

Zum Fortschritt in der Computerlinguistik

Wolfgang Wahlster

Fachbereich Informatik
Universität des Saarlandes
6600 Saarbrücken 11

Nachdem es in den letzten sechs Jahren allein in der Bundesrepublik bereits drei dokumentierte Diskussionsrunden mit einer ähnlichen Themenstellung gegeben hat (vgl. Batori et al. 1982, Görz 1984, Knorz 1985) und ich mich an zweien beteiligt hatte (vgl. Wahlster 1982 und 1984), möchte ich in diesem kurzen Beitrag meine Position nicht wiederholen, sondern die Fragestellung dieser Podiumsdiskussion aus einer anderen Perspektive betrachten, indem ich frage: Was bedeutet Fortschritt in den theoretischen Grundlagen der Computerlinguistik?

Bevor ich im folgenden zwei Teilantworten auf diese Frage skizziere, möchte ich festhalten, daß die Gegenstandsbereiche und die Methoden der sprachorientierten KI-Forschung und der Computerlinguistik in den letzten Jahren so stark konvergieren, daß es mir im Rahmen dieser Diskussion nicht gerechtfertigt erscheint, zwischen diesen beiden Forschungsgebieten zu differenzieren. Diese Auffassung wird sowohl in dem bisher umfassendsten Lehrbuch der Computerlinguistik bestätigt, in dem Allen feststellt 'The tools that the work in computational linguistics uses are those of artificial intelligence' (vgl. Allen 1987, S. 1), als auch in neueren Überblicksarbeiten z.B. von Halvorsen, der vermerkt 'Computational linguistics is best viewed as a branch of artificial intelligence' (vgl. Kapitel 3 in Halvorsen 1986).

Eine Möglichkeit, die grundlegende Zielsetzung einer wissenschaftlichen Disziplin wie der Computerlinguistik zu charakterisieren, besteht in der Beantwortung der Frage, wie wissenschaftlicher Fortschritt eigentlich definiert wird. Im folgenden wird der Versuch gemacht, anhand von zwei methodischen Leitlinien der heutigen Computerlinguistik exemplarisch den Fortschritt gegenüber dem Entwicklungsstand am Ende der 70er Jahre herauszuarbeiten.

Auf zwei methodischen Ebenen wurden in den letzten zehn Jahren erhebliche Fortschritte erzielt, die für einen höheren wissenschaftlichen Reifegrad der Computerlinguistik sprechen:

- Eine stärkere *Mathematisierung* in allen Teilbereichen
- Ein stärkerer Austausch mehrfach verwendbarer *Software-Werkzeuge/Module*

Während man sich vor zehn Jahren in der Computerlinguistik häufig noch mit Behauptungen und Vermutungen über die formalen Eigenschaften von Beschreibungs- und Verarbeitungsverfahren begnügen mußte, ist man heute hauptsächlich an *beweisbaren* Aussagen interessiert.

Man greift daher verstärkt auf Resultate aus der Mathematik und der Theoretischen Informatik zurück, wobei derzeit die folgenden Teilgebiete die wichtigste Rolle spielen:

- Mathematische Logik
- Formale Sprachen
- Komplexitätstheorie
- Algebraische Semantik

Beispiele für typische Fragen, auf die man heute von jedem Autor eindeutige Antworten erwartet, sind:

- Welches *Laufzeitverhalten* hat ein neues Parsingverfahren?

Eine typische Antwort hierzu ist: Das Parsen von Tree Adjoining Grammars (TAGs) ist in $O(n^4)$ möglich (vgl. Harbusch 1988).

- Welche *Mächtigkeit* hat ein neuer Grammatikformalismus?

Die Aussage $CFL \subset TAL \subset CSL$ stellt ein Beispiel für eine befriedigende Antwort dar.

- Ist ein neues Logik-System *vollständig*?

- Eine typische Antwort hierzu ist: Die Logik L_3 für Situationsschemata ist vollständig (vgl. Fenstad et al. 1987).

- Ist ein neues Unifikationsverfahren *entscheidbar*?

Auch negative Ergebnisse wie 'Unifikation unter Distributivität und Assoziativität ist unentscheidbar' tragen zum Fortschritt bei.

