

Informationswissenschaft und Linguistik. Ihr Verhältnis, historisch und jetzt

Volkmar Engerer

Abstrakt

Ich leite den ersten Teil meines Vortrags ein mit der Frage: Was macht/ist die Informationswissenschaft? und unterscheide drei wissenschaftstheoretische Zugänge zu einer Antwort (realistische, essentielle und problemlösungsorientierte Antworten). Ich werde hier den letztgenannten Zugang favorisieren und das mit ihm korrespondierende Szenario des Information retrieval als „Kern“ der Informationswissenschaft – und als weiterer Ausgangspunkt meiner Darlegungen – motivieren. Sprache und kommunikative Phänomene sind hohem Masse in dieses Modell der Informationssuche integriert. – Ich werde dann überblicksmäßig einige besonders wichtige Felder der Berührung von Linguistik und Informationswissenschaft ansprechen, hierunter die Computerlinguistik, Textlinguistik, Wittgensteinrezeption und mehr. – Im zweiten Teil meines Vortrags stelle ich einen Versuch einer systematischen Chronologie des Linguistik-Informationswissenschaft-Verhältnisses dar, die sich an den Hauptparadigmen in den Informationswissenschaften, der Systemphase und der kognitiven Phase, orientiert. Ich zeige hier, dass sich die aus der Sprachwissenschaft entlehnten Theorien und Gedankengebäude am informationswissenschaftlichen Bedarf in den jeweiligen Phasen richten. – Im dritten und letzten Teil meines Vortrags unternehme ich eine Standortbestimmung der Informationswissenschaft in Bezug auf die Linguistik. Als Basis hierfür führe ich die Eigenschaft der Vokabularbezogenheit ein, welche die Linguistik mit der Informationswissenschaft und zwei weiteren Disziplinen (Fachsprachenforschung und Terminologielehre) in einem homogenen „Feld“ verbindet. Lexika in der Linguistik, Thesauri in der Informationswissenschaft, Fachwörterbücher in der Fachsprachenforschung und Terminologien in der Terminologielehre sind Ausdruck für disziplinspezifische Problemstellungen und verwirklichen gleichzeitig die disziplinspezifischen formalen und bedeutungsbezogenen Anforderungen an die jeweiligen lexikalischen Zeicheneinheiten. – Eine tiefere Strukturierung des Verhältnisses dieser drei Disziplinen wird mit Hilfe des Merkmals der Vokabularkontrolle vorgeschlagen.

Manuskript für einen Plenarvortrag, zu halten am 22. Juni 2015 an der Staatlichen Universität St. Petersburg, Russland, im Rahmen der jährlichen Tagung der GeSuS.

Abkürzung: IW = Informationswissenschaft

1. Einleitung

Vorstellen; aus Dänemark; IVA.

Frage: Wer weiß (Hand aufs Herz), was IW ist? – Arme hoch!

Und von denen, die es wissen: Wissen Sie auch, was einen Informations- von einem Computerwissenschaftler unterscheidet?

Als ich am 1.1.2012 in der IW begann: Ich wusste es nicht.

Seither habe ich mich (als ausgebildeter Linguist) in die IW eingearbeitet, was will sie, was kann sie, womit arbeitet sie?

Perspektive war, ob ich wollte oder nicht: mein sprachlicher, linguistischer und humanistischer Hintergrund.

Mein bisheriges Ergebnis meiner Beschäftigung:

- a) Eine Interpretation des historischen Verhältnisses zwischen der IW und der Linguistik, etwa seit 1950.
- b) Eine Interpretation des synchronen Verhältnisses von IW zur Linguistik und anderen Disziplinen. Schlüssel ist hier: Vokabularbezogenheit.

Diese beiden Perspektiven habe ich in zwei längeren Beiträgen (erscheinen voraussichtlich dieses und nächstes Jahr) an ein informationswissenschaftliches Publikum vermittelt (s. Engerer, unpubliziert und eingereicht).

Nun sind die Linguisten dran: Mit einer neuen Perspektive, die auch das Ziel meines Vortrags bestimmt:

Ich möchte klarmachen, dass

- a) Unsere Disziplin, die Linguistik, für andere Disziplinen, hier die IW, außerordentlich attraktiv ist;
- b) Unsere Disziplin enormes Potential zur interdisziplinären Zusammenarbeit hat;
- c) Und es geht mir auch darum, Ihr Interesse an der IW zu wecken – ich meine, dass die beiden Disziplinen einander viel zu bieten haben.

Mein Vortrag hat deshalb folgende Struktur:

Teil I: Generelles

1. Einleitung
2. Was macht/ist die Informationswissenschaft? (Handbuchdefinitionen, Information Retrieval als "Kern" der Informationswissenschaft, Sprache als Komponente in Modellen des Information retrieval)
3. Überblick: Felder des Kontakts zwischen Linguistik und Informationswissenschaft (Maschinelle Textverarbeitung, Computerlinguistik, Natural Language Processing,

Casegrammatik und automatische Textanalyse, Textlinguistik, Wittgensteinrezeption: das Repräsentationsproblem, Warner: Information Retrieval und Saussure)

Teil II: Chronologie (diachron)

4. Versuch einer systematischen Chronologie: Paradigmen der Informationswissenschaft und linguistische Kontakte (das physikalische und das kognitive Paradigma, Bedeutungsskeptizismus, vom Systembenutzer zum Sprachverwender und zur Situation heute; illustrativer Überblick über die Entwicklung)

Teil III: Vokabulare

5. Synchrone Standortbestimmung der Informationswissenschaft in Bezug auf die Linguistik (Vokabularbezogenheit und Nachbardisziplinen)
6. Vokabulartypen und Nachbardisziplinen der Informationswissenschaft (Lexika in der Linguistik, Thesauri in der Informationswissenschaft, Fachwörterbücher in der Fachsprachenforschung, Terminologien in der Terminologielehre; Illustration: Linguistik und Informationswissenschaft in Feld mit 2 Nachbardisziplinen)
7. Strukturelle Analyse des Verhältnisses der 4 Disziplinen (Vokabularkontrolle)
8. Zusammenfassung und Ausblick; Syntagmatisierung

Ich weise immer darauf hin, wenn ich von einem Punkt zum nächsten übergehe. Bitte notieren Sie sich Ihre Fragen und Beiträge, wir haben anschließend ausreichend Zeit zur Diskussion.

Ich habe auf jeden Fall vor, die Teile I und II durchzugehen, aber sollte die Zeit dann schon zu fortgeschritten sein, überspringe ich Teil III und gehe direkt zur Abrundung.

Literatur:

Engerer, Volkmar, "Exploring interdisciplinary relations between linguistics and information science from the 1960s to today. A Tredinnickian perspective"; submitted to JASIST [9.7.2014]; positive review received [10.9.2014]

Engerer, Volkmar, "Vocabulary in information science and (other sign-) related disciplines", unpublished manuscript.

2. Was macht/ist die Informationswissenschaft?

Allgemein wird festgehalten, dass es keine allgemein akzeptierte Definition der IW gibt (vgl. zum Folgenden Stock & Stock):

- Junge Disziplin, ca. 50 Jahre alt
- Stark mit anderen Disziplinen verbunden, interdisziplinär

Die Frage, was die IW ist, kann verschieden beantwortet werden. Ich möchte im Folgenden 3 Arten, diese Frage zu beantworten, unterscheiden.

Zunächst Fragen wie:

- Was sind die Aufgaben der IW?
- Wie hat sich die IW historisch entwickelt?
- Mit welchen anderen Disziplinen unterhält die IW Beziehungen?

Eine Charakterisierung der IW entlang dieser Fragestellungen könnte man **realistisch** nennen. Diese ist in Bezug auf die Konstituierung der IW als Disziplin nicht-unitaristisch.

Eine andere Frage ist die:

- Welche Rolle spielt Information, Wissen und Dokument in der IW?

Diese Frage zielt auf einen „Kern“, einen intensionalen Inhalt des Ausdrucks „Informationswissenschaft“. Ich möchte diesen Zugang als „**essentiell**“ bezeichnen. Er dient der Disziplin der Stabilisierung als eigenständige Wissenschaft und ist unitaristisch für die Formierung der IW als Wissenschaft.

Eine dritte Möglichkeit der Annäherung an die IW ist **problemorientiert** und hat den Informationssuchprozess zum Gegenstand:

Die IW untersucht die Repräsentation, die Speicherung und die Versorgung sowie die Suche nach und die Wiederfindung von relevanten Dokumenten und relevantem Wissen.

