

Hanna Hedeland, Thomas Schmidt, Kai Wörner (eds.)

## Multilingual Resources and Multilingual Applications

Proceedings of the Conference of the  
German Society for Computational Linguistics and  
Language Technology (GSCL) 2011



Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Sonderforschungsbereich  
Mehrsprachigkeit





Hanna Hedeland, Thomas Schmidt, Kai Wörner (eds.)

Multilingual Resources and Multilingual Applications

Proceedings of the Conference of the German Society for  
Computational Linguistics and Language Technology (GSCL) 2011

© *Hanna Hedeland, Thomas Schmidt, Kai Wörner*  
*Hamburger Zentrum für Sprachkorpora*  
*Max Brauer-Allee 60*  
*D-22765 Hamburg*

Die „Arbeiten zur Mehrsprachigkeit – Folge B“ publizieren Forschungsarbeiten aus dem Sonderforschungsbereich 538 *Mehrsprachigkeit*, der von der Deutschen Forschungsgemeinschaft im Juli 1999 an der Universität Hamburg eingerichtet wurde. Wir danken der DFG für ihre Unterstützung.

Die „Arbeiten zur Mehrsprachigkeit – Folge B“ sind bei der Deutschen Bibliothek in Frankfurt/Main mit der Seriennummer ISSN 0176-559X eingetragen.

Redaktion:  
*Martin Elsig, Svenja Kranich, Thomas Schmidt, Manuela Schönenberger*  
Technische Umsetzung:  
*Thomas Schmidt*

**Collaborative Research Center: Multilingualism**  
**Sonderforschungsbereich 538: Mehrsprachigkeit**  
**University of Hamburg**

Founded in July 1999, the Collaborative Research Centre on Multilingualism conducts research on patterns of language use in multilingual environments, bilingual language acquisition, and the role of multilingualism and language contact for language change.

In the current, fourth funding period (2008–2011), the Centre comprises two main research branches, each of which focuses on a central set of common issues, and a third branch of projects dealing with practical application solutions. Branch E, *Multilingual Language Acquisition*, consists of four projects, with a common focus on the nature of “critical phases” in language acquisition. Branch H, *Historical Aspects of Multilingualism and Variation*, consists of eight projects, dealing with questions of language change and language contact. This branch also comprises projects of former separate branch K, *Multilingual Communication*.

Since 2007, a new Branch T, *Transfer Projects*, has been active. It consists of five projects whose goal is to develop concrete solutions for practical problems relating to multilingual situations, based on research results derived from the Centre’s research activities.

Languages currently studied at the Research Centre include the following: Danish, Catalan, English, Faroese, French, German, German Sign Language, Icelandic, Irish, Polish, Portuguese, Spanish, Swedish, and Turkish, as well as several historical or regional sub-varieties of some of these languages.

**Chair:**

Prof. Dr. Christoph Gabriel  
christoph.gabriel@uni-hamburg.de

**Co-chairs:**

Prof. Dr. Kurt Braunmüller  
braunmueller@rrz.uni-hamburg.de

Prof. Dr. Barbara Hänel-Faulhaber  
Barbara.Haenel@uni-hamburg.de



## **Local Organizing Comittee**

- Thomas Schmidt
- Kai Wörner
- Timm Lehmberg
- Hanna Hedeland

## **Program Committee**

- Maja Bärenfänger (Universität Gießen)
- Stefanie Dipper (Universität Bochum)
- Kurt Eberle (Lingnio Heidelberg)
- Alexander Geyken (Berlin-Brandenburgische Akademie der Wissenschaften)
- Ullrich Heid (Universität Hildesheim)
- Claudia Kunze (Qualisys GmbH)
- Lothar Lemnitzer (Berlin-Brandenburgische Akademie der Wissenschaften)
- Henning Lobin (Universität Gießen)
- Ernesto de Luca (Technische Universität Berlin)
- Cerstin Mahlow (Universität Zürich)
- Alexander Mehler (Universität Bielefeld)
- Wolfgang Menzel (Universität Hamburg)
- Georg Rehm (Deutsches Forschungszentrum für Künstliche Intelligenz)
- Josef Ruppenhofer (Universität Saarbrücken)
- Thomas Schmidt (Universität Hamburg)
- Roman Schneider (Institut für Deutsche Sprache Mannheim)
- Bernhard Schröder (Universität Duisburg)
- Manfred Stede (Universität Potsdam)
- Angelika Storrer (Universität Dortmund)
- Maik Stührenberg (Universität Bielefeld)
- Thorsten Trippel (Universität Tübingen)
- Cristina Vertan (Universität Hamburg)
- Andreas Witt (Institut für Deutsche Sprache Mannheim)
- Christian Wolff (Universität Regensburg)
- Kai Wörner (Universität Hamburg)

## Call for Papers

The Conference of the German Society for Computational Linguistics and Language Technology (GSCL) in 2011 will take place from 28th to 30th September 2011 at the University of Hamburg. The main conference theme is Multilingual Resources and Multilingual Applications.

