

# LILA: LINGUISTISCH INTELLIGENTE LEHRWERKSANALYSE

Johann Haller

## 1 Einleitung: Elektronische Sprachverarbeitung in Saarbrücken

Die automatische Analyse von Texten in natürlicher Sprache hat an der Universität des Saarlandes eine lange Tradition. Bereits 1967 bestand eine Arbeitsgruppe Elektronische Sprachforschung, als deren Gründer und Mentor besonders H. Eggers bekannt wurde. Er war auch der erste Sprecher des SFB 100 'Elektronische Sprachforschung', an dem sich eine Reihe von philologischen Fächern und die Informatiker beteiligten. Die Arbeit dieser Gruppe begann mit dem Programmpaket SATAN (Saarbrücker Textanalyse, für Deutsch), dem dazu notwendigen SADAW (Deutsches Analyse-Wörterbuch) und dem ersten in Deutschland entwickelten Maschinellen Übersetzungssystem SUSY, das Material für eine Reihe von Sprachen (Deutsch, Englisch, Französisch, Russisch, Esperanto) erstellte. Das SADAW enthielt morphologische und syntaktische Informationen zu Wortstämmen, während für die Ermittlung übersetzungsrelevanter Lesarten eines Wortes ein eigenes Disambiguierungswörterbuch mit semantischen Angaben notwendig war (SEDAM). Der Sonderforschungsbereich wurde nach 1986 nicht mehr weitergeführt, aber eine Reihe seiner Mitglieder hatte sich sehr aktiv an der Vorbereitung und der Forschungsphase des europäischen Forschungsprojekts EUROTRA beteiligt und setzte die Arbeit dann in diesem Rahmen fort.

Im Jahre 1984/85 wurde das Institut für Angewandte Informationsforschung an der Universität des Saarlandes (IAI) gegründet, das den deutschen Anteil an EUROTRA als erstes Großprojekt übernahm. Am IAI werden auf der Basis der im SFB 100 und in EUROTRA entwickelten Grundlagen anwendungsbezogene Projekte durchgeführt, die zu einsatzfähigen Prototypen führen. Dazu gehören Übersetzungshilfen, Korrektur- und Prüfprogramme für Rechtschreibung, Grammatik und Terminologie, mehrsprachige Informationssysteme, z.B. für die Mediengesetze in Europa oder die Bewertung kommerzieller maschineller Übersetzungssysteme.

Theoretische Vorüberlegungen zu einem Einsatz automatischer Textanalysen in verschiedenen Phasen des Fremdsprachenunterrichts sowie bei der Erstellung und Kontrolle von Lehrwerken wurden in den letzten Jahren begonnen. Ausgangspunkt waren eventuelle Einsatzmöglichkeiten eines Grammatik-Prüfprogramms in Sprachlernprogrammen sowie Unterstützungsmöglichkeiten für Autoren von fremdsprachlichen Lehrwerken. Diese Überlegungen haben zu einigen prototypischen Softwaresystemen geführt, die heute praktisch demonstriert werden können.

## 2 Linguistische Analyse von Texten in Lehrwerken für Fremdsprachen

Grundvoraussetzung für Anwendungswerkzeuge, wie sie oben angeführt sind, ist eine

automatische Analyse der textlichen Elemente eines Lehrwerks (Texte, Übungen, Vokabelverzeichnisse etc.). Hierzu muss das Werk in einer maschinenlesbaren Form vorliegen, was bei modernen Publikationsmethoden eine Selbstverständlichkeit ist. Die Extraktion des Textmaterials, d.h. seine Trennung von Formatanweisungen, Bildern, Grafiken etc., ist zwar nicht immer trivial, kann aber je nach verwendetem Publikationsprogramm meist ohne größere Probleme durchgeführt werden.

