

Nicht-Explizites Wissen in der Computerlinguistik

Walther v.Hahn,
Universität Hamburg, Fachbereich Informatik
vhahn@informatik.uni-hamburg.de

Studie im Auftrag des FAW Ulm

Zusammenfassung

Die meisten Psycholinguisten sind der Auffassung, daß sprachliches Wissen ausschließlich implizit sei, was sich darin zeige, daß der Spracherwerb nicht über Regellernen abläuft und z.B. Generierung nicht vorhersagbar und nur teilweise rationalisierbar sei. Die Linguistik hat diese Grundeinschätzung in ihren grammatischen und semantischen Disziplinen noch kaum, die Computerlinguistik solche Vorstellungen noch fast gar nicht umgesetzt. Während sich die Linguistik auf alle Sprachvarianten konzentriert, modelliert die Computerlinguistik fast ausschließlich fachsprachliches Verhalten, das seiner Natur nach eher explizit ist als Alltagssprache oder gar literarische Sprache.

Das folgende Papier stellt die Themen in Linguistik und Computerlinguistik vor, auf die der Begriff des Nicht-expliziten Wissens am ehesten abgebildet werden kann.

1. Terminologische und fachliche Abdeckung

Da der Terminus "nicht-explizit" im Sinne der Papiere (Radermacher97a) und (Radermacher97b) in der Computerlinguistik eher selten benutzt wird, bildet die Studie ihn ab auf die in der Literatur benutzten Termini "implizit", "unbewußt", "nicht-regelhaft" und "vage". Darin zeigt sich, daß in der Computerlinguistik nur wenige Facetten des Terminus "nicht-explizit" abgedeckt sind. Da die Grenzen zwischen Linguistik und Computerlinguistik häufig fließend sind (z. B. im Falle M. Pinkals), sind auch linguistische Themen aufgenommen worden.

2. Die wissenschaftliche Rolle des nicht-expliziten Wissens in der Computerlinguistik

Implizitheit ist bis in die Gegenwart ein vernachlässigtes Thema, sowohl in der Linguistik wie noch mehr in der Computerlinguistik. Einer der Gründe dafür ist, daß die Sprachbeschreibung (zu Ebene 4) auch in der Computerlinguistik sich immer noch im ganz stark

a) als regelbasiertes Paradigma versteht, und überwiegend

b) technische, informationübermittelnde Interaktion (Fachsprache) modelliert.

Daher erscheinen dann systematisch schwer beschreibbare sprachliche Phänomene, die der Vorstellung einer Datenübertragung ferner liegen, dysfunktional. Ein radikales Beispiel sind die AECMA-Definitionen, die systematisch auch die Semantik einschränken. Auch die Forschung im methodischen Paradigma der neuronalen Netze hat sich vorwiegend mit der Modellierung solcher linguistischer Phänomene befaßt, für die auch eine symbolische regelbasierte Lösung existiert oder geschrieben werden könnte (begriffliche Klassifizierung, Assoziationsstärke von Wörtern, Signalerkennung etc.)

Ob wir beim Thema "Nicht-Explizitheit" einen Forschungsansatz der (Computer-)Linguistik für relevant halten, hängt auch davon ab, wie tief er bearbeitet wurde. Unterscheiden muß man wohl zwischen

- der Bewußtheit der Problematik (z.B. beim Thema Inhouse-Terminologie),
- der systematischen Sammlung von (Trainings-) Material (z.B. beim Thema Verstehen von Segmentierungsinformation)
- einem formalen Modell (z.B. beim Thema Präsupposition und Allegation)
- einer Implementierung des Verstehensmodells (z.B. beim Thema Pronomen-Auflösung)
- einem aktiven Simulationsmodell (z.B. beim Thema Erzeugen von Ellipsen)

Forschungsgeschichtlich hat sich das Themengebiet in der Linguistik etwa folgendermaßen entwickelt: Bereits im Frühstrukturalismus ist der Gedanke entwickelt worden, daß es sprachliche Systeme gibt, deren Teile sich nicht einzeln verändern, sondern daß eine bewußte Änderung an einer Stelle systemische Veränderungen an ganz anderer Stelle bewirken (Trier, Havránek).

Die Anfänge der breiteren Beschäftigung mit nicht-explizitem Wissen liegen in der Linguistik der Nachkriegszeit, in der Assoziationen und Konnotationen als linguistische Größen am Rande eingeführt wurden. Auch die Nicht-Vorhersagbarkeit sprachgeschichtlicher Veränderungen wurde zu dieser Zeit konstatiert (als "Nebenprodukt" von Formalisierungsunternehmungen). Aus der Logik wurde die frühere Forschung

zu Vagheit und Präsupposition in dieser Zeit auch in der Linguistik wahrgenommen und in den frühen 80ern in der Computerlinguistik erstmalig modelliert. In der frühen Computerlinguistik wurden vornehmlich die logischen Aspekte in PLIDIS (IdS, in LISP) und Vagheitsphänomene mit einer Fuzzy-Repräsentation und der Sprache FUZZY, bzw. AIMDS im System HAM-RPM (Hamburg) implementiert.

Die Soziolinguistik der 70er ist vor allem deshalb gescheitert, weil die Parameter für (schichten)spezifisches Sprechen nicht erhoben werden konnten. Das gab Anlaß zu einer grundsätzlichen Skepsis gegenüber einer festen Zuordnung von bestimmten linguistischen Indikatoren zu bestimmten sozial-kognitiven Funktionen (z.B. bestimmtes Personalpronomen = restringierter Code bei Bernstein) und von bestimmten sozialen Leistungen zu bestimmten Codes.

Erst Mitte der 90er kam mit lernbasierten Verfahren, abgeleitet aus der Sprachsignalerkennung, eine Methode in die computerlinguistische Diskussion, die gleichviel, ob ein Phänomen symbolisch beschrieben werden konnte oder nicht explizierbar war, mit Lernverfahren ein gegebenes Ein-Ausgabeverhalten auf einer Maschine evozieren konnte. Bereits aus der neuronalen Forschungsrichtung bekannte Verfahren wurden dabei genutzt. In der Tat waren die Erfolge erstaunlich, von der Inhaltsanalyse bis hin zur maschinellen Übersetzung. Neuerdings werden verstärkt hybride Verfahren eingesetzt. Allerdings ist bisher noch kein Verfahren bekannt geworden, das mit einem subsymbolischen Verfahren bei gleicher Tiefe bessere Ergebnisse erzielen konnte (Wermters "scanning understanding" ist ein gutes Beispiel).

