

*Martina Schwanke*

## **Die Lemmatisierung von Goethes “Wahlverwandtschaften”**

### **Abstract**

In German studies computer-oriented methods are taking on an increasing significance and consequently should be considered in detail together with all the problems connected with them. This article deals with the automatic lemmatization of Goethe's novel “The Elective Affinities”. Following the results of the computer-system LEMMA2, central problems of lemmatization as far as manual correction is concerned will also be discussed.

### **1. Allgemeines**

Trotz seiner unbestreitbar großen Bedeutung für den Alltag unseres Lebens — auch an der Universität — spielt der Computer in der germanistischen Forschung noch eine erstaunlich geringe Rolle. In den vergangenen Jahren habe ich daher versucht, Studierende mit modernen computergestützten Arbeitstechniken vertraut zu machen, die in der Linguistik oder Literaturwissenschaft anwendbar sind und Fragen zu bearbeiten erlauben, deren Beantwortung vormals wegen des extrem hohen Zeitaufwands nicht realisierbar war. Zugleich habe ich selbst solche Methoden in meiner Forschungsarbeit eingesetzt und dabei vielfältige Erfahrungen gemacht (Schwanke 1991, Schwanke 1992). Einen Schwerpunkt bildete die Beschäftigung mit epischen Werken Goethes, wobei Deutung und Funktionsbestimmung der Namen im Vordergrund standen. Eine solche Funktionsbestimmung erfordert u.a. auch statistische Analysen des Gesamtwortschatzes der untersuchten Werke. Um diese zu leisten, reichen Wortformenindices, die mit vergleichsweise geringem Zeitaufwand erstellt werden können, nicht aus. Ausgangsbasis solcher Untersuchungen — wenn sie statistische Relevanz besitzen sollen — muß der Lemmabestand des untersuchten Textes sein. Daher habe ich mich darum bemüht, eine solche Untersuchungsbasis in Form von lemmatisierten Indices für einige ausgesuchte epische Werke Goethes exemplarisch zu schaffen (“Die Leiden des jungen Werther”, “Die Wahlverwandtschaften”, “Hermann und Dorothea”); darüber hinaus habe ich zu den “Meister”-Romanen Wortformenindices fertiggestellt, die mangels

Arbeitszeit leider nicht lemmatisiert werden konnten. Auch wenn inzwischen verschiedene Softwareprodukte zur Verfügung stehen, die entweder eigens zum Zweck der Lemmatisierung entwickelt wurden oder eine Lemmatisierungskomponente im Zusammenhang umfangreicher Textanalysen enthalten, so ist eine Lemmatisierung dennoch bisher nicht ohne eine Bearbeitung des Textes und zumeist auch nicht ohne mehr oder weniger intensive Eingriffe in den Programmablauf möglich. Ein nach meinen Erfahrungen erstaunlich korrekt funktionierendes Lemmatisierungsmodul enthält zum Beispiel der Syntaxanalyseteil des Saarbrücker Übersetzungssystems SUSY. Lemmatisiert wird im Grunde genommen auch in dem sehr exakt arbeitenden Syntaxanalysesystem PLAIN von G. Hellwig (Heidelberg), wenn der Benutzer die hierzu erforderlichen sehr umfassenden Arbeiten zum Aufbau des integrierten Lexikonsystems vorzunehmen bereit ist. Bei solchen integrierten Lemmatisierungsmodulen ist der Umfang der erforderlichen Vorarbeit des Benutzers - wenn auch in unterschiedlicher Weise - noch außerordentlich hoch. Nur zum Zweck der Lemmatisierung entwickelt und von mir zu diesem Zweck auch eingesetzt wurde das System LEMMA2 in einer Version für IBM-kompatible PC. Es wird vom Bonner Institut für angewandte Informations- und Sprachforschung e.V. vertrieben.

