es zu der „*Computerunterstützten Übersetzung*“ (*computer-aided translation (CAT)*). Die Maschinelle Übersetzung umfasst interdisziplinäres Wissen aus den Bereichen der Informatik, Linguistik, Computerlinguistik (CL), Übersetzungswissenschaft (Translatologie) und der Künstlichen Intelligenz (KI).

Herausforderungen für die Maschinelle Übersetzung

Durch die Vielseitigkeit einer Sprache treten häufig Probleme bei der Maschinellen Übersetzung auf, die negativen Einfluss auf die Qualität der Übersetzung haben. Zu diesen sprachlichen Herausforderungen zählen hauptsächlich „*Ambiguitäten*“ und „*Divergenzen*“. Diese Mehrdeutigkeiten und Strukturunterschiede können auf verschiedene Ebenen auftreten.

Die nachfolgende Aufzählung zeigt Beispiele zu diesen Ebenen und deutet die Schwierigkeiten für ein MÜ-System an.

Ambiguitäten

- **lexikalisch** = Doppeldeutigkeit von Wörtern
 Bsp. 1: Deutsch → Englisch: „*Schloss*“ → „*lock*“ oder „*castle*“
 Bsp. 2: Englisch → Deutsch: „*chair*“ → „*Stuhl*“ oder „*Sessel*“
 Bsp. 3: Englisch → Deutsch: „*know*“ → „*wissen*“ oder „*kennen*“

- **syntaktisch** = Doppeldeutigkeit der Satzsyntax
 Bsp. 1: „*Er erschlug den Mann mit dem Apfel.*“
 Bsp. 2: „*They are flying planes.*“

- **semantisch** = Doppeldeutigkeit der Semantik eines Wortes (sogenannte Polyseme)
 Bsp. 1: „*Bank*“ = „*Geldinstitut* oder *Sitzgelegenheit*“
 Bsp. 2: „*Tenor*“ = „*grundsätzliche Einstellung* oder *hohe Männersingstimme*“

- **referentiell** = Doppeldeutigkeit bei Beziehungsobjekten
 Bsp. 1: „*Paul schießt den Ball ins Tor von Max. Anschließend rollt er raus.*“
 Bsp. 2: „*In den amerikanischen Nationalparks gibt es viele Plumpsklos. Dort kann man wunderbar entspannen.*“
 Bsp. 3: „*Put the paper in the printer. Then switch it on.*“

Divergenzen

- **lexikalisch** = strukturelle Unterschiede der Wortanzahl eines Satzes
 Bsp. 1: Deutsch → Englisch: „*Magst du Fußball?*“ → „*Do you like football?*“
 Bsp. 2: Deutsch → Englisch: „*Kennst du diesen Sport?*“ → „*Do you know this sport?*“
 Bsp. 3: Englisch → Französisch: „*I don't know.*“ → „*Je l'ignore.*“

Beispiel 3 stellt ein „*lexikalisches Loch*“ dar. Das heißt es gibt keine passende Übersetzung in den Sprachen, sondern nur eine annähernde Phrase.

- **syntaktisch** = strukturelle Unterschiede in der Satzsyntax
 Bsp. 1: Deutsch → Englisch: „*Das Buch gefällt Eva.*“ → „*Eva likes the book.*“
 Bsp. 2: Deutsch → Englisch: „*Paul spielt gerne Fußball.*“ → „*Paul likes to play football.*“

Neben diesen Schwierigkeiten gibt es weitere linguistische Herausforderungen (z. B. Kollokation, idiomatische Ausdrücke, Ellipsen usw.), die bei der Entwicklung eines Übersetzungssystems beachtet werden müssen [Neu05].

Inwieweit diese sprachlichen Ambiguitäten bzw. Divergenzen für ein MÜ-System im Bereich der Gebärdensprache und einer zugehörigen Notation auftauchen, könnte in weiteren Arbeiten genauer untersucht werden.

4.2 Strategien der Maschinellen Übersetzung

Im Bereich der Maschinellen Übersetzung haben sich in den letzten Jahren zwei Strategien, die „*Regel-basierten Verfahren*“ und die „*Daten-basierten Verfahren*“, etabliert. In diesem Unterkapitel werden beide Grundprinzipien der MÜ näher betrachtet.

4.2.1 Regel-basierte Verfahren – rule-based machine translation (RBMT)

Einer der Pioniere der Maschinellen Übersetzung, Bernard Vauquois (1929 - 1985), stellte 1975 in seiner Arbeit „*La Traduction Automatique á Grenoble*“ Ideen für die Maschinelle Übersetzung basierend auf Regeln vor. Die Abbildung 39 zeigt seinen Ansatz, das sogenannte „*Vauquois Dreieck*“. Eine Quellsprache kann auf mehreren Ebenen in die Zielsprache übersetzt werden. Je besser die Analyse der Ausgangssprache ist, umso höher ist die Übersetzungsqualität bei der Generierung der Zielsprache. Dieser Ansatz gilt heute als klassischer Ansatz für Regel-basierte Verfahren in der MÜ und findet sich in vielen kommerziellen Übersetzungssystemen wieder.

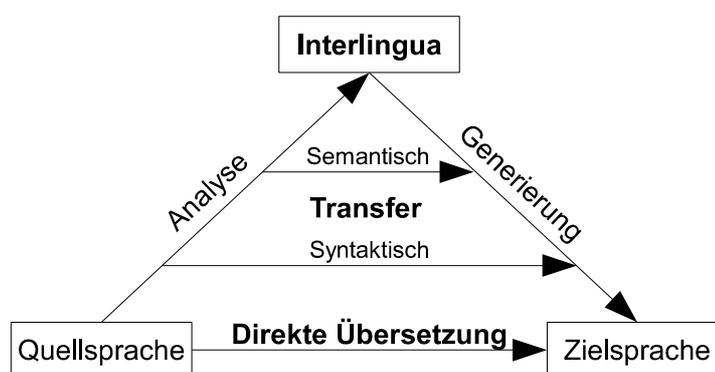


Abb. 39: MÜ-Dreieck nach Vauquois

4.2.1.1 Direkte Übersetzung

Auf der untersten Ebene des Vauquois Dreiecks befindet sich die „*direkte Übersetzung*“. Diese Ebene stellt die 1. Generation der Maschinellen Übersetzung dar. Ein Ausgangstext der Quellsprache wird hier morphologisch analysiert und mit Hilfe eines bilingualen Wörterbuchs Wort-zu-Wort in die Zielsprache übersetzt. Bei einfach strukturellen Sätzen, wie z. B. der deutsche Qs-Satz: „*Mein Name ist Paul.*“ maschinell übersetzt in einen englischen Zs-Satz: „*My name is Paul.*“, liefert dieser Ansatz gute Ergebnisse. Weitere Vorteile sind die Einfach-

heit, Robustheit und gute Übersetzungsgeschwindigkeit. Bei komplexeren Sätzen, bei denen die Satzsyntax abweichend zwischen Quellsprache und Zielsprache ist, sinkt die Qualität auf schlecht bis unverständlich ab. Grund dafür sind die fehlende Analyse und Regeln auf syntaktischer und semantischer Ebene. Des Weiteren kann die direkte Übersetzung die verschiedenen Ambiguitäten und Divergenzen nicht lösen. Deshalb benutzen heutige MÜ-Systeme diesen Ansatz nicht mehr [Höh05], [See09].

4.2.1.2 Transfer Übersetzung

Die „Transfer Übersetzung“ lässt sich in die drei Phasen „Analyse“, „Transfer“ und „Generierung“ unterteilen. Die grundsätzliche Vorgehensweise läuft folgendermaßen ab. Ein Parser analysiert die Syntax, in einigen Systemen auch die Semantik, des Ausgangssatzes. Er speichert diese Analyse in eine abstrakte Repräsentation für den Qs-Satz. Von Experten, in der Regel Linguisten, definierte Transferregeln überführen diese Repräsentation des Qs-Satzes in die abstrakte Repräsentation des Zs-Satzes. Zusätzlich übersetzt ein bilinguales Wörterbuch die entsprechenden Wörter in die Zielsprache. Ein Generierungsprozess erzeugt anschließend aus der Zwischenrepräsentation den übersetzten Zs-Satz.

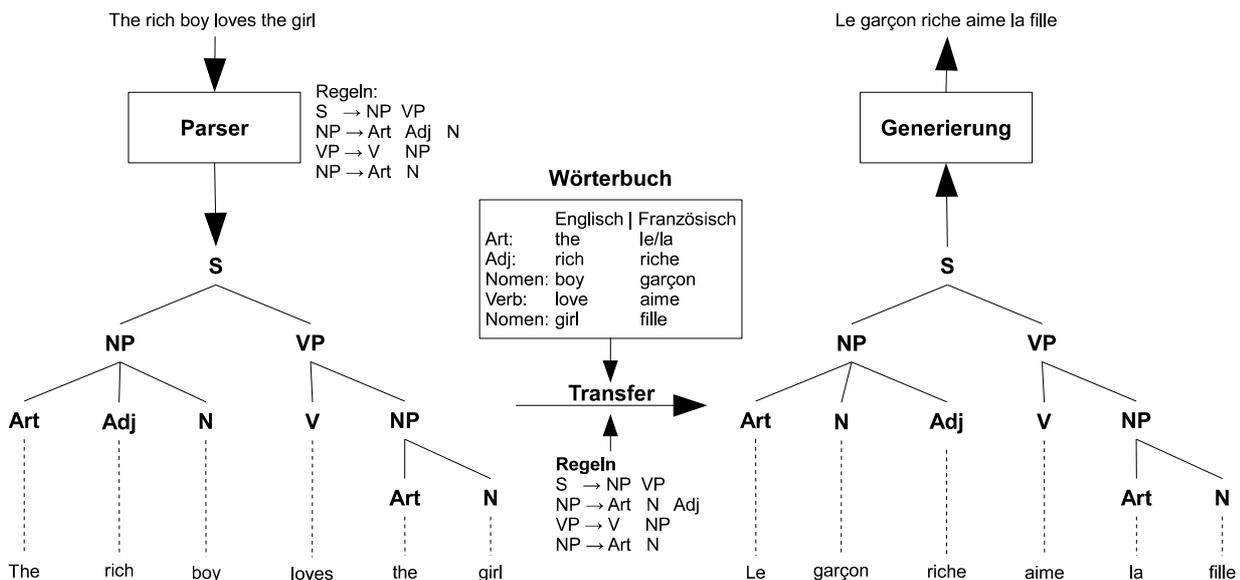


Abb. 40: Beispiel für eine Transfer Übersetzung

Die Darstellung 40 zeigt den prinzipiellen Ablauf der Transfer Übersetzung für ein Beispiel. Der englische Satz „The rich boy loves the girl.“ soll ins Französische übersetzt werden. Ein Parser kann mit Hilfe von definierten Regeln, die in der Analysephase ermittelt wurden, die Syntax des Qs-Satzes in einen Syntax-Baum abspeichern. Die selbstgewählten Bezeichnungen in den Regeln stehen für S = Satz, NP = Nominalphrase, VP = Verbalphrase, Art = Artikel, Adj = Adjektiv, V = Verb und N = Nomen. Das Wörterbuch und die Transferregeln definieren den Syntax-Baum der Zielsprache. Der Generierungsprozess erzeugt aus dem Baum den übersetzten Satz „Le garçon riche aime la fille.“.

Problematisch in diesem Beispiel kann der englische Artikel „*the*“ sein, da das Übersetzungssystem für die korrekt französische Übersetzung wissen muss, ob das zugehörige Nomen im grammatikalischen Geschlecht (Genus) weiblich oder männlich ist („*le garçon*“ oder „*la garçon*“?). Um dieses Problem zu lösen, muss der Parser den Quellsatz in eine semantische Zwischenrepräsentation abbilden. Dazu gibt es sogenannte „*Unifikationsgrammatiken*“, die Merkmalstrukturen eines Satzes syntaktisch, semantisch oder phonologisch abbilden. Diese Zwischenrepräsentation kann als Baumstruktur oder Matrixstruktur auftauchen. Häufig kommen folgende Grammatiken zur Anwendung:

- Tree-Adjoining Grammar (TAG),
- Synchronous Tree Adjoining Grammar (STAG),
- Probabilistic Context-Free Grammar (PCFG),
- Head-Driven Phrase Structure Grammar (HPSG),
- Discourse Representation Structure (DRS) und
- Lexical functional grammar (LFG).

Für das Verständnis der Funktionsweisen dieser Grammatiken befindet sich im Anhang A3 ein Beispiel für die LFG. Die anderen Grammatiken arbeiten in einer ähnlichen Form, besitzen aber ihre speziellen Eigenheiten, die an dieser Stelle nicht näher aufgeführt werden. Weitere computerlinguistische Informationen zur Grammatik und Parsen können in [Car01] nachgelesen werden. Die Transfer Übersetzung wird zusammen mit der Interlingua zur 2. Generation der Maschinellen Übersetzung gezählt [Wer04], [Höh05], [See09].

4.2.1.3 Interlingua

Das Ziel bei diesem Maschinellen Übersetzungsansatz ist es, die Quellsprache in eine universelle sprachunabhängige Zwischenrepräsentation, die sogenannte „*Interlingua*“ (Zwischensprache), zu übersetzen. Die Interlingua enthält die grundlegendsten Bausteine der Sprachen und ermöglicht es somit zwischen jeder beliebigen Sprache eine korrekte Übersetzung zu finden. Diese Idee stellt sich als ein nahezu utopischer Gedanke heraus. Eine universell sprachunabhängige Darstellung wurde bis heute nicht entdeckt. Grund dafür ist, dass sich Sprachen aus der Kultur heraus entwickelt haben. Regeln zu finden, die die strukturellen Gegebenheit aller Sprachen beinhaltet, fällt sehr schwierig aus. Sprachsysteme mit einem Interlingua-Ansatz analysieren die Quellsprache und überführen diese in eine „Interlingua“, aus der die Zielsprache generiert wird. Gegenüber der syntaktischen und semantischen Analyse der Transfer Übersetzung nutzt die Analyse der Interlingua zusätzlich „Weltwissen“. KBMT (*Knowledge-Based Machine Translation*) gilt als Spezialfall der Regel-basierten Interlingua. Das Ziel dieser Übersetzungssysteme ist es, Ambiguitäten (Mehrdeutigkeiten) durch Wissen in spezialisierten Domänen zu lösen. So kann bei dem Satz „*Das Schloss liegt auf dem 'Berg/Tisch'.*“ mit Hilfe einer Wissensdatenbank beim Übersetzungsprozess erkannt werden, ob die Mehrdeutigkeit „*Schloss*“ ein Gebäude oder eine Vorrichtung zum Verschlie-

ßen von Türen darstellt. Jedoch sind diese Systeme keine vollständigen Interlingua Systeme, da sie sich auf eine Domäne beschränken und die Zwischenrepräsentation keine sprachneutrale bzw. sprachunabhängige Darstellung ist. Der Vorteil einer Interlingua wäre, eine schnelle Integration neuer Sprachen in das Übersetzungssystem [See09].

4.2.2 Daten-basierte Verfahren – data-based machine translation (DBMT)

Die beiden grundlegenden Strategien der Daten-basierten Verfahren sind die „*Statistik-basierte*“ und die „*Beispiel-basierte*“ Maschinelle Übersetzung. Beide Methoden benötigen einen „*parallelen bzw. bilingualen Korpus*“, der eine Ansammlung von Texten in der Quellsprache und der zugehörigen übersetzten Zielsprache enthält. Die manuelle Erstellung dieser Korpora ist meist sehr aufwendig, deshalb nutzt man vorhandene offizielle EU-Texte (z. B. Europarl), die schon in mehreren Sprachen übersetzt wurden. Um einen parallelen Korpus zu analysieren, kommen „*Alinierungstechniken*“ (*Alignment*) zur Anwendung, die von Wörtern, Wortsequenzen (sogenannte Phrasen) oder kompletten Sätzen die zugehörige Übersetzung selbstständig berechnen.

4.2.2.1 Statistik-basierte MÜ – statistic-based machine translation (SBMT)

Grundlage

Die Idee der Statistik-basierten Maschinellen Übersetzung besteht darin, durch statistische Methoden zwei Sprachen in einem bilingualen Korpus automatisch zu analysieren und mit Hilfe dieser Kenntnisse für einen eingegebenen Qs-Satz eine möglichst passende Übersetzung für die Zielsprache zu finden. Der Vorteil gegenüber der RBMT ist, dass man für die SBMT keine Linguisten benötigt, die beide Sprachen erst manuell analysieren, um anschließend Übersetzungsregeln abzuleiten.

Dieses Korpus-basierte Lernverfahren wurde 1990 von IBM-Mitarbeiter in der Arbeit „*A Statistical Approach to Machine Translation*“ in [Bro90] erstmalig vorgestellt. Grundlage ist ein bilingualer Korpus, bei denen die Übersetzungen vom Menschen manuell erstellt wurden. Ein erster Schritt ist es, die Sätze der Quellsprache im Korpus zu den entsprechenden Sätzen der Zielsprache zuzuordnen (bezeichnet als „*Satzalinierung*“). Diese Übersetzungen erhalten durch die „*Maximum likelihood estimation (MLE)*“ Wahrscheinlichkeitswerte zugewiesen, die ein probabilistisches Übersetzungsmodell trainiert. Somit enthält das Übersetzungsmodell die Wahrscheinlichkeiten, dass ein Wort t aus der Zielsprache eine gute Übersetzung für das Wort s aus der Quellsprache ist (formal: $P(t|s)$). Es kann aber mehrere Einträge für ein Wort geben. Beispielsweise das deutsche Wort „*das*“ lässt sich ins Englische mit „*the*“, „*that*“, „*which*“, „*who*“ oder „*this*“ übersetzen. Die Suche nach der höchsten Wahrscheinlichkeit für die Zielsprache wird formal mit $\underset{T}{\operatorname{argmax}}$ bezeichnet.

Bei der Eingabe eines neuen Qs-Satzes S , der nicht im Korpus enthalten ist, ermittelt das Übersetzungsmodell ein Zs-Satz T , der nach den Wahrscheinlichkeiten die bestmögliche Übersetzung darstellt. Formal ausgedrückt:

$$t_{best} = \underset{T}{\operatorname{argmax}} P(T|S) \quad (1)$$

Nach Anwendung des Bayes-Theorem für $P(T|S)$ ergibt sich:

$$t_{best} = \underset{T}{\operatorname{argmax}} \frac{P(T) \cdot P(S|T)}{P(S)} \quad (2)$$

Die Wahrscheinlichkeit $P(S)$ im Nenner des Bruches sagt aus, dass die Eingabe als Quellsprache ins Übersetzungssystem erfolgen muss. Diese Voraussetzung ist sowieso gegeben und kann somit in der Formel 3 ignoriert werden. Daraus folgt die grundlegende Formel für die Statistik-basierte Maschinelle Übersetzung:

$$t_{best} = \underset{T}{\operatorname{argmax}} P(T) \cdot P(S|T) \quad (3)$$

Nach [Bro90] beinhaltet die Formel 3 die drei wesentlichen Herausforderung für die SBMT. Zu diesen gehören die Berechnung der Wahrscheinlichkeit des Sprachmodells $P(T)$, die Berechnung der Übersetzungswahrscheinlichkeiten für das Übersetzungsmodell $P(S|T)$ und das Produkt beider ergibt die Suche argmax nach der besten Übersetzung für die Zielsprache [Bon09], [Lav13].

Die Abbildung 41 zeigt das Prinzip der SBMT. Die nachfolgenden Abschnitte beziehen sich auf diese Abbildung und erläutern deren einzelne Komponenten.

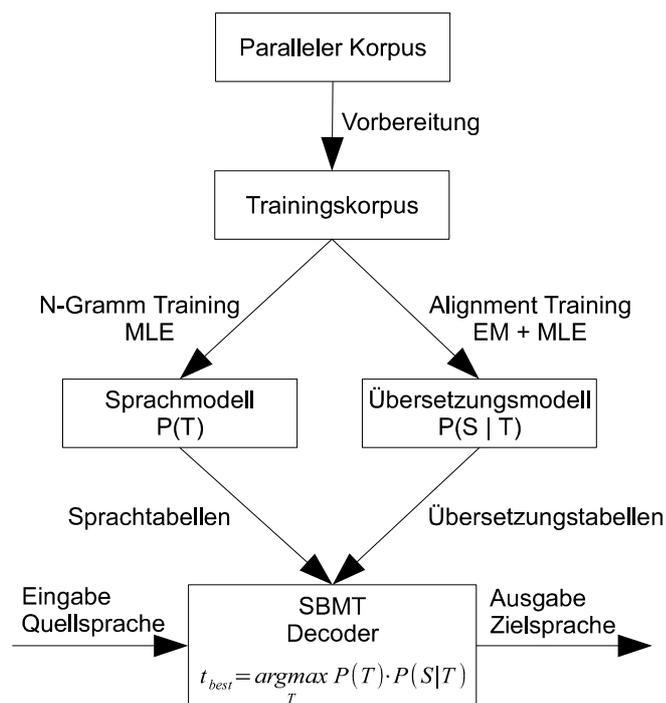


Abb. 41: Funktionsweise einer SBMT

Korpus-Vorbereitung

In der Korpus-Vorbereitung können verschiedene Anpassungen stattfinden, die den Korpus für das Training des Sprachmodells und des Übersetzungsmodells zurechtlegen. Zum einen ist nicht gewährleistet, dass in einem parallelen Korpus die Sätze der Quellsprache mit den Sätzen der Zielsprache übereinstimmen. „Denn nicht immer entspricht der n-te Satz des Quelltextes auch dem n-ten Satz des Zieltexes [Lav13].“ Die richtige Zuordnung nennt man Satzalinierung. Das kann passieren durch die Ermittlung der Satzlängen, über die Bestimmung der Wörter mit einem bilingualen Lexikon oder mit der Verwendung von Ankerpunkten. Ankerpunkte sind zusätzliche Informationen (z. B. Zahlen), die eine Zuordnung zwischen den Wörtern erleichtern.

Zum anderen können je nach Korpus weitere Veränderungen nötig sein. So können Algorithmen störende Zeichen (z. B. „|“, „#“) für die Übersetzung entfernen, Wörter auf ihre Rechtschreibung kontrollieren oder textuelle Anpassungen (z. B. Wörter kleinschreiben, Sätze zeilenweise im Korpus zurechtlegen usw.) erledigen. Ist der gewünschte parallele Korpus erstellt, können die statistischen Modelle mit diesem trainieren [Lav13].

Das Übersetzungsmodell

Die statistische Analyse der Sätze kann auf verschiedenen Ebenen ausgeführt werden. Die ersten Ansätze der SBMT bezogen sich auf die Ebene der Wörter. Aktuelle Modelle nutzen heute „Phrasen-basierte Übersetzungsmodelle“, da diese gegenüber den „Wort-basierten Übersetzungsmodellen“ Ambiguitäten vermeiden können. Beide Modelltypen werden kurz im Folgenden betrachtet.

Wort-basiertes Übersetzungsmodell

Im Jahr 1993 entwickelten die IBM-Mitarbeiter in [Bro93] die Wort-basierten IBM-Modelle 1 bis 5 zur Bestimmung der Parameter des Übersetzungsmodells. Die Idee ist, nicht die Sätze komplett zu übersetzen, sondern die Wörter eines Satzes einzeln zu übersetzen. Die folgenden Betrachtungen beziehen sich speziell auf das IBM-Modell 1. Die Formel 4 berechnet die Übersetzungswahrscheinlichkeit, dass ein Quellsatzes $S = (s_1, \dots, s_{l_s})$ mit der Länge l_s in den Zielsatz $T = (t_1, \dots, t_{l_t})$ mit der Länge l_t übersetzt wird.

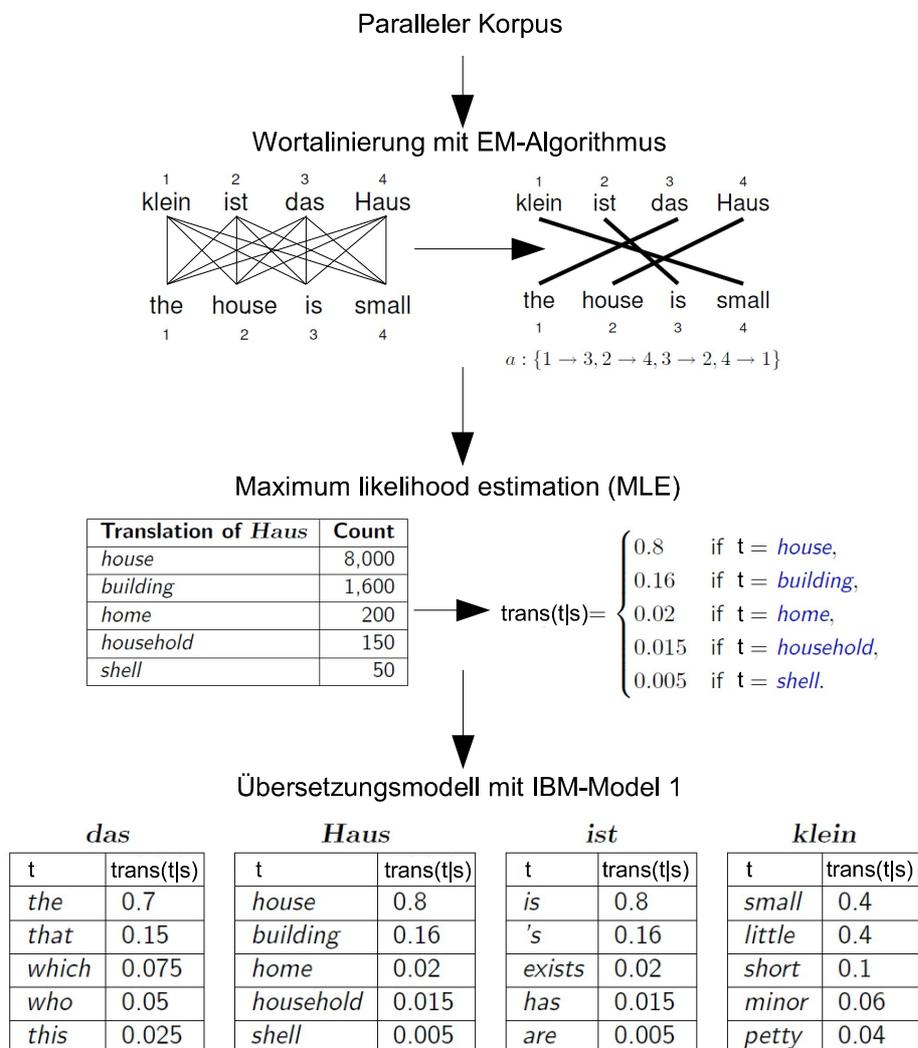
$$P(T, a|S) = \frac{\epsilon}{(l_s + 1)^{l_t}} \cdot \prod_{j=1}^{l_t} l_t \text{trans}(t_j | s_{a(j)}) \quad (4)$$

Die Alinierungsfunktion $a: j \rightarrow i$ ist eine Wort-zu-Wort Alinierung, die jedes Wort der Zielsprache an Position j auf ein Wort der Quellsprache an Position i abbildet. $\text{trans}(t_j | s_{a(j)})$ stellt dabei die lexikalische Übersetzungswahrscheinlichkeit dar, z. B. wie hoch ist die Wahrscheinlichkeit, dass das Wort „das“ mit „the“ übersetzt wird. Der Parameter ϵ ist eine Normierungskonstante.

Um die lexikalische Übersetzungswahrscheinlichkeit zu ermitteln, muss die Wort-zu-Wort Alinierung bekannt sein. Diese lässt sich durch den „Expectation-Maximization (EM)-

Algorithmus“ ermitteln. Der Algorithmus nimmt im 1. Schritt an, dass eine uniforme Wahrscheinlichkeitsverteilung im parallelen Korpus vorherrscht. Jedes Wort der Quellsprache ist mit jedem Wort der Zielsprache verbunden bzw. aligniert. Beim 2. Schritt schaut der Algorithmus, welche Wörter verbunden sind und lernt daraus häufig auftretende Verbindungen, die er im 3. Schritt durch Anpassungen der Wahrscheinlichkeiten präferiert. Durch mehrmalige Durchführung der Schritte 2 und 3 verändern bzw. verfeinern sich die Wahrscheinlichkeiten. Anhand dieser Daten kann die Maximum likelihood estimation (MLE) die Übersetzungswahrscheinlichkeit $trans(t|s)$ bestimmen. Dadurch sind die Parameter für ein Übersetzungsmodell basierend auf IBM-Modell 1 bekannt und können berechnet werden [Lav13].

Abbildung 42 zeigt die Vorgehensweise eines Wort-basierten Übersetzungsmodells in einem Beispiel. Hierbei ist die Quellsprache Deutsch und die Zielsprache Englisch.



$$p(T, a|S) = \frac{\epsilon}{5^4} \cdot trans(the|das) \cdot trans(house|Haus) \cdot trans(is|ist) \cdot trans(small|klein)$$

$$p(T, a|S) = \frac{\epsilon}{625} \cdot 0.7 \cdot 0.8 \cdot 0.8 \cdot 0.4 = \underline{0.00029\epsilon}$$

Ergebnis: Das Haus ist klein. → The house is small.

Abb. 42: Beispiel für ein Wort-basiertes Übersetzungsmodell nach IBM-Modell 1; verändert aus [Koe13]

Das Wort-basierte Übersetzungsmodell basierend auf IBM-Modell 1 lernt somit die Deutsch-Englischen Wortzusammenhänge über die Wahrscheinlichkeiten. Das IBM-Modell 1 beachtet nicht die Wortreihenfolge des Satzes. Das IBM-Modell 2 löst diese Schwäche, indem es die Wörter neu anordnen kann. Im IBM-Modell 3 basiert die Alinierung nicht nur auf Wort-zu-Wort Übersetzung, sondern es beachtet die „Fertilität“ eines Wortes (1:n Alinierungen sind möglich). IBM-Modell 4 und 5 verbessern die Sortierung der Wortreihenfolge und beheben Mängel. Ausführliche Informationen zu den Wort-basierten IBM-Modellen 1 bis 5 befinden sich in [Bro93].

Ein nützliches Tool ist „GIZZA++“ (<http://www.statmt.org/>). Das Tool ist sprachunabhängig (auch verwendbar bei einer Gebärdensprache) und erstellt die Wortalinierung automatisch.

Phrasen-basiertes Übersetzungsmodell

Im späteren Verlauf der Entwicklung der SBMT setzten sich vermehrt Phrasen-basierte Übersetzungsmodelle durch. Diese betrachten nicht einzelne Wörter bei der Übersetzung, sondern übersetzen vollständige Phrasen. Phrasen können aus mehreren Wörtern bestehen und bilden eine Übersetzungseinheit. Bei dem Wort-basierten Übersetzungsmodell kann eine 1:n Alinierung ab IBM-Modell 3 erfolgen. Dies ist nötig, wenn ein Wort in der Quellsprache sich nur durch mehrere Wörter in der Zielsprache übersetzen lässt. Ein Beispiel, das deutsche Wort „zur“ wird ins englische mit den Wörtern „to the“ übersetzt. Jedoch wäre die analoge Richtung (Englisch → Deutsch) mit einem Wort-basierten Modell nicht möglich. Diese n:1 Alinierung erlaubt aber das Phrasen-basierte Modell. Die Phrasen-basierte Alinierung in Abbildung 43 zeigt des Weiteren, dass auch n:m Alinierungen möglich sind. Dadurch können Redewendungen übersetzt und auch einige Ambiguitäten und Divergenzen gelöst werden.

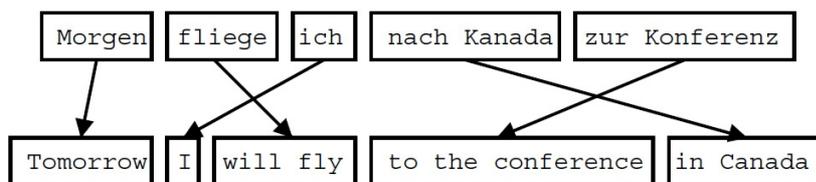


Abb. 43: Phrasen-Alinierung; aus [Koe06]

Ähnlich wie bei den Wort-basierten Modellen findet auch hier der EM-Algorithmus die korrespondierenden Übersetzungen zwischen Quellsprache und Zielsprache. Das Phrasen-basierte Übersetzungsmodell zeigt Formel 5:

$$P(\bar{s}_1^I | \bar{t}_1^I) = \prod_{i=1}^I \varnothing(\bar{s}_i | \bar{t}_i) \quad (5)$$

Ein Qs-Satz S wird dabei in I Phrasen zerlegt. Jede Phrase \bar{s}_i wird in eine zugehörige Phrase \bar{t}_i der Zielsprache übersetzt. Dabei wird eine gleichmäßige Phrasen-Übersetzung-Wahrscheinlichkeit $\varnothing(\bar{s}_i | \bar{t}_i)$ angenommen. Diese Wahrscheinlichkeit berechnet sich mit der Formel 6.

$$\theta(\bar{s}|\bar{t}) = \frac{\text{count}(\bar{t}, \bar{s})}{\sum_{\bar{s}_i} \text{count}(\bar{t}, \bar{s}_i)} \quad (6)$$

Nach der Satzalinierung wird zu jedem Satzpaar eine Anzahl Phrasenpaare extrahiert. $\text{Count}(\bar{t}, \bar{s})$ zählt die Anzahl für ein bestimmtes extrahiertes Phrasenpaar in allen Satzpaaren. Durch die Nutzung der relativen Häufigkeit berechnet sich dann die Phrasen-Übersetzungswahrscheinlichkeit. Die berechneten Wahrscheinlichkeiten speichern Übersetzungstabellen ab [Koe13], [Lav13].

Das Tool „*Phrasal*“ enthält ein Phrasen-basiertes Übersetzungsmodell für eine SBMT. Es ist ebenfalls erhältlich auf <http://www.statmt.org/>.

Das Sprachmodell

Das Sprachmodell lernt aus den Zs-Sätzen eines Trainingskorpus, wie natürlich wohlklingende bzw. wohlgeformte Sätze in der Zielsprache aussehen müssen. Dazu misst es die Wahrscheinlichkeit von regulären Wortfolgen in diesen Sätzen. Ein paralleler Korpus ist nicht zwingend, da nur die Einträge der Zielsprache relevant sind. Deshalb nutzt man oft einen monolingualen Zs-Korpus, der im Gegensatz zum bilingualen Korpus erheblich mehr Worteinträge besitzt und somit für das statistische Training/„Lernen“ genauere Ergebnisse liefern kann. Für die Berechnung kommen N-gramm-basierte Modelle zum Einsatz. „*N-Gramme*“ zerlegen den Zielsatz in Wörter und ermitteln, wie wahrscheinlich auf ein Wort ein bestimmtes Wort folgt. Beim Sprachmodell kommen häufig Trigramm-Modelle zum Einsatz. Dabei stellen zwei Wörter einen Kontext dar und durch Berechnung der Wahrscheinlichkeiten lässt sich vorhersagen, welches Wort auf diese beiden am wahrscheinlichsten folgt. Diese Trigramm-Wahrscheinlichkeit $P(w_3|w_1, w_2)$ lässt sich gemäß der Maximum Likelihood-Methode mit der Formel 7 ausrechnen:

$$P(w_3|w_1, w_2) = \frac{\text{count}(w_1, w_2, w_3)}{\sum_w \text{count}(w_1, w_2, w)} \quad (7)$$

Im Zähler des Quotienten zählt die Funktion „*count*“, wie oft die Wörter w_1 , w_2 und w_3 in dieser Reihenfolge im Trainingskorpus erscheinen und dividiert diese durch die Anzahl der Erscheinungen, dass nach den beiden Wörtern w_1 und w_2 ein beliebiges Wort w folgt. Da aber in einem SBMT-System häufig Sätze eingegeben werden, die nicht im Trainingskorpus enthalten sind, liefert das Übersetzungsmodell einen neuen Zs-Satz, welcher Trigramme enthält, die das Sprachmodell nicht kennt. Somit würden die Wahrscheinlichkeiten auf 0 fallen. Dieser Wert ist nicht sehr hilfreich für das MÜ-System, weil somit nicht überprüft werden kann, ob der Zs-Satz aus dem Übersetzungsmodell auch einen regulären Satz für die Zielsprache darstellt. Durch Vergrößerung des Trainingskorpus kann dieses Problem verringert aber nicht vermieden werden. Als Gegenmaßnahme lassen sich sogenannte „*Smoothing-Verfahren*“ an-

wenden. Einige Methoden für Smoothing sind *Good-Turing-Smoothing*, *Kneser-Ney-Smoothing*, *Add-One-Smoothing* oder *Witten-Bell-Smoothing*. Diese Glättungsverfahren verringern die Wahrscheinlichkeit für bekannte Ergebnisse und verteilen die somit gewonnene Restwahrscheinlichkeit auf unbekannte Trigramme bzw. N-Gramme [Bon09], [Koe13].

Auch hier lassen sich kostenlose Tools im Internet finden, die bei der Erstellung eines Sprachmodells helfen. Das Werkzeug „*SRILM (The SRI Language Modeling Toolkit)*“ auf der Webseite <http://www-speech.sri.com/projects/srilm/> ist ein oft gebräuchliches Programm.

SBMT-Decoder

Der SBMT-Decoder sucht die Übersetzung heraus, die die höchste Wahrscheinlichkeit vorweist. Die Suche wird jedoch nicht über den kompletten Suchraum, d. h. über alle möglichen Übersetzungen ausgeführt, sondern durch Heuristiken begrenzt. Das hat den Vorteil, dass der Rechenaufwand nicht allzu zeitintensiv ausfällt. Jedoch kann man sich nicht sicher sein, dass die gefundene Übersetzung die Beste darstellt, weil nicht alle Möglichkeiten betrachtet und bewertet wurden. Mit diesem Kompromiss lassen sich trotzdem akzeptable Ergebnisse erreichen. Einige Decoding-Algorithmen sind „*Stack Decoding*“, „*Beam Search*“ und „*A*Search*“. Für eine ausführlichere Betrachtung der Dekodierung empfehlen sich die Quellen [Bon09] und [Koe13].

Auf der Webseite <http://www.statmt.org/> finden sich einige Decoder-Tools für die SBMT. Ebenfalls auf dieser Webseite erhältlich, ist das Tool „*Moses*“, welches ein komplettes sprachunabhängiges SBMT-System darstellt.

4.2.2.2 Beispiel-basierte MÜ – example-based machine translation (EBMT)

Die Grundlage der Beispiel-basierten Maschinellen Übersetzung ist ein manuell erstellter bilingualer Korpus. Aus diesem werden Beispielübersetzungen extrahiert und beispielsweise in einer Datenbank abgespeichert. Ein Eingabesatz wird in dieser Datenbank, auf der Seite der Quellspracheinträge, gesucht. Diesen Prozess bezeichnet man als „*Matching*“ und arbeitet nach Prinzipien von „*Information Retrieval*“ Systemen.

In der Matching-Phase gibt es verschiedene Möglichkeiten. Der eingegebene Satz der Quellsprache kann entweder exakt identisch in der Korpus-Repräsentation vorkommen, teilweise in verschiedenen Einträgen in der Repräsentation enthalten sein (vgl. Abbildung 44-rechts) oder eine dritte Möglichkeit ist, dass keine Identität vorliegt. Dann muss rechnerisch ein Annäherungswert zwischen Teilen des Satzes und den quellsprachigen Datenbankeinträgen ermittelt werden. Der Eintrag mit dem geringsten Wert wird anschließend benutzt. Für das Matching können verschiedene Techniken zur Anwendung kommen. Zu diesen können „*Zeichen-basierte*“ und „*Wort-basierte*“ Verfahren gehören. Ein Vertreter der Zeichen-basierten Verfahren ist das „*Pattern-Matching*“. Es ermittelt zwischen Eingabewort und Datenbankeintrag die längste gemeinsame Teilkette, indem die einzelnen Zeichen/Buchstaben miteinander verglichen werden.

„Editierdistanz“, auch bekannt als Levenshtein-Distanz, und „Angle of Similarity“ zählen zu den Wort-basierten Verfahren. Sie berechnen einen Annäherungswert zwischen Eingabewort und den Einträgen der Datenbank. Bei der Editierdistanz zählt ein Algorithmus die Operationen (Einfügen, Löschen und Ersetzen), die durchgeführt werden müssen, um ein Wort bzw. eine Phrase aus der Eingabe in die Zeichenkette der Datenbank zu wandeln. Dabei können Probleme in der Semantik auftauchen. Zum Beispiel „Haus“ und „Maus“ haben eine Editierdistanz von „1“, sind aber semantisch nicht identisch. Die Lösung ist, Wörter in semantischen Netzen zu verbinden, bezeichnet als „Thesaurus“ (z. B. Erstellung von Synonymklassen). Der Angle of Similarity rechnet mit einer trigonometrischen Funktion einen Ähnlichkeitswinkel aus. Wie bei der Editierdistanz kommen dabei die Anzahl der Operationen (Einfügen, Löschen und Ersetzen) zur Anwendung, wobei diese hier gewichtet werden können.

Ist ein Eintrag im Korpus enthalten, weist die Alinierung die entsprechende Übersetzung zur Eingabe hinzu. Dies kann mit statistischen Verfahren funktionieren, durch manuelle Verbindungen oder grammatikalische Strukturen (Verwendung von Syntax-Bäume). In der abschließenden Rekombinationsphase werden die verschiedenen Übersetzungsteile zusammengefügt und als Zielsprache ausgegeben. Sind einzelne Wörter in der Datenbank nicht enthalten, können diese „Lücken“ in der Übersetzung mit einem bilingualen Wörterbuch übersetzt werden [Ver05],[Ver07].

Die Abbildung 44 zeigt die prinzipielle Vorgehensweise für ein EBMT-System.

