

Erschienen in Gerd Willée, Bernhard Schröder, Hans-Christian Schmitz (Eds.): *Computerlinguistik - Was geht, was kommt?* Festschrift für Winfried Lenders. Gardez! Verlag Sankt Augustin 2002, S. 138-142

Computerlinguistik: Was geht und was kommt, entscheidet die Lehre.

Peter Hellwig, Heidelberg

Die Aussichten für die Computerlinguistik erscheinen rosig. Wörter, Sätze, Texte machen den Großteil der gespeicherten Daten auf Computern aus. Sprache ist das hauptsächliche Material, aus dem das Internet besteht. Vielleicht wird die gesprochene Sprache das Medium für die sehr kleinen "hand-held"-Computer, die zu Zeit entwickelt werden. Es kann sein, daß eines Tages Computertechnologie ohne Computerlinguistik gar nicht mehr denkbar ist, wie überhaupt die Informatik-pur mehr und mehr ausgedient zu haben scheint. Standen bei der klassischen Informatik die Eigenschaften einer Rechenmaschine im Mittelpunkt, so geht es heute zunehmend um Computermodelle von Wirklichkeiten, d.h. um die Simulation biologischer, medizinischer, physikalischer, geographischer, ökonomischer und kognitiver Phänomene. Diese Simulation kann nicht ohne die jeweiligen Sachfachern vorangetrieben werden. Aufregende Perspektiven für die Computerlinguistik ergeben sich aus der menschlichen Fähigkeit, alle möglichen analogen Modelle der Wirklichkeit (z.B. eine wahrgenommene Szene) in propositionale Modelle (also sprachliche Beschreibungen) zu überführen und umgekehrt. Wenn sich die Computerlinguistik verstärkt der Verbalisierung von Modellen, die in Computersimulationen verwendet werden, widmen würde, könnte sie mit vielen Anwendungsfeldern der Informatik neue interdisziplinäre Ehen eingehen.

So rosig die Aussichten, so ernüchternd ist die Realität. Mit einigen Ausnahmen sind die Ergebnisse der computerlinguistischen Forschung und Entwicklung bescheiden. Zahllose, mit Millionen geförderte Projekte haben nie zu einem Produkt geführt. Schuld ist zweifellos die Komplexität des Gegenstandes - aber nicht allein. In der Computerlinguistik wird das Know-how nicht genügend akkumuliert. Jedes Projekt beginnt wieder bei Null, wobei die in Angriff genommene Aufgabe häufig unterschätzt wird. Würde man z.B. in die Proceedings einer aktuellen COLING-Konferenz Vorträge aus einer zehn Jahre zurückliegenden Konferenz einstreuen, es würde nicht auffallen.

Man könnte meinen, die Computerlinguistik sei noch nicht reif als akademische Disziplin. Das ist aber nicht wahr. Was die Sprache als Gegenstand betrifft, so ist wahrscheinlich alles schon einmal beschrieben worden und schlummert in den Bibliotheken. Software und Methoden der Implementierung werden von der Informatik reichlich angeboten. Und schließlich gibt es gewiß schon viele spezielle Verfahren der Sprachverarbeitung, die Standards setzen könnten. Dies alles muß nur aufbereitet, systematisiert und gezielt an den Nachwuchs vermittelt werden. Damit die Forschung und Entwicklung in der Computerlinguistik nachhaltig wird, brauchen wir also eine Ausbildungsoffensive. Der Schwerpunkt an den Universitäten muß eine Zeitlang auf der Lehre liegen.

Die Studenten dafür sind da und sie sind hochmotiviert. Als in Heidelberg der Fakultätsrat im Sommer 2001 den neuen Bachelor/Masterstudiengang Computerlinguistik beschließen sollte,

gingen über 100 Studenten dafür auf die Straße - im Grunde eine Demo für eine Studienverschärfung!

Die Computerlinguistik sollte auf die neuen Bachelor/Master(B.A./M.A.)-Studiengänge setzen, denn die bisherige Organisation der Lehre hat gravierende Nachteile. Um das zu verstehen, muß man sich das einzigartige Profil der Computerlinguistik klar machen, welches aus einer Kombination von empirisch-philologischer, formal-theoretischer und ingenieurmäßiger Orientierung besteht. Empirisch-philologisch, weil sich letztlich jedes Wort irgendwie anders verhält und entsprechend klassifiziert werden muß. Formal-theoretisch, weil es sich bei der Sprachverarbeitung um Berechnungen auf der Grundlage adäquater Verallgemeinerungen handelt. Ingenieurmäßig, weil die Algorithmisierung der empirischen und theoretischen Einsichten eine besondere Aufgabe ist, die vor allem Erfindungsgabe verlangt. Empirie, Theorie und Ingenieurarbeit kennzeichnen also die Disziplin. Methodik, formale Schulung (Denken in Systemen) und Programmieren sind die wesentlichen Lernziele. Nur wenige haben die dafür notwendige, breite, sowohl geisteswissenschaftliche wie mathematische und praktische Begabung. Diejenigen, welche sie haben, ziehen andauernde Faszination aus dem Fach.

