Grażyna Łopuszańska
ORCID: 0000-0002-8048-7431
Universität Gdańsk, Gdańsk

Linguistische Treffen in Wrocław, Vol. 15, 2019 (I)
ISSN: 2084–3062, e-ISSN: 2657–5647
DOI: 10.23817/lingtreff.15-12
S. 145–156

Maschinelle Übersetzung – Grenzen und Möglichkeiten

Die maschinelle Übersetzung, die die sofortige Übersetzung großer Textmengen ermöglicht, wird zu einem äußerst wertvollen Werkzeug für eine schnelle Kommunikation. Der vorliegende Beitrag beschreibt die Probleme, die bei der maschinellen Übersetzung auftauchen, und Versuche, diese zu lösen.

Schlüsselwörter: Kommunikation, maschinelle Übersetzung, Translatica, Google Translate

Machine Translation – Boundaries and Capabilities

Machine translation as permitting instant translation of an extensive volume of texts, is becoming an invaluable tool for quick communication. This article reviews the problems encountered in machine translation and attempts to solve them.

Keywords: communication, machine translations, Translatica, Google Translate

Author: Grażyna Łopuszańska, University of Gdańsk, Department of Applied Linguistics and Translation Studies, ul. Wita Stwosza 55 80–308 Gdańsk, Poland, e-mail: filgl@ug.edu.pl

Received: 20.11.2017 **Accepted:** 30.5.2019

Seitdem die ersten Forschungsarbeiten in den Dreißigerjahren im Bereich der maschinellen Übersetzung begonnen haben, wurden diese sowohl gelobt als auch kritisiert. Trotz ihrer kurzen Geschichte ist die maschinelle Übersetzung durch bemerkenswerte Fortschritte gekennzeichnet, die die Entwicklungen in verbundenen Wissennschaftsgebieten, wie Informatik oder Linguistik, wiederspiegeln. Obwohl die maschinelle Übersetzung anfangs häufig für eine verlorene Sache gehalten wurde, sind die maschinellen Übersetzungssysteme von den einfachsten, die Wort-für Wort-Übersetzungen lieferten, bis zu komplexen, die auf den künstlichen neuronalen Netzwerken basieren, evolviert. Die Geburt des Internets hat die Forschung im Bereich der maschinellen Übersetzung eine neue Richtung gegeben. Gleichzeitig ist aber auch eine so hohe Nachfrage nach Rohübersetzungen wie noch nie entstanden, weil heutzutage nicht nur große Unternehmen, sondern auch Privatpersonen mit fremdsprachigen Informationen überschwemmt werden. Schon früher war vielen klar, dass vollautomatische Übersetzungen von hoher Qualität nicht unbedingt das einzige Ziel der Forschung zur maschinellen Übersetzung sein muss. Die Bedürfnisse der sich ständig entwickelnden Informationsgesellschaft für schnelle und oft rohe Übersetzungen sind ein guter Beleg dafür. Die maschinelle Übersetzung spielt also heutzutage eine bedeutende Rolle und wird sicherlich auch in Zukunft nicht an Wichtigkeit verlieren. Obwohl schon seit dem Beginn der Forschungsarbeiten im Bereich der maschinellen Übersetzung die Systeme

bedeutsam weiterentwickelt wurden und viele von ihnen regelmäßig mit mehr oder weniger Erfolg benutzt werden, gibt es immer noch große Probleme bei der Vervollkommnung der Systeme.

Die Probleme in der maschinellen Übersetzung resultieren nicht nur aus der Tatsache, dass die Kunst des Übersetzens selbst besonders schwer ist und sogar für den besten Übersetzer eine große Herausforderung darstellt, sondern auch aus der Komplexität der natürlichen Sprachen und der Programmierung der Maschinen. Viele Probleme wurden schon am Anfang der Forschung zur maschinellen Übersetzung identifiziert, aber nur wenige von ihnen wurden völlig zufriedenstellend gelöst, was die Qualität der maschinellen Übersetzung bedeutsam beeinflusst. Aus diesem Grund ist das Thema ohne Zweifel aus Forscher-, Anbieter- und natürlich Anwendersicht besonders wichtig, hochaktuell und interessant.

Seit dem Beginn der Arbeiten im Bereich der maschinellen Übersetzung wurden die Systeme und Methoden bedeutsam weiterentwickelt, trotzdem gibt es immer noch zahlreiche Probleme der technischen und linguistischen Art, die die Qualität der maschinell hergestellten Übersetzungen beeinflussen. Im Folgenden werden die aktuellen semantischen, pragmatischen und syntaktischen Probleme der maschinellen Übersetzung angesprochen. Obwohl die Analyse der Probleme aus technischer Perspektive nicht weniger wichtig ist, ist sie für die Zwecke der Sprachwissenschaft nicht relevant und wird deshalb an dieser Stelle nicht besprochen.[1] Kurz nach den ersten Versuchen im Bereich der maschinellen Übersetzung in den Jahren 1948–1952, als Richard Richens und Andrew Booth (veröffentlicht erst 1955) eine Art maschinellen Wörterbuches vorschlugen, das wörtliche Übersetzungen produzierte, stellte Warren Weaver (1949) in seinem bekannten Weaver - Memorandum die Grenzen der ersten,

