

Harald H. Zimmermann
Universität des Saarlandes, Saarbrücken

WORTRELATIONIERUNG IN DER SPRACHTECHNIK Stilhilfen, Retrievalhilfen, Übersetzungshilfen

In: W. Gödert; P. Jaenecke; W. Schmitz-Esser (Hrsg., 1992): Kognitive Ansätze zum Ordnen und Darstellen von Wissen. Fortschritte in der Wissensorganisation (INDEKS Verlag) 2 (1992), S. 287-296

1. Einleitung

Es ist bekannt, dass jedes Sprachsystem - bezogen auf die reale Welt bzw. auf menschliche Vorstellungen von ihr - unterschiedliche Benennungen verwendet. Es ist ferner vorausgesetzt, dass die begrifflichen Vorstellungen einer Sprachgemeinschaft bzw. eines Einzelnen in dieser Gemeinschaft, die mit einer Benennung (d.h. einer Äußerung, einem Wort) verbunden werden, sich unterscheiden können. Damit überhaupt Kommunikation und Wissenstransfer zustandekommt, muss die Benennung - bzw. allgemeiner: der Code, den die Kommunikatoren benutzen -, vom jeweiligen Kommunikationspartner allerdings verstanden, d.h. entschlüsselt und den eigenen Vorstellungen zugeordnet werden. Das Verfahren, das der Mensch dabei benutzt, ist auch heute noch weitgehend als eine Black Box anzusehen.

Es ist jedoch sicher, dass dieser Vorgang sehr komplex und durch sprachsystematische Regeln (Morphologie, Syntax, Sprachsemantik ...) *allein* nicht vollständig zu beschreiben ist. Damit Kommunikation bzw. Wissenstransfer glücken, müssen bei der Zuordnung beispielsweise auch Verfahren wirken, die Toleranzen und Vagheiten zulassen, also trotz äußerlicher Differenzen oder auch inhaltlicher Abweichungen noch zu einem "Verstehen" führen.

Der Verstehens- und Lernprozess ist in erster Linie begrenzt von der Wissensbasis, d.h. den Kenntnissen des Betroffenen. Es gibt andererseits Phänomene, die auf fehlendes bzw. unzureichendes Zuordnungswissen zurückzuführen sind. Die fehlende Kenntnis einer fremden Sprache, einer Mundart sind hierfür Beispiele.

Den grammatisch inkorrekten Satz "Hast du genommen ein Bad?" versteht beispielsweise jeder (obwohl er auch Anlass zu einem bekannten Witz bietet); der Satz "Der Bofke alfanzt mit dem Schwiemel" wird wohl für die meisten (obwohl grammatisch korrekt und semantisch sinnvoll) unverständlich sein (nicht aber seine Paraphrasierung: "Der Narr treibt seine Possen mit dem Trunkenbold").

Jedes Sprachsystem bietet seinem (menschlichen) Nutzer schon nach kurzer Lernphase (in der der sprachspezifische Regelmechanismus *mit* gelernt wird) brillante Möglichkeiten, Benennungen zu erzeugen bzw. zu entschlüsseln.

In der Regel steht der Sprachnutzer in einer Art Bezeichnungs- oder Benennungszwang: Ein (neues) Produkt braucht einen Namen, eine Forschungserkenntnis muß benannt werden, eine Handlung muss geschildert, eine Situation beschrieben werden: Das *sprachökonomische Prinzip* führt dazu, dass auch komplexe Situationen, Handlungen und Vorstellungen schließlich "verwordet" werden: Die Benennungen Kanzlerreise, Golfkrieg, Luftlandetruppe, Herrenoberhemd, Um-

weltschutz gehören hierher, aber auch einfache Wörter wie Geher oder Host sind hier anzuführen. Dass bei der Wortbildung gelegentlich Kapriolen geschlagen werden, zeigen Wörter wie Auto (aus einer Vorsilbe entstanden) und Bus (letztlich eine Wortendung).

