

Research in English and Applied Linguistics
REAL Studies 1

**English
for Central Europe:
Interdisciplinary
Saxon - Czech Perspectives**

edited by
Josef Schmied
Christoph Haase
Katrin Voigt

Cuvillier

English for Central Europe - Interdisciplinary Saxon-Czech Perspectives

Edited by Josef Schmied, Christoph Haase
and Katrin Voigt

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

1. Aufl. - Göttingen : Cuvillier, 2005

ISBN 3-86537-420-4

Published in REAL Studies, Vol. 1

"Research in English Language and Applied Linguistics"

Chemnitz University of Technology

© The REAL Centre 2005

and



© CUVILLIER VERLAG, Göttingen 2005

Nonnenstieg 8, 37075 Göttingen

Telefon: 0551-54724-0

Telefax: 0551-54724-21

www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.

1. Auflage, 2005

Gedruckt auf säurefreiem Papier

ISBN 3-86537-420-4

PREFACE

This volume starts our new series of monographs entitled *REAL studies*. REAL is an acronym for “Research in English and Applied Linguistics”, but it also emphasises the importance of “real” empirical studies based on authentic language material and topical linguistic debate. It thus covers a wide spectrum: from corpus-based to cognitive approaches; from pragmatic to theoretical discussions in lexicography or (second) language acquisition, etc.

This volume reflects current issues in linguistic research and the teaching of English in Saxony and the Czech Republic. In an age when English studies adapt to new challenges in the Bologna framework, such a broad and open survey appears particularly appropriate. It reports recent history and current debate, the changes to effective retraining of teachers in the 1990s, and the reorientation towards more business-oriented subjects that comprise not only topics from language, literature and culture, but also from film studies and the media, particularly the new medium of the internet. Our new series offers a forum for discussion, and we hope that experiences from Saxony and the Czech Republic reported will also encourage our colleagues to continue the dialogue in their own fields.

“English for Central Europe: Saxon-Czech Perspectives” was the first meeting of English specialists from Saxony and the Czech Republic, which was made possible through the “Sächsisch-Tschechisches Hochschulzentrum” (STHZ), a project funded by the European Union and the State of Saxony to encourage the academic exchange between Saxon and Czech universities. This meeting brought together teachers and students from both sides of the *Erzgebirge* for the first time to discuss their experience and opinions. This volume cannot render the lively atmosphere and inspiring discussions that we had during the two-day conference, but it does document the breadth of the issues covered. “English for Central Europe” comprises all aspects of English at university level: language, cultural and literary studies and methodology as well as practical language teaching. The contributors also testify to some of the discussions that have taken place within the universities over the fifteen years since the Velvet Revolution and the fall of the Berlin wall. Consequently, we were aware of the similarities (the pressures of effective financial management as well as the adaptation to the Bologna framework) as well as the differences (whereas the training and retraining of English teachers is largely completed in East Germany, it is still the major emphasis of English at Czech universities).

The conference also shows that English studies at Czech universities are once again developing. However, it has been a long time since the Prague linguistic circle influenced linguistic thinking all over the world (although Josef Vachek and Jan Firbas from Brno were struggling to carry on the great tradition until the present day). But the tradition can now be taken up again in functional as well as constructive perspectives and we hope that it can develop a new “communicative dynamism”.

Of course, English has also changed. Saxony in former East Germany and the Czech Republic may have been less hit by American influences after the Second World War and the postcolonial movement from the British Empire back to the centre; but they have caught up by now and embraced the divergent movements in this vast field of study. Above all, they have experienced the arrival of the internet and in particular the world-wide web, which extends English studies to a whole new dimension.

This conference has taken stock of and, we hope, laid new foundations for better understanding in the future. The interdisciplinary discourse between German/Saxon and Czech university departments has been stimulated by the conference and by the numerous on- and off-topic discussions during the days of the conference, but it does not end there. We look forward to more fruitful co-operation and exchange of ideas in the near future.