## **2. Die automatische Lemmatisierung mit LEMMA2**

Theoretisch ist das Programmsystem LEMMA2 mit Hilfe der mitgelieferten Batch-Dateien vollautomatisch ablauffähig. In der Praxis jedoch sind immer wieder Eingriffe erforderlich, und natürlich ist auch eine bestimmte Textform für den Ablauf des Programms Voraussetzung. Ich gehe in der nun folgenden Beschreibung des Systems davon aus, daß der zu bearbeitende Text in einem selbstgewählten Format mit einem Text-editor gespeichert wurde. Es mag dahingestellt sein, ob dieses Format beispielsweise den Richtlinien der Text Encoding Initiative (TEI) mit dem dort sehr differenzierten Beschreibungsinventar entspricht (Sperberg-McQueen/Burnard 1990) oder ob beispielsweise die Weiterbearbeitung mit dem Layoutsystem TeX/LaTeX vorbereitet wurde oder ob andere einfachere Markierungen vorgesehen wurden. In jedem Fall wird die Textdatei aus einem reinen Textteil bestehen und einem weiteren Teil, der zusätzliche Informationen enthält, zum Beispiel die Angabe der Zeilen- und Seitennummern des jeweils aktuellen Textstücks. LEMMA2 erwartet einen Text, der im Extended-ASCII-Modus vorliegt und keinerlei Markierungszeichen enthält, auch keine Zeilennummern. Der vom

Benutzer gespeicherte Text muß also zunächst einmal von allen - eventuell ihm selbst nicht sichtbaren - Zeichen des verwendeten Texteditors befreit werden, und die Zeilen- und Seitenmarkierungen sind zu beseitigen.

Wenn man LEMMA2 einen so vorbereiteten Text zur Verarbeitung anbietet, wird man bei Verwendung eines Goethe-Textes schon bald mit einem Systemabbruch rechnen müssen, weil Goethe häufiger wesentlich komplexere Sätze verwendet, als sie LEMMA2 zuläßt. Wenn ein Satz mehr als einhundert Wörter umfaßt, streikt das System. Man wird daher an einer syntaktisch passenden Stelle künstlich ein Satzende setzen.

Auch der Satzanfang kann ein Problem darstellen. Standardmäßig geht das System bei seiner Analyse davon aus, daß Großschreibung ein Erkennungszeichen für Substantive darstellt. Die im Deutschen übliche Großschreibung am Satzanfang führt daher oft zu Fehlinterpretationen, sofern es sich nicht um ein Substantiv handelt. Zur Abhilfe sehe ich zwei Lösungen, die auch beide von mir erprobt wurden: Entweder werden die Analysefehler am Satzanfang gemeinsam mit sonstigen Analysefehlern in einem späteren manuellen Korrekturvorgang eliminiert, oder aber der Text muß vor dem Programmeinsatz entsprechend präpariert werden. Ein weiteres, nicht zu unterschätzendes Problem stellt die Größe der zu bearbeitenden Textdatei dar. LEMMA2 produziert aus dem vorbereiteten ASCII-Text eine neue Vertikaldatei, in der für jedes Wort und jedes Satzzeichen eine Dateizeile im Umfang von achtzig Zeichen vorgesehen ist. Das kann bei einem Goethe-Werk schon leicht zu Platzschwierigkeiten auf einer Festplatte führen, wenn man berücksichtigt, daß diese Vertikaldatei vor der eigentlichen Analyse durch LEMMA2 in eine alphabetisch geordnete Datei verwandelt werden muß, so daß während des Sortiervorgangs ein Vielfaches des originalen Dateiumfangs die Platte belastet und anschließend vorübergehend zwei inhaltlich identische und nur formal unterschiedliche Dateien auf der Festplatte Platz beanspruchen.

Ein zusätzliches Problem in diesem Zusammenhang bereitet der Sortieralgorithmus selbst. Standardmäßig kann mit dem MS-DOS-Betriebssystem-Sortieralgorithmus gearbeitet werden. Das ist bei einem umfangreichen Goethe-Werk jedoch nicht möglich. Die von LEMMA2 erzeugte Vertikaldatei zu einem Goethe-Epos ist nämlich so umfangreich, daß die Kapazität des integrierten Sortiersystems überfordert wird. Mit LEMMA2 kann ein vom Produzenten entwickeltes hauseigenes Sortiersystem (IKS-SORT) erworben werden, das recht leistungsfähig ist

und erheblich mehr Möglichkeiten bietet als die betriebssystemeigene Sortieroutine. Einem Goethe-Roman ist jedoch auch dieses System nicht gewachsen. In Kiel wurde daher ein neuer Sortieralgorithmus entwickelt, der diesen Kapazitätsansprüchen genügt und der inzwischen in verschiedenen Versionen vorliegt. Zum Ablauf des Programmsystems LEMMA2 gehört als letztes Modul ein rudimentäres, aber zu diesem Zweck effizientes Syntaxanalyse-System, durch das abschließend so manche noch ungelöste Analyseentscheidung automatisch getroffen werden kann.