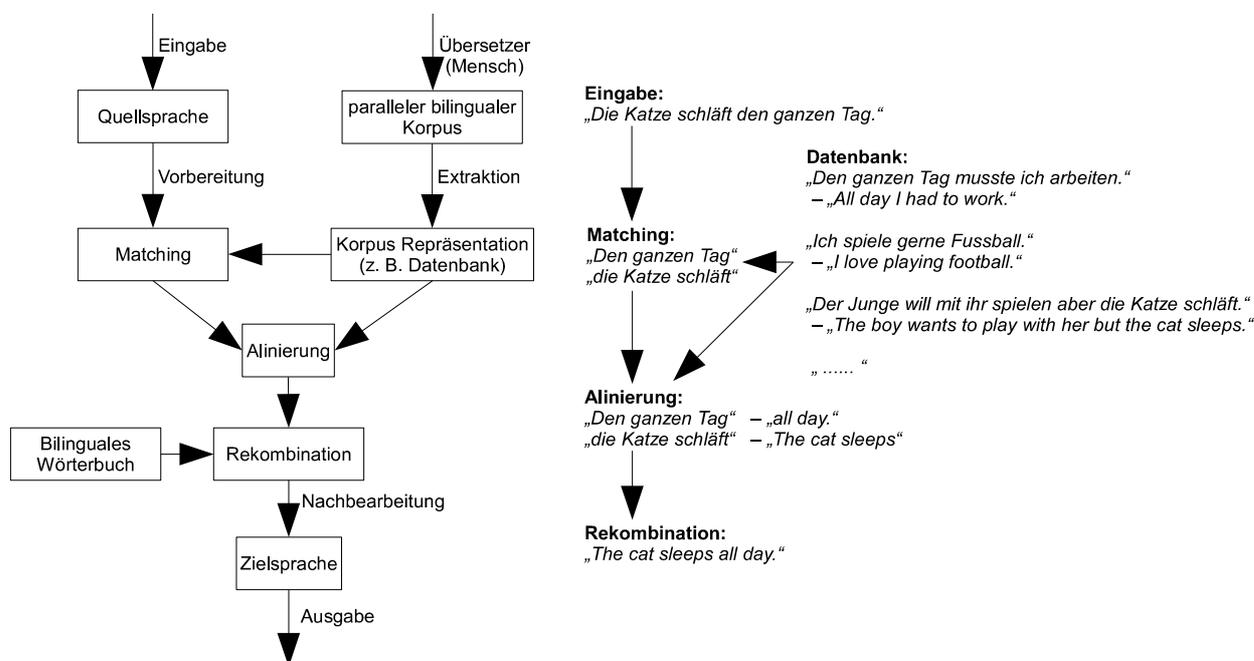


Abb. 44: links: Vorgehensweise EBMT; rechts: Beispiel in der EBMT

4.2.3 Fazit: MÜ-Strategien

Für die Regel-basierten MÜ-Strategien ist es notwendig, dass vor der Übersetzung Linguisten die Sprachen ausführlich analysieren, um mit den gewonnenen Kenntnissen Regeln für den Übersetzungsprozess zu erstellen. Für jede Sprache, welche neu ins Übersetzungssystem aufgenommen werden soll, muss diese aufwendige Prozedur vollzogen werden. Für eine gute Übersetzungsqualität müssen die morphologischen, syntaktischen und semantischen Informationen beachtet werden, um die Ambiguitäten und Divergenzen einer Sprache abzudecken. Ist das der Fall, kann eine Kombination aus Transferregeln und Unifikationsgrammatiken eine Quellsprache grammatikalisch korrekt in eine Zielsprache überführen. Je nach Sprachpaar kann die Anzahl der Regeln abweichen. So gibt es in größeren Übersetzungssystemen eine hohe Anzahl von Regeln. Das verschlechtert nicht nur die Systemübersicht, sondern kann auch Probleme mit neuen Regeln verursachen, die vielleicht bestehende Regeln negativ beeinflussen. Das erschwert die Fehlersuche bei einer Vielzahl von Regeln. Ist dieser Grad an Komplexität im System erreicht, stößt die Regel-basierte MÜ an ihre Grenzen und eine Qualitätssteigerung der Übersetzungen ist nur noch marginal und mit sehr viel Aufwand möglich.

Der große Vorteil der DBMT-Systeme ist, dass sie die Sprache ausgehend von einem Korpus selbstständig analysieren können. Dadurch werden sogar indirekt Regeln einer Sprache beachtet, die in der Linguistik noch nicht bekannt sind. Übersetzungen lassen sich automatisch durch statistische und heuristische Methoden berechnen.

Die EBMT benutzt aus einem parallelen Korpus Übersetzungsbeispiele und speichert diese in einer Datenbank ab. Übersetzungsanfragen werden aus den Einträgen der Datenbank ermittelt und gegebenenfalls neu kombiniert. Diese Strategie kann befriedigende Ergebnisse bei einem kleinen Korpus liefern. Das Problem von Divergenzen und Ambiguitäten lässt sich vermindern, indem die Beispieldatenbank vergrößert wird oder spezielle bzw. häufig verwendete Beispiele in sogenannte Übersetzungsspeicher (*translation memory (TM)*) abgespeichert werden. Neue Übersetzungsbeispiele lassen sich schnell und günstig in das System integrieren und steigern die Übersetzungsqualität. Die EBMT liefert bis zu einer bestimmten Korpusgröße bessere Ergebnisse als die SBMT.

Die Statistik-basierte Maschinelle Übersetzung hat den Vorteil der Sprachunabhängigkeit. Gegenüber den Regel-basierten Verfahren benötigt sie kein linguistisches Wissen. Die SBMT „lernt“ ein Sprachpaar aus einem bilingualen Korpus und kann dieses Wissen basierend auf Wahrscheinlichkeiten für neue Übersetzungen anwenden. Somit können schnell und kostengünstig neue Sprachen ins Übersetzungssystem integriert werden. Die Qualität hängt jedoch stark vom Umfang und der Sprachqualität des Korpus ab. Je mehr Einträge enthalten sind, umso besser sind die Übersetzungsergebnisse. Gegenüber der EBMT hat sie den Vorteil, dass sich neue Übersetzungsanfragen an den SBMT-Decoder flexibler übersetzen lassen, da die EBMT nur strikt Kombinationen aus ihrer Beispieldatenbank zur Übersetzung heranzieht. Divergenzen, Ambiguitäten und Redewendungen lassen sich effektiver und effizienter mit einer

Phrasen-basierten SBMT übersetzen als mit einer Wort-basierten SBMT. Ein Nachteil der SBMT ist die Voraussetzung eines Korpus mit einem großen Datenumfang. Ist dieser nicht vorhanden, kann sie nicht genutzt werden oder ein Korpus muss aufwändig manuell erstellt werden. Des Weiteren kann es problematisch sein, nach dem Training des Sprachmodells und des Übersetzungsmodells neue Korpuseinträge zu tätigen, da das System diese neuen Änderungen wieder mit einberechnen muss. Übersetzungsfehler lassen sich aufgrund des statistischen Ansatzes nur schwer nachvollziehen und sind demnach schwierig zu lösen. Dennoch ist die SBMT aufgrund ihrer Vorteile sehr populär und weit verbreitet. So nutzt das US-amerikanische Unternehmen „Google Inc.“ eine Phrasen-basierte SBMT für ihren „Google Translator“.

Ein neuer Ansatz kombiniert die Strategien der RBMT und der DBMT miteinander. Dieses Verfahren bezeichnet man als „*hybrides MÜ-System*“ und könnte den Sprung zur 4. Generation der MÜ einleiten. Es nutzt die Vorteile der verschiedenen Strategien aus, um ihre einzelnen Schwächen zu umgehen. So kann beispielsweise ein hybrides MÜ-System aus einer EBMT und einer SBMT bestehen und die Schwächen beider Verfahren neutralisieren. Komplizierte Einträge, wie spezielle Redewendungen oder Ambiguitäten, könnten so in einer Beispieldatenbank enthalten sein und die passende Übersetzung parat halten. Neue Übersetzungsanfragen, die in der EBMT-Datenbank nicht enthalten sind, ließen sich mit Methoden der SBMT übersetzen. Grammatikalische Regeln könnten bei der Hinzunahme einer RBMT die grammatikalische Qualität der Übersetzungen weiter steigern. Aktuelle Projekte befassen sich momentan mit der Entwicklung solch hybrider Übersetzungssysteme.

Wie sich diese verschiedenen Maschinellen Übersetzungsstrategien für die Gebärdensprache umsetzen lassen, zeigt der Abschnitt 4.4. Darauf aufbauend folgt im Abschnitt 4.5 jeweils ein Workflow in der RBMT, EBMT und SBMT für die Übersetzung eines Textes in eine animationsorientierte Notation. Zuvor folgt jedoch im nächsten Abschnitt 4.3 Evaluationsmethoden für die Bewertung der Übersetzungsqualität eines Maschinellen Übersetzungssystems.

4.3 Evaluationsmöglichkeiten für die Maschinelle Übersetzung

Eine Evaluation ist nötig, um die Ergebnisse der verschiedenen MÜ-Systeme vergleichen zu können. Dadurch lassen sich Schwachstellen der Systeme entdecken und gegebenenfalls lösen. Anfangs wurden die Systeme manuell vom Menschen bewertet. Dieser Prozess konnte sich über Wochen und Monate hinziehen. Ein weiterer Nachteil ist, dass durch Änderungen im MÜ-System die Bewertungen erneut durchgeführt werden mussten. An dieser Stelle wurden neue Verfahren benötigt, die eine Evaluation schnell, automatisch, billig und wiederholbar ausführen ließen. Zu diesen Verfahren zählen die Bewertungsmetriken WER, PER, TER, SER, BLEU, NIST und METEOR.

Die Gemeinsamkeit dieser Methoden ist, dass ihre Ergebniswerte sich auf eine manuell erstellte Übersetzung des Menschen referenzieren. Das Ziel ist, so nah wie möglich an diese Referenz heranzureichen. Die Ergebnisse der Evaluationsmetriken werden meist in Prozentzahlen angegeben. Je niedriger dieser Wert bei WER, PER, TER und SER ist, desto ähnlicher sind sich die MÜ-Ausgabe und die Referenz. Im Gegensatz dazu zählt ein höherer Wert bei BLEU, NIST und METEOR für eine hohe Übereinstimmung.

4.3.1 WER (word error rate)

Die Wortfehlerrate ermittelt die Anzahl der Fehler im Text, die im Bezug auf einen Referenztext auftraten. Als Referenztext zählt in der Maschinellen Übersetzung der von einem Menschen korrekt übersetzte Zs-Satz aus dem zugehörigen Qs-Satz. Zu den Fehlern gehören die Anzahl der Ersetzungen (*substitutions*) N_{sub} , die Anzahl der Einfügungen (*insertions*) N_{ins} und die Anzahl der Auslassungen (*Deletions*) N_{del} . Das sind die Operationen der sogenannten Levenshtein-Distanz auf Wortebene. Mit der Formel 8 lässt sich die WER berechnen:

$$WER = 100\% \cdot \frac{N_{sub} + N_{ins} + N_{del}}{N_{ref}} \quad (8)$$

Die Formel dividiert die Summe der Fehler der MÜ-Ausgabe durch die Anzahl der Wörter in der Referenz N_{ref} . Das Ergebnis ist ein prozentualer Wert, der auch einen Wert über 100 % haben kann. Dieser Fall tritt ein, wenn die MÜ-Ausgabe durch die Anzahl der Einfügungen N_{ins} im Zähler des Quotienten einen höheren Wert bekommt als die Anzahl der Wörter in der Referenz [Car01], [Bon09], [Sch12].

Das Beispiel 1 zeigt die Berechnung für eine WER-Evaluation. So könnte der englische Qs-Satz „*The cat catches four mice.*“ von der MÜ wie folgt ausgegeben worden sein.

Beispiel 1

Referenz: „*Die Katze fängt vier Mäuse.*“

MÜ-Ausgabe: „*Das Katze * wir die Mäuse.*“ $\rightarrow WER = 100\% \cdot \frac{2+1+1}{5} = \underline{\underline{80\%}}$

Fehler: Sub Del Sub Ins

Die Referenz stellt die korrekte Übersetzung dar. Die Wortfehlerrate liegt in diesem Beispiel bei 80 %.

4.3.2 PER (position-independent word error rate)

Die PER ist eine nicht so restriktive Metrik wie die WER. Bei der Wortfehlerrate WER ist die Reihenfolge der Wörter von Bedeutung. Sind die Wörter vom System richtig übersetzt, stehen aber in einer falschen Anordnung, so gibt die WER einen hohen und somit negativen Wert für die Übersetzung aus. Die PER wertet nur die Korrektheit der übersetzten Wörter und achtet

nicht auf ihre Position in einem Satz. So liegt der PER-Prozentwert in der Nähe vom WER-Wert oder darunter. Eine höhere Prozentzahl ist nicht möglich [Mau08], [Bon09].

Die PER lässt sich durch die Formel 9 berechnen:

$$PER = 100\% \cdot \frac{N_{identisch}}{N_{ref}} \quad (9)$$

PER zählt die Anzahl der übereinstimmenden Wörter $N_{identisch}$ zwischen MÜ-Ausgabe und Referenz und dividiert diese durch die Anzahl der Wörter N_{ref} in der Referenz. Auf das Beispiel 1 von WER bezogen, erhält die PER bei Anwendung der Formel 9 einen Wert von 60 %.

4.3.3 TER (Translation Error Rate)

Die Übersetzungsfehlerrate TER misst die Anzahl der Veränderungen (*edits*), die nötig sind, um die MÜ-Ausgabe in den Referenzsatz umzuwandeln. Die Idee dahinter ist ein Messwert zu bestimmen, der aussagt, wieviel Arbeit nach der Maschinellen Übersetzung aufgebracht werden muss, damit ein Mensch den vom MÜ-System übersetzten Satz korrigiert. Ähnlich wie bei der WER zählen zu den Editierungsoperationen die Anzahl der Ersetzungen N_{sub} , die Anzahl der Einfügungen N_{ins} und die Anzahl der Auslassungen N_{del} , die pro Satz auftreten. Eine zusätzliche Operation ist die Anzahl der Verschiebungen (*shift*) N_{shift} von Wörtern. Dabei können auch Blöcke, bestehend aus Wörtern, verschoben werden und als „eine“ Editierung zählen. Die Anzahl aller Editierungsschritte wird durch die Wortanzahl der Referenz N_{ref} dividiert [Pop07], [Bon09], [Sch12].

Formal lässt sich die TER mit der Formel 10 ausdrücken.

$$TER = 100\% \cdot \frac{N_{sub} + N_{ins} + N_{del} + N_{shift}}{N_{ref}} \quad (10)$$

In dem WER-Beispiel 1 ermittelt die TER ebenfalls den Wert von 80 %. Ändern wir die MÜ-Ausgabe ab, erkennt man in dem Beispiel 2 den Unterschied zwischen den beiden Verfahren.

Beispiel 2

Referenz:	„Die Katze fängt vier Mäuse.“	$TER = 100\% \cdot \frac{2}{5} = \underline{40\%}$
MÜ-Ausgabe:	„Vier Mäuse fängt die Katze.“ →	
Fehler:	Sub Sub Sub Sub	$WER = 100\% \cdot \frac{4}{5} = \underline{80\%}$

Die TER verschiebt in einem Editierungsschritt den Wortblock „Vier Mäuse“ und in einem weiteren Schritt „die Katze“ an die richtige Position und erhält dadurch den Wert von 40 %. Die WER ermittelt bei diesem Beispiel ein Wert von 80 %, da vier Ersetzungen auftreten.

4.3.4 SER (Sentence Error Rate)

Die SER, auch bezeichnet als „*String Error Rate*“, überprüft, ob Sätze der MÜ-Ausgabe identisch mit den Sätzen der Referenzen sind. Dazu vergleicht sie die einzelnen Wörter miteinander. Tritt mindestens ein Unterschied auf oder sind die Zeichenketten unterschiedlich lang, wertet sie die MÜ-Ausgabe als fehlerhaft. Die SER-Ausgabe ist eine Prozentzahl, die angibt, wie hoch der Anteil der falsch übersetzten Sätze liegt. Dieser Wert berechnet sich, indem die als falsch eingestuften Sätze durch die totale Anzahl der getesteten Sätze dividiert wird. Je höher der Wert, desto schlechter ist das MÜ-System auf der Satzebene. Aufgrund dieser restriktiven Bewertung liegt der SER-Wert höher als bei WER, PER und TER. Für das Beispiel 1 aus WER und Beispiel 2 aus TER würde der SER-Wert jeweils bei 100 % liegen [Tom03].

4.3.5 BLEU (Bilingual Evaluation Understudy)

Die BLEU-Metrik wurde 2002 in [Pap02] vorgestellt. Sie berechnet unter Verwendung von N-Grammen einen Ähnlichkeitswert zwischen der MÜ-Ausgabe, die bei [Pap02] als Kandidaten (*Candidates*) bezeichnet werden, und einer oder mehreren Referenzen.

Der BLEU-Algorithmus nimmt ein N-Gramm, wobei „N“ die Anzahl der Wörter darstellt, aus dem Kandidaten und sucht diesen in der Referenz. Das Ziel ist, die Anzahl des gemeinsamen N-Grammes zu bestimmen. Wo dieses N-Gramm aus dem Kandidaten in der Referenz vorkommt, ist nicht wichtig. Somit arbeitet BLEU positionsunabhängig. Je mehr Übereinstimmungen zwischen beiden gefunden werden, umso besser ist die Übersetzungsqualität von der Maschinellen Übersetzung. Die gefundene gemeinsame N-Gramm-Anzahl wird anschließend durch die totale Anzahl der N-Gramme im Kandidaten dividiert und ergibt den Wahrscheinlichkeitswert p_n . Das Beispiel 1 aus [Pap02] zeigt diese geschilderte Wahrscheinlichkeitsberechnung für einen Kandidaten aus der MÜ-Ausgabe und einer Referenz. Als N-Gramm-Modell dient ein Unigramm, d. h. die Vergleiche basierend auf einzelne Wörter.

Beispiel 1

Kandidat: „**the the the the the the the.**“ →
$$p_n = \frac{\text{count}(n\text{-gram})}{\text{count}(n\text{-gram}')} = \frac{7}{7} = 1$$

Referenz: „**The cat is on the mat.**“

Das erste „*the*“ aus dem Kandidat wird in der Referenz gefunden. Das zweite „*the*“ auch und die weiteren „*the*“ ebenfalls. Das ist problematisch, weil die Anzahl der gefundenen gemeinsamen N-Gramme $\text{count}(n\text{-gram})$ dividiert durch die Anzahl der Unigramme im Kandidat $\text{count}(n\text{-gram}')$ ergibt eine Wahrscheinlichkeit p_n von 1. Es ist aber erkenntlich, dass die Referenz und der Kandidat sehr unterschiedlich sind. Deshalb darf ein N-Gramm, in diesem Fall das Unigramm, nur so oft gezählt werden, wie es maximal in der Referenz vorkommt. Diese Restriktion bezeichnet man als „*clipping*“ oder „*modifizierte N-Gramme*“. Somit zählt

im Beispiel 1 das fett gedruckte erste „the“ zum fett gedruckten ersten „the“ in der Referenz. Beim zweiten „the“ aus dem Kandidat tritt dieser Fall ebenfalls für das zweite „the“ in der Referenz auf. Die nachfolgenden „the“-Unigramme im Kandidat lassen sich zu keinem weiteren Unigramm in der Referenz zuweisen. Somit ist $count(n\text{-gram}) = 2$ und die Wahrscheinlichkeit p_n ergibt im Beispiel 1 einen Wert von $2/7$.

Diese modifizierte n-Gramm-Präzision drückt [Pap02] mit folgender Formel 11 aus.

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-gram} \in C} count_{clip}(n\text{-gram})}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-gram}' \in C'} count(n\text{-gram}')} \quad (11)$$

Die Formel 11 sagt aus, dass die Anzahl der abgeschnittenen (*clip*) N-Gramme des Kandidaten, die irgendwo in der Referenz vorkommen, gezählt werden und durch die Anzahl aller N-Gramme im Kandidaten dividiert wird. Abgeschnitten bedeutet die Begrenzung auf die maximale Anzahl des N-Grammes in der Referenz.

Das sogenannte Strafmaß „*brevity penalty (BP)*“ bestraft im Bezug auf die Referenz zu kurz übersetzte Sätze. Sie wird folgendermaßen in Formel 12 definiert:

$$BP = \begin{cases} 1 & , \text{ falls } c > r \\ e^{\frac{1-r}{c}} & , \text{ falls } c \leq r \end{cases} \quad (12)$$

Ist die Anzahl der Wörter des Kandidaten c länger als die Anzahl der Wörter in der Referenz r , wird der BP-Wert auf 1 festgelegt, sonst lässt sich ein Wert zwischen 0 und 1 berechnen.

Sind die Werte aus den Formeln 11 und 12 bekannt, lässt sich der BLEU-Wert mit der Formel 13 ermitteln:

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \cdot \log p_n\right) \quad \text{mit} \quad w_n = \frac{1}{N} \quad (13)$$

Der BLEU-Wert wird häufig bei Daten-basierten MÜ-Strategien verwendet. Je höher der Wert ist, desto näher reicht die MÜ-Ausgabe an die Referenz des Menschen heran. Der Gewichtungsfaktor w_n gleicht die Zusammenfassung mehrerer N-Gramm-Längen aus. Häufig finden 4-Gramm-Modelle (genannt Tetragramm) Anwendung, die einen Gewichtungsfaktor $w_n = 1/4$ nutzen. Ein Beispiel basierend auf ein Uni- und Bigramm-Modell verdeutlicht im Anhang A4 die einzelnen Berechnungsschritte für den BLEU-Wert [Pap02], [Zin13].

4.3.6 NIST (National Institute of Standards and Technology)

Nach der Vorstellung der BLEU-Metrik von IBM beauftragte die DARPA (*Defense Advanced Research Projects Agency*) das „National Institute of Standards and Technology“ eine eigene MÜ-Evaluation zu entwickeln. Die NIST-Metrik in [Geo02] basiert auf der BLEU-Metrik, soll aber einige Verbesserungen im Bezug auf BLEU enthalten. So können in der BLEU-Metrik aufgrund des Strafmaßes BP (*brevity penalty*) ungewollte Effekte auftreten. NIST versucht diese Effekte zu reduzieren. Es enthält die Möglichkeit die N-Gramme im Bezug auf ihre Relevanz/Wichtigkeit zu gewichten. Diese Relevanz wird durch die Vorkommenshäufigkeit in der Referenz bestimmt. Tritt ein N-Gramm sehr häufig auf, fällt die Gewichtung geringer aus, als bei selten auftretenden N-Grammen. So werden im Gegensatz zu BLEU nicht zu kurze Übersetzungen negativ eingestuft. NIST kann ebenfalls wie BLEU mit mehreren Referenzen arbeiten.

Die Formel 14 aus [Geo02] berechnet unter Verwendung von N-Grammen die Informationsgewichtung $Info(w_1...w_n)$:

$$Info(w_1...w_n) = \log_2 \left(\frac{\text{Anzahl der Vorkommen von } w_1...w_{n-1}}{\text{Anzahl der Vorkommen von } w_1...w_n} \right) \quad (14)$$

Der NIST-Wert berechnet sich anschließend mit Formel 15:

$$NIST = \sum_{n=1}^N \left(\frac{\sum_{\text{all } w_1...w_n \text{ that co-occur}} Info(w_1...w_n)}{\sum_{\text{all } w_1...w_n \text{ in sysoutput}} (1)} \right) \cdot \exp \left\{ \beta \cdot \log^2 \left[\min \left(\frac{L_{sys}}{\bar{L}_{ref}}, 1 \right) \right] \right\} \quad \text{mit } N = 5 \quad (15)$$

Der Parameter β gilt hier als Strafmaß für zu kurze Sätze. Dieser bekommt einen Wert von $\beta = 0,5$, wenn die Anzahl der Wörter aus dem Ergebnis der Maschinellen Übersetzung $2/3$ der durchschnittlichen Anzahl der Wörter in der Referenz beträgt. L_{sys} steht für die Anzahl der bewerteten Wörter und \bar{L}_{ref} für die durchschnittliche Anzahl der Wörter in einer Referenz, gemittelt über alle Referenzen. Einen Vergleich zwischen NIST und BLEU ist in [Geo02] zu entnehmen [Geo02].

4.3.7 METEOR (Metric for Evaluation of Translation with Explicit Ordering)

In [Ban05] wird die Evaluationsmetrik METEOR vorgestellt. Sie soll die Schwächen von BLEU und NIST beheben. Zu diesen Schwächen zählt nach [Ban05]:

- das Fehlen von einer Recall-Berechnung, die BLEU durch das Strafmaß BP (*brevity penalty*) versucht zu kompensieren,

- das Nutzen von zu hohen N-Grammen, die eine indirekte Messung der grammatikalisch wohlgeformten Sätze darstellen,
- das Fehlen einer direkten Wortzuordnung zwischen der Ausgabe der MÜ und der Referenz, weil die N-Gramme in BLEU und NIST nur die Anzahl der Vorkommen bestimmen, aber keine direkte Zuordnung besitzen, was bei der Zählung Fehler verursachen kann und
- die Berechnung über das geometrische Mittel bei den N-Grammen.

METEOR nutzt für den Vergleich zwischen der Ausgabe der MÜ und der Referenz eine Wortalinierung. Demnach benutzt es ausschließlich Unigramme. Die Wörter der MÜ-Ausgabe werden mit den Wörtern der Referenz aligniert. Die Zuordnung kann auf drei Ebenen ablaufen. Sind die Wörter aus MÜ-Ausgabe und Referenz identisch, gehören sie zueinander und werden aligniert. Wörter die grammatikalisch nicht identisch sind wie z. B. „Haus“ und „Häuser“ werden auf der nächsten Ebene mit Hilfe des „porter stem“ Moduls gefunden. Der Porter-Stemmer-Algorithmus führt die Wörter in ihre Grund- und Stammform zurück und ermöglicht somit eine neue Zuordnung der Wörter auf ihre Stammform. Die letzte Ebene der Alinierung betrifft die Semantik der Wörter. Das „WN Synonymy“ Modul fasst Wörter in Synonymlisten zusammen. In METEOR können die drei Ebenen in beliebiger Reihenfolge durchlaufen werden. Des Weiteren können neue Ebenen definiert und hinzugefügt werden, was METEOR flexibel macht. Nach dem die Wortalinierung abgeschlossen ist, lassen sich die Unigramm-Precision P und Unigramm-Recall R mit den Formeln 16 und 17 berechnen:

$$P = \frac{m}{w_t} \qquad R = \frac{m}{w_r} \qquad (16), (17)$$

Der Parameter m steht für die Anzahl der gemeinsam gefundenen Wortalinierungen zwischen Referenz und der MÜ-Ausgabe. Je nach Berechnung des Precision-Wertes oder des Recall-Wertes bezeichnet w_t die Anzahl aller Unigramme bzw. Wörter in der MÜ-Ausgabe und w_r die Anzahl aller Unigramme die in der Referenz vorkommen.

Das F-Measure kombiniert die beiden Maße Recall und Precision in einem Wert. F_{mean} definiert sich über eine Gewichtung über das harmonische Mittel aus Recall und Precision. In METEOR berechnet sich F_{mean} über die Formel 18.

$$F_{mean} = \frac{10 \cdot P \cdot R}{R + 9 \cdot P} \qquad (18)$$

Diese Formeln beziehen sich nur auf Unigramme. Alignierte Wörter werden anschließend mit einem „Penalty“-Wert gewichtet. Dieser lässt sich mit der Formel 19 berechnen.

$$Penalty = 0.5 \cdot \left(\frac{\text{Anzahl der chunks}}{\text{Anzahl unigrams}_{\text{matched}}} \right)^3 \qquad (19)$$

Die Wörter, die in der Referenz und in der MÜ-Ausgabe aligniert sind, werden für eine Gruppierung betrachtet. Diese Gruppen bestehen aus Einheiten (*chunks*), welche aus Unigrammen oder höheren N-Grammen zusammengesetzt sind. Alignierte Wörter, die nebeneinander in der MÜ-Ausgabe stehen, müssen auch so nebeneinander in der Referenz vorkommen. Ist dies der Fall, zählt man sie als eine Einheit. Ziel ist es, möglichst wenig Einheiten zu bilden. Das eine Extremum kommt vor, wenn die Zeichenkette der MÜ-Ausgabe identisch mit der Zeichenkette der Referenz ist. Dann besteht die Gruppe aus einer Einheit und der Penalty-Wert fällt gegen 0. Das andere Extremum ist, wenn keine alignierten Wörter nebeneinander stehen, dann zählen alle alignierten Unigramme der MÜ-Ausgabe als einzelne Gruppe/Einheit und erhalten den höchsten Penalty-Wert von 0,5. Die Formel 19 dividiert die „Anzahl der chunks“ durch die Anzahl der enthaltenen Unigramme/Wörter „Anzahl unigrams_{matched}“ in diesen gefundenen Einheiten.

Ist der Penalty-Wert und F_{mean} bekannt, lässt sich mit der Formel 20 der METEOR-Wert berechnen.

$$\text{METEOR} = F_{\text{mean}} \cdot (1 - \text{Penalty}) \quad (20)$$

Die anfangs aufgezählten Probleme wurden mit Beachtung der Recall-Berechnung, ausschließlicher Nutzung von Unigrammen, Zuordnung durch Wortalinierung und die Berechnung über das harmonische Mittel gelöst. Vergleiche zwischen BLEU und METEOR zeigten, dass METEOR eine bessere Korrelation im Bezug auf menschlichen Übersetzungen erreichte, als die Bewertung mit BLEU und NIST. Die Berechnung des METEOR-Wertes wird im Anhang A5 durch ein Beispiel erläutert [Ban05], [Met13].

4.4 Stand der Maschinellen Übersetzung im Bereich der Gebärdensprache

Nach dem die grundlegenden Strategien der MÜ sowie die Bewertungsmethoden erläutert wurden, folgt in diesem Abschnitt eine Übersicht der Maschinellen Übersetzung im Bereich der Gebärdensprache. Tabelle 8 auf der nächsten Seite beinhaltet verschiedene Arbeiten, die sich speziell auf die Übersetzung von schriftlichen Texten in eine Gebärdensprache konzentrierten. Sie wurden den verschiedenen Maschinellen Übersetzungsstrategien zugeteilt. Die MÜ-Strategie der „direkten Übersetzung“ ist in der Tabelle nicht aufgeführt, weil sie als einfache Wort-zu-Wort Übersetzung für die heutigen Ansprüche an einem MÜ-System nicht mehr qualitativ ausreichend ist. Eine Beschreibung der einzelnen Systeme folgt im Anschluss der Tabelle.

MÜ-Strategie	System	Quellsprache → Zielsprache	Quellen	MÜ Evaluation	
Regel-basierte MÜ (rule-based MT (RBMT))	Transfer	Wetter Report	Englisch (Text) → ASL (Text)	[Gri99]	Keine Ergebnisse
		TEAM Projekt	Englisch (Text) → ASL (Avatar)	[Zha00]	Keine Ergebnisse
		ViSiCAST Projekt	Englisch (Text) → NGT, DGS, BSL (Avatar)	[Saf01]	Keine Ergebnisse
		ASL Workbench	Englisch (Text) → ASL (Text)	[Spe01]	Keine Ergebnisse
		SASL-MT System	Englisch (Text) → SASL (Avatar)	[Zij03] [Zij06]	Keine Ergebnisse
		Synennoese Projekt	Griechisch (Text) → GSL (Avatar)	[Kar04]	Keine Ergebnisse
		Text-To-Indian Sign	Englisch (Text) → ISL (Text)	[Das07]	Keine Ergebnisse
		Driver License System	Spanisch (Sprache) → LSE (Avatar)	[San09]	BLEU = 0,6823 NIST = 8,213 SER = 21,45 % PER = 17,24 %
	Interlingua	Zardo System	Englisch (Text/Sprache) → ISL, ASL, JSL (Avatar)	[Vea98]	Keine Ergebnisse
		CP-Generator	Englisch (Text) → ASL (Avatar)	[Hue06]	Keine Ergebnisse
Daten-basierte MÜ (data-driven MT (DBMT))	EBMT	Driver License System	Spanisch (Sprache) → LSE (Avatar)	[San09]	Keine Ergebnisse
		DCU System	Englisch (Text) → ISL (Avatar)	[Smi10]	Keine Ergebnisse
	SBMT	RWTH Stein	Deutsch (Text) → DGS (Avatar)	[Ste06]	WER = 38,9 % PER = 17,1 %
		Minimal Loss Prinzip	Tschechisch (Text) → CSE, SC	[Kan09]	BLEU = 81,29 % NIST = 11,65 SER = 38,15 % WER = 13,06 % PER = 11,64 %
		Driver License System	Spanisch (Sprache) → LSE (Avatar)	[San09]	BLEU = 0,6433 NIST = 7,700 SER = 34,46 % PER = 33,29 %
		Trainslate Projekt	Deutsch (Text) → DSGS (Avatar)	[Ebl12]	Keine Ergebnisse
Hybride MÜ	MATREX System	Englisch (Text/Sprache) → ISL (Avatar)	[Sar08]	BLEU = 39,11 % WER = 45,90 % PER = 34,20 %	
	Driver License System	Spanisch (Sprache) → LSE (Avatar)	[San09]	SER = 8,9 % WER = 4,8 %	

Tab. 8: Übersichtstabelle der Arbeiten

4.4.1 Regel-basierte MÜ-Systeme

Wetter Report System

Angus B. Grieve-Smith entwickelte ein System in [Gri99] zur Übersetzung von englischen Texten in Texte der Amerikanischen Gebärdensprache ASL (*American Sign Language*). Die Notation für die Gebärdensprache ist die „*Literal Orthography*“ nach Don Newkirk, ein Schriftsystem, das lateinische Buchstaben verwendet. Das Übersetzungsprogramm nutzt lexikalische und grammatikalische Regeln für die Translation. Das System besteht aus den vier Komponenten: lexikalischer „Analyzer“, ein Parser für die Ermittlung der Satzstruktur, ein Übersetzungsmodul und ein Generierungsmodul. Der entwickelte Prototyp wurde für die Übersetzung von Wetter Reports genutzt (Bsp.: „wind → so-bles“). Da das Literal Orthography die nicht-manuellen Komponenten einer Gebärdensprache und das Fingeralphabet nicht beinhaltet, wurden zusätzliche Zeichen für die Notation hinzugefügt.

TEAM Projekt

Bei dem TEAM Projekt (*Translation from English to ASL by Machine*) in [Zha00] geht es um die Übersetzung eines englischen Textes in die ASL, welche durch einen Avatar dargestellt wurde. Der englische Text wurde nach einer Analyse der syntaktischen, grammatikalischen und morphologischen Informationen in eine Zwischendarstellung übersetzt. Dazu wählte man eine Glossen-Transkription. Des Weiteren wurden bei der Analyse weitere Parameter, die grammatikalische Informationen wie Satztyp und Mimik beinhalten, bestimmt. Ein Syntax-Baum basierend auf STAG (*Synchronous Tree Adjoining Grammar*) übersetzte den englischen Satz in die Glossen-Darstellung. Die Ausgabe der Avatar-Animation erfolgte mit den Programmen „*Jack Toolkit*“ und „*Jack Visualizer*“. Dieses System war ein Prototyp. Weitere Arbeiten beschäftigten sich mit der entgegengesetzten Übersetzungsrichtung.

ViSiCAST Projekt

In [Saf01] wird die Vorgehensweise für die Übersetzung von englischen Texten in die Niederländische (NGT), Deutsche (DGS) oder Britische Gebärdensprache (BSL) beschrieben. Dieses multilinguale Translationssystem wurde im Verlauf des ViSiCAST Projektes (2000 - 2003) realisiert. In der Abbildung 45 ist die Arbeitsweise des Systems abgebildet.

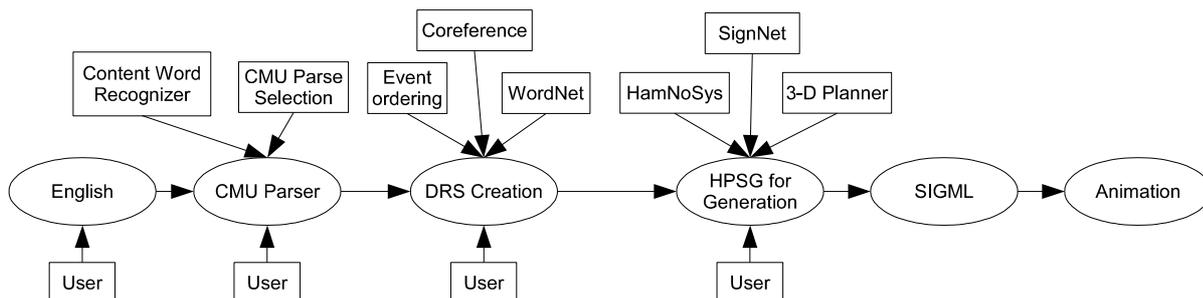


Abb. 45: Übersetzungsworkflow des ViSiCAST Projektes; aus [Saf01]

Der Übersetzungsprozess erfolgt automatisch. Falls die Qualität der Übersetzung unzureichend ist, kann der Nutzer mit dem System an verschiedenen Stellen interagieren. So kann er in der ersten Stufe den eingegebenen englischen Text kontrollieren und gegebenenfalls abändern. Der CMU (*Carnegie Mellon University*) Parser ermittelt die syntaktische Struktur der Eingabe, indem er mit Hilfe eines Wörterbuches Verbindungslinien zwischen den Wörtern der Eingabe und den Wörtern aus dem Wörterbuch erstellt. Diese vorgeschlagenen „Verbindungen“ kann der Nutzer überprüfen und bei Belieben anpassen. Im dritten Schritt erzeugt die DRS (*Discourse Representation Structure*) eine semantische Zwischendarstellung der Eingabe. Dafür nutzt das System eine Liste, die verschiedene linguistische Informationen und Regeln enthält (z. B. Synonyme eines Wortes). Auch hier kann der Nutzer eingreifen und die Darstellung korrigierend manipulieren. Für die Erzeugung der Gebärdenzeichen aus der semantischen Beschreibung (DRS) wird das HPSG (*Head-Driven Phrase Structure Grammar*) Framework genutzt. An dieser Stelle kann der Nutzer Morpheme der Gebärdenzeichen manuell korrigieren (z. B. Änderung der Handform). Diese grammatikalische Beschreibung der Gebärdensprache wird dann in die animationsorientierte Notation SIGML überführt, die für die Ansteuerung der Animation des Avatars dient.

ASL Workbench

Speers entwickelte in seiner Dissertation ([Spe01]) die Software „ASL Workbench“. Er benutzte die lexikalisch-funktionale Grammatik (LFG), um einen englischen Text in die ASL zu übersetzen. Die Ausgabe erfolgte textuell in der Movement-Hold-Modell Notation von Liddell und Johnson. Die LFG repräsentiert jeden Eingabesatz in drei Informationslevel. Die funktionale Struktur (f-Struktur), syntaktische Struktur (c-Struktur) und phonetische Beschreibung (p-Struktur) eines Satzes. Die p-Struktur beinhaltet die nicht-manuellen Komponenten der ASL. Bei der Übersetzung wird zu jedem Level eine LFG-Baumstruktur der Quellsprache erzeugt und durch Transformationsregeln in eine Baumstruktur der Zielsprache umgewandelt. Ein bilinguales Wörterbuch (Englisch : ASL-Glossen) übersetzt die einzelnen Wörter, die abschließend in einem Movement-Hold-Modell ausgegeben werden.

SASL-MT System

In [Zij03] besteht die Quellsprache aus englischen Texten, die in die Südafrikanische Gebärdensprache (SASL) übersetzt und per Avatar ausgegeben werden soll. Das SASL-MT System überprüft die Wörter des Eingabetextes mit einem Wörterbuch der Quellsprache und speichert die Satzstruktur in einem Syntax-Baum ab (XTAG-Parser). Für die Übersetzung in die Zielsprache wird der Quellsprach-Syntax-Baum in ein Zielsprach-Syntax-Baum mit Hilfe eines bilingualen Wörterbuches und definierten Grammatikregeln umgeformt. Der generierte Gebärdensprach-Baum basiert auf das Movement-Hold-Modell von Liddel und Johnson. Diese Notation dient als Eingabesprache für das Grafikmodul, das für die Darstellung des Avatars verantwortlich ist. Im späteren Verlauf des Projektes (siehe [Zij06]) wurde der XTAG-Parser durch ein STAG (*Synchronous Tree Adjoining Grammar*) Parser ersetzt. Anstatt der Liddel

und Johnson Notation kam nun die Glossen-Notation für die Zwischenrepräsentation der Gebärdensprache zum Einsatz.

Synennoese Projekt

Im Synennoese Projekt ([Kar04]) wurde eine pädagogische Lehrplattform für das Unterrichten der Griechischen Gebärdensprache (GSL) erschaffen. Als Quellsprache nutzte man griechische Texte aus Schulbüchern, die nach der Maschinellen Übersetzung ein Avatar in GSL gebärdete. Das System benutzte einen digitalen Korpus (Griechisch : GSL-Glossen), der mit dem ELAN (*EUDICO Linguistic Annotator*) Tool erstellt wurde. Ein Parser nach [Bou00] ermittelt die syntaktische und lexikalische Eingabe. Mit weiteren Tools wird die Glossen-Notation in HamNoSys überführt. Für die Ansteuerung des Avatars wird STEP (*Scripting Technology for Embodied Persona*) aus [Hua02] angewendet. Die Darstellung des Avatars erfolgt mit VRML (*Virtual Reality Modeling Language*) im Webbrowser.

Text-To-Indian Sign

[Das07] beschreibt einen Prototyp für die Übersetzung von englischen Sätzen in die Indische Gebärdensprache ISL (*Indian Sign Language*). Anfänglich wurde ein System nur mit direkter Übersetzung (Wort-zu-Wort Übersetzung) benutzt, was aber aufgrund der unterschiedlichen Satzordnung beider Sprachen eine schlechte Qualität hatte. Man entschied sich gegen ein SBMT-System, da es keinen parallelen Korpus (Englisch : ISL) gab und die Erstellung eines solchen als zu aufwendig eingeschätzt wurde. Somit nutzten die Entwickler für ihr Text-To-ISL MT-System eine Regel-basierte Maschinelle Übersetzung (vgl. Abb. 46).

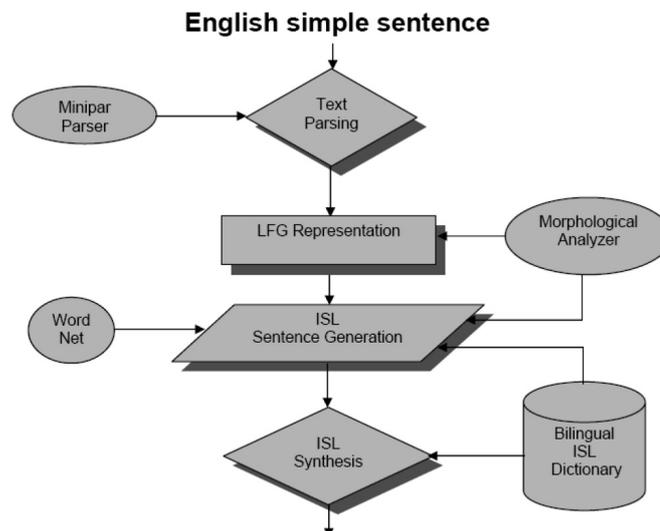


Abb. 46: Architektur des Text-To-Indian Sign Systems; aus [Das07]

Die Voraussetzung vor dem „Parsen“ ist, dass der eingegebene Satz eine einfache Struktur besitzt, d. h. er besitzt nur ein Verb, in ihm kommen keine zeitlichen Ausdrücke vor (z. B. Zeit, Tag oder Datum) und er ist frei von Idiomem oder Metaphern. Deshalb werden diese Sätze vor dem Parsen mit einer „Lookup Tabelle“ (LUT, deutsch: Umsetzungstabelle) verglichen,

welche 350 Zeitausdrücke und Phrasen beinhaltet, die problembehaftete Sätze identifiziert. Der genutzte „*Minipar Parser*“ erzeugt einen Syntax-Baum der Quellsprache und stellt diesen in LFG (Lexikalisch-funktionale Grammatik) in der „f-Struktur“, welche die grammatische Struktur ein Satzes in Attributwertpaare präsentiert, dar. Den Plural von Wörtern ermittelt der „*morphologische Analyser*“. Für die ISL-Erzeugung wird die englische LFG-Struktur durch Transformationsregeln in die ISL-LFG-Struktur konvertiert. Dabei übersetzt ein bilinguales Wörterbuch die englischen Wörter in ISL-Glossen. Die visuelle Ausgabe der Übersetzung erfolgt über zugehörige Gebärdenvideos oder Symbole. Im weiteren Verlauf des Projektes sollen die Glossen in HamNoSys transkribiert und per Avatar ausgegeben werden.

