



universität
wien

MASTERARBEIT

Titel der Masterarbeit

„Speech-to-Speech Translation Apps
Ein kritischer Vergleich“

verfasst von

Ivona Brizar, BA

angestrebter akademischer Grad

Master of Arts (MA)

Wien, 2014

Studienkennzahl lt. Studienblatt: A 060 342 345

Studienrichtung lt. Studienblatt: Masterstudium Übersetzen Englisch Französisch

Betreut von: Univ.-Prof. Dr. Gerhard Budin

Danksagung

An dieser Stelle möchte ich die Gelegenheit nutzen und mich bei allen bedanken, die mich während der Erstellung dieser Masterarbeit motiviert und unterstützt haben.

Besonderer Dank gebührt meinen Eltern Mirjana und Zdravko Brizar, ohne die ich diese Zeilen wohl nicht schreiben würde. Meiner Mutter, die mich stets durch wertvolle Ratschläge auf dem richtigen Weg gehalten hat, sowie meinem Vater, dem ewigen Optimisten und moralischen Unterstützer gilt mein herzlichster Dank. Hvala vam od srca!

Außerdem danke ich meiner Schwester Irena und meinem Freund Daniel für das kritische und konstruktive Lektorat dieser Arbeit.

Nicht zuletzt möchte ich mich bei Univ.-Prof. Dr. Gerhard Budin für die Betreuung der Arbeit bedanken.

Inhaltsverzeichnis

Inhaltsverzeichnis	i
Abbildungsverzeichnis	iii
Einleitung	1
1 Translation im Technologiezeitalter	3
1.1 Von den Anfängen der MT	3
1.1.1 Vorläufer der MT	4
1.1.2 Die Pioniere	5
1.1.3 Der Optimismus	6
1.1.4 Die Ernüchterung	7
1.1.5 Die Renaissance	8
1.2 Zur iPhone App durch Speech-to-Speech Translation.....	9
1.2.1 Phase der Einschränkungen	10
1.2.2 Phase der spontanen Sprache	11
1.2.3 Phase der tragbaren Systeme	11
1.2.4 Phase der uneingeschränkten Anwendung.....	12
1.3 Bis hin zur MT im World Wide Web.....	13
2 Die Komponenten eines Speech-to-Speech Translation Systems.....	17
2.1 Automatische Spracherkennung	18
2.1.1 Die Tücken der Spontansprache	18
2.1.2 Messung der ASR Qualität	21
2.2 Maschinelle Übersetzung.....	22
2.2.1 Regelbasierte Ansätze	22
2.2.2 Statistikbasierte Ansätze	27
2.2.3 Weitere Ansätze	30
2.3 Sprachsynthese.....	33
2.3.1 Aufgabe der Sprachsynthese.....	34
3 Pionierprojekte der SST	37
3.1 Verbmobil.....	37
3.1.1 Von Verbmobil zur iPhone-App.....	39
3.1.2 Technologische Errungenschaften von Verbmobil.....	40
3.2 Lecture Translator.....	42
3.2.1 Die Funktionsweise.....	44
4 Funktionaler Vergleich von SST Apps.....	49
5 Jibbig Translator 2.5.5.....	51
5.1 Technische Hintergründe.....	52
5.2 User Interface	53
5.3 Features.....	59
5.3.1 Sprachenangebot	59
5.3.2 Übersetzer-Modus.....	60
5.3.3 Besonderheiten.....	65
5.4 Benutzerfreundlichkeit.....	65
5.5 SWOT-Analyse.....	66

6	VoiceTra4U 3.2.....	69
6.1	Technische Hintergründe	70
6.2	User Interface	71
6.3	Features.....	75
6.3.1	Sprachenangebot	76
6.3.2	Übersetzer-Modus	77
6.3.3	Besonderheiten	79
6.4	Benutzerfreundlichkeit	83
6.5	SWOT-Analyse	85
7	SayHi Translate 3.8.2.....	87
7.1	Technische Hintergründe	87
7.2	User Interface	88
7.3	Features.....	93
7.3.1	Sprachenangebot	93
7.3.2	Übersetzer Modus	97
7.3.3	Besonderheiten	98
7.4	Benutzerfreundlichkeit	100
7.5	SWOT-Analyse	101
8	Anwendungsbeispiele der SST Apps.....	103
8.1	Testdialoge mit Jibbig 2.5.5.....	104
8.1.1	Kommentar zur Leistung von Jibbig 2.5.5.....	106
8.2	Testdialoge mit VoiceTra4U 3.2	108
8.2.1	Kommentar zur Leistung von VoiceTra4U 3.2	110
8.3	Testdialoge mit SayHi Translate 3.8.2	112
8.3.1	Kommentar zu Leistung von SayHi Translate 3.8.2	114
9	Schlussbetrachtung	117
9.1	Funktionsüberblick Apps	117
9.2	Beantwortung der Forschungsfragen.....	119
9.3	Conclusio.....	121
9.4	Blick in die Zukunft	123
	Abkürzungsverzeichnis.....	126
	Bibliographie.....	127
	Anhang	135
	Abstract	135
	Curriculum Vitae	136

Abbildungsverzeichnis

ABBILDUNG 1: DREIECK NACH VAUQUOIS (WIKIMEDIA COMMONS 2014).....	24
ABBILDUNG 2: GRAPHISCHE ÜBERSICHT ÜBER SST SYSTEME MIT EINEM STATISTISCHEN ANSATZ (A) UND EINEM INTERLINGUA ANSATZ (B) (WAIBEL & FÜGEN 2008: 71).....	29
ABBILDUNG 3: SPRACHSYNTHESESYSTEM MIT STIMMABHÄNGIGEN UND	35
ABBILDUNG 4: VERBINDUNGSaufbau EINER VERBMobil-SITZUNG (WAHLSTER 2000: 5).....	39
ABBILDUNG 5: SCHEMATISCHE ÜBERSICHT ÜBER DIE ARCHITEKTUR DES LECTURE TRANSLATORS (FÜGEN 2008: 138)	45
ABBILDUNG 6: HEADS-UP DISPLAY GOGGLES (WAIBEL 2012C: 37).....	46
ABBILDUNG 7: LECTURE TRANSLATOR AUF TABLET (WAIBEL 2012C: 39)	46
ABBILDUNG 8: ÜBERBLICK ÜBER EIN SZENARIO MIT DEM LECTURE TRANSLATOR (FÜGEN 2008: 141).....	47
ABBILDUNG 9: SCREENSHOT HAUPTMENÜ JIBBIGO 2.5.5	55
ABBILDUNG 10: SCREENSHOT EINSTELLUNGEN JIBBIGO 2.5.5	55
ABBILDUNG 11: SCREENSHOT VOLLWERTIGER ÜBERSETZER-MODUS JIBBIGO 2.5.5.....	56
ABBILDUNG 12: SCREENSHOT ÜBERSETZER-MODUS OHNE ASR JIBBIGO 2.5.5	56
ABBILDUNG 13: SCREENSHOT SPRACHAufNAHME JIBBIGO 2.5.5	57
ABBILDUNG 14: SCREENSHOT EINSTELLUNGEN ÜBERSETZER-MODUS JIBBIGO 2.5.5	57
ABBILDUNG 15: SCREENSHOT ÜBERSETZUNGSVERLAUF JIBBIGO 2.5.5.....	58
ABBILDUNG 16: SCREENSHOT INHALTE TEILEN JIBBIGO 2.5.5	58
ABBILDUNG 17: SCREENSHOT INHALTE VIA NACHRICHT WEITERLEITEN JIBBIGO 2.5.5	59
ABBILDUNG 18: SCREENSHOT RÜCKÜBERSETZUNG JIBBIGO 2.5.5	62
ABBILDUNG 19: SCREENSHOT WÖRTERBUCH JIBBIGO 2.5.5.....	62
ABBILDUNG 20: SCREENSHOT MANUELL NAMEN HINZUFÜGEN JIBBIGO 2.5.5	63
ABBILDUNG 21: SCREENSHOT ROMANISIERUNG DER CHINESISCHEN SCHRIFTZEICHEN JIBBIGO 2.5.5	63
ABBILDUNG 22: NETZWERKBASIERTER ÜBERSETZUNGSPROZESS (VGL. U-STAR 2014B)	70
ABBILDUNG 23: SCREENSHOT HAUPTMENÜ SINGLE MODE VOICETra4U 3.2	72
ABBILDUNG 24: SCREENSHOT U-STAR INFORMATIONEN VOICETra4U	72
ABBILDUNG 25: SCREENSHOT ALLGEMEINE AUSWAHL DER SPRACHEN VOICETra4U 3.2.....	73
ABBILDUNG 26: SCREENSHOT AUSWAHL DER SPRACHEN MIT SPRACHAusGABE VOICETra4U 3.2	73
ABBILDUNG 27: SCREENSHOT SPRACHAufNAHME VOICETra4U 3.2	74
ABBILDUNG 28: SCREENSHOT USERANMELDUNG CHAT MODE VOICETra4U 3.2	74
ABBILDUNG 29: SCREENSHOT BEISPIELÜBERSETZUNG SINGLE MODE VOICETra4U 3.2.....	78
ABBILDUNG 30: SCREENSHOT BEISPIELÜBERSETZUNG NICHT VERSTÄNDLICH SINGLE MODE VOICETra4U 3.2	78
ABBILDUNG 31: SCREENSHOT HAUPTMENÜ MIT KONTAKT IM CHAT MODE VOICETra4U 3.2.....	80
ABBILDUNG 32: SCREENSHOT AUSGEHENDER ANRUF CHAT MODE VOICETra4U 3.2	80
ABBILDUNG 33: SCREENSHOT ANRUF ENTGEGENNEHMEN VOICETra4U 3.2	81
ABBILDUNG 34: SCREENSHOT BEISPIELCHAT 1 VOICETra4U 3.2.....	81
ABBILDUNG 35: SCREENSHOT BEISPIELCHAT 2 VOICETra4U 3.2.....	82
ABBILDUNG 36: SCREENSHOT CHATABBRUCH VOICETra4U 3.2.....	82

ABBILDUNG 37: SCREENSHOT HAUPTMENÜ MIT ÜBERSETZUNGSBEISPIEL SAYHi 3.8.2	89
ABBILDUNG 38: SCREENSHOT SPRACHERKENNUNG LOKALISIERT SAYHi 3.8.2	89
ABBILDUNG 39: SCREENSHOT SPRACHAUSWAHL SAYHi 3.8.2	90
ABBILDUNG 40: SCREENSHOT INHALTE TEILEN SAYHi 3.8.2	90
ABBILDUNG 41: SCREENSHOT EINSTELLUNGEN SAYHi 3.8.2	91
ABBILDUNG 42: SCREENSHOT FEEDBACK & SUPPORT SAYHi 3.8.2	92
ABBILDUNG 43: SCREENSHOT PREMIUM SPRACHEN SAYHi 3.8.2	96
ABBILDUNG 44: SCREENSHOT ERKENNUNGSFEHLER ASR SAYHi 3.8.2	98
ABBILDUNG 45: SCREENSHOT MANUELLE KORREKTUR SAYHi 3.8.2	98
ABBILDUNG 46: SCREENSHOT FAVORITEN VERWALTEN SAYHi 3.8.2	99
ABBILDUNG 47: SCREENSHOT FAVORITEN ANLEGEN BZW. BEARBEITEN SAYHi 3.8.2	99
ABBILDUNG 48: SCREENSHOT FAVORITEN ANLEGEN - SPRACHAUSWAHL SAYHi 3.8.2	100

Einleitung

Heutzutage besitzt der Großteil der westlichen Bevölkerung ein Smartphone mit dem, dank sozialer Netzwerke wie Facebook oder Anwendungen wie WhatsApp, über Grenzen hinweg mit einander kommuniziert werden kann. Durch die Globalisierung und den dadurch bedingten Anstieg an interkultureller Kommunikation steigt auch der Bedarf an Translation. Professionelle TranslatorInnen bedienen sich bereits seit Jahren technischer Hilfsmittel; ihre Arbeit ist ohne den Einsatz von CAT-Tools oder computerunterstützter Übersetzung undenkbar geworden. Auch Endverbraucher haben elektronische Hilfsmittel für sich entdeckt und dem altbewährten Wörterbuch den Rücken gekehrt. Warum Zeit verlieren, wenn Google bereits automatisch die Übersetzung einer fremdsprachigen E-Mail oder Website anbietet? Auch wenn solche Übersetzungen oft mit einem schmunzelnden Auge betrachtet werden, reichen sie doch aus, um sich einen groben Überblick über den Sachverhalt zu machen. Dass diese Ergebnisse auf Basis komplizierter Algorithmen und Formeln maschinell generiert werden, ist wohl den wenigsten NutzerInnen solcher Online-Übersetzungsdienste bewusst. Selbst in der Translationswissenschaft wurden die Bestrebungen in der Forschung zur Maschinellen Übersetzung (MÜ) äußerst skeptisch beäugt und (vorwiegend von praktizierenden TranslatorInnen) nicht immer gutgeheißen.

Mit der Verbreitung von Smartphones und Tablets eröffnete sich für Silicon Valley eine weitere Möglichkeit den VerbraucherInnen bei translatorischen Problemen unter die Arme zu greifen – Translation Apps wurden entwickelt und in Stores zum Download zur Verfügung gestellt. Somit hatte jede/r Smartphone-Nutzer/in einen Dolmetscher in der Hosen- bzw. Handtasche, der überall und jederzeit einsetzbar war. Eine Vorstellung die genauso faszinierend wie zugleich besorgniserregend ist; letzteres gilt wohl vor allem für jene mit einem äußerst ausgeprägten translatorischen Gewissen. Diese für den Massenmarkt konzipierten Apps sind sogenannte Speech-to-Speech Translation Apps, da sie nicht nur auf die maschinelle Übersetzung von schriftlichen Texten beschränkt sind, sondern mithilfe einer Spracherkennungs- und Sprachausgabefunktion gesprochene Texte „dolmetschen“ können.

Der Schwerpunkt der Arbeit liegt auf dem funktionalen Vergleich der drei Speech-to-Speech Translation Apps *Jibbig Translator 2.5.5*, *VoiceTra4U 3.2* und *SayHi Translate 3.8.2*. Ausgehend von einem starken Interesse für technologische Spielereien, welche das Smartphone sowie diverse Applikationen bieten, entstand die Idee die mobilen Übersetzungsdienste im Rahmen einer Forschungsarbeit näher zu untersuchen. Um Diversität zu schaffen, fiel die Wahl auf drei Apps, die in ihrem Ursprung und ihren Funktionen doch sehr unterschiedlich sind. Die erste App, der *Jibbig Translator*, war eine der ersten Speech-to-Speech Translation Apps auf dem Markt und entstand aus einem Spin-off-Unternehmen eines langjährigen Forschungsprojektes zur Verarbeitung von Spontanspra-

che. Die zweite App namens *VoiceTra4U* wurde von einem internationalen Konsortium zur Übersetzung von gesprochener Sprache (U-STAR) konzipiert und fokussiert sich hauptsächlich auf den asiatischen Sprachmarkt. Die dritte App trägt den Namen *SayHi Translate* und stammt aus einem selbstfinanzierten Start-up-Unternehmen. Ziel dieser Arbeit ist es, eine umfassende, vergleichende Analyse der einzelnen Funktionen der drei Apps zu schaffen und diese in ihren Vor- und Nachteilen einander gegenüber zu stellen, um zu abstrahieren für welchen Zweck die einzelnen Apps sinnvoll einsetzbar sind. Zudem wurden zwei Beispieldialoge aus unterschiedlichen Kontexten erarbeitet, mithilfe welcher die übersetzerische Leistung der einzelnen Apps illustriert werden soll.

Das erste Kapitel stellt eine chronologische Aufarbeitung der Meilensteine in der Forschung zur Machine Translation dar, die sich von den Anfängen der Maschinellen Übersetzung bis zu heutigen Anwendungsbereichen erstreckt. Im zweiten Kapitel werden die Hauptkomponenten von Speech-to-Speech Translation Systemen eingehend beschrieben und die möglichen Ansätze vorgestellt, um die Funktionsweise bzw. auch die potentiellen Problemfelder nachvollziehbar zu machen. Das dritte Kapitel umfasst zwei Pionierprojekte im Bereich der Sprachtechnologien und soll deutlich machen, welche technologischen Fortschritte innerhalb von zwei Jahrzehnten möglich waren. Kapitel vier beschreibt die Herangehensweise bei der funktionalen Analyse der Apps und klärt über das Forschungsinteresse sowie die Forschungsfragen auf. Die Kapitel fünf, sechs und sieben stellen die eingehenden Analysen der individuellen App-Funktionen dar, welche sich hauptsächlich auf die Bereiche technische Fakten, User Interface, Features sowie Benutzerfreundlichkeit konzentrieren und eine abschließende SWOT-Analyse enthalten. Das achte Kapitel soll mittels zweier Testdialoge die übersetzerische Leistung der einzelnen Apps aufzeigen. Kapitel neun, die Schlussbetrachtung, bietet eine tabellarische Gegenüberstellung der Funktionen der Apps und soll der Beantwortung der Forschungsfragen dienen, sowie die Zukunftsaussichten im Bereich der Sprachtechnologien und insbesondere der Speech-to-Speech Translation Systeme aufzeigen.

Um den Forschungsgegenstand Speech-to-Speech Translation Apps zugänglicher und verständlicher zu gestalten, werden die Beschreibungen der einzelnen Apps von Screenshots begleitet und zusätzlich wurde ein Abkürzungsverzeichnis der Fachtermini am Ende der Arbeit inkludiert. Es soll an dieser Stelle darauf hingewiesen werden, dass alle in der vorliegenden Arbeit angegebenen Daten und Fakten nur von beschränkter Gültigkeitsdauer sind, da sie sich ausschließlich auf die jeweils angegebene Version der Apps beziehen.

1 Translation im Technologiezeitalter

Dass der Eroberungsfeldzug der Technologie auch in den unterschiedlichen Berufsfeldern der Translation Einzug gefunden hat, ist wohl kaum zu leugnen. KonferenzdolmetscherInnen nutzen tagtäglich das technische Equipment in den Kabinen, während seit neuestem auch in Wiener Krankenhäusern DolmetscherInnen per Video zwischen Patienten und ÄrztInnen vermitteln können. Auch bei ÜbersetzerInnen wäre der Berufsalltag ohne Computerunterstützung undenkbar – von CAT-Tools über Translation Memory Systeme – die Arbeitsprozesse werden effizienter und das Arbeitsvolumen höher. Neben den Tools für professionelle Übersetzer entwickelten sich im Laufe der Jahre auch maschinelle Übersetzungstools für die breite Masse. Diese sollen den KonsumentInnen im alltäglichen Leben dabei helfen translatorische Probleme zu lösen. Dank Google Translate können ganze Texte mit einem Mausklick in unzählige Sprachen übersetzt werden, mit einem weiteren Klick können ganze Websites in anderen Sprachen dargestellt werden. Da diese Tools webbasiert sind, war es nur eine Frage der Zeit, bis sie auch auf mobilen Geräten wie Smartphones oder Tablets der Masse zur Verfügung stehen würden. Um die einzelnen Komponenten der Translation Apps greifbarer zu machen, soll zunächst ein chronologischer Überblick über die Meilensteine der Forschung zur Machine Translation (MT) und maschinellen Sprachverarbeitung gegeben werden.

1.1 Von den Anfängen der MT

Im folgenden historischen Überblick können nur die wichtigsten Meilensteine in der Geschichte der Maschinellen Übersetzung genannt werden, da die vorliegende Arbeit nicht auf eine historische Aufarbeitung des Themas abzielt. Eine allgemeine Einführung in die Entwicklung der MT und genauere Ausführungen zu den einzelnen Projekten und Forschungen finden sich in Hutchins (1999, 2014a). Die Anfänge der MT werden in Booth (1967), Bruderer (1982) und Hutchins (2000) beschrieben, während die Entwicklungen der 1970er und 1980er in Nirenburg (1987) und King (1987) nachzulesen sind. Einen guten Überblick über die Technologien der 1990er Jahre bieten Farwell et al. (1998), Trujillo (1999) und White (2000). Für einen Einblick in die jüngsten Entwicklungen sei die Lektüre von Hutchins (2010) empfohlen, sowie die Proceedings diverser Konferenzen¹ zu MT (z.B. MT Summit) oder Computerlinguistik (z.B. COLING, ACL).

¹ Für einen chronologischen Überblick aller bereits gehaltenen Konferenzen und Workshops zu MT (inkl. Proceedings in PDF) siehe <http://www.mt-archive.info/srch/conferences-1.htm> bzw. zur Computerlinguistik siehe <http://aclweb.org/anthology/>.

1.1.1 Vorläufer der MT

Zur Geschichte der maschinellen Übersetzung finden sich in der Literatur Referenzen, die bis ins 16. oder 17. Jahrhundert zurückreichen. Gelehrte und Philosophen bemühten sich bereits damals darum, mithilfe von Geheim- und Universalsprachen die Babylonische Sprachverwirrung zu lösen. Gottfried Willhelm Leibnitz unternahm den Versuch, die Sprache in ihre kleinstmöglichen Elemente einzuteilen, um diese daraufhin in einer neuen und umfassenderen Ordnung zusammenzusetzen. Johann Joachim Becher sprach in seiner 1661² erschienenen Publikation „Verschlüsselung der Sprachen“ von einer geheimschriftlichen Erfindung, mit der jeder nach einer eintägigen Einarbeitungszeit im Stande wäre, beim Lesen der eigenen Sprache andere Sprachen deuten und verstehen zu können (vgl. Steiner 2009: 5). Im Grunde handelte es sich dabei um eine numerische Aufarbeitung von alphabetisch sortierten Termini der lateinischen Sprache. Hutchins sieht den Bezug zu Universal- und Geheimsprachen als Vorgänger der MT kritisch:

„[W]hile ‘logical’ languages can be seen as prerequisites for digital computer programs and while ‘universal’ languages can be regarded as forerunners of auxiliary languages such as Esperanto, as precursors of universal classification systems such as those of libraries and biological taxonomies, and even as early forms of interlinguas in machine translation itself, these proposals should not be considered in any way as constituting embryonic automatic translation systems.“ (Hutchins 2004: 1)

Die Erfindung solcher numerischen Kodierungen wie bei Becher, die auch als eine Art Geheimsprache genutzt werden konnten, beschreibt einen Menschen der maschinell, geleitet von den Kodierungen, agiert, jedoch nicht von Maschinen ausgeführte Prozesse. Daher sind der Franzose Georges Artsrouni und der Russe Petr Trojanskij als die wahren Vorläufer der MÜ zu sehen, da sie 1933 zeitgleich ein Patent für „Übersetzungsmaschinen“ anmeldeten. Es handelte sich dabei um Geräte, die auf Basis der elektromagnetischen Technik der 1930er Jahre als mechanische Wörterbücher verwendet werden konnten. Artsrouni hatte ein Gerät patentieren lassen, welches universell einsetzbar war und mit Zusatzausstattung auch als mehrsprachiges Wörterbuch agieren konnte. Sein sogenanntes „mechanisches Gehirn“ (*cerveau mécanique*) sollte gespeicherte Informationen abrufen und drucken – wie z.B. Zugfahrpläne oder Telefonverzeichnisse. (vgl. Hutchins 2004: 2f). Trojanskijs ging einen Schritt weiter und beschrieb nicht nur den Aspekt des mehrsprachigen Wörterbuches, sondern auch wie der gesamte Übersetzungsprozess auf Basis universeller (an Esperanto angelehnter) Symbole ablaufen könnte, die für die Kodierung und Dekodierung grammatikalischer Funktionen nötig sind. Er sah einen dreiteiligen Prozess dafür vor, bei dem die zweite Phase maschinell erfolgte. Im ersten monolignualen Schritt soll eine Person, welche nur der Ausgangsprache mächtig sein muss, den Originaltext mit festgeschriebenen Symbolen versehen, welche die grammatikalischen Beziehungen im Satz aufzeigen sollen.

² Das Werk wurde 1962 unter dem Titel „*Zur mechanischen Sprachübersetzung: allgemeine Verschlüsselung der Sprachen; ein Programmierungsversuch aus dem Jahre 1661*“ neu verlegt.

In der zweiten, der maschinellen Phase wird der mit Symbolen versehene Ausgangstext in den Zieltext verwandelt, während in der dritten, wiederum monolingualen Phase, eine der Zielsprache mächtige Person, die grammatikalisch kodierten Elemente zu einem kohärenten Text zusammensetzen soll. Er entwickelte die Maschine weiter und strebte danach auch die anderen Teile des Prozesses zu automatisieren, wobei ihm jedoch die Problematik von Homonymen, Synonymen, etc. sowie die Wichtigkeit des Post-Editings bewusst war. (vgl. Hutchins 2004: 7ff). Trojanskij war es bereits Jahre vor Erfindung der ersten Computer gelungen, einen zumindest zum Teil mechanisierten Übersetzungsprozess aufgeteilt in Analyse, Transfer und Synthese zu schaffen, ohne dabei die typischen translatorischen Probleme vollkommen außer Acht zu lassen.

1.1.2 Die Pioniere

Wenige Jahre nach der Entwicklung der ersten Computer hatte man ihr Potential auch im Bezug auf die Translation entdeckt. Der Stein des Anstoßes war ein von Warren Weaver verfasstes Memorandum aus dem Jahre 1949, in dem erstmals vorgeschlagen wurde, den Computer als Grundlage für die Erstellung automatischer Übersetzungssysteme zu nutzen. Auf Basis der Erkenntnisse aus der Dechiffrierung feindlicher Funksprüche im zweiten Weltkrieg und der Entwicklung der Informationstheorie nach Claude Shannon schlug Weaver sowohl linguistische als auch statistische Ansätze vor und sprach dabei besonders die Problematik der Sprachambiguität an (vgl. Hutchins 2001: 2; 2008:1; 2014a: 1).

Es wurde bereits an zahlreichen US-Universitäten an maschineller Übersetzung geforscht, als Yehoshua Bar-Hillel, der leitende Forscher für MÜ am Massachusetts Institute of Technology (MIT), im Juni 1952 die erste Konferenz zur MÜ einberief (vgl. Hutchins 2008: 1). In der Zwischenzeit war klar geworden, dass eine völlig automatisierte Übersetzung ohne menschliches Pre- oder vor allem Post-Editing qualitativ nicht zufriedenstellend war. Daher wurden in der Konferenz neben der bereits angesprochenen Ambiguität auch Möglichkeiten zur Integration von Mini-Glossaren und Analysen von syntaktischen Strukturen diskutiert. Im Zuge der Planung zukünftiger Forschungsbereiche, betonte Leon Dostert von der Georgetown University wie wichtig es sei der Öffentlichkeit die Realisierbarkeit von MÜ zu beweisen, um somit mehr Forschungsgelder zu erhalten (vgl. Hutchins 2001: 2).

So kam es, dass am 7. Januar 1954 das erste maschinelle Übersetzungssystem, welches aus der gemeinsamen Arbeit von Peter Sheridan von IBM und Paul Garvin von der Georgetown University, der Öffentlichkeit präsentiert wurde. Obwohl es mit einem Vokabular von nur 250 Wörtern und 6 Grammatikregeln wissenschaftlich gesehen kein großer Durchbruch war, war die Übersetzung von 49 Sätzen aus dem Russischen ins Englische

beeindruckend genug, um eine erhebliche Menge an Fördergeldern zu generieren und das Interesse für MÜ auch in anderen Teilen der Welt, vor allem der UdSSR, zu wecken (vgl. Hutchins 2001: 2). Hierbei waren die wichtigsten Forschungsgruppen jene des MIT, der Harvard, Georgetown und Washington University sowie der RAND Corporation (vgl. Hutchins 2008: 1).

1.1.3 Der Optimismus

Das erste Jahrzehnt der intensiven Forschungen im Bereich der MÜ war von drei Ansätzen geprägt. Der erste Ansatz war die *direkte Übersetzung* (direct translation vgl. Hutchins 2014b: 1) bei dem versucht wurde Regeln zu programmieren, welche die Übersetzung einer bestimmten Ausgangssprache (AS) in eine bestimmte Zielsprache (ZS) ermöglichen sollte. Dabei wurde danach gestrebt die zweisprachigen Wörterbücher insofern zu vereinfachen, als Äquivalente in der ZS gesucht wurden, die so viele Bedeutungsvarianten wie möglich zu erfüllen hatten, sodass keine Kontextanalyse nötig war und auch die Syntax des Originals nahezu unverändert blieb. Der zweite, sogenannte *interlingua* Ansatz war langfristiger angelegt und untersuchte die Nutzung interlingualer Repräsentationen in der Form von AS und ZS unabhängigen Symbolen, was einen zweiteiligen Übersetzungsprozess bedeutete (aus der AS in die Interlingua und von der Interlingua in die ZS). Es wurden zwar linguistisch orientierte Untersuchungen vorgenommen, jedoch waren viele der Meinung, dass nur die genaue Erforschung menschlicher Denkprozesse Antworten auf die vielen Fragen und Probleme der MÜ liefern würde, woraus infolge auch die Kognitionswissenschaften und die Forschung zur Künstlichen Intelligenz entstanden (vgl. Hutchins 2001: 3ff). Mit diesem Ansatz arbeiteten besonders Forschungsgruppen in Cambridge, Mailand und Leningrad. Der dritte Ansatz, der *Transferansatz* wurde vor allem am MIT praktiziert und war eine Art syntaktischer Transfer, wobei „[...] a SL grammar analyzed input sentences as phrase structure representations, a ‘structure transfer routine’ converted them into equivalent TL phrase structures, and the TL grammar rules produced output text.“ (Hutchins 2001: 5) In diesem Fall ist der Prozess dreiteilig und besteht aus der Analyse-, Transfer- und Synthesephase.

Die Forschungsgruppen, welche sich für den Wort-für-Wort Ansatz entschieden, waren vor allem jene, die auf ein zeitnahes und in der Praxis anwendbares System abzielten, wie z.B. Washington, IBM und Georgetown. Aufgrund der politischen Lage konzentrierte sich die Arbeit der US Forscher hauptsächlich auf Russisch-Englische Übersetzungen, während die sowjetische Forschung an Englisch-Russischen Systemen arbeitete, jedoch aufgrund ihres multilingualen Grundsatzes die höchste Anzahl an erforschten Sprachenpaaren aufwies (vgl. Hutchins 2001: 3f). So schuf IBM in Kooperation mit der University of Washington ein spezielles Gerät, das auf einem großen bilingualen Wörterbuch

basierte und 1958 von der US Air Force in Betrieb genommen wurde. Es produzierte bis 1970 Rohübersetzungen, bis es schließlich von Systran abgelöst wurde (vgl. Hutchins 2001: 4). Die umfangreichste Forschungsgruppe zur MÜ befand sich an der Georgetown University und deren System umfasste drei verschiedene Analyseebenen: die morphologische, syntagmatische und syntaktische Ebene. Das System (ebenfalls Russisch-Englisch) wurde 1961/1962 der Öffentlichkeit präsentiert und schließlich 1963 bei der Euratom in Italien und 1964 bei der US Atomic Energy Authority eingesetzt (vgl. Hutchins 2014a: 2). Der Forschungsbereich der MÜ hatte internationales Interesse geweckt und die Entstehung neuer Forschungszentren in Europa und Asien nach sich gezogen, jedoch sollte die anfängliche Euphorie relativ rasch der Ernüchterung weichen.

1.1.4 Die Ernüchterung

Der Optimismus des ersten Jahrzehnts begann langsam zu schwinden, da nicht nur die Komplexität der Sprache ein ernsthaftes Problem darstellte, sondern auch die technologischen Ausstattung und die begrenzten Kapazitäten bzgl. Leistung und Speichermedien für das euphorisch begonnene Unterfangen schlicht nicht ausreichend waren. Die Skeptiker häuften sich und Kritik kam sogar aus den eigenen Reihen. Yehoshua Bar-Hillel publizierte 1960 auf Grundlage persönlicher Treffen mit den größten MÜ Forschungseinrichtungen der USA einen Bericht, in dem er die Zielsetzung einer, wie er sie nennt, „fully automatic high quality translation“ (FAHQT) mit einer Illusion gleichsetzt (vgl. Bar-Hillel 1960).

„One of the reasons why we do not as yet have any translation centers, not even in the planning stage, in which electronic computers, general or special purpose, are used to automate certain parts of the translation process, in spite of the fact that such centers would fulfill a vital function in saving a considerable amount of qualified human translator time per document translated, and thereby facilitate more, quicker and, after some time, cheaper translation, is the reluctance of many MT workers to recognize that the idea of inventing a method for fully automatic high quality translation (FAHQT) is just a dream which will not come true in the foreseeable future.“ (Bar-Hillel 1960: 41)

Auffallend ist, dass Bar-Hillel bei seiner Argumentation zwar alle Forschungseinrichtungen und deren Projekte vorstellt, jedoch weder Beispiele für hervorgegangene (fehlerhafte) Übersetzungen noch eventuelle linguistische oder technische Schwierigkeiten nennt. Sein Bericht konzentriert sich auf die Frage der Methodik hinter den Ansätzen und legt mit einem simplen Beispiel nahe, dass kontextsensitives Übersetzen nicht maschinell durchführbar ist (vgl. Bar-Hillel 1960: 41ff). Sein Vorschlag war es die Ziele der MÜ niedriger anzusetzen und „thereby turning, strictly speaking, machine translation into *machine aids to translation*“. (Bar-Hillel 1960: 3) [Hervorhebung im Original]

Der Bericht hatte in der Öffentlichkeit große Wellen geschlagen und mit der Zeit stellte sich heraus, dass die intensive Forschung zur Syntaxanalyse lediglich die wahre

Komplexität syntaktischer Ambiguität ans Licht brachte - die semantische Grenze des Möglichen war erreicht worden (vgl. Hutchins 1986: 12f). 1964 äußerte sich schließlich auch die Sorge der US Sponsoren rund um den mangelnden Fortschritt in der Gründung des Automatic Language Processing Advisory Committee (ALPAC). Das Komitee sollte den allgemeinen Nutzen von MÜ-Systemen und die Brauchbarkeit von bereits existierenden Systemen untersuchen. Der 1966 publizierte Bericht besagte, dass MÜ im Vergleich zu von Menschen ausgeführten Übersetzungsprozessen ungenauer, langsamer und doppelt so teuer sei und Darüber hinaus gebe es „no immediate or predictable prospect of useful machine translation“. (ALPAC 1966: 32) Ein Punkt, der als Nachteil für die MÜ Systeme gewertet wurde, war die Tatsache, dass die generierten Übersetzungen noch eine Post-Editing-Phase durchlaufen mussten, bevor sie repräsentabel waren. Dabei ist das Lektorat bei professionellen Übersetzerinnen und Übersetzern ein selbstverständlicher und unerlässlicher Qualitätssicherungsschritt.

Obwohl das Urteil des Berichtes auch in anderen Aspekten häufig als kurzsichtig und engstirnig kritisiert wurde, hatte es schwerwiegende Folgen für die MÜ in den USA, da von weiteren Investitionen in die bisherigen Forschungsbereiche abgeraten wurde und stattdessen die Entwicklung von Systemen für maschinenunterstütztes Übersetzen empfohlen wurde (vgl. Hutchins 2014a: 2). Die Forschungsarbeiten zur MÜ standen in den USA zehn Jahre lang aufgrund der fehlenden Forschungsgelder nahezu still. Gegen Ende der 1970er zählten die TAUM Gruppe mit ihrem METEO System in Montreal, die GETA Gruppe in Grenoble und die SUSY Gruppe in Saarbrücken zu den bedeutendsten Forschungsprojekten der MÜ (vgl. Arnold et al. 1994: 14).

1.1.5 Die Renaissance

Während die Forschungsarbeiten vor dem ALPAC Bericht eher auf dem direkten Ansatz, d.h. auf der Wortebene, basierten, war die Zeit nach dem Bericht geprägt von der Weiterentwicklung des indirekten Ansatzes, also des interlingua- und besonders des transferbasierten Übersetzungsverfahrens, welches zuvor wissenschaftlich vernachlässigt worden war. Durch die Fokussierung auf konkrete Arbeitsbereiche und spezifische Fachgebiete und ein steigendes Übersetzungsvolumen in internationalen Institutionen keimte in den 1970er Jahren das Interesse an MÜ-Systemen wieder auf (vgl. Krenz & Ramlow 2008: 33). Das bereits erwähnte System SYSTRAN, welches trotz des ALPAC-Berichts in den USA umgesetzt wurde und mehrere Jahre bei der US Air Force und NASA in Verwendung war, wurde 1976 von der Europäischen Kommission gekauft und um eine Französisch-Englische und Italienisch-Englische Version erweitert (vgl. Arnold et al. 1994: 14).

Die 1980er Jahre waren geprägt von einer beachtlichen Auswahl an MT Systemen, welche aus den unterschiedlichsten Teilen der Welt stammten. Neben SYSTRAN, existierten das u.a. von Siemens geförderte METAL System (Deutsch-Englisch), Logos (Deutsch-Englisch und Englisch-Französisch), das von der Pan American Health Organisation entwickelte System (Englisch-Spanisch und Spanisch-Englisch), sowie Systeme für den japanischen Markt, wo ausgehend von Computerherstellern wie Fujitsu, Hitachi, Sharp oder Toshiba kommerzielle Software für computerunterstütztes Übersetzen entwickelt wurde (vgl. Hutchins 2001: 8-10, 2014a: 3). Mitte der 1980er Jahre begann auch die Forschung im Bereich der Übersetzung von gesprochener Sprache (Spoken Language Translation), welche neben der MT die zweite Schlüsselrolle im Untersuchungsgegenstand Translation Apps spielt.

1.2 Zur iPhone App durch Speech-to-Speech Translation

Mit den Erfolgserlebnissen der letzten Jahre stiegen auch die Anforderungen der Forscher an ihre Systeme. So gab man sich nicht mehr mit der maschinellen Übersetzung geschriebener Texte zufrieden, sondern versuchte auch mündlich dargebrachte, also gesprochene Texte für die MT Systeme bearbeitbar zu machen. Dafür bedarf es abgesehen von der Forschung zur MT auch intensiver Forschungsarbeiten auf zwei weiteren Ebenen – der Spracherkennung und der Sprachsynthese. Die wichtigsten Pionierprojekte werden in einem eigenen Kapitel etwas genauer vorgestellt; aus diesem Grund wird an dieser Stelle auf Tabelle 1 verwiesen, welche einen Überblick über die Entwicklungsphasen der Speech-to-Speech Translation Systeme bieten soll.

	Jahr	Wortschatz	Sprechweise	Anwendungsbereich	Speed	Medium	Beispiele
Erste Dialogsysteme	1989 – 1993	eingeschränkt	kontrolliert	beschränkt	2-10 x RT	Workstation	C-STAR I
One-way Phrasebooks	1997 – heute	eingeschränkt, modifizierbar	kontrolliert	beschränkt	1-3 x RT	Handgerät	Phraselator, Ectaco
Two-way Systeme mit Spontanspr.	1993 – heute	nicht eingeschränkt	spontan	beschränkt	1-5 x RT	PC/ Handgerät	C-STAR, VERBMOBIL, NESPOLE, BABYLON, TRANSTAC
Übersetzung von Nachrichten pol. Reden	2003 – heute	nicht eingeschränkt	lesend/ vortragend	unbeschränkt	offline	PC, PC-Clusters	NSF-STRDUST, EC TC-STAR, DARPA GALE
Simultane Dolmetschung von Vorträgen	2005 – heute	nicht eingeschränkt	spontan	unbeschränkt	Realtime	PC, Laptop	LECUTRE TRANSLATOR
Smartphone Apps für die breite Masse	2009 – heute	nicht eingeschränkt	spontan	unbeschränkt	Verzögerung von 1-3 Sek.	Smartphone	Jibbig, Google Translate, Say Hi, VoiceTra4U

Tabelle 1: Überblick über Entwicklung und Funktionen von Speech-to-Speech Translation Systemen (nach Waibel & Fügen 2009: 73) (Anmerk.: übersetzt und ergänzt durch Verf.)

1.2.1 Phase der Einschränkungen

Bei den ersten Speech-to-Speech Translation Systemen waren noch sehr viele Einschränkungen beobachtbar; diese reichten von sprachlichen Hürden wie Phonetik, Syntax oder Semantik bis hin zu beschränkten Anwendungsgebieten und Sprechtempo. Gesprochene Sprache ist insofern komplexer als geschriebene Sprache, als sie von Person zu Person unterschiedlich ist, Selbstkorrekturen enthält, grammatikalisch falsch und sehr kontextabhängig sein kann.

„The research in speech translation is beset with numerous problems, not just variability of voice input but also the nature of spoken language. By contrast with written language, spoken language is colloquial, elliptical, context-dependent, interpersonal, and primarily in the form of dialogues. MT has focused on well-formed, technical and scientific language and has tended to neglect informal modes of communication. Speech translation therefore represents a radical departure from traditional MT.“ (Hutchins 2009: 16)

Um diese Komplexität in den Griff zu bekommen wurden die ersten Systeme für bestimmte Anwendungsbereiche konzipiert und verfügten somit nur über einen beschränkten Wortschatz, der in dem konkreten Setting z.B. Businesskommunikation, Hotelbuchungen, Ter-

minvereinbarungen genutzt werden konnte. Es war demnach nicht möglich frei zu sprechen und eigene Sätze zu bilden; vielmehr mussten die SprecherInnen bestimmte Satzmuster einhalten oder aus einem Pool an möglichen Formulierungen wählen und diese klar, deutlich und nicht zu schnell aussprechen (vgl. Waibel & Fügen 2008: 73).

1.2.2 Phase der spontanen Sprache

Laut Waibel & Fügen (2008: 74) war bereits 1992 klar, dass diese ersten Dialogsysteme sehr realitätsfremd waren, da die AnwenderInnen sehr auf klare und deutliche Aussprache achten mussten und dabei nur eine beschränkte Anzahl von möglichen Satzkonstellationen hatten.

„The most unacceptable constraints were the vocabulary, syntax, and speaking-style limitation, as it is generally not possible for humans to speak flawlessly in a limited speaking style (effectively reading sentences) or remember a limited set of words or syntactic patterns. By contrast, it is generally possible for humans to stick to a domain of discourse when solving certain limited tasks. Many important applications are inherently domain limited, thus making spontaneous domain-limited speech translation a practically useful technology.“
(Waibel & Fügen 2008: 74)

Die Einschränkung der verwendbaren Lexik und Syntaxabfolgen versicherte zwar theoretisch bessere Übersetzungsergebnisse, jedoch wirkte die Sprechweise der AnwenderInnen unnatürlich und machte das System für den Einsatz in der Praxis eher unbrauchbar. Ein neuer Ansatz musste her. Die Forschung konzentrierte sich auf die Eliminierung syntaktischer Einschränkungen und auf den flexibleren Umgang mit unterschiedlichen Formulierungen. Die AnwenderInnen mussten zwar ihren Diskurs innerhalb eines eingeschränkten Kontextes halten, jedoch konnten sie innerhalb des Anwendungsbereiches frei und spontan formulieren.

Hierbei waren die Forschergruppen der ATR Laboratories in Japan, der Carnegie-Mellon University in den USA, des VERBMOBIL Projekts in Deutschland sowie ITC-irst und NESPOLE in Italien federführend. Im Rahmen von C-STAR (Consortium for Speech Translation Advanced Research) entstand auch eine Kooperation von ATR, der Carnegie-Mellon University und der Universität Karlsruhe, in der an einem sprecherunabhängigen Übersetzungssystem für das Telefon mit den Sprachen Japanisch, Englisch, Deutsch geforscht wurde (vgl. Hutchins 2009: 16; Hutchins 2010: 4).

1.2.3 Phase der tragbaren Systeme

Im Laufe der Jahre wurden auch die Geräte, mit denen Speech-to-Speech Translation Systeme genutzt werden konnten, immer handlicher und mobiler. Die brachte natürlich auch

gewisse Hardwarebeschränkungen mit sich, da die Speicherkapazitäten und die Rechnerleistung eines PDA wesentlich geringer waren als die eines Laptops. Zudem musste auch die Benutzerfreundlichkeit auf dem Feld beachtet werden. So entstanden eigens für bestimmte Situationen konzipierte Geräte wie der Phraselator³, der ursprünglich für das U.S. Militär entwickelt worden war. Es handelt sich um eine Art one-way Phrasebook in dem AnwenderInnen aus einer Datenbank vorgegebene Phrasen oder Sätze durch Spracherkennung aktivieren oder direkt auf dem Display auswählen können, welche dann nicht durch Sprachsynthese in der Fremdsprache generiert werden, sondern bereits im Vorhinein von einem Native Speaker als mp3 file aufgesagt wurden. Es wurde von der US Army und der US Navy in verschiedensten Operationen im Irak, in Afghanistan oder auch in Indonesien bei den Rettungsaktionen nach dem Tsunami 2005 genutzt (vgl. Hutchins 2009: 17; Waibel & Fügen 2008: 74-75).

Ein System, welches auf jedem Standard PDA lief und für Speech-to-Speech Translation im touristischen, medizinischen und militärischen Bereich spezialisiert war, war der pocket translator von Mobile Technologies (vgl. Waibel & Fügen 2008: 74). Das Interface, mit einem Fenster für den Satz in der Ausgangsprache, welcher nach der Spracherkennung noch manuell korrigiert werden kann, und einem Fenster mit dem Übersetzungsvorschlag, erinnert bereits sehr stark an das Layout von Speech-to-Speech Translation Apps für das Smartphone. Dies ist nicht sonderlich verwunderlich, da Mobile Technologies, unter der Leitung von Prof. Dr. Alexander Waibel, der Entwickler der ersten Translation App Jibbig (2009) auf dem Markt war (vgl. Waibel 2012c).

Der nächste logische Schritt war die Loslösung von beschränkten Anwendungsbereichen, sodass nicht nur die Sprache der AnwenderInnen spontan gewählt werden konnte, sondern auch das Thema des Gespräches.

1.2.4 Phase der uneingeschränkten Anwendung

Die bereits existierenden Systeme zeigten in ihren konkreten Anwendungsbereichen bereits gute Resultate, jedoch impliziert Spontansprache nicht nur spontane Selbstkorrekturen oder Füllwörter wie ähm, sondern eben auch spontane Anekdoten oder Abschweifungen vom eigentlichen Gesprächsthema. 2003 begann die Forschung sich dieser letzten großen Hürde anzunehmen und es kam zur Entstehung von diversen Projekten.

Dazu gehören das STR-DUST Projekt (Speech Translation for Domain-Unlimited Spontaneous Communication Tasks, Projekt der National Science Foundation) in den USA (2003), TC-STAR (Technology and Corpora for Speech to Speech Translation, Projekt der

³ Weitere Informationen zu technischen Details und Funktionen unter <http://www.voxtec.com/phraselator/> (10.05.2014)

Europäischen Union) in Europa (2004) und das U.S.-DARPA GALE Projekt (Global Autonomous Language Exploitation, Projekt der DARPA) in den USA (2006). Bei diesen Projekten wird auf Basis vieler unterschiedlicher Szenarien wie z.B. Nachrichtensendungen, politischen Reden oder akademischen Vorträgen gearbeitet (vgl. Waibel & Fügen 2008: 75).

Zur Phase der uneingeschränkten Anwendungsbereiche zählt auch der Untersuchungsgegenstand Speech-to-Speech Translation Apps, welche aufgrund ihrer Verfügbarkeit für die breite Masse zwingendermaßen jeden erdenklichen Kontext abdecken müssen. Inwiefern sie diese nicht zu unterschätzende Aufgabe erfüllen, wird in der Beispieldiskussion ersichtlich werden.

