

# DICTIONARY News

## KD's BLDS: a brief introduction

In 2005 K Dictionaries (KD) entered a project of developing dictionaries for learners of different languages. KD had already created several non-English titles a few years earlier, but those were basic word-to-word dictionaries. The current task marked a major policy shift for our company since for the first time we were becoming heavily involved in learner dictionaries with target languages other than English. That was the beginning of our Bilingual Learners Dictionaries Series (BLDS), which so far covers 20 different language cores and keeps growing.

The BLDS was launched with a program for eight French bilingual dictionaries together with Assimil, a leading publisher for foreign language learning in France which made a strategic decision to expand to dictionaries in cooperation with KD. The main target users were identified as speakers of French learning a foreign language at an intermediate level. To serve both production and reception purposes, it was decided that the dictionaries would be fully bilingual and bi-directional. The core for each language consisted of about 12,000 main headwords, with the total number of words and expressions running up to a maximum of 20,000. The entries typically consisted of the headword, pronunciation, part of speech, irregular forms, brief definitions, example of usage for each sense of polysemous entries, all types of compositional phrases, run-ons, usage and register labels, etc.

We started by establishing an extensive infrastructure both contentwise and technologically. The lexicographic compilation was divided into 25 projects: 1 for the French core, 8 for the dictionary cores of the eight languages, another 8 for the translation from French to eight languages, and 8 more for the translation of each language to French.

An editorial team was set up for developing each of the nine (French + 8) dictionary cores. The lexicographers worked from a distance, usually at home, all over the world. The chief editor for each language was responsible for preparing the editorial styleguide and the list of headwords. Since no corpora were publicly available for any of these languages, each editor used different means to retrieve information in order to compile the headword list. Much of the language resources for compiling the entries were found on the Internet.

The technical foundations had to be robust and elaborate enough to accommodate complex lexicographical data in any language, fitting into a single overall framework that allowed for regular exchanges with dozens of editors and translators. We configured an XML Editor for the lexicographic input and had to maintain a sub-project of having the software installed by each lexicographer, preparing proper documentation on its functioning, and providing technical support.

In its larger sense, the BLDS concerns constructing a

### 1 KD's BLDS | a brief introduction

### 3 Papers on Lexicography and Dictionaries

1 | *Perspectives in Lexicography: Asia and beyond*. Vincent B.Y. Ooi, Anne Pakir, Ismail S. Talib and Peter K.W. Tan (eds.)

2 | *DSNA Seminar on Learner's Dictionaries*. Ilan J. Kernerman and Paul Bogaards (eds.)

### 4 The Lexicographers | Michael Manis

Lexicographical poems: an introduction | Michael Adams

### 6 B.T. Sue Atkins and Michael Rundell. *The Oxford Guide to Practical Lexicography* | Rick Schutz

### 8 Seminar on Learner's Dictionaries at the DSNA XVII Meeting, Indiana University, 2009

### 9 A review of the First International Symposium on Lexicography and L2 Teaching and Learning | Zhang Yihua and Xia Lixin

### 11 Asialex 2009 | Jirapa Vitayapirak

### 12 The Center for Lexicographical Studies, GDUFS

### 14 English Dictionaries in Global and Historical Context | Janice McAlpine

### 16 Internet lexicography as a challenge: The Internet dictionary portal at the Institute for German Language | Stefan Engelberg, Annette Klosa and Carolin Müller-Spitzer

### 26 KirzeN, the international dictionary | Balázs Sipócz

### 27 TheFreeDictionary.com story | Michal Goldman

### 28 The BLDS Italian dictionary core | Palma Gallana

### 31 BLDS | sample pages

### 32 Dictionnaires Assimil Kernerman, 2009

Editor | Ilan J. Kernerman



KDICTIONARIES

© 2009 All rights reserved.

### K DICTIONARIES LTD

Nahum 8 Tel Aviv 63503 Israel

Tel: 972-3-5468102

Fax: 972-3-5468103

kdn@kictionaries.com

http://kictionaries.com

The BLDS cores currently include these languages:  
 Arabic • Chinese (Simplified • Traditional) • Czech • Dutch • English • French • German • Greek • Italian • Japanese • Korean • Latin • Norwegian • Polish • Portuguese (Brazil • Portugal) • Russian • Spanish • Turkish

database-network that can serve as a base for compiling different dictionaries. It contains more data than is needed for a single dictionary, and enables the extraction of precise components to suit individual user profiles. For example, the titles for speakers of French use sense indicators for disambiguating meanings instead of carrying the definitions, and exclude the phonetic pronunciation of the French headwords but present phonetics of the L2 headwords and translations. The appropriate balance naturally has to be fine-tuned for

each particular language pair.

Meanwhile, KD began to compile dictionary cores for more languages, and the first BLDS title was actually published already last year by Vega Forlag in Norway (Spanish/Norwegian). The eight French bilinguals are published by Assimil in France this year. Each dictionary is accompanied by an electronic application that is downloaded and activated by a personal key. This issue of KDN presents a few BLDS samples, more extensive information will be provided in the next issue.

# французско-русский

909

hôte

**hormis** *prep* *littér* кроме ['kromə], за исключением [zə iskl'u'tʃ'en'ijəm] *Il ne voit personne, hormis toi.* Он не видит никого, кроме тебя.

**hormone** *nf* гормон [ga'rmon] *m*, *prendre des hormones* принимать гормональные препараты

**horodateur** *nm* устройство для проставления даты и времени ['dati i 'vr'lem'ɛn'ɛ]

**horoscope** *nm* prévisions sur l'avenir гороскоп [gəra'skop] *m* *lire son horoscope* читать свой гороскоп

**horreur** *nf* ужас ['uzəs] *m* *1 sentiment* *Il pousse un cri d'horreur* вскрикнуть от ужаса

*2 ♦ avoir horreur de* ненавидеть *Il a horreur de la violence.* Он ненавидит жестокость.

*3 atrocité* *l'horreur d'un crime* тяжесть преступления *un film d'horreur* фильм ужасов

**horrible** *adj*

*1 effrayant* ужасный [u'zasnɛ] *une histoire horrible* ужасная история

*2 laid* некрасивый [n'ikra'sivɛ], безобразный [b'iza'braznɛ] *Cette robe est horrible.* Какое ужасное платье!

*3 mauvais* очень плохой [o'tɛ'pɛ] *plaisant*, ужасный *Il fait un temps horrible.* На улице скверная погода.

**horrifier** *vt* terrifier, épouvanter ужасать [uza'sat] *Ce drame a horrifié tout le monde.* Это трагическое событие повергло всех в ужас.

**hors** : **hors de prép**

*1 en dehors de, à l'extérieur de* вне [vɛ'ɛ], из [iz] *hors de la maison* вне дома

*2 ♦ être hors de danger* быть вне опасности

*3 ♦ hors d'usage* недействующий, вышедший из употребления *Le grille-pain est hors d'usage.* Тостер вышел из употребления.

*4 ♦ hors de prix* очень дорогой *Ces chaussures sont hors de prix.* Эти туфли слишком дорогие.

**hors-d'œuvre** *nm* *inv* plat servi en début de repas закуска [za'kuskə] *f*

**hors-jeu** *nm* *inv* sport вне игры [vn'i:'gri]

**hors-piste** *nm* *inv* катание на лыжах [ka'tan'ɛje na 'li:zə] вне установленной лыжни [vn'i usta'nov'linə] *faire du hors-piste* кататься на лыжах вне трассы

**hors service** *adj* *inv* plus en état de fonctionner недействующий [n'i'd'ejstvuʒij], вышедший из строя [vɛ'fɛtɛ'ɛ i's:tʁoʒə] *un ascenseur hors service* лифт не работает

**hospice** *nm* приют [pr'i'jut] *m*, дом престарелых [dom prɛ'sta'r'elɛx] *m*

**hospitalier, -ière** *adj*

*1 relatif à l'hôpital* больничный [ba'l'n'i'tɛ'niʒ] *un établissement hospitalier* больничное учреждение

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hospitalier**

**hormis** ['ɔʁmi] *prep* *littér* sauf *Il ne voit personne, hormis toi.*

**hormone** [ɔʁmɔn] *nf* substance active du corps *prendre des hormones*

**horodateur** [ɔʁɔdatœʁ] *nm* machine qui imprime la date et l'heure

**horoscope** [ɔʁɔskɔp] *nm* {*prévisions sur l'avenir*} prévisions sur l'avenir des personnes *lire son horoscope*

**horreur** [ɔʁœʁ] *nf* *1* {*sentiment*} violent sentiment de peur ou de dégoût *pousser un cri d'horreur*

*2 ♦ avoir horreur de* détester *Il a horreur de la violence.* *3* {*atrocité*} ce qui est horrible *l'horreur d'un crime* *un film d'horreur*

**horrible** [ɔʁibl] *adj* *1* {*effrayant*} qui fait très peur ou dégoûte *une histoire horrible* *2* {*laid*} très laid *Cette robe est horrible.* *3* {*mauvais*} très mauvais *Il fait un temps horrible.*

**horrifier** [ɔʁifje] *vt* {*terrifier, épouvanter*} provoquer un violent sentiment de peur ou de dégoût *Ce drame a horrifié tout le monde.*

**hors** : **hors de** ['ɔʁdə] *prep* *1* en dehors de, à l'extérieur de *hors de la maison* *2 ♦ être hors de danger* ne plus être exposé au danger *3 ♦ hors d'usage* qui ne fonctionne plus *4 ♦ hors de prix*

A sample page from the French-Russian section of *Dictionnaire Assimil Kernerman Russe-Français Français-Russe* (on the left), and an extract from its raw French data (below).