Die Linguistik hatte lange Zeit, vor allem vor der Zeit der Kognitionsforschung, eine starke Tendenz, implizites Wissen überhaupt dem Weltwissen zuzuordnen und damit aus der Linguistik auszugrenzen. Das hat nur teilweise den gewünschten Erfolg gehabt, denn z.B. die Steuerung einer linguistischen Generierungskomponente ist ohne massive Berücksichtigung von Weltwissen dann nur sehr partiell modellierbar.

3. Die klassischen Forschungsthemen (*mit Beispielen*)

Themen, die nur in der Linguistik, nicht aber in der Computerlinguistik ausführlicher behandelt werden, sind in diesem Kapitel mit einem * gekennzeichnet.

3.1. Impliztheit

Pinkal weist schon darauf hin, daß Expliztheit/Impliztheit eine graduelle Eigenschaft ist, d.h. jede Äußerung mehr oder weniger implizit ist. Es bestehen darüber hinaus

auch begründete Zweifel daran, ob es überhaupt Äußerungen gibt, die per se exakt sind, wie die "Philosophie der normalen Sprache" deutlich gezeigt hat.

Implizitheit gilt als nicht vollzogene Kontext-Projektion. Als Kontext soll dabei gelten, "was irgendeinen Einfluß auf den möglichen Sinn der Äußerung besitzt: die Kommunikationsteilnehmer in ihren spezifischen Rollen, Sprechzeit und Ort der Äußerung, der sprachliche Vorkontext, die gestische Begleitung der Äußerung, die unmittelbare Umgebung der Kommunikationsteilnehmer, ihre augenblickliche Verfassung, ihre Vorgeschichte, dazu beliebige und beliebig weit entfernte physische, soziale und historische Fakten - kurz gesagt, die ganze Welt relativ zum Äußerungsereignis" (weite Teile des Abschnitts basieren auf Pinkal 1985, 56ff).

Signalebene: Interpretation der Prosodie [*ich habe das nicht gesagt*], Verstehen von Segmentierungsinformation [*Pausen, Tonhöhe*]

Morphologie: *Kontext (z.B. firmen-)abhängige Expansion von Abkürzungen [*ASL = Architektur von Speech - und Language Systemen, American Sign Language*]
*Kurzterminini [*Klemmenabdeckplatte ->Klemmenplatte*], Artikelzuordnung, Semantik von Mehrfachzusammensetzungen (*Kinderlandverschickung*), Selektion von Diskriminationsdimensionen bei Zusammensetzungen [*Schweineschnitzel vs. Jägerschnitzel*]

Lexikon: Auflösung von Proformen und ihre Referenzierung [*Das muß von dieser Abteilung jetzt umgesetzt werden, oder: Der Vorstand veräußerte zwei firmeneigene Grundstücke. Davon wurden notwendige Investitionen getätigt*], Bindung von Termini an terminologische Systeme [*"Frosch" in unterschiedlichen Kontexten*], Inhouse-Terminologie [*Lopez-Lösung*], *gruppenspezifische Sprache [*z.B. bei zielgruppenspezifischer Werbung*], *kontext(z.B. firmen-)abhängiges Synonymenverstehen [*Speech = Sprache*],

Syntax: Ellipsenverstehen [*Wer von beiden?*], *Satzwertige Antworten und deren Skopus [*Ja!*], Repairs und false starts [*Also, das ist, Wir haben eben kein Geld ...*], Kognitive und instrumentelle Grammatiken

Semantik: Referenzierung [*Dieses unbedeutende Problem*], Präsuppositionen [*Haben Sie aufgehört zu trinken?*], *Implikaturen und Allegationen [*Er ist an der Kunsthochschule tätig -> er ist Künstler*], Inferenzumgebungen [*Alltagswissen vs Physik*], Naive Physik [*Vakuum entweicht aus der Flasche*], *Komparativsemantik und andere relative Ausdrücke [*die bessere Lösung, das große Haus*], Prototypenbindung [*Ein protziges Auto*], Präsuppositionen [*Haben Sie aufgehört zu trinken?*], Vererbungshierarchien [*Ein Reh ist lebendig*], terminologische Selbstdeutigkeit [*Sicherheitsvor-*

schriftenformular], Expansion von Formeln [1/2], Interdependente Wortfelder [*gelehrt, gescheit, klug, clever*], Metaphorik,

Text und Diskurs: *Verstehen von Einbettungs- und Deklarationstechniken [*Typologie, Sequenz*], Komplexe Textkohärenz [*daher, im folgenden, dieselbe*], Formularverstehen [*Logik eines Steuerformulars*], *Transitivität von Text- zu Teiltextmerkmalen [*Korrektheit von Teiltexten*], Identifikation von Diskursgrammatiken [*z.B. RST*], Multiple Übersetzungsäquivalente, Abstracting,

Pragmatik: *Empfängerangaben und damit zusammenhängende Lese- und Verarbeitungsvorschriften [*An alle Mitarbeiter*], Geltungsbereich von Dokumenten [*Nur für den internen Gebrauch!*], *Indirekte Sprechakte [*Können Sie das bitte noch einmal wiederholen?*], Grice'sche Maximen [*Sei relevant!*], *Wahl von Stilhöhe oder Vertraulichkeit, Wechsel der semantischen Granularität

Sprechermodelle: *Kognitive Landkarten, Simulation des Spracherwerbs, Begriffsbildung [*L.Steels Experiment*], *Freiheit der Wortwahl, Freie Syntaxgenerierung, Turntaking, Grice'sche Maximen.

Nach Fraas 89 begünstigt man durch einen übertriebenen System- und Exaktheitsanspruch (damit auch durch Explizierung) "einseitig eine ganz bestimmte Lehrmeinung und benachteiligt und unterdrückt sogar alle anderen hiervon abweichenden Denkansätze". Die Forderung nach Explizitheit kann also innovationsfeindlich oder indoktrinär sein.