1.3 Bis hin zur MT im World Wide Web

Nach dem kurzen Exkurs zum konkreten historischen Hintergrund der Speech-to-Speech Translation, soll nun noch ein Überblick über die Entwicklungen der traditionellen MT seit den frühen 90er Jahren gegeben werden.

Personal Computers oder kurz gesagt PCs fanden in den Achtziger Jahren ihren Einzug in Unternehmen und private Haushalte, kurze Zeit später gab es auch schon die ersten Anbieter von kommerzieller Übersetzungssoftware. Trados wurde 1984 in Stuttgart gegründet und sollte mit Multiterm (1992) und Translator's Workbench (1994) das erste Unternehmen werden, welches die neuartige Translation Memory Technologie entwickelt und vermarktet. Einen wirtschaftlich wichtigen Schub bekam das Unternehmen 1997 als Microsoft sich dazu entschloss Trados für ihre eigenen internen Lokalisierungsbedürfnisse zu nutzen (vgl. <https://www.taus.net> 2014).

Mit der öffentlichen Zugänglichkeit der Internets ab dem Jahr 1993 und dem enormen Anstieg der Zahl an Websites und somit auch online verfügbaren Texten, kam es zu einer Verlagerung der MT ins World Wide Web. Globalink bot ab 1994 einen kostenpflichtigen online Übersetzungsservice an, bei dem die Texte hingeschickt wurden und die Übersetzungen per E-Mail retourniert wurden (vgl. Hutchins 2014b: 17-18). Ein ähnliches System bot auch Systran Express an, jedoch war wohl der Launch von AltaVista's Babel-fish, der auf verschiedenen Systran MT Systemen basierte, der Grund für die rasche Verbreitung von MT in der breiten Masse. Denn die Übersetzungen waren nicht nur kostenlos, sondern auch augenblicklich vorhanden. Hutchins schneidet hierbei einen Punkt an, der auch für die vorliegende Arbeit von großer Wichtigkeit ist. Während in den bisherigen (auch kommerziellen) MT Systemen Wert auf qualitativ zufriedenstellende Übersetzungen

gelegt wurde, „[...] there was an even more rapid growth of automatic translation for direct Internet applications (electronic mail, Web pages, etc.), where the need has been for fast real-time response with less importance attached to quality“. (Hutchins: 2014a: 4) Mit Babelfish oder heute auch Google Translate ist die MT zu einem Massenprodukt geworden. Obwohl sich die Qualität der maschinellen Übersetzungen von Babelfish bzw. Google Translate im Laufe der Jahre enorm gesteigert hat, darf nicht vergessen werden, dass diese Programme in erster Linie einen schnellen und leichten Weg darstell(t)en, um an Übersetzungen zu gelangen. Wie auch Hutchins anspricht, war die Qualität hierbei zweitrangig. Genau diesen Aspekt gilt es auch im Laufe dieser Arbeit nicht aus dem Auge zu verlieren. Die damaligen Online MT Services sind ebenso wie die heutigen Apps nicht an einen Anwendungsbereich gebunden und haben dadurch mehr potentielle Hürden zu überwinden, als speziell für den beispielsweise medizinischen Bereich entwickelte MT Systeme. Dadurch liegt auch die Fehlerquote höher und die Online MT Anwendungen bieten mehr Angriffsfläche für Kritik. Während die Online MT einerseits sehr zur Popularität der MT beigetragen hat, war dieser Beitrag selbstverständlich nicht immer ausschließlich positiver Natur.

„This increased visibility of MT has had a number of side effects. It has of course increased the general public's awareness of MT, in some cases clarifying its limits but also its benefits. Informal reports suggest that users are at first impressed, then disappointed as they realize its limitations, but finally pragmatic as they learn to get the best out of raw MT. One thing to notice is how using MT to translate the unrestricted (and sometimes poorly written) material that is found on web pages goes against the general recommendations for the use of MT that have been made over the last decade or so. There is certainly a need to educate the general public about the low quality of raw MT, and, importantly, why the quality is so low.“ (Somers 2005: 523)

Immer wieder würden die Online MT Anwendungen „Evaluationen“ durch Laien durchlaufen müssen, in welchen Übersetzungsperformance bei bewusst schwer gewählten Beispielen wie idiomatischen Phrasen, doppeldeutigen Wörtern oder komplexer Syntax getestet wurde und – wenig überraschend – als nicht gut befunden wurde. Dies ist eine der Schattenseiten mit der die Online MT seit ihrer rasanten Verbreitung in den 90er Jahren auch heute immer noch zu kämpfen hat.

Laut Hutchins (2009: 15-16) weiß man sehr wenig darüber wer online MT nutzt und wofür diese genau genutzt wird. Es ist nicht bekannt wie alt die UserInnen sind, wie gut ihre Sprachkenntnisse sind, ob sie in die oder aus der Muttersprache übersetzen, ob der MT Output nur als Rohfassung verwendet wird oder nicht, usw.

„Almost all that we do know are the surprising facts that translation of web pages is very much a minor use (no more than about 15% at best), that the average length of texts submitted is just 20 words, and that more than 50% of submissions are one- or two-word phrases. It had been anticipated that longer texts would be submitted – the general maximum length of 150 words is clearly no impediment – and that much of the translation would be of web pages.“ (Hutchins 2009: 16)

Diese überraschenderweise doch relativ kurzen Texte bzw. Satzbausteine scheinen darauf hinzudeuten, dass Online MT Anwendungen wie Google Translate wohl eher als Ersatz für Wörterbücher genutzt werden, obwohl diese auch zuhause online verfügbar sind. Trotz der Anonymität der Nutzer wachsen die Zugriffszahlen von Jahr zu Jahr steil. So steigerte sich z.B. FreeTranslation von 50.000 Zugriffen im Jahr 1999 auf 3,4 Mio Zugriffe im Jahr 2006 (vgl. Hutchins 2009: 16). Im Vergleich dazu publizierte der inzwischen unangefochtene Spitzenreiter Google Translate 2012 anlässlich des 6. Jubiläums der MT-Anwendung folgende Zahlen:

„Today we have more than 200 million monthly active users on translate.google.com (and even more in other places where you can use Translate, such as Chrome, mobile apps, YouTube, etc.). People also seem eager to access Google Translate on the go (the language barrier is never more acute than when you're traveling)—we've seen our mobile traffic more than quadruple year over year. And our users are truly global: more than 92 percent of our traffic comes from outside the United States.“ (Och 2012)

Die Nachfrage nach MT ist in wenigen Jahren massiv gestiegen und nicht nur in den eigenen vier Wänden auf dem PC, sondern auch unterwegs im Alltag. Das neue Stichwort lautet Mobilität - genau diese mobilen Übersetzungsdienste bieten Translation Apps. Die MT hat sich von ihrer ursprünglichen Aufgabe der „reinen“ Übersetzung inzwischen weit entfernt und eine vielfältige Entwicklung durchlebt, um nun in einem sehr breiten Kommunikationskontext und diversen Kommunikationssituationen seine Anwendung zu finden. Das Potential der MT ist sicherlich noch nicht ausgeschöpft und die Einsatzbereiche scheinen bei der rasanten technologischen Entwicklung schier unendlich.

2 Die Komponenten eines Speech-to-Speech Translation Systems

Da nun der historische Überblick über die Entwicklungsstadien der Maschinellen Übersetzung genügend beleuchtet wurde, soll im folgenden Kapitel die technische Seite der Speech-to-Speech Translation Systeme genauer analysiert werden, um die Funktionsweise der Apps zu verstehen.

In der Literatur finden sich verschiedene Benennungen für diese bestimmte Art von Machine Translation, die mit gesprochenen Texten arbeitet. Da der Großteil der computergewissenschaftlichen Literatur und Forschungsarbeiten hierbei auf Englisch publiziert wird, hat sich die Benennung *Speech-to-Speech Translation* (SST; auch S2ST) weitestgehend durchgesetzt (vgl. Kitano 1994; Wahlster 2000; U-STAR 2014a). Es finden sich auch andere Begriffe wie *Spoken Language Translation* (vgl. Waibel & Fügen 2008) oder auch *Simultaneous Speech Translation* (vgl. Fügen et al. 2008). Im Zuge dieser Arbeit wird in Anlehnung an die Literatur auch der Begriff Speech-to-Speech Translation verwendet.

Um mündlich dargebrachte Texte maschinell übersetzen zu können, bedarf es drei Komponenten – der Spracherkennungskomponente, der Komponente für die maschinelle Übersetzung und schließlich der Sprachsynthese. Die Sprachsynthese spielt oft eine untergeordnete Rolle, da bei manchen SST Systemen der Output auch andere Formen annehmen kann. Beim *Lecture Translator* (vgl. Kap 3.2) am Karlsruher Institut für Technologie kann der Zieltext (ZT) in Form von Sprache oder in Form von Untertiteln präsentiert werden (vgl. Fügen et al. 2008: 216). Da beim Untersuchungsgegenstand SST Apps der ZT in gesprochener Form wiedergegeben und gleichzeitig auf dem Display des Smartphones schriftlich dargestellt wird, ist die Komponente der Sprachsynthese ebenso relevant wie die zwei vorangehenden Komponenten und wird deshalb in den nächsten Kapiteln auch erläutert.

Die zugrundeliegenden Technologien der drei Komponenten wurden unabhängig von einander entwickelt und finden mit all ihren individuellen Vor- und Nachteilen in der SST eine gemeinsame Aufgabe. Technologische Verbesserungen der einzelnen Elemente können zwar die Leistung des gesamten Systems steigern, jedoch könnte hierbei ein komponentenübergreifender Ansatz erfolversprechender sein als die Konzentration auf die einzelnen Bestandteile:

„[...] [A] speech translation system is not only the cascade of its parts. Since the goal is to produce output in a target language, the correctness of the components' output is of secondary concern. Uncertainty at the component level can be addressed by being noncommittal at their interface, linking components via near-miss hypothesis lattices [1]. Such a view also offers the possibility of jointly optimizing components based on the overall output rather than each component independently.“ (Waibel & Fügen 2008: 71)

Im Folgenden werden nun die einzelnen Komponenten von SST Systemen vorgestellt und ihre Stärken und Schwächen im Hinblick auf ihre Wichtigkeit für SST Apps erläutert werden.

2.1 Automatische Spracherkennung

Die Automatische Spracherkennung (automatic speech recognition, ASR) beschreibt den automatisierten Prozess der Verarbeitung von gesprochenen Wörtern in schriftlichen Texten:

„By automatic speech recognition we mean the process of speech-to-text transcription: the transformation of an acoustic signal into a sequence of words, without necessarily understanding the meaning or intent of what was spoken. Recognition without understanding is not always possible since some semantic context may be required, for instance, to disambiguate Will the new display recognize speech? from Will the nudist play wreck a nice beach?“ (Renals & Hain 2010: 299) [Hervorhebungen im Original]

Im konkreten Beispiel der SST Apps wird der Ausgangstext (AT) aufgesagt, durch die ASR Komponente analysiert und auf dem Display transkribiert dargestellt, bevor die MT Komponente einen ZT generiert und wiederum auf dem Display in der Zielsprache darstellt und im letzten Schritt der Sprachsynthese in gesprochener Sprache wiedergibt. Welche Herausforderungen gesprochene, spontane Sprache im Gegensatz zur Schriftsprache mit sich bringt, soll im Folgenden aufgezeigt werden.

2.1.1 Die Tücken der Spontansprache

Die gesprochenen Worte bzw., aus Sicht der ASR, die akustischen Signale, transportieren selbstverständlich mehr als nur lexikalische Informationen. Die menschliche Stimme impliziert durch die Intonation oder auch das Timing paralinguistische Aspekte, welche Auskunft über mit der Aussage verbundene Emotionen geben oder auch Hinweise zur Unterscheidung von Doppeldeutigkeiten liefern (vgl. Renals & Hain 2010: 299). Hinzu kommt, dass die ASR mit den typischen Merkmalen der spontan gesprochenen Sprache (auch oft als natürliche Sprache bezeichnet) umgehen muss. Hacker et al. (2012: 5-6) beschreiben gesprochene Sprache im Vergleich zur Schriftsprache als:

- spontan und flüchtig
- interaktiv
- situativ
- ökonomisch
- personenbezogen

Gesprochene Sprache ist spontan und flüchtig, weil einerseits keine Zeit für Vorausplanen gegeben ist, Gesagtes im Nachhinein nicht mehr abgeändert oder zurückgenommen werden kann und andererseits Gesagtes nur für kurze Zeit im Gedächtnis erhalten bleibt. Die ersten zwei Punkte Spontanität und Flüchtigkeit sind für die ASR Komponente der SST Apps keine allzu große Hürde, da innerhalb weniger Sekunden der eingegebene AT manuell am Display geändert werden kann oder komplett gelöscht werden und erneut eingegeben werden kann. Darüber hinaus werden die bereits übersetzten Segmente in einer Art Verlauf auf dem Display dargestellt und können auch später wiedergefunden werden.

Gesprochene Sprache wird als interaktiv beschrieben, weil sie Unterbrechungen, Nachfragen und Rückmeldungen enthält. Dieser Punkt könnte auch bei den SST Apps relevant sein, da die ASR bei Unterbrechungen o.Ä. den Sprecherwechsel nicht erkennen könnte und somit zu verzerrten Ergebnissen der Spracherkennung und in weiterer Folge der Übersetzung führen könnte.

Gesprochene Sprache zeichnet sich durch ihren situativen Charakter aus, weil sie kontextbezogen auf die konkrete Situation ist und häufig Bezüge auf zuvor Gesagtes aufweist oder auch Präsuppositionen enthält, welche auf gemeinsamen Erlebnissen, Erfahrungen oder Kenntnissen der Gesprächspartner beruhen. Dieser Aspekt ist auch für die SST Apps sicherlich eine nicht zu unterschätzende Hürde. Obwohl der Gesprächsverlauf gespeichert wird und auf dem Display zurückverfolgt werden kann, bedeutet dies nicht automatisch, dass das System auf die bereits bearbeiteten Segmente zurückgreift und Beziehungen zwischen ihnen herstellt.

Gesprochene Sprache ist ökonomisch, das bedeutet, dass auf komplizierte Syntaxkonstellationen verzichtet wird und nicht ausschweifender gesprochen wird als es für das Verständnis des/der Gesprächspartner/in nötig ist. Darüber hinaus ist die Häufigkeit von Ellipsen besonders hoch. Die Ökonomie der gesprochenen Sprache könnte auf den ersten Blick auch als Vorteil für die ASR gewertet werden, weil die Syntax einfacher ist und die Sätze kürzer sind, was das Fehlerpotential senken könnte. Jedoch relativieren die häufigen Ellipsen diese Annahme rasch, da sie als unvollständige Sätze für die MT Komponente im späteren Verlauf zu großen Hürden werden können.

Als letzten Punkt nennen Hacker et al. (2012: 6) den personenbezogenen Aspekt der gesprochenen Sprache. Dieser beschreibt ähnlich wie das situative Merkmal explizite oder implizite Hinweise auf gemeinsame Erfahrungen der GesprächsteilnehmerInnen und beinhaltet auch sprachliche Varietäten wie die Umgangssprache, den Dialekt oder auch Soziolekt. Da die SST Apps durch das Medium Smartphone ein sehr breites und diverses Zielpublikum ansprechen, spielt dieser Punkt eine wesentliche Rolle. Die App bzw. die ASR Komponente muss mit sehr vielen unterschiedlichen Dialekten und Sprachvarianten

umgehen können, um userfreundlich und sinnvoll nutzbar zu sein. Natürlich könnten die AnwenderInnen auch dazu angehalten werden die entsprechende Standardvariante der jeweiligen Sprache zu sprechen, jedoch würde auch dies nicht alle Hindernisse der Varietäten aus dem Weg räumen können und zudem die Usability der App beeinträchtigen.

Abgesehen von den bereits erläuterten Merkmalen der gesprochenen Sprache gibt es noch lexikalische und syntaktische Besonderheiten, welche bei der ASR bedacht werden müssen. Dazu zählen auf lexikalischer Ebene Klitika (hab's erledigt), umgangssprachliche Ausdrücke (hast heut Zeit?), Füllwörter (so, eben, also, halt), Platzhalter (das Ding) oder auch aus Unkenntnis resultierende falsche Nutzung von Wörtern (vgl. Hacker et al. 2012: 7-8). Bedingt durch die Globalisierung und Verbreitung von Sozialen Netzwerken sollten hier auch im Internet verbreitete Ausdrücke und Neologismen wie *liken*, *googlen*, *Hashtag* bedacht werden, die je nach Kontext nicht die ursprüngliche Wortbedeutung mehr aufweisen und auch in anderen Sprachen nicht übersetzt werden, sondern die englische Form beibehalten. Zu den syntaktischen Herausforderungen zählen Hacker et al. (2012: 7-8) die fehlende Interpunktion, welche meist nur durch die Intonation rekonstruiert werden kann, sowie verschwimmende Wortgrenzen und den freieren Satzbau. Eine nicht unwesentliche syntaktische Schwierigkeit stellen Disfluenzen dar. Dazu zählen u.a. Selbstkorrekturen und Satzabbrüche, Häitationen und Wortwiederholungen. Waibel und Fügen fassen diese Überlegungen treffend zusammen: „To err is human, and useful systems must accomodate a speaker's mistakes.“ (Waibel & Fügen 2008: 71)

Nur wenn ein SST System zumindest mit dem Großteil der genannten Schwierigkeiten der Spontansprache umgehen kann, ist es auch für den Endverbraucher sinnvoll nutzbar. Die in Kapitel 1 erwähnten ersten SST Systeme waren alle durch die Eingrenzung von Wortschatz, Anwendungsbereich oder Sprechweise in ihrer Usability beschränkt. Im Gegensatz dazu sind die auf dem Markt verfügbaren SST Apps sprecher- und domänenunabhängig, was gleichzeitig eine Vielzahl neuer Herausforderungen mit sich bringt, da jede/r Sprecher/in eine andere Stimmlage hat, in einem anderen Tempo spricht, möglicherweise heiser ist oder eine Sprechstörung hat. Zudem verändert sich die Aussprache abhängig von der emotionalen Lage oder dem situativen Kontext und es kommen akustische Störfaktoren wie Hintergrund- oder Störgeräusche hinzu.

Die Forschung um die automatische Spracherkennung hat sich bestimmte Ziele gesteckt, die nachwievor erreicht bzw. verbessert werden müssen. Hausser (2000: 31) zählt hierbei fünf verschiedene Desiderata der automatischen Spracherkennung auf:

- Sprecherunabhängigkeit
- Kontinuierlichkeit
- Domänenunabhängigkeit
- Realistischer Wortschatz
- Robustheit

Sprecherunabhängigkeit bedeutet, dass das System mit verschiedenen SprecherInnen, d.h. unterschiedlichen Stimmen, Tonlagen, Dialekten umgehen können muss; im besten Fall bedarf es keiner Anpassungsphase seitens des Systems, während der es sich an die Individualitäten der SprecherInnen adaptieren muss. Kontinuierlichkeit beschreibt die Bewältigung kontinuierlicher Spracheingaben in unterschiedlichem Tempo, ohne die AnwenderInnen zu unnatürlichen Pausen zwischen Wörtern oder Satzteilen zu zwingen. Domänenunabhängigkeit wurde bereits weiter oben angeschnitten. Das System soll für unterschiedliche Anwendungsbereiche einsetzbar sein und die eingegebene Sprache unabhängig vom Inhalt verstehen können, anstatt nur mit dem Vokabular eines bestimmten Bereiches (z.B. Hotelreservierungen) zu arbeiten. Der Wortschatz den das System versteht, soll realistisch sein, also mindestens dem Wortschatz eines durchschnittlichen Sprechers oder einer durchschnittlichen Sprecherin entsprechen. Robustheit bezeichnet die Fähigkeit des Systems auch bei verschiedenen Disfluenzen, wie Satzabbrüchen, Selbstkorrekturen oder Häsitationen, den Wortlaut sinnvoll erschließen zu können (vgl. Hausser 2000: 31).

2.1.2 Messung der ASR Qualität

Es hat sich in der Praxis etabliert die Spracherkennungsqualität in der sogenannten Wortfehlerrate (word error rate, WER) zu messen.

„The sequence of words output by the speech recognizer is aligned to the reference transcription using dynamic programming. The accuracy of the speech recognizer may then be estimated as the string edit distance between the output and reference strings. If there are N words in the reference transcript, and alignment with the speech recognition output results in S substitutions, D deletions, and I insertions, the word error rate is defined as:“ (Renals & Hain 2010: 303)

$$WER = 100 * \frac{S + D + I}{N} \%$$

$$Accuracy = (100 - WER) \% \quad (\text{Renals \& Hain 2010: 303})$$

Um die Wortfehlerrate zu erhalten, wird die durch ASR erkannte Wortsequenz der Referenzsequenz gegenübergestellt und mithilfe eines Algorithmus, der die vorgekommenen N-Wörter der Referenzsequenz, sowie die Substitutionen (S), Auslassungen (D) und Ein-

fügungen (I) beinhaltet, berechnet. Die Wortfehlerrate sollte hierbei laut Waibel & Fügen (2008: 71) niedriger als 10 % sein, um für die darauffolgende MT Komponente sinnvoll verarbeitbare Ergebnisse zu liefern. Dieser Wert ist im Vergleich zu anderen Einsatzgebieten der ASR (z.B. information retrieval) niedriger, da jene Anwendungen fehlertoleranter sein dürfen als die ASR Komponente eines SST Systems. Durch die bereits angesprochene Natur der Spontansprache, welche sich durch Disfluenzen auszeichnet, sind niedrige Wortfehlerraten umso schwieriger zu erreichen.

Eine weitere wichtige Aufgabe, die von der ASR Komponente erfüllt werden muss, bevor die MT Komponente ihre Arbeit antreten darf, ist die Segmentierung des Ausgangstextes. Da in der gesprochenen Sprache Satzzeichen beinahe nicht erkennbar sind, muss sich der Segmentationsalgorithmus natürlicher Pausen, Sprachmodelstatistiken und prosodischer Hinweise bedienen, um den Strom an Wörtern in sinnvolle Segmente zu unterteilen. Eine Optimierung der Latenzzeit (Verarbeitungszeit eines Segments) und somit des gesamten Übersetzungsvorganges ist nicht unbedeutend für die Verbesserung der Qualität (vgl. Waibel & Fügen 2008: 71f).

2.2 Maschinelle Übersetzung

Auf die bereits beschriebene ASR Komponente folgt die Machine Translation Komponente. Im Laufe der Forschungsjahre wurden verschiedene Ansätze entwickelt, die sich in ihrer Herangehensweise und Komplexität sehr unterscheiden. Zunächst wird zwischen dem linguistisch orientierten, regelbasierten Ansatz und dem empirisch orientierten statistikbasierten Ansatz unterschieden. Lange Zeit galten sie in der Literatur als konkurrierende Modelle, doch zeigt sich, dass die Forschung versucht, die Stärken der jeweiligen Ansätze in Hybridmodellen zu vereinen. Im Folgenden werden als erstes die regelbasierten Ansätze mit ihren jeweiligen Unterkategorien vorgestellt, bevor in weiterer Folge die statistikbasierten Ansätze erläutert werden.

2.2.1 Regelbasierte Ansätze

Laut Stein (2009: 7-8) handelt es sich bei den meisten heutzutage kommerziell erwerblichen MT Systemen um Modelle mit einem regelbasierten Ansatz (Rule-Based Machine Translation, RBMT). Qualitativ gesehen gibt es hier stark divergierende Übersetzungsergebnisse, abhängig von der Sprachpaarkombination, der (nicht)vorhandenen Fachterminologie etc. Die Übersetzung wird in einem dreiteiligen Prozess generiert, der die folgende Phasen umfasst: Analyse, Transfer, Synthese.

„The dominant framework of MT research until the end of the 1980s was based on essentially linguistic rules of various kinds: rules for syntactic analysis, lexical rules, rules for lexical transfer, rules for syntactic generation, rules for morphology, etc. The rule-based approach was most obvious in the dominant transfer systems of the 1980s [...].“ (Hutchins 2014b: 12)

Der Übersetzungsprozess eines RBMT-Systems beruht auf linguistischen Regeln. Zudem wird zwischen Systemen mit direktem Ansatz und indirektem Ansatz unterschieden.

Der direkte Ansatz ist historisch gesehen der Früheste und hat in der ersten Generation von MT Systemen Anwendung gefunden. Der Übersetzungsprozess basiert auf einer einzigen Phase, in der aus dem Ausgangstext direkt ein Zieltext generiert wird und ist für Übersetzungen zwischen einem bestimmten Sprachenpaar gedacht. Die Idee dahinter war die ausgangssprachlichen Termini den zielsprachlichen Termini mithilfe eines umfangreichen zweisprachigen Wörterbuchs oder Glossars zuzuordnen. Dazu durchliefen die Wörter eine Art morphologische Analyse, in der z.B. flektierte Formen zur ihrer Grundform gekürzt wurden, um so Äquivalente im Wörterbuch zu finden. Es fand jedoch keine Analyse der Syntax oder semantischen Beziehungen statt, vielmehr wurden die aufgefundenen Äquivalente in der ZS in Tempus, Genus, Numerus etc. angepasst und präsentiert (vgl. Ramlow 2008: 39; Hutchins 1992: 72). Da dieser Ansatz stark an eine simple Wort-für-Wort Übersetzung erinnert, ist es nicht verwunderlich, dass er sich in der Praxis nicht durchgesetzt hat. Die Qualität konnte nur so gut sein wie der Umfang des implementierten Wörterbuches und der morphologischen Analyseregeln. Selbstverständlich muss bei der Kritik des direkten Ansatzes bedacht werden, dass die technischen Voraussetzungen zur damaligen Zeit äußerst beschränkt waren und die IT erst in den Kinderschuhen steckte.

„The lack of computational sophistication was largely a reflection of the primitive state of computer science at the time, but it was also determined by the unsophisticated approach to linguistics in MT projects of the late 1950s.“ (Hutchins 1992: 73)

Abbildung 1 zeigt das Dreieck nach Vauquois, welches den direkten Ansatz und die beiden weiter unten beschriebenen indirekten Ansätze (Interlingua Methode, Transfer Methode) graphisch illustriert.

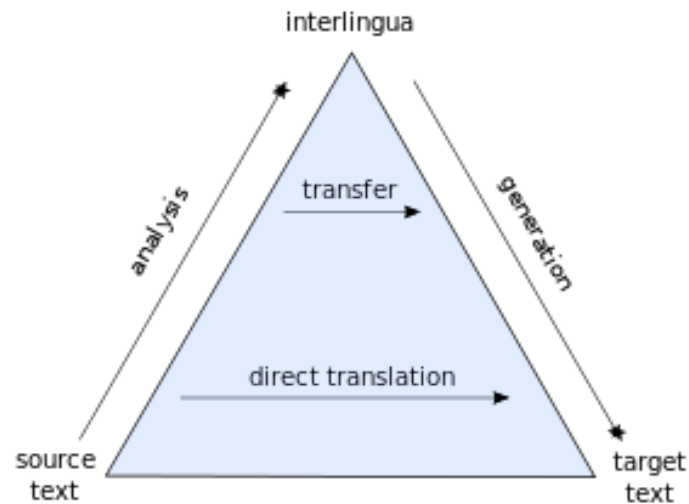


Abbildung 1: Dreieck nach Vauquois (Wikimedia Commons 2014)

Der indirekte Ansatz unterscheidet sich vom Direkten insofern als der Übersetzungsprozess mehrere Phasen umfasst. Da der direkte Ansatz unbefriedigende Resultate hervorbrachte, ging die Wissenschaft zu einer ambitionierteren Methode, der Interlingua Methode über, die sich in ihrer Komplexität als die höchste aller drei RBMT Methoden herausstellen sollte (vgl. Abb. 1).

2.2.1.1 Interlingua Methode

Stein (2009: 8) nennt die Interlingua Methode „[...] ein bis heute utopisches Ideal, das auf der Annahme beruht, es gäbe eine universelle und völlig sprachunabhängige Art der Kodierung von sprachlichen Informationen“. Der Übersetzungsprozess läuft in zwei Phasen ab. Zunächst wird der ausgangssprachliche Text analysiert und in semantischen Repräsentationen, der Interlingua, dargestellt, um im nächsten Schritt aus der Interlingua heraus einen zielsprachlichen Text zu generieren, ohne auf den ursprünglichen Ausgangstext zuzugreifen.

„The representation is thus a projection from the source text and at the same time acts as the basis for the generation of the target text; it is an abstract representation of the target text as well as a representation of the source text.“ (Hutchins 1992: 73)

Die Interlingua kann dabei verschiedene Formen annehmen wie z.B. eine natürliche Sprache, eine Plansprache wie Esperanto oder auch Repräsentationen, die sprachunabhängig und universell einsetzbar sind (vgl. Ramlow 2008: 39). Durch die Loslösung von der AS und ZS hat die Interlingua Methode den Vorteil, dass sie leicht um weitere Sprachen erweiterbar ist und beliebige Sprachen miteinander kombiniert werden können, da lediglich

zwei neue Module hinzugefügt werden müssen – ein AS Analyse Modul und ein ZS Synthese Modul. Ausgehend von einem Modell mit zwei Sprachen (Deutsch, Englisch) wird die Zahl der Übersetzungsmöglichkeiten von zwei (Deutsch-Englisch, Englisch-Deutsch), das Hinzufügen eines weiteren AS Analyse Moduls (z.B. Französisch) auf vier gesteigert (DE-EN, EN-DE, Französisch-Englisch, Französisch-Deutsch). Ein zusätzliches französisches ZS Synthese Modul würde zwei weitere Übersetzungsrichtungen (Englisch-Französisch, Deutsch-Französisch) ermöglichen (vgl. Hutchins 1992: 74).

Ein weiterer Vorteil der Interlingua Methode ist die Tatsache, dass sie sich auf die Schlüsselaussagen des AT konzentriert und somit leichter mit Disfluenzen oder umgangssprachlichen Ausdrücken umgehen kann. Darüber hinaus gibt es bei diesem Ansatz die Möglichkeit Paraphrasen aus der Interlingua zurück in die AS zu generieren, bevor in den ZT übersetzt wird, um den AnwenderInnen eine Feedback- und Kontrollfunktion zu bieten (vgl. Reichert & Waibel 2004: 1).

Die Interlingua Methode erscheint aufgrund der sprachunabhängigen Flexibilität vor allem für multilinguale Systeme sinnvoll. Jedoch bringt die Idee der semantischen Repräsentationen besonders auf der Ebene der Wortbedeutung Schwierigkeiten mit sich. So müsste das Interlingua-Vokabular für alle sprachlichen Ausdrücke und deren implizierte Konzepte Repräsentationen bereitstellen. Da es nicht immer Ein-zu-Eins-Übersetzungen von Wörtern und den dahinterliegenden Konzepten in unterschiedlichen Sprachkombinationen gibt, müssten die Repräsentationen auf ihrer Bedeutungsebene sehr detailliert sein, um richtige Übersetzungen generieren zu können. Ramlow (2008: 40) nennt hierbei als Beispiel den Unterschied zwischen den Konzepten von Fluss (dt.) und fleuve bzw. rivière (fr.).

Mit je mehr Sprachenpaaren gearbeitet wird, was auch als ein Vorteil der Methode angesehen wird, umso weniger umsetzbar erscheint die Realisierbarkeit dieser Idee. Zudem könnte die Forschung noch mit keiner vollkommen sprachunabhängigen Interlingua aufkommen (vgl. Hutchins 1992: 75; Stein 2009: 8), weshalb der Ansatz in dieser Form wohl auch weiterhin ein experimentelles Dasein fristen wird.

2.2.1.2 Transfer Methode

Die zweite Methode des indirekten Ansatzes trägt einen etwas irreführenden Namen, da so gesehen jedes MT System, unabhängig auf welcher Methode es basiert, eine mehr oder weniger transferähnliche Phase durchläuft. Der Begriff Transfer bezieht sich laut Hutchins (1992: 75) hierbei auf Modelle, die zwischen den AS und ZS Repräsentationen ein intermediäres bilinguales Modul aufweisen und somit in drei Phasen gegliedert sind. Zudem sind die Repräsentationen im Gegensatz zur Interlingua Methode sprachabhängig, d.h. das

Analyseergebnis stellt eine abstrakte Repräsentation des Ausgangstextes dar, während die Einspeisungen des Synthesemoduls abstrakte Repräsentationen des Zieltextes sind.

„Bei Transfer-Systemen geschieht die Abbildung zwischen den Sprachen nicht auf der Ebene der Lexeme, sondern der der sprachspezifischen syntaktischen und/oder semantischen Strukturen. Hierdurch können bestimmte Arten der Ambiguität aufgelöst werden (lexikalische A. wie *book* als Verb oder Nomen). Gleichzeitig werden Generalisierungen möglich, welche die MÜS erheblich transparenter werden lassen. [...] Dadurch, dass nur Merkmalsstrukturen übertragen werden, können sprachspezifische Formen (wie die Genitivform im Französischen) modular in der Grammatik angegeben werden. Für deren Generierung werden dann keine Transferregeln benötigt.“ (Carstensen 2012: 191)⁴

Verglichen mit dem direkten Ansatz arbeitet das transferbasierte System mit einer ausgefilterten morphologischen Analyse und bezieht zudem syntaktische und/oder semantische Strukturen mit ein, wodurch es von der Komplexität her höher einzustufen ist als die direkten Systeme (vgl. Abb. 1). Zunächst durchläuft der AT eine Analyse mittels einer Grammatik und eines ausgangssprachlichen Wörterbuchs, bevor ausgangssprachliche Repräsentationen des AT erstellt werden. Mittels zweisprachiger Transferregeln werden diese Repräsentationen in einer zweiten Phase zu zielsprachlichen Repräsentationen transformiert. Im letzten Schritt werden die zielsprachlichen Repräsentationen wiederum mittels einer Grammatik und eines zielsprachlichen Wörterbuches zu einem ZT generiert. Die Repräsentationen können dabei beispielsweise in Form von Phrasenstrukturbäumen⁵ auftreten (vgl. Ramlow 2008: 40f).

Im Vergleich zur Interlingua Methode ist die Transfermethode aufgrund des sprachenpaar- und sprachrichtungsabhängigen Ansatzes, anstelle des sehr ambitionierten universellen Ansatzes, weniger komplex. Gleichzeitig gestaltet sich das Hinzufügen einer zusätzlichen Sprache insofern schwieriger, als nicht nur ein Analyse- und Synthesemodul sondern auch entsprechende Transfermodule eingefügt werden müssten. So würde ein Transfermodell mit der Sprachrichtung DE-EN und EN-DE über je zwei Analyse, Transfer- und Synthesemodule verfügen. Bei der Erweiterung um die Sprache Französisch müsste das System, abgesehen von einem Analyse- und einem Synthesemodul, um vier weitere Transfermodule erweitert werden, nämlich DE-FR, FR-DE und EN-FR und FR-EN (vgl. Hutchins 1992: 75f).

Stein (2009: 8) merkt bezüglich der morphologischen, syntaktischen und semantischen Strukturregeln an, „[...] dass die Grenze nach oben offen zu sein scheint und sich zehntausende von Regeln und Kombinationen definieren lassen.“ Wie sich jedoch in der Praxis herausstellte, erreicht das System ein Plateau, an dem eine weitere Steigerung der Komplexität nicht mit einer Steigerung der Qualität einhergeht. Im Gegenteil, die Masse

⁴ MÜS = maschinelles Übersetzungssystem (vgl. Carstensen 2012)

⁵ Nähere Informationen und Beispiele für Phrasenstrukturbäumen bei Schütze & Zangenfeld (2013)

an Regeln weist einen gewissen Grad an Widersprüchlichkeit auf und generiert durch innere Konflikte nur neue Fehler. Aufgrund seiner Vor- und Nachteile wird der transferbasierte Ansatz bevorzugt bei Übersetzungen innerhalb eines Sprachpaares (in beide Richtungen), oder bei einseitigen Übersetzungen für eine begrenzte Zahl von Sprachen eingesetzt (vgl. Ramlow 2008: 42). Nach der Erläuterung der regelbasierten Ansätze, werden nun die statistikbasierten Ansätze vorgestellt.

2.2.2 Statistikbasierte Ansätze

Statistikbasierte Ansätze unterscheiden sich von den regelbasierten Ansätzen insofern, als sie empirisch belegt sind und statt auf linguistische Modelle auf Textkorpora zurückgreifen. Bis zu den 1990er Jahren war mit regelbasierten Ansätzen gearbeitet worden. Ein einschneidendes Ereignis war die Wiederaufnahme eines rein statistikbasierten Ansatzes beim Candide Projekt von IBM im Jahr 1988 (vgl. Brown et al. 1988). Ganz zu Beginn der MT Forschung wurde bereits mit statistischen Modellen gearbeitet, jedoch waren die Ergebnisse nicht befriedigend. Da sich in der Forschung zur Spracherkennung neuere stochastische Methoden bewährt hatten, begann das Forscherteam rund um Peter F. Brown bei IBM sich wieder dem statistischen Ansatz zu widmen. Das Besondere am Candide Projekt war, dass es, angelehnt an den statistischen Ansatz von Warren Weaver aus dem Jahre 1949, in der Analyse- und Synthesephase auf rein statistischen Methoden anstelle von linguistischen Regeln basierte. Dabei wurde das System anhand eines großen Korpus französischer und englischer Texte, die aus Berichten kanadischer Parlamentsdebatten stammten, bei der Second TMI Conference vorgestellt und überzeugte selbst die Anhänger des regelbasierten Ansatzes. Beinahe die Hälfte der übersetzten Phrasen stimmten mit den Übersetzung des Korpus überein, drückten den Sinn mit ähnlichen Wörtern aus, oder boten ebenso akzeptable Übersetzungen (vgl. Hutchins 2014b: 12f).

Die grundlegende Idee hinter dem Ansatz ist es durch die Verwendung eines großen Paralleltextkorpus Wahrscheinlichkeiten für äquivalente ausgangssprachliche und zielsprachliche Sätze zu erhalten und aufgrund dieser, Übersetzungsentscheidungen zu treffen. Browns Methode (vgl. Brown et al. 1990: 79-82) arbeitet auf Grundlage von Sätzen als Übersetzungssegmente und sieht dabei vor, dass ein Satz T (target) in der Zielsprache eine mögliche Übersetzung von der Satz S (source) in der Ausgangssprache sein kann. Jedoch sind manche Sätze wahrscheinlicher als andere, sodass dem Sprachenpaar (S,T) eine Wahrscheinlichkeit Pr (probability) zugeordnet wird, welche die Wahrscheinlichkeit ausdrücken soll mit der ein/e Translator/in Satz T produzieren wird, wenn er oder sie Satz S vorgelegt bekommt. Demnach steht $Pr(T|S)$ für die Wahrscheinlichkeit, dass T eine Übersetzung von S ist. Beim Satzpaar A wird die Wahrscheinlichkeit erwartungsgemäß sehr niedrig sein, während es für das Satzpaar B sehr hoch sein sollte:

Satzpaar A

„Le matin je me brosse les dents.“

„President Lincoln was a good lawyer.“

Satzpaar B

„Le président Lincoln était un bon avocat.“

„President Lincoln was a good lawyer.“

(Brown et al. 1990: 72)

In anderen Worten geht es im Übersetzungsprozess darum, den besten Wert für $\Pr(S|T)$ zu finden, indem ausgehend von Satz T, der dazugehörige Satz S gesucht wird, da er dem/der Übersetzer/in als Ausgangstext gedient hatte, und der gefundene Satz S wiederum Satz T zugeordnet wird.

„We view the problem of machine translation then as follows. Given a sentence T in the target language, we seek the sentence S from which the translator produced T. We know that our chance of error is minimized by choosing that sentence S that is most probable given T. Thus, we wish to choose S so as to maximize $\Pr(S|T)$.“ (Brown et al. 1990: 72)

Um dieses Bestreben zu realisieren, müssten alle beliebigen Sätze einer bestimmten Sprache zur Verfügung stehen, was natürlich eine Unmöglichkeit darstellt. Aus diesem Grund arbeiten statistische Ansätze mit Modellen bzw. Näherungen. Die Berechnung der Wahrscheinlichkeit beruht hierbei auf zwei Modellen. Bei Brown et al. handelt es sich um ein Übersetzungsmodell (translation model) und um ein Sprachmodell (language model) (vgl. Brown et al. 1990). Der zu Beginn erwähnte zweisprachige Korpus ist aliniert und stellt das Übersetzungsmodell dar, welches alle denkbaren und undenkbaren Übersetzungen zwischen der AS und der ZS vereint. Mit der Berechnung der Wahrscheinlichkeiten werden jene AS und ZS Sätze mit der höchsten Wahrscheinlichkeit einander zugewiesen. Zusätzlich zum Übersetzungsmodell beinhaltet das Sprachmodell ein einsprachiges Korpus in der ZS, mit allen in der ZS möglichen Sätzen. In dieser Phase ist die Zahl der möglichen Kombinationen immer noch zu hoch, sodass das Sprachmodell abstrahiert werden muss, um weiter mit Wörtern bzw. Wortsequenzen zu arbeiten. Das Übersetzungsmodell wird auch in ein Lexikonmodell und ein Alinierungsmodell eingeteilt, um einerseits die höchste Wahrscheinlichkeit der richtigen Übersetzungen für die Wortsequenzen zu berechnen und andererseits die höchste Wahrscheinlichkeit für die korrekte Satzstellung zu finden. Mittels eines Suchalgorithmus wird nun jener Satz gesucht „[...] dessen Produkt der Werte von Satzgültigkeit (Sprachmodell), Wortübersetzung (Lexikonmodell) und Satzstellung (Alinierungsmodell) am höchsten ist. Das Ergebnis ist die wahrscheinlichste Übersetzung“. (Stein 2009: 10) Um die Wahrscheinlichkeiten berechnen zu können nutzt der Computer einen Algorithmus, dazu wird üblicherweise der Satz von Bayes verwendet:

$$\Pr(S|T) = \frac{\Pr(S) * \Pr(T|S)}{\Pr(T)}$$

(Brown et al. 1990: 79)

Wie sich ein SST System mit statistikbasiertem Ansatz (a) von einem System mit Interlingua Ansatz (b) unterscheidet, wird in Abbildung 2 graphisch dargestellt.

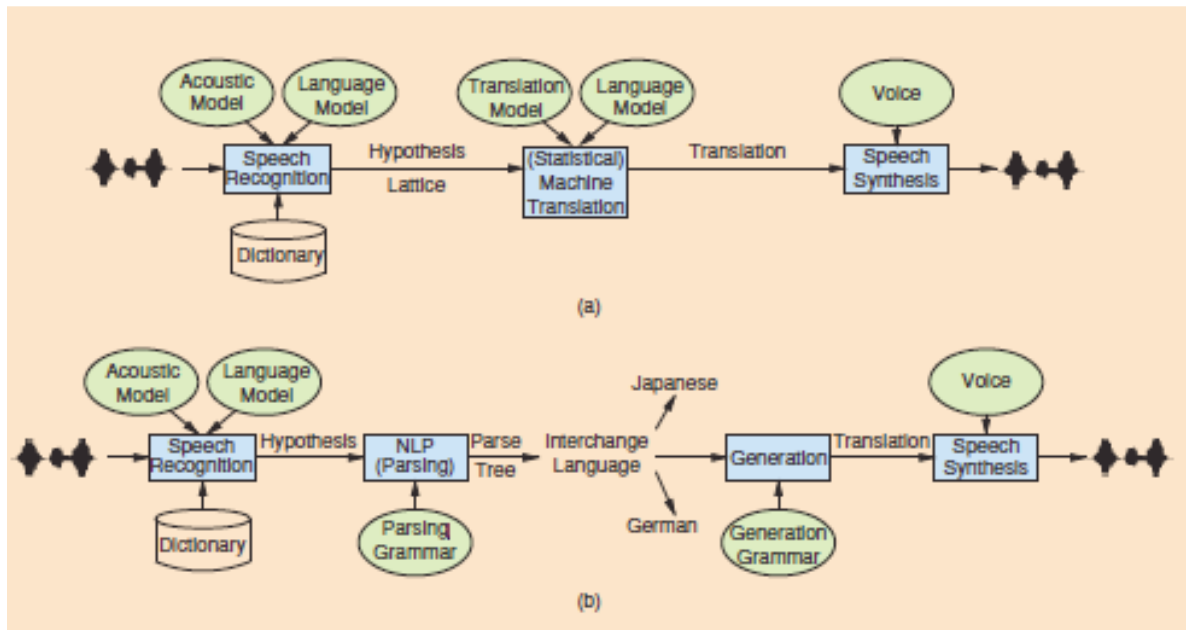


Abbildung 2: Graphische Übersicht über SST Systeme mit einem statistischen Ansatz (a) und einem Interlingua Ansatz (b) (Waibel & Fügen 2008: 71)

Die Analyse von ganzen Sätzen macht in der Realität wenig Sinn, da es sehr selten ist, dass die verwendeten Korpora den vollständigen Satz in den gewünschten Sprachen beinhalten. Solange das System nicht mit einem Korpus arbeitet der tatsächlich alle möglichen Sätze einer Sprache aufweist, ist es zielführender die zu analysierenden Segmente zu verkleinern (vgl. Stein 2009: 11). So lassen sich die statistikbasierten Systeme in wortbasierte und phrasenbasierte Systeme unterteilen. Diese sollen nun im Folgenden vorgestellt werden.

2.2.2.1 Wortbasierte MÜ

Wie bereits erwähnt orientierte sich die statistikbasierte maschinelle Übersetzung von ganzen Sätzen auf die Analyse von auf Wortebene um. In anderen Worten soll ein Wort aus der Ausgangsprache einem Wort in der Zielsprache entsprechen, diese Wort-zu-Wort Ent-

sprechungen werden Alignment genannt (vgl. Och et al. 2001). Diese Verkleinerung von der Satzebene auf die Wortebene birgt jedoch eine andere Problematik. So kann es passieren, dass ein einzelnes Wort aus Sprache A in Sprache B nur durch mehrere Wörter ausgedrückt werden kann. Stein (2009: 11f) illustriert dies an einem Beispiel aus dem Sprachenpaar Englisch-Spanisch. Das englische *slap* entspricht der spanischen Phrase *dar una bofetada*. Mit einem wortbasierten Übersetzungssystem lässt sich dieses Beispiel nur in eine Richtung übersetzen (EN-SPA). Die Konzentration auf die Wortebene macht es dem System unmöglich, mehrere Wörter in der AS in ein Wort in der ZS zu übersetzen, was in diesem Fall die Übersetzungsrichtung SPA-EN problematisch gestaltet. Ähnliche Probleme finden sich bei zusammengehörenden Wörtern, die aufgrund der Syntaxunterschiede zwischen verschiedenen Sprachen nicht als solche erkannt und übersetzt werden können. So können die einzelnen Elemente von zweiteiligen Verben wie z.B. abreisen durch die getrennte Betrachtung des Präfixes und des finiten Verbs große Bedeutungsunterschiede haben (vgl. Stein 2009: 12).