Contributions to any topic related to Computational Linguistics and Language Technology are invited, but we especially encourage submissions that are related to the main theme. The topic “Multilingual Resources and Multilingual Applications” comprises all aspects of computational linguistics and speech and language technology in which issues of multilingualism, of language contrasts or of language independent representations play a major role. This includes, for instance:

- representation and analysis of parallel corpora and comparable corpora
- multilingual lexical resources
- machine translation
- annotation and analysis of learner corpora
- linguistic variation in linguistic data and applications
- localisation and internationalisation

Conference languages are English and German; contributions are welcome in both languages. Three types of submission are invited:

- Regular talk – Submission of an extended abstract
- Poster – Submission of an abstract
- System presentation – Submission of an abstract

Only contributions in electronic form will be accepted. We do not provide style sheets for submissions at this stage; constraints on length are given below. Submissions must be submitted via the electronic conference system. Accepted abstracts will be published as a book of abstracts on the occasion of the conference. Extended versions of the papers will be published as a special issue of the Journal for Language Technology and Computational Linguistics after the conference.

# GSCL

*German Society for  
Computational Linguistics  
& Language Technology*



**Universität Hamburg**  
DER FORSCHUNG | DER LEHRE | DER BILDUNG

**hzsk** hamburger zentrum  
für sprachkorpora

## Contents

### I Invited Talks

- Constructing Parallel Lexicon Fragments Based on English FrameNet Entries:  
Semantic and Syntactic Issues* ..... 9  
Hans C. Boas
- The Multilingual Web: Opportunities, Borders and Visions*..... 19  
Felix Sasaki
- Combining Various Text Analysis Tools for Multilingual Media Monitoring*..... 25  
Ralf Steinberger

### II Regular Papers

- Generating Inflection Variants of Multi-Word Terms for French and German*..... 33  
Simon Clematide, Luzia Roth
- Tackling the Variation in International Location Information Data: An Approach  
Using Open Semantic Databases* ..... 39  
Janine Wolf, Manfred Stede, Michaela Atterer
- Towards Multilingual Biographical Event Extraction - Initial Thoughts on the  
Design of a New Annotation Scheme*..... 45  
Michaela Geierhos, Jean-Leon Bouraoui, Patrick Watrin
- The Corpus of Academic Learner English (CALE): A New Resource for the Study  
of Lexico-Grammatical Variation in Advanced Learner Varieties*..... 51  
Marcus Callies, Ekaterina Zaytseva
- From Multilingual Web-Archives to Parallel Treebanks in Five Minutes*..... 57  
Markus Killer, Rico Sennrich, Martin Volk
- Querying Multilevel Annotation and Alignment for Detecting Grammatical  
Valence Divergencies* ..... 63  
Oliver Čulo
- SPIGA – A Multilingual News Aggregator*..... 69  
Leonhard Hennig, Danuta Ploch, Daniel Prawdzik, Benjamin Armbruster,  
Christoph Büscher, Ernesto William De Luca, Holger Düwiger, Şahin Albayrak
- From Historic Books to Annotated XML: Building a Large Multilingual  
Diachronic Corpus* ..... 75  
Magdalena Jitca, Rico Sennrich, Martin Volk
- Visualizing Dependency Structures*..... 81  
Chris Culy, Verena Lyding, Henrik Dittmann
- A Functional Database Framework for Querying Very Large Multi-Layer Corpora*..... 87  
Roman Schneider