[1] Da die technischen Fragen eine bedeutende Rolle in der Entwicklung der maschinellen Übersetzung spielen, werden die wichtigsten gegenwertigen technischen Probleme kurz angedeutet. Obwohl in der maschinellen Übersetzung ein riesiger technischer Fortschritt stattfand und Datenspeicherfähigkeit sowie Rechenleistung der Computer sich bedeutsam verbessert haben, gibt es immer noch, außer Problemen, die aus der Natur der natürlichen Sprachen resultieren Einschränkungen pur technischer Art. Laut Doug Arnold (2003) sind vier wichtige Einschränkungen des Computers die Hauptursache anderer Probleme in der maschinellen Übersetzung. Erstens können die Maschinen nicht mit Aufgaben betraut werden, die in ihren Systemen nicht gut beschrieben wurden. Im Falle der natürlichen Sprachen ist es sehr schwer alle Regeln, die in den Sprachen herrschen, zu programmieren. Das nächste Problem technischer Art ist, dass die Maschinen selbst kaum etwas lernen können. Maschinen können auch nicht denken und besitzen kein Weltwissen. Sie haben auch immer noch eine begrenzte Rechenleistung. Philip Koehn (2010) konzentriert sich auf die Beschreibung der Rechen- und Modellierungsprobleme und weist u. a. auf das Problem der mangelnden Daten in Korpora hin, die notwendig für das Erhalten von Übersetzungen guter Qualität sind, oder das Problem der Sprachenbeschreibung in den Systemen. Er macht Leser darauf aufmerksam, dass auch die automatische Wahl der „besten" Übersetzung unter vielen Varianten immer noch Schwierigkeiten bereitet.

direkten Methode dar und sprach eines der wichtigsten Problemen der maschinellen Übersetzung, nämlich die Mehrdeutigkeit, an. Weaver wies darauf hin, dass die Übersetzung viel mehr als eine einfache Wortübersetzung ist und er schlug vor, das Problem der Mehrdeutigkeit durch einest statistische Analyse des Kontexts oder durch die Verwendung einer Art Universalsprache zu lösen. Dies war aber problematisch, weil die Größe des zu Disambiguierung[2] notwendigen Kontexts nicht immer dieselbe ist.

Die Frage der Mehrdeutigkeit hat später Reifler (1950, zit. nach Hutchins 1986:38) aufgenommen, der die Idee der Prä-Edition und Post-Edition einführt. Ein Prä-Editor sollte die Wortklasse der mehrdeutigen Wörter und den Text syntaktisch vereinfachen, während die Aufgabe des Post-Editors war, eine korrekte Übersetzung des Wortes aus den von dem Computer vorgeschlagenen möglichen Varianten zu wählen und die Wortstellung des Satzes an die Zielsprache anzupassen. Es muss aber unterstrichen werden, dass es viele Fälle gibt, wo die Erkennung der Wortklasse nicht hilft, das Wort zu disambiguieren. Das Konzept der Prä- und Post-Edition war eine Teillösung, die das Problem der Mehrdeutigkeit nicht völlig lösen konnte. Es war aber damals sehr innovativ und wird bis heute allgemein verwendet. Reifler's Konzept wurde auch von Bar-Hillel (1951) anerkannt. Er stellte fest, dass der menschliche Eingriff in dem Prozess des Übersetzens nötig ist, um eine Übersetzung von guter Qualität zu erhalten, weil es keine Methode gibt, die erlauben würde, das Problem der semantischen Mehrdeutigkeit eindeutig zu lösen.

Während der ersten Konferenz für maschinelle Übersetzung hat Oswald (1952) die Erstellung von Macro-Glossarien, in denen die Wörter nur bereichsspezifische Bedeutungen tragen würden als eine Lösung des Mehrdeutigkeitsproblems vorgeschlagen. Sie waren aber nur für bestimmte Bereiche zuständig. Das Konzept wurde von vielen Forschern angenommen. Es erwies sich aber nicht als mangelfrei, weil Texte nicht nur aus bereichspezifischen Wörtern bestehen. Diese Methode löste also nicht das Problem der Mehrdeutigkeit der üblichen Wörter, wie z. B. *finden*. Es muss aber betont werden, dass die maschinellen Übersetzungssysteme erfolgreich sind, wenn sie zum Übersetzen von sehr bereichsbeschränkten Texten verwendet werden, worauf auch Hutchins/Sommers (1992) hinweisen. Als Beispiel führen sie METEO an, ein System, das 1977 entstand und erfolgreich die Übersetzungen von Wettervorhersagen produzierte. Bei der in den früheren Jahren verwendeten Interlingua-Methode[3], blieb das Problem weiter ungelöst. Bar-Hillel (1960) wies damals darauf hin, dass ohne integriertes außersprachliches Wissen, die Systeme nicht imstande sind, die Mehrdeutigkeit zu lösen. Das Problem lag aber darin dass keiner wusste, wie man ein solches Wissen einer Maschine beibringen kann. In der Mitte der 70er Jahre hat Yorick Wilks (1979)

[2] Disambiguierung erfolgt dann, wenn der sprachliche Kontext nur noch eine Bedeutung für *Bank* zulässt, z. B. in: *Ich setze mich auf die Bank* oder *Ich hebe in der Bank Geld ab*.

