

# **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**

**Herausgegeben von Detlev Blanke**

**Berlin  
2007**

## **Inhalt**

<i>Detlev Blanke</i>	Vorwort.....	7
<i>Andreas Künzli</i>	Plansprachenprojekte ersten, zweiten und dritten Grades in der Schweiz. Aus welchen Projekten Sprachen wurden und was ephemer blieb.....	9
<i>Wim Jansen</i>	Wortstellungsmodelle im ursprünglichen und im heutigen Esperanto – Zeugen einer natürlichen Entwicklung der Syntax?.....	15
<i>Sabine Fiedler</i>	<i>Alice's Adventures in Wonderland</i> im Deutschen und Esperanto – Ein intra- und interlingualer Übersetzungsvergleich.....	27
<i>Toon Witkam</i>	Automatische Morphemanalyse in Esperanto macht Komposita besser lesbar auf dem Bildschirm.....	51
<i>Gunnar Fischer</i>	Esperanto-Musik – Teil der Kultur der Esperanto-Sprachgemeinschaft.....	59
<i>Fritz Wollenberg</i>	100 Jahre Esperanto in Berlin: Historiografische und interlinguistische Fakten in einer neuen Veröffentlichung.....	71
<i>Ulrich Lins</i>	Der Spanische Bürgerkrieg und das Esperanto.....	81
<i>Sebastian Kirf</i>	Über das Verhältnis zwischen Agenda 21 und Esperanto.....	83
<i>Rudolf-Josef Fischer</i>	Das Projekt KOD – ein Bericht.....	91
Anhang:	Svisa Enciklopedio Planlingva: Enhavo.....	105
Autoren.....		115
Inhalt der Beihefte 1-13.....		116

## Vorwort

Im Jahre 1887 erschien in Warschau eine kleine unscheinbare Broschüre von 40 S. in russischer Sprache. Sie enthielt vor allem eine Argumentation für eine internationale Sprache und wie diese nach Meinung des Autors aussehen könnte. Nur wenige Seiten waren der Skizze seines Sprachprojekts gewidmet, nämlich 4 S. Beispieltex-te, 6 S. Grammatik sowie ein Anhang von 900 Wörtern. Das war alles, was zu Wortschatz und Grammatik des Projekts der „Meždunarodnyj jazyk“, der „Internationalen Sprache“, in der Broschüre zu finden war. Im gleichen Jahr erschien diese in Deutsch, Französisch, Polnisch sowie 1888 in Englisch. Es folgten weitere Materialien. Doch die Broschüre kann man als Keimzelle einer neuen Sprache betrachten. Das war durchaus nicht selbstverständlich, wie Hunderte vergebliche Versuche in gleicher Richtung beweisen. Die besonderen Faktoren und Bedingungen, die den Erfolg der Sprache des Dr. Esperanto – so das Pseudonym des Augenarztes Dr. Ludwig L. Zamenhof aus Warschau – erklären, können hier nicht dargestellt werden. Aber als Resultat: Aus einem Projekt wurde eine Sprache. Ein Phänomen, das jeden Linguisten faszinieren müsste. Wir wissen, dass dem nicht so ist...

Die vorliegenden Akten der 16. GIL-Tagung, die 120 Jahre nach Veröffentlichung der Keimzelle erscheinen, illustrieren in unterschiedlicher Weise, dass wir es im Esperanto mit einer *Sprache* zu tun haben und nicht mit einem Projekt, einer Sprache, die über Anwender verfügt, also über eine Sprachgemeinschaft.

Der Schweizer Slawist und Interlinguist *Andreas Künzli*, der erst kürzlich eine ungewöhnliche Enzyklopädie über Plansprachen in seinem Land vorgelegt hat, ordnet das Esperanto in die Projektlandschaft der Schweiz ein.

Der Niederländer *Wim Jansen*, nicht nur Spezialist für Raumfahrttechnik, sondern auch einer der wenigen niederländischen Baskologen, außerdem Dozent für Interlinguistik und Esperanto an der Universität Amsterdam, vermittelt einen Einblick in seine soeben verteidigte Dissertation über die Satzgliedfolge des Esperanto. Seine Analyse realer Texte und die Ergebnisse von Experimenten mit Testpersonen zeigen sowohl Stabilität als auch Sprachwandel in der *Plansprache*.

Eine enorme Herausforderung für jeden Übersetzer, ganz gleich in welche Sprache er überträgt, ist der von Wortspielen und Andeutungen durchsetzte Text des bekannten Kinderbuches „Alice’s Adventures in Wonderland“ von Lewis Carroll. Die Anglistin *Sabine Fiedler* aus Leipzig, die ihre Habilitationsarbeit über die Rolle der Phraseologie in Esperanto verfasst hat und erst kürzlich mit einer Arbeit über die Phraseologie des Englischen hervorgetreten ist, vergleicht nicht nur drei deutsche Übersetzungen des englischen Textes untereinander, sondern diese wiederum mit zwei Übersetzungen von Alices Abenteuern in Esperanto. Ihr Ergebnis: Die Plansprache ist der Aufgabe durchaus gewachsen, mit allen Einschränkungen bei einem solchen Unterfangen, die in unterschiedlicher Weise für alle Sprachen gelten. Wer könnte da noch von einem Projekt oder einer künstlichen Sprache reden?