<i>Hybrid Machine Translation for German in taraXÜ: Can Translation Costs Be Decreased Without Degrading Quality?</i> .....	93
Aljoscha Burchardt, Christian Federmann, Hans Uszkoreit	
<i>Annotation of Explicit and Implicit Discourse Relations in the TüBa-D/Z Treebank</i> .....	99
Anna Gastel, Sabrina Schulze, Yannick Versley, Erhard Hinrichs	
<i>Devil's Advocate on Metadata in Science</i> .....	105
Christina Hoppermann, Thorsten Trippel, Claus Zinn	
<i>Improving an Existing RBMT System by Stochastic Analysis</i> .....	111
Christian Federmann, Sabine Hunsicker	
<i>Terminology Extraction and Term Variation Patterns: A Study of French and German Data</i> .....	117
Marion Weller, Helena Blancafort, Anita Gojun, Ulrich Heid	
<i>Ansätze zur Verbesserung der Retrieval-Leistung kommerzieller Translation-Memory-Systeme</i> .....	123
Dino Azzano, Uwe Reinke, Melanie Sauer	
<i>WikiWarsDE: A German Corpus of Narratives Annotated with Temporal Expressions</i> .....	129
Jannik Strötgen, Michael Gertz	
<i>Translation and Language Change with Reference to Popular Science Articles: The Interplay of Diachronic and Synchronic Corpus-Based Studies</i> .....	135
Sofia Malamatidou	
<i>A Comparable Wikipedia Corpus: From Wiki Syntax to POS Tagged XML</i> .....	141
Noah Bubenhofer, Stefanie Haupt, Horst Schwinn	
<i>A German Grammar for Generation in OpenCCG</i> .....	145
Jean Vancoppenolle, Eric Tabbert, Gerlof Bouma, Manfred Stede	
<i>Multilingualism in Ancient Texts: Language Detection by Example of Old High German and Old Saxon</i> .....	151
Zahurul Islam, Roland Mittmann, Alexander Mehler	
<i>Multilinguale Phrasenextraktion mit Hilfe einer lexikonunabhängigen Analysekomponente am Beispiel von Patentschriften und nutzergenerierten Inhalten</i> .....	157
Daniela Becks, Julia Maria Schulz, Christa Womser-Hacker, Thomas Mandl	
<i>Die Digitale Rätoromanische Chrestomathie – Werkzeuge und Verfahren für die Korpuserstellung durch kollaborative Volltexterschließung</i> .....	163
Claes Neuefeind, Jürgen Rolshoven, Fabian Steeg	
<i>Ein umgekehrtes Lehnwörterbuch als Internetportal und elektronische Ressource: Lexikographische und technische Grundlagen</i> .....	169
Peter Meyer, Stefan Engelberg	
<i>Localizing a Core HPSG-Based Grammar for Bulgarian</i> .....	175
Petya Osenova	

### III Poster Presentations

<i>Autorenunterstützung für die Maschinelle Übersetzung</i> .....	183
Melanie Siegel	
<i>Experimenting with Corpus-Based MT Approaches</i> .....	187
Monica Gavrilă	
<i>Method of POS-Disambiguation Using Information about Words Co-Occurrence (For Russian)</i> .....	191
Edward Klyshinsky, Natalia Kochetkova, Maxim Litvinov, Vadim Maximov	
<i>Von TMF in Richtung UML: In drei Schritten zu einem Modell des übersetzungsorientierten Fachwörterbuchs</i> .....	197
Georg Löckinger	
<i>Annotating for Precision and Recall in Speech Act Variation: The Case of Directives in the Spoken Turkish Corpus</i> .....	203
Şükriye Ruhi, Thomas Schmidt, Kai Wörner, Kerem Eryılmaz	
<i>The SoSaBiEC Corpus: Social Structure and Bilinguality in Everyday Conversation</i> .....	207
Veronika Ries, Andy Lücking	
<i>DIL, ein zweisprachiges Online-Fachwörterbuch der Linguistik (Deutsch-Italienisch)</i> .....	211
Carolina Flinz	
<i>Knowledge Extraction and Representation: The EcoLexicon Methodology</i> .....	215
Pilar León Araúz, Arianne Reimerink	
<i>Processing Multilingual Customer Contacts via Social Media</i> .....	219
Michaela Geierhos, Yeong Su Lee, Matthias Bargel	
<i>ATLAS – A Robust Multilingual Platform for the Web</i> .....	223
Diman Karagiozov, Svetla Koeva, Maciej Ogrodniczuk, Cristina Vertan	
<i>Multilingual Corpora at the Hamburg Centre for Language Corpora</i> .....	227
Hanna Hedeland, Timm Lehmberg, Thomas Schmidt, Kai Wörner	
<i>The English Passive and the German Learner – Compiling an Annotated Learner Corpus to Investigate the Importance of Educational Settings</i> .....	233
Verena Möller, Ulrich Heid	
<i>Register, Genre, Rhetorical Functions: Variation in English Native-Speaker and Learner Writing</i> .....	239
Ekaterina Zaytseva	
<i>Tools to Analyse German-English Contrasts in Cohesion</i> .....	243
Kerstin Kunz, Ekaterina Lapshinova-Koltunski	
<i>Comparison and Evaluation of Ontology Extraction Systems</i> .....	247
Stefanie Reimers	

