FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

GESTU

Erarbeitung eines Respeaking Trainingsprogramms und Vergleich zur automatischen Spracherkennung, um diese Techniken zur Untertitelerzeugung in E-Learning Plattformen zu verwenden und somit speziell hörbeeinträchtigte Studierende zu unterstützen.

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Magister der Naturwissenschaften

im Rahmen des Studiums

Informatikmanagement

eingereicht von

Christian Franz Hattinger, DI
Matrikelnummer 0427438

an der
Fakultät für Informatik der Technischen Universität Wien
Institut für „Gestaltungs- und Wirkungsforschung“
Zentrum für „Angewandte Assistierende Technologien“ (AAT)

Betreuung
Betreuer: ZAGLER, Wolfgang, Ao.Univ.Prof. Dipl.-Ing. Dr.techn.

Wien, 16.12.2013

(Unterschrift Verfasser) (Unterschrift Betreuer)

Eidesstattliche Erklärung

Christian Franz Hattinger
Winckelmannstraße 8/5
1150 Wien

„Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe."

Winckelmannstraße 8/5, 1150 Wien, 16.12.2013

Christian Franz Hattinger

Danksagung

Ich möchte mich an dieser Stelle bei all jenen bedanken, die zwischen den Zeilen dieser Diplomarbeit stehen und mich so kräftig unterstützt haben. Besonderer Dank gilt meinem Betreuer Wolfgang Zagler, der mir das sehr spannende Thema dieser Diplomarbeit anvertraut hat und mir die Möglichkeit gab, als Teammitglied von GESTU viel Erfahrung zu sammeln. In gleicher Weise danke ich Georg Edelmayer für die laufende Unterstützung während der Diplomarbeit. Speziell möchte mich auch beim Pablo Romero-Fresco bedanken, der während der Ausarbeitung des Trainings immer ein offenes Ohr für Fragen hatte. Darüber hinaus danke ich dem gesamten GESTU Team für die Unterstützung sowie Christoph Czurda, der ein sehr motivierter Auszubildender war.
Tausend Dank auch an Herrn Werner Nemecek, der mir während der Erstellung wiederholt sehr konstruktives Feedback gab und mich schließlich auch bei der Korrektur durch seine kritische Betrachtung und Geduld unterstütze.
Nicht zuletzt möchte ich mich von ganzen Herzen bei meiner gesamten Familie bedanken, die mir nicht nur während dem Verfassen dieser Diplomarbeit, sondern während meines gesamten Bildungsweges unterstützten.
Abschließend danke ich allen Menschen, die mich während meiner Studienzeit, ob in Wien, Groningen oder Alcalá de Henares begleitet und mich durch die kleinen und großen Erfahrungen in den vergangenen Jahren sehr bereichert haben.

Zusammenfassung

Universitäre Lehrveranstaltungen sind ohne adäquaten schriftlichen (oder gebärdeten) Inhalt für hörbeeinträchtigte Studierende eine beträchtliche bis unüberwindbare Hürde. Seit 2010 unterstützt das Projekt GESTU („Gehörlos Erfolgreich Studieren“) hörbeeinträchtigte Studierende beim zeitgerechten und erfolgreichen Absolvieren des Studiums. Eine zentrale Rolle spielt dabei die Untertitelung von Lehrveranstaltungen in Echtzeit. Die Live Untertitelung wird von einer externen Firma mit der sogenannten Respeaking Technik durchgeführt und von den hörbeeinträchtigten Studierenden allgemein als positiv beurteilt. Beim Respeaking handelt es sich um eine Mischform von maschineller und manueller Erstellung der Untertitel, mit der negative Einflussfaktoren auf die Qualität der Erkennungsrate einer Spracherkennungssoftware umgangen werden können. Um zu vermeiden, dass Akzente, Dialekte, Sprechgeschwindigkeit, Umgebungsgeräusche, etc. die Erkennungsgenauigkeit negativ beeinflussen, spricht ein Sprecher oder eine Sprecherin (der Respeaker bzw. die Respeakerin) das Gesagte, in einer für die Spracherkennung gut verarbeitbaren Weise, nach. Respeaking wird in vielen europäischen Ländern in unterschiedlichsten Bereichen zur (live) Untertitelung eingesetzt. So auch in Österreich, wo beim ORF seit 2010 Respeakerinnen und Respeaker die Echtzeituntertitelung durchführen.
Zunehmend mehr Lehrveranstaltungen werden Video- und Audio aufgezeichnet und in E-Learning Plattformen online gestellt. Die Untertitelung ist neben einer Einblendung von Gebärdensprachübersetzungen die einzige Möglichkeit für hörbeeinträchtigte Studierende, diese Aufzeichnungen zum Erlernen der Inhalte zu nutzen.
Um auch ohne (bzw. nicht ausschließlich mit) externen Firmen eine höhere Untertitelquote im E-Learning Bereich zu erreichen, liegt der Fokus dieser Diplomarbeit auf der Evaluierung von zwei thematisch verwandten und vielversprechenden Möglichkeiten der offline Untertitelerzeugung. Einerseits wurde die automatische Erzeugung von Untertitel mittels einer Sprecher bzw. Sprecherinnen unabhängigen, für spontane Sprache entwickelten Spracherkennungssoftware untersucht. Die Erkennungsraten lagen deutlich hinter den Erwartungen, wodurch diese Art der Untertitelerzeugung derzeit keine Unterstützungsmöglichkeit für hörbeeinträchtigte Studierende darstellt. Die zweite untersuchte Alternative ist die offline Untertitelerstellung mit Respeaking (Scripting). Im Gegensatz zu anderen europäischen Ländern gibt es in Österreich keine Respeaking Ausbildung an (öffentlichen) Bildungseinrichtungen. Somit legt das im Zuge dieser Diplomarbeit erarbeitete, dokumentierte und evaluierte Training den Grundstein für eine (akademische) Respeaking/Scripting Ausbildung in Österreich. Der Trainingsplan beinhaltet sieben Einheiten samt Übungen zum Erlernen des Respeakings. Die Ausbildung setzt keine spezielle Vorkenntnisse und Erfahrungen voraus und ist innerhalb von drei Monaten möglich. Im Zuge dieser Diplomarbeit wurde ein Respeaker ausgebildet. Durch qualitative, mündliche Interviews und der ständigen Evaluierung des Trainingsprozesses wurden die Einheiten ebenso wie der Zeitaufwand und das verwendete Equipment evaluiert. Eine Vorlesung wurde unabhängig von drei Seiten mittels Respeaking untertitelt: Live durch eine externe Firma und offline via Respeaking/Scripting durch den ausgebildeten Respeaker sowie den Autor dieser Diplomarbeit.
Die Beurteilung aller durch Respeaking erstellten Untertitel erfolgte erstmals in Österreich durch die NER-Analyse. Dabei wird ab einem NER-Wert von 98% von akzeptablere Qualität gesprochen. Die durch Scripting erstellten Untertitel weisen mit 98,9% bzw. 97,4% eine deutlich höhere Qualität sowie eine höhere qualitative Konstanz als die live erzeugten Untertitel mit 94,9% auf. Die Ergebnisse sind mit Kreuztabellen, Box-Whisker-Plots und Streudiagrammen visualisiert, textlich beurteilt und heben die Vorteile ebenso wie die Nachteile der verschiedenen Arbeitsweisen hervor.

Abstract

University lectures without a written or signed version of the spoken content remain difficult to impossible for hearing impaired students to access. Since 2010, the project GESTU („**G**ehörlos **E**rfolgreich **S**tudieren“, engl. „Successful Deaf Studies“) supports hearing impaired students in order that they can successfully finish their studies on time. One central part of this support is the live subtitling of lectures. The live subtitling of lectures is currently completed by an external organisation using the „respeaking“ method. Hearing impaired students generally rate this additional support positively. Respeaking involves a combination of automated voice recognition and interpretation of the original recording by the respeaker. By respeaking the original audio, accuracy rates can be improved. This is achieved by bypassing the original negative influences using human interpretation. By repeating the original content, the respeaker delivers an audio track without surrounding noise or dialect, and with consistent accents and speech rate, etc. In many european countries, respeaking is used for (live) subtitling in many different areas. In Austria, the national broadcaster ORF uses respeaking for live subtitling since 2010.
More and more frequently lectures are recorded in audio and/or video and made available on e-learning platforms. As an alternative to recorded sign language interpreters, subtitles are the only possibility for hearing impaired students to make use of recordings in order to learn the course content.
The focus of this thesis is the evaluation of two related and promising methods for offline subtitles, in order to avoid dependence on an external company to achieve a higher amount of subtitled lectures in e-learning platforms. The first method is investigated by producing subtitles with an automated speech recognition system which is speaker independent and developed for spontaneous speech. The accuracy of this method failed to meet the expectations and was found to be impractical as a support method for hearing impaired students in its current state. The second investigated method is offline subtitling with respeaking (scripting). Although respeaking training exists in other european countries, it is not currently possible to train as a respeaker within the (public) Austrian education system. For this reason, this thesis created, documented and evaluated a training programme which could be the foundation of a (academic) respeaking/scripting training in Austria. The training includes seven classroom units as well as practical homework tasks in order to learn respeaking. The training does not require any prior knowledge and is possible to complete within three months. One respeaker was trained within this thesis. The seven study units were evaluated continually throughout the learning progress by qualitative oral interviews. Further to this, the education progress was evaluated in terms of time and effort required and the suitability of the chosen equipment. For one example lecture, the performance of three independent respeakers is assessed, including live respoken subtitling by an external company and offline subtitling by the trained scripter and thirdly the author of this thesis. The evaluation of all three respoken subtitles was firstly done in Austria by the NER-analyses. The model used, defines a NER-value of 98% as acceptable. The subtitles produced with scripting achieved values of 98.9% and 97,4% for the the thesis author and the respeaking student respectively. This was significantly higher than the 94,9% reached by the live respeakers. In addition, the consistency from scripting was also higher. The results are visualized with contingency tables, box-and-whisker plots and scatter plots and explained by written analyses. Further, the advantages and disadvantages of the different methods of operation are highlighted.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

KAPITEL 1

Einleitung

1.1 Hörbeeinträchtigung

1.1.1 Einführung

In Österreich leben zwischen 400.000 und 500.000 hörbeeinträchtigte Menschen (vgl. ([LW05], [KS07, S: 397], [BB95, S: 11]). Die Dunkelziffer könnte laut Schätzungen von Ärztinnen und Ärzten jedoch bei mehr als einer Million liegen (vgl. [ORF12b]). Etwa 10.000 (gehörlose) Bürgerinnen und Bürger kommunizieren in ÖGS, der Österreichischen Gebärdensprache (vgl. [LW05]). Eine vergleichsweise geringe Anzahl von ihnen - etwa 0,3% - studiert an Österreichischen Universitäten, während es bei der österreichischen Gesamtbevölkerung rund 2,65% sind (vgl. [Nem13, Abschnitt 1.3][1]). Hörbeeinträchtigte Menschen sind im österreichischen Bildungssystem teilweise mit erheblichen Hürden konfrontiert. Diese reichen von *finanziellen*, *organisatorischen* bis hin zu *pädagogischen* Barrieren. Eine Benachteiligung von hörbeeinträchtigten Studierenden kann auch durch fehlende Sensibilisierung und zu geringes Wissen von Mitstudierenden sowie von mit der Lehre beauftragten Personen entstehen (vgl. [Hat11, S: 0, 141], [KS07, S: 74-75, 313-316, 385-397]).

1.1.2 Begriffsdefinitionen: Hörbeeinträchtigung, Gehörlos und Schwerhörig

In dieser Diplomarbeit werden die

> „Begriffe *Hörbehinderung* sowie *Hörbeeinträchtigung* für Personen verwendet werden, bei welchen eine Minderung des Hörvermögens besteht. Hörbehinderung und Hörbeeinträchtigung werden weiters als Synonym verwendet und dienen als Überbegriff für schwerhörige sowie gehörlose Menschen. Als *gehörlos* werden in dieser Diplomarbeit hörbeeinträchtigte Personen bezeichnet, für welche das bevorzugte Kommunikationsmittel eine Gebärdensprache wie ÖGS (Österreichische Gebärdensprache) ist. Jene Personen, die hörbeeinträchtigt sind und deren bevorzugte Sprache eine Lautsprache ist, werden als schwerhörige Menschen bezeichnet“ [Hat11, S: 1].

Eine ausführliche Begriffsdiskussion sowie weitere Definitionen bezüglich Hörbeeinträchtigung sind in [Hat11, S: 21-23] dokumentiert.

[1] Anm. Autor: Wie im Folgenden Abschnitt erläutert, wird bei einer Hörbeeinträchtigung häufig zwischen *Gehörlos* und *Schwerhörig* unterschieden. Diese Differenzierung ist in unterschiedlichen Veröffentlichungen nicht immer klar definiert. Da es keine statistische Erhebungen bezüglich der Anzahl von hörbeeinträchtigten Studierenden in Österreich gibt, bezieht sich die Angabe auf *gehörlose* Menschen an öffentlichen *Universitäten*. Demnach beinhalten die 0,3% nicht *alle* hörbeeinträchtigten Studierenden des gesamten *tertiären* Bildungsbereichs. Obwohl sich diese Diplomarbeit um eine Untertitelerzeugung für hörbeeinträchtigte Studierende im Allgemeinen bemüht, soll der Vergleich die sehr geringe Anzahl von gehörlosen Menschen an österreichischen Universitäten verdeutlichen.

1.2 GESTU - Gehörlos Erfolgreich Studieren

1.2.1 Pilotprojekt GESTU - Gehörlos Erfolgreich Studieren an der TU Wien

Das zwischen Juli 2010 bis Juli 2012 durchgeführte Pilotprojekt GESTU („**G**ehörlos **E**rfolgreich **S**tudieren an der **TU** Wien“) hatte das Ziel, den unterstützten hörbeeinträchtigten Studierenden ein zeitgerechtes und erfolgreiches Absolvieren ihres Studiums zu ermöglichen. Die drei Kernbereiche von GESTU gliederten sich in *pädagogische*, *organisatorische* sowie *technische Unterstützung* für die - im Laufe der Projektzeit insgesamt 15 - unterstützten Studentinnen und Studenten. Pädagogische Maßnahmen beinhalteten neben der verstärkten Bereitstellung von ÖGS Gebärdensprachdolmetschern und Gebärdensprachdolmetscherinnen u.a. auch die Entwicklung von Fachgebärden. Im Zuge von GESTU entstand auch die in Abbildung 1.1 dargestellte ÖGS Gebärde für GESTU.

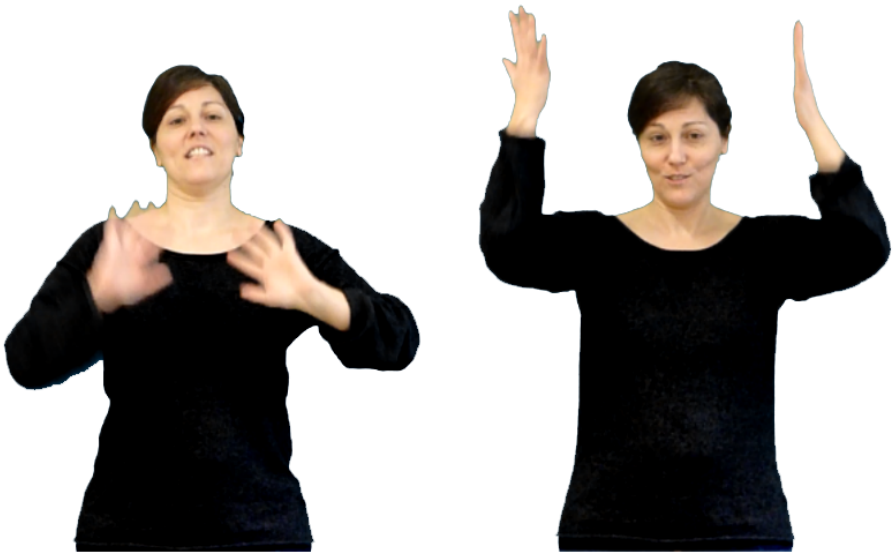


Abbildung 1.1: ÖGS Gebärde zu GESTU [Hat11, Abbildung 1.2]

Die *Servicestelle* des Projektes bildete für alle (zukünftigen) hörbeeinträchtigten Studentinnen und Studenten eine zentrale Anlaufstelle. Dabei erhielten sie von den Mitarbeiterinnen und Mitarbeitern kompetente Beratung zur Organisation des Studiums im Allgemeinen und Unterstützung für die Antragstellung finanzieller Förderungen. Für die im Projekt aufgenommenen Studierenden organisierten sie darüber hinaus Tutoren und Tutorinnen und Dolmetschleistungen.
Zusätzliche schriftliche Unterlagen sowie Maßnahmen, die hörbeeinträchtigten Studierenden einen Wissenserwerb sowie Austausch mit anderen Studierenden ermöglichen, dürfen jedoch nicht als ‚Bevorzugung’ gesehen werden (vgl. [KS07, S: 399; 369-370]). Eine solche Hilfe ist dabei „nicht ‚unfair’ den hörenden Studierenden gegenüber, sondern nur ein Ausgleichen der Bildungsbedingungen“ [KS07, S: 399].

Eine der technischen Maßnahmen, die in [Hat11] erörtert und auch getestet wurde, war das Remote Gebärdensprachdolmetschen (RGD). Dabei handelt es sich um das Dolmetschen eines Gesprächs durch einen Gebärdensprachdolmetscher oder eine Gebärdensprachdolmetscherin aus der Ferne mittels Live-Video-Schaltung. Weiters wurden die Einsatzmöglichkeiten von Spracherkennungssoftwares zur Untertitelerzeugung in [Hat11] evaluiert, erprobt und bereits im GESTU Projekt zur Erzeugung von Live Untertiteln während diverser Vorlesungen angewandt. Neben den technischen Hilfsmitteln sind auch die Lehrveranstaltungsaufzeichnungen sowie das im Zuge von GESTU entstandene ÖGS Fachgebärdenlexikon in [Hat11] und [Nem13] dokumentiert.

1.2.2 GESTU - Gehörlos Erfolgreich Studieren an Universitäten in Wien

Ein weiteres Ziel (des im vorigen Abschnitt beschriebenen Pilotprojektes) war es, in den 25 Monaten Projektlaufzeit ein zukünftiges Modell für österreichische tertiäre Bildungseinrichtungen zu entwickeln. Damit sollte langfristig die Anzahl der hörbeeinträchtigten Akademiker und Akademikerinnen erhöht werden (vgl. [Hat11, S: 3-5]). Das Projekt wurde im Wintersemester 2012/2013 in adaptierter Form und unter dem Namen GESTU - Gehörlos erfolgreich studieren an Universitäten in Wien - fortgesetzt. Darüber hinaus wurde eine Finanzierung für den Zeitraum der Leistungsvereinbarungsperiode (2013-2015) vom Bundesministerium für Wissenschaft und Forschung (BMWF) zugesichert (vgl. [Bun12]). Auf eine weitere Beschreibung des GESTU Pilotprojektes sei auf die Webseite[2], auf [Hat11, S: 3-5] sowie [Nem13, Abschnitt 1.3] verwiesen.

1.2.3 Wissenschaftliche Publikationen im Zuge von GESTU

Meine Diplomarbeit [Hat11][3] mit dem Titel „GESTU: Evaluierung von technischen Hilfsmitteln zur Förderung Studierender mit Hörbehinderung im österreichischen tertiären Bildungssektor und Einführung geeigneter Technologien an der TU Wien“ entstand während des ersten Jahres des GESTU Pilotprojektes (2010-2011). In der vorliegenden Diplomarbeit aus dem Jahr 2013 wird an einigen Stellen auf Erkenntnisse und Schlussfolgerungen von [Hat11] aufgebaut. Für Hintergründe und Diskussionen, die über den Fokus der vorliegenden Diplomarbeit hinausgehen, ist jeweils auf Abschnitte in [Hat11] verwiesen. Speziell betrifft das den Bereich der Spracherkennungssoftwares (ASR, engl. *Automatic Speech Recognition*, respektive automatische Spracherkennung), verschiedene Aspekte der Untertitelung sowie Untertitelungsmethoden abseits der *Respeaking* Technik.
Beim Respeaking „handelt es sich um eine Mischform von maschineller und manueller Erstellung der Untertitel“ [Hat11, S: 52]. Um zu vermeiden, dass „Einflussfaktoren, wie Akzente, Dialekte, Code-Switching (wechselnde Dialekte), Umgebungsgeräusche, etc. die Erkennungsgenauigkeit negativ beeinflussen [...] spricht ein Sprecher oder eine Sprecherin (der Respeaker bzw. die Respeakerin) [...] das Gesagte in einer für die ASR gut verarbeitbaren Weise nach. Die diktierten Wörter bzw. Sätze werden von der ASR transkribiert und können beispielsweise im Hörsaal angezeigt bzw. auch für den eLearning Bereich verwendet werden“ [Hat11, S: 65]. Für

[2] `http://teachingsupport.tuwien.ac.at/gestu`, letzter Zugriff: 28.09.2013.
[3] [Hat11] wurde vom Autor dieser Diplomarbeit verfasst und ist unter `www.ub.tuwien.ac.at/dipl/2011/AC07811467.pdf` aufrufbar, letzter Zugriff: 01.01.2013.

eine ausführliche Beschreibung der Respeaking Tätigkeit(en) samt Definition(en) wird auf das Kapitel 2 verwiesen.
Weiters wird in der vorliegenden Diplomarbeit auf theoretische Ausarbeitungen von [Hat11] verwiesen, die für eine Leserin oder einen Leser dieser Diplomarbeit aufschlussreich sein können bzw. Teil der im Kapitel 3 erarbeiteten Respeaking Ausbildung sind. Dies betrifft u.a. generelle Ausarbeitungen zum Rahme *Hörbeeinträchtigung* sowie *Unterstützungsmöglichkeiten für hörbeeinträchtigte Studierende.*
Nemecek widmet sich in [Nem13] neben dem Respeaking und der Spracherkennung auch Lehrveranstaltungsaufzeichnungen, Fachgebärdenlexika sowie dem Remotegebärdensprachdolmetschen. Ein direkter Zusammenhang zu dieser Diplomarbeit liegt bei der automatischen Spracherkennung der Firma EML[4], siehe Abschnitt 1.3.2 ab Seite 10. Eine genaue Abgrenzung beider Diplomarbeiten ist im Abschnitt 1.4 (ab Seite 12) beschrieben und in Abbildung 1.2 auf Seite 13 grafisch dargestellt.

1.2.4 Untertitelung im Zuge des Pilotprojektes GESTU

In dieser Diplomarbeit wird der Terminus *Untertitelung* verwendet, wenn diese „eine Tonsubstitution darstellt und somit hörbeeinträchtigten Menschen ermöglicht, einen visuellen Zugang zu einer auditiven Information (gesprochene Sprache, Lieder, Geräusche, Hintergrundstimmen, etc.) zu bekommen“ [Hat11, S: 48]. Untertitel sind für schwerhörige Menschen eine sehr wichtige Unterstützung, um beispielsweise das TV Angebot nutzen zu können (vgl. [Lis08, S: 189]). Auch im Bildungssektor stellt die Untertitelung durch Spracherkennungssoftwares ein großes Potenzial für hörbeeinträchtigte Menschen dar (vgl. [TGN$^+$10]). In [Hat11, S: 47-80] sind Möglichkeiten zur (live) Untertitelerzeugung erörtert. Dies umfasst neben Methoden zur *manuellen Untertitelerzeugung*[5] auch das Potenzial von (automatischer) Spracherkennungssoftwares und der Respeaking Technik.
Resultierend aus einer Respeaking Fallstudie (siehe [Hat11, S: 93-102]) wurde erstmals in Österreich die Live-Untertitelung einer universitären Lehrveranstaltungen durch die Firma Titelbild[7] durchgeführt. In Folge wurden während des GESTU Pilotprojektes weitere Lehrveranstaltungen durch die Firma Titelbild untertitelt. Die Resultate dieser Live-Untertitelungen sind in [Nem13, Abschnitt 4.3] dokumentiert sowie evaluiert. Den Studierenden konnten dabei die Untertitel auf einem Notebook lesen und somit dem Vortrag folgen (vgl. [Nem13, Abschnitt 4.3]).
In der jüngeren Vergangenheit wurden unabhängig von GESTU zunehmend universitäre Lehrveranstaltungen aufgezeichnet und in E-Learning Plattformen den Studierenden zur Verfügung gestellt. Zu erwähnen sei hier beispielsweise das so genannte *LectureTube*[8] Projekt der TU

[4] Anm. Autor: European Media Laboratory GmbH (EML) mit Sitz in Heidelberg, `www.eml-development.de`, letzter Zugriff: 20.12.2012

[5] Anm. Autor: Methoden zur manuellen Untertitelerzeugung wie dem Schnellschreiben mit QWERTZ bzw. QWERTY-Tastaturen[6], dem (Computer-) Stenografieren und dem Veyboard (Velotype) sind in [Hat11, S: 59-80; 116-124] erörtert und anhand der Kriterien *Verfügbarkeit* und den *Qualitätsaspekten* wie sprachliche Qualität, der Verzögerung der Untertitel sowie deren grafische Aufbereitung bewertet.

[7] TITELBILD Subtitling and Translation GmbH, Teil der Red Bee Media Group, `www.titelbild.de`, letzter Zugriff 09.03.2012.

[8] `http://teachingsupport.tuwien.ac.at/fuer_lehrende/lecturetube`, letzter Zugriff: 20.03.2012

Wien. LectureTube erlaubt Lehrenden seit dem Sommersemester 2010, deren Lehrveranstaltungen aufzuzeichnen. Ermöglicht wird dies durch in Hörsälen fest installierte Hardware sowie einer mobile Aufnahmeeinheit. Anschließend können die Aufzeichnungen in TUWEL (die E-Learning und Kommunikationsplattform der TU Wien) eingebunden und via Streaming den Studierenden online zur Verfügung gestellt werden (vgl. [Nem13, Kapitel 5], [Hru10], [Uni10]). Aufgezeichnete Lehrveranstaltungen in Audio- oder Videoform sind jedoch ohne adäquaten schriftlichen (oder gebärdeten) Inhalt für hörbeeinträchtigte Studierende eine beträchtliche bis unüberwindbare Hürde (vgl. [Hat11, S. 133]).
Nach der Analyse des mehrfach preisgekrönten, Cross-Browser[9] Open Source Annotationssystem *Synote*[10] (vgl. [Wal11], [Lew10], [Lew12]) in [Hat11] sieht der Autor dieser Diplomarbeit in der Plattform die nötigen Voraussetzungen, um den „heterogenen Bedürfnissen hörbeeinträchtigter Studierender im E-Learning Bereich gerecht zu werden“ [Hat11, S.136]. Dies begründet sich durch - die in E-Learning Plattformen übliche - Möglichkeit, Videos und somit Aufzeichnungen von Gebärdensprachdolmetscherinnen und Gebärdensprachdolmetscher einzubinden. Somit kann beispielsweise jenen hörbeeinträchtigten Studierenden ein adäquater Zugang zur gesprochenen Information gewährleistet werden, die bevorzugt in einer Gebärdensprache wie ÖGS kommunizieren. Weiters - und für den Fokus dieser Diplomarbeit ausschlaggebend - wurde Synote nicht zuletzt unter Berücksichtigung der Bedürfnisse von hörbeeinträchtigten Studierenden entworfen. Die Funktionen von Synote werden im Abschnitt 3.1.1 ab Seite 52 näher erläutert.

1.3 Motivation

Einer der technischen Schwerpunkte von [Hat11] lag bei der Evaluierung der Erzeugungsmöglichkeiten von Live-Untertiteln. Solche können durch eine ASR, der Respeaking Technik oder durch verschiedene Möglichkeiten der manuellen Untertitelerzeugung erstellt werden, siehe [Hat11, S. 59-80; 116-124]. Die Untertitelung kann, wie eingangs erläutert, eine wichtige Unterstützung für hörbeeinträchtigte Studierende darstellen. Die visuelle Informationsaufnahme kann aber nicht nur *während* des Lehrbetriebs (wie Vorlesungen) erfolgen: mit Untertiteln können hörbeeinträchtigte Studentinnen und Studenten auch außerhalb von Vorlesungssälen unterstützt werden. In Synote eingebundene Untertitel stellen darüber hinaus nicht nur für hörbeeinträchtigte, sondern auch für fremdsprachige Studierende (deren primäre Sprache nicht die Vortragssprache ist), eine wichtige Informationsquelle dar. Für Synote aufbereitete und dort eingebundene Aufzeichnungen sind nicht nur archiviert, sondern durch das Transkript auch durchsuchbar. Synote wurde nicht nur für hörbeeinträchtigte Studierende entworfen und wird auch von allen Studierenden genutzt und wertgeschätzt (vgl. [Wal10a]). Daher entsteht der in [Hat11, S: 50] dokumentierte *Mehrwert für alle*.
Wie im vorigen Abschnitt 1.2.4 ab Seite 5 bereits angeführt, wurden im Zuge von GESTU einige Lehrveranstaltungen durch die Firma Titelbild live mittels Respeaking untertitelt. Durch eine Vereinbarung zwischen GESTU und der Firma Titelbild wurden die live erzeugten Unterti-

[9] Anm. Autor: Webinhalte der Platform Synote sind in mehreren Browsern (wie Internet Explorer, Google Chrome, FireFox und Safari) lauffähig, siehe [Hat11, S: 136].
[10] `http://synote.org`, letzter Zugriff: 29.03.2012.

tel auf Wunsch anschließend samt Zeitcodes[11] dem GESTU Projekt zur Verfügung gestellt (vgl. [Nem13, Kapitel 5 und 6]). Aufgrund dieser Tatsache begründet sich die Empfehlung in [Hat11]:

> „Im weiteren GESTU Projektverlauf sollen auch Transkripte aus den live untertitelten Vorlesungen zusammen mit Videoaufzeichnungen in Synote eingebunden und für weitere Erkenntnisse herangezogen werden. Eine Anbindung der Spracherkennungssoftware des EML könnte in Zukunft [...] die Erzeugung von Transkripten für Synote vereinfachen“ [Hat11, S: 143].

Um auch ohne (bzw. nicht ausschließlich mit) der Firma Titelbild eine höhere Untertitelquote im E-Learning Bereich zu erreichen, stellte sich dem GESTU Team und auch dem Autor dieser Diplomarbeit die Frage, auf welche (im besten Fall auch ressourcenarme und kostengünstige) Ansätze zukünftig zurückgegriffen werden kann. Diese Fragestellung führte schließlich zum Thema dieser Diplomarbeit. Es wurden zwei thematisch verwandte Möglichkeiten der Untertitelerzeugung für E-Learning Plattformen ausgewählt: Die *Respeaking Technik* sowie die *(automatisch) Erzeugung von Untertitel durch eine Spracherkennungssoftware*. Beide Varianten der Untertitelerstellung bedienen sich automatischer Spracherkennungssysteme, die sich aber in ihrer Funktionsweise stark unterscheiden.
Bei der Respeaking Technik werden Spracherkennungssysteme zum *Diktieren* verwendet, die in der Regel eine alternative Eingabequelle (wie beispielsweise ein Tastaturersatz) sind und auch speziell für diesen Einsatzzweck entworfen wurden (vgl. [TGN$^+$10], [RF11, S: 63-71]). Als Spracherkennungssoftware für das Respeaking bzw. die ausgearbeitete Respeaking Ausbildung wählte der Autor dieser Diplomarbeit *Dragon NaturallySpeaking* (kurz als DNS oder *Dragon* bezeichnet)[12]. Die Spracherkennung des European Media Laboratory GmbH (EML[13]) wurde schließlich als jene der automatischen Erzeugung von Untertitel ausgewählt.
Die Spracherkennung des EML ist konträr zu Diktiersystemen speziell für *spontane Sprache* entworfen. Frei gesprochene, spontane Sprache wird im Alltag oder aber auch in Vorlesungen gesprochen. Sie unterscheidet sich in der Regel stark von geschriebener Sprache, für die Diktiersysteme konzipiert sind. So kommen bei spontaner Sprache grammatikalische Fehler und Lückenfüller (’ähm’, ’öhm’, ’hm’, etc.) ebenso vor wie Sätze die nicht zu Ende gesprochen werden (vgl. [Hat11, S: 59-70; 116-124]).

Die Auswahl der Untertitelungsmethoden ist in den folgenden Abschnitten 1.3.1 und 1.3.2 detailliert begründet.

---

[11] Anm. Autor: „So genannte Zeitcodes (engl. *timecodes*) werden dazu verwendet, die Untertiteleinblendung mit der Video/Audio-Spur zu synchronisieren“ [Hat11, S: 56], siehe [Hat11, S: 56] sowie 2.5.4 für nähere Informationen wie verbreitete Formate.

[12] Anm. Autor: Es wurde die Premium Edition von Dragon NaturallySpeaking in der Version 11 verwendet, die zum Zeitpunkt der Diplomarbeitserstellung die aktuellste Version war. Siehe Abschnitt 3.1 detaillierte Spezifikation und die Gründe für die Verwendung von DNS 11 sowie Abschnitt 2.5 bezüglich einer Übersicht bezüglich Spracherkennungssystemen.

[13] Anm. Autor: Siehe Abschnitt 1.3.2 bezüglich der Gründe für die Verwendung der Spracherkennungssoftware von des EML sowie Abschnitt 4.3.1 bezüglich der Evaluierung der Ergebnisse.

1.3.1 Respeaking Ausbildung (Scripting)

Eine Respeaking Ausbildung ist innerhalb von zwei bis drei Monaten möglich. Im Vergleich zu anderen Methoden der Live-Untertitelung ist die Trainingszeit die mit Abstand geringste. Darüber hinaus ist Respeaking die kostengünstigste Methode zur Untertitelerstellung, siehe Tabelle 2.1 auf Seite 17 (vgl. [Lam06], [RF11, S: 15; Tabelle 2.1]). Respeaking stellt neben Simultanschnellschreiberinnen bzw. Simultanschnellschreibern (dem Schriftdolmetschen mit QWERTZ Tastaturen arbeiten) die einzige Möglichkeit zur Live-Untertitelung von deutschsprachigen Vorlesungen dar (vgl. [Hat11, S: i]). Da es in Österreich seit 2010 vom ÖSB (Österreichischer Schwerhörigenbund DACHVERBAND) eine Ausbildung zum Schriftdolmetschen mittels Tastatur gibt (vgl. [Od11], [Hat11, S: 77; 102]), stellt Respeaking dabei eine Alternative mit potentiell kürzerer Ausbildung dar.
Ein ausschlaggebender Grund für die Entscheidung eine Respeaking Ausbildung zu erarbeiten beruht auf der Tatsache, dass Respeaking noch ein sehr junges akademisches Gebiet mit vielen spannenden Forschungsfragen ist[14]. Weiters findet Respeaking zunehmend mehr Beachtung und wird seit 2007 auch an einigen Universitäten in Europa unterrichtet[15] (vgl. [RF11, S: 22-42], [RF12a]), aber noch nicht in Österreich.
Arumí-Ribas und Romero Fresco erarbeiteten 2008 im Paper *A Practical Proposal for the Training of Respeakers 1* [ARRF08] die Grundlage für eine (akademische) Respeaking Ausbildung. Im Jahr 2011 veröffentlichte Romero Fresco das erste und bisher einzige Fachbuch, dass sich ausschließlich dem Respeaking[16] widmet. Das Buch trägt den Titel *Subtitling Through Speech Recognition: Respeaking*. Das Paper aus dem Jahr 2008 [ARRF08] und auch jenes aus 2012 mit dem Titel *Respeaking in Translator Training Curricula: Present and Future Prospects* [RF12b] geben teilweise Einblick in die Ausbildungen an Universitäten in Spanien und UK. Die genannten Quellen ermöglichen dem Leser bzw. der Leserin zwar einen Überblick über den Aufbau und Inhalt der Ausbildungen, stellen aber keine Lehr- und Lernunterlage dar.
Der im Zuge dieser Diplomarbeit erarbeitete, dokumentierte und evaluierte Trainingsplan soll dazu führen, dass eine (ggf. adaptierte) Ausbildung an österreichischen Universitäten Einzug findet. Somit könnten zukünftig Untertitel durch Tutorinnen und Tutoren mit abgeschlossener Respeaking Ausbildung erstellt werden. Um die Kosten für eine solche Ausbildung gering zu halten, sollte der Aufwand 3 ECTS (75 Stunden)[17] (ein an Universitäten oft üblicher Übungsumfang) nicht überschreiten. Das Training sollte weiters innerhalb der lt. Lambourne [Lam06] zwei- bis dreimonatigen Ausbildungsdauer zum Respeaker bzw. zur Respeakerin stattfinden. Daher soll langfristig die Diplomarbeit dazu beitragen, einen Grundstein für eine akademische Ausbildung zum Respeaker bzw. zur Respeakerin an österreichischen Universitäten zu legen.

[14] Anm. Autor: Viele dieser Forschungsfragen gehen über die Zielsetzung dieser Diplomarbeit hinaus.
[15] Anm. Autor: Der Abschnitt 2.1 bietet einen ausführlichen geschichtlichen Überblick zum Thema Respeaking.
[16] Anm. Autor: Respeaking wie es in Europa angewandt wird.
[17] Anm. Autor: 1 ECTS entspricht an der TU Wien 25 Arbeitsstunden (vgl. [Pou03]).

Scripting

Bei der Erzeugung von Untertitel für das E-Learning handelt es sich um so genannte *vorbereitete Untertitel*, die auch als *offline Untertitel* bekannt sind (vgl. [Hat11, S: 50]). Die Erzeugung solcher Untertitel mittels Respeaking wird auch als *Scripting* bezeichnet (vgl. [RF11, S. 23]), siehe Abschnitt 2.3 bezüglich einer Diskussion mit anschließender Definition von Respeaking. Das Ausbildungsziel Scripting als Variante des Respeakings wurde aus folgenden Gründen ausgewählt:

- **Effektive Untertitelerstellung**: Bei der traditionellen Untertitelerstellung mittels Tastatur beträgt bei Red Bee Media (RBM)[18] in UK das Verhältnis (engl. *ratio*) von Arbeitszeit der Untertitelerstellung zur Filmlänge 10:1. Beim Respeaking kann sich das Verhältnis auf bis zu 7:1 verkürzen (vgl. [RF11, S. 23]).

- **Geringe Ausbildungszeit**: Aus Sicht des Autors dieser Diplomarbeit *könnte*[19] die Scripting Ausbildungsdauer (noch) kürzer sein als die in diesem Abschnitt bereits erwähnten zwei- bis dreimonatige Ausbildung zum Live Respeaking. Dies begründet sich vor allem auf der im nächsten Aufzählungspunkt erläuterten, einfacher durchführbaren Korrektur.

- **Korrekturmöglichkeit**: Während einer Live-Untertitelung mittels Respeaking gibt es drei Möglichkeiten mit Fehlern[20] umzugehen: *keine Korrektur*, *Eigenkorrektur* oder *parallele Korrektur*. Beim Scripting bestehen weitere Korrekturmöglichkeiten die eine höhere Qualität begünstigen, Abschnitt 2.2.1 ab Seite 21.

- **Zeitlich sowie räumlich entkoppelt**: Durch die Tatsache, dass es sich beim Scripting um die Untertitelung einer Aufzeichnung handelt, entsteht eine zeitliche Entkopplung die einige Vorteile mit sich bringt. Zum einen kann die Zeit der Erstellung frei gewählt werden (bzw. mit dem Auftraggeber oder der Auftraggeberin ein Zeitrahmen vereinbart werden). Zum anderen können während des Untertitelungsprozesses auch selbständig und nach individuellen Bedürfnissen Pausen gewählt werden. Durch die zeitliche Entkopplung geht auch die räumliche einher. So können Respeaker und Respeakerinnen beim Scirpting eher von einem Heimarbeitsplatz aus arbeiten als in Live-Situationen, da bei letzteren eventuell das Risiko von technischen Problemen (Übertragungsprobleme, etc.) nicht eingegangen werden kann/will. Die Möglichkeit eines Heimarbeitsplatzes wird weiters dadurch begünstig, dass beim Scripting eine einzelne Person die Untertitelung durchführen kann und somit nicht auf einen Teampartner oder eine Teampartnerin angewiesen ist, siehe nächster Aufzählungspunkt.

[18] Anm. Autor: RBM erstellt u.a. die (live) Untertitel für den britischen Fernsehsender BBC und beschäftigte 2006 bereits 50 Respeaker und Respeakerinnen (vgl. [Mar06]).

[19] Anm. Autor: Es ist allerdings anzunehmen, dass ein intensiveres und längeres Training zu qualitativ höheren Untertiteln führt. Daher kann auch beim Scripting eine längere Ausbildung angestrebt werden.

[20] Anm. Autor: Ursachen für Fehler sind im Abschnitt 2.5.1 ab Seite 33 sowie im Abschnitt 2.2.1 ab Seite 21 erläutert.

- **Verzögerung**: Unter Verzögerung (*delay*) ist der Zeitunterschied zwischen der Anzeige der jeweiligen Untertitelstelle zu dem Gesprochenen zu verstehen, siehe Abschnitt 2.3 ab Seite 29. Bei den durch die Firma Titelbild live erstellten Untertiteln betrug die Verzögerung bis zu 30 Sekunden. Die Verzögerung wurde von einer hörbeeinträchtigten Studentin (die gleichzeitig dem Vortragenden mit Lippenlesen folgte) als störend empfunden (vgl. [Nem13, Abschnitt 4.3]). Auch beim Scripting entsteht eine Verzögerung zwischen dem Gesprochenen und den Untertiteln. Jedoch können aufgrund der zeitlichen Entkoppelung die Zeitwerte korrigiert werden. Je konstanter die ursprüngliche Verzögerung war, desto einfacher kann im Nachhinein die Synchronität hergestellt werden (z.B. durch das Verschieben der Zeitcodes um 15 Sekunden).

- **Durchführbar durch *einen* Respeaker bzw. *eine* Respeakerin**: Bei der Erstellung von Live-Untertitel mittels Respeaking arbeiten im Regelfall mindestens zwei Respeakerinnen bzw. Respeaker, die sich entweder im 20-40 Minuten Abstand abwechseln oder die wie bei Interviewsituation gleichzeitig arbeiten (vgl. [RF12b]). Durch die zeitliche Entkoppelung können beim Scripting die Pausen individuell gewählt werden. Somit ist es möglich, dass eine Person eigenständig eine ganze Lehrveranstaltungseinheit von 90 Minuten Dauer untertitelt.

- **Erlernen der Fähigkeiten zur Live-Untertitelung**: Die benötigten Kompetenzen zur Live-Untertitelung (wie die Eigenkorrektur während der Untertitelerstellung oder der Umgang mit dem erhöhten Stress) können nach erfolgter Scripting Ausbildung gezielt erlernt und geübt werden. Somit bietet Scripting aus Sicht des Autors dieser Diplomarbeit die optimalen Voraussetzungen für eine kurze Ausbildungszeit, die ggf. eine Fortbildung zur *Live*-Untertitelung ermöglicht.

1.3.2 ASR von European Media Laboratory GmbH (EML)

Entstehung der Kooperation mit EML

Im Zuge der Recherche zur Diplomarbeit [Hat11] besuchte der Autor dieser Diplomarbeit die Abschlusskonferenz des Net4Voice Projektes[21] und konnten Kontakte zur Universität Ulm (Projektpartner von Net4Voice) knüpfen. Während des Net4Voice Projektes wurde erhoben, dass es für „die deutsche Sprache keine ASR gibt, die spontane (Unterrichts) Sprache transkribieren kann“ [Hat11, S: 67]. Aus diesem Grund startete die Universität Ulm eine Kooperation mit dem EML. Daraus resultierend war bereits zum Projektende von Net4Voice eine ASR für spontane, deutsche Unterrichtssprache verfügbar (vgl. [TGN+10]). Durch den Kontakt zur Universität Ulm entstand schließlich auch jener zwischen dem EML und dem Autor dieser Diplomarbeit (bzw. im weiteren dem GESTU Pilotprojekt). Während eines Treffens[22] in Ulm wurden erste Ergebnisse vorgeführt. Dabei handelte es sich um erzeugte Untertitel einer (hochdeutschen)

[21] Net4Voice final conference: „Speech recognition supporting learning: The future“ 13. Mai 2010 in Bologna, Italien; siehe Abschnitt [Hat11, S: 69] sowie `www.net4voice.eu`, letzter Zugriff 20.12.2012

[22] Anm. Autor: Das Treffen fand am 27.09.2010 in Ulm statt. Seitens GESTU waren Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Wolfgang Zagler, Werner Nemecek (Autor von [Nem13]) und der Autor dieser Diplomarbeit anwesend.

TV Nachrichtensendung. Die Resultate waren aus Sicht des Autors dieser Diplomarbeit bereits vielversprechende[23]. Es wurde im Zuge des Treffens das Ziel gesetzt, zukünftig ähnlich gute Resultate auch für österreichische Unterrichtssprache zu erreichen. So sollte mittels dem zur Verfügung stellen von Audiomaterial der aufgezeichneten Vorlesungen sowie der dazugehörigen Unterlagen (wie Skripten, Folien, etc.) die Spracherkennung verbessert werden. Wie erwähnt sollten dabei gezielt die Anforderungen des tertiären Bildungssektors in Österreich berücksichtigt werden.
Da Nemecek in seiner Diplomarbeit ([Nem13]) die Einführung von unterstützenden Hilfsmittel für hörbehinderte Studierende behandelt und im Zuge dessen die Weiterentwicklung der Spracherkennung für GESTU durchführte, übernahm er den Kontakt zu dem EML.

Ablauf der Transkription mit der ASR des EML

Wie bereits einleitend im Abschnitt 1.3 angeführt, unterscheiden sich die beiden ausgewählten Spracherkennungssysteme stark voneinander[24]. Konträr zu Dragon NaturallySpeaking handelt es sich bei der Software vom EML um „eine offline Spracherkennung, die Sprecher bzw. Sprecherinnen unabhängig und für spontane Sprache konzipiert ist“ [Hat11, S: 69]. Bei der Software des EML wird eine Sprachdatei (also die Audioaufzeichnung) direkt via Weboberfläche durch eine Technikerin bzw. einen Techniker an einen Server übertragen. Idealerweise wird dann noch zur Verbesserung der Erkennungsgenauigkeit das verwendete Fachvokabular (beispielsweise in Form von Vortragsfolien, Skripten, Zeitungsartikeln, etc.) hochgeladen. Anschließend kann der Transkriptionsprozess gestartet und nach Fertigstellung die Untertiteldatei lokal abgespeichert werden. Konträr zu Spracherkennungssystemen die beim Respeaking verwendet werden, muss die Software des EML nicht speziell auf die vortragende Person[25] trainiert[26] werden (vgl. [Nem13, Abschnitt 3.2]).

Rechtliches

Das Hochladen von Dateien (Aufzeichnungen, Skripten, Folien, etc.) auf den Server des EML erfolgt aus rechtlichen Gründen erst nach schriftlicher Zustimmung der Urheber bzw. Urheberinnen. Nur im Falle einer Zustimmung werden die Daten „für die Ableitung von statistischen Parametern wie Worthäufigkeiten oder Wahrscheinlichkeitsverteilungen und somit zur stetigen Weiterentwicklung der ASR“ [Hat11, S: 91] verwendet. Siehe [Hat11, S: 33-35] sowie [Nem13, Abschnitt 5.3] bezüglich rechtlicher Aspekte bei der LVA Aufzeichnung bzw. der Weiterleitung von Daten an das EML.

---

[23] Anm. Autor: Dabei handelte es sich allerdings um einen subjektiven Eindruck. Die Qualität wurde nicht qualitativ (wie mittels der WER, der WRR oder dem NER-Modell, siehe Abschnitt 2.6 ab Seite 41) evaluiert.

[24] Anm. Autor: An dieser Stelle sei auf [Hat11, S: 59-70] verwiesen, wo eine detaillierte Kategorisierung von Spracherkennungssoftwares angeführt ist (Sprecher/Sprecherinnen ab- oder unabhängig, zum Diktieren oder ASR für spontane Sprache, live oder offline Transkription, etc.).

[25] Anm. Autor: Da die Software Sprecher bzw. Sprecherinnen unabhängig ist.

[26] Anm. Autor: Da die Software nicht nur Sprecher bzw. Sprecherinnen unabhängig sondern auch für spontane Sprache entworfen ist.

1.4 Abgrenzung

Das Magisterstudium Informatikmanagement[27] soll die Absolventinnen und die Absolventen für den Unterricht an Schulen und im freien Bildungssektor qualifizieren. Ein weiteres Berufsbild ist die „Vermittlung von Kenntnissen über Informations- und Kommunikationstechniken und Informationstechnologien“ [UT03] außerhalb des Bildungsbereiches (vgl. [UT03]).
Daher liegt der Schwerpunkt dieser Diplomarbeit in der Ausarbeitung, Dokumentation und Evaluierung der Respeaking Ausbildung und nicht in der (technischen) Weiterentwicklung der ASR des EMLs. Folglich wurde eine Schnittstelle zwischen der Diplomarbeit von Nemecek und dieser Diplomarbeit festgelegt: Nemecek fokussierte sich auf die Weiterentwicklung der Spracherkennungssoftware des EML, welche in [Nem13, Abschnitt 3.2] dokumentiert ist. Der Autor dieser Diplomarbeit fokussierte sich auf die Respeaking Ausbildung mit Dragon. Die Zielsetzung war es, die mit zwei unterschiedlichen Ansätzen bzw. Methoden erstellten Untertitel in dieser Diplomarbeit zu vergleichen. Dazu wurde eine Vorlesung ausgewählt, für die der auszubildende (bzw. dann ausgebildete) Respeaker Untertitel erstelle. Weiters wurde die selbe Vorlesung mit der Spracherkennungssoftware des EMLs transkribiert. Darüber hinaus sollten auch die (personellen und technischen) Ressourcen und die Praxistauglichkeit verglichen werden. Schließlich konnte jedoch mit der ASR des EMLs keine Qualität erreicht werden, die einen (sinnvollen) qualitativen Vergleich mit der im Abschnitt 2.6 erläuterten NER-Modell ermöglichte, siehe Abschnitt 4.3.1 ab Seite 102. Aus diesem Grund wurde der Schwerpunkt neben der Ausarbeitung der Respeaking Ausbildung verstärkt in die Evaluierung der Ergebnisse der (für die ausgewählte Vorlesung mittels Respeaking) erstellten Untertitel gelegt. Zu diesem Zweck wurden zwei weitere qualitative Vergleiche durchgeführt. Zum einen wurde die ausgewählte Vorlesung (live) von der Firma Titelbild untertitelt und die Resultate im Zuge dieser Diplomarbeit evaluiert. Zum anderen wurde die ausgewählte Vorlesung auch vom Autor dieser Diplomarbeit[28] mittels Respeaking untertitelt und qualitativ mit den Untertiteln vom Auszubildenden sowie den Untertiteln der Firma Titelbild verglichen.

1.4.1 Respeaking Ausbildung (Scripting)

Der Autor dieser Diplomarbeit hat die Ausbildung so konzipiert, dass sie in adaptierter Weise auch zukünftig zum Training der Live-Untertitelung von Respeakerinnen und Respeakern verwendet werden kann. Dennoch ist aufgrund der im Abschnitt 1.3.1 angeführten Motivationsgründen die Ausbildung auf eine offline Transkription mittels Respeaking (Scripting) konzipiert und müsste dementsprechend adaptiert werden.
Die Dokumentation der Ausbildung in den Kapiteln 3 und 4 soll es weiters ermöglichen, dass das Training zukünftig in ähnlicher Weise von anderen Personen (Ausbildnerinnen und Ausbildnern) durchgeführt werden kann. Allerdings setzt auch die ausführliche Dokumentation eine intensive Beschäftigung mit dem Themenbereich Respeaking (wie das Studium der erwähnten

[27] Anm. Autor: Siehe Qualifikationsprofil im Studienplan des Magisterstudiums Informatikmanagement, `www.informatik.tuwien.ac.at/lehre/studien/master/informatikmanagement`, letzter Zugriff: 26.12.2012.

[28] Anm.: Der Autor dieser Diplomarbeit hat vor bzw. auch während der Ausarbeitung der Trainingseinheiten und der dazu gehörenden praktischen Übungen selbst die Respeaking Tätigkeit erlernt.

Literatur) sowie praktische Erfahrungen der zukünftigen Trainerinnen und Trainer voraus.
Die Einheiten wurden im Einzelunterricht durchgeführt. Der Unterricht einer Gruppe von zukünftigen Respeakern bzw. Respeakerinnen wurde nicht erprobt. Demnach könnte für den Gruppenunterricht eine (vermutlich geringfügige) Anpassungen der Einheiten nötig sein.
Weiters soll das erarbeitete Training den Grundstein für eine Respeaking Ausbildung in Österreich legen, was einer fortlaufende Verbesserung der Einheiten bedarf.
Die erarbeitete Ausbildung ist als Grundausbildung zum Respeaker bzw. zur Respeakerin zu verstehen. Ein (aus Sicht des Autors dieser Diplomarbeit wichtiges) weiterführendes Training zur Verbesserung nach erfolgter Grundausbildung ist nicht Teil dieser Diplomarbeit.

Abbildung 1.2: Kernbereiche von GESTU und Themenbereiche der im Zuge des Pilotprojektes verfassten, wissenschaftlicher Publikationen

KAPITEL 2

Respeaking

2.1 Einführung und Geschichte

Wie im Abschnitt 1.2.4 auf Seite 5 sowie in [Hat11, S: 116-124] diskutiert, stellen Untertitel für hörbeeinträchtigte Menschen eine sehr wichtige Unterstützung bei der Informationsaufnahme dar. Dabei kommen Untertitel nicht nur im TV, sondern u.a. auch bei Vorträgen, im Kino, bei Theateraufführungen oder beim Karaoke zum Einsatz (vgl. [MoP11], [RF11, S: 45-55], [Hat11, S: 116-124]). Ein Vergleich von verschiedenen Methoden, die eine Untertitelung in der deutschen Sprache erlauben, ist in [Hat11, S: 116-124] dargelegt.
Wie schnell ein Fortschritt bei der Untertitelung im Allgemeinen aber speziell bei der Live-Variante vonstattengeht, ist üblicherweise eng mit gesetzlichen Rahmenbedingungen oder beispielsweise Vereinbarungen zwischen Regierungen und öffentlichen Rundfunkanstalten verbunden. Solche Vereinbarungen bzw. Gesetze folgen meist dem Druck von gehörlosen oder schwerhörigen Organisationen (vgl. [AOD07, S: 25]). Im Vereinigten Königreich (UK) verpflichtete 1990 eine gesetzliche Regelung den Fernsehsender BBC dazu, die Quote der Untertitel (live sowie vorbereitete/offline Untertitel[1]) bis 2010 auf 90% zu erhöhen. Dem Sender war es schließlich aus finanzieller Sicht nicht mehr möglich, den steigenden Bedarf an Live-Untertitel wie bisher mit den (darüber hinaus wenigen) Computerstenografen bzw. Computerstenografinnen zu decken. Daher wurde im Jahr 2001 bei der BBC eine alternative Möglichkeit der Live-Untertitelung mithilfe von Spracherkennungssoftwares (ASR, engl. *Automatic Speech Recognition*, respektive automatische Spracherkennung) bei Sportübertragungen erprobt. Neben der BBC setze auch die Rundfunkanstalt VRT in Flanders im Jahr 2001 die selbe Technik zur Untertitelerzeugung ein, die nun als *Respeaking* bekannt ist. Respeaking wird mittlerweile in vielen Ländern und in unterschiedlichsten Bereichen der (live) Untertitelung eingesetzt (vgl. [RF11, S: 22-55], [Mar06]).
Das Kapitel 1 gibt bereits einen Einblick in die Respeaking- sowie Scripting Technik. Letztere ist dabei als eine Unterkategorie des Respeakings zu verstehen, mit der vorbereitete Untertitel erstellt werden (offline Untertitelung). Das Kapitel 2 widmet sich im Weiteren detailliert dem Respeaking samt einer genaue Definition im Abschnitt 2.3 ab Seite 29. Zu Beginn soll folgendes Zitat die Respeaking Tätigkeiten verdeutlichen:

[1] Anm. Autor: Abhängig vom Zeitpunkt der Untertitelerstellung kann zwischen vorbereiteten (offline Untertiteln), Live- bzw. Echtzeit (online Untertitel) sowie Semi-Live-Untertitel unterschieden werden (vgl. [Ore06], [Hat11, S: 50-51]).

Bei der Untertitelung mittels Respeaking empfangen lt. [NH12]

> „[...] die Respeaker oder Respeakerinnen die zu untertitelnde Information über Kopfhörer, kürzen sowie formulieren diese gegebenenfalls um und diktieren schließlich das Gesagte in ein Mikrofon. Eine spezielle Spracherkennung (ASR), die auf die Stimme und Aussprache der Respeakerin bzw. des Respeakers trainiert ist, transkribiert schließlich das Nachgesprochene. Somit kann die Respeakerin oder der Respeaker die Untertitel direkt 'auf den Bildschirm sprechen'. Die ausgebildete Person achtet beim Diktieren auf eine saubere Aussprache, diktiert weiters die Satzzeichen (SZ)[2] und bereitet sich und die Spracherkennung vorab auf das Vokabular und die Thematik generell vor. Dies soll schließlich zu guten Erkennungsraten seitens der Spracherkennung führen. Eventuelle Erkennungsfehler können dabei noch vom Respeaker oder der Respeakerin korrigiert werden, bevor die Untertitel auf dem Bildschirm erscheinen."

Respeaking sollte bei der Einführung bei BBC eine kostengünstigere und praktischere Art für den steigenden Bedarf an Live-Untertitelung darstellen. Die Firma Red Bee Media (RBM) war früher Teil von BBC und ist seit Oktober 2005 eine eigenständige Firma mit Sitzen in Europa, Asien und Australien. Etwa fünf Jahre nach den ersten Respeaking Versuchen wurden bei RBM bereits zwei Drittel der jährlich erzeugten 20.000 bis 25.000 Stunden Live-Untertitel von Respeakerinnen und Respeakern erzeugt (vgl. [Mar06], [RF11, S: 22-34]). In Tabelle 2.1 ist eine

	Verzögerung	**Wörter pro Minute (WpM)**	**Wortakkuratheit**	**Ausbildungszeit**	**Kosten**	**Schwere der Fehler**
Velotype	mittel	mittel (90-120)	95%	12 Monate	mittel	hoch
Dual Tastatur	mittel	mittel bis hoch (140-150)	95-98%	6 Monate	mittel	niedrig
Computerstenografie	niedrig	sehr hoch (220- bis zu 300)	97-98%	3 Jahre	hoch	mittel
Respeaking	niedrig	hoch (160-190)	97-98%	2-3 Monate	niedrig	mittel-hoch

Tabelle 2.1: Vergleich der Live-Untertitelungsmethoden ([RF11, Table 2.1]), orig. Lambourne; Real time subtitling - extreme audiovisual translation (2007)

Übersicht verschiedener Methoden zur Live-Untertitelerzeugung anhand verschiedener Kriterien dargelegt. Wie der Tabelle zu entnehmen ist, zeichnet sich Respeaking mit der geringen

[2] Anm. Autor: In dieser Diplomarbeit sind Satzzeichen (wie Punkt, Beistrich, Anführungszeichen, etc.) eine Unterkategorie von Interpunktionszeichen (IZ). Letztere beinhalten zusätzlich sämtliche zu diktierten Zeichen (wie Bindestrich, neuer Absatz, etc.). Im angeführten Zitat ist demnach lt. dieser Definition von Interpunktionszeichen die Rede.

Verzögerung[3], der zweithöchsten Wortanzahl je Minute, der Wortakkuratheit/der Genauigkeit[4], dem kurzen Training sowie den geringen Gesamtkosten aus. Ein zu hervorhebender, negativer Aspekt ist allerdings die Schwere der auftretenden Fehler[5].
Beim Österreichischen Rundfunk kam im Jahr 2010 Respeaking erstmals als kosteneffizientere Alternative zu den bisherigen Schnellschreibern und Schnellschreiberinnen zum Einsatz. Zwei Jahre später arbeiteten ca. 20 Respeakerinnen und Respeaker beim ORF, wobei bereits einige vor dem Einsatz von Respeaking in der Untertitelungsabteilung der Rundfunkanstalt arbeiteten (vgl. [NH12], [Wal12, S: 60]). Der „Etappenplan zum Ausbau des barrierefreien Zugangs zu den ORF-Fernseh-Programm und zum ORF-Online-Angebot gemäß § 3 Abs. 1 Z 2 ORF-Gesetz“ [ORF12a] sieht für das Gesamtprogramm auf ORF eins und ORF 2 für 2013 eine Erhöhung der Untertitelquote auf 65% der Sendestunden vor (vgl. [ORF12a]), siehe Abbildung 2.1. Die im Vergleich zum Vereinigten Königreich niedrige Untertitelungsquote zeigt aus Sicht des Autors dieser Diplomarbeit, dass in Bezug auf Barrierefreiheit beim ORF noch Aufholbedarf besteht. Jedoch heißt es im Etappenplan weiters, dass der ORF als österreichisches öffentlich-rechtliches Leitmedium darauf setzt, „Themen von gesellschaftlicher, sozialer und politischer Relevanz sowie reichweitenstarke Programme barrierefrei auszustrahlen, sodass eine daran anschließende öffentliche Diskussion aktiv und erlebt mitgeführt werden kann“ [ORF12a].
Erstmals wurden auch in Österreich im Zuge des GESTU Pilotprojektes live Untertitel für universitäre Vorlesungen durch die Firma Titelbild (Teil der Red Bee Media Group) erzeugt. Dabei konnten die hörbeeinträchtigten Studierenden die Untertitel auf einem Notebook lesen und somit dem Vortrag folgen (vgl. [Nem13, Abschnitt 4.3]). Die Abbildung 2.2c[6] zeigt eine Respeakerin der Firma Titelbild während der Untertitelerstellung. In Abbildung 2.2a sowie 2.2b sind mittels Respeaking erzeugte Untertitel im Hörsaal (während einer Fallstudie im Zuge von [Hat11]) zu sehen. Einige der durch die Firma Titelbild live untertitelten Lehrveranstaltung wurden auch aufgezeichnet. Die live erzeugten Untertitel wurden anschließend zusammen mit den Aufzeichnungen in E-Learning Plattformen eingebunden (vgl. [Nem13, Abschnitt 4.3 und 5.3]). Eine für die E-Learning Platform Synote aufbereitete Präsentation ist in Abbildung 3.1a auf Seite 54 zu sehen.
Untertitel können auf sprachlicher Ebene in *intralinguale*- sowie *interlinguale* Untertitel unterteilt werden. Bei ersterer findet keine Übersetzung der Ausgangssprache (dem Gesagten) statt

[3] Anm. Autor: Unter Verzögerung (*delay*) ist der Zeitunterschied zwischen der Anzeige der jeweiligen Untertitelstelle zu dem Gesprochenen Inhalt zu verstehen, siehe Abschnitt 2.3 ab Seite 29.

