



## **Esperanto und andere Sprachen im Vergleich**

**Beiträge  
der 18. Jahrestagung  
der Gesellschaft für Interlinguistik e.V.,  
21.-23. November 2008, in Berlin**

**Herausgegeben von Sabine Fiedler**

**Berlin  
2009**



Die GIL konzentriert ihre wissenschaftliche Arbeit vor allem auf Probleme der *internationalen sprachlichen Kommunikation*, der *Plansprachenwissenschaft* und der *Esperantologie*.

Die Gesellschaft gibt das Bulletin „Interlinguistische Informationen“ (ISSN 1430-2888) heraus und informiert darin über die wichtigsten interlinguistischen und esperantologischen Aktivitäten und Neuerscheinungen international und in Deutschland.

Im Rahmen ihrer Jahreshauptversammlungen führt sie Fachveranstaltungen zu interlinguistischen Problemen durch und veröffentlicht die Akten und andere Materialien.

### **Vorstand der GIL**

- 1. Vorsitzender: Dr. sc. Detlev Blanke
- 2. Vorsitzende: Prof. Dr. habil. Sabine Fiedler
- Schatzmeister: Dipl.-Ing. Horst Jasmann
- Mitglied: Dr. habil. Cornelia Mannewitz
- Mitglied: PD Dr. Dr. Rudolf-Josef Fischer

---

Berlin 2009

Herausgegeben von der „Gesellschaft für Interlinguistik e.V.“ (GIL)

Otto-Nagel-Str. 110, DE-12683 Berlin

Tel.: +49-30-54 12 633, Fax : +49-30-54 98 16 38

[gil@blanke-info.de](mailto:gil@blanke-info.de)

[www.interlinguistik-gil.de](http://www.interlinguistik-gil.de)

