

WAS IST COMPUTERLINGUISTIK?

Einleitung. Wenige Dinge haben einen so starken Einfluss auf unseren Alltag wie Sprache und Kommunikation. Sprache ist das wichtigste Mittel zur Kommunikation und Interaktion zwischen Menschen. Obwohl Sprache äußerst komplex ist und einige sprachliche Phänomene äußerst schwer zu erklären sind, wird sie in einem äußerst kurzen Zeitraum erworben und ist eng verknüpft mit dem Denken. Auf diese Weise ist die menschliche Sprache zu einer der erstaunlichsten Erfindungen innerhalb der kulturellen Entwicklung geworden. Dafür steht auch der außerordentliche Reichtum an Sprachen, Dialekten, und Sondersprachen (bspw. Zeichensprachen), die in den vergangenen Jahrtausenden entstanden sind.

Im Rückblick auf die kulturelle Entwicklung des 20. Jahrhunderts kann der Computer als eine der entscheidenden Erfindungen angesehen werden, die einen nachhaltigen und in letzter Konsequenz noch nicht absehbaren Einfluss auf unser Berufs- und Freizeitleben ausgeübt hat. Wurden Computer anfänglich noch als Rechenmaschinen angesehen, richtete sich schon nach kurzer Zeit das Interesse auf Daten, die über den reinen Zahlenbereich hinausgingen: die menschliche Sprache in geschriebener und gesprochener Form. Und da Sprache ein äußerst effizientes Mittel zur Kommunikation ist, wird sie auch als entscheidend für die Weiterentwicklung von Computern und modernen Kommunikationsmitteln angesehen: einerseits für die Übertragung einer Sprache in eine andere Sprache, Form oder ein anderes Medium, andererseits als Interaktionsmittel zwischen Menschen und Maschinen.

Hauptteil. Was ist Computerlinguistik? In der Computerlinguistik (CL) oder linguistischen Datenverarbeitung wird untersucht, wie natürliche Sprache in Form von Text — oder Sprachdaten mit Hilfe des Computers algorithmisch verarbeitet werden kann. Computerlinguistik ist eines der Hauptanwendungsgebiete an der Schnittstelle zwischen Mensch und Computer. Das Fach Computerlinguistik ist eine eher junge Disziplin, die sehr stark mit den Entwicklungen in der Informationstechnologie verbunden ist: flexible Suchmaschinen, automatisierte Dialogsysteme, durch Sprache bedienbare Büromaschinen und Werkzeuge, e-learning Unterrichtsmaterial... Die Entwicklung und Verbesserung dieser Anwendung hängt in hohem Grad ab von der gelungenen Integration sprachlicher bzw. linguistischer Datenstrukturen. Und genau darin liegt das Hauptinteresse der Computerlinguistik, nämlich linguistische Datenstrukturen so zu modellieren, dass menschliche Sprache(n) von Computern verarbeitet und verstanden werden [1].

Computerlinguistik lässt sich als Terminus (oder dessen Umschreibung) in die sechziger Jahre des 20. Jahrhunderts zurückverfolgen. Als Aufgabenstellung war sie mit den Anfängen der künstlichen Intelligenz eigentlich schon nahegelegt und Chomskys Syntactic Structures von 1957 präsentierte die Sprache in einem entsprechend passenden neuen formalen Gerüst. Die teilweise aus dem US-Verteidigungsbudget sehr hoch geförderten Forschungen brachten jedoch nicht die erhofften Durchbrüche. Besonders die Koryphäen Chomsky und Weizenbaum beschwichtigten die Erwartungen an Automatisierungen von Sprachübersetzung. Der Wende von behavioristischen Wissenschaftskonzeptionen zu mentalistischen (Chomsky) folgte umfassende Konzipierungen in den Kognitionswissenschaften.

In den siebziger Jahren erschienen zunehmend häufiger Publikationen mit dem Begriff Computerlinguistik im Titel. Es gab bereits finanziell aufwändige Versuche exegetischer Anwendungen (Konkordanzen, Wort- und Formstatistik), aber auch schon größere Projekte zur maschinellen Sprachanalyse und zu Übersetzungen. Die ersten Computerlinguistik-Studiengänge in Deutschland wurden an der Universität des Saarlandes und in Stuttgart eingerichtet. Die Computerlinguistik bekam mit der Verbreitung von persönlich gebrauchten Computern (PC) und mit dem Aufkommen des Internets neue Anwendungsgebiete. Im Gegensatz zu einer Internetlinguistik, die insbesondere menschliches Sprachverhalten und die darüber induzierten Sprachbildungen im und mittels Internet untersucht, entstand in der Computerlinguistik eine stärkere informatisch-praktische Ausrichtung. Doch gab das Fach die klassischen philosophisch-linguistischen Fragen nicht ganz auf und wird heute in theoretische und praktische Computerlinguistik unterschieden.