[3] Interlingua Methode basiert auf Interlingua, d. h. auf universeller Zwischensprache, mithilfe von der ein Text indirekt übersetzt wird. Der Ausgangstext wird zuerst in der Form einer universalsprachlichen Repräsentation also Interlingua dargestellt, um später daraus in der Zielsprache erzeugt zu werden.

einen Künstlichen-Inteligenz-Ansatz[4] zur interlingualen maschinellen Übersetzung entwickelt, der auf der grundsätzlichen semantischen Analyse der Zusammenhänge zwischen Subjekt, Prädikat und Objekt basiert. Die Bedeutung der einzelnen Wörtern wurde durch „semantic primitives" also Basiskonzepte (wie Ding, Mensch) beschrieben, die später zur Bestimmung populärer und präferierter Verbindungen von Wörtern und Konzepten in bestimmten Sätzen dienten. Damit wird eine Einschränkung in Hinsicht auf den semantischen Kontext vorgenommen, in der ein Wort auftreten kann. Mit diesem Ansatz hat Wilks die Theorie von Bar-Hillel nachgefragt und bewiesen, dass bestimmtes alltäglichen Wissens in den Übersetzungssystemen integriert und von diesen manipuliert werden kann. Wie Wilks (2009) betont, besteht aber Schwierigkeit darin, dass es keine ausreichenden Wissensbasen erstellt wurden und es nicht möglich ist, das Potential dieses Ansatzes in einer größeren Skala zu prüfen.

Die Idee selbst, die Systeme mit bestimmtem Wissen auszustatten, um somit eine Übersetzung von besserer Qualität zu bekommen, nahm an Popularität zu. 1991 wurde das komplexe wissensbasierte[5] prototypische System KANT, ein verbesserter Nachfolger von KBMT-89 entwickelt, das laut Carbonell et al. (1992), dank der integrierten Wissensbasis imstande war, viele in den technischen Texten vorkommende Mehrdeutigkeiten zu lösen. Es muss an dieser Stelle betont werden, dass wissensbasierte maschinelle Übersetzungssysteme nur dann erfolgreich sein können, wenn sie bei stark begrenzten Fachgebieten angewendet werden, wo die Zahl der möglichen semantischen Repräsentationen reduziert ist, weil es nicht möglich ist, das gesamte Weltwissen in einer Datenbasis zu schließen.

Koehn (2010) führt an, dass während das wortbasierte statistische System nicht befriedigend mit der Mehrdeutigkeit umgeht, können die populären phrasenbasierten Systeme bessere Ergebnisse darbieten, weil sie nicht nur Wörter, sondern den nahen Kontext, ganze Phrasen analysieren. Der größte Nachteil der statistischen Systeme ist aber, dass die Qualität der Übersetzungen stark von der Größe der Corpora abhängig ist. Darüber hinaus scheint sich das Problem der lexikalischen Mehrdeutigkeit nur mithilfe des nahen Kontextes lösen zu lassen. Bei der strukturellen und referentiellen

[4] Das Hauptziel des Künstlichen-Intelligenz-Ansatzes war semantisches Wissen in einem System anzubauen, um die Qualität der Übersetzungen durch semantische Analyse zu verbessern. Es hat sich aber erwiesen, dass die Programmierung des semantischen Wissens sehr aufwändig ist und es keine genügend große Datenbasen gibt, um die Effektivität dieses Ansatzes zufriedenstellend zu prüfen.

[5] Die Idee das System mit bestimmtem Wissen auszustatten hat sich als erfolgreich erwiesen und gab den Anfang der wissensbasierten maschinellen Übersetzung. Der Ansatz besteht darin, dass ein System in diese Lage zu versetzen, das linguistische Wissen und das allgemeine Weltwissen aus Enzyklopädien, Thesauri etc. zu kodieren und zu interpretieren. Wegen bestimmter Einschränkungen und der Komplexität wird der Ansatz meistens für experimentale Zwecke verwendet. Darüber hinaus ist es anzumerken, dass dieser Ansatz nur in beschränkten Anwendungsbereichen effektiv sein kann.

Mehrdeutigkeit braucht man eine umfangreiche Textanalyse und die Anwendung des Weltwissens, womit die modernen Systeme, trotz der bedeutsamen Entwicklung, immer noch Schwierigkeiten haben (Arnold 2003; Hardmeier/Federico 2010).

Ein weiteres Problem, das eng mit dem Problem der Mehrdeutigkeit verbunden ist, ist die Idiomatizität. Da die Idiome nicht wörtlich verstanden und übersetzt werden können, konnten die ersten wörterbuchbasierten Systeme nicht mit ihnen zurechtkommen. Zwar hat Bar-Hillel (1952) vorgeschlagen, durch die Erstellung eines besonderen Phrasenwörterbuchs das maschinelle Übersetzen von Idiomen zu unterstützen, schon bald hat er aber bemerkt, dass diese Methode mit den die Größe des Wörterbuchs betreffenden Einschränkungen belastet ist, weil Idiome in verschiedenen Forman vorkommen können. Es ist auch problematisch, dass abhängig vom Kontext einige Idiome wörtlich oder nicht wörtlich interpretiert werden können. Bar-Hillel hat auch dafür eine Lösung vorgeschlagen, nämlich die Beteiligung des Posteditors, der aus den von der Maschine produzierten möglichen Übersetzungen, eine zu dem Kontext passende wählen würde. Die von ihm vorgeschlagenen Methoden wurden später angewendet. Wie Martin Volk (1998) schreibt, wurden die Systeme der Neunzigerjahren wie „Langenscheidt T1Professional" oder „Personal Translator Plus 98" mit tausenden Paaren von sich in separaten Lexikon oder im Übersetzungsspeicher befindenden Idiomen ausgestattet. Die Idiomen wurden aber wegen ihrer Komplexität nicht automatisch übersetzt. Die Paare dienten zum manuellen Anpassung der korrekten Übersetzung, was von dem Posteditor gemacht wurde. Im „Rosetta Project" (Rosetta 1994) aus derselben Zeit ist es den Forschern gelungen, die Idiome automatisch zu übersetzen. Sie haben das Problem, das sich die Formen der Idiomen ändern können dadurch gelöst, dass die Sätze während der Analyse zum Indikativ umgewandelt und die Verben zu ihrer Standardform zurückgebracht wurden. So erhielt man die Standardform der Idiomen, nach der im Lexikon gesucht werden konnte. Wichtig ist, dass das System sowohl die idiomatische, als auch die wörtliche Bedeutung der analysierten Phrase berücksichtigt hat. Die automatische Wahl des korrekten Variantes aus dem System wurde dank der Implementation der grammatischen Regeln möglich, die u. a. die Einschränkungen für das syntaktische Verhalten der Idiomen bestimmen. Die Anwendung der komplexen grammatischen Regeln hat aber dazu geführt, dass es nicht möglich war zu beweisen, inwiefern das System imstande ist, vollständige Übersetzungen zu liefern (Landsbergen 1998). Es ist bemerkenswert, dass laut Hutchins (2006) die heutzutage populärste statistische Methode die Möglichkeit gibt, die Idiome korrekt maschinell zu übersetzen, weil sie auf den Korpora basiert, die aus schon existierenden professionell angefertigten Übersetzungen bestehen. Die Phrasenbasierte maschinelle Übersetzung scheint hier von besonderer Bedeutung zu sein, weil die Idiome nicht zerlegt werden können und als Ganzes übersetzt werden müssen, da ihre Bedeutung nicht von den bestehenden Wörter abgeleitet werden kann. Salton et al. (2014) weisen aber darauf hin, dass die gegenwärtigen phrasenbasierten Systeme oft einfach die Sequenzen von Wörtern berücksichtigen, anstatt die semantischen oder grammatischen Phrasen zu analysieren, was zur fehlerhaften Übersetzung führen kann. Sie haben bewiesen, dass die