© bei den Autoren der Beiträge

ISSN: 1432-3567

---



**Esperanto und andere Sprachen  
im Vergleich**

**Beiträge  
der 18. Jahrestagung  
der Gesellschaft für Interlinguistik e.V.,  
21.-23. November 2008, in Berlin**

**Herausgegeben von Sabine Fiedler**

**Berlin  
2009**

100

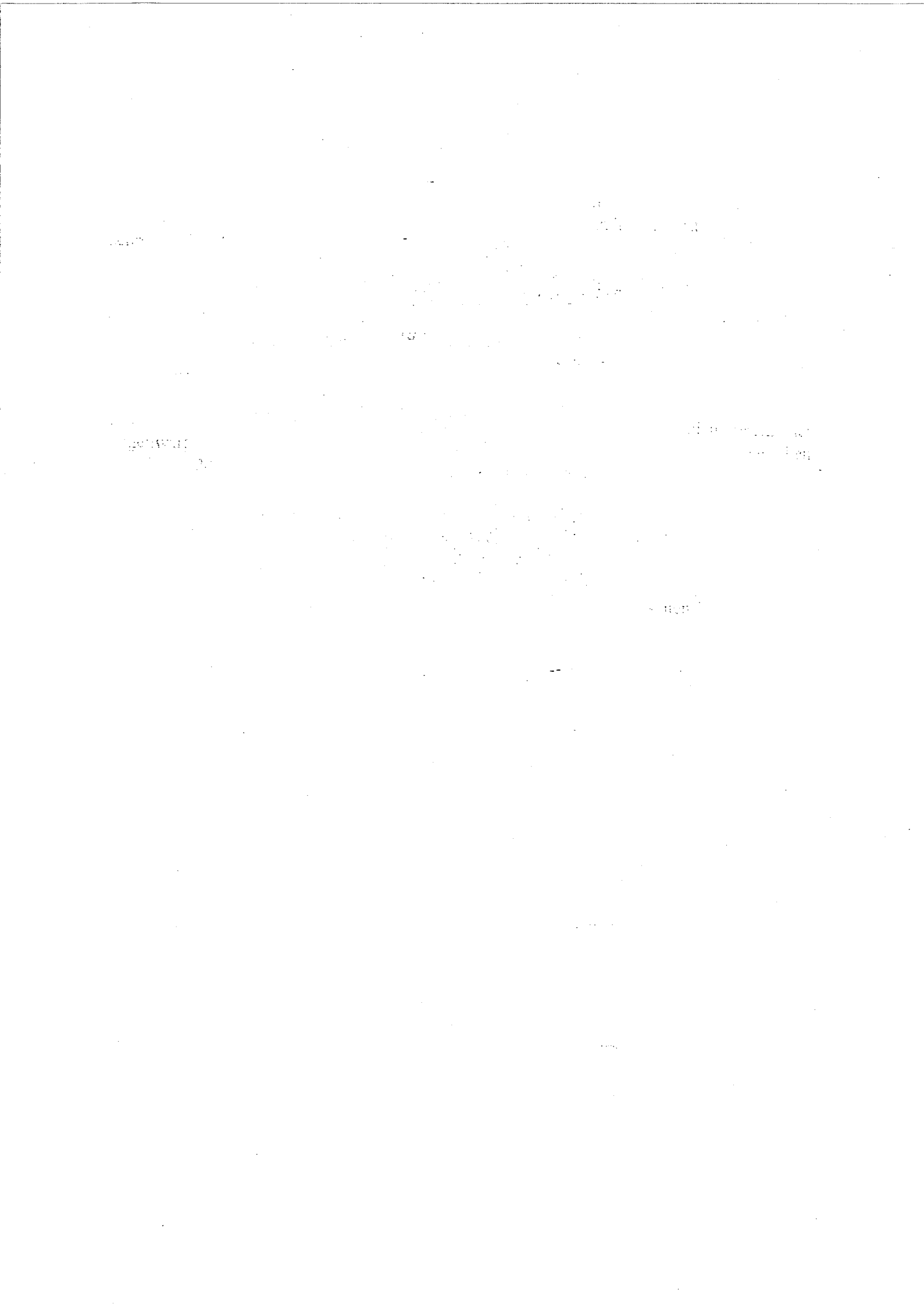
100  
100  
100

100  
100  
100

100

## Inhalt

<i>Sabine Fiedler</i>	Vorwort	7
<i>KIMURA Goro Christoph</i>	Esperanto als Minderheitensprache: Eine sprachsoziologische Betrachtung	11
<i>Otto Back</i>	Wie unvollkommen dürfen ethnische Sprachen und Plansprachen sein?	25
<i>Wim Jansen</i>	Baskisch und Esperanto: ein Strukturvergleich	33
<i>Biljana Golubović</i>	Die Phraseologie des Esperanto und des Serbischen im Vergleich	55
<i>Rudolf-Josef Fischer</i>	Der Struwwelpeter in Esperanto, Latein, Englisch Französisch, Italienisch und Spanisch: eine vergleichende Studie	65
<i>Cyril Brosch</i>	Komposition und Derivation in altindogermanischen Sprachen und im Esperanto	83
<i>Věra Barandovská-Frank</i>	Korrelativa in Esperanto und in einigen indogermanischen Sprachen	99
<i>Ilona Koutrný</i>	Esperanto im Rahmen der Sprachtypologie	117
<i>Marek Blahuš</i>	Rechtschreibprüfung für Esperanto und andere Sprachen	131
<i>Bengt-Arne Wickström</i>	Ökonomie und Sprache	137
<i>Cornelia Mannewitz</i>	Sprachplanung im Internet: Das Projekt Slovio	157
<i>Sabine Fiedler</i>	Neuere Fachliteratur zu interlinguistischen Themen (2007/2008)	165
Autoren		171
Inhalt der Beihefte 1-16		173





## Vorwort

Dieser Band vereinigt Vorträge, die im November 2008 auf der 18. Tagung der Gesellschaft für Interlinguistik e. V. (GIL) in Berlin gehalten wurden. Das Schwerpunktthema „Esperanto und andere Sprachen im Vergleich“ regte zahlreiche Teilnehmer an, sich aus unterschiedlichster Perspektive vergleichend mit dem Gegenstand Plansprache zu beschäftigen. Dieser Band enthält daher sowohl Beiträge, die gesamte Plansprachensysteme, z.B. Esperanto, anderen Sprachen gegenüberstellen oder einzelne linguistische Kategorien in Ethno- und Plansprachen miteinander vergleichen, als auch solche, die sich dem Thema aus soziolinguistischer, übersetzungswissenschaftlicher oder auch sprachökonomischer Sicht nähern. Sie machen in ihrer Gesamtheit deutlich, aus wie viel verschiedenen Blickwinkeln die Beschäftigung mit Plansprachen fruchtbringend sein kann.

Es ist besonders erfreulich, dass die Tagung 2008 die Möglichkeit des Gedankenaustausches mit einer Reihe von Interlinguisten bot, die außerhalb Deutschlands arbeiten. So z.B. mit *Goro Christoph Kimura*, Soziolinguist an der Sophia-Universität in Tokio (und seit 1994 GIL-Mitglied!), der in seinem Beitrag Unterschiede und Gemeinsamkeiten zwischen Esperanto und Minderheitensprachen beleuchtet. Er kommt zu dem Ergebnis, dass „Esperanto einige grundlegende Charakteristika mit Minderheitensprachen zu teilen [scheint]“, wobei ihm vor allem die Parallelen zu Merkmalen der sorbischen Kultur auffallen. Kimura geht von der klassischen Definition der Esperanto-Sprachgemeinschaft nach Richard E. Wood (1979) als „a voluntary, non-ethnic and non-territorial speech community“ aus und stellt fest, dass diese drei Merkmale nicht exklusiv für Esperanto, sondern zumindest graduell auch für (andere) Minderheitensprachen gültig sind. Eine Zusammenarbeit zwischen Plansprachenforschung und sprachsoziologischer Forschung zu Minderheitensprachen, so die Schlussfolgerung des Autors, könnten für beide Bereiche gewinnbringend sein.

Sprachen als „unvollkommene Werkzeuge“ stehen im Mittelpunkt des Beitrags von *Otto Back* (Wien). Während dieser Aspekt bei der Beschreibung von Ethnosprachen kaum thematisiert wird, finden Sprachstrukturbewertungen in der Plansprachenwissenschaft starke Beachtung. Der Autor behandelt das Thema auf der Grundlage von umfangreichem Beispielmateriale zu Plan- und Ethnosprachen in den Bereichen Benennungslücken, Mehrdeutigkeiten, Blockaden in der Formenbildung, verwechselbarer Wortklang und unübersichtliche Reihung und gelangt u. a. zu der Schlussfolgerung, dass „beide Arten [Plan- wie Ethnosprachen] von Vollkommenheit weit entfernt sind“.

Der niederländische Baskologe *Wim Jansen*, Professor für Interlinguistik und Esperanto an der Universität Amsterdam, vergleicht in seinem Beitrag ausgewählte morphologische, syntaktische und lexikalisch-semantische Charakteristika im Baskischen und Esperanto miteinander. Es zeigen sich, wie zu erwarten, drastische Unterschiede in Syntax und Lexik; Übereinstimmungen gibt es dagegen im Vokalsystem sowie im morphologischen Bereich, hier insbesondere in der Agglutinierung.

Der Beitrag von *Biljana Golubović* ist ein gutes Beispiel dafür, dass Arbeiten zum Esperanto über die eigentliche Beschäftigung mit den Plansprachen hinaus wirken können. Die Autorin vergleicht die Phraseologie des Esperanto mit der des Serbischen und stellt eine Reihe von Parallelen, z.B. hinsichtlich der Entstehung phraseologischer Einheiten, fest. Diese bei einer Gegenüberstellung von Plansprache und Ethnosprache eher überraschenden Übereinstimmungen erklärt sie mit Besonderheiten der Entwicklung des Serbischen. Der Aufsatz bestätigt erneut, dass Künstlichkeit und Natürlichkeit von Sprachen nicht als Dichotomie, sondern eher als Extreme eines Kontinuum zu betrachten sind. Darüber hinaus vermittelt der Beitrag

interessante Einblicke in die schwierige Entwicklung der Kroatistik und Serbistik nach dem Zerfall Jugoslawiens bzw. des Serbokroatischen.