## **IV System Presentations**

<i>New and Future Developments in EXMARaLDA</i> .....	253
Thomas Schmidt, Kai Wörner, Hanna Hedeland, Timm Lehmborg	
<i>Der VLC Language Index</i> .....	257
Dirk Schäfer, Jürgen Handke	
<i>Topological Fields, Constituents and Coreference: A New Multi-Layer Architecture for TüBa-D/Z</i> .....	259
Thomas Krause, Julia Ritz, Amir Zeldes, Florian Zipser	
<i>MT Server Land Translation Services</i> .....	263
Christian Federmann	

# DIL, ein zweisprachiges Online-Fachwörterbuch der Linguistik (Deutsch-Italienisch)

**Carolina Flinz**

Universität Pisa

E-mail: c.flinz@ec.unipi.it

## Abstract

DIL ist ein deutsch-italienisches Online-Fachwörterbuch der Linguistik. Es ist ein offenes Wörterbuch und mit diesem Beitrag wird für eine mögliche Zusammenarbeit, Kollaboration plädiert. DIL ist noch im Aufbau begriffen; zur Zeit ist nur die Sektion DaF komplett veröffentlicht, auch wenn andere Sektionen in Bearbeitung sind. Die Sektion LEX (Lexikographie), die zur Veröffentlichung ansteht, wird zusammen mit den wichtigsten Eigenschaften des Wörterbuches präsentiert.

Keywords: Fachwörterbuch, Linguistik, zweisprachig, deutsch-italienisch, Online-Wörterbuch

## 1. Einleitung

DIL (*Dizionario tedesco-italiano di terminologia linguistica / deutsch-italienisches Fachwörterbuch der Linguistik*) ist ein online Wörterbuch, das Lemmata aus dem Bereich der Linguistik und einiger ihrer Nachbardisziplinen auflistet. Es ist ein offenes Wörterbuch, nach dem Muster von *Wikipedia*, bzw. *Glottopedia*, um eine mögliche Beteiligung von Experten der unterschiedlichen Disziplinen zu fördern.

Im Handel und im Online-Medium existieren heute mehrere deutsche<sup>1</sup> und italienische<sup>2</sup> Wörterbücher der Linguistik aber kein einziges Fachwörterbuch für das Sprachenpaar deutsch-italienisch. Hingegen ist der Bedarf an einem solchen „Instrument“ in Italien, sowohl für die universitäre Didaktik als auch für die Forschung, sehr stark: in einem Zeitraum, wo das Fach „Deutsche Linguistik“ als Folge einer Universitätsreform (1999) einen starken Aufschwung erlebt hat, könnte DIL für die wissenschaftliche Kommunikation von großer Relevanz sein<sup>3</sup>. DIL könnte nämlich eine große Hilfe für die Suche

nach Äquivalenten von deutschsprachigen linguistischen Fachtermini sein.

DIL ist ein Projekt des Deutschen Instituts der Fakultät *Lingue e Letterature Straniere* der Universität Pisa (daf, 2004:37), das 2008 online veröffentlicht worden ist ([http://www.humnet.unipi.it/dott\\_linggensac/glossword](http://www.humnet.unipi.it/dott_linggensac/glossword)) und an dem weiterhin gearbeitet wird. Es handelt sich um ein monolemmatisches Fachwörterbuch<sup>4</sup> (Wiegand, 1996:46): die Lemmata sind in deutscher Sprache, während die Kommentarsprache Italienisch ist.

Ziele dieses Beitrags sind:

- 1) durch einen kurzen Überblick die wichtigsten Eigenschaften des Wörterbuches vorzustellen, wie Makro- und Mikrostruktur des Wörterbuches, Lemmabestand und Kriterien;
- 2) zu ähnlichen Arbeiten und zukünftigen Kollaborationen an diesem Projekt anzuregen, insbesondere für die geplante Sektion der Computerlinguistik;
- 3) die gerade neu erstellte Sektion LEX (Lexikographie) vorzustellen.