[4] Anm. Autor: Methoden zum Feststellen der Akkuratheit bzw. Fehlerhäufigkeit sind im Abschnitt 2.6 ab Seite 41 angeführt.

[5] Anm. Autor: Fehler können in ihrer Ursache sowie Schwere unterschieden werden. *Editierfehler* sind zum Beispiel das Weglassen oder Hinzufügen von Information (vgl. [RFM14]) und können durch den Faktor Mensch bei allen der genannten Live-Untertitelungsmethoden vorkommen. Beim Respeaking kommen weiters *Erkennungsfehler* vor. Diese können zum einen auftreten, wenn der Spracherkennung ein Wort nicht bekannt ist. Zum anderen kann es auf den Umgang mit der Spracherkennung zurückzuführen (Aussprache, der Diktiergeschwindigkeit, der Lautstärke, etc.) sein. Nicht immer können vom Publikum Erkennungsfehler als solche erkannt werden. So z.B. falsche Zahlen. 'Gefährlich' wird es lt. einem Respeaker vom ORF ([Wal12, S: 120; Interview 8]) auch dann, wenn beispielsweise während einer Parlamentsdiskussion „jüdische Flüchtlinge“ anstatt „libysche Flüchtlinge“ von der ASR erkannt wird. Darüber hinaus können Erkennungsfehler schwer lesbare oder verwirrende Untertitel zur Folge haben. Die Unterscheidung und Analyse von Erkennungs- und Editierfehlern ist im Abschnitt 2.6.1 ab Seite 42 näher erläutert.

[6] Quelle/Urheber: E-Mail der Firma TITELBILD Subtitling and Translation GmbH [Hat11, S: 71]

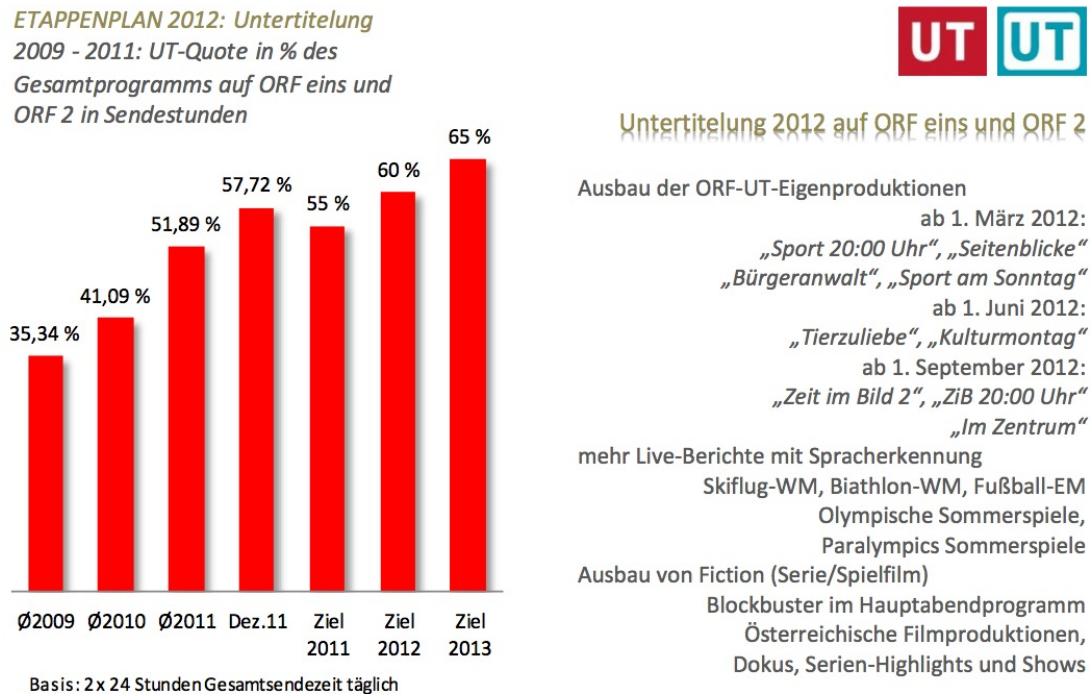
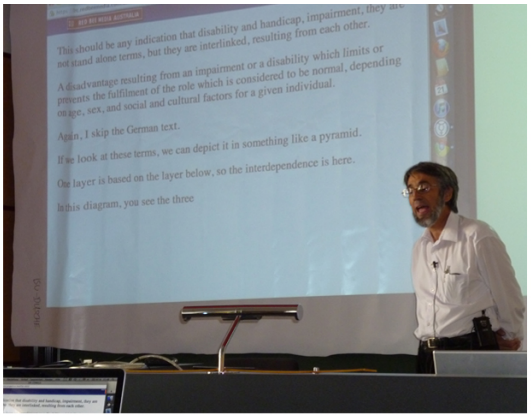
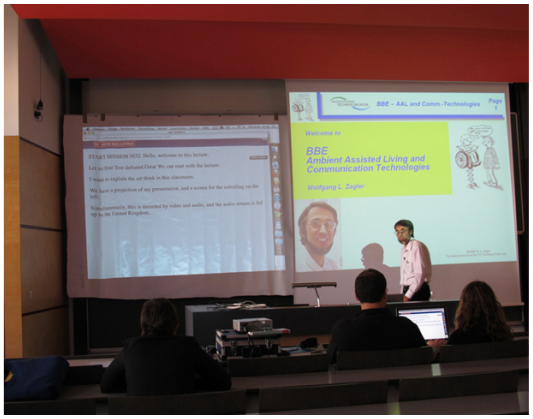


Abbildung 2.1: ORF Etappenplan zum Ausbau der Untertitelquote sowie Untertitelung 2012 auf ORF eins und ORF 2 [ORF12a]

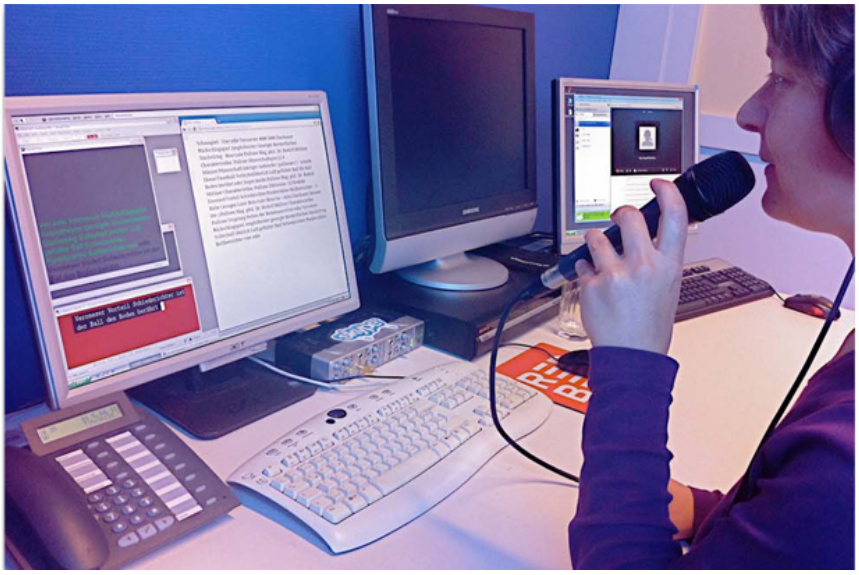
(vgl. [Bak98, S: 247], [Hat11, S: 51]). Beim Respeaking handelt es sich in den meisten Fällen um eine intralinguale Untertitelung. Jedoch kann, wie in einigen Fällen in Flanders, auch eine Übersetzung stattfinden und somit Respeaking auch für die interlinguale Untertitelerzeugung verwendet werden (vgl. [RF11, S: 22-47]).
Respeaking findet darüber hinaus auch zur Erstellung von vorbereiteten Untertiteln (offline) Anwendung. Respeakerinnen und Respeaker verbringen bei RBM in etwa die Hälfte ihrer Tätigkeit mit der Live-Untertitelung, die restliche Zeit mit der Erzeugung von vorbereiteten Untertiteln. Wie im Kapitel 1 bereits erläutert, wird die Erstellung von offline Untertitel mittels Respeaking auch als Scripting bezeichnet (vgl. [Mar06], [RF11, S. 23]). Die Untertitelung von aufgezeichneten Lehrveranstaltungen ist schließlich auch das Ausbildungsziel der erarbeiteten und im Kapitel 3 dokumentierten Respeaking/Scripting Ausbildung.



(a) Respeaking Fallstudie: vortragender Professor, dessen englischer Vortrag live untertitelt wurde (die Untertitel sind links im Bild an der Leinwand bzw. auf dem Bildschirm des Notebooks am linken unteren Bildrand zu lesen)



(b) Hörsaalübersicht während einer Respeaking Fallstudie: die Projektion der Untertitel links im Bild; die Folien zum Vortrag rechts im Bild; hörbeeinträchtigte GESTU Mitarbeiterin mit Notebook (ebenfalls mit Untertiteln) im Publikum



(c) Eine Respeakerin der Firma TITELBILD Subtitling and Translation GmbH bei der Arbeit

Abbildung 2.2: Untertitelung einer Lehrveranstaltung mittels Respeaking (vgl. [Hat11])

2.2 Respeaking Tätigkeit

2.2.1 Ablauf Respeaking: Vor- und Nachbereitung, Untertitelung

Da es sich beim Respeaking um eine komplexe Tätigkeit handelt bei der viele - und zum Teil auch simultane - Handlungen durchgeführt werden, steht üblicherweise eine Ausbildung am Beginn einer Respeaking Karriere. Lt. Romero-Fresco unterscheiden sich die jeweiligen Ausbildungen teilweise stark voneinander (vgl. [RF11, S: 22-44]).
Im wesentlichen sollen im Zuge einer Ausbildung jene Fähigkeiten erlernt werden, die für die (im besten Fall qualitativ hochwertige) Ausführung der Respeaking Tätigkeit benötigt werden. Das umfasst u.a. den Umgang mit dem Equipment (der Spracherkennung, der Untertitelsoftware, dem Mikrofon, etc.) und das Erlernen des gleichzeitigen Zuhörens, Umformulierens und Diktierens samt Fehlerkorrektur. Weiters sollen Respeakerinnen und Respeaker im Rahmen der Ausbildung die verschiedenen Anforderungen des Zielpuplikums kennen lernen und mit der qualitativen Analyse der Untertitel vertraut werden.

Romero Fresco unterscheidet in [RF11, S: 50-54] drei Phasen beim Respeaking:
Die **Vorbereitung** (engl. *prior to the process*), die **Untertitelerstellung** (engl. *during the process*) und die **Nachbereitung** (engl. *after the process*). Die drei Phasen werden im Folgenden näher betrachtet.

Vorbereitung

Die erste Phase beim Respeaking ist die (auftragsbezogene) Vorbereitung. Dabei recherchiert üblicherweise ein Respeaker oder eine Respeakerin das zu untertitelnde Thema, um inhaltlich damit vertraut zu werden. Um bei der Untertitelung von TV Sendungen Informationen über den Inhalt des Formats zu erhalten, kann eine Teilnahme an Redaktionssitzungen von Vorteil sein. Im Zuge des GESTU Pilotprojektes wurde den Respeakern bzw. Respeakerinnen der Firma Titelbild (wenn vorhanden) die Vortragsfolien und im Idealfall auch das Vorlesungsskriptum zur Verfügung gestellt. Damit sollte eine gute Vorbereitung samt Training der Spracherkennung ermöglicht werden. Dabei werden die (zu erwartende) Wörter - die nicht im Wortschatz der Spracherkennung enthalten sind - eintrainiert um später korrekt erkannt zu werden (vgl. [Nem13, Abschnitt 3.1 und 4.3], [Wal12, S: 61]).
Darüber hinaus muss in der Vorbereitung auch das Zielpuplikum und dessen Bedürfnisse an die Untertitel erörtert werden. Zusätzlich müssen ebenso gesetzliche oder unternehmesinterne Regulierungen in dieser Phase bekannt sein, um bei der Untertitelerstellung berücksichtigt zu werden. Wie beispielsweise beim ORF, wo lt. [Wal12, S: 131-141; Interview 7]) interne Guidelines/StyleGuides gibt, die auch laufend adaptiert werden.
Bei der Live-Untertitelung mittels Respeaking wird häufig im Team gearbeitet. Dabei wechseln sich die Respeaker bzw. Respeakerinnen (als R1 bzw. R2 bezeichnet) üblicherweise in 20-40 Minuten Rhythmen ab. R1 und R2 können sich auch gegenseitig bei der Korrektur von Fehlern unterstützen. Weiters ist es möglich, dass bei mehreren Sprechern bzw. Sprecherinnen (z.B. bei Interviews im TV) beide Teammitglieder gleichzeitig diktieren/Untertitel erstellen (vgl. [RF12b], [RF11, S: 29]). So kann z.B. R1 für die Erstellung der Untertitel des Reporters bzw.

der Reporterin und R2 für jene der interviewten Person zuständig sein. Auch die Absprache mit der Teampartnerin bzw. dem Teampartner über die genaue Form der gegenseitigen Unterstützung sowie den Ablauf beim Wechsel ist Teil der Vorbereitungsphase.
Bei einer Semi-Live-Untertitelung werden die Untertitel (oder Teile davon) vor der Ausstrahlung vorbereitet und dann 'manuell' zu gegebenem Zeitpunkt gesendet. Dadurch müssen beispielsweise bei internationalen Sportübertragungen die Texte der Nationalhymnen nicht live untertitelt werden (vgl. [Hat11, S: 51; 75]).

Die Vorbereitungsphase ist „ein enorm wichtiger Schritt, um in der Live-Situation gute Untertitel produzieren zu können" [Wal12, S: 63]. Die Interviews, die Walter mit Respeakerinnen und Respeakern vom ORF führte, zeigen wie unterschiedlich die Vorbereitung sein kann. So wird u.a. berichtet, dass es aus Kostengründen nicht immer eine Vorbereitungszeit gibt und es erwartet wird, dass das Trainieren der Spracherkennung 'zwischendurch' (also wenn das andere Teammitglied gerade untertitelt) geschieht (vgl. [Wal12, S: 141-147; Interview 8]).

Untertitelung ('Respeaking')

Bei der Phase des 'eigentlichen' Respeakings müssen Respeaker bzw. Respeakerinnen folgende Tätigkeiten gleichzeitig durchführen: *Zusehen* und *Zuhören*, das *Respeaken/Diktieren* des Gesagten samt Interpunktionszeichen, das *Lesen* der erzeugten Ausgabe sowie die *Korrektur* etwaiger Fehler (vgl. [RF11, S: 100-101]). Während der Untertitelung sind die Respeakerinnen und Respeaker oftmals einem großen Stress ausgesetzt, der sich bei der Live-Untertitelung vor allem durch den Zeitfaktor begründet. Das Stressempfinden kann aber mit zunehmender Routine geringer werden (vgl. [Wal12, S: 62-63]). Die Tätigkeiten sind im Folgenden näher erläutert:
Beim *Zuhören* verfolgt ein Respeaker oder eine Respeakerin - üblicherweise über Kopfhörer bzw. ein Headset - das Gesprochene. Bei der TV-Untertitelung sind meist zusätzliche optische Informationen in Form des Fernsehbildes verfügbar (*Zusehen*). Ähnlich ist es bei der Untertitelung von universitären Lehrveranstaltungen wenn Folien vorhanden sind. Durch diese hat die respeakende Person während der Untertitelung zusätzlich zur akustischen (dem Gesprochenen) eine visuelle Information verfügbar. Das kann beispielsweise bei Aufzählungen, Zahlen, Fachausdrücken, der Erläuterung von Bildern und Grafiken, etc. hilfreich sein.
Das Gehörte muss schließlich der Spracherkennung diktiert werden. Eine sehr wichtige Aufgabe besteht beim *Respeaking/Diktieren* im Umformulieren und/oder Kürzen (vgl. [RF11, S: 100-101]). Diese sprachlichen Aspekte sind im Abschnitt 2.2.2 ab Seite 26 detailliert diskutiert. Wie erläutert findet beim Respeaking meist keine Übersetzung der Ausgangssprache (dem Gesagten) statt. Allerdings können Respeakerinnen und Respeaker (wie in einigen Fällen in Flanders) zusätzlich simultan eine Übersetzung durchführen (vgl. [RF11, S: 22-47]).
Bei der Live-Untertitelung mittels Respeaking gibt es drei Möglichkeiten mit Fehlern umzugehen: *keine Korrektur*, *Eigenkorrektur* oder *parallele Korrektur*. Dabei stellt Romero Fresco fest, dass die größte Schwierigkeit beim Respeaking nicht im Multitasking selbst liegt, sondern bei den gleichzeitig stattfindenden aber nicht überlappenden Aufgaben. Das ist gerade bei der Eigenkorrektur - der üblichsten Form der Korrektur - der Fall. In Abbildung 2.3 sind die fünf Tätigkeiten beim Respeaking bei einer Eigenkorrektur in einem Zeitstrahl visualisiert. Entscheidend ist dabei, dass beim Korrigieren (mittels Tastatur) nicht simultan diktiert werden kann, da

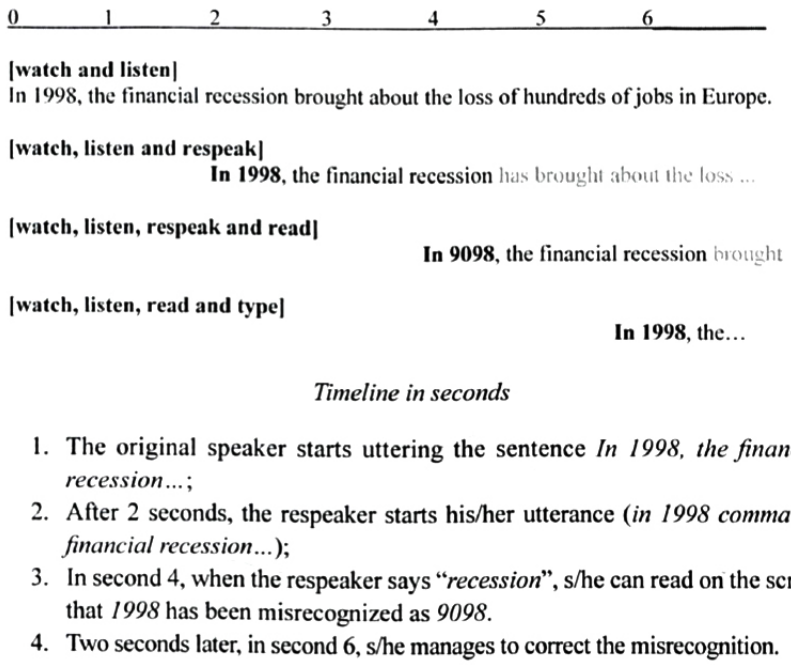


Abbildung 2.3: Ablauf einer automatischen Spracherkennungssoftware (ASR) [RF11, S: 100]

die Spracherkennung in dem Fall währen des Korrigierens an der jeweiligen Position des Cursors die erkannten Wörter einfügen würde. Während der Eigenkorrektur muss somit ein Respeaker bzw. eine Respeakerin weiterhin dem Gesagten folgen, um es nach der Korrektur zu diktieren (während zu diesem späteren Zeitpunkt wiederum bereits zugehört werden muss). Durch eine Eigenkorrektur erhöht sich dabei der Abstand zum Gesagten. Die Anforderungen beim Simultandolmetschen sind in einigen Aspekten jenen beim Respeaking ähnlich. Allerdings kann ein Dolmetscher bzw. eine Dolmetscherin bei etwaigen Problemen durch schnelleres Sprechen wieder 'Aufholen'. Das ist beim Respeaking nur begrenzt möglich, da beim Diktieren auf einen optimalen Rhythmus ebenso geachtet werden muss wie auch auf maximale Diktiergeschwindigkeiten (vgl. [RF12b, S: 9], [RF11, S: 100-101]). Ein Respeaker bringt die Problematik im Interview mit Walter auf den Punkt: „Wenn ich im Stress bin und dadurch schlecht rede, dann im Stress bin beim Ausbessern und das zu noch mehr Stress führt, dann ist es ein Teufelskreis" [Wal12, S: 95-101; Interview 2].

Beim Respeaking gibt es während einer Live-Situation neben der Eigenkorrektur zwei weitere Möglichkeiten, mit Fehlern umzugehen. Eine davon ist die *parallele Korrektur*. Dabei korrigiert eine zweite Person die Fehler des Respeakers oder der Respeakerin kurz nachdem die ASR das Diktierte transkribiert hat. Somit erfolgt die Korrektur noch bevor die Untertitel beispielsweise live im TV ausgestrahlt werden (vgl. [RF11, S: 1-17]). Im Gegensatz zur Eigenkorrektur muss dabei die gerade diktierende Person die Ausgabe der Spracherkennung nicht verfolgen.

Sehr unüblich ist die Variante in welcher *keine Korrektur* stattfindet (vgl. [RF11, S: 1-17]). Die

Möglichkeit ist aber aus Sicht des Autors dieser Diplomarbeit für Live-Untertitel im tertiären Bildungsbereich ungeeignet, da inhaltliche Fehler in Untertiteln einen erheblichen Nachteil für hörbeeinträchtigte Studierende darstellen. Auch wenn Fehler durch eine Korrektur nicht ausgeschlossen werden können, so ist das Fehlerrisiko ohne Korrektur aus Sicht des Autors dieser Diplomarbeit im (tertiären) Bildungsbereich zu hoch und die dadurch resultierenden, negativen Auswirkungen auf den Lernerfolg zu schwerwiegend.
Beim Scripting ist es möglich, dass während des Diktierens selbst keine Korrektur stattfindet und Fehler anschließend korrigiert werden. Darüber hinaus gibt es die Möglichkeit beim Scripting die Wiedergabe der Aufzeichnung zu pausieren (vgl. [RF11, S: 23]), ggf. sogar zurückzuspulen. Auch die Wiedergabegeschwindigkeit kann individuell verringert werden[7] um Beispielsweise hohe Sprechgeschwindigkeiten zu bewältigen (ohne automatisch kürzen zu müssen) und/oder um den Stress zu verringern. Eine zusätzliche Möglichkeit der Fehlerkorrektur beim Scripting sieht der Autor dieser Diplomarbeit im *Markieren* von Fehlern. Wie beschrieben kann bei der Eigenkorrektur im Live-Einsatz nicht gleichzeitig korrigiert und diktiert werden kann. Beim Korrigieren erhöht sich somit der Abstand zum Gesprochenen. Beim Scripting besteht aber die Möglichkeit, dass sich eine Respeakerin oder ein Respeaker auf das Diktieren und Beobachten der erzeugten Untertitel konzentriert. Anstatt unmittelbar eine Korrektur durchzuführen, können erkannte Fehler mittels Tastatur 'markiert' (z.B. durch das Einfügen eines Rautezeichens) und in der Nachbereitung schnell gefunden und korrigiert werden. Mit dieser Arbeitsweise kann ohne Unterbrechung diktiert werden.
Somit ergeben sich beim Scripting eine Vielzahl von kombinierbaren Möglichkeiten, um den Stressfaktor zu verringern bzw. mit Fehlern umzugehen: Anpassung der Wiedergabegeschwindigkeit, Pausieren bei Müdigkeit bzw. Fehlern, Trainieren von vorab nicht bekannten bzw. nicht zu erwartenden Vokabular, Zurückspulen, Markieren von Fehlern mit späterer Eigenkorrektur sowie generell die Korrektur in der Nachbereitungsphase.

Nachbereitung

Die Nachbereitung kann aus Sicht des Autors dieser Diplomarbeit in zwei Bereiche unterteilt werden: Zum einen in die **Aufbereitung** bzw. **Korrektur** der live erstellten Untertitel für eine etwaige weitere Verwendung. Zum anderen in die **Analyse** und damit verbunden mit einer fortlaufenden Steigerung der Qualität. Demnach soll nach erfolgter Untertitelung eine Respeakerin oder ein Respeaker in der Lage sein, die Qualität der Untertitel zu analysieren. Dabei kann die Analyse selbst oder durch andere erfolgen (vgl. [RF11, S: 54]). Hier sei auf die NER-Analyse verwiesen, die im Abschnitt 2.6 ab Seite 41 beschrieben ist. Die NER-Analyse macht die Qualität von Untertiteln - im speziellen Hinblick auf Respeaking - messbar und vergleichbar[8].

[7] Anm. Autor: Wenn die Audio- oder Videoaufzeichnung nicht nur online als Stream, sondern der Respeakerin oder dem Respeaker auch als Datei verfügbar ist, kann z.B. mit dem Mediaplayer VLC die Wiedergabegeschwindigkeit dynamisch verringert werden (VLC Version 2.0.1 für Mac OS X, `www.videolan.org/vlc`), letzter Zugriff: 09.05.2012). Auch mit der im Abschnitt 3.1.5 ab Seite 57 beschriebenen Open Source Software *Subtitle Workshop* von *UruSoft* (`www.urusoft.net/products.php?cat=sw&lang=1`, letzter Zugriff 22.10.2012) ist eine Anpassung der Wiedergabegeschwindigkeit möglich.

[8] Anm. Autor: Die Analyse der Qualität stellt auch einen wichtigen Teil des im Kapitel 3 erarbeiteten Trainings dar. So wird durch das Feststellen der Qualität bereits während des Lernprozesses der Fortschritt messbar. Durch die Analyse von Fehlern samt deren Ursachen kann somit der Lerneffekt positiv beeinflussen werden.

Aus Kostengründen wird in der Praxis teilweise auf eine (intensive) Analyse der Qualität verzichtet. Beim ORF erfolgt beispielsweise keine mit dem NER-Modell vergleichbar Analyse der Untertitel. Dennoch findet in vielen Fällen eine Korrektur der Fehler statt. Grund dafür ist die Verwendung der live erstellten Untertitel für das ORF Video-on-Demand-Portal *TVthek*[9]. Um eine effiziente Korrektur zu ermöglichen, berichten Respeaker und Respeakerinnen, dass sie während der Sendung die Fehler mitschreiben. Anschließend werden sie dann für die TVthek ausgebessert sowie nicht erkannte Wörter für zukünftige Einsätze trainiert (vgl. [Wal12, S: 85-147; Interviews 1-8]). Nichtsdestotrotz wird im Zuge der Nachbereitung aus Kostengründen eine Fehlerkorrektur nicht immer durchgeführt: „[...] 4 Stunden Parlament nachzubearbeiten, das dauert mindestens einen Tag und kostet Geld und wird deshalb nicht gemacht“ [Wal12, S: 141-147; Interview 8]. Selbst wenn das Mitschreiben von Fehlern und eine nachträgliche Korrektur für eine Respeakerin bzw. eine Respeakerin sicherlich einen gewissen Lerneffekt bewirken, so ersetzt es aus Sicht des Autors dieser Diplomarbeit die systematischen Analyse (wie mit dem NER-Modell) nicht.

Um Frust zu vermeiden, müssen Respeaker bzw. Respeakerinnen lernen mit Fehlern umzugehen. Darüber hinaus soll lt. Romero-Fresco der Zugang zu positivem und negativem Feedback sowie die persönliche Beurteilung bzw. der Umgang mit dem Feedback Teil der Nachbereitung sein (vgl. [RF11, S: 54]). Beim ORF werden vor und nach jeder Sendung die Namen der Respeakerinnen bzw. Respeaker samt einer Emailadresse eingeblendet. Es gibt auch die Möglichkeit die Untertitelabteilung via Telefon oder Fax zu kontaktieren. In den Interviews von Walter geben die Respeakerinnen und Respeaker an, dass sie relativ wenig direktes Feedback bezüglich der Qualität erhalten und wenn dann meistens nur Negatives. Die meisten Rückmeldungen kommen lt. einer Respeakerin dann, wenn „etwas nicht untertitelt wird, was halt aus Sicht des Publikums interessant gewesen wäre“ [Wal12, S: 131-141; Interview 7] oder eine angekündigte Untertitelung ausgefallen ist (vgl. [Wal12, S: 85-147; Interview 1-8]). Neben dem Erkennen von eigenen Schwächen und der daraus resultierenden Möglichkeit zur Verbesserung kann gerade auch positives Feedback vom Zielpuplikum aus Sicht des Autors dieser Diplomarbeit ein wichtiger Faktor für eine Motivationssteigerung sein. Daher ist es wichtig, den Kontakt zum Publikum direkt zu suchen und auch gezielt nach positiven Aspekten der Untertitel fragen. Dies erfolgte u.a. durch die von Nemecek erstellten Fragebögen, deren Auswertung in [Nem13, Abschnitt 4.3] dokumentiert ist. Die Evaluierung ermöglicht einen Einblick in die Sichtweise der jeweiligen hörbeeinträchtigten Studierenden bezüglich der durch die Firma Titelbild erstellten Untertitel.

[9] `http://tvthek.orf.at`, letzter Zugriff: 09.01.2013

2.2.2 Sprachliche Aspekte: Umformulieren und Kürzen beim Respeaking

Das *Umformulieren* und das *Kürzen* (engl. *editing*) gehören zu den am meisten diskutierten Themen in der Untertitelungsliteratur. Anders als bei einer Wort für Wort/'1:1' Untertitelung (engl. *verbatim*) weicht der Inhalt bei gekürzten und/oder umformulierten Untertiteln von der Originalsprachquelle (dem Gesprochenen) ab. Ein *Kürzen* und/oder *Umformulieren* soll dabei positive Auswirkungen auf die *Lesbarkeit* haben, zu besseren *Erkennungsraten* führen sowie zu einer kürzeren *Verzögerung* (engl. *delay*, siehe Abschnitt 2.3 ab Seite 29) der Untertitel beitragen. Bei hohen Sprechgeschwindigkeiten ist ein Kürzen auch durch menschliche und maschinelle Grenzen bedingt. Allerdings kann aus Sicht von hörbeeinträchtigten Menschen gerade das Kürzen als Zensur empfunden werden. Abhängig vom Grad der Umformulierung bzw. Kürzung verschlechtert sich weiters die Synchronität zum Gesprochenen. Daraus resultierend kann der Bezug zwischen Untertitelung und Visuellen verloren gehen (vgl. [ARRF08], [RF11, S: 22-44; 95-122], [RF12b]).
Nicht zuletzt aufgrund der einleitend angeführten Aspekte gibt es verschiedene Interessengruppen, die bei der Untertitelerstellung mittels Respeaking involviert sind. Sie vertreten unterschiedliche Standpunkte, die im folgenden näher erläutert werden. So liegen die Interessen von TV Sendern und Untertitelungsfirmen meist in der Erfüllung einer Quote bzw. sind ökonomischer Natur. Gleichzeitig sind von Respeakerinnen und Respeakern die Erwartungen von hörbeeinträchtigten Menschen an die Untertitelung zu berücksichtigen. Diese basieren weniger auf finanziellen als auf politischen Aspekten (wie z.B. den Anspruch auf vollen Informationszugang). Eine weitere Interessengruppe ist jene der Wissenschaftlerinnen und Wissenschaftler, die im Bereich der Untertitelung forschen und sich mit diesem politisch sensiblen Bereich beschäftigen. Dabei geht es um die Fragestellung, ob und in welcher Form das Umformulieren und Kürzen die Lesbarkeit der Untertitel beeinflusst und wie ein Informationsverlust quantifiziert werden kann (vgl. [RF11, S: 112-113]).
In [Hat11, S: 51-54] sind sprachliche Aspekte, die bei der Untertitelung generell zu berücksichtigen sind, ausführlich diskutiert. Diese umfassen u.a. den Detailgrad (Kürzen) sowie das Umformulieren des Gesagten, speziell im Hinblick auf Untertitel für den tertiären Bildungsbereich. Da es sich um ein wichtiges Spannungsfeld der Untertitelung handelt, werden im Folgenden die wichtigsten Aspekte aus [Hat11, S: 51-54] zusammen mit zusätzlichen - speziell bei der Untertitelerstellung mittels Respeaking entscheidenden - Faktoren erläutert.

Umformulieren

Wie im Abschnitt 1.3 ab Seite 6 bereits erläutert, unterscheidet sich die Alltagssprache ebenso wie die in Vorlesungen gesprochene Sprache in der Regel stark von der geschriebenen Form: So kommen bei gesprochener, spontaner Sprache grammatikalische Fehler und so genannte *Lückenfüller* ('ahm', 'öhm', 'hm', etc.) ebenso vor, wie Sätze die nicht zu Ende gesprochen werden (vgl. [RF11, S: 95-122], [PK08, S: 393], [Hat11, S: 59-70; 116-124]). Solche falschen Satzstellungen, „die in der mündlichen Sprachproduktion häufig vorkommen und bei der auditiven Rezeption eher unauffällig sind“ [Now10], können „bei schriftlicher Darstellung und der Aufnahme über den visuellen Kanal einen großen Störfaktor darstellen“ [Now10]. Weiters *kann* die Lesbarkeit der Untertitel durch kürzere Sätze bzw. das Vermeiden von verschachtelten Sät-

zen erhöht werden.
Unabhängig zur Lesbarkeit kann je nach verwendeter Spracherkennung die Verzögerung durch das Aufteilen von langen Sätzen in mehrere kurze Satzblöcke verringert werden. Die im deutschsprachigen Raum für Respeaking sehr verbreitete Spracherkennung Dragon transkribiert die Spracheingabe sobald ein Block (engl. *chunk*) fertig diktiert wurde. Solche Blöcke sind abgeschlossene - oder durch andere Satzzeichen (Komma, Bindestrich, etc.) getrennte - Sätze. Die Länge des Blocks beeinflusst somit die Verzögerung. Allerdings erzielt Dragon bei langen Sätzen bessere Erkennungsraten. Beide Aspekte müssen von Respeakern bzw. Respeakerinnen beim Umformulieren abgewogen und beachtet werden, um ein geeignetes Mittel zu finden (vgl. [ARRF08], [RF11, S: 74-94]).
Verwendet ein Sprecher oder eine Sprecherin Vokabular, welches der Spracherkennung nicht bekannt ist, kann ein Respeaker oder eine Respeakerin Erkennungsfehler durch *Paraphrasieren* (dem Verwenden von Synonymen) vermeiden. Bezüglich Erkennungsfehler sei auch erwähnt, dass oft lange sowie spezifischere Wörter von Dragon seltener falsch erkannt werden ([RF11, S: 34; 74-94]). Der Grund dafür ist, dass beispielsweise weniger Wörter im Wortschatz der Spracherkennung ähnlich wie „Muskelfaserquerschnitt“ klingen. Konträr dazu sei das Wort „sich“ angeführt, das phonetisch u.a. „mich“ ähnelt.
Ungeachtet dessen ob eine Umformulierung aus den dargelegten Gründen beim Respeaking von Vorteil ist bzw. auch einen positiven Einfluss auf Qualität der Untertitel hat, muss sie vom Zielpuplikum nicht unbedingt erwünscht sein: „So kann argumentiert werden, dass es für die gleichwertige Aufnahme von Gesprochenem durch die Untertitelung wichtig ist, *wie* der Sprecher oder die Sprecherin die Sätze formuliert und somit durch eine Umformulierung [...] nicht die gleiche Information übermittelt werden *kann* wie bei einer ’1:1’ Untertitelung“ [Hat11, S: 54].

Kürzen

Bei der Untertitelerstellung sind die Möglichkeiten des Kürzens vielseitig. Sie reichen vom Weglassen der Lückenfüller bis hin zum ausschließlichen Transkribieren der ’wesentlichen’ Informationen. Da die Bedürfnisse von hörbeeinträchtigten Menschen sehr unterschiedlich sind, kann abhängig von der jeweiligen Person ein (leichtes bis starkes) Kürzen präferiert aber auch wie erläutert nicht erwünscht sein.
Unabhängig von den Bedürfnissen ist festzustellen, dass selbst mit dem Ziel einer ’1:1’ Untertitelung mit der Respeaking Technik in den meisten Fällen keine Wort für Wort Transkription erreicht wird. In einer Studie von Romero Fresco wurden bei Sprechgeschwindigkeiten von unter 180 WpM (ohne IZ) 0-20 weniger Wörter je Minute festgestellt. Bei höheren Sprechgeschwindigkeiten waren es ca. 40 Wörter weniger je Minute. Um eine Wort für Wort Untertitelung zu erreichen, müsste man beim Respeaking aufgrund des Diktierens der Interpunktionszeichen mehr Wörter sprechen als man hört. Lt. dem Autor der Studie könnte dies unnatürlich und folglich der Grund für die geringere Wortanzahl sein. Weiters stellte sich heraus, dass bis zu einer Diktiergeschwindigkeit von 180 WpM die Wortanzahl nahezu identisch der Originalquelle ist, wenn die diktierten Interpunktionszeichen als Wörter gezählt werden. Somit diktierten bei den analysierten Untertiteln die Respeaker und Respeakerinnen in etwa gleich viele Wörter (inkl. IZ) als in der Originalquelle ohne IZ vorkamen. Generell nimmt beim Respeaking die Geschwindigkeit eine wesentliche Rolle ein. Dieser Faktor kann in drei Kategorien unterteilt werden: Die

Lesegeschwindigkeit des Publikums, die *Sprechgeschwindigkeit*[10] der zu untertitelnden Person und nicht zuletzt die *Respeakinggeschwindigkeit*[11]. Die (maximale) Arbeitsgeschwindigkeit einer Respeakerinnen bzw. eines Respeakers ist individuell verschieden (vgl. [RF11, S: 95-122], [RF12b]). Sie ist einerseits durch individuelle Faktoren (wie schlechtere Aussprache bei hoher Diktiergeschwindigkeit und damit verbundenen Erkennungsfehlern, dem lückenlosen Folgen des Gesagten bei hoher Sprechgeschwindigkeit bzw. viel Information, der Tagesverfassung) ebenso beeinflusst wie durch die Grenzen der Spracherkennung. Obwohl mit einem Training die eigenen Möglichkeiten (z.B. im Hinblick auf die Diktiergeschwindigkeit) erhöht werden können, existieren menschliche und auch maschinelle Grenzen, die ein Kürzen erfordern.
Ein Kürzen kann auch notwendig sein, wenn (gesetzlichen) Regulierungen für TV Untertitel gewisse (Mindest-)Einblendzeiten vorgeben. Diese können auch Genre abhängig sein. Daher muss gekürzt werden, sobald es zu hohen Sprechgeschwindigkeiten kommt und mehr verbale Information vorhanden ist, als in den Untertitel angezeigt werden kann/darf (vgl. [RF12b]).
Das Kürzen muss allerdings nicht unbedingt einen Informationsverlust (abgesehen vom Verlust der Information *wie* die Sätze gesprochen wurden) nach sich ziehen. Folgende Phrase einer Vorlesung[12] macht dies deutlich: „Also alles was weiß ist, ist Kompensation. Alles was hellblau ist, ist Stabilisierungsbereich. Alles was schwarz ist, sind Wettkämpfe. Alles was hellgrün ist, sind also hier Entwicklungsbereichtrainings.“. Die 27 Wörter bzw. 35 Wörter inklusive Interpunktionszeichen wurden vom ausgebildeten Respeaker[13] wie folgt gekürzt/umformuliert: „Weiß: Kompensation, Hellblau: Stabilisierungsbereich, Schwarz: Wettkämpfe, Hellgrün: Entwicklungsbereichtrainings.“. Dabei wurden 8 Wörter transkribiert (70% gekürzt), 16 inklusive der Interpunktionszeichen (54% gekürzt). Das Beispiel soll zum einen verdeutlichen, wie gekürzte Sätze die Lesbarkeit erhöhen können. Zum anderen, dass die Untertitel aufgrund der Kürzung schneller gelesen werden können und somit mehr Zeit für die Aufnahme von anderer (visueller) Information ist.

[10] Anm. Autor: Im Englischen liegt die Sprechgeschwindigkeit von spontaner Sprache zwischen 140 und 160 Wörter pro Minute ohne Interpunktionszeichen (WpM ohne IZ). Im TV ist die Sprechgeschwindigkeit abhängig vom Genre. In UK wird bei Sportveranstaltungen zwischen 124 und 182, bei Nachrichtensendungen zwischen 161 und 198 und bei Interviews und Wetterberichten zwischen 211 und 245 WpM ohne IZ gesprochen (vgl. [RF11, S: 114]). Der Autor dieser Diplomarbeit führte Stichprobenmessungen bei ORF Sendungen durch. Dabei betrug die Sprechgeschwindigkeit von Beiträgen aus Nachrichtensendungen (ZiB) 145 WpM ohne IZ und 180 WpM inklusive Interpunktionszeichen. Bei Wettervorhersagen 165 WpM ohne IZ, 205 WpM mit IZ. Die Sprechgeschwindigkeiten der Vorlesung Leistungsphysiologie lag zwischen 120 und 208 WpM ohne IZ und zwischen 135 und 234 WpM mit IZ, siehe Abschnitt 4.3.3 ab Seite 104.

[11] Wie in Tabelle 2.1 ersichtlich, überschreiten die angeführten Sprechgeschwindigkeiten zum Teil die Möglichkeiten der jeweiligen Methoden zur Live-Unteritelerstellung. Die Diktiergeschwindigkeiten die nach der erarbeiteten Ausbildung gemessen wurden, sind im Abschnitt 4.3.4 dokumentiert.

[12] Anm. Autor: Die Textpassage stammt aus einem Vortrag der Vorlesungsreihe *Trainingswissenschaft* von Mag. Manfred Zeilinger an der Universität Wien, siehe Abschnitt 4.3.1 ab der Seite 101.

[13] Anm. Autor: Siehe Abschnitt 3.2.2 auf Seite 59. Drei Fehler in den Untertiteln wurden für den Demonstrationszweck korrigiert.

2.3 Definitionen Respeaking

Es gibt zahlreiche Respeaking Definitionen und die Tätigkeit wird auch von involvierten bzw. praktizierenden Personen oft unterschiedlich beschrieben. Ohne an dieser Stelle auf die verschiedenen Definitionen im Detail einzugehen, haben alle zumindest eines gemein: Es findet eine Interaktion zwischen Mensch (dem Respeaker bzw. der Respeakerin) und Maschine (der Spracherkennungssoftware auf einem Computer) statt, die zur Erzeugung der Untertitel dient. Lt. [Hat11] ist Respeaking eine Methode,

> „[...] mit welcher negative Einflussfaktoren auf die Qualität der Erkennungsrate - wie Akzente, Dialekte, Code-Switching (wechselnde Dialekte), Umgebungsgeräusche, Sprechgeschwindigkeit, etc. - umgangen werden können. Dabei wird nicht die ASR Software hinsichtlich der erwähnten Probleme verbessert, sondern ein Sprecher oder eine Sprecherin [...] spricht das Gesagte (die ursprüngliche, akustische Originalquelle) in einer für das ASR gut verarbeitbaren Weise nach. Die durch den Respeaker oder die Respeakerin diktierten Wörter bzw. Sätze werden dann von der ASR transkribiert “ [Hat11, S: 70].

Detaillierter und allgemeiner definiert Romero Fresco Respeaking als:

> „A technique in which a respeaker listens to the original sound of a live programme or event and respeaks it, including punctuation marks and some specific features for the deaf and hard of hearing audience, to a speech recognition software, which turns the recognized utterances into subtitles displayed on the screen with the shortest possible delay“ [RF11, S: 1].

Romero Fresco geht auf die im Zitat unterstrichenen Passagen weiter ein, um das breite Einsatzspektrum von Respeaking zu beschreiben:

- **Live**: Respeaking wird aufgrund des schnellen Durchsatzes heutzutage nicht nur zur Erstellung von Live-Untertiteln angewandt. Auch für Aufzeichnungen werden nachträglich Untertitel mittels Respeaking erstellt ([RF11, S: 1]). Die als *Scripting* bezeichnete Respeaking Variante (vgl. [RF11, S: 23], [ARRF08], [Mar06]) entspricht auch der Tätigkeit, die ein Respeaker bzw. eine Respeakerin mit der im Zuge dieser Diplomarbeit erarbeiteten und evaluierten Ausbildung erlernen soll, siehe Kapitel 3 sowie Kapitel 4.
- **Respeak**: Das Verb *respeaking* definiert Romero Fresco als das Wiederholen, Umformulieren und/oder Übersetzen der Originalsprachquelle (dem Gesprochenen) durch den Respeaker oder die Respeakerin. Im Deutschen haben sich die englischen Begriffe *Re-speaker* bzw. *Re-speaking* durchgesetzt (vgl. [RF11, S: 3]). Das deckt sich mit den Erfahrungen, welche der Autor dieser Diplomarbeit im deutschsprachigen Raum gemacht hat. Aus diesem Grund werden in dieser Diplomarbeit die Substantive *Respeaker* bzw. *Respeakerin* sowie das Verb *respeaken* verwendet und nicht ins Deutsche übersetzt. Es wird, wie auch in [Nem13], [Wal12], [NH12], [Hat11] sowie [Now10], auf die Schreibweise mit Bindestrich (Re-speaking) verzichtet.

- **Some specific features**: Dabei handelt es sich um Information, die zusätzlich zum Gesprochen transkribiert wird. Beispiele dafür sind das Kennzeichnen der sprechenden Person bei mehreren Sprechern oder Sprecherinnen und untertitelte Informationen wie Klatschen, Lachen, etc. (vgl. [RF11, S: 1]). Für eine ausführliche Diskussion der sprachlichen Ebene von Untertiteln sei auf [Hat11, S: 51-54] verwiesen, wo *intra-* und *interlinguale* Untertitel ebenso erläutert und definiert sind wie verschiedene Detailgrade bei der Formulierung (Original mit Untertitel (OmU) und Hörgeschädigten-Untertitel (HG-UT)). Schließlich sind an genannter Stelle auch *paralinguistischen Eigenschaften* (Kennzeichnung von Emotionen wie Flüstern, Untertöne und Stimmlage, etc.) und die möglichen Vor- und Nachteile bei den unterschiedlichen Detailgraden und Formulierungen - speziell für den tertiären Bildungsbereich - ausführlich diskutiert. Dies betrifft u.a die Grenzen der Lesegeschwindigkeit bei gleichzeitigem Folgen von anderen Informationen an der Tafel bzw. von PowerPoint Folien, das begrenzte Zeitfenster der Untertiteleinblendungen, etc.

- **Speech recognition software**: Beim Respeaking kommen oft zwei verschiedene Softwarekomponenten zum Einsatz: einerseits die Spracherkennungssoftware (siehe Abschnitt 2.5.1 ab Seite 33), andererseits die Software welche die erzeugten Transkripte als Untertitel darstellt (vgl. [RF11, S: 2]), siehe Abschnitt 2.5.4 ab Seite 38.

- **Programme/subtitles**: Unter Programm und Untertitel ist die große Bandbreite der Einsatzmöglichkeiten von Respeaking zu verstehen. Sie geht über die Untertitelung im TV hinaus und erstreckt sich hin zum Einsatz in Museen, Theatern, Konferenzen, Karaoke und Kirchen (vgl. [RF11, S: 2; 44-55]). Weiters ist die Erzeugung von Untertitel im E-Learning Bereich und die Live-Untertitelung von universitären Lehrveranstaltungen, wie dies im GESTU Pilotprojekt der Fall war (vgl. [Nem13, Abschnitt 4.3 und 5.3]), ist ein mögliches Einsatzgebiet des Respeakings.

- **Delay**: Unter *delay* ist die Verzögerung der Untertitel zum Gesprochenen zu verstehen. Sie geht auf verschiedene Faktoren zurück und begründet sich aus der eingesetzten Software und der Korrektur Methode. Die Verzögerung stellt dabei in [Hat11, S: 142] ein wichtiges Qualitätskriterium von Live-Untertiteln dar. Lt. Romero-Frescro beträgt sie beim Respeaking zwischen 5 und 15 Sekunden (vgl. [RF12b]). In einer Stichprobenanalyse stellte Nemecek (bei den durch die Firma Titelbild im Zuge des GESTU Pilotprojektes erzeugten) Live-Untertitel eine Verzögerung von 10-15, in Extremfällen von bis zu 30 Sekunden fest (vgl. [Nem13, Abschnitt 4.3]). Auch die Regelmäßigkeit der Verzögerung ist ein wichtiges Qualitätskriterium. So kann die Lesegeschwindigkeit durch einen regelmäßigen Einblenderhythmus markant gesteigert werden (vgl. [Hei06, S: 10]). Wie problematisch sich die Verzögerung auf die Informationsaufnahme auswirken kann, zeigt das Beispiel von untertitelten universitären Lehrveranstaltungen. So kann es bei Folienwechseln bereits mit geringen Verzögerungen dazu kommen, dass sich die Informationen in den Untertiteln auf eine andere Folie beziehen (vgl. [Hat11, S: 100]). Wenn die Verzögerung jedoch gleichmäßig (z.B. immer 8-10 Sekunden) beträgt, können in der Nachbereitung die Zeitcodes um einen konstanten Zeitwert korrigiert werden (vgl. [RF11, S: 2], [Hat11, S: 63; 100; 120-124]).

2.4 Definitionen innerhalb dieser Diplomarbeit

Respeaking hat (wie in diesem Abschnitt bereits dargelegt) eine relativ junge Geschichte, das Einsatzspektrum ist allerdings mittlerweile sehr breit. Für ein einheitliches und klar definiertes Vokabular innerhalb dieser Diplomarbeit ist in diesem Abschnitt die erarbeitete Respeaking/Scripting Ausbildung ebenso definiert wie die Erzeugung der Untertitel durch die ASR des EML. Die Definitionen erfolgen dabei durch die in [RF11, S: 11-17; 22-47] verwendete Klassifizierungsmethode für Untertitelerzeugungen. Daraus resultierend sind die in dieser Diplomarbeit beschriebenen bzw. evaluierten Untertitelungsmethoden direkt zu jenen in [RF11, S: 22-47] in Relation gestellt und mit zukünftigen Entwicklungen vergleichbar.

2.4.1 Respeaking

Für *Respeaking* wird die im Abschnitt 2.3 ab Seite 29 angeführte Definition von Romero Fresco aus [RF11, S: 1] verwendet. Somit handelt es sich beim Respeaking um die Erstellung von (live) Untertiteln sowie Transkripten durch eine[14] Respeakerin oder einen Respeaker, der oder die eine Spracherkennung durch 'nachsprechen' (dem *respeaken*, wo ggf. umformuliert und gekürzt wird) bedient. Der Begriff Respeaking inkludiert darüber hinaus die Tätigkeiten während der *Vorbereitung*, der *Untertitelerstellung* und der *Nachbereitung*, siehe Abschnitt 2.2.1 ab Seite 21.

[14] Anm. Autor: Oder mehrere Respeakerinnen oder Respeaker.

2.4.2 Scripting

Beim *Scirpting* handelt es sich um die Erzeugung von Untertitel/Transkripten einer Audio- oder Videoaufzeichnung durch die Respeaking Technik. Demnach ist Scripting eine Unterkategorie von Respeaking und der in [Hat11, S: 50] erläuterten *offline Untertitelung* (auch *vorbereitete Untertitel* genannt) zuzuordnen. Bei der im Zuge dieser Diplomarbeit erarbeiteten und evaluierten Ausbildung handelt es sich demzufolge um eine Scripting (und somit auch eine Respeaking) Ausbildung, die aber den Anspruch erhebt, auf eine Ausbildung zur Erzeugung von Live-Untertitel erweitert werden zu können. Nach der in [RF11, S: 11-17; 22-47] verwendeten Klassifizierungsmethode handelt es sich bei den durch die ausgebildete Person erzeugten Untertitel um:

> **Offline erzeugte**[15] und nach Möglichkeit **nahezu einer '1:1' Transkription**[16] gehaltene, **intralinguale**[17] Untertitel, die durch die Verwendung einer Spracherkennung mittels der **Respeaking Technik**[18] erstellt und von der respeakenden Person **selbst korrigiert**[19] werden. Die Darstellung erfolgt in der E-Learning Plattform Synote, in der die **Untertitel in Blöcken dargestellt**[20] werden, deren Größe allerdings individuell vom Benutzer bzw. der Benutzerin festgelegt werden kann. Es werden **keine SDH**[21] **Features** eingebunden[22], da es sich um die Untertitelung einer **aufgezeichneten Vorlesungen**[23] handelt und weder Sprecher- oder Sprecherinnenwechsel, noch andere Gründe für SDH Features - wie paralinguistische Eigenschaften bei Ironie oder dgl. - für das Verstehen des Inhaltes wichtig sind.

[15] *Production approach: pre-recoreded* [RF11, S: 11-17]
[16] *Editing policy: near-verbatim* [RF11, S: 11-17]
[17] *Language: intralingually* [RF11, S: 11-17]
[18] *Transcription method: SR (respeaking)* in[RF11, S: 11-17]
[19] *Correction method: self-correction* [RF11, S: 11-17]
[20] *Display mode: blocks* [RF11, S: 11-17]
[21] Anm. Autor: Untertitel für gehörlose und schwerhörige Menschen (engl. *subtitling for the deaf and hard-of-hearing* (SDH)) sind Untertitel mit speziellen Informationen für hörbeeinträchtigte Menschen. Solche Untertitel können z.B. durch farbliches Hervorheben bei mehreren Sprechern bzw. Sprecherinnen die aktuell sprechende Person kennzeichnen und weitere Informationen wie 'Klatschen', etc. enthalten (vgl. [RF11, S: 1]). Sie unterscheiden sich daher von Untertiteln für hörende Menschen, wie beispielsweise bei fremdsprachigen Filmen. Im Deutschen wird je nach Aufbereitung der Untertitel für die jeweilige Zielgruppe wie in [Lis08, S: 219] zwischen *Original mit Untertitel* (OmU) und *Hörgeschädigten-Untertitel* unterschieden (HG-UT).
[22] *SDH features: none* [RF11, S: 11-17]
[23] *Programme type: pre-recoreded* [RF11, S: 11-17]

2.4.3 EML

Wie auch beim Scripting handelt es sich bei der Erzeugung der Untertitel durch die ASR des EML bezüglich der zeitlichen Ebene um offline Untertitel. Dennoch wird hier nicht der Terminus Scripting verwendet, da kein Respeaker bzw. keine Respeakerin involviert ist und die originale Sprachquelle direkt als Eingabemedium für die Spracherkennung dient.
Nach der in [RF11, S: 11-17; 22-47] verwendeten Klassifizierungsmethode handelt es sich bei Untertitelerstellung mittels der ASR des EML um:

> **Offline erzeugte**[15] **'1:1' Transkription**[24] von **intralingualen**[17] Untertiteln durch die Verwendung einer automatischen, Sprecher bzw. Sprecherinnen unabhängig und für spontane Sprache entwickelten **Spracherkennung**[25] ohne **Korrektur**[26]. Die Darstellung erfolgt in der E-Learning Plattform Synote, in der die **Untertitel in Blöcken dargestellt**[20] werden, deren Größe allerdings individuell vom Benutzer bzw. der Benutzerin festgelegt werden kann. Es werden **keine SDH Features**[22], eingebunden, da es sich um die Untertitelung einer **aufgezeichneten Vorlesungen**[23] handelt und weder Sprecher- oder Sprecherinnenwechsel, noch andere Gründe für SDH Features - wie paralinguistische Eigenschaften bei Ironie oder dgl. - für das Verstehen des Inhaltes wichtig sind.

2.5 Equipment

Aus Sicht des Autors dieser Diplomarbeit lässt sich das Equipment eines Respeakers bzw. einer Respeakerin in die Kategorien *Spracherkennungssoftware*, *Untertitelsoftware*, *Hardware (inkl. Mikrofon)* und der *Räumlichkeit* unterteilen. Der nachfolgende Abschnitt bietet einen Überblick über das Equipment beim Respeaking. Das verwendete Equipment für die im Zuge dieser Diplomarbeit erarbeiteten Ausbildung ist im Abschnitt 3.1 ab Seite 52 dokumentiert.

2.5.1 Spracherkennungssoftware

Die Spracherkennungssoftware nimmt neben dem Respeaker bzw. der Respeakerin die wichtigste Rolle beim Respeaking ein. Dabei wird meist eine handelsübliche Spracherkennungssoftware verwendet (vgl. [RF11, S: 22-36], [ARRF08], [Hat11, S: 67-71]). Diese Systeme sollen Benutzerinnen und Benutzern das Arbeiten (wie das Diktieren in Textverarbeitungsprogramme, das Verfassen von E-Mails, etc.) ohne Tastatur ermöglichen[27]. Sie sind daher nicht speziell für das Respeaking und dessen Anforderungen entwickelt, was Auswirkungen auf die Arbeitsweise beim Respeaking hat. Beispiele dafür sind das benötigte Diktieren von Interpunktionszeichen, das Umformulieren in kürzere Sätze zur Verringerung der Verzögerung, die fehlende Exportmöglichkeiten in Untertitelformate (siehe Abschnitt 2.5.4 ab Seite 38), die begrenzte Korrekturmöglichkeit in Software von Drittanbietern, etc.

[24] *Editing policy: verbatim* [RF11, S: 11-17]
[25] *Transcription method: SR* [RF11, S: 11-17]
[26] *Correction method: no correction* [RF11, S: 11-17]
[27] Anm. Autor: So bietet DNS beispielsweise spezielle Funktionen für die Bedienung von Microsoft Outlook und auch Microsoft Word (vgl. [Nua10])

Die Sprecher bzw. Sprecherinnen *abhängige*[28] ASR Systeme *Dragon NaturallySpeaking* und *ViaVoice* sind beim Respeaking in Europa die am weitesten verbreiteten Systeme (vgl. [RF11, S: 22-36], [Hat11, S: 67-67]). Andere Spracherkennungssoftares die in einzelnen Ländern beim Respeaking zum Einsatz kommen (z.B. wenn weder Dragon noch ViaVoice in der Landessprache verfügbar ist) sind u.a. in [RF11] angeführt. Ein qualitativer Vergleich von gängigen Spracherkennungssystemen ist in [Net10] zu finden. Die Funktionsweise von Spracherkennungssystemen ist in [Nem13, Abschnitt 3.1] ausführlich beschrieben und wie folgt zusammengefasst: eine (statistische) Spracherkennung wie DNS oder ViaVoice leitet aus dem digitalen Sprachsignal Phoneme ab. Darauf aufbauend werden anhand des bekannten Wortschatzes die in Frage kommenden Wörter berechnet. Schließlich wird durch das Sprachmodell mittels statistischer Verfahren die am wahrscheinlichsten diktierte Phrase transkribiert. Im Folgenden sind Dragon NaturallySpeaking sowie ViaVoice näher erläutert und die Möglichkeiten zur Verbesserung der Erkennungsraten beschrieben.

Dragon NaturallySpeaking (DNS)

Dragon NaturallySpeaking (DNS) wird von der Firma *Nuance Communications, Inc* (weiter) entwickelt und vertrieben. Das Diktierte wird von der Spracherkennung mit kurzer Verzögerung in Blöcken (mehrere diktierte Wörter auf einmal) transkribiert, wobei die Blöcke meist nach dem Diktieren der Satzzeichen am Bildschirm anzeigt werden (vgl. [Hat11, S: 67-68], [ARRF08]). DNS in der Version 10 lieferte lt. einer im Zuge des Net4Voice[29] Projektes durchgeführten Analyse für Deutsch die besten Ergebnisse unter den getesteten Spracherkennungssystemen (vgl. [Net10]). Dragon NaturallySpeaking wird von Respeakerinnen und Respeakern beim ORF, jenen beim Schweizer Fernsehen (SRF)[30] sowie zur Live-Untertitelerstellung mittels Respeaking bei der Firma Titelbild verwendet (vgl. [Hat11, S: 71], [Mon11, S: ii], [Woj12]). Weiteres kommt DNS u.a. in Flandern und in Italien zum Einsatz (vgl. [RF11, S: 22-36]). Derzeit ist DNS im deutschsprachigen Raum wohl die bedeutendste ASR beim Respeaking. So ist dem Autor dieser Diplomarbeit kein Unternehmen bekannt, dass für Respeaking eine andere Spracherkennungssoftware verwendet.

ViaVoice/ViaScribe

Die zweite verbreitete Spracherkennungssoftware beim Respeaking ist ViaVoice. Die ursprünglich von IBM (International Business Machines Corporation) entwickelte Software ist aufgrund von Verkäufen und Fusionierungen mittlerweile ebenfalls im Besitz der Firma *Nuance Com-*

[28] Anm. Autor: In [Hat11, S: 59-70] eine detaillierte Kategorisierung von Spracherkennungssoftwares (Sprecher/Sprecherinnen ab- oder unabhängig, zum Diktieren oder ASR für spontane Sprache, live oder offline Transkription, etc.) dokumentiert. Siehe auch Abschnitt 1.3 ab Seite 6.

[29] „Net4Voice Projekt wurden speziell die Möglichkeiten von Spracherkennungssoftware sowie der Untertitelung im Bildungsbereich evaluiert. Net4Voice war ein Teilprojekt aus dem europäischen „Lifelong Learning“ Programm (Key Activity 3 ICT), welches von der Europäischen Kommission finanziert wurde. Das etwa zweieinhalb Jahre dauernde Projekt startete im Dezember 2007 und wurde im Mai 2010 beendet.“ [Hat11, S: 66]; `www.net4voice.eu`, letzter Zugriff: 10.10.2012

[30] Anm. Autor: Die Untertitel für das Schweizer Fernsehen werden von SWISS TXT (Schweizerische Teletext AG) erstellt (vgl. [Mon11, S: ii]).

munications, Inc. Ebenso wie DNS benötigt ViaVoice das explizite Diktieren von Satzzeichen (vgl. [RF11, S: 65-66], [Net10]). Es wird aber - im Gegensatz zu DNS - das Gesagte unmittelbar transkribiert und somit das Diktierte wortweise und nicht in Blöcken anzeigt (vgl. [ARRF08]). 2003 war der letzte weltweite Release von ViaVoice in der Version 10. Seither ist die Software nicht mehr kommerziell erwerbbar. Dennoch gab es im Zuge des Net4Voice Projektes eine Weiterentwicklung von ViaVoice. Das so genannte *ViaScribe* verwendet die interne Spracherkennungsengine von ViaVoice und wurde speziell für Lehrveranstaltungen entwickelt . So berichtet u.a. ein Projektmitarbeiter von Net4Voice, dass es bei der Anpassung des Audio-Profils zu sofortigen Programmabstürzen kam und ViaScribe für die deutsche Sprache nie getestet wurde und demnach technisch nicht praktikabel ist. Es wurde im Zuge von Net4Voice festgestellt, dass Dragon Naturally Speaking[31] vor allem für die deutsche Sprache die besten Ergebnisse lieferte (vgl. [Net10], [Weg10]).
Ähnlich wie DNS im deutschsprachigen ist ViaVoice im englischsprachigen Europa sehr verbreitet. So verwenden die zwei bedeutendsten Unternehmen in UK diese Software zum Respeaking. Dabei handelt es sich um Red Bee Media (RBM), welche die Untertitelung für BBC vornimmt und zum anderen Independent Media Support (IMS). Allerdings wird in UK (und auch Spanien, wo ViaVoice beim Respeaking eingesetzt wird) ein Wechsel von ViaVoice auf DNS in Erwägung gezogen (vgl. [RF11, S: 22-28]).