Hiernach ist das definierende Szenario der IW grundsätzlich problematisch, beruht auf einer Defizithypothese auf Seiten des Benutzers eines Wissenssystems. Auch der problemorientierte, prozessuale Zugang zur Disziplin ist, wie der essentielle, von unitarischem Charakter.

Ich möchte diese 3 Zugänge kurz illustrieren und kommentieren.

2.1 **Essentielle Versuche einer Charakterisierung der IW**

Der essentielle Zugang zur IW, spiegelt sich in der Frage

- Welche Rolle spielt Information, Wissen und Dokument in der IW?

wider.

Eine Frage wie diese zielt auf die inhärente Bedeutung der morphologischen Komponenten des Wortes „Informationswissenschaft“ an, also

Information + Wissenschaft;

Die „Essenz“, der inhaltliche Kern der Wissenschaft, ist hier das Ziel der Übung (nomothetische Haltung?).

„Information“

Es ist in der theoretischen IW charakteristisch, dass Information in Zusammenhang mit Wissen betrachtet wird.

Die Verbindung von Wissen-Information wird z.B. an Poppers Theorie der 3 Welten gezeigt (Literaturhinweise in Stock & Stock):

Welt 1: physikalische Welt

Welt 2: Welt der bewussten Erfahrung (mental, subjektives Wissen)

Welt 3: Welt des Inhalts (objektives Wissen in z.B. Artikeln und Büchern, existiert unabhängig von Menschen)

Welt 2 und 3 hängen eng zusammen: Subjektives Wissen beruht auf Welt 3, und umgekehrt: objektives Wissen entwickelt sich aus Welt-2-Wissen heraus.

Wie kommt nun der Übergang von Welt 3 in Welt 2 zustande? Oder: innerhalb Welt 2? Hier kommt Information ins Spiel:

Information ist die FORM, die in Welt 2 verstanden wird. Wissen als solches kann nicht transportiert werden, Information ist der Träger.

Man spricht auch von Information als „Wissen in Bewegung“, Wissen kann das Ergebnis eines Informationsaktes sein (muss es aber nicht). Ein strukturierter Informationsinhalt drückt Wissen aus.

Einen rein semantischen Winkel auf „Information“ legt Buckland an. Er schließt von den 3 „Bedeutungsaspekten“ des Ausdrucks „Information“, in einem im praktischen und ontologischen Gegenzug sozusagen, auf die Identifizierung „realer Objekte der IW“:

- Information als Prozess („Peter informiert Helga“)
- Information als Wissen (Information transportiert Wissen, so dass Helga nun die besagte Information „hat“)
- Information als Ding (die Fixierung von Information in „informativen Dingen“, z.B. Dokumenten; hier hat Helga einen entsprechenden Brief von Peter bekommen)

Hier bauen also ontologische Statements auf einer semantischen Analyse von „Information“ auf.

[Interessant: Ebenso wie Information und Wissen in der IS eng verknüpft sind, ist dies der Fall in der Sprachwissenschaft in Bezug auf sprachliches Wissen und die in Sätzen enthaltene Information. Können sprachliches Wissen und enzyklopädisches Wissen in Beziehung gesetzt werden?]

„Wissenschaft“

Eine lange Diskussion ist in der IW über ihren wissenschaftlichen Status geführt worden. „Wissenschaft“ und „wissenschaftlich“ werden und wurden oft im Sinne einer naturwissenschaftlichen und positivistischen Methodologie verstanden, was es, so die Kritiker, für die Mainstream-IW schwer machte, bedeutungsbezogene Phänomene zu verstehen.

2.2 Realistische Versuche einer Charakterisierung der IW

Fragen:

- Was sind die Aufgaben der IW?
- Wie hat sich die IW historisch entwickelt?
- Mit welchen anderen Disziplinen unterhält die IW Beziehungen?

IW hat sich in enger Zusammenarbeit mit der Computerwissenschaft entwickelt und der Informationsverarbeitung. Des Weiteren besteht ein enges traditionelles Verhältnis zur Bibliothekswissenschaft (deshalb englisch LIS = Library and Information Science), Lerntheorie und Pädagogik (u.a. in Lehrplattformen).

Historisch hat sich die IW in den 50er Jahren entwickelt, also die Computer- und Datenrevolution fühlbar wurde, und historisch ist sie immer noch eine stark angewandte Disziplin.

2.3 Problemorientierte Versuche einer Charakterisierung der IW

Die IW untersucht die Repräsentation, die Speicherung und die Versorgung sowie die Suche nach und die Wiederfindung von relevanten Dokumenten und relevantem Wissen.

Bestandteile dieser Definition (aus Stock & Stock):

Repräsentation: Alles Wissen in Dokumenten (z.B. wissenschaftliche Artikel) ist kondensiert in verkürzten Texten (Abstracts) oder Deskriptoren/Schlagworten, die der Informationsfiltrierung dienen.

Speicherung und Versorgung: Strukturierung einer Datenbank geschieht so, dass die Dokumente einfach wiedergefunden werden können und inhaltlich zugänglich sind.

Suche: Die Untersuchung von Benutzern im Prozess der Deckung ihres Informationsbedarfs, der verwendeten Suchausdrücke in Bezug auf die jeweiligen Suchwerkzeuge und die Art und Weise, wie sie die gefundene Information weiterverarbeiten.

Wiederfindung/Retrieval: Entwicklung von Systemen für die Erforschung von Wissen wie Internet-Suchmaschinen oder Bibliothekskataloge.

Relevanz: Grundbegriff der IS: Nicht jegliche Information ist Gegenstand von Suchen, sondern ausschließlich solche Informationen, welchen dem Suchenden dabei helfen seinem Informationsbedarf abzuwehren.

Dokumente: Textliche und nicht-textliche Objekte (Bilder, Musik, Videos, ...).

Wissen: Statischer Inhalt, fixiert in Dokumenten oder gespeichert in einem mentalen Gedächtnis. Wissen kann digital (WWW), materiell (Bücher auf einem Bibliotheksregal) oder psychisch-mental sein (Gehirn eines Angestellten einer Firma). Information ist dagegen dynamisch.

Hiernach ist das definierende Szenario der IW grundsätzlich problematisch, beruht auf einer Defizithypothese auf Seiten des Benutzers eines Wissenssystems. Auch problemorientierte, prozessuale Zugang zur Disziplin ist, wie der essentielle, von unitaristischem Charakter.

2.4 Eine Abwägung der 3 Zugänge

Essentiell: Es sieht nicht so aus, als ob die semantisch basierten Diskussionen über Information, Wissen und Dokumente mit praktischen Problemstellungen kombiniert werden können. Will das nicht weiterverfolgen.

Realistisch: Rein deskriptiv, kein Potential für die IW als Disziplin, in einem unitaristischen Sinne. Wird nicht weiterverfolgt.

Problemorientiert: Scheint mir der vorzuziehende Umriss einer sowohl praktisch als auch theoretisch fundierten IW zu sein.

Hier einige der Eigenschaften des problemorientierten Winkels, die in diesem Sinne positiv zu werten sind:

- Das Szenario ist grundlegend problematisch und lösungsanbietend, lösungsfordernd
- Das Szenario ist allgemein und trotzdem spezifisch: Kann auf viele ähnliche Situationen angewendet werden
- Benutzerintegrierend: Der Benutzer, mit allen seinen individuellen, psychologischen und kulturellen Kennzeichen geht durch das Konzept des Informationsbedarfes als zentraler Bestandteil in das Szenario ein
- Das Szenario ist technologisch und anwendungsorientiert

- Ist nur interdisziplinär zu bearbeiten
- Das Szenario fordert durch seine logisch verbundenen Komponenten zusammenhängende Problemstellungen heraus
- Das Szenario wird Domänenvielfalt gerecht (professionell, soziale Medien, Alltagsdomänen, Freizeit, Wissenschaft und Forschung, Ausbildung, Militär, ...)
- Das Szenario ist sprach- und kommunikationsfokussiert, da es ein Interface zwischen System und dem menschlichen Benutzer modelliert.

Ich möchte daher mit dem problemorientierten Modell der IW im Folgenden weiterarbeiten.

