

EDV-gestützte Lexikographie am Beispiel idiomatischer Wörterbücher: Erfahrungen und Desiderate

Schemann, Hans

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schemann, H. (2000). EDV-gestützte Lexikographie am Beispiel idiomatischer Wörterbücher: Erfahrungen und Desiderate. *Historical Social Research*, 25(3/4), 257-263. <https://doi.org/10.12759/hsr.25.2000.3/4.257-263>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

EDV-gestützte Lexikographie am Beispiel idiomatischer Wörterbücher: Erfahrungen und Desiderata

*Hans Schemann (Heidelberg)**

Die Lexikographie – und bis zu einem bestimmten Grad auch die Lexikologie – ist mit den neuen technischen Möglichkeiten, insbesondere dem Computer, in ein grundsätzlich neues Stadium getreten.

Schon aus äußerlich-formalen Gründen läßt sich der Wortschatz einer Sprache – dessen Bestand in einer Kultursprache bis zu 250.000 und mehr Einheiten geht, von denen passiv zwischen 10 und 40/45.000, aktiv zwischen 5 und 15/20.000 "gebraucht" werden mögen (die Statistiken und Schätzungen schwanken beträchtlich!) – ungleich leichter erfassen als je zuvor: Abstände der Einheiten voneinander, Schriftgröße und Schrifttyp, Trennungen, Zusammensetzungen, Umstellungen, Nummerierungen, Kalkulationen, Simulationen u.a.m., bei einem im Umfang nahezu unbeschränkten Material, erlauben Zusammenstellungen, Übersichten, Grundlagen, Fragestellungen, für die früher das ganze Leben eines Einzelnen nicht reichte.

Was das für die strukturelle Erkenntnis des Wortschatzes bedeutet, wird am leichtesten klar, wenn man sich die unterschiedlichen "Worttypen" vor Augen hält:

geh - en	Stamm/Wurzel + Endung (Verb) - Simplex
vor-/ gehen	Simplex + Präfix
voran-/ gehen	Simplex + Präfix
aus-/gehen	Simplex + Präfix
be- geh - bar	Präfix - Stamm - Suffix + Kategorienwechsel (Verb -> Adjektiv)
kaputt - gehen	Präfixe + Suffixe + Infixe: Affixe phraseo logische Einheit
in die Brüche / gehen	Idiom
in die Binsen / gehen	Idiom
e.S. auf den Grund / gehen	Idiom
USW.	

Gibt man den Wortschatz nun sinnvoll gliedert ein – nach den sog. minimalen bedeutungstragenden Einheiten ("Morphemen"), läßt sich durch relativ einfache Prozeduren die Struktur des gesamten Wortschatzes einer Sprache transparent machen: So wie es im Deutschen Tausende von "Worten" und

* Aus dem Protokoll des 79. Kolloquiums über die Anwendung der Elektronischen Datenverarbeitung in den Geisteswissenschaften an der Universität Tübingen vom B. Juli 2000.

"Wortfügungen" gibt, die mit *geh/gehen* gebildet werden, so setzt sich die Sprache als ganze schwerpunktmäßig aus bestimmten Wurzeln und Simplexformen zusammen.

Es genügt hier anzudeuten, welche Möglichkeiten ein derart transparent gemachter Wortschatz für lexikographische, lexikologische, semantische, pragmatische, stilistische, sprachsoziologische und andere Fragestellungen bietet.

Angewandt auf mehrere Sprachen, kommt man so zu einer umfangreichen und detaillierten Kontrastierung: *gehen* und seine "Zusammenfügungen" im Deutschen, *aller* und seine "Zusammenfügungen" im Französischen usw.: Die Sprachen werden in einer Weise vergleichbar, wie das bisher nicht möglich war. Angedeutet seien nur einige Fragestellungen: Wo entsprechen sich die Bildungen, wo nicht? Welche Bedeutungen werden von welchen Wurzeln und *Simplices* her gewonnen? Welches sind die eigentlichen "Kern-Stämme" einer Sprache? Gibt es so etwas wie Kern-Stämme "*der*" Sprache? Welche Formen der "Zusammenfügungen" ("Morphologie" im weitesten Sinn) gibt es? Wo und wie entstehen neue Bedeutungen? Wo lassen sich diese neuen Bedeutungen komponentiell erklären (logisch-relational), wo nicht? Welche Faktoren spielen bei der Bedeutungsänderung eine Rolle und wie sind diese Faktoren gewichtet? usw. – Die Linguistik steht erst am Anfang, die hier skizzierten Möglichkeiten auszuschöpfen.