Übersetzungsvergleiche zwischen Plan- und Ethnosprache standen bereits mehrfach auf der Tagungsordnung von GIL-Konferenzen (vgl. z.B. Beihefte 11 und 14). *Rudolf-Josef Fischer* beschäftigt sich mit dem Struwelpeter und vergleicht dessen Übersetzungen ins Esperanto, Lateinische, Englische, Französische, Italienische und Spanische miteinander, wobei er sowohl formale Kriterien (wie Versmaß und Zeilenanzahl) als auch inhaltliche Aspekte behandelt. Während die Übersetzer in romanische Sprachen einer Reihe von Problemen mit den Jamben- und Trochäenketten des Originals begegnen, schneidet die englische Version nach seiner Einschätzung sehr gut ab, ebenso wie die neuere Fassung in Esperanto (1988).

Wie bekannt, hat der Begründer des Esperanto Zamenhof 1887 kein detailliert ausgearbeitetes Projekt, sondern nur eine Sprachskizze vorgelegt, die in der Folgezeit von den Nutzern weiter kreativ ausgebaut wurde. Es ist daher nicht verwunderlich, dass sprachliche Einzelfragen bis in die Gegenwart z.T. sehr kontrovers diskutiert werden. *Cyril Brosch* greift einige dieser Probleme im Bereich Wortbildung auf und vergleicht Wortbildungsmodelle altindogermanischer Sprachen mit denen des Esperanto. Die Plansprache verfügt über eine geringere Anzahl von Wortbildungsmodellen, die jedoch produktiver und regelmäßiger als jene in Ethnosprachen sind. Der Vergleich zeigt, dass das Esperanto mit seiner Kombination von produktiver Komposition und Derivation am ehesten dem Altindoarischen ähnelt, was den Autor am Ende seines Beitrags zu der provokatorischen Frage anregt, was der Indogermanist und bekennende Feind des Esperanto Karl Brugmann wohl dazu gesagt hätte.

Ein Vergleich zwischen Esperanto und indogermanischen Sprachen ist auch das Thema von *Věra Barandovská-Frank*, die sich in ihrem Beitrag der Zamenhof'schen Korrelativtabelle widmet. An dieser scheiden sich nach meiner Erfahrung nicht selten die Geister: Was einige Esperanto-Sprecher für einen besonderen Beweis für Zamenhofs Genialität halten, wird von anderen mit dem Argument einer der Sprachrealität fernen Künstlichkeit abgelehnt. Die Autorin sieht die Vorzüge der Tabelle in ihrer systematischen Übersichtlichkeit und belegt in ihrem Beitrag anschaulich, „dass die Korrelativität von Pronomina und Adverbien eine historisch alte Erscheinung ist, wie man in einigen indogermanischen Sprachen beobachten kann“.

*Ilona Koutny* von der Adam-Mickiewicz-Universität in Poznań betrachtet Esperanto aus sprachtypologischer Sicht. Sie bestimmt die Position der Plansprache vergleichend zu Ethnosprachen auf phonetischer, morphologischer, syntaktischer, semantischer und pragmatischer Ebene und untersucht, inwieweit sich linguistische Entwicklungen, wie z.B. Metaphorisierungsprozesse oder die Entstehung von Synonymen und Phraseologismen, auch im Esperanto vollziehen. Dabei wird auch die interessante Frage eines „sprachlichen Weltbildes des Esperanto“ berührt.

Der Beitrag von *Marek Blahuš* zeugt von den veränderten Kommunikationsbedingungen im medialen Zeitalter. Die Nutzung des Internets erleichtert internationale Kontakte in beachtlichem Maße und ist für eine diasporaähnliche Sprachgemeinschaft wie die des Esperanto von entscheidender Bedeutung. Neben zahlreichen Vorteilen bringt das Internet aber auch neue Anforderungen an die Plansprache mit sich, wenn es z.B. um die Sicherung der Sprachqualität von in großer Anzahl im Internet veröffentlichten Texten geht. Der Autor beschäftigt sich in seinem Beitrag mit Rechtschreibprüfprogrammen für verschiedene Sprachen und stellt damit gleichzeitig die Ergebnisse seiner 2008 verfassten Bachelor-Arbeit „A Spell Checker for Esperanto“ vor.

Wenn es um gleichberechtigte internationale sprachliche Kommunikation geht, sind dabei ganz unterschiedliche Aspekte einzubeziehen. Der Überblick über aktuelle interlinguistische Fachliteratur am Ende des Bandes zeigt, dass sprachenpolitische Arbeiten auch an ökonomischen Fragen nicht vorbei kommen. Die von F. Grin 2005 vorgelegte Studie, die besagt, dass Großbritannien aufgrund der gegenwärtigen Dominanz des Englischen einen finanziellen Vorteil von ca. 17 Milliarden Euro jährlich hat und die EU (einschließlich Großbritannien und Irland) beim Übergang zum „Szenario Esperanto“ insgesamt 25 Milliarden Euro jährlich sparen könnte, hat viele aufhorchen lassen. *Bengt-Arne Wickström*, Professor für Wirtschaftswissenschaft an der Humboldt-Universität zu Berlin, widmet sich in seinem Aufsatz dem Thema „Ökonomie und Sprache“. Er gibt einen Überblick über die Anwendung ökonomischer Analysen auf Probleme der Verbreitung von Sprachen und behandelt insbesondere Möglichkeiten dauerhafter Mehrsprachigkeit sowie Sprachrechte und die Auswirkungen von Sprachpolitik darauf.