2.5 *Information retrieval als "Kern" der Informationswissenschaft*

Das problemorientierte Modell der IW ist im sog. „Information retrieval“ implementiert, dessen Elemente unten zu sehen sind:

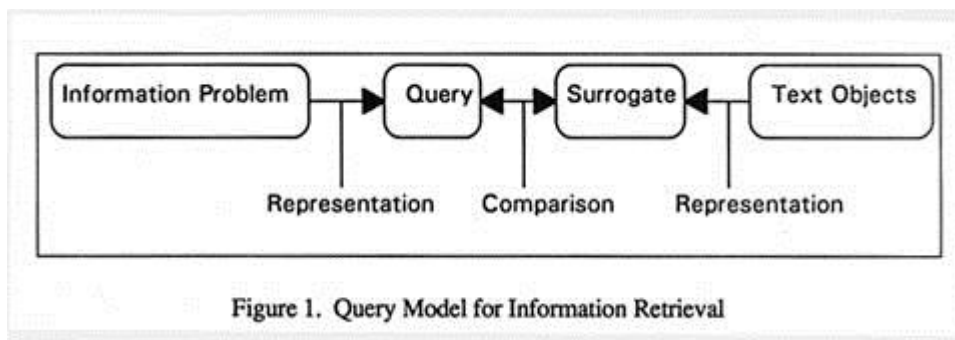


Abb. nach Kevin Cox, Information Retrieval in Browsing

Dieses Modell ist sprachlich gesehen natürlich komplizierter, mehr dazu weiter unten.

Ich möchte also für das Folgende das informationswissenschaftliche Information retrieval-Paradigma als Kernkomponente und definierendes Szenario für die IW ansehen.

2.6 *Sprachliche Repräsentationsebenen in den Komponenten des Information retrieval*

Links sind mehrere sprachliche Repräsentationsebenen anzusetzen,

- einmal der kommunikative Ausdruck und die natürlichsprache Repräsentation des Informationsbedarfs des Benutzers, und
- dann eine mehr suchmaschinenorientierte Repräsentation desselben Inhalts als „query“, Suchausdruck.
- Als dritte Ebene könnte noch die eines „intermediate“ dazukommen, also eines Informationsspezialisten (z.B. Bibliothekar), der zwischen dem Benutzer und dem System vermittelt.

- Nicht zu sprechen von iterativem sprachlichem Feedback an der Schnittstelle zwischen Benutzer und Maschine.

Repräsentationsebenen und Repräsentationsprobleme von rechts:

- Textuell-sprachliche Repräsentation von Dokumentinhalten (natürlichsprachige Repräsentation), „aboutness“
- Sprachliche Interpretation der Textbedeutung und Repräsentation dieser Textbedeutung mit Hilfe eines kontrollierten, maschinennahen Vokabulars (Indexierung, Repräsentation der „Welt“ durch sprachliche Bedeutungen)
- Repräsentation der Textbedeutung mit Hinblick auf den Sprachgebrauch eines zukünftigen Benutzers (Indexierung, benutzergerichtet)
- Syntaktisch variierende Kombination der Indexterme mit Hinblick sowohl auf die ontologische Komposition des Textreferenten und auf mögliche Übereinstimmungen der Indexierungssyntax mit der Benutzersyntax in Abfragen.

Nun sollte das Feld linguistischer und sprachtheoretischer Fragestellungen in der IW abgesteckt sein.

Ich meine, dass das problemorientierte Such-Prozess-Szenario, modelliert im informationswissenschaftlichen Information retrieval- Schema, für den Linguisten die meisten und attraktivsten Ansatzpunkte bietet.

Das problemorientierte Szenario bindet mehrere sprachliche Repräsentationsebenen in einem kohärenten Model zusammen.

Ich werde dieses Modell später unter einem etwas anderen Blickwinkel wieder aufnehmen.

Literatur:

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351-360.

Cox, Kevin, Information Retrieval in Browsing

(<http://web.simmons.edu/~chen/nit/NIT'92/069-cox.htm>). Original nach Belkin, N.J. &

Croft, W.B. (1987). "Retrieval techniques," In *Annual Review of Information Science and Technology* (ARIST), 22: 109-189.

Stock, W. G., & Stock, M. (2013). *Handbook of information science*. Berlin: deGruyter Saur. [Hier v.a. Kapitel über „Information Science“ und „Information retrieval“]

3. Überblick: Felder des Kontakts zwischen Linguistik und Informationswissenschaft

Wichtige frühe Quelle für das IW-Linguistik-Verhältnis ist:

Sparck Jones, K., & Kay, M. (1973). *Linguistics and information science*. New York, London: Academic Press.

Hier stehen die frühen Grundstatements über zwingende und mögliche Berührungsoberflächen zwischen IW und Linguistik, die, aus der Sicht der IW, immer noch gelten.

Sparck Jones & Kay (1973)

Linguistik und IW: "natural bedfellows". (S&K, 1)

Man dachte darüber nach, die IW als Teil der angewandten Linguistik anzusehen, mit folgender Logik: Die IW hat Dokumente als Gegenstand, Dokumente bestehen wesentlich aus Text, und Texte sind (auch) Gegenstand der Linguistik. (S&K, 2)

Auch damals schon wird der Hauptbereich der Zusammenarbeit im Information retrieval gesehen.

Warum?

Ein Hauptproblem des Information retrieval besteht darin, den Inhalt von Dokumenten und Informationsnachfragen so zu beschreiben, dass diese Beschreibungen in einem automatisierten Prozess benutzt werden können, und zwar in der Weise, dass für jedes Dokument seine Relevanz in Bezug auf eine konkrete Informationsnachfrage bestimmt werden kann.

(Dies ist eine einfache Umformulierung des problemorientierten Zugangs in der IW.)

Es ist dieser Match zwischen Dokumentbeschreibungen/Dokumentsurrogaten und natürlichsprachlichen Benutzernachfragen, wo Linguistik und IW zusammenkommen. (S&K, 2).

Der Status der linguistischen Theorie in den 1970ern wird von S&K eher negativ beurteilt: Sprachtheorien wechseln sich schnell ab, und für den Informationswissenschaftler bietet sich ein unüberschaubares Bild.

Als gut entwickelt wird auch die Syntax angesehen, aber ohne eine Kopplung an eine semantische Theorie ist diese auch wertlos. (Gut gesehen!)

Besonders die Semantik in den 1960ern und 1970ern wird von den Autoren skeptisch eingeschätzt: alles andere als avanciert, hochspekulativ und unterentwickelt. (S&K, 4)

Ein anderer Mangel: S&K beobachten in der Linguistik einen starken Fokus auf den Satz und kleinere Einheiten – eine Vernachlässigung des satzüberschreitenden Aspekts.

Die Autoren nennen dies die „mikroskopische“ Perspektive, welche für Informationswissenschaftler und „Dokumentforscher“, die mit kompletten Texten zu tun haben, wenig Sinn macht. (S&K, 4).

Eine Sprachtheorie, die von Information retrieval-Systemen adaptiert werden kann, muss „grobe“ Statements über große Texteinheiten zulassen. Linguisten sind nicht daran gewöhnt. (S&K, 5)

In einem eigenen Kapitel (3.) gehen S&K die Linguistik der 1960/70er Jahre aus der Sicht der Informationswissenschaft durch.

Während die Phonologie als irrelevant für das Information retrieval angesehen wurde (S&K, 36), ist die Morphologie eine zentrale Beschreibungsebene in der Computerlinguistik: Es geht hier darum, Worte in der natürlichen Sprache zu erkennen und zu analysieren (S&K, 37)

Der Beitrag der Computerlinguistik wird allgemein als hoch eingeschätzt (S&K, 43f):

Warum?

1. Natürlichkeit: Dokumente sind in natürlicher Sprache verfasst, und der suchende Benutzer wird auch die natürliche Sprache für die Formulierung seiner Suchketten vorziehen. Das bedeutet, dass alle Suchstatements eine sprachverarbeitende Prozedur durchlaufen müssen, bevor sie vom System in logischen Operationen benutzt werden können.
2. Ambiguität: Mehrdeutigkeiten müssen durch eine syntaktische Analyse eliminiert werden.
3. Syntaktisierung: Sowohl Dokumentbeschreibungen als auch Suchketten haben oft die Form einer ungeordneten Liste von Termen und sind daher prinzipiell unpräzis. Hier kann mit Hilfe der Computerlinguistik eine Syntax über diese Terme gelegt werden.
4. Semantische Struktur: Hantierung von Synonymen, sowohl in Termen als auch größeren syntaktischen Einheiten.

Grundfrage, die von S&K gestellt wird: "is the kind of drastic compression of content that must be done to provide a document with an index description a process about which linguists can reasonably be expected to have anything to say?" Kann die Linguistik hier mit den richtigen Techniken helfen? (S&K, 54)

Die semantische Kondensierung von Texten war eines der klassischen Forschungsgebiete der angewandten Textlinguistik in den 1970er Jahren.

