



Informationswissenschaft und Linguistik. Ihr Verhältnis, historisch und jetzt

Volkmar Engerer
Informationswissenschaftliche Akademie



Ziele des Vortrages

- Attraktive Partner
- Interdisziplinäre Zusammenarbeit
- Interesse



Teil I: Generelles

1. Einleitung
2. Was macht/ist die Informationswissenschaft?
3. Felder des Kontakts zwischen Linguistik und Informationswissenschaft



Teil II: Chronologie (diachron)

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

Teil III: Vokabulare

5. Standortbestimmung der Informationswissenschaft in Bezug auf die Linguistik
6. Vokabulartypen und Nachbardisziplinen der Informationswissenschaft
7. Strukturelle Analyse des Verhältnisses der 4 Disziplinen (Vokabularkontrolle)
8. Zusammenfassung und Ausblick



2. Was macht/ist die Informationswissenschaft?



Realistischer Zugang

- Aufgaben der IW?
- Historische Entwicklung der IW?
- Verbindungen zu anderen Disziplinen?

Essentieller Zugang

- Information,
- Wissen und
- Dokument?

Problemorientierter Zugang

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

Essentiell/1: Information ist "Wissen in Bewegung"

- Poppers Welten

- 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)

Essentiell/2: Information – Bucklands semantischer Winkel

- 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

Essentiell/3: "Wissenschaft" in "Informationswissenschaft"

Wissenschaft mit positivistischer Methodologie



Realistische Zugänge

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



Problemorientierter Zugang

Die IW untersucht

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



Abwägung der 3 Zugänge

- Essentiell
- Realistisch
- Problemorientiert
 - Lösungsfördernd
 - Allgemein und spezifisch
 - Benutzerintegrierend
 - Technologisch
 - Interdisziplinär
 - Zusammenhang
 - Domänenvielfalt
 - Sprachfokussiert



