

# UNTERSTÜTZENDE SPRACHÜBERSETZUNG IN TELEFONKONFERENZEN

*Martin Eckert, Stefan Feldes, Karlheinz Schuhmacher, Ralf Kirchherr, Joachim Stegmann,  
Florian Metze\**

*Voice Applications and Service Delivery Platforms, T-Systems Enterprise Services,  
Darmstadt und Berlin  
Deutsche Telekom Laboratories, Berlin \*  
martin.eckert@t-systems.com*

**Abstract:** In diesem Beitrag wird die Entwicklung eines Demonstrators für die akustische Online-Übersetzung von Nutzeräußerungen in Telefonkonferenzen von Deutsch nach Englisch beschrieben. Den Kern des Systems bildet ein domänenunabhängiges Übersetzungssystem, das - ebenso wie das Sprachmodell der Spracherkennungskomponente - auf die hier gewählte Domäne des Projektmanagements adaptiert wurde. Das Anwendungsszenario sieht den Einsatz von Sprachübersetzung zur Unterstützung von Teilnehmern vor, welche die Konferenzsprache nicht als Muttersprache beherrschen und daher gelegentlich zur Unterstützung auf ein Übersetzungssystem zurückgreifen. Der Beitrag beschreibt die Portierung des Systems auf eine aktuelle Wirkplattform, evaluiert und vergleicht verschiedene Ansätze, zeigt die dabei auftretenden Einschränkungen auf und beschreibt die für eine Anwendung in naher Zukunft kritischen Faktoren und Lösungsansätze.

## 1 Einleitung

Eine Telefonkonferenz (TK) mit mehreren Teilnehmern unterschiedlicher Muttersprache stellt eine typische Situation dar, in der automatische Sprachübersetzung nützlich sein kann. Derartige Telefonkonferenzen finden in internationalen Unternehmen häufig regelmäßig statt, um bspw. Projektbesprechungen mit Mitarbeitern an verteilten Standorten in oft verschiedenen Ländern durchzuführen. In einer Machbarkeitsstudie wurde bei T-Systems im Auftrag der Deutsche Telekom Laboratories ein Szenario untersucht, das die bedarfsgerechte Unterstützung der Teilnehmer einer derartigen Telefonkonferenz zum Ziel hat.

In vielen Fällen ist dabei ein umfassendes Simultan-Dolmetschen nicht erforderlich, da sich die Teilnehmer auf die Nutzung einer gemeinsamen Sprache (meist englisch) verständigt haben, die von allen Teilnehmern soweit beherrscht wird, dass eine grundlegende Kommunikation möglich ist. Eine Unterstützung durch ein automatisches Übersetzungssystem ist dann nur in solchen Situationen nötig, in denen ein Teilnehmer einzelne Sätze anderer nicht versteht oder Hilfe für die Formulierung eines einzelnen eigenen Redebeitrags benötigt. Vor diesem Hintergrund wird in diesem Beitrag ein als Demonstrator realisiertes System vorgestellt, das assistierend von den Teilnehmern genutzt werden kann.

Dazu wird zunächst das Anwendungsszenario genauer analysiert und eine bedarfsgerechte Lösung entwickelt. Nach der Systemarchitektur wird die domänenspezifische Adaption der Hauptkomponenten Übersetzung und Spracherkennung inkl. der dazu durchgeführten umfangreichen Sprachdatensammlungen beschrieben. Abschließend wird die Evaluierung des Systems beschrieben und deren Ergebnisse präsentiert und diskutiert.

## 2 Anwendungsszenario

Beim vorgestellten Übersetzungssystem handelt es sich um eine in ein Konferenzsystem integrierte Übersetzungshilfe, die auf Anfrage einzelne Wörter, Phrasen oder kurze Sätze ins Englische übersetzt. Der realisierte Demonstrator basiert auf der Einwahl mehrerer Teilnehmer in eine Telefonkonferenz, um miteinander in Englisch zu kommunizieren. Die integrierte Übersetzungshilfe ermöglicht dann den Konferenzteilnehmern, sich einzelne Wörter oder Phrasen von Deutsch nach Englisch übersetzen zu lassen, sofern sie mit den eigenen Sprachkenntnissen an ihre Grenzen stoßen.

