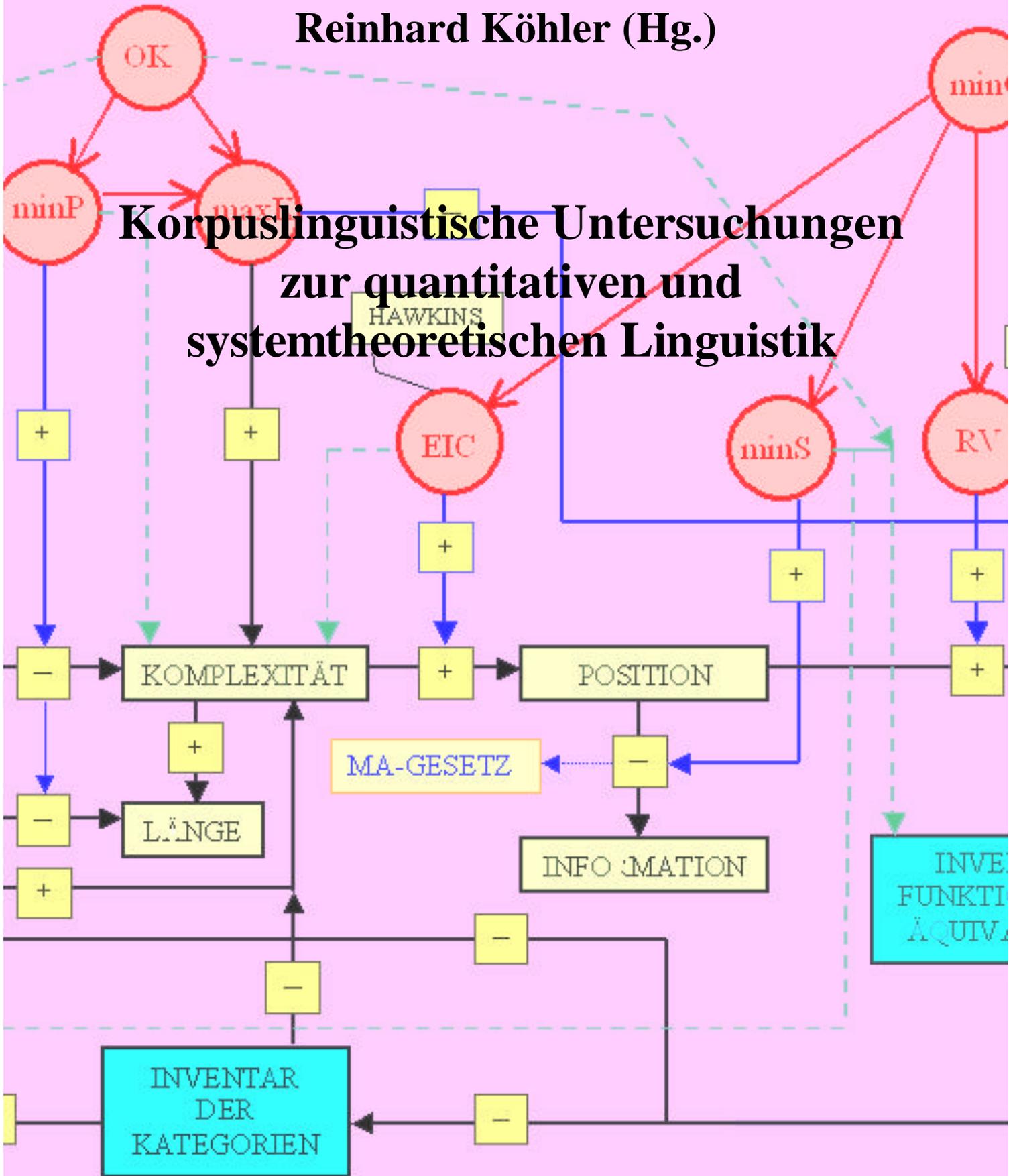


Reinhard Köhler (Hg.)