Information retrieval

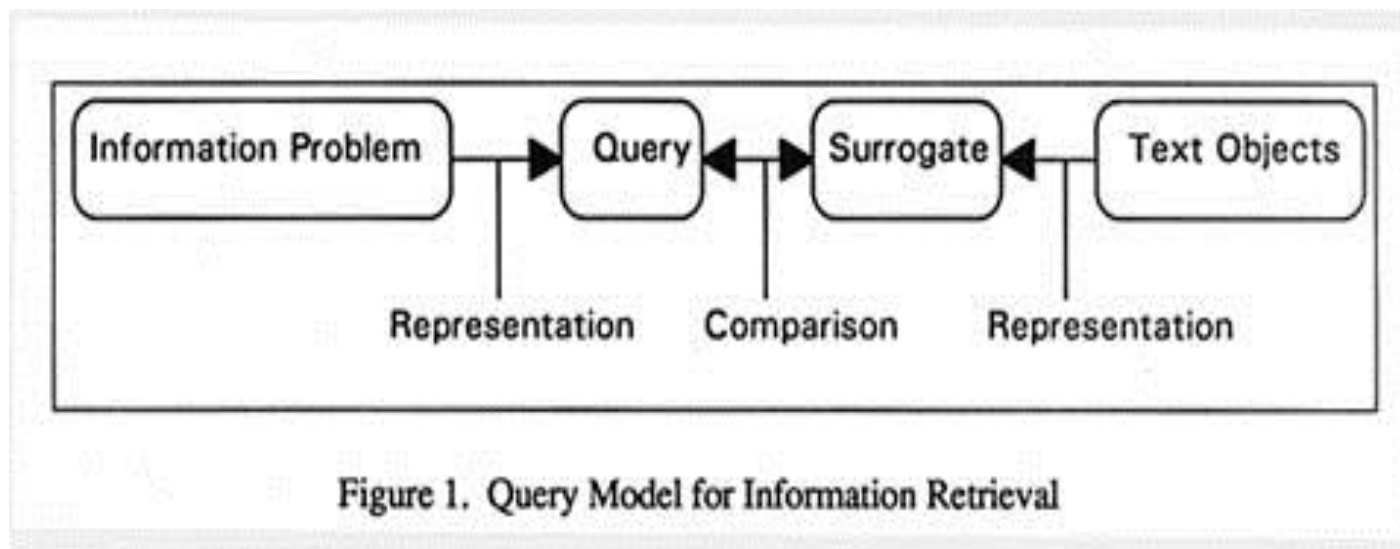


Abb. nach Kevin Cox, Information Retrieval in Browsing (<http://web.simmons.edu/~chen/nit/NIT'92/069-cox.htm>). Original nach Belkin, N.J. & Croft, W.B. (1987). "Retrieval techniques," In *Annual Review of Information Science and Technology* (ARIST), 22: 109-189.

Sprachliche Repräsentationsebenen, „von links“

- Natürlichsprache Repräsentation des Informationsbedarfs des Benutzers
- Suchmaschinenorientierte Repräsentation
- „Intermediate“
- Iteratives sprachliches Feedback

Sprachliche Repräsentationsebenen, "von rechts"

- Referentialität („aboutness“)
- Kontrolliertes, maschinennahes Vokabular
- Sprachgebrauch eines zukünftigen Benutzers
- 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

3. Felder des Kontakts zwischen Linguistik und Informationswissenschaft



Sparck Jones & Kay

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

Das 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.

Computerlinguistik

- Natürliche Sprache
- Ambiguität
- Syntaktisierung
- Semantische Struktur



Prä- und postkoordinierende Indexierung

- Präkoordinierende Systeme: stark syntaxbasiert

Subjects

[United States. Army. Cavalry--History--Civil War, 1861-1865.](#)

[Confederate States of America. Army. Cavalry.](#)

- Blake, Michael, 1943- American Civil War cavalry. London, Almark Publishing, 1973.
- Postkoordinierende Systeme: die Rolle der Syntax ist minimal

Emner: Chomsky, Noam; Sprog; Sprogvidenskab og alment; Minimalist theory (Linguistics); Generative grammar

- Seuren, P. (2004). *Chomsky's minimalism*. New York: Oxford University Press.



Andere wichtige Quellen

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



4. Versuch einer systematischen Chronologie

Das physikalische Paradigma in der IW

- 1950er/1960er Jahre
- IW als Wissenschaft des Information retrieval
- Information als Phänomen der physikalischen, „wirklichen“ Welt
- Bedeutung von Information wurde nicht thematisiert



Linguistische Theorien des physikalischen Paradigmas

- Anwendbar
- Maschinenimplementierbar
- Richtige „Körnung“

- “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)

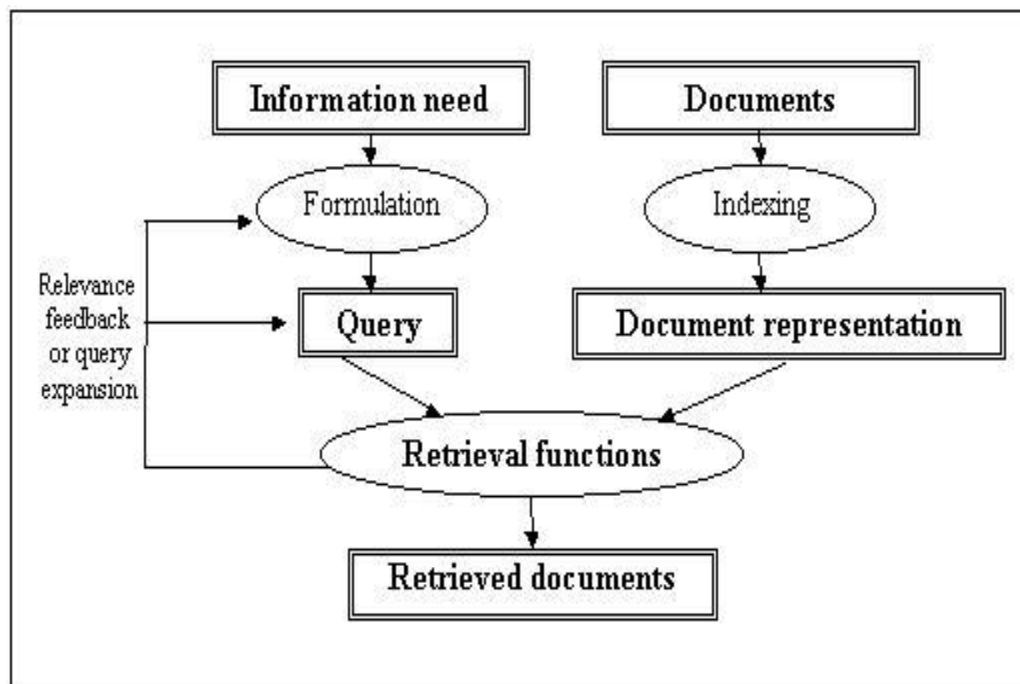
- Warner, J. (2007b). Linguistics and information theory: Analytic advantages. *Journal of the American Society for Information Science and Technology*, 58(2), 275-285.
doi:<http://dx.doi.org/10.1002/asi.20488>