Eine ständig zunehmende Verwendung findet das Esperanto im Internet. Es entstehen große Textmengen, auch mit Fehlern, wie üblich in der oft (zu) schnellen elektronischen Kommunikation. Zwar gab es immer mal Versuche, Textverarbeitungsprogramme in Esperanto zu schreiben (auf Word-Basis), doch gibt es bisher kein befriedigendes Programm, das die plansprachlichen Texte auf ihre orthographische Richtigkeit überprüft. Der niederländische Software-Experte, Ingenieur und Computerlinguist, *Toon Witkam* beschreibt Möglichkeiten und Probleme eines solchen Programms zur Prüfung von Texten.

Für Außenstehende oft schwer vorstellbar: Die Sprachgemeinschaft verfügt inzwischen über eigene Traditionen, über Elemente einer spezifischen Kultur, was *Gunnar Fischer*, der Esperanto als Muttersprache spricht, durch die Darstellung der Entwicklung des Musiklebens beweist.

## Automatische Morphemanalyse in Esperanto macht Komposita besser lesbar auf dem Bildschirm

### Gliederung

- 1 Spellcheck und Komposita
- 2 Ausgezeichnete Glieder
- 3 Trennzeichen als Lese-Hilfe
- 4 Komposita: Art und Abgrenzung
- 5 Verwendung des Wörterbuches
- 6 Die REIFLER-These und sonstige Regeln
- 7 Hülsenwörter machen es schwierig
- 8 Schluss und Vorausblick

### 1 Spellcheck und Komposita

Der Computer als Textverarbeitungsmaschine ist allmählich Teil unseres Lebens geworden. An die Verfügbarkeit automatischer Textkontrolle, mit dem sogenannten Spellingchecker, haben wir uns mittlerweile gewöhnt. Sogar für Esperanto hat sich solche Software angeboten: ‘*ĈAPELILLO*’ vor einigen Jahren, und neuerdings u.a. ‘*SIMREDO*’. Was dabei auffiel war, dass durch die Software-Hersteller gewarnt wurde vor dem möglichen Versagen ihrer Produkte bei der Behandlung von Wortzusammensetzungen. Bei dem einen hieß es “der Computer kann Sinn und Unsinn nicht unterscheiden”, und als Beispiel wurde die Zerlegung *vi-rok-at-o* neben *vir-o-kat-o* aufgeführt<sup>1</sup>. Beim anderen wurde ganz einfach eine “Abneigung” der Software gegen Komposita-Erkennung erwähnt<sup>2</sup>.

Tatsächlich stellt die automatische Erkennung von Komposita eine große Herausforderung dar. Es handelt sich dabei ja manchmal um produktive Wortbildungen, Zusammensetzungen, die nicht im Wörterbuch stehen:

*flughavenoficisto, ostaĝprenulo, vivovojdirekto, amasdetruarmiloj.*

Eine Kontrolle der Rechtschreibung dieser Wörter erfordert die Zerlegung in ihre Bestandteile (mindestens in zwei), und gerade das ist der Kern des Problems. Was für nicht so weit fortgeschrittene Benutzer des Esperanto schon schwierig ist, auf Anhieb die sinnvollen Bausteine eines längeren Wortes zu erfassen, wie soll der ‘dumme’ Computer das zustande bringen?

Sprachstatistik schien auf den ersten Blick eine Möglichkeit zu bieten. Bereits vor fünfzig Jahren hat Kawasaki in einer esperantologischen Riesenarbeit<sup>3</sup> für alle Konsonantenpaare ihre Wahrscheinlichkeit in Wortwurzeln berechnet. Ein Konsonantenpaar, das innerhalb von Wurzeln

---

<sup>1</sup> Siehe : Simono Pejno, “Dividaj dilemoj” (das Dilemma der Trennung) [ <http://come.to/simono> => ‘Pejno Simono’ => ‘Novaĵoj en Ĉapelilo 2.1’ => ‘E-dividilo’].

<sup>2</sup> “Simredo 4 havas novan literuman algoritmon, kiu malemas rekoni kunmetaĵojn” (Simredo 4 hat einen neuen Buchstabieralgorithmus, der nicht geneigt ist, Zusammensetzungen zu erkennen). Siehe [ [www4.vc-net.ne.jp/~klivo/sim/literum.htm](http://www4.vc-net.ne.jp/~klivo/sim/literum.htm) ].