---

-2-

In sehr einfacher Form (bestimmt für Texte, die ein Englisch-Lehrer zusätzlich zu einem Lehrwerk benutzen will) existiert ein solches Werkzeug in Form der 'Toolbox English G' (Cornelsen Verlag). Diese Toolbox enthält das Lexikon mit dem gesamten Wortschatz eines bestimmten Lehrwerks, wobei neben den Grundformen von Verben auch (regelmäßige und unregelmäßige) Partizipien aufgeführt sind. Beim Vorliegen einer flektierten Form (-s als Endung der 3. Person Singular oder als Plural eines Substantivs) reicht es im Englischen ja aus, wenn man im Wörterbuch an die alphabetisch entsprechende Stelle geführt wird. Komplexe Verben (*turn on*) oder andere zusammengesetzte Ausdrücke sind nicht im Demo-Wörterbuch enthalten (das nur das Vokabular der ersten beiden Jahrgänge enthält). So kann der Lehrer beispielsweise rasch einen neuen Text darauf überprüfen, welche (Einzel-)Wörter in diesem Text mit welcher Bedeutung in welcher Lektion schon gelernt wurden. Wörter, die nicht zum Wortschatz English G gehören, werden als unbekannt markiert.

Eine Voraussetzung für komplexere Analysen bzw. für die Anwendung solcher Werkzeuge auf stärker flektierende Sprachen ist das Vorliegen relativ gut entwickelter und getesteter Analyseprogramme für Morphologie und Syntax sowie ausreichender Wörterbücher der entsprechenden Fremdsprache. Satzerkennungsprogramme zerlegen zunächst den Text in einzelne Sätze und Wörter; Analyse- und Lemmatisierungsprogramme erkennen flektierte Formen und Mehrwortlexeme und führen sie auf standardisierte Darstellungen zurück, wie sie in Lexika und Vokabelverzeichnissen zu finden sind (frz. *répétons* -> *répéter*, *sac + à + dos* -> *sac\_à\_dos*, etc.). Die Programme erkennen auch, welche Konjugationsmuster hier angewandt werden. Zur Reduzierung von Mehrdeutigkeiten durch den Kontext (z.B. *fait* Partizip vs. *fait* Nomen vs. *fait* 3. Pers. Sing.) ist auch eine (zumindest flache) syntaktische Analyse notwendig, die die Wörter des Satzes in sinnvolle Gruppen (Syntagmen) zusammenfasst. Das Beispiel im Anhang zeigt einen Ausschnitt aus der Bearbeitung eines französischen Lehrbuchtextes; die Prüf- und Suchprogramme sind jedoch sprachunabhängig.

Linguistische Analysen stehen für eine Reihe von Fremdsprachen zur Verfügung; mit abnehmender Abdeckung (d.h. Größe des Wörterbuchs) sind dies derzeit:

Deutsch, Englisch, Französisch, Spanisch, Schwedisch, Russisch, Niederländisch, Bulgarisch, Italienisch, Portugiesisch, Slowenisch, Esperanto, Latein.

Vorhandene Wörterbücher können jedoch leicht integriert werden; auch weitere, zumindest westeuropäische, Sprachen bedeuten keinen größeren Aufwand.

Die aktuellen Umfänge der elektronischen Wörterbücher sowie Beschreibungen anderer Anwendungen der linguistischen Analyse finden sich im Internet unter [www.iai.uni-sb.de](http://www.iai.uni-sb.de).

### 3 Werkzeuge für den Lehrbuchautor bzw. Redakteur

Eine solche Computeranalyse des Textes bietet eine Reihe weiterer Möglichkeiten bei Erstellung und Einsatz fremdsprachlicher Lehrwerke. Als erstes sollen kurz die Möglichkeiten der Unterstützung für den Autor oder Redakteur eines Verlags dargestellt werden, die sich durch den Einsatz solcher Algorithmen ergeben.

#### 3.1 Prüfprogramm für einzelne Lektionen

Sind für die einzelnen Lektionen eines Lehrbuchs solche Analysen erstellt worden, so kann man eine Lektion gegen die "bekannten" Vokabeln und grammatischen Phänomene der vorausgehenden Lektionen (und auch vorausgehender Lehrwerksbände) prüfen lassen. Zu jedem Wort bzw. grammatischen Phänomen im Text kann das Programm bestimmen, ob dies schon gelernt wurde. Farbliche Markierungen können anzeigen, ob das Wort überhaupt nicht zur entsprechenden Sprache gehört (z.B. Schreibfehler, neue Personennamen, falsch gesetzte Übungsanweisungen - durch Vergleich mit einem allgemeinen Lexikon, s.u.), ob das Wort als solches neu ist, ob seine Form (z.B. unregelmäßiger Plural oder Futurbildung) neu ist, ob das Wort eventuell mehrdeutig interpretiert werden kann (Gefahr von Mißverständnissen beim Lerner!) etc. Dies ist im Anhang an einem (französischen) Beispiel veranschaulicht.