Fazit: Komplexes Sprachverstehen wird nicht durch Erklären, sondern durch impliziten Sprachgebrauch und Weltwissen erlernt, von vielen Sprechern muß es daher auch nicht rationalisiert werden oder kann es gar nicht (Typisches Beispiel: Präsuppositionen). Prinzipiell sind aber viele dieser Phänomene durch Regeln oder Lernverfahren darstellbar. Es kann auch zwischen Sprechern Einvernehmen über die Semantik und Pragmatik impliziter sprachlicher Äußerungen erzielt werden.

3. 2. *Unbewußtheit*

Prominentestes Beispiel ist wohl die *Nativitätsannahme Chomskys. Von computerlinguistischer Modellierung ist diese weit entfernt.

Des weiteren: *Sprecherstandpunkt (z.B. Origo-Implizierung), *Komplexe Partnermodelle (oberhalb von belief-Modellen) und alle *Stilphänomene (oberhalb von Wortwahl und grammatischen Parametern); die aktive Generierung gehört wohl auch in

diese Gruppe. Es gibt z.B. keine Vorhersagbarkeit und keine implementierten Modelle von initialen Sprechhandlungen.

Im Unterschied zu implizitem Sprechen kann zwischen Sprechern gewöhnlich kein Einvernehmen über die Semantik und Pragmatik sprachlicher Äußerungen dieser Klasse erzielt werden.

Die *Sprachgeschichte und der *Sprachwandel sind außerdem Themen, deren im Prinzip chaotisches Verhalten häufiger festgestellt wurde. Der Begriff der *Sprachgemeinschaft (oberhalb eines Code-sharing) entzieht sich ebenso der Rationalisierung wie *Modewörter oder der *Appeal von künstlichen Firmen- und Produktnamen.

Ein immer wichtiger werdendes Problem der Computerlinguistik ist die Architekturfrage. Dieses Thema hat in der Linguistik so gut wie kein Korrelat, da die Linguistik höchst selten in prozeduralen Modellen denkt.

Es geht um das Systemdesign einer Sprachverarbeitung, die Verstehen auf den unterschiedlichsten Ebenen konstituieren und dann in die anderen Ebenen propagieren kann. Verstehen entweder durch Schlüsselwörter oder durch nonverbale Clues oder durch genaue grammatische Analyse oder durch Situationsparameter werden beim menschlichen Verstehen mühelos vom selben System bewältigt und sozusagen "unterwegs" gegeneinander evaluiert. Die Pragmatik kann per Erwartung top-down das Hörverstehen steuern (oder sich sogar gegen die Höranalyse entscheiden) und die Höranalyse bottom-up für eine kumulierende kompositionelle Sinnkonstitution Werte belegen. Oberflächliches Verstehen ist in derselben Architektur möglich wie detaillierte konzeptuelle Exegese, und alle diese Prozesse menschlichen Verstehens haben any-time-Eigenschaft: Sie liefern (nach einem Offset) zu jedem Abbruchzeitpunkt konsistente Werte. Fast alle formal zufriedenstellenden Modelle der Computerlinguistik haben diese Eigenschaft nicht.

Nur mit solchen Systemen lassen sich aber genau genommen Ellipsen, Abbrüche, Unterbrechungen (Ins-Wort-fallen), False starts und ähnliche Phänomene modellieren. Bei der Maschinellen Übersetzung kristallisiert sich dieses Problem um den Begriff der variablen Analysetiefe.

Befriedigende Modelle für diese dem Sprecher unbewußte Fähigkeiten gibt es bisher in der Computerlinguistik nicht. Klassische Programmiermethoden sind zu sequentiell,

während neuronale Netze die Komplexität des Verstehens auf unterschiedlichen Ebenen (noch) nicht beherrschen können.

*Erinnern und Vergessen als eine ganze entscheidende sprachökonomische Eigenschaft von Sprechern und Hörern ist bisher psychologisch sehr breit, linguistisch nur oberflächlich und in der Computerlinguistik noch gar nicht gearbeitet worden. Aber gerade bei technischer Kommunikation spielt die Einbeziehung von Vergessen und Erinnern eine zentrale Rolle. Effiziente Partnermodelle natürlicher Sprecher beziehen das mögliche Vergessen mit ein oder bieten Gedächtnisstützen an. Mir bekannte Partnermodelle in der Computerlinguistik sind zwar vom Paradigma der Wissenskopien zugunsten von partiell geteiltem Wissen abgerückt, aber Vergessensprozesse spielen bisher keine Rolle. Daß man vergißt, ist jedem Sprecher absolut klar, was man vergessen hat, naturgemäß nicht. In industriellen Kommunikationsformen (von Gesprächsnotizen bis Marketing und Werbung) wären Vergessens- und Erinnerungsmodelle hochwichtig.

Auch weitere psycholinguistisch bereits beschriebene Phänomene unbewußten Sprachverhaltens sind in der Computerlinguistik schwer modellierbar: z.B. *Nicht-Rekonstruierbarkeit des Input nach dem Verstehen. *Nichtsequentialität der Generierung, Koexistenz unterschiedlichen Sprachwissens und -gebrauchs, natürliche Lexikonorganisation, multimodale Repräsentation und Inferenz.

Fazit: Viele sprachlich kommunikative Handlungen funktionieren, ohne daß sie exakt beschreibbar sind und ohne daß sie von den Beteiligten wahrgenommen werden. Am schwierigsten sind die Verstehensmethoden (Architektur, Kontrollfluß, Saliency) zu rationalisieren.

3.3. *Vagheit*

Klassische Arbeitsthemen sind:

Illokutive Unbestimmtheit: Hier ist der Sprechakttyp nicht eindeutig festgelegt. [*"Wir werden im nächsten Jahr unsere Einnahmen um 15% steigern!" (Versprechen, Warnung, Vermutung oder Scherz)*]. In der Computerlinguistik verschiedene Systeme zur Dialogaktzuweisung

*Uninformativität: Die Aussage ist so allgemein, daß sie keine sinnvolle Reaktion erlaubt. [*Bei "Patienten haben Krankheiten" würde man nach einer übertragenen Bedeutung suchen*].