Beim Einsatz von LEMMA2 darf man sich keiner Illusion hingeben. Bei der Bearbeitung eines modernen Textes mag eine hohe Trefferquote erreichbar sein, bei einem vergleichsweise älteren Text, zum Beispiel einem Goethe-Roman, liegt die Trefferquote deutlich unter achtzig Prozent. Dieses Resultat ließe sich theoretisch verbessern, wenn man das vom System verwendete Wortformenlexikon textspezifisch erweitern würde. Das allerdings ginge auf Kosten der Arbeitsgeschwindigkeit, ganz abgesehen von der Zeit, die für die Erweiterung des Lexikons aufgebracht werden muß. Der Analysealgorithmus des Systems basiert letzten Endes auf Häufigkeitsberechnungen zur grammatischen Funktion der Endungen deutscher Wörter. Die Wortendungen sind in der deutschen Grammatik in unterschiedlicher Intensität multifunktional; besonders viele Funktionen fallen den Endungen *-e* und *-en* zu. Das Lemmatisierungsergebnis zeigt sehr schnell, daß bei diesen Endungen die meisten Interpretationsfehler des Systems auftreten. Gegen diese Fehleranfälligkeit gibt es kein Mittel außer der Erweiterung des systemeigenen Wortformenlexikons mit dem bereits erwähnten Nachteil der Verlangsamung der Systemarbeitsgeschwindigkeit. Am Ende der automatischen Textbearbeitung durch LEMMA2 steht in jedem Falle eine lange und intensive Korrektur des automatisch erzeugten Ergebnisses.

### 3. Lexikographische Probleme

Die Grundlage des hier vorgestellten Index zu Goethes Altersroman "Die Wahlverwandtschaften" bildet die von E. Trunz herausgegebene Hamburger Ausgabe in der 11. Auflage von 1982. Entsprechend den Verfahrensansätzen von H. E. Wiegand (1983) und E. Link (1989) kann man sich bei eigener Lemmatisierungsarbeit an veröffentlichten lemmatisierten Indices, Konkordanzen oder Wörterbüchern orientieren. Dabei zeigt sich allerdings, daß nicht einheitlich vorgegangen und überdies eine

Begründung des gewählten Verfahrens nicht gegeben wird. Dennoch läßt sich in solchen Fällen der abweichenden Lemmatisierung nicht ohne weiteres sagen, der eine Lemmaansatz sei richtig, der andere falsch, denn:

*Eine (aus sprachtheoretischen Gründen) objektiv, richtige' Lemma-repräsentation kann es nicht geben. (Link 1989, 188)*

Unproblematisch sind Textbelege, die sich gemäß der Lemmatisierungskonvention auf eine Grundform zurückführen lassen, also bei Substantiven auf die Form des Nominativ Singular, bei Verben auf den Infinitiv. Diese Fälle werden bei Einsatz eines entsprechenden Rechnerprogramms bereits weitgehend automatisch gelöst, und anfallende Korrekturen lassen sich vergleichsweise einfach vornehmen. Hier bereitet allenfalls bei größeren Texten die Masse an Einträgen einen hohen Arbeitsaufwand. Eine inkorrekte Zuordnung eines Belegs zu einem Lemma führt dazu, daß er für den Benutzer nicht mehr auffindbar ist. K. Gärtner und P. Kühn haben daher die Forderung aufgestellt,

*den einzelnen Wortformen manuell Querverweise beizugeben, die das Auffinden aller Wortformen eines Lemmas garantieren sollen ..., oder durch rückläufige Indices und Verbfindelisten den Benutzer zu den verschiedenen Wortformen eines Lemmas zu leiten ... (Gärtner/Kühn 1984, 627)*

Dieses theoretisch begrüßenswerte Verfahren bringt jedoch den Nachteil mit sich, den Umfang des Index erheblich zu vergrößern, so daß sich bisweilen die Frage stellt, ob eine Veröffentlichung im Papierformat überhaupt noch sinnvoll ist, zumal viele Bibliotheken angesichts des horrenden Anstiegs der Bücherpreise stark selektiv einkaufen müssen. Obwohl ich zu Demonstrationszwecken ehemals die verschiedensten Sortierverfahren angewandt hatte, habe ich mich dazu entschlossen, lediglich den reinen Index zu publizieren. Wer weitergehende mit diesem Index nicht durchführbare Untersuchungen beabsichtigt, kann jederzeit mit Hilfe einer maschinenlesbaren Fassung der "Wahlverwandtschaften", veröffentlicht in der Reihe "Elektronische Bibliothek zur deutschen Literatur", und den vom Institut für angewandte Sprach- und Informationsforschung e.V. vertriebenen Programmen die jeweils erforderlichen weiteren Analysen computergestützt durchführen.