Kapitel 5 in S&K behandelt die Syntax und bespricht insbesondere die Phrasenstrukturgrammatik, Dependenz und die Transformationsgrammatik (S&K, 81ff) sowie Fillmores Kasusgrammatik und Theorie der semantischen Rollen (S&K, 96).

Die Hauptfrage für die Autoren ist, inwieweit syntaktische Struktur überhaupt für das Information retrieval brauchbar und relevant ist. Auch hier sind die Autoren skeptisch und bezeichnen die Miteinbeziehung von syntaktischer Information als Trade off. (S&K, 103)

Syntax kann wieder relativ zum Modell des Information retrieval gesehen werden. Die Hauptanwendung für syntaktische Regeln (auf der linken, Dokumentseite) ist eine formal ausgearbeitete Indexierungssyntax (S&K. 112):

Präkoordinierende Systeme: stark syntaxbasiert

Postkoordinierende Systeme: die Rolle der Syntax ist minimal

Oft wird die Reihenfolge der Terme als Ersatz für die Kennzeichnung für Relationen verwendet.

In den sog. ‚Subject headings‘ der Library of Congress sind syntaktische Ordnungsprinzipien wirksam, an erster Stelle die wichtigsten Terme, erst danach alle anderen. Auch das Prinzip des Kopfes und einem modifizierenden Term ist in der Indexierungstheorie bekannt:

Z.B. wird oft das Prinzip Kopfterm vor modifizierendem Term verwendet: „Leder—Schuh“ vs. „Schuh—Leder“.

Auch semantische Kategorisierungen wie ACTION und THING dienen der Festlegung der syntaktischen Reihenfolge.

Postkoordinierende Syntax macht nicht viel Gebrauch von syntaktischen Mitteln, die syntaktische Verantwortung liegt beim Benutzer (J&K, 113):

"The ideal in post-coordinate indexing is the system in which a document description consists of an essentially unordered list of keywords."

Alle Beschreibungsterme sind gleichberechtigt, und der Benutzer bedient sich u.a. Booles Operatoren.

—

Weitere Berührungspunkte

Haupteinsatzgebiete linguistischer Methoden und Theorien lagen also im Bereich der maschinellen Sprachverarbeitung, der automatischen Textanalyse, der Computerlinguistik und, noch immer ganz aktuell, im Bereich der natürlichen Sprachverarbeitung (Natural Language Processing).

Letztere auch, weil die natürliche Sprache im Zeitalter der Volltextsuchen und des Vertrauens des Benutzers in die natürliche Sprache ein ganz aktuelle Rolle spielen.

Die Textlinguistik spielte hier eine besondere Rolle, man erinnere sich an S&K's Bemerkungen zum Bedarf der Informationswissenschaftler nach „gröberen“ und „größeren“ Texteinheiten.

Ein wichtiger Beitrag hierzu stammt von Claire Beghtol:

Beghtol, C. (1986). Bibliographic classification theory and text linguistics: Aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42(2), 84-113.

Hier wird das wichtige Prinzip der Aboutness in Klassifikationssystemen mit modernem Information retrieval in Beziehung gebracht. Beghtol beruft sich hier auf die textlinguistischen Arbeiten von van Dijk. Ihr geht es hier um die sprachlichen und kognitiven Prozesse, die bei der „Übersetzung“ eines Dokumentthemas in eine klassifikatorische Sprache vorkommen.

Ein Bezug, der über das technisch-angewandte und problemorientierte Heranziehen linguistischer Konzepte und Erkenntnisse hinausgeht, und sozusagen mehr an die wissenschaftstheoretischen Grundlagen der IW rührt, ist in den Arbeiten von Donald Blair hergestellt worden.

Sein wohl wichtigstes Buch ist:

Blair, D. C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier Science.

Er nimmt hier Wittgensteinschen Ideen einer sprachgebrauchsorientierten Semantik sowie seine Theorie der Sprachspiele auf.

Diese Ideen einer Ordinary Language Philosophy wendet er auf Grundlagenprobleme des Information retrieval an, das, was er als das Repräsentationsproblem ansieht.

U.a. charakterisiert er den „Sprachgebrauch“ des Indexers bei seiner Arbeit als Kommunikationsversuch, der an einen, sehr abstrakten, suchenden Benutzer gerichtet ist. Gleichzeitig wird der Benutzer des Information retrieval-Systems als ein Teilnehmer eines Sprachspiels charakterisiert, der sich an dem Sprachgebrauch eines, ebenso abstrakten, Indexers orientiert.

Blairs Ideen sind ohne Zweifel interessant, sind aber, soweit ich das beurteilen kann, weitgehend ohne Auswirkungen auf die Mainstream-IW geblieben.

Ein letzter Versuch, linguistische Erkenntnisse intellektuell in das Information retrieval einzubauen, stammt von Julian Warner in:

Warner, J. (2010). *Human information retrieval*. Cambridge, Mass.: MIT Press.

In diesem Buch unternimmt Warner den Versuch, de Saussures bekannte Unterscheidung zwischen Syntagmatik und Paradigmatik auf eine sprachorientierte Analyse des Volltextretrievals anzuwenden.

Kurz gesagt, es ist für das Funktionieren von Information retrieval-Systemen charakteristisch, dass Indexterme aus ihrem syntaktischen Kontext im Dokument herausgenommen werden, sie sozusagen ihrer syntagmatischen Einbettung und Disambiguierung entrissen werden, und in einem Maschinen-Suchindex als paradigmatische Terme sozusagen „kaltgestellt“ werden.

Der Benutzer ist dann, folgt man Warner, gezwungen, den Weg zurück zu gehen, und in seinen Suchketten, und experimentellen Proben in Volltextdokumenten, die mehrdeutigen indexparadigmatischen Suchterme wieder einem syntagmatischen Kontext zuzuführen, der der Suchintention des Benutzers entspricht.

Dieser Ansatz ist interessant, weil er direkt die Praxis und die zugrundeliegenden Annahmen von IR-Systementwicklern deutlich macht und kritisch hinterfragt.

In diesem Abschnitt konnte ich, ohne hier feinere Abwägungen zu erörtern oder gut begründete Prioritäten zu setzen, einige wichtige Kontaktbereiche zwischen Information retrieval und der Linguistik nennen – und, wie ich hoffe, teilweise auch inhaltlich motivieren.

Die Hypothese, die ich im nächsten Teil meines Vortrages verfolgen will, beruht auf der Annahme, dass die Art des Kontakts und die Typen linguistischer theoretischer Begrifflichkeit, der sich die IW im Laufe der vergangenen 60 Jahre bediente, nicht zufällig sind. Ich möchte zeigen, dass die Art der entlehnten Elemente aus dem sprachwissenschaftlichen Lager sich an den Entwicklungsstadien der IW selbst orientiert.

Literatur:

- Beghtol, C. (1986). Bibliographic classification theory and text linguistics: Aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42(2), 84-113.
- Blair, D. C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier Science.
- Sparck Jones, K., & Kay, M. (1973). *Linguistics and information science*. New York, London: Academic Press.
- Warner, J. (2010). *Human information retrieval*. Cambridge, Mass.: MIT Press.

4. Versuch einer systematischen Chronologie: Paradigmen der Informationswissenschaft und linguistische Kontakte

Ich möchte jetzt 2 Phasen der IW aufzeigen (vgl. B. Hjørland) und damit die „entlehnten“ linguistischen Theorieelemente motivieren. Ich möchte auch kurz Stellung nehmen zur Situation der Linguistik heute in der IW.

4.1 *Das physikalische Paradigma*

Das sogenannte physikalische Paradigma in der IW:

- Geht zurück auf die 1950er und 1960er Jahre
- Definierte IW als Wissenschaft des Information retrieval

Das physikalische Paradigma:

- Betrachtete Information als ausschließliches Phänomen der physikalischen, „wirklichen“ Welt
- Mit positiven Qualitäten und distinktiven materiellen Manifestationen
- Bedeutung von Information wurde nicht thematisiert

Information retrieval: objektiver, neutraler Prozess.

Für Dokumente gilt: Das Thema eines Dokuments/Texts ist eine Funktion der Worte in diesem Dokument, d.h.: das Thema kann eindeutig dadurch beschrieben werden, Worte aus dem Dokument zu extrahieren.

Somit kam das Dogma der semantischen Kondensierung auf, das auch in der Textlinguistik in den 70er Jahren erforscht wurde.