Zardoz System

Das Zardoz System in [Vea98] beschreibt einen theoretischen Ansatz einer modular aufgebauten Systemarchitektur (vgl. Abbildung 47), die englische Sprache oder Texte in den Gebärdensprachen ISL (*Irish Sign Language*), ASL (*American Sign Language*) und JSL (*Japanese Sign Language*) übersetzen soll.

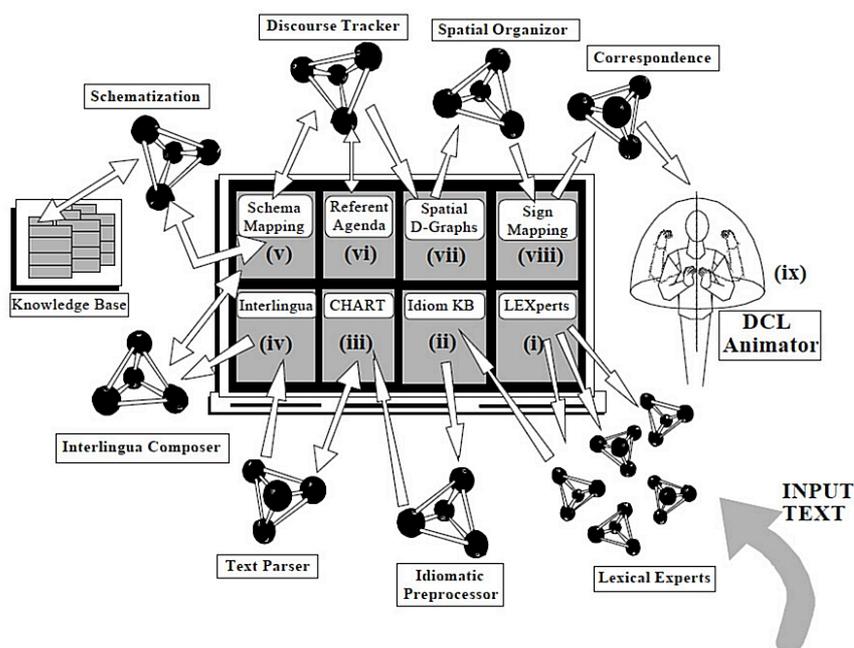


Abb. 47: Zardoz Architektur; aus [Vea98]

KI (Künstliche Intelligenz)-Komponenten, sogenannte „*LEXperts (lexical experts)*“, analysieren durch morphologische Regeln und heuristischen Methoden den eingegebenen Quelltext. Im (ii) Schritt werden die Spracheigentümlichkeiten der Eingabe mit der „*Idiom Knowledge Base*“ verglichen. Nach einer syntaktischen und semantischen Analyse des Satzes mittels eines Textparsers lässt sich die Eingabe in der Interlingua abbilden, die jedoch noch sprachabhängig ist. Die Schritte (v) und (vi) wandeln die Interlingua in eine sprachunabhängige Repräsentation, aus der anschließend die Zielsprache mit den Schritten (vii) und (viii) generiert wird. Das Übersetzungsergebnis wird durch ein DCL (*Doll Control Language*) Animator ausgegeben. Das System ist eine konzeptuelle Idee, welche nicht implementiert wurde.

CP-Generator

Einen weiteren Interlingua Ansatz zeigt Huenerfauth in seiner Dissertation ([Hue06]). Ziel war es, englische Texte in die Amerikanische Gebärdensprache (ASL) zu übersetzen. Die Übersetzung bezog sich auf Gegenstände im Raum, den sogenannte „*classifier predicates (CP)*“. Der entwickelte Prototyp ermittelt anhand syntaktischer Bäume den Inhalt der semantischen Glossen und erzeugt daraus eine hierarchische Beschreibung der CP's, die anschließend grafisch angezeigt wurden. Darstellung 48 zeigt diese grafische Visualisierung für das Beispiel: „*The car parked between the cat and the house [Mat07].*“

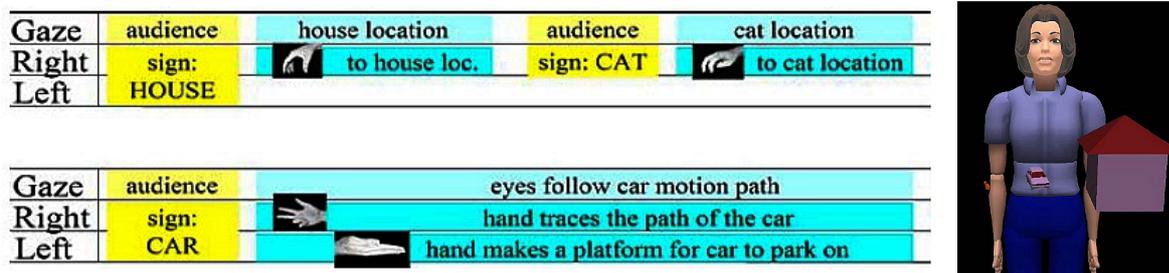


Abb. 48: links: Definition der CP's; rechts: Visualisierung der CP's; aus [Mat07]

Die grafische Zwischendarstellung lässt sich als Interlingua interpretieren. Anschließend wurden die „classifier predicates“ vor den Avatar gelegt, der dadurch die Position der Objekte kennt und die zugehörigen Gebärden im Gebärdenraum ausführen kann. Das System beinhaltet nur eine geringe Anzahl an Glossen und den zugehörigen Animationen. Für die Evaluation des Systems sahen die Nutzer nur den gebärdenden Avatar. Sie sollten in einem Multiple-Choice-Test die richtige Animation zu den gezeigten Gebärden auswählen. Eine ausführliche Beschreibung dieser Evaluation befindet sich im Anhang A6.1.

4.4.2 Daten-basierte MÜ-Systeme

DCU System

In [Smi10] ist eine Beispiel-basierte MÜ abgebildet, welche an der DCU (*Dublin City University*) entwickelt wurde. Es übersetzt englische Texte in die Irische Gebärdensprache (ISL), die anschließend ein Avatar gebärdet. Den Ablauf zeigt die folgende Abbildung 49.

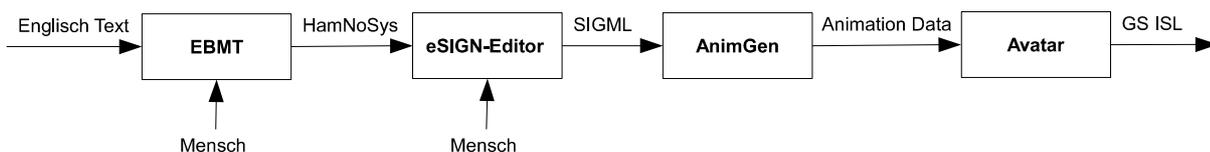


Abb. 49: DCU-Übersetzungssystem; verändert aus [Smi10]

Der Ausgangspunkt des Workflows ist eine Ansammlung von englischen Texten. Die EBMT basiert auf einen parallelen Korpus mit ca. 350 Englischen Sätzen, die manuell in ISL-Sätze übersetzt wurden und im HamNoSys abgebildet sind. Der eingegebene Text wird mit den abgespeicherten Beispielen verglichen und die zugehörige HamNoSys-Notation im bilingualen Korpus herausgesucht. Für den Schritt vom HamNoSys zu SIGML wurde der von der UEA (*University of East Anglia*) entwickelte und frei verfügbare „*eSIGN Editor*“ benutzt. In diesem kann der Benutzer zu einer Eingabe, die zugehörige Glosse und HamNoSys-Transkription manuell auswählen oder erstellen. Das Tool übersetzt dann die ausgewählten Glossen automatisch in SIGML-Code. Die Software „*AnimGEN*“ konvertiert das SIGML-Skript in eine Sequenz von animierten Bildern, den sogenannten „*Animation Data*“, die unter anderem die Position der Kamera, die Skelettstruktur bzw. das Polygonnetz des Avatars und die Bewegungsdaten beinhalten. Den übersetzten Text führt der Avatar anschließend unter Nutzung des „*SIGML Player*“ in Echtzeit aus.

RWTH Stein

Die Diplomarbeit von Daniel Stein ([Ste06]) an der RWTH Aachen handelt über die Entwicklung eines statistischen Maschinellen Übersetzungssystems von deutschen Texten in die Deutsche Gebärdensprache. Die geplante Architektur für die Übersetzung sieht wie folgt aus. Der eingegebene Text wird in eine Glossen-Notation überführt, anschließend in HamNoSys transkribiert und dann von einem Avatar (Virtual Guido aus dem ViSiCAST Projekt) ausgegeben. Der Schwerpunkt der Arbeit bezieht sich auf die bidirektionale Übersetzung zwischen deutschen Texten und Glossen. In der SBMT nutzt [Ste06] für das Sprachmodell N-Gramme und für das Übersetzungsmodell Phrasen-basierte Alinierung (Tool: GIZA++) mit einem bilingualen Korpus (Deutsch : Glossen). Es kamen zwei Korpora zum Einsatz. Der „*Nießen Korpus*“ mit 181 Sätze wurde zum Training der Algorithmen benutzt, da diese Textansammlung einfache grammatikalische Strukturen besitzt. Weil die Größe des Korpus für die Übersetzungsqualität bei der SBMT wichtig ist, wurde ein zweiter Korpus erstellt. Der „*Phoenix Korpus*“ besteht aus den Wetterberichten des Senders Phoenix, die von Experten aufgenommen und manuell in DGS übersetzt wurden. Insgesamt beinhaltet dieser Korpus 1926 Sätze. Ein dritter monolingualer Korpus (deutsche Texte) von der Tagesschau (72724 Sätze) wurde für das Training des Sprachmodells benutzt. Für die morphologische und syntaktische Analyse eines eingegebenen Textes überprüfte der „*gerCG Parser*“ den grammatikalischen Kasus, Numerus und Genus. Das SBMT-System testete man in verschiedenen Experimenten (z. B. mit und ohne Parser). Die Ergebnisse unter Nutzung des Parsers sahen wie folgt aus: Nießen Korpus: PER = 17,1 %, WER = 38,9 % und Phoenix Korpus: PER = 29,7 % , WER = 39,8 %.

Minimal Loss Prinzip

In [Kan09] kommt ein Phrasen-basiertes SBMT-System für die Übersetzung von tschechischen Texten in die Tschechische Gebärdensprache CSE (*Czech Sign Language*) bzw. in die lautsprachbegleitenden Gebärden (LBG) von Tschechien – der SC (*Signed Czech*) – zum Ein-

satz. Um zu einem Wort aus der Quellsprache das passende Wort in der Zielsprache zu finden, nutzen SBMT-Systeme Alinierungstechniken. Für die Erzeugung eines Phrasen-basierten bilingualen Korpus (Tschechisch : CSE) stellt [Kan09] eine alternative Methode zur Alinierung dar, die er als „*Minimal Loss*“ Prinzip bezeichnet. Außerdem wurde eine neue Evaluationsmethode vorgeschlagen. Die „*Semantic Dimension Overlap (SDO)*“ bewertet die semantische Gemeinsamkeit zwischen der Quellsprache und der Zielsprache. Für den Übersetzungsprozess wurde der frei verfügbare Decoder „*MOSES*“ und der eigen implementierte Decoder „*SIMPAD*“ genommen. Die Decoder nutzten für den automatischen Übersetzungsprozess einmal den mit Alinierung trainierten parallelen Korpus und beim zweiten Mal den trainierten Korpus nach dem neuen „Minimal Loss“ Prinzip. Beide Decoder lieferten für beide Strategien ähnliche Evaluationsresultate (siehe [Kan09]). Die besten Ergebniswerte sind in der Übersichtstabelle 8 enthalten.

Trainslate Projekt

Das Projekt Trainslate in [Ebl12] ist ein System, welches momentan an der Universität in Zürich entwickelt wird. Das Statistik-basierte Maschinelle Übersetzungssystem soll in Zukunft Zugansagen (z. B. Verspätungen) der SBB (Schweizerischen Bundesbahn) vom Deutschen in die Deutschschweizerische Gebärdensprache (DSGS) übersetzen. Die Ausgabe erfolgt über eine Smartphone Applikation, die den gebärdenden Avatar und den zugehörigen deutschen Untertitel anzeigt. Das Projekt ist in drei Phasen unterteilt. Diese sind: Erstellen eines parallelen Korpus, Entwicklung der SBMT und Implementierung der Applikation. Momentan befindet man sich in der ersten Phase. Für die Erzeugung des parallelen Korpus (Deutsch : DSGS) wird das DSGS-Lexikon (enthält 9000 Gebärden) von Penny Boyes Braem und ein Korpus bestehend aus Zugansagen von der SBB (enthält 180 Sätze mit 3000 Wörter) benutzt. Der SBB-Korpus wurde gebärdet und auf Video aufgenommen. Die manuelle Transkription der Videogebärden in Glossen und HamNoSys soll im Dezember 2013 abgeschlossen sein. Der Avatar soll dann durch SIGML angesteuert werden.

4.4.3 Hybride MÜ-Systeme

MATREX System

In [Mor08b] wird die Architektur des MATREX (*M*achine *T*Ranslation using *E*Xamples) Systems beschrieben, das durch Beispiel-basierte und Statistik-basierte MÜ englische Sprache in die ISL (*I*rish *S*ign *L*anguage) übersetzt. Der Anwendungsbereich des Systems konzentriert sich auf die Übersetzung von Flughafen-Ansagen. Als Datenbasis nutzte man einen parallelen Korpus. Dieser fundiert auf den ATIS (*A*ir *T*raffic *I*nformation *S*ystem) Korpus, der eine Ansammlung (insgesamt 595 Sätze) von Dialogen im Bereich von Flughafen-Durchsagen darstellt. Die Sätze wurden von Experten gebärdet, auf Video aufgenommen, in Glossen manuell übersetzt und zum parallelen Korpus aufgebaut. Dazu nutzte man das im Internet frei

verfügbare Tool ELAN (*EUDICO Linguistic Annotator*). Das MATREX System bekommt einen englischen Satz als Eingabe. Durch eine Alinierung auf der Satzebene, Satzteilenebene und Wortebene, welches dazu den parallelen Korpus nutzt, wird die passende Übersetzung in Glossen statistisch oder Beispiel-basiert ermittelt. Für die Erstellung der Animationen und eines Avatars kam die Software „Poser“ zum Einsatz. Die grundlegenden Schritte für die Übersetzung von Englisch in ISL zeigt die Darstellung 50.

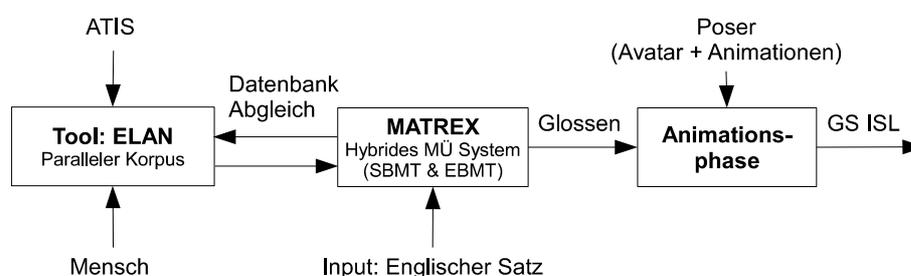


Abb. 50: Vorgehensweise für die Übersetzung

Für die Entwicklung des Systems wurde auf frühere Arbeiten von Sara Morrissey aufgebaut. So beschreibt sie erste Versuche mit einem EBMT-System in [Mor05] und [Mor06], dass aufgrund eines noch nicht vorhandenen ISL-Korpus mit einem Englischen zu NGT-Korpus arbeiten musste. In späteren Arbeiten wurde das ursprünglich auf EBMT basierte Matrex System von der Dublin City University mit dem SBMT-System der RWTH Aachen (siehe [Bau07]) kombiniert. Das beschriebene System in [Mor07] übersetzt Irische bzw. Deutsche Gebärdensprache in die englische bzw. deutsche Sprache. In der Dissertation von Sara Morrissey ([Mor08a]) sind mehrere Experimente der Beispiel-basierten und Statistik-basierten Maschinellen Übersetzung im Bereich der Gebärdensprache aufgeführt. Die untersuchten Übersetzungsrichtungen sind unter anderem: Glossen zu Text, Gebärdensprache zu Text, Glossen zu Sprache, Text zu Glossen und Text zu Gebärdensprache. Aktuelle Arbeiten sollen das System im Bereich der EBMT verbessern. Dazu wird der Korpus mit Unterstützung eines „*Translation Memory*“ ausgebaut.

Driver License System

Das Driver License System in [San09] soll gehörlosen Menschen helfen, ihre Fahrerlizenz in der Führerscheinstelle zu erneuern. Dazu wird die gesprochene spanische Sprache des Angestellten in der Führerscheinstelle automatisch in die für den Gehörlosen verständliche spanische Gebärdensprache LSE (*Lengua de Signos Española*) übersetzt und mit Hilfe eines Avatars dargestellt. Das System besteht aus den drei Komponenten: Spracherkennung, Sprachübersetzung und einen 3D-Avatar-Animationsmodul (vgl. Abb. 51).

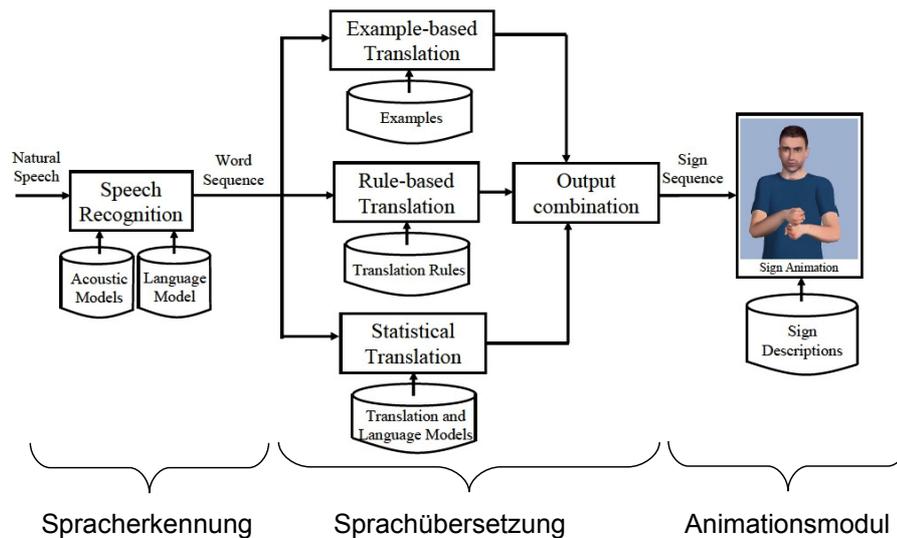


Abb. 51: Architektur des Driver License Systems; verändert aus [San09]

Das Spracherkennungsmodul konvertiert die natürlich gesprochene Sprache in eine textuelle Sequenz von Wörtern. Für die Sprachübersetzung benutzt das System drei verschiedene MÜ Strategien: die RBMT, die EBMT und die SBMT. Als paralleler Korpus (Spanisch : LSE-Glossen) dienen 547 Sätze, die für den Anwendungsfall als relevant empfunden und manuell übersetzt wurden. Die Aufgabe der Sprachübersetzung ist es, die Wortsequenz in eine Abfolge von Gebärdenzeichen, welche repräsentiert werden durch Glossen, zu übersetzen. Für die Regel-basierte Strategie definierten Experten 293 Übersetzungsregeln. Die Beispiel-basierte MÜ ermittelt mit der Levenshtein Distanz (LD) und dem parallelen Korpus die Ähnlichkeit zwischen Quellsprache und Zielsprache. Bei dem statistischen Übersetzungsprozess wurden zwei Methoden evaluiert. Die erste Methode war ein Phrasen-basiertes Übersetzungssystem mit einer Software vom NAACL Workshop (siehe: <http://www.statmt.org>), die den Korpus mit Wortalinierung trainiert und anschließend Wortgruppen (Phrasen) extrahiert. Für den Übersetzungsprozess an sich wurde der Decoder Moses verwendet. In der zweite Methode benutzte man für die Wortalinierung die Software GIZA++ und ein SFST (*Stochastic Finite State Transducer*) Modul, welches mit dem GIATI (*Grammatical Inference and Alignments for Transducer Inference*) Algorithmus arbeitet. Die besten Ergebnisse in der SBMT lieferte die SFST-basierte Methode.

Für diese Strategien wurden separate Experimente und Evaluationen ausgeführt. In der finalen Version des Systems kombiniert das Sprachübersetzungsmodul hierarchisch alle drei Strategien, wobei an erster Stelle die EBMT steht, gefolgt von RBMT und SBMT. Dadurch sollen Schwächen und Probleme der einzelnen Strategien vermieden werden. Zur Ausgabe der Zielsprache wird der Avatar „Virtual Guido“ aus dem eSIGN-Projekt benutzt und durch die animationsorientierte Notation SIGML angesteuert. In einer späteren Arbeit ([Lóp11]) wurden die beiden Methoden der SBMT weiter verbessert und erreichten für die Phrasen-basierte SBMT-Methode einen BLEU-Wert von 81,00 % sowie für die SFST-basierte Methode einen BLEU-Wert von 78,40 %.

4.5 MÜ-Strategien für die Übersetzung von Texten in eine animationsorientierte Notation

Für eine barrierefreie Kommunikation zwischen einem Hörenden und Gehörlosen muss das MÜ-System in der Lage sein, Sprache oder Texte aus der Lautsprache in eine animationsorientierte Notation der Gebärdensprache zu übersetzen. Systeme, die Sprache erfassen und textuell ausgeben, sind bereits vorhanden und liefern eine gute Qualität. Somit stellt sich die Frage: Wie können Texte der Lautsprache in eine GS-Notation automatisch überführt werden, um somit einen Avatar für die GS-Ausgabe anzusteuern? Ausgehend von den gewonnenen Kenntnissen der MÜ-Strategien aus Abschnitt 4.2 und den existierenden Systemen aus Abschnitt 4.4 zeigt dieses Teilkapitel für die Strategien RBMT, EBMT und SBMT einen möglichen Ablauf für die Maschinelle Übersetzung im Bereich der Gebärdensprache.

4.5.1 RBMT-Workflow

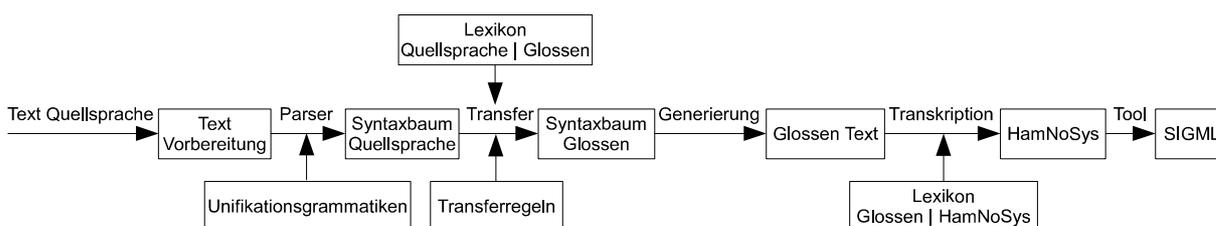


Abb. 52: RBMT von einem Text in eine animationsorientierte GS-Notation

Die Abbildung 52 stellt einen möglichen Ablauf basierend auf ein Regel-basiertes Maschinelles Übersetzungssystem dar. Ein Text aus der Lautsprache sollte vor der automatischen Übersetzung an das System angepasst werden. In dieser Vorbereitungsphase können zum Beispiel Artikel, Kopulae (z. B. sein, haben) oder Wörter, die nicht in der Gebärdensprache vorkommen, herausgefiltert oder im Fall der Wörter durch Synonyme, die auch in der GS existieren, ersetzt werden. Diese textuellen Anpassungen, zu denen auch die Überprüfung der Rechtschreibung oder Zerlegung komplexer Sätze in einfach strukturierte Sätze zählen kann, könnte ein Textparser durchführen. Ist der Ausgangstext aufbereitet, überführt ein weiterer Parser den Eingangssatz in seine syntaktische Baumstruktur. Benutzt der Parser eine Unifikationsgrammatik (z. B. ein LFG-Parser) erkennt er nicht nur die Syntax des Eingabesatzes, sondern auch die enthaltenen grammatikalischen Informationen wie Tempus, Kasus, Genus, Numerus oder Person. Diese zusätzlichen Informationen werden in einer Attribut-Wert-Matrix zusammen mit dem Syntaxbaum gespeichert. Transferregeln überführen die Syntax und Grammatik der Baumstruktur in die entsprechende syntaktische und grammatikalische Struktur der Gebärdensprache. Als Zwischenrepräsentation für die GS kommen die leicht verständlichen Glossen zur Anwendung. Während der Transformation der syntaktischen Struktur in die Zielsprache übersetzt ein Lexikon die Wörter, die in den Blättern

des Baumes enthalten sind, in Glossen. Dabei nutzt das Lexikon die Attribut-Wert-Matrix (Merkmalstruktur des Satzes), um zu den entsprechenden grammatikalischen Informationen die zugehörige Übersetzung zu finden. Das Ergebnis ist die Übersetzung der Quellsprache in einen Glossen-Satz der Gebärdensprache. Dieser Satz wird anschließend mit einem weiteren Lexikon in das HamNoSys transkribiert. Eine direkte Übersetzung von der Quellsprache in HamNoSys wäre auch möglich, aber der Zwischenschritt über Glossen ergibt den Vorteil, dass sich die Übersetzung in die GS-Struktur leichter kontrollieren lässt, da HamNoSys kryptischer und schwieriger zu lesen ist, als die Glossen-Notation. Die linguistische HamNoSys-Notation wurde gewählt, weil sie nicht nur etabliert ist und schon seit Jahren stetig weiter entwickelt wird, auch im Bereich der wichtigen nicht-manuellen Komponenten einer Gebärdensprache, sondern auch die Möglichkeit besitzt, andere Gebärdensprachen neben der DGS darzustellen. Ein weiterer Vorteil von der Notation ist, dass sie sich durch Tools (z. B. eSign Editor) in die animationsorientierte Notation SIGML umwandeln lässt, die anschließend einen Avatar steuert, der die Ausgabe des Quelltextes visuell in der Gebärdensprache darstellt.

Der Nachteil des gezeigten Workflows ist der hohe Vorbereitungsaufwand. So müssen die Lexika manuell erstellt, Transformationsregeln definiert und das Tool kompatibel zum MÜ-System implementiert werden.

4.5.2 EBM-Workflow

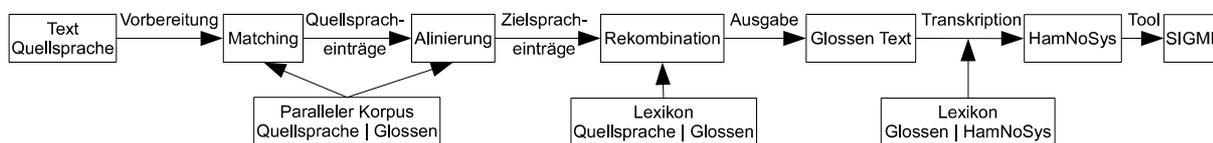


Abb. 53: EBM von einem Text in eine animationsorientierte GS-Notation

In dem Beispiel-basierten Maschinellen Übersetzungssystem in Darstellung 53 können die einzelnen Teile der Quelltext-Eingabe in einem parallelen bilingualen Korpus durch Matching-Verfahren gesucht werden. Der parallele Korpus, der aus einer Ansammlung von Texten besteht, könnte in Form einer Datenbank die enthaltenen Sätze in einzelne Einträge abspeichern. Je mehr Informationen (z. B. grammatikalische bzw. syntaktische) und Übersetzungsbeispiele in der Datenbank enthalten sind, um so leichter fällt die Übersetzung aus. Um die Suche zu erleichtern, wird die Quellsprache an den parallelen Korpus in einer Vorbereitungsphase angepasst. So ließen sich Sätze der Quellsprache mit Hilfe von Information Retrieval Techniken schneller finden. Das Matching sucht die Quelltext-Anfrage in der Datenbank des parallelen Korpus und gibt die entsprechenden Suchergebnisse auf Seite der Quellsprache im Korpus zur nächsten Phase weiter. Stehen Wörter nicht im Korpus, wird ein Annäherungswert berechnet und das Wort aus dem Korpus mit dem höchsten Wert ersetzt das fehlende Wort. Die Alinierung findet zu den Quellspracheinträgen in der Datenbank die zugehörigen Übersetzungen in der Glossen-Notation. Ist bei einem Wort kein akzeptabler Annähe-

rungswert berechnet worden, kann durch Nachschlagen in einem Wörterbuch die passende Glosse zum fehlenden Wort gefunden werden. Die Generierung des Textes aus den verschiedenen gefundenen Sucheinträgen in die Zielsprache übernimmt die Rekombinationsphase. Sie benutzt die zusätzlich grammatikalischen Informationen aus den gefundenen Einträgen für den korrekten Satzbau der Zielsprache. Die weiteren Schritte von der Glossen-Darstellung in die animationsorientierte Notation SIGML geschieht auf den selben Weg wie bei der RBMT.

4.5.3 SBMT-Workflow

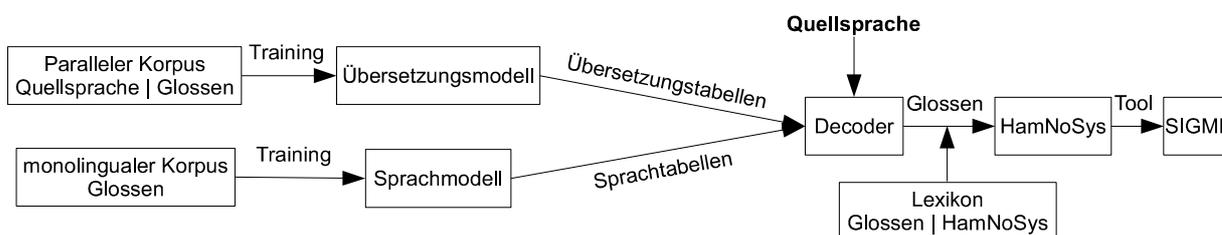


Abb. 54: SBMT von einem Text in eine animationsorientierte GS-Notation

Die Statistik-basierte Maschinelle Übersetzung in Abbildung 54 benötigt ebenfalls wie die EBMT einen parallelen Korpus. Diesen nutzt sie, um für das Übersetzungsmodell die Übersetzungswahrscheinlichkeiten zwischen den Einträgen der Quellsprache und den Glossen zu berechnen und somit zusammengehörige Einträge zu „erlernen“. Eine Phrasen-basierte Alinierung für das Training des Übersetzungsmodells minimiert die Übersetzungsprobleme (Divergenzen und Ambiguitäten). Der bilinguale Korpus sollte für das Training vorbereitet sein. Die Sätze aus den Texten stehen in separaten Zeilen und sind mit den zugehörigen Satzübersetzungen aligniert. Das Sprachmodell könnte ebenfalls diesen parallelen Korpus zum Training benutzen. Jedoch kann ein separater monolingualer Korpus der Zielsprache das Training begünstigen, indem er wesentlich mehr Einträge enthält als der bilinguale Korpus. Durch Nutzung von N-Grammen ermittelt das Sprachmodell Wörter, die in der Zielsprache häufig aufeinander folgen. Dadurch „lernt“ es, wie eine wohlklingende grammatikalisch korrekte Äußerung der Zielsprache aussieht bzw. funktioniert. Die Quellsprache wird in dem Decoder der SBMT eingegeben und dieser ermittelt die höchste Wahrscheinlichkeit für eine gute Phrasenübersetzung aus den Übersetzungstabellen und einer wohlgeformten Äußerung aus den Sprachtabellen für die Generierung der Zielsprache. Das Ergebnis ist die Übersetzung der Eingabesprache in die Glossen-Notation, die zur Ansteuerung eines Gebärdensprach-avatars wiederum in SIGML transkribiert wird.

Die Voraussetzung für die EBMT und SBMT ist ein vernünftig übersetzter paralleler Korpus. Eine Übersicht über bereits vorhandene Gebärdensprach-Korpora ist auf der Webseite <http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/gs-korpora.html> der Universität

Hamburg erhältlich (Stand: 2012). Für die Deutsche Gebärdensprache gibt es nur eine sehr geringe Anzahl von Korpora. Die Universität Hamburg will bis zum Jahr 2020 einen digitalen bilingualen Korpus aufbauen, der Einträge der DGS und der deutschen Lautsprache enthält. Erst ein paralleler Korpus mit einer hohen Anzahl von Einträgen macht die SBMT attraktiv für die automatische Übersetzung. Bei den heutigen kleinen GS-Korpora empfiehlt sich daher die RBMT oder EBMT. Vielversprechend ist die Kombination der verschiedenen Systeme in einem hybriden MÜ-System. Diese Systeme erlauben zwar die Nachteile der einzelnen MÜ-Strategien entgegenzuwirken, stehen aber im Bereich der Gebärdensprache noch am Anfang.

In den erstellten Workflows wurde als Zwischenrepräsentation die Glossen-Notation gewählt. Nach der Generierung der Glossen ist eine Bewertung der Übersetzungsqualität sinnvoll. So können mit den vorgestellten Evaluationsmethoden aus Abschnitt 4.3 die Korrektheit der Übersetzung überprüft und mögliche Übersetzungsfehler vor der Gebärdendarstellung per Avatar ausgeschlossen werden.

Die Erzeugung und Darstellung solcher Gebärdensprachavatare zeigt das nachfolgende 5. Kapitel. Aus diesen Kenntnissen lassen sich dann die weiteren Verarbeitungsschritte für die Umsetzung von Texten in eine Gebärdensprachsynthese ableiten.

5. Kapitel: Avatarsysteme

Dieses Kapitel beschäftigt sich mit der Erzeugung eines Avatar-Modells, den Anwendungsgebieten und den Bewertungsmöglichkeiten von Avatarsystemen. Der erste Abschnitt 5.1 zeigt die verschiedenen Techniken bzw. Methoden die für die Modellierung und die Animation eines Avatar-Modells in Frage kommen. Am Ende des ersten Teilkapitels werden die Erkenntnisse zum Design eines Gebärdensprachavatars zusammengefasst. Abschnitt 5.2 behandelt die generellen Anwendungsbereiche für Avatare und bietet darauf aufbauend eine Übersicht für allgemeine Avatarsysteme und spezielle Avatarsysteme im Bereich der Gebärdensprache. Abschließend gibt der Abschnitt 5.3 Methoden, Ergebnisse und Erkenntnisse aus vorhandenen Evaluationen wieder.

5.1 Design des Avatars

Für die Generierung einer animierten Figur benötigt der 3D-Artist Kenntnisse aus dem Bereich der Modellierung und der Animation. Die Abbildung 55 zeigt den Ablauf zur Erstellung einer animierten Szene.



Abb. 55: Animation Pipeline; aus [Cun12]

Die Preproduction (Vorproduktionsphase) definiert die Idee bzw. das Konzept der Animation. Hier wird die Planung des Projektes, zu der die Budgetplanung, Zeitpläne, Ressourcen, Storyboard inbegriffen sind, festgelegt. In der Objektmodellierung entsteht das geometrische Modell für die geplanten Objekte. Anschließend werden in der Szenen-Modellierung die erstellten Objekte zueinander positioniert. Hintergrund, Beleuchtung, Umgebungen (Nebel, Dunst etc.) und der Blickwinkel der Kamera sind ebenfalls festzulegen. Für die Belebung der Szene bzw. Objekte können verschiedene Techniken der Animation auftreten. In der Rendering Phase wird die erstellte Geometrie- und Merkmalsrepräsentation in die Pixelrepräsentation umgesetzt. Abschließend ermöglicht die Postproduction Phase diverse Nachbearbeitungen. Zu diesen zählen die finalen Einstellungen der Rendering Parameter (Auflösung, Bildrate), das Abspeichern der animierten Szene, Formatumwandlungen, Vertonung und Videoschnitte [Cun12].

Nach der Betrachtung der allgemeinen Animationspipeline vereinfachen und ändern wir diese auf den konkreten Anwendungsfall der Avatargenerierung ab und erhalten folgende Arbeitsphasen (vgl. Abbildung 56), die im Verlauf des Kapitels näher untersucht werden.

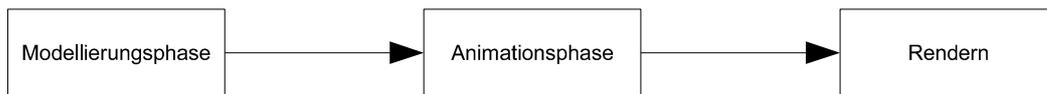


Abb. 56: Übersicht: Avatargenerierung

5.1.1 Modellierungsphase

Das Ergebnis der Modellierungsphase ist eine geometrische Abbildung des Avatars, die anschließend vollständig animiert werden kann. Dazu gibt es verschiedene Möglichkeiten bzw. Methoden Geometrien darzustellen, zu erzeugen und anzupassen.

5.1.1.1 Darstellungsarten der Geometrie

Dieser Abschnitt erläutert einige Möglichkeiten zur Darstellung von Geometrie durch Drahtgittermodelle (*wireframe modeling*), die das Aussehen eines Objektes kennzeichnen. Dazu gehören Polygonflächen, Splines, Parametrische Flächen, Unterteilungsflächen und Metaballs.

Polygonflächen

Ein Polygon, auch als Vieleck bezeichnet, besteht mindestens aus 3 Eckpunkten (genannt *Vertices*), die mit Kanten verknüpft sind. Häufig treten in der Modellierung von Objekten Drei- oder Vierecke auf (vgl. Abbildung 57 (links)). Diese werden miteinander verbunden und bilden somit ein Polygonnetz, auch als *Mesh* bezeichnet, das die Oberfläche des Objektes bildet. Je mehr Polygone benutzt werden, um so komplexere und genauere Formen lassen sich modellieren. Nachteilig zu bewerten ist der Aufwand dieser Geometrieart. Bei Rundungen/Kurven in Objekten oder sehr komplexen Formen ist eine Vielzahl von Polygonen nötig, um eine akzeptable Repräsentation des zu modellierenden Objektes zu erreichen. Des Weiteren spielt der Blickwinkel auf das Objekt eine Rolle. So kann ein Gegenstand mit wenig Polygonen von Weiten sehr gut aussehen, aber beim Heranzoomen verschlechtert sich die Qualität, da die einzelnen Polygone bei Rundungen eckig und kantig wirken (vgl. Abbildung 57 (rechts)). Somit ist die Auflösung des Objektes abhängig von der Größe und der Anzahl der Polygone [Tüm07]. Da diese Darstellungsform einer der ersten grundlegenden Techniken war, werden alle anderen Modellierungsformen früher oder später in einer Polygondarstellung umgewandelt, weil die meisten Animationssysteme für das Rendern Polygone voraussetzen [Mae99].

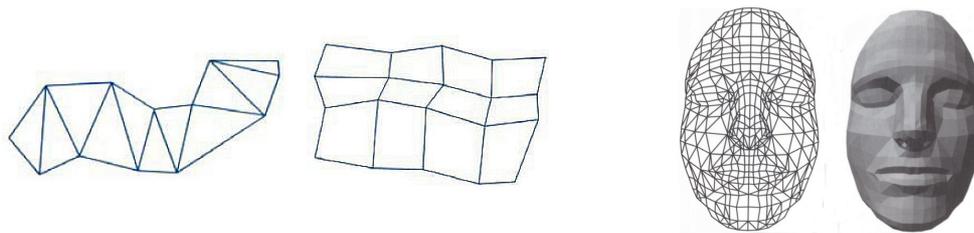


Abb. 57: links: Polygonnetz bestehend aus Dreiecke oder Vierecke; aus [Cun12]
rechts: Gesicht mit zu wenig Polygone wirkt kantig; aus [Flü08]

Splines

Eine Spline-basierte Modellierung besteht aus einer Vielzahl von kurvenförmige Linien, die durch Kontrollpunkte bestimmt werden. Gegenüber der Polygon-basierten Modellierung haben die Splines aufgrund ihrer mathematischen Definition den Vorteil, das sie von der Auflösung unabhängig sind. Beim Heranzoomen an das Modell wirkt die geometrische Struktur unverändert weich und natürlich. Eine Kantenbildung ist nicht ersichtlich, egal aus welcher Distanz ein Objekt abgebildet ist. Aufgrund dieser Eigenschaft benutzt man diese Darstellungsform häufig für organische Formen wie zum Beispiel für die Gestaltung des menschlichen Gesichtes. In der Mathematik gibt es verschiedene Arten von Splines. Für die Modellierung sind nur die B-Splines (*Basis-Spline*), Bézier-Splines und NURBS (*Non-Uniform Rational B-Splines*) relevant und geeignet. Der Unterschied zwischen diesen Linientypen zeigt Abb. 58.

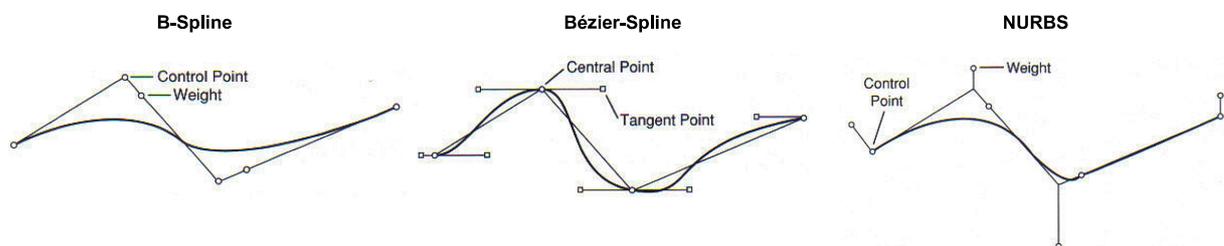


Abb. 58: Spline-Arten für die Modellierung; aus [Gia97]

Der Verlauf der Spline-Linien ist in allen 3 Arten gleich. Ihre Form wird definiert durch die Kontrollpunkte und dessen Gewichtungen (siehe Kontrolllinie). Beim B-Spline sind alle Kontrollpunkte gleich gewichtet und halten somit die kurvenförmige Linie in ihrer Balance. Die Kontrollpunkte liegen direkt über den zu verändernden Bogen der Linie. Dies ist ebenfalls bei dem Bézier-Spline zu erkennen. Zusätzlich findet man hier Tangentenpunkte, die es erlauben den Bogen, ohne das der Kontrollpunkt bzw. Zentralpunkt seine Position verliert, seitlich zu verschieben. Diese Tangentenpunkte sind voneinander unabhängig und erlauben eine höhere Kontrolle über den Verlauf der Kurve. Bei den NURBS befinden sich die Kontrollpunkte neben den Kurven, die sie beeinflussen. Jedoch befinden sich die Gewichtungspunkte über diesen Bereich und sind separat einstellbar. Für einen gewissen Abschnitt der Spline-Linie gibt es eine feste Anzahl von kontrollierenden Punkten [Gia97].