2.2.2.2 Phrasenbasierte MÜ

Wie bereits gezeigt, bringen sowohl Übersetzungssysteme die auf Satzebene arbeiten, als auch jene auf Wortebene bestimmte Komplikationen mit sich. Um diese in den Griff zu bekommen wurde ein Mittelweg gesucht – der phrasenbasierte Ansatz. Laut Stein (2009: 12) ist dies der weitverbreitetste Ansatz bei den heutigen Systemen. Dabei werden die Texte maschinell nach ihrer Größe in unterschiedliche Gruppen unterteilt und entsprechen Wortsequenzen, was die Möglichkeit eröffnet, das oben genannte Problem der Übersetzung mehrerer Wörter mit einem einzelnen Wort zu lösen. Zweiter Vorteil dieses Ansatzes ist die bessere Unterscheidung von Ambiguitäten aufgrund des miteinbezogenen Kontextes. Im Vergleich dazu hätte ein wortbasiertes System Probleme bei der Unterscheidung des Wortes *pretty* in Ausdrücken wie *pretty much* oder *pretty girl*. Durch die Kategorisierung in unterschiedliche Gruppen von Wortsequenzen können eben diese Probleme umgangen werden (vgl. Stein 2009: 12).

2.2.3 Weitere Ansätze

Neben den regelbasierten und statistikbasierten Ansätzen haben sich noch andere Ansätze entwickelt, die grob gesehen in keine der beiden Kategorien fallen. Diese sollen im Folgenden kurz angeschnitten und in ihren Grundzügen erklärt werden.

2.2.3.1 Beispielbasierter Ansatz

Der beispielbasierte Ansatz (Example-Based Machine Translation, EBMT) (vgl. Somers 2003) arbeitet genauso wie der statistische Ansatz auf einem Korpus von Paralleltexten. Während statistikbasierte Systeme ihre Analyse mit einem möglichst großen Korpus durchführen, um mittels eines Algorithmus die Wahrscheinlichkeit der Übersetzung zu berechnen, basiert das EBMT System auf dem Analogieprinzip und vergleicht die Teile des AT mit einem vergleichsweise kleinerem Korpus. Das in Abb. 1 illustrierte Dreieck von Vauquois würde, auf den EBMT Ansatz umgelegt, demnach wie folgt aussehen:

„The source-text analysis in conventional MT is replaced by the matching of the input against the example set [...]. Once the relevant example or examples have been selected, the corresponding fragments in the target text must be selected. This has been termed alignment or adaptation and, like transfer in conventional MT, involves contrastive comparison of both languages [...]. Once the appropriate fragments have been selected, they must be combined to form a legal target text, just as the generation stage of conventional MT puts the finishing touches to the output. The parallel with conventional MT is reinforced by the fact that both the matching and recombination stages can, in some implementations, use techniques very similar to [...] analysis and generation in conventional MT.“ (Somers 2003: 87f).

Ähnlich wie in der Analyse-, Transfer- und Synthesephase der traditionellen MT gleicht das EBMT System kontrastiv passende Segmente mit einander ab und versucht diese dann in zusammenhängende Sätze zu kombinieren. Durch diese Herangehensweise wird der EBMT Ansatz häufig mit kommerziellen Translation Memory Systemen verglichen, was nur zum Teil korrekt ist, da TM-Systeme als Unterstützungstool für menschliche ÜbersetzerInnen gedacht sind und EBMT Systeme rein automatischen Charakter haben (vgl. Somers 2003: 85).

2.2.3.2 Kontextbasierter Ansatz

Der kontextbasierte Ansatz (Context-Based Machine Translation, CBMT) ist eine relativ neue Herangehensweise und bedarf ebenso wie der statistische Ansatz und EBMT Korpora, um Übersetzungen anfertigen zu können. Jedoch brauchen die bisherigen MT-Ansätze entweder Unmengen an Regeln oder sehr große, vorübersetzte Parallelkorpora. Ersteres ist extrem zeitintensiv, wenn eine akzeptable Qualität erreicht werden soll und Zweiteres gestaltet sich auch bei gängigen Sprachpaaren nicht gerade einfach. Für Sprachenpaare, die seltener übersetzt werden, gibt es kaum ausreichende Mengen an Parallelkorpora. Um diese Probleme zu umgehen, stellt CBMT ein Mittel dar, welches weder auf Regeln noch auf Parallelkorpora angewiesen ist (vgl. Carbonell et al 2006: 19):

„[...] CBMT requires an extensive monolingual targettext corpus (from 50 gigabytes to 1 terabyte), a full-form bilingual dictionary, and optionally (to further improve translation quality) a smaller monolingual source-text corpus. Monolingual text can be acquired in quantities by crawling the Web, extracting pure text from HTML, and indexing such text in the manner required by the CBMT methods. The most time-consuming resource to build or acquire is the large-scale full-form bilingual dictionary for each language pair.“ (Carbonell et al 2006: 19)

CBMT Systeme beruhen demnach auf großen einsprachigen Korpora in der ZS und der Übersetzungsprozess läuft hierbei über ein zweisprachiges Vollformenwörterbuch. Eben-dieses analysiert passende Übersetzungsmöglichkeiten und gleicht sie in einem nächsten Schritt mithilfe von N-Grammen mit dem zielsprachlichen Korpus ab, um die richtigen von den falschen Übersetzungen zu trennen. Ein Decoder wählt dann jene Übersetzung als die beste aus, welche die längste N-Grammlänge und maximalen Overlap⁶ hat (vgl. Carbonell et al. 2006: 19ff).

2.2.3.3 Wissensbasierter Ansatz

Unabhängig davon ob es sich um ein regelbasiertes, statistikbasiertes oder sonstiges maschinelles Übersetzungssystem handelt, haben alle bisher Vorgestellten eine Ebene des menschlichen Wissens mehr oder weniger vernachlässigt – das Weltwissen. Weltwissen ist während des Übersetzens unerlässlich, um Ambiguitätsprobleme zu lösen, die nicht durch linguistische oder statistische Mittel gelöst werden können. Stein zeigt dies am folgenden Beispiel auf „Das Schloss liegt auf dem Berg/Tisch“ (Stein 2009: 14) und erläutert, dass es nicht möglich ist diesen Satz korrekt zu übersetzen, wenn nicht klar ersichtlich ist, von welcher Art Schloss gesprochen wird. Zu diesem Zweck wird beim wissensbasierten Ansatz (Knowledge-Based Machine Translation, KBMT) Wissen in metasprachlicher Form in Datenbanken organisiert und ähnelt daher stark dem regelbasierten Interlingua Ansatz (vgl. Nirenburg et al. 1986). Aufgrund der Komplexität der Datenbanken und ihrer Entwicklung fokussieren sich wissensbasierte Systeme auf stark beschränkte Anwendungsbereiche (vgl. Stein 2009: 14).

2.2.3.4 Hybride Formen

Hybride Ansätze der maschinellen Übersetzung sind, wie der Name bereits vermuten lässt, Mischformen der oben beschriebenen Ansätze, in denen versucht wird die Vorteile der einzelnen Herangehensweisen auszunutzen und miteinander zu kombinieren. So können je nach Bedarf unterschiedliche oben genannte Problemfelder, wie z.B. unterschiedliche Syn-

⁶ Für genauere Erklärungen und Illustrationen zu N-Grammen und Overlap empfiehlt sich die Lektüre von Carbonell et al. (2006).

taxkonstruktionen von verschiedenen Sprachen oder Mangel an Parallelkorpora von bestimmten Sprachkombinationen, umgangen werden.

Das bisher wohl bekannteste Beispiel für regelbasierte maschinelle Übersetzungen SYSTRAN, welches beispielsweise in der Europäischen Kommission genutzt wird, hat sich umorientiert und arbeitet nun mit einem hybriden System. Mit dem Launch von SYSTRAN Enterprise Servers 7 kommt erstmalig eine Hybridform der maschinellen Übersetzung für Unternehmen auf den Markt:

„With the introduction of its hybrid MT product, SYSTRAN has delivered a solution that brings together the strengths of these two approaches. SYSTRAN Enterprise Server 7 combines the predictability and language consistency of rule-based machine translation with the fluency of statistical MT.“ (SYSTRAN 2009)

Desweiteren wird damit geworben, dass das Programm durch den neuen hybriden Ansatz lernfähig gemacht wurde, indem es basierend auf Humanübersetzungen domänenspezifisch lernt wie Wörter und Phrasen übersetzt werden sollten. Durch die immer wiederkehrende Benutzung der Daten bzw. Übersetzungen hat das System einen Lerneffekt und updated sich selbst. (vgl. SYSTRAN 2009).

2.3 Sprachsynthese

Die oben beschriebenen maschinellen Übersetzungssysteme generieren den Zieltext in erster Linie in Textform. Nun bedarf ein SST System einer weiteren Komponente, die den Text in einer künstlichen, dem Menschen nachgeahmten Stimme wiedergeben soll. Diese Aufgabe übernimmt die Sprachsynthese. Dabei werden zwei verschiedene Arten von Systemen je nach Art des Inputs unterschieden (vgl. Pfister & Kaufmann 2008: 27):

Die Text-to-Speech Synthese (text-to-speech synthesis, TTS) verarbeitet orthographische Texte, was bedeutet, dass „[...] die Synthese nebst der eigentlichen Sprachsignalproduktion auch die Aussprache der Wörter, die Akzentstärke der Silben uvm. Bestimmen muss“. (Pfister & Kaufmann 2008: 27) Um dies zu erreichen muss der Text von der Sprachsynthese Komponente linguistisch analysiert werden.

Das zweite System arbeitet mit hierarchisch strukturiertem Input wie z.B. Syntaxbäumen und wird Concept-to-Speech Synthese (concept-to-speech synthesis, CTS) genannt. Diese kommt bei Systemen zum Einsatz, die über eine Komponente mit linguistischem und pragmatischem Wissen verfügen, welche eigenständig Äußerungen generiert (vgl. Pfister & Kaufmann 2008: 27).

Zusätzlich gibt es noch die Sprachausgabe, die jedoch von den beiden genannten Arten der Sprachsynthese zu unterscheiden ist. Denn hierbei beruhen die ausgegebenen Sprachsignale in erster Linie auf Aufzeichnungen von menschlichen SprecherInnen. Somit ist die Bandbreite der Sprachausgabemöglichkeiten eingeschränkt, während bei der Sprachsynthese jeder Text in ein Sprachsignal umgewandelt werden kann (vgl. Pfister & Kaufmann: 27).

2.3.1 Aufgabe der Sprachsynthese

Die Sprachsynthese soll, wie bereits erklärt, geschriebene Sprache in gesprochene Sprache umwandeln. Damit ist sie von ihren Aufgaben her mit einer Person, die vorliest, vergleichbar. Ob die vorlesende Person die Sprache beherrscht und vor allem in erster Linie versteht, ist grundsätzlich an zwei unterschiedlichen Fehlerarten erkennbar (vgl. Pfister & Kaufmann 2008: 194):

Falsche Aussprache von Wörtern: In dem Fall, dass die vorlesende Person ein oder mehrere Wörter nicht kennt, wird er oder sie versuchen, diese auf Basis des Erfahrungsschatzes und Rückschlüssen auf ähnliche Wörter auf eine bestimmte Weise auszusprechen, wobei es zu Fehlern kommen kann. „Solche Fehler äußern sich häufig in einer falschen Lautfolge, beziehen sich also auf die *segmentale* Ebene. In mehrsilbigen Wörtern kann auch die Akzentuierung der Silben falsch sein.“ (vgl. Pfister & Kaufmann 2008: 194) [Hervorhebung im Original]

Unpassende Satzakzente und abweichende Gruppierung: Jeder Sprecher und jede Sprecherin akzentuiert die einzelnen Wörter in einem Satz so, dass daraus auf ihre Relevanz geschlossen werden kann. Hierbei gibt es von Sprecher zu Sprecher unterschiedliche Herangehensweisen und allgemein einen vergleichsweise großen Entscheidungsspielraum.

„Werden die Grenzen dieses Spielraums überschritten, dann resultiert eine nicht sinnkonforme oder sogar sinnwidrige Akzentuierung, die [...] verrät, dass der Sprecher den Sinn des Textes nicht verstanden hat. Bei längeren, komplizierten Sätzen ist zudem eine Gruppierung der Wörter für den Zuhörer hilfreich, wobei die Sprechgruppengrenzen je nach Stärke durch Pausen, durch Verzögerung und durch Variieren der Tonhöhe markiert werden können.“ (Pfister & Kaufmann 2008: 194)

Ein unpassendes oder gar sinnwidriges Setzen von Satzakzenten erschwert nicht nur das Verständnis der ZuhörerInnen, sondern lässt die SprecherInnen unnatürlich wirken. Dies ist weder für humane SprecherInnen noch für das maschinelle Pendant erstrebenswert. Aufgrund der Tatsache, dass sich die Gruppierung und die Akzentuierung nicht auf einzelne Laute beziehen, sondern viel mehr auf längere Segmente wie Silben oder gar größer, sind sie laut Pfister & Kaufmann (2008: 194) in die suprasegmentale Ebene zu kategorisieren.

Die Sprachsynthesysteme sind üblicherweise in je ein (oder mehrere) stimmabhängige und stimmunabhängige Systeme geteilt. Zu Demonstrationszwecken soll im Folgenden ein System mit einer solchen Zweiteilung ausreichend sein. Die Erklärung für die Zweiteilung liegt im wesentlichen Unterschied von gesprochener und geschriebener Sprache. Während gesprochene Sprache in jedem Fall einer konkreten Stimme bedarf, ist geschriebene Sprache vollkommen von der Stimme losgelöst (vgl. Pfister & Kaufmann 2008: 196). Abbildung 3 zeigt die graphische Darstellung eines solchen zweigeteilten Systems.

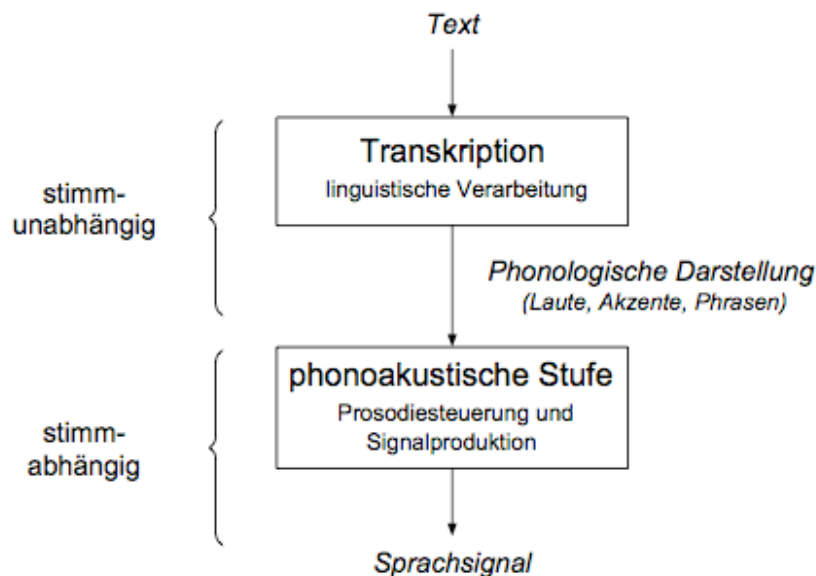


Abbildung 3: Sprachsynthesesystem mit stimmabhängigen und stimmunabhängigen Teilen (Pfister & Kaufmann 2008: 197)

Die stimmunabhängige Ebene ist die Transkriptionsstufe (vgl. Pfister & Kaufmann 2008: 196f). In dieser Stufe wird die Aussprache der Wörter festgelegt; es wird also die Lautabfolge bei der Signalproduktion ermittelt.

„Die Aussprache ist in der Regel für jedes Wort einzeln bestimmbar. Im Gegensatz dazu können die Wichtigkeit der Wörter (d.h. die Verteilung der Akzente in einem Satz) und die Unterteilung des Satzes in Sprechgruppen (sog. Phrasen) nur aufgrund von Information über die Beziehung der Wörter zueinander bestimmt werden. So kann beispielsweise aus dem syntaktischen Aufbau eines Satzes abgeleitet werden, wo Phrasengrenzen zu setzen sind.“ (Pfister & Kaufmann 2008: 197f)

Der Transkriptionsteil stellt daher die linguistische Verarbeitung des Textes mittels Wort- und Satzanalyse dar und produziert eine abstrakte, noch stimmunabhängige Darstellung des zu produzierenden Sprachsignals. Dieses abstrakte Resultat wird phonologische Dar-

stellung⁷ genannt und impliziert die „Lautfolge, die Akzente, die Phrasengrenzen und die Phrasentypen“. (Pfister & Kaufmann 2008: 198)

Die zweite Komponente der Sprachsynthese, die phonoakustische Stufe (vgl. Pfister & Kaufmann 2008: 199f), bewegt sich auf der stimmabhängigen Ebene. Ihre Aufgabe ist es die stimmunabhängige, phonologische Darstellung in ein konkretes Sprachsignal zu verarbeiten. Dazu bedarf es zweier Schritte. Zunächst definiert die sogenannte Prosodiesteuerung aus der abstrakten Darstellung prosodische Parameter wie die Intensität, Grundfrequenz und Lautdauer eines jeden Lautes. In einem nächsten Schritt erzeugt dann die sogenannte Sprachsignalproduktion die Laute „[...] in der durch phonologische Darstellung gegebenen Reihenfolge und mit dem von der Prosodiesteuerung bestimmten prosodischen Parametern [...]“. (Pfister & Kaufmann 2008: 200)

Zusammengefasst produziert die stimmunabhängige Transkriptionsstufe durch linguistische Verarbeitung eine abstrakte, phonologische Darstellung des Textes, welche hauptsächlich aus Informationen zu Akzentuierung, Phrasengrenzen und Lauten besteht und im nächsten Schritt in der stimmabhängigen, phonoakustischen Stufe - die Informationen zu Länge oder Tonhöhe der Laute, sowie die Wahl einer Frauen- oder Männerstimme enthält - zu einem Sprachsignal verarbeitet wird.

Zur vertiefenden Lektüre rund um das Thema Text-to-Speech Synthese bietet sich das umfassende Werk von Paul Taylor (2009) an, während das Werk von Pfister & Kaufmann (2008) auch andere Bereiche der Sprachverarbeitung wie z.B. Spracherkennung in Lehrbuchcharakter beleuchtet.

⁷ Anschauliche Beispiele für die Notation der phonologischen Darstellung, die auf den IPA-Lautschriftzeichen basiert (vgl. Pfister & Kaufmann 2008: 198) finden sich in eben genanntem Werk.

3 Pionierprojekte der SST

Nach der theoretischen Einführung in die technische Seite der SST Systeme sollen in diesem Kapitel zwei wesentliche Forschungsprojekte in diesem Bereich vorgestellt werden. Zunächst eines der Pionierprojekte der SST: Verbmobil, ein umfangreiches und langjähriges Projekt der 1990er Jahre und anschließend ein Neuere aus dem Jahr 2005 – der Lecture Translator. Dieses Kapitel stellt eine stark verkürzte Einführung in beide Projekte dar und konzentriert sich auf die aus übersetzungswissenschaftlicher Sicht relevanten Aspekte der Projekte. Das Pionierprojekt Verbmobil wurde ausgewählt, weil einige namhafte Mitarbeiter der Forschungsgruppe rund um Verbmobil heute in die Forschung von SST Apps involviert sind. Der Lecture Translator soll zeigen, wie weit sich die Forschung im Laufe der Jahre weiterentwickelt hat und was von der SST in Zukunft noch erwartet werden kann.

3.1 Verbmobil

„*Verbmobil: Multilinguale Verarbeitung von Spontansprache* (engl. *Verbmobil: Multilingual Processing of Spontaneous Speech*)“ war ein achtjähriges, interdisziplinäres Leitprojekt in der Sprachtechnologieforschung (vgl. Wahlster 1997: 215). Wie bereits aus dem Namen hervorgeht, handelte es sich dabei um ein „speaker-independent and bidirectional speech-to-speech translation system for spontaneous dialogs in mobile situations“ (Wahlster 2000: 3), welches die verbale Kommunikation zwischen zwei Gesprächspartnern unterstützt. Das Ziel des Projektes war es, ein mobiles maschinelles Übersetzungssystem zu schaffen, das gesprochene Spontansprache erkennt, die Eingabe analysiert, diese in die Fremdsprache übersetzt, einen Satz in der Fremdsprache generiert und diesen auch ausspricht. Die Anwendungsbereiche des Systems waren geschäftsorientiert und beschränkten sich auf Terminvereinbarung, Reiseplanung und Fernwartung von Computern (vgl. Wahlster 1997: 215).

Das in zwei Phasen (Laufzeit Phase 1: 01.01.1993–31.12.1996; Laufzeit Phase 2: 01.01.1997–30.09.2000) aufgeteilte Projekt wurde vom deutschen Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) und industriellen Investoren mit insgesamt 168,6 Mio. DM gefördert. Projektträger war das Deutsche Zentrum für Luft und Raumfahrt (DLR). In der ersten Phase lag der Fokus auf der Verarbeitung von Terminverhandlungsdialogen zwischen deutschen und japanischen Geschäftspartnern mit Englisch als Dialogsprache (vgl. Karger & Wahlster 2000: 23ff).

„Die Eingabesprachen sind Deutsch und Japanisch, wobei Englisch als Dialogsprache dient. Dabei geht man davon aus, daß beide Gesprächspartner Englisch zumindest passiv beherrschen und große Teile des Dialogs sowieso in Englisch sind. Nur wenn einer der beiden Gesprächspartner seinen Beitrag nicht auf Englisch machen kann, hat er die Möglichkeit sich in seiner Muttersprache (Deutsch oder Japanisch) an das Verbmobil-System zu wenden. Das System übersetzt dann den Beitrag ins Englische. Dieses "Dialogsprachmodell" hat den Vorteil, daß die Benutzer des Systems die Übersetzungsperformanz besser kontrollieren können als wenn sie die Zielsprache überhaupt nicht beherrschen.“ (Schmitz 1998:57)

Die „robuste und bidirektionale Übersetzung spontansprachlicher Dialoge“ (Wahlster 1997: 215) in den Bereichen Reiseplanung und Hotelreservierung für die Sprachenpaare Deutsch-Englisch (ca. 10.000 Wörter) und Deutsch-Japanisch (ca. 2.500 Wörter) stand im Mittelpunkt der zweiten Phase. An der ersten Phase waren pro Jahr 125 Personen beteiligt, die zu je 33 Forschungsgruppen von 28 Universitäten, Forschungszentren und Unternehmen an 16 unterschiedlichen Subprojekten und 135 Arbeitspaketen arbeiteten. In Phase zwei kam es zu einer Umstrukturierung⁸, sodass 8 Subprojekte mit 127 Arbeitspaketen von 24 Forschungsgruppen, die jährlich aus 100 Personen von 23 unterschiedlichen Partnern zusammengesetzt waren, bearbeitet wurden (vgl. Karger & Wahlster 2000: 24f). Verbmobil war laut Karger mit insgesamt 369 direkt beteiligten WissenschaftlerInnen und zusätzlichen 919 ForscherInnen, darunter Master Studierenden (238), Studentischen Hilfskräften (483), PraktikantInnen (18) und DoktorantInnen (164) (vgl. Karger & Wahlster 2000: 29), das „weltweit größte und erfolgreichste Leitprojekt seiner Art“. (Karger 2010).

Während der gesamten Projektlaufzeit lag die Gesamtprojektleitung beim Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI GmbH). Das Projekt endete mit einem Abschluss-symposium am 30.07.2000 im Rahmen dessen die Projektergebnisse präsentiert und der Forschungsprototyp analysiert wurde (vgl. Karger 2010). Bereits zu jenem Zeitpunkt hatten an Verbmobil beteiligte Unternehmen 18 Produkte auf den Markt gebracht, die auf den Erkenntnissen des Verbmobil Projektes basierten. Außerdem waren sechs Spin-off-Unternehmen von ehemaligen Verbmobil WissenschaftlerInnen gegründet worden (vgl. Wahlster 2000: 20).

„The sharable language resources collected and distributed during the Verbmobil project will be usefull beyond the project lifespan, since these transliterated and richly annotated corpora of spontaneously spoken dialogs can be used for building, improving or evaluating natural language and speech algorithms or systems in coming years.“ (Wahlster 2000: 20)

Wahlsters Vorhersage kann seit der Konferenz „10 Jahre Verbmobil - Rückblick und Ausblick - Multilinguale Verarbeitung von Spontansprache“ (Wahlster 2010) am 15. und 16.11.2010 als bestätigt erachtet werden, da die Verbmobil Ergebnisse inzwischen als Fundament für zahlreiche Produktinnovationen im Bereich der Sprachtechnologie, wie z.B.

⁸ Eine genauere Aufschlüsselung der Subprojekte und vor allem der organisatorischen Struktur des Verbmobil Projekts findet sich in Karger & Wahlster 2000.

in der Entwicklung von Smartphone-Apps gedient hatten (vgl. Karger 2010, Wahlster 2010).

3.1.1 Von Verbmobil zur iPhone-App

Wie bereits erwähnt war Verbmobil eine Software, die es NutzerInnen von Mobiltelefonen ermöglichte, sich mithilfe einer simultanen Dolmetschung mit ihren Gesprächspartnern zu verständigen. Die Funktionalität des Systems war auf drei Themenschwerpunkte beschränkt: Terminverhandlung, Reiseplanung und Fernwartung (vgl. Wahlster 1997).

Die folgende Graphik illustriert die einzelnen Schritte des Verbindungsaufbaus zum Gesprächspartner mit dem zwischengeschalteten Verbmobilserver.



Abbildung 4: Verbindungsaufbau einer Verbmobil-Sitzung (Wahlster 2000: 5)

Betont wurde, dass es sich bei Verbmobil um das erste "speech-only translation system" (Wahlster 2000: 4) handelte. Dies bedeutet, dass während der gesamten Gesprächszeit, von der Initiierung bis zur tatsächlichen Gesprächsführung, keine Tastenwahl notwendig ist, da das System auf Sprachbefehle reagiert. VerbmobilnutzerInnen brauchten daher lediglich ein Mobiltelefon, um mithilfe des sogenannten voice dialings den Verbindungsaufbau zum Server zu starten (siehe Abbildung 4). Dies setzt natürlich eine vorangegangene Speicherung der Telefonnummer des Verbmobilservers voraus. Da der Kontakt zum Server mittels GSM-Verbindung aufgebaut wurde, konnte Verbmobil unabhängig von Aufenthaltsort und Uhrzeit genutzt werden.

„No PC, notebook or PDA must be available to access the Verbmobil translation service, just a phone for each dialog participant. In addition, no waiting time for booting computers and keyboard or mouse input to start the Verbmobil system is needed – dialog translation can begin instantaneously.“ (Wahlster 2000: 4)

Nach erfolgreichem Verbindungsaufbau wird der Nutzer oder die Nutzerin begrüßt und dazu aufgefordert die Telefonnummer des gewünschten Gesprächspartners aufzusagen. Daraufhin wird die Nummer des Gesprächspartners gewählt und beide Seiten in den entsprechenden Sprachen mit dem Satz „*Verbmobil hat eine neue Verbindung aufgebaut. Bitte sprechen Sie jetzt.*“, begrüßt. Die Literaturrecherche ergab keine befriedigende Antwort auf die Frage, wie das Verbmobilsystem die benötigte Sprachkombination der Gesprächspartner erkennt. Da das System vom Deutschen aus in zwei Zielsprachen übersetzt, muss irgendwann entschieden werden, in welcher Sprache gearbeitet wird. Laut Graphik wird die anrufende Person an keinem Punkt des Verbindungsaufbaus nach der benötigten Zielsprache gefragt. Unter der Annahme, dass die jeweiligen GesprächspartnerInnen eine britische/amerikanische oder japanische Telefonnummer haben, könnte die Sprache auf Basis der Vorwahl erkannt werden. Bei der Nutzung einer anderen ausländischen Telefonnummer kann die Zielsprache nicht eindeutig eruiert werden. Jekat (2004: 11) erwähnt in diesem Zusammenhang die “score-based LID (Language Identification)”, bei der die Übereinstimmung zwischen akustischem Signal und der im Sprachmodell gespeicherten Muster berechnet wird, wobei die beste Übereinstimmung die Sprache bestimmt. Die Sprache einer Äußerung soll innerhalb von drei Sekunden mit einer Genauigkeit von 87 % bestimmt werden (vgl. Jekat 2004: 11). Dennoch erklärt dies nicht, wie das System bestimmt in welche Sprache der allererste deutsche Satz übersetzt werden soll, da der andere Gesprächspartner oder -partnerin noch nicht gesprochen hat.

Im konkreten Gespräch spricht Person A beispielsweise auf Deutsch und Person B auf Englisch. Dennoch kommunizieren die einzelnen GesprächspartnerInnen in ihrer Ausgangssprache, da Verbmobil die gesprochene deutsche Eingabe ins Englische übersetzt und umgekehrt. Obwohl das Hauptziel des Systems zunächst darin bestand die Face-to-Face Konversation zu unterstützen (vgl. Wahlster 2000: 4), indem alle Beteiligten das Handy als Dolmetsch-Apparat nutzen, ermöglichte die Serverlösung auch eine Art des maschinellen Remote Interpretings. Die heutigen Apps können demnach als eine technologische Weiterentwicklung in Anlehnung an diese ursprünglichen Verbmobilidee, die Face-to-Face Kommunikation zu erleichtern, gesehen werden.

3.1.2 Technologische Errungenschaften von Verbmobil

Die Besonderheit des Verbmobilsystems lag hauptsächlich darin, dass es ein Übersetzungssystem für Dialoge war, welches sich auf Spontansprache spezialisierte (vgl. Wahlster 1997: 215). Wie bereits in Kapitel 2 erwähnt, handelt es sich bei Spontansprache um

natürliche Sprache, also frei formulierte Alltagssprache, was impliziert, dass das System mit grammatikalisch inkorrekten oder unvollständigen Sätzen, Selbstkorrekturen oder Hästitionsausdrücken umgehen musste⁹.

So erklärt Walster (2000: 5): „[...] [I]n the Verbmobil corpus about 20% of all dialog turns contain at least one self-correction and 3% include false starts. Verbmobil uses a combination of shallow and deep analysis methods to recognize a speaker's slips and translate what he tried to say rather than what he actually said". Die Erkennung von solchen Selbstkorrekturen und ähnlichem erfolgt im Prosodiemodul, welches unmittelbar nach erfolgreicher Spracherkennung aktiviert wird. Das Modul identifiziert betonte Wörter und wertet die Satzmelodie aus. Die prosodischen Merkmale der menschlichen Stimme sind die Grundlage für die Erkennung von Selbstkorrekturen oder Mehrdeutigkeiten von einzelnen Wörtern, welche daraufhin in einen sogenannten „Worthypothesengraph" (Wahlster 2000: 6) eingetragen werden und mit alternativen Lösungsvorschlägen kombiniert werden, um eine Übersetzung ohne Korrektur zu produzieren. Bemerkenswert ist, dass Verbmobil zusätzlich auf Welt- und Diskurswissen zurückgriff, um eventuelle Mehrdeutigkeiten zu klären. Die Kombination aus statistischen und linguistischen Verfahren (vgl. Wahlster 1997: 215) und die Implementierung von Welt- und Diskurswissen macht das Verbmobil System zu einem hybriden Übersetzungssystem. Die „Verzahnung statistischer und linguistischer Analysestränge bei der Integration von Dialogakten/Dialogphasen in die tiefe Semantik" (Wahlster 2010: 10) sowie die „Kombination von Segmentübersetzungen aus verschiedenen Übersetzungssträngen [wie der statistischen und fallbasierten Übersetzung ebenso wie Übersetzung durch tiefe Analyse und Transfer]" (Wahlster 2010: 11), wurden zehn Jahre nach Abschluss des Projektes als Durchbruchinnovationen bezeichnet.

Das System konnte dank Verlaufsprotokollen bereits übersetzte Elemente in die noch zu übersetzenden Eingaben einbinden, um kontextsensitiv übersetzen zu können. Darüber hinaus konnten nach Gesprächsende ein Ergebnisprotokoll inklusive Ausgangs- und Zielttext gedruckt werden (vgl. <http://verbmobil.dfki.de> 2000).

Ein weiteres Merkmal spontaner Sprache, welches durch das Prosodiemodul ausgeglichen wird, ist die fehlende Interpunktion. Wahlster (1997: 215) illustriert dies anhand der Wortfolge „*Ja-zur-Not-geht-es-auch-am-Samstag*". Je nach Betonung der einzelnen Wörter, kann der Satz als Bestätigung („*Ja, zur Not geht es auch am Samstag.*") oder als Gegenvorschlag („*Ja, zur Not. Geht es auch am Samstag?*") verstanden werden. Mithilfe einer Analyse der Satzmelodie entscheidet das System über die passende Übersetzung. An

⁹ Es wurde darauf verzichtet, den exakten technischen Aufbau des Systems und die Interaktion der einzelnen Module genauer zu beschreiben, da dies einerseits den Rahmen der Arbeit sprengen würde und andererseits auch nicht Ziel dieser Arbeit ist. Das offizielle Verbmobil-Video (<http://verbmobil.dfki.de> 2000) bietet eine kurze Einführung in die technischen Hintergründe des Systems.

dieser Stelle eröffnet sich ein weiteres Hindernis - die Individualität der menschlichen Stimme.

Um vielen verschiedenen SprecherInnen gerecht zu werden, wurde Verbmobil als adaptives System konzipiert.

„Verbmobil is a speaker-adaptive system, i.e. for a new speaker it starts in a speaker-independent mode and after a few words have been uttered it improves the recognition results by adaptation. A cascade of unsupervised methods, ranging from very fast adaptation during the processing of a single utterance to a complex adaptation methods that analyse a longer sequence of dialog turns, is used to adjust to the acoustic characteristics of the speaker's voice, the speaking rate and pronunciation variats due to dialectal diversity of the user community.“ (Wahlster 2000: 6f)

Durch die Anpassung an akustische Merkmale der SprecherInnenstimme lernte das System mit jedem neuen Benutzer wortwörtlich dazu und konnte dadurch mit einem veränderten Sprechtempo und einer Vielzahl an Aussprachevarianten umgehen. Die Bewältigung all dieser Schwierigkeiten durch Verbmobil galt damals als technologische Errungenschaft.

3.2 Lecture Translator

Der Lecture Translator wurde im Jahr 2005 als gemeinsames Projekt des Anthropomatik Instituts des Karlsruher Instituts für Technologie (KIT) und der Carnegie Mellon University in Pittsburgh (USA) am International Centre for Advanced Communication Technologies (interACT) gestartet (vgl. Waibel 2012a). Es handelt sich dabei um ein Speech-to-Speech Translation System, welches im universitären Kontext integriert, gesprochene Sprache bei Vorträgen simultan von einer Sprache in die andere übertragen soll. Es ist das erste SST System, welches in Echtzeit arbeitet und dennoch Domänenadaptation aufweist (vgl. Fügen 2008: vi).

„The lecture translator combines the technologies of automatic speech recognition and statistic machine translation in an integrated system. Auxiliary components structure the text, deal with punctuation, treat German compound words, record the lecture, and display the translation result. The components are combined by an innovative service infrastructure.“ (KIT 2012)

Fügen (2008: 20f) stellt dabei in Anlehnung an humane Übersetzungen bzw. in diesem Fall wohl eher Dolmetschungen folgende vier Anforderungen an ein SST System:

- correct content (korrekter Inhalt)
- correct syntax (korrekte Syntax)
- high fluency (hohe Sprechflüssigkeit)
- low latency (niedrige Latenzzeit)

Er argumentiert dabei, dass abgesehen von der Erfüllung der wichtigsten Aufgabe eines Übersetzungssystemes – der korrekten Inhaltswiedergabe – auch die korrekte Syntax eine wesentliche Rolle spielt, da sich Syntaxfehler auf die Inhaltswiedergabe negativ auswirken können. Jedoch sei es nicht zwingend notwendig eine vollkommen korrekte Übersetzung zu liefern, um den Inhalt verstehen zu können. „Instead, it may be sufficient if the words carrying the content or meaning of a sentence are correctly translated and no misleading content do exist.“ (Fügen 2008: 21)

Die Anforderung an die Flüssigkeit der Übersetzung ist eine andere. So soll der Zieltext so natürlich wie möglich präsentiert werden. Neben der bereits angesprochenen korrekten Syntax, sollen vor allem Charakteristika der Spontansprache wie z.B. Selbstkorrekturen, Pausen oder andere Disfluenzen gar nicht erst ihren Weg in den ZT finden, sondern bereits vor oder während der Übersetzungsphase eliminiert werden. Die Anpassung an das Sprechtempo des/der Vortragenden trägt einerseits auch zu höherer Sprechflüssigkeit und Natürlichkeit bei, jedoch beeinflusst es auch die Latenzzeit, also den aus der Dolmetschwissenschaft bekannten Time Lag zwischen der Aussprache des Ausgangstextsegments und der Präsentation des Zieltextsegments. „In English, a latency of about two to six seconds is equivalent to a delay of about 4 to 12 words, since the average speaking rate is about 120 words per minute“, beschreibt Fügen (2008:21). Die Latenzzeit soll selbstverständlich so kurz wie möglich gehalten werden (vgl. Fügen 2008: 21).

Um diese Herausforderungen zu erfüllen und angesichts der Leistungsbeschränkung heutiger ASR und MT Komponenten sowie der Kosten eines solchen Systems, wurde statt eines domänenunabhängigen Ansatzes vielmehr ein domänen- und sprecheradaptiver Ansatz gewählt. Im Vordergrund stand folgende Idee: „[...] possible application scenarios for simultaneous translation are restricted to domains to which a system can be well adapted and applications in which a system can be re-used and modified or customized with less effort“. (Fügen 2008: 21) Hierfür eignet sich der Vorlesungskontext ganz besonders, da der/die Vortragende mehr oder weniger kontinuierlich über einen längeren Zeitraum spricht, der/die Vortragende im Vorhinein bekannt gegeben wurde, ebenso wie das Thema des Vortrages. Dadurch besteht die Möglichkeit auf verschiedenen Ebenen des Systems Adaptierungen durchzuführen. So kann sich die ASR Komponente an die Individualität der SprecherInnenstimme, den Sprechstil oder den Grad der Spontaneität anpassen. Die MT Komponente kann, für den Fall, dass die Trainingsdaten nicht mit den neuen Umständen übereinstimmen, die Übersetzungsqualität steigern, indem mittels einer Heuristik relevante Webseiten gesammelt werden und verwandte Publikationen aus dem Internet geladen werden, mithilfe derer die Übersetzungen angefertigt werden (vgl. Fügen 2008: vi; 21ff, Fügen et al. 2008: 228f).

Ebenso wie bei Verbmobil soll nun lediglich eine kurze Übersicht über die Funktionsweise des Systems und die darin verarbeitete Architektur gegeben werden, da die genauen technischen Spezifikationen für die vorliegende Arbeit nur zweitrangig sind.

3.2.1 Die Funktionsweise

Da das System für Vorträge und Reden konzipiert wurde, die üblicherweise vor etwas größerem Publikum gehalten werden, bedarf der/die Vortragende ein Mikrophon, welches sich nahe des Mundes des/der Sprecher/in befindet und den Ausgangstext, also den gesprochenen Vortrag, möglichst ohne Hintergrundgeräusche aufzeichnet. In der ASR Komponente wird der digitalisierte Text auf Basis verschiedener Modelle (Source Accoustic Model, Source Language Model, siehe Abbildung 5) in sogenannte partielle „first-best-hypotheses“ (Füger 2008: 16) umgewandelt. Diese Hypothesen werden durch eine zwischengeschaltete Resegmentation-Komponente in übersetzbare Segmente geteilt und an die MT Komponente weitergeleitet. Die Zwischenschaltung der Resegmentation-Komponente ist deswegen von Nöten, da statistische maschinelle Übersetzungsprogramme üblicherweise satzbasiert arbeiten und in der Spontansprache keine durch Interpunktion gekennzeichneten Sätze und Satzteile erkennbar sind.

„But when a low latency speech translation system is required, sentences are not an appropriate unit, because especially in spontaneous speech they tend to be very long – up to 20-30 words. To cope with this problem, a third component is introduced, which tries to reduce the latency by resegmenting the ASR hypotheses into smaller chunks without a degradation in translation quality.“ (Füger 2008: 16)

Aufgrund dessen teilt die Resegmentation-Komponente auf Basis akustischer und lexikalischer Merkmale den AT in semantisch abgeschlossene Segmente von maximal 10 Wörtern, bevor diese an die MT Komponente weitergeleitet werden (vgl. Füger 2008: 16; 134).

Die übersetzbaren Segmente werden in der MT Komponente schließlich übersetzt. Diese arbeitet statistikbasiert und wurde auf der Grundlage des phrasenbasierten Modells des EU Projektes TC-STAR entwickelt, welches bei Reden in den Plenarsitzungen des Europäischen Parlaments genutzt wurde (vgl. Füger et al. 2008: 225f). Jedoch handelt es sich bei den Vortragenden in den Plenarsitzungen um gut ausgebildete Redner mit vorbereiteten Reden zu einer eingrenzenden Domäne; im Gegensatz dazu war die Herausforderung im universitären Rahmen eine andere:

„Lectures, on the other hand, can go into more detail on any given topic, and, as an aggregate, cover a practically unlimited domain; a system suitable for translation general lectures must be able to cope with much more variable and more spontaneous speaking styles.“ (Füger et al. 2008: 226)

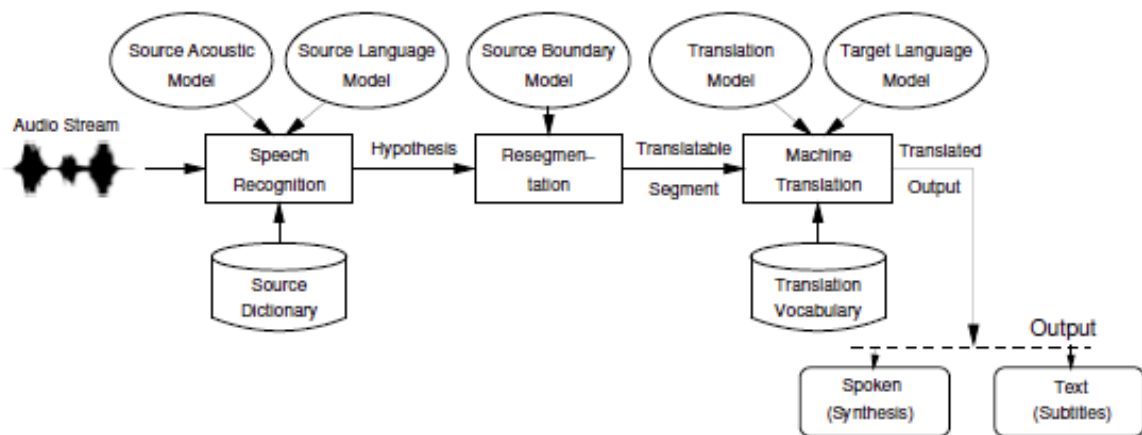


Abbildung 5: Schematische Übersicht über die Architektur des Lecture Translators (Fügen 2008: 138)¹⁰

Das Besondere an diesem SST System ist die Möglichkeit des variablen Outputs, entweder in akustischer oder in schriftlicher Form, die häufigste Output-Form sind Untertitel und Sprachsynthese (vgl. Fügen 2008: 16). Die individuellen Komponenten des Lecture Translators sind alle in einem Client-Server-Framework miteinander verbunden. Dies macht es möglich, dass mehrere Zieltextrezipienten mit dem Output eines einzelnen oder mehrerer Vortragenden versorgt werden. Statt für typischen Frontalunterricht könnte das System auch bei einer Podiumsdiskussion eingesetzt werden, in der mehrere Personen zur gleichen Zeit sprechen.

„For such a situation, the architecture makes it possible to use several speech recognizers for processing the input signals. The different outputs are then ordered, merged and resegmented per speaker in different resegmentation components, and several machine translation systems can then be used to consume the output of the resegmentation components, and to translate it into the different required languages. For the identification of the different streams and for the merging and resegmentation process, unique speaker- and segment-related identifiers are used. The different translation streams are then multiplexed for delivery by the different output technologies.“ (Fügen 2008: 137ff).

Wie bereits erwähnt ist eine der häufigsten Output-Formen jene der Projektion von Untertiteln auf einer Leinwand. Die Vorteile hierfür liegen auf der Hand: Es ist eine günstige und leicht umsetzbare Methode, die keiner zusätzlichen Hardware bedarf um mehrere RezipientInnen zu erreichen. Mit wachsendem Publikum erhöht sich auch die Zahl der Sprachen, die auf der Leinwand dargestellt werden müssen, sodass die Lesbarkeit darunter leidet. Fügen (2008: 139) zählt eine weitere Möglichkeit auf, die jedoch kostenintensiver ist; nämlich die Bereitstellung von Bildschirmen für eine Person oder kleinere Personengruppen derselben Zielsprache. Diese Variante ist heutzutage vor allem im universitären Setting nicht mehr notwendig, da Notebooks zur Grundausstattung der Studierenden gehören

¹⁰ Die Hauptkomponenten des Systems sind jene in den eckigen Umrandungen, während jene in Ellipsen die jeweils dazugehörigen Modelle darstellen. Die beiden möglichen Outputvarianten werden in Umrandungen mit runden Ecken dargestellt (vgl. Fügen 2008: 138).

und diese mit ihren eigenen Geräten in das System einsteigen können, um der Vorlesung oder dem Vortrag online auf der Website in ihrer Zielsprache zu folgen (vgl. Waibel 2012c: 39). Dasselbe gilt natürlich auch für alle anderen mobilen Geräte wie Tablets oder Smartphones (siehe Abbildung 6). Ein Nachteil den die schriftliche Darstellung der Übersetzung mit sich bringt, ist die Tatsache, dass sich die RezipientInnen nicht gleichzeitig auf den/die Vortragende/n und auf die Untertitel auf dem Display konzentrieren können. Zu diesem Zweck wurden spezielle „Heads-up Display Goggles“ (Waibel 2012c: 37) entwickelt, bei denen der übersetzte Zieltext direkt auf die Brillengläser projiziert wird (siehe Abbildung 7).



Abbildung 6: Heads-up Display Goggles
(Waibel 2012c: 37)



Abbildung 7: Lecture Translator auf Tablet
(Waibel 2012c: 39)

Die zweite Variante des Zieltext-Outputs ist die akustische Darbietung durch Sprachsynthese. Die erste Möglichkeit wäre die Nutzung von Standard-Lautsprechern, wobei dies nur sinnvoll ist, wenn das gesamte Publikum dieselbe Zielsprache verlangt. Eine andere Möglichkeit wären Kopfhörer für jede anwesende Person, jedoch stellt dies wiederum einen neuen Kostenfaktor dar und ist für die Kommunikation innerhalb des Publikums eher ungünstig, da die Kopfhörer immer wieder auf- und abgesetzt werden müssten. Um diese Probleme zu umgehen hat das Forscherteam die Anwendung von Lautsprechern mit einer Art Schallbündelung angedacht. „[...] [W]e have explored beam steered sound beams, which allow for delivering the translation to a desired group of people located in the beam using a single device only, without disturbing other people in the audience located outside of the beam.“ (Fügen 2008: 140) Damit wäre sowohl der Kostenfaktor, als auch die zusätzliche Beeinträchtigung durch Kopfhörer eingedämmt. Da für die Herstellung solcher *steered sound beams* Schallwandler nötig sind, die in ihrem Durchmesser größer als die Wellenlänge der auszusendenden Schallwellen sind, war dieser Ansatz für den Lecture Translator nicht besonders sinnvoll. So wurde eine Alternativlösung gefunden, indem auf einen

Schallwandler zurückgegriffen wurde, der mit mittels Ultraschall schmale Strahlenbündel bei deutlich kleinerem Gerätedurchmesser produziert (vgl. Fügen 2008: 137ff).