Der Aufsatz von *Cornelia Mannewitz* beschäftigt sich mit Slovio, bei dem es sich ihrer Meinung nach um „das bestausgearbeitete Projekt einer slawischen Sprache im Internet“ handelt. Sie beschreibt die linguistischen Merkmale von Slovio, wobei der Frage nachgegangen wird, inwieweit es berechtigt ist, das Projekt als slawische Sprache zu bezeichnen, und erläutert seine Spezifik als Internetprojekt sowie seine Eignung als Kommunikationsmittel. In sprachlicher wie in außersprachlicher Hinsicht lässt Slovio eine Reihe von Parallelen zum Esperanto erkennen.

Die Gesellschaft für Interlinguistik hat sich mit ihren regelmäßigen Tagungen und Publikationen in den letzten Jahren zu einem wichtigen Zentrum der Diskussion interlinguistischer Fragen entwickelt. Um einen möglichst großen Teilnehmer- und Leserkreis zu erreichen, wurde dabei im Wesentlichen an der während der GIL-Gründung getroffenen Entscheidung festgehalten, Deutsch als Arbeitssprache zu nutzen. Für die Hälfte der Beiträger in diesem Band stellt diese nicht die Muttersprache dar, und die Möglichkeit, die Aufsätze in Esperanto zu verfassen, hätte für sie vermutlich eine große Erleichterung bedeutet. Mein abschließender Dank als Herausgeberin für die gute Zusammenarbeit gilt daher insbesondere diesen Autoren.

Leipzig, September 2009

Sabine Fiedler  
(2. Vorsitzende der GIL)

## Rechtschreibprüfung für Esperanto und andere Sprachen

### Gliederung

- 1 Einleitung
- 2 Prinzipien der Rechtschreibprüfung
- 3 Rechtschreibprüfung für Esperanto
- 4 Rechtschreibprüfung für andere Sprachen  
Literatur

### 1 Einleitung

Die Entwicklung der Computertechnologien und des Internets hat starke Auswirkungen auf unsere Welt. Häufig führt sie zu wichtigen Veränderungen auf verschiedenen Gebieten menschlicher Aktivitäten oder innerhalb der menschlichen Gemeinschaften. Und in einigen Fällen wurde Instrumenten, die uns bereits vorher zur Verfügung standen, durch diese Technologien einfach völlig neues Potential verliehen.

Die Auswirkungen dieser neuen Technologien auf die Esperantogemeinschaft, eine Art Sprachdiaspora, kann als überwiegend positiv angesehen werden: Die Grenzen zwischen Ländern scheinen zu verschwinden, geographische Abstände verlieren ihren nachteiligen Charakter, es gibt mehr Gelegenheiten grenzüberschreitende Kontakte zu pflegen. Esperantosprecher und -kursteilnehmer sind in der Lage, auf einfache Art einander zu begegnen, Dokumente, Musik, Literatur und andere mit der Sprache verbundene Produkte sind einfacher zu finden. Nie zuvor war es möglich, ein so breites Publikum gleichzeitig zu erreichen, in Esperanto wie auch in anderen Sprachen. Im Internet gibt es Milliarden Webseiten, darunter Hunderttausende in Esperanto, und einen Artikel für die Esperanto-Wikipedia oder für ein eigenes Online-Tagebuch zu schreiben erfordert keine großen Anstrengungen. Einen E-Mail-Bericht an eine Online-Diskussionsgruppe zu schicken ist viel einfacher als einen Leserbrief an eine Zeitschrift zu senden oder selbst eine Postwurfsendung zu organisieren – und diese Vorteile werden genutzt.

Neben all diesen positiven Aspekten kommen jedoch auch negative Konsequenzen zum Vorschein. Die Sprachfertigkeiten eines durchschnittlichen Esperantisten haben sich in den letzten Jahrzehnten vermutlich nicht wesentlich verbessert. Es werden heute aber immer mehr Esperantotexte von durchschnittlichen Esperantisten produziert, wobei dem Sprachniveau wenig oder keine Aufmerksamkeit geschenkt wird. Das hat zur Folge, dass die Qualität der öffentlich zugänglichen Esperantotexte fällt.

Andererseits verursacht die Technik nicht nur das Problem; sie gibt uns auch die Mittel, um es lösen zu können. Im Jahr 2007 trat eine Gruppe junger Leute von E@I (Education@Internet, eine internationale Organisation, welche die Nutzung des Internets durch Esperantosprecher fördert) mit der Idee der Schaffung eines Softwarepakets zur Sprachprüfung von Esperanto-Texten hervor. Die Software soll sowohl für den Kursteilnehmer bestimmt sein, der noch seine Sprachfertigkeiten verbessern muss, als auch für den erfahrenen Esperantisten, der täglich mehrere Seiten Text produziert und dessen Fehler üblicherweise bloß Resultat von mangelnder Aufmerksamkeit sind. Eine gute Rechtschreib- und Grammatikprüfung ist

nützlich für jede Art von Textverarbeitung, E-Mail-Clients, Webbrowser und überall im Internet, wo Texte in Esperanto geschrieben werden (Foren, Chats, Blogs).

Im Rahmen meiner Bachelorarbeit „A Spell Checker for Esperanto“ (Blahuš 2008), die Teil des E@I-Projektes (Petrović-Lundberg 2007) ist, habe ich einen kurzen Überblick über Rechtschreibprüfungs-Software gegeben und den Prozess der Erarbeitung eines orthographischen Korrekturprogramms für Esperanto beschrieben, das in Form von Dateien für die *Hunspell* Rechtschreibprüfungs-Software implementiert wurde. Dieser Beitrag soll mit einigen interessanten Aspekten dieses Themas vertraut machen.