Habe ich eine Sprache in dieser Form im Computer einmal (mehr oder weniger) vollständig erfaßt, läßt sich mühelos umstellen, einfügen, ergänzen, weglassen usw. Die Wörterbücher können heute aktuell sein – d.h. den Sprachstand wiedergeben, der zur Zeit ihrer Erstellung *de facto* herrscht – während sie noch bis vor relativ kurzer Zeit hinter dem aktuellen Sprachstand hinterherhinkten.

Liegen nun die Entsprechungen zweier oder mehrerer Sprachen einmal vor, gilt *mutatis mutandis* dasselbe: Auch die zwei- und mehrsprachige Lexikographie kann heute zugleich strukturell-differenziert und aktuell sein.

Mehr: Habe ich den Wortschatz von einer Sprache zu einer anderen einmal in eine Richtung erfaßt – etwa Deutsch-Englisch – kann ich durch Umstellung eine Rohfassung für die andere Richtung – in diesem Fall: Englisch-Deutsch - gewinnen. Da jedes "Wort" einer Sprache eine sprachgebundene Eigenbedeutung hat, kann man ein Wörterbuch natürlich nicht umkehren:

gehen + "Zusammenfügungen" --> (*to go*) + "Zusammenfügungen"

Doch bildet eine solche durch Umstellung gewonnene Rohfassung eine äußerst große Erleichterung für die Erarbeitung des "Parallelbandes" und eine Basis für sehr aufschlußreiche Kontrollmöglichkeiten. Und außerdem sind nicht zuletzt die Punkte, in denen eine Umstellung zu "Ungereimtheiten" führt, im Sinn der oben angedeuteten grundsätzlichen Fragestellungen äußerst aufschlußreich.

Ordne ich die Einträge nach Sinngruppen, Wortfeldern u.ä., lassen sich mühelos ein-, zwei- und mehrsprachige Synonymwörterbücher erarbeiten – für alle Sprachen und nach allen "morphologischen" Kriterien, für die das Material in der eingangs skizzierten Form eingegeben wurde.

Diese Synonymwörterbücher lassen sich wiederum mit alphabetisch angelegten Wörterbüchern koppeln. – Eine linguistisch sinnvoll angelegte CD-Rom könnte schon heute etwa für dt. sterben alle Möglichkeiten der Übersetzung in jede der vom Computer erfaßten Sprachen liefern.

Bisher gingen wir bei unseren Überlegungen vom "Wort" und von den "Zusammenfügungen" eines Wortes aus. Koppelt man diese Perspektive nun mit Texten, erschließen sich weitere Möglichkeiten, die ebenfalls hier nur angedeutet seien:

- Wortwahl bestimmter Autoren ("Symbol", "Anschauung" bei Goethe)
Wortgebrauch in bestimmten Epochen (Wiederholungen, Modewörter, Klischees, Stereotype usw.)
- Aufschlüsselung des Wortschatzes nach Textsorten, Bereichen usw.
- Bedeutungsveränderungen je nach Text, Textsorte und weiterem Kontext (Situation, Hintergrund, Geschichte usw.)
- Text - Textgestaltung - Wort - Wortzusammensetzung - Bedeutung - Wirkung: Pragmatik, Rhetorik, Stilistik, Poetik ...