„The device uses the characteristic that, at high sound pressure levels, non-linear effects appear and the waveforms become distorted. This distortion can be calculated and used to advantage, such that the nonlinearity of the air acts as a demodulator. The ultrasound wave is used as a carrier, which is modulated by an audio signal.“ (Fügen 2008: 140)

Das Gerät besteht dabei aus vielen kleinen Ultraschalllautsprechern, die zusammen den ZT in einem Audiostrahl von 1-2 Metern Weite generieren (vgl. Fügen 2008: 141). Auch mit entsprechender technischer Ausstattung ist das sogenannte „delivery rate problem“ nicht zu verachten. Kurze Latenzzeit und hohe Qualität sind zwei Merkmale des SST Systems, die genauegenommen in Konkurrenz zu einander stehen. So muss das System, um die Latenzzeit konstant zu halten, den Input so schnell verarbeiten wie dieser produziert wird und muss demnach auch den Output in derselben Schnelligkeit generieren. Bei Sprachenpaaren in denen die Zielsprache einzelne ausgangsprachliche Ausdrücke in mehreren Wörtern wiedergeben muss, um den Sinn beizubehalten, müsste die *delivery rate* erhöht werden. Zudem sind die Übersetzungen des Lecture Translators nicht fehlerfrei, was bedeutet, dass sich das Verständnis einer nicht perfekten geschriebenen oder gesprochenen Übersetzung durch die zusätzliche Erhöhung der *delivery rate* um einiges erschweren würde (vgl. Fügen et al. 2008: 241).

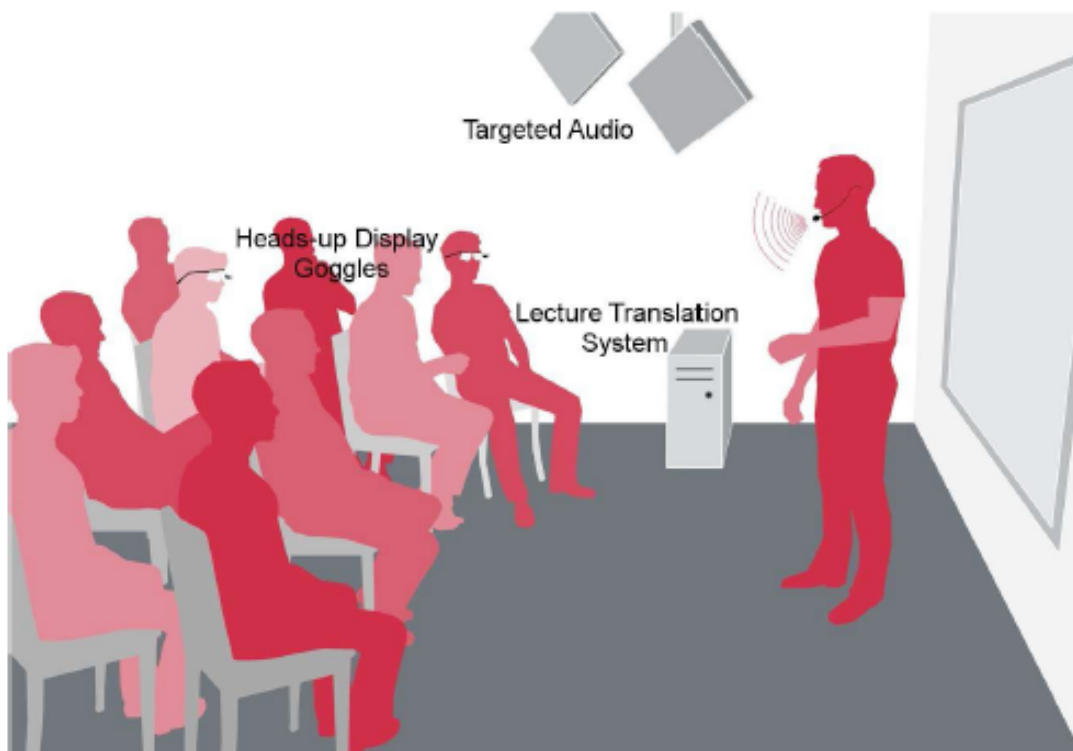


Abbildung 8: Überblick über ein Szenario mit dem Lecture Translator (Fügen 2008: 141)

Abbildung 8 stellt nochmals abschließend alle Input- und Output-Technologien dar, die der Lecture Translator für eine erfolgreiche Speech-to-Speech Translation braucht.

Am 11. Juni 2012 kam das SST System zum ersten Mal am KIT Campus offiziell zum Einsatz; seitdem wird es in verschiedenen Lehrveranstaltungen am Campus regelmäßig genutzt (vgl. KIT 2012). Eine Aufnahme der Präsentation inkl. eines Ausschnittes der Benutzeroberfläche des Lecture Translators ist auf dem Youtube-Kanal des KIT zu sehen (vgl. <https://www.youtube.com>^a 2012).

Anhand der zwei Projekte – Verbmobil und Lecture Translator – ist die lange Entwicklung der maschinellen Übersetzung von Spontansprache deutlich erkennbar. Während ersteres nur ein Forschungsprojekt war, welches bei Test durchaus gute Ergebnisse in den vorgesehenen Domänen lieferte, wurde es nie in der Praxis angewendet. Der Lecture Translator hingegen, entstanden aus der Idee der immer höher werdenden Zahl von Austauschstudierenden und Studierenden aus dem Ausland linguistisch gesehen uneingeschränkten Zugang zu Forschung und Wissen zu bieten, wird seit 2012 regelmäßig am KIT Campus eingesetzt. Obwohl das System bei weitem nicht fehlerfrei arbeitet, stellt es dennoch eine große Unterstützung für jene Studierenden dar, die den Vortrag in der Originalsprache gar nicht verstehen würden.

4 Funktionaler Vergleich von SST Apps

Im Rahmen der vorliegenden Arbeit soll ein funktionaler Vergleich von drei Speech-to-Speech Translation Apps unterschiedlicher Anbieter durchgeführt werden. Da es sich bei den Forschungsobjekten um kommerzielle SST Systeme handelt, welche für mobile Endgeräte wie Smartphones und Tablets konzipiert sind, wurde bei der Auswahl der Apps darauf geachtet, dass sie nicht auf eine oder mehrere Domänen beschränkt sind, sondern vielmehr domänenübergreifend arbeiten. Als Zielgruppe ist die breite Masse anzunehmen, von der die Apps aufgrund mangelnder Sprachkenntnisse in unterschiedlichem Ausmaß (ein Wort oder ganze Sätze) und in diversen Situationen genutzt wird. Die Apps wurden alle auf einen iPhone 5 mit iOS 7.1.2 in ihrer zum Zeitpunkt aktuellsten Version analysiert und beschrieben. Um welche Version es sich explizit handelt, wird beim jeweiligen App-Kapitel genauer beschrieben.

Der Vergleich zwischen den Apps soll vor allem funktionaler Natur sein und folgende Forschungsfragen beantworten:

- Forschungsfrage 1: Über welchen Funktionsumfang verfügen die einzelnen Speech-to-Speech Translation Apps?
- Forschungsfrage 2: Worin liegen die Stärken und Schwächen der einzelnen Speech-to-Speech Translation Apps?

Zu diesem Zweck wurde eine umfassende, vergleichende Analyse des Funktionsumfangs der drei Apps durchgeführt (Kapitel 5, 6,7), bei der folgende Punkte untersucht werden:

- Technische Hintergründe
- User Interface
- Features
- Benutzerfreundlichkeit
- SWOT-Analyse

Der Punkt technische Hintergründe enthält neben einer kurzen Vorstellung des Entwicklerunternehmens, Informationen zur Verfügbarkeit und Größe der App sowie der Kompatibilität mit bestimmten Betriebssystemen. Unter dem Punkt User Interface sollen das Erscheinungsbild der App allgemein und der einzelnen Menüpunkte, sowie der allgemeine Aufbau der App beschrieben werden. Erst im dritten Punkt werden die Features und Funktionen der im Einzelnen vorgestellt. Dazu zählen die angebotenen Sprachen, der Übersetzungsmodus sowie Besonderheiten der einzelnen Apps. Im Punkt Benutzerfreundlichkeit soll der (subjektive) Eindruck bezüglich Anpassbarkeit an persönliche Wünsche und Vorlieben, Erlernbarkeit und intuitiver Aufbau sowie Zuverlässigkeit und Robustheit der App beschrieben werden. Bei der Analyse der Benutzerfreundlichkeit der App spielt die Subjektivität eine nicht zu unterschätzende Rolle. Abschließend wird eine zusammenfassende

SWOT-Analyse (Strenghts, Weaknessess, Opportunities, Threats) erstellt, die einen übersichtlichen Einblick in die Stärken, Schwächen, Chancen und Risiken der App bieten soll.

Zudem wird in Kapitel acht ein Anwendungsbeispiel aller Apps vorgestellt. Zu diesem Zweck wurden zwei Testdialoge erstellt, welche jeweils zwei verschiedene Situationen simulieren sollen, in denen die Apps Anwendung finden könnten. Der erste Testdialog mit der Sprachkombination Deutsch-Englisch ist in einen touristischen Kontext eingebettet, während der zweite Testdialog mit der Sprachenkombination Deutsch-Französisch in einem geschäftlichen Kontext situiert ist. Bei der Erstellung der Testdialoge wurde darauf geachtet, die in der Literatur beschriebenen Charakteristika der Spontansprache, wie Selbstkorrekturen, Füllwörter, verkürzte Syntaxkonstellationen etc., zu inkludieren, um realitätsnahe Umstände und damit verbundene Leistungsergebnisse der App zu schaffen. Darüber hinaus wurde bei der Aufnahme der Testdialoge darauf Wert gelegt, so natürlich wie möglich zu sprechen, d.h. kein künstliches Verlangsamen des Sprechtempos oder überdurchschnittlich deutliches Aussprechen der einzelnen Segmente.

Abschließend wurde ein gegenüberstellender Überblick über die Funktionen der Apps erstellt, um, unter Einbezug der SWOT-Analyse, Antworten auf die Forschungsfragen zu finden.

5 Jibbig Translator 2.5.5

Die App *JibbigTranslator* entstand unter der wissenschaftlichen Zusammenarbeit zwischen der Carnegie Mellon University in Pittsburgh, dem KIT (Karlsruhe Institute of Technology) und Silicon Valley. Mit der Firmenneugründung von Mobile Technologies LLC wurde die App von Absolventen der interACT Partneruniversitäten und Dr. Alexander Waibel entwickelt. *Jibbig* wurde 2009 als erste Speech-to-Speech Translation App für das Apple iPhone 3GS auf den Markt gebracht. Mit einem Wortschatz von etwa 40.000 Wörtern war *Jibbig* als Übersetzungssoftware für humanitäre Einsätze und Touristen konzipiert worden, mit der in ca. 20 verschiedene Sprachen übersetzt werden konnte. Ein großer Vorteil den die App mit sich brachte, war die Offline-Funktion. Für die Inbetriebnahme der Software wurde kein Zugang zum Internet oder zu Servern benötigt. Für Touristen würden so keine Roaminggebühren anfallen und bei humanitären Einsätzen wird häufig in Gegenden gearbeitet, wo der Ausbau von Mobilfunkanlagen noch sehr spärlich ist (vgl. Rödder 2009).

Zu Beginn kostete die App ungefähr 25\$, später wurde dazu übergegangen eine gratis verfügbare online-only Version und offline Sprachpaare um je 4,99\$ anzubieten (vgl. <https://www.youtube.com>^b o.J.), Darüber hinaus konnten auch sogenannte Bundles um 9,99\$ erworben werden, die bestimmte Kategorie von Sprachen enthalten wie z.B. European Bundle, Asian Bundle (vgl. Waibel 2012c: 14). Die Verkauf der Sprachpaare und Bundles verlief über die in der Zwischenzeit weit verbreiteten und bekannten In-App-Käufe. So konnte die App offiziell als gratis App promotet werden, während die UserInnen je nach gewünschtem Funktionsumfang direkt in der App zusätzliche offline Sprachpaare oder Bundles erwerben konnten oder sich mit der kostenfreien online Version begnügen mussten.

Im August 2013 kam es zu einer großen Veränderung – *Jibbig* wurde, inklusive des Entwicklers Mobile Technologies von Facebook aufgekauft (vgl. <https://www.facebook.com> 2013). Zu dem Verkauf finden sich, außer einiger Artikel in diversen Blogs und Technologie-Magazinen, keine Einzelheiten oder öffentlichen Stellungnahmen seitens der Unternehmen. Facebook Director of Product Management Tom Stocky postete lediglich eine kurze Statusmeldung zum Verkauf:

„It has always been our mission to make the world more open and connected. Although more than a billion people around the world already use Facebook every month, we are always looking for ways to help connect the rest of the world as well. Voice technology has become an increasingly important way for people to navigate mobile devices and the web, and this technology will help us evolve our products to match that evolution. We believe this acquisition is an investment in our long-term product roadmap as we continue towards our company's mission.“ (<https://www.facebook.com>¹¹ 2013)

¹¹ Posting von Tom Stocky, Facebook Director of Produkt Management, vom 12.08.2013

Mobile Technologies hatte auf der *Jibbig* Website kurzzeitig eine Meldung zum Verkauf publiziert, jedoch wurde die Website seither stark verändert und enthält nunmehr lediglich einen Link zum Google Playstore und Apple App Store, eine kurze Auflistung von Frequently Asked Questions sowie je einen Link zum End User License Agreement und der Privacy Policy (Datenschutzrichtlinie) von Facebook (vgl. <http://jibbig.com> 2014). Zu Anfang der Recherchen (Juli/August 2013) zu der vorliegenden Arbeit war die *Jibbig* Website um einiges umfangreicher und informativer. Leider stehen keine Screenshots o.ä. der damaligen Seite sowie des Statements von *Jibbig* zur Verfügung, daher muss auf wörtliche Zitate in Blogbeiträgen zurückgegriffen werden:

„Mobile Technologies was founded in 2001 with the goal of breaking through language barriers to open up communication between the people of the world. With 6,000 languages in the world, the language divide presents the greatest human communication challenge. Mobile Technologies has developed a number of cross-lingual communication tools: Jibbig, the world’s first speech-to-speech translator on a phone that runs online and even off-line, independent from the Internet. Travelers around the world use Jibbig to communicate in foreign countries and health-care workers overcome the language challenge in humanitarian missions. Mobile Technologies also developed and deployed the first automatic, simultaneous interpretation service for lectures [called Lecture Translator] and deployed it in educational settings, so that ideas can transcend nations and cultures. Facebook, with its mission to make the world more open and connected, provides the perfect platform to apply our technology at a truly global scale. We look forward to continuing to develop our technology at Facebook and finding new and interesting ways to apply it to Facebook’s long-term product roadmap.“ (Lafferty 2013)

Auch der Youtube Channel von *Jibbig*, wo verschiedene Videos zur App und über deren Anwendung hochgeladen wurden, ist inzwischen eingestellt worden und es wird lediglich noch auf die aktuelle Website verwiesen (vgl. <https://www.youtube.com>^b o.J).

Im Folgenden werden die Einzelheiten der App wie in Kapitel 4 vorgesehen vorgestellt und es soll an dieser Stelle auch darauf hingewiesen werden, dass sich die folgenden Daten und Fakten auf die aktuell erhältliche Version (Stand: 23.07.2014) beziehen und daher eine vergängliche Gültigkeitsdauer haben.

5.1 Technische Hintergründe

Der *Jibbig Translator* ist im Apple App Store für alle iOS Geräte¹² in der Version 2.5.5 (letztes Update am 08.04.2014) gratis erhältlich und erfordert iOS 6.1 oder neuer. Es wird jeweils eine für das iPhone und für das iPad optimierte Version angeboten. Die Größe der App beträgt 14.6 MB und sie wird in der Kategorie Reisen aufgezählt. Es existieren keine Wertungen oder Rezensionen zu der App, weder zur aktuellen noch zu älteren Versionen.

¹² iOS Geräte sind alle Apple Produkte, die das erstmals mit dem iPhone eingeführte Betriebssystem (operating system) für mobile Geräte nutzen. Dazu gehören das iPhone, der iPod touch und das iPad. Der Name ist angelehnt an das Mac OS von Apples Macintosh Computern.

Als Entwickler wird Jibbiggo LLC angeführt, der eine weitere App namens *Record It what I read* anbietet (vgl. <https://itunes.apple.com>^a 2014). Es handelt sich laut Beschreibung um eine App, die UserInnen die Möglichkeit bietet an der Forschung zur Speech Recognition teilzunehmen. Bei Interesse solle man sich per E-Mail an Jibbiggo LLC wenden¹³ (vgl. <https://itunes.apple.com>^b 2014).

Im Google Play Store ist der *Jibbiggo Übersetzer* für Android-Versionen 2.3 oder höher gratis erhältlich und befindet sich seit dem letzten Update am 09.05.2014 in Version 3.2. Die App fällt bei Google unter die Kategorie Reise & Lokales und hat eine Größe von 11 MB. Laut den insgesamt 498 Erfahrungsberichten erhielt *Jibbiggo* eine Wertung von 3,6 von 5 möglichen Sternen (vgl. <https://play.google.com>^a 2014).

5.2 User Interface

Bei der Inbetriebnahme der App erscheint der in weiß gehaltene *Jibbiggo* Schriftzug auf dem für das Logo typischen blauen Hintergrund für 1-1,5 Sekunden, danach wird zunächst von den UserInnen verlangt den Lizenzvertrag (End User License Agreement, EULA) zu akzeptieren, bevor die Software benutzt werden darf. *Jibbiggo* behält sich das Recht vor, die Geschäftsbedingungen oder Funktionen der Software jeder Zeit zu ändern, wobei sich die UserInnen durch Weiterverwendung der Software nach Publikation von Veränderungen des EULA dazu vertraglich verpflichten diese anzunehmen (vgl. Jibbiggo, LLC 2014). Vor allem Punkt 3. Use of Speech Data sollte genauestens gelesen werden, da die UserInnen darüber aufgeklärt werden, dass sie durch Zustimmung zum Lizenzvertrag alle Audioaufnahmen, Transkriptionen und sonstige aus der Nutzung der App entstandenen Daten (speech data) dem Unternehmen *Jibbiggo, LLC* zur Verfügung stellen. Darüber hinaus darf *Jibbiggo* persönliche Daten der UserInnen und die gesammelte speech data mit Geschäftspartnern teilen:

¹³ Dass die App seit 2012 kein neues Update erhalten hat und der Link zum Support auf eine abgelaufene Domäne (<http://www.mobytrans.com>) führt, war Grund genug die App kurzerhand zu downloaden und den Sachverhalt genauer zu untersuchen. Man wird sofort aufgefordert sich als bestehender oder neuer User anzumelden und im nächsten Schritt die Terms of Agreement anzunehmen oder abzulehnen. Darin wird man aufgeklärt, dass die aufgenommene Stimme für die Forschung und Entwicklung in wissenschaftlichen und kommerziellen Projekten verwendet wird und man alle Rechte an den aufgenommenen Daten an Mobile Technologies abtritt, wobei alle persönlichen Daten anonymisiert werden und nur für interne Zwecke genutzt werden. Weitere Schritte in der App wurden nicht mehr unternommen.

„LICENSOR [Jibbiggo, LLC] collects and uses the audio files, associated transcriptions and log files provided by you hereunder or generated in connection with the SOFTWARE (collectively referred to as “SPEECH DATA”), to tune, enhance and improve the speech recognition and other components of the SOFTWARE and to develop and provide related services. In accepting the terms and conditions of this EULA, you acknowledge, consent and agree that LICENSOR may collect the SPEECH DATA and that such information shall be used by LICENSOR to tune, enhance and improve the speech recognition and other components of the SOFTWARE and to develop and provide related services. You acknowledge, consent and agree that LICENSOR may share SPEECH DATA and personal data that LICENSOR collects with business partners with whom LICENSOR jointly offers products or services; corporate affiliates; companies with whom LICENSOR is involved in connection with a merger, acquisition, asset sale or other corporate transaction; or otherwise where sharing is required or permitted by law.“ (Jibbiggo, LLC 2014) [Hervorhebungen im Original]

Der Lizenzvertrag kann auf der *Jibbiggo* Website jederzeit nachgelesen werden (vgl. Jibbiggo, LLC 2014).

Nachdem dem Lizenzvertrag zugestimmt wurde, erscheint das Hauptmenü (vgl. Abbildung 9), welches im Prinzip nur aus zwei nach oben und unten scrollbaren Listen mit Sprachen besteht. Das User Interface (UI) erscheint auf den ersten Blick durch die vielen Flaggen etwas bunt, ist jedoch sehr aufgeräumt, da es von hier aus lediglich zwei mögliche Wege gibt: zum Button *Einstellungen* oben rechts, oder der Button *Übersetzer starten* ganz unten, der sich als ganzer Balken präsentiert. Je nach ausgewählter Sprachkombination verändert sich das Erscheinungsbild des Hauptmenüs, da nicht alle Sprachen beliebig kombinierbar sind. Die Ausgangssprache wird auf der linken Seite ausgewählt, während die gewünschte Zielsprache der rechten Liste entnommen wird. In Abbildung 9 wird ersichtlich, dass die AS Deutsch relativ viele kombinierbare ZS hat; wäre Dänisch als AS gewählt, würde die rechte Liste lediglich die Sprachen Englisch International und Englisch USA/Kanada enthalten. Für die UserInnen ist außerdem sofort ersichtlich mit welchem Funktionsumfang die jeweilige Sprache ausgerüstet ist. Als Zeichen für vorhandene ASR ist neben der Sprache ein kleines, graues Mikrophon angeführt und für die TTS Komponente steht das Lautsprechersymbol. Nicht alle Sprachen verfügen zwingenderweise über eine ASR und TTS Komponente.



Abbildung 9: Screenshot Hauptmenü Jibbigo 2.5.5

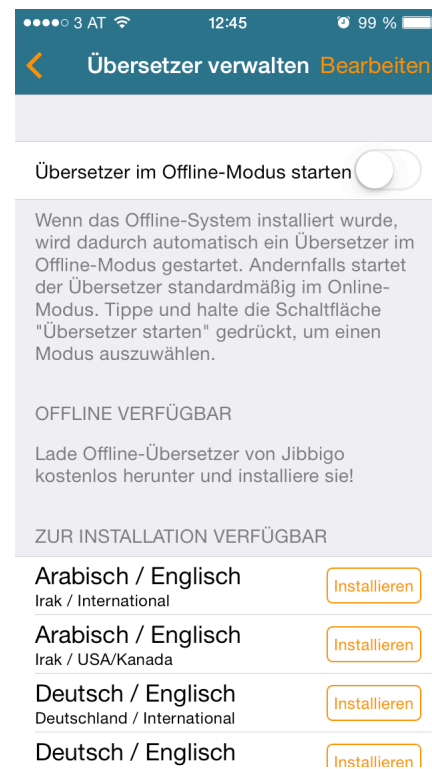


Abbildung 10: Screenshot Einstellungen Jibbigo 2.5.5

Wie bereits erwähnt haben die UserInnen vom Hauptmenü aus zwei mögliche Wege – die Einstellungen oder den Übersetzermodus. Entscheiden sich die AnwenderInnen für die Einstellungen, erscheint ein Bildschirm wie in Abbildung 10. Hier ist es möglich den Offline-Modus des Übersetzers zu aktivieren. Wenn der Schalter umgelegt wird, ändert sich die Farbe des Schalters auf orange und von diesem Moment an wird der Übersetzer automatisch im Offline-Modus gestartet, sodass kein Datenvolumen verbraucht wird bzw. keine zusätzlichen Kosten entstehen können. Dies setzt natürlich die Installation der entsprechenden Sprachkombinationen voraus. Positiv anzumerken ist, dass die UserInnen über die Einzelheiten des Online bzw. Offline-Modus durch den kurzen Informationstext unterhalb des Schalters aufgeklärt werden. Demnach können die AnwenderInnen bei aktiviertem Offline-Modus selbst entscheiden ob der Übersetzer standardmäßig im Offline-Modus gestartet werden soll oder ein online Zugriff erfolgen soll; es muss lediglich der Button *Übersetzer starten* angetippt und länger gehalten werden, um zwischen den Modi wechseln zu können. Wird auf den orangen Installationsbutton, der neben den Sprachpaaren zu finden ist, getippt, verändert sich der Text des Buttons, sodass UserInnen sofort Informationen zur Downloadgröße des Sprachenpaars erhalten. Beispielsweise für die Kombination Deutsch Deutschland/Englisch International beträgt die Größe des Pakets 212,8 MB. Der orange Pfeil links oben führt wieder zurück zum Hauptmenü, von wo aus der nächste Schritt die UserInnen zum eigentlichen Übersetzungsmodus leitet.

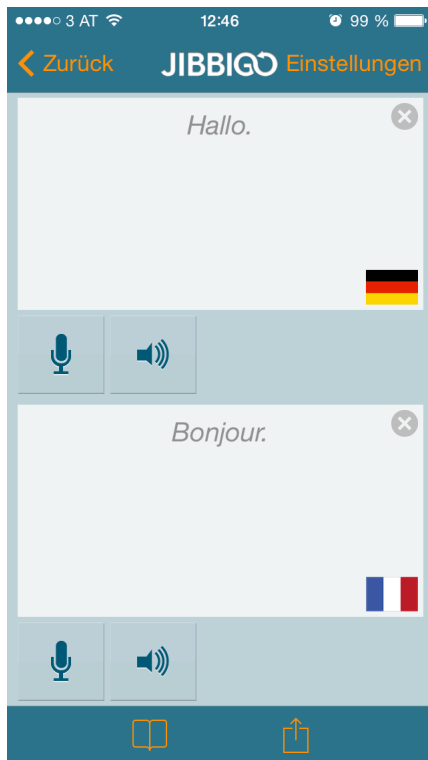


Abbildung 11: Screenshot vollwertiger Übersetzer-Modus Jibbigio 2.5.5



Abbildung 12: Screenshot Übersetzer-Modus ohne ASR Jibbigio 2.5.5

Im Übersetzer-Modus (vgl. Abbildung 11) ist der Bildschirm optisch in zwei Teile geteilt. Im oberen Teil findet sich ein weißes Textfeld, welches sowohl mit der Flagge der Ausgangssprache, als auch mit einer Begrüßung in der jeweiligen AS ausgestattet ist. Unter dem Textfeld befinden sich zwei Icons, die je zwei Funktion symbolisieren sollen. Das Mikrophon steht für die Spracherkennung und startet die Sprachaufnahme. Durch den Lautsprecherbutton kann im Nachhinein der durch die ASR Komponente analysierte und transkribierte Text als Sprachausgabe wiedergegeben werden. Dieselben Elemente – Textfeld mit Flaggen und entsprechender Begrüßung, Mikrophon und Lautsprecher – finden sich auch in der unteren Hälfte des Bildschirms und stellen die zielsprachlichen Funktionen dar. Bei einem Dialog verlieren die Benennungen Ausgangssprache und Zielsprache an Relevanz, da sich die Rollen ständig tauschen und dann die untere Sprache als Ausgangssprache bzw. die obere Sprache als Zielsprache fungiert. Es soll nur darauf hingewiesen werden, dass die in Abb. 9 aus der linken Spalte ausgewählte Sprache im Übersetzer-Modus (vgl. Abbildung 11) oben erscheint. Unabhängig davon, welche Sprache ursprünglich als Ausgangssprache gewählt wurde, kann die Übersetzung auch mit der unteren Sprache gestartet werden, sobald man den Mikrophonbutton drückt. Die Sprachaufnahme kann auch gestartet werden, indem einfach nur auf das Textfeld getippt wird. Dass die Sprach-

aufnahme in Gange ist, ist an der roten Färbung des Textfeldes und an dem erscheinenden Stimmengraphen zu erkennen (vgl. Abbildung 13).

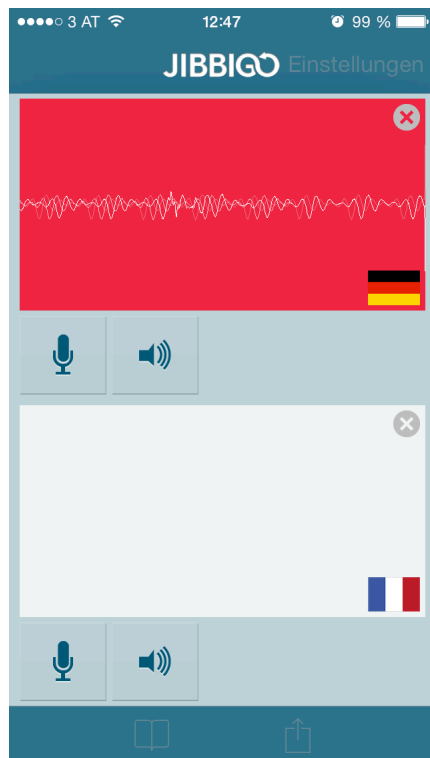


Abbildung 13: Screenshot Sprachaufnahme JibbigO 2.5.5

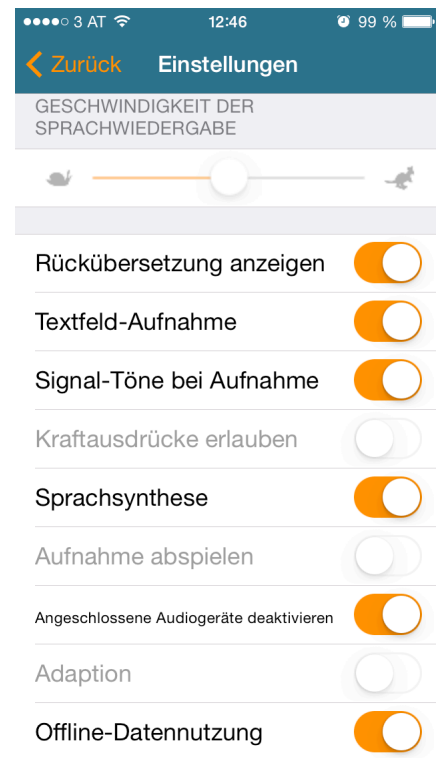


Abbildung 14: Screenshot Einstellungen Übersetzer-Modus JibbigO 2.5.5

Der Übersetzer-Modus verfügt über eigene Einstellungen, auf die über die rechten oberen Ecke zugegriffen werden kann. Die einzelnen Funktionen (vgl. Abbildung 14), die in den Einstellungen nach individuellen Wünschen aktiviert bzw. deaktiviert werden können, werden in Kapitel 5.3 genauer vorgestellt, da dieses Kapitel zur allgemeinen Beschreibung des User Interfaces und der Architektur der App dient.

Zurück im Übersetzer-Modus haben die AnwenderInnen die Möglichkeit, auf die bereits übersetzten Segmente zuzugreifen, in dem Sie auf das iOS typische Verlauf-Icon, welches einem aufgeschlagenen Buch ähnelt, in der unteren Leiste (vgl. Abbildung 11) tippen. Es öffnet sich der Verlaufs Bildschirm, der den Ausgangssatz als auch den Zielsatz (in etwas kleinerer Schriftgröße) anzeigt (vgl. Abbildung 15). Tippt man auf eines der bereits übersetzten Segmente, wird der Zielsatz mittels Sprachausgabe wiedergegeben. Jedoch nur falls, wie bereits erwähnt, eine Sprachausgabe Funktion für die jeweilige Zielsprache existiert. Nicht alle übersetzten Segmente werden in einem gemeinsamen Verlauf gespeichert, sondern abhängig von ihrer Sprachkombination. So zeigt Abbildung 15 beispielsweise nur die Segmente des Sprachenpaares Deutsch-Italienisch. Die Funktion *Bearbeiten* rechts oben erlaubt es, einzelne oder alle Einträge im Verlauf zu löschen.



Abbildung 15: Screenshot Übersetzungsverlauf Jibbiggo 2.5.5

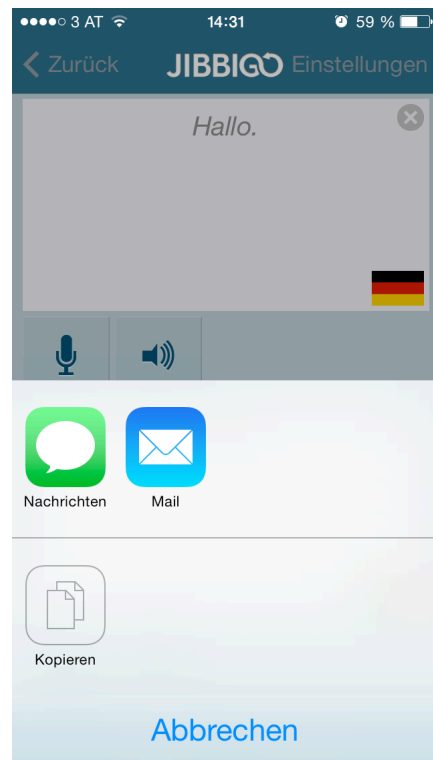


Abbildung 16: Screenshot Inhalte teilen Jibbiggo 2.5.5

Neben dem Icon für den Verlauf findet sich das iOS typische Icon für die Funktion *Inhalte teilen* (vgl. Abbildung 11). Tippt man dieses Icon an öffnet, sich eine Art Halbfenster, welches den unteren Teil des Bildschirms bedeckt, während der obere Teil des Bildschirms, in dem noch der Übersetzer-Modus zu sehen ist, mit einem dunkleren Kontrast dargestellt wird und somit optisch in den Hintergrund rückt (vgl. Abbildung 16). In dem neu geöffneten Halbfenster kann der User oder die Userin wählen, ob das aktuelle Übersetzungssegment (im oberen Teil noch sichtbar) per SMS-Nachricht oder E-Mail versendet werden soll, oder lediglich in die Zwischenablage kopiert werden soll, um es zu einem späteren Zeitpunkt an einem anderen Ort, oder in einer anderen App, einzusetzen. Entscheidet man sich für die Weiterleitung via Nachricht, öffnet *Jibbiggo* ein neues Fenster (vgl. Abbildung 17), welches dem Layout der iOS typischen Nachrichten App sehr ähnelt, jedoch in charakteristischem *Jibbiggo*-Blau gehalten ist, und fügt als Nachrichtentext das ausgangssprachliche und das zielsprachliche Übersetzungssegment ein. Der Text kann noch nach Belieben abgeändert werden, bevor die Nachricht versendet wird. Ähnlich verhält es sich auch wenn die Option per E-Mail versenden ausgewählt wird.

Allgemein kann gesagt werden, dass die *Jibbiggo* App einen klar strukturierten Aufbau hat, sodass man als Anwender und Anwenderin leicht den Überblick über die verschiedenen verknüpften Unterkapitel behält. Das User Interface wirkt aufgeräumt, nicht zu überladen und erschlägt die UserInnen nicht mit zu vielen angebotenen antippbaren Optionen. Die verwendeten Symbole sind iOS Kennern vertraut und ermöglichen damit eine leichte Navigation durch die App. Die Bewegung zwischen den verschiedenen Menüs ist sehr flüssig und es kam zu keinen Verzögerungen, Einfrierungen des Bildschirmes oder Abstürzen der App.

Da nun der Aufbau der App analysiert und präsentiert wurde, soll im nächsten Kapitel auf die konkreten Funktionen und die angebotenen Besonderheiten der App eingegangen werden.

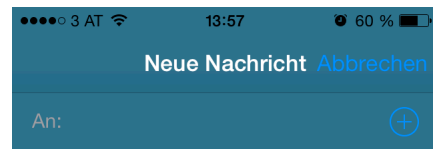


Abbildung 17: Screenshot Inhalte via Nachricht weiterleiten jibbiggo 2.5.5

5.3 Features

In diesem Kapitel werden der Funktionsumfang der App genauer vorgestellt, wobei jedenfalls das Sprachenangebot, der Übersetzer-Modus, die Besonderheiten und das Werbeversprechen der einzelnen Apps unter die Lupe genommen werden.

5.3.1 Sprachenangebot

Das wohl wichtigste Feature der Speech-to-Speech Translation Apps sind die angebotenen Sprachen. Der *Jibbiggo Translator* in der Version 2.5.5 unterstützt insgesamt 25 Sprachen, von denen 16 Sprachen mit vollem Funktionsumfang ausgestattet sind, d.h. über eine Spracherkennungs- und Sprachausgabefunktion verfügen, und 9 Sprachen ohne diese Funktionen auskommen müssen. Dies ist, wie bereits erwähnt, bei der Sprachauswahl im Hauptmenü durch das Mikrophon- und Lautstärke-Icon impliziert (vgl. Abb. 9).

Zu den angebotenen Sprachen gehören:

- Arabisch (Irak)
- Bulgarisch (Bulgarien)
- Chinesisch (China)
- Dänisch (Dänemark)

- Deutsch (Deutschland)
- Englisch (USA/Kanada)
- Französisch (Frankreich)
- Hebräisch (Israel)
- Italienisch (Italien)
- Katalanisch
- Paschtu (Afghanistan)
- Portugiesisch (Portugal)
- Russisch (Russische Föderation)
- Spanisch (Spanien)
- Thailändisch (Thailand)
- Englisch (International)
- Esperanto
- Griechisch (Griechenland)
- Indonesisch (Indonesien)
- Japanisch (Japan)
- Koreanisch (Republik Korea)
- Portugiesisch (Brasilien)
- Rumänisch (Rumänien)
- Spanisch (Mexiko)
- Tagalog (Philippinen)

Auf der Website von *Jibbig* (vgl. <http://jibbig.com>) werden lediglich 22 Sprachen angeführt. Es fehlen Indonesisch und Russisch, darüber hinaus wird Portugiesisch nur einfach aufgelistet, während es in der App eine Zweiteilung von Portugiesisch in Brasilien und Portugal gibt.

Die Wahl der Sprachkombinationen ist insofern eingeschränkt, als nicht für jede beliebige Ausgangssprache eine der restlichen 24 Zielsprachen wählbar ist. Aus dem Englischen kann beispielsweise in 20 Zielsprachen übersetzt werden, aus dem Französischen in 15 ZS, aus dem Spanischen (Spanien) in 23 ZS und aus dem Chinesischen in 12 ZS. Aus dem Deutschen kann man in 12 Zielsprachen übersetzen lassen, dazu gehören: Chinesisch, Englisch beide Varianten, Französisch, Italienisch, Japanisch, Koreanisch, Paschtu, Spanisch beide Varianten, Tagalog, Thailändisch.

5.3.2 Übersetzer-Modus

Der Übersetzer-Modus wirkt, wie bereits gesagt, sehr aufgeräumt und das User Interface ist aufgrund der überschaubaren Anzahl der zur Wahl stehenden Buttons recht intuitiv gestaltet. Eine Besonderheit des *Jibbig Translators* ist die Möglichkeit den Übersetzer-Modus auch offline nutzen zu können. Dazu muss lediglich das jeweilige Sprachenpaar gedownloadet werden. Wird der Übersetzer im standardmäßigen Online-Modus gestartet, sind die Auswahlmöglichkeiten in den Einstellungen (vgl. Abb. 14) ausgegraut und demnach nicht (de)aktivierbar. Dazu gehört zum Beispiel auch der Geschwindigkeitsregler für

die Sprachwiedergabe, der die Funktionen schnell und langsam mittels sprachunabhängiger Bilder eines Kängurus bzw. einer Schnecke darstellt. Dieser kann ebenso wie die Regler *Kraftausdrücke erlauben*, *Aufnahme abspielen* und *Adaption* nicht betätigt werden.

Wurde ein entsprechendes Sprachpaket auf das Smartphone geladen und der Offline-Modus (vgl. Abb. 10) aktiviert, verändert sich das Aussehen des Hauptmenüs und der Übersetzermodus etwas. Im Hauptmenü (vgl. Abb. 9) wird in der unteren Leiste links außen neben dem Schriftzug *Übersetzer starten* ein Pluszeichen eingefügt, sodass die UserInnen sofort erkennen, dass die gewählte Sprachkombi im Offline-Modus und somit in größerem Funktionsumfang genutzt werden kann. Auch die Einstellungen des Hauptmenüs verändern sich etwas, sobald ein offline Sprachpaket heruntergeladen wurde. So erscheint dieses nicht mehr in der Rubrik *Zur Installation verfügbar*, sondern etwas oberhalb unter der Rubrik *Offline verfügbar* anstelle des Informationstextes. Mittels der Bearbeitungsfunktion rechts oben können bereits installierte Sprachpakete wieder vom Gerät gelöscht werden.

Die verhältnismäßig größte Veränderung in der Offlinenutzung erlebt der Übersetzer-Modus. So sind die individuell anpassbaren Übersetzereinstellungen (vgl. Abb. 14) beispielsweise alle frei wählbar. Die Funktion *Rückübersetzung anzeigen* ist auch in der online Version aktivierbar, jedoch wird diese in keiner der getesteten Sprachkombinationen angezeigt. In der Offlinenutzung wird die Rückübersetzung des Zieltextes unterhalb des ausgangsprachlichen Segments in Klammer angezeigt (vgl. Abbildung 18). Die Aktivierung der Funktion *Textfeld-Aufnahme* erlaubt es, die Sprachaufnahme durch Tippen auf das weiße Textfeld anstelle des Mikrophonbuttons zu starten. *Signal-Töne bei Aufnahme* bedeutet, dass Signaltöne den Beginn und das Ende einer Sprachaufnahme markieren. Die Funktion *Kraftausdrücke erlauben* erklärt sich von selbst und ist interessanterweise nur in der Offlinenutzung auswählbar. Ist diese Funktion deaktiviert, werden Kraftaudrücke in der Transkription durch *<beep>* ersetzt. Wird die Funktion Sprachsynthese deaktiviert, wird der übersetzte Zieltext nur auf dem Bildschirm angezeigt und nicht sprachlich ausgegeben. Eine Aktivierung der Funktion *Aufnahme abspielen* bietet den UserInnen die Möglichkeit, eine Aufnahme des selbst gesprochenen Ausgangstextes bei Bedarf vorzuspielen. Ist die Funktion deaktiviert, wird bei Betätigung des Lautsprecherbuttons lediglich eine per Sprachsynthese generierte Sprachausgabe des AT wiedergegeben. Diese Funktion ist in der online Version nicht wählbar und in der offline Version standardmäßig als einzige Funktion deaktiviert. Wird die Funktion *Angeschlossene Audiogeräte deaktivieren* eingeschalten, erfolgt die Sprachausgabe trotz eventueller angeschlossener Kopfhörer oder externer Lautsprecher durch die internen Lautsprecher des Smartphones. Die auf lange Frist für die AnwenderInnen wohl nützlichste Funktion ist die *Adaption*. Sie bezeichnet die Fähigkeit der Software, sich an die Individualitäten der SprecherInnen anzupassen und mit ihnen zu „lernen“. Inwiefern dies erfolgt und worin dies sich genau äußert, konnte trotz intensiver

Recherche nicht herausgefunden werden. Jedoch ist es erwähnenswert, dass die Adaptionfunktion nur im Offline-Modus wählbar ist, was bedeuten würde, dass die Daten und Aufzeichnungen auf dem Gerät gespeichert werden und somit nicht über Server laufen und in weiterer Folge an *Jibbigo* gesandt werden. Gleichzeitig ermöglicht der letzte Punkt der Einstellungen *Offline-Datennutzung* das Miteinbeziehen von lokal gespeicherten Daten aus dem Offline-Modus während des Online-Moduses, was bedeuten würde, dass die vermeintlich lokalen Daten schließlich doch bei *Jibbigo* landen. Wie genau der Datenabgleich funktioniert und ob und welche Daten versendet werden, ist mit den aktuell recht spärlichen Informationsangebot der Website nicht erruierbar. Ein User's Manual (vgl. Jibbigo o.J.) einer frühen *Jibbigo* Version beschreibt die Adaptionfunktion wie folgt:

„It's important to keep the adaption function on when the same person is using Jibbigo. Since much of this function resets when you exit the program, begin each session by saying 4 or 5 „training“ sentences before you start using the translator in earnest. If more than one person is using the same device for the same language (i.e. then adaptation should be set to „off“.“ (Jibbigo o.J.: 5)

Im Dokument ist zwar kein Jahr angegeben, jedoch finden sich in der Kompatibilitätstabelle (vgl. Jibbigo o.J.: 2) das iPhone 4s und das iPad 2 als aktuellste Geräte, sodass anzunehmen ist, dass das Manual zwischen 2011 und 2012 erschienen ist. Es muss sich demnach um eine frühe Version von *Jibbigo* handeln, die in diesem Manual beschrieben wird.



Abbildung 18: Screenshot Rückübersetzung Jibbigo 2.5.5

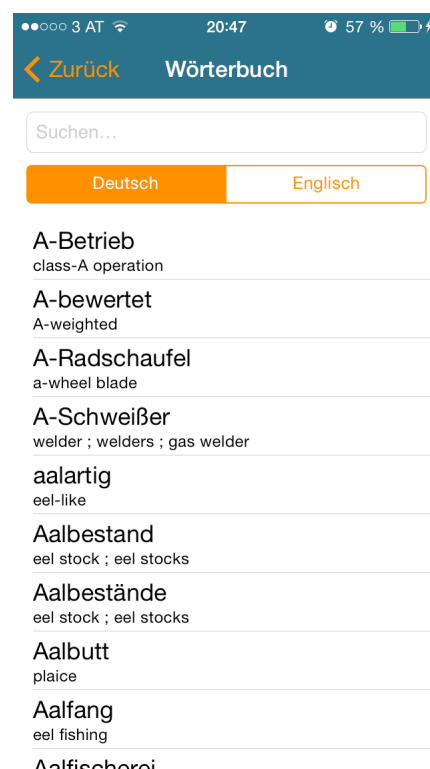


Abbildung 19: Screenshot Wörterbuch Jibbigo 2.5.5

Wie bereits erwähnt, erfährt der Übersetzer-Modus optisch die größte Veränderung, da zwei zusätzliche Icons hinzukommen (vgl. Abbildung 18). Rechts findet sich eine Art Lexikon, welche nach dem Antippen Einsicht in das heruntergeladene und nun lokal gespeicherte Wörterbuch bietet (vgl. Abbildung 19). Wird ein Eintrag angetippt, wird das ziel-sprachliche Segment per Sprachsynthese wiedergegeben. Das zweite Icon ist ein Pluszeichen, welches eng mit der Adaptionfunktion verbunden ist. Es bietet die Möglichkeit dem System bestimmte (wiederkehrende) Namen manuell „beizubringen“, indem zunächst ausgewählt wird um welche Wort-Kategorie es sich handelt (Vorname, Nachname, Ortsname) und in einem weiteren Schritt, der Name manuell eingegeben und seine Aussprache angepasst wird (vgl. Abbildung 20). Die manuell hinzugefügten Namen werden ebenso wie der Übersetzungsverlauf sprachpaarbezogen gespeichert und sind nur für die ausgewählte Sprachkombination verfügbar.

Während der Anwendung machte sich bzgl. der manuellen Namenseingabe ein Bug bemerkbar. So erscheint das Pluszeichen wie bereits gesagt nur bei Sprachkombinationen, die über ein installiertes Offline-Paket verfügen. Bei jenen Sprachen, die nur einen Online-Zugriff haben, erscheint es nicht (vgl. Abbildung 11 oder 12). Wurde der Übersetzer-Modus jedoch in einer Offline-Version einmal gestartet (auch ohne eine Übersetzung zu veranlassen), erscheint das Pluszeichen auch in den Sprachkombinationen mit ausschließ-lichem Online-Zugriff. Versucht der/die Anwender/in einen manuellen Namen einzufügen (vgl. Abbildung 20) und tippt anschließend auf *speichern*, bleibt die daraufhin angezeigte Liste mit den manuell gespeicherten Namen leer.