## 2. Makro- und Mikrostruktur

Die Makrostruktur und die Mikrostruktur von DIL wurden natürlich von der Funktion des Wörterbuches und der intendierten Benutzergruppe beeinflusst<sup>5</sup>. Die Erkundung der Benutzerbedürfnisse wurde mit Hilfe von

<sup>1</sup> Vgl. u.a. Bußmann, 2002; Conrad, 1985; Crystal, 1993; Ducrot & Todorov, 1975; Dubois, 1979; Glück, 2000; Heupel, 1973; Lewandowki, 1994; Meier & Meier, 1979; Stammerjohann, 1975; Ulrich, 2002.

<sup>2</sup> Vgl. u.a. Bußmann, 2007; Cardona, 1988; Casadei, 1991; Ceppellini, 1999; Courtes & Greimas, 1986; Crystal, 1993; Ducrot & Todorov, 1972; Severino, 1937; Simone, 1969.

<sup>3</sup> Die Relevanz von Fachwörterbüchern für die wissenschaftliche Kommunikation war Thema vieler lexikographischer Arbeiten: vgl. u.a. Wiegand, 1988; Pileegard, 1994; Schaedler & Bergenholtz, 1994; Bergenholtz & Tarp, 1995; Hoffmann & Kalverkämper & Wiegand, 1998.

<sup>4</sup> Eine bilinguistische Ergänzung des Wörterbuches ist nicht ausgeschlossen.

Fragebögen, die sowohl im Printmedium als auch im Onlineformat gesendet wurden, und einer Analyse der möglichen Benutzersituationen erforscht<sup>6</sup>. Jeder Benutzer kann weiterhin den Fragebogen von der Homepage aufrufen und beantworten, so dass ein ständiger Kontakt mit dem Benutzer vorhanden ist.

DIL wendet sich im Allgemeinen an ein heterogenes Publikum: es ist sowohl für Experten als auch für Laien gedacht, so dass die potentiellen Benutzer sowohl Lerner und Lehrender in den Bereichen Germanistik, Romanistik, Linguistik oder Deutsch / Italienisch als Fremdsprache sein können als auch Lehrbuchautoren, Lexikographen oder Fachakademiker. Das Online Medium, dank seiner Flexibilität, ist von großem Vorteil in dieser Hinsicht.

DIL kann nämlich in folgenden Benutzungssituationen verwendet werden:

- 1) Der Benutzer sucht bestimmte fachliche Informationen, und das Wörterbuch, laut seiner Werkzeugnatur, erfüllt das Bedürfnis;
- 2) Der Benutzer greift zum Wörterbuch, um ein Kommunikationsproblem in der Textproduktion, Textrezeption oder Übersetzung zu lösen. DIL erfüllt deswegen mehrere Funktionen: es kann sowohl für aktive / produktive als auch passive / rezeptive Tätigkeiten verwendet werden.
  - a. Der italophone Benutzer (primärer Benutzer) wird es als dekodierendes Wörterbuch für die Herübersetzung verwenden, d.h. wenn er ein deutsches Fachwort verstehen will oder dessen Übersetzung sucht, oder wenn er spezifischere Informationen braucht und sich weiter informieren und weiterbilden möchte;
  - b. Der deutschsprachige Benutzer wird es als enkodierendes Wörterbuch für die Hinproduktion benutzen, d.h. wenn er ins Italienische übersetzt und Fachtexte in italienischer Sprache erstellt.

Die Makrostruktur von DIL vereinigt sowohl Eigenschaften der linguistischen Printwörterbücher (1.)

<sup>5</sup> Vgl. u.a. Storrer & Harriehausen, 1998; Barz, 2005.

