



## **Verbildete, verbeamtete und überalterte Deutschschweizer? Das Schweizerhochdeutsche im Wörterbuch und sonst**

Farø, Ken Joensen

*Published in:*

Proceedings of the Eleventh EURALEX International Congress. EURALEX 2004, Lorient, France, July 6-10, 2004

*Publication date:*

2004

*Document version*

Også kaldet Forlagets PDF

*Citation for published version (APA):*

Farø, K. J. (2004). Verbildete, verbeamtete und überalterte Deutschschweizer? Das Schweizerhochdeutsche im Wörterbuch und sonst. I G. Williams, & S. Vessier (red.), *Proceedings of the Eleventh EURALEX International Congress. EURALEX 2004, Lorient, France, July 6-10, 2004* (s. 879-884). Bretagne: UBS.

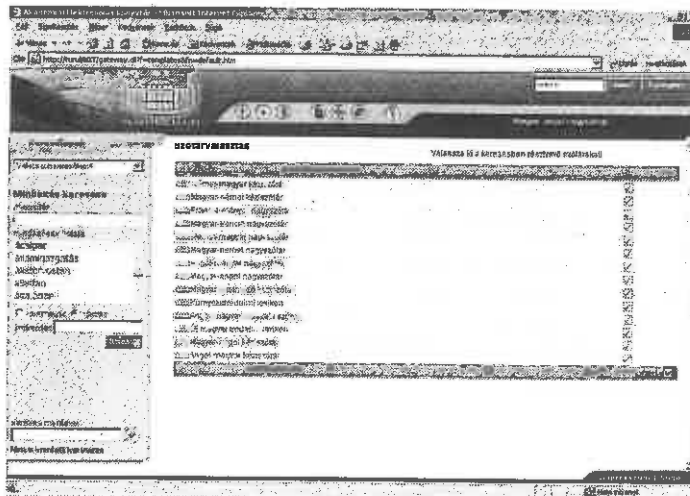


Figure 5: Akadémiai Electronic Library – Available dictionaries



Figure 6: Akadémiai Electronic Library – Search functions

## Verbildete, verbeamtete und überalterte Deutschschweizer? Das Schweizerhochdeutsche im Wörterbuch und sonst.

Ken Farø  
Universität Kopenhagen  
kenfaroe@hum.ku.dk

### Abstract

Der Beitrag beschäftigt sich mit dem Problem der Kodifizierung des Schweizerhochdeutschen (=Shd.)<sup>1</sup> in deutschen gemeinsprachlichen Wörterbüchern. Als prominente Beispiele werden das *Duden Deutsches Universalwörterbuch* (2003) und das *Wahrig Deutsches Wörterbuch* (2002) untersucht. Es wird die These vertreten, dass die Unterschiede zwischen dem Gemeindeutschen und dem Shd. wesentlich umfassender sind, als durch diese Wörterbücher vermittelt wird. Außerdem könnte man den Eindruck gewinnen, die shd. Standardvarietät sei „altmodisch“, „papierdeutsch“, „bildungssprachlich“, etc., obwohl es sich um normale Deutschschweizer Besonderheiten handelt. Dies fördert die ohnehin verbreitete Auffassung vom Deutsch der Schweizer als defizitär, was bei einer plurizentrischen Sprache wie dem Deutschen inadäquat ist. Der Artikel stellt überdies eine Methode vor, anhand deren der Begriff „Helvetismus“ operationalisiert werden kann, sowie eine rudimentäre Typologie der lexikographischen Präsentation nationaler Besonderheiten.