## 2 Prinzipien der Rechtschreibprüfung

Als **Rechtschreibprüfung** werden softwaregestützte Verfahren zur Fehlerkorrektur in elektronisch vorliegenden Texten bezeichnet. Ziel ihres Einsatzes sind die klassischen Officeprogramme, die bereits zuvor verwendeten Textverarbeitungsprogramme und Schreibmaschinen mit elektronischer Anzeige, die einen Druck nach Fertigstellung des gesamten Textes erlauben. Auch Datenbanken, Webeditoren, E-Mail-Clients und zahlreiche weitere Programme werden inzwischen mit integrierten Rechtschreibprüfungen ausgestattet.

Die Rechtschreibprüfung kann auf ihre eigene, bereits einige Jahrzehnte lange Geschichte verweisen. Am Anfang der Computerepoche wurde sie oft mithilfe einfacher Wortlisten realisiert, mit deren Hilfe der geschriebene Text verglichen wurde. Das Ausführen von komplexen Algorithmen war damals für die Computer zu anspruchsvoll. Doch die Nachteile dieses Verfahrens waren offensichtlich: Vor allem bei Sprachen, die viel Flexion enthalten, war es nicht möglich, alle potentiellen Wortformen aufzulisten, auch weil damalige Speicher über einen stark begrenzten Dateiraum verfügten. Und, auf jeden Fall, ist es mit diesem Verfahren nicht möglich Komposita richtig zu behandeln – die Wortliste würde dadurch nämlich fast unendlich. Darum musste man am Anfang Kompromisse suchen. Durch die spätere Zunahme der Computerleistung wurde schließlich die Entwicklung komplexer Rechtschreibprüfungs-Algorithmen ermöglicht.

Eines der ältesten Rechtschreibprüfprogramme mit einem Wortbildungs-Algorithmus ist *Ispell*, heute als *International Ispell* bekannt, das bereits im Jahr 1971 als Standardsoftware zur Rechtschreibprüfung unter dem Betriebssystem *Unix* entstanden ist. Aus *Ispell* wurde in den folgenden Jahren neue Software abgeleitet; die Entwicklung beginnt mit dem freier benutzbaren *GNU Aspell* und wird durch den für OpenOffice.org geschaffenen *MySpell* fortgesetzt, bis hin zum heutigen hoch entwickelten *Hunspell*. In jeder Version wurden gleichzeitig einige neue Verbesserungen eingeführt. Einen Vergleich von 6 Rechtschreibprüfprogrammen und ihrer 12 Eigenschaften habe ich in meiner Bachelorarbeit vorgenommen.

**Hunspell** (Németh 2008), das am weitesten entwickelte Produkt dieser Reihe, ist eine Rechtschreibprüfungs-Software und ein morphologischer Analysator, der für Sprachen mit reicher Morphologie und komplizierter Wortbildung oder Zeichenkodierung bestimmt ist. *Hunspell* wurde 2005 von László Németh aus Ungarn ins Leben gerufen und verdankt seinen Namen dem Umstand, der ungarischen Morphologie besser gerecht werden zu können, welche am Anfang des Programms stand. Seit dem 8. März 2006 wird *Hunspell* als Teil des Office-Pakets OpenOffice.org benutzt, und am 17. Juni 2008 wurde es auch zur Standard-Rechtschreibprüfung-Software des populären Webbrowsers Mozilla Firefox.

Die Informationen, die von *Hunspell* und derartigen Rechtschreibprüfprogrammen genutzt werden, um Texte in einer bestimmten Sprache kontrollieren zu können, werden gewöhnlich

in zwei Dateien gespeichert: In einer Wortliste und in einer Affix-Datei. Diese beiden werden zusammen als *Wörterbuch* zur Rechtschreibprüfung der betreffenden Sprache bezeichnet. Die Wortliste ist eine einfache Zusammenstellung von Wörtern und Wortteilen, von denen jedes auf einer eigenen Zeile angeordnet und eventuell durch Markierungen ergänzt wird, welche die reziproke Kombinierbarkeit dieser Elemente bestimmen und begrenzen. Die Affix-Datei beschreibt an ihrem Anfang verschiedene spezifische Eigenschaften des Wörterbuchs (wie Zeichenkodierung oder häufige Zeichenverwechslungen), aber sie enthält vor allem eine Liste von Affixen (Präfixen und Suffixen), die den einzelnen Markierungen aus der Wortliste entsprechen und bestimmen, welche Affixe sich mit welchen Wörtern verbinden können, um ein neues Wort zu bilden. Jedes Affix wird durch eine Serie von Regeln beschrieben, die sein Verhalten eindeutig bestimmen.

Ein Beispiel für ein aus mehreren Regeln bestehendes Affix ist die Bildung des Präteritums im Englischen. Falls alle Verbstämme in der Wortliste durch die Markierung „D“ bezeichnet werden, kann die Bildung von regelmäßigen Präterita durch ein Affix „D“ mit den folgenden vier Regeln beschrieben werden:

SFX D 0 d e	Beispiel: breathe <u> </u> > breathe <u>d</u>
SFX D y ied [ <sup>^</sup> aeiou]y	Beispiel: <u>cr</u> y > cr > <u>cri</u> ed
SFX D 0 ed [ <sup>^</sup> ey]	Beispiel: want <u> </u> > want <u>ed</u>
SFX D 0 ed [aeiou]y	Beispiel: play <u> </u> > play <u>ed</u>

Nach einer Abkürzung für „Suffix D“ folgen die Buchstaben, die durch die Regel vom Ende des Wortes abgetrennt werden (oder „0“, wenn das ganze Wort stehen bleibt), danach die Buchstaben, die als Resultat der Regel an das Ende des Wortes angefügt werden, und schließlich die Festlegung der Situationen, in denen die Regel benutzt werden kann, in Form der letzten Buchstaben des ursprünglichen Wortes, die vor Anwendung der Regel kontrolliert werden (die eckigen Klammern bezeichnen eine Buchstabenmenge, der Zirkumflex bedeutet eine Komplementärmenge zu der gegebenen).