Das kognitive Paradigma in der IW

- Dokumentthema und Benutzerwissen
- Information als Lückenfüller

Erweitertes Schema des Information retrieval



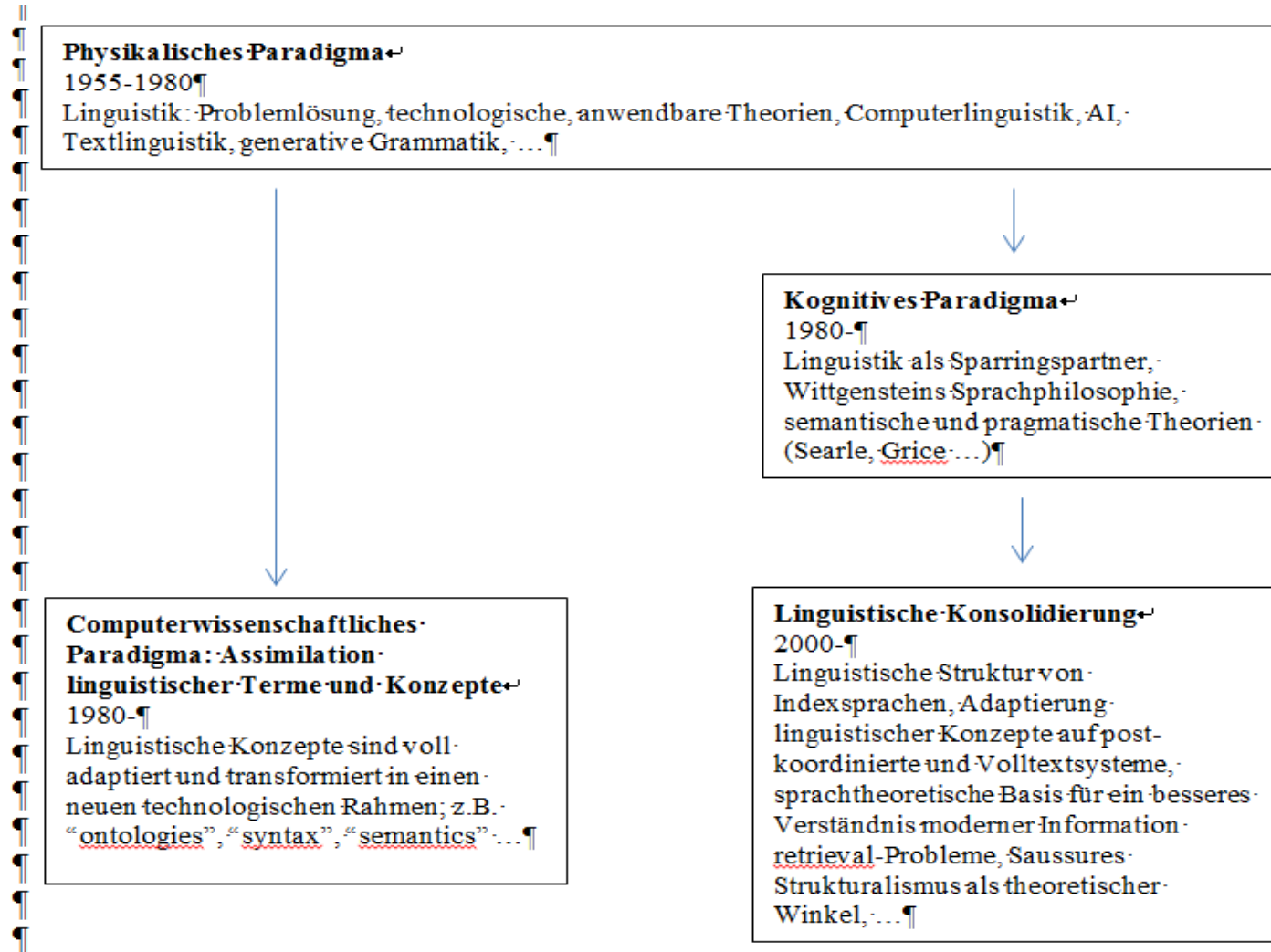
Quelle: Gharaibeh, I. K., & Gharaibeh, N. K. (2012). Towards arabic noun phrase extractor (ANPE) using information retrieval techniques. *Software Ingeneering*, 2(2), 36-42.