Zugunsten einer verbesserten Übersetzungsleistung wurde auf einen Open-Domain-Ansatz verzichtet und die Sprachmodelle auf eine festgelegte Domäne optimiert. Bei der Auswahl der Zieldomäne wurde insbesondere der potenzielle Nutzerkreis eines zukünftigen Übersetzungsdienstes in Betracht gezogen. Aufgrund der Tatsache, dass insbesondere Geschäftskunden die Konferenzsysteme der Deutschen Telekom nutzen, wurde die Thematik ‚Projektmanagement‘ als Zieldomäne gewählt. Dies bedeutet, dass Spracherkennung und Übersetzung des Demonstrators speziell auf diese Domäne trainiert sind und somit domänenfremde Äußerungen nicht in jedem Fall richtig erkannt bzw. übersetzt werden können.

Der Demonstrator kann in zwei Modi gestartet werden, dem „Open Mode“ oder dem „Hidden Mode“. Wird die Übersetzungshilfe im Open Mode gestartet, können alle Teilnehmer der Konferenz den Dialog mit der Übersetzungshilfe verfolgen und ggfs. auch mit dem System in Interaktion treten. Die eigentliche Konferenz wird für den Zeitraum der Übersetzung unterbrochen. Das Ergebnis der Übersetzung ist sofort für alle hörbar und kann so schnell von den einzelnen Parteien in den richtigen Kontext gebracht werden. Ein kompliziertes Integrieren in den eigentlichen Kontext bleibt der anfragenden Person so wahrscheinlich in den meisten Fällen erspart. Beim Hidden Mode verlässt die anfragende Person kurzzeitig den Konferenzraum, um die Übersetzungshilfe allein anzufragen. Die Konferenz kann von allen anderen Parteien ungestört fortgesetzt werden. Die anfragende Person wird für den Zeitraum der Übersetzungsunterstützung von der Konferenz getrennt und kann dieser weder aktiv noch passiv durch Zuhören folgen.

Die Steuerung und Aktivierung des Systems findet über die Tastatur des Telefons statt, um eine möglichst hohe Robustheit zu erreichen.

## 3 Systemarchitektur

### 3.1 Gesamtsystem und Sprachplattform

Abbildung 1 skizziert das Zusammenspiel der einzelnen Systemkomponenten.

In der gewählten Systemarchitektur umfasst der Voice-Server die eigentliche Sprachplattform mit Voice-Browser- und CCXML-Engine sowie TTS-Komponente zur Ausgabe der Übersetzung. Der eigentliche Sprachdialog wird dabei über VoiceXML (Voice eXtensible Markup Language) beschrieben, während CCXML (Call Control eXtensible Markup Language) die Ruf- und Telefonsteuerung des Konferenzsystems beschreibt.

Die Spracherkennung (ASR) wurde zur Sicherstellung ausreichender Rechenleistung auf einen entsprechend dimensionierten externen Server verlagert und über eine MRCP-Schnittstelle an die Sprachplattform angebunden, um so auch später größtmögliche Flexibilität in der Wahl der ASR-Technologie zu besitzen.

Der Application-Server umfasst die eigentliche Anwendungslogik mitsamt den Ressourcen zur Dialog- und Rufsteuerung (CCXML- und VoiceXML-Dokumente) sowie den benötigten

Grammatiken und Audiodateien, womit eine klare Trennung der anwendungsspezifischen Komponenten von den übrigen Plattform- und Systemmodulen gegeben ist.

Das Übersetzungs-Modul (Machine Translation, MT) ist mit dem Application-Server über eine Socketschnittstelle verbunden und tauscht die Übersetzungsparameter im XML-Format aus, wodurch sich das auf Linux ausgelegte MT-Modul in die Windows-Systemumgebung einfach integrieren ließ. Die physikalische Trennung der MT vom Application-Server ergab sich zudem durch die umfangreichen Ressourcen-Anforderungen der MT (bis zu 4 GB RAM).

Die Anbindung der Sprachplattform an das Telefonnetz wurde durch den Einsatz eines VoIP-Gateways realisiert.