Parametrische Flächen (Parametric Surfaces)

Die Parametrischen Flächen bestehen aus Spline-Kurven. Der Vorteil ist, dass die Kurven wenige Kontrollpunkte besitzen, mit denen das Aussehen des Modells schnell und einfach definiert bzw. editiert werden kann. Für die Darstellung eines Objektes nutzt man mehrere aneinander gesetzte parametrische Flächen. Aufgrund der verschiedenen Spline-Arten existieren dementsprechend unterschiedliche parametrische Flächen. Im Bereich der Computergrafik finden die NURBS Anwendung. Problematisch kann das Zusammenfügen der einzelnen Flächen sein, da die Grenzen der Flächen exakt zueinander passen müssen. Weiterhin muss vor dem Rendern das Flächennetz in eine Polygonstruktur konvertiert werden, da ein direktes Rendern mit parametrischen Flächen nicht möglich ist [Tüm07].

Unterteilungsflächen (Subdivision Surfaces)

Die Unterteilungsflächen ermöglichen ein Drahtgittermodell weiter zu verfeinern. Dazu werden in dem bestehenden Netz weitere Eckpunkte durch Approximation oder Interpolation hinzugefügt. Die Abbildung 59 zeigt, wie dadurch Oberflächenabschnitte feiner aufgelöst und glatter wirken. Bei der approximierenden Methode ergänzen weitere Stützpunkte das Netz und verschieben eventuell vorhandene Punkte. Im Gegenteil dazu ermöglicht die interpolierende Methode das Hinzufügen neuer Stützpunkte ohne das Verschieben bestehender Punkte. Dadurch können schnelle und einfache Anpassungen an dem geometrischen Netz ausgeübt werden. Die Unterteilungsflächen sind für Polygon-basierte und Spline-basierte Modelle anwendbar [Tüm07].

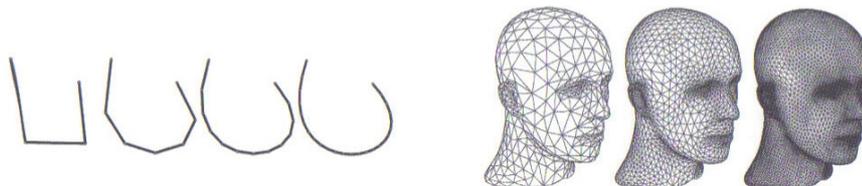


Abb. 59: links: Zunehmende Verfeinerung durch Unterteilungsflächen
rechts: Unterteilungsflächen eines Kopfes; aus [Flü08]

Metaballs

Die Metaballs, auch in der Literatur als Blobbies bezeichnet, bestehen aus grundlegenden Formen wie Kugel, Zylinder, Fläche, Ellipsoid oder Würfel. In der Regel sind sie Kugeln. Werden mindestens zwei Metaballs aneinander geführt, so verschmelzen diese miteinander und bilden ein Objekt. Den Algorithmus, der je nach Software leichte Abänderungen aufweist, entwickelte in den frühen 1980iger Jahren Jim Blinn. Der Algorithmus erzeugt eine dehnbare Oberfläche der Objekte.

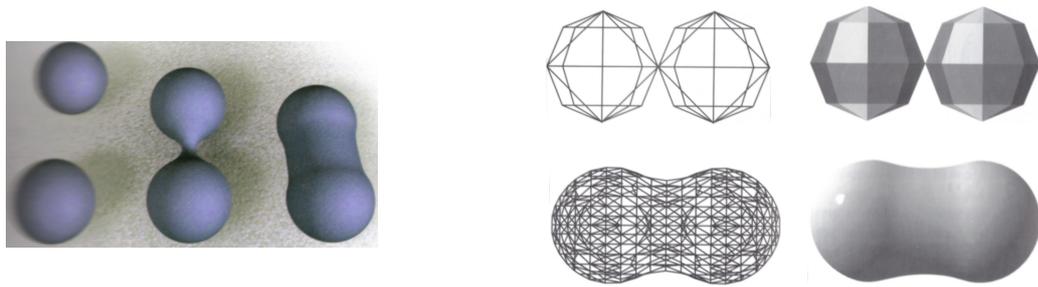


Abb. 60: links: Verschmelzung von zwei Metaballs; aus [Mae99]
rechts: Polygonnetz als dehnbare Oberfläche; aus [Rat98]

Um den Grad der Fusion zu beeinträchtigen, kann man zu jedem Metaball einen numerischen Wert festlegen. Besitzen beide Metaballs den Wert 5, so sieht ihre Form einer Sanduhr ähnlich (vgl. Darstellung 60 (links - mittlere Metaballs)). Liegt der Wert bei 1, dann ändert sich ihre Gestalt zu einem kapselförmigen Aussehen (vgl. Darstellung 60 (links - rechte Metaballs)). Demnach wird bei einem geringeren Gewichtungswert eine höhere Verschmelzung erreicht, als bei einem größeren Wert. Negative Werte sind auch möglich und bewirken, dass ein Metaball vom anderen Metaball subtrahiert wird. Eine andere Transformierungsmöglichkeit (z. B. Strecken, Extrudieren) der Metaballs ist nicht vorhanden. Nach dem die Metaballs verschmolzen sind, lässt sich ein Polygonnetz, welches in der Darstellung 60 (rechts) zu sehen ist, um die äußere Form spannen und ermöglicht somit einen homogenen weichen Übergang zwischen den Metaballs.

Diese einfache Darstellungsart ist für die Erzeugung von verschiedenen organischen Formen, wie z. B. die Erstellung von Tieren, Menschen oder Muskeln, sehr beliebt. Jedoch kann der Arbeitsaufwand sehr langwierig und ermüdend sein, da viel „try and error“ Zeit für die richtige Konfiguration und Anordnung der Metaballs vergeht. Es ist schwer das Endergebnis vorherzusehen, weil die eigentliche Form des modellierten Objektes erst zu erkennen ist, wenn sich das Polygonnetz (die Haut) darüber erstreckt (vgl. Abbildung 61 (links)). Deshalb sind Metaballs eher bei cartoonartigen Charakteren (vgl. Abbildung 61 (rechts)) und bei realistischen Abbildungen des menschlichen Körpers – mehrschichtige Modelle mit Muskelmasse – zu finden, als bei der Generierung eines Avatars [Gia97], [Rat98].

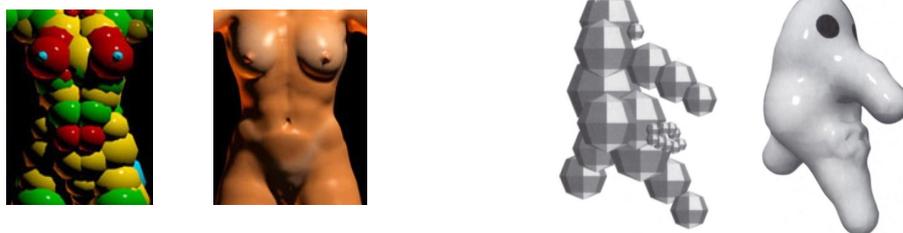


Abb. 61: links: Körper aus Metaballs; verändert aus [Plu02]
rechts: Kopf aus Metaballs; verändert aus [Rat98]

5.1.1.2 Modellierungsmethoden/-techniken

Nach dem einige grundlegende Arten zur Darstellung von Geometrien erläutert wurden, behandelt dieses Teilkapitel Methoden/Techniken zur Modellierung, bei denen diese Geometrien benutzt werden. Eine Übersicht einiger Techniken stellt die nachfolgende Abbildung 62 dar.

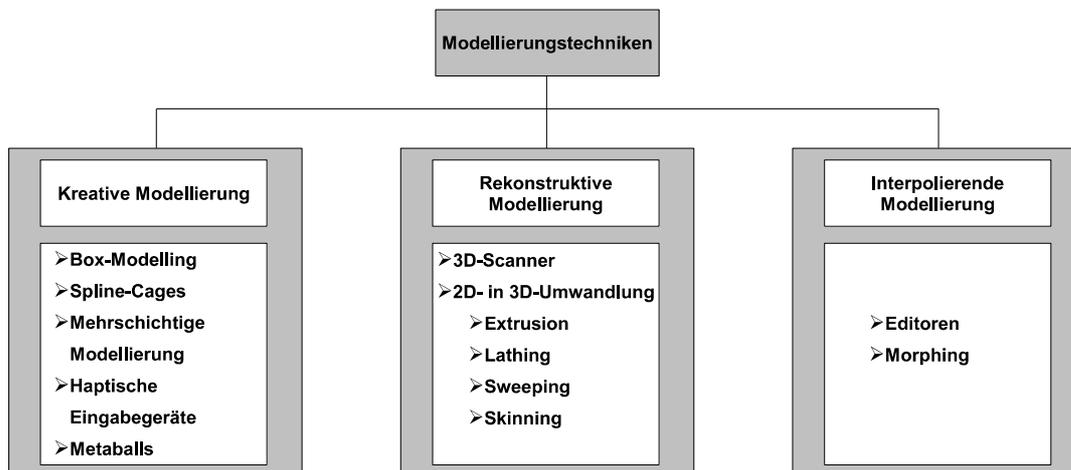


Abb. 62: Modellierungstechniken

Kreative Modellierung

Bei der kreativen Modellierung kann der 3D-Artist mit Hilfe von Software-Programmen eine Objektmodellierung realisieren. Er erzeugt das Drahtgittermodell mit den Werkzeugen der Modellierungssoftware vollständig selbst. Eine Übersicht zu ausgewählten Programmen für die Modellierung und Animation von Avataren befindet sich in [Tüm07].

Beim „*Box-Modelling*“ (vgl. Abbildung 63 (links)) modelliert man aus einem Grundelement, in dem Beispiel ein Quader, die eine Hälfte des Avatars. Durch Unterteilungsflächen, Extrudieren und Spiegeln der Vierecke wird das geometrische Modell designt. Die Kontur lässt sich durch Verschieben der Eckpunkte an die gewünschte Form anpassen. Diese Art der Modellierung lässt sich auch mit Splines als parametrische Flächen umsetzen und erhält dadurch die Bezeichnung „*Spline-Cages*“. Mit den „*mehrschichtigen Modellen*“ versucht man die Anatomie des Körpers nach zu modellieren. So gibt es Modelle, die aus einem Skelett, aus Muskeln (z. B. mit Metaballs), Fettgewebe, teilweise auch Organe und der Haut (Splines oder Polygonflächen) bestehen. Diese finden Anwendung bei Avataren für medizinische Tutorials. Sehr beliebt ist die Erstellung eines Avatars als ein zweischichtiges Modell, bei dem das Skelett zur Hilfe für die Animation genutzt wird und die Haut als äußere Hülle das Aussehen definiert (vgl. Darstellung 63 (mitte)). „*Haptische Eingabegeräte*“ wie das Phantom Omni (vgl. Abbildung 63 (rechts)) ermitteln die Position und Ausrichtung des Modellierungsstiftes. Es gibt verschiedene Methoden, wie die Eingabestifte funktionieren. So können diese mechanisch durch Trägheitssensoren, akustisch mittels Ultraschallsensoren oder elektromagnetisch über die Messung des Magnetfeldes arbeiten. Die resultierenden Abtastpunkte können dann durch ein Polygonnetz verbunden und gerendert werden. Diese Art der

Modellierung ist sehr genau, robust und schnell durchzuführen. Nachteilig ist der Wirkungsradius (tote Winkel) und die unbequeme träge Handhabung des Gerätes. „*Metaballs*“ lassen sich auch für die Modellierung von Objekte verwenden. Ihre Funktionsweise wurde schon im vorherigen Abschnitt 5.1.1.1 „Darstellungsarten der Geometrie“ erläutert.

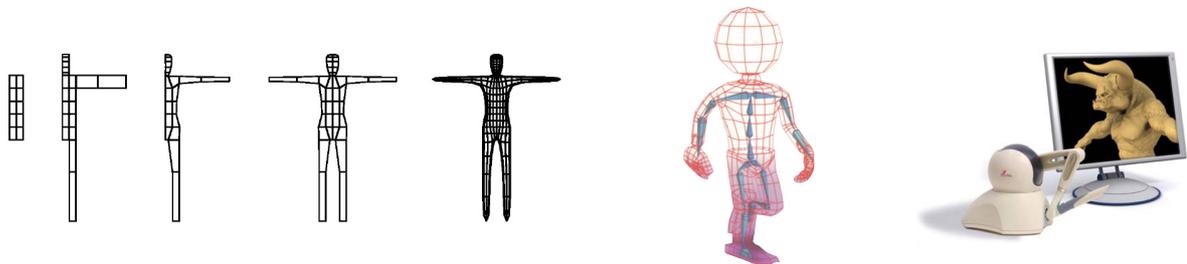


Abb. 63: links: Box-Modelling; aus [Tüm07]

mitte: zweischichtiges Modell; aus [Mae99]

rechts: Haptisches Eingabegerät Phantom Omni; aus [Tüm07]

Rekonstruktive Modellierung

Die rekonstruktive Modellierung analysiert existierende Formen und wandelt diese automatisch in eine geometrische Abbildung. Der „*3D Scanner*“ (vgl. Abbildung 64-1) ermittelt das zu digitalisierende Objekt mittels Laserlicht oder strukturiertem Licht und liefert somit die Punktdaten für das geometrische Modell. Dazu tastet der Laserstrahl das Objekt aus 360° ab. Das reflektierte Licht zeichnet eine CCD (*charge-coupled device*) auf und ermittelt somit die genauen Punkte der Oberfläche des Objektes. Die Anzahl der Punkte, die eine Software für die exakte Darstellung benötigt, liegt zwischen 100 bis 20.000 Abtastpunkte. Anschließend wird diese Punktoberfläche in ein Polygonmodell umgerechnet. Negativ an dieser Technik sind die Nachbearbeitungen. Die Reduzierung der Datenmenge, das Füllen von Löchern, Rauschunterdrückung und das automatische Erkennen des internen Skelettes, welches für die Animation von Bedeutung ist, sind die häufigsten Probleme, die bearbeitet werden müssen. Der Vorteil dieser Modellierungstechnik liegt bei der effizienten und sehr genauen Wiedergabe von detaillierten geometrischen Strukturen [Tüm07], [Flü08].

Ebenfalls lässt sich „aus Video- oder 2D-Material“ ein dreidimensionales Objekt erzeugen. Der einfachste Weg eine 2D-Form in ein 3D-Objekt umzuwandeln ist die „*Extrusion*“. Hierbei wird eine Linie, Kurve oder Fläche bestehend aus Polygonen oder Splines mittels einer Dimensionserhöhung parallel im Raum verschoben (vgl. Abbildung 64-2). So kann eine selbsterstellte Skizze oder ein abfotografiertes Objekt eingescannt, in einer Modellierungssoftware als Hintergrund eingefügt und nachgezeichnet werden. Ähnlich funktioniert das „*Lathing*“ (deutsch: Drehung) bei dem sich das Objekt nicht parallel verschiebt, sondern an einer Achse teilweise oder vollständig rotiert und somit ein Drahtgittermodell entsteht (vgl. Abbildung 64-3). Beim „*Sweeping*“ definiert ein Pfad, der aus Polygonen oder Splines besteht, die Richtung und Orientierung der Kontur des Objektes. Dieser Pfad wird anschließend extrudiert und ermöglicht somit bogenförmige auch ineinander verlaufende 3D-Drahtgittermodelle (vgl. Abbildung 64-4). Das „*Skinning*“ verbindet verschiedene Grund-

Das „*Morphing*“ ist eine Technik, bei dem zwischen zwei verschiedenen Objekten ein weicher Übergang interpoliert wird. Eine genauere Erklärung folgt im Anschlusskapitel 5.1.1.3.

5.1.1.3 Techniken zur Anpassung/Verformung der Geometrie bzw. Gestalt des Avatars

Bei der Gestaltung eines Objektes können im Nachhinein ungeplante Änderungen am Aussehen oder der Gestalt nötig sein. Für diesen Zweck gibt es Techniken, von denen eine Auswahl in diesem Kapitel aufgeführt sind. Dazu gehört die Freiformdeformation, die mehrschichtigen Modelle, das Morphing sowie Möglichkeiten zur Darstellung und Verformung der Gesichtsmimik.

Freiformdeformation (FFD)

Das Ziel der Freiformdeformation ist es, eine Geometrie mit möglichst wenigen Kontrollpunkten zu deformieren. Hierfür wird das Modell in ein umgebenes rechteckiges festes Gitter (FFD-Box) eingeschlossen. Die Schnittpunkte der Linien des Modells verbinden sich mit dem regelmäßigen Raster der umliegenden Box. Durch das Anfassen und Verschieben der Kontrollpunkte des Rasters verzerrt sich das innenliegende Modell entsprechend der Verformung der Box (vgl. Darstellung 66). Bei dieser normalen FFD ist das Gitter fest vorgegeben. Für eigene Anwendungen und zur besseren Kontrolle wurde die Erweiterte Freiformdeformation (EFFD) entwickelt, die erlaubt ein eigenes Kontrollgitter (z. B. zylinderförmig) zu erstellen [Tüm07], [Cun12].

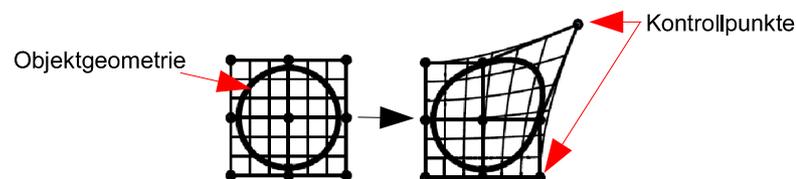


Abb. 66: Freiformdeformation; ergänzt aus [Cun12]

Mehrschicht-Modelle

Wie schon erwähnt, probieren die mehrschichtigen Modelle die Anatomie des Menschen nachzubilden. Deshalb bezeichnet man diese auch als anatomische Modellierung. Da bei den meisten Modellen aufgrund der schnellen Gestaltung und einfachen Handhabung zwei Schichten, jeweils eine für das Skelett und eine für die umliegende Haut, modelliert werden, ist es möglich dies für die natürliche Verformung bei der Animation zu nutzen. Dabei beeinflusst die unterliegende Schicht das Aussehen der höheren Schichten.

Am Ende der Modellierungsphase verbindet der 3D-Artist das Skelett mit der Haut. Hierbei hat er die Möglichkeit ein Bindungsgewicht zwischen den Eckpunkten der Hautgeometrie und den Gelenken des Skeletts festzulegen, den sogenannten „*Skinning Informationen*“. So kann zum Beispiel am Ellenbogengelenk die Hautschicht eine stärkere Gewichtung bekommen als am Unterarm, was sich darin äußert, dass sich die Haut bei einer Armbewegung am

Ellenbogengelenk stärker verformt. Ist die Gewichtung zu stark gewählt oder liegen die Eckpunkte der Geometrie ungünstig, äußert sich dieses in einer unerwünschten Deformierung, wie es in der Abb. 67 dargestellt ist. Nach dem dieser Schritt beendet ist, wird der Skelettapparat für die Bewegung des Objektes animiert und verformt die Haut bei entsprechenden Bewegungen realistisch mit [Tüm07]. „Das verformte Modell lässt sich außerdem effizient und schnell berechnen – mit Hilfe von Vertex-Shadern sogar direkt auf der Grafikhardware. Dadurch eignet sich das Verfahren besonders für die Echtzeitanimation [Tüm07].“



Abb. 67: Deformierung bei falschen Skinning-Informationen; aus [Tüm07]

Morphing

Die Technik des Morphings erlaubt es aus verschiedenen Einzelbildern oder Geometrien Zwischenübergänge zu berechnen. Dies gelingt durch das Verzerren der Kontur von typisch markanten Elementen eines Quellobjektes in das Zielobjekt. Das Ergebnis ist umso besser, je ähnlicher sich die Objekte sind. Bei dem Vorgang laufen die drei Schritte Verzerren (*warping*), Zwischenbilder/-posen (*tweening*) und Überblendung (*cross-dissolving*) ab. Dadurch ist es möglich, eine Ausgangsbasisform in eine andere Form zu „morphen“. Bei der Verformung von Geometrien definiert der Animator verschiedene Posen, die er als Schlüsselposen bezeichnet. Sind diese verschiedenen Grundformen der Geometrien erstellt, kann durch Interpolation zwischen den Posen das Aussehen verändert werden. Um eine bessere Kontrolle des Morphings zu haben, kann man Gewichtungsfaktoren zu jeder Pose festlegen. Das Erzeugen der einzelnen Schlüsselposen kann aufwendig ausfallen, da bei komplexeren Geometrien viele Eckpunkte verschoben werden müssen. Sind zu wenige oder sehr unterschiedliche Schlüsselposen vorhanden, könnte sich dies nach dem Morphing in einer unnatürlicher Verformung äußern. Angewendet wurde diese Verformungstechnik bei der Animation von Videospielen. Heute nutzt man diese Form in den meisten Produktionen für die Animation von polygonerstellten Gesichtsmimiken [Gia97], [Wik13].

Gesichtsverformung

Im 2. Kapitel wurden die nicht-manuellen Komponenten der Gebärdensprache erläutert. Dabei stellte sich heraus, dass die Mimik für das Verständnis der Gebärden eine erhebliche Rolle spielt. Des Weiteren ist sie die wichtigste Komponente für die Glaubwürdigkeit eines Avatars, denn sie definiert den Charakter der animierten Figur.

In den Anfängen probierten die Animatoren durch Verschieben der Eckpunkte des geometrischen Modells den Gesichtsausdruck des Avatars zu erstellen. Dazu mussten für jedes Bild in einer ablaufenden Animation mühselig die Eckpunkte neu bearbeitet werden. Später versuch-

ten sie die Gesichtsverformung durch die Implementierung einer Skelettstruktur im Kopf des Avatars zu verändern, wie es ähnlich bei der Animation von mehrschichtigen Modellen der Fall ist. Diese Methoden erfüllen zwar teilweise ihren Zweck, besitzen aber den Nachteil, dass die Kontrolle von subtilen Details bei der Verformung des Gesichtes fehlen. Für eine realistische Darstellung und gute Steuerbarkeit der Gestaltung des Gesichtsausdruckes setzten sich „*Morph-Targets*“, das „*Facial Action Coding System (FACS)*“ und der „*MPEG-4 Standard*“ durch.

Morph-Targets

Morph-Targets, auch als *blend shapes* bezeichnet, arbeiten nach dem Prinzip des schon vorher vorgestellten Morphings. Dadurch entsteht die Möglichkeit, eine Vielzahl von Gesichtsausdrücken zu modellieren. Bei der Erzeugung einer Geometrie des menschlichen Kopfes erhält dieser als Ausgangsbasis eine neutrale Mimik. Für einen anderen Gesichtsausdruck wird diese Geometrie durch Verschiebung der Eckpunkte abgeändert und als sogenanntes Morph-Target abgespeichert. Somit gibt es für jede Mimik ein zugehöriges Morph-Target (vgl. Abb. 68). Durch Interpolation, in der Regel einfache Transformationen der Form mittels Bewegungen, Rotationen oder Skalierungen zwischen den verschiedenen Gesichtsausdrücken, kann dann die gewünschte Mimik generiert werden.

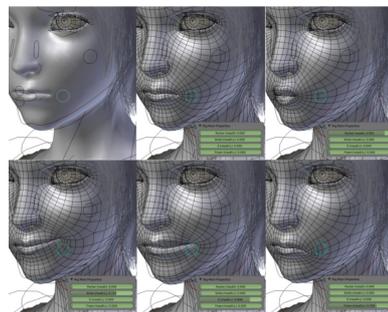


Abb. 68: Verschiedene Morph-Targets;
aus [Mor13]

Zu unterscheiden sind „*Single-Target-Morphing*“ und „*Multiple-Target-Morphing*“. In der ersten Variante kann der Animator lediglich ein Morph-Target aufrufen. Das heißt parallele Verformungen wie das Öffnen des Mundes bei gleichzeitigem Schließen des Auges sind nicht anwendbar, sondern müssen zeitlich separat voneinander ausgeführt werden. Multiple-Target-Morphing ermöglicht es hingegen, verschiedene Morph-Targets zu mixen. Durch diese Kombinationsvielfalt entsteht quasi eine Gewichtung, die es erlaubt, die unterschiedlichsten Gesichtsformen/-ausdrücke zu bilden. Die Vorteile von Morph-Targets sind die einfache Handhabung und das schnelle Rendering. So ist dieses Verfahren für Echtzeitanimation gut geeignet. Als Beispiel sei hier der Gebärdensprachavatar EMBR genannt, der mit solchen Multiple-Morph-Targets arbeitet [Mor13].

Facial Action Coding System (FACS)

Im Jahr 1976 entwickelten die Psychologen Paul Ekman und Wallace Friesen das Facial Action Coding System. Es stellt ein Kodierungsverfahren zur Beschreibung von Gesichtsausdrücken dar und gilt heute als Parametrisierungsstandard. Ekman und Friesen definierten sogenannte Bewegungseinheiten (*Action Units (AUs)*), bei der jede einzelne AU eine kleinste sichtbare Bewegung des Gesichtes zugeordnet ist. Dabei kann eine Action Unit einzelne oder mehrere Muskelbewegungen abdecken. Insgesamt existieren 46 AUs, von denen 20 der oberen Gesichtshälfte (Augen- und Nasenbereich) und 26 der unteren Gesichtshälfte (Mund- und Kinnbereich) zugeteilt sind. So bedeutet zum Beispiel die Codierung AU2: „Heben der äußeren Augenbraue“ (siehe Abbildung 69) oder AU9: „Rümpfen der Nase“. Um eine gewisse Verformungsgewichtung zu bekommen, kann der Codierung eine Ausprägungsstärke zugeteilt werden. Die Intensität ist in fünf Grade klassifiziert, die von Stärke A für „leichte Andeutung einer Bewegung“ bis Stärke E für „maximale Bewegung“ reicht. Durch die Kombination der verschiedenen Action Units lassen sich beliebige Gesichtsausdrücke realisieren. Dieses System erleichtert dem Animator bei der Vielzahl von Gesichtsmimiken/Parametern den Überblick nicht zu verlieren. So kann er mit Hilfe des Systems die richtigen Bereiche des Gesichtes für den gewünschten Gesichtsausdruck einstellen [Fac13].

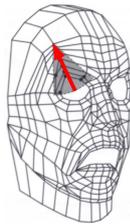


Abb. 69: Umsetzung von AU2: Heben der Augenbraue; ergänzt aus [Jac06]

MPEG-4 Standard

Bei der Einführung von MPEG-4 (*Moving Pictures Experts Group*) wurden erstmals Aspekte der Animation von Gesichtern in den Standard mit aufgenommen. Es kommt zur Anwendung bei virtuellen Köpfen, zu deren Sprachausgabe parallel die richtigen Lippenbewegungen ablaufen (z. B. virtuelle Teilnehmer in einer Videokonferenz). Dazu wird der sichtbare Teil eines Wortes (das Visem) aus den Phonemen der Sprache extrahiert. Dies dient als Ansteuerungsparameter für die Lippenbewegung. Bei der geringen Datenübertragungsrate von < 1 kbit/s müssen dann nur die Parameter für die Steuerung der Lippenbewegung und Gesichtsmimik übertragen werden. Bildinformationen sind zum Beispiel für die Kommunikation in virtuellen Konferenzen nicht nötig. Aufgrund der sehr geringen Bitrate ist MPEG-4 gut geeignet für die Steuerung der Gesichtsanimation in Echtzeit.

Für die Formung eines Gesichtes werden die *Face Definition Parameter (FDP)* benötigt, die ein Polygon-basiertes Kopfmodell mit 84 Punkten definiert. Die Darstellung 70-1 bildet diese sogenannten *Feature Points* ab. Diese erlauben neben der Definition von Gesicht, Zunge,

Lippen, Zähne und Augenbewegung auch Schulter- und Kopfbewegungen. Optionale gesichtsbeschreibende Parameter wie Textur des Gesichtes, Geschlecht, Alter, Haare und Brille sind ebenfalls einstellbar.

Für die Steuerung der Gesichtsanimation benutzt der MPEG-4-Standard sogenannte *Facial Animation Parameter (FAP)*. Insgesamt sind 68 FAPs – die in Gruppen eingeteilt sind – definiert (siehe Darstellung 70-2), die die Feature Points verschieben können und dadurch die Form des Gesichtes verändern und gewünschte Gesichtsausdrücke erzeugen. Um die FAPs für die möglich verschieden großen Polygon-Kopf-Modelle auszurichten, können diese durch sogenannte *Face Animation Parameter Units (FAPUs)* skaliert werden. Die FAPUs bestehen aus 5 Parametern (vgl. Darstellung 70-3) und geben die Distanzen zwischen Augen-, Nasen- und Mundbereich an. Sind diese auf das Kopfmodell richtig abgestimmt, können die FAPs verschoben werden [Jac06].

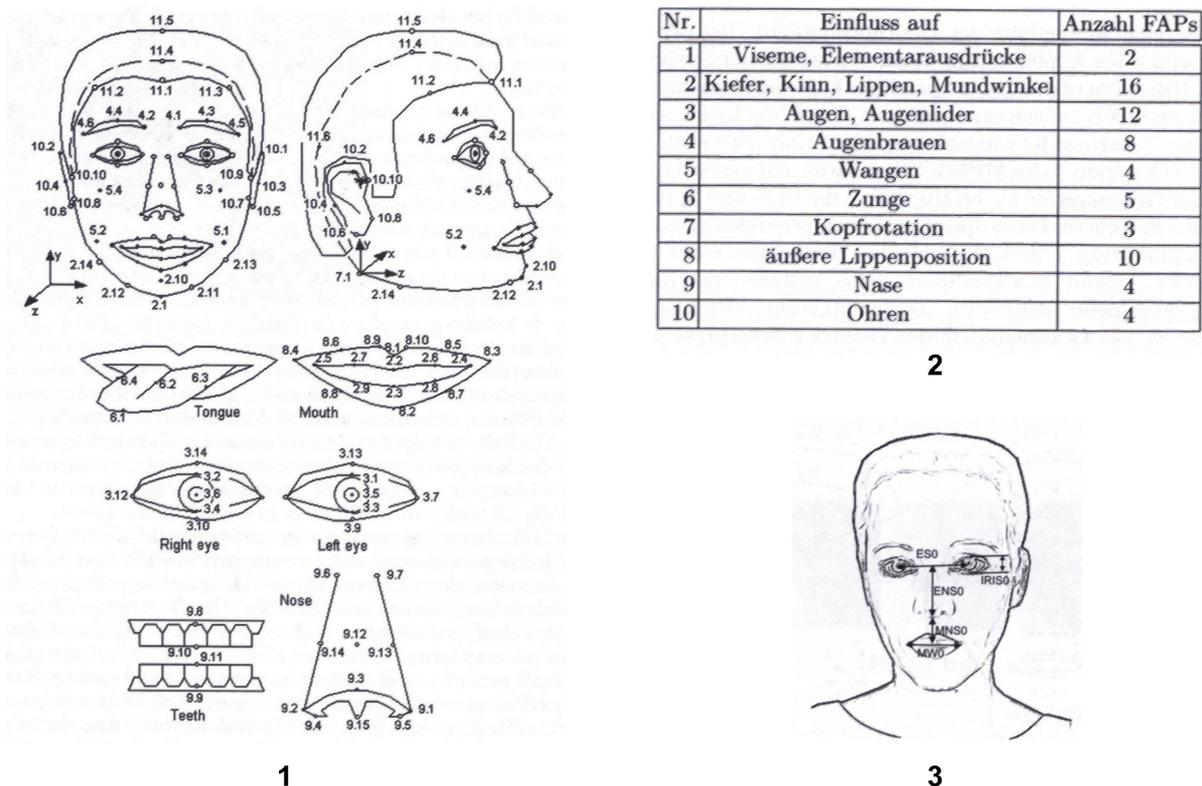


Abb. 70: 1-Feature Points legen das Drahtgittermodell fest, 2-FAP-Gruppen, 3-Positionierung der FAPUs; zusammengestellt aus [Par02], [Jac06]

Nach dem Techniken zur Darstellung, Erzeugung und Bearbeitung von Geometrien speziell für den Bereich Avatare präsentiert wurden, folgt im Bezug auf die Modellierungsphase (vgl. Abbildung 56) die Animationsphase, die im Anschluss aufgezeigt wird.

5.1.2 Animationsphase

In der Animationsphase erstellen ein oder mehrere Animatoren die Animationen eines Objektes. Der Begriff Animation stammt vom lateinischen Wort *animare/animus* ab, und lässt sich übersetzen mit „zum Leben erwecken/Geist bzw. Seele“. Um ein Objekt „zum Leben zu erwecken“ wird eine Abfolge von Bildern, die einzelne Phasen eines Bewegungsablaufs abbilden, erstellt. Um für den Betrachter die Illusion einer flüssigen Bewegung in einer Animation zu erzeugen, müssen ca. 24 erzeugte Einzelbilder pro Sekunde wiedergegeben werden. Diese Methode der Visualisierung von Abläufen findet häufig Anwendung in Filmen, Videospielen, in der Werbung, Wissenschaft und für Präsentationszwecken. Die Tabelle 9 teilt die Animationstechniken in die drei Bereiche „klassische Animation“, „2D-Animation“ und „Computeranimation“ ein.

Bereiche	Animationstechniken
Klassische Animation	<ul style="list-style-type: none"> – Stop-Motion oder Objekt-Animation → Brickfilme → Pixilation → Collagefilme → Claymation (Knetfigurenfilme) → Puppentrickfilme/-animation → Model-Animation (z. B. Raumschiffe) → Legetrickfilme/Cut-Out Animation
2D-Animation	<ul style="list-style-type: none"> – Zeichentrickfilme/Cel Animation per Handzeichnungen – traditionelles Keyframing – Daumenkino (Flip book) – Sand-auf-Glas-Animation – Nadelbrettanimation – Öl-auf-Glas-Animation – Zeichen oder Kratzen direkt auf Film – Rotoskopie oder Kinoxen – Scherenschnittfilm/Silhouettenanimation
Computeranimation	<ul style="list-style-type: none"> – Techniken aus klassische und 2D-Animationen per Softwarelösungen – Keyframe Animation/parametrisches Keyframing/Pose-To-Pose Technik – Kinematische Techniken (inverse und vorwärts) – Bild-für-Bild Animation (frame-by-frame) – Motion Capture (MOCAP) – dynamische und physikalische Animation – prozedurale (regelbasierte) Animation/parametrische Animation

Tab. 9: Übersicht Animationstechniken

Zur „klassischen Animation“ lässt sich die Stop-Motion-Technik zählen, die schon Ende des 19. Jahrhunderts benutzt wurde. Um ein Objekt zu animieren, wurde es abfotografiert, geringfügig bewegt und erneut fotografiert. Verschiedene Formen benutzen diese Technik. Einige sind Brickfilme (Figuren und Kulisse bestehen aus Legosteine), Pixilation (bildweises abfilmen von Schauspieler) oder Cut-Out Animation (z. B. Hampelmann). Die Stop-Motion-Technik wurde durch den Bereich der Computeranimation (Bild-für-Bild Animation) ersetzt. Anwendung findet sie lediglich nur noch in Filmen oder Musikvideos, um nostalgische Effekte zu erzielen.

Im Bereich der „2D-Animation“ stehen Techniken, die ebenfalls einzelne Bilder erzeugen, welche anschließend abfotografiert werden können. Im Unterschied zur klassischen Animation sind hier die Objekte keine erstellten Puppen, Figuren, Modelle oder Schauspieler, sondern per Hand hergestellte Zeichnungen, die auf Papier (Cel Animation, Daumenkino, traditionelles Keyframing), Film (Zeichen oder Kratzen, Rotoskopie), Glas (Sand- oder Öl-auf-Glas-Animation) oder als Schatten (Nadelbrettanimation, Scherenschnittfilm/Silhouettenanimation) abgebildet sind.

Der größte Teil der Techniken aus dem Bereich 2D-Animation lässt sich heutzutage mit speziellen Softwarepaketen in der „Computeranimation“ realisieren. Das Prinzip bleibt das Gleiche. Die Einzelbilder einer Bewegung können direkt am Computer gezeichnet oder mittels eines Scanners auf dem PC übertragen und gegebenenfalls verändert bzw. angepasst werden. Der Bereich der dynamischen und physikalischen Animation beinhaltet Techniken zur Kollisionsabfrage von Objekten, die Simulation von bspw. Schwerkraft, Wind, Wasser, Reibung, Trägheit usw. sowie das dynamische Verhalten der virtuellen Welt (z. B. Bewegung von Haare und Kleider im Wind). Diese Animationstechnik spielt für die Animation eines Avatars eher eine untergeordnete Rolle. Wichtiger dagegen sind „Keyframe Animation“, „Kinematische Techniken“, „Motion Capture (MOCAP)“ und „prozedurale Techniken“, die im Folgenden ausführlicher dargestellt sind.

5.1.2.1 Keyframe Animation

Eine Erleichterung für die frame-by-frame Animation ist das (parametrische) Keyframing. Anstatt jedes einzelne Bild einer Animation zu erzeugen, werden nur die markanten Posen, sog. Schlüsselposen (*keyframes*) einer Bewegung, zu festen Zeitpunkten erstellt. Dazu wird das geometrische Modell, mit den bereits vorgestellten Deformationstechniken, in die gewünschte Form gebracht. Um einen weichen Übergang zwischen den Schlüsselposen zu erhalten, berechnet der Computer mittels Interpolation die Zwischenbilder, in der Literatur auch als *in-betweens* bezeichnet. Dieses Prinzip wird auch beim traditionellen Keyframing (vgl. Tabelle 9: Bereich 2D-Animation) angewandt, wobei der Animator die Schlüsselposen und seine Assistenten die Zwischenbilder per Hand zeichnen.

Keyframe Animation ist ein universelles Animationsverfahren. Es ist anwendbar auf alle Modell-, Szenen- und Rendering-Methoden. Ein weiterer Vorteil ist die vollständige Kontrolle, die der Benutzer bei der Modellierung und Animation hat. Somit kann dieses Verfahren auch mit anderen Techniken kombiniert werden.

Ein Problem kann die Interpolation sein. In der Regel wird die lineare Interpolation, bei der ein direkter Pfad zwischen den Punkten im Raum berechnet wird, benutzt. Dadurch können Bewegungen eines Avatars sehr roboterhaft und kantig wirken, da das direkte Durchlaufen der Punkte eine Geschwindigkeitsverringerung bei Richtungsänderungen zur Folge hat (vgl. Abbildung 71-1). Eine Lösung wäre die Nutzung von Überschleiffaktoren, die aufeinanderfolgende Bewegungsabschnitte durch nicht direktes Durchlaufen der Zielkoordinaten, sondern nur durch Annäherung an diese ineinander überführen und somit eine konstante Bewegungsgeschwindigkeit ermöglichen (vgl. Abbildung 71-2). Alternativ könnte man die Interpolationsart wechseln und die Kreisinterpolation wählen, bei der die Zielpunkte direkt und mit gleichmäßiger Geschwindigkeit anvisiert werden (vgl. Abbildung 71-3).

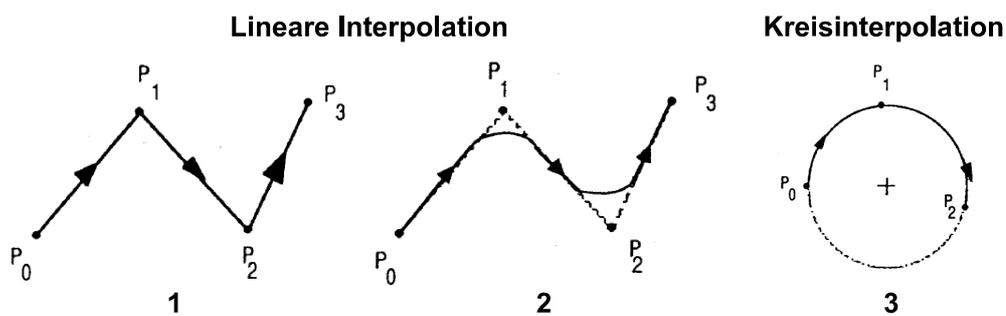


Abb. 71: 1 ohne Überschleifen, 2 mit Überschleifen, 3 ohne Überschleifen; verändert aus [Ber11]

Ein weiterer Nachteil von Keyframing ist der hohe Aufwand zur Erstellung der Schlüsselposen, die bei einer längeren Animation zahlreich ausfallen können.

5.1.2.2 Kinematische Techniken

Eine weitere sehr beliebte Methode zur Animation sind die kinematischen Techniken. Wie schon erwähnt, werden bei mehrschichtigen Modellen häufig zwei Schichten modelliert. Zum einen das Skelett und zum anderen die Hautschicht. Kinematische Techniken steuern dann das Skelett, welches die überliegenden Schichten beeinflusst. Dazu werden die Gelenke des Skeletts hierarchisch unterteilt in Elternknoten und Kinderknoten (sog. kinematische Kette). Daraus lassen sich zwei Möglichkeiten ableiten. Die Kinderknoten beeinflussen die Elternknoten oder andersherum, die Eltern beeinflussen die Kinder. In der Animation spricht man von Vorwärts Kinematik (*forward kinematic (FK)*) und Umgekehrte Kinematik (*invers kinematic (IK)*).

Bei der FK werden die einzelnen Gelenke (Kinder) vom Animator manuell definiert. Der Elternknoten (Effektor) passt sich dann durch Berechnung der Entgliedstellung der Position der Kinder an. Diese Form der kinematischen Technik ist jedoch sehr zeitaufwendig. Häufiger

in der Computeranimation genutzt ist die IK, bei der die Elternknoten manuell in die gewünschte Position gebracht werden. Die Kinderknoten folgen dann durch Berechnung – mittels des IK-Solver – den Elternknoten und verschieben sich dementsprechend in die richtige Position. Der Vorteil dieser Technik ist die schnelle Arbeitsweise. Jedoch können bei schlechtem Skelettbau oder durch zu weite bzw. ungünstige Positionierung des Elternknotens unnatürliche Posen entstehen. Um dieses Problem zu lösen, muss entweder die gewünschte Pose abgeändert oder die Gelenke manuell per FK eingestellt werden.

Kinematische Techniken erleichtern die Bewegung des geometrischen Modells durch Verschiebung weniger Punkte. Das Ergebnis der Animation wirkt jedoch etwas kantig und unnatürlich. Neben dem Bereich der Avatar-Animation findet diese Technik auch Anwendung bei der Steuerung von Industrierobotern.