Homonymie: Wörter wie "Schloß" oder "Bank" haben mehrere, nicht als zusammenhängend verstandene Bedeutungen. Implizites Verstehen muß für die Desambiguierung herangezogen werden

*Polysemie: "Gericht" ist ein polysemer Ausdruck in dem Sinne, daß er den Vorgang "zu Gericht sitzen", aber auch das Gerichtsgebäude, die Institution oder ein Gremium meinen kann. In der Computerlinguistik wird meist über künstliche Selektionsmerkmale desambiguiert.

Syntaktische Ambiguität: "*Verschrauben Sie den Schwenkarm mit der Führung am Gehäuse*" kann entweder bedeuten, daß der Schwenkarm eine Führung hat oder aber das Gehäuse. Das Problem ist in der Literatur bekannt als "PP-attachment". In eher semantische Richtung geht das Beispiel: "Alle Ärzte haben einen Feind". Dort ist der Skopus von "alle" und "ein" unklar (jeweils einen oder alle zusammen einen einzigen).

Referentielle Vieldeutigkeit: Hier sind besonders deiktische Ausdrücke als potentiell mißverständlich zu nennen. Außerhalb der originären Situation können indexikalische Ausdrücke wie "ich", "hier", "auf der vorigen Seite" oft nicht eindeutig zugeordnet werden.

Elliptische Vieldeutigkeit: "Die Firma hat geliefert" ist ohne den Kontext einer konkreten Bestellung nicht eindeutig. Mehr noch tritt die Schwierigkeit auf, wenn elliptische Nachfragen aus dem Kontext gerissen werden: Zum Verständnis von "den auch?" muß außer der Pronomenauflösung die Aktion des Vorgängersatzes verfügbar sein. Eine Ankündigung: "Tropenhölzer demnächst erheblich teurer" läßt offen, ob gegenüber einheimischen Hölzern oder gegenüber dem bisherigen Preis verteuert wird. Diese Mehrdeutigkeit ist typisch für den normalen Gebrauch von Steigerungsformen. Im Zweifelsfalle muß also immer die Bezugsgröße hinzugenommen werden.

Metaphorische Doppeldeutigkeit: "Blatt" kann in bestimmten Kontexten sowohl für eine Zeitung als auch nur eine Seite davon benutzt werden.

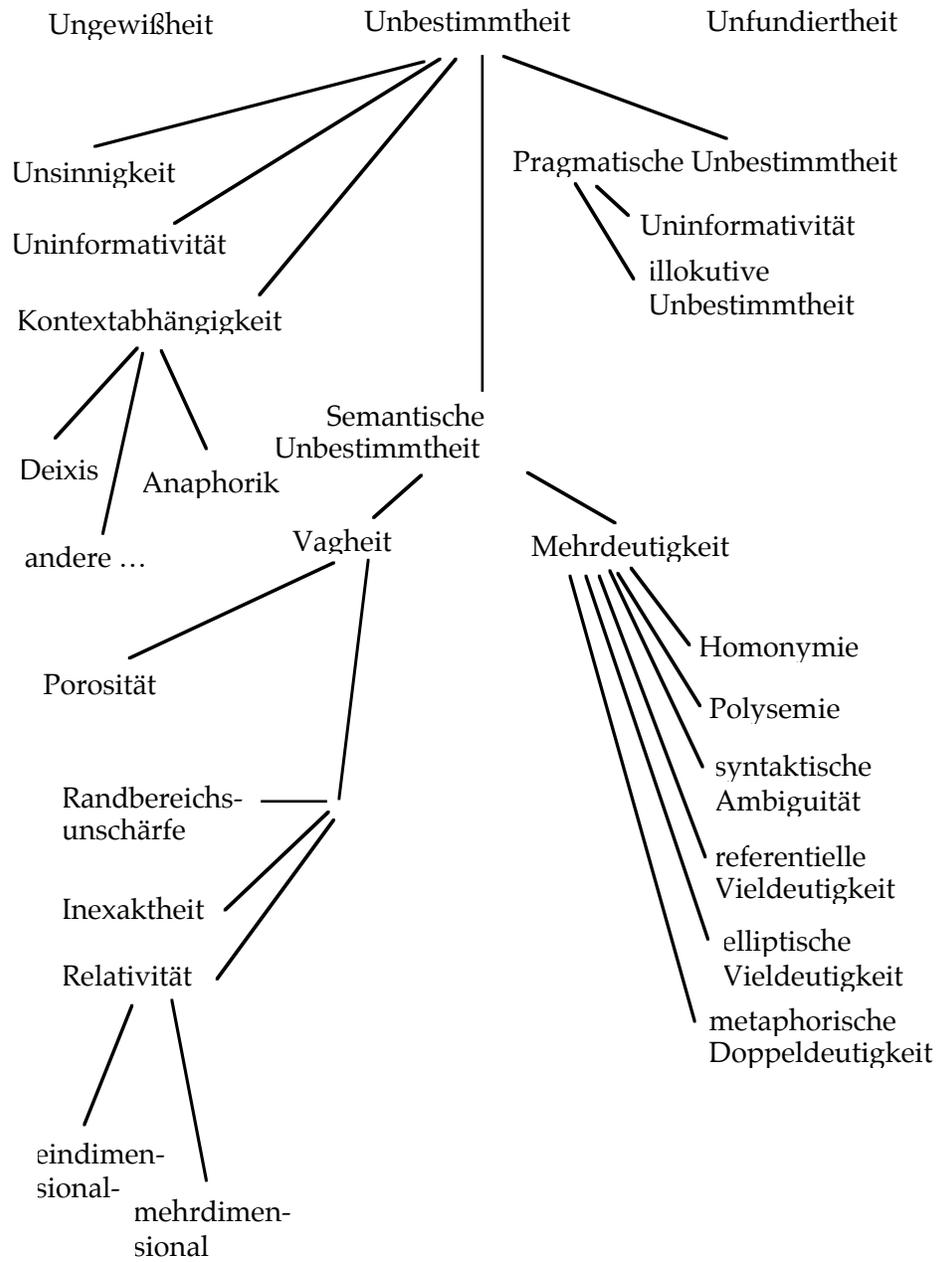
Porosität: "*Ist eine Substanz, die in allen chemischen und physikalischen Eigenschaften mit Gold übereinstimmt, dazu aber eine unbekannt Strahlung emittiert, Gold?*" (Pinal 1980, 14) In Forschung und Entwicklung kommen solche Fälle häufig vor. Die existierende Begrifflichkeit und damit die Terminologie eines Gebiets wird bei zunehmender Erkenntnis zwangsläufig porös. Der Begriff deckt also nicht mehr alle Erkenntnisse ab. Das macht sich darin bemerkbar, daß wichtige Forschungsschritte

in der Regel nicht nur zu einer reinen Auffächerung des Gebiets führen, sondern zu einer grundsätzlichen Reorganisation der Terminologie auch auf höheren Ebenen. Mit anderen Worten: Durch Forschung werden Fachbegriffe porös und entsprechend muß die Terminologie auch für die Oberbegriffe rekursiv aufsteigend neu festgelegt werden.