# Korpuslinguistische Untersuchungen zur quantitativen und systemtheoretischen Linguistik



# Inhalt

Vorwort	iii
<b>Hoffmann, Christiane &amp; Krott, Andrea</b> Einführung in die Synergetische Linguistik	1
<b>Hoffmann, Christiane</b> „Early immediate constituents“ – ein kognitiv-funktionales Prinzip der Wortstellung(svariation)	31
<b>Krott, Andrea</b> Ein funktionalanalytisches Modell der Wortbildung	75
<b>Bohn, Hartmut</b> Untersuchungen zur chinesischen Sprache und Schrift	127
<b>Menzel, Cornelia</b> Köhlers Basismodell der Lexik und die chinesische Schrift	179
<b>Steiner, Petra</b> Polylexie und Kompositionsaktivität in Text und Lexik	209
<b>Kablau, Myriam</b> Vom Konkreten zum Abstrakten? Bedeutungswandel von Wörtern	253
<b>Prün, Claudia</b> Die linguistischen Hypothesen von G.K. Zipf	271
<b>Mehler, Alexander</b> Konnotative Textbedeutungen Zur Modellierung struktureller Aspekte der Bedeutung von Texten	323
<b>Baskevic, Valentina</b> Regularitäten der lexikalischen Kombinierbarkeit	351
<b>Köhler, Reinhard</b> Syntaktische Strukturen, Eigenschaften und Zusammenhänge	365
<b>Giesecking, Kathrin</b> Untersuchungen zur Synergetik der englischen Lexik	387
<b>Aichele, Dieter</b> Bibliographie zur Systemtheoretischen Linguistik	435

# Vorwort

Dieser Band stellt eine Sammlung von neueren korpuslinguistischen Arbeiten aus der Trierer Forschungsgruppe zur quantitativen und systemtheoretischen Linguistik dar – Arbeiten, die zumeist im theoretischen Rahmen der synergetischen Linguistik angesiedelt sind und einen größeren, auf Korpusuntersuchungen beruhenden empirischen Anteil aufweisen.

Da einige der Beiträge in diesem Band ohne Kenntnis des Forschungsrahmens der synergetischen Linguistik (SL) nicht verständlich wären, beginnt die vorliegende Sammlung mit einer Einführung von Christiane Hoffmann und Andrea Krott in das Grundmodell, in die Denkweise und in die Notationskonventionen der SL, die auch eine kommentierte Auswahlbibliographie zu diesem Thema einschließt.

Das EIC-Prinzip („Early immediate constituents“) von Hawkins zur Erklärung von Präferenzen der Wortstellung (kurze vor langen Konstituenten) ist Thema des Beitrags von Christiane Hoffmann. Sie führt in die Fragestellung aus typologischer, funktional-linguistischer und kognitiv-performanzbasierter Sicht ein und fasst Hawkins' Hypothese, die diese Präferenz mit Vorteilen bei der Sprachverarbeitung erklärt, sowie seine eigenen empirischen Studien zusammen. Aufgrund ihrer Kritik an Hawkins' empirischer Methodik entwickelt die Autorin eigene empirische Ansätze und führt Untersuchungen anhand mehrerer Korpora durch. Ihre Ergebnisse bilden eine nunmehr methodisch fundierte, starke Unterstützung für die EIC-Hypothese.

Andrea Krott stellt ein morphologisches Teilmodell auf der Grundlage der SL vor, das sich auf die Wortbildung konzentriert. Das Modell umfasst eine Reihe von Hypothesen, die sie mit empirischen Untersuchungen anhand von Daten aus drei Sprachen (Deutsch, Englisch, Niederländisch) überprüft. Außerdem diskutiert sie die Frage, ob die Oszillationen, die in mehreren Studien zur Lexik (die Variable Länge zeigt auf der Frequenz-Achse einen Schwingungsverlauf, wenn die Daten mit Hilfe gleitender Mittel geglättet werden) beobachtet worden sind, real oder artefaktisch sind.