Mehr als jede andere Erfindung wird der Computer dazu beitragen, von der Wort- und Satzsemantik bzw. Wort- und Satz-Pragmatik zu einer Textsemantik und Textpragmatik zu kommen. Wenn man sich vor Augen hält, daß fast die ganze europäische Tradition der Sprachphilosophie und Sprachwissenschaft entscheidend vom Wort und dann vom Satz ausging, der Mensch aber in Texten spricht, ersieht man, daß damit nicht mehr und nicht weniger als einer "Revolution" unserer ganzen Sprachauffassung vorgearbeitet wird: Es besteht heute die Möglichkeit, dem Phänomen des Sprechens in einer grundsätzlich vertieften Form auf die Spur zu kommen, wobei natürlich die in einem engeren Sinn naturwissenschaftlichen Entdeckungen (Biologie, Medizin, Psychologie, Anthropologie usw.) einfließen (Entdeckungen, die zum größten Teil auch mithilfe des Computers erst in vollem Umfang möglich wurden). Eine sinnvolle Beschäftigung mit der Sprache wird in Zukunft Geistes- und Naturwissenschaften nicht mehr trennen können.

Auf der Basis eines zugleich detailliert und umfassend aufgearbeiteten Lexembestands sind also heute folgende "Wörterbücher bzw. "Wörterbuchtypen" zum Teil bereits erstellt, zum Teil möglich:

- ständig aktualisierte Wörterbücher jeden Umfangs für die unterschiedlichsten Bereiche, Zielsetzungen usw.
- strukturelle Wörterbücher mit spezifischer theoretischer oder/und praktischer Zielsetzung

- integrierte Wörterbücher (Integration verschiedener Wörterbuchtypen, Integration von Wörterbüchern und Texten; Integration von Wörterbüchern, Texten, Lehrmaterialien, Bildern u.a.m.)
- jeweils ein-, zwei- und mehrsprachig.

Die prinzipiell identische Erfassung und Verarbeitung des Materials macht schließlich eine Kooperation unterschiedlicher Instanzen in einer ganz anderen Weise möglich als bisher: Staat, Wirtschaft, Industrie – aus verschiedenen Ländern – können ihr Material, ihre Erfahrung, ihre Kenntnisse und ihre Zielsetzungen mühelos mit der Wissenschaft und aufeinander abstimmen. Auch hier stehen wir zurzeit am Beginn einer völlig neuen Form im Umgang mit der Sprache.

Mit dieser Entwicklung sind allerdings auch einige Schattenseiten und Gefahren verbunden, auf die es ebenfalls hinzuweisen gilt. Ich resumiere stichwortartig:

Anonymisierung des Materials

Die große Mehrheit der guten Wörterbücher war in der Vergangenheit das Ergebnis der Leistung eines Einzelnen oder einer (sehr) kleinen Gruppe. Noch der "Wahrig", der "Robert", das große spanische Wörterbuch von Maria Moliner waren Lebensleistungen Einzelner – und hatten nicht zuletzt deswegen bestimmte Qualitäten. – Heute wird die Leistung des einzelnen Lexikographen (und Lexikologen) durch "Übernahme" seines Materials, durch "Verfremdung" u.a. in einer Weise relativiert, die in vielen Fällen bis ins Kriminelle geht. Der Diebstahl von Wörterbuchmaterialien ist weithin zu einer Art Kavaliersdelikt geworden.

Staats- und Wirtschaftsinteressen

Je umfangreicher, aufwendiger und durchschlagender die technische Seite der Erstellung von Wörterbüchern (und anderen Materialien) ist – und je entpersonalisierter -, um so stärker die Tendenz der Wirtschaft, nur den Gewinn zu sehen, um so fühlbarer das Bestreben des Staates, auch an die Sprache eine technische Meßlatte anzulegen. Wie weit die Tendenz geht, Sprache als eine Art "Besitz des Staates" zu betrachten, zeigen die tristen Ereignisse um die sog. deutsche Rechtschreibreform mehr als deutlich.