Das Rechtschreibprüf-Programm analysiert den Eingangstext Wort für Wort und versucht, das Wort entweder in der Wortliste zu finden oder es mithilfe dieser Regeln aus dort stehenden Wörtern zusammensetzen. Und falls ein Wort auf keine dieser Arten erkannt wird, wird es als Rechtschreibfehler bezeichnet und es wird ein Versuch unternommen, mögliche Korrekturen vorzuschlagen. Dazu wird häufig die Levenshtein-Distanz (auch Editierdistanz) benutzt. Das ist eine Funktion aus der Informationstheorie, die den Unterschied zwischen zwei Zeichenketten als minimale Anzahl der Operationen Einfügen, Löschen und Ersetzen bezeichnet. Zum Beispiel für ein fehlerhaftes Wort „Libe“ würden im Wörterbuch zwei ähnliche Wörter gefunden – „Liebe“ und „Leib“ – von denen schließlich „Liebe“ als der beste Korrekturvorschlag gewählt würde, wegen der Einbuchstabendistanz zu „Libe“ (1 Auslassung), im Vergleich mit „Leib“, bei dem die Distanz größer ist (2 Vertauschungen).

### 3 Rechtschreibprüfung für Esperanto

Eines der ersten Rechtschreibprüf-Programme, das für Esperanto entwickelt wurde, war vermutlich **Kontrolu Literumadon**, das bereits 1992 von Klivo Lendon aus Kanada geschaffen wurde. Das Programm arbeitete unter dem Betriebssystem MS-DOS und benutzte eine pseudografische Benutzeroberfläche, um den Eingangstext zu zeigen und die unerkannten Wörter in Farbe zu markieren. Es bot jedoch keine Korrekturvorschläge und auch keine Möglichkeit, den Text auf demselben Schirm zu redigieren.

Zurzeit gibt es bereits viele Programme, die eine Rechtschreibprüfung von Esperanto-Texten anbieten können – z.B. Esperantilo, OpenOffice.org, Mozilla Firefox – aber sie alle benutzen eine einzige Rechtschreibprüfungs-Software – nämlich ein Wörterbuch, das von Sergio Pokrovskij (Pokrovskij circa 1997) geschaffen und seitdem kaum verbessert wurde; es war ursprünglich bestimmt für *Ispell* und wurde später von Dmitri Gabinski für *MySpell* adaptiert (Gabinski & Pokrovskij 2003).

Dieses Wörterbuch enthält in seiner letzten Version eine Wortliste von 19.342 Einträgen und hat einen Gesamtumfang von 58 Affixen (darunter 34 Präfixe und 24 Suffixe), die durch insgesamt 2.426 Affixregeln beschrieben werden. Doch zu den Nachteilen des Wörterbuches gehört neben der fast völlig fehlenden Dokumentation auch allein der Fakt, dass die neuen Eigenschaften von *Hunspell* überhaupt nicht genutzt werden. *Hunspell* bietet u. a. solche Vorteile an wie Unicode zur Endlösung des Problems von Zeichenkodierungen und ein ausgearbeitetes System von Wortbildungsregeln, das umfassender ist als die gewöhnlichen Affixregeln. Da eine letzte Revision des Wörterbuches nach dem Einzug von *Hunspell* noch fehlte, habe ich eine erneuerte Version für notwendig erachtet. Die letzten Veränderungen von Gabinski haben wohl einige Dinge verbessert (z.B. dass alle Zahlen jetzt als Teile von Wörtern auftreten können – „trikapa“ oder „dekseskapa“, erkannt durch Gabinski, verglichen mit lediglich „unukapa“ und „dukapa“, erkannt durch Pokrovskij), aber viele Beschränkungen der alten Versionen bleiben (z.B. dass Kombinationen von Suffixen als ein Suffix angesehen werden müssen – so findet man z.B. „-anino“ als ein selbständiges Suffix, obwohl es mit umfassenderen Regeln leicht als Kombination der Basissuffixe *-an*, *-in* und *-o* betrachtet werden könnte).

In meiner kürzlich erschienenen Bachelorarbeit habe ich einen Versuch unternommen, eine neue Rechtschreibprüfungs-Software vorzuschlagen und zu implementieren, die sich die neuen Fähigkeiten von *Hunspell* zunutze macht und gleichzeitig ein neues, stärker auf der Morphologie von Esperanto aufbauendes Verfahren anwendet. So ist als Resultat meiner Arbeit ein neues Esperanto-Wörterbuch für *Hunspell* entstanden, das den Strukturen von real existierenden Komposita folgt und diesen Wortbildungsprozess durch eine semantische Analyse aller Wortstämme unterstützt. Als Ausgangspunkt habe ich den Wortschatz des *Plena Ilustrita Vortaro (PIV)* (2005) benutzt (Grimley Evans 2005). Dessen sämtliche Wortstämme habe ich semantisch klassifiziert durch Zuweisung einer oder mehrerer von insgesamt 15 semantischen Markierungen, die verschiedenen feststellbaren Eigenschaften der Wortstämme entsprechen. Die Anregung dafür verdanke ich auch Wennergren (2005), der in seiner Grammatik *PMEG* (im Kapitel 37.1) das Thema semantische Klassifikation von Wortstämmen diskutiert, aber nur einen unvollständigen Vorschlag dazu gibt, um die Idee zu illustrieren. In meiner Arbeit habe ich diese Idee weiter ausgearbeitet, erst durch Analyse der offiziellen Esperanto-Suffixe und Suche nach den brauchbaren semantischen Kategorien, dann durch eine gründliche Gliederung aller Wortstämmen im Rahmen der vorgeschlagenen Klassifikation (mithilfe verschiedener Fachwörterbücher, Wortlisten und Korpora). Schließlich habe ich die Wortbildung in den im *PIV* vorkommenden Komposita – verarbeitet von Witkam (2006) – untersucht und daraus neue Wortbildungsregeln abgeleitet, welche die neue Fähigkeit von *Hunspell* nutzen, Regeln in Form regulärer Ausdrücke zu formulieren. Die daraus resultierende Rechtschreibprüfungs-Software arbeitet mit 15.914 Einträgen in der Wortliste und gebraucht 15 + 81 semantisch unterschiedene Affixe mit insgesamt 37.155 Affixregeln. Insgesamt ist die Software in der Lage, ungefähr  $2,8 \times 10^{18}$  verschiedene richtige Wortformen im Esperanto zu erkennen.

