

## COLING 76: Concepts, Frames, and Scripts in Aid of Semantic Networks, Knowledge Systems, and Fantasies

(November 1976)

Dies hätte der Titel der 6. Tagung für *Computational Linguistics* sein können, die vom International Committee on Computational Linguistics (ICCL) während seines inzwischen 11-jährigen Bestehens diesmal in Ottawa, Kanada, vom 28. Juni bis 2. Juli 1976 veranstaltet wurde.

Daß wissenschaftliche Termini und Begriffe zuweilen Schibboleth-Charakter annehmen, ist eine ebenso bekannte wie wenig beachtete Erscheinung. Sie wird im allgemeinen, und meist auch zurecht, als eine der kurzlebigeren Begleiterscheinungen des Wissenschaftsbetriebs abgetan. Möglicherweise aber läßt sie sich auch als Hinweis auf längerfristige Vorgänge deuten, wenn man sie als Beleg etwa thematischer, methodischer oder auch theoretischer Um- oder Neuorientierungen wertet, wie sie sich gerade im Verlauf der Konsolidierung einer wissenschaftlichen Disziplin häufiger ergeben.

Im Rahmen des *Natural Language Processing*, dem es in den zehn Jahren nach dem ALPAC-Report gelungen ist, wesentliche Teile einer früheren Aufgabenstellung aus der maschinellen Übersetzung (MT) mit Hilfe des neuen Forschungsansatzes der künstlichen Intelligenz (AI) in die Computerlinguistik (CL) hinüberzueretten, scheinen gleich mehrere Begriffe dieses Phänomen zu belegen. Deren Wirksamkeit stützt sich dabei vorerst weniger auf hohe Abstraktheit oder begriffliche Strenge, sondern vornehmlich auf die Bereitschaft, mit der ihre Verwender die eigenen mehr oder weniger vagen Grundvorstellungen mit den nicht minder unscharfen Begriffsbildungen anderer identifizieren. Vertreter zunächst noch unterschiedlicher Forschungsansätze und Fachrichtungen gelangen so zunehmend zu der Einsicht, seit langem ganz ähnliche oder gleiche Gedanken und Vorstellungen entwickelt oder mit solchen

gearbeitet zu haben, ohne sich dieser Nähe von Gegenstand, Methode oder Idee der eigenen Arbeiten zu denen anderer bewußt gewesen zu sein. Dabei steht schon jetzt der Einfluß und die Fruchtbarkeit des Neuansatzes außer Frage, der durch die AI auf dem Gebiet der Entwicklung von Modellen und Systemen zu Abbildung und Simulation von Verstehensprozessen initiiert wurde und etwa mit Begriffen wie *concept*, *frame* oder *script* verbunden ist.

As the basic atomic unit ... *Concept* is defined as representation of an object occupying space and/or time, of Attributes specifying what an object is or does, and of relation, between objects. The modular unit is a structure of concepts, a conceptualization that represents situations and relations between situations.<sup>1</sup>

A *frame* is a data-structure for representing a stereotype Situation ... Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed. We can think of a frame as a network of nodes and relations.<sup>2</sup>

A *script* is ... defined as a predetermined causal chain of conceptualizations that describe a normal sequence of things in a familiar situation.<sup>3</sup>

Die Gemeinsamkeiten in Problemstellung bzw. Forschungsansatz sind freilich noch kaum präzisiert, weder im Hinblick auf ihre begrifflichen Implikationen noch auf ihre theoretischen Grundlagen oder methodischen Konsequenzen, worauf etwa R. J. Brachman (What's in a Concept: Structural Foundation for Semantic Networks) hinzuweisen suchte. Soviel allerdings

<sup>1</sup>Colby, K. M.: „Simulation of Believ Systems“ in: Schank/Colby: Computer Models of Thought and Language, San Francisco 1973, 251-286.

<sup>2</sup>Minsky, M.: A Framework for Representing Knowledge, MITAI-Lab.-Memo Nr. 306, 1974.