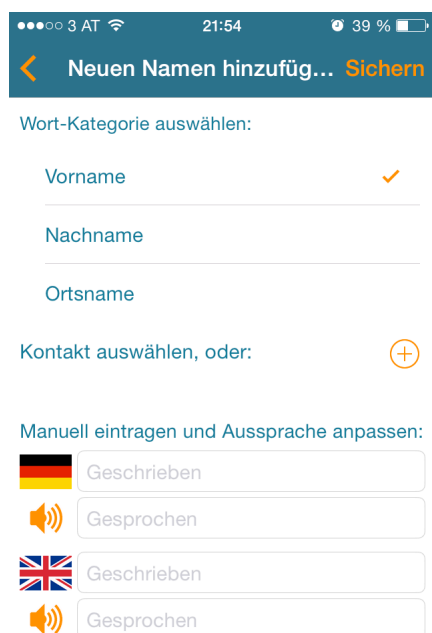


Abbildung 20: Screenshot Manuell Namen hinzufügen Jibbig 2.5.5



Abbildung 21: Screenshot Romanisierung der chinesischen Schriftzeichen Jibbig 2.5.5

Die Funktion *Offline-Datennutzung* war hierbei durchgehend aktiviert. Wird die App geschlossen und wieder neu gestartet, ist das Pluszeichen bei den online Sprachpaketen wieder verschwunden und es zeigt sich das anfängliche Bild wie bei Abbildung 11 oder 12.

Ein weiterer Bug zeigte sich erst nach tagelanger Anwendung der App. Wird die App geöffnet erscheinen alle Einstellungen des Übersetzer-Modus genau wie in Abbildung 14 mit der *Offline-Datennutzung* als letzter (de)aktivierbarer Option. Nachdem bei einem Test der App Deutsch-Chinesisch als Sprachkombination ausgewählt wurde, um die Darstellung der Zeichen genauer zu untersuchen, erschien in den Einstellungen eine weitere Option als letzter Punkt. Die Option *Romanisierung anzeigen* ist standardmäßig deaktiviert und erlaubt eine Umsetzung in lateinische Buchstaben, um damit die Aussprache der chinesischen Schriftzeichen zu verdeutlichen (vgl. Abbildung 21). Wechselt der/die Anwender/in in eine andere Sprachkombination, wird die neue Option nun auch in den Einstellungen aufgelistet, auch wenn in der gewählten Sprachkombi gar kein Bedarf für eine Romanisierung bestehen sollte. Wird die App geschlossen und wieder neu gestartet, ist die Option ebenso wie im oberen Bug nicht mehr vorhanden.

Ebenso stellte sich erst nach einer Vielzahl von Anwendungen heraus, dass bereits übersetzte Elemente und ihre entsprechenden Ausgangstexte durch einfaches Schütteln des iPhones gelöscht werden können.

In diesem Zusammenhang soll auch noch erwähnt werden, dass die für die einzelnen Sprachen nötigen Diakritika erst nutzbar sind, wenn in den allgemeinen Einstellungen des iOS die gewünschte Tastatur für die entsprechende Sprache aktiviert wurde¹⁴.

Der Übersetzungsprozess per se läuft sowohl im Online- als auch im Offline-Modus flüssig ab, wobei bemerkbar war, dass in der Offline-Anwendung sowohl die Ladezeit des Übersetzer-Modus an sich, als auch der konkrete Übersetzungsprozess minimal länger dauerte. Während der Übersetzer-Modus online nach dem Tippen auf den Button *Übersetzer starten* beinahe ohne merkbare Verzögerung startet und vergleichbar mit der durchschnittlichen Reaktionszeit von anderen (auch nicht Translations-)Apps ist, benötigt der Übersetzer-Modus in der Offline-Variante ungefähr 2 Sekunden Ladezeit. Auch die Dauer der Übersetzung gestaltet sich offline etwas länger. Um die Transkription des AT anzuzeigen, braucht die App (abhängig von der AT-Länge) online 1-2 Sekunden und offline ungefähr 2 Sekunden. Im weiteren Schritt dauert es bis zur Anzeige und Sprachausgabe der fertigen Übersetzung in der ZS online weitere 2-3 Sekunden und offline 3-4 Sekunden. Offensichtlich dauert die Online-Übersetzung mit einer durchschnittlich schnellen Internetverbindung und die Bearbeitung des AT auf den Servern weniger lang als in der

¹⁴ Beim aktuellen iOS 7.1.2 findet sich diese Funktion unter *Einstellungen > Allgemein > Tastatur > Tastaturen > Tastatur hinzufügen*.

Offline-Version, bei der die für die Übersetzung notwendigen Daten lokal auf dem Gerät gespeichert sind. Die Dauer des gesamten Übersetzungsprozesses hält sich in beiden Varianten in einem akzeptablen Rahmen, sodass damit zügig gearbeitet werden kann. Die Sprachausgabe ist verständlich, jedoch verbesserungswürdig, da die Intonation und teilweise auch die Aussprache sehr unnatürlich wirken.

5.3.3 Besonderheiten

Die Besonderheit an *Jibbig* liegt eindeutig in der Möglichkeit der Offline-Anwendung. Die App wurde damals mit der Idee auf den Markt gebracht, vor allem in Bereichen eingesetzt zu werden, wo der Zugang zum Internet eingeschränkt ist. Dazu gehören beispielsweise diverse touristische Szenarien wie Besuche im Ausland und vor allem der Einsatz in humanitären Aktionen wie in Thailand oder Kambodscha nach Umweltkatastrophen oder zur allgemeinen medizinischen Behandlung (vgl. Waibel 2012c).

Hinzu kommt, dass die App in der Offline-Version die Möglichkeit der Adaption bietet und sich somit an die Individualitäten diverser Sprecher und Sprecherinnen anpassen kann, was sich letztendlich auf die Qualität der Übersetzung positiv auswirkt. Ebenso gibt es die Möglichkeit wiederkehrende Namen wie Vornamen, Nachnamen oder Ortsnamen manuell zu speichern, sodass diese in den kommenden Übersetzungen leichter erkannt und richtig übersetzt werden können.

Die App und ihre Nutzung ist komplett kostenfrei¹⁵. Einerseits kann sie gratis aus dem App Store bzw. Google Play Store heruntergeladen und mit Internetzugang online genutzt werden. Andererseits ist auch der Download der Offline-Sprachpakete gratis und bedarf lediglich einer einmaligen Internetverbindung. Abgesehen davon, kommt die App bei Bedarf in weiterer Folge komplett ohne Internet aus.

5.4 Benutzerfreundlichkeit

Anpassbarkeit - Die individuelle Anpassbarkeit der App wurde bereits im vorherigen Kapitel angeschnitten. Abgesehen davon, dass AnwenderInnen die Wahl zwischen Online- und Offline-Nutzung haben, können sie auch die Einstellungen des Übersetzer-Modus nach individuellen Vorlieben anpassen. Den UserInnen wird also ein gewisser Grad an Individualität ermöglicht, der jedoch, abhängig davon ob der Online- oder Offline-Modus gewählt wurde, unterschiedlich hoch ausfällt. Abgesehen davon wäre die Anpassbarkeit

¹⁵ Eventuell anfallende Kosten für das Datenvolumen werden hierbei nicht beachtet, da sie nicht unmittelbar durch die App aufgezwungen werden, wie z.B. die Kosten pro Sprachpaket in den früheren *Jibbig* Versionen.

noch erweiterbar, beispielsweise durch die freie Wahl zwischen einer männlichen oder weiblichen Ausgabestimme oder der Möglichkeit den Online-Modus nur bei vorhandener W-LAN Verbindung zu nutzen, um eventuelle Kosten zu vermeiden und das mobile Datenvolumen zu schonen.

Erlernbarkeit - Der Aufbau der App ist übersichtlich gestaltet und enthält nicht zu viele Unterkategorisierungen und Menüpunkte, sodass die Navigation durch die App nicht unnötig erschwert wird. Da *Jibbig* ursprünglich für das iPhone konzipiert wurde, erinnert auch das Layout der App an andere iPhone-typische Apps und ist somit für iOS-Kenner leicht verständlich. Selbst nicht technikaffine Menschen sollten durch die intuitiv gestalteten Icons wie das Mikrophon oder den Lautsprecher rasch die Grundfunktionen der App verstehen und nutzen können.

Verlässlichkeit/Robustheit - Allgemein kann gesagt werden, dass *Jibbig* sowohl in der Online- als auch in der Offline-Version zuverlässig gearbeitet hat. Bis auf geringe Unterschiede in der Ladedauer, haben beide Versionen ohne Störungen oder Abstürze Übersetzungen erzeugt. Die App wirkte auch bei schlechter Internetverbindung, welche sich lediglich in längeren Ladeprozessen äußerte, sehr robust.

5.5 SWOT-Analyse

Stärken

- Individualität – viele (de)aktivierbare Einstellungen
- Manuelles Hinzufügen von häufig verwendeten Namen
- Adaption an SprecherInnen
- Kurze Ladezeiten
- Robustheit - kein Absturz oder sonstige Störung
- Download der App gratis
- Sprachpakete auch gratis

Schwächen

- Ladezeiten im Offline-Modus länger
- Rückübersetzung nur im Offline-Modus
- Schwierig passende Transkription für Aussprache der manuell eingegebenen Namen zu finden und bedarf langem Ausprobieren
- Mangelndes Informationsangebot seit Aufkauf durch Facebook
- Eingeschränkter Download der Sprachpakete durch Speicherkapazität des Smartphones
- Einschränkungen in der Wahl der Sprachkombination

Chancen

- Rückübersetzung auch für Online-Modus ermöglichen
- Wahlweise Online-Modus nur bei W-LAN Verbindung starten, um Datenverbrauch und eventuelle Kosten zu vermeiden
- Wahl zwischen männlicher und weiblicher Ausgabestimme
- Natürlichkeit der Ausgabestimme verbessern

Risiken

- Hoher Datenverbrauch im Online-Modus
- Keine eindeutige Transparenz bei Datennutzung
- Kursänderung nach Aufkauf durch Facebook
- Datenschutz

6 VoiceTra4U 3.2

Die App *VoiceTra4U* wurde vom U-STAR Consortium (Universal Speech Translation Advanced Research) entwickelt. Es handelt sich dabei um ein internationales Forschungskonsortium, welches sich auf Speech-to-Speech Translation konzentriert und aus dem Zusammenschluss asiatischer Forschungsgruppen A-STAR (Asian Speech Translation Advanced Research) entstanden ist (vgl. U-STAR 2014a).

„There are countless numbers of languages and dialects used around the world and overcoming the barriers between them has been a long-held dream for humankind. However over the years, research and studies on Speech-to-speech translation (S2ST) technologies have made it possible to turn this into a clear goal rather than just a dream. It is not an easy task for an individual to establish S2ST systems that cover all topics and languages, however, by connecting ASR, MT, and TTS modules distributed from organizations around the world, we strongly believe that we can establish a global S2ST system that will allow us to free ourselves from these language barriers.“ (U-STAR 2014a)

Die A-STAR hat am 29. Juli 2009 das erste netzwerkbasierte Speech-to-Speech Translation System für asiatische Sprachen eingeführt. „The system enabled real-time, location-free, multi-party communication between speakers using different Asian languages, and confirmed the feasibility of network-based S2ST protocol.“ (U-STAR 2014a) Im Jahr 2010 wurde A-STAR in U-STAR umbenannt und weitete ihren Forschungsbereich auch auf nicht asiatische Sprachen aus, um mithilfe internationaler Institutionen und Forschungsstätten ein globales Übersetzungssystem zu schaffen. Am 27. Juni 2012 wurde die *VoiceTra4U* App der Öffentlichkeit vorgestellt, welche zum damaligen Zeitpunkt mit 27 Sprachen (17 davon mit Sprachausgabe) jene Sprachen abdeckte, die von 97 % der Weltbevölkerung gesprochen wurden.

„U-STAR, currently comprised of 26 institutes from 23 countries, has been conducting an ongoing research on speech translation. U-STAR and its members have collaboratively developed the multilingual speech translation system to provide the translation service via publicly-released client application, by connecting the servers of U-STAR member institutes mutually, with the communication protocol which was implemented based on the ITU-T recommendations F.745 and H.625¹⁶. [...] We hope to see people use “VoiceTra4U-M” in various occasions when translation and interpretation are requested at the London Olympic games in July, and that “VoiceTra4U-M” may come in handy to many people wishing to overcome the language barriers.“ (U-STAR 2012: 1)¹⁷

Die App war insofern innovativ, als versucht wurde mittels eines netzwerkbasierten Systems und einer Aufteilung des technischen Aufwandes auf die verschiedenen Mitgliedsinstitutionen in diversen Ländern ein global funktionierendes SST System zu kreieren. Jedes teilnehmende Institut betreibt Server, die jeweils für die Spracherkennung, maschinelle Übersetzung und Sprachsynthese zuständig sind (vgl. Abbildung 22). Sobald die AnwenderInnen die gewünschten Sprachen ausgewählt haben, sind sie über das Netzwerk mit dem vom japanischen NICT (National Institute of Information and Communications Tech-

¹⁶ Es handelt sich hierbei um Richtlinien, welche die Architektur und netzwerkbasierten SST Systems sowie detaillierte Definitionen der Systemkomponenten und Protokolle beschreiben. Näheres in U-STAR 2012.

¹⁷ Mit dem Update auf die Version 2.3 am 18.10.2013 wurde auch der Name der App von *VoiceTra4U-M* in *VoiceTra4U* geändert (vgl. U-STAR 2013).

nology) betriebenen Kontrollserver verbunden, der wiederum mit den entsprechenden Sprachservern der einzelnen Institutionen Verbindung aufnimmt und somit die Übersetzung veranlasst (vgl. U-STAR 2012: 3).

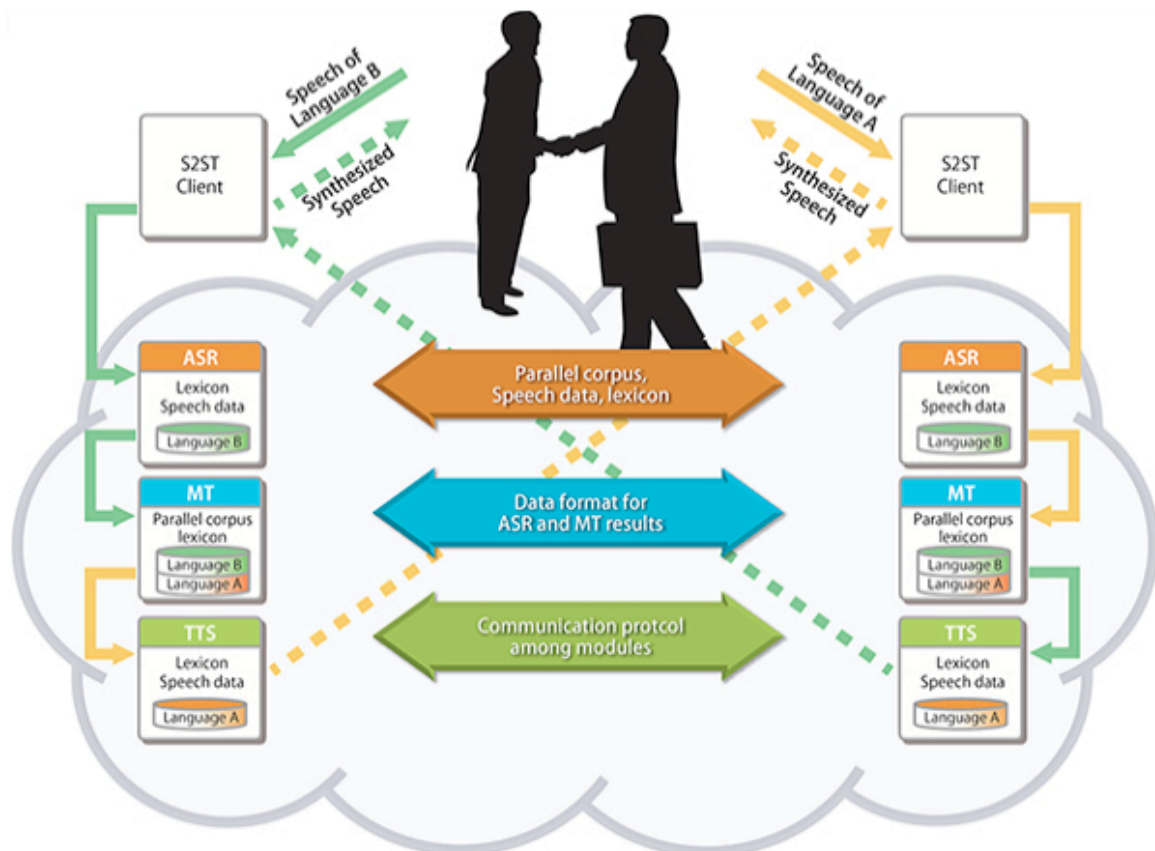


Abbildung 22: Netzwerkbasierter Übersetzungsprozess (vgl. U-STAR 2014b)

Im Folgenden werden die Einzelheiten der App, wie in Kapitel 4 vorgesehen, vorgestellt und es soll an dieser Stelle auch darauf hingewiesen werden, dass sich die folgenden Daten und Fakten auf die aktuell erhältliche Version (Stand: 30.07.2014) beziehen und daher eine vergängliche Gültigkeitsdauer haben.

6.1 Technische Hintergründe

Die App *VoiceTra4U* ist aktuell nur im Apple App Store erhältlich und erfordert iOS 6.0 oder eine noch aktuellere Version des Betriebssystems. Sie ist mit iPhone, iPad und iPod touch kompatibel, jedoch für das iPhone 5 optimiert. Ebenso wie *Jibbigo* ist die App kostenlos erhältlich und verfügt aktuell über keine Wertungen im App Store. Die aktuelle Version der App 3.2 hat eine Größe von 15,5 MB und wird in der Kategorie Reisen geführt. Sie wird auch vom Entwickler U-STAR eindeutig als App für reisebezogene Konversatio-

nen in Flughäfen, Hotels, Geschäften und Restaurants beworben, die nicht nur sprachliche, sondern auch anderweitige Barrieren zu überwinden versucht. So könne durch Sprachsynthese auch mit Blinden kommuniziert werden, während die Chatfunktion Gespräche mit hörgeschädigten Personen ermöglicht (vgl. <https://itunes.apple.com>^c 2014). Laut der U-STAR Website (2014c) ist die App sowohl in englischer als auch in japanischer Sprache erhältlich.

Am 15.08.2014 erschien erstmalig eine Android Version der App – VoiceTra4U 1.0. Die App ist 7,5 MB groß und erfordert die Android-Version 2.2 oder höher. Laut den statistischen Angaben im Google PlayStore belief sich die Anzahl der Installationen bis zum Abschluss dieser Arbeit am 20.08.14 in der Größenordnung von 10 bis 50 (vgl. <https://play.google.com>^b 2014). Nach den Screenshots im Google Play Store und der U-STAR Website (U-STAR 2014c) zu urteilen, besitzt die Android App ein stark abweichendes Layout und User Interface von jenem für iOS Geräte.

Nachtrag: Am 15.08.2014 erschien das Update 3.2.1 für iOS Geräte. Laut den Angaben im App Store wurde die Liste der nutzbaren Sprachen auf den neuesten Stand gebracht (vgl. <https://itunes.apple.com>^c 2014). Die Neuerungen konnten nicht genauer untersucht werden, da unmittelbar nach dem Download des Updates die App nicht mehr ordnungsgemäß geöffnet werden konnte, und nur der Startbildschirm mit dem U-STAR Logo angezeigt wurde. U-STAR ist der Bug bekannt und es wird laut eigenen Angaben bereits an einem neuen Update gearbeitet, um das Problem zu lösen:

„There has been a bug found in ‚VoiceTra4U for iOS ver 3.2.1‘ where the app freezes at the splash screen the first time you launch the app. Please close the app by pressing the ‚Home Button‘ on your device and restart the app. We are working diligently to fix this issue and the updated version should be released very soon. We apologize for the inconvenience and thank you for your cooperation.“ (U-STAR 2014c).

Bis zum Abschluss dieser Arbeit (20.08.2014) ist kein neues Update erschienen.

6.2 User Interface

Nachdem die App gestartet wurde, erscheint das runde U-STAR Logo und auch kurzzeitig ein kleines Fenster mit dem Text *„Verifying Server Connection. Connection to server is being verified...“*, welches den 2-3 Sekunden andauernden Verbindungsprozess zu einem der vielen Server überbrücken soll. Danach erscheint das Hauptmenü, das auch eine optische Zweiteilung aufweist. Das User Interface ist in Englisch gehalten und die App startet standardmäßig im sogenannten Single Mode (vgl. U-STAR 2014c), in dem ebenso wie bei *Jibbiggo*, per Sprach- oder Texteingabe eine maschinelle Übersetzung direkt auf dem Gerät

generiert werden kann. Mithilfe von Registerkarten wird optisch eine Zweiteilung des Bildschirmes hervorgerufen, wobei die Registerkarte der aktuell ausgewählten Ausgangssprache im Vordergrund erscheint (vgl. Abbildung 23). Zwischen den zwei Flaggen findet sich das runde U-STAR Logo, welches nach dem Antippen ein neues Fenster mit näheren Informationen zur App und zum Entwickler enthält (vgl. Abbildung 24).

Die AnwenderInnen haben die Möglichkeit über den Button *About this app* auf die U-STAR (2014c) Website zu gelangen, wo die Anwendung der App Schritt für Schritt beschrieben wird. Der zweite Button *About U-STAR* führt auch auf die U-STAR Website, jedoch zur allgemeinen Startseite. Beim Anklicken des Buttons *Tell a friend* wird die iOS typische E-Mail Maske geöffnet und es erscheint der Standardtext „*I found an application called VoiceTra4U (free) that lets you do multilingual voice chat*“, sowie ein Link zu iTunes. Beim letzten Button *Contact us* besteht die Möglichkeit, direkt an das U-STAR Consortium eine E-Mail zu senden. Außerdem ist in den Informationen auch die aktuelle Version der App zu erkennen und ebenso mit welchem Server die AnwenderInnen gerade verbunden sind. Bei der Beispielskombination Deutsch-Englisch wird eine Verbindung zum Server France-01 hergestellt (vgl. Abb. 24).

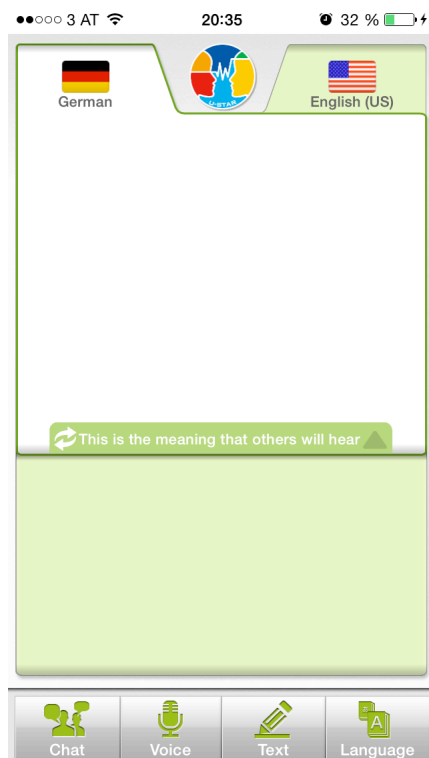


Abbildung 23: Screenshot Hauptmenü Single Mode VoiceTra4U 3.2

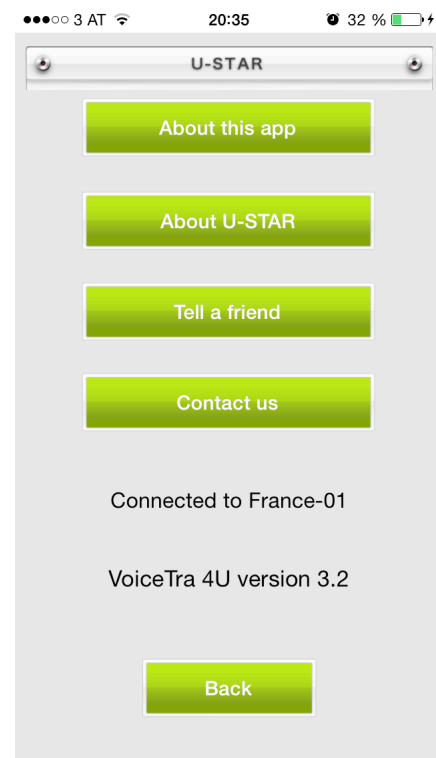


Abbildung 24: Screenshot U-STAR Informationen VoiceTra4U

Des Weiteren besteht das Hauptmenü (vgl. Abbildung 23) aus vier weiteren Buttons, die sich am unteren Ende des Bildschirms befinden. Ganz rechts findet sich das Icon *Language*, welches die AnwenderInnen in das Sprachauswahlmenü führt (vgl. Abbildung 25).

Zunächst erscheinen auf dem Bildschirm alle Sprachen der gewählten Kategorie mit den entsprechenden Flaggen zur besseren Illustration. Ganz oben in den Registerkarten werden die ausgewählte Ausgangs- und Zielsprache, sowie der Funktionsumfang der jeweiligen Sprache angezeigt. So besitzt das Englische beispielsweise sowohl eine Spracherkennungs- als auch eine Sprachsynthesefunktion, während beim Deutschen nur die Spracherkennung verfügbar ist und die fehlende Sprachsynthesefunktion mittels eines ausgegrauten Lautsprechers dargestellt wird. Unterhalb der Registerkarten wird 1-2 Sekunden lang eine kurze Information zur Auswahl der Sprachen eingeblendet (siehe Abb. 25) und wieder ausgeblendet (siehe Abb. 26). Die gewünschten Sprachen sollen mittels Drag-and-Drop-Funktion aus der Mitte des Bildschirms in die Registerkarten nach oben bewegt werden. Unterhalb der Anweisung zur Sprachauswahl befinden sich zwei Buttons, die je nach Aktivierung entweder die Auswahl der asiatischen Sprachen oder andere Weltsprachen darstellten. In Abb. 25 und Abb. 26 werden jeweils die Weltsprachen angezeigt.



Abbildung 25: Screenshot Allgemeine Auswahl der Sprachen VoiceTra4U 3.2



Abbildung 26: Screenshot Auswahl der Sprachen mit Sprachausgabe VoiceTra4U 3.2

Standardmäßig werden alle Sprachen angezeigt, wobei den AnwenderInnen auch eine Filtermöglichkeit zur Verfügung steht, mit Hilfe derer durch Antippen des Mikrophons oder des Lautsprechers unten jene Sprachen ausgeblendet werden, die nicht über die gewählte Funktion verfügen. Abbildung 26 zeigt beispielsweise, durch Aktivierung des Filters für Sprachausgabe (Lautsprecher erscheint grün), welche der Weltsprachen über eine Sprachausgabefunktion verfügen. Durch die an Windows erinnernden Buttons *OK* und *Cancel*

kann die Sprachauswahl bestätigt bzw. abgebrochen werden und die UserInnen werden wieder zurück zum Hauptmenü geführt.

Ausgehend vom Hauptmenü kann nach erfolgreicher Sprachauswahl auch der maschinelle Übersetzungsprozess gestartet werden. Im standardmäßigen Single Mode kann dies entweder mittels reiner Texteingabe durch Antippen des Buttons *Text* (siehe Abbildung 23), oder durch Spracherkennung (Button *Voice*) erfolgen. Je nach Funktionsumfang der gewählten Sprachen stehen den AnwenderInnen beide oder jeweils nur eine Funktion zur Verfügung. Entscheiden sich die UserInnen für die manuelle Texteingabe, verändert sich der Bildschirm nicht, sondern es wird lediglich am unteren Rand des Bildschirms die Tastatur eingeblendet und nach der Bestätigung des Textes wird dieser in der entsprechenden ausgangssprachlichen Registerkarte angezeigt und daraufhin übersetzt bzw. bei Vorhandensein der Funktion Sprachsynthese auch akustisch ausgegeben.

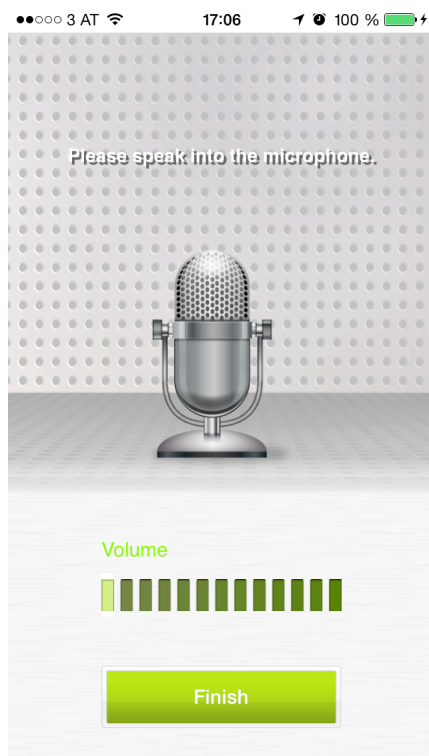


Abbildung 27: Screenshot Sprachaufnahme VoiceTra4U 3.2



Abbildung 28: Screenshot Useranmeldung Chat Mode VoiceTra4U 3.2

Entscheiden sich die AnwenderInnen für eine Spracheingabe mittels Stimme, wird beim Antippen des Icons *Voice* der Sprachaufnahme-Bildschirm dargestellt (vgl. Abbildung 27) und nach dem Ende der Aufnahme erscheint zunächst der transkribierte Ausgangstext im Hauptmenü, unmittelbar danach auch die Übersetzung und Rückübersetzung des Zieltextes in die Ausgangssprache. Auf die Einzelheiten des Übersetzungsprozesses wird im folgenden Kapitel genauer eingegangen.

Der letzte wählbare Menüpunkt im Hauptmenü ist der *Chat Mode* (vgl. Abbildung 23). Wird dieser ausgewählt, öffnet sich zunächst eine Anmeldemaske (vgl. Abbildung 28), in welchem die AnwenderInnen bestimmte Informationen angeben müssen, um den Modus nutzen zu können. Dazu gehören: Der User Name, unter welchem man von den anderen AnwenderInnen gefunden werden kann; ein Foto; die gewünschte Chat-Sprache; die Muttersprache, sowie die erste und zweite Fremdsprache; das Geschlecht; das Alter und ein optionaler Filtercode. Der Filtercode ist eine vierstellige Zahlenkombination, die dazu dient, nur von UserInnen gefunden werden zu können, welche über den persönlichen Code in Kenntnis gesetzt wurden. Die Funktionsweise des *Chat Mode* wird im folgenden Kapitel gesondert behandelt.

Allgemein kann gesagt werden, dass sich das Layout und der Menüaufbau bei *VoiceTra4U* stark von *Jibbig* unterscheidet. Während *Jibbig* eine starke Orientierung an der iOS Struktur¹⁸ und Architektur von Apps aufweist, erinnert *VoiceTra4U* durch die Registerkarten, oder die *OK* und *Cancel* Buttons eher an ein windowsinspiriertes Layout. Die *VoiceTra4U* App ist nichtsdestotrotz ohne bedeutende Vorkenntnisse leicht bedienbar und weist durch die begrenzte Zahl an Menüpunkten im Hauptmenü auch eine relativ einfache Architektur auf, sodass sich AnwenderInnen leicht darin zurechtfinden.

6.3 Features

Bevor die Features eingehend behandelt werden, soll an dieser Stelle noch erwähnt werden, dass alle NutzerInnen der *VoiceTra4U* App sich automatisch durch die Nutzung der App bereit erklären die *Terms of Use* zu akzeptieren. NutzerInnen werden nach der Installation beim ersten Starten der App auf die *Terms of Use* hingewiesen und dazu aufgefordert diese zu akzeptieren, da die App sonst nicht in Betrieb genommen werden kann. Zusätzlich gibt es wie in Abbildung 24 ersichtlich, im Informationsmenü die Möglichkeit über den Button *About this app* auf die U-STAR Website (vgl. U-STAR 2014c) zu gelangen, in der die *Terms of Use* verlinkt werden.

„All data from speech and text input, as well as translation results (hereinafter referred to as "user data"), are recorded in the server of U-STAR. U-STAR uses user data with the aim of improving and developing speech translation technologies. U-STAR may provide the user data to third parties. U-STAR is subject to the following terms and conditions for using the user data. When U-STAR distributes the user data to third parties, we require the parties to observe the terms and conditions in accordance with the 'Agreement on grants of license'.“ (U-STAR 2014d)

¹⁸ Die iOS Human Interface Guidelines stellen eine Orientierungshilfe für App-Entwickler dar und beschreiben die Ansprüche, welche an eine für iOS 7 optimierte App gestellt werden. Dazu zählen unter anderem User Interface (UI) Design Basics wie Layout, Animationen, Typographie oder Icons und Graphiken; allgemeine Design Strategien; verbindliche UI Elemente; oder auch Vorgaben zu Icon und Image Design wie Größenvorgaben, welche sich an den Bildschirmgrößen der entsprechenden Endgeräte orientieren (vgl. Apple Inc. 2014).

Die Daten werden ebenso wie bei *Jibbig* anonymisiert für Forschungszwecke aufgezeichnet und verwendet. Bezüglich der Weitergabe an Dritte müssen jedoch folgende Punkte beachtet werden:

- „The user data should not be broadcasted through media outlets, including the internet.
- Individuals should not be identified by using personal information in the corpus. Any direct contact to persons by use of personal information is prohibited.
- The treatment of user data including speech input by underage persons is subject to the laws of countries, states and regions to which the third parties belong.
- It is prohibited for the third parties who are provided user data by U-STAR to distribute it second-hand to other parties.“ (U-STAR 2014d)

6.3.1 Sprachenangebot

Die App *VoiceTra4U* in der Version 3.2 wird damit beworben, dass 30 verschiedene Sprachen und Dialekte unterstützt werden, wobei der Fokus auf Sprachen des asiatischen Raums liegt. Eine Spracherkennungskomponente ist nur für 17 Sprachen verfügbar und lediglich 14 Sprachen besitzen eine Sprachausgabekomponente. Selbstverständlich kann der Ausgangstext in allen 30 Sprachen mittels Texteingabe eingegeben werden. Zu den unterstützten Sprachen gehören (vgl. U-STAR 2014e):

- | | | |
|------------------------|--------------|----------------------|
| • Arabic | • Hindi | • Polish |
| • Brazilian Portuguese | • Hungarian | • Portuguese |
| • Danish | • Indonesian | • Russian |
| • Dutch | • Italian | • Sinhala |
| • Dzongkha | • Japanese | • Spanish |
| • English (US) | • Korean | • Taiwanese Mandarin |
| • English (UK) | • Malay | • Thai |
| • Filipino | • Mandarin | • Turkish |
| • French | • Mongolian | • Urdu |
| • German | • Nepali | • Vietnamese |

Nach einem Vergleich des online angeführten Sprachangebots mit jenem direkt in der App, stellte sich heraus, dass in der App zusätzlich die Sprache Tagalog (ohne Sprachein- und

Sprachausgabe) zu finden ist, welche in der Liste auf der Website nicht angeführt wird. Demnach bietet *VoiceTra4U* Übersetzungen in insgesamt 31 Sprachen an. Bemerkenswert an der App ist, dass die Wahl der möglichen Sprachkombinationen nicht wie bei *Jibbig* eingeschränkt ist, sondern jede Ausgangssprache mit jeder beliebigen Zielsprache kombinierbar ist. Hinzu kommen als Einschränkung lediglich die nicht für alle Sprachen verfügbaren Spracherkennungs- und Sprachausgabekomponenten.

6.3.2 Übersetzer-Modus

Die Beschreibung des Übersetzer-Modus bedarf bei *VoiceTra4U* einer Zweiteilung, da die App sowohl über einen *Single Mode* verfügt, der bidirektionale Übersetzungen auf einem Gerät für zwei Gesprächspartner produziert (ähnlich wie *Jibbig*), als auch über einen *Chat Mode*, in dem sich bis zu fünf Gesprächspartner an fünf unterschiedlichen Smartphones am Chat beteiligen können (vgl. U-STAR 2014c).

Wie bereits erwähnt, ist standardmäßig auf dem Hauptmenü der *Single Mode* aktiv. Um eine Übersetzung zu veranlassen, tippt der User oder die Userin auf den *Voice* Button (vorausgesetzt die Spracherkennungsfunktion wird für die gewählte Sprache angeboten), hört einen Signalton und gelangt nach 1-2 Sekunden zum Bildschirm der Sprachaufnahme (vgl. Abbildung 27). Besonders hervorzuheben ist, dass die Spracherkennungskomponente im Stande ist, das Ende des gesprochenen Ausgangstextes automatisch zu erkennen. Optional kann auch der oder die User/in durch Betätigen des *Finish*-Buttons das Ende der Sprachaufnahme markieren. Nach der Verarbeitung des Ausgangstextes durch die Spracherkennungskomponente erscheint der transkribierte Ausgangstext in der entsprechenden Sprachen-Registerkarte und unmittelbar danach wird auch weiter unten die Übersetzung des AT angeboten bzw. in der Mitte die Rückübersetzung des Zieltextes in die Ausgangssprache, welche vermutlich der Qualitätssicherung dienen soll (vgl. Abbildung 29¹⁹). Je nach Länge des Ausgangssegments dauert der Spracherkennungsprozess durchschnittlich 3-5 Sekunden, der maschinelle Übersetzungsprozess 3-5 Sekunden und der Rückübersetzungsprozess weniger als eine Sekunde. Die AnwenderInnen werden durch ein kleines, halbtransparentes Ladefenster über den aktuellen Prozess auf dem Laufenden gehalten. Der Informationstext im Fenster wechselt von *Recognizing* zu *Translating* und schließlich zu *Backtranslating*. Aufgrund der Serverbasiertheit kam es an einigen Tagen zu groben technischen Problemen. So wurde scheinbar eine Verbindung zum Server aufgebaut, jedoch konnten keine Übersetzungen produziert werden, da nach langen Wartezeiten (teilweise länger als 10 Sekunden) nur diverse Fehlermeldungen erschienen: „*Error speech recognition is not available now*“ oder „*Recognition complications Please try speaking once*“.

¹⁹ Der gesprochene Ausgangstext für die Beispielübersetzung lautete: „Wo befindet sich der nächstgelegene Bahnhof?“.

again“ oder „*Translation complications. Please try saying it again in a different way*“. In manchen Fällen konnte das Ladefenster (*Recognizing*) nicht geschlossen werden, sodass ein Neustart der App manuell erzwungen werden musste.

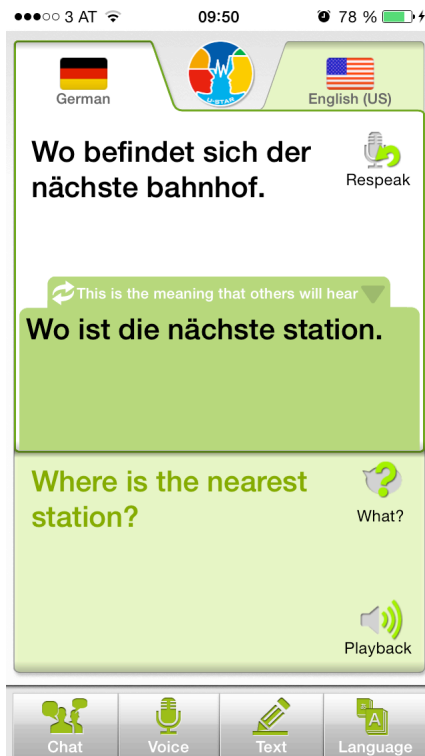


Abbildung 29: Screenshot Beispiel-
übersetzung Single Mode VoiceTra4U
3.2



Abbildung 30: Screenshot Beispiel-
übersetzung nicht verständlich Single
Mode VoiceTra4U 3.2

Die Rückübersetzung ist optional und kann durch Antippen des nach unten deutenden Pfeiles auch geschlossen werden bzw. ebenso bei Bedarf jederzeit wieder aktiviert werden. Neben dem Ausgangstext findet sich das Icon *Respeak*, welches den UserInnen die Möglichkeit gibt, das Ausgangssegment nochmals aufzusagen. In der Zieltext-Registerkarte befindet sich, abgesehen von der ZT-Wiedergabefunktion *Playback*, neben der Übersetzung ein Icon mit der Benennung *What*, das dem Zieltextpublikum eine Funktion bereitstellt, mithilfe derer sie ihr Unverständnis zum Ausdruck bringen können.

Tippen UserInnen das *What*-Icon an, wird der generierte Zieltext durchgestrichen und im AT-Feld wird mit roter Schrift der Text: „*I cannot understand. Please restate in a different way*“ angezeigt (vgl. Abbildung 30). Der oder die Sprecher/in des Ausgangstextes hat durch das bereits erwähnte *Respeak* Icon die Möglichkeit den Ausgangstext zu wiederholen, oder auch zu paraphrasieren und somit eine womöglich besser verständliche Übersetzung generieren zu lassen. Während diese Funktion per se eine Bereicherung für die App darstellt, da sich die Gesprächspartner damit auf sehr einfachem Weg mitteilen können, dass ein übersetztes Segment in der Zielsprache unverständlich ist, gestaltet sich die

Umsetzung dieser Idee etwas fraglich. Die EntwicklerInnen haben sich offensichtlich dazu entschlossen den rotgefärbten Text, unabhängig von der genutzten Sprachkombination, immer in englischer Sprache darzustellen. In Anbetracht der Tatsache, dass das vorrangige Ziel dieser SST Apps darin besteht die Kommunikation zwischen zwei (oder mehr) Personen mit unterschiedlichen Sprachen zu ermöglichen, ist es nicht sehr zielführend diese Information nur auf Englisch zu präsentieren. Dies wird besonders im Übersetzungsbeispiel in Abb. 29 bzw. 30 deutlich, da es sich um die Sprachkombination Deutsch-Englisch handelt und eine deutschsprachige Person die App in diesem hypothetischen Fall nutzen würde, weil sie der Zielsprache Englisch nicht mächtig ist. Den Informationstext dann auf Englisch anstatt Deutsch vorgesetzt zu bekommen, wirkt beinahe widersinnig. Logischer und vor allem zielführender wäre es, den Informationstext mittels maschineller Übersetzung in die jeweilige Ausgangssprache zu übersetzen, da die Funktion sonst in gewisser Weise ihren Nutzen verliert.

Ein anderes Manko des Single Modes ist die Tatsache, dass keine Korrekturmöglichkeit des von der Spracherkennung generierten Textes zur Verfügung steht. Den AnwenderInnen bleibt nur die Möglichkeit, den Text mittels der *Respeak*-Funktion zu wiederholen bzw. zu paraphrasieren oder den Text im Text-Menü manuell einzutippen.

Grundsätzlich läuft der Übersetzungsprozess in der *VoiceTra4U* App bei kurzen Segmenten flüssig und ohne Störungen oder Abstürze ab, jedoch dauert das Generieren der Übersetzung länger als bei *Jibbiggo*, was auch der Rückübersetzung geschuldet sein könnte. Je länger die Sätze werden, desto öfter kommt es zu Fehlermeldungen, dass die Spracheingabe nicht erkannt werden konnte oder der Übersetzungsdienst im Moment nicht verfügbar ist. Ein merkbarer Unterschied in der Qualität der Spracherkennungsergebnisse konnte bereits nach wenigen Anwendungen festgestellt werden. Die Ergebnisse der Spracherkennungskomponente sind bei *VoiceTra4U* auch ohne erschwerende Umstände wie Störgeräusche oder Selbstkorrekturen weniger akkurat als bei *Jibbiggo*. Eine mögliche Erklärung dafür könnte die Tatsache sein, dass einige der *VoiceTra4U* Spracherkennungssysteme auf Open Source Software basieren (vgl. U-STAR 2014e), während *Jibbiggo*, laut den direkt in der App aufgeführten Lizenzen, auf eigene Spracherkennungssysteme setzt.

6.3.3 Besonderheiten

Der Trumpf der *VoiceTra4U* App besteht sicherlich in ihrer Chatfunktion, dem sogenannten *Chat Mode*. Vor dem Wechsel vom Single Mode in den Chat Mode werden die UserInnen nochmals gefragt, ob sie auch wirklich wechseln wollen und sie müssen dies mittels eines *Yes* oder *No*-Buttons bestätigen. Um diesen zu starten, müssen die UserInnen lediglich im Anmeldebildschirm ihre Daten wie den User-Namen, das Alter und Geschlecht,

sowie die Sprachkenntnisse ausfüllen (vgl. Abb. 28). Nach erfolgter Anmeldung erscheint das Hauptmenü des *Chat Mode* (vgl. Abbildung 31), welches in seinem Aufbau, ebenso wie jenes des *Single Mode*, stark an Windows erinnert. Das runde U-STAR Logo führt, wie bereits im *Single Mode*, zum Informationsbildschirm (vgl. Abb. 24). Daneben finden sich zwei Suchfelder, in die jeweils der Name der zu suchenden Person und eventuell auch der Friends-Filter eingetragen wird, und ein *Search*-Button. Darunter werden die bereits gespeicherten Freunde gesammelt angezeigt.

In der unteren Menüleiste findet sich ganz rechts die Schaltfläche für den Wechsel in den *Single Mode* und ganz rechts eine *Config*-Schaltfläche, welche die UserInnen in das Anmeldeformular des *Chat Mode* bringt, um dort Änderungen an den zuvor angegebenen Daten vornehmen zu können. Die neben der *Config*-Schaltfläche angesiedelte *Delete*-Schaltfläche dient dazu, bereits gespeicherte Kontakte aus der Kontaktliste zu entfernen. Um den Chat zu starten, muss zunächst nach einer Kontaktperson gesucht werden. Dazu müssen der Name der Person und, falls vorhanden, der Friendsfilter in die oberen Suchleisten eingefügt und danach der *Search*-Buttons betätigt werden. Die Suche nach dem neu erstellten User gestaltete sich etwas schwierig, da die App 2-3 Minuten lang unter dem angegebenen Namen keinen User finden konnte und wiederholt eine Fehlermeldung erschien. Offensichtlich bedarf es einer kurzen Wartezeit bis neue User und Userinnen auf den Servern gefunden werden können.