Technisch gesehen kann nach P. Sappler die Bearbeitung eines Index in dreierlei Weise geschehen (Sappler 1990, 1571): Erstens durch die Nachredaktion des automatisch generierten Rohindex, zweitens durch vorherige Überarbeitung des Ausgangstextes im Hinblick auf die

späteren maschinellen Sortierungen und drittens durch Arbeitsverfahren, “die in den Herstellungsablauf integriert sind” (ebd.), die dementsprechende Flexibilität des Programmsystems vorausgesetzt. Die Korrektur der automatisch generierten Lemmatisierungsergebnisse zu den “Wahl-verwandtschaften” erfolgte durch eine Kombination von zwei Arbeits-techniken. LEMMA2 ermöglicht es dem Lexikographen, die Lemmatisierung durch die Ausgabe einer Datei, in der jedes Element in seinem Textkontext erscheint, zu überprüfen. Darüber hinaus wurde eine letzte Nachredaktion anhand des auf diese Weise erstellten alphabetisierten Index vorgenommen. Nachfolgend möchte ich einige der hierbei auftretenden Problemtypen vorstellen. Es kann an dieser Stelle nicht meine Aufgabe sein, in die umfassende Diskussion über die Definition eines Lemmas einzugreifen (Link 1989, Wiegand 1983), sondern es wird lediglich darum gehen, einige kontroverse Lösungsansätze zu diskutieren.

Unterschiedliche Entscheidungen sind grundsätzlich möglich bei der Frage, wie rigoros lemmatisiert werden sollte. So heißt es bei St. P. Sondrup und D. Chisholm:

*Die Konkordanz ist vollständig lemmatisiert. Dementsprechend findet man die flektierten Formen unter ihrer unflektierten Grundform, und zwar in derselben Reihenfolge, wie sie in Faust, Erster Teil auftreten. Alle Formen des Verbs erscheinen unter dem Infinitiv, die des Substantivs unter dem Nominativ Singular und die des Adjektivs unter der unflektierten Form des Positivs. (Ausnahme: unregelmäßige Komparativ- und Superlativformen, z.B. besser und best- erscheinen als selbständige Stichwörter). (Sondrup/Chisholm 1986, X)*

Bei dem hier zitierten Lemmatisierungsprinzip handelt es sich also lediglich um die Rückführung der unmittelbaren Flexionsendungen auf die entsprechende nächstliegende Stufe oder auf Lemmata mit Platzhalter- oder Kommentarsymbolen, wie zum Beispiel in der Konkordanz zu “Faust I” (Sondrup/Chisholm 1986):

letzt: *letzte* (Adj.), *letzten* (Adj.), *letzt* (Verb)

Licht: *Lich*, *Lichte*, *Lichter*

lieb (Adj.): *liebe*, *liebsten*, *lieben*, *liebstes*, *lieb*, *lieber*, *liebs*,

Lieb-: *Lieben*

Liebchen: *Liebchen*, *Liebchens*

Liebe: *Lieb'*, *Liebe*, *Lieb's* (substantiviertes Adjektiv mit elidiertem e)

lieben: *lieb'*, *liebt*, *liebender*, *liebt'*, *lieben*, *liebte*, *liebende*, *liebenden*

Liebend-: *Liebende*, *Liebender*

lieber (Adv.): *lieber*

Liebst-: *Liebsten, Liebste*

Hinzu kommen weitere Derivationen und Kompositabildungen wie *liebenswert, Liebeslust, Liebesträum* u.a.