### 1. Einleitung

Wer in die Schweiz kommt, dortige Zeitungen liest, dem Hochdeutsch der Deutschschweizer zuhört, und diese linguistischen Eindrücke mit der Beschreibung der shd. Besonderheiten in einem der verbreitetsten Standardwerke zum Deutschen, dem DUW (2003), vergleicht, wird feststellen, dass ein Großteil dieser Formen und Bedeutungen keine Erwähnung findet. Und zwar trotz des Versprechens des Wörterbuchs, die „Ausprägungen der deutschen Sprache in Österreich und der Schweiz“ seien auch berücksichtigt (Vorwort). Das Wörterbuch enthält zwar etwa 1800 Angaben zum shd. Sprachgebrauch; diese Zahl sollte jedoch im Verhältnis zu den etwa „140.000 Wörtern und Wendungen“ gesehen werden, die das DUW nach eigener Angabe vorweisen kann (ebd.). Nach Meyer (1989:14) machen die spezifischen Helvetismen des Shd. etwa 5% dieser Varietät aus. Doch das Shd. scheint weit weniger als 5% des Inhalts des DUW zu betragen; meiner Beobachtungen nach könnte es sogar so aussehen, als ob der wirkliche Umfang der Helvetismen Meyers Schätzung wesentlich übertrifft (s.u.). Das Ziel dieses Beitrags besteht demnach darin zu zeigen, dass sich die empirisch zu beobachtenden Divergenzen zwischen dem Deutsch der (primär) Deutschen und dem (Hoch)Deutsch der Deutschschweizer lexikographisch keineswegs auf einem zufriedenstellenden Stand befinden. Zunächst soll aber das theoretische Problem *Helvetismus* geklärt werden, denn die Beschäftigung mit der Frage nach dem spezifisch shd. Anteil an der deutschen Standardsprache kann nicht beim Einsammeln von Belegen stehen bleiben, die „offensichtlich“ Helvetismen sind. Sondern man braucht ein Werkzeug, um diese so präzise wie möglich identifizieren zu können, wenn die Untersuchung einen Anspruch auf Wissenschaftlichkeit erheben möchte. Diese methodologische Frage wurde in der Forschung wenig erörtert, dabei stellt sie ein Grundproblem der Beschäftigung mit den

nationalen Varietäten einer plurizentrischen Sprache dar. „Helvetismen“ werden in diesem Beitrag nicht als Codeswitching-ähnliche schweizerische (=schw.) Dialekteinsprengsel in der Hochsprache verstanden (vgl. Stäubli 1990:52), sondern damit sind hier Besonderheiten der schw. Varietät der deutschen Standardsprache gemeint, die durchaus dialektalen Ursprungs sein können, jedoch oft auch andere Gründe haben. Außerdem wird hier systematisch zwischen *absoluten* und *relativen* Helvetismen unterschieden.

## 2. Absolute Helvetismen

Bei den *absoluten* Helvetismen handelt es sich um Phänomene der Lexik oder Grammatik, die (fast) nur von Schweizerdeutschen produziert werden. Prozentual definiere ich dies relativ willkürlich als 91-100% der Vorkommen eines linguistischen Phänomens innerhalb des gesamtdeutschen Pools. Bei *relativen* Helvetismen dagegen handelt es sich um linguistische Merkmale, die nicht nur als schw. gelten können, die aber in schw. Texten angesichts des zahlenmäßigen Anteils der schw. Sprecher am Gemeindeutschen eine Überfrequenz erreichen. Ich definiere diese Überfrequenz als 10-90% der Vorkommen eines linguistischen Phänomens. Diese Zahlen sind wohl insofern realistisch, als sie zumindest nicht zu *viele* Helvetismen generieren, denn bei 4,5 Mio. Deutschschweizern (Bickel 2000:21) von grob gerechnet 100 Mio. Deutsch Sprechenden insgesamt müsste die produzierte Textmasse der Deutschschweizer erwartungsgemäß etwa 5% der gesamtdeutschen Textproduktion ausmachen, was gleichzeitig die Wahrscheinlichkeit eines Helvetismus innerhalb des gesamtdeutschen Textpools wäre. Sicherheitshalber aber habe ich diese Schwelle auf 10% erhöht, weil es mir keineswegs um eine Überbewertung des Phänomens geht. Die Methode, anhand deren man Hypothesen über Helvetismen testen kann, ist korpuslinguistisch, und als konkretes Korpus soll das Internet genutzt werden, das sich für dieses Problem geradezu anbietet. Wenn man z.B. das im DUW nicht verzeichnete Wort *Bancomat*<sup>2</sup> in die Suchmaschine *Google* eingibt – in seiner flektierten Form *Bancomaten*, um so u.a. italienische Formen zu vermeiden –, erhält man 1650 Belege. Um nicht sämtliche Belege analysieren zu müssen, wird eine Grenze bei 50 Belegen gesetzt. Die meisten von diesen können unmittelbar als aus schw. Homepages stammend identifiziert werden, andere erst nach Öffnung des Dokuments. Nur ein Beleg kann nicht eindeutig als shd. kategorisiert werden, was aber für die Einstufung von *Bancomat* als einem absoluten Helvetismus kaum eine Rolle spielt. Man kann also *Bancomat* als einen Helvetismus bezeichnen, weil das Wort in 98% aller Belege mit relativ großer Sicherheit von schw. Schreibern produziert wurde und sich somit innerhalb des hier aufgestellten Toleranzbereichs hält. Dass das Verfahren trotz seiner Einfachheit ganz im Sinne eines popperschen Falsifikationstests funktioniert, zeigt eine große Menge an Beispielen, die von diesem Verfasser für potenzielle Helvetismen gehalten wurden, die aber aussortiert werden mussten, weil sie die oben genannten Kriterien nicht erfüllten. Verifiziert werden konnten aber vom DUW nicht kodifizierte *absolute* Helvetismen wie:

- *Ausschaffung* („Abschiebung von Ausländern“; 98%); *Betagte* („Senioren“; 100%); *Durchflener* („jd der den Wehrdienst am Stück ableistet“; 100%); *Garagist* („Autoschlosser“; 92%); *Kapo* („Kantonspolizei“; 100%); *Logiernacht* („touristische Übernachtung“; 100%); *Natel* („Handy“; 100%); *pfaden* („Schnee räumen“; 100%); *Reiseocar* („Touristbus“; 100%); *Signalisation* („Ampelanlage“; 94%); *tief* („niedrig“, als Kollokator z.B. bei *Einkommen*; 100%); *Trottinett* („Roller“; 100%)

Beispiele für nicht kodifizierte absolute Helvetismen gibt es auch innerhalb der *Phraseologie*, etwa:

- *in (den) Ausgang gehen* („ausgehen um zu feiern“; 100%); *von Beginn weg* („gleich vom Beginn“; 96%); DUW: „von ... weg; ugs.“); *etw einen Riegel schieben* („etw verhindern“; 100%); DUW: „etw einen Riegel verschieben“); *am selben Strick ziehen* („am selben Strang ziehen“; 82%); DUW: Hinweis auf „Strang“); *gebranntes Wasser* („Schnaps“; 100%); *in stoßender Weise* („empörend“; 100%)

Schließlich soll ein wichtiger Typus hochfrequenter<sup>3</sup> absoluter Phraseohelvetismen erwähnt werden, der vom DUW generell nicht beschrieben wird, nämlich die Phrasenschablonen *Anfang/Mitte/Ende* + [ZEITANGABE IN NOM.] wie etwa *Ende Monat*, *Anfang Woche* und *Mitte Jahr*, d.h. als Juxtaposition, wie im Bundesdeutschen nur bei den Monatsnamen üblich.

## 3. Relative Helvetismen

Relative Helvetismen wurden als solche Phänomene definiert, die eine spezifisch schw. Belegfrequenz zwischen 10 und 90% haben. D.h. konkret, dass von den 50 Texten, die man herunterlädt, nachdem man die potenziell shd. Form, Bedeutung oder Konstruktion in die Suchmaschine eingegeben hat, mindestens 5 eindeutig schw. Ursprungs sein müssen. Als Beispiel kann das Wort *hospitalisieren* herangenommen werden, das vom DUW bloß als „amtsdeutsch“ markiert wird. In 80% der *Google*-Belege handelt es sich aber um schw. Texte, bei denen keineswegs immer von „Amtssprache“ gesprochen werden kann. Wir können deswegen auch hier getrost von einem relativen Helvetismus ausgehen. Andere nicht dokumentierte relative Helvetismen, die verifiziert werden konnten, waren z.B.:

- *Augenschein* („Untersuchung vor Ort“; 26%); DUW: „geh.“); *Bewohnerschaft* („sämtliche Bewohner“; 20%); *BusführerIn* („Busfahrerin“; 30%); *Erstsemestrige(r)* („Erstsemestler“; 38%); *Israeli* (pl. von „Israeli“; 33%); DUW: bloß fakultativ); *Knacknuss* („schwieriges Problem“; 84%); *Parteiobere(n)* („Parteiführer“; 20%); *Prozedere* („Verfahren“; 18%); DUW: bildungsspr.); *Studierende* („StudentIn“; 24%); *Suizid* („Selbstmord“; 18%); DUW: bildungsspr.); *zuwarten* („nicht handeln“; 38%)

Dass relative Helvetismen gleichzeitig z.B. relative *Austriazismen* darstellen können, beeinträchtigt den Wert der Methodologie kaum, sondern dies ist durchaus logisch. Es ist nämlich zu erwarten, dass relative Helvetismen auf einer höheren Klassifikationsebene manchmal allgemeine *süddeutsche* Spezifitäten darstellen. Damit wird aber deutlich, dass „relativ“ häufig eher ein Distanzbegriff zur bundesdeutschen Varietät wird, als zum Gemeindeutschen im Allgemeinen. Ein anderer übersehener Typus sind relative *phraseologische* Helvetismen, wie z.B.:

- *in den Ausstand treten* („zurücktreten“; 80%); *in allen Belangen* („was das Ganze betrifft“; 26%); DUW: „papierdeutsch“); *heute [WOCHENTAG]* („an diesem [WOCHENTAG]“; 60%); *im Nachgang zu etw* („nach etw“; 28%); DUW: „amtsdeutsch“); *in erster Priorität* („in erster Reihe“; 66%); *auf dem letzten Zucken* („mühevoll“ u.ä.; 78%)

„Helvetismus“ ist demnach häufig ein relativer Begriff. Und für absolut gehaltene Helvetismen können in Wirklichkeit relativ sein, vgl. z.B. *in den Ausstand treten* bei Meyer (1989). Trotzdem existieren offensichtlich nicht wenige Helvetismen, die wirklich nur von Schweizern benutzt werden, und die man deswegen als eigentliche schw. *Schibboleths* bezeichnen kann.

#### 4. Zur lexikographischen Präsentation von Helvetismen

Obwohl oben auf Unzulänglichkeiten der Kodifizierung von Helvetismen in einem der prominentesten Wörterbüchern des Deutschen hingewiesen wurde, so sollen hier keineswegs die grundlegenden lexikographischen Prinzipien der Behandlung dieser Varietät durch das Wörterbuch kritisiert werden. Denn wo das eigentlich Shd. erkannt wurde, wurden auch generell adäquate lexikographische Darstellungsverfahren verwendet. Es ist in diesem Zusammenhang wohl überflüssig, auf die spezifisch schw. Sonderlexik einzugehen (Beispiel: *Café crème*), d.h. nur in der Schweiz vorkommende Formen, die als Typ 1 zu kategorisieren wären. Problematischer dagegen sind die Fälle, in denen Schweizer Besonderheiten innerhalb eines auch im Gesamt- oder Bundesdeutschen vorkommenden Lexems zu verzeichnen sind. Es können hier verschiedene Typen registriert werden. Zunächst einer (Typ 2), bei dem sich eine schw. Sonderbedeutung von einer gemeindt. abhebt:

##### Agglomeration:

1. (bildungsspr.) *Anhäufung, Zusammenballung*: -en hässlicher Betonklötze
2. (schweiz.) *Ballungsgebiet*

Zwar scheint hier das Gemeindeutsche so zu sagen den Vorrang zu haben, obwohl es sich bei dieser Variante um eine markierte Form handelt. Doch das Prinzip einer selbstständigen schw. Bedeutung ohne Markierung ist eine adäquate Lösung des Problems. Der Typ 3 weist eine unmarkierte shd. Bedeutung neben stilistisch und funktional markierten Gebrauchsweisen der „gleichen“ Bedeutung, wenn auch gemeindt. Provenienz, auf:

**Knabe**: 1. (meist geh., sonst Amts- u. Geschäftsspr., schweiz.): [*größerer*] *Junge* [...]

Den Typ 4, der bei einem monosemen Lemma einen relativen Helvetismus beschreibt, soll dieses Beispiel illustrieren:

**For[ward]**: [...] (Fußball, Eishockey, bes. schweiz.): *Stürmer*

Dieser Typ wäre also nach den hier aufgestellten Kriterien aufgrund eines überfrequenten Auftretens (mehr als 10% der Belege von Schweizern verfasst, vgl. oben) zu legitimieren. Beim Typ 5, schließlich – dass es weitere geben kann, wird hier keineswegs ausgeschlossen –, handelt es sich um grammatische Besonderheiten des gleichen Lexems, wie z.B. als *Diversifikation* im Sinne von mehr Varianz im Shd. als im Gemeindt:

<sup>1</sup>**Hun[derts]tel**, das, schweiz. meist: der

Oder umgekehrt als *Restriktion*, d.h. dass bei einem Phänomen im Gemeindt. Fakultativität herrscht, die aber im Shd. eingeschränkt ist:

**Lil[ter]** [...], der (schweiz. nur so), auch: das

Diese typischen Gestaltungsweisen sind zweckmäßig, weil sie das Shd. als eigene Varietät registrieren, und können bei der noch ausstehenden exhaustiveren Erschließung verbreiteter Helvetismen übernommen werden. Würden diese Methoden auf das shd. Material konsequent angewendet, könnte der Eindruck vermieden werden, es handle sich beim Shd. um eine „archaische“, „gestelzte“ oder sonstwie defizitäre Abart des Bundesdeutschen, was sogar manchmal die Auffassung von Linguisten ist. Dies ist aber sicherlich wiederum z.T. auf die mangelhafte lexikographische Berücksichtigung des Shd. zurückzuführen.

#### 5. Schlusswort

Es wurde eine Methode vorgestellt, die es ermöglicht, Hypothesen über Helvetismen (und andere nationalen Sonderformen) relativ leicht zu testen und somit für wissenschaftliche Zwecke brauchbar zu machen. In den meisten Fällen funktioniert die Methode ausgezeichnet; sie hat aber sicherlich auch ihre Schwächen. Weniger verlässlich testen kann sie z.B. potenzielle Helvetismen, die sich auf spezifisch schw. *Inhalte* beziehen, was paradox klingen mag. Aber es ist wohl einleuchtend, dass etwa das Adjektiv *gesamtschweizerisch*, vom dem ich annehme, dass es zumindest als Adverb ein (relativer?) Helvetismus ist, natürlich in ester Linie in schw. Texten vorkommt, was aber noch kein Beweis für seinen Status als Helvetismus wäre. Denn mit dem gleichen Argument könnte man behaupten, das Syntagma *in der ganzen Schweiz* sei ein Helvetismus, weil es v.a. in schw. Texten auftaucht, was aber selbstverständlich ein unsinniger Schluss wäre. Potenzielle Fehlerquellen des Verfahrens sind übrigens die Möglichkeit, dass Shd. Schreibende im Vergleich zur gesamtdeutschen Tendenz erheblich über- oder unterrepräsentativ Texte im Internet veröffentlichen, oder dass schw. Homepages stark von Nichtschweizern verfasst würden. Das Letztere halte ich für wenig wahrscheinlich, während das Erstere nicht ganz von der Hand zu weisen ist. Trotzdem stellt das Verfahren sicherlich eine Verbesserung der Bestrebungen nach der Operationalisierung von Helvetismen dar.

Die andere Komponente dieses Beitrags hat bestätigt, dass Helvetismen aller Typen von einem der prominentesten deutschen Wörterbücher in großem Maß übersehen werden. Sowohl in der Lexik, darunter nicht zuletzt als Phraseme und Sonderbedeutungen, als auch grammatisch, scheinen sie immer noch ein zu wenig beachtetes Dasein zu fristen. Wichtig ist dabei die Erkenntnis, dass sich in nicht wenigen Fällen hinter lexikographisch als stilistisch, diachron, diasituativ u.a.m. markierten Lexemen oder Bedeutungen Helvetismen verstecken (vgl. Fenske 1973:160). Das DUW wird übrigens in dieser Hinsicht vom *Wahrig Deutsches Wörterbuch* (2002) (=WDW) Gesellschaft geleistet, das eine fast genau so geringe Kodifizierungsfrequenz der untersuchten Helvetismen aufweist. Zwar sind hier *Trottinett* und *gebranntes Wasser* vertreten, dafür fehlen aber mehrere der Helvetismen, die (unter 4.) immerhin im DUW verzeichnet waren. Und das obwohl schon laut dem Vorwort der 1975er Ausgabe „auf die Aufnahme österreichischer, schw. und landschaftlicher Besonderheiten [großer Wert] gelegt“ worden sei (op.cit.).