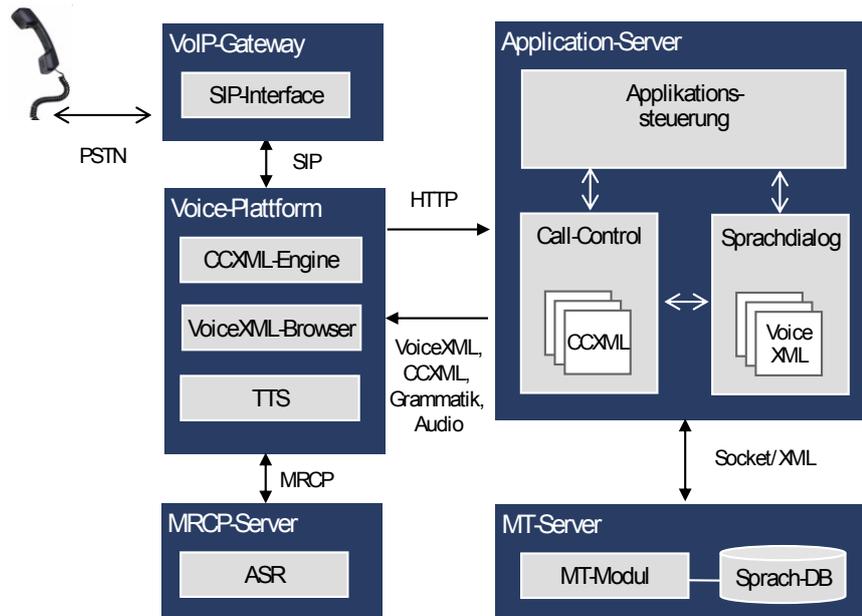


Abbildung 1 – Systemarchitektur

### 3.2 Spracherkennung und -synthese

Für die Spracherkennung wurde die kommerziell verfügbare Nuance-ASR 8.5 verwendet. Es handelt sich hierbei um eine Spracherkennungssoftware, die speziell für den Einsatz in telefonbasierten Sprachapplikationen entwickelt wurde. Die akustischen Modelle für die deutsche Sprache gehören zum Lieferumfang und mussten daher für das System nicht extra erstellt werden. Voraussetzung für den Einsatz der ASR der Fa. Nuance war Unterstützung von statistischen Sprachmodellen (statistical language models, SLM).

Zur Wiedergabe des ins Englische übersetzten Textes wurde die Sprachsynthese (text-to-speech, TTS) der Fa. Retorical verwendet.

### 3.3 Sprachübersetzung

Nach [1] lassen sich Sprachübersetzungssysteme in direkte Ansätze („Statistical machine translation“, SMT) und interlingua-basierte Ansätze aufteilen. Da eine geeignete sprachenunabhängige semantische Repräsentation der Domäne nicht zur Verfügung stand und auch nicht im Rahmen des Projektes erstellt werden konnte, wurde ein statistisches Übersetzungsmodell implementiert. Das System stellt eine Weiterentwicklung und Spezialisierung der Übersetzungskomponente des in [2] beschriebenen domänenunabhängigen phrasen-basierten Sprachübersetzers dar, welche auf neuen Daten trainiert wurde.

Die Übersetzung verwendet wie in Abbildung 2 dargestellt eine Strahlsuche und optimiert die gesamten gewichteten Modellkosten unter Berücksichtigung verschiedener Nebenbedingungen, insbesondere in begrenztem Umfang Wort- und Phrasenumstellungen im Zielsatz [3].

Zahlen, Uhrzeiten und Datumsangaben werden vom Spracherkennung in ausgeschriebenen Worten dargestellt. Um solche Ausdrücke zuverlässig übersetzen zu können und für die nachfolgende Sprachsynthese aufzubereiten, werden sie durch eine zusätzliche Komponente mittels regulärer Ausdrücke erkannt und in der Ausgabe speziell gekennzeichnet.