<sup>6</sup> Für einen Überblick über mögliche Techniken zur (Benutzerbedürfnissen-Erforschung: meglio zur Erforschung von Benutzerbedürfnissen) vgl. u.a. Barz, 2005; Ripfel & Wiegand, 1988; Schaefer & Bergenholtz, 1994; Wiegand, 1977.

als auch der Onlinewörterbücher (2.):

1. Die Strukturierung der Umtexte im Printmedium beeinflusste den aktuellen Stand. DIL verfügt nämlich über folgende nach wissenschaftlichen Kriterien verfasste Umtexte: Einleitung, Abkürzungsverzeichnis, Benutzerhinweise, Redaktionsnormen, Register der Einträge<sup>7</sup>;
2. Die Vorteile der Online-Wörterbücher wurden auch zum größten Teil ausgenutzt:
  - a. Neue Einträge und neue Sektionen können sehr schnell veröffentlicht werden;
  - b. DIL kann ständig erneuert, ergänzt und korrigiert werden;
  - c. Es verfügt über ein klar strukturiertes Menu, in dem die wichtigsten Umtexte verlinkt sind, so dass der Benutzer schnell die gewünschten Informationen erreichen kann;
  - d. Es verwendet sowohl interne<sup>8</sup> als auch externe<sup>9</sup> Hyperlinks;
  - e. Es bietet dem Benutzer nützliche Informationen, wie die TOP 10 (vgl. u.a. die „zuletzt gesuchten“ oder „die am meisten geklickten Lemmata“);
  - f. Es bietet wichtige Instrumente, wie die Suchmaschine, die Feedbackseite, das Login Feld etc.

Die Mikrostruktur von DIL bietet sowohl sprachliche als auch sachliche Informationen und ist auf der Grundlage, dass der Erst-Adressat der italophone Benutzer ist, strukturiert worden. Jeder Eintrag wird von folgenden Angaben komplettiert:

- 1) grammatische Angaben (Genus und Numerus);
- 2) das Äquivalent / die Äquivalente in italienischer Sprache;
- 3) die Markierung als Information zum fachspezifischen Bereich des Lemmas;
- 4) die enzyklopädische Definition;
- 5) Beispiele<sup>10</sup>;

<sup>7</sup> Eine empirische Analyse linguistischer Online-Fachwörterbücher zeigte, wie „unwissenschaftlich“ oft Online-Wörterbücher mit Umtexten umgehen. Nur 45% der analysierten Werkzeuge verfügte über solche Texte und nur in seltenen Ausnahmen wurden wissenschaftlichen Kriterien gefolgt (Flinz, 2010:72)

<sup>8</sup> Der Benutzer kann von einem Eintrag zu thematisch verbundenen Lemmata springen.

<sup>9</sup> Es sind sowohl sprachliche Wörterbücher, wie *Canno.net* und *Grammis*, als auch sachliche, wie *Glottopedia* und *DLM*, verlinkt.

<sup>10</sup> Alle Lemmata folgen im Prinzip dem gleichen Schema, da

- 6) Angaben zur Paradigmatik, wie Synonyme; thematisch verbundene Lemmata;
- 7) bibliographische Angaben.

### 3. Lemmabestand und Kriterien

Der Lemmabestand von DIL kann nur „eingeschätzt“ werden. Die Gründe dafür können wie folgt zusammengefasst werden:

- 1) erstens handelt es sich um ein Online-Wörterbuch, das noch in Projekt und Testphase ist;
- 2) zweitens soll das Werk, wie es sein Format vorgibt, nicht als etwas Statisches und Vollendetes gesehen werden, sondern in ständiger Erweiterung und Erneuerung. Aus einem Vergleich der existierenden linguistischen Fachwörterbücher kann aber eine ungefähre Zahl von ca. 2.000 Lemmata ausgerechnet werden, die allerdings ständig erweitert oder geändert werden kann.

Primärquellen waren allgemeine Wörterbücher der Linguistik (deutsch- wie italienischsprachige), sowie spezifische deutsche und italienische Glossare der Disziplin Lexikographie und Fachlexikographie. Es wurden Quellen sowohl im gedruckten als auch im Online-Medium herangezogen. Sekundärquellen waren Handbücher aus dem Bereich der jeweiligen Disziplin (für die Sektion **Lex** waren es zum Beispiel Standardwerke der Disziplin Lexikographie und Fachlexikographie) sowohl in deutscher als auch in italienischer Sprache.