5.1.2.3 Motion Capture (MOCAP)

Motion Capture ist eine sehr verbreitete Form zur Erzeugung von Animationen, die häufig in Film- oder Videospieldproduktionen auftauchen. Es fundiert auf aufgezeichneten Bewegungsdaten. Dazu muss ein Schauspieler den gewünschten Bewegungsablauf vorspielen.

Die Bewegungsdaten lassen sich erfassen durch optische, magnetische oder mechanische Systeme (vgl. Abbildung 72). Bei der Nutzung von optischen Systemen werden reflektierende oder selbstleuchtende Marker an markanten Körperstellen angebracht. In einem gleichmäßig gut ausgeleuchteten Raum bewegt sich der Akteur und wird von verschiedenen digitalen Kameras (üblicherweise 8 - 24), die leicht erhöht in 360° angeordnet sind, erfasst. Eine Software ermittelt per Bildvergleich und Triangulation die von den Kameras erkannten Markerpunkte und wandelt diese dann in eine dreidimensionale weiße Punktdatenwolke, welche anschließend als Grundlage für ein Drahtgittermodell dient, um die dadurch erzeugten Bewegungsdaten können dann bearbeitet und für die Animation wiedergegeben werden. Diese Methode ist sehr teuer, aber das Resultat ist eine natürlich flüssig ablaufende Bewegung. Probleme, die bei optischen Systemen auftauchen, sind das Fehlen von Daten (z. B. durch Verdeckung der Marker bei Bewegungen) und der beschränkte mit Kamera überwachte Erfassungsraum.

Die magnetischen Systeme arbeiten mit Hall-Sensoren, die am Körper des Akteurs platziert werden. Eine elektronische Einheit empfängt die gesendeten Daten der Sensoren und kann somit ihre Position im Raum bestimmen. Bei dieser Methode treten keine Überdeckungsprobleme auf, wie bei der Nutzung mit optischen Marker. Störend kann jedoch die Umgebung sein. So können Metallgegenstände die Sensordaten beeinträchtigen.

Für die mechanische Variante muss der Akteur ein äußeres Skelett tragen, welches die Bewegungen des Trägers durch elektronische Sensoren verfolgt und aufzeichnet. Bei den anderen beiden Möglichkeiten wurde die Position der Sensoren im Raum erfasst. Für diese Methode sind nur die Rotationsdaten des Skeletts wichtig. Dadurch ergibt sich der Vorteil, dass diese Technik keine räumliche Beschränkung hat. Hinderlich ist hingegen das Tragen des äußeren

Skeletts, welches den Bewegungsfreiraum des Akteurs begrenzt. Lösen lässt sich dieser Nachteil durch die Verwendung von Gyroskop Sensoren (vgl. Abb. 72), die ihre Lage und Orientierung im Raum ermitteln.

Zusammenfassend besitzt MOCAP den Vorteil der exzellent natürlich wirkenden Animation. Negativ zu sehen ist der hohe Bedarf an Speicherplatz für die Aufzeichnungsdaten, die aufwändige Arbeit der Datenerfassung, die Nachbearbeitungen (Ausfiltern von Rauschen, Fehler bei der Aufnahme restaurieren) und die somit verbundenen hohen Produktionskosten [Tüm07].



Abb. 72: MOCAP-Arten zur Datenerfassung: von links nach rechts: optisch, magnetisch, mechanisch und mit Gyroskope; aus [Tüm07]

5.1.2.4 Prozedurale Animation

Der Bereich der prozeduralen Animation beinhaltet eine Reihe von Animationstechniken, die sich mit anderen Techniken kombinieren lassen. Zu diesen zählen (stochastische) Partikelsysteme (z. B. Feuer, Funken, Luftblasen, Wolken) oder Flocking-Systeme (Tierherden, Menschenmassen, Vogel- oder Fischschwärme).

Bei einer prozeduralen Animation wird der Bewegungsablauf von einer Prozedur oder ein System von Regeln vorgegeben. Diese regelbasierten Verhalten sind in Algorithmen beschrieben. Ein Programm oder Benutzerfenster erlaubt es dem Nutzer die Parameter, welche als Kontrollregler für die Animation dienen, des Animationsalgorithmus abzuändern. Ein Beispiel wäre ein Laufalgorithmus, der die Gangart eines Avatars beschreibt. In diesem Algorithmus ist die Verhaltensbeschreibung mit sämtlichen Details der Animation beschrieben. Die Parameter könnten dann zum Beispiel spezifische Anweisungen sein, wie „bewege dich von Punkt A nach B“ oder die Art der Bewegung „gehen, laufen oder rennen“ (siehe Abbildung 73).



Abb. 73: Laufanimation durch prozedurale Animation; aus [Tüm07]

Der Vorteil der prozeduralen Animation ist das Vorhandensein einer Vielzahl von verhaltensbeschreibenden Algorithmen, die lediglich in ihren Parameter für die zur verwendeten Situation angepasst werden müssen. Somit erspart sich der Animator die zeitaufwendige Methode des Keyframing. Weiterhin benötigt man weniger Speicherplatz für die Algorithmen als für die anderen Techniken. Durch eine geeignete Parameterwahl und -benennung besitzt der Animator eine hohe und kompakte Kontrolle über die Animation.

Negativ ist das sogenannte Kontrollproblem. Nach der Festlegung der Parameter können diese während der Animation nicht mehr verändert werden, was zur Folge hat, dass eine direkte Beeinflussung bei der Berechnung der Animation nicht möglich ist. Ein Lösungsansatz ist die Parameter nach dem Bewegungsablauf neu einzustellen und erneut zu testen, ob die Bewegung die gewünschte Qualität erreicht. Eine andere Lösungsmöglichkeit ist das Benutzen von Beschränkungen (*Constraints*). Diese Zwangsbedingungen können zusätzliche Funktionen im Programmcode sein. Voraussetzung dafür ist der freie Zugriff auf den Algorithmus und ausreichende Programmierkenntnisse. Ein Beispiel wäre ein Algorithmus, der ein Avatar auf einen festgelegten Punkt im Raum schauen lässt. Durch den Constraint „Folge den Punkt“ könnte ein Avatar einen rollenden Ball oder den Mauszeiger folgen. Dazu könnten weitere Parameter nötig sein, die für eine benutzerfreundliche Übersicht eher hinderlich sind.

Ein weitere Nachteil ist, dass bei speziellen Anwendungen (wie z. B. Gebärdensprachavatar) ein hoher Entwicklungsaufwand für Spezialfunktionen/-verhalten nötig sein kann. Außerdem sehen die Bewegungen aufgrund des Algorithmus ständig gleich aus (z. B. Laufbewegung), was dem Avatar steril und roboterhaft wirken lässt. Um dieses Problem zu kaschieren, gibt es das Improv-System von Perlin und Goldberg (siehe [Gol96]), das zufällige leichte Störungen einbaut und den Eindruck der wiederholenden Bewegungen vermindert.

Wie in den vorherigen beiden Abschnitte zu erkennen ist, gibt es mittlerweile für die Nachteile verschiedene Methoden zur Lösung. Somit lässt sich diese Animationstechnik bei Computerspielen, Filmen (Menschenmassen), Wettereffekte oder bei komplexen Bewegungsabläufen, die ein hohen Grad an Realismus fordern – sowie es beim Gebärdensprachavatar der Fall ist – anwenden [Tüm07].

5.1.3 Rendern

Um das endgültige Resultat der Animationserstellung in einer Szene zu sehen, muss es gerendert werden. Dazu berechnet ein Render-Algorithmus die Farbwerte der einzelnen Pixel, aus denen sich das Bild zusammensetzt. Für die dreidimensionale Darstellung der Szene muss das Programm – der Renderer – die Geometrien der Modelle, das Aussehen von Objektoberflächen und Materialeigenschaften, die Lichtquellen und deren Schattenwürfe sowie die Position und Blickrichtung des Betrachters berechnen.

Zur Bewertung der Qualität des Bewegungsablaufs ist es ratsam, mit niedrigen Einstellungen den Render-Prozess zu starten, da somit die aufwändige Berechnung zeitlich kürzer ausfällt.

Ist der Animator mit dem Ergebnis zufrieden, können die Parameter wie Anti-Aliasing (Kantenglättung), die endgültige Auflösung, Lichteinstellungen, Schatten, Reflexionen, Motion Blur (Bewegungsunschärfen) usw. hochgestellt und das finale Rendern durchgeführt werden. Um diesen Prozess zu beschleunigen, lassen sich mehrere Rechner zusammenschalten und führen die Berechnung gemeinsam durch.

Für die Erzeugung von realistisch hochwirkenden Objekten (z. B. Architektur-Modelle von Gebäude) benötigt der Renderer eine sehr lange Zeit und würde für die Anwendung von interaktiven Avataren nicht sinnvoll sein. Für den Anwendungszweck von Gebärdensprachavatare empfiehlt sich das Echtzeitrendern. Voraussetzung ist dafür eine Beschleunigung der Hardware mittels einer Grafikkarte. Diese enthält eine Vielzahl von Shadern, die als kleine Recheneinheiten auf dem Grafikchip dienen und somit das Rendern beschleunigen.

Das Echtzeitrendern basiert auf einer Grafikpipeline bzw. Rendering-Pipeline, zu der folgende Schritte ablaufen. Der „*Anwendungsschritt*“ ermittelt alle interaktiven Änderungen einer Szene und leitet diese zur Bearbeitung an die nächste Phase weiter. Techniken, die hier zu finden sind, können Kollisionserkennung, Animation und Morphing sein. Der nachfolgende „*Geometrieschritt*“ ist verantwortlich für Operationen mit den Polygonen und deren Eckpunkte. Je nach Implementierung können folgende Aufgaben in der Phase abgearbeitet werden: Modell- und Kameratransformation, Beleuchtung, Schneiden des gewünschten Bildausschnittes (Clipping) und weitere Bildtransformationen zur richtigen Darstellung. Im letzten Schritt, den „*Rasterungsschritt*“, werden die Pixel eingefärbt und anschließend auf dem „*Bildschirm*“ angezeigt [Gra13].

5.1.4 Fazit: Gebärdensprachavatar

5.1.4.1 Design des Avatars

Bevor ein 3D-Artist den Avatar erstellt, sollte die Frage nach der äußerlichen Erscheinung geklärt sein, denn diese kann die Akzeptanz des Systems beim Endnutzer mit beeinflussen. Diese Aufgabe übernimmt der Designer. Es gibt zwei Möglichkeiten zur Darstellung von Gebärdensprachavatare. Erstens der Charakter soll „*photorealistisch*“ wirken oder zweitens eher „*cartoonhaft*“. In [Kip11] wurde zu dieser Problematik eine Untersuchung durchgeführt, bei der verschiedene Avatare gezeigt wurden. Das Meinungsbild wurde dazu in einer empirischen Studie festgestellt. Eine Fokusgruppe, bestehend aus Gehörlosen, Ertaubten oder Schwerhörigen im Alter zwischen 25 bis 50 Jahre, betrachtete die abgebildeten Avatare in Darstellung 74 in einem Video.



Abb. 74: von links nach rechts: Video „The Forest“ von Vcom3D, Avatar „Max“ von Televirtual/UEA, ein spanischer Geschichtenerzähler, spanischer Junge, PR-Video der deafworld 2000 und ein Avatar für mobile Endgeräte von Vcom3D; aus [Kip11]

Die Teilnehmer bewerteten das Aussehen und es stellte sich Folgendes heraus: „Am positivsten schnitt die cartoonartige Gebärdenanimation ab. Dort wurden die Charaktere als sehr warm und verständlich bewertet. Auch die Umsetzung und der Einsatz der Emotionen und des Mundbilds gefiel den Teilnehmern [Kip11].“ Die Anzahl der befragten Personen lag lediglich bei acht. Um das Ergebnis zu bekräftigen, wurde in einer weiteren Umfrage per Internet (siehe [Stu11]), die insgesamt 320 Teilnehmer ansprach, durchgeführt. Eine Bewertung der Parameter: Verständlichkeit, Mimik, Natürlichkeit, Ausstrahlung, Bewegung, Emotionen, Zusammenspiel Gebärde und Mimik, Mundbild, Aussehen und Handform ergab, dass der speziell für Gebärdensprache entwickelte Avatar Max am schlechtesten bewertet wurde und der cartoonartige Avatar von „deafworld“ die meisten positiven Urteile erhielt. Dies lässt sich mit dem Effekt des „Uncanny Valley“ in Abbildung 75 begründen.

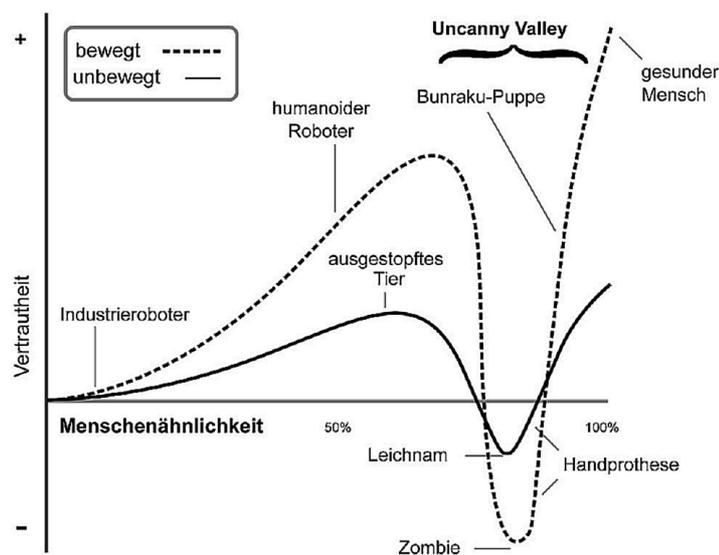


Abb. 75: Effekt des Uncanny Valley (unheimliches Tal); aus [Poh09]

Dieser Effekt wurde 1970 von Masahiro Mori bei der Untersuchung in der Robotik entdeckt und bildet den Verlauf der Akzeptanz des Menschen zu einer künstlichen Figur ab. In der Grafik ist zu erkennen, je menschlicher (x-Achse) eine künstliche Figur ist, desto vertrauenswürdiger (y-Achse) ist sie in unserem Empfinden. Bei einem beweglichen Charakter erhöht

sich die Akzeptanzkurve. Jedoch nimmt diese Akzeptanz schlagartig bei einem bestimmten Punkt – dem Uncanny Valley – ab. Dieser Effekt lässt sich durch die Psychologie des Menschen erklären. Besitzt zum Beispiel ein Avatar ein realistisches Aussehen und einen natürlichen Bewegungsverlauf, aber keine Animation für Augenzwinkern oder Atembewegungen, so wirkt er auf den Menschen nicht lebendig oder abstoßend und fällt somit ins Uncanny Valley. Bei einer Figur, die wie ein Mensch aussehen und wirken will, erwartet unser Gehirn auch ein menschliches Verhalten und bewertet jede kleinste Abweichung negativ. Erst wenn die Figur vollständig menschlich wirkt, steigt die Kurve der Akzeptanz wieder an. Versucht der Avatar erst gar nicht 100 % realistisch auszusehen, indem er beispielsweise cartoonartig dargestellt wird, so versucht das Gehirn nicht die Verbindung und Erwartung zwischen künstlicher Figur und realistischen Menschen herzustellen. Deshalb sollte ein Designkonzept eines Avatars entweder einen Menschen mit allen seinen Details widerspiegeln oder eine cartoonartige Darstellung vorweisen, um vom Benutzer des Avatarsystems akzeptiert zu werden.

5.1.4.2 Erstellung des Avatars

Ein endgültiges Fazit für die Erstellung eines Gebärdensprachavatars kann nicht pauschalisiert werden. Es gibt viel zu viele Faktoren. Wichtige Fragen, die vorher geklärt werden müssen, sind zum Beispiel:

- Wie sieht die genaue Zielgruppe aus?
- In welchen Bereichen soll der Gebärdensprachavatar zur Anwendung kommen?
- Welche Darstellungsformen unterstützt die Animationssoftware?
- Welche Softwarepakete sind für die Geometrieerzeugung, -verformung, Animation und Rendering überhaupt anwendbar?
- Gibt es Werkzeuge (Tools), die die Arbeit erleichtern bzw. unterstützen?
- Gibt es Probleme bei dem Austausch der Formate zwischen den Softwarepaketen?

Unabhängig vom Anwendungsgebiet und den Fragen lassen sich aber grob die Techniken in den Bereichen Modellierung und Animation klären.

Für den Bereich der Modellierung sind Polygone oder die Spline-basierten parametrischen Flächen gut geeignet. Der Vorteil der Polygone ist, dass eine Vielzahl von Rendering Programme diese berechnen können. Die parametrischen Flächen erlauben es dagegen, den Avatar auflösungsunabhängig mit wenigen Kontrollpunkte schnell zu erstellen. Ein weiterer Vorteil ist, dass sich parametrische Flächen in eine Polygondarstellung umwandeln lassen und somit den Vorteil der guten Rendering-Kompatibilität mit annehmen können. Metaballs und Unterteilungsflächen können auch gute Ergebnisse liefern, haben aber den Nachteil der aufwändigen komplexen Arbeit.

Als Modellierungstechnik empfiehlt sich eine Kombination aus der kreativen und rekonstruk-

tiven Modellierung. Der Avatar könnte somit auf ein zweischichtiges Modell, bestehend aus einer unteren Skelettschicht und einer darüberliegende Hautschicht, basieren. Für die Umsetzung der Geometrie des Avatars kann eine per Hand gezeichnete Skizze eingescannt und in einem Modellierungsprogramm mittels Box-Modelling oder Spline-Cages, die wiederum Techniken wie Extrusion, Lathing, Sweeping oder Skinning nutzen können, kreiert werden. Bei wenig Erfahrung mit geometrischer Modellierung sind Charakter-Editoren eine gute Alternative, besitzen aber nicht die gestalterischen Freiheiten einer Modellierungssoftware. Um die Geometrie eines Avatars selber zu realisieren, beschreiben die Quellen [Rat98] und [Mae99] detailliert die einzelnen Teilschritte.

Wichtig bei der Modellierung von Gebärdensprachavataren ist die Mimik. Hier bietet das FACS ein Parametrisierungsstandard, an den sich der Modeller für die Modellierung der Gesichtsgeometrie und des richtigen gewünschten Gesichtsausdrucks orientieren kann. Eine Alternative ist der MPEG-4 Standard, der ebenfalls Parameter für die Gesichtsform zur Verfügung stellt und Kontrollpunkte für die Erzeugung von Mimik definiert. Sein Vorteil besteht in der geringen Datenrate und wäre somit für Echtzeitanwendungen wie zum Beispiel Internet-Applikationen geeignet. Mit Hilfe von Morph-Targets können dann einzelne Körperposen oder Gesichtsausdrücke separat abgespeichert werden. Für eine facettenreiche Gesichtsmimik kann Morphing eine Gewichtung zwischen den Morph-Targets erreichen.

Im Bereich der Animation kann mittels Keyframing die Animierung zwischen den Morph-Targets erfolgen. Denkbar ist auch, verschiedene Körperposen als Morph-Targets abzuspeichern und zwischen diesen zu interpolieren. Üblicher ist es jedoch den Körper durch kinematische oder per prozedurale Techniken zu steuern. Bei einem zweischichtigen Modell kann die Skelettschicht durch die Vorwärts oder Umgekehrte Kinematik animiert werden. Um mehr Einfluss auf die Parameter der Animation zu haben, ist die auf constraints-basierte prozedurale Animation geeignet. So kann ein Skript-basiertes Notationssystem diese Constraints beinhalten (siehe z. B. EMBRScript) und dient somit als Kontrollsprache für das Animations-system. Die besten Ergebnisse für die Gesicht- bzw. Körperanimation bietet die MOCAP-Technik, besitzt aber den Nachteil der hohen Kosten. Zusammenfassend zeigt die Abbildung 76 einen möglichen Ablauf zur Erzeugung eines Gebärdensprachavatars.

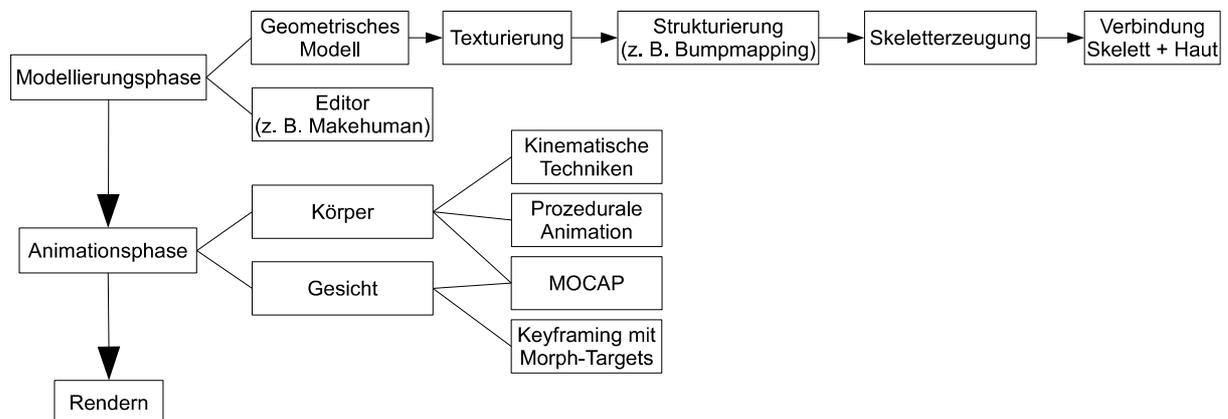


Abb. 76: Generierung eines Gebärdensprachavatars

Das Ergebnis der Modellierungsphase ist ein vollständiger Avatar, der bereit für die Animationserstellung ist. Diese Phase kann mehrere Schritte durchlaufen. Als erstes wird ein geometrisches Drahtgittermodell erstellt. Um dieses geometrische Modell detaillierter und natürlicher wirken zu lassen, bekommt es mehrere Texturen. Diese werden auf die Geometrie gelegt und bewirken keine Änderungen der Form, sondern nur das Aussehen der Oberfläche. Um Strukturen in den Texturen zu erzeugen, wie zum Beispiel Gesichtsfalten, kann der Modellierer Mapping-Techniken (z. B. Bump-Mapping oder Displacement-Mapping) nutzen. Bump-Mapping und Displacement-Mapping arbeiten mit Graustufenbildern. Der Unterschied ist, dass beim Bump-Mapping der Eindruck der Oberflächenunebenheiten/Tiefenwirkung ohne Veränderung der geometrischen Form funktioniert. Beim Displacement-Mapping wird die unterliegende Geometrie hingegen an das Graustufenbild angepasst. Nachdem das Aussehen des Charakters fertiggestellt ist, wird ein Skelett speziell für diesen Avatar entworfen. In der letzten Modellierungsphase verbindet der Modellierer es per Skinning mit dem texturierten und strukturierten geometrischen Drahtgittermodell. Da diese Schritte sehr zeitintensiv sind, kann alternativ ein Charakter-Editor (z. B. Makehuman) genutzt werden. Der Vorteil dieser Editoren ist, dass der Anwender eine Auswahl an Charakteren hat und durch verschiedene Parameter wie Größe, Geschlecht, Augen, Mund usw. das Aussehen des Avatars nach seinen Vorstellungen gestalten kann. Anschließend werden diese Daten abgespeichert und in einem Animationsprogramm importiert, in dem die abgebildeten Animationstechniken in den Bereichen Körper und Gesicht zum Einsatz kommen können. Abschließend folgt das Rendering.

Manchmal ist es nötig verschiedene Methoden in der Praxis auszuprobieren, um für den geforderten Anwendungsbereich die richtige Darstellungsform, Modellierungsart oder Animationstechnik zu wählen. Durch Kombination der verschiedenen Methoden lassen sich die Nachteile der verschiedenen Techniken für die Erstellung eines Gebärdensprachavatars beseitigen.

Eine weitere Erkenntnis der Untersuchung der Techniken in diesem Kapitel zeigt, dass die Anforderungen (vgl. 2. Kapitel, Abschnitt 2.4) für die Animation eines Avatars im Bereich der Gebärdensprache technisch alle realisierbar sind. So können Handformen, Handstellungen, Körperbewegungen und Mimik ausreichend modelliert und animiert werden. Kollisionserkennung lässt sich in einigen Modellierungsprogrammen automatisch anschalten oder könnte durch physikalische Animation umgesetzt werden.

5.2 Anwendungsbereiche Avatarsysteme

Dieses Unterkapitel befasst sich mit möglichen Anwendungsgebieten, bei denen Avatare zum Einsatz kommen können. Die Darstellung 77 zeigt diese vielseitigen Anwendungszwecke, die grob in „Kommunikation“, „Trainings-simulation“, „Entertainment“, „Interaktion“, „Informationsdarstellung“ und „Linguistik“ unterteilt sind.

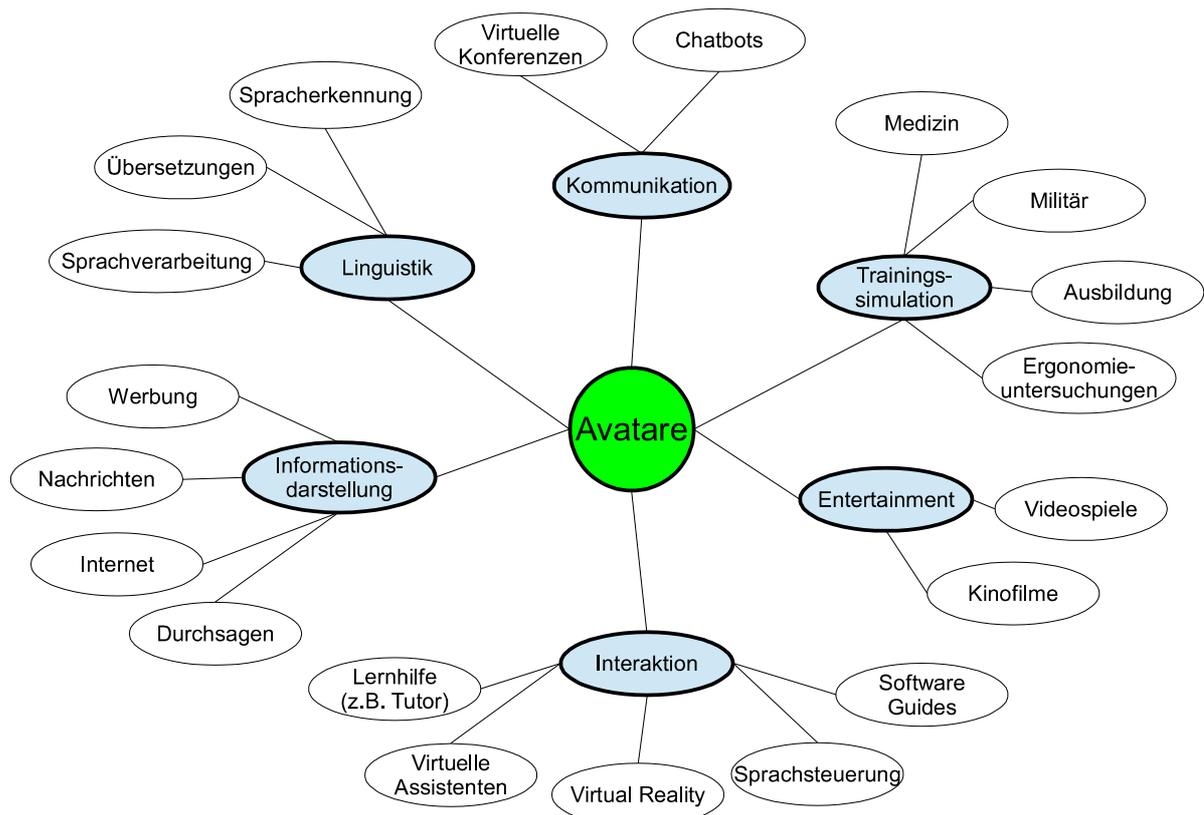


Abb. 77: Mögliche Anwendungsbereiche für Avatare

Avatare dienen im Bereich der *Kommunikation* als virtuelle Gesprächspartner und erleichtern den Nutzern das Gespräch visuell zu folgen. So kann ein Internetnutzer einen Avatar in einem Forum oder einer Freundesliste verwenden, der in Form eines (animierten) Bildes (sog. Chatbots oder Avatare), welches ein Hobby, ein Lieblingsspiel, einen Verein, eine Band, usw. darstellt, den Internetnutzer repräsentiert. Bei virtuellen Konferenzen (vgl. Bild 78) sitzen die Teilnehmer innerhalb einer virtuellen Umgebung (z. B. Konferenzraum) und können sich mit Hilfe ihrer Avatare unterhalten. Sie können sich gegenseitig lokalisieren, identifizieren und gegebenenfalls Aktionen (z. B. Visualisierung von Interesse, Austausch von Dokumente) ausführen, was die Wahrnehmung und Intensität der Unterhaltung aufwertet. Der Vorteil virtueller Meetings ist, dass Kosten (Mieten von Tagungsräume oder Anfahrtskosten) entfallen und spontane bzw. kurzfristige Treffen möglich sind.

Eine andere Einsatzmöglichkeit für Avatare ist die *Trainingssimulation*. Die Simulation erlaubt es verschiedene Situationen in verschiedenen Schwierigkeitsstufen abzubilden. Dabei sind Avatare häufig im Bereich des Militärs zu finden. Hier können Einsätze, die in der realen Welt zu teuer oder zu gefährlich sind, nachgestellt werden. Durch das Wiederholen, das Aufzeichnen und der Auswertung der hier simulierten Übungen kann der Anwender seine Vorgehensweise und Reaktionen verbessern und im realen Einsatz anwenden. Dieses Prinzip lässt sich auch in der Medizin finden, bei der Avatare als Patienten zum Beispiel für eine Operation (vgl. Abb. 79) in Frage kommen. Zur Untersuchung und Entwicklung von Fahrzeuginnenräumen, Flugzeugcockpits, Fahrgasträumen oder Arbeitsplätzen lassen sich ebenfalls Avatare, wie die Abbildung 80 zeigt, benutzen. Ziel dieser Ergonomieuntersuchungen ist es, die Arbeitsabläufe und -bedingungen für den Menschen zu optimieren. Andere Ausbildungen (Flugzeug-/Bahn-/Fahrzeugsimulator usw.), in denen Avatare vorkommen, lassen sich ebenfalls zu den Bereich der Trainingssimulation zählen.

Der Bereich, in dem die Avatare am weitesten entwickelt sind, ist wahrscheinlich der des *Entertainment*. Hier kommen die teuersten Techniken z. B. das MOCAP vor, was die Qualität der Produkte erheblich steigert. So sehen die Animationen der Avatare in Kinofilmen, wie beispielsweise Shrek 2 (vgl. Darstellung 81), natürlicher und realistischer aus als in den anderen Bereichen. In Videospielen erhält der Nutzer die Möglichkeit, einen eigenen Avatar in einer virtuellen Umgebung zu steuern. Andere Charaktere, die durch den Computer kontrolliert werden (vgl. Abbildung 82), bezeichnet man als NPC (*non-player character*). Sie dienen als gestalterisches Element der virtuellen Welt und können zur Erzählung der Geschichte, Interaktion mit dem Spieler oder einfach zur Belebung der Umgebung dienen.

Avatare lassen sich auch für *Interaktionen* verwenden. Ein Anwendungsgebiet dafür sind die Software Guides. Diese animierten Figuren geben den Anwendern Erklärungen und Hilfestellungen für die gerade verwendete Software. Einer der bekanntesten Avatare in diesem Gebiet ist die animierte Büroklammer „Karl Klammer“ (vgl. Darstellung 83), die als Hilfeassistent in dem Schreibprogramm Microsoft Office dem Anwender zur Verfügung steht. Diese Art von Hilfestellungen treten auch bei virtuellen Assistenten und Lernhilfen auf. So können diese den Nutzer durch Programmmenüs oder Internetseiten mittels einer Gebrauchsanleitung (Tutorial), die die Bedienung und Funktionen erläutert, leiten. Durch Interaktion kann der Anwender auf Aktionen der Avatare reagieren oder eigene Interessen kundtun. Intelligente Avatarsysteme, wie der Avatar Max (vgl. Abbildung 84) der Uni Bielefeld, werden in Virtual Reality Umgebungen dargestellt. In seiner Aufgabe als Museumsführer kann er sich Gesichter merken und wieder erkennen und sogar auf sprachlich geäußerte Fragen reagieren. In der Tabelle 10 sind weitere Spezifikationen zu diesen Avatar aufgeführt.

Ein weiterer Nutzen von virtuellen Charakteren befindet sich in der *Präsentation/Darstellung von Informationen*. In diesem Bereich lässt sich auch ein Gebärdensprachavatar sinnvoll ein-

gliedern. So können Informationen auf Internetseiten, Durchsagen am Bahnhof/Flughafen oder Nachrichten visuell an den Zuschauer übermittelt werden. In der Werbung/Marketing tragen sie dazu bei, ein Produkt anzupreisen. In einigen Fällen repräsentieren sie sogar eine Firma oder ein ganzes Unternehmen, wie der Avatar „Robert“ (vgl. Abb. 85) bei der Bewerbung des Börsenganges von T-Online, einem Tochterunternehmen von der Deutschen Telekom, im Jahr 2000 oder die Spielfigur „Mario“ von dem japanischen Hersteller von Videospiele und Spielekonsolen Nintendo.

Der Bereich der *Linguistik* spiegelt die wissenschaftlichen Anwendungen wieder (siehe Avatar Max von der Uni Bielefeld). Ein Avatar, der mit seiner Umgebung interagieren soll, muss Anweisungen in Form von Sprache oder Schrift verstehen und verarbeiten können. So ist es denkbar, dass Avatarsysteme als Dolmetscher Informationen (z. B. Arztgespräche, Fernsehsendungen, Internetseiten usw.) aus einem Sprachsystems in ein anderes übersetzen.

Bilder Avatare in den Anwendungsbereichen



Abb. 78: Virtuelle Konferenzen; aus [Tüm07]

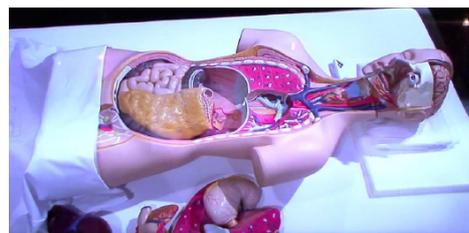


Abb. 79: Avatarmodell im Bereich der Medizin; aus [Cun12]



Abb. 80: Ergonomieuntersuchung Fahrzeuginnenraum; aus [Tüm07]



Abb. 81: Avatare aus Shrek 2; aus [Cun12]



Abb. 82: NPCs in die SIMS 2; aus [Tüm07]

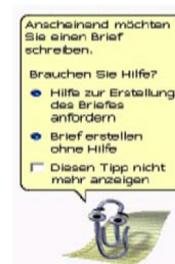


Abb. 83: Interaktiver Hilfeassistent von Microsoft Office



Abb. 84: Max als Museumsführer;
aus [Tüm07]



Abb. 85: Robert von T-Online;
aus [Dus11]

In der Machbarkeitsstudie [Kip11] sahen die gehörlosen Teilnehmer nach einer Diskussionsrunde die Anwendungszwecke von Gebärdensprachavatare im Wesentlichen im Bereich der Informationsdarstellung. So schlugen sie folgende Anwendungen vor.

Übersetzung von:

- Ansagen auf Bahnhöfen oder Flughäfen
- festen Dialogsequenzen (z. B. Steckenbleiben im Fahrstuhl)
- Formulare, Prüfungsfragen
- kurzen Informationen, wie News-Ticker, Verkehrsinfos, Ansagen, Versicherungen, Kino...
- Merkblätter (Bank, öffentliche Einrichtungen, Versicherung)
- Partei- und Wahlkampfinformationen
- Informationstafeln in Städte

Übersetzung von (bezogen auf das Internet):

- Synonyme, Lexikon, Duden
- Internetseiten (Versicherungen, Behörden, Nachrichten, Bildung)
- Online-Live-Ticker
- rechtlichen Texten (z. B. AGB = Allgemeine Geschäftsbedingungen)
- Diskussionsforen
- Videoportale (z. B. Youtube)

Für zwischenmenschliche Kommunikation (z. B. Beratungsgespräche) bevorzugen sie jedoch weiterhin in Zukunft menschliche Dolmetscher, da sie gerne diese Auswahlmöglichkeit weiterhin nutzen wollen und den Wegfall der Arbeitsplätze von Dolmetschern durch Avatare nicht unterstützen würden [Kip11].

Die folgenden zwei Abschnitte 5.2.1 und 5.2.2 enthalten eine Übersicht über allgemeine und über speziell für die Gebärdensprache erstellte Avatarsysteme. Sie sollen dazu dienen, um zu zeigen, wie Avatarsysteme in der Praxis und in diesem 5. Kapitel sowie die im 3. Kapitel gezeigten Techniken zusammenwirken. Eine ausführliche Beschreibung im Anschluss erklärt die Systeme im Bereich der Gebärdensprache genauer.

5.2.1 Übersicht: ausgewählte allgemeine Avatarsysteme

	SmartBody	Max (Bielefeld)	Greta	Elckerlyc
Hersteller/ Entwickler	ISI (<i>Information Sciences Institute</i>) / ICT	Uni Bielefeld	Catherine Pelachaud	Universität Twente
Lizenz	LGPL-Lizenz	Keine Angaben	Keine Angaben	GPL v3 Lizenz
Abbildung	86	87	88	89
Kontrollsprache/ Eingabesprache	BML	MURML	Angepasster BML-Dialekt	BML/ Angepasste Variante BMLT
Animations- technik	Gestenanimation: MOCAP oder Keyframe Technik/ Blickbewegung: prozedurale Animation	Animationssystem: ACE (<i>Articulated Communicator Engine</i>)/ prozedurale Animation	Gesichtsanimation: MPEG4-Standard/ Körper: Keyframe Technik/ prozedurale Animation	Physikalische Animation/ prozedurale Animation
Einsatzbereich	interaktives Training für das Militär (<i>Projekt MRE: Mission Rehearsal Exercise</i>)	Museumsführer in Paderborn	Ausbildungszwecke/ Lernassistent/ sprachliche Interaktionen	Sprachliche Interaktionen mit Benutzer
Verfügbarkeit	Open Source	intern	Verfügbar auf Anfrage	Verfügbar auf Anfrage
Anmerkung	Avatar mit verschiedene Rendering-Engines kompatibel	beherrscht die meisten DGS Handformen/ Max (<i>the Multimodal Assembly eXpert</i>)	ursprünglich bestehend nur aus dem Kopf → ausgeprägte Mimik/ spricht 6 Sprachen	Gesichtsverformung/ -animation durch MPEG4- Standard/ FACS/ Morph- Targets
Internetseite	http://smartbody.ict.usc.edu	http://techtak.uni-bielefeld.de/~skopp/max.html	https://trac.telecom-paristech.fr/trac/project/greta	http://elckerlyc.ewi.utwente.nl/

Tab. 10: Ausgewählte allgemeine Avatarsysteme

Bilder der Avatare

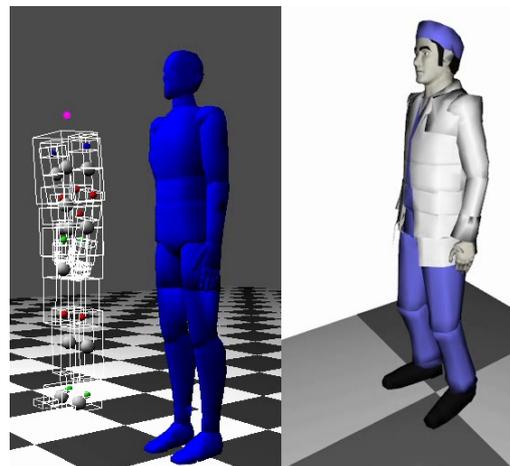
*Abb. 86: Avatar aus Projekt MRE;
aus [Len02]*



*Abb. 87: Interaktiver Museumsführer Max;
aus [Wac09]*



*Abb. 88: Avatar Greta;
von der Webseite*



*Abb. 89: Demonstrationsvideo Elckerlyc;
von der Webseite*

5.2.2 Übersicht: ausgewählte Gebärdensprach-Avatarsysteme

	Virtual Guido	Max	Vcom3D	LIMSI	EMBR
Hersteller/ Entwickler	Televirtual Ltd./ UEA	BMAS	Vcom3D	Forschungsinstitut	DFKI GmbH
Lizenz	SYS Consulting Ltd.	SYS Consulting Ltd.	© Vcom3D, Inc.	lizenziert	lizenzfrei
Gebärdensprache	DGS, BSL, NGT	DGS	ASL	LSF	DGS
Kontrollsprache/ Eingabesprache	SIGML	SIGML	Glossen	eigene Skriptsprache; Ähnlichkeiten zu XML	BML/ EMBRScript
Animations- technik	Prozedurale Animation	Prozedurale Animation	MOCAP/ IK/ FK/ Prozedurale Animation	Rotoskopie	Rotoskopie/ IK
Software-Tools	ARP/ Aningen/ JASigning-Paket	ARP/ Aningen/ JASigning-Paket	Sign Smith Studio/ Vcommunicator Gesture Builder/ Vcommunicator Studio	Keine Angaben	Blender/ ANVIL/ BehaviorBuilder
Einsatzbereich	Dolmetscher/ Gebärdensprachsynthese/ Forschung	Informationen wiedergeben	Gebärdensprache/ Weiterbildung/ Werbung/ Informationsdarstellung	Durchsagen am Bahnsteig	Gebärdensprachsynthese
Verfügbarkeit	Weiterentwicklung im Dicta-Sign Projekt	kommerziell aber auf Webseite nutzbar	kommerzielle Produkte	intern	Open Source
Anmerkung	Lippenbewegungen basieren auf SAMPA	Lippenbewegungen basieren auf SAMPA	Hohe Benutzerfreundlichkeit bei der Bedienung	Weiterentwicklung im Rahmen des EU- Projektes Dicta-Sign	Auftraggeber: BMAS Renderer: Panda3D
Internetseite	http://aningen.com	http://bmas.de	http://vcom3d.com	http://limsi.fr	http://embots.dki.de/EMBR

Tab. 11: Ausgewählte Gebärdensprach-Avatarsysteme

Der Gebärdensprachavatar „*Virtual Guido*“ (vgl. Darstellung 90) wurde im europäischen eSIGN Projekt entwickelt. Dieses Projekt ist der Nachfolger des ViSiCAST Projektes, bei dem die Animationen auf Motion Capture Technik setzte. Zur Erzeugung eines beliebigen Avatars wurde das Werkzeug *ARP (Avatar Research Platform)* entworfen. Die Geometrie basiert hier auf zweischichtige Modelle (Skelett+Haut). Für die Verformung/Anpassung des Avatars nutzt ARP Morph-Targets. Die Virtual Human (VH) Group an der University of East Anglia (UEA) konstruierte für Virtual Guido die zu ARP kompatible Animationssoftware *Animgen*, die es erlaubt mit Hilfe der Eingabesprache SIGML (vgl. Kapitel 3.3.1) Animationen basierend auf Algorithmen zu generieren. Um den Avatar in einer Anwendung wie zum Beispiel auf eine Webseite zu platzieren, wurde das auf Java basierende *JASigning-Paket* zur Verfügung gestellt. Diese drei Programme stellen ein komplettes Avatarsystem dar, weil sie die Bereiche Avatargenerierung, Animation und die Anwendungs-Platzierung abdecken. Auf der Webseite des Bundesministeriums für Arbeit und Soziales (BMAS) lässt sich die Gebärdensprach-Qualität des Avatars „*Max*“ (vgl. Darstellung 91), der ebenfalls die Softwareumgebung von Virtual Guido nutzt, betrachten.