Relativität: "groß" ist für eine Nadel etwas anderes als für ein Sonnensystem. An absoluten Meßdaten läßt sich der Ausdruck "groß" sicher nicht vereinheitlichen, sondern eher an Typen oder typischen Handlungssituationen.

Inexaktheit: Adjektive wie "rund" oder "rechteckig" sind nur in Alltagszusammenhängen implizit exakt. Unter dem Mikroskop mag ein runder Gegenstand alles andere als rund sein. Dasselbe gilt für die meisten Zahlenangaben: "Der Sperrbügel ist 1,20 m breit" ist sicher nicht exakt, wenn man zehnstellige Maße angeben muß. Die Aussage: "Die Firma hat uns ein neues Angebot geschrieben" ist nur dann exakt, wenn über dem Schreiben genau steht: Neues Angebot.

Randbereichunschärfe: "Komplexe Sachverhalte, speziell Prozeßabläufe, lassen sich durch einfache Graphen nicht mehr ohne weiteres darstellen, hier bieten z.B. Petri-Netze eine geeignete Methode" Eine solche Aussage ist in Bezug *auf Komplexe Sachverhalte, einfache Graphen* und *ohne weiteres* randbereichunscharf: Was genau noch darunter gefaßt werden kann, ist nicht festlegbar.



Pinkals Typologie der Vagheit

Ein Mißverständnis ist weit verbreitet: Vagheit behindere die Kommunikation. Das ist nicht einmal als Tendenzaussage zutreffend, denn der Vorteil oder Nachteil von Vagheit bestimmt sich aus den kommunikativen Randbedingungen, nicht aus den Eigenschaften isolierter linguistischer Einheiten. So besteht oft sogar ein positiver Zusammenhang zwischen Verständlichkeit und Vagheit: In einer Fachdiskussion kann es ausgesprochen nützlich sein, zunächst zur Herstellung eines einheitlichen Kenntnisstandes sehr pauschal zu bleiben und dann erst im weiteren Verlauf der Diskussion zu immer präziseren und weniger vagen Formulierungen überzugehen.

Gerade Arbeitsgespräche in Gruppen gehen typischerweise so vor, daß ausgehend von sehr allgemeinen Aussagen die Redeweise immer enger und durch Vereinbarungen ("Ich nenne das jetzt einmal die Schmidt'sche Lösung 2") immer exakter wird. Diese trichterartige Vorgehen innerhalb derselben Sprache wird gerade ermöglicht dadurch, daß auch die Fachsprache unscharfe, vage oder unexakte Ausdrücke zuläßt und zwischen verschiedenen Exaktheitsebenen virtuos gewechselt werden kann. Man könnte die Vermutung äußern, daß innovatives problemlösendes Sprachverhalten genau mit dieser Fähigkeit flexibler Vagheit verbunden ist. Oft macht auch der Fachmann die Erfahrung, daß, befindet er sich auf einer zu tiefen Ebene der Exaktheit, die Sicht auf Innovationen geradezu verstellt ist.

4. Methoden der Bearbeitung

4.1. zum Verstehen von Impliztheit

1. vorherige prozedurale Explizierung (z.B. Ausnahmebehandlung für Ellipsen in der Syntax)
2. vorherige deklarative Explizierung (z.B. Semantische Repräsentation zur Transitivität von Begriffshierarchien ,
3. Neuronale Netze (z.B. unscharfe Klassifizierungsangaben)
4. Lernverfahren (z.B. Übersetzungsäquivalente durch Alignment, Bestimmung von Dialogakten durch unüberwachtes Lernen

4.2. zur Ein- und Abgrenzung

Es gibt in der Computerlinguistik keine klare Ein- oder Abgrenzung des Themengebiets. Allenfalls M. Pinkals Definitionsumgebung von Expliztheit/Impliztheit bzw. Vagheit grenzt diese Begriffe gegen ähnliche ab, wobei das wissenschaftlich solide ist, im Bewußtsein eines Sprechers aber kein Korrelat hat.

Unter Ausdrücklichkeit verstehen wir die redundante Nennung des gleichwohl vorhandenen Kontexts, um dessen Bindung gegenüber Mißverständnis, Mißinterpretation oder Übersehen zu sichern ("Es sei ausdrücklich betont, daß Ausdrücklichkeit nicht synonym mit Expliztheit ist")

Als Vollständigkeit bezeichnen wir neben der (syntaktischen oder semantischen) Lückenlosigkeit die angemessene Expliztheit eines Textes. Ein vollständiger Text ist also ein in einer speziellen Kommunikationssituation ausreichend expliziter (oder: nicht zu impliziter) Text.

Präzise ist ein Text, wenn er nicht vage ist. Vage kann ein Text dadurch werden, daß er in seinem Wahrheitswert trotz genauer Kenntnis der Umstände nicht festlegbar ist. Das Ideal der Präzision von Fachtexten wird also erreicht durch Entfernung vager Ausdrücke, nicht durch Explizierung (Pinkal 1985, 48). Allerdings ist aus der Definition auch ersichtlich, daß vage Texte oft unvermeidlich sind.

Ein Text höherer Spezifität (geringerer Allgemeinheit) entsteht, wenn seine Aussagen an eine niedrigere Klasse von Entitäten gebunden wird, z.B. Hund -> Dackel, Schiff -> Segelschiff -> Segelschulschiff.