<sup>3</sup>Schank, R. C.: „Using knowledge to understand“ in: Theoretical Issues in Natural Language Processing, Cambridge/Mass. (BBN) 1975, 117-121.

ist deutlich: die neuen Begriffe – nach innen vereinheitlichend, nach außen wie Erkennungs- und Markenzeichen wirkend – werden zu einer semantischen Umorientierung führen, die es erlaubt, Bedeutung übergreifend als Resultat eines Prozesses zunehmender Einschränkung von Wahlmöglichkeiten nicht nur zu charakterisieren, sondern praktisch zu simulieren innerhalb des Rahmens *frame*, der von einer Situation *script* vorgezeichnet wird.

Die Auswahl der Papers durch die für COLING 76 zuständige Programm-Kommission unter Leitung von M. Kay ließ eine offenbar nicht auf Repräsentativität abzielende Schwerpunktsetzung in drei Bereichen erkennen: maschinenunterstützte bzw. automatische Übersetzungssysteme, anwendungsorientierte Frage-Antwort-Systeme, sowie der durch den Einfluß der AI geprägte Komplex der im weiteren Sinne semantischen Ansätze. Diese Schwerpunkte ergaben sich nach Auskunft der Programm-Kommission einmal aus den am Ort (Ottawa, Montreal) vorhandenen Übersetzungsprojekten, sowie durch die in den USA (Stanford, Yale, MIT, Bolt-Beranek-Newman) zur Zeit verstärkt betriebenen Projekte zur praxisnahen Entwicklung von World-Knowledge-Modellen und Systemen zur Simulation von Verstehen in begrenzten situativen Kontexten.

Nur auf den letzteren soll im folgenden kurz eingegangen werden, da er sich – schon jetzt erkennbar – auf die anderen Bereiche auszuwirken beginnt und vermutlich einmal das Kernstück einer zukünftigen Computerlinguistik bilden können.

Sieht man von den Vorträgen C. J. Fillmores (The Need for a Frame. Semantics in Linguistics) und Y. Wilks (Frames, Scripts, Stories, and Fantasies) ab, die eine umfassende, einmal auf pragmalinguistische Zusammenhänge, zum anderen auf methodologische Konsequenzen verweisende theoretische Reflexion dieser neuen Gemeinsamkeiten einleiten könnten, so wurde das Bild der Konferenz weithin bestimmt von Beiträgen, die – auf implementierbare Lösungen konzentriert – praxisnahe Entwicklungen funktionsfähiger Systeme vorstellten. So etwa zum Aufbau und zur Organisation von Da-

tenstrukturen und Netzwerken: A Compiling System for Augmented Transition Networks (R. R. Burton/W. A. Woods), Organizing Knowledge for English-Chinese Translation (W. J. Stutzman); oder zur Modellentwicklung des Sprachverstehens: Organizing and Control of Syntactic, Semantic, Inferential, and World Knowledge for Language Understanding (F. Hayes-Roth/D. J. Mostow), A Multi-Processing Model for Natural Language Understanding (R. Smith/F. Rawson), The Application of Script-Based World Knowledge in an Integrated Story-Understanding System (R. E. Cullingford), Evolving Uses of Knowledge in a Speech Understanding System (B. Bruce/Nash-Weber); oder zum Komplex der Frage-Antwort-Systeme: Dynamic Processing and Question Answering (W. Lehnert), Query Optimization for Question-Answering Systems (R. Reiter), Supporting a Computer-Directed Natural Language Dialog for Automatic Business Programming (G. E. Heidorn).