Die US amerikanische Firma „*Vcom3D*“ befasst sich mit dem Bereich der virtuellen Kommunikationscharaktere (vgl. Abb. 92). Ihre Produkte sind zu finden in Schulen, Familien, Unternehmen und Regierungsbehörden in den Bereichen Sprache, Kultur, Schulung und pädagogische Betreuung. Ihre Software-Palette umfasst die Programme Sign Smith Studio 3.0, Vcommunicator Gesture Builder 2.0 und Vcommunicator Studio 2.0. Mit dem *Sign Smith Studio* kann der Nutzer durch Texteingabe oder direkter Auswahl von Glossen (vgl. Kapitel 3.3.1) aus einer Liste eine Abfolge von Gebärdensprachgesten ausführen. Hinter den Glossen befindet sich eine gespeicherte Animation, die im unteren rechten Rand des Programms von einem Avatar in Echtzeit ausgeübt wird. Gibt der Nutzer ein unbekanntes Wort (z. B. aus einer anderen Sprache) ein, buchstabiert der Gebärdensprachavatar dieses mit Hilfe des Fingeralphabets. Benutzer, die Kenntnisse in der Gebärdensprache haben, können die Animation einer Glosse nach belieben anpassen. So können Parameter wie Zeitdauer, Geschwindigkeit, Blickrichtung, Mimik/Emotionen (wütend, überrascht, Varianten von Lächeln), Grammatik (Mimik zu: Konditionalsätze, Relativsätze, Rhetorische Fragen, W-Fragen, Varianten des Kopfschüttelns) frei abgeändert werden. Ebenfalls bietet das Programm verschiedene Avatarmodelle, die sich aber nicht in ihrem Aussehen anpassen lassen. Preislich muss der Nutzer für dieses Produkt 495 \$ (Stand: 01/2013) zahlen. Für eine höhere Kontrolle der Gebärdenzeichen oder zur Erzeugung neuer Gebärden wurde der *Vcommunicator Gesture Builder* entwickelt. Hier kann der Benutzer aus einer Vielzahl von Handformen und Ausführungsstellen (mittels Kinematische Techniken oder vordefinierten Punkte) die Hände, Finger und Arme selber positionieren und somit eine eigene Animation für eine Glosse erstellen, die nach dem Exportieren kompatibel mit dem Sign Smith Studio ist. Der Preis beträgt bei diesem Produkt 2.500 \$ (Stand: 01/2013). Mit dem *Vcommunicator Studio* lassen sich ebenfalls Gesten erzeugen. Das Programm erzeugt keine Gesten für die Gebärdensprache, sondern Verhaltensweisen wie Laufanimationen, verschiedene Posen und Körperhaltungen. Durch Aneinanderreihung

dieser Posen können zum Beispiel natürliche Bewegungsanimationen bei Wartepausen (z. B. bei Benutzertexteingaben) als Übergang dienen. Die Kosten des Produktes liegen bei 3.500 \$ (Stand: 01/2013). Die Programme sind alle untereinander kompatibel und können auf der Vcom3D Webseite kostenlos heruntergeladen und für eine Zeitdauer von 30 Tagen in vollem Umfang getestet werden.

Ebenfalls für die Entwicklung von Gebärdensprachavatara ist die Forschungsgruppe ILES (*Information, Language, writtEn and Signed Group*) am „LIMSI-Institut“ in Frankreich zuständig. Sie entwickelten in den Jahren 2005 – 2007 einen Avatar der Durchsagen an Bahnhöfen tätigen sollte. Ihr Auftraggeber war die französische Bahngesellschaft SNCF. Der Avatar, dargestellt in Abbildung 93, kam an den beiden Pariser Bahnhöfen „Gare Montparnasse“ und „Gare de l'est“ zum Einsatz. Für die Animation wurde die Technik des Rotoskopierens genutzt, bei der aufgenommene Bewegungsabfolgen am PC in einem 3D-Programm nachgezeichnet wurden (vergleichbar mit der frame-by-frame Animationstechnik). Aufgrund der geringen Kosten und aufwändigen Arbeit wird diese Technik auch als das Motion Capture des armen Mannes bezeichnet. Die aktuelle Forschungslinie beschäftigt sich mit der neuen Eingabesprache Zebedee (siehe Anhang A2.2) zur Steuerung der Animationen. Anstatt der Technik der Rotoskopie kommt jetzt die prozedurale Animation zum Einsatz.

Der Gebärdensprachavatar „EMBR“ (*Embodied Agent Behavior Realizer*) wurde im Jahre 2008 vom DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz) in Saarbrücken entwickelt (vgl. Abb. 94). Der ursprüngliche Aufgabenbereich von EMBR lag in der sprachlichen Interaktion. Das Bundesministerium für Arbeit und Soziales (BMAS) lies in einer Machbarkeitsstudie ([Kip11]) die Nutzungsmöglichkeit von Gebärdensprachavatara untersuchen. Ein Ergebnis dieser Studie ist, dass EMBR und das System als Open Source (seit 2011) auf der gegebenen Internetseite (siehe Tabelle 11) frei verfügbar ist. Es benutzt als Animationstechnik die Rotoskopie. Hierfür wurde speziell das Werkzeug ANVIL (*The Video Annotation Research Tool*) von Michael Kipp programmiert. Dieses Werkzeug wurde genutzt um Gebärdensprachvideos zu analysieren. Dazu unterteilte es die ablaufende Gebärden in Sätze und anschließend grenzte es die einzelnen Gebärdenzeichen (Glossen) aus und speicherte diese als separate Videos ab. Zur Erstellung von Gesten/Animationen musste zuerst der Avatar, der für sprachliche Interaktion ausgelegt war, in ein Gebärdensprachavatar umgewandelt werden. Mit dem kostenlosen 3D-Modellierungs- und Animationsprogramm *Blender* erstellte die DFKI 50 neue Handformen speziell aus dem Bereich der GS. Des Weiteren erhielt das zweischichtige (Skelett+Haut) Avatar-Modell zusätzliche Halterungsposen. Für die geometrische Verformung der Gesichtsmimik kamen Morph-Targets zum Einsatz. Zwischen den erstellten Handformen, vordefinierten Körperposen und den Morph-Targets wird durch Interpolation die Animation erzeugt. Durch diese Arbeitsschritte wurde sicher gestellt, dass die Bewegungen natürlicher wirken und das geometrische Modell nicht ineinander verläuft (Stichwort: Kollisionskontrolle/-vermeidung). Nun konnte man mit der eigens hergestellten Software *BehaviorBuilder* eigene Gesten aufbauen. Die erzeugten Posen basierten auf die mit ANVIL abgespeicherten Einzelgebärdenzeichen-Videos und wurden im EMBRScript (siehe dazu

Kapitel 3.3.3) abgespeichert. Es beinhaltet somit die präzise Abfolgen von Körperbewegungen und dient als Kontrollsprache für die Animation. Somit kann z. B. durch Anweisungen im EMBRScript der Avatar sich nach vorne zu einem Punkt beugen. Dabei wendet der Avatar die inverse Kinematik an. Von den oben genannten Software-Tools ist der BehaviorBuilder nicht erhältlich. Jedoch lässt sich das Java Applet EMBR von der Internetseite kostenlos herunterladen und mittels EMBRScript-Befehle steuern [Kip11].

Bilder der Avatare



Abb. 90: Virtual Guido aus dem eSIGN Projekt; aus [Kip11]



Abb. 91: Avatar Max; von der BMAS Internetseite



Abb. 92: Vcom3D-Avatar aus dem Werbevideo: „The Forest“ by Jason Stewart; aus [Kip11]



Abb. 93: LIMSI Avatar für Bahndurchsagen; aus [Seg08]



Abb. 94: Avatarsystem EMBR aus dem Java Applet

5.3 Evaluation von Gebärdensprachavataren

Evaluationen sind nötig, um die Qualität eines Systems zu überprüfen. Aus den Ergebnissen dieser Testversuche lassen sich die Schwachstellen des Systems aufdecken und ermöglichen es Rückschlüsse zu ziehen, die die Gesamtqualität des Systems verbessern können. Zum Abschluss des Kapitels werden in den nächsten drei Unterkapiteln Methoden, Ergebnisse und Auswertungen von Gebärdensprachdarstellungen durch Avatare betrachtet.

5.3.1 Methoden

Nach aktuellem Stand gibt es noch keine Standardmethode zur Evaluation von animierter Gebärdensprachdarstellung. Das könnte an den verschiedenen Anwendungen und somit an den verschiedenen Herangehensweisen der Evaluationen für den Avatar liegen. Jedoch ähneln sich die eingesetzten Mittel zur Bewertung. Diese sind häufig:

„*Fragebögen*“, die für die Ermittlung der demografischen Informationen der Versuchspersonen (VP) dienen. Des Weiteren können vorhandene Kenntnisse über Avatare abgefragt und Tendenzen durch Umfragen per Internet für das Projekt gesammelt werden. In manchen Evaluationen wurden teilweise Fragebögen am Ende benutzt, um Kriterien, wie Korrektheit der Avatardarstellung, Aussehen/Ausstrahlung/natürliche Bewegungen des Avatars, Verständlichkeit der Aussage, Mimik, Zusammenspiel Gebärde und Mimik, Mundbild, Handform, Emotionen usw., zu bewerten. Diese Methode liefert meistens subjektive Ergebnisse/Bewertungen. Fragebögen können sowohl von den Versuchspersonen als auch von den Testbeobachtern ausgefüllt werden.

„*Multiple-Choice-Tests*“ wurden bis jetzt nur bei den Versuchen von Huenerfauth angewendet (siehe Anhang A6.1). Durch eine Auswahlmöglichkeit von mehreren Szenarien muss die Testperson den Satzinhalt einer Gebärdensprachdarstellung des Avatars erkennen. Die Ergebnisse fielen im Vergleich mit anderen Methoden relativ hoch aus. Fraglich ist, ob durch die Auswahlmöglichkeiten das Ergebnis verzerrt wird. So könnte die Versuchsperson nur wenige Worte aus dem Satz verstehen und durch die verschiedenen Auswahlmöglichkeiten sich den semantischen Inhalt des Satzes ableiten. Ähnlich wie beim Lesen eines Absatzes in einer fremden Sprache, bei dem nicht jedes Wort verstanden werden muss, um den Kontext zu verstehen. Andere objektive Methoden, wie die Befragung von Versuchspersonen, sollte vor dieser Methode bevorzugt werden.

„*Vergleich mit menschlichen Gebärdendarstellungen*“ ist eine neuere Methode. Dabei werden die Verständniswerte der einzelnen Glossen bzw. der Sätze für die Avatardarstellung und für ein Referenzvideo ermittelt. Mit Hilfe des Dreisatzes kann ein Vergleichswert berechnet werden, wobei das Video mit den menschlichen Gebärdens-Ausführungen als Referenzobjekt

dient. Diese Methode fand vermehrt Anwendung im Dicta-Sign Projekt und in der Delta-Evaluation bei dem Avatar EMBR, die ausführlich im Anhang A6.2 dargestellt wird.

„*Befragung von Versuchspersonen*“ ist die bessere Methode zur Bestimmung des Verständnisses eines Gebärdenzeichens oder einer Satzaussage. Ein Moderator befragt nach der Avatargebärdendarstellung die Versuchsperson nach dem Inhalt der Glosse oder des Satzes und notiert diese. Alternativ kann die Versuchsperson die verstandenen Glossen selber in einem Fragebogen mitschreiben. Dies sollte aber aufgrund der abgelenkten Aufmerksamkeit vermieden und von externen Person ausgeführt werden. Am Ende lässt sich die Verständlichkeit berechnen, indem die Anzahl der verstandenen Glossen gezählt und durch die Gesamt-Glossen-Anzahl geteilt wird. Abschließend kann der Moderator direktes Feedback von den Teilnehmern bekommen. In Kombination mit einem Fragebogen kann dieses noch spezieller ausgewertet werden.

In der Regel werden bei einer Evaluation immer subjektive und objektive Messwerte bestimmt. So sollen die subjektiven Bewertungen als Referenz zu den objektiven Bewertungen dienen. Die objektiven Werte können zum Beispiel verzerrt werden, weil die Versuchsperson die Bewegungen des Avatars nachahmte und somit mehr richtige verstandene Glossen gezählt wurden. Dies lässt sich überprüfen, indem ein Beobachter seine subjektive Einschätzung zum Verständnis der Versuchsperson (Mittel der Befragung) abgibt. Sind objektive und subjektive Werte danach weit auseinander, könnte das Nachahmen der Gebärden ein Grund dafür sein.

Üblich ist auch den Testpersonen vor Beginn der Evaluation eine Eingewöhnungszeit mit den Avatar zu gewähren. Somit können sie sich an den Gebärdenstil des Avatars gewöhnen. Bei der Durchführung ist es außerdem ratsam Wiederholungsmöglichkeiten der Gebärdendarstellung im Avatarsystem zu ermöglichen, da so die Ergebnisse erheblich gesteigert wurden.

Um genauer zu verstehen, wie diese Methoden zur Anwendung kommen, befindet sich im Anhang A6 eine ausführliche Beschreibung ausgewählter Experimente und deren Ergebnisse. Diese sind zur Übersicht in den Punkten „Versuchspersonen“, „Datenbasis“, „Durchführung“ und „Ergebnisse“ unterteilt.

5.3.2 Ergebnisse

Nachdem einige Methoden zur Bewertung von Gebärdensprachdarstellungen durch Avatare aufgezeigt wurden, fasst dieser Abschnitt kurz die Ergebnisse der drei großen europäischen Projekte ViSiCAST, eSIGN und Dicta-Sign zusammen, die mehrere Evaluationen zu ihren Projekten durchführten.

ViSiCAST Projekt

In dem europäischen Projekt ViSiCAST entwickelten Universitäten, Institute und Unternehmen aus verschiedenen Ländern, Systeme für drei unterschiedliche Anwendungen mit Gebärdensprachavataren. Um die Qualität dieser Systeme zu verbessern, wurden jährlich Evaluationen (siehe [She03]) durchgeführt. Im Anhang A6.3 ist der Ablauf der ersten Evaluation für den Anwendungsfall Dialogsystem in einer Poststelle aufgeführt. Dieser Abschnitt fasst die Ergebnisse des Gesamtprojektes zusammen. Insgesamt nahmen in dem 3-jährigen Projekt (2000-2003) 133 Personen an den Evaluationen teil. Die drei Anwendungsbereiche waren:

- ➔ Gebärdensprachavatar in einem Dialogsystem für die Poststelle,
- ➔ Gebärdensprachavatar für die Darstellung von Wetterauskünfte auf einer Internetseite,
- ➔ Gebärdensprachavatar zur Darstellung von Informationen im Fernseher.

Jedes Avatarsystem wurde zweimal von Fokusgruppen getestet. Die Bewertungskriterien waren:

- ➔ Verständlichkeit der Gebärdenzeichen,
- ➔ Akzeptanz der Gebärdenzeichendarstellung,
- ➔ Akzeptanz für das Erscheinungsbild des Avatars.

Die Ergebnisse sind in der Tabelle 12 zusammengefasst. Um diese drei Anwendungsbereiche im Projekt umsetzen zu können, wurde ein System entwickelt, welches den Text in die Britische, Holländische und Deutsche Gebärdensprache übersetzt (siehe 4. Kapitel, Abschnitt 4.4.1). Dieses System wurde ebenfalls getestet. Der Abschlussbericht zum Projekt ViSiCAST ist unter [Wak02] zu finden. Weitere Informationen zum Ablauf der Evaluationen und weitere Ergebnisse sind in den Bewertungsreports 1 bis 3 (siehe [She03]) des Projektes dokumentiert.

Fokusgruppen Evaluation	Gebärdenzeichenverständnis	Akzeptanz Gebärdenszeichendarstellung	Akzeptanz Erscheinungsbild Avatar
Avatar Dialogsystem	49 % (Bandbreite: 13 % - 90 %)	46 % (Bandbreite: 5 % - 100 %)	47 % (Bandbreite: 5 % - 90 %)
Avatar Internet	39 % (Bandbreite: 6 % - 86 %)	32 % (Bandbreite: 4 % - 86 %)	36 % (Bandbreite: 4 % - 85 %)
Avatar Fernseher	15 % (Bandbreite: 0 % - 60 %)	16 % (Bandbreite: 0 % - 84 %)	15 % (Bandbreite: 0 % - 73 %)
Text in GS	29 % (Bandbreite: 0 % - 79 %)	34 % (Bandbreite: 0 % - 79 %)	30 % (Bandbreite: 0 % - 91 %)

Tab. 12: Ergebnisse des europäischen ViSiCAST Projektes; Werte aus [She03], Evaluation Report 3

eSIGN Projekt (Essential Sign Language Information on Government Networks)

Im eSIGN Projekt (2002-2004) wurde der Schwerpunkt auf die Entwicklung eines Avatars für die Darstellung von Informationen auf Internetseiten gelegt. Insgesamt wurden acht Evaluationen, zwei in Deutschland und jeweils drei in England und den Niederlanden, durchgeführt. Das Ziel war es, Feedback für die Entwicklung des Gebärdensprachavatars zu bekommen. Für die Bewertung der Avatardarstellungen kamen Gruppen mit geringer Teilnehmeranzahl zur Anwendung. Die Anzahl lag zwischen 6 bis 10 und einmal bei 15 Versuchspersonen. Das Alter der Teilnehmer rangierte von 16 bis 66 Jahre. Der Ablauf war bei allen 8 Evaluationen ähnlich. Nach einer kurzen Erklärung und Einführung mit dem Avatar wurden die Aufgaben abgearbeitet und am Ende ein Fragebogen ausgefüllt.

Die Ergebnisse für die Verständlichkeit *einzelner Gebärdenzeichen* lag in England bei 70 % und in den Niederlande bei 75 %. Nach dem Auswerten des Feedbacks der Versuchspersonen und weiteren Verbesserungen des Avatars „Virtual Guido“ (siehe Tabelle 11) erreichten die Niederländer bei der letzten Evaluation eine Gebärdenzeichen-Verständlichkeit von 95 %. Jedoch kamen diese Werte nur Zustände bei Gebärdenzeichen, die hauptsächlich aus manuellen Komponenten bestanden.

Den ersten Test, bei dem die *Satzverständlichkeit* von „Virtual Guido“ getestet wurde, führte man in Deutschland durch und erhielt eine Verständlichkeit von 40,4 %. Weitere Test in England mit 46 % und Niederlande mit 35 % bestätigten annähernd den Wert aus Deutschland. Im Laufe der Entwicklung des Projektes verbesserte sich der Wert der Satzverständlichkeit weiter auf 58 % bis zu 62 %. In England erreichte man sogar 71 %, wobei die Testpersonen neben dem Avatar auch den englischen Text auf dem Bildschirm sahen [Ken07].

Nach dem Ende des ViSiCAST und eSIGN Projektes fassten die Forscher ihre Ergebnisse zusammen und waren der Auffassung:

„The evaluations show that comprehension of synthetic signing has improved during the project as a result of improvements made to the animation software resulting from the earlier tests and also from informal feedback. However, the project duration was too short to be able to incorporate all of the desired improvements. We are aware of some gaps in the avatar’s repertoire of nonmanual bodily behavior and facial animation, and there are some issues concerning manual components that need to be further developed. We believe that with these techniques we would be able to raise the recognition rates for continuous signed text to the levels of above 90 % that we have achieved for isolated signs [Ken07].“

Dicta-Sign Projekt

Das EU-finanzierte Forschungsprojekt Dicta-Sign beschäftigte sich mit dem Thema der Online-Kommunikation für gehörlose Gebärdensprachnutzer. Das Ziel dieses 3-jährigen Projekts (2009-2012) war es, die Möglichkeiten der Web 2.0-Technologien leichter zugänglich zu machen. Um eine gebärdensprachliche Interaktion im Web 2.0 nutzen zu können, wie zum Beispiel Mitteilungen in Blogs bzw. Diskussionsforen zu veröffentlichen, Beiträge zu

ergänzen oder abzuändern (z. B. Wikipedia) oder Informationen einfach zu teilen (z. B. Facebook), mussten neue Technologien entwickelt werden. Mit diesen konnte der Nutzer über seine Webcam Gebärden erfassen lassen, die der Computer in eine interne Repräsentation konvertierte und über einen Avatar wiedergab. Diese Inhalte konnten dann im Internet verbreitet werden. Der Abschlussbericht des Projektes ist unter [Eft12] zu finden.

Innerhalb des Projektes wurden neun Evaluationen durchgeführt. Drei davon werden kurz aufgeführt. Für den Test der *Gebärdenzeichenerkennung* konnten die Versuchspersonen den Avatar fünfmal die Gebärdenzeichen wiederholen lassen. Anschließend folgte im sechsten Durchlauf ein Video bei dem ein Gebärdender das Zeichen darstellte. Das Ergebnis der Verständlichkeit für den ersten Durchlauf ergab 58 %. Bis zur dritten Wiederholung stieg der Wert auf 83 %. Der Gebärdende in der sechsten Darstellung erreichte ein Wert von 98 %.

Bei dem Test der *Satzverständlichkeit* wurde die Geschwindigkeit der Avatardarstellung um 50 % verlangsamt. Der Wert für die Erkennung einzelner Gebärden lag bei 80 % und stieg beim fünften Mal Anschauen auf 95 %. Das Verständnis des Satzinhaltes hatte eine Bandbreite von 33 % bis 62 %.

In einer weiteren Evaluation wurde ein Avatar auf Wikipedia genutzt. Dieser gebärdete einen kurzen Satz, der von Versuchspersonen bewertet und verändert werden konnte. Die Verständlichkeit der Satzaussagen betrug 75 % bis 85 % [Web11]. Die Tabelle 19 im Anhang A6.4 zeigt einen Überblick über alle neun Evaluationen. Ein ausführlicher Bericht über die Durchführung, Ergebnisse und Auswertungen dieser Beurteilungen ist in [Web11] zu finden.

Fazit

Der höchste Wert für die Verständlichkeit einzelner Gebärdenzeichen liegt bei 83 %. Bestehen diese Gebärden größtenteils aus manuellen Komponenten ist sogar ein Wert von 95 % möglich. Bei der Satzverständlichkeit liegt das beste Ergebnis bei 62 %. Durch die Möglichkeit der Geschwindigkeitsregulierung beim Avatar konnten die Werte ebenfalls verbessert werden. Außerdem hat sich gezeigt, dass durch Wiederholung der Gebärdendarstellung oder Untertitel beim Abspielen der Avataranimation die Werte positiv verbesserten. Negativ ist die große Bandbreite der Ergebnisse. So traten bei Avatarsystemen mit guten Durchschnittsergebnissen auch sehr schlechte Verständlichkeitswerte in den Evaluationen auf. Jedoch lässt sich beim Verlauf der europäischen Projekte eine kontinuierliche Steigerung der Akzeptanz- und der Verständlichkeitswerte erkennen.

5.3.3 Auswertungen und Erkenntnisse

Bei den Auswertungen der verschiedenen Evaluationen traten meistens die gleichen Erkenntnisse auf. Es gibt zwei Hauptgründe, warum die Werte in der Satzverständlichkeit noch nicht im Bereich von 90 % bis 100 % liegen.

Der eine Grund ist die Vielfalt der Gebärdensprache. So ergab sich nach Diskussionen mit Versuchspersonen die Erkenntnis, dass einige Gebärdenzeichen von den Gebärdenden in den Versuchstest nicht erkannt wurden, weil sie die Gebärden gar nicht oder nur in einer anderen Bedeutung kannten. Es gibt kein standardisiertes Sprachwerk in der Gebärdensprache, wie zum Beispiel in der deutschen Sprache den Duden. Somit variieren Gebärdenzeichen von Region zu Region innerhalb eines Landes oder einer Gebärdensprache. Die Hamburger Universität will sich diesem Problem annehmen und möchte ein Wörterbuch (Deutsch : DGS) bis zum Jahr 2020 in elektrischer Form produzieren.

Das zweite Problem lag in der technischen Darstellung und Ausführung der Avatare. Nach den Verständniswerten sind die manuellen Komponenten nahezu alle verständlich. Jedoch wurden die nicht-manuellen Komponenten stark kritisiert, im Speziellen die Mimik des Avatars. So bemängelten Versuchspersonen, dass die Mimik nicht ausgeprägt war oder an der falschen Stelle – in Bezug auf die manuellen Komponenten – ausgeführt wurde. Weiterhin störten die unnatürlichen Bewegungen, die in Zukunft wesentlich verbessert werden müssen. So sahen viele Bewegungen unflexibel aus und wirkten roboterhaft. Die äußerliche Gestalt des Avatars gefiel einigen Versuchspersonen auch nicht. Eine Lösung wäre ein Avatarsystem, bei dem der Nutzer das Aussehen (Geschlecht, Kleidung, Farben, Kamerawinkel usw.) selber einstellen kann. Lösungen für diese Probleme sind schon vorhanden, wurden aber alle noch nicht in einem System zusammengefügt.

Weitere Gründe, die eine schnelle Entwicklung im Bereich der Gebärdensprachavatare abmildert, sind:

- ➔ die Forschungsgemeinde fällt sehr klein aus,
- ➔ die Fördermittel sind schwer zu akquirieren und sind nicht für längerfristige Projekte ausgelegt (Problem der fehlenden Langzeitförderung),
- ➔ Gebärdensprachavatar ist ein Querschnittsthema, d. h. Wissen aus der Linguistik, Gebärdensprache, Computeranimation, Modellierung und Maschinelle Übersetzung sind nötig [Kip11].

Der linguistische Aspekt ist größtenteils korrekt. Es kam lediglich zu leichten Diskrepanzen in der Grammatik, da zum Beispiel Fragen auch anders in der GS formuliert werden können und einzelne Gebärden mit mehrfacher Bedeutung in den Tests falsch gewählt wurden. Maschinelle Übersetzungen kamen in diesem Bereich eher selten zur Anwendung. Die Sätze wurden meistens manuell erstellt und in einem System einprogrammiert.

Interessant ist auch, dass bei der Bewertung im Anhang A6.3 die gehörlosen Versuchspersonen ein Dialogsystem in einer Poststelle mit Avatar schlechter bewerteten als ein Dialog direkt mit hörenden Menschen. Ähnliche Erkenntnisse gibt es auch in der Machbarkeitsstudie

[Kip11], bei der sich aus einer Diskussionsrunde und Internetbefragung herausstellte, dass die Gebärdende persönliche Gespräche lieber direkt oder über einen Dolmetscher führen wollen. Dies begründen sie mit der Aussage, dass sie Angst haben, den Beruf der menschlichen Dolmetscher durch die Verwendung von Avatare zu gefährden. Deshalb wünschen sie sich für die Zukunft die Möglichkeit der Wahl zwischen Avatar und Dolmetscher.

Auffällig sind die geringen Anzahlen der Teilnehmer und die geringe Datenbasis bei den Versuchen. Die geringe Datenbasis lässt sich durch den hohen Aufwand für die Animation einer Glosse begründen. Tools ermöglichen zwar schon eine kürzere Zeit für die Erstellung, sind aber noch am Anfang der Entwicklung und müssten dementsprechend noch optimiert werden. Alternativ könnte die Datenbasis mit Hilfe von der gehörlosen Gemeinschaft vergrößert werden, jedoch müssten dazu mehr Tools als Open Source veröffentlicht werden. Des Weiteren fällt auf, dass es noch keine Versuche gab, bei dem kleinere Geschichten oder Textabsätze vom Avatar gebärdet wurden. Die Ergebniswerte beziehen sich entweder auf einzelne Gebärdenzeichen oder einzelnen Sätze.

6. Kapitel: Zusammenfassung der Ergebnisse und Ausblick

6.1 Zusammenfassung der Ergebnisse

Der erste Abschnitt in der Arbeit betrachtete die Phonologie und Grammatik der Deutschen Gebärdensprache. Als Ergebnis wurden drei Übersichtsgrafiken zu den Bereichen „manuelle Komponenten“, „nicht-manuelle Komponenten“ und „Grammatik der DGS“ erstellt. Sie geben einen schnellen Überblick über die Funktionsweise und den Aufbau der Deutschen Gebärdensprache. Basierend auf diesen Zusammenfassungen ließen sich die Anforderungen für das Bewegungsrepertoire und der Animation eines Gebärdensprachavatars ableiten. Diese können ohne Weiteres für eine andere Gebärdensprache als die DGS übernommen werden. Der Avatar sollte sich demnach so nah wie möglich an den Bewegungsablauf des Menschen orientieren. Die manuellen Komponenten sind größtenteils erforscht und können bei vernünftiger detaillierter Modellierung des Avatars ohne größere Schwierigkeiten umgesetzt werden. Problematisch sind die nicht-manuellen Komponenten. Insbesondere die Mimik mit der Augen- und Mundpartie muss sehr genau modelliert und anschließend animiert werden, um eine gute Gebärdensverständlichkeit des Avatars zu erreichen. Nicht zu unterschätzen ist die Akzeptanz der dargestellten Animation. Wirkt der Avatar nicht sympathisch, z. B. wegen der Umgebung oder Gestaltung des Avatars selber, oder ist sogar unnatürlich in seinem Bewegungsablauf, fällt die Akzeptanz und somit indirekt die Konzentration des Nutzers. Das kann zur Folge haben, dass die Verständlichkeit des Avatars negativ beurteilt wird. Deshalb ist es wichtig, den Nutzer Möglichkeiten zu geben, die Animationsdarstellung zu personalisieren oder interaktiv anzupassen. So kann die Auswahl der Umgebung, die Wahl zwischen verschiedenen Avatar-Modellen, die Möglichkeit die Perspektive frei zu ändern sowie die Geschwindigkeitsregulierung und die Wiederholungsmöglichkeit der Gebärdendarstellung die Verständlichkeit des Avatars positiv beeinflussen. Ein natürliches Verhalten wie fließende Bewegungen und sanfte Übergänge zwischen den Animationen oder natürliches Verhalten bzw. Reflexe (z. B. Atmen, Blinzeln, Gewichtsverlagerung beim Stehen) steigern das Verständnis und den positiven Eindruck eines Gebärdensprachavatars.

Der zweite Abschnitt behandelte Schriftsysteme für die Gebärdensprache. Die erste Erkenntnis war, dass für die Gebärdensprache eines Landes kein einheitlich genutztes Schriftsystem existiert. Somit wurde eine Übersicht erstellt, die 22 Schriftsysteme in den drei Bereichen „Menschliche Kommunikation“, „Linguistische Notation“ und „Animationsorientierte Notation“ einteilt. Ausgewählte Schriftsysteme aus diesen Bereichen wurden kurz vorgestellt und erläutert.

Es stellte sich heraus, dass die animationsorientierten Notationen, die für die Steuerung eines Gebärdensprachavatars geeignet sind, sich in zwei Gruppen einteilen ließen. Zum einen die „Notations-basierten“ Schriftsysteme, deren grundlegender Aufbau sich an eine linguistische Notation orientieren, und zum anderen die „Verhaltens-basierten“ Notationen, die unabhängig von anderen Notationen sind. Beide Arten können für einen Gebärdensprachavatar benutzt werden, wobei bei den „Verhaltens-basierten“ Notationssystemen der Aufwand für die Anpassungen an die Gebärdensprache größer ausfällt.

Im dritten Teil der Arbeit wurde untersucht, wie Maschinelle Übersetzungstechniken genutzt werden können, um eine automatische Übersetzung von einem Text aus der Lautsprache in eine animationsorientierte Notation zu überführen. Eine Recherche über die Maschinellen Übersetzungsstrategien ergab, dass die gebräuchlichsten Verfahren für eine Übersetzung die „Regel-basierten“ Verfahren und die „Daten-basierten“ Verfahren sind. Die „Regel-basierten“ Verfahren ließen sich wiederum unterteilen in „Direkte Übersetzung“, „Transfer Übersetzung“ und „Interlingua“ und die „Daten-basierten“ Verfahren in „Statistik-basierte MÜ“ und „Beispiel-basierte MÜ“. Zu diesen Strategien wurden wissenschaftliche Arbeiten im Bereich der Gebärdensprachavatare recherchiert und in einer Übersicht zu diesen zugeordnet. Es stellte sich heraus, dass bereits in den verschiedenen MÜ-Strategien Versuche mit Gebärdensprachavataren vollzogen wurden. Des Weiteren wurden MÜ-Strategien kombiniert, die sogenannten „hybriden MÜ-Systeme“, um die Schwächen der einzelnen Strategien zu umgehen und somit die Übersetzungsqualität zu steigern. Auch hier gab es schon erste Versuche mit Gebärdensprachavataren. Um die verschiedenen MÜ-Systeme vergleichen zu können, wurden die Evaluationsmetriken WER, PER, TER, SER, BLEU, NIST und METEOR vorgestellt. Aus den bereits vorhandenen Avatarsystemen wurden Kenntnisse für die Übersetzung von Texten aus der Lautsprache in eine animationsorientierten Notation gewonnen, die ein Workflow jeweils für die RBMT, EBMT und SBMT Strategie beispielsweise darstellt.

Eine animationsorientierte Notation steuert anschließend die Animation eines Avatar-Modells. Der vierte Abschnitt der Arbeit betrachtete die Techniken für die Gestaltung/Modellierung und der Animation eines solchen Avatar-Modells. Wie schon erwähnt, kann das Aussehen des Modells die Akzeptanz der Gebärdendarstellung beeinflussen. Für das Design eines Avatars wäre eine fotorealistische oder Cartoon-artige menschliche Figur sinnvoll. Die cartoonhafte Darstellung des Avatars wäre jedoch empfehlenswert, da die Ansprüche des Betrachters geringer sind, als wenn der Avatar fotorealistisch wirkt. Sollte dennoch der Avatar dem Menschen fotorealistisch nachempfunden sein, muss das Modell den Menschen in allen seinen Details abbilden, weil sonst die Akzeptanz und somit die Konzentration des Betrachters negativ beeinträchtigt wird und darunter die Verständlichkeit der Gebärdendarstellung leidet. Des Weiteren beinhaltet der vierte Abschnitt einen Vorschlag für die Modellierung und Animation eines Avatars. Außerdem zeigte sich, dass die ermittelten Anforderungen für einen Gebärdensprachavatar aus dem ersten Teil der Arbeit alle bereits technisch realisierbar sind.

So wurden im 5. Kapitel neben den Anwendungsgebieten für Avatarsysteme auch „Allgemeine-Avatarsysteme“ und „Gebärdensprach-Avatarsysteme“ vorgestellt. Für die Bewertung dieser Systeme wurden die Evaluationsmethoden „Fragebogen“, „Multiple-Choice-Test“, „Vergleich mit menschlichen Gebärdendarstellungen“ und „Befragung von Versuchspersonen“ ermittelt. Nach der Betrachtung vorhandener Evaluationen stellte sich heraus, dass die bis jetzt von Avataren dargestellten Gebärdensanimationen eine Verständlichkeit für einzelne Gebärdenzeichen von 83 % erreichten. Die Satzverständlichkeit liegt bei 62 %. Die Interaktionsmöglichkeiten der Geschwindigkeitsregulierung und der Wiederholung der Gebärdensanimationen sowie zusätzlich eine textuelle Darstellung in einer Gebärdennotation auf dem Bildschirm beim Gebärdensprachavatar ermöglichten eine Steigerung der Verständlichkeitswerte. Die schwankende Bandbreite der Ergebnisse und die geringe Anzahl der Probanden bei den Versuchen sowie Evaluationen mit längeren Sätzen bzw. Absätzen wurden noch nicht durchgeführt und könnten in Zukunft verbessert werden. Positiv ist jedoch, dass sich die Verständlichkeitswerte in den letzten 10 Jahre für die Gebärdensprachavatare kontinuierlich steigern ließen.

Nach der Zusammenfassung der Ergebnisse, kann das eingangs gestellte Problem, welches nochmals in Abbildung 95 ersichtlich ist, gelöst werden.

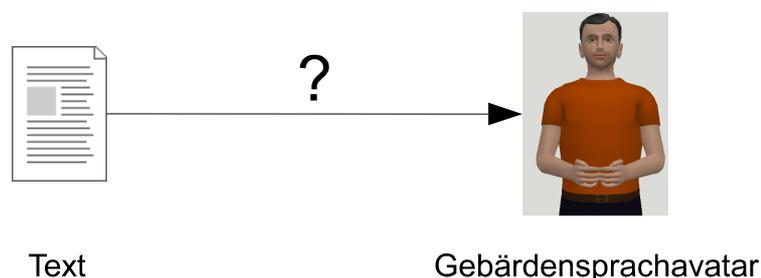


Abb. 95: Problemdarstellung der Arbeit

Ein Lösungsvorschlag für die Verarbeitungsschritte von Texten in eine Darstellung mit einem Gebärdensprachavatar zeigt die Abbildung 96. Ein Text aus der Lautsprache wird mittels einer Maschinellen Übersetzung in die Glossen-Notation übersetzt. Dabei werden die Grammatik und die Wörter der Ausgangssprache in die Grammatik und Wörter der Gebärdensprache umgewandelt. Für die detaillierte Betrachtung des Übersetzungsprozesses sei auf den Abschnitt 4.5 im 4. Kapitel verwiesen. Die Zwischenrepräsentation der Glossen wurde gewählt, weil diese leicht verständlich/lesbar ist und sich an den Wörtern der Lautsprache orientiert. Nach der Maschinellen Übersetzung kann eine MÜ-Evaluation die Qualität der Übersetzung überprüfen. Diese Zwischenevaluation ist sinnvoll, um Probleme vor der weiteren Verarbeitung im nächsten Schritt zu entdecken und zu beseitigen. In der Transkriptionsphase wird die Glossen-Notation in eine animationsorientierte Notation umgeschrieben. Dazu müssen geeignete Wörterbücher erstellt und Computerprogramme implementiert werden, die die Transkription durchführen. Ein „Notations-basiertes“ Schriftsystem wie zum Beispiel SIGML, welches auf HamNoSys aufsetzt, könnte akzeptable Ergebnisse erreichen.

Die animationsorientierte Notation steuert anschließend die Gebärdendarstellung des Avatar-Modells, welches auf einem Bildschirm ausgegeben werden kann. Eine abschließende Bewertung überprüft die Gesamtqualität der Gebärdendarstellung des Avatars. Ist die Ausgabe der Gebärdenzeichen durch den Avatar unzureichend, müssen Anpassungen bzw. Nachbearbeitungen in der Transkriptionsphase oder an dem Avatar-Modell (z. B. bei schlechter Modellierung einer Handform) ausgeführt werden. Durch die Zwischenrepräsentation der Glossen-Notation ist die Fehlersuche auf bestimmte Phasen des Entwicklungsprozesses beschränkt. Alternativ könnte auch die Maschinelle Übersetzung eingesetzt werden, um den Text direkt in eine animationsorientierte Notation zu übersetzen, was aber bedeutet, dass die Fehlersuche aufwändiger ausfallen kann.

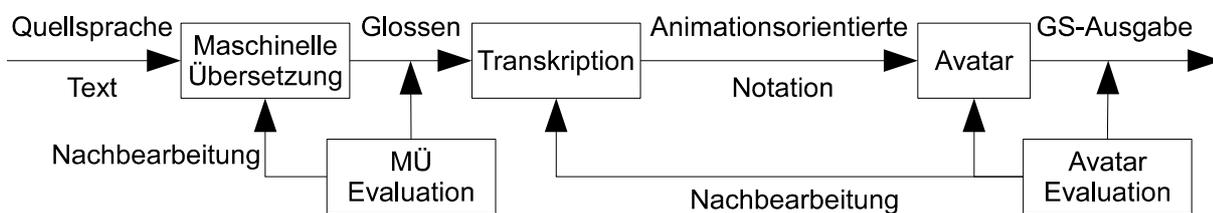


Abb. 96: Verarbeitungsschritte zur Umsetzung von einem Text in eine Gebärdensprachavatar-Darstellung

Jedoch ist ein vollständig automatischer Übersetzungsprozess mit sehr viel Aufwand verbunden. So muss je nach gewählter MÜ-Strategie ein Regelwerk festgelegt oder ein paralleler Korpus erstellt werden. Zusätzliche Tools oder Datenbanken müssen den Schritt der Transkription erledigen. Eine Alternative wäre, Interaktionsmöglichkeiten in den Verarbeitungsphasen eines Avatarsystems zu implementieren, die erlauben, dass der Nutzer an verschiedenen Stellen (z. B. Auswahlmöglichkeiten in der Übersetzung) korrigierend eingreifen kann.

6.2 Ausblick

Die Arbeit hat gezeigt, dass die Nutzung von Gebärdensprachavataren zur Informationsdarstellung sinnvoll ist. Jedoch gibt es noch einige Schwierigkeiten, die untersucht und gelöst werden müssen. Eines der Hauptprobleme ist die Gebärdensprache selber. So gibt es in einem Land kein übergeordnetes anerkanntes System zur Darstellung der einzelnen Gebärden, die alle Gehörlosen nutzen. Die Gebärden und deren Bedeutung unterscheidet sich von Region zu Region. Die Einigung auf einen Gebärdensstil und auf ein Schriftsystem würde bei der Entwicklung eines Avatarsystems hilfreich sein.

Weitere Arbeiten könnten sich speziell mit dem Thema der Maschinellen Übersetzung beschäftigen. Divergenzen, Ambiguitäten und weitere linguistische Mehrdeutigkeiten können Probleme für den MÜ-Prozess bereiten. Inwieweit diese Herausforderungen für die Maschinelle Übersetzung in der Gebärdensprache auftauchen, wurde noch nicht ausreichend untersucht.

In dem vorgestellten Lösungsansatz arbeitet das Übersetzungssystem mit Glossen. Der Avatar benötigt jedoch eine animationsorientierte Notation. Die Verbindung zwischen beiden Notationen wurde im Lösungsansatz mit der Phase der Transkription beschrieben. Eine genauere Untersuchung dieses Schrittes wäre nötig. So müsste geklärt werden, ob eine Datenbank oder ein speziell erstelltes Wörterbuch diese „Umschreibung“ erledigen kann oder ob es andere Möglichkeiten gibt.