Explizit machen kann man vor allem den Kontext einer bivalenten (wahren oder falschen) Aussage. Dennoch kann man auch vage Aussagen explizieren. Man kann z.B. den Charakter der polyvalenten Semantik klären ("Unter dieser Bedingung ist das so, unter einer anderen nicht so, das kommt drauf an, das ist nicht entscheidbar", etc).

M. Pinkals Definitionsumgebung von Vagheit:

Unsinnigkeit im umgangssprachlichen Gebrauch ist eine "unsinnigen Frage", also solche Äußerungen, die Präsuppositionsverletzungen oder Sortenverstöße enthalten.

[*"Hat Abteilung 2b aufgehört, für dieses Produkt zu werben?" (aber Abt. 2b ist der Einkauf), "Ist das Lösungsmittel groß oder klein"*]

Derartige Äußerungen oder Fragen sind nicht vernünftig direkt zu beantworten. Man würde in beiden Fällen, in der Annahme, daß eine kommunikativ wirksame Äußerung beabsichtigt ist, mit einer Rückfrage oder Richtigstellung reagieren.

Ungewißheit, z.B. Aussagen, die erst in der Zukunft liegen ("*Unsere Firma erzielt im nächsten Jahr 5% mehr Umsatz*") oder aus Gründen ungeklärter oder nicht klärbarer Fakten

Allgemeinheit. Spezifischer im Gegensatz zu präziser ist ein Ausdruck, in dem niedrigere Klassen verwendet werden: Hund - Dackel, Schiff - Segelschiff. Dasselbe gilt von dem Gegensatzpaar "vage/allgemein". Vage ist: nicht präzise und allgemein ist: nicht spezifisch

Fazit: Als Bearbeitungsmethoden werden prozedurale oder deklarative Explizierung, neuronale Netzmodelle oder Lernverfahren angewendet. Es gibt in der Computerlinguistik keine klare Ein- oder Abgrenzung des Themengebiets. Die nach Pinkal außerhalb liegenden Randbegriffe liegen sicher auch außerhalb des Skopus von "Nichtexplizitem Wissen".

4.3. zur Klassifikation

Verschiedene Arten der Nicht-Expliztheit kann man nach folgender Ad-hoc-Schematik unterscheiden. Eine wissenschaftlich begründete oder breiter diskutierte Typologie liegt nicht vor:

1. Unabhängiges implizites Wissen
 - (a) (noch) nicht explizites regelhaftes Wissen Methode: Explizierung
 - (b) (noch) nicht entdecktes Wissen Methode: Data Mining
 - (c) strukturell nicht explizierbares Wissen Methode: Neue Repräsentationen, Lernverfahren, Neuronale Netze
 - (d) Wissen der Ebenen 1 und 2 Methode: Explizierung durch Hebung

2. mitgewußtes Wissen
 - (a) Assoziationen Methode: Neue Repräsentationen Neuronale Netze
 - (b) Konnotationen Methode: ??
 - (c) Präsuppositionen Methode: Neue Repräsentationen
 - (d) Implikaturen Methode: ??
 - (e) Allegationen Methode: Neue Repräsentationen
 - (f) Beziehungswissen Methode: Meta-Wissensbasis

3. strukturell nicht regelhaftes Wissen
 - (a) nicht wahrheitswertfähig Methode: Neuronale Netze
 - (b) chaotische Prinzipien Methode: Neuronale Netze
 - (c) überlappendes/widersprüchliches Wissen Methode: Neue Repräsentationen (Kontexte, constraints)

- | | |
|-------------------------------------------|-----------------------------------------------------|
| 4. (überwiegend) Unbewußtes Wissen | |
| (a) sprach- evolutionäres Wissen | Methode: Wissenschaft, Statistik,
Theoriebildung |
| (b) Kontrollstruktur verstehender Systeme | Methode: ?? |
| (c) Kommunikationsverhalten | Methode: Lernverfahren |
| (d) Multiple Repräsentation | Methode: Medienunabhängige
Repräsentation |

Fazit: Eine anerkannte Typologie gibt es nicht. Der hier vorgestellte ad-hoc-Versuch orientiert sich an Radermachers Ebenenmodell ebenso wie an der Dimension abhängig/unabhängig.

5. National und international führende Wissenschaftler des Gebiets

Es gibt keine Forscher oder Forschungseinrichtungen, die sich ausdrücklich diesem Thema gewidmet haben. Einige wichtige Forschungen in der Vagheitsfrage wurden in der Vergangenheit in Deutschland von Manfred Pinkal, Saarbrücken ausgeführt. Rieger, Aachen hat früh einige eher formale Beiträge zur Vagheit in Informatiknähe geliefert. Die Vagheitsproblematik wird heute eher als ein Repräsentationsproblem, denn als kognitives Prinzip gesehen. Nicht-explizites Wissen wird in der Kognitionswissenschaft vorwiegend im Bereich räumliche Orientierung angesiedelt, dort spielt aber die (computer-) linguistische Frage der Erzeugung sprachlicher Performanz eine untergeordnete Rolle. NN-Methoden und hybride Lösungen werden etwa von Stefan Wermter (jetzt Univ. of Sunderland) angewendet und ausführlich beschrieben.

Fazit: Nicht-Explizites Wissen wird nicht als einheitliches computerlinguistisches Forschungsfeld angesehen und daher auch nicht schwerpunktmäßig an bestimmten Institutionen bearbeitet.

Architekturfragen sind im BMBF-Projekt ASL intensiv exploriert worden. Computerlinguistische Modelle sind dabei von G. Görz, W. v.Hahn und W. Menzel in größerer Breite bearbeitet worden.

6. Computerlinguistische Forschungsprojekte über Implizites Wissen

Es gibt m.W. zur Zeit kein Projekt, das sich genau diesem Thema widmet. Einige der oben genannten Themen werden erforscht im Zusammenhang von sprachlichen Agenten (SFB Bielefeld). Vor allem multimodale Integration, multiple Repräsentation und der Komplex "Sprache und Handlung" werden dort intensiv erforscht. Verschiedene Saarbrücker Projekte (Wahlster) haben sich erfolgreich mit der Frage multimodalen Wissens beschäftigt. Das BMBF-Projekt ASL hatte die Fragen nichtsequentieller Sprachverstehenssysteme thematisiert

Fazit: Es gibt m.W. kein Projekt, das sich genau dem computerlinguistischen Aspekten des Themas "Nicht-Explizites Wissen" widmet.