Die Computerlinguistik beschäftigt sich mit der Verarbeitung natürlicher Sprache in Computersystemen. Zu ihren Schwerpunkten gehören sowohl anwendungsbezogene wie grundlagenforschungsorientierte Fragestellungen, darunter:

- die Schaffung von (Software-)Werkzeugen, die die Mensch-Maschine-Kommunikation (MMK) durch die Verarbeitung sprachlicher Daten erleichtern (z. B. natürlich sprachliche Dialogsysteme zur Kommunikation mit Datenbanken und anderen Informationssystemen, automatische Beantwortung von Emails, sprachgesteuerte Computer und andere Systeme, Sprachausgabe);
- maschinelle, vor allem sprach- und texttechnologische Hilfsmittel zur Verarbeitung von Texten in natürlicher Sprache zu verschiedensten Zwecken (z. B. Diktierprogramme, Rechtschreibungs-

Korrekturprogramme, Vorlesemaschinen, Übersetzungssysteme, Übersetzungshilfen, Inhaltsanalyse, Dokumenten-Management, Textgenerierung usw.), ein Bereich der seit Jahren in der „Informations- und Kommunikations-Gesellschaft“ immer größere Bedeutung erlangt;

- die Untersuchung sehr großer Mengen von Sprachdaten (z. B. Textkorpora) vor allem mit Hilfe statistischer und anderer quantitativer Methoden, um möglichst fehlerfrei und komfortabel komplexe Beschreibungen von Sprachen anfertigen zu können (Wörterbücher, Grammatiken), bzw. das Anlegen und Pflegen großer Sammlungen textueller Daten;

- die Ausnutzung von Eigenschaften der natürlichen Sprache mittels Computer für andere Zwecke, z. B.: Erkennung der Sprache, in der ein Text geschrieben wird/wurde, Autorenbestimmung für philologische, forensische und andere Zwecke;

- die Erforschung der Zusammenhänge von Eigenschaften sprachlicher Strukturen und Prozesse bis hin zur Entdeckung universeller Sprach- und Textgesetze. Die computerlinguistischen Verfahren dienen hier vor allem der Überprüfung entsprechender Hypothesen anhand von großen Mengen authentischen Sprachmaterials.

Worin besteht die größte Herausforderung der Computerlinguistik? Menschen benutzen zur Kommunikation sog. „natürliche Sprachen“, ein äußerst komplexes System aus Lauten und Zeichen, das einem ständigen Wandel unterliegt und ausreichend robust ist, um in äußerst vielen, störanfälligen Situationen gebraucht zu werden. Computer hingegen sind auf sog. „formale Sprachen“ beschränkt, die anfällig sind gegenüber Mehrdeutigkeit, Veränderungen und Spontanität. Die wichtigste Aufgabe der Computerlinguistik besteht demnach darin, adäquate formale Modelle bereitzustellen, die diese besonderen und gegensätzlichen Eigenschaften natürlicher Sprachen innerhalb der formalen Einschränkungen von Rechnersprachen erfassen [2].

Auf den ersten Blick erscheint Computerlinguistik häufig als Verbindung der Disziplinen Linguistik und Informatik; jedoch ist das Spektrum der Wissenschaften, die für die Computerlinguistik von Belang sind, weitaus vielfältiger. Es bestehen enge Verbindungen zur Kognitiven Psychologie, zur Mathematik, zu den Einzelphilologien und letztlich zur Philosophie. Der Stoff mit dem sich Studierende der Computerlinguistik auseinandersetzen können, reicht durchaus von den frühesten Überlegungen zur menschlichen Sprache (bspw. den ersten Sanskrit-Grammatiken) bis hin zu den neuesten Entwicklungen der Informationstechnologie.

Computerlinguistik stellt eine Verbindung zwischen Sprachwissenschaften und Informatik her und wird seit den 80ern an Hochschulen gelehrt. Es geht darum natürliche Sprache (also Sprache, die von uns Menschen geäußert wird) in Algorithmen zu verarbeiten, die ein Computer verstehen kann. Im Studium analysiert man Sprache, lernt Programmiersprachen und statistische Grundlagen.