Zur Situation heute

- Verselbständigung linguistischer Terminologie und Theorie in einem Zweig der Computerwissenschaft
- Konsolidierung linguistischer Konzepte in der informationswissenschaftlichen Grundlagenforschung



Illustration: Linguistik und IW seit den 1950er Jahren



5. Synchrone Standortbestimmung



Synchrone Fragestellungen

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



Vokabularbezogenheit des Information retrieval

- Index"sprachen"
- Kontrollierte Vokabulare (Thesauri)
- Semantische Strukturierung von Vokabulartermen



6. Nachbardisziplinen (im Sinne der Vokabularbezogenheit)



Beteiligte Disziplinen

- Linguistische Lexikologie und Wortsemantik
- Fachsprachenforschung
- Terminologielehre

Thesauri in der IW

- Anforderungen:
 - effektiv in Suchanfragen, suchbar
 - relevantes Feedback an den Benutzer des Systems („interactive information retrieval“).



Eigenschaften von Thesauruszeichen

- Relevante und gängige Denotationen
- Semantische Relation charakteristisch für das terminologische System einer Disziplin
- Einbettung in rudimentäre Syntagmen
- Kompositionalität und parallele Strukturierung sprachlicher Bedeutungskomplexe und komplexer Problemsettings



Lexika in der Sprachwissenschaft

Anforderungen:

- speicherbar
- wiederfindbar
- lernbar, sowohl im primären als auch im sekundären Spracherwerb



Eigenschaften von Lexikonzeichen

- Morphologische Transparenz
- Vorbereitet für den Übergang in Syntagmen und kommunikative Einheiten (z.B. Flexion, Aspekt und Tempus, Deiktika, ...)



Fachsprachenforschung

„Postulate“:

- Genauigkeit/Exaktheit
- Explizitheit
- Ökonomie
- Verständlichkeit

Terminologielehre

- Standardisierung
- Harmonisierung
- Unifikation von Terminologien

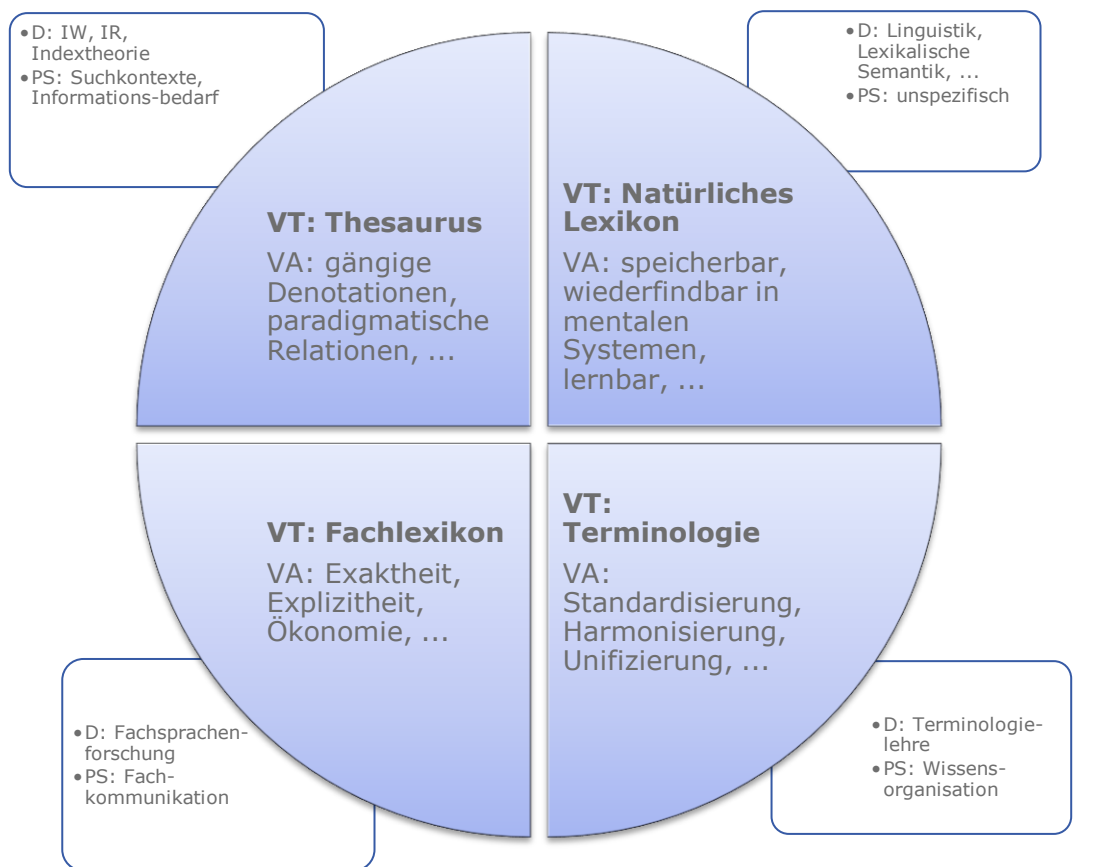
Linguistik und IW im Feld mit 2 Nachbardisziplinen

4 Parameter:

- 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, ...)