Konträr dazu bietet sich das Verfahren an, einem einzigen Hauptlemma verschiedene Sublemmata unterzuordnen, insbesondere alle rechtserweiternden Wortbildungen, also:

lieb: *Liebhaber, Liebhaberin, Liebhaberei, liebäugeln, liebkosen, lieblich, lieblich, Liebling, Lieblingin, Lieb, Liebchen, Liebe, Liebesapfel, Liebesbrief, Liebeserklärung, Liebesgott, Liebeslied, Liebestrank, Liebelei, Liebde, lieben* (Paul 1992, 531-532).

Sobald ein ursprünglich anderes Etymen zugrunde liegt, wird ein neues Hauptlemma angesetzt, z.B. *Liebstöckel* (ebd., 533).

Aufgrund der Vorgaben des eingesetzten Programmsystems ist der Lexikograph bei der automatischen Lemmatisierung zunächst einmal angewiesen auf die vom System erzeugten Lemmata. Eine eigenständige Erarbeitung von Lemmatisierungsprinzipien bringt entweder einen enorm hohen Aufwand an Korrekturarbeit mit sich oder erfordert die Programmierung eines eigenen automatischen Lemmatisierungssystems. Aufgrund der von LEMMA2 generierten Ergebnisse verbot sich von vornherein eine Entscheidung für nur ein Hauptlemma. Gleichwohl ergab sich aus der alleinigen Tilgung von Flexionsendungen eine solche Vielzahl an Lemmata, daß ich mich dazu entschloß, für diese aus meiner Sicht nicht akzeptablen Fälle eigene Lösungen zu erarbeiten. Damit ist auch das von den Bearbeitern der beiden "Faust"-Konkordanzen angewandte Verfahren verworfen; stattdessen habe ich mich für eine Lemmatisierung entschieden, wie sie etwa von G. Wahrig, Ch. Dill und den Lexikographen des Goethe-Wörterbuchs praktiziert wird. In bezug auf die Form *lieb* ergeben sich somit folgende Lemmata:

Liebe: *Liebe*

lieben: *geliebt, geliebte, geliebten, geliebtes, lieben, Liebende, liebenden, Liebenden, liebender, Liebender, Liebens, liebst, liebt, liebte, liebten, liebtest*

lieb: *lieb, liebe, Liebe, lieben, lieber, Lieber, Lieberes, liebes, liebsten, Liebsten, liebster, Liebster, liebstes*

Hinzu kommen Derivationen wie *lieblich, liebevoll* oder *Liebling* und Kompositabildungen wie *Liebenswertigkeit, Liebhaberei*. Somit werden substantivierte Partizipien bis auf wenige Ausnahmen ( etwa *Geliebte(r)*) nicht als selbständige Lemmata aufgeführt. In der Forschungsliteratur

wird häufig darauf hingewiesen, daß es entscheidend ist, Lemmatisierungsprinzipien konsequent anzuwenden. Allerdings sollte die letzte Entscheidung über die Anwendung einer Regel im Einzelfall dem kompetenten Lexikographen überlassen bleiben. Dies gilt beispielsweise für das durchaus anfechtbare Ansetzen eines eigenen Lemmas für substantivierte Infinitive (*Bedauern, Bedenken, Behagen, Belieben, Bemühen*). Vom Verbstamm abgeleitete Substantive, die eine selbständige Bedeutung erlangt haben, sind als eigenes Lemma aufgenommen, etwa *Beschließerin* (,Verwalterin’).

Außerdem treten in geringer Frequenz Belegformen auf, die durch die Verwendung spezifischer Wortbildungselemente zur Kompositabildung auffallen. Um den besonderen Charakter dieser Bildungen etwa als substantivierte Infinitive oder Partizipien zu bewahren, wurden hier lediglich die unmittelbaren Flexionsendungen getilgt, so daß sich Lemmata ergeben wie *gegenscheltend, doppeltverwandelt, Wiederniedersitzen*. Über-dies wurden Syntagmen wie *Besser- oder Anderswissen, Um- und Wiedermeinen* als ein Lemma akzeptiert. Dadurch konnte auf Konstruktemmata verzichtet werden.

Häufig treten verschiedene Schreibvarianten eines Lexems auf oder heute ungebräuchliche Graphien, etwa *ungefähr/ohngefähr, ehlich/ehe-lich* oder *wechselsweise*. In diesen Fällen wurde für das Lemma die Orthographie des Dudens gewählt. Diese Entscheidung hat meines Erachtens den Vorteil, daß auch Benutzer, denen die ältere Schreibweise nicht geläufig ist, die gesuchten Belegstellen finden können.