# PAPERS ON LEXICOGRAPHY AND DICTIONARIES

1

## *Perspectives in Lexicography: Asia and beyond*

Vincent B.Y. Ooi, Anne Pakir, Ismail S. Talib and Peter K.W. Tan (eds.)

ISBN 978-965-90207-1-3

290 pages

K Dictionaries, Tel Aviv. August 2009

Introduction • Vincent B.Y. Ooi

### **Asian perspectives**

Lexical concerns about neologism in Chinese lexicography

• *Yihua Zhang*

A survey of pragmatic information in bilingual

English-Chinese learners' dictionaries • *Saihong Li*

Grammatical class, tags and lemmas: A corpus-based study of the Malay lexicon • *Gerry Knowles and Zuraidah Mohd Don*

Analysing sorrow and grief: A contrastive-semantic perspective • *Ruihua Zhang*

Lexical innovations in the multimodal Corpus of Asian

Magazine Advertising • *Andrew Moody and Azirah Hashim*

### **Pedagogical perspectives**

EFL dictionaries on the Web: Students' appraisal and issues in the Cambridge, Longman, and Oxford dictionaries

• *Shigeru Yamada*

The potential of learner corpora for pedagogical lexicography • *Yukio Tono*

A trilingual dictionary for learners of Cantonese, English and Putonghua

• *Jacqueline Kam-mei Lam, Lan Li and Tom McArthur*

Electronic dictionaries in the classroom

• *Shinya Ozawa and James Ronald*

The benefits of CD-ROM dictionaries in teaching

• *Monika Szirmai*

Dictionary use strategies by EFL learners in Taiwan

• *Da-Fu Huang*

English learners' dictionaries: An undervalued resource

• *Julia Miller*

Multiple word class entries in advanced learner's

dictionaries of English • *Sadayuki Nakane*

Representation of word combinations in illustrative

examples in English learners' dictionaries • *Hai Xu*

Not quite first language, not quite second language either:

Dictionary entries for learners caught in between

• *Peter K.W. Tan*

### **General perspectives**

Interlingual lexicography, with special reference to research priorities • *R.R.K. Hartmann*

Internet-based communication and the ecology of dictionaries • *Wengao Gong*

Relational network notation and the intelligent web

• *Jonathan J. Webster and Ian C. Chow*

Putting the corpus into the dictionary • *Adam Kilgarriff*

Towards a multimodal dictionary of narration

• *Ismail S. Talib*

2

## *DSNA Seminar on Learner's Dictionaries*

Ilan J. Kernerman and

Paul Bogaards (eds.)

Publication by K Dictionaries,

Tel Aviv. December 2009

Introduction

• *Ilan J. Kernerman*

The evolution of learners' dictionaries and the *Merriam Webster Advanced Learner's English Dictionary*

• *Paul Bogaards*

I've heard so much about you: Introducing the native-speaker lexicographer to the learners' dictionary

• *Wendalyn Nichols*

From the ground up: Making a new learner's dictionary from scratch

• *Peter Sokolowski*

Three collegiate dictionaries: Preferences by college students: A comparison of reading comprehension test scores and survey results for university students using three collegiate dictionaries: MW, AHD, and NOAD

• *Don R. McCreary*

FEARful words in learners' dictionaries

• *Henri Béjoint*

Folk defining strategies vs. comprehension of dictionary definitions: An empirical study

• *Mateusz Fabiszewski-Jaworski and Marta Grochocka*

Corpus, lexicography and language learners

• *Mari Carmen Campoy-Cubillo*

Why we need bilingual learners' dictionaries

• *Arleta Adamska-Sałaciak*

Writing a bilingual learner's dictionary:

A case study of Kirundi

• *Teddy Bofman, Jeanine Ntahirageza and Paul Prez*

An ideal EFL dictionary environment

• *Shigeru Yamada*

What future for the learner's dictionary?

• *Michael Rundell*

Proceedings of the Seminar on Learner's Dictionaries held during the DSNA 17th Biennial Meeting at Indiana University, Bloomington, on 29 May 2009.

Orders from: [www.CreateSpace.com](http://www.CreateSpace.com)

# The Lexicographers

Michael Manis



**Michael Manis** was born in Akron, Ohio and is currently a graduate student at Indiana University. He is perusing his Master of Fine Arts in creative writing with an emphasis in the short story, though he also writes poetry. His interests include the persona poem, the history of lexicography, and the craft of the historical fictional narrative.  
mjmanis@indiana.edu

---

## Johnson Upon the Priapism of Lexicography

"A preternatural tension"—ah yes between our throats creating sounds endless and the pen's wet running upon the page. Such early morning thoughts arise to wag my confidence at my bedside as if to tantalize, to offer and then stiff me of my resolve to suffer English fixed like a tent pitched to serve whim and wish of Pope and Dryden. Then, outside my home, the chatter rises from the street, the drone, the hum, which should comfort my stultified, wooden mind and limpen the stature of my worries. But the muffled sounds that shove their way into my room causes worry once again to bloom. I close my ears and try to train my mind on Pope's poems, but against the street, not all Pope's verses could stay defeat. There my dictionary stands on a shelf and outside the language busies itself. Who am I but an old man with an ear through which to soften the stubborn fear that men do not need monuments erect to all Shakespeare, Dryden, and Pope's affect.

---

## Murray On His Pigeon Holes

Today I look up from my work, and I am astounded by the pigeon-holes lining the scriptorium wall like wooden pores, from which the language exhales its secrets. I have spent near half my life listening to the lungs of English, measuring its intervals. I wonder if those pores will swallow me one day and inside there sits my doppelganger, waiting for his chance to make sense of the last of my breaths.

---

## Picture of a Lexicographer

Words  
alone  
make  
for  
lonely  
men.

---

## Memo One: From Philip Gove to America

Ain't is a word like any other, though substandard, but all words are substandard in their own way.

For example, I came home late one evening and found my house empty except for strangers reading stacks of memos like this one—all trying to figure out who it was that lived here.

---

## Memo Two: From Phillip Gove to the Copy of Webster's Third on His Desk

Congratulations for being finished, though I don't mean "finished" in the completed sense. I mean through or dead or irrelevant. We both know you'll never be finished in the first way but you'll be forever finished in the second.

Last night, I dreamt I opened you. Inside: all the bones of anyone who ever tried to say anything and mean it.

### Lexicographical poems: an introduction

Indiana University has an outstanding graduate creative writing program. Some students in that program were members of a seminar I taught last spring, titled "Reading Dictionaries." While doctoral students wrote traditional research papers at the end of the term, the creative writers were encouraged to find a way to make sense of the course material in fiction or poetry. Michael Manis, primarily a writer of fiction, submitted the poems published here. In them, he conveys memorably the burdens and joys of lexicography. I am aware of only a few poems devoted to dictionaries and their makers; it seems fitting that *Kernerman Dictionary News*, devoted as it is to the art and craft of lexicography, would once in a while publish art about lexicography, too. Through lyric, we see our familiar world with unexpected clarity; we are reminded of the humanity that underlies all lexicographical enterprise.

**Michael Adams**  
Indiana University

---

## Dictionary as Lexicographer

If Webster's Third were a lexicographer,  
she would be a tall woman with  
glasses—popular glasses like something  
Sarah Palin would wear. She would speak  
eloquently, though in long-winded sentences  
that lose people's attention at parties, but  
those people would invariably believe  
she was very smart, despite the fact  
that she was known for repeating herself.

If the OED were a lexicographer,  
he would be a skinny man in his late forties  
prone to wearing tweed suits and bow-ties.  
At parties, the OED would entertain with  
trivia that, while precise, would annoy the guests  
by the sheer weight of it. Once the OED  
gets talking, he will never stop.

If the American Heritage Dictionary were a lexicographer,  
he would be a portly man in his thirties,  
though he would dress as if he were in his twenties:  
casually, with no tie, open collar and camel hair blazer.  
He would be jealous of the attention Webster's gets  
and would constantly look out for the politically  
incorrect, if only to butt his way into a conversation,  
which was hard enough to do with the OED around.

---

## Gove's Wife, Cataloguing Johnson's Cards

The words on the cards,  
like fossils buried in bedrock—  
they are easy to disregard.

Many words come from the bard  
and some vulgarities are a shock  
to find on the cards.

Most people I show are dullards  
and care for no words and balk,  
but they are easy to disregard.

The quotations are like barnyards,  
here a cow and there a cock  
domesticated on the cards.

The tedium makes it hard,  
and Philip does not take stock—  
my work, he disregards.

I see my husband in Johnson's awkward  
script—small letters like pocks  
fall into line on the cards,  
which I find difficult to disregard.