Qualität des Übersetzens idiomatischer Wendungen stark von der Art und Größe der Korpora abhängt. Dies bedeutet, dass das, was in dem Korpus nicht beinhaltet ist, nicht übersetzt werden kann. Daher werden weitere Forschungsarbeiten geführt, die sich auf das Umgehen mit den Wörtern, die nicht in den Korpora auftreten, fokussieren. Als die Lösung des Problems der neuen Wörter im Korpus schlagen Okuma et al. (2008) z. B. die Integrierung von bestimmten separaten Wörterbüchern mit dem statistischen System vor. Sie haben bewiesen, dass diese Idee einen positiven Einfluss auf die Qualität des Übersetzens auch von Idiomen hat. Auch die korrekte Erkennung und Berücksichtigung ganzer Idiomen, die später von dem statistischen System übersetzt werden, ist von großer Bedeutung und wird nach wie vor weiter erforscht (Ren et al. 2009).

Nicht weniger problematisch scheint die Frage der Komposition zu sein. Da Komposita neue Wörter sind, die aus vielen schon bereits vorhandenen Wörtern bestehen, fällt es den maschinellen Übersetzungssystemen schwer diese Einheiten zu übersetzen, besonders dann, wenn die neuen Wörter in dem Korpus oder im Lexikon des Systems nicht vorkommen. Schon in den Fünfzigerjahren hat Reifler (1955) darauf hingewiesen, dass es unmöglich ist, alle Komposita in das Lexikon einzuführen, weil sich die Sprache ständig weiterentwickelt. Hinzu kommt, dass das System umso langsamer wird, je größer das Lexikon ist. Er hat vorgeschlagen, die Bedeutung der Komposita durch die mechanische Identifizierung ihrer Bestandteile erkennbar zu machen, d. h. dass sie zuerst in die Lexikon vorkommenden Wörter zerlegt werden müssen, damit sie dann übersetzt werden können. Er hat auch auf das Problem der Mehrdeutigkeit der Komposita hingewiesen und bemerkt, dass die Bedeutung der Bestandteile der Zusammensetzungen vom Kontext abhängig ist. Das Problem der Komposita erforschten Nießen/Ney (2000) und stellten fest, dass statistische Systeme mit bestimmten Phänomenen natürlicher Sprache nicht erfolgreich umgehen können, weil sie über ein ungenügendes linguistisches Wissen verfügen. Dies ist auch ein Grund dafür, dass Komposita nicht grammatisch korrekt in die Zielsprache übersetzt werden können, obwohl sie durch das maschinelle Übersetzungssystem zerlegt werden können und ihre Bestandteile sich im Korpus befinden. Nießen/Ney (2000) haben bestimmte morpho-syntaktische Regeln für Deutsch und Englisch in das statististische System implementiert und das Zerlegen der Komposita angewendet, was sich als erfolgreich erwies und zur besseren Übersetzungsergebnissen führten. Im Gegensatz zu dieser linguistischen Methode haben Koehn/Knight (2003) eine mehr auf dem Korpus basierte Methode vorgeschlagen. Bei dieser Methode spielt die Häufigkeit des Auftretens der Wörter im Korpus eine wichtige Rolle. Da die Zusammensetzung verschieden zerlegt werden kann, was nicht immer zu den gewünschten Ergebnissen führt, werden die Bestandteile des Kompositums nach der Häufigkeit ihres Auftretens bestimmt. Mithilfe der Informationen aus dem übersetzten Text und dem daraus gewonnenen Übersetzungslexikon, wird die wahrscheinlichste Übersetzung hergestellt. Um die Zerlegung der in den Komposita häufig vorkommenden und verschieden übersetzten Präfixe und Suffixe zu vermeiden, wurden in das System mithilfe von part-of-speech tagging (das Markieren der Wortarten) Informationen über die Wortarten implementiert, was die möglichen

Bestandteile der Komposita zu bestimmen erlaubt. Obwohl diese Methode die Qualität der statistisch produzierten Übersetzungen verbessert hat, ist sie im großen Maße von der Größe der Korpora abhängig. Koehn/Knight (2003) haben sich hauptsächlich auf die Komposita konzentriert, die eine eindeutige Entsprechung im Englischen haben. Obwohl die Methoden, die auf die Zerlegung der Komposita basieren sehr erfolgreich sind, brauchen sie noch weitere Forschungen und Verbesserungen, besonders wenn es um die Übersetzung eines in einer Phrase auftretenden Kompositums geht.