**Zum Vergleichen und Beurteilen** von Rechtschreibprüfungs-Software werden zwei wichtige mathematische Angaben verwendet, nämlich *Recall* und *Precision*. Der *Recall* ist der prozentuale Anteil der gefundenen Rechtschreibfehler unter allen Rechtschreibfehlern, die im

untersuchten Text vorhanden sind. *Precision* ist der Anteil wirklicher Rechtschreibfehler unter allen Fällen, die als Rechtschreibfehler durch die Software bezeichnet wurden. Der Idealfall ist ein hoher *Recall* und eine hohe *Precision*. Im Vergleich mit dem Werk von Pokrovskij zeigte meine Arbeit in Tests einen höheren *Recall*, aber eine geringere *Precision*. Das kann leicht durch das viel umfassendere Wortbildungssystem in meiner Arbeit erklärt werden, das für den hohen *Recall* verantwortlich ist, das jedoch auch fehlerhafte Identifizierungen von falsch geschriebenen Wörter als andere and häufig nicht sehr wahrscheinliche Stammkombinationen verursacht (z.B. wird das Wort „cirkonstanco“ – deutsch „Umstand“ – falsch geschrieben als „cirkonstanto“ als „cir-konstanto“ – also eine „Schuhcremekonstante“ – erkannt).

#### 4 Rechtschreibprüfung für andere Sprachen

Die ersten Rechtschreibprüfprogramme sollten ausschließlich den Ansprüchen des Englischen genügen, denn in der Zeit ihres Entstehens war die Computerwelt noch stark mit dieser Sprache verbunden. Darum gebrauchte *Ispell* allein die Zeichenkodierung ASCII, die nur für westeuropäische Sprachen geeignet ist. Und auch darum war eine der wichtigsten Neuerungen von *GNU Aspell*, seinem Nachfolger, ein Algorithmus für phonetisch gegründete Korrekturvorschläge – eine Sache, die für viele andere Sprachen mit phonetischer Rechtschreibung ein Problem darstellt.

Jüngste Entwicklungen in der Rechtschreibprüfung werden aber stark vom Bedarf an anderssprachigen Prüfprogrammen beeinflusst. Ein Höhepunkt ist zurzeit die oben erwähnte Software *Hunspell*, die ursprünglich als eine verbesserte Version von *MySpell*, mit zusätzlichen Fähigkeiten zur Verarbeitung von morphologisch reichen Sprachen, wie Ungarisch, entstanden ist. So sind Eigenschaften, die früher nur für das Englische zutreffend gewesen sind, jetzt auch für den Bedarf anderer Sprachen generalisiert worden, und neue Eigenschaften wurden hinzugefügt.

Die Generalisierung und Ausarbeitung des Wortbildungssystems in *Hunspell* hat viele verschiedene Folgen für alle möglichen Sprachen gehabt, die nicht ausschließlich nur mit dem Schaffen von Komposita verbunden sind. Das Zubehör von *Hunspell* enthält auch 84 kleine Testwörterbuchdateien, die als Antworten auf viele Aufforderungen der Rechtschreibprüfung für verschiedene Sprachen dienen. Dadurch werden z.B. die folgenden Problemen berührt:

**Im Deutschen:** Die Großschreibung von ß als SS: „Weiß“ wird zu „WEISS“.

**Im Ungarischen:** Der Superlativ der durch ein Zirkumfix von „leg-“ (Präfix) und „-abb“, „-ebb“ oder „-bb“ (Suffix) entsteht: *Aus der Basisform „magas“ wird der Komparativ „magasabb“ und der Superlativ „legmagasabb“ abgeleitet, aber „legmagas“ ist kein korrektes Wort.*

**Im Finnischen und Schwedischen:** Ein Doppelpunkt kann in dieser Sprachen als Apostrophe benutzt werden: *Im Schwedischen kann „Sankt Petersburg“ als „S:t Petersburg“ geschrieben werden. Im Finnischen ist „USA:han“ der Illativ von „USA“.*

**Im Englischen:** Die fortlaufenden Zahlen haben verschiedene Endungen: *Die fortlaufende Form von 1 ist „1st“, von 2 ist „2nd“, von 3 ist „3rd“, von 4 ist „4th“.*

**Im Deutschen:** Zahlen werden manchmal als Teile von Wörtern benutzt: *Ein System für das Schaffen von korrekten Wörtern wie „1-jährig“ oder „24-stündig“ muss entwickelt werden.*



**Im Deutschen:** Viele Aspekte der Komposition müssen berücksichtigt werden: *Der erste Buchstabe des zweiten Teils eines Kompositums muss verkleinert werden („Computer“ + „Arbeit“ = „Computerarbeit“), einige Komposita brauchen ein „s“ zwischen den Teilen („Arbeit“ + „Computer“ = „Arbeitscomputer“), einige können auch einen Bindestrich gebrauchen („Arbeit“ + „Computer“ = „Arbeits-Computer“), einige Komposita sind verbotene Ausnahmen („Arbeitsnehmer“ ist nicht korrekt), Präfixe sind nur am Anfang von Komposita erlaubt („Prä-“).*

Dies und vieles andere mehr sind Anforderungen, denen die heutige Rechtschreibprüfung gerecht werden muss. Einigen kann bereits heute entsprochen werden, für andere sind jedoch die Algorithmen noch nicht stark genug, oder es ist überhaupt schwierig, eine Lösung für das Problem vorzuschlagen. Auch die Leistung von Rechtschreibprüfungen, durch *Recall* und *Precision* gemessen, kann immer noch verbessert werden. Eine perfekte Rechtschreibprüfung wird es aber vermutlich nie geben, weil manchmal der Mensch selbst nicht weiß, was er eigentlich schreiben wollte.