Im Verlauf ihrer Benutzung sind Benennungen bzw. Bezeichnungen (die beiden Wörter werden im folgenden synonym verwendet) verschiedensten Einflüssen ausgesetzt:

- Sie veralten (mit ihrem Gegenstand z.B.).
- Sie erhalten eine neue äußere Gestaltung (Schreibform).
- Sie stehen neben anderen Benennungen (teilen sich also gleichsam die Sprachnutzer, z.T. in einem regelrechten Verdrängungswettbewerb (neueres Beispiel: Broiler Brathähnchen).
- Sie erhalten ein "Korsett" (z.B. über eine Definition bzw. graphische Darstellungen), v.a. in Fachsprachen.
- Ihre Bedeutung (d.h. das, was sie benennen) verändert sich, die Benennung bleibt jedoch bestehen.

Diese Aufzählung ließe sich beliebig fortsetzen; all dies sind Phänomene, mit denen beispielweise der Terminologe bzw. Lexikograph ständig konfrontiert ist.

Es ist hier nicht der Ort, allgemein auf die Aufgaben der Lexikographie einzugehen. Jedenfalls ist die Funktion der Normierung (Präskribierung) nicht die wesentliche, in keinem Falle die entscheidende Aufgabe. In der (traditionellen) Lexikographie wird man diesen Problemen - wie ich meine - zunehmend besser gerecht. Von einem *präskriptiven* Konzept geht man über zu eher *deskriptiven* Darstellungen, orientiert sich am täglichen Sprachgebrauch. Dies setzt allerdings im Grunde voraus, dass repräsentative Materialien (der gesprochenen wie geschriebenen Sprache) verfügbar sind; insgesamt wird jedoch nicht mehr (zumindest nicht mehr in so rigider Weise) der Lexikograph zum obersten Richter einer Sprachgemeinschaft. Ein für mich sehr prägnantes Beispiel ist die jüngste Einführung des Wortes "editieren" in den neuen (Universal-)DUDEN (1989; swv. "Daten in ein Terminal eingeben, ändern, löschen ...), obwohl es ja bereits das (allgemeine) Wort "edieren" gab. Alphabetisch ganz in der Nähe findet man übrigens die Einträge ECU / Ecu. Wer weiß allerdings noch, das Ecu die Bezeichnung für eine alte französische Silbermünze ist / war? Das in der deutschen Sprachnorm bislang nicht vorgesehene Apostroph beim sog. «sächsischen Genitiv" greift (aus dem Englischen übernommen) - wie man überall beobachten kann - heftig um sich (Beispiel: Karin's Kinderladen statt standardmäßig Karins Kinderladen).

Die Terminologie hat (als eine Art spezielle Lexikographie) zunächst zum Ziel, den Wortschatz eines Faches bzw. Fachgebiets zu sammeln und zur Nutzung bereitzustellen. Es ist jedoch unbestritten, dass *Fachkommunikation* und -wissenschaft präzise Begrifflichkeiten braucht, die sich in entsprechenden (einheitlichen) Benennungen niederschlagen sollten. Je eindeutiger sich ein Sachverhalt (v.a. in den Naturwissenschaften) darstellen lässt, desto eher wird man sich auch (als Nutzer) bereit finden, Standards (Normen) in der Benennung zu akzeptieren.

Die folgenden Überlegungen stellen die Bedeutung klassischer lexikographischer und terminologischer Arbeit nicht in Frage. Dies gilt zunächst grundsätzlich und allgemein, besonders aber für die "papierorientierte" Dokumentation. Bei der Anwendung *elektronischer* Speicherungs- und Distributionsformen erhält jedoch die terminologisch-lexikographische Arbeit einen neuen Aspekt. Da ein Teil der oberflächenorientierten Phänomene bereits "computerseitig" behandelt wer-

den kann, kann der menschliche Nutzer durch darauf ausgerichtete elektronische Wörterbücher und damit verbundene Verfahren v.a. von *trivialen* Problemen bei der Wahl bzw. Heranziehung von Benennungen entlastet werden, etwa beim Zugriff auf Datenbanken, aber auch bei der Textgenerierung.