<sup>3</sup> Naokazu Kawasaki, “Radiktipoj laŭ aranĝo de vokaloj kaj konsonantoj” [Fremdlingva Universitato de Osaka, 1952].

als unwahrscheinlich gilt, hat also in einer Zusammensetzung eine große Chance, Schnittstelle zu sein.

Im Fall von *amasdetruarmiloj* ist *sd* ein solches Paar, und übrigens auch *ua*. Es zeigt sich aber, dass verfeinerte sprachstatistische Methoden aus der Vergangenheit heutzutage überholt werden durch die rohe Gewalt des Computers, der imstande ist riesige Vokabulare in Millisekunden zu durchsuchen. Dann macht es praktisch nichts aus, wenn sämtliche Zeichenpositionen eines unbekanntes Wortes als mögliche Schnittstellen betrachtet werden (außerhalb derjenigen am Rande des Wortes; bei *amasdetruarmiloj* ergibt das z.B. 11 Schnittstellen). Zwar muss dabei immerhin eine Strategie hinzugefügt werden, um eine Fehlanalyse (*vi-rok-at-o*) herauszufiltern, aber im ganzen kann in dieser Weise die automatische Analyse von Komposita überraschend verbessert werden. Ein erster Versuch zur Zerlegung von 1500 Zusammensetzungen aus der Esperanto-Ausgabe der *Le Monde diplomatique*<sup>4</sup> zeigte eine Erfolgsrate von 97%.

## 2 Ausgezeichnete Glieder

Gerade für Esperanto liegt es nahe, ein zu analysierendes Wort in Morpheme zu zergliedern. Mit einem Agglutinationsindex von 100% liegt diese Sprache ja unumstritten an der Spitze<sup>5</sup>. Die Unveränderlichkeit der Morpheme macht diese zu ausgezeichneten Gliedern – zu goldenen Bausteinen, die man ausnützen sollte bei der Entwicklung von Software für die Textkontrolle.

Ein Textwort das nicht im Wörterbuch steht, muss – falls es sich um Esperanto handelt – zumindest in Morpheme zergliedert werden können. Nehmen wir als Beispiel wieder das Wort *amasdetruarmiloj*. Eine exakte Zergliederung in Morpheme ist:

*am-as-de-tru-ar-mil-oj*

(Liebel|Spielkartel|von|Loch|Mengel|tausend|Substantiv-Plural)

aber ebenfalls:

*amas-detru-arm-il-oj*

(Masse|Vernichten|bewaffnen|Werkzeug|Substantiv-Plural)

sowie weitere ‘Mischungen’ zwischen diesen zwei (*amas-de-tru-ar-mil-oj*, *amas-de-tru-arm-il-oj*, *am-as-detru-ar-mil-oj*, usw).

Eine Zergliederung in Morpheme ist an sich noch kein Wundermittel, um die richtige Struktur eines Wortes, gemäß seiner zutreffenden Bedeutung in einem Text, herauszufinden. Sicher ist jedenfalls, dass die Lösung eine Morphemkette sein muss. Ist es dann nicht angebracht, bei Komposita auch die Morphemkette zu zeigen, auf der die automatische Textkontrolle ihre Begutachtung begründet hat?

---

<sup>4</sup> Siehe dazu: Toon Witkam, “La ekscito de vortstatistiko: Kiel krudforta kunmet-analizo kompletigas tekstkontrolon” [In: Blanke, Detlev (2007, Hrsg.): *Lingvaj kaj historiaj analizoj. Aktoj de la 28-a Esperantologia Konferenco en la 90-a Universala Kongreso de Esperanto, Vilno 2005*. Rotterdam: UEA, S. 23-46..<sup>5</sup> John Wells (1978) : “Lingvistikaj aspektoj de Esperanto”. Rotterdam. UEA.

Eine Textkontrolle wird niemals vollkommen sein! Erstens können Tippfehler manchmal ein unvermutetes und unsinniges Kompositum erzeugen, wobei nur ein fortgeschrittener Spellingchecker eine Chance haben wird, das zu entdecken.

Zweitens gibt es, relativ gesehen, nicht viele, aber immerhin einige Fälle, worin dasselbe Wort zwei sinnvolle Zergliederungen hat (z.B. *heroin-o*, *hero-in-o*; *sublim-a*, *sub-lim-a*; *vek-ant-a*, *ve-kant-a*).

Verdeutlichung der Wortkontrolle durch Sichtbarmachung der Morphemgrenzen bringt darin Klarheit: sowohl für PC-Benutzer als auch für die Software-Entwickler.

Die "Intelligenz" des Computers, wie gut sie auch werden kann, wird immer beschränkt sein. Umso wichtiger ist es, dass die zugrundeliegende Morphemanalyse deutlich dokumentiert wird.