## Literatur

Blahuš, Marek (2008): A Spell Checker for Esperanto. Brno: Masaryk-Universität, Fakultät für Informatik, 40 S. Nachzulesen im Internet unter: [https://is.muni.cz/auth/th/172464/fi\\_b/](https://is.muni.cz/auth/th/172464/fi_b/).

Gabinski, Dmitri & Pokrovskij, Sergio (2003): Esperanta literumilo por OpenOffice.org. Nachzulesen im Internet unter: <http://www.esperanto.pisem.net/literumilo.html>.

Grimley Evans, Edmund (2005): Kapvortoj de PIV. Versio 1.4. Nachzulesen im Internet unter: <http://web.archive.org/web/20061012192331/rano.org/pivkap>.

Németh, László (2008): Hunspell : open source spell checking, stemming, morphological analysis and generation under GPL, LGPL or MPL licenses. Budapest : SourceForge.net. Nachzulesen im Internet unter: <http://hunspell.sourceforge.net/>.

Petrović-Lundberg, Sonja (2007): Lingvohelpilo : Projektpropono kun buĝeto. Stockholm.

Pokrovskij, Sergio (circa 1997): Vortaro por ISpell. Nachzulesen im Internet unter: <http://www.esperanto.mv.ru/Download/ispell/ispelleo.tgz>.

Wennergren, Bertilo (2005): Plena manlibro de Esperanta gramatiko. El Cerrito : ELNA. 696 S. Nachzulesen im Internet unter: <http://bertilow.com/pmeg/>.

Wikipedia-Autoren (2008): Rechtschreibprüfung. Wikipedia, Die freie Enzyklopädie. Nachzulesen im Internet unter: <http://de.wikipedia.org/wiki/Rechtschreibprüfung>.

Witkam, Toon (2006): Automatische Morphemanalyse in Esperanto macht Komposita besser lesbar auf dem Bildschirm. In Blanke, Detlev (2007, Hrsg.). Esperanto heute – Wie aus einem Projekt eine Sprache wurde : Beiträge der 16. Jahrestagung der Gesellschaft für Interlinguistik e.V., 1.-3. Dezember 2006 in Berlin. Berlin: GIL.

## Autoren

*Otto Back* (Laudongasse 20/11, A-1080 Wien), Dr. phil., vormals Honorarprofessor für Graphematik und Orthographieforschung an der Universität Wien.

*Věra Barandovská-Frank* (Kleinenberger Weg 16, D-33100 Paderborn, vera.barandovska@uni-paderborn.de), Dr., Latinistin, PDoc. der AIS San Marino. Redakteurin der Zeitschrift „Grundlagenstudien aus Kybernetik und Geisteswissenschaft“.

*Marek Blahuš* (Rudy Kubíčka 1002, CZ-68605 Uherské Hradiště, marek@blahus.cz), Bc., ist Student der Angewandten Informatik (Spezialisierung Computerlinguistik) an der Masaryk-Universität in Brno, Tschechien. Er ist außerdem Generalsekretär von E@I und Vorsitzender der Tschechischen Esperanto-Jugend.

*Cyril Brosch*, M.A. (Kaiser-Friedrich-Str. 60, 10627 Berlin; info@cyrilbrosch.net), ist Doktorand im Fach Altorientalistik an der Freien Universität Berlin.

*Sabine Fiedler* (Am Zollamt 5, 04838 Jesewitz OT Gordemitz, sfiedler@uni-leipzig.de), Prof. Dr. phil. habil., Sprachwissenschaftlerin am Institut für Anglistik der Universität Leipzig.

*Rudolf-Josef Fischer* (Gustav-Adolf-Str. 2a, 48356 Nordwalde, fischru@uni-muenster.de), Dipl.-Mathematiker, Dr. rer. medic., Dr. phil., M.A., Privatdozent am Inst. f. Med. Informatik und Biomathematik der Universität Münster; Lehrbeauftragter für Esperanto am Institut für Allgemeine Sprachwissenschaft der Universität Münster.

*Biljana Golubović* (Gebhardstr. 29, D-76135 Karlsruhe, golubovic@t-online.de), Dr. phil., Lektorin an der Eberhard-Karls-Universität in Tübingen.

*Wim Jansen* (Emmaplein 17A, NL-2225 BK Katwijk, wimjansen@casema.nl), Diplomingenieur (Luft und Raumfahrttechnik), Dr. phil. (Baskologe; Esperantologe), Professor für Interlinguistik und Esperanto an der Universität Amsterdam.

*Kimura Goro Christoph* (Nogaya-mati 619-42, Matida-si, 195-0053 Tokio, Japan, gkimura@sophia.ac.jp), Dr. phil., Soziolinguist, außerordentlicher Professor an der Sophia-Universität in Tokio.

*Ilona Koutny* (Pobiedziska 14 / 22, PL-61-052 Poznań, ikoutny@amu.edu.pl), Dr. habil., ungarische Sprachwissenschaftlerin (Hungarologie und Esperantologie) an der Adam-Mickiewicz-Universität in Poznań (Polen) und Leiterin der dortigen Interlinguistischen Studien, Mitglied der Akademie des Esperanto, Dozentin von AIS.

*Cornelia Mannewitz* (Parkstr. 26, 18059 Rostock, cornelia.mannewitz@gmx.de), Dr. phil. habil., Slawistin an der Universität Greifswald.

*Bengt-Arne Wickström* (Weseler Str. 17, 10318 Berlin, wickstr@wiwi.hu-berlin.de), Ph.D., ist Professor für Wirtschaftswissenschaft an der Humboldt-Universität zu Berlin sowie an der Akademio Internacia de la Sciencoj (AIS) San Marino.