Komparative und Superlative wurden in der Regel jeweils ihrer Positivform zugeordnet; Diminutivformen wurden nicht als eigenständige Lemmata verzeichnet, es sei denn, daß sie durch eine spezifische Semantik gekennzeichnet sind. So wurden die Belegformen *Blättchen* und *Büchlein* zu den Lemmata *Blatt* bzw. *Buch*, während die Textform *Fräulein* als Lemma erhalten blieb.

Überaus kontroverse Lösungen finden sich beim Ansetzen der genusbedingten Endungen von Pronomina und einigen Adjektiven, letztere besonders substantiviert. St. P. Sondrup und D. Chisholm (1986) haben sich für eine neutrale Position entschieden und verwenden zumeist Bindestrichformen, zum Beispiel *erst* (Adv.), *erst-* (Ordnungszahlwort, für *erste, ersten, erster*), *Erst-* (für *Erst’, Erste*), *Allerschlimmst-* (für *Allerschlimmste*), oder *Klug-* (für *Kluge*). In der Verskonkordanz zu “Faust II” hingegen wurden von St. P. Sondrup und R. L. Jones (1989) konsequent die Bindestrichformen vermieden, dementsprechend haben

sie Lemmata gewählt wie *erst* (Adv./Ordnungszahlwort), *jeder*, *dies*, *Kluge*, *leicht*, *Leichte*, *link*, *Linke*, *links*. Eine weitere Notationsmöglichkeit besteht darin, die drei Genusformen nebeneinander zu setzen, eine Lösung, die sich etwa bei G. Wahrig (1982) findet. Zur Veranschaulichung der Problematik mögen einige ausgewählte Beispiele in tabellarischer Übersicht genügen:

Wie die Übersicht zeigt, gilt beim Lemmaansetzen für bestimmte Pronomina und Adjektive die Verwendung einer flexionsfreien Form. Diese findet sich jedoch bei den Pronomina recht selten, und auch vielen Adjektiven fehlt eine solche endungslose Grundform, in der sie prädikativ gebraucht werden. Bei der Verwendung flektierter Formen divergieren die gewählten Endungen, ebenso bei aus Adjektiven abgeleiteten Substantiven (etwa *Geistliche/Geistlicher*). Bei G. Wahrig (1991) hingegen finden sich neben Mehrfachlemmata (etwa *äußere(r,-s)*, *andere(r,-s)*) auch ausschließlich maskuline Flexionsformen (etwa *jeder*), ein Vorgehen, das sich schwer nachvollziehen läßt. In der älteren Taschenbuchausgabe dieses Wörterbuchs ist demgegenüber das Mehrfachlemma *jede(r, -s)* verzeichnet. Bedingt durch das häufige Fehlen flexionsloser Formen habe ich mich in den hier angeführten Fällen für Mehrfachlemmata entschieden (etwa *andere(r,-s)*, *welche(r,-s)*). Bei Substantiven sollte unterschieden werden zwischen Formen, die nur als Maskulinum auftreten (etwa *Geistlicher*) und solchen, die mehrere Genera zulassen (etwa *Nächste(r)*, *Neuankommende(r)*, *Neueintretende(r)*).

Zusammenfassend läßt sich festhalten: Lemmatisierte Indices können heute mit Hilfe computergestützter Arbeitsverfahren in angemessener Zeit hergestellt werden. Die gewählten Lemmatisierungsprinzipien sollten bei der vielfach notwendigen manuellen Nachredaktion möglichst einheitlich angewandt werden, wobei dem Lexikographen allerdings noch ein Spielraum für bisweilen notwendige Einzelentscheidungen zugebilligt werden sollte.

## Literatur

- Dill, Ch. (1987): *Wörterbuch zu Goethes West-östlichem Divan*. Tübingen.
- Drosdowski, G. (Hrsg.) (1989): *DUDEN. Deutsches Universalwörterbuch. 2, völlig neu bearbeitete und stark erweiterte Auflage*. Wien, Zürich: Mannheim.
- Gärtner, K. / Kühn, P. (1984): Indices und Konkordanzen zu historischen Texten des Deutschen: Bestandsaufnahme, Typen, Herstellungsprobleme, Benutzungsmöglichkeiten. In: W. Besch, O. Reichmann. *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Handbücher zur Sprach- und Kommunikationswissenschaft, 2, Band 1. Sprachgeschichte. Berlin, New York, 620-641.
- Goethe, J. W. von: *Werke*. Hamburger Ausgabe in 14 Bänden, herausgegeben von E. Trunz. München 1981-1986.
- Goethe-Wörterbuch, herausgegeben von der (vormaligen) Akademie der Wissenschaften der DDR, der Akademie der Wissenschaften in Göttingen und der Heidelberger Akademie der Wissenschaften, Bd. 1-3. Stuttgart, Berlin, Köln, Mainz 1978-1992.