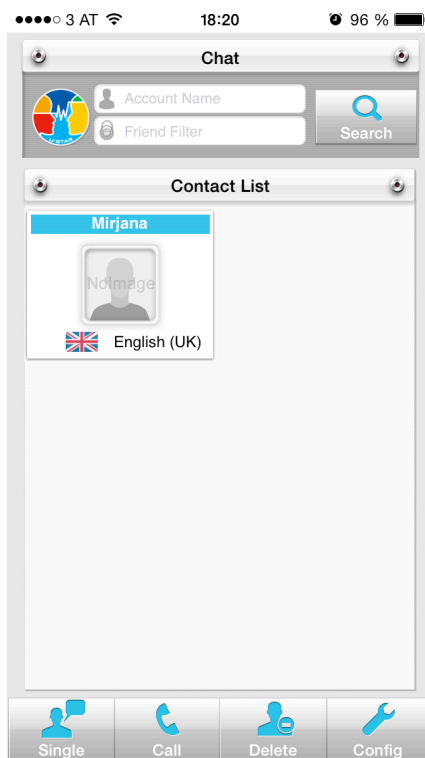


Abbildung 31: Screenshot Hauptmenü mit Kontakt im Chat Mode VoiceTra4U 3.2

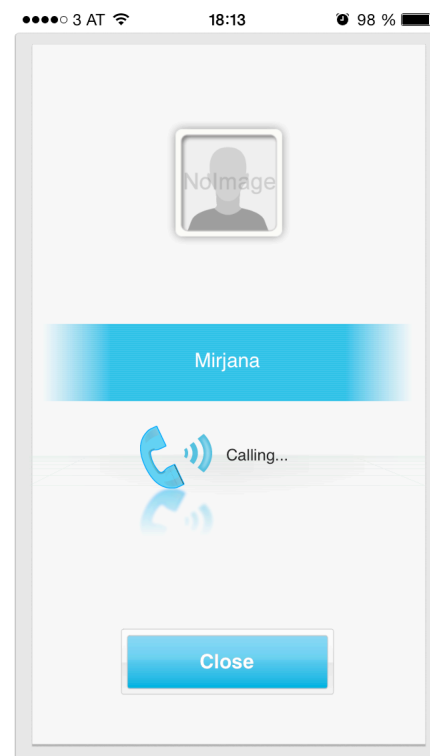


Abbildung 32: Screenshot Ausgehen-der Anruf Chat Mode VoiceTra4U 3.2

Der hinzugefügte Kontakt erscheint schließlich in der Kontaktliste mit seinem User-Namen, falls vorhanden einem Bild und der zuvor ausgewählten Chatsprache inklusive Flagge. Soll nun ein Gespräch mit einem der Kontakte geführt werden, muss dieser ausgewählt (weißer Hintergrund des Kontaktes wird blau und Kontakt erhält neben der Sprache ein blaues Häkchenfeld) und anschließend mittels der *Call*-Funktion kontaktiert werden (vgl. Abbildung 31).

Die anrufende Person sieht während des Gesprächsaufbaus den Bildschirm in Abbildung 32 und jene Person, die angerufen wird, kann auf dem Bildschirm (vgl. Abbildung 33²⁰) durch die Wahl der Tasten *Answer* bzw. *Decline* entscheiden, ob der Anruf entgegengenommen werden soll oder nicht. Nach der Annahme des Anrufes kann das Chatgespräch gestartet werden. Je nach Verfügbarkeit in der ausgewählten Sprache, kann der Ausgangstext per Spracherkennung (*Voice-Button*) oder Texteingabe (*Text-Button*) eingegeben werden. Der AT erscheint transkribiert als blaue Sprechblase, während der übersetzte Zieltext mit einer grauen Umrandung dargestellt wird (vgl. Abbildung 34).

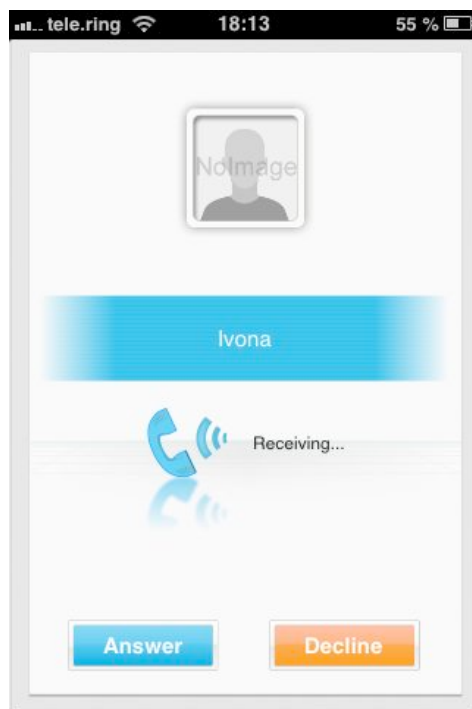


Abbildung 33: Screenshot Anruf entgegennehmen VoiceTra4U 3.2

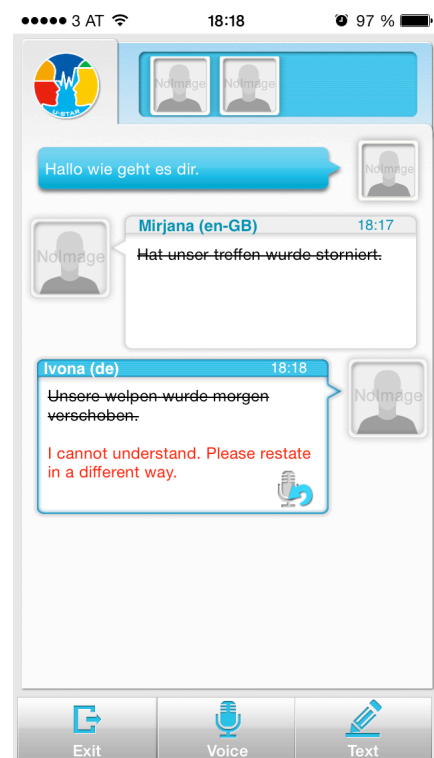


Abbildung 34: Screenshot Beispielchat 1 VoiceTra4U 3.2

Die Abbildungen 34 und 35 zeigen ein Beispielgespräch in dem die *What*-Funktion von jeder Gesprächsteilnehmerin je ein Mal genutzt wurde. Teilnehmerin Ivona erhielt um

²⁰ Der Screenshot wurde auf einem iPhone 3GS gemacht, welches eine im Vergleich zum iPhone 5 kleinere Bildschirmgröße aufweist. Daraus resultieren auch die abweichenden Maße des Screenshots und die veränderte Farbe der Statusleiste oben.

18:17 eine Nachricht von Mirjana, die lautete: *Hat unser treffen wurde storniert* und entschied sich den in der rechten, unteren Ecke befindlichen *What*-Button zu nutzen, woraufhin der empfangene Ziltext durchgestrichen wurde (vgl. Abb. 34). Innerhalb von 2-3 Sekunden wurde auch der transkribierte Ausgangstext auf Mirjanas Gerät durchgestrichen und es erschien rote Schriftzug, dass der Gesprächspartner nicht im Stande ist, die Übersetzung zu verstehen (vgl. Abb. 35). In diesem konkreten Fall stört es nicht, dass der Infotext auf Englisch präsentiert wird, da dies die gewählte Sprache der Teilnehmerin Mirjana war. Es gäbe nun die Möglichkeit mithilfe des *Respeak*-Buttons den Ausgangssatz neu aufzunehmen. Da es sich jedoch lediglich um ein Testbeispiel handelte und der englische Ausgangssatz von der Spracherkennungskomponente ohne Fehler verarbeitet wurde, wurde auf eine Wiederholung des Originals verzichtet. Demnach muss der Ursprung der schlechten Übersetzungsleistung in der maschinellen Übersetzungskomponente liegen. Warum der AT in Abb. 35 in Großbuchstaben dargestellt wird, konnte bis dato nicht herausgefunden werden. Dass dies nicht an der Ablehnung der Übersetzung durch den Gesprächspartner liegen kann, wird in Abb. 34 deutlich, da der AT aus der Nachricht um 18:18 auch abgelehnt und dennoch in Kleinbuchstaben dargestellt wurde.

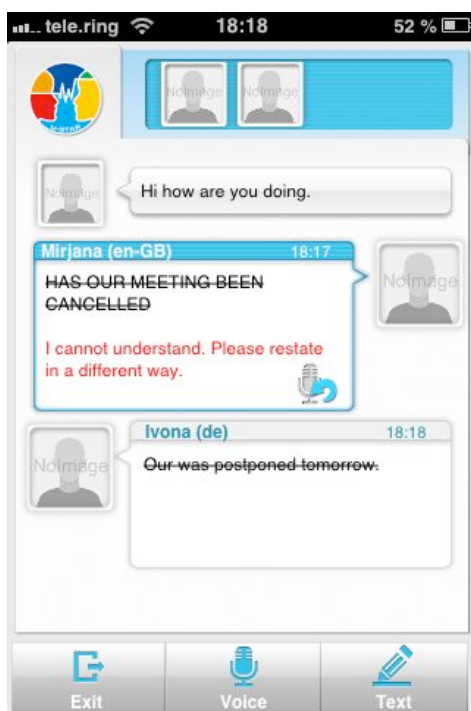


Abbildung 35: Screenshot Beispielchat 2
VoiceTra4U 3.2

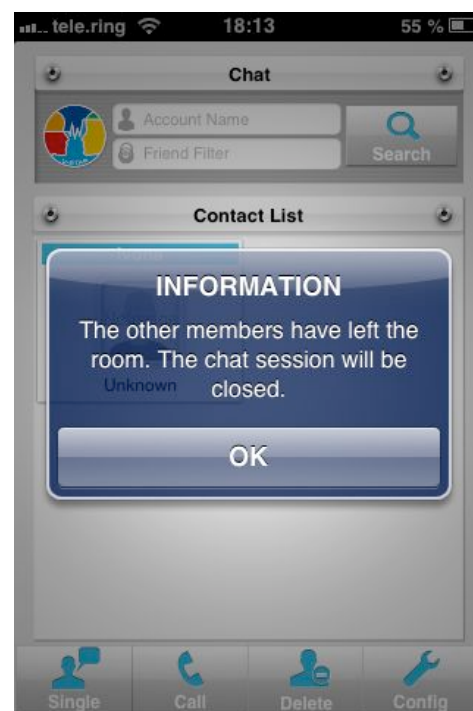


Abbildung 36: Screenshot Chatabbruch
VoiceTra4U 3.2

In der von Teilnehmerin Ivona versandten Nachricht um 18:18, welche von Teilnehmerin Mirjana mittels des *What*-Buttons abgelehnt wurde, wird ersichtlich, dass der Infotext auch

im *Chat Mode* nur in englischer Sprache verfügbar ist (vgl. Abb. 34)²¹. In diesem Fall wird abermals deutlich, dass dieser Infotext unabhängig von der Sprachkombination nur auf Englisch verfügbar ist, was bei der Nutzung von SST Apps aufgrund mangelnder Sprachkenntnisse sehr sinnwidrig erscheint. Da die App nur ins Englische und Japanische lokalisiert wurde, könnten zwar auch bei der Interpretation des User Interface und der Funktionen Verständnisprobleme entstehen, jedoch wird versucht durch den starken Gebrauch von allgemein verständlichen Icons dem entgegenzuwirken. Dieses Argument verliert jedoch an Gewicht, da auch diverse Einblendungen, wie jene in Abbildung 36, nur auf Englisch verfügbar sind. Der Infotext in Abb. 36 erscheint, wenn der oder die Gesprächspartner den Chat mittels der *Exit*-Funktion (vgl. Abb. 35) verlassen oder die App geschlossen wird.

Die Chatzeit wurde auf maximal eine Stunde beschränkt. „As the number of computers for this experiment is limited, we would like to restrict the chat time to one hour so that we’re able to offer our service to as many users as possible.“ (U-STAR 2014c). Darüber hinaus zeigten sich während der Nutzung des *Chat-Mode* im Vergleich zum *Single Mode* verstärkt längere Verbindungszeiten zum Server und es kam in einigen Fällen auch zu Abstürzen der App bzw. zu einem Einfrieren des Bildschirms, welches nur durch manuelles Schließen der App gelöst werden konnte.

Eine weitere Besonderheit ist der bereits erwähnte international angelegte, netzwerk-basierte Ansatz. Dieser wird auch während der Anwendung sehr deutlich, da jedes Mal wenn die App in den Hintergrund gelegt wird, oder auch nach Eintritt der automatischen Tastensperre aufgrund langer Inaktivität ein kleines Ladefenster mit der Information, dass gerade eine Serververbindung aufgebaut wird, erscheint. Die Anzeigedauer dieses Fensters gestaltet sich unterschiedlich und kann je nach Verbindungsqualität von 2-10 Sekunden dauern.

6.4 Benutzerfreundlichkeit

Anpassbarkeit - Dem Individualisierungsgedanken wurde bei der Konzeption der *VoiceTra4U* keine Bedeutung beigemessen. Die App lässt sich nur sehr wenig bis gar nicht individualisieren und an persönliche Vorlieben anpassen. Die NutzerInnen haben die Möglichkeit im *Chat Mode* ein Bild von sich selbst einzufügen und im *Single Mode* steht es ihnen frei sich die Rückübersetzung anzeigen zu lassen oder diese auszublenden. Die anderen Test-Apps bieten in diesem Bereich viel mehr Auswahlmöglichkeiten.

²¹ Der gesprochene Ausgangstext für die Übersetzung lautete: *Unser Treffen wurde auf morgen verschoben*. Es entstand bereits bei der Spracherkennung ein grober Fehler, der in weiterer Folge zu einer unverständlichen Übersetzung führte. In beiden Beispielchats (Abb. 34-35) wurde die App in einer ruhigen Umgebung frei von Störgeräuschen getestet.

Erlernbarkeit - Die App ist übersichtlich gestaltet und enthält nicht zu viele Unterkategorien und Menüpunkte, sodass die Navigation durch die App nicht unnötig erschwert wird. Das Layout und auch die User Interface Elemente sind nicht an die für Apple typischen Empfehlungen für EntwicklerInnen angepasst, sondern erinnern aufgrund der Schaltflächen (Ja, Nein, OK, Abbrechen) optisch stark an Windows. Mittels intuitiv gestalteter und allgemein verständlicher Icons wie dem Mikrophon oder einem Block und Stift (für die Texteingabe-Funktion) soll die Bedienung der App erleichtert werden. Leider wurde sie nur ins Japanische und Englische lokalisiert, sodass die potentiellen Problemfelder für NutzerInnen anderer Muttersprachen verhältnismäßig groß sind, im Vergleich zu den anderen Test-Apps, bei denen mehr Lokalisierungssprachen zur Verfügung stehen. Auf der Website (vgl. U-STAR 2014c) wird davon berichtet, dass es eine app-interne Bedienungsanleitung auf Englisch, Japanisch und Chinesisch gibt:

„We have the preinstalled language service for operation instructions in Japanese, English, and Chinese within the application. When you configure a foreign language as an input language, and tap the question mark, "?" in the lower section of the screen, it shows the instructions in the language you selected. For instance, if you select Chinese (standard) as input language, the instructions are shown in Chinese (standard), while the explanation is in English for other languages.“ (U-STAR 2014c)

Diese Funktion konnte leider in der Version 3.2 nicht gefunden werden. Möglicherweise bezieht sich diese Erläuterung auf die Version 2.6, ebenso wie die mit Bildern illustrierte Anwendungsanleitung der App auf derselben Seite.

Verlässlichkeit/Robustheit - Allgemein kann gesagt werden, dass *VoiceTra4U* sowohl im *Single Mode* als auch im *Chat Mode* recht zuverlässig gearbeitet hat. Im *Single Mode* kam es nur bei langen Segmenten (mehr als 10 Wörter) zu technischen Schwierigkeiten, wie Verbindungsproblemen mit dem Server oder Abstürzen. Im Vergleich dazu kamen besagte Schwierigkeiten im *Chat Mode* vermehrt vor und konnten nur durch manuelles Schließen der App behoben werden. Dabei muss festgehalten werden, dass in den Test-Chats lediglich zwei Personen an der Konversation beteiligt waren, ob mit einer höheren Anzahl an Gesprächsteilnehmern auch mehr technische Probleme auftreten wurde nicht weiter getestet. Die App war zusammenfassend in ihrer Leistung mittelmäßig robust, da die Probleme nicht dermaßen häufig vorkamen bzw. schwerwiegend waren, dass ein Chat unmöglich gewesen wäre, sondern sich vielmehr negativ auf den Gesamteindruck und die User Experience ausgewirkt haben.

6.5 SWOT-Analyse

Stärken

- Sowohl App Download als auch Nutzung kostenlos
- Spezialisierung auf asiatische Sprachen
- Spracherkennung erkennt Ende der Rede automatisch
- Sowohl Single Mode als auch Chat Mode verfügbar
- What-Icon um Unverständlichkeit der Übersetzung auszudrücken
- Dezentralisierte Server-Lösung bei technischen Problemen von Vorteil, da nur einzelne Sprachen und Regionen betroffen sind
- Sprachkombination ohne Einschränkungen frei wählbar

Schwächen

- Ständige Internetverbindung nötig
- Spracherkennungsqualität auffällig schlechter als bei anderen Apps
- Keine manuelle Korrekturmöglichkeit der Sprachaufnahme durch Texteingabe (Korrektur entweder durch Respeak-Funktion oder vollkommen neue Texteingabe)
- What-Funktion: Aufforderung zur Paraphrase wird unabhängig von Sprachkombination immer auf Englisch dargestellt
- Wenig Individualisierungsspielraum und Anpassungsmöglichkeiten an die UserInnen
- Begrenzte Chatzeit (1h)
- Keine Verlaufs- und Favoritenfunktion (Verlauf nur in Chat Mode)
- App wurde nur ins Japanische und Englische lokalisiert, dies kann zu Verständnisproblemen bei Personen mit abweichender Erstsprache führen
- Teilweise lange Lade- und Verbindungsdauer im Chat Mode
- Vermehrte Abstürze im Chat Mode
- Bei längeren Segmenten auch häufige Fehlermeldungen im Single Mode

Chancen

- Großes Verbesserungspotential der einzelnen SST-Komponenten aufgrund internationaler Zusammenarbeit und gemeinsamer Ressourcen
- Spezialisierung auf asiatische Sprachen erweiterbar, um dem wachsenden Markt gerecht zu bleiben
- Integration von Verlaufs- und Favoritenfunktion, um auf häufig gebrauchte Phrasen zurückgreifen zu können

Risiken

- Hoher Datenverbrauch
- Dezentralisierung und Ressourcenverteilung eröffnet die Gefahr (zu) vieler kleinerer „Baustellen“, worunter die allgemeine Qualität der Übersetzungen leiden könnte
- Datenschutz

7 SayHi Translate 3.8.2

Die App *SayHi Translate* wurde von einem im Jahr 2009 gegründeten Startup-Unternehmen erfunden und war ab 29. Juni 2011 im Apple App Store in der Version 1.1 verfügbar (vgl. <https://m2.facebook.com/sayhitranslate> 2011). Die Idee zu der App entstand aus Beobachtungen des Krankenhausalltages, in dem viele Patienten und Patientinnen nicht im Stande waren dem Krankenhauspersonal ihre Beschwerden verständlich mitzuteilen. Nach einer drei-jährigen Entwicklungsphase wurde versucht, die Dienste der App an Krankenhäuser und andere Institutionen des Gesundheitswesens zu verkaufen, bis man sich schließlich dazu entschied die App auch im App Store der breiten Maße zur Verfügung zu stellen (vgl. <http://www.hartfordbusiness.com> 2012).

„The self-funded app [...] uses the Amazon cloud, Dragon's medical and standard libraries for interpretation and a few other proprietary systems created in house. Security is addressed with full data 250-bit encryption protecting in transit and at rest conversations.“ (<http://www.hartfordbusiness.com> 2012)

Die zwei Hauptargumente, die als Vorteil der App gegenüber beispielsweise Google Translate galten, waren die Sicherheit und die Genauigkeit der App. Die Leistung der App würde sich stetig verbessern, je mehr die App genutzt werde. Darüber hinaus hätten AnwenderInnen die Möglichkeit durch manuelles Korrigieren, die Treffersicherheit der App zu verbessern (vgl. <http://www.hartfordbusiness.com> 2012).

Im Folgenden werden die Einzelheiten der App, wie in Kapitel 4 vorgesehen, vorgestellt und es soll an dieser Stelle auch darauf hingewiesen werden, dass sich die folgenden Daten und Fakten auf die aktuell erhältliche Version (Stand: 03.08.2014) beziehen und daher eine vergängliche Gültigkeitsdauer haben.

7.1 Technische Hintergründe

Die App *SayHi Translate* (vgl. <https://itunes.apple.com>^d 2014) ist im Apple App Store für alle iOS Geräte (iPhone, iPad, iPod touch) in der Version 3.8.2 um 1,79€ erhältlich und erfordert iOS 6.1 oder neuer. In der deutschsprachigen Version des App Stores wird die App unter der Bezeichnung *Übersetzer Sprach (SayHi Translate)* aufgeführt. Es wird jeweils eine für das iPhone und für das iPad optimierte Version angeboten. Die Größe der App beläuft sich auf 16.1 MB und findet sich in der Kategorie Wirtschaft wieder. Es gibt insgesamt 61 Rezensionen, die, bis auf eine Ausnahme, alle durchweg positiv sind und vier oder fünf von insgesamt fünf Sternen vergeben. Als Entwickler wird SayHi LLC angeführt, der noch eine weitere App namens *Sprach Übersetzer für Tischplatte (TableTop Translator)* anbietet (vgl. <https://itunes.apple.com>^e 2014). Diese ist eine SST App für das iPad,

welche der *SayHi* App vom User Interface sehr ähnlich sieht, jedoch optisch so konzipiert ist, dass zwei sich gegenüberstehende GesprächspartnerInnen die App bedienen können, da sie je auf einer Hälfte des Bildschirms dem Gesprächsverlauf folgen können. Die deutschen Beschreibungen beider Apps im App Store scheinen durch maschinelle Übersetzung generiert worden zu sein und hinterlassen marketingtechnisch gesehen keinen guten Eindruck. Es existiert bis dato keine App für NutzerInnen von Android Geräten.

7.2 User Interface

Beim Starten der App erscheint zunächst der Startbildschirm mit dem *SayHi Translate* Logo bzw. Schriftzug und darunter der Slogan *The Interpreter In Your Pocket* für ungefähr 1-2 Sekunden. Hier wird darauf Bezug genommen, dass die Speech-to-Speech Translation Apps durch die Erweiterung der maschinellen Übersetzungskomponente um die beiden Spracherkennungs- und Sprachsynthesekomponenten im Grunde dieselbe Form von Translation durchführen wie DolmetscherInnen. Nach dem Startbildschirm erscheint das in typischen Blautönen gehaltene Hauptmenü (vgl. Abbildung 37). Bei *SayHi* fehlt die bisher für SST Apps typische optische Zweiteilung des Bildschirms. Hier wird vielmehr in Form von Sprechblasen, die von diversen Instant-Messaging Diensten bekannt sind, gearbeitet. Auf Abb. 37 ist dank des Übersetzungsbeispiels die Zuordnung der Sprechblasen zu den jeweiligen Sprechern oder Sprecherinnen deutlich erkennbar. Der Ausgangstext wird in einer transparenten Sprechblase dargestellt, während unmittelbar darunter, der Zieltext in einer der Sprache entsprechend eingefärbten Sprechblase erscheint. Die rechte Sprache ist immer blau eingefärbt (siehe auch Sprachaufzeichnungsbutton) und die links ausgewählte Sprache wird grün dargestellt. Durch die farbliche Kodierung ist die Zuordnung der Sprechblasen leicht durchführbar. Wie in Kapitel 7.3.1 ersichtlich, ist die App in eine Vielzahl von verschiedenen Sprachen lokalisiert worden. Dies erklärt auch warum sich die Sprache des User Interface in Abb. 38 von Deutsch auf Französisch ändert. Sobald die Spracherkennungsfunktion mittels des blauen oder grünen Buttons am unteren Rand des Bildschirms gestartet wird, erscheint der Text auf dem Spracherkennungsbildschirm in der jeweiligen Sprache. Darüber hinaus wird auch das farbliche Schema in der Sprachaufnahme-Animation wieder aufgegriffen und in diesem Fall in dem Grünton, welcher der Sprache Französisch bereits im Startbildschirm zugewiesen wurde, dargestellt. Wäre die Sprachaufnahmefunktion mittels des blauen Button gestartet worden, wäre die Animation in blau gehalten und es wären die Wörter *Sprechen* und *Erledigt* statt *Parler* und *Fini* zu lesen gewesen.



Abbildung 37: Screenshot Hauptmenü mit Übersetzungsbeispiel SayHi 3.8.2



Abbildung 38: Screenshot Spracherkennung lokalisiert SayHi 3.8.2

Auf dem Hauptmenü sind, abgesehen von den Sprachaufnahmebuttons und der Sprechblasen, am oberen Rand Icons für zusätzliche Funktionen und erweiterte Einstellungen zu finden. Darüber hinaus ist die Einstellung der Sprachkombination ebenfalls im Hauptmenü aufzufinden. Tippen die UserInnen auf eine der beiden Sprachen, öffnet sich ein scrollbares Auswahlfenster für die gewählte, blau unterlegte Sprache. Eine Vielzahl der Sprachen sind mit dem Erwerb der App bereits automatisch installiert und können sofort genutzt werden. Mit dem Kauf der Premium Variante können alle anderen angebotenen Sprachen aktiviert werden (vgl. Abbildung 39). Auf die Sprachauswahl wird erst in Kapitel 7.3.1 näher eingegangen. Darüber hinaus gibt es den UserInnen frei zwischen einer männlichen und einer weiblichen Wiedergabestimme zu wählen und das Tempo der Sprachwiedergabe zu bestimmen. Um das Sprachauswahlfenster zu schließen, muss lediglich die blau markierte Sprache bzw. der nach oben zeigende Pfeil angetippt werden. Zurück auf dem Hauptmenü finden sich am oberen Rand des Displays vier Icons, welche vier verschiedene Funktionen symbolisieren. Das erste Icon, ganz rechts in der Leiste, zeigt das iOS-typische *Teilen*-Symbol. Die *Teilen*-Funktion (vgl. Abbildung 40) ermöglicht es den UserInnen die bereits übersetzten oder wahlweise auch die ausgangssprachlichen Segmente an andere Personen per E-Mail oder SMS-Nachricht zu versenden oder in den Zwischenspeicher zu kopieren und in einer anderen App den Text direkt einzusetzen und weiter zu verarbeiten. Dazu müssen die gewünschten Elemente lediglich am linken Rand markiert werden und können sofort mittels des *Teilen*-Buttons weitergesandt werden. In allen Funktionsfeldern

der oberen Leiste ist am unteren Ende des Bildschirmes der Ansatz eines blauen Fensters zu sehen. Wird diese blaue Fläche angetippt, so rutscht automatisch das Hauptmenü nach oben und erfüllt das Display. Da kein traditioneller *Zurück*-Button existiert, ist dies die einzige Möglichkeit von den Einstellungen und Funktionen der oberen Leiste wieder zurück auf das Hauptmenü zu gelangen.

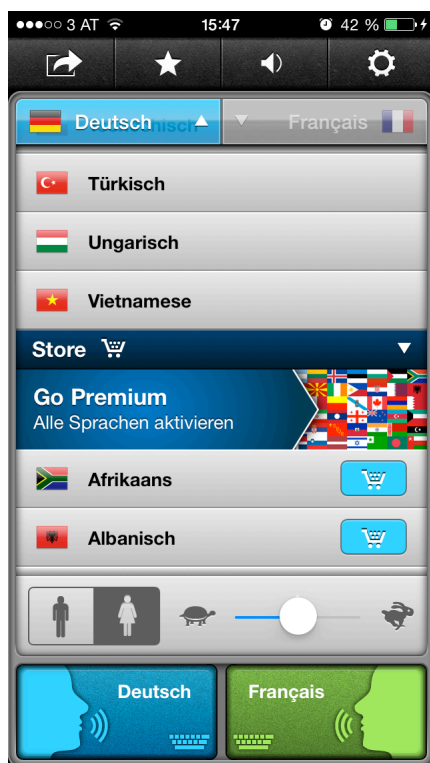


Abbildung 39: Screenshot Sprachauswahl SayHi 3.8.2

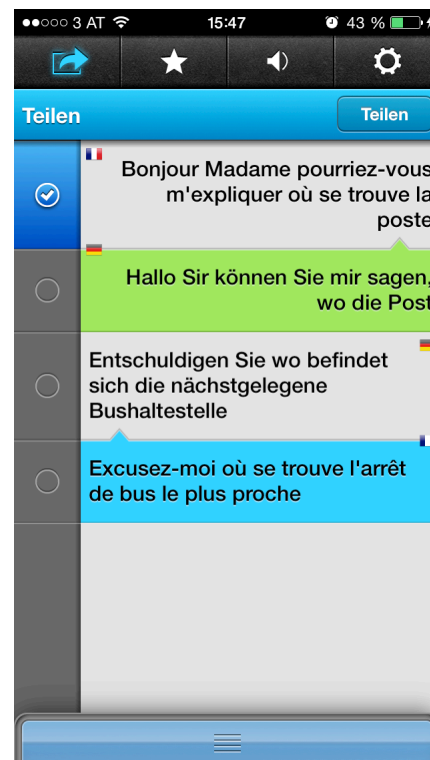


Abbildung 40: Screenshot Inhalte teilen SayHi 3.8.2

Das zweite Icon von rechts, der Stern, symbolisiert die favorisierten AT und ZT Segmente. Auf diese Funktion soll im Kapitel 7.3 näher eingegangen werden, da es sich um ein besonderes Feature von *SayHi* handelt, welches in den anderen untersuchten Apps in dieser Form nicht existiert.

Das dritte Icon von rechts, der bereits bekannte Lautsprecher, dient selbstverständlich der Lautstärkeregelung. Wird dieses angetippt, wird auf Höhe der Sprachauswahlstasten (vgl. Abb. 37) ein senkrechter Lautstärkereglер eingeblendet, der zur manuellen Anpassung der Lautstärke dienen soll. Diese können UserInnen auch durch die üblichen, direkt am Gerät verbauten Lautstärketasten nach Belieben verändern.

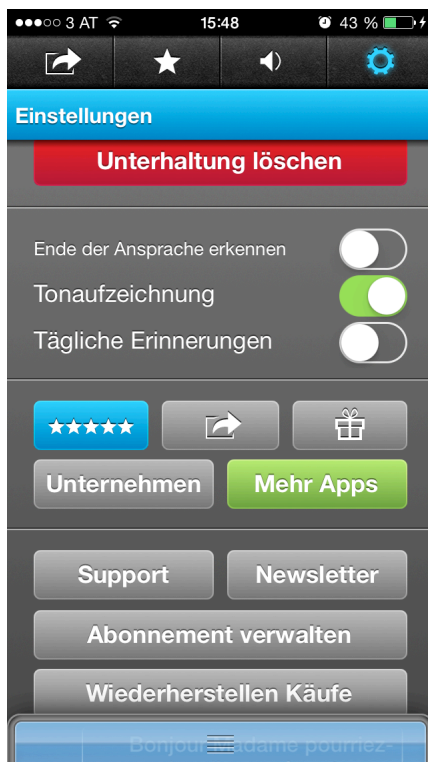


Abbildung 41: Screenshot Einstellungen SayHi 3.8.2

Das letzte Icon, das Zahnrad, steht für die erweiterten Einstellungen (vgl. Abbildung 41). Hier können beispielsweise die bisher übersetzten Segmente (in *SayHi Unterhaltungen* genannt) gelöscht werden. Darüber hinaus können drei verschiedene Einstellungen aktiviert bzw. deaktiviert werden. Die UserInnen können entscheiden, ob das Ende des gesprochenen Textes automatisch durch die App erkannt und definiert werden soll oder, ob dies durch manuelles Betätigen des *Erledigt*-Buttons (vgl. Abb. 38) von Seiten der NutzerInnen geschehen soll. Des Weiteren kann, wie bei *Jibbigo*, der Ausgangstext aufgenommen werden und bei Bedarf wiedergegeben werden. Wenn die Funktion *Tonaufzeichnung* deaktiviert ist, wird die Wiedergabe des AT per Sprachsynthese generiert.

Die erste Kategorie der Schaltflächen, welche sich unterhalb der Funktionen befindet, steht in direktem Zusammenhang mit der App, ihren Entwicklern und dem App Store. So kann zum Beispiel durch Tippen auf die blaue, mit Sternen versehene Schaltfläche eine Wertung für die App abgegeben werden. Die AnwenderInnen werden gefragt, ob sie die Anwendung im App Store bewerten wollen und werden dann nach erfolgter Zustimmung in den App Store weitergeleitet. Der zweite Button steht für die bereits bekannte *Teilen*-Funktion und öffnet automatisch ein E-Mail Fenster mit dem vorgefertigten Text: „*Von SayHi Übersetzer. Ein genialer, universeller Übersetzer, der Ihre Sprache buchstäblich in eine andere Sprache übersetzt. SayHi Übersetzer im App Store (Link)*“. Der dritte Button dieser Reihe ist der aus dem App Store bekannte *Geschenke*-Button und dient dazu, die App als Geschenk an andere Personen versenden zu können. Die UserInnen werden, wie auch bei dem Bewertungsbutton, explizit gefragt, ob sie die App verschenken sollen und ob eine Weiterleitung zum App Store erfolgen soll. Mittels des *Unternehmen*-Buttons kann mit dem Sales Department von *SayHi* per E-Mail Kontakt aufgenommen werden. Durch Antippen des grünen *Mehr Apps*-Buttons wird eine Werbeanzeige für die zweite von *SayHi* angebotene App *TableTop Translator* angezeigt.

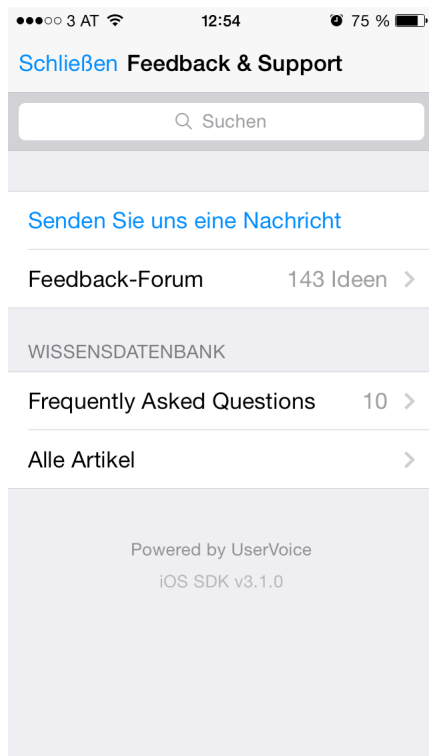


Abbildung 42: Screenshot Feedback & Support SayHi 3.8.2

Die zweite Kategorie der Schaltflächen hat diverse Funktionen (vgl. Abb. 41). Der erste Button, *Support*, ermöglicht eine Feedbackfunktion von Seiten der UserInnen an die EntwicklerInnen, und bietet gleichzeitig eine Plattform für den Support (vgl. Abbildung 42). Hier können AnwenderInnen eine Nachricht mit Feedback oder Fragen an das Entwicklerteam senden. Diese werden dann im Feedback-Forum gepostet und andere UserInnen können wiederum darauf reagieren und kommentieren, sowie der Frage ihre Stimme geben, für den Fall, dass sie das Anliegen als wichtig erachten. Darüber hinaus gibt es eine FAQ-Sammlung. Dieser Dienst wird durch UserVoice (vgl. <https://www.uservoice.com> 2014), eine Anwendung für verbesserte Kundenbetreuung in webbasierten und mo-

bilen Applikationen, ermöglicht. Zusätzlich ist auf dieser Seite erkennbar mit welcher iOS SDK²² Version die App *SayHi* erstellt wurde.

Der zweite Button (vgl. Abb. 41), *Newsletter*, bietet die Möglichkeit sich für den Newsletter anzumelden. Es öffnet sich ein kleines Fenster mit der Aufforderung: „*Bleiben Sie informiert! Geben Sie Ihren E-Mail über Gewinnspiele und Neuigkeiten aus SayHi lernen.*“ und einer Zeile für die Eingabe der E-Mail Adresse. Der Button *Abonnement verwalten* leitet die UserInnen direkt zu iTunes weiter, wo das Abonnement des Premium Sprachpaktes angefordert bzw. verwaltet werden kann. Der vorletzte Button *Wiederherstellen Käufe* dient dazu, zum Beispiel nach einem Reset des Geräts oder der Installation der App auf einem neuen Gerät, mithilfe des iTunes Accounts bereits gekaufte Sprachpakete wiederherzustellen, um eine wiederholte Zahlung zu vermeiden. Mit dem letzten Button namens *Über* (in Abb. 41 nicht erkennbar) öffnet sich ein Bildschirm mit dem *SayHi Translate* Logo und Copyright Angaben, sowie Links zum Datenschutz, dem Twitter Account oder der Facebook Seite von *SayHi Translate*. Darüber hinaus ist im unteren Teil das *Nuance* Logo zu erkennen, was darauf hindeutet, dass *SayHi*, ebenso wie *Jibbigo*, für die Sprachsynthese die Dienste von *Nuance Communications, Inc.* in Gebrauch hat. *SayHi* nutzt Darüber hinaus auch für die Spracherkennungskomponente Software von *Nuance* (vgl. Nuance Communications, Inc. 2011). Auch andere bekannte Apps verwenden die Sprachtechnologien von *Nuance Communications*:

²² Mit SDKs (Software Development Kit) wird den EntwicklerInnen ein Paket mit nötigen Tools und Programmierschnittstellen zur Verfügung gestellt, mithilfe dessen (in diesem Fall auf das iPhone) angepasste Anwendungen entwickelt werden können (vgl. Apple 2008).

„With the Dragon Mobile SDK, developers can rapidly bring any voice-enabled app to market, even if they have never implemented speech technology before. The Nuance Mobile Developer Program has helped integrate the voice recognition and text-to-speech capabilities of some of the most popular apps on the market, including Siri, Price Check by Amazon, Ask for iPhone, Merriam-Webster, Dictionary.com, OnStar RemoteLink, Aisle411, AirYell, Yellow Pages by Avantar, Taskmind, EZ CM by EZ Connector, Bon App and others.“ (Nuance Communications, Inc. 2011)

Allgemein kann festgehalten werden, dass der Aufbau und das Layout der *SayHi Translate* App stark von jenen der beiden zuvor beschriebenen Apps abweicht. Auffällig ist hierbei das Layout der Unterhaltungen, die in Sprechblasen dargestellt werden und somit leicht zu vorherigen Segmenten gescrollt werden kann, was einen erheblichen Vorteil in Bezug auf die Benutzerfreundlichkeit darstellt. Ebenso wie die anderen beiden Apps weist auch *SayHi* keine komplizierte Architektur auf und ist durch iOS-bekannte Icons leicht verständlich und bedienbar.

7.3 Features

Ebenso wie bei den anderen beiden Apps soll an dieser Stelle der Aspekt des Datenschutzes beleuchtet werden, bevor auf die Features der App eingegangen wird. *SayHi LLC* schreibt in seinen Datenschutzbestimmungen, dass jegliche Installation, Registrierung oder anderweitige Nutzung der Dienste der *SayHi* App einer Zustimmung zu den Nutzungsbestimmungen gleichgesetzt wird:

„By installing, using, registering to or otherwise accessing the services, you agree to this privacy policy and give an explicit and informed consent to the processing of your personal data in accordance with this privacy policy. If you do not agree to this privacy policy, please do not install, use, register to or otherwise access these services.“ (SayHi LLC 2014a).

SayHi speichert und verarbeitet unpersönliche Daten wie beispielsweise Geräteeigenschaften, Gerätesoftware, Firmware, Mobilfunkanbieter, etc. sowie weitere unpersönliche Daten, die für die Leistungsoptimierung der App von Nöten sind. Persönliche Daten, die von UserInnen beispielsweise durch Abonnementanmeldung, Newsletteranmeldung, Supportanfragen oder die Teilen-Funktion, an *SayHi* übermittelt werden, werden auch gespeichert und verarbeitet, jedoch wird die Weitergabe der Daten an Dritte nicht explizit erwähnt. (vgl. SayHi LLC 2014a).

7.3.1 Sprachenangebot

Die App *SayHi Translate* verfügt über ein standardmäßiges Sprachangebot von 40 Sprachen und Dialekten. Die Kombinationsmöglichkeit der Sprachenpaare ist nicht wie bei *Jibbiggo* begrenzt, sondern kann wie bei *VoiceTra4U* beliebig gewählt werden. Die im vorherigen Kapitel erwähnte Möglichkeit zur individuellen Einstellung einer männlichen oder weiblichen Ausgabestimme ist nicht bei allen Sprachen gegeben.

Folgende Sprachen sind laut dem *SayHi* Press Kit in der Standard Version der App nutzbar (vgl. SayHi LLC 2014b):

- Arabic (Egypt),
- Arabic (Saudi),
- Arabic (UAE),
- Basque,
- Cantonese,
- Catalan,
- Czech,
- Danish,
- Dutch,
- English (Australia),
- English (UK),
- English (USA),
- Finnish,
- French (Canada),
- French (France),
- German,
- Greek,
- Hindi,
- Hungarian,
- Indonesian,
- Italian,
- Japanese,
- Korean,
- Malay,
- Mandarin (China),
- Mandarin (Taiwan),
- Norwegian,
- Polish,
- Portuguese (Brazil),
- Portuguese (Portugal),
- Romanian,
- Russian,
- Slovak,
- Spanish (Mexico),
- Spanish (Spain),
- Spanish (USA),
- Swedish,
- Thai,
- Turkish,
- Vietnamese

Diese Aufzählung deckt sich auch mit den in der App angeführten Sprachen, wobei festgehalten werden muss, dass für drei Sprachen keine Sprachausgabe verfügbar ist. Dazu gehören Baskisch, Hebräisch und Malaiisch. Die ASR Komponente ist bei allen Sprachen vorhanden.

Durch ein Abonnement der Premium Version von *SayHi* werden 40 neue Sprachen, sowie 25 neue Dialekte erworben (siehe weiter unten), sodass insgesamt in 105 Sprachen und Dialekte übersetzt werden kann. Die AnwenderInnen haben die Wahl zwischen zwei Abonnementoptionen: Eine 3-monatige Variante um 2,69€ oder eine 6-monatige um 4,49€ (vgl. <https://itunes.apple.com>^d 2014).

Folgende Sprachen sind nur in der kostenpflichtigen Premium Version verfügbar (vgl. SayHi 2014b):

- | | | |
|----------------|--------------|---------------|
| • Afrikaans, | • Filipino, | • Latvian, |
| • Albanian, | • Galician, | • Lithuanian, |
| • Armenian, | • Georgian, | • Macedonian, |
| • Azerbaijani, | • Gujarati, | • Maltese, |
| • Belarusian, | • Haitian | • Marathi, |
| • Bosnian, | • Creole, | • Serbian, |
| • Bulgarian, | • Hebrew, | • Slovakian, |
| • Bengali, | • Hmong, | • Swahili, |
| • Cebuano, | • Icelandic, | • Telugu, |
| • Croatian, | • Irish, | • Ukranian, |
| • Esperanto, | • Javanese, | • Urdu, |
| • Estonian, | • Kannada, | • Welsh, |
| • Farsi, | • Latin, | • Yiddish |
| | • Lao, | |

-
- | | | |
|---|--|--|
| • Arabic (Bahrain, Kuwait, Libya, Oman, Palestine, Qatar) | • English (Canada, India, New Zealand, South Africa) | • Spanish (Argentina, Bolivia, Columbia, Costa Rica, Dominican Republic, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Uruguay, Venezuela) |
|---|--|--|

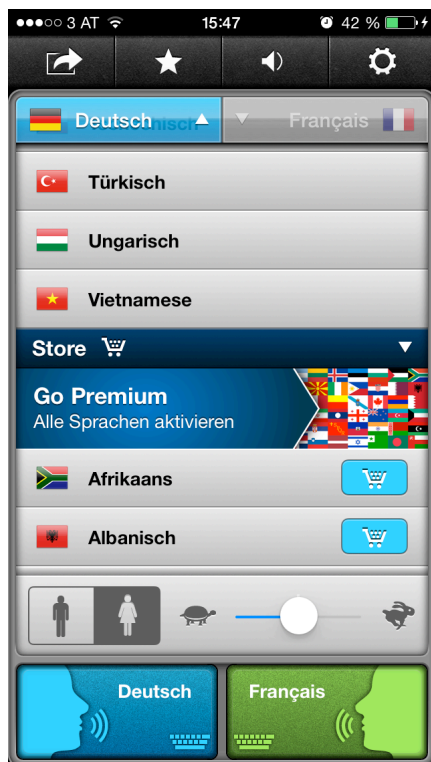


Abbildung 43: Screenshot Premium Sprachen SayHi 3.8.2

Abbildung 43 zeigt die Premium Version von *SayHi* die mittels eines kostenpflichtigen Abonnements bezogen werden kann. An dieser Stelle soll darauf hingewiesen werden, dass nicht alle Premium Sprachen zwingend über eine ASR Komponente oder eine Sprachsynthese Komponente verfügen.

Mit insgesamt 105 verschiedenen Sprachen und Dialekten hat *SayHi* das größte Sprachenangebot von allen drei untersuchten Apps. Darüber hinaus wurde die App in eine Vielzahl der angebotenen Sprachen lokalisiert und bietet den UserInnen bei der Sprachauswahl im Hauptmenü auch einen akustischen Input, da die gewählte Sprache in der jeweiligen nationalen Bezeichnung ausgesprochen wird, sodass die UserInnen sich einen Eindruck von der Stimme und der Sprachausgabe machen können. Zu den insgesamt 33 lokalisierten Sprachen gehören (vgl. SayHi 2014b):

- English,
- Arabic,
- Bokmal,
- Norwegian,
- Catalan,
- Chinese,
- Czech,
- Danish,
- Dutch,
- Finnish,
- French,
- German,
- Greek,
- Hebrew,
- Hindi,
- Hungarian,
- Indonesian,
- Italian,
- Japanese,
- Korean,
- Malay,
- Polish,
- Portuguese,
- Romanian,
- Russian,
- Simplified Chinese,
- Slovak,
- Spanish,
- Swedish,
- Thai,
- Traditional Chinese,
- Turkish,
- Vietnamese

7.3.2 Übersetzer Modus

Der Übersetzer-Modus ist, wie bereits erwähnt, optisch etwas anders aufbereitet als die zwei vorangegangenen Apps. Mit der Darstellung des Dialoges in Sprechblasen (vgl. Abb. 37), die bei Bedarf hin und her gescrollt werden können, haben UserInnen die Möglichkeit einfach und rasch auf bereits übersetzte Segmente zurückzugreifen.

Indem auf eine der unteren Sprachaufnahmetasten gedrückt wird, lässt sich der Übersetzungsprozess direkt vom Hauptmenü (vgl. Abb 37) aus starten. Es ertönt ein Signalton mit dem auch der Sprachaufnahmebildschirm (vgl. Abb. 38) eingeleitet wird. Besonders anzumerken ist, dass dieser Bildschirm in der jeweiligen Sprache lokalisiert angezeigt wird, sodass sich fremdsprachliche NutzerInnen leichter zurechtfinden. Wenn in den Einstellungen die Funktion *Ende der Ansprache erkennen* aktiviert ist, erkennt die App automatisch das Ende des Zieltextes. Als zweite Option kann die Sprachaufnahme auch manuell beendet werden, indem auf den *Erledigt*-Button getippt wird. Das Ende der Aufnahme wird auch von einem Signalton begleitet. Die Transkription des Zieltextes benötigt je nach Länge 1-2 Sekunden und unmittelbar danach erfolgt auch die Übersetzung und Sprachausgabe. Ist das Umfeld zu laut oder konnte aus einem anderen Grund kein verwertbarer Ausgangstext erkannt werden, erscheint eine Fehlermeldung, in der die UserInnen aufgefordert werden das Gesagte zu wiederholen (vgl. Abbildung 44). Wie bei den anderen Apps gibt es auch bei *SayHi* die Möglichkeit den transkribierten AT bzw. den generierten ZT manuell zu verändern oder zu korrigieren. Dazu muss lediglich die gewünschte Sprechblase angetippt werden. Unmittelbar danach öffnet sich ein kleineres Fenster mit dem Text und verschiedenen Funktionen (vgl. Abbildung 45). In diesem Fenster kann der Text manuell nachbearbeitet werden oder mithilfe des Sternsymbols als Favorit gespeichert werden. Darüber hinaus kann das Segment mittels des *Teilen*-Buttons an andere versendet werden oder nochmals abgespielt werden. Abhängig davon ob in den Einstellungen (vgl. Abb. 41) die Funktion *Tonaufzeichnung* aktiviert wurde, spielt die App (für den Fall, dass es sich um ein ausgangssprachliches Segment handelt) die Originalaufnahme des Gesagten ab. Handelt es sich um eine generierte Übersetzung wird das Segment per Sprachsynthese ausgegeben. Mithilfe der letzten Funktion, des Papierkorbs, kann das entsprechende Segment gelöscht werden. Der blaue Balken am linken Rand des Fensters deutet entsprechend der Farbkodierung auf die Sprache hin.