#### 3.2 Automatische Erstellung und Kontrolle von Vokabelverzeichnissen

Das LiLa-System verfügt über relativ umfangreiche Wörterbücher der behandelten Sprache und vermag deshalb auch Wörter und Ausdrücke zu klassifizieren, die nicht im Lehrbuch enthalten sind. Für den Autor kann LiLa dadurch einen ersten Entwurf eines Vokabelverzeichnisses in der gewünschten Form erstellen; natürlich muss eine intellektuelle Nachkontrolle erfolgen, die aber mit sehr viel geringerem Aufwand verbunden ist. Bei fertigen Lehrwerken kann eine systematische Kontrolle durchgeführt werden. Wenn ein (z.B. im selben Verlag vorhandenes) Wörterbuch als Basis oder Bezugspunkt genommen werden soll, kann auch hiermit ein automatischer Abgleich vorgenommen werden.

### 4 Einsatz bei Wörterbüchern auf CD-ROM

Bei der Verwendung eines Wörterbuchs auf dem PC ist es oftmals nur möglich, Grundformen nachzuschlagen. Zwar bringen Suchsysteme (wie in der oben geschilderten Toolbox oder in der Duden CD) den Benutzer eines solchen Wörterbuches wenigstens an einen dem gesuchten Wort ähnlichen Eintrag (bei der Suche nach dem Verb *ahmen* immerhin zu *ahnen* - im Text steht *ahnen .... nach*), aber bei der Suche nach *gegangen* auf den Eintrag *gehen*.

Beim Einsatz einer linguistischen Analyse kann der Benutzer im Falle von *gegangen* auf *gehen* geführt werden, im Falle von *gehört* mindestens auf die beiden Einträge *hören* und *gehören*, und im Falle von *ahmen ... nach* auf den richtigen Eintrag *nachahmen*.

## 5 Einsatz in Sprachlernsoftware-Paketen

### 5.1 Nachschlagen von unbekanntem Wörtern in Texten und Übungen

Die im vorigen Absatz geschilderte Methode kann auch bei Erklärungen von Wörtern in Sprachlernsoftware-Paketen verwendet werden, bei denen dann nicht alle Flexionsformen von Hand mit der Grundform verbunden werden müssen; wenn die Zuordnung eindeutig erkennbar ist, kann auch in Zweifelsfällen die richtige Erklärung automatisch präsentiert werden.

Beispiel: *Das Auto gehört mir : gehören*

### 5.2 Komplexe Analyse von Lernerantworten

In den meisten der heute verfügbaren 'Grammatiktrainer' und verwandter Software werden die Antworten der Lerner mit vorgegebenen Möglichkeiten verglichen oder sogar nur mit richtig/falsch gekennzeichnet. Aus diesem Grund können bei Lückentexten meist auch nur ein bis zwei Wörter eingesetzt werden; Rechtschreibfehler werden hier von vornherein als falsche Antworten interpretiert.

Die Integration eines Rechtschreibprüfprogramms, wie sie in einzelnen Trainingspaketen vorgenommen ist, bringt hier bereits wesentliche Verbesserungen.

Das IAI hat einen Formalismus entwickelt, mit dem auf der Basis der oben geschilderten linguistischen Analyse individuelle syntaktische Patterns geprüft und mit individuellen Fehlermeldungen und Hinweisen auf Grammatikparagraphen und Übungen verbunden werden können. Das Prüfprogramm in LiLa erfüllt einerseits die Funktion eines Rechtschreibprüfprogramms auf Wortebene und überprüft andererseits grammatische Korrektheit auf Satzebene. Als fehlerhaft erkannte grammatische Strukturen werden mit einer Fehlerdiagnose versehen. Behandelte Fehlerklassen umfassen Kongruenz, Wortstellung, Kasusrektion und Zeichensetzung, z.B.:

- *Das Programm suchen den Fehler.*

“*Das Programm*”: Subjekt und Verb müssen im Numerus übereinstimmen.

- *Sie findet den Fehler, nachdem ihr hilft das System.*

“*nachdem*”: Nach dieser Konjunktion muß das finite Verb am Satzende stehen.