Durchgängig fällt bei allen diesen Modellen und Systemen auf, daß von der Verfügbarkeit der primär benötigten und etwa in *frames*, *scripts* oder *Netzen* enthaltenen vielfältigen Informationen immer schon ausgegangen wird. Je nach Aufgabenstellung und (linguistischem) Vermögen ad hoc zusammengestellt, werden in diesen Daten Vorstellungen über Aufbau und Funktion sprachlicher und außersprachlicher Mittel vorgegeben, die von deren tatsächlichem Gebrauch in konkreten Situationen verbaler Interaktion vermutlich nicht immer abgedeckt sind. Die Tatsache jedenfalls, daß *frame-* oder *script-based* Netzstrukturen sich als fruchtbar und überaus leistungsfähig bei der Entwicklung und Konstruktion von *Frage-Antwort-Systemen* erweisen, rechtfertigt nicht schon die Annahme (etwa Cullingfords und Lehnerts), damit die Grundprinzipien auch der natürlich-sprachlichen Bedeutungskonstitution erfaßt zu haben. Zumal alle bisherigen Versuche, das natürlich-sprachliche Generieren/Verstehen beliebiger neuer Bedeutungen innerhalb dieser Netzstrukturen abzubilden, die Zahl der benötigten *scripts* oder *frames* explodieren läßt (Salton).

Vor diesem Hintergrund sollte die Forderung nach Formulierung empirisch überprüfbarer Hypothesen, wie sie für den Bereich der AI ausdrücklich von Wilks erhoben wurde, auf den Bereich der CL ausgedehnt werden. Wenn deren Aussagen über natürlich-sprachliches Verhalten keine im Prinzip unüberprüfbar behauptungen im Rahmen bestenfalls plausibler Phantasien bleiben sollen, bedarf es einer analytisch-quantitativ arbeitenden, empirischen Komponente dieser Disziplin, die anhand performanz-orientierter Textanalysen die Adäquatheit und Erklärungsstärke modelltheoretischer Hypothesen aus kompetenz-orientierten Theorien empirisch zu überprüfen gestattet. Freilich kann davon, trotz der in diese Richtung weisenden Beiträge von S. Allen (Text-Based Lexicography and Algorithmic Text Analysis), J. King/I. McMasters/J. R. Sampson (Computer Acquisition of Natural Language: Experimental Tests of a Proposed System) oder W. Moskovich, bisher noch nicht eigentlich die Rede sein.

Dabei würden sich zahlreiche im Bereich des *Document Retrieval* verwendete Techniken, die – wie G. Salton (A Comparison of Term Value Measurements for Automatic Indexing) einmal mehr nachwies – semantische Textcharakteristika erfolgreich zu ermitteln erlauben, unmittelbar umsetzen und nutzen lassen für eine quantitative Analyse bedeutungskonstituierender Regularitäten in nun allerdings pragma-linguistisch zu definierenden Korpora natürlich-sprachlicher Texte. Die damit erst sich bietende Möglichkeit zu wechselseitiger Kontrolle von theoretisch-formaler Modellbildung einerseits und empirisch-experimenteller Datenerhebung andererseits würde den Prozeß der Konsolidierung der CL als Disziplin sicherlich beschleunigen. Solange allerdings Fragen der Organisation von Datenmengen und der Konstruktion von Netzen, der Kodierung und Repräsentation von Informationen und die praktische Entwicklung oder anwendungsbezogene Optimierung funktionsfähiger Systeme die Aktivitäten dieser Disziplin ausmachen, wird ihr Verhältnis zur Linguistik vorerst eher wie das einer Ingenieurwissenschaft beschrieben werden

müssen, der die Theorienbildung der Physik entbehrlich scheint. Die Hoffnung, daß CL vom bloßen Anwendungsbereich der Computerwissenschaft zum so notwendigen, weil harte Daten verarbeitenden Teilbereich einer wahrhaft empirisch zu nennenden Linguistik werden möge, fand auf der 6. Tagung für „Computational Linguistics“ noch wenig Nahrung.