Eine Gefahr, die Schwartz treffend mit 'bad theory with a mathematical passport' (vgl. Schwartz 1962) gekennzeichnet hat, ist mit dieser Tendenz der Mathematisierung allerdings verbunden: Man stößt ab und zu auf pseudomathematisch präsentierte 'Theorien', die bei Nicht-Mathematikern durch eine aufwendige Formelsprache den Eindruck der Seriosität erwecken wollen. Im Gegensatz zur Theoretischen Linguistik, wo z.B. im Zusammenhang mit der Montague-Rezeption gelegentlich bizarre Formalisierungsversuche, die weder mathematische Tiefe noch linguistische Aussagekraft besaßen, veröffentlicht wurden, ist dieses Phänomen in der Computerlinguistik jedoch recht selten anzutreffen.

Ein weiteres Problem im Zusammenhang mit der Mathematisierung der Computerlinguistik ist in einigen Fällen die extreme Fixierung auf einzelne Formalismen, die innovationshemmend wirken kann. In der Computerlinguistik steht noch zu oft die Frage im Vordergrund, ob und wie in einem vorgegeben Formalismus eine sprachliche Regularität erfaßt werden kann, anstatt zunächst das inhaltliche Ergebnis zu betrachten und dann nach einem geeigneten Formalismus zu suchen. Hier muß sich die in der modernen mathematischen Logik übliche Vorgehensweise, zunächst die Strukturen des zu modellierenden Gegenstandsbereiches genau zu untersuchen und dann einen passenden Kalkül mit den gewünschten Eigenschaften zu entwerfen, erst noch allgemein durchsetzen.

Während in den 70er Jahren nur ATN-Entwicklungsumgebungen (Interpreter, Compiler, Editoren, Browser) in größerem Umfang zwischen computerlinguistischen Forschungsgruppen ausgetauscht, weiterentwickelt und in verschiedenen Projekten genutzt wurden, existieren heute für verschiedene Ebenen des Sprachverstehens und der Sprachproduktion universelle, portable und effiziente Software-Werkzeuge/Module in den Standardprogrammiersprachen der Computerlinguistik (CommonLisp und Prolog).

Dies stellt einen bemerkenswerten Fortschritt dar, weil man sich nun in einem computerlinguistischen Projekt beim Aufbau der experimentellen Umgebung für die Untersuchung einer ganz bestimmten Fragestellung im Kontext eines natürlichsprachlichen Gesamtsystems im Idealfall mehrerer vorgefertigter Software-Pakete bedienen kann, um dadurch schneller zum eigentlichen Forschungsthema eines Projektes zu kommen. Die- Investitionen in Basissoftware für die Analyse und Generierung natürlicher Sprache bringen hohe Dividenden, da es so möglich wird, neue Methoden schneller zu entwickeln und zu testen (vgl. auch Wahlster 1988). Da in den 70er Jahren

noch erheblich mehr Entwicklungsarbeit in die jeweiligen Rahmensysteme gesteckt werden mußte, konnten in vielen Projekten nur Systemfragmente entstehen.

Wenn auch die derzeitigen Modelle menschlichen Sprachverhaltens noch viel zu lückenhaft sind, um übertriebene Erwartungen in Hinblick auf schnelle Erfolge im Bereich standardisierter Basissoftware zu rechtfertigen, so stellt doch der Durchbruch, der auf dem Gebiet der Expertensysteme durch die Einführung sog. Expertensystem-Shells erreicht werden konnte, eine ermutigende Analogie dar.

Als Beispiele für solche oft lizenzfrei und kostenlos erhältlichen Software-Pakete, die z.T. aufgrund ihrer guten Dokumentation und ihrer hohen Zuverlässigkeit bereits weit verbreitet sind, seien hier genannt:

- Morphologie-Pakete (z.B. KIMMO, MORPHIX)
- Grammatik-Werkbänke (z.B. LFG-Werkbänke, PATR, TAG-DevEnv)
- Benutzermodellierungskomponenten (z.B. GUMS, TRUMP)
- Wissensrepräsentationssprachen (z.B. KL-ONE/TWO, SB-ONE, QUIRK/QUARK)
- Sprachgenerierungsmodule (z.B. MUMBLE-86, SUTRA)

Einzelne dieser Werkzeuge sind in Versionen für alle gängigen LISP-Dialekte und Rechnertypen erhältlich. So ist z.B. das in unserem Saarbrücker KI-Labor entstandene MORPHIX-System in CommonLisp, Interlisp-D, ZetaLisp und FranzLisp auf Symbolics und TI Lispmaschinen sowie Arbeitsplatzrechnern von DEC, HP, SUN, Xerox und Siemens verfügbar (vgl. Finkler/Neumann 1988).