Diese wissenschaftstheoretischen und methodischen Voraussetzungen der IW begründete die Ausprägung des Kontakts zur Linguistik:

Linguistische Theorien mussten

- Anwendbar sein
- Maschinenimplementierbar sein
- Und die richtige grobe „Körnung“ für größere Einheiten, wie ganze Texte oder Bücher, besitzen

Hierzu ein Zitat von Warner:

“Linguistically very crude procedures seemed to work quite well for retrieval, with retrieval primarily understood as the transformation of a query into a set of records, and it was not clear what contribution could be obtained from more sophisticated procedures.” (S. 282)

Mögliche Kandidaten aus der Linguistik waren hier:

- Formale Sprachen der Computerlinguistik
- Forschung der Künstlichen Intelligenz

- Textlinguistische Theorien mit ausgesprochenen strukturalistischen und formalisierenden Eigenschaften

Die IW suchte bei der Linguistik Hilfe zur konkreten Problemlösung.

Vorherrschend eine positivistische Grundhaltung zum Text sowohl in der IW als auch in Teilen der Linguistik und v.a. Textlinguistik (z.B. Petöfi).

Text wurde in beiden Wissenschaften als Einheit mit abgegrenzten, physikalischen strukturellen Merkmalen angesehen, die wissenschaftlich und objektiv untersucht werden konnten.

4.2 Das kognitive Paradigma

Das nächste Paradigma, das ich vorstellen will, wird in der Literatur allgemein das „kognitive“ Paradigma genannt. Es folgte als Phase auf das physikalische Paradigma, aber beide Paradigmen haben wohl immer nebeneinander existiert.

- Das Thema eines Dokuments ist mit dem Wissen des Systembenutzers verknüpft
- Information ist etwas, das eine Lücke im Wissen eines Individuums füllen kann

Laut dem kognitiven Paradigma: Die IW soll Systeme konzipieren, die den Inhalt eines Dokuments mit dem individuellen Benutzerbedarf in Einklang bringen.

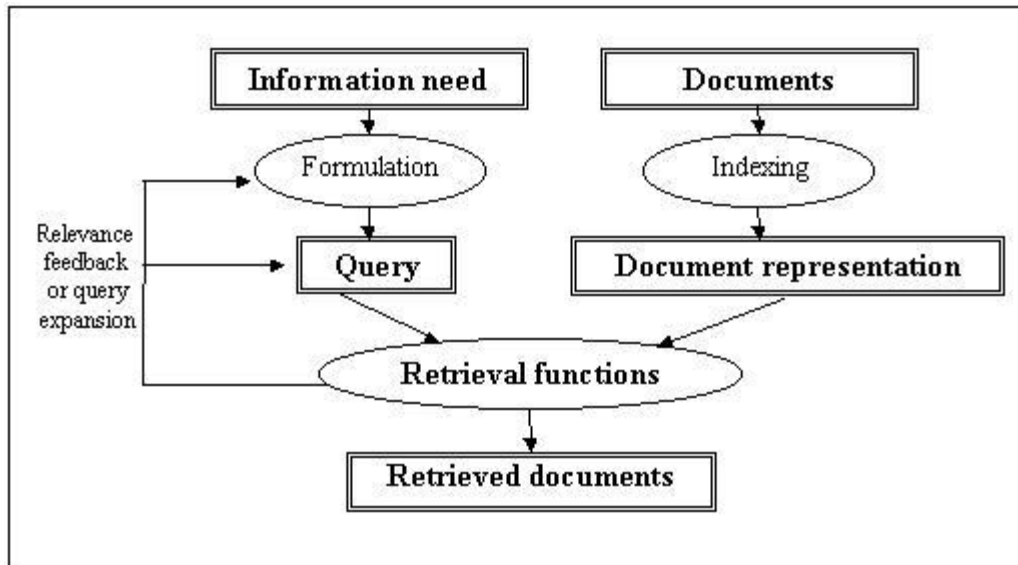
Diese Umorientierung in der theoretischen IW hat dazu geführt, die vorherrschende positivistische Haltung gegenüber Information zu hinterfragen.

Auch der Information retrieval hat sich geändert: Weg vom Prinzip des „präzisen“ Korrelierens von Indexdaten mit kontext- und überindividuellen Suchanfragen hin zu einem mehr individualistischen, vageren, und subjektiveren Begriff des Information retrieval-Prozesses.

Der Prozess der Informationssuche wird zur Verhandlung zwischen einer informationssuchenden Person und einem Informationssystem.

Welche linguistische „Schützenhilfe“ lässt sich aus der kognitiven Umorientierung in der IW ableiten?

Ich verwende zur Illustration meines Vorschlages wieder ein Schema des Information retrieval-Prozesses, diesmal aber bereichert mit den kognitiven Elementen des benutzerrelatierten Informationsbedarfs, der sprachlichen Formulierung desselben und der abschließenden Suchkette („query“):



Quelle: Gharaibeh, I. K., & Gharaibeh, N. K. (2012).

Das kognitive Umdenken in der IW setzte nach meiner Interpretation an den beiden Ästen, dem linken (Benutzerast) und dem rechten (Repräsentationsast) an. Beide Kritiken lenken die Aufmerksamkeit auf die Linguistik, da beide Äste direkt auf sprachliche Problemstellungen hinweisen.

Repräsentationsast (rechts): Das kognitive Umdenken leistete einem ausgeprägten Bedeutungsskeptizismus Vorschub, wo die naive Vorstellung einer einfachen semantischen Relation zwischen Dokumenten und ihren Repräsentationen, wie sie in der traditionellen Indexierungstheorie vertreten wurde, in Frage gestellt wurde.

Dies hat mit einer steigenden Skepsis gegenüber dem objektivistischen Erbe zu tun, das das kognitive Paradigma von den physikalischen Informationswissenschaftlern übernommen hatte und zu überwinden suchte.

Benutzerast (links): Das kognitive Umdenken beinhaltete den Schritt vom Systembenutzer zum Sprachbenutzer, was zu einer radikalen und kommunikativen Uminterpretation des Index-System-Benutzer-Komplexes führt.

Das kognitive An-sagen des linken (Rolle des Benutzers) und des rechten Asts (Rollen von Dokumentrepräsentationen) hatte eine Hinwendung zu spezifischen linguistischen Theorien und Traditionen geführt.

Die Hinwendung zur Linguistik (und Sprachphilosophie) lag nahe, da sowohl die kommunikative Neuinterpretation des Benutzers als auch die Verkomplizierung des Text-Repräsentations-Verhältnisses auf linguistischen Konzepten beruht.

Innerhalb des rechten Asts bezog man sich auf Konzepte der Ordinary Language Philosophy und Wittgenstein, wie schon zuvor an Blairs Wittgenstein-Rezeption beschrieben.

In Bezug auf den linken Ast wurde z.B. Searles Sprechakttheorie oder Grices Konversationsmaximen als Rahmentheorien für den handelnden Benutzer diskutiert. Auch hier ist die Arbeit des Informationswissenschaftlers Blair maßgebend.