Josef Schmied

CONTENTS

PREFACE	i
The Internet and English: new perspectives for language learning and research JOSEF SCHMIED	1
The Development of the Chemnitz InternetGrammar KATRIN VOIGT AND JOSEF SCHMIED	13
Learner Autonomy and Information Technology in language area studies ISABEL HELLER	25
Good manners in East African English: Corpus evidence for a problem in lexical semantics CHRISTOPH HAASE	35
English as a Foreign Language Needs to be Contemporary: A Message for Saxons and Czechs MARIE ERNESTOVÁ	45
Some Linguistics-Oriented Courses and Their Contribution to the Preparation of Future Teachers of English RENATA POVOLNÁ	63
The Rise and Fall of ELT MARY HAWKER	69
Teaching Practice as a Motivational Factor LUCIE BETAKOVÁ	81
Encouraging Teachers of English to Use Problem-Based Learning Strategies with Students of Different Levels and Age Groups ANNA KINOVIČOVÁ	91

Prospective Teachers' Professional Development through E-mail Communication NATALIA ORLOVÁ	103
Student-Centred Teaching of Literature in English: The Application of the Reader-Response Model in the Czech Republic MARKÈTA HEZINOVÁ	111
Teaching and Learning Children's Literature KAMILA VRÁNKOVÁ	117
Louis George Alexander – A Portrait VERONIKA NOVOTNÁ	123
Cultural Anxieties in a Globalised World DANA SEDLMAJEROVÁ	133
Teaching Risky Business Through Fiction: “A Deal in Wheat” by Frank Norris CHRISTOPHER E. KOY	139

The Internet and English: New Perspectives for Language Learning and Research

JOSEF SCHMIED (Chemnitz)

INTRODUCTION

This contribution looks at two trends that have affected English Departments world-wide over the last decade or so.

On the one hand, the pressure of modernisation and cost-effectiveness at universities has threatened English like other humanities since they are perceived as traditional and less immediately profitable than other subjects. The view that English is not a modern subject is based on old notions of the importance of English and of what English specialists do. In fact, English caters for a wide variety of students and interests, as will be shown in the discussion. For centuries, universities have served two functions in varying proportions, allowing some students to acquire a broad general education and others to achieve more specialised job-oriented training. The German contrast between *Bildung* and *Ausbildung* (translatable freely as education and job training) can be rendered only less elegantly in other languages, but the demand remains for English to demonstrate that it can achieve a more immediate relevance than other subjects, bridging university teaching with language applications in schools and beyond. This contribution illustrates that English as a subject encourages students to combine theoretical rigour of thought and practical “employable skills”. On the other hand, the internet and particularly the World Wide Web (WWW) have made many English users and others (sometimes painfully) aware of the world-wide functions of English as a *lingua franca* in international cooperation of government and non-government institutions, of tourist host organisations and guests, of sellers and buyers, and of many others participating in global exchange, particularly between non-native speakers of English. The main development of forms and functions of English nowadays is not among English native speakers but among the rest, since in recent years the number of non-native speakers has surpassed the number of native speakers and this has major consequences for the market value of languages and language teaching¹ (cf. Fig. 1).

¹ The predicted declining trend by L2 speakers for EFL after a few decades has to be seen parallel to the predicted increase of International English. This may change attitudes towards native speakers as models in English teaching and the importance of the WWW as a reference model for (sub-) varieties of English.

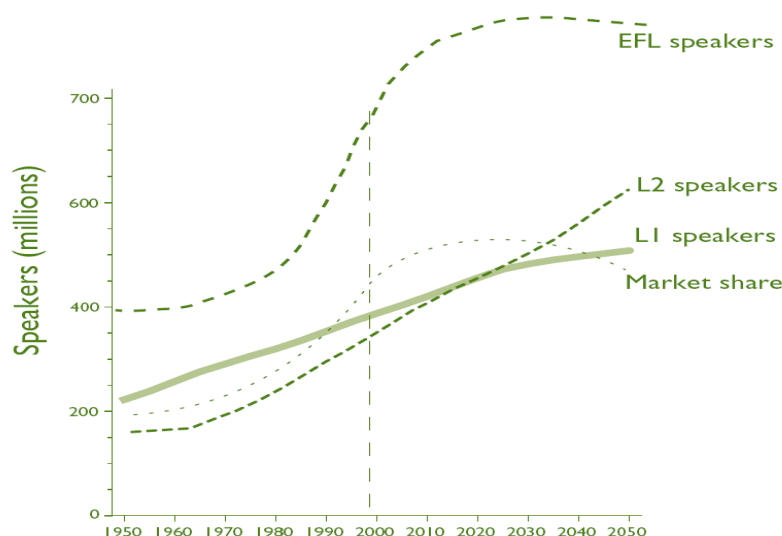


Fig. 1: Estimates of English speakers between 1950 and 2050 (from Graddol 1997: 60)

