

# **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 teaching<sup>1</sup> (cf. Fig. 1).

---

<sup>1</sup> 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.

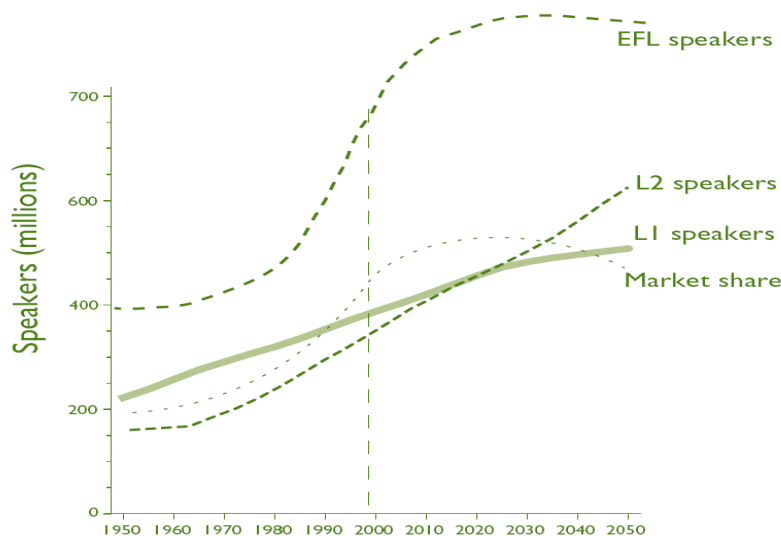


Fig. 1: Estimates of English speakers between 1950 and 2050 (from Graddol 1997: 60)

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-hacks<sup>2</sup>, 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.

---

<sup>2</sup> 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, Webcorp<sup>3</sup> 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).<sup>4</sup>

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

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

<sup>4</sup> 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)