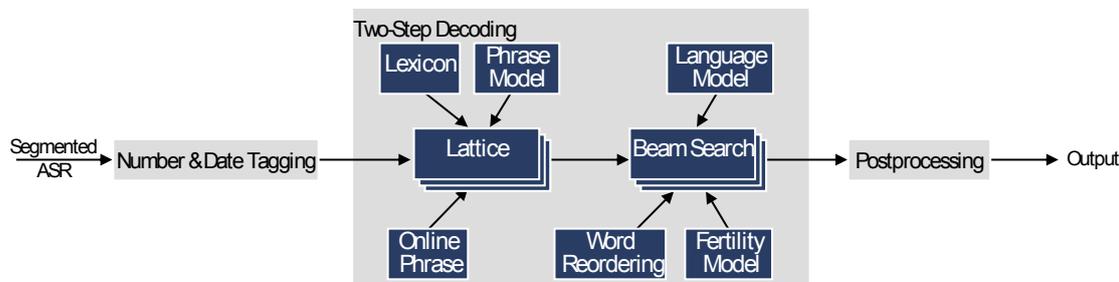


Abbildung 2 – Übersicht der Komponenten des Übersetzungsmoduls

## 4 Sprachmodelle für Spracherkennung und -übersetzung

Zum Training der Sprachmodelle ist neben einer Datenbasis für den allgemeinen Sprachwortschatz auch umfangreiches domänenspezifisches Sprachmaterial notwendig. Da für die Ziel-domäne des Übersetzungssystems (Projektmanagement) keine kommerziell verfügbaren Sprachdatenbanken gefunden werden konnten, wurden 40 Personen akquiriert, mit denen 30 Telefonkonferenzen (je 3 Teilnehmer plus Moderator) a 30 min zum Thema Projektmanagement durchgeführt wurden. Die Aufzeichnungen der Telefonkonferenzen im Umfang von ca. 15h wurden von Projektmitarbeitern manuell transkribiert und ins Englische übersetzt.

Als weitere Quellen für domänenspezifisches Material standen ein Projektmanagement-Handbuch in Deutsch und Englisch sowie ein Projektmanagement-Glossar in Deutsch zur Verfügung.

### 4.1 Sprachmodellierung für Spracherkennung

Um die Erkennung allgemeiner Phrasen und Redewendungen zu ermöglichen, wurden zum genannten domänenspezifischen Material folgende Sprachdaten hinzugezogen: Terminabsprachen [4], Parlamentsdebatten aus dem europäischen Parlament (EPPS) und Radionachrichtensendungen [5] sowie Zeitungstexte.

Domänenspezifisches Sprachmaterial				Allgemeines Sprachmaterial			
Telefonkonferenzen	PM-Glossar	PM-Book	Web	Termin	Radio	EPPS	Zeitung
77k	67k	13k	345M	520k	840k	24M	100M

Tabelle 1 – Größe der einzelnen Korpora in Worten

Zusätzlich zu den genannten Korpora wurde weiteres Material aus dem Web gesammelt. Hierzu wurden ca. 2000 N-Gramme aus dem vorhandenen allgemeinen und domänenspezifischen Material extrahiert, aus denen 10.000 Anfragen für eine Websuche generiert wurden. Über die Suchmaschine Google wurden für jede Anfrage mit Hilfe der Open-Source-Skripte der University of Washington [6] die ersten vier Ergebnisseiten heruntergeladen.

Vor dem Training des Sprachmodells wurden die Textdaten der Korpora vorverarbeitet, d.h. satzweise resegmentiert, Interpunktion und Sonderzeichen wurden entfernt und Abkürzungen durch die gesprochene Form ersetzt. Bestimmte Wörter wurden für eine hohe Generalisierung des Modells durch semantische Klassen ersetzt, z.B. Namen, Zahlen, Uhrzeiten und Daten.

Mit Hilfe des SRI Language Modeling Toolkit [7], [8] wurde das Vokabular aus den Trainings-Korpora durch Minimierung der Perplexität auf einen Testkorpus selektiert. Die in zwei Selektionen - einmal mit den Telefonkonferenzen und einmal mit PMBook und PM-Glossar als Testkorpus - ermittelten Vokabulare wurden um alle Wörter aus den Telefonkonferenzen, PMBook und PM-Glossar ergänzt, wodurch sich ein Umfang von etwa 30.914 Wörter und eine OOV-Rate auf den Evaluationsdaten von 2.6% ergab. Zur weiteren Optimierung der Spracherkennung wurden abschließend die noch fehlenden Wörter aus den Evaluationsdaten hinzugefügt, wodurch das endgültige Vokabular eine Größe von 31.155 Wörter hatte.

Für das finale Trigramm-Sprachmodell wurden mehrere Sprachmodelle, die auf verschiedenen Korpora trainiert wurden, interpoliert und zusammen gemischt. Die Interpolationsgewichte wurden hierbei durch Minimierung der Perplexität auf den Testdaten optimiert. Alle Sprachmodelle wurden mit dem modifizierten Kneser-Ney-Discounting unter Verwendung des SRI Language Modeling Toolkits berechnet.