For English specialists at university level this raises the question whether they can “use” this awareness of globalisation and the internet (“we are the internet subject”) to ensure that their subject is taken seriously and given a minimum of resources, at least, and whether they can use the internet for simplifying and modernising their teaching. This does not mean that English specialists have to teach their students basic internet skills and it does not mean that they take WWW materials uncritically as “the truth”, either. It means rather that they encourage their students to exploit the new opportunities of global communication. This includes:

- to prepare them for the exchange of information via email contacts and to support their critical awareness of languages and cultures involved (including their own),
- to allow and encourage them to supplement their English materials from school with other materials obtained from the internet,
- to enable them to evaluate WWW material by comparing the information and presentation styles with their own experience and by cross-checking key issues on the WWW, and thus
- to combine skills of text production, reception and evaluation, which have always been at the core of language studies.

The following project descriptions illustrate the opportunities and limitations of these new perspectives for language learning and teaching. It does not demonstrate how the internet can be used as a resource, as a classroom tool, as a basis for language-learning activities and as a coursebook (like Teeler/Gray 2000, for instance), but how the internet can lead to a reorientation of the subject, enabling university teachers to demonstrate how “technical” their subject is to administrators and how “cool” their subject is to students.

As usual, the opportunity is at the same time a challenge - not only because it forces English teachers to keep abreast with their students as “web power users”, but also because it challenges old language issues, like the openness and linearity of texts (cf. Schmied 2005). Thus the internet combines surface-skills of immediate use and meta-skills of long-term applications. This helps English specialists to argue that they can fulfil demands not only from those who are more interested in immediate global communicative exchange, but also from those who wish to prepare students for the globalised communication of the future. This future may be as unpredictable as the WWW was only 20 years ago, but the language principles involved can be maintained and adapted.

THE INTERNET AS A DICTIONARY

The idea that the WWW can be used as the great source of information is not new. Even the application to language studies has been common for a long time (cf. Ide/Reppen/Suderman, 2002; Volk, 2001 or Kilgarif, 2001). With the advent of powerful search engines like Google, the search for the occurrence of language forms on the WWW has become part of the daily routine of most professional language users, linguists, translators and other language service providers. For translators, for instance, the site <http://www.multilingual.ch> not only leads to a service provider in Ticino but gives useful hints on exploiting Google to find webpages with expressions in two languages, which can be very useful for rendering similar meanings in another language, be it as direct translations or paraphrases. Such piggybacking on Google has become popular and effective with the help of so-called Google-hacks², which do not use the Google presentation of results but take the Google API format and apply a Perl script to present the figures in a different form. Generally, Google provides three types of information: the frequency of occurrence of the search phrase in the part of the WWW indexed by Google, the URL of the most “relevant” occurrences, and the context lines of the search phrase from the WWW or the Google archives. This can be illustrated by “googling” for an “unknown” word like *palacinky*. It occurs over 7,000 times on the Google-indexed WWW, mostly on pages not in English, but you can easily find one that gives you an explanation and even the price, which also shows you that this may be the American tourist gaze, not necessary the Czech “native” one: “Homemade Crepes - Palacinky Filled with fruit preserves, topped w/ chocolate and whipped cream, \$2.00.”

However, you may overlook a lot of useful (?) spelling variants and even misleading (?) cultural variants:

Entrees of rabbit legs and sirloin of pork were typical Mitteleuropa fare. We couldn't leave Prague without tasting the famous dessert crepes, palacinki, which the chef obligingly prepared and served with fruit, whipped cream and ice cream.

² Calishain/Dormfest (2003) offer a wide array of suggestions and even a webpage with many interesting proposals on <http://WWW.oreilly.com/catalog/googlehks/chapter/>.

And for dessert you will probably be recommended to try the traditional syrupy baklava or different kinds of fine pastries. A favored dessert is the palachinki- crepe stuffed with chocolate or nuts and honey. Good appetite, or, as the Bulgarians say DOBAR APETIT!