Das Problem des Wort-zur-Phrase-Übersetzens, das sehr oft bei Komposita auftritt, besteht auch bei den Eins-zu-Null-Entsprechungen, also Ausdrücke, die in der Zielsprache keinen direkten Äquivalent haben. Schon relativ früh erkannten die Forschern, dass es Wörter gibt, die nicht direkt in eine andere Sprache übersetzt werden können, sondern die Verwendung des komplexen Ausdrucks benötigen. Nagao/Tsujii (1986) haben den Einsatz formenorientierter Wörterbücher vorgeschlagen, in denen den Wörtern die lexikalischen Regeln der beiden Sprachen zugeordnet werden, damit das System auch mithilfe der bestimmten Transferregeln imstande sei, mehr grammatische Phrasen zu produzieren und damit effektiver mit den Unterschieden zwischen den Sprachen, die u. a. aus lexikalischen Lücken resultieren, umzugehen. Nach Santos (1990) führt diese Lösung zur Entstehung überfüllter Wörterbücher, die mit Redundanzen belastet sind. Er ist der Meinung, dass sich in den bilingualen Wörterbüchern nur Wörter und ihre Übersetzungen befinden sollen, die nach dem lexikalischen Transfer mithilfe von speziellen Regeln schon in der Transferphase der Grammatik der Zielsprache angepasst werden sollen. Eine ähnliche Methode wurde von Hai et al. (1997) angewendet. Sie haben sich stärker auf die letzte Phase konzentriert, also auf das Umkonstruieren der aus dem Wörterbuch erhaltenen Phrasen, so dass sie in der Zielsprache grammatisch korrekt sind. Mithilfe der syntaktischen Parser und zusätzlichen Umstrukturierungsregeln werden die in der Zielsprache stufenweise Phrasen hergestellt, die von den einfachsten bis zu komplexeren reichen. Die Geburt der phrasenbasierten statistischen maschinellen Übersetzung, die auf der Analyse der Korpora von der schon produzierten Übersetzungen basiert, scheint zum Teil das Problem der Eins-zu-eins Entsprechungen zu lösen, denn solange sich ein bestimmtes Wort in dem Korpus befindet, ist es möglich, eine entsprechende Übersetzung dafür zuzuordnen. Dies weist darauf hin, dass es außer den semantischen Unterschieden zwischen den Sprachen auch strukturelle Unterschiede gibt, die im Bereich der maschinellen Übersetzung schon von Anfang an viele Probleme bereiten. Der Umfang der Frage ist sehr breit und kann aus verschiedenen von den gewählten Sprachpaaren abhängigen Standpunkten betrachtet werden. Im Folgenden werden nur wichtigste Entwicklungsrichtungen und Hauptmethoden berührt.

Die ersten maschinellen Übersetzungssysteme waren sehr einfach und basierten auf dem direkten Ersatz der Wörter, was zu Wort-für-Wort-Übersetzungen von niedrigerer Qualität führte. Um die Qualität dieser Übersetzungen zu verbessern, schlug Warren Weaver (1949) Forschungen und die Verwendung der universellen Sprachen für die Zwecke

der maschinellen Übersetzung vor. Reifler (1950, zit. nach Hutchins 1986:38) dagegen hat die Idee der Post- und Prä-Edition eingeführt. Die Rolle der Posteditoren war u. a. die Wortreihenfolge der maschinell erhaltenen Sätze an die Zielsprache anzupassen. Obwohl dieses Konzept des maschinellen Eingriffs das Problem der Syntaxunterschiede nicht völlig löste, ist es bedeutsam und wiederholt sich in der einen oder anderen Form durch die ganze Entwicklung der Arbeiten an der maschinellen Übersetzung. Bar-Hillel (1951) sah die Lösung des Problems der syntaktischen Unterschiede in der Implementierung der morphologischen und syntaktischen Analyse des Satzes und der automatischen Umstellung der Satzteile, so dass sie mit der Standardreihenfolge der Zielsprache übereinstimmen.