7. Forschungsbedarf

Die oben genannten offenen Fragen stellen eine kaum zu übersehende Forschungsarbeit in Linguistik und Computerlinguistik dar. Der Forschungsbedarf der Computerlinguistik ist besonders hoch, weil sich neben den Fragen expliziter Modellierung (oder sonstiger Beherrschung nicht-expliziter Phänomene) noch die Frage stellt, wie Mensch und Maschine in Zukunft in den Bereichen nicht-expliziten Wissens zusammenarbeiten sollen, eine Frage, die in absehbarer Zeit beantwortet werden muß.

In vielen Gebieten der Sprachverarbeitung sind viele der oben genannten offenen Fragen eine eklatante Technologiehürde, von der man zur Zeit nicht weiß, wie und wann man sie wird überspringen können. Auf dem Technologiesektor sind beispielhaft zu nennen:

- Qualitätsmessung maschineller Übersetzung und maschinellen Dolmetschens,
- Multimodale Repräsentation und Transformation von modalem Wissen,
- Architekturen sprachverstehender Systeme,
- Speziell: Speech-Language-Integration,
- Variable Analysetiefe von Sprachverarbeitungsprozessen,
- Vergessens- und Erinnerungsmodelle (von Gesprächsnotizen bis zur Werbung)
- Dynamisch wachsende sprachverarbeitende Systeme (von Terminologiebank bis Abstracting)
- Integration von expliziten sprachverstehenden Systemen in die implizite menschliche Kommunikationsumgebung,
- Systematisch multiple Repräsentation aspektgebundenen Wissens in Firmen,
- Integration von expliziter Terminologieschöpfung in die implizite dynamische Arbeitsumgebung;

und mehr theoretisch:

- Was ist nicht-explizites linguistisches Wissen?
- Was ist Ganzheitlichkeit des Sprachverstehens?
- Sprachliches Erinnern und Vergessen,
- Was löst sprachliche Aktivität aus (selbständige sprachliche Agenten)?
- Spontanes Metapherngenerieren (z.B. Räumliche, Szenario- und Personmetaphern)

Fazit: Der angesprochene Forschungsbereich "Nicht-explizites Wissen" wird gerade erst in seinen Teilen und seiner Bedeutung sichtbar. Viele Themen sind theoretisch abzuklären und gleichzeitig sind eine Reihe dringender Technologiehürden zu beseitigen.

8. Forschungsperspektiven

In Forschungsgruppen zusammen mit Psycholinguisten, Betriebspsychologen, Sprachtechnologien und Informatikern muß das Thema "Nicht-explizites Wissen" in der Computerlinguistik erforscht und technologisch umgesetzt werden.

Es ist das in der Wissenschaftsentwicklung jetzt anstehende Thema nach einer neunzig-jährigen Ära des expliziten Paradigmas aus dem Strukturalismus. In den heute zwischen Informatik, Linguistik, Psychologie, Philosophie und Betriebswirtschaft neu entstehenden Forschungsgebieten spielt das nicht-explizite Wissen eine herausragende Rolle. Gerade moderne betriebswirtschaftliche Themen wie "die Börsendynamik", "Kommunikative Eigenschaften von Führungspersönlichkeiten", "Trendorientierung" "Corporate Identity" oder "Erkennungswert von Werbung", um nur wenige zu nennen, weisen auf die eminente Bedeutung nicht-expliziten Wissens hin.

Alle oben aufgeführten Forschungsprojekte in diesem Themenfeld haben die interessante Eigenschaft, daß sie computerlinguistische FuE , wie auch kognitionswissenschaftliche Theorie und technologische Anwendung in organischer Weise verbinden (müssen).

Auch von den informatischen Paradigmen her ist die Entwicklung von den symbolischen, rein regelbasierten Verfahren hin zu hybriden Systemen unter Einschluß neuronaler Netze oder von Lernverfahren ein Hinweis darauf, daß eine Integration der unterschiedlichen Methoden der Wissensverarbeitung in bearbeitbare Nähe gerückt ist.

Fazit: Wir befinden uns am Ende einer Ära der Erforschung von vorwiegend positivem, explizitem, bewußtem und persistentem Datenwissen. Die nächste Runde muß das nicht-explizite Wissen bearbeiten und dazu ein neues Paradigma entwickeln, das nicht nur der Explizierung solchen Wissens (und damit wieder der Anwendung des alten Paradigmas) dient: Neue Methoden der Beschreibung und Handhabung müssen entwickelt werden, die heute erst in Konturen erkennbar sind. Forschungsprogramme in diesem Themengebiet sind für den Fortschritt in der Theorie wie in der Technologie gleich entscheidend. Sie würden eigentlich erst den lange schon angekündigten Schritt von der Datenverarbeitung (der Verarbeitung expliziten Wissens) zur Wissensverarbeitung (der Einbeziehung nicht-expliziten Wissens) vollziehen.