Zu den Aufgabenbereichen eines/ r Computerlinguisten/in gehören: gesprochenen Sprache in Text zu verarbeiten wie bei der Spracheingabe eines Smartphones oder andersrum Text in gesprochener Sprache zu verarbeiten wie bei der Vorlesefunktion eines Smartphones.

Die Computerlinguistik ist ein Forschungsgebiet, das Kenntnisse und Methoden aus sehr unterschiedenen Forschungsdisziplinen benötigt:

- aus der Allgemeinen und der Theoretischen Sprachwissenschaft die Begriffe zur Beschreibung von Sprachen auf den Ebenen der Wortbildung (Morphologie), des Satz- und Textaufbaus (Syntax), der Bedeutung (Semantik) und der Verwendungsweise (Pragmatik);

- aus der Informatik verschiedene Programmiersprachen und Methoden ihrer Übersetzung in maschinennahe Sprachen, Datenstrukturen, effiziente Speicher- und Suchverfahren;

- aus der Mathematischen Logik — zur Modellierung von Bedeutungen- verschiedene logische Formelsprachen und deren Interpretationen sowie Beweisverfahren;

- aus der Mathematik und Phonetik statistische Verfahren, nicht nur für die akustische Sprachverarbeitung;

- aus der Kommunikationswissenschaft und Kognitionspsychologie Erkenntnisse über Dialogstrategien und Aufmerksamkeitssteuerung.

Bisher liegt der Schwerpunkt der Forschung auf der Verarbeitung geschriebener Sprache, aber erste akustische Spracherkennung werden schon industriell entwickelt und erprobt. Wegen der zunehmenden Bedeutung der elektronischen Informationsverarbeitung werden in einigen Softwarefirmen große Anstrengungen unternommen, auf den genannten Gebieten anwendungsreife Produkte, in denen sprachliches Wissen verfügbar gemacht wird, anzubieten.

Fazit. Vereinfacht gesagt, ist Computerlinguistik das wissenschaftliche Studium der Sprache aus der Perspektive des Computers. Die Computerlinguistik ist von daher darin interessiert, unterschiedlichste sprachliche Phänomene so zu modellieren, dass sie durch Computer verarbeitet werden können. Diese Modelle können wissensbasiert und von Hand eingegeben oder datenorientiert, d. h. statistischer bzw. empirischer Natur sein. Arbeiten in der Computerlinguistik sind in einigen Fällen aus einer wissenschaftlichen Perspektive motiviert, derart, dass versucht wird, eine formal-rechnerische Erklärung für ein besonderes linguistisches oder psycholinguistisches Phänomen zu geben. In anderen Fällen ist die Motivation rein technologischer Natur, in dem Komponente für ein sprachbasiertes System erstellt wird. In der Tat ist die Arbeit von Computerlinguisten in vielen erprobten Systemen zu finden, Modulen zur Sprachsynthese, automatisierten Dialogsystemen, Suchmaschinen, Text-Editoren, Sprachlehrmaterialien usw.

1. Computerlinguistik und Sprachtechnologie / Kai-Uwe Carstensen [et Al.]. — Heidelberg : Spektrum Akademischer Verlag, 3. Auflage, 2010. — S. 18.
2. Henning Lobin. Computerlinguistik und Texttechnologie. Paderborn / Lobin Henning. — München : Fink, 2010. — S. 3.

UDK: 811.112

G. A. Ignatowitsch

Ausbildungsinstitution «Grodnoer Staatliche Inka-Kupala-Universität», Grodno

DIE ÜBERSETZUNGSÄQUIVALENZ DER EUPHEMISMEN IN DEN WERKEN DER RUSSISCHEN KLASSISCHEN LITERATUR

Einleitung. Die Erscheinung der Euphemisation stellt ein wissenschaftliches Interesse dar, weil die Euphemismen in verschiedenen Sphären der Sprachkommunikation, besonders in der schönen Literatur, gebraucht werden. In den Werken vieler Schriftsteller der Vergangenheit und der Gegenwart wird das Verfahren der Euphemisation als eine der stilistischen Quellen des Russischen gebraucht. Man kann die schöne Literatur als sichere Quelle für die Charakteristik der Euphemisation in der Sprache betrachten. Als Erforschungsstoff wurden die Werke der russischen Klassiker des 19. Jahrhunderts verwendet (N. Gogol, F. Dostojewski, I. Gontscharow, L. Tolstoi, A. Puschkin, I. Turgenew, A. Tschechow).