Nach Hinzunahme der Webdaten zum Sprachmodell zeigte sich, dass sich die Interpolationsgewichte stark zugunsten der Webdaten verändern. Gleichzeitig ist zu erkennen, dass PM-Glossar, PM-Book und Radionachrichtensendungen nahezu keinen Beitrag am gesamten Sprachmodell leisten. Dies ist damit zu erklären, dass es sich um relativ spezielle Textdaten handelt, die nicht zum Stil der Telefonkonferenzen passen. Neben den Webdaten und Telefonkonferenzen sind vor allem die Terminabsprachen am wichtigsten. Die Perplexität auf den Testdaten sinkt von 188 auf 161 und die auf den Evaluationsdaten auf von 215 auf 185. Nachdem die Testdaten noch hinzugefügt wurden, vermindert sich die Perplexität auf den Evaluationsdaten auf 183.

## **4.2 Sprachmodelle für die Sprachübersetzung**

Für die Übersetzung wurde ein statistisches Verfahren umgesetzt, bei dem halbautomatisch aus großen, parallelen bilingualen und monolingualen Textkorpora statistische Modelle für die Übersetzung erstellt werden. Dadurch kann im Gegensatz zur Vorgehensweise bei traditionellen grammatik- und regelbasierten Ansätzen auf linguistisches Expertenwissen verzichtet werden und es können innerhalb kurzer Zeit Übersetzungssysteme für neue Sprachpaare und Zieldomänen trainiert und adaptiert werden.

Während der Trainingsphase wurden Phrasenpaare aus dem bilingualen Trainingskorpus extrahiert, indem ein eingeschränktes Wort-zu-Wort-Alignment für ein ganzes Satzpaar optimiert wird. Für hochfrequente Phrasenpaare kurzer Länge werden die besten Übersetzungsalternativen während der Trainingsphase extrahiert und zusammen mit allen relevanten Modellparametern als Phrasenmodell abgespeichert. Aus Speicherbedarf- und Laufzeitgründen ist dies jedoch nicht für sämtliche während der Übersetzungsphase auftretenden Phrasen möglich; daher werden seltenere und längere Phrasen online, d.h. zur Übersetzungszeit, mittels eines Suffixarrays im Korpus lokalisiert und ein On-the-fly-Phrasenalignment durchgeführt.

Die allgemeine Datenbasis für das Training bestand aus zwei größeren bilingualen Textkorpora (Deutsch-Englisch) mit gesprochener Sprache, dem EPPS-Korpus mit aufbereiteten Parlamentsdebatten aus dem europäischen Parlament sowie dem BTEC-Korpus mit einem hohen Anteil an allgemeiner Konversationssprache. Für die Anpassung an die Zieldomäne wurden die bereits genannten Telefonkonferenzen, PM-Book und PM-Glossar (ins Englische übersetzt) sowie kurze bilinguale Wortlisten verwendet.

Diese Daten wurden nach Aufarbeitung für das Training der statistischen Modelle verwendet und dabei gegenüber der allgemeinen Datenbasis durch optimierte Modellinterpolation höhergewichtet. Mit Daten aus weiteren Korpora sowie gesammelten monolingualen Webdaten wurde experimentiert, aber im ersten Ansatz keine Verbesserung erzielt.

	Domänenspezifisch					Allgemein	
	Telefonkonferenzen	PM-Glossar	PM-Book	Phrases	Doppelwörter	BTEC	EPPS
<b>Deutsch</b>	76.748	66.364	12.450	435	112	802.911	21.292.259
<b>Englisch</b>	76.085	78.758	14.182	557	151	803.761	22.406.324

**Tabelle 2** – Verwendete bilinguale Trainingsdaten (Größe in Worten)

## 5 Evaluierung und Ergebnisse

### 5.1 Spracherkennung

Um Aussagen über die Erkennungsgenauigkeit der Spracherkennung machen zu können, musste weiteres Sprachmaterial für Testzwecke beschafft werden. Zu diesem Zweck wurden 2 Testsets definiert: Testset A mit 50 Phrasen aus den domänenspezifischen Trainingskorpora sowie Testset B mit 50 Phrasen, die zwar zur Domäne Projektmanagement gehörten, jedoch nicht direkt im Trainingskorpus vorhanden waren.

Ca. 30 Personen wurden gebeten, ein telefonbasiertes Aufnahmesystem anzurufen und jeweils 10 Phrasen aus Testset A und B nachzusprechen. Vor der Aufnahme einer Phrase wurde der Anrufer gebeten, die Phrasen-ID über die Telefontastatur einzugeben, so dass die einzelnen Wav-Files automatisch dem zugehörigen Text zugeordnet werden konnten.