Das Chinesische ist durch eine Reihe von im Vergleich zu anderen Sprachen extremen Merkmalsausprägungen gekennzeichnet. Dadurch ist diese Sprache besonders interessant im Hinblick auf die Überprüfung von Gesetzhypothesen, die ja Gültigkeit für alle Sprachen und alle Zeiten beanspruchen. Die Arbeit von Hartmut Bohn untersucht zum ersten Mal die Gültigkeit des Menzerath-Altmann-Gesetzes auf verschiedenen linguistischen Analyse-Ebenen an chinesischen Textdaten. Ein weiterer Gegenstand ist die Rang-Frequenz-Verteilung von Wörtern in abgeschlossenen chinesischen Texten. Schließlich geht Bohn der interessanten Vermutung eines Zusammenhangs zwischen der Frequenz chinesischer Schriftzeichen und ihrer Phonetizität nach. Mit dieser Korpusuntersuchung gelingt erstmals ein objektiver Nachweis der Existenz dieses in der Sinologie lange diskutierten Phänomens.

Ebenfalls mit dem Chinesischen befasst sich Cornelia Menzel, die das lexikalische Basismodell der SL auf das chinesische Schriftzeichensystem überträgt. Ähnlich der Lexik bildet dieses ein Inventar mit analogen Eigenschaften und Zusammenhängen. Funktionskomplexität, graphische Kom-

ventar mit analogen Eigenschaften und Zusammenhängen. Funktionskomplexität, graphische Komplexität (operationalisiert als Strichzahl, Graphemzahl und durch ein Aufwandsmaß für die manuelle Herstellung eines Zeichens) und Frequenz sind die untersuchten Größen; ihre direkten und indirekten Abhängigkeiten untereinander Gegenstand der empirischen Überprüfung. Als Datenmaterial wird ein 1,8 Mio Kurzzeichen umfassendes Korpus verwendet.

Ausgangspunkt für Petra Steiners Studie ist G. Altmanns 1989 veröffentlichte Hypothese zur Abhängigkeit der Zahl der Komposita, die ein Wort bildet, von seiner Polylexie. Steiner untersucht anhand von deutschsprachigem Wörterbuch- und Korpusmaterial (VDI-Nachrichten, Spiegel und eine Novelle) die funktionalen Zusammenhänge zwischen der Polylexie einerseits und der Komposita-Zahl bzw. der Kompositionsaktivität andererseits.

Myriam Kablau setzt sich theoretisch und empirisch mit der Hypothese der Gerichtetheit des Bedeutungswandels von Wörtern auseinander, die eine Veränderung vom Konkreten hin zum Abstrakten im Laufe der Zeit behauptet. Wichtig ist in diesem Zusammenhang die Klarstellung des Unterschieds zwischen Abstraktheit/Konkretheit gegenüber Allgemeinheit/Spezifität. Bedeutungswandel und Bedeutungsübertragung deutscher Wörter werden in diesem Beitrag an Material aus dem Alt-, Mittel- und Neuhochdeutschen untersucht, Typen von Bedeutungsveränderungen klassifiziert und Verteilungen ihrer Spezifizierungen festgestellt. Das Ergebnis bestätigt die Hypothese nicht.

Eine Gesamtschau der linguistischen Hypothesen von George Kingsley Zipf bietet Claudia Prün mit ihrer exzellenten Einführung in die Gedankenwelt des Begründers der quantitativen Linguistik, in seine wichtigsten Prinzipien (principle of least effort und principle of relative frequency), seine Vermutungen über sprachliche Größen, Einheiten und Konfigurationen, sein Werkstatt- und das Glockenmodell, in die Hypothesen über Gleichgewichtsbeziehungen, die grundlegenden Kräfte Unifikation und Diversifikation, die sprachformenden Prozesse wie Spezifikation, Kristallisation, Kürzung, Intensivierung und Erhöhung der Polysemie. Sie zeigt, wie Zipf bestrebt ist, über bloße Beschreibungen hinaus zu gelangen und zu Erklärungen zu kommen, wozu er vor allem Ökonomieprinzipien wie optimale Inventarausnutzung und Inventarminimierung sowie die Trägheit heranzieht. Besonders die vielfältige Vernetzung seiner Ideen mit psycho- und soziologischen Konzepten ähnelt in vielem bereits modernen systemtheoretischen Ansätzen, was vor allem durch die Zusammenschau seiner Hypothesen in der Notation der synergetischen Linguistik verblüffend augenfällig wird.