---

## Johnson Upon Craigie

"It was a railway porter"—what have I  
with my methods brought about? Is it dry  
witticism, for surely no man thinks  
that the object here is by its link  
to existence (ah "it was!") distinguished  
from the atoms composing a bonefish  
or those of the ink running black over  
the presses of Grub Street or of Dover?  
Perhaps I mistake the nature of this  
porter and it is truly formal bliss  
the likes of which might leave Plato pleased  
to see exist. Or better Democrates,  
attempting to divide the object in twain,  
might find the process trying and refrain  
from attempting to define—excuse me—  
divide it any further. Bad to lose  
a pure proof of ancient-logic to those  
poets who'd sentence divinity's word  
to context low! Perhaps to protect and gird  
it best, the lexicographer chose to  
leave this sacred porter alone in lieu  
of finding it a wordly epitome,  
which the language in brief could not condone,  
or if it could, why can't the writer shirk  
his readers to do philosophy's work?

---

## Burchfield on Nonce words

The ground beneath  
the tree of words grows  
softer with common rinds,  
and I do the work of  
a botanist, picking up here  
a rough and yellowed peel  
and sketching the seeds  
still stuck to the husk  
on my steno-pad with an  
old fountain pen.

Have not I earned, in all  
these long years among the rinds,  
the swarming flies, the sweet  
and sticky juices, the exposed,  
oxidized flesh in browns  
—I know I've earned the right  
to climb the tree and pick  
the fruit that never falls but  
is so light, it evaporates  
into the rarefied air.

---

## Memo Three: Philip Gove to His Employees

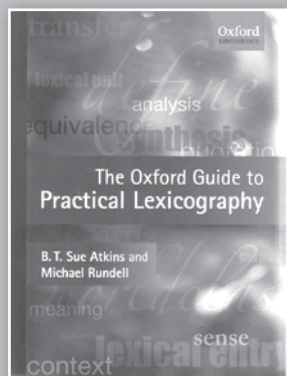
Murray died before  
he got to see the  
complete OED,  
and I began to  
wonder—if I hired  
one of you to  
write these memos  
for me, could I  
embark on a voyage  
to somewhere warm  
and distant?

None of you  
harmless drudges  
would ever know.



## B.T. Sue Atkins and Michael Rundell.

### *The Oxford Guide to Practical Lexicography*



The Oxford Guide  
to Practical lexicography  
B.T. Sue Atkins and Michael  
Rundell  
Oxford University Press, 2008  
978 0 19 927770 4  
(hardbound)  
978 0 19 927771 1 (paperback)

#### Double practical

In a previous edition of this publication, I observed that the book *Practical lexicography, a reader*, edited by Thierry Fontenelle, was maybe not as practical as the title suggested (KDN16, 2008: 14-16). In any case, it did not give the answers to practical questions like:

- Under which entry does the user of my dictionary find fixed phrases and idiomatic expressions?
- What does the blueprint of an empty dictionary look like? Which building blocks are universal and essential?
- Where can I find information on tools/software to build a dictionary?

A little later in 2008 OUP published a companion book with a nearly similar title, *The Oxford Guide to Practical Lexicography* (OGPL). This title is well chosen indeed: the book does discuss, and in many cases provide, the answers to, these questions and many, many more.

#### Very, very practical

After a brief and crystal clear introduction that tells us what the book is about, what lexicographers do and how the book works, the first chapter begins with 'The birth of a dictionary'. In three major parts, *Pre-lexicography*, *Analysing the data* and *Compiling the entry*, OGPL guides us – lexicographers to be – along practically every issue that anyone who decides on making a dictionary encounters. On page 499, we are ready for the final paragraph 'Writing the entry'.

The style of the book is pleasantly down to earth and on every page one senses the abundant experience of the authors. The basic piece of advice, repeated throughout, is: keep the user of your dictionary in mind with every decision you make. That seems an obvious thing to do, but many existing dictionaries prove that somewhere along the road the maker forgot about the (limitations in the) skills and needs of the users.

The practical approach of OGPL shows in the many remarks, often pieces of very practical advice, preceded by a blue arrow. A few examples:

“Don't rely too much on labels in your entry: they usually mean more to you than to the user.” and “When you're deciding how to handle MWEs (multiword expressions), it's a good idea to look at a lot of other dictionaries, think

about your user profile, then choose the way that best fits the needs of your most vulnerable user.”

Here and there, the reader gets an explicit warning, like: “Building the database is a wholly monolingual exercise” and “When in doubt, don't leave it out.” Each chapter ends with references to more information, divided into 'Recommended reading', 'Further reading and related topics' and 'Websites'. The lists with references are short and up to date.

A quote from a paragraph that discusses grades of idiomaticity may illustrate the stress OGPL lays on dictionaries as realistic inventories of current words and phrases. The following remark is made on idioms that present no problems of identification as true idioms:

“...these and only these are used to illustrate the many papers on idiom by theoretical linguists, who are single-handedly keeping alive old favourites like *to rain cats and dogs* and *to kick the bucket*. It is a very long time since either of us heard these in day-to-day discourse.” (note 13, p. 360).

Apparently the old favourites are hard to suppress, since they also show up in OGPL on page 167 (*kick the bucket*) and page 181 (*raining cats and dogs*) respectively.

#### Headword of a multiword expression

It would be impossible for this review to comment on every aspect of dictionary making that OGPL deals with. I focussed on the answer to the very first question in the beginning of this review: “under which entry does the user of my dictionary find fixed phrases and idiomatic expressions?”

In my lifetime, I have spent a lot of time looking in vain for multiword expressions (MWEs) under the wrong headword. Very few (English) dictionaries bother to mention the way they distribute MWEs among entries in the front matter. Consequently it is impossible for a motivated (front matter reading) user to save time by following the lexicographer's system, however counter-intuitive it might be.

In a recent concise dictionary by a renowned editor,<sup>1</sup> the distribution of MWEs

1 *Essential Dictionary for Learners of English*, edited by Michael Rundell, Macmillan, 2007.

was probably not an item in the styleguide. The same MWE occurs in several articles under different entries, obviously without intention. The lexicographical treatment varies in wording, labeling, definition and the use of examples. Here are two such doubles (with the entry word printed in bold):

- at/in the **back** of your mind: if something is at the back of your mind, you are slightly conscious of it all the time

- at the back of your **mind**: if something is at the back of your mind, you are not thinking about it now, but you still remember it or know about it: *At the back of the mind, she knew he was lying.*

- take sth with a **pinch** of salt: *informal* to doubt that something is completely true

- take sth with a pinch of **salt**: to doubt the truth or value of something is completely true: *If I were you, I'd take his advice with a pinch of salt.*

Would consultation by this dictionary's editor of his later work have prevented this? The disappointing answer is probably no.

OGPL discusses the relevant question 'under which entry does the user of my dictionary find fixed phrases and idiomatic expressions?' in chapter 9.2.6, 'Multiword Expressions'. The authors refer to "a lot of academic research with a view to discovering where dictionary users expect to find various types of MWE". The only more or less concrete statement is "it is often said that German users will look for a phrase first under the noun". In fact, the question is left unanswered. OGPL then offers a number of possible strategies for the lexicographer to follow ('the principal options for English'):

- 1 under the first or only lexical word
- 2 under the least frequent lexical word
- 3 under the first or only noun
- 4 under the first or only verb
- 5 as a headword in its own right (in dictionaries of idioms)

In the rest of OGPL there is no sign of the application of any of these options, nor of any other strategy. Further on, in chapter 10.2.1 OGPL discusses the tricky issue of the relative fixedness of many expressions. For example, when we look at 'with a pinch of salt' in a large corpus we may find many instances without the word 'pinch', like *with a huge lump of salt*. The authors argue convincingly that the canonical form in a dictionary should nonetheless be 'with a pinch of salt'. However, without any clarification they present *salt* as the entry for this MWE in the dictionary. Yet *salt* is neither the first or only lexical word, nor the least frequent lexical word, nor the first or only noun, nor the first or only verb.

### Automated selecting and ordering

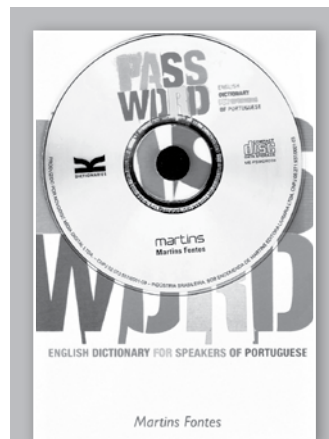
It strikes me that the frame of mind of the authors of OGPL still seems to be the traditional printed dictionary, in spite of the distinction in the very structure of the book between building the database and editing the final product. The authors even say (on p. 363) "What you do with an MWE in the database is not necessarily what will eventually be done with it in the dictionary proper." So much for the theory.

In fact, the book is suffused with the spirit of the traditional idea that every single dictionary article is the creation of a skilled craftsperson. The way Atkins and Rundell describe the arrangement of meanings and MWEs in a dictionary entry smells like arranging flowers in a bouquet. The fundamental distinction between database and resulting product is mentioned, but is neither discussed nor illustrated and thus seems to be of little substance.

In their own words, the authors state that the transformation from database to dictionary essentially consists of selecting and ordering. The notion that if the database is well structured, it can be turned into a dictionary – into a number of different dictionaries – by using a range of programmed rules, seems to be too unorthodox to deserve attention in OGPL. In my view, it is precisely the task of a modern lexicographer to design the database and the rules for derivation and ordering in such a way that ultimately composing lemmas is an automated process rather than a human creation per article.