Nach einer langen Anlaufphase steht die Computerlinguistik jetzt vor einer Entwicklung, bei der es auch im Softwarebereich immer häufiger möglich werden sollte, auf die Ergebnisse anderer Forscher aufzubauen.

Zum Schluß möchte ich noch einen weiteren Bereich erwähnen, in dem Fortschritte zu verzeichnen sind, die allerdings eher die anwendungsorientierte Seite der Computerlinguistik als ihre theoretischen Grundlagen betreffen. Seit 1985 wurden mit Q & A, Language Craft, NL-Menu, DataTalker und Parlance fünf natürlichsprachliche Schnittstellen auf dem kommerziellen Markt eingeführt, die Resultate der sprachorientierten KI-Forschung der 70er Jahre nutzbringend für die Mensch-Maschine-Kommunikation umsetzen (vgl. auch Wahlster 1984). Dadurch, daß diese Systeme heute weltweit von mehr als 10.000 Anwendern benutzt werden, wird es erstmals vertretbar, durch empirische Untersuchungen beim Einsatz natürlichsprachlicher Systeme die Priorisierung der Bearbeitung bestimmter Theorie- und Methodendefizite in der Computerlinguistik auch aus der Anwendungspraxis heraus zu motivieren.

Literatur:

Allen, J. (1987): Natural Language Understanding. Menlo Park: Benjamin/Cummings.

Batori, I., Krause, J., Lutz, H.-D. (eds.) (1982): Linguistische Datenverarbeitung. Versuch einer Standortbestimmung im Umfeld von Informationslinguistik und Künstlicher Intelligenz. Tübingen: Niemeyer.

Fenstad, J., Halvorsen, P.-K., Langholm, T., van Benthem, J. (1987): Situations, Language and Logic. Dordrecht: Reidel.

- Finkler, W., Neumann, G. (1988): MORPHIX: A Fast Implementation of a Classification-Based Approach to Morphology. In: Trost, H. (ed.): Wiener Workshop Wissensbasierte Sprachverarbeitung. Heidelberg: Springer.
- Görz, G. (ed.)(1984): Linguistik und Künstliche Intelligenz. Podiumsdiskussion mit Beiträgen von Th. T. Ballmer, E. Lehmann, K. Morik, M. Pinkal, W. Wahlster. In: Laubsch, J. (ed.): GWAI-84. 8th German Workshop on Artificial Intelligence. Heidelberg: Springer, S. 265-282.
- Halvorsen, P.-K. (1986): Computer Applications of Linguistic Theory. In: Newmeyer, F., Ubell, R. (eds.): Linguistics: The Cambridge Survey. Cambridge: University Press.
- Harbusch, K. (1988): Effizientes Parsing mit TAGs. In: diesem Band.
- Knorz, G. (ed.)(1985): Ansätze der sprachorientierten KI-Forschung. Eine Thesensammlung mit Beiträgen von H.-D. Lutz, A. Fauser, P. Scheffe, U. Hahn, K. Morik und B. Endres-Niggemeyer. In: LDV-Forum, 3, 2, S. 77-90.
- Schwartz, J. (1962): The Pernicious Influence of Mathematics on Science. In: Nagel, E., Suppes, P., Tarski, A. (eds.): Logic, Methodology and Philosophy of Science. Stanford: Univ. Press, S. 356-360.
- Wahlster, W. (1982): Aufgaben, Standards und Perspektiven sprachorientierter KI-Forschung. Einige Überlegungen aus informatischer Sicht. In: Bátori et al., S. 13-24.
- Wahlster, W. (1984): Zur Rolle der Linguistik bei der Entwicklung natürlichsprachlicher KI-Systeme. In: Görz, S. 267-269.
- Wahlster, W. (1986): The Role of Natural Language in Advanced Knowledge-Based Systems. In: Winter, H. (ed.): Artificial Intelligence and Man-Machine Systems. Heidelberg: Springer, S. 62 - 83.
- Wahlster, W. (1988): Natural Language Systems - Some Research Trends. In: Schnelle, H. (ed.): Logic and Linguistics, Hillsdale: Lawrence Erlbaum.