Um die Darstellung der Möglichkeiten der Sprachtechnik nicht durch zu hohe Ansprüche zu belasten, wurde der zu behandelnde *praktische* Anwendungsbereich mit dem Attribut "Hilfe" versehen. Damit sollen weitergehende Verfahren (etwa Systeme zur vollautomatischen Übersetzung, zur Erkennung beliebiger gesprochener Sprache, zum automatischen Abstracting) nicht grundsätzlich in Frage gestellt werden. Sie bleiben aber außerhalb der Betrachtung.

Als Anwendungsbereiche dieser "Hilfen", für die ein größeres Marktinteresse vorausgesetzt wird und die sich mit der bestehenden Technik verwirklichen lassen, werden angesehen:

- a) wortbezogene Stilhilfen (v.a. bezogen auf Worterklärung und Synonymik)
- b) Retrievalhilfen beim Zugang zu textuellen Datenbanken
- c) Elektronische Wörterbücher als Hilfen bei der intellektuellen Übersetzung.

Dabei werden - am Tagungsthema gemessen - Begriffsbeziehungen (genauer gesagt: Beziehungen zwischen Benennungen) und deren Nutzungsmöglichkeiten in den Vordergrund gestellt.

2. Explizite und implizite Darstellungen von Beziehungen

Im folgenden werden explizite und implizite Darstellungen von Beziehungen als äquivalent betrachtet, sozusagen als unterschiedliche *Repräsentationsformen* des Sprachwissens. Das (elektronische) Lexikon ist nicht isoliert zu sehen, sondern steht mit dem Verarbeitungssystem (= Algorithmus) in Wechselwirkung bzw. ist Teil dieses Systems. Die "Relation" "IST NOMINATIV PLURAL ZU" kann beispielsweise explizit in einem Lexikon gespeichert sein (HÄUSER "IST NOMINATIV PLURAL ZU" HAUS) oder aber implizit in der paradigmatischen Information enthalten sein: HAUS gehört zur Flexionsklasse XY usf. Im letzteren Falle würde das System diese Beziehung analytisch erschließen.

Ein *Synonym* ist gerade dadurch gekennzeichnet, dass die Eigenschaften / Merkmale des repräsentierten Gegenstands usf. identisch sind.

Da es jedoch nicht immer möglich (bzw. angemessen) ist, die Merkmale alle aufzuführen, die es dann möglich machen, derartige Beziehungen in den *algorithmischen* Teil (oder interpretativen Regelteil) zu verlagern, wird häufig die lexikalische Variante genutzt: die Benennungen werden direkt durch eine Relation miteinander verknüpft: SAMSTAG "IST SYNONYM ZU" SONN-ABEND usf. Derartige "verkürzte" Verfahren sind später allerdings nicht mehr unbedingt nachvollziehbar, wenn sie dem Nutzer nicht einleuchten (er also Zuordnungsschwierigkeiten hat). Die "Erklärung" müßte nämlich gerade anhand der Merkmale erfolgen.

Im Grunde unterliegen alle bekannten Thesauri diesem Phänomen. Definition und Erläuterungen haben u.a. den Nebenzweck, diese Problematik zumindest praktisch gesehen zu umgehen. Besonders schwierig ist die Situation bei der Assoziationsrelation bzw. der "SIEHE-AUCH"-Relation.

Sieht man einmal davon ab, dass es zumindest problematisch, wenn nicht vielleicht unmöglich ist, Sprache so weit zu formalisieren, dass die "Relationierung" völlig anhand impliziter Merkmale erfolgt, so bleibt auch in einem solchen Falle das Problem der *Explikation* gegenüber dem menschlichen Nutzer. Sieht man nämlich von Grenzfällen ab (Beispiel: ein absoluter Laie will in einer Rechts- oder Chemiedatenbank recherchieren), so kann man mit Fug und Recht erwarten, dass der Nutzer die *explizite* Beziehungsdarstellung nachvollzieht (d.h. entscheiden kann, inwieweit sie für seine Belange nützlich ist).

Für Anwendungen in der Praxis erscheint es weitgehend unerheblich, inwieweit eine Beziehung zwischen Benennungen systemseitig anhand von Merkmalen oder durch explizite Lexikalisierung dargestellt wird. Hierbei spielen speicher- und rechenökonomische Faktoren, aber auch Kodierungskosten eine besondere Rolle.