We know that there is no universal strategy for users to select the headword for a random MWE. We also know that building a database as a source for various products can save a considerable amount of editorial time and can guarantee consistency throughout these products. If one inserts an MWE in the database only once, with a link to every single-word lexical unit (i.e. each meaning of a headword), all options for the selection of the headword in a final product would still be open. This procedure would avoid double, but non-identical, treatment like the examples above. In each entry in which the MWE would be included, we would see the same information (wording, labeling, definition, translation). It would also allow for different choices for different products, and each choice would guarantee a 100% consistency. For example, the choice for either of the strategies that OGPL offers could easily be programmed and automatically produced. The results would be as follows:

- 1 *take with a pinch of salt* (first or only lexical word, if a support verb like 'take'



### PASSWORD

English Dictionary for  
Speakers of Portuguese  
+ ePASSWORD CD-ROM

MARTINS EDITORA  
LIVRARIA

São Paulo, Brazil

July 2009 (book+cd edition)

783 pp, 190 x 128 mm

Paperback, incl. CD-ROM

ISBN 978-85-61635-37-4

<http://martinseditora.com.br>

[http://kdictionary.com/  
products/medium/pwbr.html](http://kdictionary.com/products/medium/pwbr.html)

From the series

KERNERMAN SEMI-  
BILINGUAL DICTIONARIES

were to be excluded from the list of 'lexical words')

2 *take with a **pinch** of salt* (least frequent lexical word, assuming that relative frequency of words is available in the database)

3 *take with a **pinch** of salt* (the first or only noun)

4 ***take** with a pinch of salt* (first or only verb)

Lexicographers could present the various results to potential users for testing. Probably they would formulate an additional rule that results in *salt* as a headword. For example, the occurrence of variation in the first lexical word (*take with a pinch/grain of salt*) could be a reason for an additional rule that marks the second lexical word as the headword. Of course, it would also be possible to place the MWE in several entries, if the size of the dictionary allows it. An in-between solution would be a cross-reference. OGPL briefly mentions the possibility of referring from a potential headword to the headword under which the MWE is treated as a common feature in a DWS (dictionary writing system). However, in 'Building the dictionary entry' (chapter 12) we do not learn how to deal with this feature. It is simply not mentioned again.

In my view, the use of cross-references for MWEs is an attractive time-saving feature. For example in Cobuild we find a reference to *salt* under *pinch* for *take something with a pinch of salt*. In a concise dictionary, a reference to just the headword without the full MWE is already helpful to a user who will recognise the reference as part of the phrase he or she has in mind.

The distribution of MWEs among articles offered a case for looking into the way database and final product relate to each other in OGPL. I am sure that this book will be a success and that it will go through a large number of reprints and I am curious to see if further automation of the dictionary making process will receive more attention in future editions.

### Non-natives read English too

Finally, I would like to make a remark on the blind spot that many native speakers – even those who write for an international public – seem to have for the problems that non-natives may have with infrequent idiomatic English. They should realise that we already have to deal with a double handicap. Not only do we read/study in a language that is not our mother tongue, in addition to that all the example material is drawn from a language that is not our own. What is immediately instructive or illustrative to a native speaker often requires some additional study from a non-native speaker. A little consideration would be appropriate.

In general the language in OGPL is plain and lively English, but here and there an infrequent idiom suddenly enforces the consultation of a dictionary. On page 5 the authors quote the great Dr. Samuel Johnson from 'The plan of an English Dictionary', 1747. Then follows: "Crudely paraphrased this tells us that no amount of theoretical rigour is worth a hill of beans if the average user of your dictionary can't understand the message you are trying to convey." It is ironic that many readers will only be able to understand the paraphrase because the original text is perfectly understandable for an advanced EFL-student. The paraphrase – ideally intended to clarify the quotation – introduces the rare and opaque idiom 'a hill of beans'. No problem for those who remember the famous ending of the film *Casablanca*, in which Humphrey Bogart says '... it doesn't take much to see that the problems of three little people don't amount to a *hill of beans* in this crazy world'. For many others the Americanism introduces a puzzle.

**Rik Schutz**

rik.schutz@onderwoorden.nl

## Seminar on Learner's Dictionaries

### DSNA XVII Biennial Meeting, Bloomington, 2009

The 17th biennial meeting of the Dictionary Society of North America, held at Indiana University in Bloomington on 27-30 May 2009, included a first-of-its-kind seminar on learner's dictionaries. The seminar was an experiment in program structure and was generally restricted to its participants, although some 20 other persons attended in the audience. The participants all read each other's papers in advance, and the seminar served as a forum for further discussion. Three participants were unable to attend the conference and took part in the discussion using Skype. The participants included Arleta Adamska-Sałaciak, Henri Béjoint, Paul Bogaards, Mari Carmen Campoy-Cubillo, Don R. McCreary, Wendalyn Nichols, Michael Rundell, Peter Sokolowski and Shigeru Yamada, and the organizer was Ilan J. Kernerman. The seminar proceedings, along with two more papers on this topic that were presented at the DSNA meeting, will be published by K Dictionaries, as the second volume in its new Papers on Lexicography and Dictionaries series. For the list of contents, see p.3.



# A review of the First International Symposium on Lexicography and L2 Teaching and Learning

Zhang Yihua and Xia Lixin

The First International Symposium on Lexicography and L2 Teaching and Learning was held from 26 to 28 November 2008 at Guangdong University of Foreign Studies (GDUFS). Sponsored by the Chinalex Bilingual Committee, the event was co-hosted by the Center for Lexicographic Studies and the National Key Research Center for Linguistics and Applied Linguistics of GDUFS with support from the Commercial Press, Shanghai Foreign Language Education Press, Foreign Language Teaching and Research Press, Casio (Shanghai) Trade Co. Ltd and Beijing University Press.

## 1. Purpose

The zeal for learning second languages (L2, esp. ESL) has grown in recent decades all over the world, and the number of L2 learners in China continues to increase. With the trade openness and economic growth in China, more and more people around the world are interested in China and start to learn Chinese. The global economic integration and the needs for international exchanges stimulate people to acquire a second language in a short time. One of the results is that the research of learner's dictionaries is drawing growing attention from lexicographers, linguists and language educators. However, due to the feeble exchange between Chinese and international scholars in lexicography and L2 teaching and learning, few learner's dictionaries outside China target specifically Chinese L2 learners, and many L2 dictionaries published in China still cling to traditional approaches to lexical definition and show no consideration for the learner's real needs for L2 acquisition.

Therefore, it is necessary for Chinese and international scholars to form a platform for information exchange. The purpose of the symposium was to bring together scholars of lexicography, second language acquisition, linguistics and dictionary publishing from around the world in an interdisciplinary forum, so that Chinese scholars may be acquainted with new theories and practices of lexicography integrated in learner's dictionary-making abroad, and international scholars and publishers may investigate Chinese users' needs for L2 learner's dictionaries. As stated at the opening ceremony by the organizer, Zhang Yihua, the symposium has great realistic significance for promoting international

exchanges among lexicographers, dictionary publishers and second language teachers from home and abroad, and improving the quality of learner's dictionaries in the coming future.

## 2. Topics

The main topics of the symposium were as follows:

- Learner's dictionary research, compilation and publishing.
- Learner's dictionary and L2 (esp. ESL and C[hinese]SL) teaching and learning.
- Learner's needs analysis and dictionary making.
- Learner corpus and dictionary making.
- Computer-aided dictionary making and dictionary data access.

The conference included an exhibition of dictionaries and lexicographic works that was held alongside the symposium, including Beijing University Press, the Commercial Press, Foreign Language Teaching and Research Press, Shanghai Foreign Language Education Press, Casio, K Dictionaries, Pearson Education (Beijing), and Oxford University Press, and a workshop on electronic dictionaries and L2 teaching in the classroom.

## 3. Participants

The conference was well attended with 136 participants from universities, publishing houses and research institutes in China, the UK, Denmark, Japan, Spain, Poland, Israel, Canada, Australia, Singapore, Hong Kong and Taiwan.

The opening ceremony was chaired by Zhang Yihua and addressed by Professor Sui Guangjun, president of GDUFS, Professor Li Yuming, vice-chairman of the China State Language Committee and Director of the Administration Department of Chinese Language and Information of the Ministry of Education, Professor Wang Tiekun, vice-director of Administration Department of Chinese Language and Information of the Ministry, and Yoshida Shusaku from Casio (Shanghai) Co.