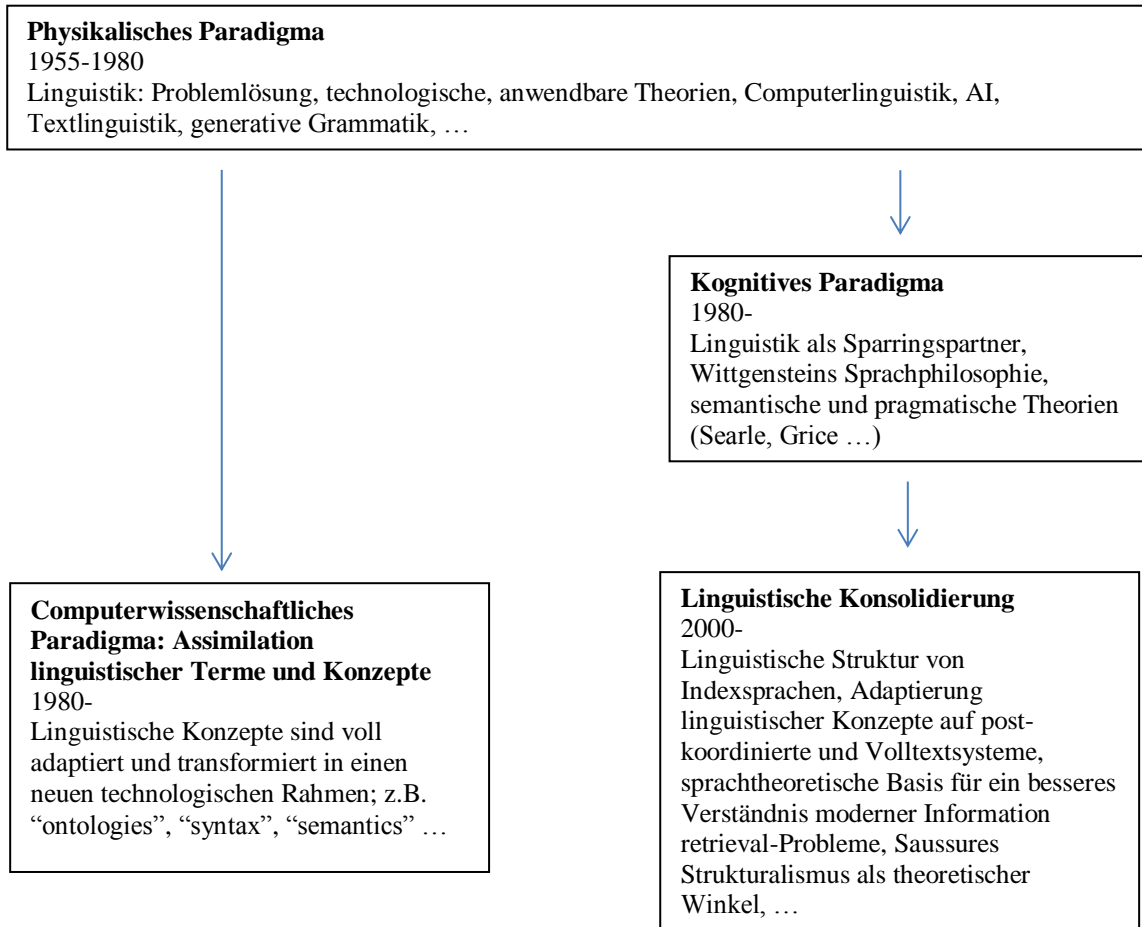
4.3 Zur Situation heute

Es ist schwer, zum „Ist-Zustand“ Stellung zu nehmen, wir befinden uns ja alle mittendrin. Doch meine ich, zwei Tendenzen im Verhältnis der IW zur Linguistik feststellen zu können.

1. Große Teile der angewandten Linguistik sind, als Erbe des physikalischen Paradigmas, in der Computerwissenschaft aufgegangen. Das zeigt sich an einer Verselbständigung der linguistischen Terminologie in diesem Zweig, die eine neue, linguistik-unabhängige und technologische Interpretation erhalten hat. Linguistische terminologische Ausdrücke wie ‚semantics‘, ‚morphological analysis‘, ‚intertextual semantics‘, ‚ontology‘, ‚semantic roles‘, ‚predicate-argument structures‘, ‚lexical analysis‘ haben in diesem Zweig der Computerwissenschaft neue Bedeutungen erhalten, die nichts mehr aus ihrer Ausgangsdomäne, der klassischen Linguistik, gemein haben. Ich habe hier eine kleine Untersuchung angestellt.
2. Die andere Tendenz ist eine Konsolidierung linguistischer Konzepte in neuerer informationswissenschaftlicher Grundlagenforschung. Traditionelle Konzepte wie „Syntax“, „Semantik“, „syntagmatische Relation“, „Paradigmatik“ u.a. haben in diesem linguistik-treuen Kontext informationswissenschaftlicher Forschung ihre linguistische Kernbedeutung erhalten und bringen diese in einen neuen informationswissenschaftlichen Kontext ein. Ein Beispiel hierfür ist der schon erwähnte Julian Warner.

4.4 Illustrativer Überblick über die Entwicklung

Ich möchte an folgender Illustration das Gesagte noch einmal im Zusammenhang klarmachen:



Literatur:

- Gharaibeh, I. K., & Gharaibeh, N. K. (2012). Towards arabic noun phrase extractor (ANPE) using information retrieval techniques. *Software Engineering*, 2(2), 36-42.
- Hjørland, B. (1998b). Theory and metatheory of information science: A new interpretation. *Journal of Documentation*, 54(5), 606-621.
- Petöfi, J. S., & Bredemeier, J. (1977). *Das Lexikon in der Grammatik - die Grammatik im Lexikon*. Hamburg: Helmut Buske.
- Warner, J. (2007b). Linguistics and information theory: Analytic advantages. *Journal of the American Society for Information Science and Technology*, 58(2), 275-285.
doi:<http://dx.doi.org/10.1002/asi.20488>

5. Synchrone Standortbestimmung der Informationswissenschaft in Bezug auf die Linguistik

Ist es möglich, eine „synchrone“ Standortbestimmung der IW und der Linguistik, zueinander, strukturell sozusagen, vorzunehmen?

Mein Vorschlag ist folgender: Ebenso wie man eine strukturelle Analyse von interagierenden Varietäten, z.B. Dialekten mit der Hochsprache, vornehmen kann, muss es möglich sein eine Disziplin, hier die IW, in ihrem Verhältnis zu einer anderen oder mehr Disziplinen, hier die Linguistik und andere „benachbarte“ Wissenschaften zu untersuchen.

Diese strukturelle Untersuchung setzt eine vergleichbare Grundeinheit voraus, die den zu untersuchenden Disziplinen gemeinsam ist und, um dem Kriterium der Abgrenzung zu genügen, systematisch in den einzelnen Disziplinen variiert.

Wieder gehe ich, wie schon zuvor in meiner „diachronen“ Analyse, von einer Information retrieval-Version als informationswissenschaftlicher Kernbereich aus.

Wie wird die IW mit der Linguistik und eventuell anderen Disziplinen sozusagen „zusammengehalten“, was ist diesen Disziplinen gemeinsam und womit können wir ihr gegenseitiges Verhältnis beschreiben?

Meine These ist: Es ist die grundlegende Vokabularbezogenheit des Information retrieval, das ihr Verhältnis zur Linguistik verstehbar macht.

Im Rahmen des Information retrieval werden

- Index“sprachen“ entwickelt und ausgebaut.
- Es werden kontrollierte Vokabulare mit lexikalischen Einheiten zum Zwecke der Informationsfindung entwickelt. Diese Vokabulare heißen Thesauri.
- Thesauri weisen eine semantische Strukturierung von Vokabularterminen auf: hierarchische Beziehungen (Vogel-Spatz), Synonymrelationen oder sogenannte assoziative Relationen, z.B. Gegenstand und Wissenschaft, die diesen Gegenstand untersucht (Sprache-Linguistik).

Diese Vokabularbezogenheit in Form von Thesaurusvarianten bringt die IW mit mindestens 3 anderen vokabularbezogenen Disziplinen in ein systematisches Verhältnis:

- Linguistische Lexikologie und Wortsemantik
- Fachsprachenforschung
- Terminologielehre.

Ich möchte zeigen, dass es eben diese Vokabularbezogenheit ermöglicht, die vier Disziplinen systematisch anhand ihrer spezifischen Vokabularkonzeptionen und ihrer differierenden Anforderung an die lexikalischen Zeichen, mit denen die jeweilige Wissenschaft arbeitet, in Beziehung zu setzen.

6. Vokabulartypen und Nachbardisziplinen der Informationswissenschaft

Ich möchte die 4 Disziplinen jetzt durchgehen und die spezifische Ausprägung ihres Vokabulars kurz besprechen.

6.1 *Thesauri in der IW*

Lexikalische Zeichen in den Thesauri der IW sind anhand ihrer Gebrauchsanforderungen in Suchkontexten zu charakterisieren:

Thesauruseinheiten müssen, um den Bedingungen des Information retrieval zu genügen,

- effektiv in Suchanfragen benutzt werden können, also suchbar sein.
- Darüber hinaus müssen Suchbegriffe dazu designt sein, relevantes Feedback an den Benutzer des Systems zu unterstützen („interactive information retrieval“).

Aus diesen fundamentalen Gebrauchsanforderungen ergeben sich mehrere formale Eigenschaften von informationswissenschaftlichen Zeichen, d.h. Thesaurustermen:

1. Thesauruszeichen müssen relevante und gängige Denotationen in Bezug auf ihre Benutzergruppe einführen.
2. Vokabulareinheiten in Thesauri müssen die semantischen Relation kodieren, welche für das terminologische System einer Disziplin charakteristisch sind. Nur das unterstützt Suchanfragen von Spezialisten.
3. Thesauruszeichen müssen die Einbettung in rudimentäre Syntagmen unterstützen, die, mit einer gewissen Wahrscheinlichkeit, auch in Benutzeranfragen eingehen.
4. Die Semantik von Zeichen in einem Thesaurus sollte der Eigenschaft der Kompositionalität von komplexen, zusammengesetzten Zeichen in der Weise genügen, dass eine parallele Strukturierung sprachlicher Bedeutungskomplexe auf der einen Seite und die Formulierung komplexer Problemsettings von Forschungsthemen in Suchanfragen auf der anderen Seite unterstützt wird.