Training einer Spracherkennung

Vor der Nutzung der Spracherkennung erstellt ein Respeaker oder eine Respeakerin im so genannten *Trainingsprozess* ein individuelles Profil. Dies erfolgt in der Regel durch das Vorlesen von Texten. Weiters ist es bei den meisten Systemen möglich, durch die Analyse von Dokumenten die Spracherkennung auf den Schreibstil anzupassen. Darüber hinaus kann das Profil laufend verbessert werden, wie durch das Trainieren von neuen Wörtern oder der Korrektur im Falle von Erkennungsfehlern (vgl. [RF11, S: 74-94]). Das Korrigieren von Fehlern zur Verbesserung der Erkennungsgenauigkeit kann bei DNS in der Version 11 durch eine spezielle Korrekturfunktion erfolgen. Mittels Vorschlagsliste kann aus alternativen Wörtern das beabsichtigte ('korrekte') Wort ausgewählt werden. In dem Fall 'lernt' die Spracherkennung die individuelle Aussprache des jeweiligen Wortes und erkennt es beim nächsten Mal idealerweise richtig. Weiters ist die Korrektur mittels Tastatur sowie durch Sprachbefehle möglich. Die Funktion kann genutzt werden, wenn z.B. das diktierte Wort nicht in der Vorschlagsliste enthalten ist. Mit den Korrekturmethoden können darüber hinaus neue Wörter trainiert werden. Die unterschiedlichen Korrekturfunktionen sind allerdings nur innerhalb von Dragon (im so genannten *DragonPad* sowie im *Diktierfenster*) bzw. in Programmen verfügbar, die von Dragon unterstützt werden. Dabei handelt es sich u.a. Microsoft Outlook und Microsoft Word (vgl. [Nua10, S: 67-88]). Wird jedoch in ein Fenster einer nicht unterstützen Software diktiert, können Erkennungsfehler nicht mit der Funktionen korrigiert und daher die Genauigkeit von Dragon nicht verbessert werden. So zum Beispiel bei den Programmen zur Untertitelerstellung, die im Abschnitt 2.5.4 ab Seite 38 erläutert sind. Dieses Schnittstellenproblem kommt somit beim Respeaking zum Tragen.

[31] Anm. Autor: Damals war DNS in der Version 10 verfügbar.

Benutzung einer Spracherkennung

Aus Sicht des Autors dieser Diplomarbeit ist es für Respeakerin bzw. Respeaker entscheidend zu wissen, wie die jeweilige Spracherkennung prinzipiell aufgebaut ist und funktioniert. Darüber hinaus sollen die Stärken einer Spracherkennung genutzt werden und ein Wissen über den Umgang mit Problemen vorhanden sein. Romero Fresco empfiehlt, die Spracherkennung als Partnerin zu sehen. So soll während der Trainingsphase erlernt werden, mit welcher Diktierweise und Aussprache die Spracherkennung gute Ergebnisse liefert. Aufgrund der Funktionsweise von Dragon werden die besten Erkennungsraten durch ein klares und natürliches, jedoch monotones Sprechen mit gleicher Lautstärke erzielt. Jedes Wort soll dabei deutlich betont werden und die Diktiergeschwindigkeit konstant sein. Darum sollen Respeakerinnen und Respeaker in einer konstanten Geschwindigkeit diktieren, selbst wenn ein Sprecher oder eine Sprecherin das Tempo wechselt. Romero Fresco vergleicht die Diktierweise dabei mit dem bei Nachrichtensendungen üblichen Sprechen (vgl. [RF11, S: 45-94]). Weiters stellen meist längere, zusammengesetzte sowie fachspezifische Wörter (wenn sie im Vokabular der Spracherkennung DNS vorhanden sind) weniger Probleme hinsichtlich der Erkennungsraten als kurze Worte dar. Dies begründet sich darin, dass im Wortschatz einer Spracherkennung weniger Wörter existieren die phonetisch dem Wort 'Leistungssportkarriere' ähnlich sind als beispielsweise dem Wort 'kennen' (vs. 'kenne', nenne', 'nennen', etc.).

2.5.2 Hardware, Software, Räumlichkeit

Rechner

Für die Anforderungen an den Rechner verweist der Autor dieser Diplomarbeit auf die jeweiligen Angaben seitens der Herstellerfirmen. Im Falle von Dragon NaturallySpeaking wird für die Version 11 einen 1,8 GHz Intel Dual Core oder gleichwertiger AMD-Prozessor (SSE2-Befehlssatz erforderlich) mit 2 MB Prozessor-Cache empfohlen. Abhängig vom verwendeten Betriebsystem werden zwischen 2 GB RAM (Windows XP und Windows Vista) und 4 GB (Windows 7 und Windows Server 2003/2008 64-Bit) empfohlen. Für die Version 12 wird eine leistungsstärkere CPU - ein Intel Pentium 2.4 GHz (Dual 1.8 GHz Core Prozessor) oder gleichwertiger AMD-Processor (SSE2-Befehlssatz erforderlich) - mit ebenfalls 2 MB Prozessor-Cache empfohlen. Die Anforderungen bezüglich des Arbeitsspeichers sind gleichgeblieben (vgl. [Nua12a], [Nua12b]). Da der letzte Release von ViaVoice im Jahr 2003 war, sind die Systemanforderungen dementsprechend geringer: Intel® Pentium® Prozessor mit 300 MHz und 256K L2 Cache oder gleichwertiger AMD-Prozessor sowie zwischen 64 und 192 MB RAM (je nach Betriebssystem - Windows XP ist das 'aktuellste' Betriebssystem für das ViaVoice entwickelt wurde) (vgl. [RF11, S: 65-66]).

Mikrofon

Die Auswahl eines geeigneten Mikrofons wird von den meisten Experten bzw. Expertinnen als essentiell bezeichnet, da es einen hohen Einfluss auf die Erkennungsraten einer Spracherkennungssoftware hat. Es gibt verschiedene Mikrofon Typen, u.a. Headsets, Desktop- und Handmikrofone. Mikrofone können mit dem Audioeingang (Audio-In), USB oder Bluetooth mit dem

Rechner verbunden werden. Für Respeaking in UK werden Desktop- sowie Handmikrofone verwendet, in den meisten anderen Ländern Headsets (vlg. [RF11, S: 74]). Beim ORF arbeiten die Respeaker und Respeakerinnen mit einem Headset. Bei der Firma Titelbild kommen Handmikrofone mit einem On-Eear Kopfhörer zum Einsatz, siehe Abbildung 2.4[32] auf Seite 37.
Die richtige Benutzung des jeweiligen Mikrofons ist darüber hinaus ein wesentliches Kriterium für gute Erkennungsraten. So ist neben der Positionierung (seitlich links oder rechts vs. vorm Mund) auch der optimale Abstand zum Mund vom jeweiligen Mikrofontype abhängig (vlg. [RF11, S: 74-75]). Neben den qualitativen Unterschieden in Bezug auf die Erkennungsraten sind bei der Auswahl eines Mikrofons auch der Aspekt Ergonomie sowie die jeweilige Respeaking Tätigkeit an sich zu beachten. So wäre beispielsweise ein Handmikrofon ungeeignet, wenn ein Respeaker oder eine Respeakerin Erkennungsfehler im Live-Einsatz mittels Tastatur korrigieren möchte.
Darüber hinaus ist bei der Auswahl des Mikrofons auch der Einsatzort zu beachten, da Mikrofone unterschiedlich empfindlich auf Umgebungsgeräusche reagieren (vgl. [RF11, S: 74-75]). Eine Liste von empfohlenen Mikrofonen[33] ist auf der DVD[34] zum Buch [RF11] enthalten. Weiters ist auf der DVD eine umfangreiche Analyse[35] von Knowbrainer[36] zu finden, die verschiedene Mikrofone mit DNS in der Version 11 getestet und bewertet haben.

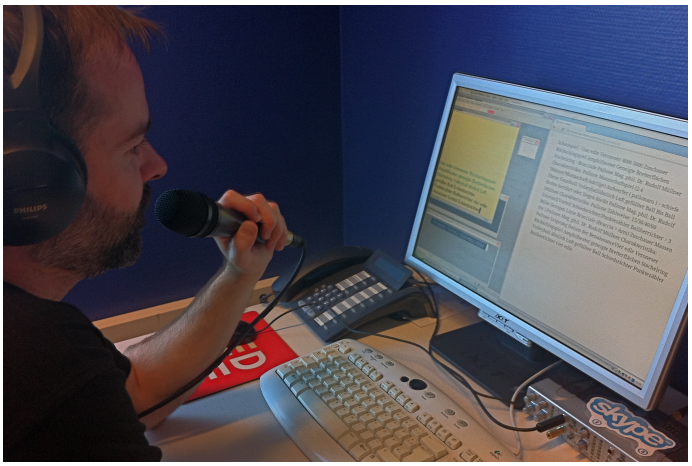


Abbildung 2.4: Respeaker der Firma TITELBILD Subtitling and Translation GmbH

[32] Quelle/Urheber: E-Mail von der Firma TITELBILD Subtitling and Translation GmbH [Hat11, S: 71]
[33] Anm. Autor: Empfohlene Mikrofone der Firma Nuance (Dragon NaturallySpeaking Herstellerfirma).
[34] [DVD]/Chapter 6/6.1.1/Recommended Microphones.pdf
[35] [DVD]/Chapter 6/6.6.2/Review of Dragon 11.pdf
[36] Anm. Autor: Knowbrainer vertreibt u.a. Dragon NaturallySpeaking und berät deren Kundinnen und Kunden bezüglich der richtigen Hardwareauswahl (vgl. [Kno12]), u.a. auch in einem Forum: `www.knowbrainer.com/forums/forum/index.cfm`, letzter Zugriff: 04.10.2012;

2.5.3 Räumlichkeit

Idealerweise arbeiten Respeaker und Respeakerinnen in schalldichten Kabinen[37] oder schalldichte Räumen, wobei es bei letzteren zwischen mehreren Respeakern bzw. Respeakerinnen zu gegenseitigen Störungen kommen kann (vgl. [RF11, S: 24]). Das betrifft einerseits die Konzentration, andererseits auch die im Abschnitt 2.5.2 angeführte unterschiedliche Empfindlichkeit der Mikrofone in Hinblick auf Umgebungsgeräusche.

2.5.4 Untertitelsoftware

Wie im Abschnitt 2.5.1 ab Seite 33 dargelegt, sollen Spracherkennungssysteme Benutzerinnen und Benutzern das Arbeiten ohne Tastatur ermöglichen und wurden nicht für das Respeaking und dessen Anforderungen entwickelt. So muss für die Erzeugung von Untertiteln - im Gegensatz zur Benutzung von automatischer Spracherkennung im Office Bereich - nicht nur das Gesprochene transkribiert, sondern das Transkript zusätzlich zu den Untertiteln verarbeitet werden. Ein wesentlicher Bestandteil der Untertitel sind die so genannten Zeitcodes (engl. *timecodes*). Sie ermöglichen das Synchronisieren der Untertiteleinblendung mit der Video/Audio-Spur. Solche Zeitcodes werden mit einer Untertitelsoftware erstellt, da sie weder von DNS noch ViaVoice erzeugt bzw. exportiert werden können.
Es existieren verschiedene Zeitcode Dateiformate, die alle zumindest eine eindeutige Nummer, die Start- sowie Endzeit der Einblendung und den Text des Untertitels beinhalten. Ein solches Format ist SubRip mit der Dateiendung „.srt“ (vgl. [AOD07, S: 95], [Hat11, S: 56]). Ein beispielhafter Auszug aus einer solchen Datei ist im Folgenden angeführt:

```
1
00:00:00,000 --> 00:00:05,000
Dies ist die Untertiteleinblendung für die ersten fünf Sekunden.

2
00:00:05,000 --> 00:00:09,000
Dies ist die Einblendung für die darauf folgenden 4 Sekunden.
```

Weiters wird von diversen Untertitelformaten eine optischen Aufarbeitung der Untertitel (z.B. verschiedene Farben zur Unterscheidung bei mehreren Sprechern bzw. Sprecherinnen) unterstützt. Bei der TV Untertitelung bzw. bei jener von Filmen für Kino und DVDs ist die Einblenddauer sowie die maximale Zeichenanzahl je Untertiteleinblendung ein wesentliches Kriterium.

Um das vom Respeaker bzw. der Respeakerin Diktierte (und von der ASR transkribierte) als Untertitel an die TV Geräte senden zu können, verwendet RBM in UK das von BBC entwickelte *K-Live*. Die Software erlaubt es den Respeakern bzw. den Respeakerinnen von zu Hause zu arbeiten. Bei IMS in UK sowie Flandern wird *Wincaps* verwendet (vgl. [RF11, S: 24-31], [Mar06]). Der ORF und das Schweizer Fernsehen (SWISS TXT) setzen dazu eine Software der

[37] Anm. Autor: Solche Kabinen kommen auch beim Konferenzdolmetschen zum Einsatz (Dolmetschkabinen).

Firma FAB[38] ein. Über eine spezielle Schnittstelle zur Spracherkennung können Respeaker bzw. Respeakerinnen direkt in die Software von FAB diktieren. Bevor die Untertitel von der Software via Teletext an die TV Zuseherinnen und Zuseher gesendet werden, können noch eventuelle Korrekturen (manuell durch die Tastatur) durchgeführt werden. Der Ablauf ist in Abbildung 2.5 auf Seite 40 schematisch dargestellt sowie in [Hat11, S: 73-77] dokumentiert. Ein Problem zwischen der Schnittstelle FAB und DNS ist allerdings, dass Dragon eventuelle Korrekturen von Erkennungsfehlern, die durch den Respeaker bzw. der Respeakerin in FAB am Transkript vorgenommen werden, nicht zur zukünftigen Vermeidung von Erkennungsfehlern verwendet (vgl. [Hat11, S: 73-77]).
Wie im Abschnitt 2.4.2 ab Seite 32 erläutert, widmet sich diese Diplomarbeit der intralingualen Untertitelerzeugung und Darstellung in E-Learning Plattformen. Im Speziellen handelt es sich um die im Abschnitt 3.1.1 ab Seite 52 erläuterten E-Learning Plattform *Synote*. Mittels einer SubRip Untertiteldatei können die (offline) erzeugten Untertitel synchron zur Aufzeichnung (Video/Audio) in Synote eingebunden und somit die in Abbildung 3.1 auf Seite 54 farbliche (gelbe) Hervorhebung im Transkript ermöglicht werden. Da es sich bei den erwähnten Systemen fürs Fernsehen um propretäre bzw. kostenpflichtige Systeme handelt und wie erläutert DNS keine Zeitcodes erzeugen/exportieren kann, wurde für die im Zuge dieser Diplomarbeit erarbeitete Ausbildung eine Open Source Software zum Erstellen der Untertitel verwendet, siehe Abschnitt 3.1.5 ab Seite 57.

2.5.5 Spezielle Hardware

Zusätzliche Hardwarekomponenten finden zum Teil beim Respeaking Einsatz. So wird in Italien beispielsweise ein Touch Screen verwendet, mit welchem die Respeakerinnen bzw. Respeaker Interpunktionszeichen setzen können und sie somit nicht diktieren müssen. Einem Touch Screen können darüber hinaus andere Funktionalitäten hinzugefügt werden, wie beispielsweise das Einfügen von themenspezifischen Wörtern. Romero Fresco hebt weiters hervor, dass ein Touch Screen generell das individuelle Anpassen von Funktionalitäten ermöglicht (vgl. [RF11, S: 7; 36-37]). Weiters werden in Kanada Joysticks verwendet, mit dem auch Soundinformationen[39] erstellt werden können.

[38] FAB Teletext & Subtitling Systems. Eine Firma, die div. Softwarelösungen zum Erzeugen und Senden von Untertiteln anbietet (vgl. [Hat11, S: 176]).

[39] Anm. Autor: Soundinformationen sind paralinguistischen Eigenschaften zuzuordnen. Paralinguistischen Informationen können „Emotionen wie Flüstern, Untertöne und Stimmlage [...] “ [Hat11, S: 51] vermitteln.

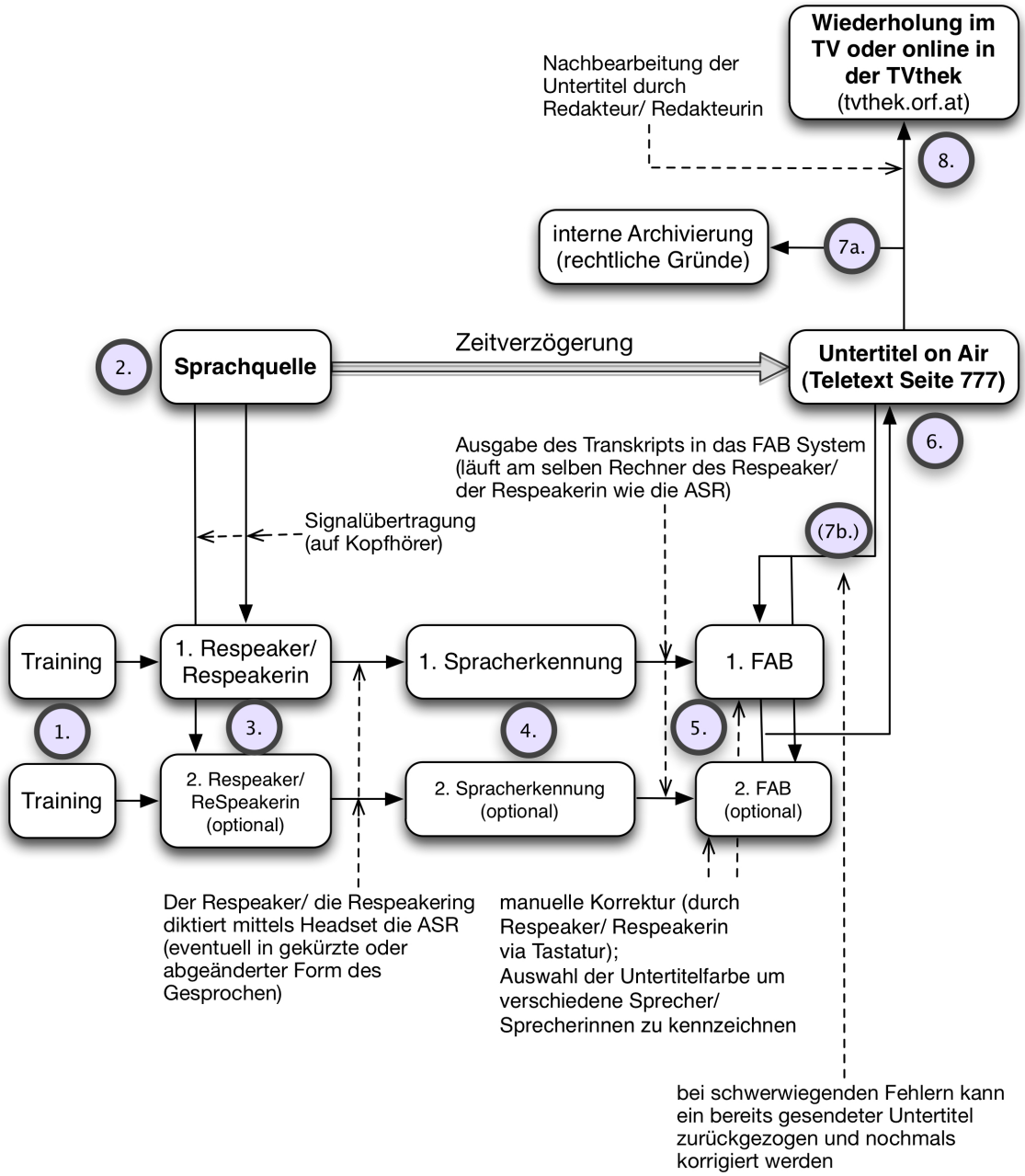


Abbildung 2.5: Ablaufdiagramm der Untertitelerstellung mittels Respeaking beim ORF [Hat11, S: 74]

2.6 Evaluierung der Qualität (WER, WRR, NERD sowie NER)

„Die Erkennungsgenauigkeit (Akkuratheit) bzw. Fehlerhäufigkeit einer ASR dient für Nutzerinnen und Nutzer oft als wichtiges Qualitätsmerkmal. Auch zum Messen des Fortschrittes gegenüber Vorgängerversionen eines ASR Systems sind die Werte wichtig“ [Hat11, S: 71]. Die Wortfehlerrate WER (engl. *word error rate*) sowie die Wortakkuratheit WRR (engl. *word recognition rate*) werden häufig in Zusammenhang mit der Qualität von Spracherkennungssoftwares bzw. zur Evaluierung der Untertitelqualität genannt bzw. als Messwert verwendet.
Die Berechnung der WER (Wortfehlerrate) ist in Formel 2.1 dargestellt, jene für die WRR (Wortakkuratheit) in 2.2.

$$WER = \frac{I + D + S}{N} (*100\%) \tag{2.1}$$

$$WRR = 1 - WER = \frac{N - I - D - S}{N} \tag{2.2}$$

In den Formeln repräsentiert der **N** Wert die Anzahl der gesprochenen Worte (engl. *total number of spoken words*). **S** steht für Ersetzungen (*S* für engl. *substitutions*), demnach für ein falsch erkanntes Wort. Der Wert **D** repräsentiert Auslassungen (*D* für engl. *deletions*), wenn also anstatt eines gesprochen Wortes nichts erkannt wurde. Schließlich steht der Wert **I** für Einfügungen (*I* für engl. *insertions*), also dem irrtümlichen 'Erkennen' von nicht gesprochenen Worten (vgl. [RF11, S: 151-152] [PK08, S: 292-293], [Eul06, S: 21-22], [O'S08], [Hat11, S: 63-65]). Kurzum stellt der WER Wert die Gesamtanzahl der Fehler, dividiert durch die gesprochenen Wörter bzw. die Wortanzahl dar. An der Stelle sei auf [Hat11, S: 63-65] verwiesen, wo WER bzw. WRR ausführlich diskutiert und erläutert ist. An genannter Stellte wurde auf verschiedene Schwächen dieser beiden häufig verwendeten Berechnungsmethoden hingewiesen, nicht zuletzt im Hinblick auf gekürzte bzw. umformulierte Untertitel, wie sie gerade beim Respeaking oft vorkommen. Daraus resultierend entwickelte der Autor dieser Diplomarbeit zur Evaluierung einer Respeaking Fallstudie in [Hat11, S: 116-124; 144-153] eine Methode zum Evaluieren der Qualität dieser Untertitel.
Wenig später veröffentlichte Romero Fresco in [RF11, S: 150-161] (im ersten und bisher einzigen Fachbuch zum Thema Respeaking) das NERD Modell, das in zum Teil ähnlicher, aber detaillierter Art den Problemen von WER bzw. WRR gerecht werden soll und über die in [Hat11] verwendeten Methodik hinausgeht. Im folgenden Abschnitt ist das NER-Modell (der Nachfolger des NERD Modells) erläutert.

2.6.1 Weiterentwicklung Erkennungsratenmessung (NERD und NER)

Untertitelungsfirmen stellen seit einiger Zeit den Auftraggebern (TV-Sendern) die Erkennungsraten der Live-Untertitel zur Verfügung. Die Berechnungen unterscheiden sich aber zwischen den verschiedenen Ländern stark, ja sogar innerhalb der Konzerne. Trotz der unterschiedlichen Berechnungsmethoden werden die Werte oft miteinander verglichen. Ausgehend von dieser Problematik entwickelte Romero Fresco in [RF11, S: 150-161] das NERD Modell, dass den Anforderungen für vergleichbare Erkennungsraten gerecht werden soll. Das Modell wurde anhand folgender Anforderungen entwickelt: es soll funktionell und einfach anzuwenden sein und auf den seit langen erprobten WRR Prinzipien beruhen. Zusätzlich soll es die Originalquelle (das Gesprochene) berücksichtigen, um Auslassungen, Änderungen, etc. identifizieren zu können. Darüber hinaus soll in Betracht gezogen werden, dass es sich bei den Untertiteln (wie beim Respeaking) auch um umformulierte sowie gekürzte handeln kann. Das Modell soll auf spezielle Eigenschaften von unterschiedlichen Sprachen eingehen können, dennoch für alle Sprachen anwendbar bleiben. Auch soll es die Live-Korrektur (durch den Respeaker oder die Respeakerin) berücksichtigen. Schließlich soll die Qualität im Modell nicht nur anhand der Erkennungsrate (Zahlenwert) beschrieben werden, sondern Auskunft über zukünftige Möglichkeiten zur Verbesserungen geben sowie idealerweise in welcher Weise die Verbesserung durchgeführt werden kann (vgl. [RF11, S:27; 150-161], [RFM14]). Das NER-Modell ist eine Weiterentwicklung des NERD Models und in der noch unveröffentlichten Publikation [RFM14][40] von Romero-Fresco vorgestellt. Es soll einerseits das zukünftige Modell zum Vergleich der Qualität der durch Respeaking erstellten Untertitel dienen, aber auch für automatisch erzeugte Untertitel durch eine ASR anwendbar sein (vgl. [RFM14]).
In Abbildung 2.3 ist die Berechnungsformel für NER-Werte angeführt.

$$NER = \frac{N - E - R}{N} * 100 \tag{2.3}$$

Die Qualität der Live-Untertitel wird als akzeptabel bezeichnet, sobald der NER-Wert die Genauigkeit von 98% erreicht (vgl. [RFM14]) bzw. übersteigt.

[40] Anm. Autor: Zum Zeitpunkt der Veröffentlichung dieser Diplomarbeit.

Im NER-Modell wird der errechnete Zahlenwert durch zwei schriftliche Teile (*Korrekte Editierung* sowie *Kommentar*) ergänzt (vgl. [RFM14]). Diese Ergänzungen werden ebenso wie die Parameter der Formel nachfolgend erläutert.

- **N: Gesamtzahl der erzeugten Wörter** (*N* für engl. **N**umber of words in the respoken text[41]): zur Gesamtanzahl der erzeugten Wörter zählen neben den transkribierten Wörtern auch Interpunktionszeichen, die Identifizierung der sprechenden Person[42], etc. (vgl. [RFM14]).

- **E: Editierfehler** (*E* für engl. **E**dition errors): dies sind Fehler, die üblicherweise aufgrund der gewählten Strategie des Respeakers bzw. der Respeakerin verursacht wurden. Beispiele sind das Weglassen von Informationen, das Hinzufügen von Information und (inhaltlich) falsch transkribierte Passagen in den Untertiteln. Wie in diesem Abschnitt erläutert, wird im NER-Modell je nachdem wie schwerwiegend ein Fehler ist, zwischen drei Fehlerarten unterschieden (vgl. [RFM14]).

- **R: Erkennungsfehler** (*R* für engl. **R**ecognition errors): Erkennungsfehler entstehen Aufgrund falscher Aussprache bzw. sind sie auch auf die verwendete Technologie[43] zurückzuführen. Wie auch Editierfehler werden Erkennungsfehler in drei Kategorien - je nachdem wie schwerwiegend ein Fehler ist - unterteilt. Dabei kann es sich um Ersetzungen, Auslassungen und Einfügungen handeln (vgl. [RFM14]). Demnach werden die Parameter **I**, **D** sowie **S** vom WRR bzw. WER Modell im NER-Modell unter Erkennungsfehler (**R**) zusammengefasst.

- **Korrekte Editierung** (engl. *Correct editions*): in textueller Form werden im NER-Modell jene Veränderungen/Editierungen erfasst, die vom Respeaker bzw. der Respeakerin durchgeführt wurden und zu keinem Informationsverlust im Vergleich zum Gesprochen bzw. der Originalquelle geführt haben. Im Fall einer **nahezu '1:1' Transkription**[44] ist beispielsweise das Auslassen von Redundanzen eine korrekte Editierung, solange die Kohärenz sowie Kohäsion bewahrt wird (vgl. [RFM14]).

- **Beurteilung** (engl. *Assessment*): Abschließend wird eine allgemeine Beurteilung und Analyse abgegeben sowie auf andere wichtige Qualitätskriterien eingegangen. Dabei sind Aspekte wie Sprechgeschwindigkeit, Verzögerung, die benötigte Zeit für Korrekturen oder die Kohärenz zwischen Untertitel und anderen optischen Bildquellen Teil der Beurteilung. Zusätzlich kann erfasst werden, wie die respeakende Person mit den Anforderungen umgegangen ist (vgl. [RFM14]).

Editierfehler sowie Erkennungsfehler werden wie erwähnt - abhängig von der Abweichung zum Gesprochenen - einer von drei Fehlerklassen zugeordnet. Die Fehlerkategorisierung nimmt

[41] Anm. Autor: Bzw. wie in [RFM14] angeführt auch durch andere Methoden der Untertitelerzeugung.
[42] Anm. Autor: Wie das Wechseln der Untertitelfarbe bei verschiedenen Sprecherinnen bzw. Sprechern oder das Einfügen vom Namen der sprechenden Person.
[43] Anm. Autor: Wenn beispielsweise die ASR ein diktiertes Wort falsch 'erkennt'.
[44] *(6) Editing policy: near-verbatim* [RF11, S: 11-17]

durch eine dreistufige Gewichtung Einfluss auf den NER-Wert. Somit haben gewisse Fehler einen höheren Einfluss auf den NER-Wert als andere:

- **Gravierender Fehler** (engl. *Serious errors*): Editier- bzw. Erkennungsfehler dieser Klasse haben den Faktor 1. Als Beispiel für einen gravierenden Erkennungsfehler nennt Romero Fresco „in Island ist der Zinssatz von 3,5 auf 2% gesunken“ anstatt „in Irland ist der Zinssatz von 3,5 auf 2% gesunken“. Ein gravierender Editierfehler wäre lt. Romero-Fresco „es wird nützen, dass die USA bekannt gemacht haben, dass sie ihre Regierungscomputer besser schützen lassen will“ anstatt „es wird wenig nützen, dass die USA bekannt gemacht haben, dass sie ihre Regierungscomputer besser schützen lassen will“. Als gravierende Fehler sind jene einzustufen, mit denen nicht nur wichtige Information in den Untertitel verloren gehen, sondern darüber hinaus falsche Information transportiert wird bzw. im Kontext falsch interpretiert werden könnte (vgl. [RFM14]). So können in den genannten Beispielen die Fehler von hörbeeinträchtigten Menschen nicht als solche identifiziert werden, da die Untertitel trotz des Fehlers einen Sinn ergeben, jedoch eine falsche Aussage transportieren.

- **Normaler Fehler** (engl. *Standard*): Fehler dieser Kategorie werden mit dem Faktor 0,5 multipliziert und wirken sich demnach weniger stark auf die Erkennungsrate aus. Als Beispiel für einen normalen Erkennungsfehler nennt Romero Fresco „sie haben Maschinenpistolen eisig getragen“ anstatt „sie haben Maschinenpistolen bei sich getragen“. Solche normale Erkennungsfehler machen es der Untertitel lesenden Person schwer, die original Botschaft zu rekonstruieren, selbst wenn der Fehler als solcher erkennbar ist. Ein normaler Editierfehler wäre lt. dem Autor „die Zinsen sind heute weiter gestiegen“ anstatt „die Zinsen für irische Staatsanleihen sind heute weiter gestiegen“. Normale Editierfehler verändern im Gegensatz zu gravierenden nicht die Bedeutung des Gesagten, stellen allerdings einen nicht feststellbaren Informationsverlust dar (vgl. [RFM14]).

- **Geringfügiger Fehler** (engl. *Minor*): Fehler dieser Klasse werden mit dem Faktor 0,25 gewichtet. Als Beispiele für geringfügige Erkennungsfehler nennt Romero Fresco „zweite“ anstatt „zweiten“ oder „bereit haben“ anstatt „bereits haben“. Fehler dieser Art werden unter Umständen beim Lesen nicht bewusst wahrgenommen. Sie erlauben das Folgen der Untertitel und eventuell sogar das Rekonstruieren des Gesprochenen. Weiters ist eine falsche Groß- und Kleinschreibung sowie das Fehlen von unbedeutenden Wörtern ist als ein geringfügiger Erkennungsfehler zu werten. Wie der Zugang zum Kürzen von Untertiteln ist, hängt einerseits vom Land, dem Unternehmen sowie vom Programm ab. Anhand der Kriterien erfolgt die Abgrenzung zwischen einer korrekter Editierung und einem geringfügigen Editierfehler. So kann beispielsweise „der frühere Notenbankchef hat versichert“ anstatt „der frühere Notenbankchef, Alan Greenspan, hat versichert“ bei einer geforderten nahezu ’1:1’ Transkription als geringer Editierfehler gewertet werden. Liegen die Anforderungen weniger bei einer ’1:1’ Transkription, so kann dies (abhängig vom Land, Unternehmen, Programm) auch eine korrekte Editierung darstellen (vgl. [RFM14]). Da aber einem Respeaker oder einer Respeakerin die Anforderungen in der Vorbereitungsphase (siehe Abschnitt 2.2.1 ab Seite 21) bekannt sein müssen/sollten und nach diesen eine spätere Beurteilung erfolgt, sind damit die NER-Werte vergleichbar.

Vergleicht man das WRR Modell mit dem NER-Modell, so sind drei Aspekte hervorzuheben: Zum ersten werden bei der NER-Formel die tatsächlich transkribierten Worte samt Interpunktionszeichen zur Berechnung herangezogen und nicht wie bei der WRR lediglich die tatsächlich gesprochenen. Zweitens wird durch die Gewichtung der Fehler beim NER-Modell berücksichtigt, dass nicht jeder Fehler gleichermaßen die Qualität der Untertitel beeinflusst. Und zum Dritten werden bei der NER-Analyse andere Aspekte in textlicher Form (Verzögerung, Art der Umformulierung, etc.) erfasst. Das ermöglicht nicht zuletzt ein gezieltes Training zur zukünftigen Vermeidung von Fehlern.

Die Analyse mit dem NER-Modell ist ein wichtiger Bestand der im Kapitel 3 erarbeiteten und dokumentierten Respeaking bzw. Scripting Ausbildung. Nach erfolgter Ausbildung transkribierte der ausgebildete Student eine Vorlesung. Weiters wurde die Vorlesung durch die Firma Titelbild live untertitelt sowie vom Autor dieser Diplomarbeit Untertitel mittels Scripting erzeugt. Die Qualität der jeweiligen Untertitel wurde mit dem NER-Modell analysiert und beurteilt, siehe Abschnitt 4.3.4 ab Seite 106. An erwähnter Stelle sind weitere Beispiele von *gravierenden*, *normalen* und *geringfügigen Editier-* und *Erkennungsfehlern* angeführt. Darüber hinaus ist auch dort die Abgrenzung zwischen *geringen Fehlern* und *korrekter Editierung* bezüglich der Untertitelung von universitären Vorlesungen diskutiert.

2.7 Ähnlichkeiten zu anderen Professionen

Die Anforderungen und Tätigkeiten beim Respeaking werden oft mit jenen des Simultandolmetschens und mit der traditionellen Untertitelerstellung verglichen (vgl. [RF11, S: 45-55]). In [Hat11, S: 70-73] sind basierend auf [RV06] und [ARRF08] Überlappungen zum Simultandolmetschen ebenso diskutiert wie jene Fähigkeiten und Tätigkeiten, die speziell beim Respeaking zum Tragen kommen. Romero Fresco hat zusammen mit Arumí Ribas in [ARRF08] und später als Autor von [RF11] die Anforderungen beim Respeaking erarbeitet. Dabei wurden Überschneidungen zum Simultandolmetschen und jene zur klassischen Untertitelerstellung erörtert. In den genannten Publikationen sind die Anforderungen beim Respeaking den jeweiligen Phasen (der Vorbereitung, Untertitelerstellung und Nachbereitung) zugeordnet. Darüber hinaus erfolgte eine Klassifizierung jener Anforderungen, die speziell das Respeaking betreffen und die keine Überlagerungen zur klassischen Untertitelung bzw. zum Simultandolmetschen aufweisen (vgl. [RF11, S: 45-55]). Im Folgenden wird ein Überblick in die Thematik gegeben und jene Aspekte hervorgehoben, die für das erarbeitete Training (siehe Kapitel 3) und somit den Fokus dieser Diplomarbeit relevant sind.

2.7.1 Audiovisuell Übersetzung: Untertitelung

Wie die Synchronisierung (engl. *dubbing*) und die Tonspurüberlagerung (engl. *Voice-Over*) stellt die Untertitelerzeugung eine Unterkategorie der Audiovisuellen Übersetzung dar (engl. *Audiovisual Translation* (AVT)) (vgl. [Dia09, S: 4-5], [Bak98, S: 245]). Verschiedenste Anwendungsgebiete innerhalb der Untertitelung sind im Abschnitt 2.1 ab Seite 16 angeführt, wobei abhängig vom Einsatzgebiet die Anforderungen unterschiedlich sind bzw. sein können.

Die Anforderungen beim Respeaking sind dabei jenen der vorbereiteten/offline *Hörgeschädigten-Untertitel* (HG-UT) am ähnlichsten. In beiden Disziplinen werden die Untertitel üblicherweise in der eigenen Sprache formuliert (intralinguale Untertitelung, bei der die Ausgangssprache die Untertitelsprache ist)[45] und aufgrund von Kenntnissen über die Bedürfnisse der Zielgruppe(n) dementsprechend gekürzt und umformuliert. Das setzt gute grammatikalische Kenntnisse voraus, speziell im Hinblick auf das Setzen von Interpunktionszeichen. Hinzu kommt häufig ein fachspezifisches Vokabular wie Namen, geografische Orte, etc. Ebenso müssen Respeaker und Respeakerinnen mit dem Transkribieren von sich abwechselnden - und zum Teil überlappenden - Sprechern bzw. Sprecherinnen umgehen. Auch das Arbeiten mit Technologien verbindet das Respeaking mit der Untertitelung. Wenn bei der Erstellung von HG-UT Skripte oder der Ausgangstext zur Verfügung steht, handelt es sich um eine Transformation von schriftlichen zu schriftlichen Informationen. Beim Respeaking findet konträr eine Transformation einer oralen zu einer schriftlichen Information statt. Respeaking kann zur Untertitelung somit das sein, was Dolmetschen zur Übersetzung ist: Der Sprung von der schriftlichen zur oralen Information ohne den Schutz der zeitlichen Entkoppelung (vgl. [RF11, S: 47-48]).
Der angesprochene Live Aspekt beim Respeaking hat überlappende Anforderungen zu den Untertitelungsmethoden, die in [Hat11, S: 77-80; 116-124] beschrieben sind und in Tabelle 2.1 auf Seite 17 verglichen sind. Beim Scripting hingegen - also der offline Untertitel Erstellung mittels Respeaking - ist eine zeitliche Entkoppelung gegeben.

2.7.2 Simultandolmetschen

Die wesentlichen Überlappungen zwischen dem Simultandolmetschen und Respeaking stellt das gleichzeitige *Zuhören*, *Sprechen* und das Achten auf die eigene *Aussprache/Stimme* dar. Dabei unterscheiden sich die Anforderungen der Aussprache allerdings insofern, als ein Respeaker oder eine Respeakerin zu einer Maschine spricht und damit die im Abschnitt 2.5.1 ab Seite 33 beschriebene Sprechweise (klar, natürlich, monoton, gleiche Lautstärke, etc.) entscheidend für gute Erkennungsraten ist. Beim Simultandolmetschen ist hingegen ein für das Zielpuplikum angenehmes Sprechen gefordert. Wie beim Simultandolmetschen ist beim Respeaking die thematische Vorbereitung, inklusive des fachspezifischen Vokabulars, ein Teil der Tätigkeit. In beiden Professionen wird die Tätigkeit unter Verwendung von Mikrofonen/Headsets durchgeführt. Weitere Gemeinsamkeiten sind das Arbeiten in Echtzeit[46] und die geringen Korrekturmöglichkeiten während der Tätigkeit. Auch die Arbeitszeiten (Wochenenden, etc.) und die Teamarbeit bei längeren live Einsätzen stellen Ähnlichkeiten beider Disziplinen dar. Aus den dargelegten Gründen gibt es lt. Romero Fresco beim Respeaking mehr Überschneidungen mit dem intralingualen Simultandolmetschen als mit der Untertitelung. Der größte Unterschied zum Simultandolmetschen liegt in der fehlenden Interlingualität, da wie beschrieben in den meisten Fällen beim Respeaking die Ausgangs- und Zielsprache identisch sind. In Bezug auf Multitasking ist Respeaking anspruchsvoller als das Simultandolmetschen. Das betrifft einerseits das *Lesen* der von der Spracherkennung erzeugten Ausgabe und andererseits die *Korrektur* etwai-

[45] Anm. Autor: Beim Respeaking findet in den meisten Fällen keine Übersetzung des Gesprochenen statt. Respeaking wurde weiters, wie in einigen Fällen in Flanders, bereits zur intralingualen Untertitelung eingesetzt (vgl. [RF11, S: 22-47]).

[46] Anm. Autor: Die offline Untertitelung mittels Scripting stellt eine Ausnahme dar.

ger Fehler. Wie bereits im Abschnitt 2.2.1 ab Seite 21 erläutert, stellt beim Respeaking nicht das Multitasking selbst die Schwierigkeit dar, sondern dass nicht alle Aufgaben (z.B. Diktieren und Korrigieren) gleichzeitig stattfinden können (vgl. [RF11, S: 45-94]), siehe Abbildung 2.3 auf Seite 23.

2.8 Respeaking Ausbildungen: Respeaking als Profession

2.8.1 Anforderung beim Respeaking/Auswahl der Kanditen u. Kandidatinnen

In [RF11, S. 22-44] ist erstmals dokumentiert, wie in den verschiedenen europäischen Ländern bzw. ebenfalls in den USA und Kanada die Bewerbungsprozesse ablaufen bzw. welche Vorbildung die dort arbeitenden Respeaker bzw. Respeakerinnen vorweisen sollen. Ein detaillierter Einblick in die genauen Prozesse (wie Fragebögen, Gewichtung der Schwerpunkte, etc.) wird aber nicht angeführt. Generell kann aus Sicht des Autors dieser Diplomarbeit jedoch gesagt werden, dass sich die Auswahlverfahren in den einzelnen Ländern bzw. in den verschiedenen Unternehmen zum Teil stark unterscheiden.
So werden in Spanien und Norwegen keine speziellen Qualifikationen im Sinne von Vorbildung/Studium von den Bewerbern und Bewerberinnen verlangt. In der Schweiz hingegen hatten die Respeaker und Respeakerinnen (anfangs) konträr großteils einen Abschluss in Translationswissenschaften mit vier Jahren Ausbildung in Übersetzung und ein gewisses Maß an Ausbildung im Dolmetschen[47]. Darüber hinaus wird von Respeakern und Respeakerinnen in der Schweiz erwartet, perfekt in Wort und Schrift in den Sprachen Deutsch, Französisch und Italienisch zu sein. Neben Multitasking Fähigkeiten sollen sie weiters vertraut im Umgang mit hoch technischen Ausstattungen zu sein. Es können sich jedoch Theorie und Praxis unterscheiden und die Anforderungen von den tatsächlichen Qualifikationen der arbeitenden Respeaker und Respeakerinnen abweichen, wie das Beispiel der Firma IMS in UK zeigt. Der tatsächliche Ausbildungshintergrund deckt sich in diesem Unternehmen nicht immer mit den Anforderungen, die üblicherweise ein Abschluss in Sprachen und idealerweise eine absolvierte Lehrveranstaltung zur Untertitelung im Masterstudium[48] vorsehen (vgl. [RF11, S. 22-44]). Beim ORF, wo seit Beginn 2010 ein Teil der Live-Untertitel mit Respeaking erzeugt werden, arbeiten derzeit Respeaker und Respeakerinnen mit unterschiedlichsten Vorwissen und ohne vorangegangene Ausbildung im akademischen Bereich. Fünf, der acht von Walter interviewten Respeaker und Respeakerinnen, arbeiteten zum Zeitpunkt der Respeaking Einführung bereits jahrelang in der Untertitelungsabteilung des Senders (vgl. [Wal12, S: 60-69], [Hat11, S: 73-77]).
Neben den Qualifikationen im Sinne von Vorbildung sind auch die Aufnahmeverfahren selbst in den einzelnen Ländern bzw. in den verschiedenen Unternehmen unterschiedlich. Oftmals findet Eingangs ein (telefonisches) Interview statt, gefolgt von einem oder mehreren Tests. Die Aufnahmetests unterscheiden sich weiters in deren Inhalt und Schwerpunkten. Würde man die verschiedenen Inhalte vereinen, so würden folgende Gebiete abgedeckt werden: *Linguistische Fä-*

[47] Anm. Autor.: „Initially, they were mostly translations graduates with four years training in translation and some training in interpreting“ [RF11, S. 31].
[48] Anm. Autor: „[...] ideally a postgraduate course in subtiling“ [RF11, S. 24] würde in Österreich beispielsweise eine Masterlehrveranstaltung in Untertitelung entsprechen.

higkeiten (in Bezug auf Wortschatz, Aussprache, Rechtschreibung, Grammatik, das Verständnis des Inhalts, die Fähigkeit des Umformulierens, etc.), eine hohe *Konzentrationsfähigkeit*, *Multitaskingfähigkeit*, *Technisches Wissen*, *Fachwissen* (in Bezug auf *Untertitelung*, Themen wie Sport oder Nachrichten und Allgemeinwissen), *Interesse an Aspekten* von TV *in Bezug auf hörbeeinträchtigte Menschen*, die *Motivation* und *Lernbereitschaft*, *Kommunikations-* und *Teamfähgikeit*, *Flexibilität* sowie ein *Improvisationstalent*. Auch erste Respeaking Übungen, jedoch ohne Spracherkennung aber bereits mit dem Diktieren von Interpunktionszeichen[49], können - wie bei RBM - Teil der Aufnahmeprozedur sein (vgl. [RF11, S. 22-44]).

2.8.2 Ausbildung im akademischen Bereich

Respeaking wurde erstmals 2001 in UK und Flanders zur Untertitelung eingesetzt (vgl. [RF11, S: 22]). Das Jahr 2006 stellt im Hinblick auf akademische Forschung einen Wendepunkt in der kurzen Respeaking Geschichte dar: An der *Universität von Bologna* fand das *First International Seminar on New Technologies in Real Time Intralingual Subtitling* statt, seither ist Respeaking in vielen akademischen Konferenzen und wissenschaftlichen Veröffentlichungen präsent (vgl. [RF12b]).
Der erste Respeaking Kurs an einer europäischen Universität fand 2007 in Antwerpen (Belgien) statt. Bereits 2008 folgte an der *Autonomen Universität Barcelona* (Spanien) ein Respeaking Kurs, wo 2009 der weltweit erste online Respeaking Kurs gestartet wurde. Die beiden Kurse in Spanien etablierte Romero Fresco, nachdem er als Co-Autor von [ARRF08] einen Ausbildungsvorschlag zum Respeaker bzw. zur Respeakerin erarbeitete. Nach den Kursen in Spanien führte er ebenfalls 2009 an der *Roehampton Universität* (UK) einen Respeaking Kurs ein (vgl. [RF12a], [RF11, S: 40-42]) und publizierte mit [RF11] und [RF12b] weitere Werke bezüglich akademischer Respeakingausbildungen. Die Intensität der Kurse bzw. Module reicht von 4 ECTS (On Campus Modul „Subtitling for the deaf and Hard of Hearing“ in Barcelona), 10 ECTS (E-Learning Modul „European MA in Audiovisual Translation“, ebenfalls in Barcelona) bis zum 20 ECTS Modul mit dem Namen „Media Accessibility“ in Roehampton (vgl. [RF12b]).
In Österreich gibt es derzeit keine Ausbildung zum Respeaker oder zur Respeakerin an öffentlichen Bildungseinrichtungen (vgl. [Wal12, S: 76]). Es sei erwähnt, dass in den letzten Jahren auch in Österreich erste wissenschaftliche Arbeiten veröffentlicht wurden, die sich entweder direkt ([Wal12] und [Kel07]) oder teilweise ([Nem13], [Hat11] sowie [Now10]) dem Thema Respeaking widmen.

2.8.3 Ausbildung im nicht akademischen Bereich

Begründet durch den erläuterten geschichtlichen Vorsprung der Industrie beim Respeaking wird und wurde die Ausbildung meist in den jeweiligen Unternehmen (TV Sender, Untertitelungsfirmen) selbst durchgeführt (vgl. [RF12b]). Wie die Auswahlverfahren unterscheiden sich die Ausbildungen und Trainings in den einzelnen Ländern bzw. in den verschiedenen Unternehmen.

[49] Anm. Autor: Es kann von Shadowing Übungen gesprochen werden: „Unter 'Shadowing' versteht man das Anhören und gleichzeitige Nachsprechen von akustisch dargebotenem Material in der Ausgangssprache - ein 'Mitsprechen' gewissermaßen. Es ist eine häufig verwendete Methode zur Untersuchung der selektiven Aufmerksamkeit in der kognitiven Psychologie“ [Kur96, S: 102], siehe Abschnitt 4.2.3 ab Seite 90.

In Flanders zum Beispiel folgt nach dem erwähnten dreiwöchigen Aufnahmeverfahren ein dreimonatiges Training. In England kann das Training zwischen drei Wochen und drei Monaten dauern, in der Schweiz zwischen eineinhalb und zwei Monaten. Die Respeaker und Respeakerinnen in Dänemark haben durchschnittlich 75 Stunden Training. In Spanien dauert die Ausbildungszeit zwischen drei und sechs Monaten, wobei es bis zu einem Jahr dauern kann, bis die Arbeit als voll effektiv angesehen wird (vgl. [RF11, S. 22-44]). Beim ORF gibt es keinen standardisierten Trainingsprozess. So unterschieden sich die Art und der Umfang des Trainings bei den von Walter interviewten Respeakerinnen und Respeakern. Einige von ihnen, die zum Zeitpunkt der Respeaking Einführung bereits jahrelang in der Untertitelungsabteilung des Senders arbeiteten, wurden in zwei bis drei Tagen in einer Einschulung von Kollegen aus der Schweiz die Technologie erläutert. Währenddessen wurden auch Übungseinheiten zu den jeweiligen Sendungen durchgeführt. Andere Respeaker und Respeakerinnen beim ORF nahmen an achtstündigen (internen) Einschulungen teil. Gefestigt wurde das kurze Training direkt live auf Sendung (vgl. [Wal12, S: 68-69], [Hat11, S: 73-77]).
Aufgrund der im vorigen Abschnitt beschriebenen Module an europäischen Universitäten arbeiten bereits akademisch ausgebildete Respeakerinnen und Respeaker in den jeweiligen Ländern (vgl. [RF12b]).

2.8.4 Bedeutung für die erarbeitete Respeaking/Scripting Ausbildung

Anhand der im Abschnitt 2.2.1 ab Seite 21 erläuterten Vielseitigkeit und Komplexität der Respeaking Tätigkeiten ist es naheliegend, dass Respeaking (schrittweise) erlernt werden muss. Theoretisches Wissen, praktische Übungen und die laufende Weiterbildung sollen aus Sicht des Autors dieser Diplomarbeit eine hohe Qualität der Untertitel sicherstellen. Weiters soll dadurch erreicht werden, dass die Tätigkeit(en) für die Respeaker und Respeakerinnen selbst eine geringe psychische und physische Belastung darstellen. In diesem Kapitel wurde erörtert, welche Fähigkeiten beim Respeaking besonders wichtig sind. Weiters wurde dargelegt, wie unterschiedlich die Ausbildungen und das Vorwissen von den derzeit arbeitenden Respeakerinnen und Respeakern ist.
Für den Autor dieser Diplomarbeit stellt sich die Frage, ob bestimmen Menschen - abhängig von ihrer Vorbildung, ihres Berufes, ihrer Talente - die Respeaking Fähigkeiten besser bzw. leichter erlernt werden können sowie gewisse Personen bessere Grundvorraussetzungen zum Respeaking haben als andere. Speziell betrifft dies Vorwissen aus dem im Abschnitt 2.7 ab Seite 45 erläuterten Simultandolmetschen sowie der (Hörgeschädigten) Untertitelung.
In [RF12b] wurden Fragebögen von 30 Studierenden ausgewertet, welche das im Abschnitt 2.8.2 erwähnte *On Campus* Modul „Subtitling for the deaf and Hard of Hearing“ in Barcelona absolvierten. Alle Teilnehmerinnen und Teilnehmer der Studie gaben an, bereits Training/Erfahrung im Bereich der Untertitelung (HG-UT) zu haben. Jene Studierende, die darüber hinaus Erfahrung mit dem Dolmetschen hatten, schnitten beim Kurs durchschnittlich mit besseren Noten ab. Innerhalb dieser Gruppen hatten darüber hinaus jene, die ihre Dolmetschfähigkeiten als *gut* oder *sehr gut* bezeichneten, die besseren Noten. Der Studienautor hebt die geringe Anzahl der Studierenden, die für die Auswertung herangezogen werden konnten, hervor. Es scheint jedoch, dass Studierende mit Dolmetschhintergrund weniger Probleme mit dem Multitasking hatten. Anderseits fanden sie das Diktieren der Interpunktionszeichen sowie das Achten auf die Aus-

sprache (in Bezug auf den Umgang mit der Spracherkennung) schwierig. Aus dem Grund hält der Studienautor fest, dass Studierende die Respeaking im Zuge einer Simultandolmetschausbildung erlernen, neben Grundkenntnisse der Audiovisuellen Übersetzung vor allem einige Dolmetschfähigkeiten 'ablegen' bzw. 'verlernen' müssen. So zum Beispiel das Sprechen in einer angenehmen und freundlichen Weise, da diese Sprechweise nicht den effizientesten Umgang mit der Spracherkennung darstellt. In Bezug auf die benötigten Fähigkeiten aus dem Bereich der Untertitelung hatten die Befragten einen großen Konsens darüber, dass für das Respeaking die Fähigkeiten des Umformulierens, des Kürzens, die Verwendung und des Setzens von Interpunktionszeichen sowie der Untertitelformatierung am wichtigsten sind. Für das Training von Studierenden, die Respeaking im Zuge einer Ausbildung zur Audiovisuellen Übersetzung erlernen, empfiehlt Romero-Fresco eine generell Einführung in die Untertitelung sowie speziell in die HG-UT. Hinzu kommt dann im Zuge des Respeaking Moduls das Erlernen der Fähigkeiten aus dem Simultandolmetschbereich (vgl. [RF12b]).
Aufgrund der beschriebenen Überlappungen der Anforderungen stellte sich bezüglich dem Training von (zukünftigen) Respeakerinnen und Respeakern für den Autor dieser Diplomarbeit die Frage, welche Teile und Übungen der Simultandolmetsch- sowie der (Hörgeschädigten) Untertitelungsausbildung (ggf. in adaptierter Form) sich gut für das Erlernen des Respeakings eignen. Das im folgenden Kapitel erarbeitete und dokumentierte Respeaking/Scripting Training basiert schließlich auf Übungen und theoretischen Inhalten aus dem Simultandolmetschen, der Untertitelung sowie spezifischen Respeaking Inhalten. Bei den (zukünftigen) Auszubildenden ist dabei kein spezielles Vorwissen im Sinne einer Dolmetsch- oder Untertitelausbildung nötig.

KAPITEL 3

Ausarbeitung der Respeaking/Scripting Ausbildung

3.1 Verwendetes Equipment/Räumlichkeit

3.1.1 Synote: Untertitelung im E-Learning

Wie im Abschnitt 1.2.4 ab Seite 5 bereits erläutert, bietet die E-Learning Plattform Synote aus Sicht des Autors dieser Diplomarbeit die nötige Funktionalität, um den Bedürfnissen - vor allem in Bezug auf die Untertitelung - von hörbeeinträchtigten Studierenden gerecht zu werden. Dabei stellt die Einbindung eines Transkripts des Gesprochenen ein Kernelement von Synote dar. Die Darstellung des Transkripts/der Untertitelung ist bei der Wiedergabe der Aufzeichnung ständig synchron[1] mit der Audio- oder Videoaufzeichnung sowie ebenfalls ständig synchron mit anderen Materialien, die in Synote eingebunden werden können (PowerPoint Vortragsfolien, Notizen, etc.). Darüber hinaus fallen für Bildungseinrichtungen keine Lizenzkosten bei der Verwendung von Synote an (vgl. [Syn12], [WWM+09], [Hat11, S: 133-137]). Auch wenn in E-Learning Plattformen wie TUWEL Untertiteleinblendungen bei Videoaufzeichnungen eingebunden werden können, so sind sie im Gegensatz zu Synote „nicht optimal für Untertiteleinblendungen bzw. deren einfache Einbindung konzipiert“ [Hat11, S. 133-137]. Einige Merkmale der E-Learning Plattform Synote sind im Folgenden hervorgehoben (vgl. [Wal10b], [Wal10a], [Hat11, S: 133-137]):

- **Die in der Größe individuell anpassbaren vier Synote Bereiche** (engl. *panels*). Die vier Bereiche sind in Abbildung 3.1a auf Seite 54 (einem Screenshot von einer in Synote eingebundenen Aufzeichnung) zu sehen:

  - Linker oberer Bereich: Im Multimedia Bereich kann die Wiedergabe einer Audio- oder Videoaufnahme (Play, Stop, Pause, Lautstärke, etc.) gesteuert werden. Im Falle einer Videoaufzeichnung ist in diesem Bereich das Video des Vortragenden, der Gebärdensprachdolmetscherin bzw. des Gebärdensprachdolmetschers, etc. zu sehen.
  - Rechter oberer Bereich: Die als *Synmarks* bezeichneten Lesezeichen/Notizen können von Benutzerinnen und Benutzern der E-Learning Platform (also den Studierenden) selbst erstellt und auf Wunsch mit anderen Synote Benutzern bzw. Benutzerinnen geteilt werden. Sie enthalten neben der zeitlichen Zuordnung durch *Start-* und *Endzeit* auch einen *Titel*, *Notizen*, *Tags*, etc. In einer Notiz kann darüber hinaus HTML eingebunden werden und somit können Synmarks externe Webinhalte beinhalten. Weiters können Studierende durch das Feature selbst Anmerkungen zum Inhalt erstellen und mit Kolleginnen und Kollegen sowie Lehrenden den Inhalt diskutieren und ergänzen. Mittels Twitter[2] können darüber hinaus bereits während der Aufzeichnung Synmarks erstellt werden (z.B. wichtige, unklare, schwierige, etc. Stellen), die später mit der Aufzeichnung synchronisiert werden können. In Abbildung 3.1b auf Seite 54 ist das Erstellen einer Synmark zu sehen.

[1] Anm. Autor: Screenshots eines in Synote eingebundenen Vortrags und dessen synchrone Elemente sind in Abbildung 3.1a zu sehen: Die aktuelle Stelle im Transkript ist in der E-Learning Platform mit gelbem Hintergrund hervorgehoben.

[2] Anm. Autor: Twitter ist ein soziales Netzwerk mit dem kurze Nachrichten im Internet verbreitet werden können; `http://twitter.com`, letzter Zugriff: 01.07.2012.

  - Linker unterer Bereich: Der Bereich beinhaltet (wenn vorhanden) das gesamte Transkript der Video- oder Audioaufzeichnung. Auf Wunsch ist durch automatisches Scrollen das Transkript synchron mit der Wiedergabe. Während der Wiedergabe kann im Transkript gescrollt und somit beispielsweise der Inhalt wie ein Buch durchgeblättert werden, um dann an einer bestimmten Stelle die Wiedergabe fortzusetzen. Die aktuelle Wiedergabestelle ist optional gelb dargestellt. In Abbildung 3.1c auf Seite 54 ist zu sehen, wie ein Transkript manuell erstellt bzw. bearbeitet werden kann.
  - Rechter unterer Bereich: Die Vortragsfolien, bei welcher die aktuelle Folie gelb umrandet ist. Beim Klick auf eine Folie wird die Wiedergabe (Audio- bzw. Video, das Transkript sowie die Synmarks) an der Stelle der Folie fortgesetzt.

- **Die schnelle Navigation an eine gewünschte Stelle der Aufzeichnung** ist durch die Verwendung der Browsersuche gewährleistet. Damit können die Vortragsfolien, der Synmarks und das gesamte Transkript durchsucht werden. Damit ist es dem Benutzer oder der Benutzerin möglich, schnell alle Vorkommnisse eines gewissen Wortes bzw. einer Wortfolge zu finden und anschließend an der gewünschten Stelle die Wiedergabe fortsetzen. Zum Navigieren kann weiters direkt auf Synmarks oder Folien geklickt werden.

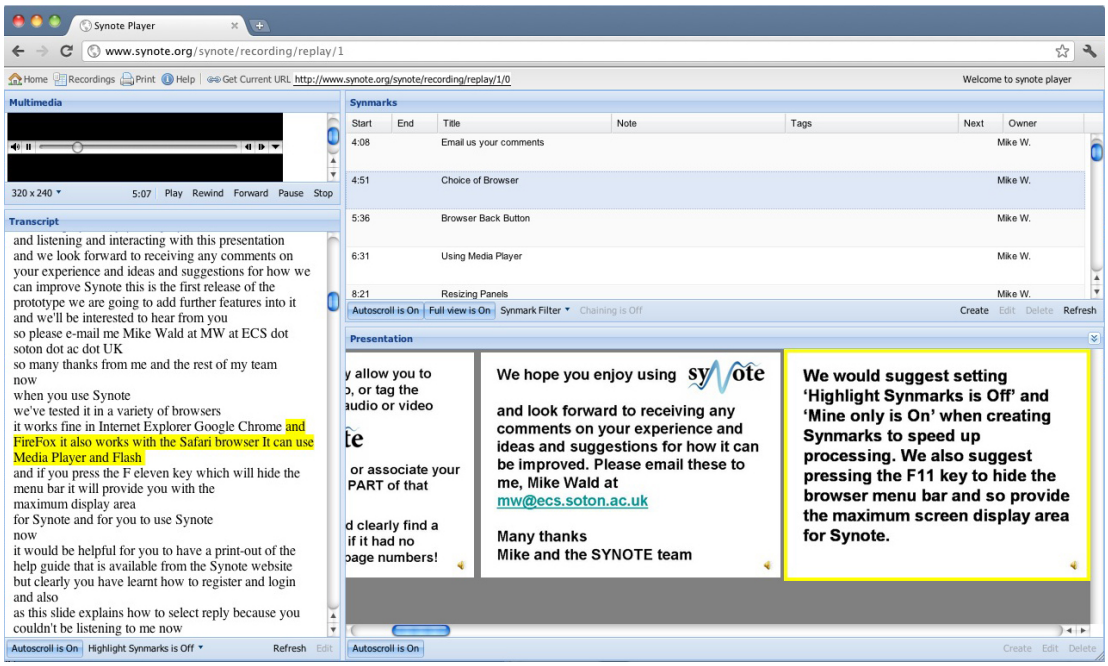
- **Die Druck- und Speicherfunktion** ermöglicht das individuelle Speichern und Drucken der Inhalte (Ausnahme: Video und Audio) und somit das Erstellen von offline Lernunterlagen.