The invited speakers included Li Yuming, Wang Tiekun, Zhang Yihua, Dr. Zheng Ding'ou from the City University of Hongkong, Professor Chen Guohua from Beijing Foreign Studies University, Michael Rundell from Macmillan dictionaries (UK), Dr. Yukio Tono from Tokyo University of Foreign Studies (Japan), Dr. Vincent



**Zhang Yihua** has a PhD in linguistics and applied linguistics. He is the director of the Center for Lexicographical Studies at Guangdong University of Foreign Studies, and is concurrently the vice-president of the China Lexicography Association, chairman of Chinalex Bilingual Committee, vice-chairman of China National Standardization Committee for Lexicographical Terminology, and executive director of the State Committee of Modern Technology for Lexicography. His main research interests include lexicography, semantics and translation, and he has authored numerous publications in lexicography, including more than 70 academic papers, 6 monographs, 2 translation works, and 7 dictionaries. In recent years his research has focused on theoretical issues at the leading edge in lexicography, involving the integration of cognitive linguistics and cyber-linguistics theories, computational lexicography, and the multilingual dictionary generation system. Professor Zhang was the organizer of the First International Symposium on Lexicography and L2 Learning and Teaching. [bilex@mail.gdufs.edu.cn](mailto:bilex@mail.gdufs.edu.cn)



**Xia Lixin** is a research fellow and M.A. candidate supervisor at the Center for Lexicographical Studies, GDUFS. His main research interests are lexicography, translation and semantics. He is also secretary of the Chinalex Bilingual Committee, co-editor of *The English-Chinese Dictionary for Intermediate Learners* (2008), author of more than ten papers on lexicography and translation, and collaborator on several major research projects. xlx710@sina.com

B.Y. Ooi from the National University of Singapore, Dr. Robert Lew of Adam Mickiewicz University (Poland), Dr. Pedro A. Fuertes-Olivera from Valladolid University (Spain), and Dr. Sandro Nielsen of Aarhus University (Denmark).

#### 4. Overview

Over the three conference days, participants shared their experience in dictionary making and language teaching in a variety of forms, on the basis of their research work and vision for the development of lexicography and L2 acquisition. The symposium helped to promote international exchanges among lexicographers from different countries and dictionary publishers and L2 teachers, and it enabled Chinese lexicographers and teachers to learn about the new theories and methods in overseas learner dictionary research and compilation. Likewise, foreign lexicographers and publishers could enrich their knowledge of Chinese learners' needs for L2 dictionaries and of publications of the Chinese language and Chinese dictionaries for foreign learners.

#### 5. Speeches

On each day of the conference, two hours were scheduled for plenary and hot-topic sessions, including eleven invited speakers.

Li Yuming made a keynote speech entitled *The great influence of informatization on Chinese lexicography*. He pointed out that one of the major purposes of this symposium was to call on experts and scholars to abide by the State language norms and standards in the process of dictionary compilation, and emphasized that because dictionaries constitute an important part in language standardization, lexicography plays a key role in the system. In order to create a harmonious language life, we should adhere to the philosophy of "regarding language as a resource", and attach importance also to resolving "the current language conflicts" and realizing language standardization and informatization. Professor Li described the relationship between language standardization and lexicography from five aspects: modernization of dictionary making; multimedia functions of dictionaries; the subtle changes of the relationship among lexicographers, users and publishers; the retrieval capacity; and the function of dictionaries. He stated that knowledge acquisition and creation might become the new dictionary functions, which would be of great significance to future dictionary making.

Wang Tiekun pointed out in his speech, entitled *Investigation on actual language use and some reflections about it*, that it

is the first time for China to hold such a conference, so the Ministry of Education has valued it highly and gave it its full support. He argued that lexicographers engaged in Chinese language dictionaries had little contact with international lexicographical circles and knew little about what foreign dictionary compilers and publishers had done, especially concerning learner's dictionaries. As a result, the making of Chinese dictionaries continued in a conventional way and failed to meet current users' needs. He offered to help Chinese and foreign experts and scholars with the Ministry's language resources, and presented and analyzed important data from *The Annual Report on the Language Situation in China* issued by the State Language Commission with regard to Chinese characters, words, catch phrases and the neologisms of the year.

Zhang Yihua made a speech entitled *A theoretical proposal for meaning driven multi-dimensional definition for a new generation learner's dictionary*. Based on cognitive semantics and L2 acquisition, he set out the interaction between L2 acquisition and learner's dictionaries, and put forward his proposal in a bid to settle the problems in conventional learner's dictionaries and meet the needs of users from a specific language community. He noted that L2 acquisition follows a developmental sequence from *formula* through *low-scope* pattern to *construction*, but existing learner's dictionaries attach importance only to the first two developmental phrases and pay little attention to the third phase of construction. The new generation learner's dictionary should make up for this deficiency by highlighting the description of construction, so its focus should shift from "basic material" to "prefabricated modular", from isolated items to frame-based information, from individual relevance to network relevance. This gives expression to the latest research achievements of cognitive and functional linguistics, i.e. the Cognitive Economy and Idiom Principle, and conforms to the laws of language cognition and learning of human beings.

After a detailed analysis, Zhang advocated that new generation learner's dictionaries should take the meaning-driven communicative pattern as the basis of semantic representation, so the lexicographical definition should be meaning-driven instead of grammar-driven, the description of lexical units should be focused on lexical sequence or chunks instead of on single words, and the lexical meanings should be represented through a semantic and lexical network instead of

by isolated definitions and annotations. Thus the phonology, morphology, syntax and concept are all representational forms of lexical meaning under semantic and pragmatic cognitive constraints, integrated to represent lexical meaning in a multi-dimensional way. Due to the anisomorphism between different cultures and languages, learners from different language communities have different cognitive patterns in L2 acquisition, and a single learner's dictionary cannot satisfy all learners from different communities, so it is essential to compile different dictionaries targeted at different learners.

Michael Rundell began his speech by reviewing the current state-of-the-art in the use of corpora (including learner corpora) in the production of learner's dictionaries, with particular reference to examples of usage, collocations, and the role of metaphor. Then he argued that the use of corpus data in dictionaries is implicit rather than explicit: what users see in their dictionaries is not the primary language data that lexicographers work with, but generalizations from this data made by the editors. Finally he explored some of the possibilities of giving dictionary users more direct access to corpora with the help of emerging technologies.

Yukio Tono argued that information from learner corpora has been used for learner's dictionaries for more than a decade, but it seemed that the analysis was rather limited in scope and the information provided was fragmental in nature, showing only individual lexico-grammatical usage problems of learners. Therefore, a more systematic approach should be adopted in the analysis of learner corpora in order to obtain information that is useful for improving the content of language teaching materials, including dictionaries. He also discussed the potential benefits of such information for pedagogical lexicography.

Zheng Ding'ou discussed the problems concerning the translation of polysemous verbs in a bilingual learner's dictionary (Chinese-English) under compilation, centering on valency items in the light of the French lexicon-grammar. Vincent Ooi introduced the tenets of the 5-circle lexicographic model originally used to relate Singaporean-Malaysian English to British and American English, and argued that such a model served not only to 'harmonize' the notion of international intelligibility but also to 'balance' Asian and Western realities in the dictionary. Chen Guohua held that one of the most important factors that hinder the development of collocation studies is the lack of a proper definition of collocation. He defined a construction as two words that

combine to form a unit, and a collocation in one language in terms of a corresponding construction in another language. Robert Lew discussed new ways of indicating meaning in electronic dictionaries, such as pictorial illustrations, animation and videos, and examined whether the dictionary users always appreciate and benefit from them. Pedro Fuertes-Olivera discussed some basic and general requirements that specialized monolingual dictionaries must have when aiming to solve the needs of learners enrolled in ESP courses. He argued that specialized dictionaries for ESP students must be all-inclusive, i.e. adding encyclopaedic/conceptual information to the linguistic information normally given. Sandro Nielsen defined pedagogical lexicography by the objective to develop principles and guidelines that help practitioners to produce lexicographical tools that fulfil the needs of specific types of users in specific types of situations in the real world, and argued that the dictionary should be analysed in terms of three significant features, namely lexicographic functions, lexicographic data and lexicographic structures. Finally, he believed that the modern theory of dictionary functions opens up exciting new possibilities for theoretical and practical lexicography and encourages lexicographers to adopt a new way of thinking when planning and compiling dictionaries.

The closing ceremony was devoted to these three questions: What should the next-generation learner's dictionaries be like? What could a corpus further offer for dictionary making? and, What is the relationship between printed and electronic dictionaries?

The Second International Symposium on Lexicography and L2 Teaching and Learning will be held at Sichuan International Studies University in 2010.

### **ASIALEX 2009: The Sixth ASIALEX Conference**

ASIALEX 2009 is the 6<sup>th</sup> biennial conference of the Asian Association for Lexicography. The conference is hosted by King Mongkut's Institute of Technology Ladkrabang in Bangkok, Thailand from 20 to 22 August 2009. Its purpose is to promote activities in various areas of lexicography by offering a forum for the exchange of ideas, the presentation of research achievements, and the discussion of future directions. With the theme 'Dictionaries in Education', the main aims of this three-day conference are to examine research into dictionary use and users in Asia, training in dictionary use and dictionary skills in education.

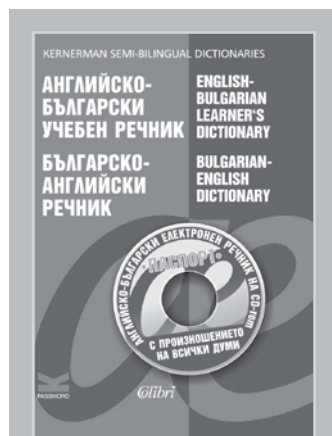
We are delighted to have five distinguished keynote speakers: Michael Hoey (University of Liverpool), Pam Peters (Macquarie University), Gregory James (Hong Kong University of Science and Technology), Theraphan Luangthongkum (Chulalongkorn University) and Suwilai Premrirat (Mahidol University). Speakers include Heming Yong (China), Shigeru Yamada (Japan), Z. Iqbal (Pakistan), G-M. de Schryver (Belgium/South Africa), Amy Chi and Lan Li (Hong Kong). The highlights of the social program include the conference reception and the Bangkok temple and city tour.