### 3 Trennzeichen als Lese-Hilfe

Bekanntlich wurden in manchen Esperanto-Lehrbüchern, insbesondere bereits in Zamenhof's berühmtem *Ekzercaro*, Morphemgrenzen systematisch durch Apostrophe angegeben, z.B.:

*Ni ĉiu\_j kun'ven'is, por pri'parol'i grav'a'n afer'o'n; sed ni ne pov'is ating'i i'a'n rezultat'o'n, kaj ni dis'ir'is.*

In der modernen Zeit ließ sich niemals ein so kräftiges Plädoyer für *morfeĵlim-montriloj* hören wie 2005 von Prof. Dr. med. Horst H. Renemann<sup>6</sup>, einem ehemaligen Piloten bei der Lufthansa, der sich auf Flugmedizin und Ergonomie spezialisiert hatte und sich außerdem für die Verwendung von Esperanto statt Englisch im internationalen Flugverkehr einsetzt.

Hinzu kommt, dass heutzutage im Cockpit die Kommunikation von der Flugverkehrsleitung nicht nur im Kopfhörer, sondern zu gleicher Zeit auf einem Bildschirm wahrnehmbar ist. Das Ziel ist selbstverständlich die Vermeidung von irgendwelchen Missverständnissen. Beim Lesen vom Bildschirm ist dann – gerade unter Zeitdruck – das Verstehen längerer Wörter äußerst wichtig.

Das erklärt Renemanns spezielles Interesse an einer Sache, die übrigens auch außerhalb der Luftfahrt nützlich sein kann. Wo das Drucken von Trennzeichen in Büchern immerhin eine heikle Sache ist – es kann Lesern helfen, aber sie genauso gut irritieren – bietet dagegen ein PC die Möglichkeit, sämtliche Trennzeichen durch einen Mausklick erscheinen oder wieder verschwinden zu lassen. Und natürlich lässt sich eine Software machen, die dem Benutzer die Wahl gibt, statt Apostrophe andere Trennzeichen einzusetzen, oder z.B. die Grenzen zwischen *finaĵo* und *postfiniĵo* nicht zu markieren:

*Ni ĉiu\_j kun\_ven\_is, por pri\_parol\_i grav\_an afer\_on; sed ni ne pov\_is ating\_i i\_an rezultat\_on, kaj ni dis\_ir\_is.*

Wie das Beispiel zeigt, werden auch die Grenzen zwischen Wurzeln und Endungen bei einfachen Wörtern (*ĉiu\_j*, *venis*, *gravan*, *aferon*, ...) angegeben. Wir konzentrieren uns in diesem Artikel aber auf die zusammengesetzten.

### 4 Komposita: Art und Abgrenzung

Der Wortstamm eines Kompositums kann aus zwei (*merkato\_ekonomio*) aber auch aus mehreren Morphemen zusammengesetzt sein (*ek\_respond\_ec\_i*, *kamp\_ar\_an\_ar\_o*), wobei manchmal Affixe (Präfixe und Suffixe) verwendet werden. Nicht immer ist ein Kompositum ein längeres Wort; so sind z.B. auch *av\_in\_o*, *po\_pas\_e*, *iu\_kial\_e* Komposita.

---

<sup>6</sup> Persönliche Kommunikation mit dem Autor während des Esperanto-Weltkongresses 2005, Vilnius.

Im Gegensatz zu anderen Sprachen besteht in Esperanto kein prinzipieller Unterschied zwischen lexikalischen Stammmorphemen und sogenannten ‘gebundenen’ Morphemen (Affixen). Wörter wie *dis\_er\_iĝ\_o* (Zerfall) und *aĉ\_aj\_ar\_o* (Plunder, Schweinerei) sind ausschließlich aus Affixen gebildet.

Die Definition von ‘Kompositum’ in Esperanto lautet: jedes Wort das, abgesehen von einer grammatischen Endung, aus mehr als einem Morphem besteht.

Die grammatischen Endungen sind *-a, -e, -o, -i*, eventuelle *postfinaĵoj -n, -j, -jn*, und die Verbalendungen *-as, -is, -os, -us, -u*. Affixe (z.B. *-et, -in, -ant, -int, -at, -it, ...*) erzeugen immer ein ‘Kompositum’, es sei denn, der ganze Wortstamm besteht aus nur einem Affix (*aj\_o, er\_ojn*).

## 5 Verwendung des Wörterbuchs

Abgesehen von Eigennamen stehen im großen Esperanto-Wörterbuch NPIV<sup>7</sup> ungefähr 22.500 Komposita, gegenüber etwa 18.500 ‘einfachen’ Wörtern. Von den nicht-zusammengesetzten sind 13.500 Hauptwörter eines Lemmas; die übrigen 5000 sind lediglich grammatische Variationen eines Hauptwortes, z.B. *reĝa, reĝe, reĝi* im Eintrag *reĝo*).