Das Shd. sollte aus soziolinguistischen Gründen lexikographisch als eine gleichberechtigte Varietät des Deutschen betrachtet werden, denn als plurizentrische Sprache kennt das Deutsche prinzipiell keine besseren Standardvarietäten, was spätestens seit Ammon (1995) klar geworden ist, und übrigens auf der IdS-Tagung 2004 bestätigt wurde. Stattdessen kennt sie einen großen gemeindeutschen Pool, wovon u.a. jeweils österreichische, bundes- und sd. Besonderheiten abweichen. Der v.a. geschriebene, aber auch von so gut wie allen gesprochene (vgl. Ris 1990:43) Standard der Deutschschweizer ist an sich weder *veraltet*, *amtssprachlich*, *papierdeutsch*, *gehoben* oder sonstwie defizitär im Vergleich zum Deutsch des großen Bruders. Sondern er ist eben *schweizerisch*. Nach Meyer (1989:14) wird ein größeres Wörterbuch des gesamten Shd. wohl nie erscheinen, weil die Redundanz im Vergleich zu den bereits existierenden Wörterbüchern damit zu groß würde. Man könnte aber einwenden, dass diese vielleicht weniger umfassend ist, als Meyer und andere (z.B. Ris 1990:43) annehmen. Denn von den hier verifizierten Helvetismen sind vom bisher einzigen shd. „Differenzwörterbuch“ (=Meyer 1989) nur 14% kodifiziert, was auf ein weiterhin

großes unbeschriebenes Material hindeutet. Insbesondere scheinen nicht zuletzt die phraseologischen und relativen Helvetismen eine Gruppe zu bilden, deren künftige Beobachtung im Hinblick auf eine mögliche „helvetische Verbindung“ lohnenswert erscheint.

Das DUW und das WDW sind immer noch keine Wörterbücher des Deutschen im Sinne seiner nationalen Varietäten, sondern sie sind zum großen Teil noch Wörterbücher des „Deutschländischen“, neben dem unsystematisch (quasi als Feigenblatt?) und verstreut Helvetismen und Austriazismen auftreten. Hoffentlich wird das im Herbst 2004 erscheinende „*Wörterbuch der nationalen und regionalen Varianten der deutschen Standardsprache*“ ([www.sprachwissenschaft.ch/prolex](http://www.sprachwissenschaft.ch/prolex)) wesentlich zur Änderung dieser Lage beitragen können.

### Schlussnoten

<sup>1</sup> Ich verwende mit Absicht nicht die Bezeichnung „Schriftdeutsch“, da dies zur bekannten Merkwürdigkeit wie bei Christen (2001:120): „Schriftdeutsch reden“ führen würde. „Hochdeutsch“, wie diese Varietät von den Deutschschweizern auch genannt wird (vgl. op.cit.:121), wäre in diesem Zusammenhang mit ‚Bundesdeutsch‘ (d.h. dem Deutschen Deutschlands) zu leicht verwechselbar.

<sup>2</sup> Bei den angeführten Beispielen handelt es sich sämtlich um bei der Lektüre der Zeitungen *20 Minuten*, *Neue Zürcher Zeitung* und *Der Tagesanzeiger* gefundene Helvetismen.

<sup>3</sup> Selbstverständlich werden hier nur hochfrequente Phänomene behandelt.

### Literatur und Quellen

20 Minuten 2003/04.

Ammon, U. (1995). Die deutsche Sprache in Deutschland, Österreich und der Schweiz: das Problem der nationalen Varietäten. Berlin etc.: de Gruyter.