- *Rettet dem Dativ!*

“*dem Dativ*”: Diese Nominalgruppe ist vom Verb nicht erlaubt.

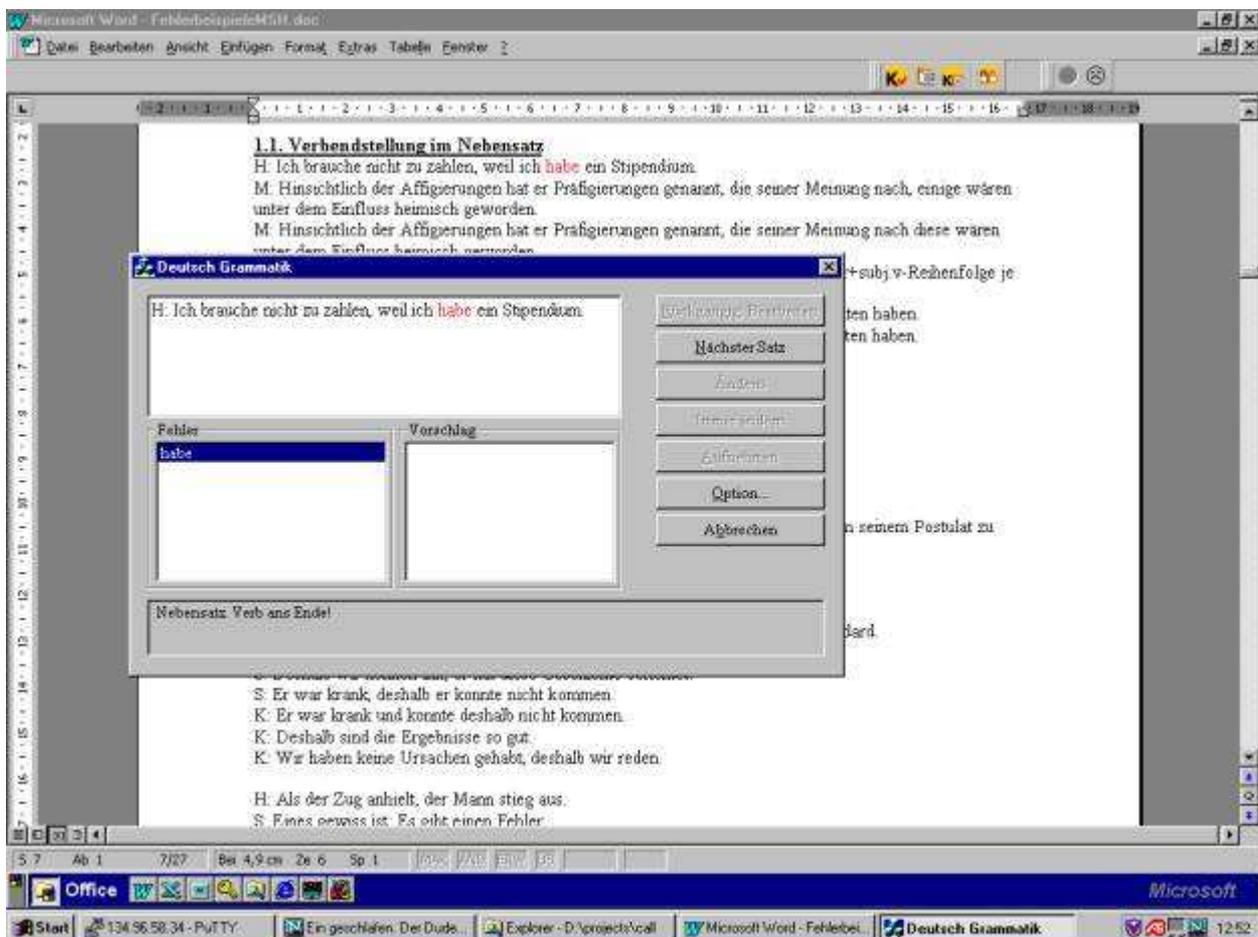
Integriert in ein CALL System kann LiLa zur automatischen Überprüfung von schriftlichen Übungen am Computer eingesetzt werden. Es ermöglicht eine wesentlich flexiblere und genauere Fehlerdiagnose als gängige Pattern-Matching-Techniken. Auch von Lernern relativ frei formulierte Texte können auf grammatische Korrektheit geprüft werden. Dies eröffnet die

Möglichkeit, auch anspruchsvollere Übungsformen anzubieten, wie z.B. Fragen zum Text, Zusammenfassungen, Bildergeschichten oder Übersetzungen, da eine erste Fehlerdiagnose automatisch durchgeführt werden kann.