Of course, the food and language interface is usually demonstrated with other national dishes (haggis, neeps ‘n’ tatties) or national idiosyncracies (marmite, kippers) in English teaching today. The example demonstrates that Google is not a linguistic tool for analysis and needs to be complemented by other web-extracting software.

For linguists, Webcorp³ has been one of the most useful applications for the last few years – and thus deserves special consideration. Like all searches for lexical strings, Webcorp can only retrieve language forms, it cannot provide specific meanings of polysemous or even homograph words (like the general meaning of *fall* and the specific meaning of *fall* as *autumn* in AmE), but the option to select site, newspaper and text domains helps to restrict search and processing. Site domain refers to the w3c.org categorisation of TLDs (top level domains) that are basically country-specific with well-known exceptions like .tv, which is not Tuvalu in the Pacific, or the US custom of having .com or .edu as world-wide, not specifically American domains, or the almost exclusively American .mil used by the US forces and .gov by the US administration.

Other WWW restrictions are equally language- and style-relevant. The newspaper domains are extremely useful for distinguishing between different types of “journalese”, which have been established in many traditional corpus-compilations and analyses (since the Brown and LOB corpora distinguished between reportage and editorials in the press section, i.e. between informative and persuasive text types, respectively), as well as other genres used in, although not exclusively by newspapers (like narrative texts in short stories or instructive texts, e.g. on gardening).⁴

Webcorp is not only a lexeme- and phrase-retrieval tool that allows a quick selection of appropriate examples, it is also a collocation tool since it offers a presentation of key words in context and a brief statistical matrix showing lexemes in front of and behind the key words, and a list of their most significant collocates (defined here as lexemes that co-occur with each other significantly in natural texts). Of course, collocates depend on meaning and thus they can also be used for word disambiguation. The case of *fall* mentioned above illustrates options and problems: Table 1 below includes not only the adverbs “back” and “forward” and “the”, but also the time-specific collocate “Berlin Wall” and the US-specific “thru” and even “fall waiting” in the collocate list, which is revealed

³ The web interface <http://www.webcorp.org.uk> has been updated recently. Background information can be found on the WWW.

⁴ For a discussion of the classification of the International Corpus of English (ICE) see Schmied (1990). Nowadays, web newspapers provide a wide range of text-types from reportage to letters-to-the-editor and even related chats and background links.

by a click through to the concrete sentences as occurring in the college context of “fall waiting lists”. Both occur in the calculation because obviously .com WWW pages have been included in the search. The semantic continuum from BrE (*drop*) to AmE (*autumn*) can be seen in the collocation “leaf fall”. “Rise” demonstrates the collocation with the antonym and the problem of capitalisation at the same time.

Word	Total	L4	L3	L2	L1		R1	R2	R3	R4	Left Total	Right Total
rise	12			11				1			11	1
back	9			2			7				2	7
Rise	7			5			1			1	5	2
cent	7				2				1	4	2	5
forward	7			2			3			2	2	5
thru	7				1		4			2	1	6
Wall	6									6	0	6
Berlin	6								6		0	6
love	6							4	2		0	6
Scotland	5	1						2		2	1	4
leaf	4				4						4	0
education	4			3					1		3	1
shoe	4			4							4	0

leaf fall concerns fall fall back fall thru fall waiting

Table 1: External collocates of “fall” in domain .uk (excluding stopwords)

The list from the American .gov administration (List 2) obviously includes not only variety-specific cases like *fall color(s)* and *fall foliage*, but also the geriatric *fall hazard/risk*, and *fall protection/prevention* for soldiers and *ash/rock fall* after explosions, etc.

ash fall rock fall substantial fall personal fall free fall fall protection fall arrest
fall color fall prevention fall foliage fall colors fall hazards fall risk fall arresting
fall injuries fall 2002

List 2: External collocates of “fall” in domain .gov (excluding stopwords)

The corresponding military list (3) unsurprisingly displays similar collocates, but also the rather common and long-expected *free fall* and *fall asleep*.

free fall fall protection fall asleep fall transition fall hazards fall arrest fall
meeting fall back

List 3: External collocates of “fall” in domain .mil (excluding stopwords)

None of the lists above contains *fall short of* or *fall through*; *fall flat*, *fall sharply* and *fall dramatically*; to say nothing of *fall by the wayside*, *fall into place* or *fall between two stools*, which can all be expected in a good collocation list in a learner dictionary. *Fall for* and *fall out* occur in the general Webcorp list, however. Obviously, a lot of computer linguistic work still has to be done to measure strengths of collocation in larger databases.