Bickel, H. (2000). Deutsch in der Schweiz als nationale Varietät des Deutschen. In: Sprachreport 16/4.

Christen, H. (2001). ‚Die regionalen Besonderheiten der deutschen Standardsprache in der Schweiz‘. In: Knipf-Komlósi/Berend.

Duden (2003). Duden Deutsches Universalwörterbuch (CD-ROM-Version). Mannheim etc.: Dudenverlag.

Fenske, H. (1973). Schweizerische und österreichische Besonderheiten in deutschen Wörterbüchern. Tübingen: Narr.

Google (Internet-Suchmaschine: [www.google.com](http://www.google.com)).

Knipf-Komlósi, E./Berend, N. (Hg.) (2001). Regionale Standards. Sprachvariationen in den deutschsprachigen Ländern. Budapest-Pécs: Dialóg Campus Kiadó.

Meyer, K. (1989). Wie sagt man in der Schweiz? Wörterbuch der schweizerischen Besonderheiten. Mannheim etc.: Dudenverlag.

Neue Zürcher Zeitung 2003/04.

Ris, R. (1990). ‚Diglossie und Bilingualismus in der deutschen Schweiz: Verirrung oder Chance?‘. In: Vouga et al.

Stäubli, E. (1990). ‚Muttersprache“ und „Vatersprache“‘. In: Vouga et al. Der Tagesanzeiger 2003/04.

Vouga, J.-P./Hodel, M. E. (1990). La Suisse face à ses langues/Die Schweiz im Spiegel ihrer Sprachen/La Svizzera e le sue lingue. Aarau etc.: Verlag Sauerländer.

Wahrig-Burfeind, R. (Hg.) (2002). Wahrig Deutsches Wörterbuch. Gütersloh: Wissen Media Verlag. [www.sprachwissenschaft.ch/prolex](http://www.sprachwissenschaft.ch/prolex)

## ANCR – The Adjective-Noun Collocation Retriever (A Tool for Generating a Bilingual Dictionary from a German-English Parallel Corpus)

Günther Fliedl, Andreas Homa, Philippa Maurer-Stroh, Georg Weber

University of Klagenfurt

Universitätsstrasse 65-67

9020 KLAGENFURT

AUSTRIA

[philippa.maurer-stroh@aon.at](mailto:philippa.maurer-stroh@aon.at)

Sebastian Maurer-Stroh

IMP - Institute of Molecular Pathology

Bioinformatics - Group Eisenhaber

Dr. Bohrgasse 7

1030 VIENNA

AUSTRIA

### Abstract

ANCR is a tool which enables the user to automatically generate a bilingual dictionary of adjective-noun combinations from a sentence-aligned parallel corpus of German and English. The underlying idea of ANCR is to help translators and/or linguists retrieve correct translations of adjective-noun collocations without having to engage in time-consuming linguistic pre-processing tasks, which can depend on the availability of the appropriate software.

### 1 Introduction – Or the Motivation for ANCR

“Our knowledge of a language is not only a knowledge of individual words, but of their predictable combinations” (Stubbs 2001: 4). In fact, words in isolation can have different meanings from words combined as phrases. Take, *green light*, for example. The combination of the adjective *green* and the noun *light* does not only mean that there is a light which is green in colour, but also ‘a permission for a project etc. to start or continue’. When it comes to contrastivity this phenomenon shows even more strongly. The English lemma POSITION can have various German translation equivalents, like *Lage*, *Standpunkt*, *Stellungnahme*, *Position*, etc. and it can only be correctly translated when its context, e.g. its preceding adjective, is taken into consideration. Therefore, *political position* is translated as *politische Position* and *geographical position* as *geographische Lage*, etc.

The way words combine in a language in order to produce natural-sounding, native-like speech or writing can be defined by the concept of collocation (Deuter et al. eds. 2002). Thus, *lasting peace* will be more acceptable in a native-speaker environment than *durable peace*, although both *lasting* and *durable* are translated as *dauerhaft* in German. Starting in the 1980s, high-storage capacities and increased data-processing speed have made it possible to capture this *language in use*. Sophisticated software enables users, e.g. lexicographers, to