Die Funktionalitäten von Synote können nicht nur für hörbeeinträchtigte bzw. speziell für schwerhörige Studierende sehr nützlich sein, sondern auch für fremdsprachige Menschen. Durch die Konservierung von Lehrveranstaltungen, deren effiziente Durchsuchbarkeit und der Möglichkeit interaktiv den Lehrinhalt zu ergänzen bzw. zu diskutieren, stellen in Synote aufbereitete Lehrveranstaltungen für *alle* Studierenden und Lehrende einen Mehrwert dar. Um einen ausführlicheren Überblick über Synote und seine Funktionen zu bekommen, sei an dieser Stelle auf [Wal10b] [Hat11, S: 133-137] verwiesen. Die Kosten bzw. der Aufwand für die Erstellung von Untertiteln und der dabei entstehende Mehrwert ist in [Hat11, S. 50] diskutiert.

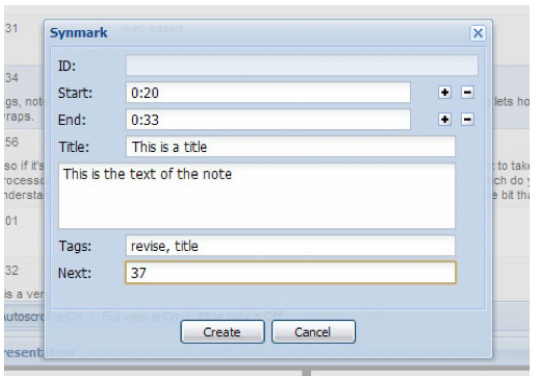
3.1.2 Spracherkennungssoftware: Dragon NaturallySpeaking Premium (DNS)

Wie im Abschnitt 2.5.1 ab Seite 33 erläutert, nimmt die Spracherkennungssoftware neben dem Respeaker bzw. der Respeakerin die wichtigste Rolle beim Respeaking ein (vgl. [Hat11, S: 71], [ARRF08]). Aufgrund der zentralen Rolle der Spracherkennung beim Respeaking beeinflusst diese viele Aspekte der erarbeiteten Ausbildung. Um nicht während des Trainings (aufgrund von möglicher auftretender Probleme) einen Softwarewechsel durchführen zu müssen, war es für den Autor dieser Diplomarbeit wesentlich, bereits vor dem Beginn der Ausbildung eine geeignete Wahl zu treffen. Dragon NaturallySpeaking Premium[3], Version 11.50.100.039, Deutsch wurde schließlich als Spracherkennungssoftware verwendet.

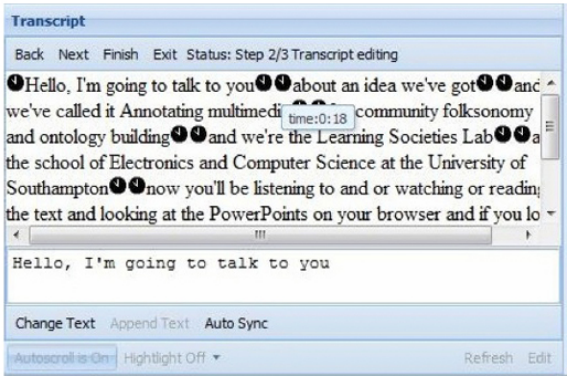
[3] Anm. Autor: Dragon NaturallySpeaking Premium hieß früher Dragon NaturallySpeaking Preferred (vgl. [Orc10])



(a) Synote annotation system, Screenshot aus: Synote Guide [Wal10b])



(b) Erstellen von Synmarks (Figure 11. [LWK+09])



(c) Bearbeiten des Transkripts (Figure 11. [LWK+09])

Abbildung 3.1: Synote

Dragon in der Version 11 stellte (bzw. stellt) aus Sicht des Autors dieser Diplomarbeit aus folgenden Gründen die geeignetste Spracherkennungssoftware für die erarbeitete Ausbildung dar:

- **Fehlende Alternativen**: Die in Europa am meisten verbreiteten Spracherkennungssysteme beim Respeaking sind Dragon NaturallySpeaking (DNS) und ViaVoice (vgl. [RF11, S: 22-36], [Hat11, S: 67-67]). ViaVoice wird jedoch seit 2003 nicht mehr weiter entwickelt und ist nicht mehr kommerziell erwerbbar. Weiters ist ViaScribe (eine Weiterentwicklung von von ViaVoice) für Deutsch technisch nicht praktikabel, siehe Abschnitt 2.5.1 ab Seite 33 (vgl. [Net10], [Weg10]). Anzunehmenderweise ist dies ein (Mit)Grund für die im nächsten Aufzählungspunkt erläuterte große Verbreitung von DNS im deutschsprachigen Raum.

- **Große Verbreitung von DNS**: Ein wesentlicher Grund für die Wahl von DNS ist dessen große Verbreitung beim Respeaking, vor allem im deutschsprachigen Raum (ORF, Schweizer Fernsehen, Titelbild), siehe Abschnitt 2.5.1 ab Seite 33. Weiters kommt beim Respeaking Kurs an der britischen Universität von Roehampton auch die Deutsche Version von DNS zum Einsatz (vgl. [RF12a]).

- **Möglicher Erfahrungsaustausch**: Aufgrund der großen Verbreitung von DNS konnte der Autor dieser Diplomarbeit während der Ausarbeitung des Trainings auf die bereits gesammelten Erfahrungen von Respeakern und Respeakerinnen, u.a. des ORFs, zurückgreifen. Der Kontakt zum ORF entstand bei einem Lokalaugenschein im Jahre 2010. Vor Ort konnte sich der Autor dieser Diplomarbeit mit Respeakern über die Funktionen, Möglichkeiten und aktuelle Probleme mit Dragon austauschen, siehe Abschnitt 3.2.3 in [Hat11, S: 73-77]. Weiters entstand ein reger E-Mail Kontakt mit Romero Fresco, dessen Erfahrungen als Respeaker, Ausbildner und Autor im Bereich des Respeakings sehr hilfreich bei der Ausarbeitung waren, siehe Abschnitt 2.8.2 ab Seite 48.

- **Mögliche Berufschancen**: Der ORF beschäftigte 2012 ca. 20 Respeaker und Respeakerinnen (vlg. [NH12]) und ist aktuell das einzige Unternehmen in Österreich, bei dem Respeaker und Respeakerinnen arbeiten. Die Ausbildung soll u.a. zukünftig (in gleicher oder adaptierter Weise) zur Ausbildung von Respeakern bzw. Respeakerinnen an österreichischen Bildungsstätten verwendet werden können. Das war ein Grund für die Wahl von DNS. Somit können im Falle einer solchen Ausbildung die zukünftigen Respeaker und Respeakerinnen ihr Wissen und ihre erlernten Fähigkeiten beruflich anwenden.

- **Version 11 von DNS**: Zum Zeitpunkt der Erarbeitung und Durchführung der Respeaking/Scripting Ausbildung wurde von der Firma Titelbild (vgl. [Woj12]) und beim ORF Dragon in der Version 10 verwendet.
  Das Schweizer Fernsehen stellt allerdings bereits auf die aktuelle Version 11 um (vgl. [Mar12]). Es mag zwar für Unternehmen sinnvoll sein, nicht (sofort) auf neue Versionen umzusteigen (Umschulung von Mitarbeitern und Mitarbeiterinnen, unvorhersehbare Probleme mit neuer Software, etc.). Trotzdem war es für den Autor dieser Diplomarbeit klar, für die Ausbildung die aktuelle Version zu verwenden. Einerseits ist sie bereits im deutschsprachigem Raum erprobt und andererseits können so in Österreich Erfahrungen

mit der Version 11 gesammelt werden. Seit Sommer 2012 ist Dragon in der Version 12 verfügbar.

- **DNS wird in Österreich bereits zum Respeaking verwendet**: Wie erwähnt wird beim ORF die Version 10 von DNS verwendet. Es war trotzdem im Vorfeld anzunehmen, dass auch die nachfolgende Version von Dragon prinzipiell zum Respeaking Einsatz mit österreichischen Akzenten verwendet werden kann.

- **Kaufpreis**: Der Kaufpreis der aktuellen Version 12 von Dragon NaturallySpeaking Premium beläuft sich auf 149 Euro (inkl. Mehrwertsteuer) (vgl. [Nua13]). Gerade im Hinblick auf ein mögliches Training an Universitäten ist aus Sicht des Autors dieser Diplomarbeit der Preis im Rahmen der finanziellen Möglichkeiten.

3.1.3 Hardware: Rechner, Mikrofon und Headset

Um die Einheiten, Übungen bzw. das Respeaking unter geeigneten sowie vergleichbaren Bedingungen durchführen zu können, wurde vom „Zentrum für Angewandte Assistierende Technologien“ (ATT) ein handelsüblicher Rechner[4] (mit Windows 7[5], Bildschirm, Tastatur und Maus) zur Verfügung gestellt. Der Rechner erfüllt die im Abschnitt 2.5.1 ab Seite 33 angeführten Hardwareanforderungen von DNS 11.
Wie am Abschnitt 2.5.2 ab Seite 36 erläutert, wird die Auswahl eines geeigneten Mikrofons von den meisten Expertinnen bzw. Experten als essentiell bezeichnet (vlg. [RF11, S: 74]). Nuance - die Herstellerfirma von Dragon - gibt an mit der Version 11 um 15% weniger Erkennungsfehler zu haben als mit der Vorgängerversion. Bei dem im Abschnitt 2.5.2 angeführten umfangreichen Review von Knowbrainer[6] konnte die Steigerung der Erkennungsraten mit dem von Nuance mitgelieferten Mikrofon nicht verifiziert werden. Jedoch konnten unter der Verwendung von High-End Mikrofonen bis zu 30% weniger Erkennungsfehler gemessen werden. So wird von Knowbrainer u.a. das *Samson Airline 77* empfohlen. Darüber hinaus wird darauf hingewiesen, dass bei DNS 11 im Gegensatz zu Vorgängerversionen die Verwendung von qualitativ hochwertigen Mikrofonen eine signifikante Erhöhung der Erkennungsraten ermöglichen (vgl. [Orc10]). Aus qualitativen Gründen entschied sich daher der Autor dieser Diplomarbeit, das Mikrofone *Airline 77* von *Samson*[7] (anstatt dem mitgelieferten Headset) zu verwenden. Die Spracherkennungssoftware DNS sowie das genannte Mikrofone wurde vom GESTU Pilotprojekt zur Verfügung gestellt. Das Funkmikrofon von Samson ist aber kein Headset und hat demnach nur die Funktion eines Mikrofons. Daher müssen beim Respeaking zusätzlich Kopfhörer verwendet werden. Am Beginn der Ausbildung kamen die (privaten) In-Ear Kopfhörer vom Auszubildenden zum Einsatz. Anschließend wurde mit einem Over-Ear-Kopfhörer (getragen nur auf einem Ohr) gearbeitet, siehe Abbildung 3.2 auf Seite 57.

[4] Intel(R) Core(TM)2 CPU 6600 @ 2,4GHz 2400MHz, 2 Kerne 2 logische Prozessoren, 4GB Arbeitsspeicher (2,82GB verfügbarer realer Speicher, 6,60GB verfügbarer virtueller Speicher)
[5] Microsoft Windows 7 Enterprise (Version 6.1.7600 Build 7600)
[6] Auf der DVD zum Buch [RF11]: [DVD]/Chapter 6/6.6.2/Review of Dragon 11.pdf
[7] Anm. Autor: Das Mikrofone kostet ca. 300 USD (inkl. Mehrwertsteuer) (vgl. [Kno13]).

Abbildung 3.2: Der Auszubildende beim Respeaken

3.1.4 Räumlichkeit

Als Räumlichkeit diente ein Besprechungsraum vom GESTU Pilotprojekt an der TU Wien[8]. Dort wurden zum einen die im Abschnitt 3.3 ab Seite 60 beschriebenen sieben Einheiten abgehalten. Bei den (Respeaking) Übungen während der Einheiten war lediglich der Autor dieser Diplomarbeit im Raum anwesend (um die Übung zu beobachten bzw. anschließend mit der auszubildenden Person zu analysieren). Abgesehen von Besprechungsterminen konnte das Besprechungszimmer jederzeit zum Absolvieren der eigenständig durchzuführenden Übungen genutzt werden. Auch wenn es sich bei dem Besprechungszimmer um keinen schalldichten Raum handelte, ermöglicht die geräuscharme Umgebung ein konzentriertes Arbeiten. Daher wurde auf die Installation einer Dolmetschkabine verzichtet.

3.1.5 Untertitelsoftwares

Wie im Abschnitt 2.5.4 ab Seite 38 beschrieben, kann Dragon ohne zusätzliche Software keine Zeitcodes erzeugen bzw. exportieren. Solche Zeitcodes sind jedoch für das Einbinden von Untertiteln in Synote nötig[9]. Da die von den Fernsehanstalten verwendeten Systeme propretär bzw.

[8] 1040 Wien, Favoritenstrasse 9/Stiege 3/1. Stock

[9] Anm. Autor: In Synote können auch Transkripte ohne Zeitcodes eingebunden werden. Es werden die Zeitcodes dann von Synote interpoliert, damit die farbliche Hervorhebung synchron zu der Video- oder Audioauzeichnung stattfindet, siehe Abbildung 3.1a auf Seite 54. Hat eine Sprecherin oder ein Sprecher eine sehr konstantes Sprechtempo und über die gesamte Aufnahme kaum oder gleichmäßig verteilte Pausen, kann eine Interpolation zu (mehr oder weniger) synchronen Untertitel führen. Der Autor dieser Diplomarbeit machte allerdings die Erfahrung, dass Vorträge an Universitäten oft starken Tempowechseln unterliegen bzw. weiters Sprechpausen üblich sind.

kostenpflichtig sind[10], wurden für die Ausarbeitung der Ausbildung im Zuge dieser Diplomarbeit einige Open Source bzw. gratis nutzbare Alternativen untersucht. Die Anforderungen an eine solche Open Source Software überlappen sich zwar aus Sicht des Autors dieser Diplomarbeit in vielen Aspekten mit den Anforderungen an Softwares die von Rundfunkanstalten verwendetet werden[11], sind aber dennoch geringer. Anders als im TV (bzw. generell bei Untertiteln, die direkt im Bild eingeblendet werden) ist bei Synote der Multimedia Bereich[12] von den Untertiteln bzw. dem Transkript im linken unteren Bereich getrennt. Dadurch gibt es in Synote keine Anforderungen an die minimale bzw. maximale Einblenddauer und folglich keine maximale Zeichenanzahl je Untertiteleinblendung. Es ist zwar anzumerken, dass sich kurze Untertitelblöcke (beispielsweise ein bis drei Sätze) aus Sicht des Autors dieser Diplomarbeit auch in Synote gut auf die Navigation, Synchronität, Durchsuchbarkeit und letzten Endes auf den Lerneffekt auswirken. Anzunehmenderweise sind die negativen Auswirkungen von langen Blöcken für die hörbeeinträchtigte Person jedoch bei weitem nicht so groß wie bei Untertiteln im Fernsehn. So wird bei Synote die aktuelle Stelle im Transkript gelb dargestellt (ggf. auch mehrere Sätze), siehe Abbildung 3.1a auf Seite 54. Im TV würden lange Untertitelblöcke jedoch nur kurz oder unter Umständen gar nicht vollständig angezeigt werden. Weiters ist im Vergleich zum TV die optische Aufarbeitung (z.B. verschiedene Farben zur Unterscheidung bei mehreren Sprechern bzw. Sprecherinnen) für das Untertiteln von Vorlesungen im E-Learning eher sekundär und keine primäre Anforderung an die Untertitelungssoftware.
Das wesentliche Kriterium für eine Untertitelungssoftware zur Erstellung von offline Untertiteln stellt aus Sicht des Autors dieser Diplomarbeit eine einfache Handhabung dar, damit der Respeaker bzw. die Respeakerin während der Arbeit möglichst wenig Konzentration auf die Erstellung der Zeitcodes aufbringen muss. Ideal wäre es demnach, wenn die Untertitelsoftware automatisch das Diktierte mit Zeitcodes versieht, beispielsweise in definierbaren Zeitabständen (z.B. alle fünf Sekunden) oder nach einer gewissen Wort- oder Zeichenanzahl. Darüber hinaus sollte eine Untertitelsoftware den Export in das SubRip Format unterstützen. Eine ohne Lizenzkosten benutzbare Software, die all diese Kriterien erfüllt, konnte nicht gefunden werden. Schließlich wurde die Open Source Software *Subtitle Workshop* von *UruSoft*[13] verwendet. Einzig die automatische Erstellung von Zeitcodes - abhängig von definierbaren Wort- oder Zeichenanzahlen - ist in Subtitle Workshop nicht möglich, alle anderen erwähnten Kriterien werden von der Software jedoch erfüllt. So kann ein Respeaker oder eine Respeakerin die Spracherkennung dazu benutzen, um direkt in Subtitle Workshop zu diktieren und manuell mittels Tastenkombination die Intervalle der der Untertitel während der Respeaking Tätigkeit festlegen. So ist es nach kurzem Training möglich, die Länge der Untertiteldauer während des Respeakings auf ca. ein bis drei Sätze festzulegen.
Wie auch bei FAB und DNS besteht bei Subtitle Workshop das Problem, dass die Korrektur der Erkennungsfehler nicht zur Verbesserung der Erkennungsgenauigkeit führt.

[10] Anm. Autor: Und somit nicht für die erarbeitete Ausbildung verfügbar waren bzw. darüber hinaus das finanzielle Budget bei weitem überschritten hätte.
[11] Anm. Autor: Wie K-Live, Wincaps oder der vom ORF verwendeten Software der Firma FAB, siehe Abschnitt 2.5.4 ab Seite 38.
[12] Anm. Autor: Linker oberer Bereich, in dem die Video- bzw. auch Audioaufzeichnung eingebunden werden, siehe Abschnitt 3.1.1 ab Seite 52.
[13] `www.urusoft.net/products.php?cat=sw&lang=1`, letzter Zugriff 22.10.2012

3.2 Die Auszubildenden

Wie im Abschnitt 2.7 ab Seite 45 erläutert, werden die Anforderungen und Tätigkeiten beim Respeaking oft mit denen des Simultandolmetschens sowie mit der traditionellen Untertitelerstellung verglichen. Derzeit sind an europäischen Universitäten nicht zuletzt aus diesem Grund Respeakingkurse bei Instituten der Translationswissenschaften[14] angesiedelt. In der Praxis arbeiten allerdings Respeakerinnen und Respeaker mit unterschiedlichen Ausbildungen und beruflichen Vorkenntnissen, siehe Abschnitt 2.8.1 ab Seite 47. Wie erläutert unterscheiden sich darüber hinaus die Auswahlverfahren von Respeakern und Respeakerinnen bei den jeweiligen Unternehmen teilweise stark ([RF11, S: 22-55], [RF12b], [Wal12, S: 60-69], [Hat11, S: 73-77]). Daher ist es für den Autor dieser Diplomarbeit nicht ausschlaggebend, dass der oder die (zukünftige) Auszubildende ein Vorwissen oder Studium im Bereich der Translationswissenschaften besitzt. Vielmehr ist die Intention und das Ziel des Trainings, die praktischen Fähigkeiten und das theoretische Wissen ohne spezielle Vorkenntnisse und Erfahrungen (wie aus den Bereichen Untertitelung sowie Simultandolmetschen) zu erlernen.
Dem Autor dieser Diplomarbeit wurden vom GESTU Projekt die räumlichen sowie finanzielle Ressourcen zur Ausbildung von *einer* Respeakerin oder *einem* Respeaker zur Verfügung gestellt. Auf der Suche nach einer geeigneten und motivierten Person entstand der Kontakt zum Auszubildenden.

3.2.1 Christian Hattinger (R1)

Der Autor dieser Diplomarbeit erarbeitete das in diesem Kapitel dokumentierte Respeaking bzw. Scripting Training und leitetet als Ausbildner die sieben Einheiten. Die entworfenen praktischen Übungen wurden im Eigenversuch erprobt und darauf folgend (nach teilweiser Adaptierung) mit dem Auszubildenden (R2) durchführt. Daher sammelte der Autor dieser Diplomarbeit selbst Erfahrung im Bereich des Respeakings und führte wie auch R2 die abschließende Respeaking Übung durch. Die Ergebnisse sind im Kapitel 4 dokumentiert.

3.2.2 Auszubildender (R2)

Der Auszubildende war und studierte zum Zeitpunkt der Ausbildung (Sommersemester 2012) im sechsten Semester das Bachelorstudien „Software & Information Engineering“ an der Technischen Universität Wien. Neben Deutsch als Muttersprache maturierte er in Englisch als erste, in Französisch als zweite Fremdsprache. Seine Deutschkenntnisse in Bezug auf Rechtschreibung schätzt er selbst als sehr gut ein. Vor Beginn der Ausbildung hatte er keine Erfahrungen mit Untertitelung bzw. demnach nicht mit Respeaking. Er arbeitete bereits als Tutor[15] im GESTU

[14] Anm. Autor: In Studiengängen der Audiovisuellen Übersetzung (engl. *Audiovisual Translation* (AVT)) oder dem Simultandolmetschen.

[15] Anm. Autor: „Die im GESTU Projekt als *Tutorinnen* und *Tutoren* angestellten Mitschreibhilfen [...] begleiten meist schwerhörige Studierende zu Lehrveranstaltungen und verfassen [...] Mitschriften bzw. ergänzen Skripten. Weiters vermitteln sie die Inhalte in vereinbarten Treffen. Diese ausführlichen, elektronischen Mitschriften und deren Ergänzungen sowie Hinweise gehen dabei über den reinen LVA-Inhalt hinaus. Auch setzen sie hörbeeinträchtigte Studierende über informelle Informationen von Mitstudierenden (Informationen über Prüfungsgewohnheiten von Professoren und Professorinnen, aktuelle Veranstaltungshinweise, etc.) in Kenntnis“ [Hat11, S: 47].

Projekt, wodurch der Kontakt zum Autor dieser Diplomarbeit entstand. In einem ersten Treffen, bei dem der Autor dieser Diplomarbeit die Ziele und den geplanten Ablauf der Respeaking Ausbildung erläuterte, bekundete der Auszubildende starkes Interesse an der Ausbildung. Im Interview nach Ende der Ausbildung gab er an, dass sich seine Motivation zum einen aus dem Interesse an der Respeaking Tätigkeit, zum anderen in der Chance auf eine zusätzliche Ausbildung begründete.

3.3 Erarbeitete und durchgeführte Ausbildung

Das erarbeitete Respeaking bzw. Scripting Training besteht aus sieben Einheiten zu je 90 Minuten. Das Ziel des Trainings ist es einerseits, der auszubildenden Person theoretisches Wissen im Themengebiet der Untertitelung zu vermitteln. Andererseits soll in praktischen Übungen schrittweise das Respeaking (die Vorbereitung, die Untertitelerstellung und die Nachbereitung) erlernt werden. Nach absolvierter Ausbildung soll es möglich sein, eine ausgewählte Vorlesung (siehe Kapitel 4 ab Seite 85) mit einem NER-Wert von mindestens 98% (nahezu '1:1') zu transkribieren und die Untertitel für eine Vorlesung in Synote aufbereiten zu können. Dabei sollte der Gesamtaufwand für die Ausbildung die für Übungen an Universitäten oft üblichen 3 ECTS (entspricht 75 Arbeitsstunden)[16] nicht überschreiten.
Die Einheiten wurden im Einzeltraining abgehalten, sind allerdings so gestaltet, dass sie künftig eine gleichzeitige Ausbildung von mehreren Personen zu Respeakern bzw. zu Respeakerinnen ermöglichen sollen. Darüber hinaus soll die Ausbildung zukünftig in ähnlicher Weise von anderen Personen durchgeführt und geleitet werden können. Wie im Abschnitt 1.3 ab Seite 6 erläutert, soll aus den gewonnen Erfahrungen damit der Grundstein für eine akademische Ausbildung zum Respeaker bzw. zur Respeakerin an österreichischen Universitäten gelegt werden.
Sämtliche praktische Übungen wurden vorab vom Autor dieser Diplomarbeit im Eigenversuch erprobt und (nach teilweiser Adaptierung) mit dem Auszubildenden durchgeführt sowie dokumentiert. Die Dauer der Ausbildung betrug ca. drei Monate, siehe Abschnitt 4.2.9 ab Seite 100. In diesem Zeitraum leitete der Autor dieser Diplomarbeit als Ausbildner die sieben Einheiten. Neben den praktischen Übungen zum Umgang mit der Spracherkennung sowie dem Erlernen des Respeakings beinhalten die Einheiten auch jeweils theoretisches Wissen. Beispielsweise über das Thema Hörbeeinträchtigung, die verschiedene Aspekte der Untertitelung, der Funktionsweise von Spracherkennungssoftwares, etc. Damit soll den zukünftigen Respeaker(n) bzw. Respeakerin(nen) eine Einordnung ihrer Tätigkeit im breiten Kontext ermöglicht und die Grundlage für tiefergehende Spezialisierungen geschaffen werden. Darüber hinaus sollen die erwähnten theoretischen Blöcke die eineinhalbstündigen Einheiten abwechslungsreicher gestalten, da die praktischen Übungen stimmlich und geistig belastend und somit ermüdend sein können. Zwischen den Einheiten sind klar definierte **Übungsaufgaben** - ähnlich wie begleitende Übungen bei universitären Vorlesungen - von der auszubildenden Person eigenständig durchzuführen und zu dokumentieren. Damit soll das Wissen jeder Einheit gefestigt, vertieft sowie das Erlernte geübt und dokumentiert werden. Die dadurch erzielten Fortschritte sind dann jeweils zusammen mit der auszubildenden Person in der Folgeeinheit zu evaluieren. Der Stundenaufwand außerhalb der Einheiten wurde durch eine Liste samt Angaben der genauen Tätigkeit dokumentiert,

[16] Anm. Autor: An der TU Wien entspricht 1 ECTS-Punkt 25 Arbeitsstunden (vgl. [Pou03]).

siehe Abschnitt 4.2.9 ab Seite 100. Die Einheiten sowie die Übungsaufgaben wurden alle mit dem im Abschnitt 3.1 beschriebenen Equipment durchgeführt.

Die Dokumentation der einzelnen Einheiten ist an jene von Romero-Fresco angelehnt, die er in [RF12b] für die Beschreibung einer Einheit des Respeaking-Moduls *online European MA in Audiovisual Translation* (METAV) der *Universitat Autònoma de Barcelona* wählte. Die Einheitsbeschreibungen haben einerseits den Zweck, eine Hilfestellung für die auszubildende Person zu geben, um Inhalte eigenständig wiederholen zu können sowie weiterführende Literatur zu finden. Darüber hinaus dienen die Inhaltsangaben und die Literaturverweise dazu, dass zukünftige Respeaking Ausbildnerinnen und Ausbildner den Inhalt jeder Einheit (nach erfolgtem Lesen/Studium und Praxis mit Respeaking) lehren können. Damit könnten die Ergebnisse dieser Diplomarbeit mit zukünftigen Arbeiten verglichen sowie Einheiten angepasst, überarbeitet und ggf. verbessert werden.
Jede Einheit trägt dabei einen **Titel** (die Kapitelüberschrift), der neben der Nummerierung die Grundinhalte der Einheit in wenigen Worten beschreibt. Unter dem Punkt **Ziel** sind in wenigen Sätzen die Inhalte angeführt, welche der oder die Auszubildende nach dem Absolvieren der Einheit bzw. dem nachfolgenden Einzeltraining erreicht haben soll. Unter **Matrialien/Quellen** sind sämtliche verwendete Unterlagen zu finden. Für theoretisch vermitteltes Wissen sind dabei die Materialien angegeben, welche dem Autor dieser Diplomarbeit als Grundlage für die jeweilige (verbal geführte) Einheit dienten bzw. zukünftigen Ausbildnerinnen und Ausbildnern ein Abhalten in gleicher bzw. ähnlicher Form ermöglichen sollen. Darüber hinaus sind jene Quellen angegeben, die einerseits als Grundlage für den inhaltlichen Aufbau der jeweiligen Einheit dienten bzw. auch jene, die Ideen für die erarbeiteten Übungen lieferten. Sämtliche Informationen der bewusst kompakt gehaltenen Einheitsbeschreibungen (Definitionen, Zahlen, etc.) sind den dort angeführten Quellen entnommen. Bei Verweisen auf Abschnitte dieser Diplomarbeit sowie auf [Hat11], sind die Originalquellen die dort (wissenschaftlich) zitierten Unterlagen. Ähnlich wie bei universitären Vortragsfolien, sind bei den Einheitsbeschreibungen die Quellenangaben dann aus Übersichtsgründen nicht als wissenschaftliche Zitate geführt. Nach den Angaben der Materialien beinhaltet jede Einheitsbeschreibung unter dem Punkt **Inhaltsübersicht/Ablauf** in detaillierterer Form und in chronologischer Reihenfolge die speziell hervorgehobenen und gelehrten Inhalte. Abschließend ist unter **Übungsaufgabe bis zur nächsten Einheit** festgehalten, welche Tätigkeiten von der auszubildenden Person bis zur Folgeeinheit selbstständig zu absolvieren sind.
Der Abschnitt 4.2 ab Seite 86 beinhaltet schließlich eine Diskussion sowie die Evaluierung des erarbeiteten und durchgeführten Trainingsprozesses. So sind an erwähnter Stelle weitere Informationen zu den Einheiten zu finden, wie beispielsweise theoretische Hintergründe zu Übungen und das Feedback der ausgebildeten Person.

3.3.1 Einheit 1 - Einführung in Hörbeeinträchtigung, Untertitelung, Respeaking sowie Dragon

Ziel: Einen Überblick über Hörbeeinträchtigung und die dabei wichtigsten Begriffe zu erlangen. Weiters die Arten der Untertitelung und die Grundbegriffe beim Respeaking zu kennen und einen Einblick in die allgemeinen Ziele der Ausbildung zu bekommen. Schließlich soll die Einführung in Dragon samt ersten Diktierübungen die Einheit abrunden und die Basis für die Übungsaufgabe bilden.

Matrialien/Quellen: In der Respeaking Ausbildung in UK wird anfangs den zukünftigen Respeakern bzw. Respeakerinnen eine Einführung in die Untertitelung (Methoden zur Erstellung, Bedürfnisse des Zielpuplikums, Methoden zur Erstellung, etc.) gegeben (vgl. [RF12b]). Ausgehend davon ist folgendes Material die Grundlage des theoretischen Teils der Einheit: Für die erläuterten Begriffe und Definitionen in Bezug auf Hörbeeinträchtigung dient der Abschnitt 2.1 (Behinderung) aus [Hat11, S. 9-22]. Der Abschnitt 2.3 aus [RF11, S. 11-17] und der Abschnitt 3.2.1 [Hat11, S.47-59] sind die Literatur zum Thema Untertitelung. Das vermittelte Ziel der Respeaking Ausbildung entstammt dem Abschnitt 1.3 sowie 3.3 dieser Diplomarbeit.
Wie in der Ausbildung in Spanien wird in der ersten Einheit Dragon NaturallySpeaking (DNS) vorgestellt und das Sprachprofil erstellt (vgl. [RF12b]). Wie in [RF11, S: 81-84] vorgeschlagen, wird nach dem Initialtraining ein Text mit nicht all zu speziellem Vokabular zuerst diktiert und anschließend korrigiert, sowie die auszubildende Person mit den wichtigsten Dragon Kommandos vertraut gemacht. Der zu diktierende/abzulesende Text entstammt dem deutschsprachigen Wikipedia Artikel über Spracherkennung[17], wobei lediglich der einleitende Absatz sowie die ersten beiden Absätze im Abschnitt *Geschichtliche Entwicklung*, mit insgesamt 214 Wörtern (exklusive IZ), für die Diktierübung herangezogen werden.
Als Referenz zu den vorgestellten Funktionalitäten von Dragon und dessen Bedienung dienen das Kapitel 3 der DNS-Bedienungsanleitung ([Nua10, S. 33-48]) sowie das Kapitel 6 aus [RF11, S: 74-93]. Der 1024 Wörter umfassende Text für die Diktierübung als Vorbereitung zur zweiten Einheit ist der Abschnitt *Geschichte und Definition* aus dem Kapitel 3.2.3 Respeaking ([Hat11, S. 70-73]), demnach bis zur Überschrift *Lokalaugenschein beim ORF*. Dabei sind das englischsprachige Zitat, wissenschaftliche Literaturhinweise, Fußnoten sowie Bildbeschreibungen nicht zu diktieren.
Die Idee zum Sammeln von Eindrücken durch das Ansehen von verschiedenen untertitelten Programmen stammt aus [RF12b][18].

[17] Herausgeber: Wikipedia, Die freie Enzyklopädie; `http://de.wikipedia.org/w/index.php?title=Spracherkennung&oldid=101421780`, Datum der letzten Bearbeitung: 29.03.2012, 09:22 UTC; Versions-ID der Seite: 101421780

[18] Anm. Autor: Beim E-Learning Modul „European MA in Audiovisual Translation“ in Barcelona werden die Eindrücke von den Studierenden darüber hinaus in einem Wiki gesammelt (vgl. [RF12b]).

Inhaltsübersicht/Ablauf:

- Erklärung folgender Termini im Zusammenhang mit Hörbeeinträchtigung: *Lautsprache*, *Gebärdensprache*, *ÖGS*, *gehörlos*, *schwerhörig*, *Erstsprache* (L1), *Muttersprache*, *bevorzugte Sprache* sowie die früher übliche Bezeichnung und heute als diskriminierend geltende Bezeichnung *taubstumm*. Hervorhebung der heterogenen Bedürfnisse von hörbeeinträchtigten Menschen in Bezug auf Hilfsmittel mit speziellem Fokus auf die Untertitelung.
- Einführung in die Untertitelung:
  - Die Untertitelerzeugung wird meist als eine Unterkategorie der *audiovisuellen Übersetzung* (engl. *Audiovisual Translation* (AVT)) bezeichnet. Die Synchronisierung (engl. *dubbing*) sowie die Tonspurüberlagerung (engl. *Voice-Over*) sind weitere Kategorien der AVT.
  - Geschichte der Untertitelung: BBC erstellte bereits 1972 Untertitel für Teile ihrer TV-Ausstrahlungen (via Teletext). 1990 wurde in Großbritannien gesetzlich festgelegt, dass mit 2010 90% des TV-Programms verpflichtend zu untertiteln sind. Beim ORF werden Untertitel seit 1980 mit laufender Erhöhung (1985 waren es 2,75%, Ende 2011 55% des gesamten Programms) eingesetzt.
  - Arten der Untertitelung:
    * Zeitliche Ebene bei Erzeugung: *Vorbereitete Untertitelung* (*offline/pre-recorded subtitles*), *Live-* oder *Echtzeit Untertitelung* (*online Untertitel*) sowie *Semi-Live-Untertitelung*
    * Sprachliche Aspekte: *Intralinguale* sowie *interlinguale Untertitelung*
    * Verschiedene Möglichkeiten zur manuellen Untertitelerzeugung: *Schnellschreiben* (mit QWERTZ bzw. QWERTY-Tastaturen), *(Computer-) Stenografie*, *Veyboard* (*Velotype*) und *Respeaking*
    * Detailgrad und Formulierung der Untertitelung: *Original mit Untertitel* (OmU) und *Hörgeschädigten-Untertitel* (HG-UT). Bei HG-UT werden paralinguistischen Eigenschaften (Flüstern, Untertöne, Stimmlage, etc.) hinzugefügt (daher menschliches Zutun nötig).
    * '1:1' Transkription vs. Kürzungen: Letzteres ist im Fernsehen oft üblich, um auch dem visuellen Inhalt folgen zu können. Kürzungen müssen nicht automatisch einen Informationsverlust darstellen und können die Lesbarkeit erhöhen. Keinen Informationsverlust gibt es allerdings nur bei '1:1' Transkription, was zusätzlich das Folgen des Lippenbildes ermöglicht. Beim Fernsehen liegt die Anzahl der Wörter bei Untertiteln meist bei 140-150 WpM, auf DVD's ca. bei 180 WpM. Aufgrund der Grenzen der Spracherkennungssysteme sowie der Respeakerin bzw. des Respeakers sind im TV meist Umformulierungen und Kürzungen üblich. Auch bei den innerhalb von GESTU live erstellten Untertitel von Vorlesungen wurde der Inhalt umformuliert und gekürzt.

    - Technische Aspekte: Erläuterung von *offenen* und *geschlossenen/optionalen* Untertiteln sowie verschiedener Darstellungsformen (*Blockdarstellung*, *fließend/ 'scrollend'*, etc.). *Zeitcodes* ermöglichen u.a. das Einbinden von Untertiteln in E-Learning Plattformen.

- Einführung und Einrichtung von Dragon: Das Equipment wurde vor der Einheit von der ausbildenden Person vorbereitet (Dragon installiert, Mikrofon angeschlossen, keine anderen Anwendungen wie Virenscanner gestartet/laufend, etc.). Es wird folgendes erläutert bzw. durchgeführt:
  - Richtiges Positionieren und Verwendung des Mikrofons
  - Erstellen eines Benutzerprofils sowie abschließen des initialen Trainings von Dragon (jedoch keine Datensammlung durchführen)
  - Dragon starten und die wichtigsten Merkmale der grafischen Oberfläche erläutern: *Lautstärken Pegel*, *Mikrofon Symbol*, *Hilfe-Funktionalität*
  - Interpunktionszeichen *Punkt* und *Beistrich* und erste kurze Diktierversuche
  - Sprachkommandos *neue Zeile* sowie *neuer Absatz*
  - Die Sprachbefehle wie: *was kann ich sagen?*, *geh schlafen* und *wach auf* sowie das Mikrofon ein- und ausschalten (mit Maus- und Tastatur)
  - *Diktiermodus* vs. *Standardmodus*: Diktiermodus wird ab nun zum Respeaking verwendet
  - *DragonPad*, *Diktierfenster* und *Erkennungsfenster* sowie die gesamte Dragon-Leiste (z.B. die Wiedergabe eines Wortes)

- Erste Diktierübung (freier Text): Diktierversuche mit frei wählbarem Inhalt (z.B. ein fiktiver Brief, ein fiktives Rezept, etc.) unter Verwendung der vorgestellten Interpunktionszeichen. Anschließend - und nicht während des Diktierens - die Korrektur von etwaigen Fehlern mittels der Korrekturfunktion von Dragon (Training der Software).

- Ziel der Scripting Ausbildung: Die Aneignung von theoretischem Wissen zu Untertitelung, Spracherkennung, dem Respeaking, etc. und mit praktischen Übungen die Respeaking Technik erlernen. Abschließend soll die auszubildende Person eine ausgewählte Vorlesung transkribieren. Die gewonnenen Erfahrungen sollen als Grundstein für eine akademische Ausbildung zum Respeaker bzw. zur Respeakerin an österreichischen Universitäten dienen.

- Erweiterte Nutzung von Dragon. Es wird folgendes erläutert bzw. durchgeführt:
  - Weitere Interpunktionszeichen: *Rufzeichen*, *Doppelpunkt*, *Fragezeichen*, *Bindestrich*, *Klammer auf* und *Klammer zu*, *einfaches Anführungszeichen auf*, *einfaches Anführungszeichen zu*
  - Einführung in das Korrigieren: Wiedergabe des Gesagten (*Auswahl wiedergeben*) um Fehlerquelle zu lokalisieren (ASR oder Aussprache) und Verwendung des *Korrekturmenüs*. Ggf. über *korrigier-das* korrigieren und trainieren

  - Einführung in das Vokabular Training: Die verschiedenen Möglichkeiten des vorab Trainings und die DNS Kernfunktionalitäten *Neues Wort oder neuen Ausdruck hinzufügen...*, *Liste von Wörtern oder Ausdrücken importieren...* und *von bestimmten Dokumenten lernen*
  - *Dokumentation des Trainings*: Aus *DragonPad* Transkripte speichern sowie die zugehörigen Zeitstempeldateien mit dem *Erkennungsfenster* erstellen

- Zweite Diktierübung: Den unter Materialien angeführten Text zur Spracherkennung aus Wikipedia diktieren. Vorab die Spracherkennungssoftware bezüglich des neuen Vokabulars trainieren und nach dem Diktieren das Transkript korrigieren. Das mittels Dragon erstellte Transkript samt Zeitstempeldatei vor und nach dem Korrigieren für spätere Auswertungen abspeichern.

Übungsaufgaben bis zur nächsten Einheit: Die unter Materialien angeführte Textpassage über die Geschichte und Definition des Respeakings diktieren. Vor dem Diktieren Dragon mit dem speziellen Vokabular trainieren: Mittels *Neues Wort oder neuen Ausdruck hinzufügen...* die Wörter bzw. Wortfolgen *Red Bee Media* und *Interlinguale* und mittels *Liste von Wörtern oder Ausdrücke importieren...* die Wörter *Fresco*, *Ribas* und *Lambourne* trainieren. Das verbleibende (neue) Vokabular über *Von bestimmten Dokumenten lernen* hinzufügen und die ASR trainieren. Das Transkript samt Zeitstempeldatei archivieren (jeweils vor und nach der Korrektur).
Zusätzlich soll die auszubildende Person für ca. 10 Minuten ORF Sendungen mit Untertiteln ansehen. Dabei soll neben einer Sitcom (es handelt sich dabei um vorbereitete Untertitel, die mittels QWERTZ-Tastaturen erstellt werden), einer Sportübertragung (die durch Respeaking erstellten Live-Untertitel bzw. Semi-Live-Untertitel bei Hymnen, Aufstellungen, etc.) eine ZiB (Semi-Live-Untertitelung und Live-Untertitelung durch Schnellschreiber bzw. Schnellschreiberinnen bei Live-Schaltungen) angesehen und die Eindrücke in der nächsten Einheit geschildert werden.

3.3.2 Einheit 2 - Einführung in den Themenbereich Spracherkennung sowie erweitertes Training und Anpassung von Dragon

Ziel: Einen Einblick in die Arten und Anwendungsgebiete von Spracherkennungssoftwares und in die Funktionsweise von Dragon zu erlangen. Weiters soll Respeaking definiert werden können und dessen Methoden und Einsatzbereiche erlernt werden. Darüber hinaus soll der erweiterte Umgang mit der Hard- und Software trainiert werden, um der auszubildenden Person das selbstständige Durchführen der Trainings-, Diktier-, Korrektur- sowie Shadowing Übungsaufgaben zu ermöglichen.

Matrialien/Quellen: Angelehnt an die Respeaking Ausbildung in den UK (vgl. [RF12b]) werden neben dem theoretischen Wissen über die Untertitelung in der Einheit der erweiterte Umgang mit der Spracherkennungssoftware und dem Equipment vermittelt. Der Abschnitt 3.2.2 aus [Hat11, S: 59-70], der Abschnitt 2.5.1 dieser Diplomarbeit, [Nem13, Abschnitt 3.1] und schließlich Kapitel 5 sowie 6 aus [RF11, S: 56-93] dienen als Literatur zur Spracherkennung. Die Erläuterung zum Respeaking stammen aus [RF11, S: 1-5] und dem Kaptiel 2. Das Video *Respeaking at Swiss TXT- SF Sports*[19] sowie *BBC Item on Respeaking*[20] entstammen der beiliegenden DVD des Buches [RF11]. Die Bedienungsanleitung des *Samson AirLine 77 Headset System* ist die Grundlage für die Einstellungen des Mikrofons und dessen Empfänger. Die Referenz zu den vorgestellten Funktionalitäten von Dragon und dessen Bedienung sind die Kapitel 3 und 5 der DNS Bedienungsanleitung ([Nua10, S. 33-48; 67-91]). Die vom Autor dieser Diplomarbeit erstellte Diktierübung über die erlernten Funktionen (Interpunktions- und Sonderzeichen, Buchstabieren, Datum, etc.) ist in Abschnitt 6.1 zu finden. Die Idee zur mehrstufigen Shadowing-Übung stammt aus dem Kapitel 7.1.1 aus [RF11, S: 96-97].

Inhaltsübersicht/Ablauf:

- Automatische Spracherkennungssoftwares (ASR):
  - Grundproblem: Jeder Mensch hat eine individuelle Stimme und kann Worte bzw. Sätze nicht zweimal exakt gleich wiedergegeben. Auch Pausen zwischen Wörtern, Umgebungsgeräusche, Emotion, Lautstärke, Akzente, Dialekte, Sozialeke, etc. beeinflussen die Sprechweise. Weiters ist gerade im universitären Bereich Code-Switching üblich.
  - Einsatzgebiete von Spracherkennung samt den Einsatzgebieten: *Keyword-Spotter*, *Einzelworterkenner*, *Verbundworterkenner*, *Sprecherinnen-* bzw. *Sprechererkennung*, *kontinuierliche Spracherkennung* (ASR)

[19] [DVD]/Chapter 3/3.4.3/Respeaking at Swiss TXT- SF Sports.flv
[20] [DVD]/Chapter 11/11.4.3/BBC Item on Respeaking.flv

- Geschichte: 1950er Jahre bereits Experimente mit einzelnen Zahlen und einsilbigen Wörtern; 1960er Jahre erste Forschung mit kontinuierlicher Spracherkennung; 1970er Jahre tiefergehende Forschung (mit Sprachmustervergleich); seit 1980er Jahre statistischen Modelle im Einsatz. Solche Wahrscheinlichkeitsdichtefunktion (z.B. Hidden Markov Model (HMM)) sind noch heute in der Spracherkennung von hoher Bedeutung. Seit den 1990er Jahre sind Dikitersysteme kommerziell erwerbbar.
- Funktionsweise: Moderne ASR Systeme basieren auf den Kenntnissen der Phonetik, Signalverarbeitung, Statistik, Algorithmentheorie bis hin zur Linguistik (Sprachmodelle).
- Dreistufiger Aufbau: (a) Akustik Model, (b) Grammatik / Vokabular / Lexikon / Wörterbuch, (c) Sprachmodel(e)
- Kategorien von ASR:
  * *Sprecher/Sprecherinnen ab-* oder *unabhängig* sowie *adaptive Systeme*
  * ASR zum *Diktieren* oder für *spontane Sprache*: ASR zum Diktieren sind häufig Sprecher/Sprecherinnen abhängig und erzielen bessere Ergebnisse beim Diktieren von Interpunktionszeichen sowie geringen Umgebungsgeräuschen; spontane Sprache wird im Alltag gesprochen, unterscheidet sich meist stark von geschriebener Sprache. Die Erkennung von Alltagssprache ist technisch anspruchsvoller als von diktierter Sprache
  * *Live* oder *Offline Transkription*. Bei Live Systemen muss die jeweilige Verzögerung (engl. *delay*) berücksichtigt werden (z.B. Dragon vs. Viavoice)
- Verbreitete sowie für die Untertitelung zukünftig relevante Produkte: Dragon (DNS), ViaVoice, European Media Laboratory GmbH (EML), Google, Inc. (YouTube)

- Respeaking:
  - „A technique in which a respeaker listens to the original sound of a live programme or event and respeaks it, including punctuation marks and some specific features for the deaf and hard of hearing audience, to a speech recognition software, which turns the recognized utterances into subtitles displayed on the screen with the shortest possible delay“ [RF11, S: 1].
  - Derzeitiger Einsatz von Respeaking: in UK Independent Media Support (IMS erzeugt ca. 200h Live-Untertitel wöchentlich) und Red Bee Media (RBM erzeugt wöchentlich zwischen 256 und 320 Stunden Untertitel durch Respeaking); Schweiz (185h Respeaking wöchentlich); weiters in den Ländern: Spanien, Belgien (Flandern), Österreich, Dänemark, Frankreich, Italien, Kanada (seit 2008 auch offline) und US (wenngleich in US die zweite Wahl).
  - Training an Universitäten: Barcelona (als online Kurs mit 10 ECTS Punkten oder als Teil eines Moduls mit 4 ECTS Punkten), London, Belgien (3 ECTS Punkte) sowie in den US
  - Korrektur Methoden: ohne Korrektur, Eigenkorrektur (die üblichste Methode) oder parallele Korrektur

  - Editier Strategie: '1:1' Transkription vs. Kürzungen: In der Praxis meist nahezu '1:1' Transkription ('near-verbatim')
  - Videobeitrag *Respeaking at Swiss TXT- SF Sports*

- Fortgeschrittene Bedienung und Verwendung des Mikrofons/Empfängers[21]. Es wird folgendes erläutert bzw. durchgeführt:
  - Richtige Positionierung des Empfängers: idealerweise Aufrechterhaltung der Sichtlinie
  - Sender: Ein- und Ausschalten (*Power*), Stummschaltung (*Mute*), Leuchtsignal bei verbleibender Lebensdauer von 2 Stunden der Batterie sowie den Batteriewechsel
  - Empfänger: Anschluss am PC mittels *unbalanced output*, das Ein- und Ausschalten (Überprüfung durch *Power*-LED), Ausrichtung der Antennen und Erläuterung der beiden Antennen LEDs, die Funktionen des *Volume*- und des *Squelch*-Reglers[22] sowie die Überprüfung des Signals (korrekte Adjustierung des Volume- sowie Squelch-Reglers)
  - Windows Audio-Settings: Änderung auf 2 Kanal, 16 Bit, 96.000 Hz (Studioqualität)
  - Funktion und Lautstärke des Mikrofons prüfen (mittels der Dragon Funktionalität *Mikrofon prüfen...*).

- Fortgeschrittene Bedienung von Dragon im Diktiermodus. Es wird folgendes erläutert bzw. durchgeführt:
  - Erweiterte Interpunktions- und Sonderzeichen: *Eckige Klammer auf*, *Eckige Klammer zu*, *Geschweifte Klammer auf*, *Geschweifte Klammer zu*, *Schrägstrich*, *Anführungszeichen auf*, *Anführungszeichen zu*, *Punkt Punkt Punk*, *Pluszeichen*, *Minuszeichen*, *Paragrafzeichen*, *Prozentzeichen*, *Urheberrechtssymbol*, etc.
  - Großschaltung: *Großschaltung anfangen* und *Großschaltung beenden*
  - Buchstabieren von Abkürzungen: *Anton*, *Berta*, *Cäsar*, [...], *Eszet*, *Ärger*, *Österreich*, *Übermut*
  - Zahlen, römischen Zahlen und Währungszeichen: *0617/9 65 32*, *12.530*, etc.; *I*, *V*, *X*, *C*, *M*, *XXIV*, etc.; Euro- sowie Dollar-Symbol
  - Datumsangaben und Uhrzeit: *22. Januar 1999*, *9. April 2001*, *3.11.2002*, *am 1. April*; *8.30 Uhr*
  - Detaillierte Funktionsweise von *Wort oder Ausdruck bearbeiten...*: Makros, Wörter trainieren, löschen, Worteigenschaften (*vor dem Wort*, *Folgewort einstellen*, *Alternative Schreibweise*, etc.) sowie das Selektieren der Anzeige (*Alle Wörter*, *Nur benutzerdefinierte Wörter*, etc.)

[21] Anm. Autor: Das *Samson AirLine 77 Headset System*, bestehend aus dem *AH1 Wireless Transmitter* und dem *CR77 Half-rack UHF receiver*, siehe Abschnitt 3.1.3

[22] Anm. Autor: Respektive der Lautstärken- und der Rausch-Regler

  - Erkennungsraten verbessern: *Erkennung eines Wortes oder Ausdrucks verbessern...* sowie *Genauigkeitsoptimierung starten...*

- Diktierübung über die erlernten Funktionen (Interpunktions- und Sonderzeichen, Buchstabieren, Datum, etc.) mit anschließendem Korrigieren. Nicht zuletzt soll die auszubildende Person in der Übung eine individuelle Taktik entwickeln, in welchen Situationen besser diktiert und in welchen besser getippt wird (z.B. kann es schneller sein bei E-Mail-Adressen die Tastatur zu verwenden, etc.)

Übungsaufgaben bis zur nächsten Einheit: Den Videobeitrag *BBC Item on Respeaking* ansehen, um dessen Inhalt in der nächsten Einheit zu diskutieren. In Dragon den vorgegeben Text *Der Stechlin (Roman)*[23] lesen, um die Genauigkeit für das Benutzerprofil weiter zu trainieren. Dabei ist es wichtig, bei falscher oder undeutlicher Aussprache ggf. einzelne Passagen zu wiederholen. Zusätzlich soll die auszubildende Person Shadowing[24]-Übungen durchführen. Dazu eignen sich TV Übertragungen mit geringer Redegeschwindigkeit (z.B. Sportübertragungen, Kindersendungen, etc.). Dabei wird das Gesagte so schnell wie möglich 'nachgesprochen'. Die Übungen erfolgen ohne den Einsatz von ASR. Die Verzögerung beim Nachsprechen sollte dabei so gering wie möglich und die Aussprache deutlich sein. Darauf folgend soll das Gesagte ohne lautes Sprechen gedanklich nachgesprochen werden, dabei allerdings die Verzögerung einen halben bis eineinhalb Sätze betragen. Nach einiger Übung sollen gedanklich auch Interpunktionszeichen (im Kopf beim 'Diktieren') hinzugefügt werden. Als letzte Übung soll das Gesagte bereits samt Interpunktionszeichen in der genannten Verzögerung nachgesprochen werden. Die vier Übungen sollen jeweils für mindestens 15 Minuten durchgeführt werden, bevorzugt aber deutlich länger (mit Erholungspausen während und zwischen den Übungen). Am Besten werden die Übungen über mehrere Tage verteilt durchgeführt.
Zur Verbesserung der Erkennungsgenauigkeit ist ab Einheit 2 von der auszubildenden Person in regelmäßigen Abständen (z.B. wöchentlich) eine Modelloptimierung durchzuführen.

[23] Anm. Dragon stellt mehrere vorgefertigte Texte zum Ablesen zur Verfügung, um die Erkennungsraten zu erhöhen.
[24] Anm. Autor: Siehe Abschnitt 4.2.3 bezüglich Shadowing.

3.3.3 Einheit 3 - Vorbereitung sowie Multitasking beim Respeaking I: Übungen zur Vorbereitung, zum Shadowing, zum Verständnis und Diktieren

Ziel: Weitere Shadowing- sowie Diktierübungen (ohne ASR) zum Üben und Feststellen der Aussprache und des Sprechrhythmus. Darüber hinaus eine detaillierte Erläuterung der (teilweise) parallel ablaufenden Respeaking Tätigkeiten. Eine Übung soll die Respeaking-Vorbereitungstätigkeiten festigen und eine weitere Übung das Verständnis, das Diktieren und die Analyse verbessern.

Matrialien/Quellen: Als Literatur zum Shadowing sowie Multitasking beim Respeaking dienen die Abschnitte 7.1.1 und 7.1.2 aus [RF11, S: 96-101] sowie die im Abschnitt 4.2.3 dieser Diplomarbeit angeführten Quellen. Wie in Einheit 2 stammen die Erläuterungen bezüglich ASR sowie Respeaking aus dem Kapitel 2 dieser Diplomarbeit, [RF11, S: 1-5; 56-93], [Hat11, S: 59-70] sowie [Nem13, Abschnitt 3.1]. Die Idee zur Übung der (auftragsbezogenen) Vorbereitung stammt aus [ARRF08]. Die zum Zeitpunkt der Ausbildung aktuelle Verhandlung über die Anschläge in Norwegen, bei denen im Jahr 2011 77 Menschen ums Leben kamen, diente als Thema für die Übung. Als Video-Material wurden die ein- bis zweiminütigen ZiB Beiträge verwendet, die während der Zeit der Verhandlung rund um den Angeklagten Anders Behring Breivik im ORF ausgestrahlt wurden, siehe Abschnitt 4.2.4. Für die darauf aufbauende Übung zum Training von wichtigen Respeaking Skills diente [ARRF08] als Inspiration. Als Video-Material wurden dazu ebenfalls ZiB-Beiträge verwendet, die über die Verhandlung von Behring Breivik berichteten.

Inhaltsübersicht/Ablauf:

- Beim Respeaking ist eine gute Betonung und Aussprache sehr wichtig. Auch auf einen gleichmäßigen Rhythmus und kurze Pausen nach Interpunktionszeichen ist zu achten.
- Reflexion über die Übungen der Einheit 2 sowie erneute Shadowing-Übung:
  - Shadowing soll beim Erlernen des gleichzeitigen Hörens und Sprechens helfen
  - Shadowing ist/war Teil von Simultandolmetschausbildungen, ist aber nicht unumstritten
  - Feststellung des Übungsfortschrittes: Die auszubildende Person spricht einen Videobeitrag (ohne den Einsatz einer ASR) so schnell wie möglich nach und hält die Verzögerung beim Nachsprechen dabei so gering wie möglich ('Phonemic Shadowing'). Anschließend wird das Gesagte nun laut und samt Interpunktionszeichen mit einer Verzögerung von einem halben bis eineinhalb Sätze nachgesprochen ('Phrase Shadowing') und die Stimme der auszubildenden Person dabei aufgezeichnet. Die Aufzeichnung wird darauf folgend zusammen mit der auszubildenden Person analysiert, um im Hinblick auf die ASR eventuelle Probleme zu lokalisieren. So werden die Aussprache und der Sprechrhythmus analysiert und ggf. weitere Übungsaufgaben bis zur Folgeeinheit vereinbart.

- Multitasking beim Respeaking: Zuhören, Respeaken, Überprüfen, Zusehen, Tippen
  - Die Tätigkeiten können parallel ablaufen, sind aber nicht notwendigerweise überlappend. Erläuterung anhand der Abbildung 2.3
  - Einzig das Tippen und Respeaken können *nie* parallel stattfinden
  - Ein Respeaker bzw. eine Respeakerin kann nicht (wie oft Simultandolmetscher bzw. Silmuntandolmetscherinnen) kurzzeitig schneller sprechen, um ggf. 'aufzuholen', da das die Erkennungsraten beeinflussen würde. Dies ist beim Korrigieren (Tippen) zu beachten.
- Übung zur (auftragsbezogenen) Vorbereitung:
  - Für das Training der Spracherkennung wird das Thema der auszubildenden Person bekanntgegeben. Zum Thema sollen ein bzw. mehrere TV-Beiträge vorhanden sein (insgesamt von mind. 5 Minuten), zu dem weiters die Untertitel des TV-Senders verfügbar sind.
  - Die Recherche und das Training der ASR wird in der etwa 20 Minuten dauernden Vorbereitungszeit durchgeführt. Dabei soll der Fokus auf dem zu erwartenden Vokabular liegen. Dies umfasst das Trainieren von neuen Wörtern und Wortfolgen ebenso wie das Üben von vorhandenen Wörtern, bei denen jedoch Fehler zu erwarten sind.
- Übung wichtiger Respeaking Skills: Verständnis, Diktieren und Analyse samt Korrektur zum vorbereiteten Thema. Dabei werden die Schritte (a) bis (f) in mehreren Iterationen durchgeführt:

  (a) Einen 8-15 Sekunden Block[25] der Aufzeichnung(en) über das vorbereitete (trainierte) Thema wird abgespielt. Die Dauer des Blocks kann im Schritt (f) erörtert werden.

  (b) Die auszubildende Person folgt dem Gesprochenen, ohne Notizen zu nehmen.

  (c) Die auszubildende Person diktiert mit der ASR den Inhalt des abgespielten Blocks aus dem Gedächtnis. Es soll versucht werden, beim Diktieren dem originalen Text so weit wie möglich zu folgen. Damit soll sichergestellt werden, dass keine (wesentlichen) Informationen vergessen oder weiters keine falschen Information transkribiert werden. Da die Diktiertätigkeit - im Gegensatz zum Respeaking in einer Live-Situation - nicht parallel zum Gesprochenen erfolgt, sind in der Phase Aspekte wie Diktiergeschwindigkeit und Rhythmus (noch) nicht ausschlaggebend. Der Fokus liegt beim Zuhören und Wiedergeben (Steigerung der Konzentrationsspanne) sowie dem Umgang mit der ASR.

[25] Anm. Autor: Bei einer Redegeschwindigkeit von 145 WpM entspricht das ca. 19 bis 36 Wörtern. Die Angabe der Sekunden bzw. Wörter gilt als Richtlinie und soll einen Anfangswert für die ersten Iterationen geben, um schließlich die Konzentrationsspanne nach mehreren Iterationen feststellen zu können.

(d) Die auszubildende Person beurteilt nun die Qualität des Transkripts, ohne das Gesprochene nochmals zu hören (ob es Passagen gibt, an die sie sich nicht mehr genau erinnern kann, ob die auszubildende Person davon ausgeht dass alles korrekt transkribiert wurde, etc.). Ggf. kann auch die ASR noch mit weiteren Wörtern trainiert werden.

(e) Der Beitragsblock wird erneut abgespielt und Fehler werden lokalisiert, um deren Ursache zu begründen (falsche Information, nicht transkribierte Information, Fehler im Umgang mit der ASR, etc.). Im Anschluss daran werden die Fehler korrigiert.

(f) Die Konzentrationsspanne der auszubildenden Person wird festgestellt (z.B. ob Probleme bei Blöcken ab 10 Sekunden bzw. ab einer gewissen Wortanzahl auftreten, etc.). Darauf hin wird wieder beim Punkt (a) begonnen, bis schließlich ein Beitrag von einigen Minuten korrekt transkribiert wurde. Bei den jeweiligen Iterationen wird die Konzentrationsspanne ermittelt, in der die auszubildende Person das Gesagte korrekt (ggf. umformuliert) wiedergeben kann.

(g) Das erstellte Transkript wird mit jenem der TV-Untertitel verglichen und Unterschiede werden diskutiert.

Übungsaufgaben bis zur nächsten Einheit: Die Verständnis-, Diktier- und Analyseübung aus der Einheit selbständig mit frei wählbaren TV-Aufzeichnungen durchführen. Empfohlen werden Aufzeichnungen mit einer konstanten Redegeschwindigkeiten (vergleichbare Geschwindigkeit zu jener des Beitrags aus der Übung während der Einheit) und vorbereiteten Reden wie beispielsweise ZiB Beiträge. Dabei soll gezielt auf die Vermeidung der (während der Einheit) lokalisierten Aussprachefehler geachtet und somit Erkennungsfehler der ASR minimiert werden. Die - während der Einheit festgestellte - Konzentrationsspanne soll weiter gefestigt und im besten Fall kontinuierlich erhöht werden. Es werden mindesten zwei Stunden für die Übung empfohlen. Weiters soll zur Verbesserung der Erkennungsgenauigkeit eine Modelloptimierung durchgeführt werden (Erläuterung erfolgte in Einheit 2).

3.3.4 Einheit 4 - Feststellung von verschiedenen Diktiergeschwindigkeiten und Messung mit Wortfehlerraten samt erster Respeaking Übung

Ziel: Die Feststellung der *angenehmen*, *maximalen* sowie *optimalen Diktiergeschwindigkeit* unter Berücksichtigung der erläuterten und zu messenden Wortfehlerraten bzw. Worterkennungsraten. Darauf aufbauend soll eine erste Respeakingübung die Einheit abrunden und gemeinsam mit der Diktierübung die Basiskenntnisse für die Übungsaufgabe bilden.