„Methoden“ statt sinnvoller Zielsetzungen

Wie in so vielen Bereichen, zeigt sich auch auf dem Wörterbuchmarkt, daß die Fragen nach sinnvollen Zielsetzungen nicht selten verstellt werden nicht nur von Nutzeffekten der Wirtschaft und des Staates, sondern auch von Gesichtspunkten des "Machbaren". – Es gibt leider zu denken, daßbisher kein einziges Kultusministerium die Funktionen von Wörterbüchern und die neuen Möglichkeiten der Bildung durch eine sinnvolle Nutzung von Wörterbüchern neuen Typs an allen Schultypen entschlossen als Teil einer Bildungspolitik aufgegrif-

fen hat. In einer Zeit, in der der Computer zur Normalausrüstung der Schulen und der Schüler gehört, sollte es darauf ankommen zu fragen, was man damit und daraus im Sinne einer umfassenden Bildung macht. Es scheint aber ähnlich wie beim Auto zu werden: Das Gerät wird immer stärker zum (wirtschaftlichen) Selbstzweck.

Technik versus Benutzer

Es zeigt sich immer mehr, daß die übergroße Mehrheit der Benutzer das, was man mit einem (guten!) EDV-gestützten Wörterbuch "alles machen kann", gar nicht aufgreift, ja in der Regel gar nicht aufgreifen "will". Man hat nicht selten den Eindruck, daß sich auf dem Markt nur das durchsetzt, was erstens nicht schwer ist, und zweitens "Anreize bietet". – Der Lexikograph und Lexikologe kann vor einer solchen Entwicklung gar nicht genug warnen: Die Banalisierung des Markts würde ihn zunehmend in seine "Sklavenherrschaft" nehmen.

Es geht heute, in allen Bereichen, um Inhalte. Technische Möglichkeiten sind an ihnen zu messen – nicht umgekehrt. – Stellt man die heute möglichen Wörterbücher gegen die heute angebotenen und (noch mehr) gegen die heute verbreiteten Wörterbücher, ergibt sich eine tiefe Kluft.

Dieselbe Kluft ergibt sich, wenn man das gegeneinander hält, was durch die Integration von Wörterbüchern und Texten möglich ist, und das, was realiter in den Ausbildungsstätten gemacht wird. Ja, schon die Methoden der Sprachausbildung sind – oder wären – auf die neuen Möglichkeiten durch Wörterbücher und andere Lehrmaterialien neuen Typs präzise abzustimmen. Wir kommen mit diesen allgemeinen Erwägungen zu den Problemen der Lexikographie und Lexikologie, die der Computer prinzipiell nicht lösen kann.

Die Linguistik unterscheidet beim Wort zwischen den sog. "gebundenen" und "freien" Formen. Eine gebundene Form ist ein Element einer (höheren) Einheit aus mehreren Elementen, an die sie in ihrer Funktion "gebunden" ist. Bei kaputt gehen hat gehen eine spezifische Funktion oder Bedeutung, die ihm nur in dieser Verbindung – oder, wenn noch in anderen, dann auch in ganz spezifischen, d.h. lexematisch gegebenen Verbindungen – zukommt. Gehen als freie Form dagegen erlaubt Sätze wie: Der Junge geht über die Straße / Die alte Dame geht langsam die Allee hinunter / der Hund geht über den Marktplatz / ...: Alles, was Beine hat, geht (oder läuft). Die freie Form wurzelt also in einem Begriff: Der Junge geht anders als die alte Dame, und beide gehen wieder anders als ein Hund; doch diese unterschiedlichen Modalitäten des "Gehens" werden im Begriff aufgehoben, das Ähnliche wird als gleich behandelt, das Unterschiedliche wird unterschlagen. – Der Mensch "verstehet" die begriffliche Sphäre – und kann eben deswegen, syntaktisch-frei "treffende", "richtige" "Verbindungen" mit gehen bilden. Die begriffliche Verbindung stellt also die wesentliche Grundlage der Syntax dar – und damit eine der wesentlichen Grundlagen der menschlichen Sprache. Die vieldiskutierte Möglichkeit, mit

einer endlichen Zahl von Elementen ("Worten") unendlich viele Sätze zu bilden, gründet wesentlich hierin (wenn auch nicht nur).

Genau dieses Verständnis hat nun der Computer nicht; er kann die freien Verbindungen daher nicht herstellen. Man muß ihn deswegen auf jede einzelne syntaktische Verbindung programmieren – d.h. jedes einzelne Element der Sprache so auszeichnen, daß klar wird, mit welchen anderen Elementen es kombinierbar ist.