Die Komposita in NPIV sind immer einem Lemma untergeordnet. So findet man im Eintrag vom Stichwortartikel *ven/i* u.a. die Zusammensetzungen *nerē~igebla, ne~o, post~i*. Die Tilde gibt nur die Morphemgrenzen um die Wurzel herum an, aber weitere Morphemgrenzen (z.B. in *nerē~igebla*) werden nicht dargestellt!

Genauso verhält es sich in der von Grimley Evans digitalisierten NPIV-Version<sup>8</sup>. Ein Software-Hersteller, der die systematische Darstellung von Morphemgrenzen ermöglichen will, muss sich deshalb einige Mühe geben, schon für die im Lexikon enthaltenen 22.500 Komposita.

Um auch ‘produktive’, frei gebildete Komposita mit den zutreffenden Morphem-Trennzeichen ausstatten zu können, muss die Software zumindest über ein vollständiges Computerlexikon von Wortwurzeln (*radikoj*) verfügen. Zusätzlich aber ist folgendes erforderlich:

- Die Fähigkeit zum Zurückgreifen auf die bereits im Lexikon enthaltenen (‘geprägten’) Komposita, welche ja wieder als Bausteine dienen können für die Bildung von längeren Komposita: z.B. *flug\_haven* und *ofic\_ist\_o* zwecks *flug\_haven\_ofic\_ist\_o*.
- Eine Einrichtung die dafür sorgt, dass auf irgendeine Weise die Gebrauchsfrequenz einer Wurzel berücksichtigt werden kann. Je reicher eine Sprache an (kurzen) Wortwurzeln ist, umso größer ist die Chance für mehrere (Fehl-)Zergliederungen bei Komposita. Es hilft dabei, wenn die in der Software eingebaute Entscheidungsstrategie kontrollieren kann, ob eine Wurzel zum Allgemeinwortschatz gehört oder zu einem begrenzten Fachgebiet (Botanik, Jura, Mythologie, ...). In dieser Hinsicht nützt schon eine Abstufung mittels separater Liste von zirka 2200 Basalwurzeln<sup>9</sup>. Ein weiteres Mittel ist die Nutzung von Fachgebietsvignetten, wie sie in den Druckversionen des PIV und NPIV existieren.

<sup>7</sup> Nova PIV [Waringhien, Gaston/Duc Goninanz, Michel (2002,Hrsg.): La Nova Plena Ilustrita Vortaro de Esperanto. Paris : SAT].

<sup>8</sup> Edmund Grimley Evans, “Kapvortoj de PIV”, Version 1.2, 2004-05-12 [das betrifft hier NPIV, einschließlich die Korrekturen für die provisorische Version von 2004/01/05]. Der Name ‘Kapvortoj’ (gemeint sind Stichwörter-Red.) ist etwas irreführend, weil neben Stichwörtern auch sämtliche untergeordneten Wörter in dieser digitalisierten Wörterbuchversion enthalten sind.

<sup>9</sup> BRO [Baza Radikaro Oficiala, Akademio de Esperanto 1966], ebenfalls digitalisiert von Edmund Grimley Evans [1998/02/04].

Glücklicherweise ist im Internet eine Vielzahl von lexikalischem Material zugänglich. Was in der von Grimley Evans digitalisierten NPIV-Version noch fehlt, sind Informationen darüber, ob ein Wort allgemeinsprachlich ist oder zu einem Fachgebiet gehört. Das findet man (oder besser gesagt: die oben erwähnte Software) in einer computerisierten und PIV-basierten Wortliste Esperanto-Japanisch. Die erste Ausgabe von 1997 dieses sogenannten DENTAN-Projekts<sup>10</sup> war das Ergebnis einer mehrjährigen Zusammenarbeit zwischen 17 japanischen Esperantisten, unter Leitung von HIROTAKA Masaaki und ONO Takao.

DENTAN ist PIV-basiert, aber nicht PIV-identisch. Die Kennzeichnung der *Fundamenteco*<sup>11</sup> der *Oficialaj Aldonoj* (Offizielle Ergänzungen)<sup>12</sup> aus dem PIV sind zuverlässig in DENTAN übernommen. Die Wortklassifizierung mit Vignetten für etwa 60 Fachgebiete ist im Großen und Ganzen dieselbe. Einen erheblichen Unterschied gibt es aber zwischen der relativ wenig verwendeten Wortgruppe mit der Bezeichnung *komunuza*<sup>13</sup> in der Druckausgabe des PIV und der in DENTAN häufig auftretende Klasse 一般語彙 (Allgemeinsprachlicher Wortschatz). Obwohl diese Klasse aus globaler Sicht einzelne Wörter vielleicht ungerechtfertigt enthält [z.B. ein Wort wie *dodoico*, japanischer Folksong], ist sie trotzdem nützlich. Sie lässt sich auch mit der Kennzeichnung *Fundamenteco/Oficiala Aldono* kombinieren, zwecks Entscheidungsstrategie in der Software für die Zergliederungen von Komposita.