#### **Jirapa Vitayapirak**

President of ASIALEX (2008-2009)

<http://www.kmitl.ac.th/asialex/>

# The Center for Lexicographical Studies, GDUFS



**PASSPORT**  
**English-Bulgarian Learner's Dictionary**  
 + Passport Electronic Dictionary  
 COLIBRI PUBLISHERS  
 Sofia, Bulgaria  
 January 2009 (book+cd edition)  
 800 pp, 170 x 128 mm  
 Hardcover, incl. CD-ROM  
 ISBN 978-954-529-613-0  
<http://colibri.bg>  
<http://kictionaries.com/products/junior/ppbg.html>

From the series  
 KERNERMAN SEMI-BILINGUAL DICTIONARIES

## 1. Introduction

The study of lexicography at Guangdong University of Foreign Studies (GDUFS) started in the 1970s, and the main figure was Professor Huang Jianhua, former president of GDUFS and well-known among the pioneers in modern theoretical lexicography in China. In the early 80s lexicography became the key study area in the former Institute of International Languages and Cultures (IILC). The Center for Lexicographical Studies (CLS) was established in the mid-90s as a subordinate to IILC. From then on, Huang, Professor Chen Chuxiang and Professor Zhang Yihua obtained significant achievements in their research, exercising great impact on contemporary lexicography in China. With the development of lexicographical studies and the growth of the academic team, the CLS became an independent research institution of the university in 2001.

At present, lexicographical study at GDUFS is one of the three well-established research areas of the National Key Research Center for Linguistics and Applied Linguistics, and concurrently appraised to be a National Key Discipline. CLS functions as the seat of Chinalex Bilingual Committee. Huang and Zhang are respectively the Committee former and present chairman. The CLS faculty includes 8 full-time members and 10 guest or part-time ones, among them 7 have professorship and 5 associate professorship. Ten of the professors or associate professors are graduate supervisors and two are Ph.D. supervisors. The Center is composed of the following sections: Lexicographical Research Section; Dictionary Compilation Section; Laboratory for Computer-Aided Dictionary Compilation; Lexicography Teaching and Researching Section; and a Reference Room.

## 2. Leaders

Huang Jianhua, the first president of the Asian Association for Lexicography (Asialex, 1997-99), was vice president of the Lexicographical Society of China (Chinalex), chairman of Chinalex Bilingual Committee, vice chairman of China-French Literature Society, and a member of the editorial board of the International Journal of Lexicography. At present, he is adviser of Chinalex, honorary chairman of Chinalex Bilingual Committee, and president of Guangzhou Translators' Association.

Professor Huang's major research interest has been in lexicography. Many of his publications are referred to by Chinese

lexicographers in their studies. In 1989, his monograph *Studies in Lexicography* won the Second Prize of Guangdong Excellent Achievements in Social Science. He edited *An Investigation into Language Dictionaries in English, Russian, German, French, Spanish or Japanese* and collaborated with Chen Chuxiang on *An Introduction to Bilingual Dictionaries*, which both won the second prize of Guangdong Excellent Achievements in Social Science respectively in 1994 and 1999, and is editor-in-chief of *A New Concise French-Chinese Dictionary*, *Dictionary of Biblical Personages*, and *Contemporary Chinese-French Dictionary* (the biggest in China).

Zhang Yihua is currently the CLS president (see page 18). He places emphasis on cutting-edge theoretical research and often brings the latest findings in linguistics into theoretical lexicographical and practical dictionary-making. In recent years, Professor Zhang focuses his research on the theoretical issues at the leading edge in lexicography, involving the integration of cognitive linguistics and cyber-linguistics theories in lexicographical research, computational lexicography, and the multilingual dictionary generation system. His theory of "Meaning Driven Multidimensional Definition" is innovative and widely accepted in monolingual lexicography in China.

Since 1998 Zhang has undertaken many research projects subsidized by provincial and state governments and GDUFS; authored numerous publications in lexicography, including more than 70 academic papers, 6 academic works, 2 translation works and 7 dictionaries. Among them, *Semantics and Lexicographical Definition*, *Computational Lexicography and New-Type Dictionaries* and *Contemporary Lexicography* have exercised great influence on Chinese lexicography in teaching and research; and two of the dictionaries won respectively the first and second prize of China National Dictionary Award.

## 3. Achievements

Over the past seven years, the Center has undertaken more than ten social science research projects sponsored by the Provincial Government (PG), National Education Ministry (NEM) and National Social Science Foundation (NSSF), as follows:

- Cognitive Semantic Structure in Lexicographical Definition (Zhang and Huang, NEM).



- Bilingual Dictionary Generation System — Based on Lexicographical Micro-Data (Zhang and Huang, NEM).
- Investigation on New-Type Dictionaries in the 1990s (Zhang, PG).
- Internetized Dictionary Compilation and Generation System (Zhang, PG).
- Lexicographical Definition of Foreign Oriented Chinese Dictionary Based on User's Perspective (Zhang and Xia Lixin, PG).
- Selection and Definition of Lexicographical Terminology (Zhang, NSSF).
- Semantics and Lexicographical Definition (Zhang, PG).
- Theory and Practice of Modern Lexicography (Zhang, PG).
- Contemporary Chinese-French Dictionary and Its Corpus Construction (Huang, NEM).
- The Construction of the Mode of English Learner's Dictionaries Based on an Empirical Study (Xu Hai).

More than 200 papers and over 10 works, monographs, textbooks, translated books and proceedings on lexicography have been published by the CLS staff members, some of which are listed as follows:

- *Studies in Lexicography* (revised, Huang).
- *A Study on Lexicography* (Huang).
- *Introduction to Bilingual Lexicography* (Huang and Chen).
- *Semantics and Lexicographical Definition* (Zhang).
- *Computational Lexicography and New-Type Dictionaries* (Zhang).
- *Contemporary Lexicography* (Zhang and Yong Heming).
- *Meaning – Cognition – Definition* (Zhang).
- *Dictionaries: The Art and Craft of Lexicography* (by Sydney Landau, translated to Chinese by Zhang and Xia Lixin).
- *Proceedings of the First ASIALEX Regional Symposium* (eds. Huang and Zhang).
- *Collection of Bilingual Lexicographical Studies* (eds. Huang and Zhang).
- *Proceedings of the Sixth Annual Conference on Bilingual Lexicography in China* (Zhang and Sheng Peilin).
- *Index of Publications for Chinese Lexicographical Studies in the 20th Century* (eds. Zhang and Xu Hai).

CLS has also completed 7 dictionary projects, as follows:

- *An Illustrated English-Chinese Dictionary for Primary School Learners* (Zhang and Xu Hai).
- *A Spoken English Dictionary for Middle School Students* (Zhang and Xu Hai).

- *The English-Chinese Medical Dictionary* (Wang Xiaoying and Zhang).
- *The Chinese-English Medical Dictionary* (Wang Xiaoying and Zhang).
- *Dictionnaire Français-Chinois des Locutions et des Expressions* (Xie Pinru and Zhang).
- *English-Chinese & Chinese-English Medical Dictionary* (Wang Xiaoying and Zhang).
- *The English-Chinese Dictionary for Intermediate Learners* (Zhang and Xia Lixin).

#### 4. Exchange

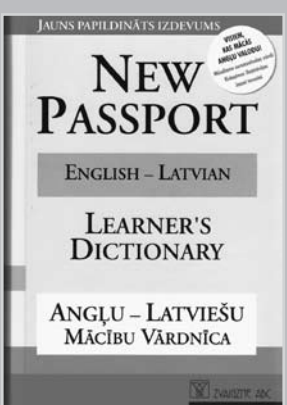
The CLS, as the seat of Chinalex Bilingual Committee, plays an active role in organizing academic activities in China, including national and international conferences and symposiums such as The First Annual Conference of Chinalex and the Fourth National Conference of Lexicography (both in 1993), the First Regional Symposium of Asialex (1999) and the First International Symposium on Lexicography and L2 Teaching and Learning (2008), as well as six National Symposiums for Bilingual Lexicography and four Evaluation Meetings of Bilingual Dictionaries for the China National Dictionary Award.

The members of the Center are frequently invited to national and international academic conferences. CLS has built a Website of lexicographical information, providing an index of academic papers on lexicography in China throughout the twentieth century.

#### 5. Education

The enrollment of Ph.D students majoring in lexicography began in 1994. The year of 1998 saw the enrollment of M.A. students. Annual enrollment takes place in spring, and the process begins half a year earlier, with 16-18 M.A. students and 2-3 Ph.D students are enrolled each time. The main courses for postgraduates are: Semantics, Syntax, Pragmatics, Lexicography, Semantics and Lexicographical Definition, Translation and Bilingual Dictionaries, Practical Lexicography, Principles and Methods of Dictionary-Making, Computational Linguistics and Computational Lexicography, Corpus Processing and Program Design, Linguistic Research Methodology, Guided Reading of Classics in Lexicographic Literature, Contemporary Translation Theory and Practice, Lexicology and Lexical Acquisition, Leading Lectures on Linguistic and Lexicographical Studies, and so on.