Hauptkriterien für die Auswahl der Lemmata sind Frequenz und Relevanz (Bergenholtz, 1989:775)<sup>11</sup>:

- 1) Es wurde eine entsprechende Analyse der existierenden lexikographischen Wörterbücher sowohl im Print- als auch im Onlineformat hinsichtlich der dort aufgeführten Lemmata des jeweiligen Bereiches durchgeführt;
- 2) Es wurde ein kleiner Korpus von Fachtexten des betreffenden Faches hergestellt. Die im Endregister enthaltenen Termini wurden in Excell-Tabellen

eingetragen, und die entstehenden Listen wurden auf Grund von Frequenzkriterien verglichen. Die aus diesem Prozess entstehende Endliste wurde zusätzlich auf der Basis des Relevanzkriteriums ergänzt.

Die Einträge sind in strikt alphabetischer Reihenfolge, und die typischen Nachteile dieser Ordnung können dank des Online-Formats teilweise aufgehoben werden, da die begriffsystematischen Zusammenhänge durch verlinkte Verweise oft verdeutlicht werden.

Das Wörterbuch enthält zurzeit eine vollständige Sektion (DaF) mit 240 Einträgen, während andere Bereiche in Erarbeitung sind:

- 1) Historische Syntax;
- 2) Wortbildung;
- 3) Textlinguistik;
- 4) Fachsprachen.

Eine neu erstellte Sektion LEX (Lexikographie) wurde gerade fertiggestellt und steht zur Veröffentlichung an. Sie enthält Lemmata aus dem Bereich der Lexikographie und Fachlexikographie sowie Metalexikographie und Metafachlexikographie.

### 4. Die Sektion: LEX

Die Sektion LEX wird voraussichtlich ca. 120 Einträge (Stand Juni 2011) enthalten, die sich auf die wichtigsten Aspekte des Fachbereiches der Lexikographie konzentrieren. Es wird dabei auf folgende Themen Aufmerksamkeit gelegt:

- a. Lexikographie;
- b. Fachlexikographie;
- c. Wörterbuchforschung;
- d. Wörterbuchtypologie;
- e. Wörterbuchbenutzer und Wörterbuchbenutzung;
- f. Wörterbuchfunktionen;
- g. lexikographische Kriterien;
- h. Makrostruktur;
- i. Umtexte;
- j. Mediostruktur;
- k. Mikrostruktur.

Im Folgenden wird ein Beispiel eines Eintrags aus dem Bereich **Lex** gezeigt. Es kann als Muster für die Erarbeitung von neuen Einträgen gelten. Jeder Autor kann die produzierten Lemmata an die Redaktion des Wörterbuches senden; nach der redaktionellen Prüfung wird der Eintrag veröffentlicht und mit der Abkürzung des Autorennamens vermerkt.

---

die Standardisierung der Mikrostruktur eine wichtige Voraussetzung war. Da aber Beispiele nur in bestimmten Kontexten behilflich sind, wurden sie nur gelegentlich eingefügt.

<sup>11</sup> Korpusanalysen, im Sinne von automatischen Analysen von Textkorpora mit anschließender Korpusauswertung (Frequenzwerte) wurde bis jetzt ausgeschlossen. Jedoch wäre es interessant zu sehen, inwiefern eine solche Analyse mit einer Integrierung des Relevanzkriteriums die erhaltenen Ergebnisse widerspiegeln könnte oder nicht.

## Fachlexikographie

(LEX); (die, PLURAL UNÜBLICH)

Lessicografia specialistica

Termine che indica l'attività scientifica, il cui obiettivo primario è la produzione (ossia la pianificazione, la redazione, la revisione, la stampa ecc.) di dizionari specialistici linguistici, di dizionari enciclopedici e dizionari linguistico-enciclopedici. Pertanto si distingue in:

- lessicografia specialistica linguistica;
- lessicografia specialistica enciclopedica;
- lessicografia linguistica-enciclopedia.

Da coloro che la praticano la lessicografia specialistica pretende non solo una solida base lessicografica, ma anche un competenza nel linguaggio specialistico oggetto del dizionario.

La lessicografia specialistica si attesta come disciplina scientifica a partire da metà degli anni '90, infatti fino ad allora l'interesse per i dizionari scientifici era marginale e spesso si lamentava la mancanza di lavori che si occupassero in modo scientifico di dizionari specialistici. I primi articoli pubblicati trattano ancora la disciplina in modo generico; il primo articolo ufficiale viene considerato quello di Wiegand, H. E. (1988): *Was ist eigentlich Fachlexikographie?*. Dagli anni '90' vengono organizzati le prime giornate di studio e vengono pubblicate le prime bibliografie. (cf)

Vedi anche: [Lexikographie](#)