Diese Prinzipien 1-4, die das „richtige“ Zeichenformat von Thesauruseinheiten regeln, sind in der Indextheorie (vgl. z.B. Lancaster), in der Theorie der Wissensorganisation, der Klassifikationstheorie und den Richtlinien für das Erstellen von Thesauri niedergelegt.

6.2 *Lexika in der Linguistik*

Ich möchte nun, ganz parallel, dieselben Fragen an linguistische Lexika für natürliche Sprachen stellen.

Sowohl Thesauri als auch natürlichsprachliche Lexika sind strukturierte Sammlungen von Zeichen, unterscheiden sich jedoch in ihren kommunikativen Anforderungen. Das hat wiederum Auswirkungen auf die formalen und designmäßigen Eigenschaften der jeweiligen Zeichen.

Natürlichsprachliche Zeichen müssen z.B.

- speicherbar und
- wiederfindbar sein in einem mentalen, biologischen System
- Lexikonzichen müssen genügend komplex und flexibel für strukturelle Variation sein.
- Sie müssen lernbar sein, sowohl im primären als auch im sekundären Spracherwerb.

Die sich daraus ergebenden formalen Anforderungen lexikalische Zeichen sind u.a. von der Morphologie oder der Wortbildungs- und Flexionstheorie untersucht worden.

Was macht sprachliche Zeichen z.B. lernbar? Lernbarkeit wird z.B. unterstützt durch einen gewissen Grad von morphologischer Transparenz, welche die Relationen paradigmatisch verbundener Zeichen in einem natürlichen Lexikon durchschaubar widerspiegelt.

Beispiele hierfür sind systematische Ableitungsbeziehungen zwischen adjektivischen Basen wie *dumm* und suffigierten Substantiven wie *Dumm-heit*.

Die womöglich wichtigste Eigenschaft natürlichsprachlicher Lexikoneinheiten sind solche formale und semantische Eigenschaften, welche einen gleitenden Übergang statischer Lexikoneinheiten (Wörter) in grammatisch komplexe Syntagmen (Sätze) und letztendlich kommunikative Strukturen wie Sprechakte unterstützen.

Es ist wohl das herausragende Kennzeichen natürlichsprachlicher Lexikoneinheiten, dass sie schon im Lexikon darauf vorbereitet sind, in syntagmatische und kommunikative sprachliche Einheiten einzugehen. Ich werde diesen Punkt später noch aufgreifen.

Die restlichen beiden Disziplinen behandle ich nur noch kurz.

6.3 Fachwörterbücher in der Fachsprachenforschung

Fachsprachen haben viele Funktionen, sie unterstützen z.B. die Professionsidentität. Sehen wir auf Forschung im Bereich der Fachlexika im Besonderen, ist da neben der Gruppenfunktion v.a. die Funktion von Fachvokabularen hervorgehoben worden, professionelle Fachkommunikation und kommunikative Kooperation zwischen Menschen, die in derselben Fachdomäne arbeiten, zu fördern.

Hiervon wurden eine Reihe mehr spezifische Kriterien für Fachsprachen und fachliche Vokabulare abgeleitet (Man spricht auch von „Postulaten“, vgl. Hoffmann et al.):

- Genauigkeit/Exaktheit
- Expliztheit
- Ökonomie
- Verständlichkeit

Diese Eigenschaften sind natürlich strukturell nicht ganz einfach zu fixieren, zudem bedingen sie einander (ein maximaler Wert kann immer nur auf Kosten eines niedrigen Wertes auf anderen Parametern erreicht werden).

Aber der spezifische Vokabularbezug dieser Disziplin sollten auch diese kurzen Bemerkungen schon begründen können.

6.4 Terminologien in der Terminologielehre

Die Terminologielehre umfasst die Systematisierung von Konzepten und ihrer Ausdrücke, die „Terminologien“ genannt werden (vgl. Hoffmann et al.).

Terminologien strukturieren fachliches und Domänenwissen und dienen dazu, Probleme in der Fachkommunikation zu lösen.

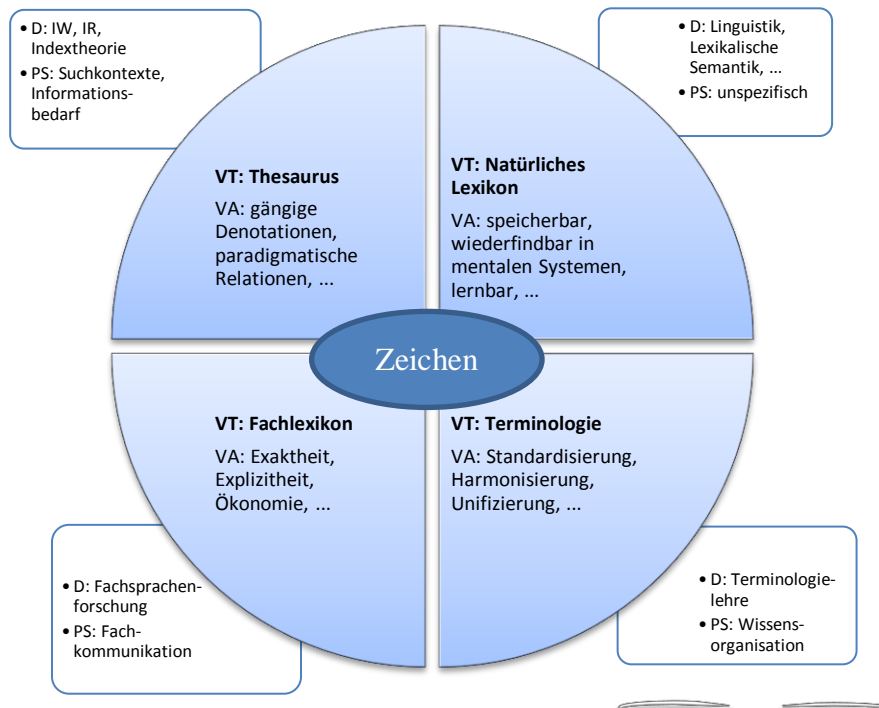
Terminologien können daher als extrem kontrollierte Vokabulare betrachtet werden, mit den Hauptzielen:

- Standardisierung
- Harmonisierung
- Unifikation von Terminologien.

6.5 Illustration: Linguistik und Informationswissenschaft im Feld mit 2 Nachbardisziplinen

Ich möchte jetzt das Verhältnis dieser 4 Disziplinen anhand von 4 Parametern überschaubar illustrieren:

- D: Disziplin (IW, Linguistik, ...)
- PS: Disziplinäres Problemsetting (Suchkontexte, professionelle Kommunikation, ...)
- VT: Vokabulartyp (Thesaurus, Lexikon, ...)
- VA: Vokabularanforderungen in Bezug auf VT und PS (gängige Denotate, wiederfindbar, ...)



Literatur:

Hoffmann, L., H. Kalverkämper & H. E. Wiegand (1998/1999) (Eds.), *Fachsprachen: Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin, New York: Gruyter.

Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice* (3rd ed.). London: Facet.

7. Strukturelle Analyse des Verhältnisses der 4 Disziplinen: Vokabularkontrolle

Können wir in diesem Schema ein organisierendes Prinzip entdecken? Ist es möglich, eine erklärenden Hintergrund auszumachen, warum die 4 Disziplinen so und nicht anderes miteinander verwoben sind?

Ich möchte hier eine mögliche Dimension diskutieren, welche den Disziplinen eine skalare Ordnung auferlegt, nämlich Vokabularkontrolle als die organisatorische Struktur, in die die jeweiligen Vokabulare eingebettet sind.

Vokabularkontrolle: absichtliches Design von Vokabularen durch menschliche Agenten, z.B. durch Regeln, Prinzipien und Richtlinien.

Hier befinden sich natürliche Lexika am unteren Ende der Kontrollskala (vgl. deskriptive vs. normative Zugänge zur Sprache)

Am oberen Ende: Terminologien, die den höchsten Grad von Vokabularkontrolle aufweisen.

Die beiden anderen Vokabulardisziplinen sind „Mischtypen“.

Fachvokabulare: Hier ist ein Kontrollaspekt zugegen, der von den 3 „Postulaten“ (Exaktheit, Ökonomie, Expliztheit) motiviert ist.

Dennoch: Die kommunikative Seite von Fachkommunikation scheint im Vordergrund zu stehen.