FREQUENCY, NORM AND VARIATION

The applications of using the WWW as a domain-specific frequency dictionary are abundant. Of course, the frequency of certain phrases in Google varies, since new pages are added (or taken out) all the time. For variation specialists, Webcorp is interesting because it can distinguish between individual native- and non-native speaker countries (i.e. domains), but it does not display the results in comparative tables.

WebPhraseCount (WPC) uses the same principle as Webcorp, but it offers a more quantitative approach, since it presents the statistical results in comparison. It uses Google Web APIs, which have been made available recently in a beta version. This service enables programmers to develop software that accesses the Google web index, but it is limited to 1,000 queries a day.

Through a combination of queries we can either do a WPC multi-phrase search by entering two or more search phrases and comparing their occurrence in a certain domain, which is usually culture-specific (.uk for Britain, .au for Australia or .za for South Africa, for instance), or a WPC multi-domain search for the same phrase in different domains. From a linguistic perspective the first use is more language learner (EFL) related, i.e. we can find out whether a certain expression is used more commonly than another; this enables us to verify or falsify collocations, for instance. The second use is rather variation (ESL) related, i.e. we compare phrases in different countries, so that we can distinguish expressions that are country-specific (or region-specific like *ugali*) from world-wide expressions of English. Collocational analyses can be pursued in lexis and in grammar.

As an example, for the non-native speaker the question of prepositions after verbs or adjectives can be quite tricky. Dictionaries often indicate only standard prepositions after lexemes (like *different + from*) and may happen to show specific usage in sample sentences, but they hardly comment on frequencies or on variety and style differences. Thus, a simple WWW search for *different from*, *different to* and *different than* shows that only a fraction of the over 92,000,000 occurrences of *different* is followed by a preposition. *From* is chosen in 71 %, *to* in 8 % and *than* in 20 % of the cases; but these proportions are distributed quite differently in the British, Australian, Canadian, US Government and South African domains. The statistical measurements *intrasite share* and *intrasite*

factor offer help in interpreting the results. Of course, *different from* is the standard collocation, whereas *different than* is clearly American usage. This becomes particularly clear when we look at the intrasite factor, which takes the least frequent search phrase as 1 and displays a factor that indicates how many times more frequent the others are, because in all three North American domains *different than* is used more often than *different to*.

Similarly, a search for *Tesco is / has* in contrast to *are / have*⁵ reveals very quickly that in two thirds of the cases the superstores are considered as singular, since they would like to appear as a unit that gives a standard appearance to its customers (Table 4). However, in some cases they would like to appear as a multifaceted and ubiquitous entity, which may be brought across more easily in the plural. But this does not explain why (the much less frequent) *does/do* does not follow that pattern nor why this is less extreme with other superstore names (in Table 5).

no.	phrase	total sites	intra-phrase percentage	inter-phrase share
1a	Tesco has	1,250	0.01 %	30.34 %
1b	Tesco have	596	0.005 %	14.466 %
2a	Tesco is	1,520	0.012 %	36.893 %
2b	Tesco are	581	0.004 %	14.102 %
3a	Tesco does	65	0 %	1.578 %
3b	Tesco do	108	0.001 %	2.621 %

Table 4: Singular vs. plural verb forms after TESCO

no.	phrase	total sites	intra-phrase percentage	inter-phrase share
1	Tesco has	1,250	0.01 %	40.349 %
2	Tesco have	596	0.005 %	19.238 %
3	Sainsbury's has	279	0.002 %	9.006 %
4	Sainsbury's have	198	0.002 %	6.391 %
5	Safeway has	324	0.002 %	10.458 %
6	Safeway have	171	0.001 %	5.52 %
7	Waitrose has	176	0.001 %	5.681 %
8	Waitrose have	104	0.001 %	3.357 %

Table 5: Singular vs. plural verb forms after UK superstore names

⁵ The most frequent verb forms have to be used, because other verbs like *want*, *need*, etc. do not occur often enough for a rewarding WWW analysis.