Im folgenden wird auf speicher- und zugangstechnische Fragen (aus ähnlichen Überlegungen) nicht weiter eingegangen, obgleich auch diese aus einer Reihe von Gründen nicht trivial sind (Konsistenzfragen, Problem der Merkmalsvererbung, Portierbarkeit, gleichzeitige Nutzung in einem Netz, Zugangsgeschwindigkeit, Volumen, Modifizierbarkeit).

3. Anwendungsbeispiele für Wortrelationierungen

Die folgenden Ausführungen beschreiben ausschnitthaft, d.h. ohne Anspruch auf Vollständigkeit, Ansprüche und Anforderungen der "Praxis", die mit Hilfe von Wortrelationierungen behandelt bzw. gelöst werden können. Um den v.a. in der Dokumentation relativ festgelegten Terminus "Thesaurus" nicht zu belasten, wird in der Regel die Bezeichnung "Relationenwörterbuch" benutzt. Die Differenzierung der "Relationen" ist vorwiegend anwendungsbezogen. Allgemeine Zielsetzung ist, zu zeigen, welche Vielfalt in der Nutzung derartiger Relationen sich heute bereits auftut.

Ein Relationseintrag in einem Relationenwörterbuch wird formal wie folgt beschrieben:

REL = <RELATIONSEINGANG (RE-ZUSATZ) > <RELATIONSTYP> <RELATIONSAUSGANG (RA-ZUSATZ) >

Es ist dabei darauf hinzuweisen, dass bei der Relationierung zusätzliche Merkmale über die graphematische Zeichenkette hinaus benutzt werden müssen (obligatorisch sind), wenn die Graphemkette allein nicht eindeutig ist. Zu diesen Zusätzen gehören:

- Wortklassenangaben
- Bedeutungsdifferenzierungen (im einfachsten Falle eine Indexziffer, ggf. auch eine Fachgebiets- oder Themenangabe)

Als freie Zusatzangaben (solche, die nicht notwendig zur Vereindeutigung einer Graphemkette benutzt werden) können angeführt werden:

- Erläuterungen, Beispiele
- Definitionen

Eine Relationierung ist also beispielsweise zu "lesen": "MUTTER (Substantiv Femininum, PERSON)" "IST VON DEUTSCH NACH ENGLISCH ZU ÜBERSETZEN MIT" "MOTHER (Substantiv,...)" oder aber (verkürzt):

"MUTTER_1" "IST VON DEUTSCH NACH ENGLISCH ZU ÜBERSETZEN MIT"
"MOTHER_1"

3.1 Wortbezogene Stilhilfen mit Bezug zu Worterklärung und Synonymik

Ausgangspunkt

Als Standard-Rahmensituation sei das Schreiben mit einer Textverarbeitung bzw. die Interpretation (das "Lesen") eines Textes vorgegeben. Die Verfügbarmachung von Alternativen und Entscheidungshilfen auf Wortebene wird als "Stilhilfe" bezeichnet. Wird "nur" die Alternative angeboten, so handelt es sich um eine lexikalische Hilfe. Es erscheint jedoch möglich, die Attribute bzw. Differenzierungen in der Relationierung zu Vorschlägen bzw. Restriktionen bei der Verwendung auszunutzen.

Folgende Relationierungen sind zu nennen bzw. zu unterscheiden:

Synonymie

Die Synonymie ist durch die Invertierbarkeit gekennzeichnet, d.h. Relationsausgang und Relationseingang können getauscht werden (Darstellung: Schrägstrich). In der Anwendung kann zwischen den Alternativen gewählt werden.

- *'Echte' Synonymie* (Beispiel: Samstag / Sonnabend).