Schließlich enthält DENTAN auch Angaben über die Transitivität von Verben, was insofern nützlich ist, als eine als Spellchecker fungierende Komposita-Analyse damit auf eine fehlerhafte Verwendung der Passiv-Suffixe *-at* oder *-it* (wie in *\*venita*, *\*okazata*, *\*naskiĝita*...) aufmerksam machen kann.

## 6 Die REIFLER-These und sonstige Regeln

In einer Strategie, um aus mehreren Zergliederungsmöglichkeiten die beste zu wählen, steht die Reifler-These an erster Stelle. Der Sprachwissenschaftler Erwin Reifler arbeitete in den USA in den Anfangsjahren der maschinellen Übersetzung, wo er mit seiner Forschung zur automatischen Zergliederung deutscher Komposita beitrug<sup>14</sup>. Von ihm stammt der sogenannte “Grundsatz des längsten Vergleichs”, der nicht nur für die deutsche Sprache effektiv ist. Für Esperanto lässt die Reifler-These sich frei übersetzen als ‘die Tendenz zum längsten Morphem’.

Schauen wir uns die möglichen Zergliederung des Worts *aventurismo* an. Die rohe Gewalt des Computers hat in wenigen Millisekunden sein Vokabular durchsucht und darin die Morpheme *av* (Großvater), *en* (Präposition / japanische Währung), *tur* (Turm), *ent* (Abstrahierung in der Scholastik), *ur* (ausgestorbenes Rind / etwas mit Urin), *aven* (Hafer), *aventur* (Abenteuer) und *ism* (Suffix) aufgespürt, was insgesamt vier Zergliederungen ergibt:

av en tur ism o  
av ent ur ism o  
aven tur ism o  
aventur ism o

<sup>10</sup> Siehe: [www.s-w.co.jp/~taon/dentan](http://www.s-w.co.jp/~taon/dentan) .

<sup>11</sup> Gemeint sind die Wortstämme, die in der 1905 beschlossenen Systemurkunde des Esperanto (Fundamento de Esperanto) aufgeführt sind.

<sup>12</sup> Es handelt sich um Wortlisten, die von der Akademio de Esperanto sanktioniert sind. Die dort registrierte Lexik, (gleichsam Ergänzungen zum Fundamento) besitzt damit einen Status, der dem im Fundamento registrierten Wortschatz vergleichbar ist, also „offizielles“ Esperanto darstellt-Red.

<sup>13</sup> ‘allgemein verwendet’ mit der Bedeutung ‚gemeinsprachlich‘.

<sup>14</sup> “Mechanical determination of the constituents of German substantive compounds”, Mechanical Translation 2, 1955.



Auf Grund der Reifler-These ist hier die untere Lösung die richtige, weil sie die längsten (oder anders formuliert: die wenigsten) Morpheme enthält. Beispiele sind zahlreich, einige andere sind: *amasdetruarmiloj*, *interkultureco*, *amuzita*. Aber passen Sie auf: die Reifler-These ist - jedenfalls für Esperanto – eine Tendenz, kein Gesetz! Das zeigt folgendes Beispiel:

an tim on di gist o  
anti mon di gist o  
anti mon dig ist o  
anti mond ig ist o  
antimon dig ist o

Nach dem Grundsatz des längsten Vergleichs würde wieder die untere Lösung die richtige sein, aber das stimmt hier nicht (zumindest solange es noch keine Deiche von Antimon gibt!). Ein anderes Gegenbeispiel ist *suprecitita*, das man ja nicht als *sup\_recit\_it\_a* zergliedern möchte. Die Reifler-These bleibt wichtig, aber sie muss mit einem ziemlich komplexen System von weiteren Regeln und Punktwertungen unterstützt werden. Elemente darin sind:

- Punktbewertung von regulären Affixen;
- Kontrolle auf Affixe an falschen Stellen (z.B. *-ant* vorne, *po-* hinten);
- Punktbewertung für gemeinsprachliche Wurzeln (wird negativ sein für z.B. *antimon*);
- Verbot für...-a....-o und ....-i....-o (bei *grandaspiro* bleibt nur *grand\_aspir\_o* übrig);
- Verbot für...-i....-a (es sei denn mit *-pova*, *-deva*, *-vola*);
- Beschränkung von Funktionswörtern (kein Personalpronomen wie in *vi\_rok\_at\_o*).

Zu den Regeln gibt es auch wieder Ausnahmen, sogar in Esperanto (*surdamutulo*, *vaksimaŝino*, *neforgesumino*, ...).

## 7 Hülsenwörter machen es schwierig

Innerhalb eines Kompositums kann ein Esperanto-Wort mit oder ohne Fugen-Vokal stehen (z.B. *prezodiferenco* oder *prezdiferenco*), das heißt: es können dort Wörter auftreten ohne Endung, also Wortstämme und Wortwurzeln.