Insgesamt konnten auf diese Art 533 Phrasen aufgezeichnet werden, die zum Test der Spracherkennung verwendet wurde. Da es aufgrund der Projektlaufzeit nicht möglich war, einen größeren Testkorpus aufzuzeichnen, sollten die Tests zunächst dazu dienen einen groben Eindruck von der Erkennungsleistung des Systems zu erhalten.

Von den 533 Samples kam es bei 27 zu einer Rückweisung durch die Spracherkennung, d.h. es konnte kein Ergebnis ermittelt werden. Bei den verbleibenden 506 Phrasen konnte eine Satzerkennungsrate von 40,51% ermittelt werden. Hierbei wurde ein Erkennungsergebnis nur dann als korrekte Hypothese gewertet, wenn die gesamte erkannte Phrase exakt der Referenz entsprach.

Mit Hilfe des NIST Scoring Toolkit (sclite Version 2.2) wurde die Worterkennungsrate mit 82,7% ermittelt.

Anzahl	Korrekte Hypothesen	Vertauschungen	Rückweisungen
506	40,51%	59,49%	0,00%

**Tabelle 3** – Satzerkennungsraten

Anzahl Worte	Korrekte Ergebnisse	Vertauschungen	Einfügungen	Löschungen
3670	82,7%	14,7%	3,0%	2,60%

**Tabelle 4** – Wortkennungsraten

## 5.2 Sprachübersetzung

Für das Übersetzungssystem wurde eine Testumgebung aufgebaut, die auch zur automatisierten Verifikation zukünftiger Optimierungen dienen soll. Als Metrik wurde das etablierte BLEU-Verfahren [1] gewählt, das trotz verschiedener Nachteile das zurzeit am weitesten verbreitete Evaluierungsverfahren darstellt. Aufgrund der komplizierten Berechnung ist eine einfache Interpretation nicht möglich. BLEU eignet sich also nur zum relativen Vergleich unter gleichen Testbedingungen, was hier aber ausreicht.

Für einen ersten Test wurden zunächst zwei spezifische Testkorpora erstellt. Dazu wurden deutsche Texte aus verschiedenen vorhandenen Korpora aufbereitet und eine übersetzte englische Referenz manuell erstellt. Die Texte wurden auf einheitliche Schreibweise hin normiert. Im einzelnen wurden Groß-Kleinschreibung und Zeichensetzung entfernt, Abkürzungen aufgelöst sowie Zahlen- und Datumsangaben normiert. Synonyme wurden nicht ersetzt. Weiterhin wurden für das Anwendungsszenario untypische Sätze und Wiederholungen häufig gebrauchter kurzer Floskeln entfernt.

Korpus	Sätze EN	Wörter EN	Quelle
TELCO	1664	12278	Subset aus dem Trainingskorpus
ASR100	100	808	Texte aus der ASR-Evaluation

**Tabelle 5**– Korpora für den MT-Test (Parallel deutsch - englisch)

Beim TELCO-Korpus handelt es sich um Mitschnitte aus realen Konferenzszenarien. Die Audio-Daten wurden manuell transkribiert und übersetzt.

Der ASR100-Korpus wurde aus dem Material der ASR-Tests gewonnen. Dabei waren 100 deutsche Phrasen vorgegeben. Im Referenztest wurde das ursprüngliche Textmaterial genutzt, im Basistest die von der Spracherkennung erzeugten Transkriptionen der tatsächlich gesprochenen Sprachproben.

Testkandidaten waren die MT-Komponente alleine sowie das Gesamtsystem bestehend aus ASR und MT. Die Testergebnisse sind in folgender Tabelle zusammengestellt:

Test	Kandidat	Korpus	Beschreibung	Bleu
telcoVer	MT	TELCO	Trainingsverifikation : Qualität der MT	0,95
asrRef	MT	ASR100	Referenz ASR: Referenzqualität der MT für ASR-Korpus	0,44
asrBas	ASR & MT	ASR100	Basistest ASR : Gesamtqualität ASR/MT	0,27

**Tabelle 6**– Testergebnisse für MT

Der Verifikationstest (*telcoVer*) zeigt erwartungsgemäß einen hohen BLEU-Wert, da die MT-Komponente mit diesem Material trainiert wurde.