Fonte: SCHAEFER, B. – BERGENHOLTZ, H. (1994): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*, WIEGAND, H.E. (1988): *Was ist eigentlich Fachlexikographie?*. In Munske, H.H. – Von Polenz, P. – Reichmann, O. – Hildebrandt, R. (Hrsg.) 1988: S. 729-790., WIEGAND, H.E. (1976): *Die Wahrheit der Wörterbücher. In: Probleme der Lexikologie und Lexikographie. Jahrbuch 1975 des Instituts für deutsche Sprache, Düsseldorf (Sprache der Gegenwart XXXIX)*. S. 347-371., WIEGAND, H.E. (1988): *Was ist eigentlich Fachlexikographie?*. In Munske, H.H. – Von Polenz, P. – Reichmann, O. – Hildebrandt, R. (Hrsg.) 1988: S. 729-790., PILEGAARD, M. (1994): *Bilingual LSP Dictionaries. User benefit correlates with elaborateness of „explanation“*. In: Bergenholtz, H. - Schaefer, B. 1994. S. 211-228., KUCERA, A. (1984): *Aus der Werkstatt der praktischen Verlagslexikographie. Übersetzungswörterbücher der Fachsprachen*. In: Mitteilungen für Dolmetscher und Übersetzer 1/30. S. 3-6.

Bild 1: Das Lemma "Fachlexikographie"

### 5. Literatur

- Abel, A. (2006): Elektronische Wörterbücher: Neue Wege und Tendenzen. In San Vincente, F. (Ed.) Akten der Tagung "Lessicografia bilingue e Traduzione: metodi, strumenti e approcci attuali" (Forlì, 17.-18.11.2005). Polimetrica Publisher (Open Access Publications). S. 35-56.
- Almind, R. (2005): Designing Internet Dictionaries. *Hermes*, 34, S. 37-54.
- Barz, I., Bergenholtz, H., Korhonen, J. (2005): Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch. Frankfurt a. M.: Peter Lang.
- Bergenholtz, H. (1989): Probleme der Selektion im allgemeinen einsprachigen Wörterbuch. In Hausmann, F. J. et al. (Hg). *Wörterbücher: ein internationales Handbuch zur Lexikographie*. Band 1. Berlin & New York: de Gruyter. S. 773-779.
- Bergenholtz, H., Tarp, S. (1995): *Manuel of LSP lexicography. Preparation of LSP dictionaries - problems and suggested solutions*. Amsterdam, Netherlands & Philadelphia: J. Benjamins.
- Foschi-Albert, M., Hepp, M. (2004): Zum Projekt: Bausteine zu einem deutsch-italienischen Wörterbuch der Linguistik. In *daf Werkstatt*, 4, S. 43-69.
- Hoffmann, L., Kalverkämper, H., Wiegand, H.E (Eds.) (1999): *Fachsprachen. Handbücher zur Sprach- und Kommunikationswissenschaft (HSK 14.2.)*. Berlin & New York: de Gruyter.
- Pilegaard, M. (1994): *Bilingual LSP Dictionaries. User benefit correlates with elaborateness of „explanation“*. In Bergenholtz, H. & Schaefer, B. S. 211-228.
- Schaefer, B., Bergenholtz, H. (1994): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: G. Narr.
- Ripfel M., Wiegand, H.E. (1988): Wörterbuchbenutzungsforschung. Ein kritischer Bericht. In *Studien zur Neuhochdeutschen Lexikographie VI*. 2. Teilb. S. 482-520.
- Storrer, A., Harriehausen, B. (1998): *Hypermedia für Lexikon und Grammatik*. Tübingen: G. Narr.
- Wiegand, H.E. (1977): *Nachdenken über Wörterbücher. Aktuelle Probleme*. In Drosdowski, H., Henne, H. & Wiegand, H.E. *Nachdenken über Wörterbücher*. Mannheim: Bibliographisches Institut / Dudenverlag. S. 51-102.
- Wiegand, H.E. (Ed.) (1996): *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger Lexikographie-Kolloquium*. Tübingen: Lexicographica Series Major 70.
- Wiegand, H.E. (1988): *Was ist eigentlich Fachlexikographie?*. In Munske, H.H., Von Polenz, P. & Reichmann, O. & Hildebrandt, R. (Hg.). *Deutscher Wortschatz. Lexikologische Studien*. Berlin & New York: de Gruyter. S. 729-790.