Die Kombinierbarkeit lexikalischer Einheiten ist eine quantitative Eigenschaft, die in den letzten Jahren immer mehr Beachtung findet. Valentina Baskevic war eine Zeitlang als Gastwissenschaftlerin Mitglied der Trierer Forschungsgruppe. Sie gehört zu den Wissenschaftler(inne)n um Prof. Viktor Levickij in Cernivcy (Ukraine), die, vor allem am Deutschen, intensive Studien zu dieser Frage betreiben, um ein Begriffssystem zur lexikalischen Kombinierbarkeit zu schaffen und empirische Daten zu erheben. Die vorliegende Arbeit entstand in Baskevics Trierer Zeit; die Daten stammen aus DDR-Zeitungskorpora.

In meinem Beitrag zu diesem Band wird ein erster Versuch beschrieben, ein synergetisch-linguistisches Teilmodell syntaktischer Größen und Prozesse aufzustellen und zu testen. In das vorgestellte Modell werden u.a. bereits bestehende Hypothesen (Hawkins' EIC-Prinzip über Wortstellungspräferenzen, Yngves Hypothese zur Rechtsverzweigung) aufgenommen und integriert. Die empirische Überprüfung des Modells geschieht anhand des englischsprachigen Susanne-Korpus.

Ein Modell des konnotativen Aspekts der Bedeutung von Texten auf der Grundlage der unscharfen Mengen ist Gegenstand der Arbeit von Alexander Mehler. Er stellt ein Maß für die Ähnlichkeit von Texten in Bezug auf Kookkurrenzen lexikalischer Einheiten unter Bezugnahme auf den Distributionalismus nach Harris auf. Wesentlich für den Ansatz ist die Auffassung, dass die Bedeutung eines Textes nicht nach dem Kompositionalitätsprinzip zustande kommt, sondern aufgrund des Kontext-

prinzips, der wechselseitigen Abhängigkeit der Bedeutungen aller Textkonstituenten. Im Rahmen seiner Untersuchung entwickelt Mehler einen Algorithmus zur automatischen Textanalyse für die Wort- und Textebene unter vorläufiger Aussparung der Zwischenebenen, der inhaltlich verwandte Texte einander zuordnet. Grundlage für den empirischen Teil der Arbeit ist ein Korpus aus 502 Artikeln der Süddeutschen Zeitung.

Kathrin Giesecking überprüft anhand von drei Korpora und einem längeren Einzeltext das lexikalische Basismodell der SL am Englischen. Sie betrachtet als Größen Länge, Polylexie, Polytextie und Frequenz von Wörtern und testet die zwölf möglichen Abhängigkeiten zwischen diesen Größen (außer den direkten auch indirekte und zweifach indirekte Abhängigkeiten). Der Beitrag enthält außerdem eine ausführliche Diskussion der Ergebnisse und methodische Grundsatzüberlegungen.

Den Band schließt eine sorgfältige, 500 Titel umfassende, von Dieter Aichele zusammengestellte Bibliographie.

Mit dieser Sammlung wird nur ein kleiner Teil der einschlägigen Forschungsbeiträge veröffentlicht, die in den letzten Jahren in Trier entstanden sind, aber sie bietet einen repräsentativen thematischen und methodischen Überblick über die aktuellen Arbeiten dieser Gruppe.

Trier im Juli 2002

Reinhard Köhler