Illustration: Vokabulardisziplinen



7. Vokabularkontrolle

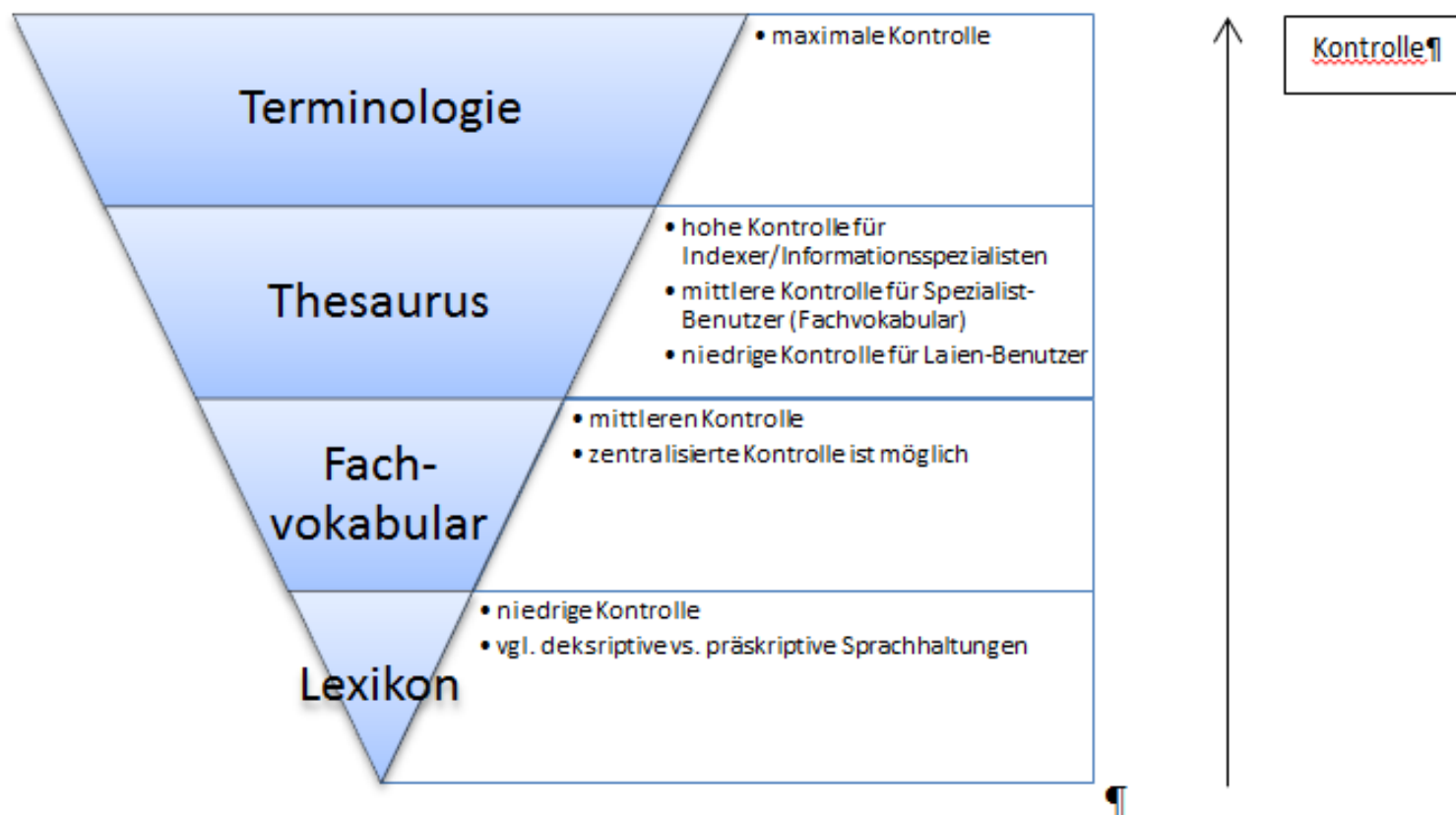


Motivation einer strukturellen Analyse

- Können wir hier ein organisierendes Prinzip entdecken?
- Ist es möglich, einen erklärenden Hintergrund auszumachen, warum die 4 Disziplinen so und nicht anders miteinander verwoben sind?



Kontrolle von 4 Vokabularen



8. Zusammenfassung, Teil I

1. Essentielle, realistische und problemorientierte Versuche. Entscheidung für problemorientierten Winkel.
2. Information retrieval mit Indexierung auf der einen Seite und dem Benutzer auf der anderen Seite als Grundkonstellation der IW
3. Information retrieval-Modell bezieht sprachliche Phänomene in einem wesentlichen Sinne mit ein.
4. Felder der Zusammenarbeit zwischen IW und Linguistik



Zusammenfassung, Teil II

1. Übernahme linguistischen Gedankenguts ist an die physikalische und die kognitive Phase der IW gebunden ist.
2. Heute: a) eigene, technologische Disziplin der Computerwissenschaft vs. b) Bestrebungen in der Mainstream-IW, die Grundlagen der Wissenschaft durch sprachtheoretische Konzepte kommunikativ zu untermauern.
3. Illustration