<http://bilex.gdufs.edu.cn>



**NEW PASSPORT**  
English-Latvian Learner's Dictionary  
ZVAIGNZE ABC  
Rīga, Latvia  
March 2009 (new edition)  
679 pp, 220 x 150 mm,  
Hardcover  
ISBN 978-9984-34-914-3  
<http://www.zvaigzne.lv>  
<http://kdictionaries.com/products/junior/pplv.html>

From the series  
KERNERMAN SEMI-BILINGUAL DICTIONARIES

# English Dictionaries in Global and Historical Context

Janice McAlpine



**Janice McAlpine** has been Director of the Strathclyde Language Unit (SLU) at Queen's University in Kingston, Ontario, since 2000. She graduated from Queen's University and has a master's degree from the University of Toronto (in English Language and Literature), was Acting Director of the SLU from 1993 to 1996 and an ESL writing instructor at the Queen's School of English from 1996 to 2000. She is co-editor of the *Guide to Canadian English Usage* (Oxford University Press, Toronto, 1997, rev. 2007), edits an annual undergraduate working papers series on Canadian English and is currently lending a hand with two dictionary projects: the revision of the historical *Dictionary of Canadianisms* (ed. Stefan Dollinger), and the English-Chinese *Dictionary of Canadian English* (Commercial Press, Beijing, forthcoming 2009), which began as a project by the dictionary's editor, Xu Hai, while he was a visiting scholar at the SLU in 2003-2004. She is one of the organizers of the conference on English Dictionaries on Global and Historical Context. [jm27@queensu.ca](mailto:jm27@queensu.ca)

Ross Finnie, an economics professor at the University of Ottawa, recently crunched demographic data on Canadian households to find accurate predictors of children's going to university.<sup>1</sup> Not surprisingly, there was a positive correlation between total family income and children's rates of university education. Children whose homes contained books or whose parents had gone to university were also more likely to go to university. But surprising many, and no doubt delighting dictionary publishers, was Finnie's finding that the presence of a dictionary in the home was the single best predictor of children's participation in university education.

From June 3-5, 2010, the English Department and the Strathclyde Language Unit of Queen's University, Kingston, Ontario, Canada, will host a conference on dictionaries where the focus is wide angle. In other words, the aim of this conference—English Dictionaries in Global and Historical Context—is to bring to light the social significance and cultural import of English dictionaries, as well as their impact on our understanding of language. Historical and contemporary English and English-bilingual dictionaries and other lexically oriented pedagogical and reference texts will be under scrutiny.

Most literate people, lexicographers included, take the existence of dictionaries as a class of book for granted. Obviously, people in oral cultures do not. In societies that don't have writing, lexicon cannot be atomized and codified around a sequencing of graphemes as inherently meaningless as it is pervasive and conventionalized—the alphabet, of course. And, neither can the lexicon be documented in quite the same way in societies with ideographic writing systems or highly agglutinative morphology. Is there a relationship between the institutionalized organization of English word hoard and culture, social organization, or ideology of language? What happens when “dictionaryed” and “non-dictionaryed” societies confront each other?

At the interface of English and Aboriginal languages, for example, we might ask under what circumstances, by whom and for whom were bilingual dictionaries made? We can look at their real and putative authors, the texts themselves, their publication history, their European reception, their use or lack of use by Aboriginal people. We can look,

perhaps only indirectly, at the effect they had on Aboriginal worldview, identity, language, language acquisition and retention, creativity, storytelling and literature. Were dictionaries tools of oppression or cultural bridges? And why recently have some bilingual English dictionaries been “de-bilingualized” and reconceived, resulting in publications such as *He Ptaka Kupu*,<sup>2</sup> the first adult monolingual dictionary of the Māori language?

These are just some of the many questions that could be explored at this conference, and the organizers, while we would like to suggest possibilities, do not want to delimit them for potential participants.

Other topics that could be explored include the relevance of the concept of plagiarism to dictionaries, the relationship between lexicographers and their publishers, and between lexicographers and their “readership.” And what about the financial life of dictionaries? Have English dictionaries been cash cows, money pits, or loss leaders for their publishers? How have dictionaries been marketed and where and to whom have they been sold? Has the commercial nature of most dictionary publication kept dictionaries relevant to a wide literate public and prevented them from falling prey to academic fashions and excesses? Or has the commercial nature of dictionary publishing hampered the open exchange of resources and ideas and held back the development of innovative electronic formats and online delivery. For, as Adam Kilgarriff quipped in the pages of an earlier number of this newsletter, “If dictionaries are free, who will buy them?”<sup>3</sup>

The keynote speaker at English Dictionaries in Global and Historical Context will be the author Mark Abley. Having written a widely acclaimed book on endangered languages, *Spoken Here: Travels Among Threatened Languages* (2003),<sup>4</sup> Abley turned his attention to the world's least threatened language—English. *The Prodigal Tongue: Dispatches from the Future of English* (2008) is meticulously researched and global in outlook. The book has raised two fascinating questions with respect to dictionaries and society. Can the authority that we attribute to dictionaries—or that lexicographers have arrogated to themselves—withstand the spontaneous, unsanctioned, neologistic and

lexicographical melee of the Internet? (Abley is fascinated by the popular, user-compiled online dictionary of slang *urbandictionary.com*.) And can it withstand what we might call the “anti-standardization” of the English language—the hutzpah, hypercreolization and creative language mixing that is an essential and globally emulated aspect of the hip-hop movement?<sup>5</sup>

At this conference, some presenters may look at what words (and what people and what things) have been systematically excluded from traditional dictionaries. Others will tackle the historical relationship between linguistic prescriptivism and nation-building as well as the relevance and nature of national dictionaries of English in the twenty-first century. Others still will examine the historical link between language standardization and literary movements and creativity.

Sinfrey Makoni and Alastair Pennycook argue that language is performance and that languages—certainly standard languages—are a fiction, a fiction that has been imposed at great cost to the indigenous languages of Africa and to speakers of creoles. Pennycook derides “the myth of English as an international language” with the comment: “This thing called English colludes with many of the pernicious processes of globalisation, deludes many learners through false promises it holds out for social and material gain, and excludes many people by operating as an exclusionary class dialect, favouring particular people, countries, cultures and forms of knowledge. ...There is something rather bizarre in the belief that if everyone learned English, everyone would be better off.”<sup>6</sup>

This critique of standard English from the point of view of foreign language learners, and by extension of the lexical reference resources that support or reify it, is difficult to reconcile with the view that many English speakers have of their dictionaries—that these books carry cultural knowledge forward through the centuries and across linguistic frontiers, allowing them to read historically and cross-culturally and keeping them from being governed entirely by the politics of the here and now. Intergenerational data from the study by Ross Finnie mentioned above suggest that parents value their literacy and their formal education as more than a gateway to relative prosperity. So would a world without English dictionaries be a better one?

At this conference, we hope to start with English dictionaries and lexical reference texts and let them take us to broader questions of society and governance, culture and creativity, literacy and language. Participation will not be limited to those

presenting papers; others are welcome to register and attend. We look forward to welcoming you at the conference at Queen’s University next year.

- 1 Finnie, R., Mueller, R.E., Sweetman, A. and Usher, A. (eds.), 2009. *Who Goes? Who Stays? What Matters? Accessing and Persisting in Postsecondary Education in Canada*. Queen’s Policy Studies Series 63. Montreal and Kingston: McGill-Queen’s Press.
- 2 Edited and published by Te Taura Whiri i te Reo Māori (Māori Language Commission, New Zealand) in 2008.
- 3 Kilgariff, A. 2005. If dictionaries are free, who will buy them? *Kernerman Dictionary News* 13.
- 4 Abley, M. 2003. *Spoken Here: Travels Among Threatened Languages*. Boston and New York: Houghton Mifflin; Toronto: Random House.
- 5 Abley, M. 2008. *The Prodigal Tongue: Dispatches from the Future of English*. New York and Toronto: Random House, 49-53 and 144-166.
- 6 Pennycook, A. 2007. The Myth of English as an International Language, in S. Makoni and A. Pennycook (eds.), *Disinventing and Reconstituting Languages*. Clevedon, England: Multilingual Matters, 101-102.

#### The conference **English Dictionaries in Global and Historical Context** will take place at Queen’s University, Canada, from 3 to 5 June 2010.

The university was established in 1842 on the shore of Lake Ontario in the heart of Kingston, a historic city of 117,000 inhabitants, midway between Toronto and Montreal. The hosting departments are the Department of English Language and Literature and the Strathy Language Unit, a research unit endowed to produce an authoritative guide to Canadian English usage. The deadline for submission of abstracts is 31 August 2009. Both university residence and hotel accommodation will be reserved for conference attendees, and there are also charming B&Bs in the quarter near the university. Early booking of accommodation will be essential.