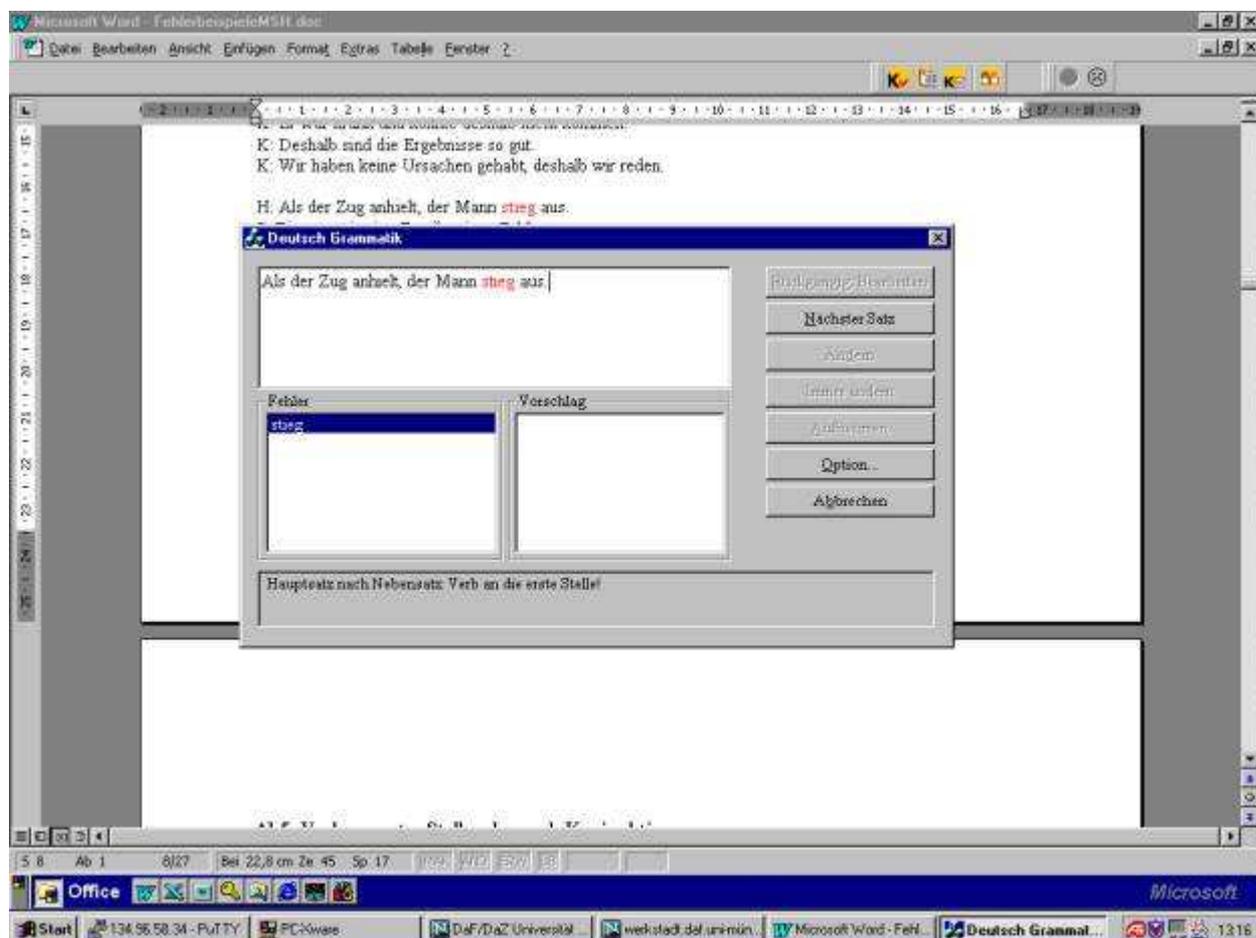
-5-

Einige einfache, aber syntaktisch interessante Beispiele aus der unten zitierten Fehlersammlung sollen dies veranschaulichen.