Matrialien/Quellen: Als Literatur zur Erkennungsgenauigkeit dient der Abschnitt 2.6. Die Idee zur Übung für die Feststellung der verschiedenen Diktiergeschwindigkeiten stammt aus dem Kapitel 7.4 aus [RF11, S: 112-120]. Der dabei zu diktierende/abzulesende Text ist der deutschsprachige Wikipedia Artikel über Untertitelung[26], wobei lediglich der einleitende Absatz sowie die ersten beiden Absätze von *Zeitdruck und Beschränkungen*, mit gesamt mit 151 Wörtern (ohne Interpunktionszeichen) verwendet werden. Als Video-Material zur ersten Respeaking Übung dient ein knapp zweiminütiger ZiB Beitrag (*Doppelagent vereitelt Anschlag*. Als Videomaterial für die Übungsaufgaben bis zur nächsten Einheit dienen drei ZiB Beiträge über den österreichischen Nationalratsabgeordneten Martin Graf. Dieser wurde beschuldigt, als Vorstand einer Privatstiftung zum Nachteil der Stifterin gehandelt zu haben. vom 09.05.2012).

Inhaltsübersicht/Ablauf:

- Erkennungsgenauigkeit (Akkuratheit):
  - Erläuterung der Berechnung der Wortfehlerrate (WER) bzw. der Worterkennungsrate (WRR).
  - Hintergrund: WER bzw. WRR sind für klassische ASR Systeme entworfen.
  - WER bzw. WRR in Bezug auf Respeaking: WER bzw. WRR sind nicht gut für die Evaluierung der Qualität der durch Respeaking erstellten Untertitel geeignet. Eine speziell für das Respeaking entworfene Berechnung der Worterkennungsrate wird in der Einheit 5 vorgestellt.
- Analyse über den Fortschritt der Respeaking Skills *Verständnis*, *Diktieren* und *Analyse* samt *Korrektur*: Wiederholung der Übung aus der Einheit 3[27], um die Konzentrationsspanne zu messen und ggf. Probleme und Fehler zu analysieren.

[26] Herausgeber: Wikipedia, Die freie Enzyklopädie; `http://de.wikipedia.org/w/index.php?title=Untertitel&oldid=103002941`, Datum der letzten Bearbeitung: 09.05.2012, 08:11 UTC; Versions-ID der Seite: 103002941

[27] Anm. Autor: Es sollten 3-5 Iterationen ausreichend sein, um einen Überblick über den Übungsfortschritt zu erlangen.

- Feststellung der (derzeitigen) *angenehmen*, *maximalen* und *optimalen* Diktiergeschwindigkeit: Die auszubildende Person liest zu Beginn der Übung einen vorgegebenen, kurzen Text (in etwa 150 bis 200 Wörter) und trainiert ggf. die Spracherkennung bezüglich unbekanntem Vokabular. Dann werden die folgenden Schritte (a) bis (d) in mehreren Iterationen[28] durchgeführt, bis die in Schritt (b) festgestellte Worterkennungsrate für die *angenehme* und im Anschluss daran die *maximale* Diktiergeschwindigkeit mindestens 95% beträgt. Als angenehm wird jene Diktiergeschwindigkeit bezeichnet, in der sich die zukünftige Respeakerin bzw. der zukünftige Respeaker nach aktuellem Übungsstand 'wohl' fühlt (Zeit für klare Aussprache, ohne Hast, etc.) und von sich selbst annimmt, über einen längeren Zeitraum mit konstant wenigen Diktierfehlern die Spracherkennung bedienen zu können. Beim Diktieren mit der maximalen Geschwindigkeit hingegen wird versucht, so schnell wie möglich zu diktieren, jedoch in ähnlicher oder leicht geringerer Qualität der angenehmen Diktiergeschwindigkeit. Schließlich wird eine *optimale* Diktiergeschwindigkeit festgelegt, die nach der Erfahrung von angenehmer und maximaler Diktiergeschwindigkeit von der auszubildenden Person selbst definiert und für die nächste Übung (erste Respeaking Übung) weiterverwendet wird. Die optimale Diktiergeschwindigkeit kann sich mit der angenehmen decken, jedoch auch zwischen angenehmer und maximaler liegen. Auch bei der angenehmen Diktiergeschwindigkeit soll die WRR 95% nie unterschreiten. Die Schritte/Iterationen zum Feststellen der Geschwindigkeiten laufen wie folgt ab:

  (a) Den Text lesen/diktieren und dabei die benötigte Zeit messen.

  (b) Die WRR berechnen. Sie wird mit sowie ohne Einbeziehung von Interpunktionszeichen (Punkt, Beistrich, Neuer Absatz, etc.) berechnet, Fehler in der Groß- und Kleinschreibung können vernachlässigt werden sofern sie nicht eine andere Interpretation des Satzes zulassen.

  (c) Anhand der gemessenen Zeit und der Wortanzahl die Diktiergeschwindigkeit berechnen (z.B. 140 WpM mit Interpunktionszeichen und 115 WpM ohne Interpunktionszeichen bei maximaler Diktiergeschwindigkeit).

  (d) Die Fehler analysieren (ggf. Spracherkennung erneut trainieren) und die Probleme von WRR diskutieren (z.B. keine Unterscheidung der Fehlerschwere, etc.).

[28] Anm. Autor: siehe Abschnitt 4.2.5 bezüglich dem Hintergrund und den möglichen Konsequenzen die beim mehrfachen Lesen des selben Textes zu beachten sind.

- Erste Respeaking Übung mit leicht reduzierter, *optimaler* Diktiergeschwindigkeit samt Analyse und kritischer Betrachtung von WER/WRR:
  - Das Thema des zu untertitelten Beitrags wird der auszubildenden Person bekanntgegeben. Eine vorbereitete Wortliste dient zum Trainieren der ASR.
  - Der Beitrag wird mit einer Wiedergabegeschwindigkeit von ca. 20 WpM unter der optimalen Diktiergeschwindigkeit abgespielt und die auszubildende Person dazu angehalten, auch dem Inhalt zu folgen.
  - Die auszubildende Person transkribiert den Beitrag mittels Respeaking. Aufgrund der gewählten Wiedergabegeschwindigkeit sollte nahezu eine '1:1' Transkription möglich sein, die auch die Feststellung des WRR Wertes erleichtert. Prinzipiell ist es aber auch möglich, das Gehörte umzuformulieren bzw. zu kürzen[29].
  - Nach dem Respeaken soll die auszubildende Person den Inhalt soweit wie möglich wiederholen und feststellen, dass beim Respeaken das Erfassen des Inhaltes wichtig ist.
  - Analyse des transkribierten Inhalts und eine Diskussion über Probleme mit der WRR-Methode.

Übungsaufgaben bis zur nächsten Einheit: Die Diktierübung aus der Einheit mit frei wählbaren Text erneut durchführen. Dabei soll der Text deutlich länger sein als in der Einheit und aus mind. 700 Wörter (ohne IZ) bestehen. Beginnend mit der optimalen Diktiergeschwindigkeit soll gemessen werden, ob auch bei einem längeren Text die Geschwindigkeit gehalten werden kann und somit bereits ein Gefühl für die während der Einheit festgestellte, optimale Geschwindigkeit vom Respeaker bzw. der Respeakerin vorhanden ist. Weiters soll überprüft werden, ob dabei die angestrebte WRR aus der Einheit (mind. 95%) erreicht wird und ob eine Fehlerkonzentration zu verzeichnen ist (z.B. am Ende des Textes wegen mangelnder Konzentration, etc.). Ggf. soll die in der Einheit festgestellte optimale Diktiergeschwindigkeit neu festgelegt werden. Anschließend wird erneut mit maximaler Diktiergeschwindigkeit der selbe Text diktiert und die selben Kriterien überprüft und ggf. die maximale Diktiergeschwindigkeit aus der Einheit korrigiert.
Die Respeaking Übung aus der Einheit soll mit der optimalen Diktiergeschwindigkeit mehrfach wiederholt werden. Empfohlen werden Aufzeichnungen mit einer konstanten Redegeschwindigkeiten (vergleichbare Geschwindigkeit zu jener des Beitrags aus der Übung während der Einheit) und vorbereiteten Reden wie beispielsweise Nachrichten Beiträge. Dabei soll die Wiedergabegeschwindigkeit wie in der Einheit ca. 20 WpM unter der optimalen Diktiergeschwindigkeit liegen. Beim Respeaken soll gezielt auf die Vermeidung der (während der Einheit und der erneuten Diktierübung) lokalisierten Aussprachefehler geachtet und somit Erkennungsfehler der ASR minimiert werden. Es sollen mindestens 20 Minuten transkribiert werden, wobei bei verschiedenen Themenbereichen ggf. mehrmals die Spracherkennung vorab trainiert werden muss. Weiters soll zur Verbesserung der Erkennungsgenauigkeit eine Modelloptimierung durchgeführt werden (Erläuterung erfolgte in Einheit 2).

[29] Anm. Autor: Das Umformulieren und Kürzen sind wichtige Respeaking Techniken, die erst in späteren Übungen trainiert werden. Allerdings sollen sie nicht unterbunden werden, wenn sie bereits bei ersten Übungen 'automatisch' durchgeführt werden.

3.3.5 Einheit 5 - Theorie und praktische Übungen zur Interpunktion beim Respeaking sowie beim NER-Modell

Ziel: Fundiertes Wissen über Hintergründe sowie Anwendung des NER-Models (das zur Messung der Qualität von (Live) Untertitel verwendet wird) zu erlangen. Nach einer Übung zur Interpunktion beim Respeaking steht die praktische Anwendung vom NER-Modell im Vordergrund.

Matrialien/Quellen: Als Literatur zur NER-Analyse dient der Abschnitt 2.6 bzw. die dort angeführten Literaturquellen ([RF11, S: 150-161], [RFM14]). Das Video-Material zur Analyse über den Fortschritt nach den Respeaking Übungen ist ein ZiB Beitrag mit eineinhalb Minuten Dauer (*Auch in Österreich gibt es Erdbeben-Linien* vom 29.05.2012). Nach dem Respeaking werden gezielte Fragen - Entscheidungsfrage (ja/nein) und Denkfragen (Warum-Fragen) - über den Inhalt gestellt. Die Idee stammt aus Erfahrungen, die im Zuge von Ausbildungen zum Simultandolmetschen gesammelt wurden, und in [Kur96, S: 100-106] dokumentiert sind. Als Literatur zur Interpunktion beim Respeaking dienen die Kapitel 7.2 bis 7.4 des Buches [RF11, S: 101-119] sowie der Abschnitt 2.2.2 ab der Seite 26. Resultierend aus dieser Quelle entstand die Idee zur Übung zum Setzen von Interpunktionszeichen. Die Beispiele für 'richtiges' Setzen von IZ wurden vom Autor dieser Diplomarbeit für deutschsprachige Untertitel adaptiert. Die Vorlesungseinheit der LVA Trainingswissenschaft vom 08.05.2012 (siehe Kapitel 4) dient als Material für die Übung zum Setzen von Interpunktionszeichen sowie für die Respeaking Übungsaufgabe, die bis zur nächsten Einheit durchzuführen ist.

Inhaltsübersicht/Ablauf:

- Das NER-Modell (Erkennungsgenauigkeit/Akkuratheit):
  - Gründe und Motivation für die Entwicklung des NER-Modells: vergleichbare Werte; funktionell und einfach anwendbar; für gekürzte/umformulierte Untertitel geeignet; für unterschiedliche Sprachen nutzbar; die Live-Korrektur berücksichtigen; weitere Informationen im Zusammenhang mit der Qualität (wie Verzögerung, Einfluss der Sprechgeschwindigkeit, Verbesserungspotenzial, etc.) enthalten, etc.
  - Erläuterung der Berechnungsformel (Editierfehler, Erkennungsfehler, Korrekte Editierung, Beurteilung, gravierende vs. normale vs. geringfügige Fehler).
  - Vergleich vom NER-Modell und WRR/WER.

- Analyse über den Fortschritt der Respeaking Übung mit *optimaler* Respeakinggeschwindigkeit[30] samt Analyse durch das NER-Modell:
  - Das Thema des zu untertitelten Beitrags wird der auszubildenden Person bekanntgegeben. Eine vorbereitete Wortliste dient zum Trainieren der ASR.
  - Der Beitrag wird mit der Respeakinggeschwindigkeit abgespielt und die auszubildende Person dazu angehalten, auch dem Inhalt zu folgen.
  - Die auszubildende Person transkribiert den Beitrag mittels Respeaking. Es ist möglich, das Gehörte umzuformulieren bzw. zu kürzen (nahezu '1:1' Transkription). Es werden kurz die Möglichkeiten zur Umformulierung/Kürzung erläutert, auf die dann in der nächsten Einheit eingegangen wird.
  - Nach dem Respeaken soll die auszubildende Person den Inhalt so gut wie möglich wiederholen und somit üben, dass beim Respeaken das Erfassen des Inhaltes wichtig ist. Ergänzend müssen Fragen[31] zum Beitrag beantwortet werden.
  - Abschließend erfolgt die Analyse des transkribierten Inhalts durch das NER-Modell. Dabei wird ebenfalls die Form der Dokumentation (vorbereitete Tabellen und Hervorhebung der verschiedenen Fehlerarten im Transkript) erläutert.
- Interpunktion beim Respeaking:
  - In der Praxis ist die Wortanzahl in den Untertiteln aufgrund der Interpunktion selbst bei niedrigen Sprechgeschwindigkeiten geringer als in der Originalquelle (ein Respeaker oder eine Respeakerin müsste 'schneller sprechen' als die original Sprachquelle).
  - Bei der Arbeit mit Dragon beeinflusst die Interpunktion die Zeitverzögerung sowie die Erkennungsgenauigkeit .
  - Folgende Interpunktionen können die Lesbarkeit der Untertitel erhöhen sowie deren Verzögerung verringern: Beistriche bei Aufzählungen; Beistriche bei Nebensätzen; Beistriche nach einleitenden Sätzen; ggf. vor einem *und* oder *oder* einen Beistrich setzen; ein *und*, dass zwei Sätze verbindet kann ggf. dazu verwendet werden lange Sätze in zwei kürzere zu Teilen; Zitate hervorheben („Anführungszeichen“ oder $\ll Dragon Zitat \gg$), etc.
  - Eine Respeakerin oder ein Respeaker muss Interpunktionszeichen (Beistrich, Punkt, etc.) ohne langes Nachdenken (nahezu 'intuitiv') setzen.
- Übung zum Setzen von Interpunktionszeichen: Bei einem vorbereiteten '1:1' Transkript eines Vortrages (z.B. Vorlesung), bei dem sämtliche Interpunktionszeichen entfernt wurden (sowie die Großschreibung an Satzanfängen) muss die auszubildende Person die Interpunktionszeichen setzen.

[30] Anm. Die Geschwindigkeit wird von der auszubildenden Person selbst festgelegt und soll anhand der bisherigen Erfahrung (der unterschiedlichen Diktiergeschwindigkeiten und den ersten Respeaking Messungen) festgelegt werden. Gute Erkennungsraten (WRR über 95%) sollen angestrebt werden und in der gewählten Geschwindigkeit möglich sein.

[31] Anm. Autor: Entscheidungsfrage (ja/nein) und Denkfragen (Warum-Fragen).

Übungsaufgaben bis zur nächsten Einheit: Respeaking wird mit dem Vorlesungsmaterial weiter geübt. Im Vordergrund steht dabei, dass nach den Übungen ein NER-Wert von mindestens 98% erreicht wird und die Lesbarkeit der Untertitel durch das Setzen von Interpunktionszeichen erhöht wird. Dabei soll vor allem auf Erkennungsfehler geachtet werden, um sie im Laufe der Übung zu minimieren. Primär soll mehr Erfahrung mit dem Respeaking gesammelt werden. Somit ist es nicht notwendig, sämtliche transkribierte Passagen mit dem NER-Modell zu analysieren. Repräsentative Stichproben sollen während der Übung dennoch die Qualität der Untertitel sicherstellen und der auszubildenden Person bei der Selbsteinschätzung helfen. Abschließend werden zehn Minuten der Vorlesung transkribiert und schließlich mit dem NER-Modell analysiert. Wie bereits in den letzten Einheiten soll zur Verbesserung der Erkennungsgenauigkeit eine Modelloptimierung durchgeführt werden (Erläuterung erfolgte in Einheit 2).

3.3.6 Einheit 6 - Multitasking beim Respeaking II: Korrekte Editierung und Erkennen von (gravierenden) Erkennungsfehlern

Ziel: Theoretisches Wissen zum Décalage[32] sowie zum Editieren der Untertitel (Umformulierung, Kürzen und den Umgang mit nicht vorhandenen Vokabular) zu erwerben. In weiteren Respeaking Übungen wird das Editieren sowie das aktive Beobachten der erzeugten Untertitel (zum Erkennen von Fehlern) trainiert.

Matrialien/Quellen: Als Literatur zu Décalage und den Sinneseinheiten beim Respeaking (engl. *units of meaning used by respeakers*) dient das Kapitel 7.3 aus [RF11, S: 107-112]. Der Abschnitt 2.2.2 bzw. die dort angeführten Literaturquellen ([ARRF08], [RF11, S: 22-44; 74-122], [Hat11, S: 51-70; 116-124], [PK08, S: 393]) dienen als Quellen für das Umformulieren, Kürzen und den Umgang mit nicht vorhandenem Vokabular. Die Ausarbeitung der Übung zum Training der Respeaking Skills *Umformulieren*, *Kürzen*, *Umgang mit nicht vorhandenem Vokabular* und *Beobachten* basiert auf den Erkenntnissen und Empfehlungen die im Kapitel 7 von [RF11, S: 94-122] angeführt sind und einer Konversation mit Ao. Univ.-Prof. Dipl.-Dolm. Dr. Ingrid Kurz (vgl. [Kur12]), u.a. Autorin von [Kur96]. In den Übungen während der Einheit sowie in der Einzelarbeit zur nächsten Einheit wird wie bereits in Einheit 5 die Vorlesung Trainingswissenschaft vom 08.05.2012 (siehe Kapitel 4) verwendet.

Inhaltsübersicht/Ablauf:

- Als *Décalage* wird die Verzögerung zur Originalquelle (z.B. Professorin oder Professor) bezeichnet, mit der ein Respeaker oder eine Respeakerin zu diktieren beginnt. Es gibt zwei grundsätzliche Ansätze, die sich beide nicht an einem Zeitwert (Sekunden) orientieren, sondern an den zu transkribierenden Äußerungen:
  - Erster Ansatz: Sobald eine Sinneseinheit gesprochen wurde. Als solche wird die kleinste brauchbare Einheit für den Respeaker oder die Respeakerin bezeichnet. Eine Sinneseinheit kann aus einem einzelnen Wort bis hin zu einer langen Phrase bestehen.
  - Zweiter Ansatz: Mit dem Respeaken warten bis man genügend Information hat, dass man das Ende des Satzes beim Beginnen des diktierens sicher weiß. Der Ansatz ist gerade beim live Untertiteln bezüglich der Verzögerung problematisch.
  - Gute Respeakerinnen und Respeaker sollen mehr oder weniger einen konstanten Décalage halten können, ggf. aber auch flexibel verändern können.

[32] Anm. Autor: Als Décalage wird der Abstand vom Diktierten zum Originalsprecher bzw. zur Originalsprecherin bezeichnet (vgl. [RF11, S: 107-112]).

- Umformulierung, Kürzen und Umgang mit nicht vorhandenem Vokabular:
  - Als *Umformulieren* wird die Veränderung der grammatikalischen Struktur von Sätzen bezeichnet, bei der dennoch die Aussage des Satzes nicht verändert wird. Dies kann das Umwandeln von passiven in aktive Verben ebenso sein wie das Ausgliedern von Nebensätzen und das Verändern der Satzstruktur. Das Umformulieren soll der Lesbarkeit von Untertiteln dienen.
  - Als *Paraphrasieren* wird das Umschreiben von unbekanntem Vokabular (Wörter die die Spracherkennung nicht kennt) bezeichnet. Um Erkennungsfehler zu vermeiden, soll eine Respeakerin oder ein Respeaker Wörter umschreiben können, die von der Spracherkennung nicht erkannt werden würden.
  - Beim *Kürzen* werden Redundante Informationen in den Untertiteln weggelassen bzw. bei hohen Geschwindigkeiten Gedankeneinheiten zusammengefasst.
- Beobachten der erzeugten Untertitel: Beim (live) Korrigieren ist es wichtig, dass man weiß welche Wörter eventuell nicht richtig erkannt werden. Speziell ist dabei auf Zahlen, Namen, und (spezielle) Ausdrücke zu achten, die noch nicht trainiert wurden bzw. wo häufig Erkennungsfehler auftreten.
- Übung der Respeaking Skills *Umformulieren*, *Kürzen*, *Umgang mit nicht vorhandenem Vokabular* und *Beobachten* als Vorbereitung zur (Echtzeit) Korrektur: Die auszubildende Person transkribiert einen Vortrag in mehreren Schritten von je drei Minuten (beispielsweise werden bei fünf Iterationen werden somit 15 Minuten transkribiert). In der ersten Iteration wird mit der derzeitigen Respeakinggeschwindigkeit (nach den Übungen in der Einheit 5, mit welcher mind. 98% NER-Wert erreicht wird) begonnen und die Iterationen (siehe Schritte (a) bis (d)) durchgeführt. Nach jeder Iteration wird die Wiedergabegeschwindigkeit erhöht, bis sie höher ist als die maximalen Diktiergeschwindigkeit (siehe Einheit 4). In jeder Iteration werden folgende Schritte durchgeführt:
  - (a) Respeaken des Vortrags (dabei wenn möglich die erlernten Interpunktions- und Umformulierungstechniken anwenden. Kürzen und Umformulieren soll zu einer nahezu '1:1' Transkription führen.
  - (b) Beobachten der erstellten Untertitel (Aufmerksamkeit auf Bild und Ton *und* die erzeugten Untertitel lenken). Dabei speziell auf gravierende Erkennungsfehler achten und diese bereits während des Respeakings mit der Hand bzw. einen Stift am Bildschirm zeigen. Dabei wird der Courser nicht bewegt, jedoch die Aufmerksamkeit auf typische schwere Erkennungsfehler gelenkt (Namen, Zahlen).
  - (c) Nach ca. 3 Minuten die Wiedergabe pausieren. Darauf folgend die wichtigsten Gedanken/Inhalte des gehörten Wiedergabeblocks wiedergeben und den letzten Satz in Gedanken weiterführen.
  - (d) Feststellung des NER-Wertes: dabei überprüfen, ob bereits während des Respeakings sämtliche schweren Erkennungsfehler beobachtet wurden.
  - (e) Analyse der Arbeit in Bezug auf Theorie (Umformulieren, Kürzen, etc.) und in wie weit Fehler vermieden werden hätten können.

Übungsaufgaben bis zur nächsten Einheit: Respeaking wird mit Vorlesungsmaterial weiter geübt, um mehr Routine und Erfahrung zu sammeln. Dabei soll speziell auf die praktische Anwendung des *Umformulierens*, des *Kürzens* und der *Umgang mit nicht vorhandenem Vokabular* geachtet werden. Im Vordergrund steht auch das *Beobachten* der Untertitel und das Erkennen (durch zeigen) von Fehlern. In NER-Analysen soll einerseits der tatsächliche NER-Wert festgestellt werden, darüber hinaus zusätzlich jener unter Berücksichtigung der durch hinzeigen erkannten Fehler. Somit kann verglichen werden, wie sehr der NER-Wert bei einer (live) Korrektur der erkannten Fehler verbessert würde. Nicht zuletzt dient die NER-Analyse weiters zur kontinuierlichen Kontrolle der Qualität und soll der auszubildenden Person bei der Selbsteinschätzung helfen. Wie bereits in den letzten Einheiten soll zur Verbesserung der Erkennungsgenauigkeit eine Modelloptimierung durchgeführt werden (Erläuterung erfolgte in Einheit 2).

3.3.7 Einheit 7 - Überprüfen, Zusehen, Korrigieren sowie Aufbereitung von Vorlesungen fürs E-Learning

Ziel: Den praktischen Umgang mit den Softwarekomponenten zur Erzeugung und Aufbereitung der Untertiteln für die E-Learning Plattform Synote zu erlernen. Dies umfasst theoretisches Wissen über Zeitcodes, ihre Erstellung in *Subtitle Workshop* sowie die Aufbereitung einer Vorlesung in Synote.
Matrialien/Quellen: Als Grundlage für den theoretischen Teil über Zeitcodes und Untertitelformate dient der Abschnitt 3.2.1 aus [Hat11, S. 56]. Auf *Subtitle Workshop* wurde der Autor dieser Diplomarbeit durch die Erwähnung dessen Funktionalität in [KDH10] aufmerksam. Die Erläuterungen wurden eigenständig erarbeitet, siehe weiters Abschnitt 3.1.5. In den Übungen während der Einheit sowie in der Einzelarbeit werden die Vorlesungen der Trainingswissenschaft vom 08.05.2012 sowie vom 15.05.2012 verwendet, siehe Kapitel 4. Der Abschnitt 2.2.1 dient als Quelle für die erweiterten Korrekturmöglichkeiten beim Respeaking. Der Abschnitt 3.1.1 beinhaltet die Motivation für die Verwendung von Synote, die Erarbeitung der Erläuterungen wurden basierend auf [Wal10b] erarbeitet.

Inhaltsübersicht/Ablauf:

- Zeitcode und Untertitelformate:
  - Zeitcodes (engl. *timecodes*) sind von Zeitstempeln (engl. *timestamps*) zu unterscheiden.
  - Zeitcodes werden dazu verwendet, die Untertiteleinblendung mit der Video/Audio-Spur zu synchronisieren.
  - Es gibt viele Formate, aber in den Zeitcodes werden zumindest eine eindeutige Nummer, die Start- sowie Endzeit der Einblendung und der Text des Untertitels gespeichert. Optional können zusätzlich weitere Daten wie Meta-, Formatierungs- Informationen, etc. gespeichert werden.

- Subtitle Workshop:
  - Die verwendete Dragon Version erlaubt keinen Export von Zeitcodes. Daher wird zur Untertitelerzeugung direkt in Subtitle Workshop diktiert. Die Software erlaubt das Erstellen von Untertiteln in verschiedenen Formaten und das Wiedergeben von Videos[33] um die Synchronität der Erstellten Untertitel zu erreichen.
  - Die Bedienung von Subtitle Workshop:
    * Benutzeroberfläche (*Video öffnen*, *Neue Untertitel*, *Wiedergabegeschwindigkeit*, etc.).
    * Shortcuts und Workflow: *Neues Item beenden* ('Alt + X') und *Item unterhalb Einfügen* (Einfg.) verwenden um die Zeitcodes mit dem Video zu synchronisieren. Ca. alle 5 bis 30 Sekunden einen Eintrag erstellen.
    * Mittels Doppelklick in die erstellten Untertitel kann zur gewünschten Stelle im Video navigiert werden (z.B. zur Korrektur).
    * Mittels 'Strg+D' können nach dem Erstellen Zeitcode-Einträge verschoben werden (z.B. um 5 Sekunden, um den *Delay* zu korrigieren der beim Respeaken entsteht).
    * Export in verschiedene Formate möglich. Für Synote wird SubRip verwendet. Nachträglich ist die Änderung in die UTF-8 Kodierung notwendig.
    * Der Export in einem „Custom“ Format ermöglicht das Exportieren des Transkripts ohne Zeitstempel (u.a. zur Feststellung der Wortanzahl nützlich, die für die NER-Analyse benötigt wird).
- Übung: in ca. drei Minuten einen Teil eines Vortrags in dem erläuterten Workflow transkribieren. Dabei kommt die in der zweiten Einheit erläuterte Eigenkorrektur zum Einsatz.
- Erweiterte Möglichkeiten für die (individuelle) Erstellung von offline Untertiteln:
  - Variante A (original Geschwindigkeit): Die Untertitel ohne Unterbrechung in gleicher Form wie beim Live Respeaken (z.B. in 20 bis 30 Minuten Blöcken) erzeugen.
  - Variante B (individuelle Geschwindigkeit): Mit der bevorzugten Wiedergabegeschwindigkeit untertiteln (z.B. 90%).
  - Option für Variante A/B: Nach dem Erstellen der Untertitel kann die Aufnahme bei nochmaligen Wiedergeben mit den Untertiteln verglichen und Fehler korrigiert werden. Hierbei kann es nützlich sein, Fehler während der Erstellung mittels Tastatur zu markiert (z.B. Raute für beobachteten Fehler, der später korrigiert werden soll).
  - Option für Variante A/B: Es kann die Aufnahme jederzeit gestoppt werden (Erkennungsfehler sofort korrigieren, die ASR mit neuem Vokabular trainieren, etc.) um die Qualität der Untertitel zu verbessern. Auch das Zurückspulen an gewünschte Stellen ist möglich.

[33] Anm. Autor: Mit dem AVI Format wurde die reibungslose Wiedergabe erprobt.

- Synote:
  - Gründe für die Verwendung von Synote:
    - Die Platform erfüllt speziell die Bedürfnisse von hörbeeinträchtigten Studierenden
    - Sämtliche Materialien (Untertitel, Audio- bzw. Video, Folien, Synote) werden immer synchron wiedergegeben
    - Keine Lizenzkosten für Bildungseinrichtungen
  - Verwendung von Synote:
    - Benutzerkonto erstellen und anmelden
    - Aufnahme einbinden (samt Titel, Beschreibung, Tags, Berechtigungen, etc.)
    - Folien zum online Stellen und Synchronisierungspunkte setzen
    - Die mittels Subtitle Workshop exportierte SubRip Datei einbinden

Abschließende Einzelarbeit/Übungen: In der abschließenden Übung sollen sämtliche praktische Fähigkeiten weiter geübt und gefestigt werden. Das sind im Speziellen das *Beobachten* der Erzeugten Untertitel sowie die in der Einheit erstmals durchgeführte *Eigenkorrektur*. Dabei werden die Zeitcodes mit Subtitle Workshop erstellt und der Umgang mit der Software vertieft, um mehr Routine und Erfahrung zu sammeln. Die auszubildende Person soll abschließend eine individuelle Scripting Technik anwenden bzw. entwickeln, mit der ein NER-Wert von 98% erreicht wird. Dabei kann je nach Präferenz die Wiedergabegeschwindigkeit verringert werden, ggf. gestoppt sowie beobachtete Fehler markiert werden. Wie bereits in den letzten Einheiten soll zur Verbesserung der Erkennungsgenauigkeit eine Modelloptimierung durchgeführt werden (Erläuterung erfolgte in der zweiten Einheit).

KAPITEL 4

Evaluierung

4.1 Aufbau der Evaluierung

In diesem Kapitel sind die gewonnenen Erkenntnisse der Respeaking/Scripting Ausbildung evaluiert und dokumentiert. Die Ausbildung wurde vom Autor dieser Diplomarbeit entworfen und mit *einem* Auszubildenden durchgeführt. Der erarbeitete Trainingsplan bzw. dessen Evaluierung soll den Grundstein für eine Respeaking Ausbildung in Österreich legen. Eingangs ist im Abschnitt 4.2 das verwendete Equipment beurteilt. Im folgenden Abschnitt ab Seite 87 sind die im Kapitel 3 dokumentierten sieben Einheiten diskutiert bzw. evaluiert. Diese Evaluierung soll dazu beitragen, dass das Training zukünftig in ähnlicher Weise von anderen Personen (Ausbildnerinnen und Ausbildnern) durchgeführt und weiter verbessert werden kann. Daher sind die Gründe für den jeweiligen Inhalt und Aufbau bzw. die dahinter stehenden Überlegungen dokumentiert. In die Diskussion der einzelnen Einheiten fließen darüber hinaus die gewonnenen Erfahrungen - basierend auf einem mündlichen Interview mit dem Auszubildenden sowie dem direkten mündlichen Feedback nach bzw. während Einheiten - ein. Das qualitative Interview wurde nach erfolgter Ausbildung bzw. nach der Abschlussübung (dem Erstellung von Untertiteln für eine Vorlesung) durchgeführt. Positive Erfahrungen sind ebenso wie Verbesserungsvorschläge dokumentiert. Die vom Auszubildenden aufgewandte Zeit (Einheiten und Übungsaufgaben) ist schließlich ab Seite 100 im Abschnitt 4.2.9 dokumentiert. Weiters sind Verbesserungsvorschläge an genannter Stelle angeführt.
Im darauf folgenden Abschnitt 4.3 ab Seite 101 sind die Resultate der Evaluierung einer untertitelten Vorlesung aufbereitet. Somit ist abschließend dargelegt, welche Qualität durch die Ausbildung erreicht werden konnte. Hinzu kommen zwei weitere qualitative Vergleiche. Jener zu den live erstellten Untertiteln der Firma Titelbild und jener zu den Untertiteln, die der Autor dieser Diplomarbeit mittels Scripting erstellte.

4.2 Diskussion und Evaluierung der Ausbildung

4.2.1 Equipment

Das im Abschnitt 3.1 ab Seite 52 beschriebene Equipment war aus Sicht des Autors dieser Diplomarbeit für das durchgeführte Training geeignet. So konnten in vielen Fällen die erforderten Erkennungsraten mit der verwendeten Spracherkennung - in Kombination mit dem verwendeten Mikrofon - erreicht werden (siehe Abschnitt 4.3 ab Seite 101).
Beim verwendeten Mikrofon handelt es sich um kein Headset. Aus dem Grund mussten bzw. müssen beim Respeaking zusätzlich Kopfhörer verwendet werden. Das hat den Vorteil, dass ein Respeaker oder eine Respeakerin individuell präferierte Kopfhörer (On-Eear oder In-Eear) verwenden kann.
Aus ergonomischer Sicht könnte ein größerer Monitor, ein höhenverstellbarer Tisch sowie ein geeigneter Sessel eine Verbesserung darstellen.
Sollte das verwendete Mikrofon unter anderen Ausgangsbedingungen verwendet werden, müssen weitere Aspekte berücksichtigt werden. Zum Beispiel, wenn mehrere Personen im selben Raum sprechen (oder respeaken). In diesem Fall muss zuerst evaluiert werden, ob das bei diesem Mikrofon einen negativen Einfluss auf die Erkennungsraten hat. Im Falle eines live Einsatzes mit diesem Mikrofon muss weiters bedacht werden, dass es sich um ein Funkmikrofon handelt. Da-

her muss sichergestellt sein, dass die Batterien nicht während der Tätigkeit gewechselt werden müssen.

4.2.2 Einheit 1

Ziel der Einheit ist es, einen „Überblick über Hörbeeinträchtigung und die dabei wichtigsten Begriffe zu erlangen. Weiters die Arten der Untertitelung und die Grundbegriffe beim Respeaking zu kennen und einen Einblick in die allgemeinen Ziele der Ausbildung zu bekommen. Schließlich soll die Einführung in Dragon samt ersten Diktierübungen die Einheit abrunden und die Basis für die Übungsaufgabe bilden."
Wie auch die folgenden sechs Einheiten besteht die Erste aus theoretischen Inhalten und praktischen Übungen. Wenn möglich und sinnvoll, sind die praktischen Übungen (z.B. Diktierübungen zum Training der Spracherkennung) immer thematisch mit den theoretischen Teilen abgestimmt. Somit kann zusätzlich während der praktischen Übungen theoretisches Wissen vermittelt werden. Die Einheit beinhaltet in etwa 50% Theorie und 50% praktische Übungen. Der Anteil der praktischen Übungen wird in den folgenden Einheiten sukzessive erhöht, da diese (vor allem zu Beginn) rasch stimmlich und geistig ermüdend sind.
Die theoretische Einleitung und Erläuterung der Termini im Zusammenhang mit Hörbeeinträchtigung stellen eine gute Einführung für die Ausbildung dar. Darauf folgt die Erläuterung der Untertitelung - als spezielle Unterstützungsmöglichkeit für hörbeeinträchtigte Menschen. Der folgende praktische Teil, bestehend aus der Einführung in Dragon bzw. das Equipment, bietet bereits in der ersten Einheit die Grundlage für Diktierübungen - auch im Hinblick auf die erste Übungsaufgabe. Das Diktieren eines freien Texts dient nach dem Erstellen des Sprachprofils, dem Sammeln von ersten Erfahrungen und Eindrücken mit Dragon. Dabei waren die Erkennungsraten beim Auszubildenden noch relativ gering. Eine detaillierte Analyse der Probleme ist jedoch noch nicht Teil in dieser Anfangsphase. Vielmehr stehen dabei die ersten Eindrücke mit dem Umgang der Hard- und Software im Vordergrund, um ein Gefühl für das Diktieren entwickeln zu können. Nichtsdestotrotz stellen Ratschläge vom Trainer bzw. der Trainerin eine wichtige Unterstützung in dieser Phase dar, um ggf. Probleme mit Hard- und Software zu besprechen.
Nach dem Initialtraining und ersten Versuchen mit Dragon wird die auszubildende Person mit den wichtigsten Dragon Kommandos sowie der gesamten grafischen Oberfläche vertraut gemacht. Mit diesen Kenntnissen wird die erste Diktierübung durchgeführt. Neben dem Erweitern der Spracherkennung mit neuem Vokabular dient die Übung zum Erlernen des Diktierens. Weiters wird themenbezogenes Fachwissen bezüglich Spracherkennungssystemen vermittelt. Zur Verbesserung der Erkennungsgenauigkeit werden nach dem Diktieren die Erkennungsfehler mit der Korrekturfunktion von Dragon korrigiert. Bei dieser kurzen Diktierübung (244 Wörter inkl. IZ) erzielte die auszubildende Person bereits einen NER-Wert von 99,5%. Die Diktiergeschwindigkeit betrug 88,7 WpM (inkl. IZ). Hier zeigte sich, dass schon während der ersten Einheit die auszubildende Person in der Interaktion mit dem vorhandenen Equipment ausgezeichnete Erkennungsraten erzielen konnte. Diese guten Erkennungsraten sind selbstverständlich als sehr positive und motivierende erste Erfahrungen zu bewerten. Es ist jedoch hervorzuheben, dass es sich hier *nicht um Respeaking* handelte und Erkennungsraten beim Diktieren nicht auf jene beim Respeaking übertragbar sind. Die Gründe dafür sind u.a, dass sämtliche Wörter durch das

vorangegangene Training im Vokabular von Dragon vorhanden waren. Weiters konnte die Diktiergeschwindigkeit frei gewählt werden und im Gegensatz zum Respeaking wurde ein (grammatikalisch korrekter) Text *abgelesen*. Im Hinblick auf die geringeren Worterkennungsraten bei späteren Einheiten wird hiermit darauf hingewiesen, dass gerade durch das gleichzeitige Hören und Sprechen beim Respeaking die Worterkennungsraten aus verschiedenen Gründen (Aussprache, Konzentration, Diktiergeschwindigkeit, Satzbau, etc.) geringer sein können. Aus dem Grund sind die Übungen der folgenden Einheiten als schrittweises Herantasten an die komplexe Respeaking Tätigkeit zu sehen.
Bereits ab der ersten Einheit stellt die Korrektur der Fehler eine wesentliche Rolle in der Ausbildung dar. Dabei wird die Funktion von Dragon genutzt, die das Diktierte der Respeakerin bzw. des Respeakers aufnimmt und die Wiedergabe des Gesagten ermöglicht (*Auswahl wiedergeben*). Die diktierten Wörter und Wortfolgen samt Interpunktionszeichen können so angehört ggf. als Grund von auftretenden Erkennungsfehlern (wie mangelnde Aussprache vs. Problem bei der Spracherkennung) lokalisiert werden. Romero Fresco stellt fest, dass Respeaker bzw. Respeakerinnen bei Erkennungsfehlern oft das Problem bei der Software sehen. Allerdings haben mittlerweile die verwendeten Spracherkennungssysteme eine Qualität erreicht, bei der Erkennungsfehler zum gleichen Teil auf die Respeaker und Respeakerinnen zurückzuführen sind (vgl. [RF11, S: 76]). Daher ist diese Korrekturfunktion aus zwei Aspekten essentiell für die Ausbildung. Einerseits, um eigene Probleme in der Aussprache zu erkennen und sie zukünftig zu vermeiden. Andererseits für das Trainieren und Anpassen der Spracherkennung an die individuelle Aussprache. So sollen lt. dem Dragon Benutzerhandbuch ([S: 67][Nua10]) alle Erkennungsfehler mit der Korrekturfunktion korrigiert/trainiert werden. Damit kann die Genauigkeit der Software kontinuierlich verbessert werden. Aus dem Grund sollen während der gesamten Ausbildungszeit[1] Erkennungsfehler mit der Korrekturfunktion korrigiert werden.
Auch für die Diktierübungsaufgabe ist das Thema inhaltlich für die Ausbildung relevant. Im Gegensatz zum Text aus der Einheit ist dieser Artikel über Spracherkennung und Respeaking bedeutend länger (1180 Wörter inkl. IZ) und enthält eine größere Variation von Interpunktionszeichen (*Bindestrich*, *Klammer auf* und *Klammer zu*, etc.). Somit wird die Schwierigkeit in der Übungsaufgabe auf zwei Ebenen erhöht: Der Länge (Konzentration) und der Verwendung von Interpunktionszeichen. In der Analyse der durchgeführten Übung stellte sich heraus, dass beim Diktieren der Interpunktionszeichen vermehrt Erkennungsfehler auftraten. Trotzdem konnte der Auszubildende einen NER-Wert von 98,0% erzielen und die Diktiergeschwindigkeit auf 90 WpM (inkl. IZ) steigern. Bei der Korrektur konnte die auszubildende Person die Probleme mit den Interpunktionszeichen näher betrachten. Darauf aufbauend konnte er das Diktieren von Satzzeichen weiter üben. Im Feedback nach der Ausbildung merkte der Auszubildende an, dass er es als besonders gut empfand bereits in der ersten Einheit mit der Spracherkennung zu arbeiten.

[1] Anm. Autor: Mit Ausnahme der Einheit 7.

4.2.3 Einheit 2

Ziel dieser Einheit ist es, einen „Einblick in die Arten und Anwendungsgebiete von Spracherkennungssoftwares und in die Funktionsweise von Dragon zu erlangen. Weiters soll Respeaking definiert werden können und dessen Methoden und Einsatzbereiche erlernt werden. Darüber hinaus soll der erweiterte Umgang mit der Hard- und Software trainiert werden, um der auszubildenden Person das selbstständige Durchführen der Trainings-, Diktier-, Korrektur- sowie Shadowing Übungsaufgaben zu ermöglichen. “
Die Einführung in die Geschichte, Funktionsweise und aktuelle Probleme mit Spracherkennungssoftwares stellt den ersten Schritt der Überleitung vom Diktieren zum Respeaking dar. Aufbauend auf dem vermittelten Wissen bezüglich derzeitiger Möglichkeiten und Grenzen von Sprecher und Sprecherinnen unabhängigen Spracherkennungsystemen für spontane Sprache wird die Verbreitung von Dragon erläutert. Der theoretische Teil zum Respeaking wird in der Einheit durch zwei Videos ergänzt. Sie dienen dazu, der auszubildenden Person einen Einblick in die Tätigkeiten beim Respeaking zu geben. Das erste Video (*Respeaking at Swiss TXT- SF Sports*), dass Teil der Einheit ist, zeigt Respeaker und Respeakerinnen bei der Arbeit. Der etwa zehnminütige Videobeitrag dient als Zusammenfassung und Ergänzung zum vorhergegangenen theoretischen Teil über die Respeaking Tätigkeit. Der Beitrag beinhaltet einerseits die Aufgaben beim Respeaking während TV Live-Übertragungen und hebt weiters Probleme und Schwierigkeiten hervor. Darauf aufbauend werden anhand der Definition aus [RF11, S: 1] die verschiedenen Ausprägungen und Anwendungen von Respeaking erläutert. Auf diese Weise soll zukünftigen Respeakern und Respeakerinnen ein Überblick über die Profession vermittelt werden. Der zweite - ebenfalls etwa zehnminütige Videobeitrag (*BBC Item on Respeaking*) - diskutiert Probleme von Fehlern in Untertiteln aus Sicht von hörbeeinträchtigten Menschen. Er dient zur Übung und ist bis zur dritten Einheit von der auszubildenden Person anzusehen. Ein Respeaker erläutert darüber hinaus seine Tätigkeit bei der Live-Untertitelung im TV. Geschichtliche Aspekte werden im Beitrag ebenso erläutert wie alternative Möglichkeiten zur Untertitelerzeugung. Der Auszubildende gab im Feedback nach der Ausbildung an, dass die Videos für ihn besonders interessant waren.
Der praktische Teil der Einheit dient dem Vertiefen der Kenntnisse aus der ersten Einheit. Die detaillierte Beschreibung der Bedienung und Verwendung des Mikrofons/Empfängers soll es der auszubildenden Person ermöglichen, ggf. auftretende Problemen selbstständig zu beheben. Die vom Autor dieser Diplomarbeit entworfene Diktierübung ist speziell für das Erzeugen von Interpunktions- und Sonderzeichen, das Buchstabieren, das Diktieren von Datumsangaben, etc. entworfen. Romero Fresco stellt fest, dass eine effiziente Nutzung der ASR nicht nur durch das Diktieren, sondern sehr oft durch eine Kombination mit manuellen Kommandos (Tastatur und Shortcuts) erzielt werden kann (vgl. [S: 84][RF11]). Es ist daher eine weitere Intention in der Übung, dass zukünftige Respeakerinnen und Respeaker eine individuelle Technik entwickeln. Diese kann die Verwendung der Tastatur berücksichtigt. Der Auszubildende hat beispielsweise im Zuge dieser Übung festgestellt, dass bei ihm bei Telefonnummern und E-Mail-Adressen viele Erkennungsfehler auftraten, die er effizient mittels Tastatureingabe vermeiden kann.
Die Übungsaufgabe bis zur nächsten Einheit besteht aus zwei Teilen. Der erste Teil ist der ersten Phase des Respeakings - der Vorbereitung - zuzuordnen und besteht aus dem Verbessern der Erkennungsgenauigkeit. Dies erfolgt durch das Lesen eines von Dragon zur Verfügung gestellten

Textes. Der zweite Teil der Übungsaufgabe dient zur Vorbereitung zur zweiten Respeakingphase - der Untertitelerstellung. Diese Vorbereitung erfolgt mit einer so genannten Shadowing Übung. Diese durchaus kontrovers diskutierte Übungsform ist im folgenden Abschnitt näher erläutert.

Shadowing und Dual-task Training (Doppelleistungsaufgaben)

Ein wesentliches Merkmal beim Simultandolmetschen ist das Sprechen[2] und gleichzeitig zuhören der redenden Person. Diese Fertigkeit - die nur ein Teil der komplexen kognitiven Tätigkeit beim Simultandolmetschen darstellt - ist eine, die „im Rahmen der Dolmetschausbildung geübt werden soll“ [Kur96, S: 101] und nicht unbedingt vorauszusetzen ist (vgl. [Kur96, S: 100-106]). Auch beim Respeaking stellt das gleichzeitige Hören und Sprechen eine wesentliche Fertigkeit dar, die erlernt und demnach in der Ausbildung geübt werden muss. Dabei stellt sich die Frage, mit welchen Übungen diese Fähigkeit effizient und gut erlernt werden kann.
„Unter ’Shadowing’ versteht man das Anhören und gleichzeitige Nachsprechen von akustisch dargebotenem Material in der Ausgangssprache - ein ’Mitsprechen’ gewissermaßen. Es ist eine häufig verwendete Methode zur Untersuchung der selektiven Aufmerksamkeit in der kognitiven Psychologie“ [Kur96, S: 102]. Dabei kann Shadowing noch detaillierter unterschieden werden. Das „’Phonenmic Shadowing’ ist das sofortige Nachsprechen ohne Abwarten der vollständigen Präsentation einer Sinneseinheit“ [Kur96, S: 101] wo im Gegensatz zum ’Phrase shadowing’ versucht wird, dem Redner oder der Rednerin „dicht auf den Fersen“ [Kur96, S: 101] zu bleiben. Beim genannten ’Phrase shadowing’ wird hingegen versucht einen zeitlichen Abstand (eine Sinneseinheit oder ein Chunk) zwischen der Rednerin bzw. dem Redner und dem Wiederholen/Nachsprechen zu haben. Ein wesentlicher Kritikpunkt vom Shadowing ist, dass dabei nicht (zwingend) dem Inhalt des Gesagten gefolgt wird. So hält Kurz fest, dass das Erfassen des Sinns des Gehörten das oberste Gebot ist (vgl. [Kur96, S: 100-106]): „Übungen, die nur das mechanische Wiederholen von akustischem Material ohne Sinnerfassung beinhalten, erscheinen demnach nicht empfehlenswert“ [Kur96, S: 102].
Der Autor dieser Diplomarbeit entschloss sich trotz der berechtigten Kritikpunkte dazu, Shadowing einzusetzen. Um der Problematik des ’blinden’ Nachsprechens entgegenzuwirken, sind zukünftige Respeaker und Respeakerinnen dazu angehalten, während des Shadowings bewusst dem Inhalt zu folgen. Durch das anschließende Wiederholen des Gesagten in eigenen Worten kann festgestellt werden, wie detailliert und inhaltlich korrekt die auszubildende Person den Sinn des Gehörten erfassen konnte. In der Shadowing Übung der Einheit 3 sind weiters gezielte Fragen über den Inhalt - Entscheidungsfrage (ja/nein) und Denkfragen (Warum-Fragen) - Teil der Übung. Ein wesentlicher Grund für den Einsatz von Shadowing im Zuge der erarbeiteten Scripting Ausbildung ist das Üben vom gleichzeitigen Sprechen/Diktieren und Hören. Da bei den erarbeiteten Shadowing Übungen die auszubildende Person keine Spracherkennung verwendet, kann sie sich alleinig auf das gleichzeitige Sprechen/Diktieren konzentrieren und muss somit weder auf die korrekte Bedienung des Mikrofons noch auf die Ausgabe der Spracherkennung achten. Im Feedback nach der Ausbildung hob der Auszubildende (R2) diesen Aspekt als sehr positiv hervor. Wie im Abschnitt 2.5.1 ab Seite 33 erläutert, ist bei der Verwendung von Spracherkennungssystemen wie Dragon die Diktierweise und Aussprache entscheidend für

[2] Anm. Autor: Mit Ausnahme des Simultandolmetschens einer Gebärdensprache.

gute Erkennungsraten. Wie in diesem Kapitel angeführt, erreichte der Auszubildende bereits nach zwei Einheiten hohe Erkennungsraten beim Diktieren (ablesen von Texten) der Spracherkennung. Jedoch kann sich die Aussprache beim gleichzeitigen Hören verändern, was negative Auswirkungen auf die Erkennungsraten hätte. Die Kontrolle der Aussprache beim gleichzeitigen Hören ist somit ein weiterer Grund für den Einsatz von Shadowing. In der Einheit 3 ist das Aufzeichnen der Shadowing Übung essentiell. Dabei können beim anschließenden Analysieren der Aufnahme mögliche Veränderungen der Aussprache festgestellt werden. Zukünftigen Respeaker und Respeakerinnen können somit bereits vor der ersten Respeaking Übung das (in den ersten beiden Einheiten erlernte) Diktieren auch beim gleichzeitigen Hören üben. Dies soll schließlich zu hohen Erkennungsraten bei den ersten Respeaking Übungen führen. Dieser Ansatz ist an [ARRF08] angelehnt, wo u.a. Shadowing als Vorbereitungsübung zum Respeaking am Beginn der Ausbildung empfohlen wird.
Ein weiterer Grund für den Einsatz von Shadowing am Beginn der Ausbildung ist das Erhöhen der Konzentrationsspanne. Nachdem die auszubildende Person im 'Phonenmic Shadowing' geübt ist, folgt der Übergang zum 'Phrase Shadowing'. Im Gegensatz zur 'Phonenmic Shadowing' Übung startet diese anstatt des lauten Nachsprechens mit rein gedanklichem Wiederholen des Gesagten, allerdings zusätzlich mit dem 'gedanklichen Diktieren' der Interpunktionszeichen. Diese Übung soll den Übergang zum Respeaking (Einheit 4) erleichtern. Das Hinzufügen der Interpunktionszeichen führt automatisch dazu, dass der bzw. die Auszubildende den Abstand zum Gesagten variieren muss. Schließlich wird die Übung zusätzlich mit lauten Nachsprechen (allerdings nach wie vor ohne das Diktieren in die Spracherkennung) durchgeführt und der Abstand zum Gesagten sukzessiv gesteigert. Somit kann die Konzentrationsspanne erhöht werden. Dies ist später beim Respeaking wichtig, wo beim Diktieren ein gewisser Abstand zum Gesagten gehalten werden muss (um ggf. umformulieren/kürzen zu können oder um Fehler zu korrigieren).

4.2.4 Einheit 3

Ziel der Einheit sind „weitere Shadowing- sowie Diktierübungen (ohne ASR) zum Üben und Feststellen der Aussprache und des Sprechrhythmus. Darüber hinaus eine detaillierte Erläuterung der (teilweise) parallel ablaufenden Respeaking Tätigkeiten. Eine Übung soll die Respeaking-Vorbereitungstätigkeiten festigen und eine weitere Übung das Verständnis, das Diktieren und die Analyse verbessern."
Der Auszubildende (R2) führte als Einzelarbeit der zweiten Einheit für insgesamt 75 Minuten die beschriebenen vier Shadowing Übungen[3] durch. Dabei wählte er einen Beitrag der ORF Kinder- und Jugendserie 'Miniversum'. Die Sprechgeschwindigkeiten betrugen dabei ca. 110 WpM ohne IZ bzw. 125 WpM mit Interpunktionszeichen. Der Auszubildende stellte fest, dass er sich während des Shadowings zunehmend mehr auf die Aussprache konzentrieren und sie im Weiteren auch verbessern konnte. Probleme hatte er kurzfristig bei höheren Sprechgeschwindigkeiten der Sprecherin. Diese konnten aber lt. eigenen Angaben nach einigen Übungen beseitigt

[3] Anm. Autor: ohne Verzögerung; gedanklich mit Verzögerung jedoch ohne lautes Sprechen; gedanklich mit Verzögerung samt Interpunktionszeichen; sowie gleichzeitiges Hören und Sprechen samt Interpunktionszeichen und Verzögerung, siehe Abschnitt 3.3.2 ab Seite 66.

werden. Nach den ersten drei durchgeführten Shadowing Übungen fiel ihm das Shadowing samt dem Sprechen der Interpunktionszeichen leicht. Da der Auszubildende durch diese Übungen das gleichzeitige Sprechen (samt IZ) und Hören erlernte, erfüllten die Übungen aus Sicht des Autors dieser Diplomarbeit ihren Zweck.
Die Einheit 3 beinhaltet die Wiederholung der Shadowing Übungen mit einem ZiB Beitrag. Die Sprechgeschwindigkeit dieser Nachrichtensendung liegt mit ca. 145 WpM ohne IZ und 180 WpM inklusive Interpunktionszeichen und somit deutlich über jener des Miniversums. Aufgrund der höheren Geschwindigkeit hatte der Auszubildende vor allem beim 'Phrase shadowing' (Verzögerung von einem halben bis einem Satz) Probleme. Diese traten auf, wenn im Beitrag keine (kurzen) Sprechpausen (wie z.B. zwischen Sätzen) vorkamen. So erhöhte sich der Abstand zum Gesprochenen laufend, bis der Auszubildende schließlich 'den Faden verlor' und das Gesagte nicht mehr wiedergeben konnte. Die Steigerung der Konzentrationsspanne und der Sprechgeschwindigkeit ist nicht Teil dieser Übung. Deshalb wurde schließlich die Wiedergabegeschwindigkeit so gewählt, dass die Sprechgeschwindigkeit in etwa gleich wie jene der Miniversum Sendungen war. Wie im vorigen Abschnitt erläutert, sind bei der Shadowing Übung dieser Einheit gezielte Fragen über den Inhalt ein essentieller Bestandteil der Übung. Damit soll der Problematik des 'blinden' Nachsprechens entgegengewirkt werden. Es zeigte sich hierbei, dass der Auszubildende dem Inhalt der Nachrichtensendung sehr gut folgen konnte und weder bei den Fragen noch beim Wiedergeben des Inhalts in eigenen Worten Probleme hatte. Die Analyse der Audioaufzeichnung diente abschließend dem gemeinsamen Identifizieren von Ausspracheproblemen. Diese Evaluierung stellte sich aus Sicht des Autors dieser Diplomarbeit als sehr wichtig heraus. So konnten zusammen mit dem Auszubildenden einige Schwächen in der Aussprache lokalisiert werden. Beispielsweise bei kurzen Wörtern wie 'sich' und 'mich'. Das Identifizieren dieser Probleme stellt schließlich den Abschluss der Shadowing Übungen dar und dient dem Übergang zu den folgenden Diktierübungen bzw. den Respeaking Übungen der nächsten Einheit(en).
Nach den praktischen Shadowing Übungen folgt ein theoretischer Teil. Dabei werden die parallel ablaufenden Tätigkeiten beim Respeaking erläutert und auf jene eingegangen, die nicht parallel durchgeführt werden können (Korrigieren und Diktieren). Diese Aspekte dienen als Vorbereitung für die erste Respeaking Übung in der Folgeeinheit. Anschließend werden in einer praktischen Übung die Tätigkeiten der (auftragsbezogenen) Vorbereitung erlernt. Wie im Abschnitt 2.2.1 ab Seite 21 erläutert, stellt die Vorbereitungsphase einen wichtigen Bestandteil der Respeaking Tätigkeit dar. In der Praxis werden allerdings den Respeakern und Respeakerinnen nicht immer genügend zeitliche Ressourcen für die Vorbereitung zugesprochen. Gerade deshalb ist es aus Sicht des Autors dieser Diplomarbeit essentiell, dass das professionelle Erlernen der Vorbereitung ein fester Bestandteil einer Respeakingausbildung ist. Damit sollen zukünftige Respeakerinnen und Respeaker den Nutzen dieser Vorbereitung erkennen, sie effizient durchführen können und bei Arbeitgebern und Arbeitgeberinnen im besten Fall erfolgreich die wichtigen zeitlichen Ressourcen dafür einfordern können. In dieser Einheit wird für die Vorbereitung eines Themas ein Zeitraum von 20 Minuten veranschlagt. Aktuelle Themen[4] (über die neben dem

[4] Anm. Autor: Beim Videomaterial handelte es sich um verschiedene Beiträge der ZiB, die sich der Verhandlung des Angeklagten Anders Behring Breivik widmeten. Der Norweger musste sich des Todes von 77 Menschen verantworten, die seinen Anschläge in Norwegen 2011 zum Opfer vielen.

ORF auch Print- und Onlinemedien berichten) eigenen sich aus folgenden Gründen für die Verwendung der Einheit(en): Die auszubildende Person recherchiert das zu erwartende Vokabular einmalig und verbessert die ASR für einen Themenbereich. Aufgrund der vielen Medienberichte kann für das Training der Spracherkennung aus einer Vielzahl von Quellen gewählt werden. Weiters sind zu Nachrichtenbeiträgen des ORFs die Untertitel des Senders verfügbar. Sie werden in einer weiteren Übung zum Vergleich der von der auszubildenden Person erstellten Untertitel herangezogen. Hinzu kommt, dass der Sprecher oder die Sprecherin bei Nachrichtensendungen keinen freien Text spricht und das Umformulieren (welches erst Teil in späteren Einheiten ist) noch nicht zwingend notwendig ist. In der Übung suchte der Auszubildende Zeitungsartikel aus dem Internet und erstellte in der ersten Hälfte der veranschlagten Zeit eine Wortliste (Ortsnamen, Name des Täters, etc.) und trainierte damit die Spracherkennung. Die verbleibenden Zeit begann er mit dem Training von Dragon anhand des (gesamten) Wikipedia Artikels. Die Zeit reichte allerdings nicht zum Trainieren sämtlicher Wörter. Nach der Einheit kam er zur Erkenntnis, dass er zukünftig mehr Zeit in Wortlisten verwenden wollte und lange Artikel mit einer großen Anzahl neuer Wörter bei kurzen Vorbereitungszeiten vermeiden würde. Er stellte die Überlegung an, neben der Wortliste ausschließlich die Einleitung von Wikipedia Artikeln zu verwenden.
Darauf aufbauend folgt das Üben von wichtigen Respeaking Skills: Dem Verständnis, dem Diktieren und der Analyse. Dabei handelt es sich jedoch noch nicht um Respeaking sondern um eine schrittweise Annäherung an das Respeaking bzw. den Tätigkeiten beim Respeaking. Dabei liegt der Fokus beim Zuhören und Wiedergeben bzw. Diktieren sowie der Steigerung der Konzentrationsspanne. Letzteres betrug beim verwendeten Nachrichtenbeitrag in etwa acht Sekunden. Die festgestellte Konzentrationsspanne soll in der Übungsaufgabe bis zur nächsten Einheit gefestigt und kontinuierlich erhöht werden.