Des Weiteren können mit den genannten Computerprogrammen in der Arbeit die verschiedenen Maschinellen Übersetzungsstrategien für die Gebärdensprache praktisch umgesetzt werden. Dazu müssten für ein RBMT-System Übersetzungsregeln definiert werden, die die Grammatik der genutzten Lautsprache in die Grammatik der Gebärdensprache umwandelt. Die Verwendung eines EBMT- oder SBMT-Systems setzt voraus, dass ein ausreichend großer paralleler Korpus, denkbar wäre hier ein bilingualer Korpus von Text und animationsorientierte Notation oder Text und HamNoSys, für die Gebärdensprache erstellt wird. Sind diese Punkte erfüllt, lassen sich die verschiedenen Übersetzungsstrategien in der Praxis vergleichen und Vor- und Nachteile feststellen.

Ebenfalls interessant wäre die Übersetzung in der umgekehrten Richtung. Welche Schritte sind nötig, um die visuelle Sprache des Gehörlosen in eine textuelle oder sprachliche Darstellung der Lautsprache zu wandeln? Ist die Kommunikation in beiden Richtungen möglich, würde sich die Frage von Dialogsystemen stellen. Könnten diese Systeme die zwischenmenschliche Kommunikation zwischen den Gehörlosen und den hörenden Menschen erleichtern? Und wie hoch ist ihr Nutzen im Alltag eines gehörlosen Menschen?

Ein weiteres Thema könnte sich mit dem Aussehen und die damit verbundene Nutzerakzeptanz des Avatar-Modells beschäftigen. Mit unterschiedlich modellierten Avatar-Modellen ließen sich in einer Studie die Präferenzen von Probanden feststellen und so Design-Parameter ableiten, die positive oder negative Einflüsse in der Darstellung des Avatars haben. Zusätzlich könnte in dieser Untersuchung eines der Avatar-Modelle die in dieser Arbeit festgelegten Anforderungen beachten. So würde der Schwerpunkt bei der Modellierung und der Animation auf die nicht-manuellen Komponenten, insbesondere Mimik, Blickbewegung und Mundbewegung, liegen. Evaluationen könnten dann die Anforderungen praktisch bestätigen bzw. widerlegen und in einem Ranking die relevanten Anforderungen einordnen. Des Weiteren kann untersucht werden, an welchem Punkt ein fotorealistischer Avatar vom Betrachter höherwertiger eingeschätzt wird als ein cartoon-artiger Avatar.

Die bereits vorhandenen Evaluationsergebnisse wurden größtenteils auf einzelne Gebärden oder kurzen Sätzen und mit einer geringen Anzahl von Teilnehmern durchgeführt. In einer weiteren Untersuchung könnten verschiedene Avatarsysteme mit den unterschiedlichen vorgestellten Methoden evaluiert werden, wobei die Datenbasis größer und die Teilnehmeranzahl höher ausfallen sollte. Anbieten würde sich eine Evaluation über das Internet.

Literaturverzeichnis

- [Bak09] Baker, Anne; Woll, Benice: Sign Language Acquisition. John Benjamins Publishing Co-Verlag, 2009.
- [Ban05] Banerjee, Satanjeev; Lavie, Alon: METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. Language Technologies Institute, Carnegie Mellon University, Pittsburgh, 2005.
- [Bau07] Bauer, Britta; Nießen, Sonja; Hienz, Hermann: Towards an Automatic Sign Language Translation System. Department of Technical Computer Science, Aachen University of Technology (RWTH), 2007.
- [Ber11] Berger, Ulrich: Vorlesung: NC- und Robotertechnik – Bewegungsprogrammierung. BTU Cottbus, Lehrstuhl für Automatisierungstechnik, 2011.
- [Blu06] Blume, Claudia: Gebärdensprache: Eine phänomenale oder eine normale Sprache?. Artikel in: ADÜ NORD (Assoziierte Dolmetscher und Übersetzer in Norddeutschland e.V.), 2006.
- [BML12] o.V.: BML 1.0 Standard. URL: <http://www.mindmakers.org/projects/bml-1-0/wiki#BML-10-Standard>, Datum der Recherche: 11/2012.
- [Bon09] Bond, Christine: Statistische maschinelle Übersetzung von Dari nach Englisch unter Einbeziehung linguistischer Faktoren bei begrenztem Trainingsmaterial. Magisterarbeit, Philosophische Fakultät, Universität Bonn, 2009.
- [Bou00] Boutsis, Sotiris; et al.: A Robust Parser for Unrestricted Greek Text. Institute for Language and Speech Processing, National Technical University of Athens, 2000.
- [Boy95] Boyes Braem, Penny: Einführung in die Gebärdensprache und ihre Erforschung. 3. überarbeitete Auflage, SIGNUM-Verlag, 1995.
- [Bro90] Brown, Peter F.; et al.: A Statistical Approach to Machine Translation. 1990.
- [Bro93] Brown, Peter F. et al.: The Mathematics of Statistical Machine Translation: Parameter Estimation. IBM T.J. Watson Research Center, 1993.

- [Car01] Carstensen, Kai-Uwe; et al.: Computerlinguistik und Sprachtechnologie. Eine Einführung, Spektrum Akademischer Verlag, 2001.
- [Cos01] Costa, Antônio Carlos da Rocha; Dimuro, Graçaliz Pereira: A SignWriting-Based Approach to Sign Language Processing. Universität Católica de Pelotas, 2001.
- [Cox02] Cox, Steven; et al.: The Development And Evaluation Of A Speech To Sign Translation System To Assist Transaction. ViSiCAST Projekt, 2002.
- [Cun12] Cunningham, Douglas W.: Vorlesung Modellierung, Bearbeitung und Visualisierung von 3D Objekten. BTU Cottbus, Lehrstuhl Grafische Systeme, 2012.
- [Das07] Dasgupta, Tirthankar; Dandpat, Sandipan; Basu, Anupam: Prototype Machine Translation System From Text-To-Indian Sign Language. Indian Institute of Technology, Kharagpur, 2007.
- [Dot01] Dotter, Franz: Gebärdensprachforschung. URL: http://www.uni-klu.ac.at/zgh/downloads/Dotter_GSforschung_2001.pdf, Datum der Recherche: 10/2012.
- [Dot12] Dotter, Franz; Holzinger, Daniel: Typologie und Gebärdensprache: Sequentialität und Simultanität, URL: http://www.uniklu.ac.at/zgh/downloads/Dotter_Typol_GS_Sequenti.pdf, Datum der Recherche: 10/2012.
- [Dus11] Duschek, Alexander; Schuster, Philipp; Tu, Xi: Avatar Frameworks. Fachstudie Nr.130, Institut für Visualisierung und Interaktive Systeme (VIS), Abteilung Graphisch-Interaktive Systeme (GIS), Universität Stuttgart, 2011.
- [Ebl12] Ebling, Sarah: Maschinelle Übersetzung von Deutsch in die Deutschschweizerische Gebärdensprache. PowerPoint Präsentation, Institut für Computerlinguistik, Universität Zürich, 2012.
- [Eft12] Efthimiou, Elni: Sign Language Recognition, Generation and Modelling with application in Deaf Communication. Abschlussbericht: Dicta-Sign Projekt, 2012.
- [Ell12] Elliott, R.; et al.: SiGML Notation and SiGMLSigning Animation Software. School of Computing Sciences, UEA Norwich, UK, URL: http://www.visicast.cmp.uea.ac.uk/eSIGN/Images/Poster_sigmlSigningREA1.pdf, Datum der Recherche: 11/2012.

- [Erl00] Erlenkamp, Sonja: Syntaktische Kategorien und lexikalische Klassen. Typologische Aspekte der Deutschen Gebärdensprache, LINCOM EUROPA Verlag, 2000.
- [Fac13] o.V.: Facial Action Coding System. URL: http://en.wikipedia.org/wiki/Facial_Action_Coding_System, Datum der Recherche: 01/2013.
- [Fil09] Filhol, Michael: Zebedee: a lexical description model for Sign Language synthesis. URL: <http://perso.limsi.fr/filhol/research/files/Filhol-2009-Zebedee.pdf>, 2009.
- [Flü08] Flückiger, Barbara: Visual Effects. Filmbilder aus dem Computer, Schüren Verlag, 2008.
- [Geo02] George, Doddington: Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics. ARPA Workshop on Human Language Technology, 2002.
- [Gia97] Giambruno, Mark: 3D Graphics and Animation: From Starting Up to Standing Out. Indianapolis, New Riders Publishing, 1997.
- [Gol96] Goldberg, Athomas; Perlin, Ken: Improv: A System for Scripting Interactive Actors in Virtual Worlds. Media Research Laboratory, Department of Computer Science, New York University, 1996.
- [Gra13] o.V.: Grafikpipeline. URL: <http://de.wikipedia.org/wiki/Grafikpipeline>, Datum der Recherche: 01/2013.
- [Gri99] Grieve-Smith, Angus B.: English to american sign language machine translation of weather reports. University of New Mexico, 1999.
- [Gro99] Grobel, Kirsti: Videobasierte Gebärdenspracherkennung mit Hidden-Markov-Modellen. VDI Verlag, 1999.
- [Hän04] Hänel, Barbara: Der Erwerb der Deutschen Gebärdensprache als Erstsprache. Die frühkindliche Sprachentwicklung von Subjekt- und Objektverbkongruenz in DGS, Gunter Narr Verlag, 2004.

- [Ham12] o.V.: HamNoSys im Überblick. URL: <http://www.sign-lang.uni-hamburg.de/projekte/hamnosys/hns3.0/hamnosysintroduce.html>, Datum der Recherche: 11/2012.
- [Han04] Hanke, Thomas: HamNoSys-Representing Sign Language Data in Language Resources and Language Processing Contexts. Institut Deutsche Gebärdensprache und Kommunikation Gehörloser, Uni Hamburg, 2004.
- [Hel09] Heloir, Alexis; Kipp, Michael: Realtime Animation of Interactive Agents: Specification and Realization. DFKI, Embodied Agents Research Group, 2009.
- [Hel10] Heloir, Alexis: EMBR Documentation. A Realtime Animation Engine for Interactive Embodied Agents. DFKI, Embodied Agents Research Group, 2010.
- [Hel11] Kipp, Michael; Heloir, Alexis; Nguyen, Quan: Sign Language Avatars: Animation and Comprehensibility. DFKI, Embodied Agents Research Group, Saarbrücken, 2011.
- [Hoh07] Hohenberger, Annette: The phonology of sign language. Powerpoint Präsentation, 2007.
- [Höh05] Höfeld, Stefanie; Kwiatkowski, Melanie; Nather, Juliane: Maschinelle Übersetzung. PowerPoint Präsentation, URL: http://user.phil-fak.uni-duesseldorf.de/~petersen/Einf_CL/CL_MaschUbersetzung.pdf, Universität Düsseldorf, 2005.
- [Hua02] Huang, Zhisheng; Eliëns; Visser, Cees: STEP: A Scripting Language for Embodied Agents. Vrije University, Amsterdam, 2002.
- [Hue06] Huenerfauth, Matt: Generating American Sign Language Classifier Predicates for English-to-ASL Machine Translation. Dissertation in Computer and Information Science, University of Pennsylvania, 2006.
- [Hue07] Huenerfauth, Matt; et al.: Design and Evaluation of an American Sign Language Generator. Computer Science Department, CUNY Queens College, New York, 2007.
- [Jac06] Jackèl, Dietmar; Neunreither, Stephan; Wagner, Friedrich: Methoden der Computeranimation. Berlin [u.a.], Springer Verlag, 2006.

- [Kan09] Kanis, Jakob; Müller, Luděk: Advances in Czech - Signed Speech Translation. Faculty of Applied Sciences, University of West Bohemia, 2009.
- [Kar04] K. Karpouzis; et al.: Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture. Video and Multimedia Systems Lab, Athen, 2004.
- [Kar09] Karnagel, Kathleen: Vergleichende Untersuchungen zur Sprach- und Gebärdensprache. BTU Cottbus Lehrstuhl Kommunikationstechnik, 2009.
- [Kat08] Kato, Mihoko: A Study of Notation and Sign Writing Systems for the Deaf. Toyohashi University of Technology, 2008.
- [Ken02] Kennaway, Richard: Synthetic Animation of Deaf Signing Gestures. School of Information Systems, University of East Anglia, Norwich, 2002.
- [Ken07] Kennaway, J.R.; Glauert, J.R.W.: Providing Signed Content on the Internet by Synthesized Animation. University of East Anglia, UK and I. ZWITSERLOOD Viataal, NL, 2007.
- [Kip11] Kipp, Michael; Nguyen, Quan; Heloir, Alexis: Machbarkeitsstudie zur Abschätzung der Nutzungsmöglichkeiten von Gebärdensprachavataren. Forschungsbericht Sozialforschung 417, URL: <http://www.bmas.de/DE/Service/Publicationen/Forschungsberichte/Forschungsberichte-Teilhaber/fb417.html>, 2011.
- [Koe06] Koehn, Philipp: Statistical Machine Translation: the basic, the novel, and the speculative. PowerPoint Präsentation, University of Edinburgh, 2006.
- [Koe13] Koehn, Philipp: Statistical Machine Translation. Chapter 4: Word-based models, Chapter 5: Phrase-Based Models, Chapter 6: Decoding, Chapter 7: Language Models, Präsentationsfolien zum Buch, URL: <http://www.statmt.org/book/>, Datum der Recherche: 03/2013.
- [Kop06] Kopp, Stefan; et al.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. Artificial Intelligence Group, Universität Bielefeld, 2006.

- [Kop07] Kopp, Stefan: Surface Realization of Multimodal Output from XML representations in MURML. Artificial Intelligence Group, Fakultät der Technologien, Universität Bielefeld, 2007.
- [Kra01] Krammer, Klaudia: Schriftsprachkompetenz gehörloser Erwachsener. Forschungszentrum für Gebärdensprache und Hörgeschädigtenkommunikation, Universität Klagenfurt, 2001.
- [Lav13] Lavanchy, Claire: Statistische Maschinelle Übersetzung. Institut für Computerlinguistik, Universität Zürich, 2013.
- [Len02] van Lent, Mike; et al.: ICT Mission Rehearsal Exercise. USC Institute for Creative Technologies, 2002.
- [Leu04] Leuninger, Helen; Happ, Daniela: Gebärdensprachen: Struktur, Erwerb, Verwendung. Linguistische Berichte Sonderheft 13, HELMUT BUSKE Verlag, 2004.
- [Lex13] o.V.: Lexikalisch-funktionale Grammatik. URL: http://de.wikipedia.org/wiki/Lexikalisch-funktionale_Grammatik, Datum der Recherche: 03/2013.
- [Lóp11] López-Ludeña; et al.: Automatic categorization for improving Spanish into Spanish Sign Language Machine Translation. 2011.
- [Mae99] Maestri, George: [digital] Character Animation 2. Volume 1 - essential techniques, Indianapolis, New Riders Publishing, 1999.
- [Mar03] Martin, Joe: A Linguistic Comparison. Two Notation Systems for Signed Languages: Stokoe Notation & Sutton SignWriting, Western Washington University, 2003.
- [Mau08] Mauser, Arne; Hasan, Saša; Ney, Hermann: Automatic Evaluation Measures for Statistical Machine Translation System Optimization. Lehrstuhl für Informatik 6, Computer Science Department RWTH Aachen University, 2008.
- [Mat07] Huenerfauth, Matt: Generating American Sign Language animation: overcoming misconceptions and technical challenges. Published online, Springer-Verlag, 2007.

- [Met13] o.V.: METEOR. URL: <http://en.wikipedia.org/wiki/METEOR>, Datum der Recherche: 03/2013.
- [Mor05] Morrissey, Sara; Way, Andy: An Example-Based Approach to Translating Sign Language. National Centre for Language Technology, School of Computing, Dublin City University, 2005.
- [Mor06] Morrissey, Sara; Way, Andy: Lost in Translation: the Problems of Using Mainstream MT Evaluation Metrics for Sign Language Translation. School of Computing, Dublin City University, 2006.
- [Mor07] Morrissey, Sara; et al.: Combining Data-Driven MT Systems for Improved Sign Language Translation. School of Computing, Dublin City University, 2007.
- [Mor08a] Morrissey, Sara: Data-Driven Machine Translation for Sign Languages. Dissertation an der School of Computing, Dublin City University, 2008.
- [Mor08b] Morrissey, Sara: Assistive Translation Technology for Deaf People: Translating into and Animating Irish Sign Language. School of Computing, Dublin City University, 2008.
- [Mor13] o.V.: Morph target animation. URL: http://en.wikipedia.org/wiki/Morph_target_animation, Datum der Recherche: 01/2013.
- [Neu05] Neuschild, Elena; Simon, Christian: Übersetzungsprobleme. Seminar für Computerlinguistik, Institut für Allgemeine und Angewandte Sprachwissenschaft, 2005.
- [Ngu11] Kipp, Michael; Heloir, Alexis; Nguyen, Quan: Signing Avatars: a Feasibility Study. DFKI, Embodied Agents Research Group, Saarbrücken, 2011.
- [Pap02] Papineni, Kishore; et al.: BLEU: a Method for Automatic Evaluation of Machine Translation. IBM T. J. Watson Research Center, Yorktown, New York, 2002.
- [Pap08] Papaspyrou, Chrissostomos; et al.: Grammatik der deutschen Gebärdensprache aus der Sicht gehörloser Fachleute. Gebärdensprachlehre Band 6, SIGNUM-Verlag, 2008.

- [Par02] Parent, Rick: Computer Animation. Algorithms and Techniques, San Francisco [u.a.], Morgan Kaufman Publishers, 2002.
- [Plu02] Plugge, Katharina: Simulation und Graphik in Computerspielen. Character Animation, PowerPoint Präsentation, 2002.
- [Poh09] Schwind, Valentin; Pohl, Norman: Modellierung, Darstellung und interaktive Animierung virtueller Charaktere in 3D-Echtzeitanwendungen. Die Entstehung eines Charakters und die dynamische Animierung eines ped-Systems für eine interaktive 3D-Welt, Diplomarbeit im Studiengang Medieninformatik, Hochschule der Medien, Stuttgart, 2009.
- [Pop07] Popović, Maja; Ney, Herman: Word Error Rates: Decomposition over POS Classes and Applications for Error Analysis. Lehrstuhl für Informatik 6, RWTH Aachen University, 2007.
- [Pri07] Prillwitz, Siegmund; et al.: Grundkurs Deutsche Gebärdensprache. DGS-Grundkurs Stufe 2, SIGNUM-Verlag, 2007.
- [Rat98] Ratner, Peter: 3-D Human Modeling and Animation. New York [u.a.], Wiley Verlag, 1998.
- [Saf01] Safar, Eva; Marshall, Ian: The Architecture of an English-Text-to-Sign-Languages Translation System. School of Information Systems, University of East Anglia, 2001.
- [San06] Sandler, Wendy; Lillo-Martin, Diane: Sign Language and Linguistic Universals. Cambridge University Press Verlag, 2006.
- [San09] San-Segundo, R.; et al.: Design, Development and Field Evaluation of a Spanish into Sign Language Translation System. Departamento de Ingeniería Electrónica, ETSI Telecomunicación, Universidad Politécnica de Madrid, 2009.
- [Sch00] Schneider, Gerold: LFG: Einführung, c- und f-Strukturen. URL: <https://files.ifi.uzh.ch/cl/gschneid/SyntaxVorlesung/Vorl7.LFGI.html>, Department of Informatics, Universität Zürich, 2000.
- [Sch09] Scherrers, Svenja: Verbalklassen der deutschen Gebärdensprache. Universität Wuppertal, 2009

- [Sch12] Schade, Ulrich: Computerlinguistik. 13 Vorlesung, PowerPoint Präsentation, Fraunhofer-Institut für Kommunikation, Informationstechnik und Ergonomie, 2012.
- [Sch13] Schneider, Gerold: LFG: Einführung, c- und f-Strukturen. Vorlesung: Syntax und Formale Grammatiken, URL: https://files.ifi.uzh.ch/cl/gschneid/Syntax_Vorlesung/Vor17.LFGI.html, Datum der Recherche: 03/2013.
- [See09] Seewald-Heeg, Uta; Stein, Daniel: JLCL Journal for Language Technology and Computational Linguistics. Maschinelle Übersetzung – von der Theorie zur Anwendung Machine Translation – Theory and Applications, GSCL Gesellschaft für Sprachtechnologie und Computerlinguistik, Volume 24 – Number 3, 2009.
- [Seg08] Segouat, Jérémie; Braffort, Annelies: Toward the Study of Sign Language Coarticulation: Methodology Proposal. LIMSI Institut, Frankreich, 2008.
- [She03] Sheard, Mary; Beresford, J.; et al.: Evaluation Reports 1 to 3 of the EU ViSiCAST project Essential Sign Language Information on Government Networks, 2001 - 2003.
- [Sie07] Prillwitz, Siegmund; et al.: Grundkurs Deutsche Gebärdensprache. DGS-Grundkurs Stufe 1, SIGNUM-Verlag, 2007.
- [Smi10] Smith, Robert; Morrissey, Sara; Somers, Harold: Realistic Avatars for Sign Language Synthesis. CNGL, School of Computing, Dublin City University, 2010.
- [Spe01] Speers, Armond L.: Representation of American Sign Language for Machine Translation. Dissertation, Faculty of the Graduate School of Arts and Sciences, Georgetown University, 2001.
- [Ste06] Stein, Daniel: Morpho-syntax Based Statistical Methods for Sign Language Translation. Diplomarbeit am Lehrstuhl für Informatik VI, RWTH Aachen, 2006.
- [Sto12] o.V.: Stokoe notation. URL: http://en.wikipedia.org/wiki/Stokoe_notation, Datum der Recherche: 11/2012.

- [Stu11] Kipp, Michael; Nguyen, Quan; Heloir, Alexis: Umfrage der Machbarkeitsstudie zur Abschätzung der Nutzungsmöglichkeiten von Gebärdenavataren. URL: <http://embots.dfki.de/evaluation/index.php?sid=28195%E3%80%88=de>, Datum der Recherche: 01/2013.
- [Tom03] Tomás, Jesús; Mas, Josep Àngel; Casacuberta, Francisco: A Quantitative Method for Machine Translation Evaluation. Escola Politècnica Superior de Gandia, Universitat Politècnica de València, 2003.
- [Tüm07] Tümmler, Jörn: Avatare in Echtzeitsimulationen. Dissertation im Fachbereich Elektrotechnik/ Informatik, Universität Kassel, 2007.
- [Vea98] Veale, Tony; Collins, Bróna; Conway, Alan: The Challenges of Cross-Modal Translation: English to Sign Language Translation in the Zardoz System. Dublin City University, 1998.
- [Ver05] Vertan, Cristina: Beispiel-basierte Maschinelle Übersetzung. PowerPoint Präsentation, Fakultät für Sprache, Literatur und Medien, Universität Hamburg, 2005.
- [Ver07] Vertan, Cristina: Einführung in die Maschinelle Übersetzung II. Prinzipien und Algorithmen beispielbasierter Übersetzung, Fakultät für Sprache, Literatur und Medien, Universität Hamburg, 2007.
- [Vil07] Vilhjalmsson, Hannes; et al.: The Behavior Markup Language: Recent Developments and Challenges. Reykjavik University, Iceland, 2007.
- [Wac09] Wachsmuth, Ipke: MAX - eine Maschine, mit der man sprechen kann. Universität Bielefeld, Technische Fakultät - AG WBS, 2009.
- [Wak02] Wakefield, Michele: ViSiCAST Milestone: Final Report. 2002.
- [Web11] WebSourd: Dicta-Sign. Sign Language Recognition, Generation and Modeling with application in Deaf Communication. Deliverable D7.4: Prototype evaluation synthesis, 2011.
- [Wer04] Werthmann, Antonia; et al.: Maschinelle Übersetzungssysteme. PowerPoint Präsentation, URL: http://www.cl.uni-heidelberg.de/courses/archiv/ws04/ecl/tutorium/Glossar_A12.pdf, Universität Heidelberg, 2004.

- [Wik13] o.V.: Morphing. URL: <http://de.wikipedia.org/wiki/Morphing>, Datum der Recherche: 01/2013.
- [Wil10] Wilcke, Anne: Vergleichende Untersuchungen zur Nutzung von Laut- und Gebärdensprache im Telekommunikationsbereich. BTU Cottbus Lehrstuhl Kommunikationstechnik, 2010.
- [Zha00] Zhao, Liwei; et al.: A Machine Translation System from English to American Sign Language. Center for Human Modeling and Simulation, University of Pennsylvania, 2000.
- [Zij03] Van Zijl, Lynette; Barker, Dean: South African Sign Language Machine Translation System. Stellenbosch University, 2003.
- [Zij06] Van Zijl, Lynette; Combrink, Andries: The South African Sign Language Machine Translation Project: Issues on Non-manual Sign Generation, Stellenbosch University, 2006.
- [Zin13] Zinsmeister, Heike: Maschinelle Übersetzung: SMT-Vertiefung. Seminarfolien SS 2011, Universität Konstanz, URL: <http://ling.uni-konstanz.de/pages/home/zinsmeister/MT2011/Seminarplan.html>, Datum der Recherche 2013.

Anhang

A1: Symbole Stokoe-Notation

Ausführungsstelle (tab):

Symbol	Bedeutung
∅	neutral location
∩	face, or whole head (symbol is superimposed \wedge and \smile)
∩	forehead, brow, or upper face
⊏	eyes, nose, or mid face
∪	lips, chin, or lower face
3	cheek, temple, ear, or side face
∏	neck
∏	torso, shoulders, chest, trunk
∖	non-dominant upper arm
✓	non-dominant elbow, forearm
ⓐ	inside of wrist
ⓑ	back of wrist

Tab. 13: Ausführungsstellen nach Stokoe; aus [Sto12]

Handform (dez):

Symbol	Bedeutung
A	fist (as ASL 'a', 's', or 't')
B	flat hand (as ASL 'b' or '4')
5	spread hand (as ASL '5')
C	cupped hand (as ASL 'c', or more open)
E	claw hand (as ASL 'e', or more clawlike)
F	okay hand (as ASL 'f'; thumb & index touch or cross)
G	pointing hand (as ASL 'g' 'd' or 'l')
H	index + middle fingers together (as ASL 'h,' 'n' or 'u')
I	pinkie (as ASL 'i')
K	thumb touches middle finger of V (as ASL 'k' or 'p')
L	angle hand, thumb + index (as ASL 'l')
3(3)	vehicle classifier hand, thumb + index + middle fingers (as ASL '3')
O	tapered hand, fingers curved to touch thumbtip (as ASL 'o')
R	crossed fingers (as ASL 'r')
V	spread index + middle fingers (as ASL 'v')
W	thumb touches pinkie (as ASL 'w')
X	hook (as ASL 'x')
Y	horns (as ASL 'y', or as index + pinkie)
8(8)	bent middle finger; may touch thumb (as ASL '8' in the latter case; this is a common allophone of Y)

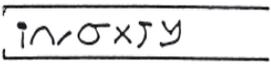
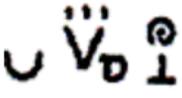
Tab. 14: Handformen nach Stokoe; aus [Sto12]

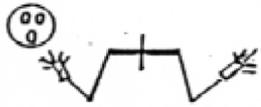
Bewegung (sig):

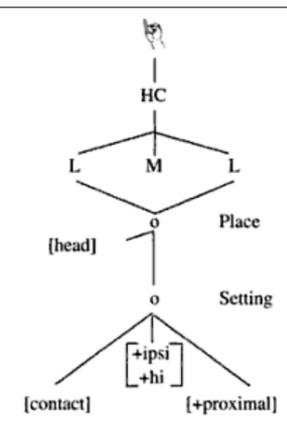
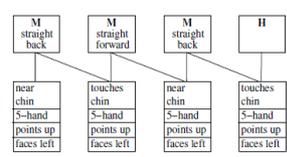
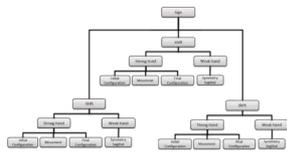
Symbol	Bedeutung
∧	moving upward
v	moving downward
N	moving up and down
>	to the dominant side
<	to the center or non-dominant side
≃	side to side
┘	toward signer
└	away from signer
I	to and fro
α	supinate (turn palm up)
β	pronate (turn palm down)
ω	twist wrist back & forth
η	nod hand, bend wrist
□	open up
#	close
⌘	wriggle fingers
⊙	circle
∫	approach, move together
×	contact, touch
∅	link, grasp
‡	cross
⊙	enter
÷	separate
“	exchange positions

Tab. 15: Bewegungen nach Stokoe; aus [Sto12]

A2: Übersichtstabelle: Notationssysteme der Gebärdensprache

	Notation	Anmerkung	Beispiel
Notationen für Menschliche Kommunikation	SignFont	Schreibsystem, das nicht alle Details der Gebärdensprache abdeckt; beschränkt sich auf phonemische Informationen	
	Schnelle Signschrift	An der Uni Hamburg 2002 von Dr. Alejandro Oviedo entwickelt; geeignet für die DGS; sehr schnell erlernbar; besteht überwiegend aus Zeichen, die man im Alltag benutzt (lateinische Buchstaben, Zahlen, deutsche Wörter, Pfeile usw.); reduziert die Gebärden auf ihre wesentlichen Merkmale; Ziel: keine vollständige Abbildung einer GS, sondern dient als Erinnerungsstütze beim Lernen; nicht am PC geeignet	Keine Angaben
	Glossen-Notation	vollständig großgeschriebene Wörter, die sich an der Lautsprache orientieren; leicht erlernbar; basiert nicht auf die Parameter der GS; keine feste Codierungsvorschrift; zusätzliche Informationen (z. B. Emotionen) durch weitere Symbole (z. B. „+“, „#“, „-“, „∞“ usw.) möglich	HALLO; IDEE; GUT; RECHNEN; ANGST; WER
Linguistische Notationen	Stokoe-Notation	erste linguistische Gebärdensprach-Notation; 1960 von Stokoe für die ASL entwickelt; basiert auf lateinischen Buchstaben, Zahlen für die Fingerstellung und ikonische Glyphen für die Position (tab), Bewegung (sig) und Handform (dez); umfasst insgesamt 55 Symbole	 SNAKE

	Notation	Anmerkung	Beispiel
Linguistische Notationen	SignWriting/ Gebärdenschrift	SignWriting für ASL; Gebärdenschrift für DGS; 1974 von Valerie Sutton entwickelt; stetige Weiterentwicklung bis heute; einfache Lesbarkeit; schnell zu erlernen; Gebärdenalphabet international anwendbar; ursprüngliche Notation für Tänzer → nicht von den linguistischen Notationen motiviert; basiert auf Piktogramme; Software zum Erstellen von Transkriptionen am PC	 Hallo
	Kakumasu- Notation	Notation für die USL (<i>Urubu Sign Language</i>) im Bundesstaat Maranhão (Brasilien); entwickelt 1968 von Jim Kakumasu; besteht aus lateinischen Buchstaben, Zahlen, „+“ (= gefolgt von) und „*“ (= linke Hand)	$L^{10}(AE^{21}) + (AE^{12})^{29}$ ZUCKER
	Friedman- Notation	Notation aus der phonologischen Sicht für die ASL; entwickelt 1977 von Lynn Friedmann; präsentiert phonetische Alternativen im Gegensatz zu Stokoe; Parameter: Handkonfiguration, Ausführungsstelle, Handorientierung (neu), Bewegung	Keine Angaben
	Lloyd B. Anderson Notation	alternatives Schriftsystem für die ASL; entwickelt 1981; Notation wurde auch für Gebärden-Wörterbüchern z. B. für die Thai Gebärdensprache genutzt; nutzt das lateinische Alphabet → Idee: höhere Akzeptanz, da Symbole bekannt sind	Keine Angaben
	Japan Research Group Notation	untersuchten 1984 die JSL (<i>Japanese Sign Language</i>); orientierten sich an den Notationen von Stokoe und Friedmann; Gebärdenzeichen basieren auf Handform, Handposition, Handbewegung; Ziel: universelles Notationssystem für alle Gebärdensprachen; Zeichensatz: Buchstaben, Zahlen, Piktogramme, Pfeile usw.	 GROß

	Notation	Anmerkung	Beispiel
Linguistische Notationen	HamNoSys - Hamburger Notationssystem	phonetisches Transkriptionssystem für Gebärdensprachen; seit 1984 stetige Weiterentwicklung (aktuelle Version 4.0); basiert auf Piktogrammen; Parameter: Handform, Handstellung, Ausführungsstelle und Bewegung; gute Skalierbarkeit; großer Anwendungsbereich	 <p>NEUNZEHN</p>
	Hand-Tier Modell	beeinflusst vom Movement-Hold-Modell und Stokoe-Notation; 1986 entwickelt von Wendy Sandler und 1989 modifiziert; Modell basiert auf einer Baumstruktur; erstes Modell, welches die Handkonfiguration auf eine separate Ebene legt; → Handform spielt eine größere Rolle als beim Movement-Hold-Modell; Parameter: Handkonfiguration (HC), Ausführungsstelle (L) und Bewegung (M)	 <p>IDEE</p>
	Movement-Hold-Modell	entwickelt 1989 von Liddell & Johnson; großer Einfluss auf nachfolgende Notationen; Gebärden als Abfolge von Bewegungsphasen (move) und Haltephasen (hold); „move“ und „hold“ bilden ein Segment; Parameter: Handform, Handorientierung, Ausführungsstelle und nicht-manuelle Komponenten; in Bewegungsphase Parameter änderbar, in Haltephase nicht	 <p>Mother</p>
	LV-Notation	1998 entwickelt von Losson und Vannobel; basiert auf dem Movement-Hold-Modell; Gebärden werden als hierarchische Baumstruktur dargestellt; sehr flache und leicht zu lesende Struktur; Verwendung zu rein linguistischen Zwecken	 <p>BRÜCKE</p>

	Notation	Anmerkung	Beispiel
Linguistische Notationen	BTS – Berkeley Transcription System	speziell für die niederländische Gebärdensprache (NGT) und ASL im Jahr 2001 entwickelt; Anwendung: Videos mit Gebärdensprache auf die Semantik zu transkribieren; Transkriptionsformat: CHAT (lineare ASCII-Kodierung); Notation dient der Konversationsanalyse; nicht geeignet für die Avataranimation, da Gebärden nicht präzise genug formuliert werden	(sit_on)-pm'PL_VL- pm'TL-loc'PL_VL_TOP- pst'STR 'sit on a horse'
Animationsorientierte Notationen	SWML – SignWriting Markup Language	XML-basierte Variante von SignWriting; dient zur Entwicklung von Wörterbüchern und zur Förderung der Barrierefreiheit im Web; Anwendung im Projekt „Vsigns“ zur Animation von Gebärdensprachavataren; Projekt wurde eingestellt	Beispiel siehe Anhang A2.1
	SIGML – Signing Gesture Markup Language	XML-Variante von HamNoSys; ca. 2002 entwickelt; Anwendung in den Projekten VisiCAST und eSIGN; entwickelt zur Ansteuerung von Gebärdensprachavataren; nutzt SAMPA für die Abbildung des Mundbildes	Beispiel im 3.Kapitel, Abschnitt 3.3.1
	PDTS-SIGML - posture/ detention/ transition/steady shift – Signing Gesture Markup Language	Erweiterung von SIGML; beinhaltet Ideen von Liddell und Johnson; kompatibel zum HamNoSys; präzisere Definitionsmöglichkeiten im Raum (Orientierung, Distanzen, Punkte) und im Timing; Vorteil: unabhängiger vom HamNoSys als SIGML	<sigml> <pdts> <Anweisungen/> </pdts> </sigml>
	Zebedee-Notation	2009 entwickelt von Filhol am LIMSI-Institut; soll Defizite der parametrischen Ansätze von Stokoe beheben; arbeitet mit sog. geometrischen Constraints und verwendet die Gelenke des menschlichen Skeletts; ähnelt einer Programmiersprache → sehr schwer zu lesen und zu verstehen	Beispiel siehe Anhang A2.2

	Notation	Anmerkung	Beispiel
Animationsorientierte Notationen	Emote-Modell (Expressive MOTion Engine)	entwickelt im Jahr 2000 von Diane Chi et al.; keine 100 % GS-Notation im Sinne der Definition; ein System, welches Bewegungsqualität (z. B. schnelle, langsame, träge Bewegungen) bei interaktiven Avataren beschreibt und versucht umzusetzen; entwickelt an der „University of Pennsylvania’s Center for Human Modeling and Simulation (HMS)“ ; orientiert sich an die Tanznotation (Labanotation) des Tanztheoretikers Rudolf von Laban (1879 – 1958); Zweck: synthetische Gebärden wirken individueller, emotionaler und natürlicher	Keine Angaben
	MURML - Multimodal Utterance Representation Markup Language	kompatibel zu BML, aber besitzt mehr Möglichkeiten Gesten zu definieren; 2002 entwickelt von Kranstedt et al.; die Struktur ähnelt dem Movement-Hold-Modell; Spezifikation der Handform orientiert sich am HamNoSys; Baumstrukturen wie beim LV Modell; sehr komplex und schwer lesbar für den Menschen	Beispiel siehe Anhang A2.3
	BML - Behavior Markup Language	2005 realisiert im SAIBA-Framework; Konzentration auf zeitliche Koordination von Gesten bzw. Verhalten und Sprache; wenig Einstellmöglichkeiten an der Form der Geste	Beispiel im 3.Kapitel, Abschnitt 3.3.2
	EMBRScript - Embodied Agent Behavior Realizer Script	Alexis Heloir und Michael Kipp erwähnten es 2009 erstmalig; Zwischenschicht zwischen Verhaltensbeschreibung wie BML und dem Animationssystem; ist Theorie unabhängig; beschreibt Sequenzen von Posen; Posen definieren eine Konfiguration des Körpers (z. B. Mimik, Handstellung usw.); Sequenzen können simultan ablaufen	Beispiel im 3.Kapitel, Abschnitt 3.3.3

Tab. 16: Anmerkungen zu Notationen; aus [Boy95], [Mar03], [Han04], [San06], [Kat08], [Kip11], [Dot12]

A2.1: Beispiel SWML

```
<?xml version="1.0" ?>
<swml version="1.0-d2" symbolset="SSS-1995">
  <generator>
    <name>SignWriter</name>
    <version>4.3</version>
  </generator>
  <sw_text>
    <sw_text_defaults>
      <sign_boxes>
        <unit> pt </unit>
        <height> 60 </height>
      </sign_boxes>
      <text_boxes>
        <box_type> graphic_box </box_type>
        <unit> pt </unit>
        <height> 60 </height>
      </text_boxes>
    </sw_text_defaults>
    <new_line/>
    <sign_box>
      <!-- the B hand -->
      <symbol x="8" y="13">
        <shape number="21" fill="1" variation="1" />
        <transform flop="0" rotation="0" />
      </symbol>
      <symbol x="7" y="25">
      <!-- the movement -->
        <shape number="108" fill="0" variation="1" />
        <transform flop="1" rotation="4" />
      </symbol>
    </sign_box>
  </sw_text>
</swml>
```

Abb. 97: SWML-Gebärde für das Wort „Brasilien“; aus [Cos01]

A2.2: Beispiel Zebedee-Notation

```

1. SEQUENCE "ball"
2.   DEP loc = @AB + 10*FWD
3.   DEP size = 2
4.   Point p0 := [loc] + [size]*UP
5.   KEY_POSTURE (10) {
6.     KEEP:
7.     For $h = s, w
8.       #round($h, [size])
9.       Orient NRM!palm($h) along <@PA($h), [loc]>
10.    End
11.  HERE:
12.    Place @PA(s) at p0
13.    Place @PA(w) at @I_KN(s, 1)
14.  }
15.  TRANSITION (70) {
16.    Accel 1
17.    Arc @PA(s) : [size]*LAT
18.    Arc @PA(w) : -Arc@PA(s)
19.  }
20.  KEY_POSTURE (20) {
21.    HERE:
22.    Place @PA(s) at SYM p0 WRT [loc]
23.    Place @PA(w) at @L_KN(s, 1)
24.  }
25. End "ball"

```

Abb. 98: Zebedee-Gebärde für das Wort „BALL“; aus [Fil09]

A2.3: Beispiel MURML

```

<definition><utterance>
  <specification>
    And now take <time id="t1"/> this <time id="t2"/> bar <time id="t3" chunkborder="true"/>
    and make it <time id="t4"/> this big. <time id="t5"/>
  </specification>

  <focus onset="t1" end="t2" accent="H"/>

  <behaviorspec id="gesture_1">
    <gesture id="pointing_to">
      <affiliate onset="t1" end="t3"/>
      <param name="refloc" value="$Loc-Bar_1"/>
    </gesture>
  </behaviorspec>

  <behaviorspec id="gesture_2">
    <gesture>
      <affiliate onset="t4" end="t5"/>
      <constraints>
        <symmetrical dominant="right_arm" symmetry="SymMS">
          <parallel>
            <static slot="HandShape" value="BSflat (FBround all o) (ThCpart o)"/>
            <static slot="PalmOrientation" value="DirL"/>
            <static slot="ExtFingerOrientation" value="DirA"/>
            <dynamic slot="HandLocation">
              <dynamicElement type="linear">
                <value type="start" name="LocShoulder LocCenterRight LocNorm"/>
                <value type="direction" name="DirR"/>
                <value type="distance" name="125.0"/>
              </dynamicElement>
            </dynamic>
          </parallel>
        </symmetrical>
      </constraints>
    </gesture>
  </behaviorspec>
</utterance></definition>

```

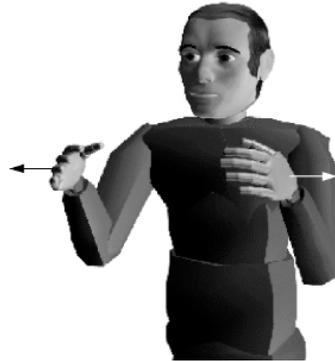


Abb. 99: MURML-Beispiel; aus [Kop07]

A3: Erklärung der LFG

Die Lexikalisch-funktionale Grammatik LFG ist eine Unifikationsgrammatik. Das heißt, sie bildet die Merkmalstruktur, speziell die syntaktischen, morphologischen und semantischen Informationen, eines Satzes ab. Dazu nutzt sie zwei Darstellungsebenen. Einen „*Konstituentenbaum*“ (c-Struktur, engl. *constituent*) für die Darstellung der Satzsyntax und die „*Merkmalstruktur*“ (f-Struktur, engl. *feature*) für die grammatikalische Darstellung. Zu diesen grammatikalischen Informationen können der Kasus (Nominativ, Genitiv, Dativ, Akkusativ), Genus (Maskulinum, Femininum, Neutrum), Numerus (Singular, Plural), Person (ich, du, er/sie/es, wir, ihr, sie) und der Tempus (Zeitform) gehören. Diese Eigenschaften stehen in einem Lexikon zu den zugehörigen Wörtern. Der Parser hat Zugriff auf dieses Lexikon und kann somit die Semantik eines Satzes „verstehen“. Die Abbildung 100 zeigt, wie der englische Satz „*The child worshipped the elephant.*“ in der LFG dargestellt wird.