LIMITATIONS

The Web as a corpus?

Corpus linguists know, of course, that the results of an analysis depend largely on a well-stratified sample. Even if the name Webcorp, for instance, suggests that the Web can be used as a corpus, it does not claim that it is actually a corpus in the sense that it is stratified according to productive or receptive usage or according to similar principles in all TLDs. Since it seems not quite clear whether the World-Wide Web is really stratified enough to represent a certain national variety, for instance, we have called our local research tool WebPhraseCount (WPC), which is more neutral and emphasises that the statistical procedure is more important than finding individual occurrences. Many corpus linguists have pointed out that such simple solutions are tempting because they are fast and provide masses of data, but they are also aware that these data have to be checked thoroughly, because they are always skewed in some direction. Understandably, putting cookery recipes up on the internet is not perceived an urgent need in Africa, in contrast to Europe or the US. Thus web texts are clearly culture-specific and still the web does not reflect reality or language use as a whole.

Statistical operations

Of course, the occurrence or even absolute frequency of a phrase is much less interesting than its occurrence in proportion to other phrases or in other domains. Unfortunately, the size of domains varies a lot, and for small countries like Malawi or Tanzania, the statistical basis for more than the most general words is not big enough yet. Since WPC subcategorisation is based on TLDs, the user must be familiar with them or be able to consult a list easily. Thus WPC users are able to simply click on a map to obtain automatically the TLD attributed by the WWW consortium.

Technical limitations

The program developed is relatively quick and efficient, but a few restrictions apply. The total number of sites varies constantly by as much as three percent, but this should not affect the statistics' outcome since the search words should not necessarily occur on "unstable" pages more frequently than on others.

The major limitation of WPC is imposed by Google, which specifies that not more than 1,000 search results can be obtained by a single user per day. This quota is of course used up quickly if, for instance, more than ten multi-domain complex searches with ten phrases and ten domains are carried out. Thus with WPC search results are stored in a cache and irrelevant options can be clicked away easily, so that the analyst does not have to submit all the queries again, if only a few provide unsatisfactory results (e.g. because the figures are too low).

By this method, changing display options does not use up the limited queries available.

Web language limitations

Finally it is worth summarising the limitations of the web that are specific to language studies. As mentioned above, the frequency of an English phrase can be distorted if it also occurs in other languages, even in domain sites where the vast majority of texts is expected to be published in English (e.g. information in French, which has many forms identical with English).

Whether the comparison of non-standard English in English-language web pages from non-ENL countries like Italy or Spain is very useful has to be seen. Contrastive (parallel) texts may be analysed if the language marker “+the” is replaced by others (e.g. the German articles). The relative frequency of equivalent forms like *church*, *uncle*, *school marks*, etc. could also be used in cross-cultural studies.

Obviously, the WWW is not a corpus in the sense that it is stratified according to everyday language usage. If we use it because it is so nicely machine-readable and available, we have to be aware of the text types that it is biased towards or against. The most obvious is that oral communication is completely missing, but this may be compensated for partly by including texts from discussion groups, email, chats and other exchanges outside the WWW. Then, the WWW is obviously a public domain, thus more personal pages are rare; even “personal, private” homepages are public, sometimes even exhibitionistic – and rather culture-specific. This does not only reflect “development standards” but partly also functional domains in multilingual societies, where English may only be used in more formal contexts and hence shows, as a second language, much less variation and specificity than in first-language usage.

A CORPUS FROM THE WEB?

If the ICE corpora are too small and the web is too “untidy”, maybe a compromise can be found. If we concentrate on text-types that are available world-wide, like different types of newspaper texts (cf. above), economic texts (like marketing and company-related texts) or academic web pages (with scholarly articles as well as student advice), we may be able to replicate the WWW with a corpus-like stratification. Such a WWW corpus would normalise the size differences between TDLs and the text-types included. This corpus could be used as a monitor corpus, against which other search results from smaller or supposedly uneven corpora could be measured. Thus the national ICE corpora of one million words will soon prove too small for collocational analyses (Schmied 2004); but a quick check in WPC and our WWW corpus would verify or falsify hypotheses quickly enough.