4.2.5 Einheit 4

Ziel dieser Einheit ist die „Feststellung der *angenehmen*, *maximalen* sowie *optimalen Diktiergeschwindigkeit* unter Berücksichtigung der erläuterten und zu messenden Wortfehlerraten bzw. Worterkennungsraten. Darauf aufbauend soll eine erste Respeakingübung die Einheit abrunden und gemeinsam mit der Diktierübung die Basiskenntnisse für die Übungsaufgabe bilden."
Die Einheit beginnt mit einem Theorieteil über das Messen der Erkennungsgenauigkeit (Akkuratheit) bzw. Fehlerhäufigkeit von Untertiteln. Aufgrund der großen Verbreitung von WRR bzw. WER sollen zukünftige Respeaker und Respeakerinnen diese Berechnungsmethode kennen und anwenden können. Es werden die Schwächen dieser Berechnungsmethoden - vor allem im Hinblick auf gekürzte bzw. umformulierte Untertitel, wie sie beim Respeaking meist vorkommen - verdeutlicht und aufgezeigt. Aufbauend auf dieser Problematik wird schließlich in der nächsten Einheit das NER-Modell erläutert. Als erste praktische Übung der vierten Einheit dient die Wiederholung der Verständnis-, Diktier- und Analyseübung aus der Einheit 3 bzw. aus der Übungsaufgabe zwischen dritter und vierter Einheit. Durch die Übungsaufgabe konnte der Auszubildende (R2) seine Konzentrationsspanne von ca. acht Sekunden auf zehn bis zwölf Sekunden steigern. Dabei diktierte er das Gesagte nicht Wort für Wort sondern formulierte immer wieder Passagen um. Das Umformulieren während der Übung ist aus Sicht des Autors dieser Diplomarbeit als positiv zu werten, da es das Verständnis des Gesprochenen ebenso voraussetzt wie Wissen über den Umgang mit der Spracherkennung. Probleme stellten allerdings Passagen

mit mehreren Zahlen und Prozentangaben innerhalb kurzer Zeit (z.B. bei drei Zahlenangaben innerhalb von zwei Sätzen) dar. Da die Diktiergeschwindigkeit in dieser Verständnis-, Diktier- und Analyseübung frei gewählt werden kann, stellt die Folgeübung zum Feststellen der *angenehmen*, *maximalen* und *optimalen* Diktiergeschwindigkeit eine wichtige Brücke zum Respeaking dar.
Wie viel Information in der Praxis in den Untertiteln transportiert werden kann, hängt u.a. von der individuellen (maximalen) Arbeitsgeschwindigkeit eines Respeakers bzw. einer Respeakerin ab (vgl. [RF12b]). Faktoren die diese Arbeitsgeschwindigkeit beeinflussen sind die Aussprache, der Sprechrhythmus bzw. im Allgemeinen der Umgang mit der Spracherkennung bei verschiedenen Diktiergeschwindigkeiten. Hohe Erkennungsgenauigkeit kann - bis zu einem gewissen Maß - durch gezieltes Training bei hohen Diktiergeschwindigkeiten erreicht werden. Dennoch ist zu berücksichtigen, dass menschliche Grenzen ebenso existieren wie Grenzen bei Spracherkennungssystemen. Die Übung zum Feststellen der *angenehmen*, *maximalen* und *optimalen* Diktiergeschwindigkeit ermöglicht der auszubildenden Person das Feststellen der (derzeitigen) eigenen Möglichkeiten und Limits. Weiters dient die Übung der Verbesserung der Aussprache beim Diktieren und der schrittweisen Steigerung der Diktiergeschwindigkeit(en) unter Beachtung hoher Erkennungsraten. Die vom Auszubildenden als *angenehm* empfundene Diktiergeschwindigkeit betrug in der Einheit 107 WpM inklusive der zu diktierenden Interpunktionszeichen. Er erreichte dabei eine WRR von 96,7%. In einer weiteren Iteration sollte die *maximale* Diktiergeschwindigkeit festgestellt werden. Dabei diktierte er mit 152,5 WpM (inkl. IZ), was allerdings zu hohen qualitativen Einbußen führte (WRR von 90,1%). In der letzten Iteration der Übung wurde die *optimale* Diktiergeschwindigkeit mit 129,2 WpM (inkl. IZ) festgestellt. Dabei konnte eine WRR von 98,9% erreicht werden. Interessant ist, dass bei 129,2 WpM eine höhere WRR als bei 107 WpM erreicht wurde. Dies verdeutlicht einerseits, dass mit Dragon bei niedrigen Diktiergeschwindigkeiten nicht zwingend hohe Erkennungsraten erzielt werden. Und andererseits zeigt die Auswertung, dass bei hohen Diktiergeschwindigkeiten auch gute Erkennungsraten möglich sind. Im speziellen Fall dieser Übung trägt sicherlich auch das mehrfache Lesen des gleichen Textes zu diesem Phänomen bei. Einerseits ist der auszubildenden Person der Text bereits bekannt und somit kann gezielt auf die Aussprache geachtet und vorangegangene Fehler vermieden werden. Aus persönlichen Erfahrungen des Autors dieser Diplomarbeit liefert die Spracherkennung Dragon zum anderen beim Diktieren von thematisch ähnlichen (oder wie in diesem Fall einem identischen Text) bessere Resultate. Das ist jedoch Ziel der Übung. Es wird festgestellt, welche Erkennungsraten die auszubildende Person zu diesem Zeitpunkt des Trainings im besten Fall erreichen kann. Darüber hinaus können vorhandene Fehler während des Diktierens festgestellt und zukünftig vermieden werden. Aus Zeitgründen wurde auf das Feststellen der maximalen Diktiergeschwindigkeit (unter Berücksichtigung einer WRR von $\geq 95\%$) verzichtet. Das erfolgte in der zu dieser vierten Einheit definierten Übungsaufgabe mit einem zweiten, längerem Text.

4.2.6 Einheit 5

Ziel der Einheit ist „fundiertes Wissen über Hintergründe sowie Anwendung des NER-Models (das zur Messung der Qualität von (Live) Untertitel verwendet wird) zu erlangen. Nach einer Übung zur Interpunktion beim Respeaking steht die praktische Anwendung vom NER-Modell im Vordergrund."
Aufbauend auf der Einheit 4, in der u.a. die Messung der Erkennungsgenauigkeit (Akkuratheit) bzw. Fehlerhäufigkeit von Untertitel mit der weit verbreiteten WRR bzw. WER erläutert wird, beginnt diese Einheit mit dem Theorieteil des NER-Modells. Beispiele stellen die jeweiligen Fehlerklassen und Fehlerursachen dar und bilden die Basis für die praktische Anwendung. Durch die Erläuterung des Modells sollen darüber hinaus zukünftige Respeaker und Respeakerinnen die Relevanz sowie die Vorteile vom NER-Modell für ihre Tätigkeit erkennen. Dies betrifft vor allem die Möglichkeit zur laufenden Verbesserung der eigenen Arbeit.
Im Anschluss an den theoretischen Teil folgt eine Respeaking Übung und die Evaluierung des erstellten Transkripts mit dem NER-Modell. Dadurch können eventuelle Unklarheiten in der Anwendung des Modells - wie z.B. die Einteilung in die drei Fehlerklassen - noch während der Einheit besprochen werden. Die in der Übungsaufgabe der vierten Einheit vom Auszubildenden selbst festgestellten, *optimalen* Diktiergeschwindigkeit ist mit 128 WpM (inkl. IZ) nahezu gleich jener, die während der Einheit 4 festgestellt wurde. Mit dieser Geschwindigkeit erreichte er - wohlgemerkt nicht beim Respeaking sondern beim Ablesen eines Textes - die angestrebte WRR von $\geq 95\%$. Bei der Respeaking Übung (Untertitelung von ZiB Beiträgen) in der Einheit 5 begann er anfangs mit einer Diktiergeschwindigkeit von ca. 110 WpM (inkl. IZ) und steigerte sie schließlich bis zu seiner momentanen, optimalen Diktiergeschwindigkeit. Auffällig war dabei, dass der Auszubildende beim Respeaking zur besseren Konzentration die Augen verschloss. Zu diesem Zeitpunkt der Ausbildung ist das Beobachten der Untertitel noch kein Bestandteil der Respeaking Tätigkeit, da eine Fehlerkorrektur noch nicht durchgeführt wird. Im Hinblick auf die kommenden Einheiten wurde jedoch der Auszubildende darauf hingewiesen, dass es zukünftig während des Respeakings wichtig sein wird, visuelle Informationen (Ausgabe der Spracherkennung, Vortragsfolien, Videoaufzeichnung, etc.) aufnehmen zu können. Bei der im Anschluss an die Respeaking Übung durchgeführte Analyse mit dem NER-Modell wurden noch kleinere Unklarheiten bezüglich der Fehlerklassen und Fehlerursachen geklärt. Anschließend folgte eine Einführung in die Arbeitsweise mit einer vom Autor dieser Diplomarbeit erstellten Excel-Datei. Die Datei ermöglicht eine schnelle Berechnung von NER-Werten und dient weiters zur Dokumentation.
Der dritte Fokus der Einheit liegt erneut bei der Interpunktion und beginnt mit einem theoretischen Teil. Dabei wird hervorgehoben, dass aufgrund des zusätzlichen Diktierens von Interpunktionszeichen in der Praxis beim Respeaking (auch bei geringen Sprechgeschwindigkeiten) meist keine Wort für Wort Untertitelung erreicht wird. Darauf folgend wird erläutert, dass Dragon das Diktierte in Blöcken (engl. *chunks*) transkribiert. Somit kann die Verzögerung durch das Aufteilen von langen Sätzen in mehrere kurze Satzblöcke verringert werden. Das ist nicht zuletzt für die Folgeeinheiten relevant, wo bei einer geringeren Verzögerung das Beobachten sowie das Korrigieren von Fehlern leichter ist. Kurze Satzblöcke können darüber hinaus die Lesbarkeit von Untertitel erhöhen. Diese Aspekte stehen allerdings im Konflikt mit der Erkennungsgenauigkeit, da Dragon bei langen Sätzen bessere Erkennungsraten erzielt. Dieses Hintergrundwissen soll der

auszubildenden Person innerhalb der weiteren Ausbildungszeit beim Abwägen zwischen kurzen und langen Satzblöcken helfen, um ein geeignetes Mittel zu finden.
In den vorangegangenen Shadowing und Respeaking Übungen wurde mit TV Material des ORFs gearbeitet. Die gesprochenen Beiträge basierten dabei in der Regel auf redaktionell vorbereiteten Texten. Eine Alltagssprache kam in wenigen Fällen (u.a. bei Interviews) vor. Die in universitären Vorlesungen gesprochene Sprache unterscheidet sich (zum Teil stark) von der geschriebenen Form. Hinzu kommt, dass das verwendete Vokabular fachspezifischer als das der bisher verwendeten Beiträge ist. Daraus resultierend entstehen andere Herausforderung an eine Respeakerin oder einen Respeaker. Da das Ausbildungsziel das Erstellen von Untertiteln einer Vorlesung ist, wird ab dieser Stelle der Ausbildung ausschließlich mit Vorlesungsmaterialien gearbeitet. Somit wird während der Ausbildung gezielt auch mit gesprochener, spontaner Sprache trainiert.
In der letzten praktischen Übung der fünften Einheit müssen fehlenden Satzzeichen in ein vorbereitetes '1:1' Transkript einer Vorlesung eingefügt werden. Dabei soll auf eine gute Lesbarkeit der Untertitel ebenso geachtet werden wie auf die Länge der Satzblöcke. Ohne Zeitdruck steht bei dieser Übung die intensive Auseinandersetzung mit Satzzeichen im Vordergrund, was zukünftig das 'intuitive' Setzen während des Respeakings erleichtern soll.
In der Übung bis zur nächsten Einheit erfolgt das weitere Festigen sowie Verbessern der Respeaking Fähigkeiten. Für die Analyse wird von der auszubildenden Person erstmals eigenständig das NER-Modell angewandt.

4.2.7 Einheit 6

Das Ziel der Einheit ist „theoretisches Wissen zum Décalage sowie zum Editieren der Untertitel (Umformulierung, Kürzen und dem Umgang mit nicht vorhandenen Vokabular) zu erwerben. In weiteren Respeaking Übungen wird das Editieren sowie das aktive Beobachten der erzeugten Untertitel (zum Erkennen von Fehlern) trainiert."
Im Folgenden wird auf die Ergebnisse der Übungsaufgabe der fünften Einheit eingegangen. Der Auszubildende diktierte mit einer Geschwindigkeit von 88,1 WpM (inkl. Interpunktionszeichen). Er analysierte das mittels Respeaking erzeugte Transkript und stellte einen NER-Wert von 97,8% fest[5]. Durch die Fehleranalyse setzte sich der Auszubildende intensiv mit den verschiedenen Fehlerklassen und Ursachen auseinander und konnte deren unterschiedliche Auswirkung auf die Qualität erkennen. Wie bereits in der Evaluierung der Einheit 5 angeführt, verschloss der Auszubildende bei den vorangegangenen Respeaking Übungen aus Konzentrationsgründen die Augen. Das Beobachten dient jedoch als erster Vorbereitungsschritt zur live Korrektur von Fehlern und ist daher ein wichtiger Bestandteil der Respeaking Tätigkeit. Resultierend aus dem Hinweis beobachtete R2 lt. eigenen Angaben (zumindest großteils) die Spracherkennungsausgabe sowie die Videoaufzeichnung in der Übungsaufgabe.
Die sechste Einheit beginnt mit einem theoretischen Teil über *Décalage* - der Verzögerung zur Originalquelle, mit der ein Respeaker oder eine Respeakerin zu diktieren beginnt. Ein solcher Abstand ist für das Umformulieren, das Kürzen aber auch zum Korrigieren von Fehlern notwendig. Eine Intention der Shadowing Übungen aus vorangegangen Einheiten ist daher das Üben des Diktierens mit variierendem Abstand und die Steigerung der Konzentrationsspanne.

[5] Anm. Autor: Die Angaben basieren auf einer analysierten Stichprobe von ca. 25 Minuten.

Der Theorieteil zum Themenbereich Décalage soll die Erfahrungen der vorangegangenen praktischen Übungen ergänzen. Dabei werden der auszubildenden Person zwei Décalage Ansätze erläutert: Beim ersten beginnt eine Respeakerin oder ein Respeaker zu Diktieren, sobald eine Sinneseinheit gesprochen wurde. Beim zweiten Ansatz wird mit dem Diktieren erst dann begonnen, wenn genügend Information vorhanden ist um das Ende des Satzes sicher zu wissen. Während dieser Einheit werden die jeweiligen Vor- und Nachteile der unterschiedlichen Herangehensweisen erläutert. Zusammen mit der auszubildenden Person wird festgestellt, welche Décalage Technik bei den vorangegangenen Respeaking Übungen intuitiv angewandt wurde. Aus Trainingszwecken soll bei den weiteren Respeaking Übungen der bisher nicht praktizierte Ansatz angewandt werden. Somit soll es zukünftigen Respeakerinnen und Respeakern in der Praxis möglich sein, auf verschiedene Anforderung flexibel zu reagieren. Der Auszubildenden R2 wandte bis zur sechsten Einheit hauptsächlich den ersten Ansatz an, den er als angenehmer bezeichnete.
Im Anschluss folgt ein theoretischer Teil zum Umformulieren, Kürzen und dem Umgang mit nicht vorhandenem Vokabular. Dabei wird auf den Unterschied zwischen Umformulieren und Kürzen eingegangen und verschiedene Taktiken beim Umgang mit unbekanntem Vokabular erläutert (Paraphrasieren sowie Tippen mit der Tastatur). Diese theoretischen Grundlagen ermöglichen zusammen mit der Respeaking Übungsaufgabe aus der vorigen Einheit 5 die Überleitung zum praktischen Teil der sechsten Einheit. Dabei wird das Umformulieren, das Kürzen und auch der Umgang mit nicht vorhandenem Vokabular geübt. Weiters wird während der Übung das Beobachten der erzeugten Untertitel trainiert. Erkannte Fehler werden von der auszubildenden Person mit dem Finger (bzw. alternativ mit einem Stift) am Bildschirm angezeigt. Dabei sollen Passagen beobachtet werden, in denen eine hohe Fehlerhäufigkeit vermutet wird (z.B. Zahlen, Wörter die nicht deutlich diktiert wurden, etc.). Obwohl vorerst das Erkennen von gravierenden Fehlern im Vordergrund steht, ist es aus Sicht des Autors dieser Diplomarbeit anzustreben, sämtliche Fehler zu erkennen. Über eine (live) Korrektur kann dann eine Respeakerin oder ein Respeaker situationsabhängig (Zeitfaktor, schwere des Fehlers, etc.) entscheiden.
In der Übung fand es der Auszubildende bei hohen Geschwindigkeiten leichter mit einem geringeren Abstand zum Gesprochenen zu Diktieren (erster Décalage Ansatz). Dabei kürzte er vor allem inhaltlich nicht relevante Stellen, wie z.B. Füllwörter und redundante Formulierungen. Weiters formulierte er vermehrt jene Sätze des Vortragenden um, die dieser nicht zu Ende gesprochen hatte. Das führte aber teilweise zu Konzentrationsschwierigkeiten und in Folge zu erhöhten Fehlerraten sowie einem höheren Abstand zum Gesagten. Um den Abstand wieder zu verringern, kürzte er auch inhaltlich. Bei Redewendungen im Dialekt wandte er den zweiten Décalage Ansatz an. Damit konnte er den Inhalt grammatikalisch und inhaltlich korrekt diktieren. Bei Dialektwörtern und Eigennamen (Nachnamen von Professoren, LVA-Namen an der Universität Wien) praktizierte der Auszubildende bereits die in dieser Einheit vorgestellten Paraphrasiertechniken. Während der Übung diktierte er mit 69,6 WpM (inkl. IZ) und erreichte (ohne Fehlerkorrektur) einen NER-Wert von 96,1%. Er erkannte bzw. beobachtete dabei zwei der drei gravierenden Fehler.
Als Übungsaufgabe dient die Wiederholung dieser Übung mit dem Fokus auf die Steigerung der Arbeitsgeschwindigkeit sowie der Qualität. Dabei wird zusätzlich jener NER-Wert errechnet, der sich aus einer (erfolgreichen) Korrektur der beobachten Fehler ergeben würde.

4.2.8 Einheit 7

Ziel der Einheit ist es, den „praktischen Umgang mit den Softwarekomponenten zur Erzeugung und Aufbereitung der Untertiteln für die E-Learning Plattform Synote zu erlernen. Dies umfasst theoretisches Wissen über Zeitcodes, ihre Erstellung in *Subtitle Workshop* sowie die Aufbereitung einer Vorlesung in Synote."
Die Zielsetzung der durchgeführten Respeaking Übungsaufgabe der sechsten Einheit ist die Steigerungen der Arbeitsgeschwindigkeit und der Untertitelqualität. Der Auszubildende R2 erhöhte seine Respeaking Diktiergeschwindigkeit von 69,6[6] auf 72,5 WpM (inkl. IZ). Hervorzuheben ist die qualitative Verbesserung durch die Übungsaufgabe. So lag der NER-Wert (ohne Korrektur) während der Einheit 6 mit 96,1% deutlich unter dem Ausbildungsziel von $\geq 98,0\%$. Gegen Ende der Übungsaufgabe erreichte er einen NER-Wert von 98,1% (ohne Korrektur). Dabei erkannte bzw. beobachtete er während des Respeakings zwei der neun normalen sowie zwei der sieben geringfügigen Fehler. Nicht beobachtet hatte er allerdings den einzigen gravierenden Fehler ('Blitztraining' anstatt 'Splittraining')[7]. Bei einer erfolgreichen Korrektur der vier beobachteten Fehler würde der NER-Wert mit 98,5% noch deutlicher über dem Ausbildungsziel von $\geq 98,0\%$ liegen. Somit konnte der Auszubildende zu diesem Zeitpunkt bereits vier der fünf Respeaking Tätigkeiten (die während der Phase der Untertitelerstellung ausgeführt werden) in hoher Qualität praktizierten: das *Zusehen*, das *Zuhören*, das *Diktieren* sowie das *Lesen*. Jedoch stellte zu diesem Zeitpunkt die Steigerung der Geschwindigkeit noch ein großes Verbesserungspotential dar.
In dieser letzten siebten Einheit werden aufbauend auf den bereits erlernten Fähigkeiten weitere praktische Tätigkeiten trainiert. Am Beginn der Einheit werden Zeitcodes und Untertitelformate erläutert. Darauf aufbauend wird die Software Subtitle Workshop[8] vorgestellt und der Umgang mit ihr praktisch geübt. Dabei liegt der Fokus auf Tastenkombinationen, die für die Respeaking Tätigkeiten benötigt werden. Vor allem die Erstellung der Zeitcodes mittels Tastenkombinationen soll einer Respeakerin oder einem Respeaker wenig Konzentration abverlangen[9].
In den ersten sechs Einheiten wurden sukzessive vier der fünf Respeaking Tätigkeiten erlernt, die während der Phase der Untertitelerstellung ausgeführt werden. Die *Korrektur* stellt die fünfte und noch zu erlernende Tätigkeit dar. Während der Einheit wird in einer dreiminütigen Übung erstmals die *Eigenkorrektur* praktisch durchgeführt. Das Erstellen von Zeitcodes mit Subtitle Workshop ist ebenfalls Teil der Übung. Eventuelle Probleme und Fragen können während der letzten Einheit geklärt werden, um ein intensives und erfolgreiches Trainieren in der abschließenden Übungsaufgabe zu gewährleisten. Der Auszubildende hatte während der Einheit kaum

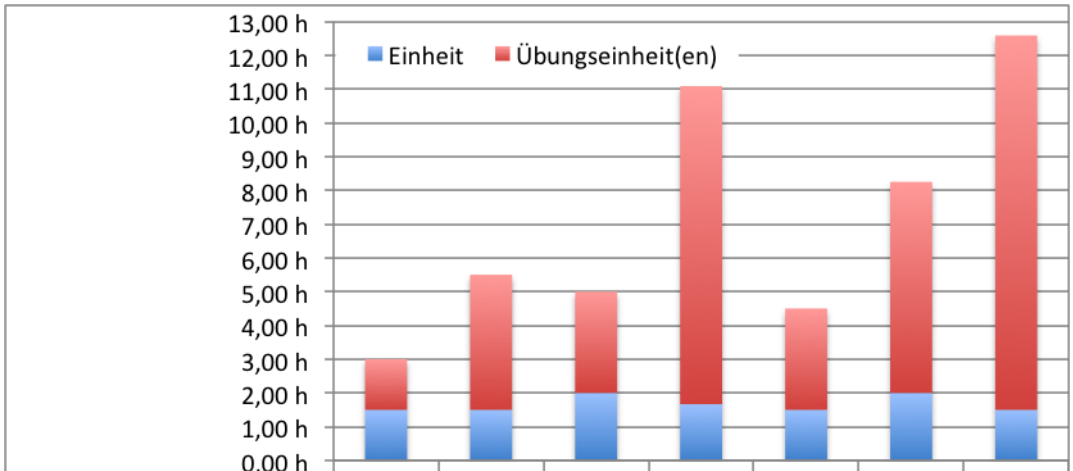
[6] Anm. Autor: Diktiergeschwindigkeit *während* der sechsten Einheit.
[7] Anm. Autor: Die Angaben basieren auf einer Stichprobe von ca. fünf Minuten. Die Analyse erfolgte durch den Auszubildenden R2.
[8] Anm. Autor: Dragon kann ohne zusätzliche Software keine Zeitcodes erzeugen bzw. exportieren. Daher wird die Software Subtitle Workshop verwendet, siehe Abschnitt 3.1.5 ab der Seite 57.
[9] Anm. Autor: In den Diktier- und Respeakingübungen der ersten sechs Einheiten wurde DragonPad verwendet. Im Gegensatz zu Subtitle Workshop ermöglicht DragonPad durch eine spezielle Korrekturfunktion die Verbesserung der Sprachprofils von Dragon. Die daraus resultierende Steigerung der Erkennungsgenauigkeit ist ein sehr wichtiger Bestandteil der Ausbildung. Da der Umgang mit Subtitle Workshop im Vergleich zu den anderen Respeaking Tätigkeiten weniger Übung bedarf, wurde die Verwendung von DragonPad einer längeren Übungsphase mit Software Subtitle Workshop vorgezogen.

Schwierigkeiten im Umgang mit Subtitle Workshop bzw. mit dem Erstellen der Zeitcodes. Das Erstellen der Zeitcodes in kürzere sowie regelmäßigere Zeitfenster stellte jedoch noch Verbesserungspotenzial dar. Erste Versuche der Eigenkorrektur von Fehlern während des Respeakings führte (wie zu erwarten) zu einem höheren Abstand zur Originalquelle. Dadurch hatte der Auszubildende Probleme mit dem Folgen des Gesagten während der Korrektur.
Bei der Erzeugung von offline Untertitel gibt es weitere Möglichkeiten für die (individuelle) Erstellung von Untertiteln sowie deren Korrektur. Aufbauend auf der Übung werden diese erläutert. Der Auszubildende soll schließlich in der abschließenden Übungsaufgabe eine individuelle Scripting Technik entwickeln. Diese kann verschiedene Korrekturtechniken beinhalten und soll zu einem NER-Wert von $\geq 98,0\%$ führen.
Das Aufbereiten sowie das Einbinden der Untertitel in die E-Learning Plattform Synote im Zuge der Nachbereitungsphase stellt den finalen Teil dieser Einheit dar. Für den Auszubildenden stellte das Erstellen des Benutzerkontos sowie das Einbinden der Aufnahme samt Folien und Untertitel keine Schwierigkeit dar.
In der Abschlussübung der Einheit 7 übte und festigte der Auszubildende die erlernten Scripting Techniken mit dem zur Verfügung gestellten Vorlesungsmaterial. Dabei führte er die Übung sieben mal - jeweils zwischen drei und sieben Minuten - durch und experimentierte u.a. mit der Wiedergabegeschwindigkeit. Diese betrug in der Übungsaufgabe der sechsten Einheit 69%. Neben einem Versuch mit der Originalsprechgeschwindigkeit (100%) wählte er schließlich für die letzten drei Übungsdurchläufe eine Wiedergabegeschwindigkeit von 88% und diktierte mit 85,2 WpM (inkl. IZ). Dabei stellte er in der Analyse dieser letzten drei Übungsdurchläufe einen NER-Wert von 97,5% fest. Durch das Markieren der Fehler und anschließender Korrektur konnte der NER-Wert auf 98,6% gesteigert und somit das Ausbildungsziel von $\geq 98,0\%$ erreicht werden. Die Steigerung um mehr als ein Prozent ist vor allem durch das Erkennen der gravierenden Fehler (vier von sechs) sowie der normalen Fehler (13 von 27) zurückzuführen. Er korrigierte weiters neun der 41 geringfügigen Fehler. Die in der Einheit 7 erstmals durchgeführte Eigenkorrektur wandte er kaum an, siehe Abschnitt 4.3.1 ab Seite 103.

4.2.9 Übersicht des Zeitaufwandes

Die Ausbildung ist in sieben Unterrichtseinheiten zu je 90 Minuten gegliedert, siehe Abschnitt 3.3 ab Seite 60. In den Einheiten wurde theoretisches Wissen vermittelt und praktische Übungen durchgeführt. Weiters erläuterte der Autor dieser Diplomarbeit jeweils die Übungsaufgaben, die von der auszubildenden Person eigenständig durchgeführt wurden. Im Anschluss an die sieben Einheiten wurde vom ausgebildeten Respeaker eine Vorlesung transkribiert und die Untertitel für Synote aufbereitet.
Der Gesamtaufwand für den Auszubildenden sollte dabei 75 Stunden[10] nicht überschreiten. Sämtliche Unterrichts- und Übungszeiten wurden in einer Stundenliste dokumentiert. In Abbildung 4.1 ist der Aufwand der sieben Einheiten sowie der dazugehörigen Übungsaufgaben dargestellt. Der Gesamtaufwand dafür betrug ca. 50 Stunden, was 2 ECTS entspricht. Somit konnten wie geplant die verbleibenden 25 Stunden für das Transkribieren der Vorlesung (siehe Abschnitt 4.3 ab Seite 101) sowie zum Einholen von Feedback aufgewandt werden, ohne den Gesamtaufwand von 3 ECTS zu überschreiten.
Wie in Abbildung 4.1[11] ersichtlich, wurden in der dritten, vierten und sechsten Einheit die geplanten 90 Minuten Unterrichtsdauer nicht eingehalten. Insgesamt wurden die Einheiten um 70 Minuten überzogen. Es wäre daher zukünftig sinnvoll eine zusätzliche, achte Einheit abzuhalten. Inhalte der erwähnten drei Einheiten könnten in diese Einheit verschoben werden. Zusätzlich könnte der erhöhte Aufwand der Übungsaufgaben der vierten und sechsten Einheit besser verteilt werden. Beide Adaptierungen würden den geplanten Gesamtaufwand von 75 Stunden weiterhin ermöglichen.
Die Termine der Einheiten wurden individuell mit dem Auszubildenden vereinbart. Dabei wurde auf die zeitlichen Ressourcen des Studenten Rücksicht genommen, damit es ihm möglich war die definierten Übungsaufgaben durchzuführen. Gerade zwischen vierter und fünfter, sechster und siebter Einheit lag aufgrund des hohen Übungsaufwandes verhältnismäßig viel Zeit. Für zukünftige Ausbildungen empfiehlt der Autor dieser Diplomarbeit die gleichmäßige Aufteilung dieser Einzelübungen auf die zusätzliche achte Einheit. Das würde einen regelmäßigeren Ablauf ermöglichen. Unverändert würde der Autor dieser Diplomarbeit den arbeitsintensiven Übungsteil der letzten Einheit lassen, der sich wie folgt begründet: Nach sämtlichen Einheiten hatte der ausgebildete Respeaker neben abschließenden Übungen noch Zeit, sich auf die zu transkribierende Vorlesung vorzubereiten. Weiters gehörte das gezielte Training bezüglich verbleibender Schwächen dazu. Das ist in dieser Intensität erst nach der letzten Einheit sinnvoll. Somit kann der hohe Aufwand nach der letzten Einheit mit der Vorbereitung auf eine Prüfung verglichen werden.

[10] Anm. Autor: 75 Arbeitsstunden entsprechen 3 ECTS an der TU Wien (vgl. [Pou03]).
[11] Anm. Autor: Siehe blauer Balkenbereich bzw. angeführte Stundenanzahl in der Datentabelle (Zeile 'Einheit').



	Einheit 1	Einheit 2	Einheit 3	Einheit 4	Einheit 5	Einheit 6	Einheit 7
Übungseinheit(en)	1,50 h	4,00 h	3,00 h	9,42 h	3,00 h	6,25 h	11,08 h
Einheit	1,50 h	1,50 h	2,00 h	1,67 h	1,50 h	2,00 h	1,50 h
Tage bis zur Folgeeinheit	9	20	8	22	10	23	

Abbildung 4.1: Zeitaufwand der Ausbildung für den Auszubildenden R2

4.3 Untertitel einer Vorlesung: Vergleich der Resultate

4.3.1 Datenbasis der Evaluierung

Lehrveranstaltung Trainingswissenschaft

Die Vorlesungsreihe *Trainingswissenschaft* wurde im Sommersemester 2012 von Mag. Manfred Zeilinger an der Universität Wien gehalten. Sie wurde von einer hörbeeinträchtigten Studentin des GESTU Projektes besucht. Weiters wurde sie von Titelbild in Echtzeit untertitelt und zusätzlich in Ton und Bild aufgezeichnet. Die zweistündige Vorlesung vom 05.06.2012[12] wurde vom Ausgebildeten (*R2*) und vom Autor dieser Diplomarbeit (*R1*) mit der erlernten Scripting Technik transkribiert. Die jeweilige Qualität der (für die angeführte Vorlesung erstellten) Untertitel ist in diesem Kapitel evaluiert.
Wie im Folgenden Abschnitt erläutert, wurde die Audioaufzeichnung der Vorlesung weiters mittels der Spracherkennung des EML transkribiert. Aufgrund der hohen Fehlerraten war kein qualitativer Vergleich zu den durch Respeaking erstellten Untertiteln sinnvoll. Im weiteren sind in diesem Abschnitt die Ausgangssituationen und Arbeitsweisen der Respeaker und Scripter erläutert. Diese Dokumentation ist Grundlage für die Interpretation der Ergebnisse, siehe Abschnitt 4.3.4 ab Seite 106.

[12] Anm. Autor: Um eine praxisnahe Ausgangssituation beim Scripting zu erreichen, erfolgte die Auswahl *einer* Vorlesung durch den Autor dieser Diplomarbeit ohne die Aufzeichnung vorab zu öffnen.

Ergebnisse der ASR des EML

Im Zuge von [Nem13] begann Nemecek das Spracherkennungssystem des EMLs mit aufgezeichneten Vorlesungen, erstellten '1:1' Transkripten und weiteren Materialien zu trainieren. Letztlich betrug die Wortakkuratheit (WRR) für Vorlesungen aus Österreich maximal 50%. Auch für die ausgewählte Vorlesung vom 05.06.2012 wurde mit der Spracherkennung ein Transkript erstellt. Im Abschnitt 6.1.1 ab Seite 139 ist ein Auszug davon in tabellarischer Form dem Gesagten des Vortragenden gegenübergestellt. Die erzeugten Untertitel in dieser Qualität „lassen nur sehr schwer, wenn überhaupt, auf den Inhalt des Vortrages schließen" [Nem13, Abschnitt 3.2] und sind für eine Unterstützung von hörbehinderten Studierenden unbrauchbar (vgl. [Nem13, Abschnitt 3.2]). Auch der Autor dieser Diplomarbeit vertritt diese Sichtweise und sieht für Untertitel mit dieser Qualität keine Unterstützungsmöglichkeiten für hörbeeinträchtigte Studierende. Selbst als Ausgangsbasis für eine Überarbeitung (z.B. für die Korrektur durch einen Tutor oder eine Tutorin) sind die Erkennungsraten zu niedrig.
Infolgedessen wurde die Mitwirkung an der Verbesserung der Spracherkennung Seitens GESTU vorerst eingestellt (vgl.[Nem13, Abschnitt 3.2]). Aufgrund der dokumentierten Qualität des Transkripts ist es aus Sicht des Autors dieser Diplomarbeit nicht sinnvoll die Ergebnisse - im Gegensatz zu den durch Respeaking erstellten Untertitel - qualitativ mit dem NER-Modell zu analysieren.

Christian Hattinger (*R1*) und der Auszubildende (*R2*)

Der Ausgebildete (*R2*) und der Autor dieser Diplomarbeit (*R1*) erstellten offline Untertitel mittels Respeaking (Scripting). Dazu hatten sie die Videoaufzeichnung der Vorlesung sowie die Vortagsfolien zur Verfügung. *R1* und *R2* arbeiteten unabhängig voneinander und jeweils alleine. Der Prozess wurde dokumentiert (Arbeitszeiten, Pausen, Kommentare, etc.) und je Scripter zwei Transkripte zur qualitativen Analyse herangezogen. Die Analyse ist in diesem Kapitel dokumentiert und beinhaltet je Scripter das Transkript direkt nach dem Respeaking (*vor Korrektur*) und jenes nach erfolgter Korrektur (*nach Korrektur*). In einem Zeitraum von ca. zwei Wochen arbeiteten *R1* und *R2* die zweistündige Vorlesung in sechs, je ca. 20 Minuten Blöcken, ab.
Die Vorlesungen der *Trainingswissenschaft* fanden im Sommersemester 2012 ca. wöchentlich statt und wurden aufgezeichnet. Daher stand *R1* und *R2* ausreichend Material für das Erlernen und Üben des Respeakings zur Verfügung. Während der Ausbildung wurden zu Übungszwecken zwei dieser Vorlesungsaufzeichnungen (vom 08.05.2012 und 15.05.2012) verwendet, siehe angeführte Materialien im Kapitel 3. Für die Vorbereitung zur Untertitelung der Vorlesung vom 05.06.2012 wurden den Scriptern die Folien bereitgestellt. Jedoch hatte weder *R1* noch *R2 während* der Ausbildung die Aufzeichnung vom 05.06.2012 zur Verfügung, wodurch keine Übungsmöglichkeiten gegeben waren. Somit konnte sichergestellt werden, dass ihnen keine detaillierten Kenntnisse über den Inhalt dieser Vorlesungseinheit bekannt war. Dadurch konnte eine Ausgangssituation wie beim Scripting in der Praxis hergestellt werden.

Titelbild (R3+R4)

Bei der Firma Titelbild wurde die live Untertitelung mittels Respeaking im Team, bestehend aus zwei Personen (*R3* und *R4*) durchgeführt. Die Beiden wechselten sich im 15 bis 20 Minuten Rhythmus ab. Die genauen Zeiten der Wechsel sind nicht bekannt. Infolgedessen wird in der Analyse des Transkripts nicht zwischen Untertitelpassagen von *R3* und *R4* unterschieden und das Team als *R3+R4* bezeichnet.
Wie auch die beiden Scripter *R1* und *R2* bekam das Respeaking Team der Firma Titelbild vorab die Folien zum Vortrag übermittelt. Die Vorlesung wurde akustisch (ohne Video) in Echtzeit von Wien nach Berlin übertragen. Technisch wurde die Übertragung mit einem Funkmikrofonset und einem Notebook im Hörsaal realisiert. Der Vortragende trug dazu das Funkmikrofon. Der Funkempfänger war mit dem Notebook verbunden. Das Audiosignal wurde mit der Software Skype[13] durch die Nutzung der im Hörsaal vorhandenen Internetverbindung übertragen (vgl. [Nem13, Abschnitt 4.3], [Woj12]).
Durch eine Webapplikation wurden die Untertitel am Notebook der hörbeeinträchtigten Studentin (die sich im Hörsaal befand) in Echtzeit angezeigt. Die Fehlerkorrektur erfolge live durch die gerade untertitelnde Person (Eigenkorrektur). Die Vorbereitungszeit für Vorlesungen beträgt lt. einer Auskunft von der Firma Titelbild je Respeaker und Respeakerin, abhängig vom Umfang des vorab verfügbaren Materials, zwischen zehn Minuten und einer Stunde (vgl. [Nem13, Abschnitt 4.3], [Woj12]). Die Durchführung der Echtzeituntertitelung durch die Firma Titelbild ist in [Nem13, Abschnitt 4.3] detailliert dokumentiert und evaluiert.

Vergleichbarkeit der Resultate von *R1, R2, R3+R4*

Die in Echtzeit erzeugten und nachträglich nicht korrigierten Untertitel der Firma Titelbild sind in diesem Kapitel mit den offline erstellten Untertiteln von *R1* und *R2* verglichen. Die Arbeitsweisen von *R1* und *R2* unterscheiden sich erheblich von der von *R3+R4.* Einerseits stand *R1* und *R2* eine Videoaufzeichnung (und nicht nur die Audioübertragung) des Gesagten zur Verfügung. Einen weiteren Unterschied stellt die Team (*R3+R4*) vs. Einzelarbeit (*R1* bzw. *R2*) dar. Entscheidend ist auch der Unterschied von live vs. offline Untertitelung. Daraus ergeben sich erweiterte Korrekturmöglichkeiten (z.B. die nachträgliche Korrektur, die Möglichkeit des Pausierens, etc.) sowie die individuelle Wahl der Wiedergabegeschwindigkeit, siehe Abschnitt 2.2.1 ab Seite 21. Darüber hinaus ist anzunehmen, dass der Stressfaktor bei der offline Untertitelung geringer ist. Weiters muss berücksichtigt werden, dass der Vortragende im 'gehobenen Wiener Dialekt' und an einigen Stellen einen stärker ausgeprägten Dialekt sprach, siehe Abschnitt 4.3.3 ab Seite 104. Die beiden österreichischen Respeaker *R1* und *R2* hatten dadurch eventuell einen Vorteil gegenüber den deutschen *R3+R4.* All diese Faktoren zusammen führten somit zu schwierigeren Arbeitsbedingungen für die aus Berlin arbeitenden *R3+R4.* Das muss bei der Interpretation der Ergebnisse berücksichtigt werden.
Auch die Arbeitsweisen von *R1* und *R2* unterschieden sich. So arbeitete *R1* ohne Veränderung der Wiedergabegeschwindigkeit, *R2* hingegen bevorzugte und arbeitete mit einer Wieder-

[13] `www.skype.com`, letzter Zugriff: 10.03.2013.

gabegeschwindigkeit von 90%[14]. Es kam bei *R1* sowie *R2* zu Fehlerkorrekturen während des Respeakings, siehe Eigenkorrektur im Abschnitt 2.2.1 ab Seite 21. Generell lag der Fokus auf der Korrektur und Verbesserung der Untertitel *nach* dem Respeaking. Aus dem Grund haben *R1* und *R2* während des Respeakings die Ausgabe der Spracherkennung beobachtet und Fehler zum effizienten Korrigieren markiert. Das Augenmerk richtetet sich dabei auf das Markieren von gravierenden und normalen Fehlern, siehe Abschnitt 2.6 ab Seite 41. *R2* wendete diese Technik intensiver an[15]. Das erstellte Transkript wurde in der Korrekturphase nach den Markierungen durchsucht und die markierten Fehler korrigiert. *R1* hingegen überprüfte in der Korrekturphase das gesamte Transkript auf Fehler und korrigiert diese. *R1* wie *R2* pausierten in unterschiedlichem Ausmaß die Wiedergabe der Aufzeichnung während des Respeakings. Resultierend aus den beiden Arbeitsweisen unterscheidet sich auch der jeweilige Aufwand für die Erstellung der Untertitel, siehe Abschnitt 4.3.6 ab Seite 128.

4.3.2 Stichproben der mittels Respeaking erstellten Untertitel

Es wurden fünf Transkripte analysiert und deren Qualität evaluiert. Dabei handelt es sich um die Transkripte von *R1* und *R2* jeweils *vor* und *nach* erfolgter Korrektur, sowie das von *R3+R4* live erstellte und nachträglich nicht korrigierte Transkript. Für die Analyse wurde die Vorlesung vom Autor dieser Diplomarbeit in sechs Blöcke (á 20 Minuten) unterteilt. Diese Unterteilung ist an die Arbeitsweise von *R1* und *R2* angelehnt, die die Vorlesung in diesen Blöcken transkribierten. Bei der Analyse ist dadurch nicht nur die qualitative Beurteilung der Vorlesung als Ganzes möglich, sondern können ggf. auch eventuelle Konzentrationsschwankungen von *R1* und *R2* (während der jeweiligen 20 Minuten dauernden Arbeitsblöcke) festgestellt werden. Für eine effektive[16] und trotzdem repräsentative Analyse wurden von jedem der sechs Blöcke die erste, die zehnte und die letzte Minute mit der NER-Analyse qualitativ beurteilt. Somit wurden je Transkript á 18[17] ca. einminütige Passagen ausgewertet. Die Auswahl ermöglicht aus Sicht des Autors dieser Diplomarbeit eine repräsentative Gesamtanalyse.
Die erläuterten unterschiedlichen Ausgangsbedingungen und Arbeitsweisen von *R1*, *R2* und *R3+R4* müssen wie erläutert bei der Interpretation der Ergebnisse berücksichtigt werden.

4.3.3 Vortrag: Sprechgeschwindigkeit und Stil

Aus Sicht des Autors dieser Diplomarbeit ist anzunehmen, dass die Sprechgeschwindigkeit und der Vortragsstil einen Einfluss auf die Schwierigkeit beim Respeaking haben. Daher wird im Folgenden auf die Sprechgeschwindigkeit(en) und auf einzelne Stilelemente des Vortragenden eingegangen.
Bezüglich des Vortragsstils handelte es sich um einen Frontalvortrag. Demnach gab es wenig

[14] Anm. Autor: *R2* transkribierte unbeabsichtigt für fünf Minuten mit 100% Wiedergabegeschwindigkeit. Die restliche Stunde und 55 Minuten mit 90%.
[15] Anm. Autor: *R1* markierte 26 Fehler, *R2* 115.
[16] Anm. Autor: Eine detaillierte NER-Analyse ist zeitintensiv. Die Analyse der 120 Minuten aller fünf Transkripte würde mindesten 150 Arbeitsstunden betragen.
[17] Anm. Autor: *R2* arbeitete wie beschrieben mit 90% Wiedergabegeschwindigkeit. Einzig die erste Stichprobe vom 4. Block berücksichtigt die Stelle, an welcher *R2* er mit einer Wiedergabegeschwindigkeit von 100% transkribierte.

Interaktion mit den Studierenden im Hörsaal. Trotzdem kam es zu einigen Publikumsfragen. Die gestellten Fragen wurden vor der Beantwortung vom Vortragenden nicht immer wiederholt. Daher konnten Publikumsfragen nur begrenzt untertitelt werden[18]. Der Vortragende sprach großteils im 'gehobenen Wiener Dialekt'. Teilweise verwendete er aber auch einen stärker ausgeprägten Dialekt. So wechselte er bei Beispielen aus der Praxis (Erfahrungsberichten) häufig zu einer sehr umgangssprachlichen Ausdrucksweise.

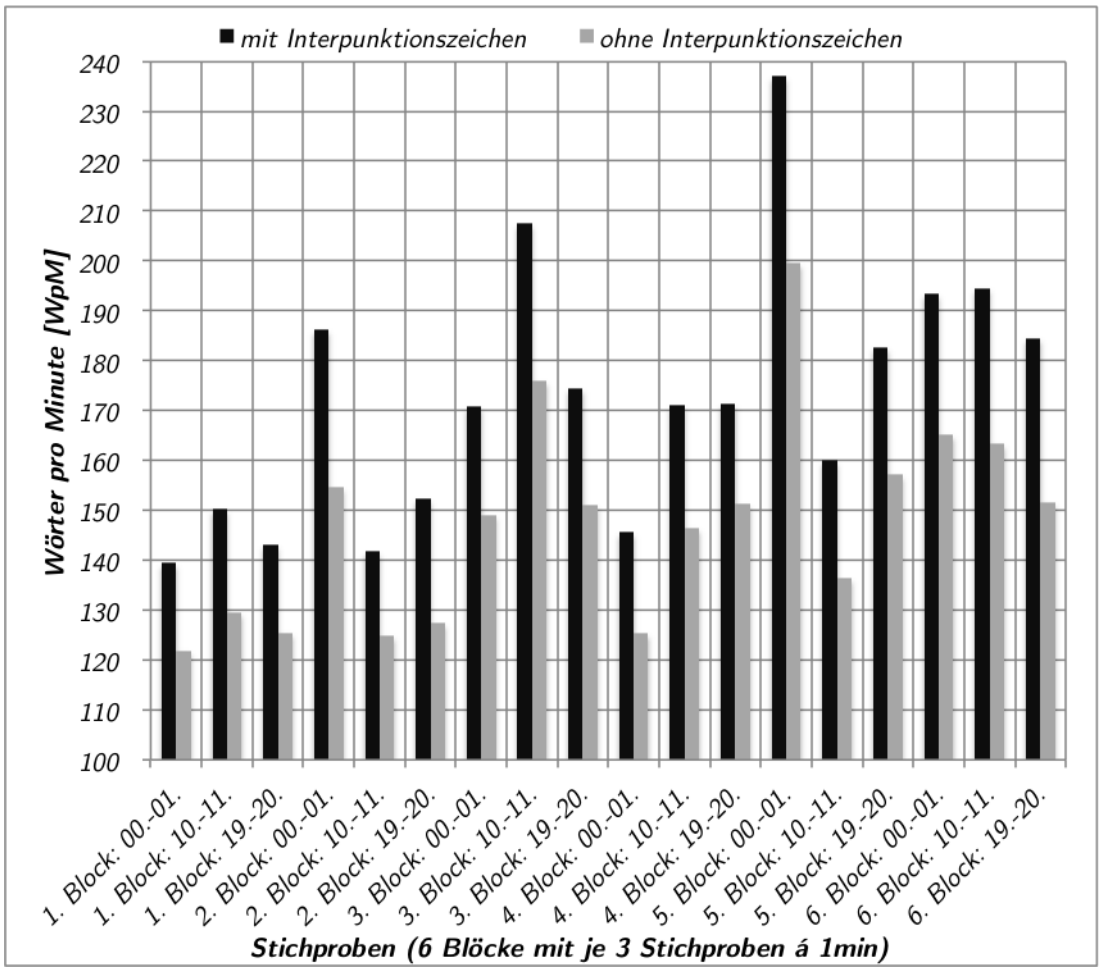


Abbildung 4.2: Sprechgeschwindigkeit des Vortragenden in Wörter pro Minute (WpM)

[18] Anm. Autor: Das Mikrofon wurde in Mundnähe an der Kleidung des Vortragenden angebracht. Daher mussten Publikumsfragen vom Vortragenden wiederholt werden, damit einerseits das Respeaking Team von Titelbild die Fragen transkribieren konnte sowie andererseits die Fragen aufgezeichnet werden konnten.

Der Vortragende sprach durchschnittlich[19] mit 173 WpM mit IZ, 148 Wörter pro Minute ohne Interpunktionszeichen. Der Median der Sprechgeschwindigkeit beträgt 171 WpM mit IZ bzw. 150 WpM ohne IZ. Berücksichtigt man also die beim Respeaking zu diktierenden Interpunktionszeichen, erhöht sich im Fall dieser Vorlesung die Sprechgeschwindigkeit durchschnittlich[19] um 16,8% (Median 16,5%; minimaler Wert 13,3%; maximaler Wert 21.7%).
Aus Sicht des Autors dieser Diplomarbeit können neben hohen Sprechgeschwindigkeiten auch Rhythmuswechsel das Respeaken erschweren. In der analysierten Vorlesung war die Sprechgeschwindigkeit des Vortragenden nicht gleichmäßig, wie der Abbildung 4.2 auf Seite 105 zu entnehmen ist. So lag die geringste Wortanzahl je Minute bei einer der 18 analysierten Minuten bei 139 WpM mit IZ bzw. 122 WpM ohne IZ. Der maximale Wert der Sprechgeschwindigkeit betrug 237 WpM mit IZ bzw. 200 WpM ohne IZ. Diese Werte beziehen sich auf die betrachteten Minuten als Ganzes, im Detail unterscheiden sich die jeweiligen Geschwindigkeiten innerhalb der Stichproben stärker. Der Satz „So, das war natürlich dann am nächsten Tag lustig, weil der Sportler ist zu mir gekommen und hat gesagt: Na du hast kein Vertrauen zu mir, wenn in der Zeitung schon steht dass ich e schlecht bin. Und so weiter.“ wurde beispielsweise vom Vortragenden innerhalb von sieben Sekunden gesprochen, was ca. 394 Wörtern pro Minute mit Interpunktionszeichen entspricht. Dabei handelt es sich um einen Spitzenwert, der nur in einer sehr kurzen Zeitperiode festgestellt wurde. Allerdings soll das Beispiel verdeutlichen, dass die Sprechgeschwindigkeit kurzzeitig sehr stark variierte. Das stellt wiederum eine Herausforderung für die Respeaker und Respeakerinnen dar.
Wenn nicht explizit anders angeführt, werden bei Angaben zu Sprech- und Diktiergeschwindigkeiten in diesem Kapitel die Wörter pro Minute *inklusive* Interpunktionszeichen angegeben, da das für das Respeaking die entscheidende Kennzahl ist.

4.3.4 NER-Analyse

Fehlerklassen

Die qualitative Beurteilung erfolgt anhand der im Abschnitt 2.6 ab Seite 41 beschriebenen NER-Analyse. Die Analyse wurde in einer vom Autor dieser Diplomarbeit erstellten Excel-Datei durchgeführt. Die Datei befindet sich auf einer CD, die zusammen mit einem gedrucktem Belegexemplar der vorliegenden Diplomarbeit beim ’Zentrum für angewandte assistierende Technologien (AAT)’ am ’Institut für Gestaltungs- und Wirkungsforschung’[20] hinterlegt ist.
In Bezug auf das inhaltliche Kürzen sowie das Umformulieren wird eine nahezu ’1:1’ Transkription als Ziel und somit Grundlage der Bewertung herangezogen. Wie im Abschnitt 2.6 ab Seite 41 erläutert, ist dies für die Unterscheidung von *geringfügigen Editierfehlern* und *korrekter Editierung* wichtig. Die Grenze zwischen den Fehlerklassen ist im Gegensatz zur Unterscheidung zwischen *gravierenden-*, *normalen-* und *geringfügigen Fehlern* abhängig vom Land, Unternehmen, Programm, etc. (vgl. [RFM14]). Aus dem Grund sind im Folgenden die angewandten Bewertungskriterien - im Speziellen bezüglich korrekter Editierung - erläutert. Zusätzliche Beispiele sollen die weiteren Fehlerklassen verdeutlichen.

[19] Anm. Autor: Als durchschnittlich wird in diesem Kapitel das arithmetisches Mittel bezeichnet. In dem Fall jenes der 18 Stichproben mit á ca. einer Minute.
[20] `www.aat.tuwien.ac.at`, letzter Zugriff: 10.03.2013.

- **Korrekte Editierung**:
  - Sämtliche Umformulierungen, die einerseits die Lesbarkeit verbessern bzw. nicht verschlechtern und gleichzeitig zu keinem Informationsverlust führen.
  - Das Weglassen von Wörtern, die keinen oder nur einen sehr geringen Informationsverlust bedeuten (z.B.: „dann werde ich weiter trainieren“ anstatt „dann werde ich einfach weiter trainieren“).
  - Grammatikalische Fehler im Vortrag, die vom Respeaker oder der Respeakerin nicht korrigiert wurden.
  - Das Weglassen von Bemerkungen, die keinen Informationsverlust der vermittelten Inhalte nach sich ziehen (z.B.: „Eine Wettkampfanalyse gibt es in Österreich eigentlich nicht.“ anstatt „Wettkampfanalyse, habe ich schon erwähnt, wo es in Österreich eigentlich nichts gibt.“; „Und da war klar, dass die Leistungsfähigkeit [...]“ anstatt von „Und da war natürlich klar, dass die Leistungsfähigkeit [...]“).
  - Das Paraphrasieren von Orten, Namen, etc. ohne Auswirkung auf den Inhalt (z.B.: „Der Torwart kommt teilweise bis zur Mittellinie heraus und schießt auch Elfmeter wenn es sein muss.“ anstatt „Neuer[21] [...] kommt teilweise fast bis zur Mittellinie hinaus. Und wenn es sein muss schießt er auch einen Elfer. Ja.“).
  - Nicht transkribierte Publikumsfragen, wenn sie nicht oder nur kaum verständlich sind.

[21] Anm. Autor: Manuel Neuer ist ein Torwart.

- **Geringfügige Fehler**:
  - Ein Informationsverlust in den Untertitel wird bei der angestrebten nahezu '1:1' Transkription als normaler Fehler gewertet. Ist die Information jedoch auf den Folien der Vorlesung klar ersichtlich und der Bezug auf die Folien im Zusammenhang gegeben, wird der Informationsverlust als geringer Editierfehler gewertet. (z.B.: „Bei einer langen Wettkampfperiode, geht die Wochenbelastung zurück." anstatt „Und was man dann auch sieht, ist natürlich im Bereich der Wettkampfperiode, wenn man da Wochenbelastungen haben irgendwo um 20 bis 24, dann gehen natürlich hier die Wochenbelastungen zurück auf zehn bis zwölf Stunden pro Woche." Die Angaben von 20 bis 24 sowie zehn bis zwölf Stunden sind auf den Folien vorhanden und wurden daher nicht als normaler, sondern jeweils als ein geringer Editierfehler gewertet.
  - Fehlende, falsch gesetzte oder doppelte Interpunktions- und Leerzeichen (z.B.: „Das wird uns heute beschäftigen und eventuell in der letzten Einheit in 14 Tagen.").
  - Rechtschreib- und Tippfehler (z.B.: durch manuelle Korrektur).
  - Falsch transkribierte Wörter, die jedoch im Zusammenhang klar das richtige Wort rekonstruieren lassen bzw. ggf. beim Lesen nicht als Fehler erkannt werden („Dort muss sich im Hintergrund die Jahrestrainingsstruktur, die ich selber habe [...] nehmen." anstatt „Dort muss ich im Hintergrund die Jahrestrainingsstruktur, die ich selber habe [...] nehmen.").
  - Doppelt transkribierte Wörter.
  - Wenn ein Wort doppelt - und zwar einmal korrekt und einmal falsch - transkribiert ist und aus dem Zusammenhang das korrekte Wort klar hervor geht.
  - Falsche Artikel und Präpositionen (z.B.: „der stehen einerseits die [...]" anstatt „da stehen einerseits die [...]").
  - Falsche Worttrennung (z.B.: „Einige Bemerkungen zu Problem stellen [...]" anstatt „Einige Bemerkungen zu Problemstellen [...]").
  - Flasche Verwendung von Singular und Plural (z.B.: „Auch die Talent bereits in diesem Umfeld zu trainieren." anstatt „Auch die Talente bereits in diesem Umfeld zu trainieren.").

- **Normaler Fehler**:
  - Fehler die im Zusammenhang klar als solche erkennbar sind, jedoch der korrekte Inhalt nicht in den Untertiteln enthalten ist bzw. sich das richtige Wort nicht eindeutig 'erahnen' lässt (z.B.: „Weltstandsanalyse, das habe ich schon erwähnt, was es in Österreich nicht gibt [...] " anstatt „Wettkampfanalyse, habe ich schon erwähnt, wo es in Österreich eigentlich nichts gibt.").
  - Das Fehlen von Information (z.B.: „Also wird hier die Kalenderwoche festgelegt." anstatt „Mit dem Hintergrund um eben in der Trainingsauswertung Zeitabschnitte vergleichbar zu machen, werden die also hier von der Kalenderwoche her festgelegt.").
- **Gravierender Fehler**:
  - Falsch transportierter Inhalt, der in den Untertitel jedoch nicht als solcher zu identifizieren ist (z.B.: „Darum spielen nicht mehr viele Franzosen mit und haben soviel Franzosen eine Wildcard bekommen." anstatt „Darum spielen auch so viele Franzosen mit und haben so viele Franzosen eine Wildcard gekriegt."; „Insoweit, dass wir von den Ebenen hier ein nationales oder internationales Wettkampfsystem haben." anstatt „Insoweit, dass wir von den Ebenen her ein nationales Wettkampfsystem haben und ein internationales.").

Beispiele von NER-Werten

Wie im Abschnitt 2.6 ab Seite 41 erläutert, wird die Qualität der Untertitel ab einem NER-Wert von $\geq 98,0\%$ als akzeptabel bezeichnet. Aus Sicht des Autors dieser Diplomarbeit könnte allerdings alternativ ab $\geq 97,0\%$ von akzeptabler, ab $\geq 98,0\%$ von guter und ab einem NER-Wert $\geq 99\%$ von ausgezeichneter Qualität gesprochen werden.
Um die Aussagekraft und Bedeutung der NER-Werte für die Qualität der Untertitel zu verdeutlichen, sind im Folgenden einige Beispiele mit verschiedenen Werten angeführt.

Beispiel 1:
Der Vortragende sagte:
„Also, das ist natürlich eine Auswirkung, die sich[22] jetzt das gesamte Arbeiten in dieser Mannschaft von Haus aus mit einem Schlag fast unmöglich macht. Ja. Das kann dazu führen, dass Sportler plötzlich völlig von den Trainern weg gehen. Das hier Berater von außen hereinkommen müssen, um genau diese Schnittstellen wieder zu zu machen. Und und und. Ja."

R1 erreichte *nach* erfolgter Korrektur für den Block einen NER-Wert von **99,6**%:
„Das ist natürlich eine Auswirkung, die das gesamte Arbeiten in dieser Mannschaft von Haus aus mit einem Schlag fast unmöglich macht. Das kann dazu führen, der[23] Sportler plötzlich völlig von den Trainern weg gehen. Das Berater von außen hereinkommen müssen, um diese Schnittstellen zu schließen. Und und und."

R2 erreichte *nach* erfolgter Korrektur einen NER-Wert von **98,1**%:
„das[24] ist eine Auswirkung, die jetzt das gesamte Arbeiten in der wahren[25] Mannschaft von Haus aus mit einem Schlag fast unmöglich macht. Das kann dazu führen, dass Sportler plötzlich von den Trainern Weg[26] gehen, dass Berater von außen kommen müssen um diese Schnittstellen wieder zu zu machen usw."

R1 erreichte *vor* erfolgter Korrektur einen NER-Wert von **96,8**%:
„Das ist natürlich eine Auswirkung, die das gesamte arbeiten[27] dieser Mannschaft von Haus aus mit einem Schlag fast unmöglich macht. das[24] kann dazu führen, der[23] Sportler plötzlich völlig von den Tränen[28] Weg[26] gehen. Das Berater von außen kommen müssen, um diese Schnittstellen zu schließen. Und und und. "

[22] Anm. Autor: Hier handelt es sich um einen Grammatikfehler im Vortrag.
[23] Anm. Autor: Es wurde anstatt 'dass' das Wort 'der' transkribiert; geringer Editierfehler (Faktor 0,25).
[24] Anm. Autor: Es wurde das Wort 'das' am Satzbeginn klein transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[25] Anm. Autor: Es wurde das Wort 'waren' mit Dehnungs-h transkribiert; normaler Erkennungsfehler (Faktor 0,50).
[26] Anm. Autor: Es wurde das Wort 'weg' groß anstatt klein transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[27] Anm. Autor: Es wurde das Wort 'arbeiten' klein transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[28] Anm. Autor: Es wurde das Wort 'Tränen' anstatt 'Trainern' transkribiert; normaler Erkennungsfehler (Faktor 0,50).

Beispiel 2:
Der Vortragende sagte:
„Und auch einige Bemerkungen zu Problemstellen: Wettkamfanalyse, habe ich schon erwähnt, wo es in Österreich eigentlich nichts gibt. Weil wir uns eben diese Struktur nicht leisten können oder wollen oder weil man diesen Bereich nicht wirklich ernst nimmt."

R1 erreichte *nach* erfolgter Korrektur für diesen Block einen NER-Wert von **96,9**%:
„Und auch einigen[29] Bemerkungen zu Problemstellen. Weltstandsanalyse[30], das habe[31] schon erwähnt, was es in Österreich nicht gibt weil wir uns diese Struktur nicht leisten können oder wollen oder weil man dem[32] Bereich nicht wirklich ernst nimmt."

R2 erreichte *nach* erfolgter Korrektur für diesen Block einen NER-Wert von **97,1**%:
„und[33] einige Bemerkungen zu Problem stellen[34], wie zum Beispiel Weltstandsanalyse[30]. In Österreich gibt es das nicht, weil wir uns diese Struktur nicht leisten können oder wollen oder sie nicht ernst nehmen."

R2 erreichte *vor* erfolgter Korrektur für diesen Block einen NER-Wert von **95,7**%:
„und[33] einige Bemerkungen zu Problem stellen[34], wie zum Beispiel Weltstandsanalyse[30]. In Österreich gibt es das nicht, weil wir uns diese Struktur nicht leisten können oder wollen oder Sinn[35] nicht ernst nehmen."

[29] Anm. Autor: Es wurde das Wort 'einigen' anstatt 'einige' transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[30] Anm. Autor: Es wurde das Wort 'Weltstandsanalyse' anstatt 'Wettkampfanalyse' transkribiert; normaler Erkennungsfehler (Faktor 0,50).
[31] Anm. Autor: Es wurde das Wort 'ich' nicht transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[32] Anm. Autor: Es wurde das Wort 'dem' anstatt 'diesen' transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[33] Anm. Autor: Es wurde das Wort 'und' am Satzbeginn klein transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[34] Anm. Autor: Es wurde das Wort 'Problem stellen' anstatt 'Problemstellen' transkribiert; geringer Erkennungsfehler (Faktor 0,25).
[35] Anm. Autor: Es wurde das Wort 'Sinn' anstatt "sie' transkribiert; normaler Erkennungsfehler (Faktor 0,50).

Erzielte NER-Werte

Im Folgenden sind die NER-Werte der jeweiligen Transkripte angeführt. Sie basieren auf der durchgeführten Analyse der jeweils 18 Stichproben:

- NER-Wert Transkript von *R1 nach* erfolgter Korrektur: **98,9**%
- NER-Wert Transkript von *R1 vor* erfolgter Korrektur: **96,8**%
- NER-Wert Transkript von *R2 nach* erfolgter Korrektur: **97,4**%
- NER-Wert Transkript von *R2 vor* erfolgter Korrektur: **96,7**%
- NER-Wert Transkript von *R3+R4* (*live* erstellt und *ohne* anschließende Korrektur): **94,9**%

Wie im Abschnitt 4.3.1 auf Seite 103 erläutert, sind beim Vergleich der jeweiligen Ergebnisse die unterschiedlichen Arbeitsweisen sowie die daraus resultierenden zeitlichen Aufwände für die Erstellung der Transkripte/Untertitel zu beachten. Wie im Abschnitt 4.3.6 dokumentiert, beträgt die Dauer für die Erstellung des Transkripts *vor* der *Korrektur* bei *R1* die zweifache Zeit der Vorlesung bzw. die 4,3-fache für das Transkript *nach* der *Korrektur*. Bei *R2* ist der Arbeitsaufwand für das Transkript *vor* der *Korrektur* das 2,3-fache der Vorlesungszeit sowie das 3,2-fache nach der Korrektur. Da *R3+R4* im Team arbeiteten und eine Vorbereitungszeit zwischen zehn Minuten und einer Stunde je Vorlesung benötigten, beträgt die Arbeitszeit das 2,2 bis 3,0-fache der Vorlesung.
In den folgenden Abschnitten sind die Werte detailliert analysiert, grafisch aufbereitet und die Ergebnisse sowie Erkenntnisse abschließend im Abschnitt 4.3.7 ab Seite 129 zusammengefasst.