Hierbei kann durch Attribuierung (Hinzufügung spezifischer Merkmale weiter unterschieden werden zwischen "fremdsprachig" und "eingedeutscht", auch Merkmale wie "veraltet" oder "hochsprachig", "regionalsprachig" "Fachgebiet X" usf. gehören hierher bzw. können zur weiteren Differenzierung genutzt werden. Es ist eine Frage der Expliziertheit bzw. Differenziertheit der Relationierung, diese "Attribuierung" unmittelbar in einen eigenen Relationstyp umzusetzen - die dann nicht mehr reziprok ist (z.B. "Veloziped" "VERALTET FÜR" "Fahrrad") oder aber als Zusatz in Merkmalsform bereitzustellen und die Differenzierungen ggf. algorithmisch zu lösen: "Veloziped (VERALTET)" "SYNONYM ZU" "Fahrrad" (s.o.).

- *Abkürzung /Langform* (Beisp.: GmbH - Gesellschaft mit beschränkter Haftung).

Auch hier kann das Attribut ("Abkürzung" bzw. "Langform") in der Relation selbst ausgedrückt werden: "LKW" "IST ABKÜRZUNG FÜR" "Lastkraftwagen" (mit dem Relationstypenpaar "IST ABKÜRZUNG FÜR" / "IST LANGFORM VON"),

- *Schreibvarianten*

Zu den Schreibvarianten rechnen beispielsweise im Deutschen die unterschiedlichen Schreibweisen PH / F und C / K (Photo / Foto ...; Kuba / Cuba ...), aber auch E-Auslassungen (bummlich -

bummelig). Im weiteren Sinne kann man die zeichenorientierten sprachregionalen Besonderheiten (schweizerisch: ss deutsch: ß) hierher rechnen (vgl. aber die Rubrik "Regionalsprache").

- *Regionalsprache*

Regionalsprachige Besonderheiten können regelhaft sein (dann handelt es sich eher um entsprechend attribuierte Schreibvarianten) oder sie müssen lexikalisiert werden (Beispiel: "Allfälliges (...; ÖSTERR:)" "SCHREIBVARIANTE ZU" "Verschiedenes (...; DEUTSCH)").

- *Worterläuterungen*

Es ist sinnvoll, die Worterläuterung (wie beispielsweise auch die Übersetzung) als eine Relation zu betrachten. Im Grunde genommen handelt es sich um eine paraphrastische Synonymie: statt eines Wortes wird ein komplexer Ausdruck (Satz, Umschreibung) gebraucht.

Die Definition kann der Erläuterung gleichgesetzt werden mit dem Unterschied, dass ihr ein (anderes) Attribut zugeordnet werden kann, etwa DEF oder QUAL

- *Anwendung*

Die Worterläuterung wird in erster Linie bei der Textinterpretation (d.h. beim "Lesen") genutzt. Bei der Textgenerierung spielen die unterschiedlichen Synonymieformen eine weitaus größere Rolle.

Vor allem in der betrieblichen Information und Kommunikation (vom Firmenbericht bis zur Bedienungsanleitung) ist eine einheitliche Präsentationsform von Bedeutung. Hierbei wird die prinzipielle Freiheit, die sich beispielsweise in der Synonymie ausdrückt, durch Festlegungen (und entsprechende Attribuierung, etwa im Sinne einer "Vorzugsbenennung") eingeschränkt. Dies kann "lexikalisch" geschehen (z.B. "*verwende* PHOTO *statt* FOTO"), aber auch algorithmisch gestützt werden (*ersetze* CUBA *durch* KUBA).

3.2 Retrievalhilfen beim Zugang zu textuellen Datenbanken

Im folgenden werden formale Zugangsformen und Zugangsformen über Deskriptoren ausgeklammert. Im letzteren Falle kann auf die Ausführungen unter 3.1 verwiesen werden.

Ausgangspunkt

Heutige Textdatenbanken haben in der Regel nur eine schwache Zugangsmöglichkeit im Bereich der Freitextsuche. Wesentliche Merkmale beim Freitextretrieval - d.h. der Ermittlung relevanter Textstellen anhand von im Text vorkommenden Wörtern - sind Techniken der *Trunkierung* (etwas vereinfacht umschrieben mit: Suche über Wortanfänge bzw. Teilwörter) und über *Wortabstände*. In beiden Fällen hat man bei der *Texterschließung* (der sog. Freitext-Indexierung) das Problem der Ermittlung von Grundformen und Mehrwortbegriffen umgangen; man überlässt es der Geschicklichkeit des Nutzers, sich der o.a. Techniken zu bedienen.