Hauptteil. Infolge der praktischen Forschung von der Erscheinung der Euphemisation in den Werken der russischen klassischen Literatur wurden die Euphemismen aus solchen Funktionssphären wie übernatürliche Mächte, Krankheiten und Tod, körperliche und geistige Defekte, die Euphemismen der Kategorie der Unbestimmtheit gesammelt. Aus dem nationalen Korpus des Russischen wurden die Beispielabschnitte aus den Werken der russischen klassischen Literatur in den Deutschübersetzungen eingeführt.

So, zum Beispiel, sieht man in den Werken von N. Gogol „Die toten Seelen“ («Мёртвые души») und „Abende auf dem Weiler bei Dikanka“ («Вечера на хуторе близ Диканьки») bei der Übersetzung, dass die Euphemismen wie «лукавый», «чёрт», «нечистый», «бес» durch das Wort „der Böse“ ersetzt werden. Das weist auf die Tatsache hin, dass in der russischen und der deutschen Kultur der Teufel als ein Geist bezeichnet wird, der die Menschen zum Begehen der Sünde, böser und schlechter Taten aufwiegelt:

‘Лукавый, что ли, миром ворочает, ей Богу!’ *Als ob der Böse die Welt regiert, bei Gott!* [1].

‘Может быть, это и правда, что ты ничего не скажешь худого, — подумала про себя красавица, — только мне чудно... верно, это лукавый!’ *Vielleicht ist es auch wahr, dass du mir nichts Schlechtes sagen wirst, — dachte sich die Schöne, — aber es ist mir so wunderbar zumute... das macht gewiss der Böse!* [1].

Die gleiche Erscheinung kann man in den Werken von A. Tschechow „Der Zweikampf“ («Дуэль») I. Turgenew „Väter und Söhne“ («Отцы и дети») beobachten: ‘— Нечистый попутал: иди да иди... Вот и пошел, и чуть в кукурузе не помер от страха’. *Der Böse hat mich verführt, und ich bin ihm gefolgt und wäre vor Schrecken gestorben dort im Maisfeld* [1].

Das Tabu für die Namen der übernatürlichen Mächte gilt als das altertümlichste Tabu. Im Werk von F. Dostojewski „Die Brüder Karamasow“ («Братья Карамазовы») erfüllt die Wortgruppe «ewiger Richter» die Funktion des Euphemismus, die die Benennung *der Gott* «Бог» ersetzt, und wird als substantiviertes Adjektiv „der Ewige“ ins Deutsche übersetzt: ‘Прости ему вечный судья за сорокалетние прелести, как и я прощаю, но ведь это... ведь это знаете кто?’ *Möge ihm der Ewige Richter die vierzigjährigen Reize verzeihen, so wie ich sie ihm verzeihe! Aber wissen Sie, wer es war, wer das geschrieben hat?* [1]. In vielen Quellen wird das Wort *der Gott* als das ewige Wesen bezeichnet, als das Wesen, das außerhalb der Zeit existiert. Das Wort „der Ewige“ bedeutet derjenige, der ständig existiert, derjenige, der weder einen Anfang noch ein Ende hat.

Als Beispiel der Verwendung der Euphemismen im Bereich von Krankheiten und Tod kann man den Abschnitt aus dem bekannten Roman von L. Tolstoi „Krieg und Frieden“ («Война и мир») anführen: ‘Ах, мой друг, религия, и только одна религия, может нас, уже не говорю утешить, но избавить от отчаяния; одна религия может объяснить нам то, чего без ее помощи не может понять человек: для чего, зачем существа добрые, возвышенные, умеющие находить счастье в жизни, никому не только не вредящие, но необходимые для счастья других, — призываются к Богу, а остаются жить злые, бесполезные, вредные, или такие, которые в тягость себе и другим.’ *Dennoch, meine liebe Freundin, die Religion und nur die Religion allein kann uns, wenn ich auch nichtsagen möchte trösten, so doch vor der Verzweiflung retten; nur die Religion allein kann uns das erklären, was der Mensch allein und ohne ihre Hilfe niemals begreifen würde: warum gute, hoch herzige Menschen, die ihr Glück auf Erden finden würden und nicht nur niemandem schaden, sondern auch für das Glück anderer unentbehrlich sind, warum solche Menschen von Gott abberufen werden und solche, die böse, unnützlich und schädlich sind und anderen und sich selber nur zur Last fallen, am Leben*