Die Qualität der Übersetzungen versuchte man auch mit dem Einbau von syntaktischen Regeln zu erreichen, was die Georgetown-IBM Gruppe in ihrem 1954 der Öffentlichkeit präsentierten System verwirklichte (Hutchins 1986). Obwohl die Einschränkungen des Systems riesig waren, war es das erste direkte System, das über die Wort-für-Wort-Übersetzung hinausging. Da die direkten Systeme nicht befriedigende Ergebnisse liefern, wurden anspruchsvolle linguistische Modelle für die maschinelle Übersetzung entwickelt. Man hoffte, dass die Verwendung der indirekten Interligua-Methode, die auf dem Übersetzen via eine universale und abstrakte Repräsentation der Bedeutung basierte, zu einer weniger wörtlichen Übersetzungen führt. Nach Hutchins (1986) war die Qualität der diese Methode benutzenden Systeme schwankend. Somers (1998) betont aber, dass es sich in der Praxis als zu kompliziert erwies, Interlingua und alle mit ihr verbundenen Regeln zu bilden. Überdies waren tiefe Repräsentationen immer noch Repräsentationen der Texte, nicht der Bedeutung selbst. Es schiene also, dass Systeme auf einem Mechanismus basieren müssten, das imstande wäre, erfolgreicher linguistische Strukturen einer Sprache in eine andere zu verwandeln. Aus diesem Grund wurde die Transfermethode[6] vorgeschlagen, die auf einer komplexen strukturellen Analyse des Textes und dem Transfer der Strukturen basiert. Eines der ersten Systemen mit dem man automatische Übersetzungen durch eine sprachliche Analyse durchzuführen versuchte, war TAUM. Die späteren Systeme, wie das GETA-System, das auf zwei Transfern, dem lexikalischen und dem strukturellen, basierte oder das METAL-System, das zu einem kommerziellen System wurde, waren laut Hutchins (1986) erfolgreicher. Darüber hinaus haben Kitteredge und Lehrberger (1982) bemerkt,

[6] Die statistische maschinelle Übersetzung ist die populärste und beruht auf bilingualen parallelen Korpora, die Rohdaten in der Form von Texten und deren Übersetzungen beinhalten. Das Hauptziel der statistischen maschinellen Übersetzung ist, den Text durch eine Analyse von Millionen von Wörtern in zweisprachigen Korpora zu übersetzen, und zwar durch die Berechnung der wahrscheinlichsten Übersetzung in Bezug auf lexikalische Entsprechungen und Wortreihenfolge. Unter den statistischen Systemen unterscheidet man vor allem wörterbasierte Systeme, die nur einzelne Wörter in Betracht ziehen und phrasenbasierte Systeme, die Sätze in Phrasen zerlegen und diese für das Kalkulieren der wahrscheinlichsten Übersetzungen benutzen.

dass die Begrenzung der Anwendungsdomäne sehr günstig die Übersetzungsqualität beeinflussen kann, weil die domänenspezifischen Texte durch bestimmte Normen regiert werden, die auch eine bestimmte Grammatik und einen bestimmten Stil aufzwingen, was aus präziser und einfacherer Beschreibung der Regeln resultiert. Ein gutes Beispiel dafür ist das METEO System, das seit 1977 erfolgreich zum französisch-englischen Übersetzen der Wettervorhersagen benutzt wird und als einer der größten Erfolge der maschinellen Übersetzung gilt (Chan 2015). Es hat sich aber erwiesen, dass je größer die strukturellen Unterschiede zwischen den Sprachen sind, desto problematischer für die transferbasierten Systeme ist, die Strukturen entsprechend umzuwandeln und grammatisch kohärente Übersetzungen zu produzieren. 1984 versuchte Nagao (1984) dieses Problem zu lösen und entwickelte eine beispielbasierte Methode, die auf dem Prinzip der Analogie basiert, was bedeutet, dass die Übersetzung anhand von ausgewählten schon übersetzten Beispielen generiert wird.

Viele Forschungen haben die Effizienz der Methode und ihre Modifikation bestätigt und auch Somers (1998) stellt fest, dass diese Methode zur einer weniger wörtlichen Übersetzung und besserer Lesbarkeit führt. Auch den Entwicklern der wörterbasierten statistischen maschinellen Übersetzung, die nicht auf einer tiefgründigen linguistischen Analyse, sondern auf schon übersetzten Texten beruhte, war bewusst, dass das System mit den strukturellen Unterschieden zwischen den Sprachen erfolgreicher umgehen muss, um Übersetzungen von besseren Qualität zu liefern. Aus diesem Grund haben Brown et al. (1993) als die ersten eine Serie von Ausrichtungsmodellen (IBM alignment models) für das wörterbasiertes statistisches System entwickelt, die um einen grammatisch korrekten Satz zu erhalten für die Auswahl der Wörter aus bilingualen Korpora und deren Umstellung im Satz der Zielsprache, verantwortlich sind. Ein noch effiziertes System als das wörterbasierte, nämlich das phrasenbasierte statistische System haben Koehn et al. (2003) eingeführt, das auch ein Modell für Umstellung der Phrasen beinhaltete. Den angeführten Modellen wird aber vorgeworfen, dass sieWissen verfügen und nur nicht über bei ähnlichen Sprachpaaren erfolgreich sein können. Um mit dem Problem der linguistischen Unterschiede in der statistischen maschinellen Übersetzung zurecht zu kommen, haben Yamada/Knight (2001) eine syntaxbasierte Methode vorgeschlagen, in der mithilfe bestimmter Regeln und Operationen - wie die Umstellung und die Einführung von Wörtern - die Syntaxbäume in der Muttersprache in eine Folge von Wörtern in die andere Sprache transformiert werden. Diese Methode benutzt syntaktisches Wissen und hat sich als erfolgreicher als die früher eingeführten Methoden erwiesen, besonders wenn es um Sprachen geht, die sich stark in Bezug auf die Wortstellung voneinander unterscheiden. Diese Methode wurde später erfolgreich von Galley et al. (2004) erweitert, indem größere Baumfragmente analysiert und komplexere Übersetzungsregeln angewendet wurden. Die phrasenbasierte Methode wurde dagegen von Chiang (2005) verbessert. Seine hierarchische phrasenbasierte Methode stützt sich ebenfalls auf syntaktisches Wissen, das aber aus den Phrasen in der Ausgangs- und Zielsprache gewonnen wurde, was zu besseren Übersetzungen führte. Hayashi

et al. (2010) haben vorgeschlagen, dass diese neue und versprechende Methode mit einem wörterbasierten Umstellungsmodell bereichert werden sollte, um eine bessere Transformation und darüber hinaus eine effiziente Kalkulation der Korrektheit der erhaltenen Wortstellung erreichen zu können.