Ziel *weitergehender* Verfahren ist es, den Nutzer bei diesem Suchvorgang zu entlasten. Es ist natürlich auch möglich, die im folgenden beschriebenen Komponenten in ein automatisches In-

dexierungsverfahren einzubringen, doch zeigt die Erfahrung, dass - international gesehen - davon kaum Gebrauch gemacht wird. Eine Ausnahme stellt - in gewissen Grenzen - die Erschließungskomponente PASSAT von GOLEM (SIEMENS) für Deutsch dar.

Im folgenden werden nur noch solche Bereiche beschrieben, die nicht schon im Abschnitt 3.1 behandelt wurden. Es ist jedoch einleuchtend, dass die dort beschriebenen Relationierungen (v.a. die Synonymie-Relationen) beim Retrieval in Textdatenbanken eine wichtige Rolle spielen. Im übrigen wird der Begriff "Textdatenbank" nicht allein auf Online-Systeme bezogen, vielmehr gilt für Bürosysteme und Archive Entsprechendes.

Der Retrievalbereich ist besonders gekennzeichnet durch das Phänomen, dass Hilfen zum sog. Browsing gegeben werden müssen. Hierbei werden im "klassischen" Thesaurus hierarchische Relationen bereitgestellt bzw. genutzt: Ober- / Unterbegriff; Teil / Ganzes usf.

Relationierungen

Für das Retrieval im Freitextbereich lassen sich diese "klassischen" Relationen, wie sie in Fachthesauri bereitgestellt werden, natürlich nutzen. Das Problem liegt weniger in der Bereitstellung solcher (meist fachspezifisch stark limitierter) Wörter und Wortrelationierungen als in der Zusammenführung von Text und Thesaurus. Die klassischen Thesauri setzen in der Regel schon eine hohe Abstraktions- und Verdichtungsleistung des (menschlichen) Indexierers voraus; ein Vorkommen der Deskriptoren in einem Text ist eher zufällig, zumindest aber ohne komplexe (statistisch basierte) Verfahren kaum zu nutzen (vgl. zu derartigen - erfolgreichen - Ansätzen zur automatischen thesaurusbasierten Indexierung die Arbeiten von Lustig / Knorz aus Darmstadt).

Für die *Freitextsuche* (i.S. einer Identifikation von Wörtern im Text) müssen daher andere - einfachere - Methoden gefunden werden. Hierzu rechnen v.a.

1. Relationen zwischen Mehrwortbegriffen und ihren Teilwörtern
(Beispiel: juristische Person - > Person)
2. Relationen zwischen Komposita und den (sinnhaften) Wortteilen
(Beispiel: Haustür -> Tür) - (Dekompositionsrelation)
3. Relationen zwischen Wörtern und Wortableitungen
(Beispiel: Unterschied / unterscheiden / unterschiedlich) - (Derivationsrelation)

Anwendungen

Es ist kaum zu erwarten, dass der Nutzer beim Retrieval derartige Hilfen ausreichend in Anspruch nimmt: So wie ihn häufig die Trunkierung überfordert bzw. "stört", würde ihn die Relationierung "erschlagen" (angesichts der vielfältigen Möglichkeiten, die ein solches Relationensystem letztlich bietet).

Es sind daher Lösungen anzustreben, die entweder die lexikalischen Daten auf das zugrundeliegende Material abstimmen (etwa wie es PASSAT schon tut) oder aber den Aufbau der Beziehungen in automatische Verfahren beim Retrieval integrieren (vgl. dazu ZIMMERMANN 1991).

3.3 Elektronische Wörterbücher als Hilfen bei der intellektuellen Übersetzung

Ausgangspunkt

Ein Übersetzungswörterbuch stellt - allgemein betrachtet - ein spezielles Relationenwörterbuch dar, da die Übersetzung als eine spezielle Synonymrelation angesehen werden kann. Dies ist jedoch trivial und ' nicht Gegenstand des vorliegenden ' Abschnitts. Zu den spezifischen Problemen, die bei der Bereitstellung und Nutzung von Übersetzungswörterbüchern zu lösen sind, rechnen:

- die Behandlung von Synonymen,
- der Zugang über Alternativstämme,
- die Vermeidung falscher Übersetzungen (im Gegensatz zur Zulassung bei der Identifikation).