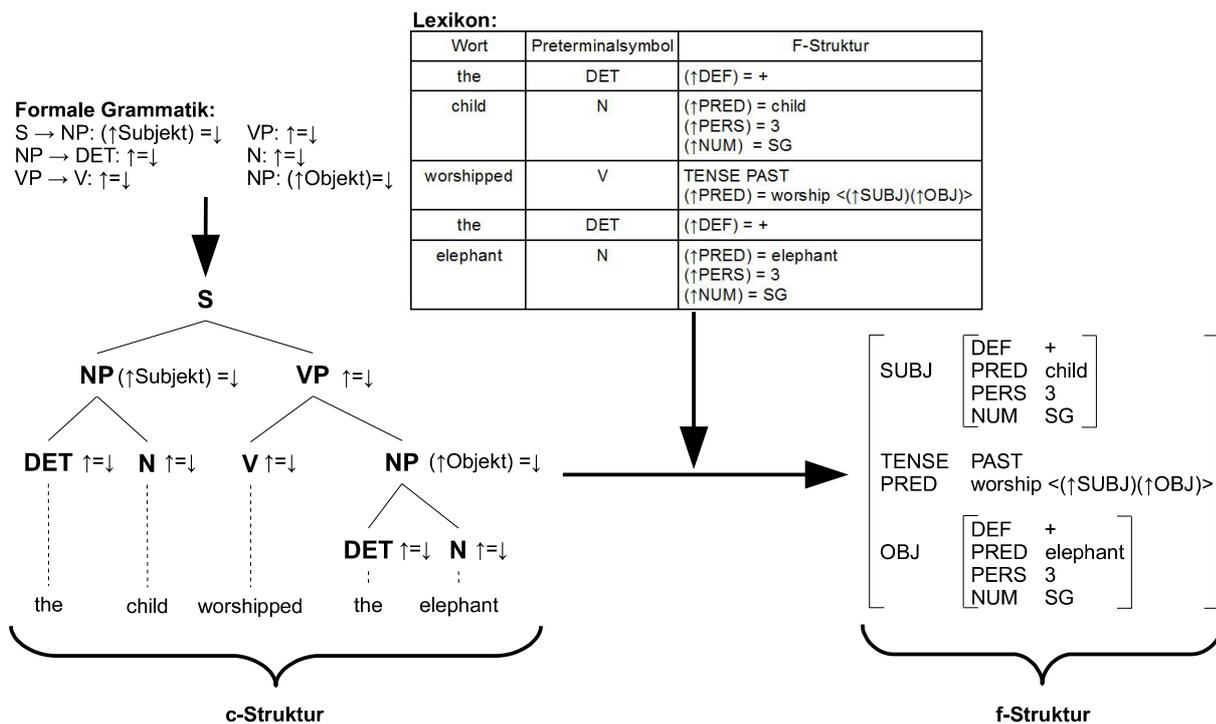


Abb. 100: Beispiel einer LFG; verändert aus [Sch13]

Die formale Grammatik beschreibt die Syntax einer Sprache. Eine Annotation in den Regeln gibt an, wie sich die Attribut-Wert-Matrix für die f-Struktur (Merkmalstruktur) bildet. Der Pfeil nach oben „ \uparrow “ zeigt auf den zugehörigen Mutterknoten und der Pfeil nach unten „ \downarrow “ auf den Knoten selbst. Die Bezeichnung „*Subjekt*“ und „*Objekt*“ sagt aus, welcher Teil im Satz das entsprechende Subjekt und Objekt ist. So wird zu der Regel „*DET: $\uparrow = \downarrow$* “ im linken Teilbaum das Wort „*the*“ im Lexikon gesucht, mit der zugehörigen f-Struktur ersetzt und zum Mutterknoten „*NP*“ weitergegeben.

In diesem Beispiel enthält das Lexikon Informationen zum Prädikat (PRED), Person (PERS), Numerus (NUM) und Tempus (TENSE). Der sogenannte Determinierer (DET), zu dem Artikel und Pronomen zählen, besitzt in der f-Struktur eine Definitheit (DEF), welche Ausdrücke zueinander referenziert. Nach der Abarbeitung des Baumes enthält die f-Struktur die Semantik des Satzes. Kasus- und Genus-Einträge würden das Lexikon in der f-Struktur vervollständigen und somit alle grammatikalisch relevanten Informationen beinhalten, die für eine qualitativ gute Semantik-basierende Transfer Übersetzung nötig sind [Sch00], [Lex13].

A4: Beispielberechnung für BLEU-Evaluation

Ein verkürztes Beispiel nach [Zin13] zeigt die Berechnung des BLEU-Wertes unter Nutzung eines Unigramm-Modells und eines Bigramm-Modells. Die MÜ-Ausgabe stellt das Testsystem dar. Die korrekte Übersetzung des zugehörigen Qs-Satzes zeigt die Referenz.

Referenz

„Ich danke der Delegation des Europäischen Parlaments im Konvent und dem Präsidenten des Konvents, Herrn Roman Herzog, für ihre engagierte Arbeit.“

Testsystem

„sollte ich der Parlamentdelegation in der Versammlung und im Präsidenten der Versammlung, Herr Roman Herzog, für ihre festgelegte Arbeit danken mögen.“

1. Schritt: Bestimmung der Unigramme und Bigramme zwischen der Referenz und dem Testsystem.

Unigramme:

modifizierte Unigramme: *ich; der; und; im; Präsidenten; Roman; Herzog; für; ihre; Arbeit;*
 $= \text{count}_{\text{clip}}(1\text{-gram}) = \underline{10}$

Unigramme: *sollte; ich; der; Parlamentdelegation; in; der; Versammlung; und; im; Präsidenten; der; Versammlung; Herr; Roman; Herzog; für; ihre; festgelegte; Arbeit; danken; mögen;* = $\text{count}(1\text{-gram}') = \underline{21}$

Bigramme:

modifizierte Bigramme: *Roman Herzog; Herzog für; für ihre;* = $\text{count}_{\text{clip}}(2\text{-gram}) = \underline{3}$

Bigramme: *sollte ich; ich der; der Parlamentdelegation; Parlamentdelegation in; in der; der Versammlung; Versammlung und; und im; im Präsidenten; Präsidenten der; der Versammlung; Versammlung Herr; Herr Roman; Roman Herzog; Herzog für; für ihre; ihre festgelegte; festgelegte Arbeit; Arbeit danken; danken mögen;*
 $= \text{count}_{\text{clip}}(2\text{-gram}') = \underline{20}$

2. Schritt: Berechnung der p_n jeweils für Uni- und Bigramm.

$$p_n = \frac{\sum_{C \in \{\text{Candidates}\}} \sum_{n\text{-gram} \in C} \text{count}_{\text{clip}}(n\text{-gram})}{\sum_{C' \in \{\text{Candidates}\}} \sum_{n\text{-gram}' \in C'} \text{count}(n\text{-gram}')}$$

Unigramm

$$p_{n=1} = \frac{\text{count}_{\text{clip}}(1\text{-gram})}{\text{count}(1\text{-gram}')} = \frac{10}{21} \approx \underline{\underline{0,48}}$$

Bigramm

$$p_{n=2} = \frac{\text{count}_{\text{clip}}(2\text{-gram})}{\text{count}(2\text{-gram}')} = \frac{3}{20} \approx \underline{\underline{0,15}}$$

3. Schritt: Berechnung des Brevity Penalty-Wertes.

$$BP = \begin{cases} 1 & , \text{ falls } c > r \\ e^{1-\frac{r}{c}} & , \text{ falls } c \leq r \end{cases} \quad \begin{array}{l} \text{mit } c = \text{Anzahl der W\u00f6rter des Testsystems} = \underline{\underline{21}} \\ r = \text{Anzahl der W\u00f6rter der Referenz} = \underline{\underline{21}} \end{array}$$

$$BP_{\text{Testsystem}} = e^{(1-\frac{21}{21})} = \underline{\underline{1}}$$

4. Schritt: Berechnung des BLEU-Wertes.

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \cdot \log p_n\right) \quad \text{mit } w_n = 1/N \text{ und } N = 2$$

$$BLEU_{\text{Testsystem}} = 1 \cdot \exp\left(\left(\frac{1}{2} \cdot \log \frac{10}{21}\right) + \left(\frac{1}{2} \cdot \log \frac{3}{20}\right)\right) \approx \underline{\underline{0,56}}$$

Der BLEU-Wert wird h\u00e4ufig in einer Prozentzahl angegeben. Somit hat das Testsystem einen Wert von 56 % erreicht.

A5: Beispielberechnung für METEOR-Evaluation

Das gewählte Beispiel entspricht dem Beispiel der gezeigten BLEU-Rechnung aus dem Anhang A4.

Referenz

„Ich danke der Delegation des Europäischen Parlaments im Konvent und dem Präsidenten des Konvents, Herrn Roman Herzog, für ihre engagierte Arbeit.“

Testsystem

„sollte ich der Parlamentdelegation in der Versammlung und im Präsidenten der Versammlung, Herr Roman Herzog, für ihre festgelegte Arbeit danken mögen.“

1. Schritt: Alinierung nach identischen Wörtern und nach Wortstamm.

Ich danke der Delegation des Europäischen Parlaments im Konvent und dem Präsidenten des Konvents, Herrn Roman Herzog, für ihre engagierte Arbeit.
 Sollte ich der Parlamentdelegation in der Versammlung und im Präsidenten der Versammlung, Herr Roman Herzog, für ihre festgelegte Arbeit danken mögen.

2. Schritt: Berechnung von Precision und Recall.

- m = Anzahl der Wortalinierungen
- w_t = Anzahl der Unigramme/Wörter im Testsystem bzw. die Ausgabe der MÜ
- w_r = Anzahl der Unigramme/Wörter in der Referenz

$$P = \frac{m}{w_t} = \frac{12}{21} \approx \underline{\underline{0,571}} \qquad R = \frac{m}{w_r} = \frac{12}{21} \approx \underline{\underline{0,571}}$$

3. Schritt: Berechnung von F_{mean}

$$F_{\text{mean}} = \frac{10 \cdot P \cdot R}{R + 9 \cdot P} = \frac{10 \cdot \frac{12}{21} \cdot \frac{12}{21}}{\frac{12}{21} + 9 \cdot \frac{12}{21}} = \frac{4}{7} \approx \underline{\underline{0,571}}$$

4. Schritt: *chunks* ermitteln und *Penalty* ausrechnen.

Alle Unigramme aus der Referenz und der MÜ-Ausgabe werden in möglichst kleinen Einheiten (sogenannte *chunks*) gruppiert. Dabei werden nebeneinanderstehende alignierte Unigramme der MÜ-Ausgabe, die ebenfalls nebeneinander in der Referenz vorkommen, als eine Gruppe angesehen.

Chunks:

Referenz: *Ich; danke; der; im; und; Präsidenten; Herrn Roman Herzog für ihre; Arbeit*

MÜ-Ausgabe: *Ich; der; und; im; Präsidenten; Herr Roman Herzog für ihre; Arbeit; danken*

Die unterstrichene Einheit ist in diesem Beispiel die einzige Einheit, die größer ist als ein Unigramm und nebeneinander steht. Insgesamt beträgt die *Anzahl der chunks* 8. Die *Anzahl für die enthaltenen Unigramme*: „Ich; danke; der; im; und; Präsidenten; Herrn; Roman; Herzog; für; ihre; Arbeit;“ beträgt 12. Eingesetzt in der Formel ergibt sich ein Strafwert von 0,148.

$$Penalty = 0.5 \cdot \left(\frac{\text{Anzahl der chunks}}{\text{Anzahl unigrams}_{\text{matched}}} \right)^3 = 0.5 \cdot \left(\frac{8}{12} \right)^3 = \frac{4}{27} \approx \underline{\underline{0,148}}$$

5. Schritt: Berechnung des METEOR-Wertes.

$$METEOR = F_{\text{mean}} \cdot (1 - Penalty) = \frac{4}{7} \cdot \left(1 - \frac{4}{27} \right) \approx \underline{\underline{0,487}}$$

Die METEOR-Evaluation bewertet den übersetzten Satz mit ca. 48,7 % etwas schlechter als die BLEU-Evaluation, die einen Wert von 56 % erreichte.

A6: Zusammenfassung ausgewählter Evaluationen

A6.1: Huenerfauth'schen „multiple choice“ Test

Dieses Verfahren wurde für die Bewertung eines Übersetzungssystems vom Englischen in die ASL (*American Sign Language*) und in die „Signed English“ angewendet.

Versuchspersonen

An diesem Versuch nahmen 15 Personen teil, bei denen alle die ASL von der Geburt oder frühen Kindheit an lernten.

Datenbasis

Als Grundlage für dieses Verfahren dienten 10 Sätze. Sie beschreiben die Bewegung oder die Position eines Objektes bzw. zwischen Objekte (vgl. Tabelle 17 (linke Spalte)). Diese 10 Aussagen wurden einmal in die Amerikanische Gebärdensprache (ASL) und in die „Signed English“ Sprache transkribiert. Die „Signed English“ Sprache ist eine Lautsprachbegleitende Gebärde (LBG), bei der die englischen Sätze mit Glossen dargestellt werden. Für Wörter (z. B. Artikel), die nicht in der ASL existieren, wurden neue Gebärdenzeichen entwickelt. Die Animationen zu den ASL und „Signed English“ Sätzen führte ein Avatar vor.

English Gloss	ASL Sentence with Classifier Predicates (CPs)	Signed English Sentence
The car parks between the house and the cat.	ASL sign HOUSE; CP: house location; sign CAT; CP: cat location; sign CAR; CP: car path.	THE CAR PARK BETWEEN THE HOUSE AND THE CAT
The man walks next to the woman.	ASL sign WOMAN; CP: woman location; sign MAN; CP: man path.	THE MAN WALK NEXT-TO THE WOMAN
The car turns left.	ASL sign CAR; CP: car path.	THE CAR TURN LEFT
The lamp is on the table.	ASL sign TABLE; CP: table location; sign LIGHT; CP: lamp location.	THE LIGHT IS ON THE TABLE
The tree is near the tent.	ASL sign TENT; CP: tent location; sign TREE; CP: tree location.	THE TREE IS NEAR THE TENT
The man walks between the tent and the frog.	ASL sign TENT; CP: tent location; sign FROG; CP: frog location; sign MAN; CP: man path.	THE MAN WALK BETWEEN THE TENT AND THE FROG
The man walks away from the woman.	ASL sign WOMAN; CP: woman location; sign MAN; CP: man path.	THE MAN WALK FROM THE WOMAN
The car drives up to the house.	ASL sign HOUSE; CP: house location; sign CAR; CP: car path.	THE CAR DRIVE TO THE HOUSE
The man walks up to the woman.	ASL sign WOMAN; CP: woman location; sign MAN; CP: man path.	THE MAN WALK TO THE WOMAN
The woman stands next to the table.	ASL sign TABLE; CP: table location; sign WOMAN; CP: woman location.	THE WOMAN STAND NEXT-TO THE TABLE

Tab. 17: Datenbasis für das Experiment; aus [Hue07]

Des Weiteren erstellte man zu jedem Satz drei Videoanimationen. Ein Video bildete den exakten Inhalt des Satzes ab. Die anderen beiden dagegen zeigten die gleichen Objekte wie in der korrekten Animation, wurden aber mit kleinen Änderungen in der Bewegung und Orientierung versehen [Hue07].

Ein Beispiel wäre:

Der originale Satz lautet: „*Das Auto parkt zwischen dem Haus und der Katze [Hue07].*“

1. Videoanimation zeigt: „*Ein Auto fährt einen kurvigen Weg entlang und parkt zwischen dem Haus und der Katze [Hue07].*“
2. Videoanimation zeigt: „*Das Auto fährt zwischen dem Haus und der Katze hindurch und fährt weiter aus der Szene heraus [Hue07].*“
3. Videoanimation zeigt: „*Das Auto fährt zwischen dem Haus und der Katze los und parkt an einem anderen Punkt, der nicht zwischen beiden Objekten liegt [Hue07].*“

Somit gab es insgesamt 10 Sätze mit 30 Vergleichsvideos. Bei späteren Versuchen vereinfachte Huenerfauth das Experiment, indem er anstatt Videoanimationen entsprechende Piktogramme mit Beschriftungen einsetzte.

Durchführung

Ein Moderator erläuterte den Versuchspersonen (VP) den Ablauf und stand während des Tests für Fragen zur Verfügung. Die Versuchspersonen sahen auf einem Bildschirm den Avatar, der einen Satz gebärdet. Anschließend sollten sie mit einer 10-Punkte-Skala die Kriterien Verständlichkeit (1 = verwirrend; 10 = leicht zu verstehen), natürliche Bewegung (1 = roboterhaft; 10 = reibungslos, wie eine reale Person) und Korrektheit der ASL Grammatik (1 = gemischtes Englisch; 10 = perfekt) bewerten.

Zu jedem Satz gab es eine Aufgabe (*matching task*), bei der aus drei Auswahlmöglichkeiten (Videoanimationen A, B, C) diejenige gewählt werden sollte, die den exakten Inhalt der Avatardarstellung wiedergab. Das Benutzerinterface des Multiple-Choice-Tests ist in der Darstellung 101 zu sehen.

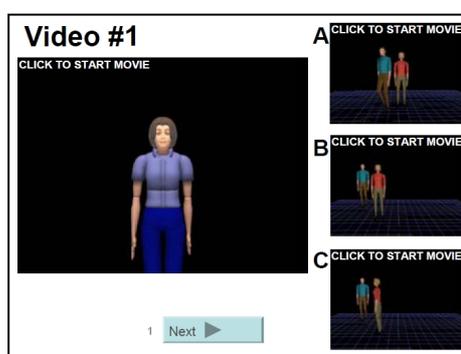


Abb. 101: Auswahlmöglichkeiten A, B und C für die Versuchsperson; aus [Hue07]

Dieses Verfahren wurde für alle 10 ASL Sätze und für alle 10 „Signed English“ Sätze durchgeführt. Die Sätze wurden zufällig gewählt und die Versuchspersonen konnten die Animationen beliebig oft ausführen. Abschließend wurden den Teilnehmern drei zufällig

gewählte Avatarvideos mit den zugehörigen korrekten Videoanimationen gezeigt, zu denen sie ihre Verbesserungsvorschläge in den Bereichen Animationsgeschwindigkeit, Farben/Lichter, Sichtbarkeit und Korrektheit der Handformen, Gesichtsausdruck sowie Blickbewegung äußern konnten [Hue07].

Ergebnisse

Das Ergebnis für dieses Experiment ist in der Abbildung 102 ersichtlich. Die ersten drei Balkenpaare stellen die subjektiven Werte der Versuchspersonen in den Bereichen Grammatik, Verständlichkeit und Natürlichkeit der Bewegungen dar. Die objektiven Verständniswerte für die Multiple-Choice-Aufgabe sind mit dem letzten Balkenpaar abgebildet.

Die gelb bzw. rot eingefärbten Balken stehen für die Gebärdensprache bzw. Lautsprachbegleitende Gebärdensprache. Die Verständlichkeit des Inhalts (*matching task* und *Understandable*) betrug in etwa 83 % für die Gebärdensprache, wobei kritisch hinterfragt werden kann, wie repräsentativ dieser Wert ist im Vergleich zu den anderen Methoden, bei denen es keine Auswahlmöglichkeiten gab.

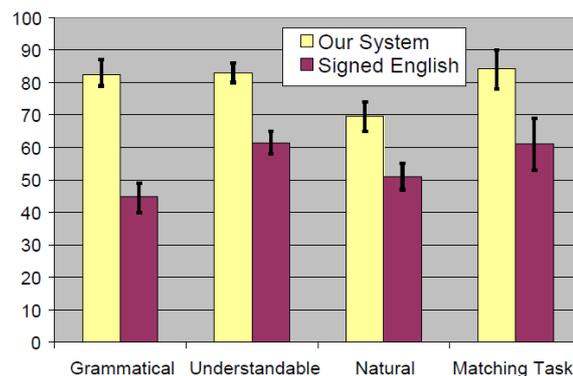


Abb. 102: Balkendiagramm für Huenerfauth'schen „multiple choice“ Test; aus [Hue07]

A6.2: Delta-Evaluation (EMBR)

Die Delta-Evaluation wurde in [Kip11] erfunden und erstmalig angewendet. Ziel der Bewertungsmethode war es, die Verständlichkeit des Gebärdensprachavatars durch einen Vergleich mit einem menschlichen Sprecher zu ermitteln.

Versuchspersonen

An dem Experiment nahmen 13 Versuchspersonen (6 männliche und 7 weibliche) im Alter von 33 – 55 Jahre teil. Sie beherrschten alle die Deutsche Gebärdensprache. Pro Teilnehmer betrug die Dauer des Versuches 1,5 bis 2 Stunden.

Datenbasis

Als Ausgangsmaterial nutzte man Gebärdensprachvideos von der e-Learning Plattform Vibelle. In diesen zwei Videos, einmal zum Thema „Gelbe Seiten“ und „Was ist Vitamin B?“, wurde der Inhalt vom Menschen gebärdet (Material: *original*). Aus diesem Material produzierte der Animator die Animationen des Avatars. Dieser Versuch scheiterte, da die DGS-Experten keine einzige Gebärde verstanden.

Nach dem Analysieren der Probleme wurden diese Videos von einem Gebärdensprach-Experten nach gebärdet, wobei er die Gebärden ausdrucksstärker, einfacher, langsamer und klarer getrennt darstellte (Material: *remake*). Daraus entstanden anschließend die Animationen für den Avatar (Material: *Avatar*).

	1	2	3	4	5	6	7	8	9	10	11
<i>contained glosses</i>	17	18	21	10	8	7	12	21	13	14	13
<i>Video (original)</i>	10s	7s	11s	7s	5s	3s	7s	11s	13s	3s	7s
<i>Video (remake)</i>	14s	14s	20s	14s	9s	6s	16s	19s	19s	7s	13s
<i>Avatar</i>	17s	20s	25s	16s	12s	11s	15s	23s	25s	8s	22s

Tab. 18: Datenbasis der Evaluation zu dem Avatar EMBR; aus [Hel11]

Die Tabelle 18 gibt die Anzahl der verwendeten Glossen und die genaue Länge der Videos wieder. Insgesamt wurden 11 Äußerungen mit 154 Glossen erstellt. Da einige Glossen für die Aussagen mehrfach benutzt werden konnten, lag die genaue Anzahl der Glossen-Animationen bei 95. Für die Erstellung einer Glossen-Animation benötigten die Entwickler 5 – 25 Minuten nach einer ein- bis zweiwöchigen Einarbeitungszeit. Sie verwendeten dafür den Behavior-Builder von EMBR [Hel11].

Durchführung

Jede Versuchsperson musste vor dem Experiment einen Fragebogen ausfüllen, in denen neben demografischen Informationen auch Erfahrungen mit Avataren und Einstellungen zu dieser Technik geklärt wurden. Danach gab es eine kurze Eingewöhnungszeit mit dem Avatar „Max“. Er stellte sich den Teilnehmern vor und ermöglichte ihnen somit einen ersten Eindruck seines Gebärdensprachstils. Anschließend begann die Evaluationsphase mit dem Abspielen der Videos in folgender Reihenfolge (vgl. Abb. 103):

- 1.) Darstellung der Avataramationen (Material: *Avatar*); bis zu 6 Wiederholungen möglich
- 2.) Original Video (Material: *original*); bis zu 3 Wiederholungen waren möglich
- 3.) vereinfachte Video Remake (Material: *remake*); bis zu 3 Wiederholungen waren möglich

Die Versuchspersonen sahen die erste Äußerung aus der Datenbasis, welche in alle drei Videos abgebildet wurde. Nach jedem Video befragte ein Moderator die VP nach dem Inhalt des Videos. Jeweils beim 2.) und 3.) Video wurde zusätzlich gefragt, ob und was die VP im Vergleich zum vorherigen Video mehr verstanden hat. Die Versuchsperson beurteilte nach jedem Video per Fragebogen die Kriterien: Natürlichkeit, Qualität und Verständlichkeit für das Avatarvideo und für die Videos 2.) und 3.) die Verständlichkeit und Schwierigkeiten der gezeigten Aussage.

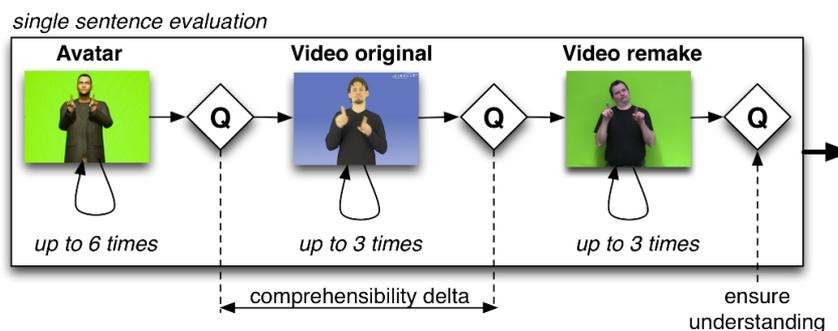


Abb. 103: Ablauf der Delta-Evaluation; aus [Ngu11]

Bevor dieses Verfahren mit dem nächsten Satz aus der Datenbasis wiederholt werden konnte, musste am Ende nochmal das 3.) Video mit Untertitel gezeigt werden. Dadurch wurde sichergestellt, dass der Proband die Aussage verstanden hat, weil die Inhalte der Sätze voneinander abhängig waren. Während der Durchführung wurden die Teilnehmer von zwei DGS-Experten analysiert. Diese notierten bei der Frage nach dem Inhalt des Videos an die VP die Anzahl der richtig genannten Glossen im Video. Um das Nachahmen der Gebärden (aus dem Video) zu limitieren, wurde auch ein subjektives Urteil der DGS-Experten abgefragt. Dazu gaben die Experten nach jedem gezeigten Satz auf einer Skala von 1 (gar nicht) bis 7 (alles) an, wie gut die VP den Satz nach ihrer subjektiven Beurteilung verstand. Abschließend konnten die Teilnehmer Verbesserungsvorschläge und Kritiken zum Avatar äußern [Kip11].

Ergebnisse

Die Darstellung 104 zeigt die Ergebnisse für die objektiven und subjektiven Verständniswerte.

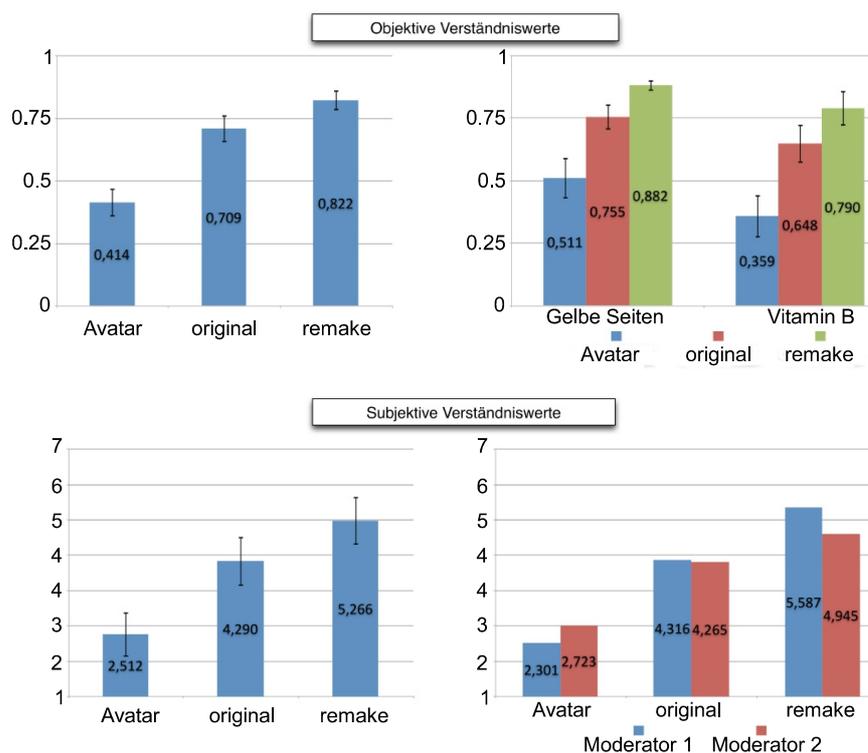


Abb. 104: Ergebnisse der Delta-Evaluation zu EMBR; aus [Kip11]

Für die Ermittlung der objektiven Verständniswerte in der oberen Grafik wurde die Anzahl der verstandenen Glossen im Satz durch die Gesamtzahl der Glossen im Satz geteilt. Wurde eine Glosse in einem vorherigen Video ermittelt, so zählte diese bei den nachfolgenden Videos mit. Somit musste die Anzahl der richtig erkannten Glossen beim nachfolgenden Video gleich oder ansteigend sein. Das obere linke Balkendiagramm zeigt den Durchschnitt der richtig verstandenen Glossen bzw. das Satzverständnis an. Oben rechts ist die Grafik in den beiden Themenvideos „Gelbe Seiten“ und „Vitamin B“ aufgeteilt. Mit Hilfe der beiden oberen Grafiken lässt sich ein objektiver Verständniswert für den Avatar berechnen.

$$\text{Verständlichkeit} = \frac{\text{Avatarwert}}{\text{Videowert}} \cdot 100\%$$

Wird der Avatar im Bezug auf das Vibelle-Video und dem vereinfachten Video betrachtet, so erreicht er ein Verständnis von 58,39 % bzw. 50,36 %. Bei dem Themen lag das durchschnittliche Verständnis bei den „Gelben Seiten“ (59,02 %) höher als bei dem Thema „Vitamin B“ (50,42 %).

Die subjektive Einschätzung der beiden DGS-Experten ähnelt dem obigen Ergebnis. In der linken unteren Grafik lag die Einschätzung der VP-Verständlichkeit bei 58,55 % bzw. 47,70 %. Die rechte untere Grafik zeigt, dass die Meinungen beider Moderatoren in ihrer Einschätzung recht nah bei einander lagen.

A6.3: ViSiCAST Evaluation (speziell: Dialogsystem Poststelle)

Als Beispiel der Vorgehensweise soll hier die Bewertung des Dialogsystems in der Poststelle näher gebracht werden. Dabei sollten Gehörlose eine Transaktion mit den Mitarbeitern in der Post durchspielen. Die verwendete Gebärdensprache war die BSL (*British Sign Language*). Ziel war es, die Akzeptanz und die Verständlichkeit des Systems zu bestimmen.

Versuchspersonen

An dem Versuch nahmen 6 gehörlose Versuchspersonen und 3 hörende Mitarbeiter des „UK Post Office“ teil. Die Postmitarbeiter besaßen 10 Jahre Berufserfahrung in diesem Bereich und die Gehörlosen sprachen die BSL als Muttersprache. Die Evaluierung dauerte zwei Tage und bestand aus zwei Phasen. Bei der ersten Phase bewerteten die 6 Gehörlosen die Verständlichkeit der Sätze und Gebärdenzeichen der Gebärdendarstellung des Avatars. Für die zweite Phase – der Dialog – wurden die Teilnehmer in drei Gruppen eingeteilt. Eine Gruppe bestand jeweils aus zwei Gehörlosen und einem Postmitarbeiter.

Datenbasis

Verwendet wurde der Gebärdensprachavatar Tessa, dessen Geometrie aus 5000 Polygone bestand und die Animation auf ein magnetisches MOCAP Verfahren basierte. In der Vorarbeit wurden 350 Sätze, die über 90 % der Transaktionen in der Poststelle abdeckten, einprogrammiert. In der ersten Testphase wurden von diesen 133 Sätze mit insgesamt 444 Gebärdenzeichen benutzt. In der zweiten Testphase wählte die Post 18 von 30 Transaktionsgespräche, die in drei Schwierigkeitsbereiche (sechs einfache, sechs mittel schwere und sechs komplexe) unterteilt waren, aus.

Durchführung

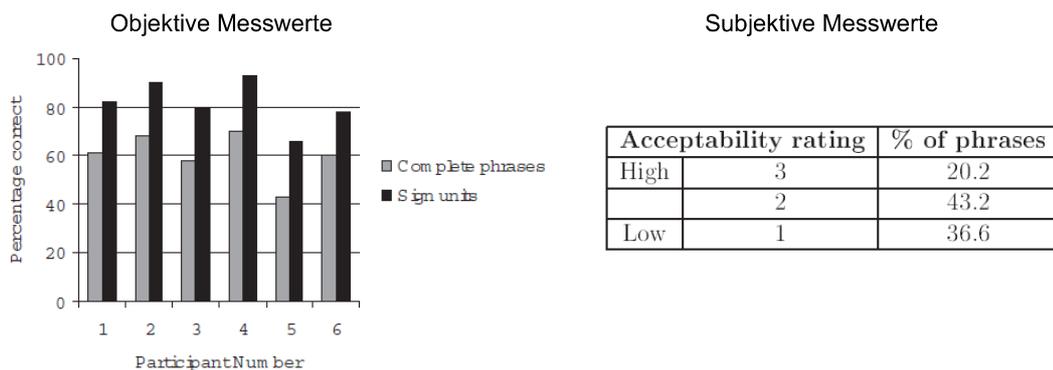
Am Anfang füllten die Teilnehmer den ersten Teil eines Fragebogens aus. In der ersten Phase gebärdete der Avatar Sätze, die in Hauptkategorien wie Postgebühr, Bezahlung usw. unterteilt wurden. Die Gehörlosen konnten eine Satzanimation bis zu 5 mal wiederholen lassen und sollten nach jeder Animation aufschreiben, was sie verstanden haben. Längere Aussagen wurden blockweise (20 – 24 Zeichen) wiedergeben. Die Genauigkeit des Verständnisses ermittelte man mit zwei objektiven Methoden. Zum Einen die exakte Wiedergabe des Inhalts durch die genauen Glossen (siehe Delta-Evaluation: Anzahl der richtigen Glossen durch Gesamt-Glossen-Anzahl) und zum Anderen durch eine Approximation der Genauigkeit auf semantischer Ebene. Dazu mussten die Versuchspersonen die semantisch relevanten Gebärdenzeichen erkennen. Als Beispiel dient der Satz: „It should arrive by Tuesday but it's not guaranteed [Cox02].“ Eine 100 % richtige Erkennung wäre gewesen: „should arrive Tuesday not guaranteed [Cox].“ Wurden in diesem Beispiel von den fünf Wörtern nur „should arrive Tuesday“ erkannt, zählte es als ein Verständniswert von 66 %.

Am Ende der ersten Phase mussten die Teilnehmer die Akzeptanz der Satzdarstellung des Systems subjektiv mit Hilfe eines Rankings von 1 (schlecht) bis 3 (gut) bewerten.

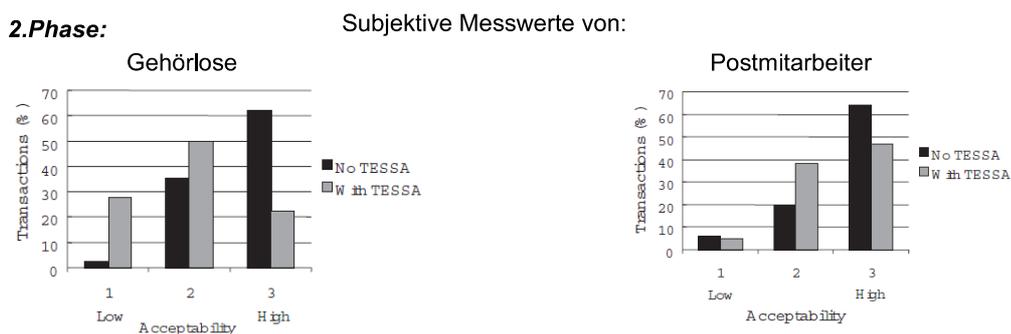
Anschließend folgte in der zweiten Phase der Dialog zwischen dem Postmitarbeiter und den beiden Versuchspersonen (= 1 Gruppe) in den drei verschiedenen Schwierigkeitsgraden. Der Mitarbeiter trug ein Headset, welches permanent für den Empfang von Sprachbefehlen aktiv war. Auf einem Bildschirm vor sich sah er Schlüsselwörter/Hauptkategorien (z. B. „Postage“, „Bill Payments“, „Passport“ usw.), die er mit seiner Stimme aktivieren konnte und somit in ein Untermenü kam, welches eine Auswahl von Sätzen, die der Avatar gebärden kann, abbildete. Bei der Aussprache der einprogrammierten Sätze stellte der Avatar diese in BSL dar. Die eine Hälfte der Transaktionsgespräche wurde mit dem Avatar Tessa durchgeführt und die andere Hälfte ohne. Dabei wurde die Zeit für die Transaktionen gemessen. Abschließend füllten die Teilnehmer den zweiten Teil des Fragebogens aus, welcher die subjektive Einschätzung (ebenfalls das Ranking 1 bis 3) des Dialogsystems abfragte [Cox02].

Ergebnisse

1.Phase:



2.Phase:



Objektive Messwerte: Dialogdauer

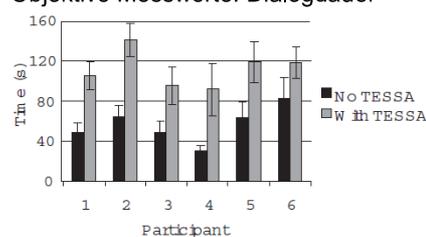


Abb. 105: Ergebnisse der Evaluierung des Poststellen Systems; zusammengestellt aus [Cox02]

Die Abbildung 105 zeigt die Ergebnisse der Evaluation. In der ersten Phase wurde die Qualität der Gebärdenzeichen und des Satzinhaltes bewertet. Die Verständlichkeit der Sätze lag im Durchschnitt bei 61 % und die der einzelnen Gebärdenzeichen bei 81 %. Die Akzeptanz der Darstellungsqualität des Avatars lag im Durchschnitt bei einer Bewertung von 2.2.

Die Ergebnisse der zweiten Phase zeigten, dass die Benutzung mit dem Avatar Tessa (Durchschnittszeit: 112 s) zeitlich ca. 2 mal langsamer war, als wenn die Transaktion ohne Avatar (Durchschnittszeit: 57 s) vollzogen wurde. Die subjektive Bewertung der Akzeptanz ergab, dass die Postmitarbeiter mit ca. 2,62 Punkte im Durchschnitt die Nutzung mit dem Avatar besser bewerteten, als die Gehörlosen Teilnehmer mit ca. 1,94 Punkte. Transaktionen, bei denen der Avatar nicht benutzt wurde, bewerteten die Gehörlosen im Durchschnitt mit ca. 2,83 Punkte etwas besser als die Mitarbeiter der Post mit ca. 2,76 Punkten [Cox02].

Dies waren die Ergebnisse der ersten Evaluation des „Post Office“ Systems. Weitere Fokusgruppen testeten dieses System nach diesem Schema. Die Ergebnisse sind in [She03] (Evaluation Report 1) einsehbar.

A6.4: Die 9 Evaluationen des Dicta-Sign Projektes

Evaluation	Goal	Protocol	Panel	Main results	Exploitation	Future work
Isolated sign comprehension	Measure isolated sign comprehension Identify sources of incomprehension	Several isolated sign presentations with a virtual signer (default speed, front view)	6 Deaf Signing evaluators	58 % comprehension with the virtual signer. Major criticisms of movement animation.	Adjust the visualization parameters in the prototypes	Find the best visualization parameters for the comprehension of isolated signs performed by a virtual signer.
Full utterance comprehension	Measure isolated sign & Full sentence comprehension Identify sources of incomprehension	Several isolated sign presentations with a virtual signer and a real signer slow speed (20° tilted side view). Several avatar sentence presentations with and without non-manual features. Presentation of the same sentences performed by a real signer.	5 Deaf and 5 Hearing signing evaluators	Comprehension loss caused by the virtual signer ranging from 33 % to 62 %. Non-manual features help in understanding only if they are accurate and synchronized with manual attributes.	Adjust the visualization parameters in the prototypes	Find the best visualization parameters for the comprehension of a full sentence realized by a virtual signer.
Translation Tool	Measure the recognition rate in a real context Observe the use of the tool for an SL to SL translation Get Feedback on the interface	Use the tool to translate isolated signs and full sentences.	5 Deaf and 5 Hearing signing evaluators	In average 24 % of the good answers are ranked first in the presentation. The use of the look up tool is comparable to a dictionary.	Kinect based interaction to avoid switching from the mouse to the Kinect interface.	Effectuate the same evaluation with a larger corpus to see whether the approach is scalable.
Sign Wiki evaluation	Evaluate sign wiki usability Get feedback from the Deaf community.	Watch and understand a short sentence signed by a virtual signer Modify and create content in the sign wiki. Complete an online questionnaire about the wiki.	18 evaluators from across Europe	Main interface is related to the SL video display. The outline display was well understood but the low level item presentation should be enhanced.	Suggestions not implemented because the wiki is only a prototype.	Implement the functionalities and evaluate the wiki for more complex tasks like collaborative content creation.
Test about initial virtual signer Rendering	Evaluate virtual signer acceptability among elderly Deaf community	Watch a sentence performed by a virtual signer and formulate free comments.	20 Deaf signing elderly evaluators	Animation speed should be slowed down. Criticisms of the lack of mouthing and the inaccuracy of movements.	Creation of a new virtual signer (Françoise)	Define a list of criteria to create a user-friendly and understandable virtual signer

Evaluation	Goal	Protocol	Panel	Main results	Exploitation	Future work
Interaction with the virtual signer	Evaluate the interaction with the virtual signer Get feedback from the Deaf community.	Conversation with virtual signer following the principle of a Wizard of Oz.	33 Deaf signing European evaluators	70 % of good interactions. Criticisms of the lack of face expression.	Focus on non-manual features in the framework of grammatical structures	Create more grammatical structures adapted to interactive situations.
Search-by-example interface	Get Feedback on the Search-By-Example interface	Evaluators only view the interface and are asked how they would perform the action of searching for a sign, translating it and examining its definition.	2 Deaf LSF experts.	Need to add an icon to explain WordNet. Need of a function to select the sign by clicking on the corresponding virtual signer. The source and the target SL should be easier to identify.	Reorganization of the prototype interface.	Evaluate the look up tool with a larger database in order to see if the approach is scalable. Use information from hand shape.
Evaluation of Françoise Virtual Signer	Get Feedback on the skinning of Françoise	The evaluators are invited to speak freely about their first impression of the skinning.	12 signing testers	Main criticisms about facial expression, body proportions and clothing texture.	Creation of the new virtual signer (Françoise)	Define a list of criteria to create an user-friendly and understandable virtual signer
Sign Wiki first module interface evaluation	Get Feedback on the visualization and composition modules	View one animation using the interface. Create a symmetrical posture presented by the experimenter using the composition interface.	3 Deaf LSF experts.	Need of options to adjust the speed of the animation and pause it. Criticisms of excessive mouthing by the virtual signer. Icons should replace HamNoSys symbols.	Signing speed adjustment in the last prototype. Interface restructuring in order to use more explicit icons.	Enhance the sign editor in order to create dynamic non-symmetrical signs.

Tab. 19: Zusammenfassung der Evaluationen aus dem Dicta-Sign Projekt; aus [Web11]

Erklärung

Ich erkläre, dass ich die vorliegende Masterarbeit selbständig, ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Cottbus, 30.04.2013

Marcel Linke