Kreuztabellen

In den zwei Kreuztabellen (Tabelle 4.1 sowie 4.2) ist ersichtlich, bei welchen der 18 Stichproben im jeweiligen Transkript der NER-Wert 98,0% bzw. 97,0% beträgt bzw. übersteigt. Durch die Korrektur von *R1* ist eine qualitative Steigerung von vier auf 17 Stichproben beim NER-Wert von $\geq$ 98,0% bzw. von sieben auf 18 beim NER-Wert von $\geq$ 97,0% zu verzeichnen. Im Vergleich ist die Steigerung durch die Korrektur bei *R2* von zwei auf sechs beim NER-Wert $\geq$ 98,0% bzw. von sieben auf zwölf bei jenem von $\geq$ 97,0% ersichtlich. Demnach erhöhte *R1* die Qualität der Untertitel durch die Korrektur deutlicher als *R2*. Zu beachten sind hierbei jedoch die benötigten zeitlichen Aufwände. Bei der live Untertitelung durch *R3+R4* hat eine der 18 Stichproben einen NER-Wert von mindestens 98,0% bzw. drei einen NER-Wert von mindestens 97,0%.

Box-Whisker-Plot

Der Box-Whisker-Plot in diesem Abschnitt dient zur grafischen Darstellung der Qualität und Streuung. Die Streuung bezieht sich auf die variierenden NER-Werte der 18 Stichproben und ermöglicht die Interpretation der qualitativen Konstanz des jeweiligen Transkriptes.
Jedes der fünf analysierten Transkripte ist durch eine Box/Spalte repräsentiert. Die Boxhöhe ist der Bereich, in dem 50% der Daten liegen und wird durch das untere Quartil (Q1) und das obere

	R1 nach Korrektur	*R1* vor Korrektur	*R2* nach Korrektur	*R2* vor Korrektur	R3+ R4
Block 1: 00.-01. min	1	0	1	1	0
Block 1: 10.-11. min	1	0	1	0	0
Block 1: 19.-20. min	1	0	0	0	0
Block 2: 00.-01. min	1	0	0	0	0
Block 2: 10.-11. min	1	1	0	0	0
Block 2: 19.-20. min	1	0	1	1	0
Block 3: 00.-01. min	0	0	0	0	0
Block 3: 10.-11. min	1	0	0	0	0
Block 3: 19.-20. min	1	0	0	0	0
Block 4: 00.-01. min	1	0	0	0	1
Block 4: 10.-11. min	1	0	0	0	0
Block 4: 19.-20. min	1	0	1	0	0
Block 5: 00.-01. min	1	0	0	0	0
Block 5: 10.-11. min	1	0	1	0	0
Block 5: 19.-20. min	1	0	0	0	0
Block 6: 00.-01. min	1	1	0	0	0
Block 6: 10.-11. min	1	1	1	0	0
Block 6: 19.-20. min	1	1	0	0	0
Summe	**17** von 18	**4** von 18	**6** von 18	**2** von 18	**1** von 18

Tabelle 4.1: NER-Wert von mind. 98,0% erreicht ('1' wenn Wert $\geq 98,0\%$; '0' wenn < 98,0%)

Quartil (Q3) begrenzt. Die Höhe (Q3-Q1) wird auch als Interquartilsabstand (engl. *interquartile range*, IRQ) bezeichnet und gibt Auskunft über das Maß der Streuung. Der Wert Q2 ist der Median der NER-Werte und als horizontale Linie innerhalb in der jeweiligen Boxen visualisiert. Diese Linie trennt optisch den grünen (Werte innerhalb der IRQ über dem Median) und den rosa (Werte innerhalb des IRQ unter dem Median) Bereich der Boxen. Die Antennen (engl. *whisker*) stellen die Werte außerhalb der Box dar. Ausgenommen sind sogenannten obere- und untere Ausreißer. Ausreißer sind jene Werte, die um das 1,5-Fache von IRQ über Q3 bzw. unter Q1 liegen und sind als rote Quadrate dargestellt.

Der Box-Whisker-Plot der Abbildung 4.3 auf Seite 114 stellt die Qualität sowie Streuung der jeweiligen Transkripte dar und beruht auf den 18 Stichproben der zweistündigen Vorlesung.
Das Transkript von *R1 nach* erfolgter *Korrektur* ist am konstantesten (IRQ von 0,63) und weist die höchste Qualität der fünf verglichenen Transkripte auf (Q2 NER-Wert von 98,9%). Der einzige Ausreißer dieses Transkripts ist in der Kreuztabelle 4.1 ersichtlich (in der ersten Stichprobe des 3. Blocks beträgt der NER-Wert weniger als die angestrebten 98,0%.) *Vor* erfolgter *Korrektur* ist die Qualität von *R1* wie zu erwarten weniger konstant (IRQ von 2,01) und niedriger

	R1 nach Korrektur	R1 vor Korrektur	R2 nach Korrektur	R2 vor Korretkur	R3+R4 (live)
Q3-Q2	0,34	0,86	0,65	0,69	0,85
Q2-Q1	0,29	1,15	0,67	0,94	0,94
Q1	98,6% NER-Wert	95,4% NER-Wert	96,9% NER-Wert	95,9% NER-Wert	94,0% NER-Wert
Q2 (Median)	98,9% NER-Wert	96,6% NER-Wert	97,6% NER-Wert	96,8% NER-Wert	94,9% NER-Wert
IQR	0,63	2,01	1,32	1,63	1,79
U. Ausreißer #1	97,4% NER-Wert		94,7% NER-Wert		86,3% NER-Wert
O. Ausreißer #1					
U. Ausreißer #2					91,3% NER-Wert

Abbildung 4.3: NER-Werte *R1*, *R2*, *R3+R4*; Streuung der 18 Stichproben (á ca. einer Minute)

	R1 nach Korrektur	*R1* vor Korrektur	*R2* nach Korrektur	*R2* vor Korrektur	R3+ R4
Block 1: 00.-01. min	1	0	1	1	0
Block 1: 10.-11. min	1	0	1	1	0
Block 1: 19.-20. min	1	1	1	1	0
Block 2: 00.-01. min	1	0	0	0	0
Block 2: 10.-11. min	1	1	1	1	0
Block 2: 19.-20. min	1	1	1	1	1
Block 3: 00.-01. min	1	0	0	0	0
Block 3: 10.-11. min	1	0	0	0	0
Block 3: 19.-20. min	1	0	1	0	0
Block 4: 00.-01. min	1	0	1	0	1
Block 4: 10.-11. min	1	0	1	0	0
Block 4: 19.-20. min	1	0	1	1	1
Block 5: 00.-01. min	1	1	0	0	0
Block 5: 10.-11. min	1	0	1	0	0
Block 5: 19.-20. min	1	0	0	0	0
Block 6: 00.-01. min	1	1	1	0	0
Block 6: 10.-11. min	1	1	1	1	0
Block 6: 19.-20. min	1	1	0	0	0
Summe	**18** von 18	**7** von 18	**12** von 18	**7** von 18	**3** von 18

Tabelle 4.2: NER-Wert von mind. 97,0% erreicht ('1' wenn Wert $\geq 97,0\%$; '0' wenn < 97,0%)

(Q2 NER-Wert von 96,6%). Das verdeutlicht die Qualitätssteigerung durch die Korrektur. So ist eine Steigerung des NER-Wertes um 2,38% zu verzeichnen und die Streuung um das 3,19-fache geringer (demnach qualitativ besser) als *vor* der *Korrektur*.
Das Transkript von *R2 nach* erfolgter *Korrektur* hat die zweit konstanteste (IRQ von 1,32) und die zweit höchste Qualität (Q2 NER-Wert von 97,6%) der fünf Transkripte. Wie in der Kreuztabelle 4.1 ersichtlich, beträgt in vier der 18 Stichproben der NER-Wert die angestrebten $\geq 98,0\%$. Der einzige Ausreißer im Transkript *R2 nach* erfolgter *Korrektur* ist bei der dritten Stichprobe des 5. Blocks. *Vor* erfolgter *Korrektur* ist die Qualität des Transkripts von *R2* wie zu erwarten niedriger (Q2 NER-Wert von 96,8%), was einer Steigerung von 0,83% durch die Korrektur entspricht. Der Wert für die Konstanz des Transkripts von *R2 nach* erfolgter *Korrektur* ist um das 1,23-fache geringer (demnach qualitativ besser) als *vor* erfolgter *Korrektur*.
Im direkten Vergleich der Transkripte *vor* erfolgter Korrektur sind die erreichten NER-Werte von *R1* und *R2* nahezu ident (96,7% vs. 96,8% NER-Wert), jedoch ist die Qualität von *R2* konstanter. Durch die unterschiedlichen Korrekturansätze ist die Qualität beim Transkript von *R1 nach* der *Korrektur* höher bezüglich des erreichten NER-Wertes und der Konstanz. Wie erwähnt, wirkte sich das negativ auf den Aufwand aus, siehe Abschnitt 4.3.6 ab Seite 128.

Das live erstellte Transkript von *R3+R4* hat die niedrigste Qualität (Q2 NER-Wert von 94,9%) der fünf analysierten Transkripte und eine Abweichung in der Konstanz von 1,79. Wie auch in der Kreuztabelle 4.1 ersichtlich, beträgt in einer der 18 Stichproben der NER-Wert mindestens die angestrebten 98,0%. Die beiden Ausreißer sind aus der ersten Stichprobe des 2. Blocks sowie der zweiten Stichprobe des 3. Blocks. Sie sind im nächsten Abschnitt näher diskutiert.
Wie im Abschnitt 4.3.1 ab Seite 103 erläutert, sind beim Vergleich der jeweiligen Ergebnisse die unterschiedlichen Arbeitsweisen und sowie die daraus resultierenden zeitlichen Aufwände für die Erstellung der Transkripte/Untertitel zu beachten, siehe voriger Abschnitt 4.3.4 bzw. Abschnitt 4.3.6.

Fehlerklassen und Fehlerursachen

In Tabelle 4.3 auf Seite 117 ist die Anzahl der Fehler je nach *Fehlerursache* (*Editier-* bzw. *Erkennungsfehler*) und nach *Fehlerklasse* (*gravierend*, *normal*, *geringfügig*) dargestellt. Die Anzahl bezieht sich auf die jeweils 18 (ca. einminütigen) Stichproben der analysierten Transkripte. Dabei ist die Fehlerursache vor allem für das Training bzw. die Fortbildung von Respeakerinnen und Respeakern und für die Verbesserung der Spracherkennung interessant. Generell ist jedoch aus Sicht des Autors dieser Diplomarbeit für die Interpretation der Qualität die Summe der jeweiligen Fehlerklassen entscheidend. Es ist daher weniger wichtig, ob die Ursache beim Respeaker bzw. der Respeakerin oder der Spracherkennung liegt.
Bei *R1* ist die Gesamtanzahl der Fehler 228, bei *R2* 191 und beträgt sie *R3+R4* 165. Bei der Gegenüberstellung der gesamten Fehleranzahl - ohne die Unterscheidung der Fehlerklassen - mit den NER-Werten ist ersichtlich, dass höhere Fehleranzahlen *nicht* mit geringer Qualität korrelieren. So ist der NER-Wert im Median (siehe Abbildung 4.3 auf Seite 114) von *R3+R4* (geringste Fehleranzahl) am niedrigsten. Das verdeutlicht die Auswirkung der drei Fehlerklassen und deren Gewichtung auf die NER-Werte. So sind im Transkript von *R3+R4* die höchste Anzahl von gravierenden und normalen Fehlern, was sich negativ in der Gesamtqualität widerspiegelt.
Den Tabellen ist auch zu entnehmen, welche Fehler in welchem Ausmaß von *R1* sowie *R2* in der Korrekturphase gefunden bzw. korrigiert wurden. Hier zeigt sich die in den Kreuztabellen bzw. im Box-Whisker-Plot dokumentierte Verbesserung der Qualität durch die Korrektur. Während *R1* und *R2* prozentuell eine vergleichbare Anzahl von gravierenden Fehlern korrigieren konnten, ist ein großer Unterschied in der Korrektur von normalen und geringfügigen Fehlern zu sehen. So korrigierte *R1* 77,9% der normalen und 42,86% der geringfügigen Fehler. *R2* hingegen 24,0% der normalen und 14,0% der geringfügigen. Das unterstreicht die verschiedenen Korrekturmethoden. Wie im Abschnitt 4.3.1 ab Seite 103 beschrieben, markierte er während der zwei Stunden *R2* 115 Fehler (demnach im Durchschnitt alle 57,5 Sekunden) und das Transkript in der Korrekturphase nach den Markierungen durchsucht bzw. die Fehler korrigiert. *R1* überprüfte hingegen in der Korrekturphase das gesamte Transkript auf Fehler und korrigierte diese. Durch die verschiedenen Arbeitsweisen unterscheidet sich wie erwähnt der jeweilige Aufwand für die Erstellung bzw. speziell für die Korrektur der Untertitel, siehe Abschnitt 4.3.6 ab Seite 128.
Bei der Schnittstelle zwischen Dragon und der Untertitelsoftware kam es vor allem bei *R1* und *R2* (*Subtitle Workshop*) häufig zu Erkennungsfehlern am Satzbeginn (besonders Klein- anstatt Großschreibung). Von den 122 geringfügigen Erkennungsfehlern beim Transkript von *R1 vor*

der *Korrektur* beträgt die Anzahl von Kleinschreibungsfehlern am Satzbeginn 35. Alle diese Fehler wurden im Zuge der Korrektur von *R1* korrigiert. Bei *R2* sind ebenfalls 35 Fehler (der insgesamt 89 geringfügigen Erkennungsfehler) im Transkript *vor* der *Korrektur* auf diese Problematik zurückzuführen. Aufgrund der anderen Korrekturmethode von *R2* sind die 35 Fehler auch im Transkript *nach* der Korrektur vorhanden. Ohne sie würde sich der NER-Wert im Transkript von *R2 nach* erfolgter Korrektur anstatt der 97,4% (siehe Tabelle 4.5 auf Seite 127) auf 97,8% erhöhen. Bei *R3+R4* kam es insgesamt drei mal zur Kleinschreibung am Satzbeginn.
Im Folgenden sind die Ursachen der im vorigen Abschnitt dokumentierten Ausreißer (siehe weiters Abbildung 4.3 auf Seite 114) erläutert. Bei *R1 nach* der *Korrektur* lässt sich der Ausreißer auf die - im Vergleich zu den anderen Stichproben - hohe Anzahl von normalen (vier) und geringfügigen (sieben) Erkennungsfehlern zurückführen. Der Ausreißer im Transkript von *R2 nach* erfolgter Korrektur begründet sich in einem gravierenden Erkennungsfehler. In den zwei Stichproben von *R3+R4*, welche die beiden unteren Ausreißer beinhalten, liegen fünf der insgesamt 12 gravierenden Fehler. In der zweiten Stichprobe des 3. Blocks sind darüber hinaus elf normale Editierfehler zu verzeichnen. Ein Grund für den hohen Informationsverlust innerhalb einer Minute könnten technische Probleme gewesen sein, mit denen eventuell *R3+R4* konfrontiert waren. Dabei könnte es sich um akustische Probleme bei der Übertragung, aber ebenso um Probleme mit Dragon oder der Untertitelsoftware gehandelt haben. Auch inhaltliche Verständnis- sowie Konzentrationsschwierigkeiten könnten die Ursache für diese Ausreißer sein. Unabhängig der Ursachen heben die Ausreisser deutlich die leichtere Ausgangsbedingung bei der offline Transkription hervor. So pausierte *R1* die Wiedergabe der zweistündigen Vorlesung insgesamt 13 Minuten, *R2* für 72 Minuten, siehe Abschnitt 4.3.6 ab Seite 128. Die Möglichkeit zum Pausieren hatten *R3+R4* durch die live Situation nicht, wodurch sich unterschiedliche (auch kurzfristige) Probleme rasch negativ auf die Qualität auswirkten.

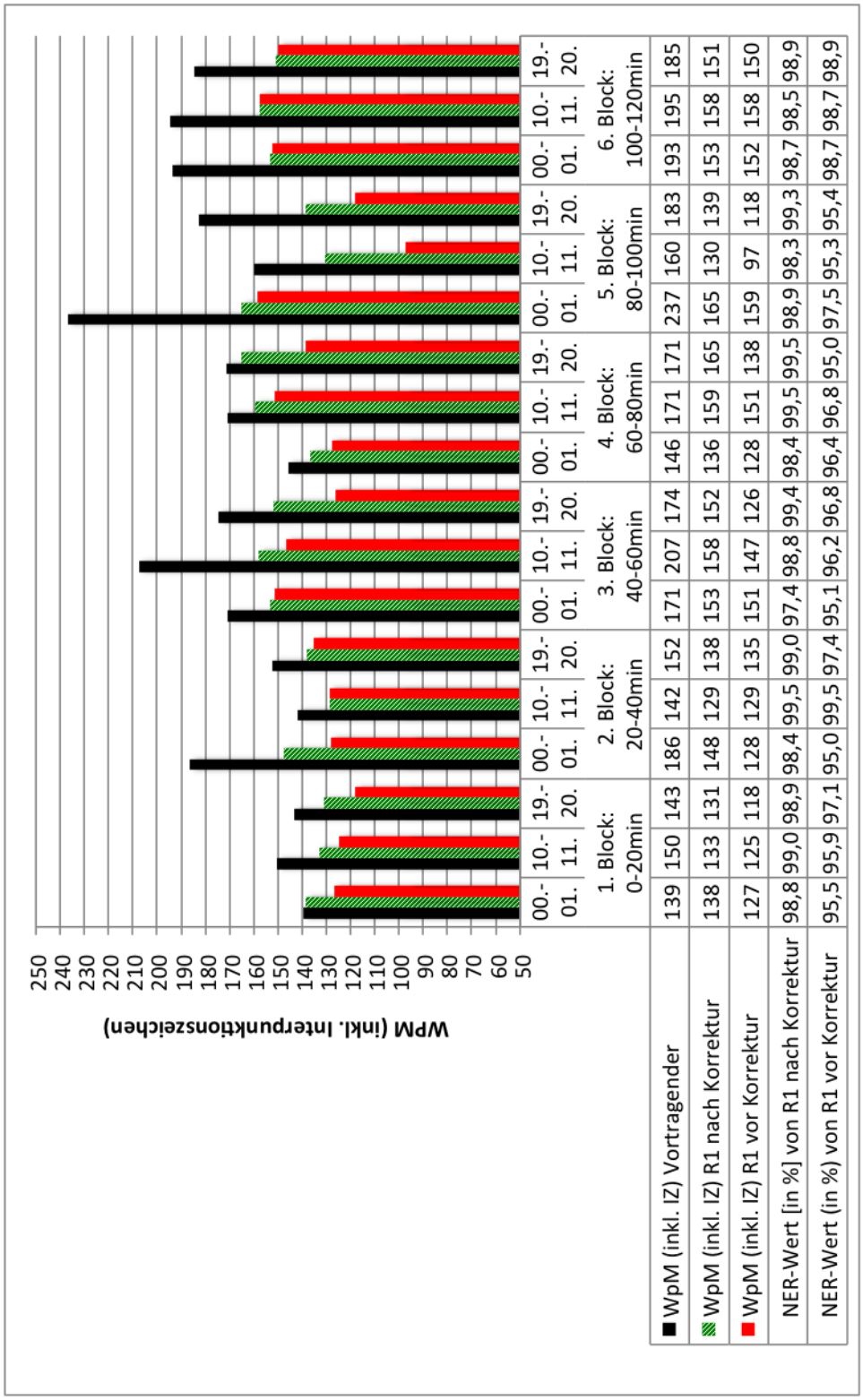
		geringfügig	normal	gravierend	geringfügig	normal	gravierend	geringfügig	normal	gravierend
		R1			R2			R3+R4		
Vor Editierung	Editierfehler	18	16	2	18	20	6	11	101	11
	Erkennungsf.	122	70	1	89	55	3	26	15	1
	Summe	**140**	**86**	**3**	**107**	**75**	**9**	**37**	**116**	**12**
Nach Editierung	Editierfehler	20	5	1	18	19	4	/	/	/
	Erkennungsf.	60	14	1	74	38	2	/	/	/
	Summe	**80**	**19**	**2**	**92**	**57**	**6**	/	/	/
Verb. durch die Korrektur (%)		42,86	77,9	33,4	14,0	24,0	33,4	/	/	/

Tabelle 4.3: NER: Anzahl der Fehler je Fehlerursache und Fehlertyp; Verbesserung durch die Korrektur

Diktiergeschwindigkeiten/Wortanzahl in den Transkripten

In den analysierten Stichproben unterscheidet sich jeweils die von *R1*, *R2*, *R3+R4* transkribierte Anzahl von Wörtern je Minute. Durch das Vergleichen der Wortanzahl in den Untertiteln mit jener des Vortragenden (Sprechgeschwindigkeit), lassen sich verschiedene Ansätze des Kürzens sowie eventuelle Probleme während des Respeakings erkennen. In Abbildung 4.4 auf Seite 119 ist die Wortanzahl von *R1* des mittels Scripting erzeugten Transkripts (*vor* der Korrektur) der Sprechgeschwindigkeit des Vortragenden in einem Balkendiagramm gegenübergestellt. Auch die Erhöhung der Wortanzahl im Transkript durch die anschließende Korrektur ist im Balkendiagramm dargestellt. Darüber hinaus sind die NER-Werte der jeweiligen Stichprobe *vor* und *nach* erfolgter Korrektur in der Legende (des Balkendiagramms) angegeben. In ähnlicher Weise sind in der Abbildung 4.5 auf Seite 120 die Werte von Respeaker *R2* visualisiert. Wie im Abschnitt 4.3.1 auf Seite 103 erläutert, transkribierte *R2* die Aufzeichnung mit einer Wiedergabegeschwindigkeit von 90%[36]. Aus diesem Grund beziehen sich die Werte der Abbildung 4.5 auf Seite 120 auf die tatsächlich transkribierten Wörter pro Minute und somit nicht (wie bei *R1* und *R3+R4*) auf die Diktiergeschwindigkeit. Der Wert von 107 WpM der ersten Stichprobe im ersten Block bedeutet demnach, dass *R2* innerhalb dieser Minute ca. 92 Wörtern diktierte. Wie in diesem Kapitel erläutert, wurden die Untertitel von *R3+R4* live erstellt. Dabei wurde - im Gegensatz zu *R1* und *R2* - keine anschließende Korrektur der Untertitel durchgeführt. Daher ist in der Abbildung 4.6 auf Seite 121 ausschließlich die Wortanzahl der von *R3+R4* erstellten Untertitel in Relation zur Sprechgeschwindigkeit des Vortragenden dargestellt bzw. sind die erzielten NER-Werte in der Legende des Balkendiagramms angeführt.
Das durch Respeaking erstellte Transkript von *R1* (*vor* der Korrektur) enthält durchschnittlich[19] um 21,6% (Median von 19,0%) weniger Wörter pro Minute[37] als von Mag. Zeilinger vorgetragen wurden. Das vom Respeaker *R2* - ebenfalls *vor* der Korrektur - erstellte Transkript enthält durchschnittlich[19] um 31,8% (Median von 31,6%) weniger Wörter als vorgetragen. Das lässt darauf schließen, dass sich *R1* mehr an dem Gesagten orientierte und *R2* mehr kürzte und umformulierte. Bei *R3+R4* kam es zu deutlich mehr Kürzungen. Im Durchschnitt enthält das Transkript um 52,1% (Median 52,2%) weniger Wörter als vom Vortragenden gesprochen wurden. Das führt unweigerlich auch zu einem höheren Informationsverlust, was sich sehr deutlich in den NER-Werten (siehe u.a. Tabelle 4.1 auf Seite 113) widerspiegelt. Wie in diesem Kapitel bereits erläutert, handelt es sich bei *R3+R4* nicht ausschließlich um gezieltes Kürzen. Die Analyse lässt auf auch Ausfälle - vermutlich aufgrund technischer Probleme - schließen. Auch die Steigerung der Wortanzahl durch die anschließende Korrektur macht die unterschiedlichen Korrekturansätze von *R1* und *R2* deutlich. So editierte *R1* deutlich mehr und verringerte die durchschnittlich[19] geringere Wortanzahl im Vergleich zum Gesprochenen von 21,6% auf 11,2%. *R2* verringerte im Vergleich zu *R1* die Wortanzahl zum Gesprochenen weniger stark: von 31,8% *vor* der Korrektur auf 29,3% *nach* der Korrektur. Daraus resultierte der höhere Aufwand für die Untertitelerstellung von *R1*, siehe Abschnitt 4.3.6 ab Seite 128.

[36] Anm. Autor: Mit Ausnahme von fünf Minuten, in denen *R2* mit 100% Wiedergabegeschwindigkeit transkribierte, siehe Abschnitt 4.3.1 ab Seite 103.
[37] Anm. Autor: In diesem Abschnitt inkludierten sämtliche Wortanzahlangaben die Interpunktionszeichen.

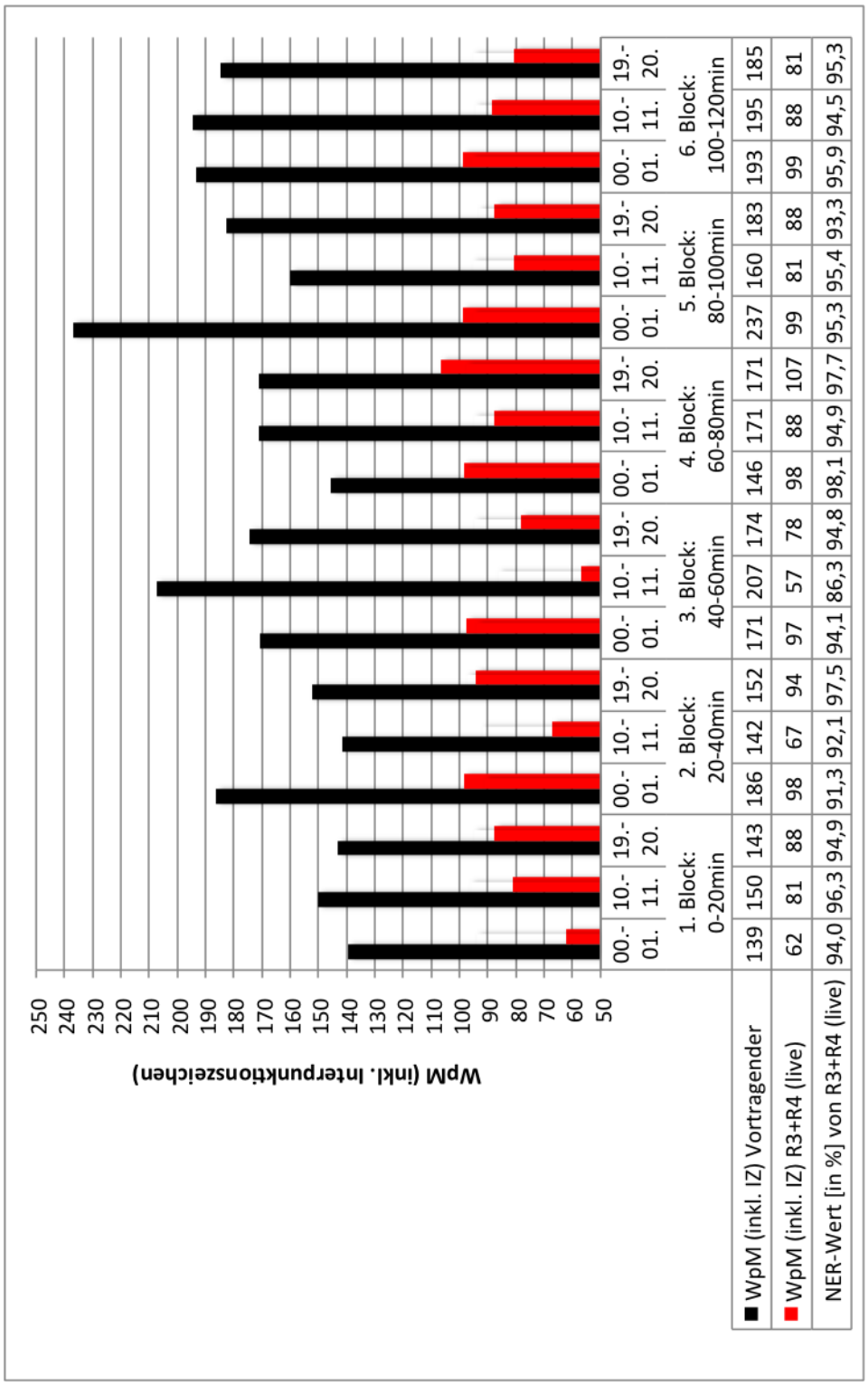


	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.
	1. Block: 0-20min			2. Block: 20-40min			3. Block: 40-60min			4. Block: 60-80min			5. Block: 80-100min			6. Block: 100-120min		
WpM (inkl. IZ) Vortragender	139	150	143	186	142	152	171	207	174	146	171	171	237	160	183	193	195	185
WpM (inkl. IZ) R1 nach Korrektur	138	133	131	148	129	138	153	158	152	136	159	165	165	130	139	153	158	151
WpM (inkl. IZ) R1 vor Korrektur	127	125	118	128	129	135	151	147	126	128	151	138	159	97	118	152	158	150
NER-Wert [in %] von R1 nach Korrektur	98,8	99,0	98,9	98,4	99,5	99,0	97,4	98,8	99,4	98,4	99,5	99,5	98,9	98,3	99,3	98,7	98,5	98,9
NER-Wert (in %) von R1 vor Korrektur	95,5	95,9	97,1	95,0	99,5	97,4	95,1	96,2	96,8	96,4	96,8	95,0	97,5	95,3	95,4	98,7	98,7	98,9

Abbildung 4.4: Diktiergeschwindigkeit von *R1*

	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.
	1. Block: 0-20min			2. Block: 20-40min			3. Block: 40-60min			4. Block: 60-80min			5. Block: 80-100min			6. Block: 100-120min		
WpM (inkl. IZ) Vortragender	139	150	143	186	142	152	171	207	174	146	171	171	237	160	183	193	195	185
WpM (inkl. IZ) R2 nach Korrektur	107	109	103	93	99	110	129	128	110	102	117	134	138	116	120	141	148	108
WpM (inkl. IZ) R2 vor Korretkur	107	109	104	94	99	110	129	128	105	101	118	133	140	117	122	141	148	105
NER-Wert [in %] von R2 nach Korrektur	98,4	98,3	97,4	98,4	97,5	98,1	96,6	97,0	97,7	97,6	97,7	98,3	96,9	98,3	94,7	97,2	99,5	95,2
NER-Wert [in %] von R2 vor Korretkur	98,4	97,9	97,2	94,4	97,5	98,1	96,6	97,0	96,3	96,4	95,4	97,5	95,0	96,9	94,8	96,8	97,9	95,8

Abbildung 4.5: Wortanzahl in den Transkripten von *R2*



	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.	00.-01.	10.-11.	19.-20.
	1. Block: 0-20min			2. Block: 20-40min			3. Block: 40-60min			4. Block: 60-80min			5. Block: 80-100min			6. Block: 100-120min		
WpM (inkl. IZ) Vortragender	139	150	143	186	142	152	171	207	174	146	171	171	237	160	183	193	195	185
WpM (inkl. IZ) R3+R4 (live)	62	81	88	98	67	94	97	57	78	98	88	107	99	81	88	99	88	81
NER-Wert [in %] von R3+R4 (live)	94,0	96,3	94,9	91,3	92,1	97,5	94,1	86,3	94,8	98,1	94,9	97,7	95,3	95,4	93,3	95,9	94,5	95,3

Abbildung 4.6: Diktiergeschwindigkeit von *R3+R4*

Streudiagramme

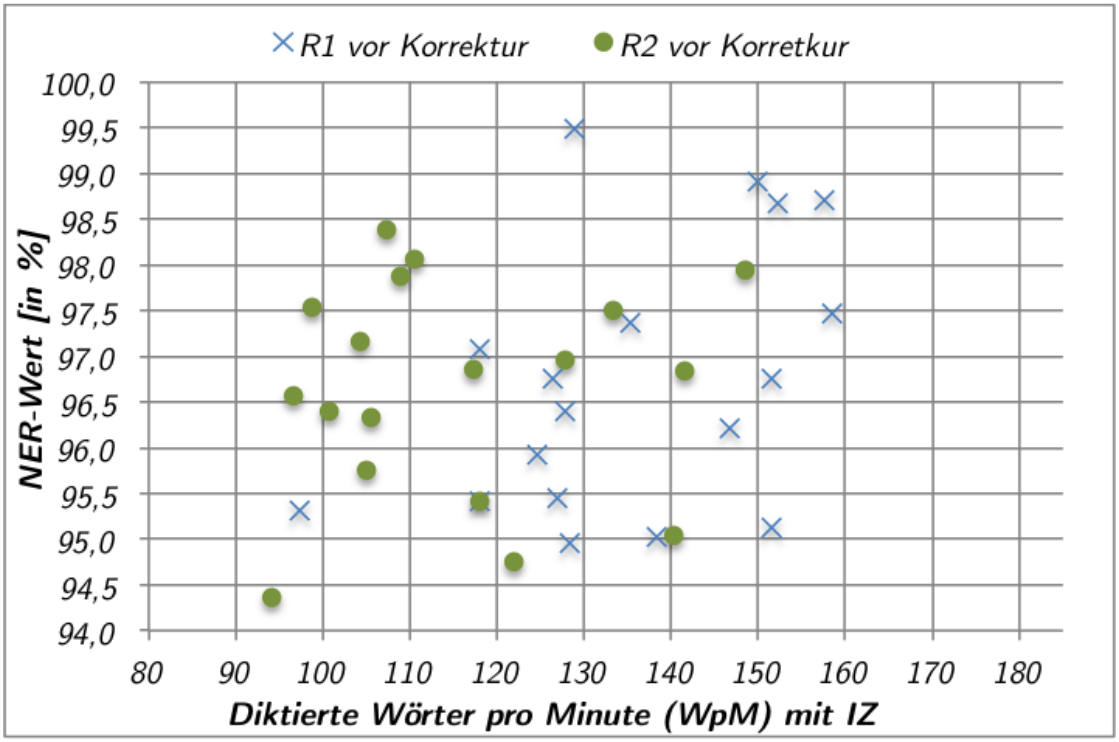
Die Streudiagramme in diesem Abschnitt dienen zur grafischen Darstellung der Abhängigkeit von Wertepaaren. Neben der Auswirkung der Sprechgeschwindigkeit des Vortragenden auf die NER-Werte, ist die Auswirkung der Wortanzahl auf die NER-Werte in den Diagrammen visualisiert. Wie im Abschnitt 4.3.1 auf Seite 103 sowie im Abschnitt 4.3.4 auf Seite 118 erläutert, transkribierte *R2* die Aufzeichnung mit einer Wiedergabegeschwindigkeit von 90%[14]. Die Werte in den Streudiagramme von *R2* beziehen sich daher auf die tatsächlich transkribierten Wörter pro Minute und nicht wie bei *R1* und *R3+R4* auf die Diktiergeschwindigkeit.
Die zwei Diagramme in Abbildung 4.7 auf Seite 124 verdeutlichen die Steigerung der NER-Werte durch die Korrektur von *R1* und *R2* ebenso wie die dadurch zu verzeichnende Veränderung der Wortanzahl. Durch die Korrektur erhöhte sich die Wortanzahl im Transkript von *R1*. Weiters ist eine Qualitätssteigerung erkennbar. Es konnte die Qualität der Stichproben mit weniger als 130 WpM[38] und einem NER-Wert unter 98,0% durch die Korrektur und das Ergänzen von fehlender Information auf jeweils über 98,0% erhöht werden. Betrachtet man die Wertepaare von *R2*, so ist die Steigerung der Qualität weniger durch eine Ergänzung des Transkripts und daraus resultierend einer Erhöhung der Wortanzahl gegeben als vielmehr durch die gezielte Fehlerkorrektur. Weiters verdeutlichen die beiden Streudiagramme auch, dass bei *R2* durch die gewählte Korrekturmethode die Steigerung der NER-Werte geringer ist.
In Abbildung 4.8 ab Seite 125 zeigen die beiden Streudiagramme das Verhältnis des Sprechtempos des Vortragenden zu den NER-Werten von *R1* sowie *R2*. Die Diagramme heben hervor, dass eine geringe Sprechgeschwindigkeit des Vortragenden nicht automatisch zu höheren NER-Werten führte. Ebenso zeigt sich, dass aus hohen Sprechgeschwindigkeiten nicht zwingend niedrige NER-Werte resultierten. Aus Sicht des Autors dieser Diplomarbeit ist das Phänomen auf mehrere Ursachen zurückzuführen. So kann auch bei langsamem Sprechen viel Inhalt transportiert werden und umgekehrt bei hohen Sprechgeschwindigkeiten eine höhere Redundanz des Gesagten vorhanden sein (und demnach ein Kürzen ohne inhaltlichen Informationsverlust ermöglichen). Darüber hinaus pausierte *R1* die Wiedergabe der zweistündigen Vorlesung für insgesamt 13 Minuten, *R2* für 72 Minuten, siehe Abschnitt 4.3.6 ab Seite 128. Es ist nicht dokumentiert zu welchen Zeitpunkten die Unterbrechungen stattfanden. Daher wäre es möglich, dass sie (teilweise) bei hohen Sprechgeschwindigkeiten von über 195 WpM erfolgten. Weiters ist die Sprechgeschwindigkeit des Vortragenden nur ein Parameter, welcher die Schwierigkeit beim Respeaking definiert. Auch die momentane Konzentration des Respeakers oder der Respeakerin kann die Qualität beeinflussen.
Wie im Abschnitt 4.3.1 beschrieben, markierte *R2* während des Respeaking die Fehler und korrigierte sie nachträglich. Betrachtet man die Steigerung der NER-Werte durch die Korrektur, so ist zu sehen, dass *R2* sehr viele Fehler im Bereich der niedrigeren Sprechgeschwindigkeiten von 135-165 WpM korrigierte. Das lässt darauf schließen, dass er bei hohen Sprechgeschwindigkeiten weniger effizient Fehler erkannte bzw. markierte. Ein gezieltes Training könnte hier dazu beitragen, dass *R2* auch bei hohen Sprechgeschwindigkeiten Fehler in den Untertiteln besser erkennt und markiert.
Aus Übersichtsgründen sind die Werte von *R3+R4* in eigenen Diagrammen visualisiert und nicht

[38] Anm. Autor: In diesem Abschnitt inkludierten sämtliche Wortanzahlangaben die Interpunktionszeichen.

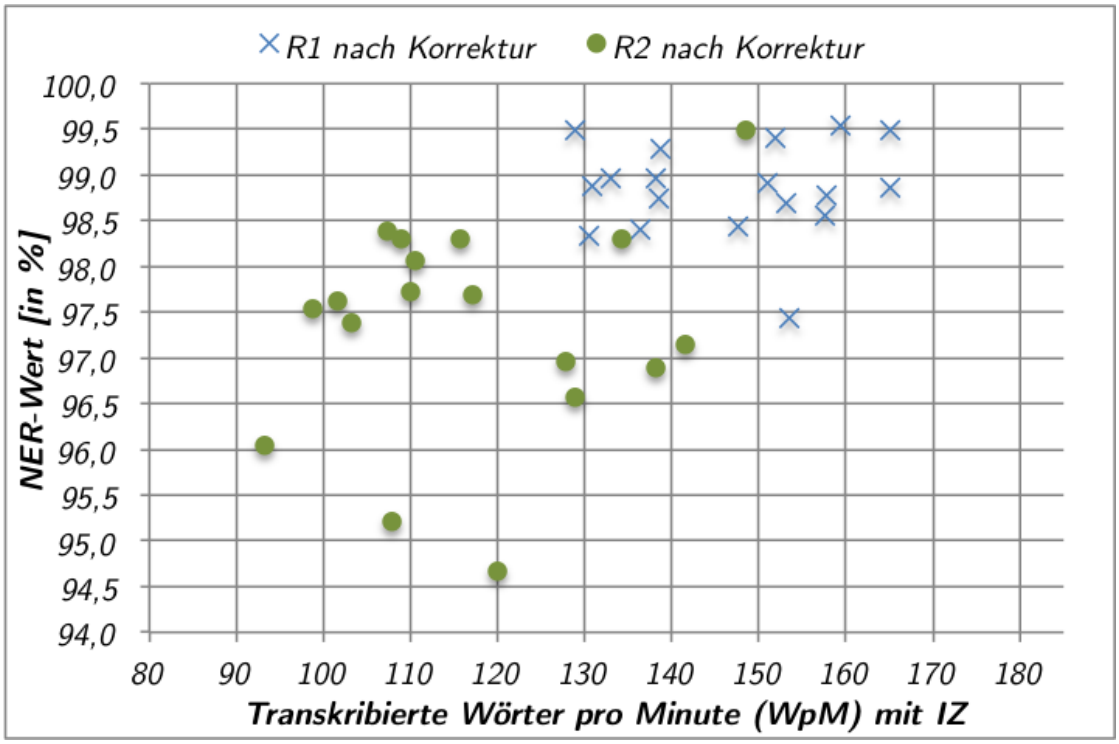
zusammen mit jenen von *R1* und *R2* dargestellt. Es ist zu beachten, dass die Abbildung 4.9 auf Seite 126 andere Skalen für X und Y-Werte aufweist als die Abbildung 4.7 und Abbildung 4.8 auf den Seiten 124-125. In Abbildung 4.9 stellt das obere Streudiagramm die Abhängigkeit des Sprechtempos des Vortragenden zu den NER-Werten des Transkripts von *R3+R4* dar. Das untere Diagramm visualisiert die Abhängigkeit der Respeaking Diktiergeschwindigkeiten zu den NER-Werten. Hervorzuheben ist, dass die drei höchsten NER-Werte bei Sprechgeschwindigkeiten von 145-175 WpM erzielt wurden, jedoch bei Sprechgeschwindigkeiten von unter 145 WpM eine geringere Akkuratheit erreicht wurde. Wie auch bei *R1* und *R2* zeigt sich, dass eine geringe Sprechgeschwindigkeit des Vortragenden nicht automatisch zu höheren NER-Werten führt. Betrachtet man die Auswirkung der Diktiergeschwindigkeit von *R3+R4* auf die NER-Werte, so ist zu erkennen, dass die Qualität tendenziell bei höheren Diktiergeschwindigkeiten (also geringerem Kürzen) steigt. Vergleicht man die beiden Streudiagramme von *R3+R4*, sieht man im Vergleich zu *R1* und *R2* weiters den höheren Grad der Kürzung.

4.3.5 Zeitcodes der Untertitel

Wie im Abschnitt 2.5.4 ab Seite 38 erläutert, kann Dragon ohne zusätzliche Software keine Zeitcodes erzeugen bzw. exportieren. Aus diesem Grund verwendeten *R1* und *R2* die Software Subtitle Workshop, siehe Abschnitt 3.1 ab Seite 52. Mit dieser Software erzeugten sie während der Respeaking Tätigkeit die Intervalle für die Zeitcodes manuell mittels Tastenkombination. Kürzere Untertitelblöcke wirken sich positiv auf die Navigation, Synchronität und Durchsuchbarkeit bei einer in Synote aufbereiten Vorlesung aus, siehe Abschnitt 3.1.1 ab Seite 52.
R1 erstellte für die zweistündige Vorlesung insgesamt 358 Intervalle. Die Dauer der Intervalle beträgt dabei im Median 16,4 Sekunden, das längste Intervall ist 48,7 Sekunden. Im Vergleich dazu erstellte *R2* insgesamt 119 Intervalle. Der Median davon beträgt 51,5 Sekunden, die maximale Intervalldauer ist drei Minuten und 56 Sekunden.

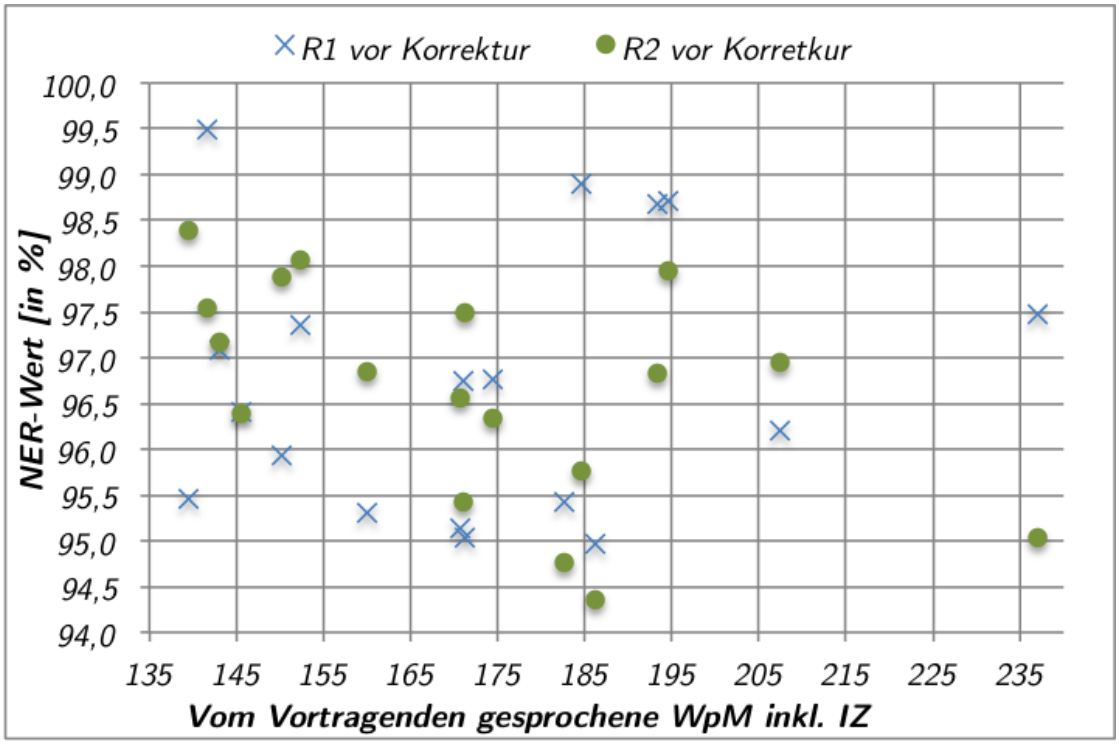


(a) Vor der Korrektur

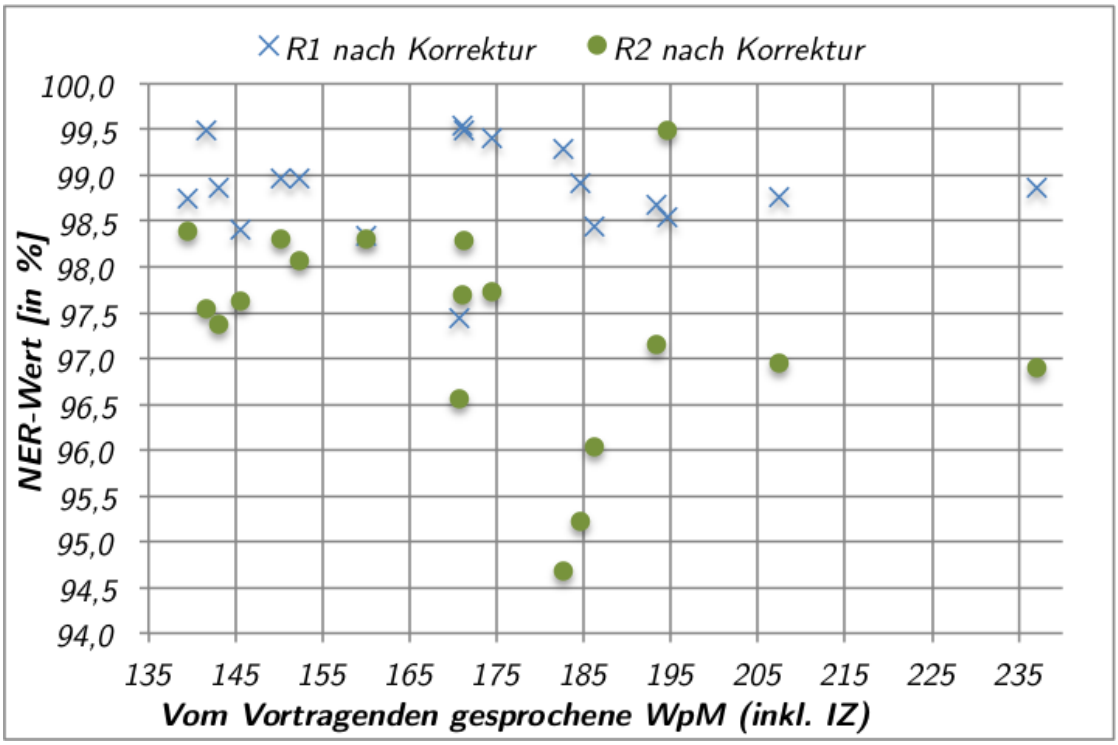


(b) Nach der Korrektur

Abbildung 4.7: Auswirkung der Wortanzahl auf NER-Werte (*R1* und *R2*)

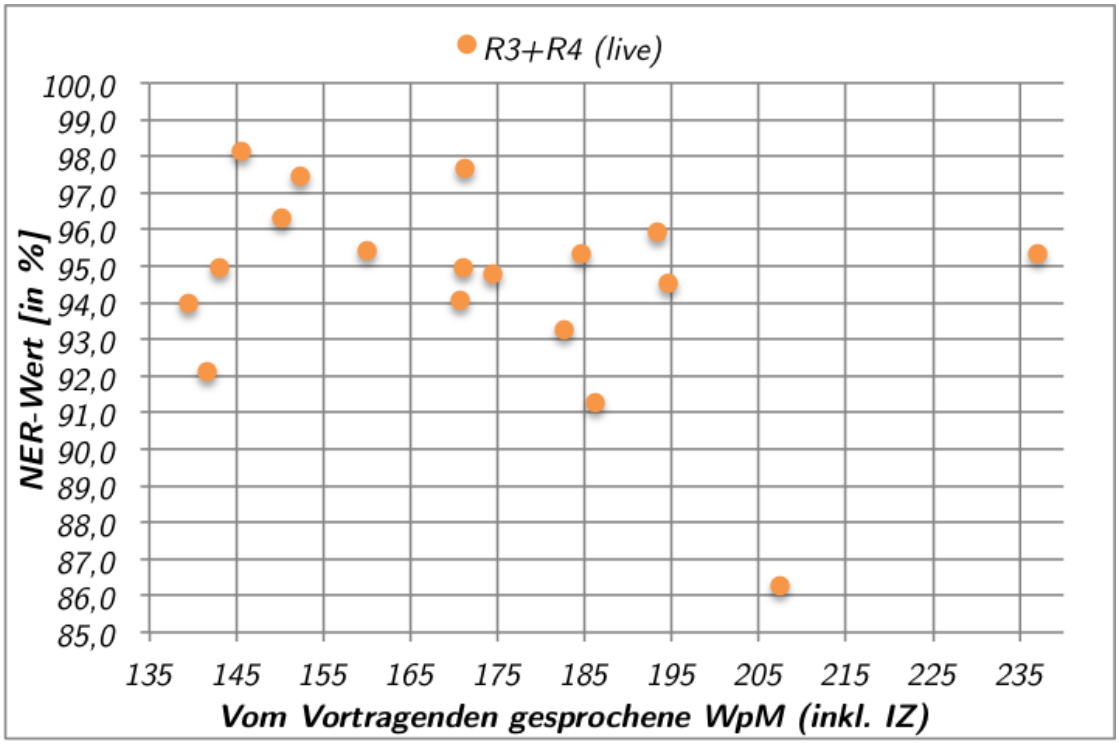


(a) Vor der Korrektur

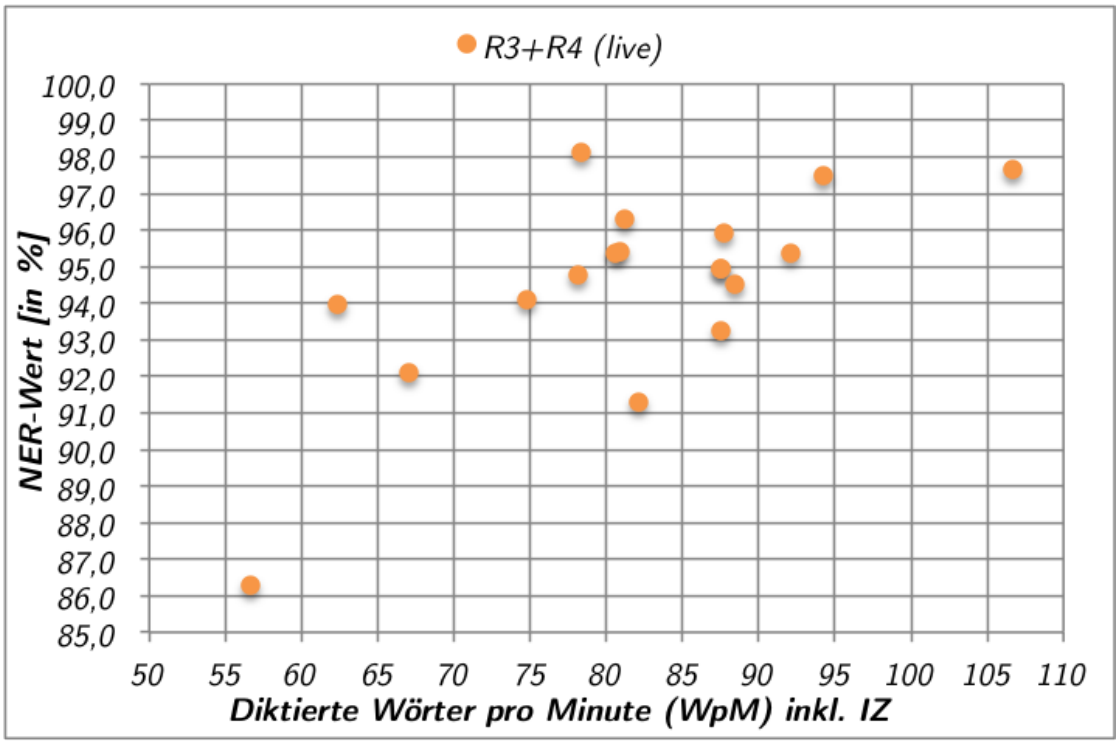


(b) Nach der Korrektur

Abbildung 4.8: Auswirkung der Sprechgeschwindigkeit auf NER-Werte (*R1* und *R2*)



(a) Auswirkung der Sprechgeschwindigkeit auf NER-Werte



(b) Auswirkung der Diktiergeschwindigkeit auf NER-Werte

Abbildung 4.9: Auswirkung der Sprech- und Diktiergeschwindigkeit auf NER-Werte (R3+R4)

Stichprobe	Vorbereitung [min]	Respeaking [min]	Korrektur [min]	Erholungspause [min]	Gesamt [min]	Verhältnis excl. Korrektur (ratio)	Verh. inkl Korrektur (ratio)	NER-Wert [in %]	NER-Wert Selbsteinschätzung [in %]
1. Block: 0-20 min	0,5	21	50	15	86,5	1,8:1	4,3:1	98,9	98,5
2. Block: 20-40 min	0,5	22	70	15	107,5	1,9:1	5,4:1	98,9	99,0
3. Block: 40-60 min	0,5	22	40	15	77,5	1,9:1	3,9:1	98,5	99,0
4. Block: 60-80 min	10,5	22	45	15	92,5	2,4:1	4,6:1	99,2	99,5
5. Block: 80-100 min	0,5	22	40	15	77,5	1,9:1	3,9:1	98,8	99,5
6. Block: 100-120 min	0,5	24	35	15	74,5	2,0:1	3,7:1	98,7	99,5
Gesamt	**13**	**133**	**280**	**90**	**516**	**2,0:1**	**4,3:1**	**98,9**	**99,2**

Tabelle 4.4: Aufwand *R1* je Block (inkl. Korrektur und Pausen)

Stichprobe	Vorbereitung [min]	Respeaking [min]	Korrektur [min]	Erholungspause [min]	Gesamt [min]	Verhältnis excl. Korrektur (ratio)	Verh. inkl Korrektur (ratio)	NER-Wert [in %]	NER-Wert Selbsteinschätzung [in %]
1. Block: 0-20 min	0,5	25	13	15	53,5	2,0:1	2,7:1	98,0	98,7
2. Block: 20-40 min	0,5	30	13	10	43,5	2,0:1	2,7:1	97,1	98,5
3. Block: 40-60 min	0,5	30	13	10	43,5	2,0:1	2,7:1	97,1	98,5
4. Block: 60-80 min	0,5	35	26	10	61,0	2,3:1	3,6:1	97,9	98,2
5. Block: 80-100 min	0,5	38	21	10	59,0	2,4:1	3,5:1	96,7	98,0
6. Block: 100-120 min	0,5	53	13	10	66,0	3,2:1	3,8:1	97,7	98,7
Gesamt	**3**	**211**	**99**	**65**	**378**	**2,3:1**	**3,2:1**	**97,4**	**98,4**

Tabelle 4.5: Aufwand *R2* je Block (inkl. Korrektur und Pausen)

4.3.6 Aufwand und Selbsteinschätzung

Aufwand

Bisher wurden verschiedene qualitative Aspekte der Untertitel dargelegt und analysiert. Im Folgenden sind die Aufwände für die Erstellung der Transkripte dokumentiert. In Tabelle 4.4 auf Seite 127 ist der Zeitaufwand von *R1* aufgelistet, jener von *R2* in Tabelle 4.5 auf Seite 127.
Wie im Abschnitt 4.3.1 ab Seite 103 dokumentiert, wurden zur Analyse die Transkripte in sechs (á 20 Minuten) Blöcke unterteilt. Die einmalige Vorbereitungszeit von *R1* und *R2* wurde gleichmäßig auf die sechs Blöcke aufgeteilt. Im 4. Block sind bei *R1* zehn Minuten zusätzlich als Vorbereitungszeit angeführt, die für einen Batteriewechsel beim Funkmikrofon benötigt wurden.
Der Aufwand der Respeaking Tätigkeit (demnach ohne Vorbereitung, Korrektur, Pausen, etc.) betrug bei *R1* 133 Minuten, demnach 13 Minuten mehr als die reine Vorlesungsdauer. Die Differenz begründet sich durch das Pausieren (und ggf. dem nochmaligem Wiedergeben) der Aufnahme, um unmittelbar Fehler zu korrigieren bzw. auch um sie für die spätere Korrektur zu markieren. Zusammen mit den Pausen (insgesamt 90 Minuten) betrug der Aufwand für das Respeaking bei *R1* 223 Minuten. Im Vergleich dazu benötigte *R2* 211 Minuten, inklusive der Erholungspausen (65 Minuten) 276 Minuten. *R2* transkribierte die Aufzeichnung mit einer Wiedergabegeschwindigkeit von 90%[14]. Dadurch ergab sich eine Zeit von 72 Minuten, in der er die Aufnahme pausierte. *R3+R4* arbeiteten im Team und somit insgesamt 240 Minuten für die zweistündige Vorlesung. Die Vorbereitungszeit von zehn Minuten bis zu einer Stunde je Respeakerin bzw. Respeaker ergibt einen Gesamtaufwand von 260 bis 360 Minuten.
Die Aufwände für die Korrektur unterschieden sich bei *R1* und *R2* sehr deutlich. So benötigte *R1* 280 Minuten (also das 2,33-fache der Vorlesungszeit) und *R2* 99 Minuten für die anschließende Korrektur. Der zeitliche Unterschied verdeutlicht die verschiedenen Arbeitsweisen der beiden Scripter.
Für das online Stellen der Vorlesung (Aufbereitung der Folien, das Hochladen des Transkripts, etc.) benötigte *R1* 55 Minuten, *R2* insgesamt 65 Minuten.

Selbsteinschätzung

Aus Sicht des Autors dieser Diplomarbeit ist es wichtig, dass Respeaker und Respeakerinnen selbst ohne eine ausführliche Analyse die erreichte Untertitelqualität gut einschätzen können. Deshalb sind in den Tabellen auf Seite 127 jeweils zwei NER-Werte angeführt. Die tatsächlich erreichten Werte sind jenen NER-Werten gegenübergestellt, die *R1* bzw. *R2* schätzten. Die Selbsteinschätzungen erfolgten direkt *nach* erfolgter Korrektur des transkribierten Blocks und vor einer NER-Analyse. *R1* schätzte den NER-Wert durchschnittlich um 0,3% geringer als den tatsächlich erreichten Wert ein. *R2* erreichte einen um 1,0% geringeren Wert im Transkript als er selbst schätzte. *R2* führte nach der Schätzung selbständig eine Analyse des Transkripts durch. Die von ihm dokumentierten Werte liegen dabei näher an seiner Selbsteinschätzung. Er schätzte im Durchschnitt den NER-Wert des Transkripts auf 98,4% ein und erzielte lt. seiner folgenden Analyse einen NER-Wert von 98,9% (anstatt der vom Autor dieser Diplomarbeit festgestellten 97,4%). Die Unterschiede zwischen der vom Autor dieser Diplomarbeit durchgeführten Analyse und der von *R2* ergeben sich durch Fehler im Transkript, die *R2* nicht fand bzw. teilweise anders bewertete. Bereits während der Ausbildung übersah *R2* bei der NER-Analyse einige Fehler

bzw. beurteilte diese anders. Da seine Selbsteinschätzung auf diesen Analysen beruht, kann sie aus Sicht des Autors dieser Diplomarbeit als zufriedenstellend eingestuft werden.

4.3.7 Zusammenfassung der Interpretation

Ausgangssituation

Die in diesem Kapitel bereits aus verschiedenen Blickwinkeln betrachteten Ergebnisse sind im Folgenden zusammengefasst und interpretiert.
Wie erläutert, wurde eine zweistündigen Vorlesung der Vorlesungsreihe *Trainingswissenschaft* durch die Firma Titelbild (*R3+R4*) in Echtzeit untertitelt. Die Videoaufzeichnung der Vorlesung wurde weiters vom Ausgebildeten (*R2*) und vom Autor dieser Diplomarbeit (*R1*) mittels Scripting transkribiert. Die qualitative Beurteilung aller erstellten Untertitel erfolgte durch die NER-Analyse, die der Autor dieser Diplomarbeit durchführte. Dabei wurden von *R1* und *R2* je zwei Transkripte (*vor* und *nach* erfolgter Korrektur) betrachtet. Eine Zielsetzung der Diplomarbeit war auch der Vergleich zu den Untertiteln, die automatisch (ohne Respeaking) von der Spracherkennungssoftware des EML erzeugt wurden. Aufgrund mangelnder Qualität war es weder sinnvoll noch praktisch möglich, diese Untertitel qualitativ mittels des NER-Modells zu analysieren. Somit standen fünf Transkripte für die Evaluierung zur Verfügung. Für eine effektive und trotzdem repräsentative Analyse wurden die Transkripte in sechs 20 Minuten Blöcke unterteilt. Von jedem Bock wurde jeweils die erste, die zehnte und die letzte Minute ausgewertet.