Abbildung 44: Screenshot Erkennungsfehler ASR SayHi 3.8.2

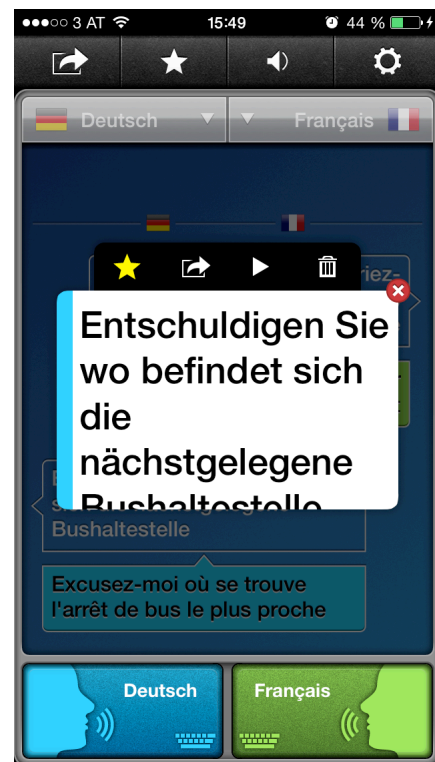


Abbildung 45: Screenshot manuelle Korrektur SayHi 3.8.2

Subjektiv betrachtet macht *SayHi* den robustesten Eindruck von allen getesteten Apps. Es kam nie zu Abstürzen der App oder dergleichen. Auch in der Generierung des ZT arbeitet die App (je nach Länge der Segmente) minimal schneller als die anderen beiden. Das Layout, welches UserInnen bereits von diversen Instant Messaging Diensten kennen (wie z.B. der eigenen SMS-App), ist übersichtlich gestaltet und aufgrund der Farbkodierung leicht verständlich. Der große Vorteil besteht darin, dass bereits übersetzte Segmente schnell und einfach wiedergefunden, rasch abgeändert und neu übersetzt oder einfach nur wiederholt vorgespielt werden können. Mit 40 angebotenen Sprachen und sehr flexiblen Kombinationsmöglichkeiten in der Standardversion ist *SayHi* die umfangreichste App im Bezug auf das Sprachangebot. Durch das kostenpflichtige 3 bzw. 6-monatige Abonnement können die UserInnen zwischen 105 Sprachen und Dialekten wählen. Leider ist die App auf eine stabile Internetverbindung angewiesen und kann nicht offline verwendet werden.

7.3.3 Besonderheiten

Der große Vorteil der *SayHi* App liegt unbestritten in ihrer, bereits erwähnten, großen Sprachenvielfalt. Verglichen mit *Jibbigo* und *VoiceTra4U*, überzeugt *SayHi* mit großteils natürlicher Sprachausgabe. Lediglich bei den Sprachen Arabisch, Indonesisch und Rumänisch lässt sich die Tatsache, dass es sich um automatisierte Sprachausgabe handelt, nicht leugnen. Dennoch wurde darauf geachtet, auch das User Interface sprachlich so gut als

möglich an die AnwenderInnen anzupassen. Nicht nur, indem die App in eine Vielzahl von Sprachen lokalisiert wurde, sondern auch dadurch, dass der Spracherkennungsbildschirm app-intern nochmals an die ausgewählte Sprachkombination angepasst wird. Darüber hinaus wird die Sprachauswahl durch akustisches Feedback begleitet, indem die gewählte Sprache in der nationalen Benennung wiedergegeben wird.

Ein anderer großer Vorteil der App ist ihre Favoritenfunktion, da sie in dieser Hinsicht einzigartig ist. Wie bereits in Abb. 45 aufgezeigt wurde, können einzelne Segmente durch Tippen auf den Stern als Favoriten gespeichert werden. Im Hauptmenü ist die Favoritenfunktion ebenfalls mit einem Stern in der oberen Funktionsleiste markiert. Existieren bereits favorisierte Segmente, so werden diese nach Sprachen kategorisiert dargestellt (vgl. Abbildung 46). Die UserInnen sehen eine Vorschau des Segmentes und können dieses durch Antippen auswählen. Das Ausgangssprachliche Segment wird daraufhin im Hauptmenü übersetzt und per Sprachsynthese ausgegeben. Ist eine Bearbeitung der bereits gespeicherten Favoriten gewünscht, muss der *Bearbeiten*-Button angetippt werden, woraufhin sich ein weiteres Fenster öffnet (vgl. Abbildung 47).

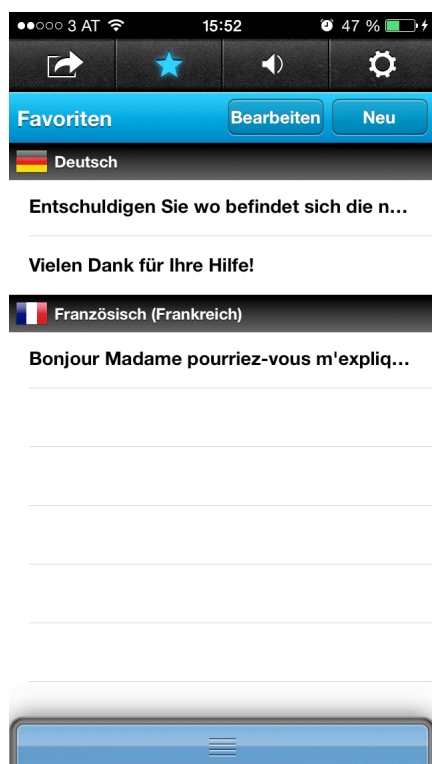


Abbildung 46: Screenshot Favoriten verwalten SayHi 3.8.2

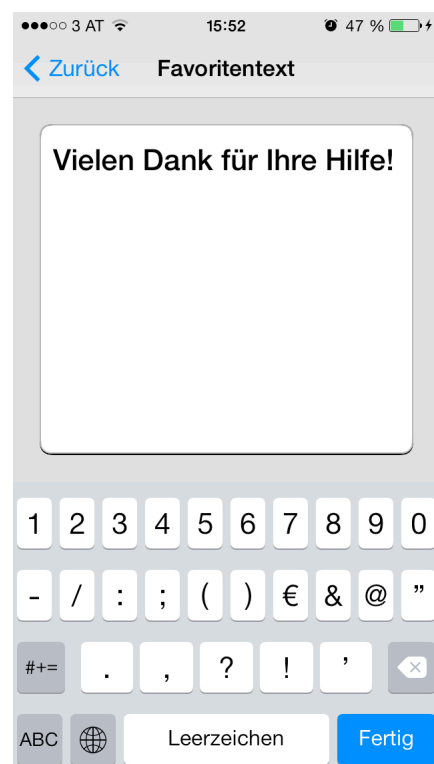


Abbildung 47: Screenshot Favoriten anlegen bzw. bearbeiten SayHi 3.8.2

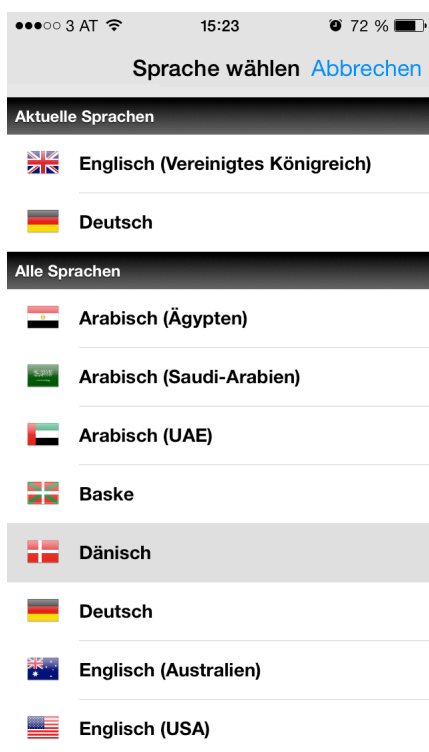


Abbildung 48: Screenshot Favoriten anlegen - Sprachauswahl SayHi 3.8.2

In der Favoritenfunktion können auch gänzlich neue Segmente, die also nicht bereits im Hauptmenü übersetzt wurden, als Favoriten definiert werden. Dazu wird mittels des *Neu*-Buttons (vgl. Abb. 46) zunächst ein Fenster mit der Sprachauswahl geöffnet (vgl. Abbildung 48), bevor der neue Text im Favoritentext-Fenster (vgl. Abb. 47) eingegeben werden kann. Die Sprachauswahl in Abb. 48 gestaltet sich etwas kompliziert, da, trotz der Auflistung aller verfügbaren Sprachen, nur Favoriten für die aktuell ausgewählte Sprachkombination erstellt werden können. Wird eine der Sprachen aus der Kategorie *Alle Sprachen* ausgewählt, (z.B. Dänisch) reagiert die App nicht. Markiert man z.B. Dänisch und daraufhin eine der Sprachen aus der Kategorie *Aktuelle Sprachen* (z.B. Deutsch) und gibt einen dänischen Text ein, erscheint dieser in der Favoritenliste unter der Kategorie Deutsch und wird im Hauptmenü dennoch mit der anderen Sprache

aus der Kategorie *Aktuelle Sprachen* (in diesem Fall Englisch) übersetzt. In anderen Worten: Um einen Favoritentext für eine gewünschte Sprachkombination anzulegen, muss diese Kombination zunächst im Hauptmenü ausgewählt werden, da die App sonst fälschlicher Weise den Text für die aktuell gewählte Sprachkombination anlegt.

Die Favoritenfunktion ist gut angedacht, jedoch nicht optimal umgesetzt worden. Durch die Anzeige aller verfügbaren Sprachen (vgl. Abb. 48) werden UserInnen insofern in die Irre geführt, als vermutet werden kann, dass auch aus dieser Liste eine Sprache gewählt werden kann. Stattdessen stehen im Prinzip nur die zwei im Hauptmenü definierten Sprachen für das Anlegen eines neuen Favoriten zur Verfügung.

7.4 Benutzerfreundlichkeit

Anpassbarkeit - Die individuelle Anpassbarkeit der App wurde bereits im vorherigen Kapitel angeschnitten. In den Einstellungen ist die *SayHi* individualisierbarer als die *VoiceTra4U* App (bei der dieser Aspekt vollkommen fehlt) jedoch etwas weniger anpassbar als *Jibbigo*. Dazu zählen z.B. die Wahl zwischen weiblicher und männlicher Ausgabestimme, das Sprechtempo, das Erkennen des Textendes etc. Eine Komponente, die nur *SayHi* anbietet, die jedoch nicht unbeträchtlich zum Individualisierungsgrad der App bei-

trägt, ist die Favoritenfunktion. Wie bereits im vorherigen Kapitel angemerkt wurde, ist diese Funktion leider etwas verwirrend umgesetzt und daher noch verbesserungswürdig.

Erlernbarkeit - Der Aufbau der App ist übersichtlich gestaltet und enthält nicht zu viele Unterkategorisierungen und Menüpunkte, sodass die Steuerung der App nicht unnötig erschwert wird. Durch das auf Sprechblasen basierende Layout bringt *SayHi* eine übersichtliche Verlaufsfunktion mit sich und erleichtert die Zuordenbarkeit der Segmente zu den jeweiligen SprecherInnen. Universell verständliche Piktogramme bzw. Icons steigern, ebenso wie bei den anderen beiden Apps, die Verständlichkeit und den Grad der Erlernbarkeit. Unterstützt werden die AnwenderInnen zudem durch akustisches Feedback bei der Sprachauswahl (Aussprache der Sprachebezeichnung in der jeweiligen Sprache) und durch die app-interne Lokalisierung des Spracherkennungsbildschirmes. Zudem wurde die App in 33 verschiedene Sprachen lokalisiert.

Verlässlichkeit/Robustheit - Allgemein kann gesagt werden, dass *SayHi* zuverlässig gearbeitet hat. Es kam bei den Tests zu keinen Störungen, Verbindungsproblemen oder sonstigen technischen Schwierigkeiten. Die App wirkte auch bei schlechter Internetverbindung, welche sich lediglich in längeren Ladeprozessen äußerte, sehr robust.

Negativ ist jedoch der deutsche Werbetext auf iTunes aufgefallen. Es muss sich dabei um eine maschinelle Übersetzung des englischen Werbe- und Informationstextes handeln. Alleine die Benennung *Übersetzer Sprach* wirft kein gutes Licht auf die Dienste und Funktionen der App und könnte unter Umständen potentielle Kunden von einem Kauf abschrecken, zumal bereits die Standard App nicht kostenlos ist.

7.5 SWOT-Analyse

Stärken

- In-App Lokalisierung bei Spracherkennung
- Favoritenfunktion und deren manuelle Erweiterbarkeit
- Verlauf leicht einsehbar
- Akustisches Feedback z.B. bei Sprachauswahl
- Erweiterbarkeit auf 105 verschiedene Sprachen und Dialekte
- Keine Einschränkung in der Wahl der Sprachkombination
- Natürlicher Klang der Sprachausgabe
- Lokalisiert in 33 Sprachen

- Robustheit – Keine Abstürze oder technische Störungen

Schwächen

- Drei Sprachen klingen sehr mechanisch
- Kostenpflichtiger Download der App
- Kostenpflichtiges Abonnement für erweiterte Sprachauswahl
- Favoritenfunktion etwas verwirrend

Chancen

- Verbesserung des iTunes Werbetextes
- Offline Option anbieten
- Optimierung der Favoritenfunktion

Risiken

- Trotz kostenlosem Download der App, sind In-App Käufe möglich und können zur Kostenfalle werden
- Hoher Datenverbrauch
- Datenschutz

8 Anwendungsbeispiele der SST Apps

In diesem Kapitel soll die individuelle Leistung der drei Apps mittels zweier Musterdialoge aufgezeigt werden. Die Musterdialoge spielen in zwei unterschiedlichen Kontexten und wurden mit zwei unterschiedlichen Sprachkombinationen durchgeführt. Der deutsch-englische Dialog ist in ein touristisches Setting eingebettet, während der deutsch-französische Dialog in einem geschäftlichen Umfeld angesiedelt ist. Zu Testzwecken wurde Person A ein männlicher Sprecher und Person B eine weibliche Sprecherin zugewiesen. Es wurden bewusst Charakteristika der Spontansprache eingefügt, wie beispielsweise Sprech- bzw. Denkpause (...), Selbstkorrekturen, syntaktisch unvollständige oder falsche Segmente, etc. Die Apps werden chronologisch behandelt und es wird am Ende eines jeden Testlaufs ein Kommentar zur Leistung der App geboten, jedoch keine exakte linguistische Analyse der Fehler, da dies nicht das Ziel dieser Arbeit ist. Die untenstehenden Tabellen enthalten eine Transkription des gesprochenen Ausgangstextes, der Spracherkennung und des generierten Zieltextes.

Musterdialog – Touristischer Kontext	Musterdialog – Geschäftlicher Kontext
A1: Guten Morgen! Ich hätt' gern drei Tomaten ... einen Salat ... und ein halbes Kilo Äpfel.	A1: Guten Tag, Frau Dubois. Ich möchte Sie im Namen des Unternehmens herzlich willkommen heißen.
B1: Good morning Sir. Would you like to have the large or the medium-sized tomatoes?	B1: Bonjour Monsieur Müller. Merci pour votre invitation.
A2: Ähm ... Ich denke ich nehme die Mittelgroßen. Nein, lieber nicht, geben Sie mir bitte die Großen.	A2: Haben Sie die Unterlagen für das Meeting bereits erhalten?
B2: Good choice. As for the apples, which variety would you like?	B2: Non, je n'ai pas encore reçu les documents malheureusement.
A3: Diese hier, Granny Smith.	A3: Oh, das tut mir aber leid. Ich werde mich sofort darum kümmern.
B3: OK, then we have three tomatoes, some lettuce and half a dozen apples. Here you are Sir. That's 7.60€.	B3: Merci Monsieur et pourriez-vous inclure le compte-rendu de la dernière réunion, s'il vous plaît?
A4: Machen wir 8€ daraus. Vielen Dank!	A4: Natürlich. Bitte nehmen Sie Platz. Ich bin sofort wieder da.

8.1 Testdialoge mit Jibbig 2.5.5

Musterdialog (DE-EN Int.)	Resultat Spracherkennung Jibbig 2.5.5	Resultat Zieltext Jibbig 2.5.5
Touristischer Kontext		
A1: Guten Morgen! Ich hätt' gern drei Tomaten ... einen Salat ... und ein halbes Kilo Äpfel.	A1: Guten morgen ich hätte gerne drei tomaten salat und ein halbes kilo äpfel.	A1: Good morning, I would like to have a 3 tomato salad and a half a kilo apples.
B1: Good morning Sir. Would you like to have the large or the medium-sized tomatoes?	B1: Good morning, sir. Would you like to have the large or the medium-sized tomatoes.	B1: Guten morgen mein herr. Möchten sie haben große oder der mittelgroße tomaten.
A2: Ähm ... Ich denke ich nehme die Mittelgroßen. Nein, lieber nicht, geben Sie mir bitte die Großen.	A2: Kam ich denke ich nehme den mittelgroßen nein lieber nicht geben sie mir bitte die großen.	A2: And I think I will take the medium one, better not please give me the great.
B2: Good choice. As for the apples, which variety would you like?	B2: Good choice as for the apples which variaty would you like?	B2: Gute wahl in bezug auf die äpfel die vielfalt möchten sie?
	B2.1: Good choice. As for the apples which kind would you like?	B2.1: Gute wahl. In bezug auf die äpfel welche art gefällt ihnen?
A3: Diese hier, Granny Smith.	A3: Dieser hier meins mir.	A3: This one of my me.
	A3.1: Diese hier. Granny Smith.	A3.1: This one granny Smith.
B3: OK, then we have three tomatoes, some lettuce and half a dozen apples. Here you are Sir. That's 7.60€.	B3 ²³ : Okay then we have the tomatoes, some lettuce and half a dozen apples. Here you are sir that's seven euro sixty.	B3: Okay dann haben wir die einige tomaten kopfsalat und ein halbes dutzend äpfel. Bitte schön das ist 7 60 euro ²⁴ .
A4: Machen wir 8€ daraus. Vielen Dank!	A4: Machen wir 8 euro daraus vielen dank.	A4: We do 8 euros from it, thank you.

²³ Teilung in zwei Segmente, weil zu lang für Aufnahme - Fehlermeldung.

²⁴ Preis wurde wie folgt gesprochen: *Sieben Euro sechzig*.

Musterdialog (DE-FR Fra) Geschäftlicher Kontext	Resultat Spracherkennung Jibbig 2.5.5	Resultat Zieltext Jibbig 2.5.5
A1: Guten Tag, Frau Dubois. Ich möchte Sie im Namen des Unternehmens herzlich willkommen heißen.	A1: Guten tag frau die bauer ich möchte sie im namen des unternehmens jetzt nicht willkommen heißen.	A1: Bonjour, je veux vous y a heure au nom de l'entreprise, pas d'accueillir maintenant.
	A1.1 : Guten tag frau Dubois. ich möchte sie im namen des unternehmens herzlich willkommen heißen. (manuelle Korrektur)	A1.1: Bonjour, mme dubois. Je veux vous au nom de l'entreprise, bienvenue chaud.
B1: Bonjour Monsieur Müller. Merci pour votre invitation.	B1: Bonjour monsieur mais la merci pour votre invitation.	B1: Hallo herr aber die vielen dank für ihre einladung.
	B1.1: Bonjour monsieur Müller. Merci pour votre invitation. (manuelle Korrektur)	B1.1: Hallo herr müller. Vielen dank für ihre einladung.
A2: Haben Sie die Unterlagen für das Meeting bereits erhalten?	A2: Haben sie die unterlagen für das meeting bereits erhalten?	A2: Avez vous la réunion documents déjà reçu?
B2: Non, je n'ai pas encore reçu les documents malheureusement.	B2: Non je n'ai pas encore reçu les documents malheureusement.	B2: Nein ich habe noch nicht bekommen die dokumente leider.
A3: Oh, das tut mir aber leid. Ich werde mich sofort darum kümmern.	A3: Gut das tut mir aber leid ich werde mich sofort darum kümmern.	A3: Bien mais ça me fait mal je vais m'en occuper immédiatement.
B3: Merci Monsieur et pourriez-vous inclure le compte-rendu de la dernière réunion, s'il vous plait?	B3: Merci messieurs et pourriez vous inclure le compte-rendu de la dernière réunion s'il vous plait?	B3: Danke herren und dann können sie das konto der letzten versammlung besucht bitte.
A4: Natürlich. Bitte nehmen Sie Platz. Ich bin sofort wieder da.	A4: Natürlich bitte nehmen Sie platz ich bin sofort wieder da.	A4: Bien sûr, asseyez vous s'il vous plait je suis de retour immédiatement.

8.1.1 Kommentar zur Leistung von Jibbig 2.5.5

Dialog 1 (DE-EN International) – Touristischer Kontext

Der erste Dialog wurde in der Sprachkombination Deutsch-Englisch International im Online-Modus durchgeführt. *Jibbig* war im Stande manche Charakteristika der Spontansprache bereits in der Spracherkennungsphase zu eliminieren. Einerseits wurden so unvollständige Wörter wie *hätt'* und *gern* in A1 durch die Spracherkennung grammatikalisch korrekt verbessert und Darüber hinaus auch die Denkpausen (...) gekürzt. Andererseits wurde das Füllwort *Ähm* in A2 fälschlicherweise von der Spracherkennung als *kam* identifiziert, was sich jedoch nicht weiter negativ auf die Übersetzung auswirkte. Die Spracherkennungskomponente lieferte eine solide Leistung, da der überwiegende Großteil fehlerfrei erkannt wurde. Lediglich in A3 und B3 kam es zu schwerwiegenden Defekten. In A3 konnte der Eigenname *Granny Smith* nicht als solcher erkannt werden und musste manuell korrigiert werden. Bei B3 schien die Segmentlänge nicht bewältigbar zu sein, was sich in wiederholten Fehlermeldungen wie folgender äußerte: „Überlastung. Unsere kostenlosen Übersetzer sind gerade nicht verfügbar. Bitte versuchen Sie es später erneut.“ Die Interpunktion wurde von *Jibbig* beachtet, was beispielsweise an der korrekten Verwendung von Fragezeichen und Punkt erkennbar ist. Zwei aufeinanderfolgende Sätze werden selten als solche identifiziert und stattdessen zu einem Satz zusammengefasst, was bei der Selbstkorrektur in A2 zu Verwirrung und Missverständnissen in der Übersetzung führen könnte.

Die Abhängigkeit des maschinellen Übersetzungsmoduls von der automatischen Spracherkennungskomponente wird in A1 deutlich sichtbar. In der Spracherkennung ist die Mengenangabe *einen Salat* vermutlich dem Kürzungsprozess der vorangehenden Denkpause zum Opfer gefallen und generiert dadurch eine Sinnverzerrung im ZT. Aus *drei Tomaten, einen Salat* wurde fälschlicherweise *a 3 tomato salad*. Auffällig ist, dass sich die Übersetzung sehr an der ausgangssprachlichen Syntax orientiert. Ganz besonders wird dies in A3 deutlich, wo die Spracherkennung den Eigennamen nicht erkannt hat und stattdessen andere, willkürlich erscheinende Wörter produzierte. Diese wurden wiederum in einer Form wiedergegeben, welche stark an die ausgangssprachliche Reihung angelehnt ist, und machen die Übersetzung unbrauchbar. Die Anmerkung zur Syntax gilt auch für B1, wo die falsche Positionierung des Verbs *haben* zwar einen verständlichen, jedoch syntaktisch falschen Satz entstehen lässt. Bei B2 musste in der ausgangssprachlichen Formulierung umdisponiert werden, da zwischen den Worten *apples* und *variety* kein Bezug hergestellt werden konnte und statt *Apfelsorte* die Entsprechung *Vielfalt* angeboten wurde. B2.1 liefert zwar kein kritikfestes Ergebnis, ist jedoch in der Übersetzung verständlicher als B2. An der umgangssprachlichen Formulierung in A4 ist *Jibbig* vollkommen gescheitert. Positiv anzumerken ist, dass die Sprachsynthesekomponente der App den Preis in B3 trotz falscher Transkription richtig und verständlich wiedergibt (*sieben Euro sechzig*). Während der

Aufnahme kam es, abgesehen von der Fehlermeldung aufgrund eines zu langen Segments, zu keinen Störungen, Abstürzen oder sonstigen technischen Schwierigkeiten.

Dialog 2 (DE-FR Frankreich) – Geschäftlicher Kontext

Der zweite Dialog wurde mit der Sprachkombination Deutsch-Französisch (Frankreich) durchgeführt und ist in einen geschäftlichen Kontext eingebunden. Die Spracherkennungskomponente hat beim zweiten Dialog ein zufriedenstellendes Ergebnis geliefert, da, bis auf zwei Segmente, alle korrekt analysiert und transkribiert wurden. Ein fehlerhaftes Segment ist A3, wo das Füllwort *Oh* von der Spracherkennung als *Gut* ausgegeben wurde und im Folgenden in der Übersetzung einen etwas verzerrten Sinn nach sich trägt. Bei A1 ist der App jedoch ein grober Fehler unterlaufen, da die Phrase *herzlich willkommen heißen* von der Spracherkennung als *jetzt nicht willkommen heißen* deklariert wurde und somit eine vollkommen sinnwidrige Übersetzung entstehen lässt. Zudem konnte der Eigenname *Dubois* nicht korrekt erkannt werden (*die bauer*) und wurde in der Übersetzung gar nicht inkludiert. Die gesamte Übersetzung des Segments A1 war unbrauchbar und aus diesem Grund musste eine manuelle Korrektur A1.1 erfolgen. Diese brachte auch nur begrenzten Erfolg, da nur die Anrede korrekt übersetzt wurde und der zweite Teil nach wie vor nicht zufriedenstellend war. Ebenso wurde bei B1 der Eigenname Müller nicht erkannt, jedoch lieferte die manuelle Korrektur B1.1 ein zufriedenstellendes Übersetzungsergebnis. Auffällig ist, dass der Terminus *Bonjour* in allen drei Apps mit *Hallo* übersetzt wurde. Je nach Kontext mag diese Entsprechung ihre Berechtigung haben, jedoch wäre es weitaus naheliegender die formelle Entsprechung *Guten Tag* der informellen Entsprechung *Hallo* vorzuziehen. Interpunktion wird zwar angewendet, jedoch nicht gleichbleibend konsequent. So wird A2 bereits von der Spracherkennung als Frage erkannt und in der Übersetzung beibehalten, jedoch zeigt B3, dass Fragen in der Übersetzungskomponente nicht immer als solche identifiziert werden.

Kürzere (B1.1) sowie längere (A4) Segmente können fehlerfrei und ohne weitere Segmentierung übersetzt werden. Jedoch kommt es häufig auch zu Syntaxfehlern, in denen der Satzbau viel zu sehr an die Ausgangssprache angelehnt ist (B2). Es wären alle Wörter für einen korrekten Satz vorhanden, jedoch ist die Aneinanderreihung nicht gelungen. Wie B3 beweist, gibt es auch terminologische Schwierigkeiten, weil Termini wie *compte-rendu* nicht korrekt übersetzt werden und den Sinn des Zieltextes somit stark verfälschen. Die Groß- und Kleinschreibung wurde in beiden Testdialogen von der App nicht beachtet. Während der Aufnahme kam es zu keinen Störungen, Abstürzen oder sonstigen technischen Schwierigkeiten.

8.2 Testdialoge mit VoiceTra4U 3.2

Musterdialog (DE-EN US) Touristischer Kontext	Resultat Spracherkennung VoiceTra4U 3.2	Resultat Zieltext VoiceTra4U 3.2
A1: Guten Morgen! Ich hätt' gern drei Tomaten ... einen Salat ... und ein halbes Kilo Äpfel.	A1: Guten morgen ich hätte gerne drei tomaten einen salat und ein halbes kilo äpfel.	A1: Good morning I'd like three tomato salad and a half a kilo apples
B1: Good morning Sir. Would you like to have the large or the medium-sized tomatoes?	B1: Recognition could not be performed. Please try speaking once again. ➔Manuelle Texteingabe	B1: Guten morgen mein herr. Möchten sie die groß oder mittelgroßer tomaten.
A2: Ähm ... Ich denke ich nehme die Mittelgroßen. Nein, lieber nicht, geben Sie mir bitte die Großen.	A2 ²⁵ : kann ich denke ich nehme die mittelgroßen. Nein lieber nicht geben sie mir bitte die großen.	A2: Can I think I'll have the middle size? No I'd rather not give me the big.
B2: Good choice. As for the apples, which variety would you like?	B2: Good choice. as for the apples whichever I eat tea would you like.	B2: Gute wahl. Wie äpfel für welches ich esse tee möchten sie.
	B2.1: Good choice. As for the apples, which variety would you like? (manuell)	B2.1: Gute Wahl für die wie Äpfel welche auswahl möchten sie.
A3: Diese hier, Granny Smith.	A3: Diese hier eine zehn min.	A3: These a ten.
	A3.1: Diese hier, Granny Smith. (manuelle Eingabe)	A3.1: These smith.
B3: OK, then we have three tomatoes, some lettuce and half a dozen apples. Here you are Sir. That's 7.60€.	B3 ²⁶ : Okay then we have three tomatoes some lettuce and half a dozen apples stay here you are sir. That's seven euros sixty.	B3: Okay wir haben drei einige tomaten eisbergsalat und ein halbes dutzend äpfel hier bleiben sie mein herr. Das ist sieben sechzig.
A4: Machen wir 8€ daraus. Vielen Dank!	A4: Machen wir acht euro daraus vielen dank.	A4: Let's take eight total that thank you very much.

²⁵ Teilung in zwei Segmente, da Sprachaufnahme automatisch abbricht.

²⁶ Teilung in zwei Segmente, da Sprachaufnahme automatisch abbricht.

Musterdialog (DE-FR) Geschäftlicher Kontext	Resultat Spracherkennung VoiceTra4U 3.2	Resultat Zieltext VoiceTra4U 3.2
A1: Guten Tag, Frau Dubois. Ich möchte Sie im Namen des Unternehmens herzlich willkommen heißen.	A1: Guten tag frau deborah ich möchte sie im namen des unternehmens leipzig willkommen heißen.	A1: Bonjour mademoiselle je voudrais vous au nom de la société du bienvenue.
	A1.1: Guten Tag, Frau Dubois. Ich möchte Sie im Namen des Unternehmens herzlich willkommen heißen. (manuelle Eingabe)	A1.1: Bonjour mademoiselle je voudrais vous dans le nom de la société du bienvenu.
B1: Bonjour Monsieur Müller. Merci pour votre invitation.	B1: bonjour , monsieur Muller , merci de votre invitation .	B1: Hallo herr danke für ihre einladung.
A2: Haben Sie die Unterlagen für das Meeting bereits erhalten?	A2: Haben sie die unterlagen für das medizin bereits erhalten.	A2: Avez-vous déjà eu les documents pour le médicament.
	A2.1: Haben Sie die Unterlagen für das Meeting bereits erhalten ? (manuelle Eingabe)	A2.1: Avez-vous déjà eu le document pour la réunion.
B2: Non, je n'ai pas encore reçu les documents malheureusement.	B2: non , je n'ai pas encore reçu les documents malheureusement.	B2: Nein ich noch nicht erhalten die dokumente ich fürchte.
A3: Oh, das tut mir aber leid. Ich werde mich sofort darum kümmern.	A3: Oh es tut mir leid <.> ich werde mich sofort darum kümmern.	A3: Oh je suis désolé <.> je vais m'occuper de ça tout de suite.
B3: Merci Monsieur et pourriez-vous inclure le compte-rendu de la dernière réunion, s'il vous plait?	B3: Merci messieurs et pourriez-vous à inclure le compte-rendu de la dernière réunion , s'il vous plait.	B3: Danke herren und könnten sie die beinhaltet der compte-rendu letzten treffen bitte.
A4: Natürlich. Bitte nehmen Sie Platz. Ich bin sofort wieder da.	A4: Natürlich bitte nehmen sie platz ich bin sofort wieder da.	A4: Bien sûr je vous en prie prenez un siège je suis de retour tout de suite.

8.2.1 Kommentar zur Leistung von VoiceTra4U 3.2

Dialog 1 (DE-EN US) – Touristischer Kontext

Der Testdialog wurde mit der Sprachkombination Deutsch-Englisch UK begonnen und musste aufgrund schwerwiegender technischer Probleme, in der Annahme, dass es sich um Verbindungsschwierigkeiten mit dem Server handelt, zunächst auf einen anderen Tag verschoben werden, bevor diese Sprachkombination komplett aufgegeben wurde. Die Bearbeitungsdauer der Spracherkennung war ungewöhnlich lang und es dauerte 10 bis 15 Sekunden bis die App meist mit einer Fehlermeldung reagierte: „*Error speech recognition is not available now* oder *Recognition complications Please try speaking once again.*“ Nach einigen Strategiewechseln, wie z.B. Segmentteilung, und Versuchen an unterschiedlichen Tagen wurde schließlich die Sprachkombination auf Deutsch-Englisch US geändert, womit auch die Fehlermeldungen ein Ende hatten. Allgemein kann gesagt werden, dass die *VoiceTra4U* App in beiden Testdialogen merkbar mehr Zeit beanspruchte, als die beiden anderen Apps. Die Ergebnisse der Spracherkennungskomponente sind sehr durchwachsen und reichen von Totalausfällen wie in B1, aufgrund eines zu langen Segments, und A3, aufgrund eines Eigennamens, bis hin zu nahezu perfekten Resultaten wie in A1 und B3. Die Beispiele A1 und B1 spiegeln die Natur der Spracherkennung von *VoiceTra4U* sehr gut wieder. Während das Segment in A1, trotz der Denkpausen und spontansprachlichen Formulierung, auf der Wortebene perfekt erkannt wurde, konnte das ca. gleich lange Segment in B1 (trotz späterer Segmentierung) partout nicht erkannt werden und erforderte eine manuelle Texteingabe. Der Eigenname in A3 konnte ebenso wie bei *Jibbiggo* von der Spracherkennungskomponente nicht erkannt werden, jedoch ist das Übersetzungsergebnis selbst nach der manuellen Eingabe nicht zufriedenstellend.

Im Laufe des Tests wurde ersichtlich, dass *VoiceTra4U* vor allem mit der Übersetzung von Geldbeträgen Probleme hat. So wurde die Währung in B3 und A4 von der Spracherkennung in beiden Sprachen zwar richtig erkannt, jedoch sehr willkürlich übersetzt. In B3 wurde die Währung im ZT schlicht und ergreifend ausgelassen und bei einem späteren Test stellte sich heraus, dass die App mit einer Fehlermeldung reagiert, wenn versucht wird *Euro* vom Englischen (US & UK) ins Deutsche zu übersetzen: „*Translation complications. Please try saying it again in a different way.*“ In der umgekehrten Sprachrichtung schlägt die App für *Euro* die englische Übersetzung *total* vor (vgl. A4). Bedingt durch die zahlreichen Fehler der Spracherkennung waren auch die Resultate der maschinellen Übersetzungskomponente qualitativ wenig zufriedenstellend, da häufig unbrauchbar wie z.B. bei A1, B2, A3, A3.1, B3, A4. Gerade die korrekte Übersetzung von Geldbeträgen und Währungen sollte in einer universal einsetzbaren App gewährleistet sein. Es ist auffällig, dass in der App prinzipiell alles, außer Satzanfänge, klein geschrieben wird und die Interpunktion, trotz manueller Eingabe in der Übersetzung nicht korrekt übernommen wird (vgl. B2.1).

Dialog 2 (DE-FR) – Geschäftlicher Kontext

Der zweite Dialog der *VoiceTra4U* App wurde mit der Sprachkombination Deutsch-Französisch durchgeführt. Da sich die App hauptsächlich auf asiatische Sprachen und Dialekte konzentriert, ist es nicht verwunderlich, dass es keine Spezifikation bezüglich der französischen Sprache gibt. Allgemein kann gesagt werden, dass es während dieser Aufnahme zu keinen Störungen, Abstürzen oder sonstigen technischen Schwierigkeiten kam, jedoch wieder lange Wartezeiten bei der Spracherkennung und folglich bei der Erstellung von Übersetzungen eintraten. Die Leistung der Spracherkennung ist zum wiederholten Male durchgewachsen. So wurden B2, A3, B3, A4 gut analysiert und korrekt erkannt, während A2 durch den Fehler in der Spracherkennung eine Sinnveränderung in der Übersetzung mit sich bringt (*Medizin* anstelle von *Meeting*). Interessanterweise wurde der Eigenname Dubois (A1) nicht erkannt, der Name Müller (B1) jedoch sehr wohl. Umso überraschender ist die Tatsache, dass der bereits transkribierte Nachname Müller keinen Weg in die Übersetzung findet. Die Interpunktion, der von der Spracherkennung transkribierten französischen Sätze, ist insofern interessant, als Leerzeichen vor und nach dem Komma aufzufinden sind (vgl. B1, B2, B3). Dies ist wohl eine Anlehnung an die französischen Typographienormen, die üblicherweise vor einem Fragezeichen, Ausrufezeichen, Doppelpunkt und Semikolon eine Leertaste verlangen, jedoch ist dies bei einem Komma nicht der Fall. Eine sinnvolle Antwort auf die Frage, warum der Punkt im deutschen A3 Segment zwischen ein Kleiner-als-Zeichen (<) und ein Größer-als-Zeichen (>) gesetzt wurde, konnte nicht gefunden werden.

Der Großteil der Fehler entstand dennoch im maschinellen Übersetzungsprozess. So wird das Segment A1 bzw. A1.1 trotz manueller Eingabe ohne den Nachnamen und stark angelehnt an die Ausgangssprachliche Syntax übersetzt. Darüber hinaus wäre es ratsam den Terminus *Madame* dem Terminus *Mademoiselle* als Entsprechung für die Anrede *Frau* vorzuziehen. In B3 sind beispielsweise grobe terminologische Mängel erkennbar, da der Terminus *compte-rendu* erst gar nicht übersetzt wurde, sondern einfach in den Zielsprachlichen Satz eingefügt wurde. Abgesehen davon ist die Übersetzung unbrauchbar, trotz der guten Leistung von Seiten der Spracherkennung. Auch Segment A4 beinhaltet trotz guter Analyse durch die Spracherkennung an einer Stelle eine wortwörtliche Übersetzung, was die Aussage des gesamten Zielsatzes nicht eindeutig verständlich wirken lässt.

8.3 Testdialoge mit SayHi Translate 3.8.2

Musterdialog (DE-EN UK) Touristischer Kontext	Resultate Spracherkennung SayHi Translate 3.8.2	Resultate Zieltext SayHi Translate 3.8.2
A1: Guten Morgen! Ich hätt' gern drei Tomaten ... einen Salat ... und ein halbes Kilo Äpfel.	A1: Guten Morgen ich hätte gern drei Tomaten einen Salat und ein halbes Kilo Äpfel	A1: Good morning, I would like three tomatoes a salad and half a kilo of apples
B1: Good morning Sir. Would you like to have the large or the medium-sized tomatoes?	B1: Good morning sir would you like to have the large or the medium-sized tomatoes	B1: Guten Morgen, Sir möchten Sie die große oder die mittelgroße Tomaten zu haben,
A2: Ähm ... Ich denke ich nehme die Mittelgroßen. Nein, lieber nicht, geben Sie mir bitte die Großen.	A2: Ich denke ich nehme die mittelgroßen nein lieber nicht geben Sie mir bitte die großen	A2: Please I think I'll take the midsize no better not give me the big
B2: Good choice. As for the apples, which variety would you like?	B2: Good choice as for the apples which variety would you like	B2: Gute Wahl, da für die Äpfel, die Vielfalt möchten Sie
	B2.1: Good choice. As for the apples which kind would you like? (manuelle Korrektur)	B2.1: Gute Wahl. Wie für die Äpfel, die Art hätten Sie gerne?
A3: Diese hier, Granny Smith.	A3: Diese hier granny Smith	A3: This here granny Smith
B3: OK, then we have three tomatoes, some lettuce and half a dozen apples. Here you are Sir. That's 7.60€.	B3: Okay then we have three tomatoes some leathers and half a dozen apples here you are sir that is €7.60	B3: Okay, dann haben wir drei Tomaten einige Leder und ein halbes Dutzend Äpfel hier, Sir, das sind € 7,60 ²⁷
A4: Machen wir 8€ daraus. Vielen Dank!	A4: Machen wir acht Euro daraus vielen Dank	A4: Let's make it eight euro thanks

²⁷ Preis wurde wie folgt gesprochen: Sieben Euro und sechzig Cent.

Musterdialog (DE-FR Fra) Geschäftlicher Kontext	Resultat Spracherken- nung SayHi Translate 3.8.2	Resultat Zieltext SayHi Translate 3.8.2
A1: Guten Tag, Frau Dubois. Ich möchte Sie im Namen des Unternehmens herzlich willkommen heißen.	A1: Guten Tag Frau die Bohr ich möchte Sie im Namen des Unternehmens herzlich willkommen heißen	A1: Bonjour Mme Bohr Je vous souhaite la bienvenue au nom de la société
	A1.1: Guten Tag Frau Dubois ich möchte Sie im Namen des Unternehmens herzlich willkommen heißen	A1.1: Bonjour Mme Dubois, je voudrais vous souhaiter la bienvenue au nom de la société
B1: Bonjour Monsieur Müller. Merci pour votre invitation.	B1: Bonjour Monsieur Müller merci pour votre invitation	B1: Hallo Herr Müller ich danke Ihnen für Ihre Einladung
A2: Haben Sie die Unterlagen für das Meeting bereits erhalten?	A2: Haben Sie die Unterlagen für das Meeting bereits erhalten	A2: Avez-vous déjà reçu les documents de la réunion
B2: Non, je n'ai pas encore reçu les documents malheureusement.	B2: Non je n'ai pas encore reçu les documents malheureusement	B2: Nein, ich habe noch nicht die Dokumente leider eingegangen
A3: Oh, das tut mir aber leid. Ich werde mich sofort darum kümmern.	A3: Oh das tut mir aber leid ich werde mich sofort darum kümmern	A3: Oh, mais je suis désolé, je vais m'en occuper immédiatement
B3: Merci Monsieur et pourriez-vous inclure le compte-rendu de la dernière réunion, s'il vous plait?	B3: Merci messieurs et pourriez-vous inclure le compte rendu de la dernière réunion, s'il vous plait	B3: Vielen Dank, dass Sie, meine Herren und könnten Sie das Protokoll der letzten Sitzung geben Sie bitte
A4: Natürlich. Bitte nehmen Sie Platz. Ich bin sofort wieder da.	A4: Natürlich bitte nehmen Sie Platz ich bin sofort wieder da.	A4: Bien sûr, s'il vous plaît prendre un siège, je serai bientôt de retour

8.3.1 Kommentar zu Leistung von SayHi Translate 3.8.2

Dialog 1 (DE-EN UK) – Touristischer Kontext

Der erste Dialog wurde bei *SayHi* Translate mit der Sprachkombination Deutsch-Englisch UK durchgeführt. Die Leistungen der Spracherkennungskomponente von *SayHi* waren deutlich besser als die der beiden anderen Apps. *SayHi* hat von den sieben gesprochenen Segmenten sechs fehlerfrei transkribiert, wobei das nicht vollkommen korrekt transkribierte Segment B3 lediglich ein falsches Wort in dem doch vergleichsweise langen Segment aufwies. Die App war als einzige im Stande den Eigennamen *Granny Smith* zu erkennen und auch korrekt zu übersetzen. Zudem wurde auch der gesprochene Geldbetrag korrekt transkribiert. Die spontansprachlichen Elemente wie Wortkürzungen und Denkpausen in A1 sowie Füllwörter in A2 konnten bereits in der Spracherkennungsphase eliminiert werden und hatten keinen weiteren Einfluss auf die Übersetzung. Darüber hinaus muss positiv angemerkt werden, dass *SayHi* als einzige App Rücksicht auf Groß- und Kleinschreibung nimmt. Nicht nur Nomen sondern auch die Höflichkeitsform *Sie* wird korrekterweise großgeschrieben. Lediglich der Eigenname *Granny Smith* wurde zwar korrekt von der Spracherkennungskomponente verarbeitet, jedoch wurde nur *Smith* mit einem großen Anfangsbuchstaben versehen.

Die maschinelle Übersetzungskomponente liefert durchaus zufriedenstellende Ergebnisse. Die Resultate von A1, A3 A4 sind grammatikalisch betrachtet vollkommen korrekt und eindeutig verständlich. Auch B3 ist, abgesehen von dem Fehler in der Spracherkennung (*lettuce* – *leather*), einwandfrei und enthält in der Preisangabe sogar das deutschsprachige Komma anstelle des englischen Dezimalpunkts. Darüber hinaus wurde auch die Preisangabe von der Sprachsynthese vorbildlich wiedergegeben (*sieben Euro sechzig Cent*). Lediglich die Übersetzung in B2 und B2.1 ist als wenig verständlich einzustufen und müsste in der Ausgangsprache paraphrasiert werden, um einen verständlichen Zieltext zu generieren. Auffällig ist, dass in der Spracherkennung von Beistrichen oder Punkten kein Gebrauch gemacht wird, während im ZT Beistriche inkludiert werden, jedoch keine Punkte. Bei einer manuellen Korrektur werden die Satzzeichen sowie das Fragezeichen übernommen (vgl. B2.1). Während der Aufnahme kam es zu keinen Störungen, Abstürzen oder sonstigen technischen Schwierigkeiten.

Dialog 2 (DE-FR Frankreich) – Geschäftlicher Kontext

Der zweite Dialog wurde mit der Sprachkombination Deutsch-Französisch (Frankreich) durchgeführt. Die Leistung der Spracherkennung hat sich im Vergleich zum ersten Dialog gesteigert. Lediglich der Eigenname *Dubois* in A1 konnte nicht eindeutig erkannt werden. Stattdessen transkribierte die Spracherkennungskomponente den Wortlaut *die Bohr*, welcher in der Übersetzung zu *Mme Bohr* verarbeitet wurde. Trotz des per se falschen Über-

setzungsergebnis ist es verblüffend, dass *SayHi* das fehlerhaft transkribierte Segment zu einem grammatikalisch und syntaktisch einwandfreien Zielsatz verarbeitet. Ein anderer Defekt der von der Spracherkennungskomponente generiert wurde und auch bei den anderen Apps vorgekommen ist, ist das Wort *Messieurs* anstelle von *Monsieur*. Dieses wird auch in der Übersetzung im Plural übernommen und verzerrt den Sinn des bereits syntaktisch nicht optimalen Zielsatzes. Wie in A4 ersichtlich, hat *SayHi* mit dem Segmentteil *nehmen Sie Platz* das gleiche Problem wie *VoiceTra4U* und übersetzt das Element wortwörtlich. *SayHi* liefert insgesamt vier völlig korrekte Übersetzungen (A1.1, B1, A2, A3).