Relationierungen

Die wesentlichen Relationierungsmöglichkeiten sind bereits in den Abschnitten 3.1 und 3.2 aufgezeigt worden. Im Übersetzungswörterbuch treffen gleichsam alle Phänomene zusammen. Ein elektronisches Übersetzungswörterbuch könnte jedoch (im Gegensatz zu seiner gedruckten Repräsentationsform) weitaus mehr über logische Strukturen abbilden als phasisch zu realisieren. Hierzu einige Beispiele:

- Verweisrelation (von einer graphematischen Variante zu dem "lexikalisierten" Eintrag),
- Nutzung der Regionalsprachenkennung (-relation) zur Aktivierung bzw. Unterdrückung einer Alternante,
- Realisierung einer Verweisrelation von einer falschen Schreibform zu der korrekten Form (Beispiel: Lybien -> Libyen).

Anwendungen

Die wesentlichen Anwendungen liegen im Bereich des elektronischen Blätterns (inkl. Browsing). Hierbei können Zugangsprobleme gelöst werden, aber auch Präsentationsformen beeinflusst werden. Im letzteren Falle kann z.B. optisch (am Bildschirm) eine "Nesterbildung" (wie in klassischen Übersetzungswörterbüchern üblich) realisiert oder aber eine explizitere Präsentation gewählt werden (etwa zur Erleichterung der Auswahl / Markierung einer Übersetzung zur Insertion in den Text).

Eine interessante Variante ist m.E. die Nutzung derartiger Relationierungen beim Vokabeltest. Durch Relationierung (etwa zu Schreibvarianten) wird beim Test eine Alternative zugelassen, die bei der Präsentation selbst nicht auftaucht (den Nutzer also nicht unnötig belastet). Ähnliches gilt für Regionalvarianten (britisches / amerikanisches Englisch usw.). Ein solches Verfahren ist in den Vokabeltrainern MiniVoc und EuLex von SOFTEX, Saarbrücken, eingebracht.

4. Zusammenfassung und Ausblick

Die vielfältigen Möglichkeiten, die sich durch Anwendungen von Wortrelationierungen in der Sprachtechnik ergeben, konnten hier nur angedeutet werden. Es ist vielleicht deutlich geworden, dass es eine wahre Sisyphusarbeit darstellt, das Material zu einer Sprache entsprechend aufzubereiten. Automatisierungen durch statistische Auswertungen können eine gewisse Hilfestellung geben, reichen aber nicht aus.

Es wird in diesem Zusammenhang derzeit häufig die Frage der Verwendbarkeit vorhandener Materialien ("Wiederverwendbarkeit") - etwa klassischer Lexika und bestehender Thesauri - in der Sprachtechnik betrachtet. Nach den bisherigen eigenen Erfahrungen decken diese Materialien (selbst wenn sie maschinenlesbar vorliegen) nur einen kleinen Ausschnitt ab. Hinzu kommt, dass die meisten (Groß-)Verlage, die über ein entsprechendes Grundmaterial verfügen, zwar ansatzweise in den Markt einsteigen, aber selten die notwendigen Software-Lösungen dazu anbieten. Vielfach wird das elektronische Lexikon als eine Variante der Buch-*Präsentation* gesehen (man vgl. etwa die neueren Anwendungen im CD-ROM-Bereich bzw. des DATA DISK MAN von Sony) und weniger als eine integrierte Funktion in der Textverarbeitung, beim Retrieval oder beim Übersetzen. Dies kann und wird sich aber ändern.

Der vorliegende Beitrag ist ein Versuch, den Blick der Beteiligten in Richtung auf die vielfältigen Anwendungsaspekte und Präsentationsmöglichkeiten zu richten, die sich heute schon ergeben.

(Leicht revidierte Version der Fassung vom 14. Oktober 1991 (T21VW1))