An der vom Department of Linguistics der Universität Ottawa ausgerichteten und im ganzen wohl vorbereiteten und organisierten Konferenz nahmen rund 200 Wissenschaftler aus etwa 20 Ländern teil. Sie machten zirka 75 Prozent der ursprünglich gemeldeten Interessenten aus, von denen allerdings viele durch einen Streik der kanadischen Fluglotsen und Air Canada am Kommen gehindert wurden. Daß – wie schon bei den früheren Tagungen – die Länder des Ostblocks nicht vertreten waren (Ausnahme: Tschechoslowakei), ist dagegen wohl nicht auf den Streik zurückzuführen. Wegen der bedauerlichen, überdies sehr kurzfristigen Ausfälle – z. B. war keiner der mit zugesagten Beiträgen angemeldeten sowjetischen Wissenschaftler erschienen – konnten von den ursprünglich vorgesehenen eingeladenen Vorträgen (lectures), eingereichten und akzeptierten Vorträgen (papers) und den praktischen Vorführungen (demonstrations) nur etwa 60 zum Teil in bis zu vier parallelen Veranstaltungen während der vier Kongreßtage stattfinden. Ein zur Halbzeit eingeschobener Exkursionstag (Upper Canada Village) bot eine willkommene Unterbrechung. Allgemein wurde am Tagungsprogramm von Teilnehmern kritisiert, daß ein (privates) Forschungsinstitut (BBN, Cambridge, Mass.) offenbar besondere Bevorzugung (6 papers) genoß. Rücksichten der Programm-Kommission auf die in Kanada notwendige, bi-linguale (französisch/englische) Ausgewogenheit führten überdies im Rahmen der in Ottawa vertretenen europäischen Forschung zu einer gewissen Überbetonung (9 papers) von Projekten aus Frankreich (Grenoble, Paris, Rouen, Straßburg). Dabei war auffallend, daß sowohl England (1 paper) als auch die Bundesrepublik (2 papers) – gemessen an den jeweiligen For-

schungsaktivitäten in diesen Ländern deutlich unterrepräsentiert waren. Für den mit der computerunterstützten linguistischen Forschung bei uns weniger vertrauten Beobachter konnte daher der Eindruck entstehen, daß es in der Bundesrepublik – mit ihrer (nach den USA höchsten) Zahl und Dichte an Großrechenanlagen – eine im internationalen Vergleich beachtliche Computerlinguistik nicht gäbe. Ob diese Unterrepräsentation allerdings ausschließlich auf die Programm-Kommission zurückzuführen ist oder sich teilweise auch aus der geringeren Bereitschaft westdeutscher Linguisten erklärt, überhaupt Papers anzubieten, muß dahingestellt bleiben.

Dagegen waren die für COLING 76 herrschenden Auswahlprinzipien entscheidend dafür, daß jene neueren Ansätze in der computerunterstützten Linguistik noch völlig ausgeklammert blieben, die über die Theorie der unscharfen Mengen (fuzzy sets) bzw. die Toleranztopologie empirisch wie theoretisch angemessenere Modelle und Systeme zu entwickeln suchen auf Gebieten wie der Knowledge Representation, der Inferenzprobleme bei unvollständiger Kenntnistage, sowie der natürlich-sprachlichen Bedeutungskonstitution. So kamen weder die Forschungsgruppe von L. A. Zadeh (Berkeley, Cal.) noch die von B. R. Gaines (Sussex, Engl.) zu Wort, die im Bereich der computerunterstützten Entwicklung von Systemen des logischen Schließens mit natürlich-sprachlichen Aussagevariablen, der Repräsentation von natürlich-sprachlichem Wissen und der Problemlösung bei unscharfer Aufgabenstellung arbeiten. Von dem derzeitigen Stand dieser gerade für die Computerlinguistik interessanten Entwicklung, die etwa auf der Tagung für „Cybernetics and Systems Research“ (EMCSR 76) in Wien deutlich geworden war, fand sich im offiziellen Programm von COLING 76 nichts.

Als Tagungsort für COLING 78 wurde, um den Wissenschaftlern der Ostblockländer die Teilnahme zu erleichtern, Varna (Bulgarien) vorgesehen, alternativ auch noch Bergen (Norwegen) und Bonn (Bundesrepublik).