Transkripte von *R1*

Mit einem NER-Wert von 98,9% weist das Transkript von *R1 nach* erfolgter Korrektur die höchste Qualität aller Transkripte auf. Dabei erreichten 17 der 18 Stichproben den angestrebten NER-Wert von mindestens 98,0%. Darüber hinaus beinhaltet das Transkript in allen drei Fehlerkategorien (*gravierende*, *normale* und *geringfügige* Fehler) die geringste Anzahl von Fehlern. Auch in Bezug auf die qualitative Konstanz zwischen den einzelnen Stichproben erzielte das Transkript von *R1 nach* erfolgter Korrektur das beste Ergebnis. Weiters ist die Anzahl der erstellten Untertitelzeilen im direkten Vergleich zu denen von *R2* höher und konstanter. Den Untertiteln in dieser Qualität steht aber auch der zeitlich größte Aufwand[39] gegenüber. Der zeitliche Aspekt stellt demnach das größte Verbesserungspotential bei der Arbeitsweise von *R1* dar.

Transkripte von *R2*

Das Transkript von *R2 nach* erfolgter Korrektur weist mit einem NER-Wert von 97,4% die zweithöchste Qualität der fünf Transkripte auf. Dabei beträgt der NER-Wert in 12 der 18 Stichproben mindestens die angestrebten 98,0%. Auch in Bezug auf die Konstanz der Qualität zwischen den einzelnen Stichproben erzielte das Transkript das zweitbeste Ergebnis. Weiters weist das Transkript die zweit geringste Fehleranzahl von *gravierenden* sowie *normalen* Fehlern auf.

[39] Für die 120 Minuten dauernde Vorlesung wurden 516 Minuten (inkl. Pausen) aufgewandt und zusätzliche 55 Minuten für die Einbindung in die E-Learning Plattform Synote.

Die vergleichsweise hohe Anzahl von *geringfügigen* Fehlern ist im wesentlichen auf zwei Aspekte zurückzuführen. 38% der Fehler dieser Fehlerklasse sind auf das beschriebene Schnittstellenproblem zwischen Dragon und Subtitle Workshop zurückzuführen. Der zweite Faktor ist die angewandte Korrekturmethode von *R2*. So fokussierte er sich während des Respeakings auf das Markieren von gravierenden und normalen Fehlern, die dann in der Korrekturphase korrigiert wurden. Die von *R2* angewandte Methode stellt eine sehr zeiteffiziente[40] Methode zur Untertitelerstellung dar. Aus Sicht des Autors dieser Diplomarbeit ist jedoch mit dieser Korrekturmethode durchaus auch ein NER-Wert von 98,0% möglich. Wäre die Kleinschreibung am Satzbeginn korrigiert worden bzw. durch die Verbesserung der Schnittstelle nicht aufgetreten und zusätzlich ein Viertel der normalen Erkennungsfehler während des Respeakings erkannt und später korrigiert worden, so wäre bereits ein NER-Wert von 98,0% erreicht worden. Daher wäre ein weiteres, gezieltes Training auf das Erkennung und Markieren von Fehlern anzustreben. Damit könnte mit einem vergleichbar geringen Aufwand die angestrebte Qualität erreicht werden. Für eine gute Navigation, Synchronität und Durchsuchbarkeit der aufbereiteten Vorlesung in Synote ist die Intervalllänge der Zeitcodes entscheidend. Aus Sicht des Autors dieser Diplomarbeit ist Anzahl der von *R2* erstellten Untertitelzeilen zu gering bzw. die Intervalle zu groß, um die Vorteile von Synote zu gewährleisten. Ein gezieltes Training von wenigen Stunden könnte diese Schwäche in den Untertiteln anzunehmender Weise ohne wesentliche Auswirkung auf die NER-Werte entgegenwirken.

Transkript von *R3+R4* (live)

Das in Echtzeit erstellte Transkript von *R3+R4* (*ohne Korrektur*) weist in vielen Aspekten die geringste Qualität der fünf verglichenen Transkripte auf. Der NER-Wert beträgt 94,9%, eine der 18 Stichproben erreichte den angestrebten NER-Wert von mindestens 98,0%. Gerade die Anzahl der *gravierenden* und *normalen* Editierfehler ist für die vergleichsweise niedrige Qualität verantwortlich. So hätten *R3+R4* um 74% weniger *gravierende* sowie *normale* Editierfehler erzielen müssen, um die angestrebten 98,0% NER-Wert zu erreichen. In Bezug auf die Konstanz der Qualität zwischen den einzelnen Stichproben erzielte lediglich das Transkript von *R1 vor* erfolgter Korrektur ein schlechteres Ergebnis. Allerdings ist der NER-Wert von *R1 vor* der Korrektur gesamt sowie im Median höher als jener von *R3+R4*. Den vergleichsweisen geringen Qualität der Untertiteln steht allerdings der geringste Aufwand[41] für die Erstellung gegenüber. Abschließend wird nochmals darauf hingewiesen, dass *R3+R4* unter schwierigeren Ausgangsbedingungen gearbeitet haben (sie hatten nur die akustische Übertragung des Gesagten, arbeitenden live und im Team und waren demnach einem höheren Stresspotential ausgesetzt, etc.).

[40] Anm. Autor: Die Erstellung der Untertitel war mit 378 Minuten um ca. 26,7% geringer als bei *R1*. Hinzu kommen 65 Minuten Aufwand für die Einbindung in die E-Learning Plattform Synote.

[41] Anm. Autor: Für die 120 Minuten dauernde Vorlesung wurden zwischen 260 und 360 Minuten (abhängig davon, ob die benötigte Vorbereitungszeit mit á 10 Minuten oder einer Stunde kalkuliert wird) aufgewandt.

KAPITEL 5

Schlusswort und Ausblick

In Österreich leben mindestens 500.000 hörbeeinträchtigte Menschen. Hörbeeinträchtigte Personen sind im österreichischen Bildungssystem zum Teil mit erheblichen Hürden konfrontiert. Diese reichen von **finanziellen**, **organisatorischen** bis hin zu **pädagogischen Barrieren**. Im tertiären Bildungssektor ist die Anzahl der Studierenden mit Hörbeeinträchtigung vergleichsweise gering. Das Pilotprojekt GESTU unterstütze zwischen Juli 2010 und Juli 2012 die teilnehmenden hörbeeinträchtigten Studentinnen und Studenten dabei, ihr Studium zeitgerecht und erfolgreich zu absolvieren. Eine Servicestelle übernahm u.a. die Organisation von Gebärdensprachdolmetscherinnen und Gebärdensprachdolmetschern, von Tutoren und Tutorinnen sowie Mitschreibhilfen. Weiters leistete das GESTU Team eine wichtige Arbeit in der Sensibilisierung von Mitstudierenden und den mit der Lehre beauftragten Personen. Als Teammitglied konnte ich im ersten Pilotprojektjahr von GESTU im Zuge meiner vorangegangenen Diplomarbeit zur Evaluierung von technischen Hilfsmitteln und deren Einführung beitragen. Bei der im Jahr 2011 veröffentlichten Diplomarbeit lag der Fokus bei Hilfsmitteln, welche die Informationsaufnahme von Lehrinhalten und der Partizipation *während* Vorlesungen, Diskussionen, etc. ermöglichen. Aufbauend auf den Empfehlungen der erwähnten Diplomarbeit widmete sich Nemecek im zweiten Pilotprojektjahr von GESTU in seiner Diplomarbeit der weiteren Evaluierung und Einführung von technischen Hilfsmitteln. Insgesamt fünf Lehrveranstaltungen wurden durch die **Respeaking** Technik in Echtzeit untertitelt. Diese Untertitelung durch die Firma Titelbild wurde allgemein als **positiv beurteilt** und ermöglichte den gehörlosen und schwerhörigen Studierenden, dem Inhalt der Vorlesungen zu folgen. Darüber hinaus intensivierte Nemecek u.a. auch die Videoaufzeichnungen von Lehrveranstaltungen, die in der Österreichischen Gebärdensprache gedolmetscht wurden. Die **Lehrveranstaltungsaufzeichnung** war das am häufigsten eingesetzte technische Hilfsmittel im zweiten GESTU Jahr. Die Aufzeichnungen wurden von den Studierenden als besonders positiv bewertet und stellten einen wichtigen **Beitrag zum Abbau bestehender Barrieren** dar. Neben einer Einblendung von Gebärdensprachübersetzungen sind Vorlesungstranskripte bzw. Untertitel die einzige Möglichkeit für hörbeeinträchtigte Studierende, Lehrveranstaltungsaufzeichnungen zum Erlernen und Vertiefen der Inhalte zu nutzen und somit davon zu profilierteren. Versuchsweise wurden in Echtzeit erzeugte Untertitel nachträglich zusammen mit der Aufzeichnung in die **E-Learning** Platform Synote eingebunden. In Synote aufbereitete Vorlesungen bieten eine gute Möglichkeit, den heterogenen Bedürfnissen hörbeeinträchtigter Studierender im E-Learning Bereich gerecht zu werden. Von solchen in Synote eingebundene Untertitel können zukünftig nicht nur hörbeeinträchtigte, sondern auch fremdsprachige Studierende (deren primäre Sprache nicht die Vortragssprache ist), profitieren. Generell entsteht durch den archivierten und durchsuchbaren Inhalt ein **Mehrwert für alle** Studierenden.
Das Ziel des Pilotprojekts GESTU lag in der Entwicklung eines zukunftsfähigen Modells für den gesamten österreichischen tertiären Bildungssektor und langfristig in der Erhöhung der Anzahl von hörbeeinträchtigten Akademikerinnen und Akademiker in Österreich. Basierend auf den Erfahrungen des Pilotprojekts wird GESTU in adaptierter Form fortgesetzt. In Zusammenarbeit mit Lehrenden und Studierenden konnte das GESTU Team einige der evaluierten und erprobten technischen Maßnahmen fortsetzen und nahtlos in den täglichen Universitätsbetrieb integrieren. Durch pädagogische, organisatorische und technische Unterstützung trägt GESTU maßgeblich der Inklusion von hörbeeinträchtigten Studierenden im tertiären Bildungsbereich im Raum Wien bei.

Auch abseits von GESTU werden zunehmend mehr Lehrveranstaltungen aufgezeichnet und in E-Learning Plattformen zur Verfügung gestellt. Aufgrund von fehlenden effektiven und kostengünstigen Alternativen zur Einbindung von live erzeugten Untertiteln, konnte im Laufe des zweijährigen Pilotprojekts die Untertitelquote im E-Learning Bereich nicht maßgeblich erhöht werden. Die Erstellung von Untertitel durch die Unterstützung von **Spracherkennungssystemen** (ASRs) stellt ein **großes Potenzial** dar. Um eine Steigerung der Untertitelquote im E-Learning Bereich mittelfristig erreichen zu können, wurden im Zuge dieser Diplomarbeit zwei vielversprechende Ansätze untersucht. Die dabei eingesetzen Spracherkennungssysteme unterscheiden sich stark in ihrer Funktionsweise und Anwendung. Die erste untersuchte Option ist die offline Untertitelerstellung mit Respeaking (Scripting) mit der Spracherkennung Dragon. Die **Erarbeitung**, **Dokumentation** und **Evaluierung eines Scripting Trainings** stellt den Schwerpunkt dieser Diplomarbeit dar. Das Ausbildungsziel liegt in der Erzeugung einer nahezu '1:1' Untertitelung von aufgezeichneten Lehrveranstaltungen in hoher Qualität. Als zweite vielversprechende Alternative wurde die Erstellung eines '1:1' Transkripts durch die Sprecher bzw. Sprecherinnen unabhängige, für spontane Sprache entwickelte automatische Spracherkennungssoftware des EML evaluiert.

Eine intensive Literaturrecherche ermöglichte mir die Einarbeitung in die noch sehr junge Geschichte des Respeakings und die Ausarbeitung des theoretischen Teils dieser Diplomarbeit. Der Theorieteil bietet einen detaillierten Einblick in die Respeaking Geschichte, die komplexe Tätigkeit des Respeakings und die zum Teil simultan stattfindenden Handlungen und Arbeitsabläufe. Ebenso sind die existierenden Ausbildungen im europäischen Raum erläutert. Durch die intensive Auseinandersetzung mit den verwandten Tätigkeiten des Simultandolmetschens sowie der Audiovisuellen Übersetzung konnte ich darüber hinaus Erkenntnisse und Erfahrungen aus diesen Professionen in das Training einfließen lassen. Dies betrifft u.a. die Themenbereiche des Umformulierens und des Kürzens, die zu den am meisten diskutierten Themen in der Untertitelungsliteratur gehören. Neben der Literaturrecherche ermöglichte mir ein reger Austausch mit führenden europäischen Forschern im Bereich des Respeakings und der Respeaking Ausbildung die Ausarbeitung des Scripting Trainings sowie die Auswahl des Equipments. Im Eigenversuch erlernte ich schrittweise das Respeaking und vermittelte die Kenntnisse in **sieben dokumentierten Einheiten** einem Auszubildenden. Neben praktischen Übungen beinhaltet das entworfene Training auch theoretisches Wissen über Hörbeeinträchtigung, dem Themengebiet der Untertitelung sowie der Funktionsweise von Spracherkennungssoftwares. Der Anteil der praktischen Übungen erhöht sich im Laufe der Ausbildung sukzessive, da diese zu Beginn rasch stimmlich und geistig ermüdend sein können. Weiters sind die praktischen Übungen so aufgebaut, dass sie ein schrittweises Herantasten an die komplexe Respeaking Tätigkeit ermöglichen. So wird beispielsweise in den Shadowing Übungen am Beginn der Ausbildung keine Spracherkennung verwendet. Damit ist es der auszubildenden Person möglich, sich einzig auf das gleichzeitige Sprechen/Diktieren zu konzentrieren und anschließend Schwächen in der Aussprache zu lokalisieren. In ähnlicher Weise dienen weitere Shadowing Übungen zum gezielten Erhöhen der Konzentrationsspanne. Gegen Ende der Ausbildung ermöglicht die Anpassung der Wiedergabegeschwindigkeit die stetige Steigerung der Arbeitsgeschwindigkeit. Das **schrittweise Erlernen der Respeaking Tätigkeiten** hob der Auszubildende im Feedback als sehr positiv hervor. Wei-

ters wird die in dieser Diplomarbeit hervorgehobene Relevanz der drei Respeaking Phasen - der Vorbereitung, der Untertitelerstellung sowie der Nachbereitung - im Training Rechnung getragen. Jede Phase ist ein fundamentaler Bestandteil der Ausbildung. Die Auseinandersetzung mit den konträren Standpunkten der verschiedene Interessengruppen bezüglich des Umformulierens und Kürzens ermöglicht während der Ausbildung das Entwickeln der nötigen Sensibilität. Praktische Übungen dienen dem Erlernen des Kürzens und Umformulierens um speziell den Anforderungen bei der Untertitelung von universitären Vorlesungen gerecht zu werden. Ein zentrales Ziel der Ausbildung ist die Erstellung von Untertiteln für eine Vorlesung in hoher Qualität. Die laufende Evaluierung der Qualität ist eine wesentliche Säule der Ausbildung. Sie ermöglichte dem Auszubildenden während des Lernprozesses Probleme zu identifizieren und sich fortlaufend zu verbessern.

Mit dem geplanten **Ausbildungsaufwand entsprechend von 3 ECTS** konnte der ausgebildete Respeaker innerhalb von ca. drei Monaten eine individuelle Scripting Technik entwickeln. In der Abschlussübung der Einheit 7 erreichte er einen NER-Wert von 98,6% und somit das Ausbildungsziel von $\geq 98,0\%$. Das gewählte Equipment ist demnach - vor allem in Bezug auf die Spracherkennung Dragon und das gewählte Mikrofon - für die Scripting Anwendung geeignet.
Durch die Evaluierung der Einheiten konnten Verbesserungsvorschläge erarbeitet werden. Diese können in zukünftigen Ausbildungen berücksichtigt werden. Beispielsweise könnte durch eine zusätzliche achte Einheit der Aufwand der Übungen gleichmäßiger verteilt werden. Obwohl die Ausbildung nur mit einer Person durchgeführt wurde, ist mit geringfügigen Adaptierungen auch das Training von mehreren Respeakerinnen und Respeakerin möglich. Weiters ist der Trainingsplan so konzipiert, dass er mit geringfügigen Änderungen und vor allem weiteren Übungen zur Echtzeitkorrektur auch die Ausbildung zur Live Untertitelung möglich macht. Obwohl Respeaking auch in Österreich beim ORF seit 2010 zur live Untertitelung eingesetzt wird, gibt es im Gegensatz zu anderen europäischen Ländern keine Respeaking Ausbildung an (öffentlichen) Bildungseinrichtungen. Somit legt das im Zuge dieser Diplomarbeit erarbeitete, dokumentierte und evaluierte **Training** den **Grundstein für** eine (akademische) Respeaking/Scripting **Ausbildung in Österreich**. So könnten zukünftig Studierende in universitären Modulen im Umfang von 3 ECTS als Scripter und Scripterinnen ausgebildet werden und Untertitel für E-Learning Platformen wie Synote erstellen.
Der zweite Schwerpunkt dieser Diplomarbeit stellt der qualitative Vergleich von unterschiedlichen Methoden der Untertitelerzeugung dar. Ich wählte eine zweistündige Vorlesung aus, für die unabhängig voneinander die Untertitel von vier verschiedenen Quellen erstellt wurden. Einerseits wurde die Vorlesung von der Firma Titelbild in Echtzeit untertitelt. Der Vortrag wurde zu diesem Zweck akustisch (ohne Video) vom Hörsaal in Wien nach Berlin übertragen. Weiters wurde die Videoaufzeichnung der Vorlesung nach erfolgter Ausbildung vom ausgebildeten Scripter sowie vom mir (offline) untertitelt. Damit konnten zusätzlich die in der Ausbildung erlernten Fähigkeiten des Respeakings analysiert werden. Als vierte Quelle diente das Transkript der Vorlesung, das durch die automatische Spracherkennung von EML erstellte wurde.
Die **Erkennungsraten** der **automatischen Spracherkennungssoftware** von EML lagen hinter den Erwartungen. Sie erreichten maximal 50% Wortakkuratheit und stellen daher **keine Unterstützungsmöglichkeiten** für hörbeeinträchtigte Studierende dar. Die qualitative Evaluierung der

drei weiteren Quellen wurde mit dem NER-Modell durchgeführt. Im Zuge dieser Diplomarbeit erfolgte somit erstmals in Österreich die qualitative Analyse von Untertiteln mit NER-Modell. Insgesamt wurden fünf durch Respeaking erstellte Transkripte ausgewertet und die Ergebnisse in vielen qualitativen Aspekten verglichen. Zum einen das von Titelbild live erstellte und nachträglich nicht korrigierte Transkript. Zum anderen die Untertitel vom ausgebildeten Scripter sowie die von mir erstellten Untertitel. Um die Auswirkung der nachträglichen Korrektur beim Scripting beurteilen zu können, wurden die offline erzeugten Untertitel jeweils *vor* und *nach* erfolgter Korrektur evaluiert.
Das von mir erstellte Transkript nach erfolgter Korrektur weist mit einem NER-Wert von 98,9% die höchste Qualität auf. In 17 der 18 Stichproben konnte der angestrebten Wert von mindestens 98% erreicht werden. Allerdings benötigte ich für die Erstellung in dieser Qualität mit dem 4,33-fachen (inkl. Pausen und Korrektur) der Vorlesungsdauer auch die meiste Zeit. In der Literatur wird das Verhältnis für die Untertitelerstellung mittels Tastatur auf die zehnfache, beim Scripting mit der siebenfachen Zeit der Originalquelle beziffert. Somit kann bezüglich des Zeitaufwandes von einer effektiven Untertitelerstellung in hoher Qualität gesprochen werden. Der im Zuge dieser Diplomarbeit ausgebildete Scripter konnte in der 3,2-fachen Vorlesungszeit einen NER-Wert von 97,4% erreichen. In sechs der 18 Stichproben wurden mindestens 98% erreicht. Die **vom Ausgebildeten angewandte** Scripting **Methode** stellt somit eine **sehr zeiteffiziente Untertitelerstellung** dar. Aufgrund der Schnittstelle zwischen Dragon und der Untertitelsoftware Subtitle Workshop kam es beim Scripting häufig zu Erkennungsfehlern am Satzbeginn (besonders Kleinanstatt Großschreibung). Wären vom Ausgebildeten diese Fehler korrigiert worden bzw. könnte das Schnittstellenproblem behoben werden, so würde dies eine Erhöhung des NER-Wertes auf 97,8% bedeuten. Der Auszubildende erkannte bei hohen Sprechgeschwindigkeiten weniger Fehler als bei niedrigen Geschwindigkeit. Ein weiteres, gezieltes Training auf das Erkennen und Markieren von Fehlern auch bei höheren Sprechgeschwindigkeiten könnte schließlich die angestrebte Qualität ermöglichen. So hätte die Korrektur von zusätzlich einem Viertel der normalen Erkennungsfehler einen NER-Wert von 98,0% ergeben.
Die von der Firma Titelbild live erzeugten Untertitel erreichten einen NER-Wert von 94,9%. Eine der 18 Stichproben erreichte den angestrebten NER-Wert von mindestens 98,0%. Der vergleichsweisen niedrigen Qualität der Untertitel steht der geringste Aufwand der 2,2 bis 3,0-fachen Vorlesungszeit gegenüber.
Die **Ergebnisse verdeutlichen** die **Vorteile** einer **offline Untertitelung** durch Respeaking gegenüber der Echtzeit Variante. Durch die zeitliche und räumliche Entkoppelung konnte somit eine signifikant höhere und konstanterer Qualität im Vergleich zu den live erstellten Untertitel erreicht werden. Diese Ergebnisse heben die besseren Ausgangsbedingung bei der offline Transkription hervor. Resultierend aus der Wahl der individuellen Wiedergabegeschwindigkeit und der Möglichkeit des Pausierens kann beim Scripting auch bei hohen Sprechgeschwindigkeiten und Rhythmuswechsel im Vortrag dessen Inhalt ohne (erheblichen) Informationsverlust transkribiert werden. Dies erklärt u.a. die Qualität *vor* erfolgter Korrektur bei den beiden Scriptern. Mit der 2,0-2,3-fachen Zeit konnte schon ohne nachträgliche Korrektur mit 96,8 bzw. 96,7% ein höherer NER-Wert als bei der Echtzeituntertitelung erreicht werden. Durch die nachträgliche Korrektur konnten darüber hinaus Fehlerhäufigkeiten vermieden und die Qualität generell deutlich erhöht werden. Kurzfristige Probleme wie technische Schwierigkeiten, Konzentrations-

schwierigkeiten, anspruchsvolle Passagen beim Vortrag, etc. führen bei der offline Untertitelung weniger stark zu Qualitätseinbußen als bei der Echtzeit Untertitelung. Somit kann in der Praxis von den Studierenden zusammen mit dem GESTU Team abgewogen werden, ob durch eine Echtzeit Untertitelung die Partizipation an einer Vorlesung als essentieller angesehen wird als die **höhere Qualität** durch die **offline Erstellung**.

Um den steigenden Bedarf an mittels Respeaking erstellten Untertiteln im Bildungs- aber auch im Rundfunkbereich gerecht zu werden, wäre eine **Etablierung** einer **Respeaking Ausbildung** in **Österreich** einer der wichtigsten **nächsten Schritte**. Dies könnte auch das Interesse an diesem spannenden Forschungsgebiet erhöhen und dazu beitragen, den Forschungsrückstand zu anderen europäischen Ländern zu verringern. Seit 2010 gibt es in Österreich vom ÖSB eine Ausbildung zum Schriftdolmetschen. Die Ausbildung zur Untertitelerstellung mittels Tastatur dauert zehn Monate und die Schriftdolmetscherinnen und Schriftdolmetscher arbeiten in der live und der offline Transkription. Mit den ca. drei Monaten Ausbildungszeit zum Scripting stellt die entworfene Ausbildung eine Alternative mit potentiell kürzerer Ausbildungsdauer dar. Eine Kooperation mit dem ÖSB für eine Kombination aus Respeaking- und Schriftdolmetschausbildung könnte zur Erhöhung der Untertitelquote in Österreich beitragen. Dabei könnten individuell, je nach Einsatzgebiet, die jeweiligen Vorteile der beiden Methoden in der Praxis genutzt werden.
Durch eine umfangreichere Evaluierung aktueller Untertiteleinblendungen könnte ähnlich wie in anderen europäischen Ländern der **Fokus** von Quantität auch **vermehrt auf** die **Qualität** gerichtet werden. Sehr hilfreich dabei könnte der im Frühjahr 2013 vorgestellte NERStar[1] sein, der die Auswertung von Untertitel mit dem NER-Modell ermöglicht. Diese Software ist speziell für eine effektive Anwendung des NER-Modells entworfen und für nicht kommerzielle Forschungszwecke frei verfügbar. Im Rahmen des „4th International Symposium on Live Subtitling: Live subtitling with respeaking and other respeaking applications“ konnte ich im März 2013 die Ergebnisse des erarbeiteten Respeaking Trainings in Barcelona ebenso präsentieren wie sbeim IKT-Forum für Menschen mit Behinderungen im Juli in Linz. Ebenfalls im Juli konnte ich als Teilnehmer einer Podiumsdiskussion zum Barrierefreier Rundfunk beim Bildungskongress des Österreichischer Gehörlosenbundes die Notwendigkeit der Evaluierung der Untertitelqualität zur fortlaufenden Verbesserung der Untertitelqualität hervorheben und die Möglichkeiten und Grenzen beim Umformulieren sowie Kürzen mit Fachleuten und hörbeeinträchtigten Menschen diskutieren. Im Zuge dieser Diplomarbeit wurde gezeigt, dass bei den derzeitigen Möglichkeiten der Live Untertitelung mittels Respeaking keine Perfektion möglich ist. Ich bin davon überzeugt, dass im gemeinsamen Diskurs zwischen Respeakern und Respeakerinnen, dem Zielpuplikum und den Auftraggebern und Auftraggeberinnen gemeinsam individuell ein Optimum gefunden werden kann.

[1] `www.speedchill.com/nerstar`, letzter Zugriff: 20.10.2013

KAPITEL 6

Anhang

6.1 Material zur Einheit 2: Interpunktions- und Sonderzeichen, Buchstabieren, Datum, etc.

Es folgt ein Text zum Üben der Interpunktions- und Sonderzeichen, dem Buchstabieren und Schreiben mit Großbuchstaben, das Diktieren von Datums- und Zeitangaben sowie von Nummern, Emailadressen, et cetera im Diktiermodus. Dabei sollen sämtliche Zeichen diktiert werden, um den Umgang mit Dragon zu üben:

Warum ist es wichtig, Interpunktionszeichen zu üben? Die Spracherkennung Dragon ist nicht in der Lage, aus der Betonung automatisch die Satzzeichen zu erkennen (wie „PUNKT“ oder „FRAGEZEICHEN“). Auch müssen ‚Hervorhebungen wie diese‘ explizit diktiert werden.
Es kann manchmal gewünscht sein, Eigennamen wie folgt zu schreiben: ÖSTERREICHISCHER RUNDFUNK, nur um ein Beispiel zu nennen und diese Funktionalität zu demonstrieren. Diese geschriebenen Wörter können auch als ORF abgekürzt werden. Dragon kennt viele Akronyme wie ORF, GmbH, USA, Dr., cm und viele mehr.
Auch das Buchstabieren und das Diktieren von Datumsangaben soll nun geübt werden...
„Bitte buchstabieren Sie Ihren Namen“ HATTINGER CHRISTIAN und nun Ihr Geburtsdatum in verschiedenen Schreibweisen: 2. August 1982 oder 02.08.1982. Buchstabieren Sie nun bitte das heutige Datum: 12. April 2012. Und nun die aktuelle Uhrzeit: 11:47 Uhr bzw. 11:47 oder, um genau zu sein 11:47:03 Uhr. Manchmal möchte man jedoch auch römische Zahlen buchstabieren. So kann 2012 auch als MMXII geschrieben werden.
Dragon kann auch mit großen und kleinen Zahlen umgehen. Die Zahl 57 ist ebenso leicht zu diktieren die 5,37 oder 5.768.200. Auch Währungsangaben wie € 5087 oder $700 stellen kein Problem dar. Dragon erkennt auch viele Sonderzeichen: %, ≪, ≫, #, &, et cetera.
Manchmal ist es nötig, Telefonnummern, Emailadressen und Webseiten zu diktieren. Eine Wiener Telefonnummer könnte wie folgt aussehen: 01 534 / 3215-12. Um diese Nummer außerhalb Österreichs zu wählen, müsste man diese wie folgt angegeben: +43 (0) 534 / 3215-12.
Auch E-Mail-Adressen können diktiert werden, wie info@google.com. Die Suchmaschine Google ist unter www.google.com zu finden. Es könnte jedoch auch sinnvoller sein, bei schwierigen Adressen oder Webseiten die Tastatur oder ein Makro zu verwenden.

6.1.1 Auszug des Transkripts der trainierten ASR von EML

Wie im Kapitel 4.3 erläutert, lagen die Erkennungsraten der Spracherkennungssoftware von EML hinter den Erwartungen. Im folgenden wird das Transkript der ersten fünf Minuten der Vorlesung Trainingswissenschaft vom 05.06.2012 (siehe Abschnitt 4.3.1), dem von manuell erstellten '1:1' Text des Vortrages gegenübergestellt. Die Gegenüberstellung stammt aus [Nem13, Abschnitt 9.1] und soll unterstreichen, dass ein Transkript in dieser Qualität keine Hilfestellung für hörbeeinträchtigte Studierende ist. An dieser Stelle sei an die Empfehlungen in [Nem13, Abschnitt 3.2] verwiesen.

Transkript der EML ASR	**1:1 Transkript (manuell erstellt)**
und dann gibt es das nächste Kapitel des Gebietes Wettkämpfe dass es beschäftigen wird sogar gewertet in der letzten Einheit in 14 Tagen und zum Schluss wird derzeit noch offen bleibt überlebt kommen Bemerkungen zur Thematik aus der Konzepte nur und besonders belegte des Kindertrainings	Und dann gibt es noch das nächste Kapitel. Das Kapitel Wettkämpfe. Das uns heute beschäftigen wird. Und möglicherweise dann in der letzten Einheit in 14 Tagen. Und zum Abschluss, je nachdem wie viel Zeit noch offen bleibt, übrig bleibt, kommen noch ein paar Bemerkungen zur Thematik Ausdauerkonzepte-neu und besondere Aspekte eines Klimatrainings.
für für aber hat die Übernahme daran glaubt verschiedener Sportarten Hinblick auf das eigene diesen letzten gegen werden oder auch wieder mit verspielen geht dann in entsprechender Weise stattfinden oder und ein Sprecher mitwirken hatte Bedingungen beginnt und natürlich verbittert sind so wird das wir das gesamte begrenzte Wettkampfvorbereitung den Wettkampf oder sich genauer anschauen	Das sind also Dinge, die man wahrscheinlich nicht immer hört. Die aber momentan gerade in verschiedenen Sportarten in Hinblick auf Großereignisse, wie sie in den letzten Jahren mit Peking waren oder auch immer wieder mit Winterspielen, die dann in entsprechender Höhenlage stattfinden oder unter entsprechenden mikroklimatischen Bedingungen, wie in Peking, natürlich von Bedeutung sind. Insoweit, dass man das gesamte Paket der Wettkampfvorbereitung, den Wettkampfort betreffend, sich genauer anschaut.

das Kürzel zu aber das ist immer mehr der werden über das Thema der Stelle und Trainingsplanung oder zur wenigstens durch Wiederherstellung der Trainingsplanung gewissermaßen ein Paket zu konservativ oder erschlagen wird diese Darstellung von Professor war

Ja und ganz kurz zur Erinnerung an das letzte Mal. Wir haben uns beim letzten Mal unterhalten über das Thema Trainingssteuerung und Trainingsplanung. Ich hab dazu erwähnt, dass eigentlich die Trainingssteuerung und die Trainingsplanung gewissermaßen ein Paket sind und wenn wir auf die Folie drauf schauen, auf diese Darstellung von Professor Baron.

das natürlich der Trainingsplanung der komplexen der ist man allerdings ging zur Feinheiten zu unterscheiden und hat das Ganze insoweit zu Kapitel aufgeteilt werden aber das technisch Trainingswissenschaft der SPD für den Trainer war besonderer Abschnitte zu beurteilen ist zu und die muss man natürlich immer in der lettischen Konzept bei und Sprecher der wird dann kann man wird es bleibt ist das ist dann wieder zusammen fließen lassen

Dann sehen wir natürlich dass die Trainingsplanung ein Teil der komplexen Trainingssteuerung ist - „nonanet“. Allerdings gibt es hier einige Feinheiten zu unterscheiden und ich habe das Ganze insoweit in zwei Kapitel aufgeteilt, weil rein arbeitstechnisch für den Trainingswissenschaftler respektive für den Trainer besondere Abschnitte zu beurteilen sind und die muss man natürlich einmal im theoretischen Konzept verinnerlichen und entsprechend aufarbeiten und dann kann man in der Trainingspraxis das System dann wieder zusammenfließen lassen.

oder sogar der Planung wenn sich der Staat von Beginnern normale jetzt mit einer Bestimmung des Ist-Zustand wird ist der Ist-Zustand für die Planung und findet nicht inhaltlich 10.

Gut sowohl die Planung wie die Trainingssteuerung beginnen einmal - „nonanet“ - mit einer Bestimmung des Istzustandes, wobei klarerweise der Istzustand für die Planung und für die Trainingssteuerung inhaltlich ident sind.

da haben 2 Kernbereiche die uns zu Beginn seiner erarbeiten müssen diese sondern die Fragestellungen der Welt ist ein Athlet gesund

Da haben wir zwei Kernbereiche, die wir uns zu Beginn erarbeiten müssen. Das ist einmal die Fragestellung, inwieweit ist mein Athlet gesund.

ist die Stadt für die Basisuntersuchungen wird der Planung zumindest bis Sportarten die so ein Sprecher konditionsintensiven Sehenswertes ganze 11. bis werde es dreimal

Das ist die Sportärztliche-Basisuntersuchtung, die im Rhythmus der Planung zumindest jährlich passiert. In Sportarten, die also entsprechend konditionsintensiv sind, wird das ganze öfter passieren - 2-3x.

dass das oder wichtige Basisuntersuchungen die vor das Geld das ist betreffen und verschiedener Parameter geht den Sonderstatus Planung bestreiten oder bestätigen und das ist jetzt hier und da Belastbarkeits diagnostik angeführt der Bereiche

Das sind also immer wieder wichtige Basisuntersuchtungen, die vor allem das Herz-Kreislaufsystem betreffen und verschiedene Parameter die den Gesundheitsstatus dann unterschreiten oder eben bestätigen und was immer wieder vergessen wird, hier unter Belastbarkeitsdiagnostik angeführt. Das ist der Bereich hier.

besichtigen Fragestellungen da ist man letztendlich betreibe aber wirklich gesund und geht es um den orthopädischen Gesundheitszustand und also wieder auch genannt die Auswertung von Großereignissen ist dann kann man hört damals wirklich versteht dass viele Athleten ist so vor einer Personal für die Koalition zusammen und dann von der ist der Qualifikation aufgrund von beinahe Verletzungen und dann spricht mich Ausfall	Das ist die Grundfragestellung: inwieweit ist mein Athlet, den ich betreue auch wirklich gesund. Und da geht es um den orthopädischen Gesundheitszustand und wenn man also immer wieder im Nachhinein die Auswertungen von Großereignissen liest, dann kann man dort sehr häufig feststellen, dass viele Athleten, die zu Großereignissen fahren, ich habe es das letzte Mal angesprochen, eigentlich gar nicht gesund sind und dann eigentlich im Vorfeld bereits bei der Qualifikation aufgrund von Banalverletzungen unter Anführungsstrichen scheinbar ausfallen.
umgekehrt kann es passieren dass der Sportler Krippenplätze Überzeugungen haben und dann ist man Deutschland müssen wir gesund nicht in der Lage sind derart große Turnier dann zu bestehen	Umgekehrt kann es passieren, dass Sportler Quotenplätze bereits errungen haben und dann im letzten Moment ausscheiden müssen, weil sie eben gesundheitlich nicht in der Lage sind, ein derart großes Turnier dann zu bestehen.
von der Situation geringer ist man immer ganz sicher bis jetzt die Spiele der Gegend werden wo früher aus ich werde sollten und das ist doch der die Spiele ist wurde das stecken	Und wer sich da zurück erinnert, ich bin mir nicht mehr ganz sicher ob das jetzt die Spiele in Peking waren, wo im Vorfeld aus einem Beachvolleyball Team durch den Olympiaarzt, durch den Doktor Engel, eine Spielerin sozusagen eliminiert wurde, unter Anführungsstrichen.
ist also ganz böse Schlagzeilen Zeitungen über unseren Spielerin offensichtlich an haben wir uns dann Problematik gelitten hatte offensichtlich wird	Das hat also ganz böse Schlagzeilen in den Zeitungen gebracht. Wo also eine Spielerin offensichtlich an einem Wirbelsäulenproblematik gelitten hat, offensichtlich noch immer leidet.
und die Situation und das schon gehört wurde ist das Bestehen eines wird das 1. passen würde wenn stellen Formate Gesundheitszustand nicht möglich ist und er ist der 1. Van wird nicht dass wir für wirklich stellen und darf wurde die gesperrt sozusagen spricht ausgeschlossen	Und die Situation vom Olympiaarzt eben so beurteilt wurde, dass das Bestehen eines Olympiaturniers inklusive aller Strapazen die dort entstehen. Vom aktuellen Gesundheitszustand her nicht möglich ist und er als Olympiaarzt die Verantwortung nicht übernimmt, dass hier Folgeschäden entstehen und daraufhin wurde die Sportler sozusagen Anführungsstrichen ausgeschlossen.
wird das natürlich die Betroffene spart absolut Peter Situation und umgekehrt belegt so dass das offensichtlich gehört und dann stecken kann Sportarten dieses Zusammenwirken Mannschaft oder als Trainer Trainings Umfeld das Trainer-Berater-System mit einer ausreichend qualitativ hochwertige und läuft	Gut das ist natürlich für die betroffene Sportlerin eine absolut bittere Situation. Umgekehrt belegt also das genau das, dass offensichtlich gerade in unter Anführungsstrichen Randsportarten dieses Zusammenwirken Mannschaftsarzt, Trainer, Trainingsumfeld, also Trainer-Beratersystem nicht immer ausreichen qualitativ hochwertig abläuft.

Literaturverzeichnis

[AOD07] ALINE, Remael ; ORERO, Pilar ; DIAZ CINTAS, Jorge: *Audiovisual Translation: Subtitling (Translation Practices Explained)*. Amsterdam, New York : Saint Jerome Publications, 2007. – 272 S. – ISBN 978–1900650953

[ARRF08] ARUMÍ-RIBAS, Marta ; ROMERO-FRESCO, Pablo: A Practical Proposal for the Training of Respeakers 1. In: *JoSTrans: The Journal of Specialised Translation* 10 (2008)

[Bak98] BAKER, Mona: *Routledge Encyclopedia of Translation Studies*. Routledge, 1998. – ISBN 978–0415093804

[BB95] BRAUN, Julius ; BURGHOFER, Birgitt: *Gehörlose Menschen in Österreich : ihre Lebens- und Arbeitssituation*. 1. Auflage. Linz : Instut für Sozial- und Wirtschaftswissenschaften, 1995. – ISBN 3–901320–03–2

[Bun12] BUNDESMINISTERIUM FÜR WISSENSCHAFT UND FORSCHUNG: *Bestmögliche Rahmenbedingungen für gehörlose Studierende - GESTU wird fortgesetzt*. www.ots.at/presseaussendung/OTS_20120712_OTS0112, letzter Zugriff: 29.12.2012, Juli 2012

[Dia09] DIAZ CINTAS, Jorge: *New Trends in Audiovisual Translation (Topics in Translation)*. Multilingual Matters, 2009. – ISBN 978–1847691545

[Eul06] EULER, Stephan: *Grundkurs Spracherkennung: Vom Sprachsignal zum Dialog - Grundlagen und Anwendungen verstehen - Mit praktischen Übungen*. Viewg+Teubner Verlag, 2006. – 200 S. – ISBN 978–3834800039

[Hat11] HATTINGER, Christian F.: *GESTU: Evaluierung von technischen Hilfsmitteln zur Förderung Studierender mit Hörbehinderung im österreichischen tertiären Bildungssektor und Einführung geeigneter Technologien an der TU Wien*, Technischen Universität Wien, Diplomarbeit, 2011. – 188 S.

[Hei06] HEIMGARTNER, Cornelia: *Die Übersetzung als Tor zur Informationsgesellschaft: Tonsubstitution für Hörgeschädigte am Fernsehen*, Université de Genève, Diplomarbeit, 2006. `http://www.deafzone.ch/file/file_pool/action/download/file_id/227/`. – 74 S.

[Hru10] HRUSKA, Andreas: *E-Learning Impuls TUWEL News & LectureTube, Präsentation vom 09.09.2010*. Wien, 2010

[KDH10] KOSEC, Primož ; DEBEVC, Matjaž ; HOLZINGER, Andreas: Sign Language Interpreter Module: Accessible Video Retrieval with Subtitles. In: *Computers Helping People with Special Needs* 6180/210 (2010), S. 221–228

[Kel07] KELEN, Balint: *Spracherkennung: Grundlagen und dolmetschrelevante Anwendung beim Respeaking und Simultandolmetschen*, Universität Wien, Diplomarbeit, 2007. – 79 S.

[Kno12] KNOWBRAINER: *KnowBrainer: The Speech Recognition Experts*. www.knowbrainer.com, letzter Zugriff: 24.10.2012, 2012

[Kno13] KNOWBRAINER: *Samson Airline 77 microphone*. www.knowbrainer.com, letzter Zugriff: 05.03.2013. `http://www.knowbrainer.com/NewStore/pc/viewPrd.asp?idproduct=76`. Version: 2013

[KS07] KRAUSNEKER, Verena ; SCHALBER, Katharina: Sprache Macht Wissen - Zur Situation gehörloser und hörbehinderter SchülerInnen, Studierender & ihrer LehrerInnen, sowie zur Österreichischen Gebärdensprache in Schule und Universität Wien. Abschlussbericht des Forschungsprojekts 2006/2007 / Innovationszentrum der Universität Wien; Verein Österreichisches Sprachen-Kompetenz-Zentrum. 2007 (November). – Forschungsbericht. – 517 S.

[Kur96] KURZ, Ingrid: *Simultandolmetschen als Gegenstand der interdisziplinären Forschung*. WUV - Universitätsverlag, 1996. – ISBN 978–3–85114–262–4

[Kur12] KURZ, Ingrid: *Emailauskunft Sabine Kurz. 31.03.2012*. 2012

[Lam06] LAMBOURNE, Andrew: *Subtitle respeaking: A new skill for a new age*. www.intralinea.it/specials/respeaking/eng_more.php?id=447_0_41_0_C, letzter Zugriff: 20.10.2012, 2006

[Lew10] LEWIS, Joyce: *Synote developer wins third major award of the year*. www.ecs.soton.ac.uk/about/news/3389, letzter Zugriff: 14.08.2011, 2010

[Lew12] LEWIS, Joyce: *Synote wins ICT Initiative of the Year in prestigious THE Awards*. http://www.ecs.soton.ac.uk/about/news/3874, letzter Zugriff: 28.03.2012, 2012

[Lis08] LISCHKA, Katharina: *Medien gehören nicht nur gehört! Rechtlicher Status der europäischen Gebärdensprachen sowie deren mediale Repräsentation in der EU.*, Universität Wien, Dissertation, 2008

[LW05] LAPP, Christine ; WITTMANN, Peter: *Bericht des Verfassungsausschusses: 1029 der Beilagen zu den Stenographischen Protokollen des Nationalrates XXII. GP*. www.parlament.gv.at/PG/DE/XXII/I/I_01029/fnameorig_045253.html, letzter Zugriff: 20.10.2012, 2005

[LWK+09] LI, Yunjia ; WALD, Mike ; KHOJA, Shakeel ; WILLS, Gary ; MILLARD, David ; KAJABA, Jiri ; SINGH, Priyanka ; GILBERT, L.: Enhancing Multimedia E-Learning with Synchronised Annotation. In: *Proceedings of the first ACM international workshop on Multimedia technologies for distance learning*, ACM, 2009, S. 9–18

[Mar06] MARSH, Alison: *Respeaking for the BBC*. www.intralinea.it/specials/respeaking/eng_more.php?id=484_0_41_0_M, letzter Zugriff: 13.07.2010, 2006

[Mar12] MARTINEZ, Juan: *Emailauskunft Speedchill, 14.03.2012*. 2012

[Mon11] MONSORNO, Stefania S.: *Il respeaking in televisione: strategia di condensazione e valutazione del tasso di accuratezza dei sottotitoli*, LUSPIO Libera Università, Diss., 2011. – 278 S.

[MoP11] MOPIX/MOTION PICTURE ACCESS: *What is the Rear Window Captioning System? How does it work?* http://ncam.wgbh.org/mopix/faq.html, letzter Zugriff: 12.08.2011, 2011

[Nem13] NEMECEK, Werner: *GESTU: Evaluierung von technischen Hilfsmittel zur Förderung Studierender mit Hörbehinderung im österreichischen tertiären Bildungssektor und Einführung geeigneter Technologien an der TU Wien*, Technischen Universität Wien, Diplomarbeit, 2013

[Net10] NET4VOICE: Net4Voice: Document on Methodology Specification. In: *www.net4voice.eu* WP 2 (2010), Nr. 1: Definition of a new learning methodology, S. 78

[NH12] NOWAK, Selina ; HATTINGER, Christian: Neues Berufsprofil: Live-Untertitelung durch Respeaking. In: *Universitas* Mitteilung (2012), S. 20–21. – ISSN 1996–3505

[Now10] NOWAK, Selina: *Live-Untertitelung: Die Simultandolmetschung am Bildschirmrand*, Universität Wien, Diplomarbeit, 2010

[Nua10] NUANCE: *Dragon NaturallySpeaking und Dragon Medical Version 11 Benutzerhandbuch*. 2010

[Nua12a] NUANCE: *Systemvoraussetzungen für Dragon NaturallySpeaking 11*. http://nuancede.custhelp.com/app/answers/detail/a_id/6474/˜/systemvoraussetzungen-f%E3%BCr-dragon-naturallyspeaking-11, letzter Zugriff: 10.02.2013, 2012

[Nua12b] NUANCE: *Systemvoraussetzungen für Dragon NaturallySpeaking 12*. http://nuancede.custhelp.com/app/answers/detail/a_id/6936/˜/systemvoraussetzungen-f%E3%BCr-dragon-naturallyspeaking-12, letzter Zugriff: 10.02.2013, 2012

[Nua13] NUANCE: *Spracherkennungssoftware Dragon NaturallySpeaking Premium für den PC*. www.nuance.de, letzter Zugriff: 05.03.2013. http://www.nuance.de/for-individuals/by-product/dragon-for-pc/premium-version/index.htm. Version: 2013

[Od11] OESB-DACHVERBAND.AT: *ÖSB-Schriftdolmetschausbildung*. www.oesb-dachverband.at/neues/oesb-statements, letzter Zugriff: 18.04.2011, 2011

[Orc10] ORCUTT, Lunis: KnowBrainer Review of NaturallySpeaking Ver. 1. In: *Audio* (2010)

[Ore06] ORERO, Pilar: Real-time subtitling in Spain: An overview. In: *inTRAlinea, online Translation journal* (2006). http://www.intralinea.it/specials/respeaking/ita_more.php?id=450_0_41_0_M

[ORF12a] ORF: *Etappenplan zum Ausbau des barrierefreien Zugangs zu den ORF-Fernseh-Programm und zum ORF-Online-Angebot gemäßParagraph 3 Abs. 1 Z 2 ORF-Gesetz*. 2012

[ORF12b] ORF, Zeit im B.: *Zeit im Bild vom 25.04.2012: Beitrag 'Hörstörungen werden oft unterschätzt'*. 2012

[O'S08] O'SHAUGHNESSY, Douglas: Invited paper: Automatic speech recognition: History, methods and challenges. In: *Pattern Recognition* 41 (2008), Nr. 10, S. 2965–2979. http://dx.doi.org/10.1016/j.patcog.2008.05.008. – DOI 10.1016/j.patcog.2008.05.008. – ISSN 00313203

[PK08] PFISTER, Beat ; KAUFMANN, Tobias: *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. 1. Auflage. Springer Berlin Heidelberg, 2008. – 483 S. – ISBN 978–3540759096

[Pou03] POUSEK, Wolfgang: *Technische Universität Wien : ECTS-Punkte*. www.tuwien.ac.at/aktuelles/news_detail/article/3602/, letzer Zugriff: 12.02.2013, 2003

[RF11] ROMERO-FRESCO, Pablo: *Subtitling Through Speech Recognition: Respeaking (Translation Practices Explained)*. Manchaster : St Jerome Publishing, 2011. – ISBN 978–1905763283

[RF12a] ROMERO-FRESCO, Pablo: *Emailauskunft Pablo Romero-Fresco, 22.02. 07.05.2012, 08.05.2012*. 2012

[RF12b] ROMERO-FRESCO, Pablo: Respeaking in Translator Training Curricula: Present and Future Prospects. In: *The Interpreter and Translator Trainer* 6 (2012), Nr. 1

[RFM14] ROMERO-FRESCO, Pablo ; MARTÍNEZ, Juan: Accuracy Rate in Live Subtitling - the NER Model. In: *Media for All 4 - Taking Stock (vorzeitiger Titel)* (2014)

[RV06] REMAEL, Aline ; VEER, Bart Van D.: *Real-Time Subtitling in Flanders: Needs and Teaching*. `http://www.intralinea.it/specials/respeaking/ita_more.php?id=492_0_41_0_M`. Version: 2006

[Syn12] SYNOTE: *Synote: Terms and Conditions*. http://synote.org/synote/user/termsAndConditions, letzter Zugriff: 29.03.2012, 2012

[TGN+10] TIBALDI, Daniela ; GARLASCHELLI, Luca ; NARDONE, Mariarosaria ; SANTINI, Marco ; RIGONI, Matteo ; WALD, Mike ; HOWARD, Chris ; WEGGERLE, Alexander ; SCHULTHESS, Peter: Net4Voice Final Report. In: *Education, Audiovisual & Culture Executive Agency* (2010)

[Uni10] UNIVERSITÄT WIEN: CENTER FOR TEACHING AND LEARNING: *Audio- und/oder Videoaufnahme von Lehrveranstaltungen*. http://ctl.univie.ac.at/fileadmin/user_upload/elearning/100915_Vorlesungsstreaming.pdf, letzter Zugriff: 26.10.2010, 2010

[UT03] UNIVERSITÄT WIEN ; TECHNISCHEN UNIVERSITÄT WIEN: *Studienplan Bakkalaureats- und Magisterstudien Informatikmanagement*. `www.informatik.tuwien.ac.at/lehre/studien/master/informatikmanagement`. Version: 2003

[Wal10a] WALD, Mike: Synote: Designed for all Advanced Learning Technology for Disabled and Non-Disabled People. In: *Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies* 10 (2010), Nr. Sousse Tunisia, S. 716–717

[Wal10b] WALD, Mike: *Synote Guide*. www.synote.org/synote/recording/replay/1, letzter Zugriff: 12.08.2011, 2010

[Wal11] WALD, Mike: Crowdsourcing Correction of Speech Recognition Captioning Errors. In: *W4A 2011: 8th International Cross-Disciplinary Conference on Web Accessibility* W4A 11 (2011), Nr. 2. ISBN 9781450304764

[Wal12] WALTER, Cornelia: *Respeaking - Intralinguales Simultandolmetschen für die Untertitelung*, Universität Wien, Masterarbeit, 2012. – 151 S.

[Weg10] WEGGERLE, Alexander: *Emailauskunft Net4Voice Projektpartner (Universität Ulm), 25.03.2010 und 13.04.2010*. Ulm, 2010

[Woj12] WOJCZEWSKI, Thomas: *Emailauskunft Titelbild, 11.09.2012, 12.09.2012, 13.12.2012*. 2012

[WWM+09] WALD, Mike ; WILLS, Gary ; MILLARD, Dave ; GILBERT, Lester ; KHOJA, Shakeel ; KAJABA, Jiri ; LI, Yunjia ; SINGH, Priyanka: Enhancing Learning Using Synchronised Multimedia Annotation. In: *EUNIS 2009: IT: Key of the European Space for Knowledge* (2009)

Ein gedrucktes Belegexemplar der vorliegenden Diplomarbeit liegt beim 'Zentrum für angewandte assistierende Technologien (AAT)' am 'Institut für Gestaltungs- und Wirkungsforschung' auf: Favoritenstraße 11/187-2b, A-1040 Wien. Dort ist auch eine CD beigelegt, auf welcher die Daten der NER-Analyse gespeichert sind.

Glossar

Glossary

ASR **A**utomatic **S**peech **R**ecognition, respektive automatische Spracherkennung; siehe Seite(n): 4, 6, 10–12, 16–18, 23, 29, 31, 33, 34, 38, 41–43, 65, 66, 69–73, 75, 77, 89, 93, 133, 152

ATT Zentrum für **A**ngewandte **A**ssistierende **T**echnologien an der Fakultät für Informatik der Technischen Universität Wien, Institut für Gestaltungs- und Wirkungsforschung; siehe Seite(n): 56

AVI **A**udio **V**ideo **I**nterleave. AVI ist ein Multimedia Format; siehe Seite(n): 83

AVT **A**udio**v**isual **T**ranslation, respektive audiovisuelle Übersetzung; siehe Seite(n): 45, 59, 63

BBC **B**ritish **B**roadcasting **C**orporation. BBC ist eine britische Rundfunkanstalt die große Teile ihres TV-Programms untertitelt; siehe Seite(n): 9, 16, 17, 35, 38, 63, 151

BMWF **B**undes**m**inisterium für **W**issenschaft und **F**orschung; siehe Seite(n): 4

DNS **D**ragon **N**aturally**S**peaking. DNS (kurz *Dragon* genannt) ist eine Spracherkennungssoftware der Firma Nuance Communications, Inc.; siehe Seite(n): 7, 34–39, 55, 56, 62, 65–67

DVD **D**igital **V**ersatile **D**isc. DVD ist ein digitales Speichermedium; siehe Seite(n): 38

ECTS **E**uropean **C**redit **T**ransfer **S**ystem. ECTS Punkte sind eine europäische Maßeinheit für den Arbeitsaufwand von Studierenden, wobei ein Studienjahr 60 ECTS Punkte vorsieht. 1 ECTS-Punkt entspricht an der TU Wien 25 Arbeitsstunden (vgl. [Pou03]); siehe Seite(n): 8, 48, 60, 67, 100, 134

EML **E**uropean **M**edia **L**aboratory Heidelberg. Die Firma EML entwickelt Software, u.a. Spracherkennungssysteme; siehe Seite(n): 5, 7, 10–12, 31, 33, 101, 102, 129, 133, 134, 139

FAB FAB Teletext & Subtitling Systems. Eine Firma, welche div. Softwarelösungen zum Erzeugen und Senden von Untertiteln (z.B. via Teletext bei Fernsehsendern) vertreibt; siehe Seite(n): 39, 58

GESTU Zwischen Juli 2010 bis Juli 2012: **G**ehörlos **E**rfolgreich **S**tudieren an der **TU** Wien, Barrierefreiheit für gehörlose Studierende (vgl. [Hat11, S: 3-5]), seit Juli 2012: **G**ehörlos **e**rfolgreich **stu**dieren an Universitäten in Wien (vgl. [Bun12]); siehe Seite(n): i, ii, 3–7, 10, 11, 18, 21, 30, 56, 57, 59, 63, 101, 102, 132

HG-UT **H**ör**g**eschädigten-**Un**t**e**rtitel; siehe Seite(n): 30, 32, 46, 49, 50, 63, 151

HMM **H**idden **M**arkov **M**odel. HMM ist ein stochastisches Modell, das häufig in der Spracherkennung eingesetzt wird; siehe Seite(n): 67

HTML **H**yper**t**ext **M**arkup **L**anguage. HTML ist eine Auszeichnungssprache zur Erstellung von Dokumenten, die meist im World Wide Web mittels eines Webbrowsers angezeigt werden; siehe Seite(n): 52

IBM **I**nternational **B**usiness **M**achines. Ein großes IT Unternehmen, da u.a. auch Spracherkennungssoftware entwickelt; siehe Seite(n): 34

ICT **I**nformation and **C**ommunication **T**echnology, respektive Informations- und Kommunikationstechnologie; siehe Seite(n): 34

IMS **I**ndependent **M**edia **S**upport Group Limited. IMS ist eine Firma, welche (Live) Untertitelung für Medienkonzerne durchführt; siehe Seite(n): 35, 38, 47, 67

IRQ **Inter**quartile range, respektive der Interquartilsabstand. Bei Box-Whisker-Plots ist der Interquartilsabstand ein statistisches Maß der Streuung von Daten. Die obere und untere Grenze bei IRQ wird als Q1 und Q3 bezeichnet; siehe Seite(n): 113, 115

IZ Als **I**nterpunktions**z**eichen werden in dieser Diplomarbeit sämtliche Zeichen (wie Bindestrich, etc.) inclusive Satzzeichen (wie Punkt, Beistrich, Komma, Doppelpunkt, Anführungszeichen, etc.) bezeichnet. 17, 27, 28, 62, 75, 76, 87, 88, 91, 92, 94, 95, 97–99, 106

L1 first language, respektive die Erstsprache; siehe Seite(n): 63

LED **L**ight-**E**mitting **D**iode, respektive lichtemittierende Diode oder Leuchtdiode genannt; siehe Seite(n): 68

LVA **L**ehr**ver**a**n**staltung; siehe Seite(n): 59, 76, 97

NER **N**umber of words in the respoken text (respektive die Gesamtzahl der durch Respeaking erzeugten Wörter), **E**dition errors (respektive Editierfehler), **R**ecognition errors (respektive Erkennungsfehler), **D**educted marks (respektive erkannte und korrigierte Erkennungsfehler); NER ist ein Modell zur Feststellung der Wortakkuratheit beim Respeaking; siehe Seite(n): i, ii, 11, 12, 24, 25, 41–45, 60, 76–78, 80, 81, 83, 84, 87, 88, 93, 95–99, 102, 104, 106, 110–113, 115–118, 122, 123, 128–130, 135, 136, 148

NERD **N**umber of words in the respoken text (respektive die Gesamtzahl der durch Respeaking erzeugten Wörter), **E**dition errors (respektive Editierfehler), **R**ecognition errors (respektive Erkennungsfehler), **D**educted marks (respektive erkannte und korrigierte Erkennungsfehler); NERD ist ein Modell zur Feststellung der Wortakkuratheit beim Respeaking; siehe Seite(n): 41, 42

Net4Voice Net4Voice war ein Teilprojekt aus dem europäischen „Lifelong Learning“ Programm; siehe Seite(n): 10

ÖGS **Ö**sterreichische **G**ebärden**s**prache; siehe Seite(n): 2–4, 6, 63

ÖSB **Ö**sterreichischer **S**chwerhörigen**b**und DACHVERBAND; siehe Seite(n): 8, 136

OmU **O**riginal **m**it **U**ntertitel; siehe Seite(n): 30, 32, 63

ORF **Ö**sterreichische **R**und**f**unk; siehe Seite(n): i, ii, 18, 22, 25, 28, 34, 37, 38, 40, 47, 49, 55, 58, 63, 65, 70, 91, 93, 96, 134, 152

PowerPoint Microsoft PowerPoint ist ein Präsentationsprogramm; siehe Seite(n): 30, 52

Q1 Bei Box-Whisker-Plots wird das untere Quartil als Q1 bezeichnet und definiert die untere Grenze des Wertebereichs, in dem 50% der Daten liegen; siehe Seite(n): 112, 113, 150

Q2 Bei Box-Whisker-Plots wird der Median als Q2 bezeichnet; siehe Seite(n): 113, 115, 116

Q3 Bei Box-Whisker-Plots wird das obere Quartil als Q3 bezeichnet und definiert die obere Grenze des Wertebereichs, in dem 50% der Daten liegen; siehe Seite(n): 113, 150

R1 Im Falle von Teamarbeit beim Respeaking wird das erste Teammitglied als Respeaker 1 (R1) bzw. Respeakerin 1 (R1) bezeichnet; Im Kapitel 4 wird allerdings *R1* für Christian Hattinger und den Auszubildenden*R2* für (beide Respeaker/Scripter, die nicht im Team arbeiteten) verwendet; siehe Seite(n): 21

R2 Im Falle von Teamarbeit beim Respeaking wird das erste Teammitglied als Respeaker 2 (R2) bzw. Respeakerin 2 (R2) bezeichnet; Im Kapitel 4 wird allerdings *R1* für Christian Hattinger und *R2* für den Auszubildenden (beide Respeaker/Scripter, die nicht im Team arbeiteten) verwendet; siehe Seite(n): 21, 22

RBM **R**ed **B**ee **M**edia. RBM ist eine internationale Firma, welche u.a. zu großen Teilen die (Live) Untertiteln für BBC durchführt; siehe Seite(n): 9, 17, 19, 35, 38, 48, 67, 152

RGD **R**emote **G**ebärdensprach**d**olmetschen; siehe Seite(n): 4

SDH **S**ubtitling for the **d**eaf and **h**ard-of-hearing, siehe HG-UT; siehe Seite(n): 32, 33

SRF **S**chweizer **R**adio- und **F**ernsehen; siehe Seite(n): 34

Titelbild TITELBILD Subtitling and Translation GmbH ist Teil der Red Bee Media Group (RBM) und führt u.a. (Live) Untertitelung mittels Respeaking durch; siehe Seite(n): 5–7, 12, 18, 21, 25, 30, 34, 37, 55, 86

TUWEL die E-Learning- und Kommunikationsplattform der Technischen Universität Wien; siehe Seite(n): 6, 52

UK **U**nited **K**ingdom, respektive das Vereinigte Königreich; siehe Seite(n): 16

USB **U**niversal **S**erial **B**us. USB ist ein serielles Bussystem, mit welchem Geräte mit einem Computer verbunden werden können; siehe Seite(n): 36

USD **US**-**D**ollar; siehe Seite(n): 56

VRT **V**laamse **R**adio- en **T**elevisieomroep. VRT ist eine flämische Rundfunkanstalt; siehe Seite(n): 16

WER **W**ord **E**rror **R**ate, respektive Wortfehlerrate (ein Qualitätsmerkmal für ASR); siehe Seite(n): 11, 41, 43, 73, 75, 76, 93, 95

WpM **W**ords **p**er **M**inute, respektive Wörter pro Minute; siehe Seite(n): 17, 27, 28, 63, 71, 74, 75, 87, 88, 91, 92, 94–99, 106, 118, 122, 123

WRR **W**ord **R**ecognition **R**ate, respektive Wortakkuratheit (ein Qualitätsmerkmal für ASR); siehe Seite(n): 11, 41–43, 45, 73–77, 93–95, 102

ZiB **Z**eit **i**m **B**ild, Hauptnachrichten des ORF; siehe Seite(n): 28, 65, 70, 72, 73, 76, 92, 95