Zweites Beispiel: die Situation und der Hintergrundkontext. – Gespräch zwischen A und B. – 1. Hintergrund: B ist eilig. Er sagt: Ich gehe. A versteht: "Er geht aus Zeitmangel / weil er eilig ist". – 2. Hintergrund: Es gibt Streit. Dieselbe Äußerung von B. A versteht: "B ist verschnupft". – 3. Hintergrund: A und B bilden ein Liebespaar, das sich zerstritten hat. Dieselbe Äußerung von A. B versteht: "A löst die Beziehung". – Kurz: In jeweils unterschiedlichen Kontexten verstehen wir dieselbe Äußerung jeweils anders, neu. – Der Computer faßt den Kontext als notwendigen Rahmen des Verstehens nicht mit auf, er "verstehet" nur, was man ihm "eingibt".

Drittes Beispiel: Eine unwillige Äußerung von A zu B über C: "Dieser Mann hat den Kopf auch nur, um einen Hut darauf zu stülpen". B versteht: "A hält C für einen Dummkopf" – aufgrund der Präsupposition: "der Kopf ist "ein Instrument zum Denken" und nicht – oder jedenfalls weniger (wesentlich) – "eine Kugel für Hüte". – Wir verstehen aufgrund von "Bildern", die wir von unseren Mitmenschen, von unserer Umwelt, von uns selbst – und nicht zuletzt auch von unseren Organen haben. Der Computer "verstehet" hier gar nichts. Bei jeder Äußerung, die von solchen Präsuppositionen – bildhaften und anderen – bestimmt ist, ist er "blind".

Viertes Beispiel: Nehmen wir an, wir kennen den Ausdruck einer Sache auf den Grund gehen nicht und hören den Satz: "Wenn du der Frage auch nur ein wenig auf den Grund gegangen wärest, hättest du dir all die Unannehmlichkeiten sparen können". Das Bild "auf den Grund gehen" gibt uns eine (ungefähre) Anleitung, in welcher Richtung die Bedeutung (im gegebenen Kontext) zu suchen ist.

Der Mensch "verstehet" – in Begriffen und Bildern, innerhalb spezifischer Situationen und vor einem jeweils spezifischen Hintergrund. Nicht zuletzt dies macht unsere Sprache zu einer menschlichen Sprache – im Gegensatz zu den "Sprachen" der Tiere, die – zumindest, soweit wir das bisher "verstehen" – in ihrem "Verständnis" in einen festen Rahmen "eingeschlossen" sind, die also jeweils Unterschiedliches, Neues nicht auf Identisches beziehen können.

Der Computer "verstehet" das, was man ihm beibringt; Neues, von dem Bisherigen Abweichendes auf das bisher Gegebene zu beziehen, mag ihm nach statistischen o.ä. Gesetzmäßigkeiten gelingen – aufgrund von "Verstehen" gelingt es ihm nicht.

Aufgrund dieser Zusammenhänge sagten wir eingangs: Die neuen Möglichkeiten, Worte in ihren Zusammenfügungen und in Texten präzise zu übersehen,

ist eine völlig neue Chance für die Linguistik – doch nur, wenn man Schritt für Schritt herausarbeitet, was der Computer kann und was er nicht kann. Unsere Bemerkungen hierzu sind natürlich nur skizzenhaft. Sie führen indessen von einer anderen Seite aus zurück auf den oben berührten Grundgedanken: Geistes- und Naturwissenschaften sind heute – in der Beschäftigung mit der Sprache wie in der Beschäftigung mit anderen grundlegenden Problemen – aufgerufen, wie vielleicht noch nie in der Geschichte, gemeinsam zu arbeiten. Gelingt das nicht, wird eine "naturwissenschaftlich-technisierte" Welt die "geistige" Dimension unterhöheln. Die Lexikologie und Lexikographie wird durch den Computer somit in nie geahnter Form gefördert; nicht weniger aber wird sie durch ihn gefordert.