- Kempcke, G. (Hrsg.) (1984): *Handwörterbuch der deutschen Gegenwartssprache*. In zwei Bänden. Berlin.
- Link, E. (1989): Was ist eigentlich ein Lemma? oder: Gehört z.B. das “-” bzw. “...” in “meta-, Meta-” bzw. “meta..., Meta...,” zum Lemma oder nicht? Anmerkungen zu einem Beitrag H.E. Wiegands zur Theorie der lexikographischen Sprachbeschreibung. In: H. E. Wiegand *Wörterbücher in der Diskussion. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*. Lexicographica: Series Maior, 27. Tübingen, 155-190.
- Merker, E. (Hrsg.) (1966): *Wörterbuch zu Goethes Werther*. Berlin.
- Paul, H. (1992): *Deutsches Wörterbuch*. 9., vollständig neu bearbeitete Auflage. Hrsg. von Helmut Henne und Georg Objartel unter Mitarbeit von Heidrun Kämper-Jensen. Tübingen.
- Sappler, P. (1990): Der Index/Das Belegstellenwörterbuch. In: F. J. Hausmann, O. Reichmann u.a. *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Zweiter Halbband, Handbücher zur Sprach- und Kommunikationswissenschaft, 5.2. Berlin, New York, 1567-1573.
- Schwanke, M. (1991): *Maschinelle Übersetzung. Ein Überblick über Theorie und Praxis*. Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest.
- : *Name und Namengebung bei Goethe. Computergestützte Studien zu epischen Werken*. Beiträge zur Namenforschung, Beih., 38. Heidelberg 1992.
- : *Index zu Goethes “Wilhelm Meisters Lehrjahre”*. Stuttgarter Arbeiten zur Germanistik, 270. Stuttgart 1993 (erscheint).
- : *Index zu Goethes “Wilhelm Meisters Wanderjahre”*. Stuttgarter Arbeiten zur Germanistik, 271. Stuttgart 1993 (erscheint).
- : *Lemmatisierter Index zu Goethes “Die Leiden des jungen Werther”*. Stuttgarter Arbeiten zur Germanistik, 272. Stuttgart 1993 (erscheint).
- : *Lemmatisierter Index zu Goethes “Die Wahlverwandtschaften”*. Stuttgarter Arbeiten zur Germanistik, 273. Stuttgart 1993 (erscheint).
- : *Lemmatisierter Index zu Goethes “Hermann und Dorothea”*. Stuttgarter Arbeiten zur Germanistik, 27. Stuttgart 1993 (erscheint).
- Sondrup, St. P. / Chisholm, D. (1986): *Verskonkordanz zu Goethes “Faust, erster Teil”*. Indices zur deutschen Literatur, 18. Tübingen .
- Sondrup, St. P. / Jones, R. L. (1989): *Verskonkordanz zu Goethes “Faust, zweiter Teil”*. Indices zur deutschen Literatur, 21. Tübingen.
- Sperberg-McQueen, C. M. / Burnard, L. (Hrsg.) (1990): *Guidelines For the Encoding and Interchange of Machine-Readable Texts*. Document Number: TEI P1, Draft Version 1.1. Chicago, Oxford: ACH, ACL, ALLC.
- Wahrig, G. (Hrsg.) (1982): *dtv-Wörterbuch der deutschen Sprache*, 5. Aufl.. München.
- : *Deutsches Wörterbuch*. Mit einem “Lexikon der deutschen Sprachlehre”. Jubiläumsausgab. Gütersloh, München 1991.
- Wiegand, H. E. ( 1983): Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung. In: ders. (Hrsg.), *Studien zur*

*neuhochdeutschen Lexikographie III*, Germanistische Linguistik 1-4/82. Hildesheim, Zürich, New York, 401-474.