Bisher ist die Computerlinguistik vielerorts nur als Schwerpunkt in andere Studiengänge eingebunden. Hier und da ist sie Haupt- oder Nebenfach in einem Magisterstudiengang oder wird als Diplomstudiengang angeboten. Aufgrund ihres besonderen Profils ist die Computerlinguistik bei den Nachbarwissenschaften aber nicht gut aufgehoben. Soweit Sprachverarbeitung Gegenstand an Informatikfakultäten ist, fehlt der breite empirische und theoretische Hintergrund der Linguistik. Typisch ist, daß hier die Probleme bei weitem unterschätzt werden und immer wieder versucht wird, das Rad neu zu erfinden (oder, noch lieber, den "lernenden" Computer das Rad selbst erfinden zu lassen).

Häufiger ist die Computerlinguistik jedoch in Fakultäten für Sprach- und Literaturwissenschaften angesiedelt und ist als Magisterfach eines von mehreren. Geisteswissenschaften sind - wahrscheinlich zu recht - gekennzeichnet durch einen sehr geringen Organisationsgrad. Das Angebot an Lehrveranstaltungen entstammt dem aktuellen Forschungsinteresse der Lehrenden. Man lernt exemplarisch, geht seinen Interessen nach, vertieft, verknüpft, läßt sich anregen. Ideen kommen einem vielleicht in der Vorlesung, aber genauso gut irgendwann später. Man wird sehen. Mit dieser Haltung wird dann auch Computerlinguistik studiert. Manche sitzen im Programmierkurs wie in einer Rilke-Vorlesungen. Sie erwarten interessante Anregungen und daß sie später irgendwann einmal das Gehörte vertiefen und verbinden werden. Natürlich verlieren sie schon in der zweiten Stunde den Anschluß. Es fehlt ihnen die ingenieurmäßige, handwerksartige Ader, die für die Computerlinguistik unbedingt notwendig ist. Im Hauptstudium zeigt sich dann dasselbe Bild wie in manchen Drittmittelprojekten: An sich geht es um ein Übersetzungssystem, ein Programm zur Textzusammenfassung, ein Touristeninformationssystem oder was auch immer. Aber jetzt wird erst einmal ein Morphologieformalismus, ein Parser, eine lexikalische Datenbank usw. entworfen, statt daß jeder seit dem Grundstudium quasi eine Tasche voller Lösungen für diese Details mit sich führt. Das Vertrauen des Geisteswissenschaftlers, wenn es darauf ankommt, jedwede Aufgabe mit seinem gesunden Menschenverstand lösen zu können, ist der Grund dafür, daß das Wissen in der Computerlinguistik zu wenig aufgeschichtet wird und daher keine ausreichende Höhe erreicht.

Die B.A./M.A.-Studiengänge bieten die Chance, an dieser Situation etwas zu ändern. Für die Computerlinguistik mit ihrem handwerklichen Einschlag ist die intensive Ausbildung im BA-

Studium genau das Richtige. Der Schwerpunkt der Ausbildung liegt in diesem ersten dreijährigen Abschnitt auf der Methodik empirischen Arbeitens und den Kriterien für eine exakte Beschreibung sprachlicher Phänomene, gepaart mit extensiver Programmiererfahrung. Dem Berufsbild der Absolventen entspricht der qualifizierte Mitarbeiter in industriellen Projekten mit Anteilen maschineller Sprachverarbeitung oder Sprachdatenakquisition.

Auf den relativ verschulten B.A. folgt der M.A., der relativ frei und projektorientiert sein sollte. Ziel des zweijährigen Studiums ist die Befähigung zur Generalisierung, Abstrahierung und Modellbildung. Auch Erfahrungen in Organisation und Management sollten erworben werden. Das Berufsbild der Absolventen mit M.A. ist der verantwortliche Projektleiter in der Softwareentwicklung oder bei einem Anbieter der "Language Industry".

In den herkömmlichen Magisterstudiengängen besuchen viele Studenten nur die wenigen Lehrveranstaltungen, die für die Zwischenprüfung oder für das Endexamen die juristische Voraussetzung bilden. Manche schaffen es nicht einmal, diese wenigen Scheine in den regulären Fristen zusammenzubekommen, und zwar nicht, weil sie faul sind, sondern weil sie sich zu sehr in einige wenige Seminare vertiefen und wohlmöglich die ganzen Semesterferien damit verbringen, eine einzige Hausarbeit anzufertigen. Vielleicht ist das in manchen geisteswissenschaftlichen Fächern nicht schlimm: Qualität vor Quantität! Aus computerlinguistischer Sicht ist es nicht gut. Im BA-Studiengang wird der Studienaufwand durch ECTS-Punkte (credit points) kalkulierbar gemacht. Ein Punkt entspricht 20 bis 30 Stunden Arbeitszeit. Für den Bachelor-Abschluß müssen 180 Punkte vorgewiesen werden. Wichtig ist, daß bei der Festlegung der ECTS-Punkte nicht nur die Anwesenheitszeit berücksichtigt, sondern auch die Zeit für Vorbereitung, Nachbereitung, Hausarbeiten usw. eingeplant wird. Daß in herkömmlichen Studienordnungen an deutschen Universitäten nur die Präsenzzeiten (Semesterwochenstunden, SWS) angegeben werden, mag mitverantwortlich sein für lange Studienzeiten. Indem dem Studierenden - ebenso übrigens wie dem Dozenten (!) - eine klare Vorstellung von dem zu leistenden zusätzlichen Arbeitseinsatz fehlt, kommt es nicht zu einem vernünftigen Zeit-Management, d.h. häufig wird viel zu lange an zu wenigen Themen gearbeitet.