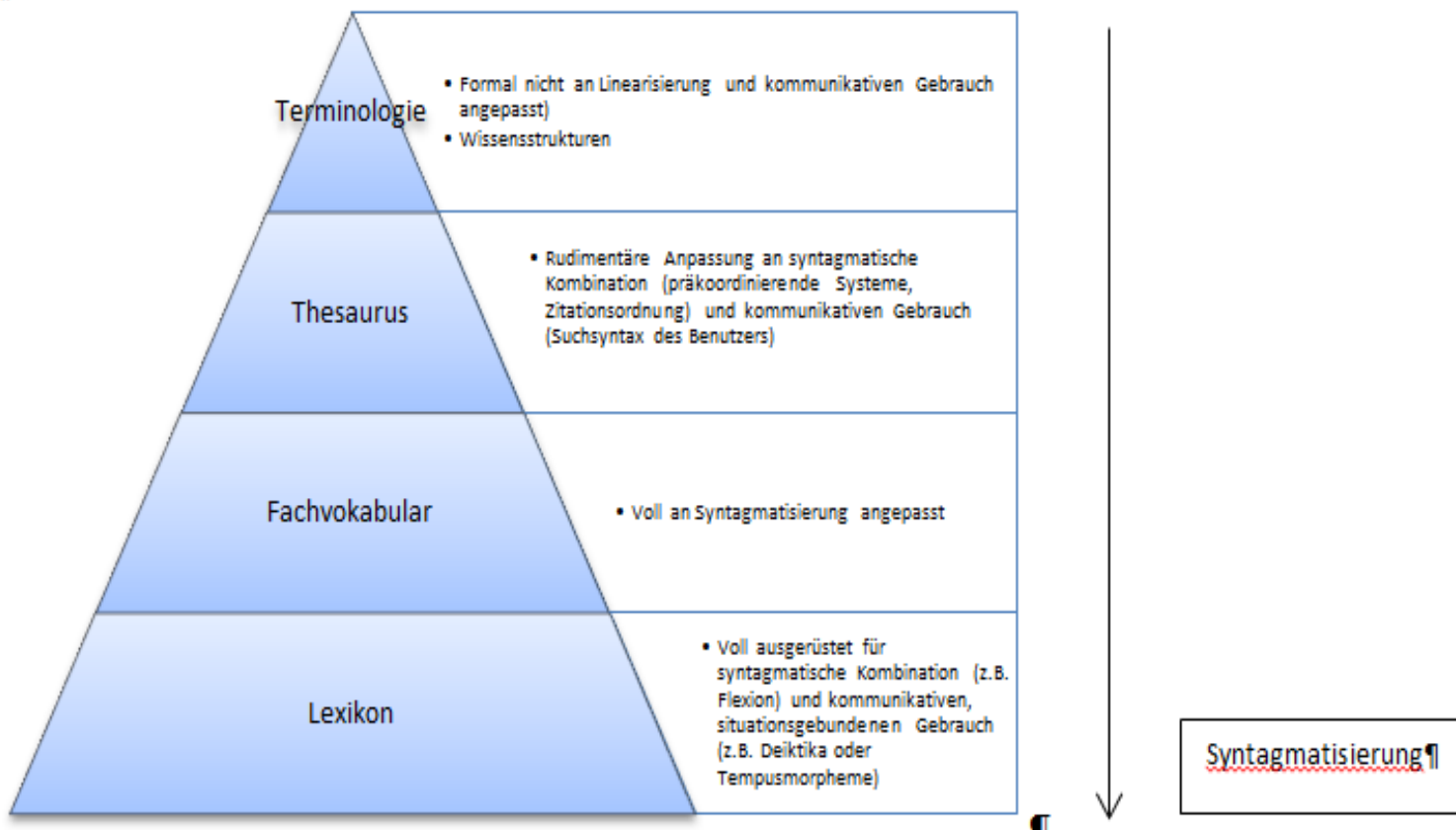


Zusammenfassung, Teil III

1. IW und die Linguistik unterhalten als vokabularbezogene Disziplinen in ein strukturelles Verhältnis zur Fachsprachenforschung und Terminologielehre
2. 4 Wissenschaften, 4 Problemsettings und 4 verschiedene Sets von Anforderungen an 4 verschiedene Vokabulartypen
3. Konstellationen in (2.) bestimmen die Anforderungen an Form und den Inhalt der jeweiligen Vokabularzeichen
4. Skalare Dimension der Vokabularkontrolle als Strukturierungsprinzip



Ausblick: Syntagmatisierung



Schluss

Danke.