9. Literatur

- AECMA/AIA 1989= Association Européenne des Constructeurs de Matériel Aérospatial and Aerospace Industries Association: Simplified English. A Guide for the Preparation of Aircraft Maintenance Documentation in the international Aérospatial Maintenance Language, Chance 5. Paris 1989.
- Ballmer/Pinkal 1983= Thomas T. Ballmer und Manfred Pinkal: Approaching Vagueness. Amsterdam/New York/Oxford 1983. (North Holland Linguistik Series 50)
- Berman/Hestvik 1992 = Steve Berman, Arild Hestvik: Proceedings of the Stuttgart Ellipses Workshop 1992. Stuttgart 1992. (Sprachtheoretische Grundlagen für die Computerlinguistik. Papiere des SFB 340, Bericht 29).
- Biere 1989 = Bernd Ulrich Biere: Verständlich-Machen. Hermeneutische Tradition - Historische Praxis - Sprachtheoretische Begründung. Tübingen 1989.
- Fraas 1989 = Claudia Fraas: Terminologiebetrachtung - Sache der Fachleute oder der Linguisten? In: Fachsprachen 11. 1989. 106-113 .
- Görz 1996 = Günther Görz., Kessler, M., Spilker, J., Weber, H.: Research on Architectures for Integrated Speech/Language Systems in Verbmobil. Proceedings of COLING-96, August 1996, Copenhagen [ps.gz]
- Grice 1975 = H.P. Grice: Logic and Conversation. In: P. Cole ,J.J.Morgan (Hrg.): Syntax and Semantics 3.41-58. New York 1975.
- v.Hahn 1992 = Walther v.Hahn: Von der Verknüpfung zur Integration: Kontrollstrategie oder kognitive Architektur. Görz, Günther (Hrsg.), Proceedings of KONVENS92. Berlin 1992. S. 1 - 10.
- v.Hahn 1995 = Walther v.Hahn und Claudius Pyka: System Architectures for Speech Understanding and Language Processing. In: G. Heyer und H. Haugeneder: Language Engineering. Wiesbaden 1995. S. 139 - 157
- v.Hahn 1998a = Walther v.Hahn, Das Postulat der Expliztheit für den Fachsprachengebrauch. In: H.E. Wiegand, L. Hoffmann (Hrsg.), HSK Fachsprachen. Berlin 1998. S.
- v.Hahn 1998a = Walther v.Hahn, Vagheit bei der Verwendung von Fachsprachen In: H.E. Wiegand, L. Hoffmann (Hrsg.), HSK Fachsprachen. Berlin 1998. S.
- Hörmann 1970 = Hans Hörmann, Psychologie der Sprache. Berlin 1970
- Korn 1982 = Monika Korn: Untersuchungen zur graphischen Darstellung wissenschaftlicher Theorie. Dargestellt am Beispiel von Kommunikationsmodellen. Hamburg 1982. Manuskript
- Menzel 1998 = Wolfgang Menzel, Ingo Schröder (1998) Model-based Diagnosis under Structural Uncertainty. Proc. 13th European Conference on Artificial Intelligence, ECAI'98, Brighton UK, p. 284-288.
- Neubert 1986 = Albrecht Neubert: Translatorische Relativität. In: Snell-Hornby 86. 85-105.
- Möller 1999 = Jens-Uwe Möller, DIA_MOLE: Modellierung gesprochen-sprachlicher Dialoge unter Zuhilfenahme eines maschinellen Lernverfahrens. St. Augustin 1999.
- Pinkal 1980 = Manfred Pinkal: Semantische Vagheit: Phänomene und Theorien. In: Linguistische Berichte 70. 1980. 1-26. und 72. 1981. 1-26.
- Pinkal 1985a = Manfred Pinkal: Logik und Lexikon - Die Semantik des Unbestimmten. Berlin, New York 1985.
- Pinkal 1985b= Manfred Pinkal: Kontextabhängigkeit, Vagheit, Mehrdeutigkeit. In: Christoph Schwarze und Dieter Wunderlich (Hrsg.): Handbuch der Lexikologie. 27-63. Königstein/Ts. 1985.
- Pinkal 1989= Manfred Pinkal: Imprecise Concepts And Quantification. Hamburg 1989.
- Pinkal 1991= Manfred Pinkal: Vagheit und Ambiguität. In: Arnim von Stechow/Dieter Wunderlich: Semantik. Handbuch der Sprach- und Kommunikationswissenschaft, Band 6. S.250-269. Berlin, New York 1991.

- Radermacher 1997a = Franz Joseph Radermacher: Intelligenz - Kognition - Bewußtsein: Systemtheoretische Überlegungen. Technische Möglichkeiten, philosophische Fragen. In; Stadelhofer, C. (Hrsg.), Interdisziplinäre Beiträge zur Kommunikation und zum Mensch-Technik-Verhältnis. S. 146 - 193
- Radermacher 1997b = Franz Joseph Radermacher: Komplexe Systeme und lernende Unternehmen. Manuskript 1997.
- Rieger 1989= Burkhard B., Unscharfe Semantik: Die empirische Analyse, quantitative Beschreibung, formal Repräsentation und prozedurale Modellierung vager Wortbedeutungen in Texten. Frankfurt a.M. 1989
- Roelcke 1991 = Thorsten Roelcke: Das Eineindeutigkeitspostulat der lexikalischen Fachsprachensemantik In: Germanistische Linguistik 19. 1991.194-208.
- R° ,u°,z ič ka 1975 = Rudolf R° ,u°,z ič ka: Sprachwissenschaft und Wissenschaftssprache.. In: Albrecht Neubert/Rudolf R° ,u°,z ič ka (Hrsg.): Sitzungsberichte der Akademie der Wissenschaften der DDR.18-30. Berlin 1975.
- Schmitt 1986 = Peter A. Schmitt: Die "Eindeutigkeit" von Fachtexten: Bemerkungen zu einer Fiktion. In: Snell-Hornby 86. 252-282.
- Schröder 1988 = Hartmut Schröder: Fachtext und interkulturelle Autor-Leser-Kommunikation: Überlegungen zum Übersetzen wissenschaftlicher Fachtexte aus dem Bereich der Gesellschaftswissenschaften, In: C. Laurén/M. Nordman(Hrg.): From Office to School: Special Language and Internationalisation. 21-38. Clevedon, Philadelphia 1988.
- Snell-Hornby 1986 = Mary Snell-Hornby (Hrsg.): Übersetzungswissenschaft - eine Neuorientierung: Zur Integrierung von Theorie und Praxis.Tübingen 1986 (UTB 1415).
- Wermter1996 = Stefan Wermter, E. Riloff, G. Scheler (Ed). Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing Springer Verlag, Berlin, 1996
- Wichter 1988= Sigurd Wichter: Signifikantgleiche Zeichen. Untersuchungen zu den Problembereichen Polysemie, Homonymie und Vagheit auf der Basis eines kommunikativen Zeichenbegriffs am Beispiel deutscher Substantive, Adjektive und Verben.Tübingen 1988.
- Wolski 1980 = Werner Wolski: Schlechtbestimmtheit und Vagheit - Tendenzen und Perspektiven: Methodologische Untersuchungen zur Semantik.Tübingen 1980.