Das Problem liegt aber darin, dass die derzeitigen phrasenbasierten Systeme die Sätze in Phrasen und Wörtern brechen, die weitgehend unabhängig voneinander übersetzt werden, was zu den schon angeführten Problemen, wie dem Mangel an grammatische Kohärenz und korrekte Wortstellung oder den Schwierigkeiten mit der Lösung verschiedener Mehrdeutigkeitsarten führt und verursacht, dass die Qualität der Übersetzung mit der Länge der Sätze senkt. Viel Hoffnung auf die Verbesserung der generellen Performanz der maschinellen Übersetzung und die Überwindung der Grenzen der früheren Methoden wird in den neusten den auf neuronalen Netzwerken basierenden Ansatz zur maschinellen Übersetzung gesetzt. Wie Sutskever et al. (2014) bemerken, sind tiefe neuronale Netzwerke extrem leistungsstarke maschinelle Lernmodelle, die bei den komplexen und multidimensionalen Problemen wie dem Erkennen der natürlichen Sprachen, sehr erfolgreich sein können.

Es gab viele Versuche das Potential der neuronalen Netzwerken in der maschinellen Übersetzung zu benutzen. Sie wurden schon verwendet, um die Leistung der statistischen phrasenbasierten maschinellen Übersetzung zu verbessern. Wie Microsoft-Forscher Microsoft Translator (2016) anführen, liegt die Stärke der neuronalen Netzwerke darin, dass sie die einzigartigen Eigenschaften der Wörter in einem bestimmten Sprachpaar repräsentieren können. Darunter gibt es einfache Konzepte, wie Genus, Wortart oder Höflichkeitsformel oder auch andere weniger offensichtliche Merkmale, die aus den Trainingsdaten abgeleitet werden, zu berücksichtigen. Es scheint wichtig zu sein, dass die ganzen Sätze in der neuronalen maschinellen Übersetzung analysiert werden und die Bedeutung der Wörter im Kontext der ganzen Sätze in bestimmt ist. Wu et al. (2016) haben ein Google Neural Machine Translation System geschaffen, das auf neuronalen Netzwerken basiert und vielversprechende Ergebnisse liefert. Laut den Forschern wurden Übersetzungsfehler beim bestimmten Sprachpaaren um mehr als 60% reduziert. Sie weisen darauf hin, dass ihr System immer noch zahlreiche Fehler zu verzeichnen hat, die ein menschlicher Übersetzer niemals begehen würde. Dazu gehören das Löschen der Wörter, das Übersetzen der Sätze ohne umfangreichere Kontextberücksichtigung sowie Fehler aufgrund von Schwierigkeiten mit Eigennamen und seltenen Begriffen. Die neuronale Methode wurde bisher nur in wenigen öffentlichen Systemen verwendet aber die Ergebnisse sind schon befriedigender als diejenigen, die die früheren Systeme lieferten. Die Methode wird gegenwärtig intensiv erforscht und entwickelt, um die Probleme in der maschinellen Übersetzung besser bewältigen zu können. Die neuronale maschinelle Übersetzung ist zwar nicht mangelfrei aber Forschungsarbeiten in diesem, viel versprechenden Bereich können dazu beitragen, dass sich die Qualität der maschinell hergestellten Übersetzungen, nach denen eine riesige Nachfrage besteht, erheblich verbessert.

Literaturverzeichnis

Arnold, Doug. „Why translation is difficuld for computers". *Computers and Translation: A translator's guide*. Hrsg. Harold Somers. Amsterdam: John Benjamins, 2003, 119–142. Print.

Bar-Hillel, Yehoshua. „The present state of research on mechanical translation". *American Documentation*, Vol. 2 (4) (1951): 229–237. Print.

Bar-Hillel, Yehoshua. *The Treatment of "idioms by a Translating Machine, presented at the Conference on Mechanical Translation at Massachusett Institute of Technology*. 1952. http://mt-archive.info/MIT- 1952-Bar-Hillel-2.pdf. 15.2.2017.

Bar-Hillel, Yehoshua. „The Present Status of Automatic Translation of Languages". *Advances Advances in Computers* 1 (1960): 91–163. Print.

Brown, Peter, Pietra, Stephen A. Della, Pietra, Vincent J. Della und Robert L. Marcer. „The mathematics of machine translation: Parameter estimation". *Computational Linguistics*, 19(2), 1993, 263–311. Print.

Carbonell, Jaime G., Mitamura, Teruko und Eric H. Nyberg. „The KANT Perspective: A Critique of Pure Transfer (and Pure Interlingua, Pure Statistics, ...)". *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*. Montreal, 1992, 225–235. Print.

Chan, Sin-wai. *Routledge Encyclopedia of Translation Technology*. New York: Routledge, 2015. Print.

Chiang, David. „A hierarchical phrase-based model for statistical machine translation". *Proceedings of the 43rd Annual Meeting of the ACL*. New Brunswick: University of Michigan, 2005, 263–270.

Hai, Le Manh, Kawtrakul, Asanee und Yuen Poovorawan. „Phrasal transfer model for Vietnamese-English machine translation". *Natural Language Processing Pacific Rim Symposium, NLPRS 97*. 1997. http://www.cs.jhu.edu/nguyen/share/phrasal-transfer-model-for.pdf. 21.4.2017.