Je kleiner ein Wort ist, umso mehr "Hülsenwörter" hat es. Zum Beispiel hat die Wurzel *di* vom Substantiv *dio* 6 Hülsenwörter: *did*, *dig*, *dik*, *dil*, *din*, *dir*. Die Wurzel *te* von *teo* hat 8 Hülsenwörter: *ted*, *teg*, *tek*, *tel*, *tem*, *ten*, *ter*, *tez*; und die Wurzel *al* von *alo* sogar 22: *bal*, *fal*, *gal*, *hal*, *kal*, *mal*, *ral*, *sal*, *tal*, *val*, *alb*, *ald*, *ale*, *alf*, *alg*, *ali*, *alk*, *alm*, *aln*, *alo*, *alp*, *alt*, noch abgesehen von der Präposition *al*.

Bei zunehmender Wortlänge verringert sich die Hülsenwortzahl: bei der Wurzel *plan* von *plano* gibt es vier Hülsenwörter: *pland*, *plane*, *plank*, *plant*; bei *vesper* von *vespero* nur eins *vespert*. Den Begriff Hülsenwort muss man hier definieren als: ein um einen Buchstaben erweitertes Wort, egal ob die Erweiterung am Wortanfang oder am Wortende stattfindet.

Hülsenwörter bilden ein Risiko, weil sie in einem Kompositum in Paaren auftreten können und so außer einer erwünschten auch eine unerwünschte Kompositum-Zergliederung verursachen:

sen tas kul o  
sen task ul o

Die Lösung besteht hier darin, dass in der Punktwertung dem Suffix *ul* Priorität gegeben wird vor dem Hülsenwort *kul*. Die Priorität von *task* über *tas* folgt daraus. In vielen Fällen aber bestimmt die Allgemeinsprachlichkeit eines Wortes die Punktbewertung: *oneco* ist ein Fachwort (Chemie), hat

daher ein niedrigeres Gewicht als das gemeinsprachliche *boneco*. Ähnliches gilt für das Paar *omaso* (Blättermagen, Omasus) und *maso*.

<i>korb on ec o</i>	<i>varmo maso</i>
<i>kor bon ec o</i>	<i>varm omaso</i>

Für eine gute Wirkung der Punktbewertung lohnt es sich, alle Hülsenwörter (von Wörtern bis zu 7 Buchstaben) systematisch zu inventarisieren und ihre Angaben zur Allgemeinsprachlichkeit im digitalisierten Wörterbuch kritisch zu kontrollieren.

Dem Software-Hersteller, der das Ziel hat, die Komposita-Analyse zu optimieren, steht es frei, einem Wort diese Angabe zu entziehen, falls sie sowieso diskutabel ist und außerdem ein erhebliches Risiko bildet. Ganz kurze Wörter bilden ja noch ein zusätzliches Risiko, weil sie im Falle von Tippfehlern leicht zu einer formell richtigen jedoch unsinnigen Zergliederung des Kompositums führen können.

## 8 Schluss und Vorausblick

Eine automatische Analyse von Esperanto-Komposita ist möglich, vorläufig bis zu einer Erfolgsrate von schätzungsweise 99,5 Prozent. Das bedeutet, dass man eine Software herstellen kann, die in einem Esperanto-Text Morphemtrennzeichen automatisch erscheinen lässt – bis auf einige Ausnahmen – auch in produktiven Komposita.

Hundertprozentig einwandfrei wird die automatische Analyse von zusammengesetzten Wörtern wahrscheinlich nie. Schließlich gibt es auch mehrdeutige Fälle, wo eben der Kontext nicht genügend Auskunft bringt. Im Großen und Ganzen aber kann man mit den heutigen Möglichkeiten des Computers, mit dem digitalisierten Esperanto-Wortschatz und mit einem vernünftig hergestellten Softwaresystem gut zurechtkommen.

Eine solche Software wird es auch ermöglichen, dass eine vom Benutzer abgelehnte und korrigierte Zergliederung sofort dem eingebauten Wortschatz hinzugefügt wird, gegebenenfalls an einem fach- oder kundenspezifischen Teil. Dasselbe kann übrigens eingerichtet werden für akzeptierte Zerlegungen von produktiven Komposita. Auf diese Weise bereichert man allmählich das eingebaute Esperanto-Wörterbuch. Die Aufnahme eines 'Problemworts' (wie z.B. *supr\_e\_cit\_it\_a*, eine Ausnahme der Reifler-These) ins Wörterbuch entlastet außerdem die Verarbeitungssoftware.

Das größte Problem bleibt immerhin die Signalisierung von Tipp- und Rechtschreibfehlern. Die automatische Morphemanalyse ist zugleich eine Kontrolle der Orthografie, oder besser gesagt: sie sollte das sein. Aber bei vielen Schreibfehlern produziert die Software stattdessen bloß ein unerwartetes, sinnloses Kompositum, z.B. *labor\_ip\_ret\_a* wenn *laborpreta* gemeint war; *pra\_vit\_it\_a* wenn *pravigita* gemeint war; *reg\_lament\_ad\_o* wenn *reglementado* gemeint war; *ek\_ter\_aj* wenn *eksteraj* gemeint war<sup>15</sup>.