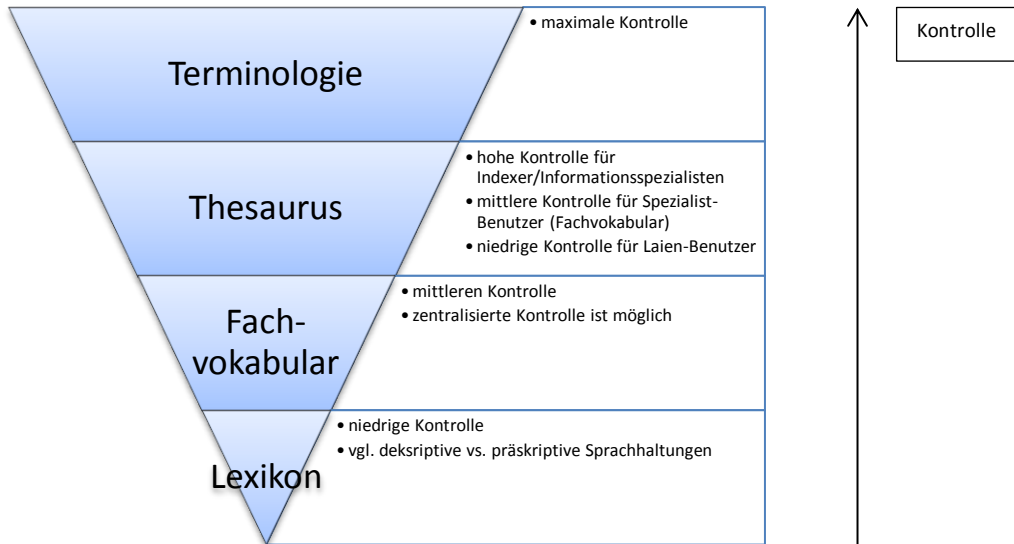
Thesauri: Sind definitorisch kontrollierte Vokabulare – kontrolliert in Bezug auf eine bestimmte Benutzergruppe.

Hier drehen sich also die Vorzeichen um in Bezug auf Fachlexika, und der Kontrollaspekt kann als dominierend angesehen werden. Dies gilt für die **professionellen Gruppen** der Indexer und Informationsspezialisten.

Von der **Benutzerseite** gibt es hier ein unkontrolliertes/unkontrollierbares Gegengewicht. Der Benutzer von Systemen wendet Ausdrücke aus seiner „normalen“ Sprache an, sei sie nun aus der **allgemeinen Standardsprache** oder aus einer **Fachsprache** (beim Fachspezialisten).

Diese beiden Seiten, das Kontrollinteresse der Systementwickler und Informationsprofessionellen und das Festhalten der Benutzer an unkontrollierten lexikalischen Spracheinheiten, machen die IW zu einem Hybrid mit kontrollierendem Übergewicht.

Folgende Illustration macht den Zusammenhang der 4 Disziplinen anhand des Kontrollparameters noch einmal deutlich:



8. Zusammenfassung und Ausblick; das Prinzip der Syntagmatisierung

Zusammenfassung

Ich hoffe, dass ich meine Ambitionen aus der Einleitung teilweise erfüllen konnte:

- Linguistik und IW sind attraktive Partner
- Es gibt ein großes Potential der Zusammenarbeit
- Sie finden die IW ist eine interessante Disziplin?

Teil I: Was ist die IW? Ich habe unterschieden:

- Essentielle Versuche
- Realistische Versuche
- Problemorientierte Versuche

Ich habe mich hier für den problemorientierten Winkel entschieden.

Ich habe dafür argumentiert, das Information retrieval mit Indexierung auf der einen Seite und dem Benutzer auf der anderen Seite als Grundkonstellation der IW anzusehen (die IW umfasst in der Realität, vgl. die realistische Zugänge, natürlich sehr viel mehr Bereiche).

Ich habe gezeigt, dass das Information retrieval-Modell der IW sprachliche Phänomene in einem wesentlichen Sinne miteinbezieht. Ohne linguistische Beteiligung und Erkenntnisse ist das Modell nicht hinreichend verstehbar.

Ich habe einige frühe und neuere Felder der Zusammenarbeit zwischen IW und Linguistik aufgezeigt wie Computerlinguistik, natürliche Sprachverarbeitung und mehr.

Teil II

Ich habe gezeigt, dass die Übernahme linguistischen Gedankenguts an die physikalische und die kognitive Phase bzw. Paradigma der IW gebunden ist.

Heute ist ein Teil des linguistischen Erbes in einer eigenen, technologischen Disziplin der Computerwissenschaft aufgegangen, wogegen in der Mainstream-IW Bestrebungen zu sehen sind, die Grundlagen der Wissenschaft durch sprachtheoretische Konzepte kommunikativ zu untermauern.

Teil III

Ich setzte im 3. Teil meines Vortrages die IW und die Linguistik als vokabularbezogene Disziplinen in ein strukturelles Verhältnis zur Fachsprachenforschung und Terminologielehre.

Ich arbeitete die Hypothese heraus, dass die 4 Wissenschaften jeweils aus ihren spezifischen Problemsettings heraus 4 verschiedene Sets von Anforderungen an ihre Vokabulare stellen,

und dass diese Anforderungen die Form und den Inhalt der jeweiligen Vokabularzeichen bestimmen.

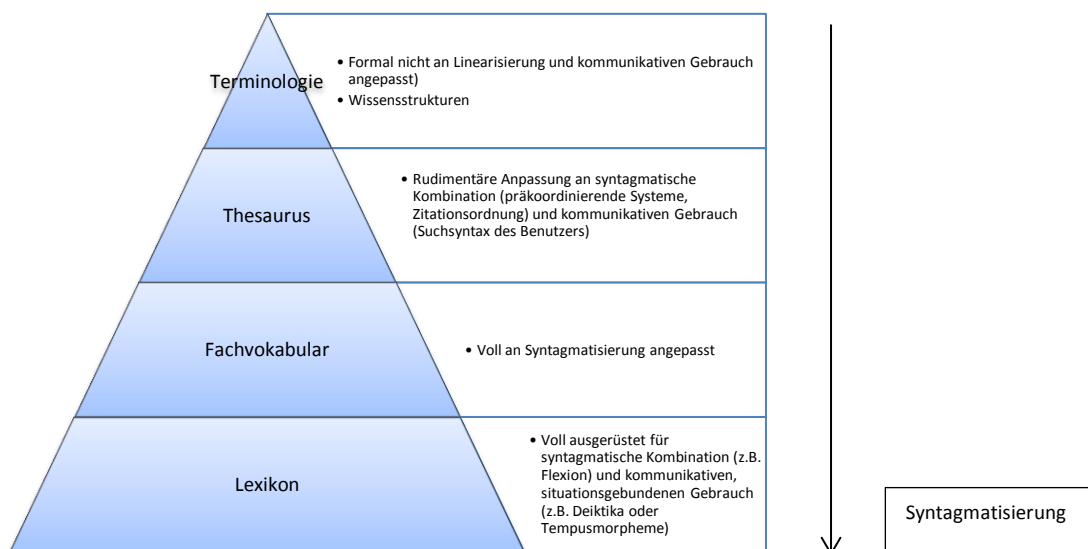
Ich diskutierte die skalare Dimension der Vokabularkontrolle als Strukturierungsprinzip für die 4 Vokabulartypen.

Ausblick

Ich meine, es gibt eine zur Vokabularkontrolle gegenläufig gerichtete Skala der Ausgeprägtheit von zeichenkombinatorischen Aspekten von Vokabulareinheiten.

Damit ist die Eigenschaft von Vokabulareinheiten gemeint, zeichentheoretisch dafür ausgerüstet zu sein, in lineare und syntaktisch strukturierte sowie kommunikativ verwendbare Ketten einzugehen.

Ich möchte hier nicht weiter in die Einzelheiten gehen, aber in einem abschließenden Diagramm den Grundgedanken zeigen:



Dieser Gedanke sollte noch weiter ausgearbeitet werden.

Ich hoffe, dass der Beitrag zur Ausweitung der interdisziplinären Beziehungen zwischen der Linguistik, der IW, Fachsprachenforschung und Terminologielehre beitragen kann. Ich meine, dass die gemeinsame Vokabularbezogenheit dieser Disziplinen an der Basis eines interdisziplinären Netzwerks liegen kann.

Es ist, aus dieser Perspektive gesehen, sicherlich nicht überraschend, dass eben diese Disziplinen traditionell sich den Konzepten und theoretischen Zugängen zu Sprache und dem Lexikon der jeweiligen anderen Disziplinen intensiv bedient haben. Jetzt wissen wir zumindest, warum.

Danke für Ihre Aufmerksamkeit.