Hardmeier, Christian und Marcello Federico. „Modelling Pronominal Anaphora in Statistical Machine Translation". *Proceedings of the 7th International Workshop on Spoken Language Translation Paris*. Paris. 2010, 283–289. Print.

Hayashi Katsuhiko, Tsukada, Hajime, Sudoh, Katsuhito, Duh, Kevin und Seiichi Yamamoto. „Hierarchical Phrase-based Machine Translation with Wort-based Reordering Model". *Proceedings of the 23rd International Conference on Computational Linguistics* (Coling 2010). Beijing, 2010, 439–446. Print.

Hutchins, William John. *Machine translation: past, present, future*. Chichester: Ellis Horwood Limited, 1986. Print.

Hutchins, William John und Harold L. Somers. *An Introduction to machine translation*. London: Academic Press, 1992. Print.

Kitteredge, Richard und John Lehrberger. *Sublanguage: studies of language in restricted semantic domains*. Berlin: De Gruyter, 1982. Print.

Koehn, Philipp und Kevin Knight. „Empirical methods for compound splitting". *EACL'03 Proceedings of the tenth conference on European chapter of the Assotiation for Computational Linguistics – Volume 1*. 2003, 187–193. Print.

Koehn, Philipp. *Statistical Machine Translation*. New York, USA: Cambridge University Press, 2010. Print.

Landsbergen, Jan. „Compositional translation revisted". *Proceedings of the 10th European Summer School on Logic, Linguistics and Information*. Saarbrücken, 1998, 1–3. http://www.mt-archive. info/ESSLLI-1998-Landsbergen.pdf. 22.3.2017.

Nagao, Makoto. „A Framework of a Mechanical Translation between Japanese and Englisch by Analogy Principle". *Artificial and Human Intelligence*. Hrsg. Alick Elithorn und Ranan Banerji. Amsterdam: Elsevier Science Pub. Co, 1984: 173–180. Print.

Nagao, Makoto und Junichi Tsujii. „The transfer phase of the Mu machine translation system". *Proceedings of the 11th Conference on Computational linguistics*. 1986, 97–103. Print.

Niessen, Sonja und Hermann Ney. „Improving SMT quality with morpho-syntactic analysis". *Proceedings of the 18th International Conference on Computationnal Linguistics*. 2000, 1081–1085. Print.

Okuma, Hideo, Yamamoto, Hirofumi und Eiichiro Sumita. „Introducing Translation Dictionarz Into Phrase-based SMT". *IEICE-Transactions on Information and Systems* 7, 2008, 2051–2057.

Reifler Erwin. *Studies in mechanical translation*, no. 1, MT. Mimeo, Jan 10, 1050, Seattle: Univ. Washington, 1950. Print.

Ren, Zhixiang, Lü, Yajuan, Cao, Jie und Qun Liu. „Improving Statistical Machine Translation Using. Domain Bilingual Multiword Expressins". *Proceedings of the 2009 Workshop on Multiword Expressions, ACL-IJCNLP*. Singapore, 2009, 47–54. Print.

Salton, Giancarlo D., Ross, Robert J. und John D. Kellehr. „An Empirical Study of the Impact of Idioms on Phrase Based Statistical Machine Translation of English to Brazilian-Portuguese". *Proceedings of the 3rd Workshop on Hybrid Approaches to Translation* (HyTra 2014). Gothenburg: Association for Computational Linguistics, 2014, 36–41. Print.

Santos, Diana. „Lexical gaps and idioms in machine translation". *Proceedings of the 13th Conference on Computational linguistics*. 1990, 330–335. Print.

Somers, Harold L. „Machine Translation: Methodology". *Routledge Encyclopedia of Translation Studies*. Hrsg. Mona Baker. London, New York: Routledge. 1998, 143–149. Print.

Sutskever, Ilya, Vinyals, Oriol und Quoc V. Le. „Sequenze to sequenze learning whit neural networks". *NIPS'14, Proceedings of the 27th International Conference on Neutral Information Processing Systems*. 2014, 3104–3112. Print.

Volk, Martin. „The automatic translation of idioms. Machine translation vs. translation memory systems". *Machine translation: theory, applications, and evaluation. An assessment of the state of the art*. Hrsg. Nico Weber. St. Augustin: Michael Itschert, 1998, 167–192. Print.

Weaver, Warren. „Translation". 1949. Nachgedruckt in: *Machine transation of languages: fourteen essays*. Hrsg. William N Locke und Andrew Donald Booth. Cambridge: M.I.T. Press, 1955, 15–23. Print.

Wilks, Yorick. „Machine translation and artificial intelligence". *Translating and the Computer*. Hrsg. Barbara M. Snell. Amsterdam: North-Holland Publishing Company, 1979, 27–43. Print.

Wilks, Yorick. *Machine Translation: Its Scope and Limits*. New York: Springer, 2009. Print.

Wu Yonghui, Schuster, Mike, Chen, Zhifeng et al. „Google's Neural Machine Translation System". *CoRR abs/1609.08144*. 2016. https://arxiv.org/pdf/1609.08144.pdf. 6.5.2017.

Yamada, Kenji und Kevin Knight. „A syntax-baset statistical translation model". *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, 2001, 523–530. Print.

ZITIERNACHWEIS:

Łopuszańska, Grażyna. „Maschinelle Übersetzung – Grenzen und Möglichkeiten." *Linguistische Treffen in Wrocław* 15, 2019 (I): 145–156. DOI: 10.23817/lingtreff.15-12.