Zur recht zuverlässigen Fehlersignalisierung sollte der Computer nicht nur ein Lexikon und sprachliches Wissen sondern auch folgende Kenntnisse besitzen:

- Kenntnis typischer und häufiger Orthographie-Fehler;
- Kenntnis der Tastatur (Fehlanschläge auf naheliegenden Tasten);
- Kenntnis sonstiger Eingabeverfahren (Klarschriftlesen, Spracherkennung);
- Kenntnis der Textdomäne und des Textzusammenhangs;
- Kenntnis der Eigennamen-Listen (und adjektivierter Eigennamen ohne Großschreibung).

---

<sup>15</sup> Die Beispiele sind der MondeDiplo entnommen (= die Esperanto Version von "Le Monde Diplomatique", 2002-2006, <http://eo.monediplo.com>).

Solange das alles noch fehlt, muss man sich mit einer Markierung (z.B. mittels der Farbe blau) sämtlicher analysierter produktiver Komposita begnügen, sowohl der richtig geschriebenen als auch der falsch geschriebenen. Der Benutzer kann dann selber kontrollieren, welche von den markierten Komposita richtig sind, wo Fehler im Text gemacht wurden, und wo ausnahmsweise ein korrekt geschriebenes Kompositum von der Software falsch zergliedert worden ist. In diesem letzten Fall gibt es übrigens noch die Möglichkeit mittels Mausclick die Software um eine alternative Zerlegung zu bitten. Man kann dazu die Software so einrichten, dass sie ein blau markiertes Kompositum zusätzlich unterstreicht, falls es alternativ zergliedert werden kann.

## Autoren

*Sabine Fiedler* (Am Zollamt 5, DE-04838 Gordemitz, [sfiedler@rz.uni-leipzig.de](mailto:sfiedler@rz.uni-leipzig.de)), Privatdozentin, Dr. phil., Linguistin, unterrichtet englische Sprachwissenschaft und Interlinguistik am Institut für Anglistik der Universität Leipzig, stellvertr. Vorsitzende der Gesellschaft für Interlinguistik e.V.

*Gunnar Rudolf Fischer* (Steinfurter Straße 2a, 48149 Münster, [Gunnar.Fischer@esperanto.de](mailto:Gunnar.Fischer@esperanto.de)), Diplom-Wirtschaftsinformatiker, IT-Consultant in Münster (Westfalen). Er ist Esperanto-Muttersprachler und seit vielen Jahren aktiv in der Esperanto-Musikszene.

*Rudolf-Josef Fischer* (Gustav-Adolf-Str. 2a, DE-48356 Nordwalde, [fischru@uni-muenster.de](mailto:fischru@uni-muenster.de)), Dipl. Math., Dr. rer. medic., Dr. phil., Privatdozent am Institut für Med. Informatik und Biomathematik der Univ. Münster, dort Lehrbeauftragter für Interlinguistik am Institut für Allgemeine Sprachwissenschaft.

*Wim Jansen* (Emmaplein 17A, NL-2225 BK Katwijk, Niederlande, [wimjansen@casema.nl](mailto:wimjansen@casema.nl)), Diplomingenieur (Luft- und Raumfahrttechnik), Diplomphilologe (Baskologie), Dr. phil. (Esperantologie), Dozent für Interlinguistik und Esperanto an der Universität Amsterdam.

*Sebastian Kirf* (Neutorstr. 65, 26721 Emden, [ePost@kirf.de](mailto:ePost@kirf.de)), Diplomsozialpädagoge, wiss. Mitarbeiter für virtuelle Lehre und Medienpädagogik, FH Emden.

*Andreas Künzli* (Blockweg 8, CH-3007 Bern, [kuenzli@osteuroopa.ch](mailto:kuenzli@osteuroopa.ch)), Lic. phil., Slawist, Webmaster ([www.plansprachen.ch](http://www.plansprachen.ch)).

*Ulrich Lins* (Lindenallee 13, DE-53173 Bonn, [u.lins@gmx.net](mailto:u.lins@gmx.net)), Dr. phil., Historiker, Mitarbeiter des Deutschen Akademischen Austauschdienstes.

*Toon Witkam* (Ramstr. 6, NL-3581 HH Utrecht, Niederlande, [toon.witkam@planet.nl](mailto:toon.witkam@planet.nl)), Diplomingenieur, Informatiker und Computerlinguist.

*Fritz Wollenberg* (Normannenstr. 9, 10367 Berlin, [F.W\\_PrB@t-online.de](mailto:F.W_PrB@t-online.de)), Pädagoge, Erzieher in einem Kinderfreizeithaus.