B.A./M.A.-Studiengänge sind modular aufzubauen. Dadurch soll das Studium überschaubar und kohärent gemacht werden. Es stellt sich die Frage, welche Modulgliederung diesem Ziel am besten dient. Systematisch läßt sich die Computerlinguistik am leichtesten nach Disziplinen gliedern, nämlich in allgemeine Linguistik, Informatik und den speziellen Bereich der maschinellen Sprachverarbeitung. Man kann auch gut drei Aufgaben unterscheiden, die sich einem Computerlinguisten typischerweise stellen: die empirische Untersuchung der Sprache, die formale Repräsentation und die Entwicklung eines Algorithmus. Man könnte einen Studiengang so konzipieren, daß die Module jeweils diesen Teilgebieten und Teilkompetenzen entsprechen. Didaktisch geschickter erscheint es, wenn die meisten Module durch eine Aufgabe definiert werden und dann Lehrveranstaltungen aus den verschiedenen Teildisziplinen enthalten, die zur Lösung der Aufgabe zusammenwirken müssen. Dies entspricht eher dem Lernziel einer Problemlösungskompetenz anstatt eines reinen Wissenserwerbs.

Eine Neuerung gegenüber dem bisherigen System ist die weitgehend studienbegleitende Erfolgskontrolle. Der schriftlichen und mündlichen Prüfung am Ende des Studiums sollten höchstens 20% Anteil an der Endnote eingeräumt werden. Die übrigen 80% beruhen auf studienbegleitenden Prüfungen, in der Regel Klausuren, mündliche Prüfungen, Hausarbeiten zu einzelnen Lehrveranstaltungen. In die Endnote sollen jedoch nicht alle Studienleistungen eingehen. Das würde

einer völligen Verschulung gleichkommen. Deshalb kann der Studierende teilweise selbst entscheiden, in welcher Veranstaltung er nicht nur eine Studienleistung, sondern auch eine Prüfungsleistung erbringen will.

In der Master-Phase muß der Studierende in die aktuelle Forschung einbezogen werden. Dies geschieht am intensivsten durch Teilnahme an Projekten. Der Umfang dieser Beschäftigung sollte einer halben studentischen Hilfskraftstelle für ein Jahr entsprechen. Für hervorragende Studierende sollten tatsächlich Stellen in Drittmittelprojekten zur Verfügung stehen.

Eine Motivation für die Einführung von M.A./B.A.-Studiengängen (und auch für die Wahl der englischen Bezeichnungen) ist die Internationalität. Möglichst alle Studierenden sollten ein Semester im Ausland verbringen. Es sollte in der Computerlinguistik nicht schwerfallen, mit ausländischen Universitäten feste Vereinbarungen über einen Studierendenaustausch zu treffen. Der Club der Universitäten, welche eine Akkreditierung für den "European Masters in Language and Speech" besitzen, sollte ausgeweitet werden. Einige obligatorische Leistungspunkte sollten für den Besuch von wissenschaftlichen Konferenzen vergeben werden, weil der Kontakt zur wissenschaftlichen "Szene" ein wichtiges Ausbildungsziel ist. Auch für die Teilnahme an Sommerschulen oder an Fernkursen über das Internet kann man Leistungspunkte bekommen. Die Fähigkeit, eine mündliche Präsentation in englischer Sprache und mit professioneller Technik zu halten, wird selbstverständlich im Curriculum berücksichtigt.

Meines Erachtens hat die Computerlinguistik ein PR-Problem. Die Öffentlichkeit, ebenso wie Rektorate und Ministerien sind sich nicht bewußt, daß die Computerlinguistik eine Schlüsseltechnologie der IT-Branche ist. Zum Teil steht die Vorstellung dahinter, daß Sprachverarbeitung nebenbei von den Informatikern erledigt werden kann. Die Computerlinguistik wird daher nicht besonders gefördert, sondern die Ressourcen müssen mühsam aus den Philologien herausgepreßt werden, die sich verständlicherweise beharrlich wehren. So vollzieht sich der Ausbau der Computerlinguistik in der Lehre unendlich langsam und muß praktisch mühsam gegen alle durchgesetzt werden. Es gibt nur einen Ausweg. Die Computerlinguistik muß sich am eigenen Schopf aus dem Sumpf ziehen. Sie muß im Kleinen hervorragende Ausbildung bieten. Dann werden Forschung und Entwicklung zunehmend erfolgreich sein, was eine Vermehrung der Ressourcen zur Folge haben wird.