Der Referenztest (*asrRef*) liefert den Qualitätswert der isolierten MT-Komponente. Der Bleu-Wert 0,44 kann als relativ gut bewertet werden. Es gilt jedoch zu beachten, dass trotz repräsentativem domänenspezifischem Inhalt der Korpus nur 100 Sätzen umfasste, die zum Teil auch im großen Trainingskorpus vorhanden waren.

Beim Basistest (*asrBas*) wird die Gesamtqualität des Übersetzungssystems bewertet. Hier wird ersichtlich, wie die Spracherkennung die Übersetzungsqualität beeinflusst.

## 6 Zusammenfassung und Ausblick

Das hier vorgestellte Projekt zeigt, dass mit den heute verfügbaren Technologien innerhalb kurzer Zeit ein Prototyp für die Übersetzung von Nutzeräußerungen in Telefonkonferenzen von Deutsch nach Englisch erstellt und auf die Zieldomäne Projektmanagement adaptiert werden kann. Dazu wurde ein vorhandener domänenunabhängiger Sprachübersetzer auf eine neue Domäne und Systemumgebung adaptiert.

Das entstandene Übersetzungsmodul erreicht eine Geschwindigkeit von durchschnittlich einer Sekunde pro Äußerung und enthält eine Komponente zur Identifikation von Zahlen, Uhrzeiten und Datumsangaben, um diese in der Ausgabe zu kennzeichnen und für die nachfolgende Sprachsynthese aufzubereiten.

Im Bereich der Spracherkennung sind Verbesserungen besonders durch den optimierten Einsatz von zusammengesetzten Wörtern zu erwarten, die derzeit noch zu Fehlern bei der Übersetzung führen können („Projekt Leiter“ vs. „Projektleiter“). Die Gesamtleistung sollte durch zusätzliches und gezielter ausgewähltes Trainingsmaterial verbessert werden können. Gerade für die Übersetzung sind Verbesserungen durch den Einsatz von mehr und besser aufbereiteten Daten, sowie durch spezialisierte Behandlung semantischer Ausdrücke, z.B. Zeitangaben, zu erwarten.

Im Rahmen dieser Arbeit wurden Referenzimplementierungen zur automatisierten Durchführung von Test u.a. nach dem BLEU-Verfahren entwickelt, Tests wurden mit und ohne Spracherkennung durchgeführt, sowie Empfehlungen für die Aufbereitung von Referenzdatenmaterial erstellt. Für einen Produktiveinsatz wären darüber hinaus Verbesserungen im Hinblick auf Robustheit, Ressourcenbedarf und Skalierbarkeit erforderlich, die bei der Erstellung dieses Prototyps nicht berücksichtigt werden konnten.

*Danksagung:* ...

### Literatur

- [1] Schultz, T. und Kirchhoff, K. (Ed.): Multilingual Speech Processing. Academic Press, Oxford. 2006.
- [2] Fügen, C., Kolss, M., Bernreuther, D., Paulik, M., Stüker, S., Vogel, S. und Waibel, A.: Open Domain Speech Recognition & Translation: Lectures and Speeches. Proc. ICASSP 2006; Toulouse, Frankreich; IEEE, Mai 2006.
- [3] Vogel, S.: SMT Decoder Dissected: Word Reordering. In Int. Conf. on Natural Language Processing and Knowledge Engineering (NLP-KE), Beijing, China, 2003.
- [4] Verbmobil: Sprachkorpora <http://www.phonetik.uni-muenchen.de/Bas/BasVM1deu.html> und <http://www.phonetik.uni-muenchen.de/Bas/BasVM2deu.html>
- [5] Köhn, P.: Parallel Corpora, <http://people.csail.mit.edu/koehn>
- [6] University of Washington: Collect Data Tools, [http://ssli.ee.washington.edu/projects/ears/WebData/web\\_data\\_collection.html](http://ssli.ee.washington.edu/projects/ears/WebData/web_data_collection.html)
- [7] Venkataraman A. and Wang, W.: Techniques for effective vocabulary selection. In Proceedings of Eurospeech, Geneva, 2003.
- [8] Stolcke, A.: “SRILM – An Extensible Language Modeling Toolkit,” in ICSLP, 2002.
- [9] Papineni, K.A., Roukos, S., Ward, T. und Zhu, W.J.: BLEU: a Method for Automatic Evaluation of Machine Translation. In Proceedings of 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, PA, July 2002.