Wie auch beim ersten von *SayHi* bearbeiteten Dialog finden sich in der Spracherkennungsphase wenige bis keine Satzzeichen, während im Zieltext eine sinnvolle Verwendung von Kommas zu beobachten ist. Die Groß- und Kleinschreibung wurde auch in diesem Dialog fehlerfrei umgesetzt. Während der Aufnahme kam es zu keinen Störungen, Abstürzen oder sonstigen technischen Schwierigkeiten. Ein nicht zu verachtender Vorteil der *SayHi* App im Hinblick auf die User-Experience ist die Natürlichkeit der Sprachausgabe.

9 Schlussbetrachtung

9.1 Funktionsüberblick Apps

Ziel dieser Arbeit war es, einerseits die drei Speech-to-Speech Translation Apps auf ihren Funktionsumfang hin genau zu analysieren und andererseits die Stärken und Schwächen der einzelnen Apps herauszufiltern. Dazu wurde ein Funktionsüberblick erstellt:

	Jibbig Translator 2.5.5	VoiceTra4U 3.2	SayHi Translate 3.8.2
Logo	 Abbildung 49: Logo Jibbig Translator 2.5.5 (vgl. https://itunes.apple.com)	 Abbildung 50: Logo VoiceTra4U 3.2 (vgl. https://itunes.apple.com)	 Abbildung 51: Logo SayHi Translate 3.8.2 (vgl. https://itunes.apple.com)
Entwickler	JIBBIGO LLC. Alex WAIBEL – Gründer und Chairman, ehemaliger Mitarbeiter und Deputy Scientific Director von <i>VerbMobil</i>	U-STAR Consortium Universal Speech Translation Advanced Research Consortium Internationale Forschungsvereinigung	SayHi LLC CEO Leo BOSSIO – konzipierte die Software ursprünglich für den Einsatz im Krankenhaus.
Betriebssystem	- ab Android 2.3 - ab iOS 6.1	- ab Android 2.2 - ab iOS 6.0	ab iOS 6.1
Erscheinungsjahr	2009 (als erstes seiner Art)	2012	2011
Größe	14,6 MB	15,5 MB	16,1 MB
Preis	App kostenlos Download der offline nutzbaren Sprachpakete ebenfalls kostenlos	App kostenlos Nutzung kostenlos	Standard App € 1,49 Premium Version: € 2,69 für 3 Monate € 4,49 für 6 Monate
FEATURES:	JIBBIGO Translator 2.5.5	VoiceTra4U 3.2	SayHi Translate 3.8.2
Sprachen	25 Sprachen und Dialekte	31 Sprachen und Dialekte	Standard: 40 Sprachen und Dialekte Premium: 105 Sprachen und Dialekte
Input:			
- Text - Spracheingabe	JA JA – für 16 Sprachen	JA JA – für 17 Sprachen	JA JA – für 40 Sprachen (Standard) JA – für 72 Sprachen (Premium)
- Handschrift - Foto (OCR²⁸)	NEIN NEIN	NEIN NEIN	NEIN NEIN

²⁸ OCR = Optical caracter reader (optische Zeichenerkennung über die Handykamera)

Output:

• Text	JA	JA	JA
• Sprache	JA – für 16 Sprachen	JA – für 14 Sprachen	JA – für 37 Sprachen (Standard) JA – für 78 Sprachen (Premium)

Übersetzer-Modus

(2 SprecherInnen mit einem einzigen Gerät)

	JA	JA	JA
--	----	----	----

Chat-Modus (mehrere Geräte)	NEIN	JA	NEIN
---------------------------------------	------	----	------

Offline-Nutzung	JA	NEIN	NEIN
------------------------	----	------	------

Post-Editing	JA	JA – jedoch wird neues Übersetzungssegment erstellt, nicht vorheriges korrigiert	JA
---------------------	----	--	----

Gesprächssituation:

• Telefongespräch (Verbomobil)	NEIN	NEIN	NEIN
• SMS-Konversation	NEIN	JA (Chatfunktion)	NEIN
• Face-to-face-Gespräch	JA	JA	JA

History	JA (jedoch nur ZT Ausgabe)	NEIN (Single Mode) JA (Chat Mode)	JA
----------------	----------------------------	--------------------------------------	----

Favoriten	NEIN	NEIN	JA
------------------	------	------	----

Zusatzfunktionen	• Einziger Anbieter einer Offline-Funktion • Sprechtempo und Lautstärke regulierbar • Teilen-Funktion: Inhalte per E-Mail oder SMS weiterleiten • Rückübersetzung • AT und ZT verwerfen durch Schütteln des Geräts • Textfeld-Aufnahme • Signal-Töne bei Aufnahme • Sprachsynthese deaktivierbar • Kraftausdrücke (nicht) erlauben • Aufnahme abspielen • Adaption	• Einziger Anbieter einer Chatfunktion für bis zu fünf Personen • Rückübersetzung • What-Icon: Feedback an GesprächspartnerIn • Automatisches Erkennen der Ansprache	• Individuell anpassbare Favoriten • Sprechtempo und Lautstärke regulierbar Verknüpfung • Teilen-Funktion: Inhalte per E-Mail oder SMS weiterleiten • Wahl zwischen männlicher und weiblicher Ausgabestimme • Signaltöne bei Aufnahme • Wiedergabe der nationalen Sprachbezeichnung bei Sprachauswahl • Aufnahme abspielen • Tonaufzeichnung • Ende der Ansprache erkennen
-------------------------	--	---	--

HOME PAGE	http://www.jibbig.com	http://www.ustar-consortium.com/app/app.html	http://www.sayhitranslate.com
------------------	---	---	---

9.2 Beantwortung der Forschungsfragen

Bei der vorliegenden Arbeit handelt es sich um eine vergleichende Analyse von drei verschiedenen Speech-to-Speech Translation App, die Antworten auf folgende Forschungsfragen finden sollte:

- Forschungsfrage 1: Über welchen Funktionsumfang verfügen die einzelnen Speech-to-Speech Translation Apps?
- Forschungsfrage 2: Worin liegen die Stärken und Schwächen der einzelnen Speech-to-Speech Translation Apps?

Zur Beantwortung der Fragen, wird einerseits die gegenüberstellende Übersicht über die Funktionen der einzelnen Apps herangezogen, sowie die individuellen SWOT-Analysen.

Ad Forschungsfrage 1

Die grundlegende Funktion der drei SST Apps ist dieselbe: Einen gesprochenen Text in der Ausgangssprache per automatischer Spracherkennung analysieren, durch maschinelle Übersetzungsprozesse in eine Zielsprache übersetzen und im nächsten Schritt mithilfe von Sprachsynthese in der Zielsprache wiedergeben. Jedoch unterscheiden sich die Apps in der Art und Weise der Umsetzung, ihrer Individualisierbarkeit und den Zusatzfunktionen sehr.

Der *Jibbig Translator 2.5.5* ist für Smartphones mit den Betriebssystemen iOS und Android kostenlos erhältlich. Zudem ist auch die spätere Nutzung der App vollkommen kostenlos. *Jibbig* hat insgesamt 25 Sprachen und Dialekte im Angebot, jedoch ist die Kombinationsmöglichkeit der einzelnen Sprachen insofern eingeschränkt, als jede Ausgangssprache nur eine begrenzte Anzahl von kombinierbaren Zielsprachen hat. Weit verbreitete Sprachen wie Englisch haben mehr Zielsprachen zur Auswahl als kleinere Sprachen wie Deutsch. Die Ausgangstexteingabe kann für alle Sprachen per manueller Texteingabe erfolgen, jedoch nur für 16 Sprachen über die Spracherkennung. Die Sprachausgabe erfolgt jedenfalls per Textausgabe und wiederum nur für 16 Sprachen per Sprachsynthese. Es handelt sich dabei um dieselben 16 Sprachen, was bedeutet, dass die restlichen 9 Sprachen nur über eine maschinelle Übersetzungskomponente verfügen. Die App ist für die Anwendung von einer oder zwei Personen, die sich in unmittelbarer Nähe befinden, konzipiert. Es sind auch mehr Personen möglich, vorausgesetzt das Gerät wird herumgereicht. Es existiert kein Chat Modus per se, jedoch können die Segmente mittels der Teilen-Funktion per E-Mail oder SMS versendet werden. Die App besitzt eine Verlaufsfunktion, bei der der ZT wiedergeben werden kann, jedoch keine Favoritenfunktion. *Jibbig* bietet im Vergleich zu den anderen Apps den höchsten Grad an Individualisierbarkeit, durch eine Vielzahl von (de)aktivierbaren Einstellungen. Dazu gehören: Sprechtempo und

Lautstärkeanpassung, Rückübersetzung, Textfeld-Aufnahme, Signaltöne bei Aufnahme, Sprachsynthese, Kraftausdrücke erlauben, Aufnahme abspielen, Adaption. Darüber hinaus ist *Jibbiggo* der einzige Anbieter einer vollwertigen Offline Übersetzungsfunktion. Ebenfalls können im Offline-Modus häufig vorkommende Namen manuell hinzugefügt und deren Aussprache definiert werden.

Die App *VoiceTra4U 3.2* ist ebenfalls für Geräte mit dem Betriebssystem Android oder iOS kostenlos verfügbar. *VoiceTra4U* bietet 31 verschiedene Sprachen und Dialekte, die untereinander beliebig kombinierbar sind und eine starke Fokussierung auf den asiatischen Sprachmarkt aufweisen. Die Spracheingabe kann per manueller Texteingabe oder bei 17 Sprachen auch per Spracherkennung erfolgen. Die Sprachausgabe erfolgt jedenfalls per Textausgabe am Display und bei 14 Sprachen auch per Sprachsynthese. Eine Sprache, die über eine Spracherkennungskomponente verfügt, besitzt nicht zwangsläufig eine Sprachsynthesekomponente. Manchen Sprachen steht keine der beiden Komponenten zur Verfügung. Die App bietet die Möglichkeit eines Übersetzer-Modus, in dem zwischen zwei oder mehreren Personen mittels eines Geräts übersetzt wird. Darüber hinaus wird auch ein Chat-Modus angeboten, in dem bis zu fünf Personen auf fünf unterschiedlichen Geräten miteinander kommunizieren können. Die App kann offline nicht genutzt werden und verfügt auch über keine Favoritenfunktion. Eine Verlaufsfunktion existiert nur im Chat-Modus, jedoch nicht im Übersetzer-Modus. Die Anzeige einer Rückübersetzung des zielsprachlichen Segments in die Ausgangssprache ist optional wählbar. Zudem kann durch das *What*-Icon die Unverständlichkeit der Übersetzung ausgedrückt werden.

SayHi Translate 3.8.2 ist nur für das iPhone (iOS) verfügbar, wobei es eine extra für das iPad konzipierte App gibt. Der Download der App ist kostenpflichtig und darüber hinaus kann app-intern ein Abonnement (3 oder 6 Monate) abgeschlossen werden, welches ein erweitertes Sprachangebot mit sich bringt. Die App verfügt in der Standard Version über 40 Sprachen und Dialekte, und in der Premium Version insgesamt 105 Sprachen und Dialekte. Die Spracheingabe kann sowohl per Texteingabe als auch für 40 Sprachen (Standard) bzw. insgesamt 72 Sprachen (Premium) per Spracherkennung erfolgen. Die Sprachausgabe erfolgt wie bei den anderen beiden Apps immer per Textausgabe und für 37 (Standard) bzw. insgesamt 78 Sprachen (Premium) auch per Sprachsynthese. Die App verfügt, ebenso wie *Jibbiggo*, nur über den traditionellen Übersetzer-Modus, kann diese Resultate jedoch ebenfalls per E-Mail oder SMS teilen. Durch das von den anderen beiden Apps abweichende Layout hat *SayHi* eine inkludierte Verlaufsfunktion und eine stark individualisierbare Favoritenfunktion. Zu den weiteren Zusatzfunktionen, die zum Teil (de)aktivierbar sind, gehören: Anpassung von Sprechtempo und Lautstärke, Wahl zwischen männlicher und weiblicher Ausgabestimme, Signaltöne bei Aufnahme, Nennung der nationalen Sprachbezeichnung bei Sprachauswahl, Aufnahme abspielen, Tonaufzeichnung, Ende der Ansprache erkennen. *SayHi* bedarf ebenfalls einer ständigen Internetverbindung.

Ad Forschungsfrage 2

Die größte Stärke des *Jibbig Translator 2.5.5* liegt unbestritten in der Verfügbarkeit eines offline Übersetzungsangebots. Während die anderen Apps auf eine stabile Internetverbindung angewiesen sind, kann *Jibbig* nach der Installation der entsprechenden Sprachpakete auch offline jederzeit vollkommen kostenlos genutzt werden. Zudem besticht *Jibbig* durch die Individualisierungsmöglichkeiten in den Einstellungen und eine robuste Leistung. Als Schwäche der App kann die Einschränkung in der Wahl der Sprachkombinationen gesehen werden, da alle anderen Apps eine beliebige Kombination zulassen. Darüber hinaus hat auch die Offline-Version eine Schattenseite, die der Natur des Mediums Smartphone geschuldet ist. Smartphones haben nur eine begrenzte interne Speicherkapazität, die beim Download der offline Sprachpakete bedacht werden muss.

Der Vorteil der *VoiceTra4U* App liegt eindeutig in der Chat-Funktion, die es ermöglicht mit bis zu fünf Personen ausgehend von fünf verschiedenen Geräten gleichzeitig und ortsunabhängig zu chatten. Zudem verfügt die App über eine Funktion, die es ermöglicht, den GesprächsteilnehmerInnen mitzuteilen, dass der Zieltext unverständlich ist und sie auffordert den Ausgangssatz zu paraphrasieren. Das größte Manko der *VoiceTra4U* App ist wohl der Zwang zu einer stabilen Internetverbindung, da sonst immer wieder Fehlermeldungen angezeigt werden und keine sinnvollen Übersetzungen generiert werden können. Zudem wurde die App nur ins Japanische und Englische lokalisiert, was für fremdsprachige Menschen ein grundlegender Nachteil sein kann.

Die Stärke von *SayHi* liegt in der Sprachvielfalt, da bereits die Standard Version mit 40 Sprachen das größte Angebot hat, jedoch muss bedacht werden, dass es auch die einzige App mit kostenpflichtigem Download und kostenpflichtigen Erweiterungen ist. Ein weiterer Vorteil ist, dass die App in 33 der angebotenen Sprachen lokalisiert wurde und somit eine breitere Masse anspricht. Zudem war die App äußerst robust und sicherte dadurch eine angenehme User-Experience. Dazu trägt auch die Natürlichkeit der Sprachausgabe bei. Der größte Schwachpunkt ist eindeutig der Kostenfaktor.

9.3 Conclusio

Das Ziel dieser Arbeit war es drei verschiedene Speech-to-Speech Translation Apps in ihrem Funktionsumfang miteinander zu vergleichen und die Stärken und Schwächen der einzelnen Apps zu erarbeiten. Zu diesem Zweck wurden zunächst im ersten Kapitel die Meilensteine in der Forschung zur Spracherkennung, Sprachsynthese und vor allem zur maschinellen Übersetzung vorgestellt. Nach dem historischen Abriss stellt das zweite Ka-

pitel eine detaillierte Einführung in die einzelnen Komponenten eines Speech-to-Speech Translation Systems dar, um die Arbeitsweisen, die Zusammenhänge zwischen den Phasen und potentielle Problemfelder nachvollziehen und verstehen zu können. Das dritte Kapitel beschreibt zwei Pionierprojekte der SST und illustriert den rasanten Fortschritt der Sprachtechnologien. Mit dem vierten Kapitel findet der theoretische Teil sein Ende und es wird die Herangehensweise beim funktionalen Vergleich der SST Apps beschrieben, der der Beantwortung der beiden Forschungsfragen dienen soll:

- Forschungsfrage 1: Über welchen Funktionsumfang verfügen die einzelnen Speech-to-Speech Translation Apps?
- Forschungsfrage 2: Worin liegen die Stärken und Schwächen der einzelnen Speech-to-Speech Translation Apps?

Die Kapitel fünf, sechs und sieben beinhalten die funktionale Analyse der drei Apps *Jibbiggo Translator 2.5.5*, *VoiceTra4U 3.2* sowie *SayHi Translate 3.8.2*. Dabei werden die Aspekte technische Fakten, User Interface, Features sowie Benutzerfreundlichkeit beleuchtet und abschließend mittels einer SWOT-Analyse zusammengefasst. Das achte Kapitel soll mithilfe zweier Testdialoge je aus dem touristischen und dem geschäftlichen Kontext die übersetzerische Leistung der drei Apps aufzeigen. Dabei wurden jeweils Kommentare und Eindrücke formuliert, jedoch keine konkrete linguistische Analyse gegeben, da dies nicht das vorrangige Ziel dieser Arbeit ist. Das letzte und abschließende neunte Kapitel bietet eine Gegenüberstellung der Funktionen der einzelnen Apps und beinhaltet zudem der Beantwortung der Forschungsfragen sowie weiterführende Gedanken und Zukunftsaussichten für im Bereich der Sprachtechnologien und insbesondere der Speech-to-Speech Translation Systeme.

Abschließend kann gesagt werden, dass alle untersuchten Apps solide Produkte für den Massenmarkt sind. Da die Erwartungen an die Leistung der Apps aufgrund meiner professionellen Tätigkeit als Übersetzerin und negativer Erfahrungen mit Google Translate niedrig angesetzt waren, konnten mich die Apps dennoch positiv überraschen. Besonders hervorgestochen ist die allgemeine Leistung der Spracherkennung, wobei angemerkt werden muss, dass *Jibbiggo* und *SayHi* bessere Ergebnisse lieferten als *VoiceTra4U*. Der Großteil der Ursachen für schlechte Übersetzungsergebnisse lag in der maschinellen Übersetzungskomponente. Positiv aufgefallen ist auch die Natürlichkeit der Sprachausgabe bei der App *SayHi*. Je nach Anwendungsbereich und Anforderungen ist eine der Apps passender als die anderen beiden. So bringt *Jibbiggo* beispielsweise den höchsten Grad an Individualisierbarkeit und Zusatzfunktionen mit und bietet zudem als einzige App einen offline nutzbaren Übersetzer-Modus. Soll mit mehreren Personen gleichzeitig und ortsunabhängig kommuniziert werden, ist *VoiceTra4U* die beste Wahl, da auch ein Chat Mode für bis zu fünf GesprächsteilnehmerInnen geboten wird. Darüber hinaus bietet *VoiceTra4U* eine be-

sondere Spezialisierung auf asiatische Sprachen. Jedoch muss darauf hingewiesen werden, dass eine ständige und vor allem stabile Internetverbindung nötig ist, da der netzwerkbauierte Serveransatz dies unbedingt erforderlich macht und es sonst vermehrt zu Abstürzen und Fehlermeldungen kommen kann. Generell hat die lange Testphase der Apps gezeigt, dass eine manuelle Texteingabe bessere Übersetzungsergebnisse liefert als der Input durch die Spracherkennungskomponente, zumal der transkribierte Text nicht einfach manuell nachkorrigiert werden kann, sondern erneut eingetippt werden muss. *SayHi* lieferte sowohl auf Ebene der Spracherkennung als auch der maschinellen Übersetzung die brauchbarsten Ergebnisse. Es stellte sich zudem als robusteste der drei Apps und jene mit der natürlichsten Sprachausgabe heraus. Die Startkosten von € 1,49 spiegeln sich meiner Meinung nach in der Leistung der App wieder. Der Kauf der Premium Option ist nicht zwingend notwendig, da mit 40 angebotenen Sprachen in der Standardversion genügend Sprachen zur Verfügung stehen. Falls eine exotische Sprache oder ein bestimmter Dialekt benötigt wird, wäre das Abonnement empfehlenswert.

Die Apps sind für die beworbenen Kontexte und Situationen wie auf dem Flughafen, im Hotel oder sonstigen Reiseszenarien durchaus brauchbar und die generierten Zieltexte können trotz Syntaxfehler o.ä., in vielen Fällen dennoch verstanden werden. Bei *VoiceTra4U* wird in den FAQs (vgl. U-STAR 2014c) sogar explizit darauf hingewiesen, dass die App für Konversationen rund um das Thema Reisen konzipiert wurde und unverständliche oder spezielle Ausdrücke und Formulierungen nicht zufriedenstellend bearbeiten kann. Es muss bedacht werden, dass die Apps für den Massenmarkt produziert wurden und somit auch die breite Masse als Zielgruppe zu definieren ist. Es ist anzunehmen, dass diese Zielgruppe mit Übersetzungsfehlern viel weniger kritisch umgeht als professionelle TranslatorInnen und daher auch einen größeren Mehrwert aus den Apps ziehen kann.

9.4 Blick in die Zukunft

Im Jahr 2014 feiert die Machine Translation ihr 60. Jubiläum seit der ersten öffentlichen Präsentation des Georgetown-IBM MT Systems in New York (vgl. <https://www.taus.net> 2014). Jahrzehntelange interdisziplinäre Forschung ermöglichte es, maschinelle Übersetzungssysteme für die breite Masse zugänglich zu machen und in wenigen Augenblicken Speech-to-Speech Übersetzungen anzufertigen. Woran die Forschung heutzutage arbeitet und welche neuen Entwicklungen uns in naher Zukunft erwarten, soll in diesem Kapitel aufgezeigt werden.

Der wohl bekannteste Anbieter von maschinellen Übersetzungssystemen ist *Google* mit seiner Software *Google Translate*, welche mittlerweile auch als App existiert. Mit dem Aufkauf von *Quest Visual*, dem Entwickler von *Word Lens* zeigt *Google* eine mögliche Weiterentwicklungslinie von SST Apps auf. Die App *Word Lens* kann mittels einer OCR Funktion gedruckte Termini aus einer Sprache in eine andere übersetzen und bedient sich dabei der im Smartphone verbauten Kamera (vgl. <http://questvisual.com> 2014). Mit der Implementierung dieser Technologie in *Google Translate*, wäre die SST App auch im Stande gedruckte Zeichen zu übersetzen, was vor allem bei Sprachen mit anderen Schriftzeichen wie dem Chinesischen oder Russischen ein großer Vorteil wäre. Abseits des Anwendungsgebiets Apps und Smartphones wird diese Technologie bereits für Google Glass verwendet, sodass mithilfe der eingebauten Kamera ohne Netzwerkverbindung eine Übersetzung von gedruckten Texten angezeigt werden kann (vgl. <https://glass.google.com> 2014).

Eine vollkommen andere Art der Übersetzung hat Microsoft in Zusammenarbeit mit chinesischen WissenschaftlerInnen im Oktober 2013 vorgestellt. Der *Kinect Sign Language Translator* ist ein Prototyp, der an die speziellen Anforderungen der Kommunikation zwischen gehörlosen und nicht-gehörlosen Personen angepasst ist. Das System analysiert Gebärdensprache und übersetzt die Zeichen in Echtzeit in schriftlichen Text oder gibt diesen per Sprachsynthese wieder. Umgekehrt kann auch geschriebene oder gesprochene Sprache in Gebärdensprache übersetzt werden (vgl. Microsoft Research 2013).

Wie ein weltweites Großereignis wie die FIFA Fussball-Weltmeisterschaft 2014 in Brasilien für wissenschaftliche Zwecke genutzt werden kann, zeigt der *Brazilator*. Das irische CNGL Centre for Global Intelligent Content und *Microsoft Research* stellten mit dem *Brazilator* ein „live tweet translation streaming service“ (CNGL 2014) für den Zeitraum der Weltmeisterschaft zur Verfügung. Tweets, mit bestimmten Hashtags wie #WC2014 wurden Tweets aus 12 verschiedenen Sprachen ins Englische übersetzt und umgekehrt, sodass Twitter-NutzerInnen beispielsweise englische Tweets in ihrer jeweiligen Sprache verfolgen konnten. Diese Art von Ausgangstext birgt so einige Hürden. Tweets sind aufgrund ihrer auf 140 Zeichen begrenzten Länge oftmals inkohärent, grammatikalisch und syntaktisch unstrukturiert und enthalten häufig Links, Emoticons oder Hashtags (vgl. CNGL 2014).

Noch Ende des Jahres 2014 soll für Windows 8 der *Skype Translator* auf den Markt kommen. Es handelt sich dabei um ein Speech-to-Speech Translation System, welches direkt in die bereits auf dem Markt bekannte IP-Telefonie Software integriert wird. Ein Dialog zwischen zwei Menschen mit unterschiedlichen Ausgangssprachen soll mittels Videotelefonie ermöglicht werden, indem der gesprochene Ausgangstext in Echtzeit beim Gesprächspartner oder bei der Gesprächspartnerin in übersetzter Form und per Sprachaus-

gabe wiedergegeben wird. Der *Skype Translator* ist ein gemeinsames Projekt der *Skype* und *Microsoft Translator Teams* (vgl. Microsoft Research 2014). Um das jahrzehntelange Ziel einer sprecherunabhängigen Spracherkennung zu erreichen, wurde beim *Skype Translator* ein seit den fünfziger Jahren bekannter Ansatz der „artificial neural networks (ANNs)“ (Microsoft Research 2011) verfolgt. Es handelt sich dabei um eine künstliche, mathematische Nachbildung der im menschlichen Gehirn vorhandenen neuronalen Netze. Der ANN-Ansatz wurde bereits in den achziger Jahren in der Spracherkennungsforschung genutzt. Durch die Konstruktion kontextabhängiger, tiefer neuronaler Netze²⁸ konnte eine Reduktion der relativen Fehlerquote um 16 % erreicht werden. (vgl. Microsoft Research 2011).

So faszinierend die Werbevideos zu den diversen Erfindungen sein mögen, kommt man nicht umher, gesellschaftskritische Aspekte zu hinterfragen. So ist wohl die Frage, ob und wie gut eine Maschine menschliche Gedankengänge nachvollziehen und imitieren kann, noch weitestgehend unbeantwortet. Apples Sprachassistentin Siri und der neulich erschienene Google Now Launcher für Android Geräte können gesprochene Befehle erkennen und ausführen, und somit beispielsweise nach der nächsten Busstation suchen, das Wetter für morgen ansagen oder eine SMS oder E-Mail an bestimmte Personen versenden. Diese Technologien werden stetig weiterentwickelt und verbessern sich zusehens durch den Input der NutzerInnen. Diesbezüglich ist es jedoch fraglich, welches Ausmaß an Mensch-Maschine Kommunikation im Alltag tatsächlich nötig bzw. gewinnbringend ist und wie sich diese auf menschliches Sprachverhalten im Allgemeinen auswirkt.

²⁸ Nähere Informationen zu Context-Dependent Deep Neural Networks finden sich in Seide et al. 2011.

Abkürzungsverzeichnis

AS	Ausgangssprache
ASR	Automatic Speech Recognition
AT	Ausgangstext
CBMT	Context-Based Machine Translation
EBMT	Example-Based Machine Translation
EULA	End User License Agreement
KBMT	Knowledge-Based Machine Translation
MÜ	Maschinelles Übersetzen
MT	Machine Translation
OCR	Optical Character Reader
RBMT	Rule-Based Machine Translation
S2ST	Speech-to-Speech Translation
SLT	Spoken Language Translation
SST	Speech-to-Speech Translation
UI	User Interface
WER	Word error rate
ZS	Zielsprache
ZT	Zieltext

Bibliographie

- ALPAC. 1966. Language and Machines: Computers in Translation and Linguistics. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. In: http://www.books.nap.edu/html/alpac_lm/ARC000005.pdf (06.03.2014).
- Apple. 2008. iPhone SDK mehr als 100.000 mal heruntergeladen. In: <https://www.apple.com/de/pr/library/2008/03/12iPhone-SDK-Downloads-Top-100-000.html> (05.08.2014).
- Apple Inc. 2014. iOS Human Interface Guidelines. In: <https://developer.apple.com/library/ios/documentation/userexperience/conceptual/mobilehig/> (17.07.2014).
- Arnold, Douglas & Balkan, Lorna & Meijer, Siety & Humphreys Lee R. & Sandler, Louisa. 1994. *Machine Translation: An Introductory Guide*. London: NCC Blackwell.
- Bar-Hillel, Yehoshua. 1960. The Present Status of Automatic Translation of Languages. In: <http://mt-archive.info/Bar-Hillel-1960.pdf> (06.03.2014).
- Booth, Andrew Donald. 1967. *Machine translation*. Amsterdam: North-Holland Publishing Company.
- Brown, Peter F. & Cocke, John & Della Pietra, Stephen A. & Della Pietra, Vincent J. & Jelinek, Fredrick & Lafferty, John D. & Mercer, Robert L. & Roossin, Paul S. 1990. A Statistical Approach to Machine Translation. In: <http://dl.acm.org/citation.cfm?id=92860> (20.05.2014).
- Bruderer, Herbert E. 1982. *Automatische Sprachübersetzung*. Darmstadt: Wissenschaftliche Buch-Gesellschaft.
- Carbonell, Jaime & Klein, Steve & Miller, David & Steinbaum, Michael & Grassiany, Tomer & Frey, Jochen. 2006. Context-Based Machine Translation. In: https://www.cs.cmu.edu/~jgc/publication/Context_Based_Machine_AMTA_2006.pdf (30.06.2014).
- Carstensen, Kai-Uwe. 2012. Sprachtechnologie. Ein Überblick. In: <http://kai-uwe-carstensen.de/Publikationen/Sprachtechnologie.pdf> (20.05.2014).
- CNGL. 2014. CNGL and Microsoft Research deliver ‚Brazilator‘ FIFA World Cup Live Translation Service. In: <http://www.cngl.ie/cngl-and-microsoft-research-develop-fifa-world-cup-live-translation-service/> (11.08.2014).
- Farwell, David & Gerber, Laurie & Hovy, Eduard (Hg.) 1998. *Machine translation and the information soup*. Third conference of the Association for Machine Translation in the Americas, AMTA ‘98, Langhorne, PA, USA, October 1998: proceedings. Berlin: Springer.

Fügen, Christian & Waibel, Alexander & Kolss, Muntzin. 2008. Simultaneous Translation of Lectures and Speeches. In: http://i13pc106.anthropomatik.kit.edu/fileadmin/publication-files/861_lt2007.pdf (15.05.2014).

Fügen, Christian. 2008. A System for Simultaneous Translation of Lectures and Speeches. In: http://d-nb.info/1014223113/34?origin=publication_detail (12.07.2014).

Hacker, Martin & Ludwig, Bernd & Görz, Günter. 2012. Automatisches Verstehen gesprochener Sprache. In: <http://www.dh.cs.fau.de/IMMD8/Lectures/AVGS/AVGS01.3x2.pdf> (15.05.2014).

Hausser, Roland. 2000. Grundlagen der Computerlinguistik. Mensch-Maschine-Kommunikation in natürlicher Sprache. In: <http://www.linguistik.uni-erlangen.de/tree/pdf/slides/ganz.pdf> (20.05.2014).

Hutchins, William John. 1992. *An introduction to machine translation*. London: Academic Press.

Hutchins, William John. 1999. The development and use of machine translation systems and computer-based translation tools. In: Zhaoxiong, Chen (Hg.) *International Conference on Machine Translation & Computer Language Information Processing*. Proceedings of the conference, 26-28 June 1999, Beijing, China. 1-16. <http://www.hutchinsweb.me.uk/Beijing-1999.pdf> (05.03.2014).

Hutchins, William John. 2000. *Early years in machine translation: memoirs and biographies of pioneers*. Amsterdam/Philadelphia: John Benjamins.

Hutchins, William John. 2001. Machine Translation over fifty years. <http://www.mt-archive.info/00/HEL-2001-Hutchins.pdf> (05.03.2014).

Hutchins, William John. 2004. Two precursors of Machine Translation – Artsrouni and Trojanskij. In: <http://www.mt-archive.info/00/IJT-2004-Hutchins.pdf> (05.03.2014).

Hutchins, William John. 2008. Timeline of MT. <http://www.mt-archive.info/05/Aslib-2008-timeline.pdf> (05.03.2014).

Hutchins, William John. 2009. Multiple Uses of Machine Translation and Computerised Translation Tools. In: <http://www.hutchinsweb.me.uk/Besancon-2009.pdf> (10.05.2014).

Hutchins, William John. 2010. Outline of Machine Translation Developments in Europe and America. In: <http://www.hutchinsweb.me.uk/JAPIO-2010.pdf> (05.03.2014).

Hutchins, William John. 2014a. The history of machine translation in a nutshell. In: <http://www.mt-archive.info/10/Hutchins-2014.pdf> (05.03.2014).

Hutchins, William John. 2014b. Machine translation: History of Research and Applications. In: <http://www.hutchinsweb.me.uk/Routledge-2014.pdf> (05.03.2014).

Jekat, Susanne J. 2004. Ausgewählte Techniken der maschinellen Übersetzung. <https://files.ifi.uzh.ch/cl/jekat/ws0405/folien7.pdf> (30.06.2013).

Jibbigo LLC. 2014. End User License Agreement. In: <http://jibbigo.com/EULA-en.html> (13.07.2014).

Jibbigo LLC. o.J. Jibbigo User's Manual. In: <http://www.docstoc.com/docs/140620405/JIBBIGO-USER-S-MANUAL---APPLE> (16.07.2014).

Karger Reinhard & Wahlster Wolfgang. 2000. Facts and Figures about the Verbmobil Project. In: Wahlster, Wolfgang (Hg.), 22–31.

Karger Reinhard. 2010. Verbmobil - vom Maschinellen Dolmetscher zur iPhone App. http://www.dfki.de/web/presse/pressemitteilungen_intern/2010/verbmobil-2013-vom-maschinellen-dolmetscher-zur-iphone-app (20.06.2014).

King, Margaret. 1987. *Machine translation today: the state of the art*. Edinburgh: Edinburgh University Press.

KIT. 2012. Fokus: Lecture Translation. In: <http://isl.anthropomatik.kit.edu/english/1520.php> (09.07.2014).

Kitano, Hiroaki. 1994. *Speech-to-Speech Translation*. Boston: Kluwer Academic Publications.

Lafferty, Justin. 2013. Facebook acquires startup Mobile Technologies, maker of mobile translation app Jibbigo. In: <http://www.insidefacebook.com/2013/08/12/facebook-acquires-startup-mobile-technologies-maker-of-mobile-translation-app-jibbigo/> (13.07.2014).

Microsoft Research. 2011. Speech Recognition Leaps Forward. In: <http://research.microsoft.com/en-us/news/features/speechrecognition-082911.aspx> (11.08.2014).

Microsoft Research. 2013. Kinect Sign Language Translator expands communication possibilities. In: <http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx> (11.08.2014).

Microsoft Research. 2014. Enabling Cross-Lingual Conversations in Real Time. In: <http://research.microsoft.com/en-us/news/features/translator-052714.aspx> (11.08.2014).

Nirenburg, Sergei & Raskin, Victor & Tucker, Allen. 1986. On Knowledge-Based Machine Translation. In: <http://aclweb.org/anthology/C/C86/C86-1148.pdf> (30.06.2014).

Nirenburg, Sergei. 1987. *Machine translation: theoretical and methodological issues*. Cambridge: Cambridge University Press.

Nuance Communications, Inc. 2011. Nuance gives SayHi Translate App a Voice. In: <http://www.nuance.de/company/news-room/press-releases/sayhinmdp.doc> (05.08.2014).

Och, Franz Josef & Nicola, Ueffing & Ney, Hermann. 2001. An Efficient A* Search Algorithm for Statistical Machine Translation. In: <http://mt-archive.info/ACL-2001-Och.pdf> (30.05.2014).

Och, Franz Josef. 2012. Breaking down the language barrier – six years in. In: <http://googleblog.blogspot.co.at/2012/04/breaking-down-language-barriersix-years.html> (10.05.2014)

Pfister, Beat & Kaufmann, Tobias. 2008. *Sprachverarbeitung. Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Berlin/Heidelberg: Springer Verlag.

Renals, Steve & Hain, Thomas. 2010. Speech Recognition. In: Clark, Alexander & Fox, Chris & Lappin, Shalom (Hg.) *The Handbook of Computational Linguistics and Natural Language Processing*. Malden/Oxford/Chichester: Wiley-Blackwell, 299-332.

Rödter, Margit. 2009. Das Mobiltelefon als Sprachübersetzer. In: http://www.informatik.kit.edu/309_3685.php (13.07.2014).

SayHi LLC. 2014a. Privacy Policy. In: <http://www.sayhitranslate.com/privacy/> (05.08.2014).

SayHi LLC. 2014b. Press Kit. In: <http://www.sayhitranslate.com> (05.08.2014).

Schmitz, Birte. 1998. *Pragmatikbasiertes Maschinelles Dolmetschen*. Heidelberg: Julius Groos Verlag und Berlin: Technische Universität, Dissertation.

Schütze, Hinrich & Zangenfeind Robert. 2013. Einführung Computerlinguistik – Konstituentensyntax II. In: <http://www.cis.uni-muenchen.de/~hs/teach/13w/intro/pdf/07syn.pdf>

Seide, Frank & Li, Gang & Yu, Dong. 2011. Conversational Speech Transkription Using Context-Dependent Deep Neural Networks. In: <http://research.microsoft.com/pubs/153169/CD-DNN-HMM-SWB-Interspeech2011-Pub.pdf> (11.08.2014).

Somers, Harold. 2003. An Overview of EMBT. In: Cral, Michael & Way, Andy (Hg.) *Recent advances in Example-Based Machine Translation*. Dordrecht: Kluwer, 3-57.

Somers, Harold. 2005. Machine Translation: Recent Developments. In: Mitkov, Ruslan (Hg.) *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press. 512-528.

Stein, Daniel. 2009. Maschinelle Übersetzung – ein Überblick. *JLCL - Journal for Language Technology and Computational Linguistics*, Volume 24, Number 3, 5-18.

SYSTRAN. 2009. SYSTRAN introduces hybrid machine translation solution for enterprises. In: <http://www.systransoft.com/systran/news-and-events/press-release/hybrid-machine-translation-solution-for-enterprises/> (30.06.2014).

Taylor, Paul. 2009. *Text-to-Speech Synthesis*. Cambridge: Cambridge University Press.

Trujillo, Arturo 1999. *Translation engines: techniques for machine translation*. London: Springer Verlag.

U-STAR. 2014. About us. <http://www.ustar-consortium.com/aboutus.html> - objective (15.05.2014).

U-STAR. 2012. International Press Release of VoiceTra4U-M. In: http://ustar-consortium.com/app/presskits/International-press-release-of-VoiceTra4U-M_E.pdf (15.07.2014).

U-STAR. 2013. News & Events. In: <http://ustar-consortium.com/newsevents.html> (16.07.2014).

U-STAR. 2014a. About us.. In: <http://www.ustar-consortium.com/aboutus.html> (15.07.2014).

U-STAR. 2014b. Standardization.. In: <http://ustar-consortium.com/standardization.html> (17.07.2014).

U-STAR. 2014c. Speech Translation App „VoiceTra4U“. In: <http://www.ustar-consortium.com/app/app.html> (15.07.2014).

U-STAR. 2014d. Terms of Use.. In: <http://www.ustar-consortium.com/app/agreement.html> (20.07.2014).

U-STAR. 2014e. Languages & Tools. In: <http://www.ustar-consortium.com/app/other.html> (20.07.2014).

Wahlster, Wolfgang. 1997. Verbmobil. Erkennung, Analyse, Transfer, Generierung und Synthese von Spontansprache. In: Jarke, Matthias & Pasedach, Klaus & Pohl, Klaus (Hg.) *Informatik 97 - Informatik als Innovationsmotor*. Heidelberg: Springer, 215-224.

Wahlster, Wolfgang. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin: Springer.

Wahlster, Wolfgang. 2000. Mobile Speech-to-Speech Translation of Spontaneous Dialogs: An Overview of the Final Verbmobil System. In: Wahlster, Wolfgang (Hg.): *Verbmobil*. Berlin: Springer. 3-21.

Wahlster, Wolfgang. 2010. 10 Jahre Verbmobil - Rückblick und Ausblick: Eröffnung und Einleitung. <http://www.dfki.de/web/aktuelles/10-jahre-verbmobil-multilinguale-verarbeitung-von-spontansprache/10%20Jahre%20Verbmobil%202013%20Ruckblick%20und%20Ausblick%20Eroffnung%20und%20Einleitung.pdf> (08.07.2014).

Waibel, Alexander & Fügen, Christian. 2008. Spoken Language Translation. *IEEE Signal Processing Magazine* 25(3), 70-79.

- Waibel, Alexander. 2012a. Interview bei LT-Innovate Summit 2012. In: <https://www.youtube.com/watch?v=PpFsOe7XLPk> (10.05.2014).
- Waibel, Alexander. 2012b. Präsentation: University without Language Barriers Introducing the First Simultaneous Translation Service by Computer - Simultaneous Speech Translation. In: <https://www.youtube.com/watch?v=x5IL0wpr-88> (09.07.2014).
- Waibel, Alexander. 2012c. Simultaneous Machine Interpretation – Utopia? In: http://www.europarl.europa.eu/interp/rectorsconference2012/docs/pdf/waibel_presentation.pdf (12.07.2014).
- White, John S. 2000. *Envisioning machine translation in the information future*. 4th conference of the Association for Machine Translation in the Americas, AMTA 2000, Cuernavaca, Mexico, October 2000. Berlin: Springer.
- Wikimedia Commons. 2014. Dreieck von Vauquois. In: http://en.wikipedia.org/wiki/File:Direct_translation_and_transfer_translation_pyramid.svg (15.06.2014).

Internetquellen

Konferenzen Computerlinguistik

<http://aclweb.org/anthology//>

Facebook Posting Stocky

<https://www.facebook.com/tstocky/posts/10100115269490818>

Facebook Timeline von SayHi LLC

https://m2.facebook.com/sayhitranslate?v=timeline&timecutoff=1335501046&page=2§ionLoadingID=m_timeline_loading_div_1325404799_1293868800_8_2&timeend=1325404799×tart=1293868800&tm=AQACxHDdxP3vrxtU&_rdr

Google Glass – Word Lens App

<https://glass.google.com/u/0/glassware/16789448588362059188>

Jibbigo Website

<http://jibbigo.com>

Konferenzen MT

<http://www.mt-archive.info/srch/conferences-1.htm>

Quest Visual

<http://questvisual.com>

TAUS – Translation Automation Timeline

<https://www.taus.net/timeline/a-translation-automation-timeline>

UserVoice

<https://www.uservoice.com/powered-by/>

Verbmobil Website

<http://verbmobil.dfki.de>

KITinformatik – Youtube Kanal – Präsentation Lecture Translator

^a<https://www.youtube.com/watch?v=x5lL0wpr-88>

Jibbiggo Youtube Kanal

^b<https://www.youtube.com/user/jibbiggo>

Apple iTunes/App Store

^a<https://itunes.apple.com/at/app/jibbiggo-translator/id528691646?mt=8>

^b<https://itunes.apple.com/at/app/record-it-what-i-read/id448295196?mt=8>

^c<https://itunes.apple.com/us/app/voicetra-4u/id537952299?mt=8>

^d<https://itunes.apple.com/at/app/ubersetzer-sprach-sayhi-translate/id437818260?mt=8>

^e<https://itunes.apple.com/at/app/sprach-ubersetzer-fur-tischplatte/id576422798?mt=8>

Google Play Store

^a<https://play.google.com/store/apps/details?id=com.jibbiggo.player&hl=de>

^b<https://play.google.com/store/apps/details?id=com.ustar.multi>

Alle Internetquellen wurden zuletzt am 20.08.2014 eingesehen.

Anhang

Abstract

Ziel dieser Arbeit ist es, einerseits die drei Speech-to-Speech Translation Apps *Jibbig Translator 2.5.5*, *VoiceTra4U 3.2*, *SayHi Translate 3.8.2*. auf ihren Funktionsumfang hin genau zu analysieren und andererseits die Stärken und Schwächen der einzelnen Apps herauszufiltern. Ausgehend von umfassenden Beschreibungen der Apps wurden SWOT-Analysen erstellt und schließlich eine tabellarische Gegenüberstellung der Funktionen angefertigt. Darüber hinaus wurden zwei Testdialoge aus zwei verschiedenen Kontexten durchgeführt, welche einen unmittelbaren Vergleich der übersetzerischen Leistung der Apps ermöglichen sollen. Die ersten Kapitel der Arbeit beinhalten eine historische Einführung in die Forschung zur maschinellen Übersetzung, sowie der damit verbundenen Spracherkennung und Sprachsynthese. Zudem wurden die einzelnen Komponenten von Speech-to-Speech Translation Systemen genau analysiert und zwei Pionierprojekte vorgestellt.

The aim of this master thesis is on the one hand to accurately analyse the functionality of the three speech-to-speech translation apps *Jibbig Translator 2.5.5*, *VoiceTra4U 3.2*, *SayHi Translate 3.8.2*. and on the other hand to determine the strengths and weaknesses of the individual apps. Based on comprehensive descriptions of each app, SWOT analyses as well as a tabular comparison of features were created. Moreover, two test dialogues from two different contexts were conducted, in order to provide a direct comparison of the translation-related performance. The first chapters of the thesis include a historical introduction into research on machine translation, speech recognition and speech synthesis. In addition, it includes a detailed analysis of the individual components of speech-to-speech translation systems and a chapter about two pioneer projects.

Curriculum Vitae

Persönliche Daten

Vorname und Zuname	Ivona Brizar
Geburtsdatum und -ort	23. August 1989 Slavonski Brod (Kroatien)
Staatsbürgerschaft	Österreich

Studium und Ausbildung

02/2012 – (10/2014)	Master Übersetzen (Fachübersetzen) Universität Wien (Abschluss Oktober 2014)
07/2012 – (06/2015)	Master Dolmetschen (Konferenzdolmetschen) Uni- versität Wien (voraussichtlich bis Juni 2015)
10/2008 – 01/2012	Bachelor Transkulturelle Kommunikation Universi- tät Wien Sprachkombination: Deutsch – Englisch – Franzö- sisch
06/2008	Matura mit ausgezeichnetem Erfolg
2003 – 2008	BHAK Gmunden Schwerpunkt: Marketing & internationale Ge- schäftstätigkeit

Berufliche Erfahrung

2010 – ...	Lernquadrat Wien Nachhilfe (Unter- /Oberstufe) (Deutsch, Englisch und Französisch)
07 – 08/2014	WIFI Gmunden „Ein neuer Weg – Ihr neuer Job“ Trainerin für Perfektionstraining Business English
07 – 09/2013	WIFI Gmunden „Ein neuer Weg – Ihr neuer Job“ Trainerin für Englisch B1, B2
10/2012 – 04/2013	Praktikum bei OKTO Community TV GmbH Qualitätssicherung der Untertitelungen
07 – 09/2012	WIFI Gmunden „Ein neuer Weg – Ihr neuer Job“ Trainerin für Englisch A1, B1, B2

04/2012	Website-Lokalisierung www.hospitals.at
02/2012	WIFI Gmunden „Büro Intensiv“ Trainerin für Business English
07 – 09/2011	WIFI Gmunden „Ein neuer Weg – Ihr neuer Job“ Trainerin für Englisch B1, B2

Weiterbildung

10/2012	WIFI Weiterbildungsveranstaltung „Diversity Management und Gender Mainstreaming“
11/2012	WKO Weiterbildungsveranstaltung für Sprachdienstleister „CAT-Systeme richtig anwenden“

Kenntnisse & Fähigkeiten

Sprachen	Deutsch: Bildungssprache (C2) Englisch: fließend in Wort und Schrift, 1. Fremdsprache (C1) Französisch: fließend in Wort und Schrift, 2. Fremdsprache (C1) Kroatisch: gute Kenntnisse (B1)
----------	--