Der jeweils rot gekennzeichnete Fehler wird unten im Kommentarfeld erläutert; so wird im ersten Beispiel gegen die „Verbendstellung im Nebensatz“ verstossen:



Auch Verstöße gegen die Verbstellung in einem Hauptsatz können erkannt werden, dem ein Nebensatz vorausgeht:



Natürlich lassen sich auch Links zu einer Hypertext-Grammatik oder Verweise auf entsprechende Kapitel in einem konventionellen Lehrwerk bzw. auf Übungssequenzen einbauen. Die von LiLa erstellten Fehlerdiagnosen können an die Bedürfnisse (z.B. Muttersprache, Kenntnisstand) der Lerner angepasst werden. Eine optimale Anpassung ist zu erreichen, indem über die technische Integration hinaus Grammatik und Lexikon von LiLa gezielt auf ein bestimmtes computergestütztes Lehrwerk zugeschnitten werden. Bei einer Verbindung mit den in den ersten Kapiteln beschriebenen Methoden kann der Prozess der Anpassung automatisch unterstützt werden.

-6-

Eine Weiterentwicklung und erste Erprobung dieses Konzepts finden innerhalb des Projekts [uni-deutsch.de](http://www.iai.uni-sb.de/pilot/pilot.html) (<http://www.iai.uni-sb.de/pilot/pilot.html> existiert leider nicht mehr. Stand: 1. Januar 2004) statt, das unter der Federführung der „Werkstadt“ Deutsch als Fremdsprache (<http://werkstadt.daf.uni-muenchen.de/home2.htm>) der Universität München ein Lehrwerk zum Online-Unterricht in der Wissenschaftssprache Deutsch zur Studienvorbereitung und Forschungsbetreuung ausländischer Studierender und Wissenschaftler an deutschen Hochschulen entwickelt.

Als erste Schritte wurden neben dem Abgleich der Vokabeln mit der Liste des Zertifikats „Deutsch als Fremdsprache“ die Erzeugung von Lexika in bestimmten Formaten sowie die

systematische Analyse der bei Heringer gesammelten und klassifizierten typischen Fehler solcher Deutschlerner (<http://www.philhist.uni-augsburg.de/Faecher/GERMANIS/daf/forschung/fehler/index.html> [Der Link existiert leider nicht mehr. Stand: 3. Januar 2005]) in Angriff genommen.

## **6 Zukunftsvision: Das integrierte Lehrwerk auf CD-ROM und/oder Internet**

Wie in diesem Projekt schon versuchsweise angedacht, könnte sich in der Zukunft ein Szenario ergeben, wo das gesamte Lehr- und Lernmaterial nicht mehr auf Papier existiert, sondern wo alle Komponenten nur in elektronischer Form erstellt und über CD und/oder Internet vertrieben werden: Lehrbuch, Arbeitsbuch, Grammatik, Lernsoftware, Wörterbuch. Eine solche Konstellation wird dann optimal genutzt werden können, wenn sie maximal miteinander vernetzt ist, und wenn eine intelligente Verbindung von einer Lernerproduktion (Einzelwort, Wortliste beim Vokabeltest, Satz bei der grammatischen Übung, Übersetzung, freie Nacherzählung, Rollenspiel, Aufsatz) zu dem vom Lehrbuch- und Grammatikautor geschaffenen Material besteht. Eine solche intelligente Verbindung ist die linguistische Analyse. Sie wird auch für die Schaffung und Erhaltung der Vokabel- und Grammatikkonsistenz innerhalb eines solchen 'Hyperlehrwerks' unabdingbar sein.

---

*Copyright © 2002 Zeitschrift für Interkulturellen Fremdsprachenunterricht*

Haller, Johann. (2002). LILA: Linguistisch intelligente Lehrwerksanalyse. <i>Zeitschrift für Interkulturellen Fremdsprachenunterricht</i> [Online], 7(1), 6 pp. Available: <a href="http://www.spz.tu-darmstadt.de/projekt_ejournal/jg_07_1/beitrag/haller1.htm">http://www.spz.tu-darmstadt.de/projekt_ejournal/jg_07_1/beitrag/haller1.htm</a>
--

---

[Zurück zur [Leitseite](#) der Nummer im Archiv]