[http://post.queensu.ca/~strathy/topics/dic\\_conf.html](http://post.queensu.ca/~strathy/topics/dic_conf.html)



# Internet lexicography as a challenge: The Internet dictionary portal at the Institute for German Language

Stefan Engelberg, Annette Klosa and Carolin Müller-Spitzer



**Stefan Engelberg** is head of the department 'Lexik' at the Institute for German Language and professor for German linguistics at the University of Mannheim.  
engelberg@ids-mannheim.de



**Annette Klosa** is manager of the project 'elexiko' at the Institute for German Language.  
klosa@ids-mannheim.de



**Carolin Müller-Spitzer** is manager of the project 'OWID' at the Institute for German Language.  
mueller-spitzer@ids-mannheim.de

## 1 Introduction

The Internet has become the major challenge for lexicography in the 21st century. Not only does it present lexicography with new possibilities of data integration and crosslinking, it has also changed the demands made on the competences that have to be gathered within lexicographic projects, reaching from the development of corpus analysis methods to text technology and web technology. In connection with these developments, the lexicographic work proper is becoming more complex: The ability to handle and analyse mass data and the need for Internet-adequate lexicographic concepts is changing the profession. In this article, we will take a look at the Internet dictionary portal developed at the Institut für deutsche Sprache ('Institute for German Language') in Mannheim and illustrate the lexicographic practice. In Section 2, we present the Online German Lexical Information System OWID. Section 3 illustrates the lexicographic practice and structure of *elexiko*, the largest dictionary published in OWID. In Section 4, the basic principles of data structuring and administration are discussed, and Section 5 gives an overview on the workflow and work organization within the department *Lexik*, which hosts the Internet-lexicographic projects of the Institute for German Language.

## 2 The lexicographic information system OWID

**Dictionaries in OWID:** The *Online-Wortschatz-Informationssystem Deutsch* (OWID; Online German Lexical Information System) is a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the Institute for German Language (Institut für Deutsche Sprache, IDS; cf. OWID 2008ff.). The main emphasis of OWID is on academic lexicographic resources of contemporary German. Presently, the following dictionaries are included in OWID.

- "*elexiko*" (2003ff., cf. Haß 2005, Klosa et al. 2006) consists of an index of about 300,000 short entries with information on spelling, spelling variation, and word division. In addition, many entries contain citations automatically chosen from the *elexiko*-corpus. Furthermore, *elexiko* contains more than 1,000 fully elaborated entries of high-frequency headwords, focussing on extensive semantic-pragmatic

descriptions of lexical items in actual language use. The dictionary is being extended continuously by further elaborated entries.

- The *Neologismenwörterbuch* (Dictionary of Neologisms) (2005ff., cf. Herberg et al. 2004) describes in detail about 800 new words and new meanings of established words added to the German vocabulary since the 1990s. This dictionary is also constantly upgraded.

- *Feste Wortverbindungen online* (Multiword Expressions Online) (2007ff.) publishes the research results of the project *Usuelle Wortverbindungen* (Fixed Multiword Combinations). Twenty-five detailed entries for fixed multiword combinations and 100 shorter entries dealing with additional collocations are currently available.

- The *Diskurswörterbuch 1945-55* (Discourse Dictionary 1945-55) (2007) is a reference work covering the lexemes that establish the discourse about "guilt" in the early post-war era 1945-1955. It subsumes the lexical-semantic results of the discourse-theoretical studies in Kämper (2005, 2007).

In the near future, the *Handbuch deutscher Kommunikationsverben*, a handbook of German communication verbs which consists of two volumes, a dictionary containing approximately 350 entries and a volume representing the lexical structures of German communication verbs by means of lexical fields (cf. Harras et al. 2004, Harras et al. 2007), the "VALBU - Valenzwörterbuch deutscher Verben" (Schumacher et al. 2004) (Valency Dictionary of German Verbs), and about 300 articles from a corpus-based project on proverbs will be published in OWID.

**Access structure:** The main function of OWID is to provide a common access structure in the form of search options across the individual dictionaries. This is the typical function of lexicographic portals (cf. Engelberg and Müller-Spitzer, forthcoming).

To give an example: if a user types "global\*" in the search box of OWID, he/she gets the results displayed in Figure 1.

The entries from *elexiko* are presented in bold black, the entries from the neologism dictionary in bold black italics, those from the multiword dictionary (see below) in small capitals. (In the online presentation, information coming from





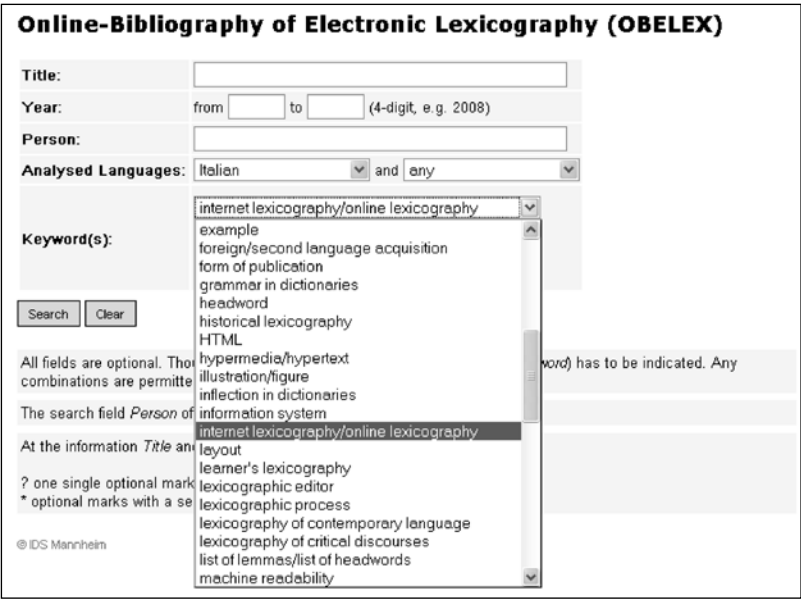


Figure 5 The Online Bibliography of Electronic Lexicography (OBELEX)

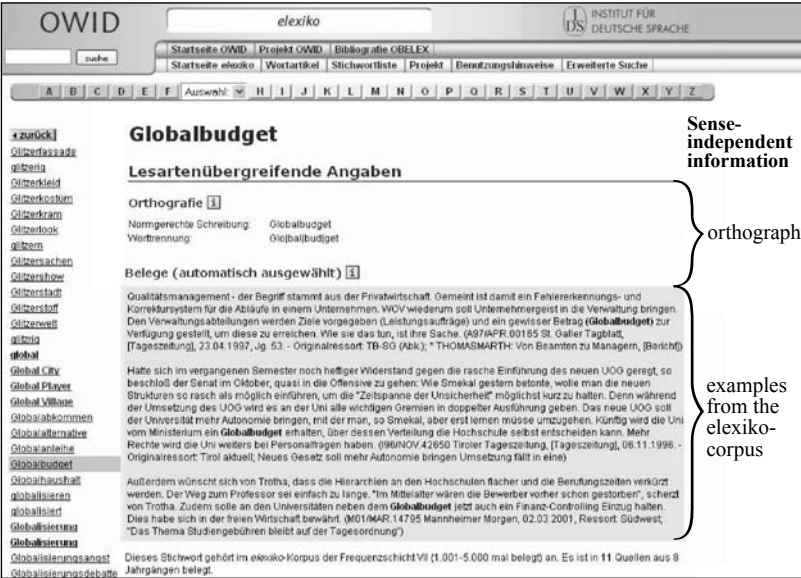


Figure 6 Entry Globalbudget with automatically compiled sense-independent information

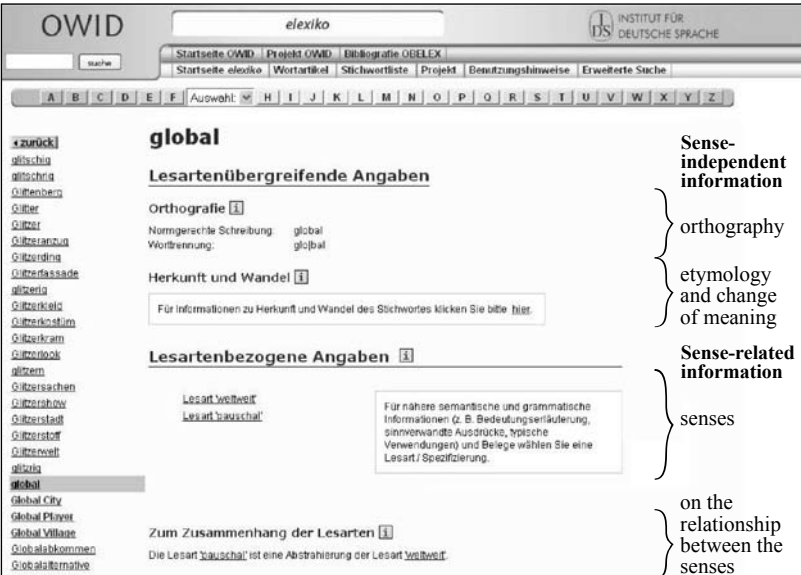


Figure 7 Entry global with information on the senses.

on the Internet and is published within the lexicographic information system OWID. If one refers to *ellexiko* as an Internet “dictionary,” one should keep in mind that *ellexiko* is more than a dictionary in its traditional sense although, of course, it contains descriptions of the meaning and use of a lexeme just as any traditional dictionary. It is both, a hypertext dictionary and a lexical data information system (Storjohann 2005b, Klosa et al. 2006).

The primary and exclusive basis for lexicographic interpretation in *ellexiko* is an extensive German corpus. The compilation of the dictionary does not proceed letter by letter. Instead, the headwords are grouped into word classes or word fields for the compilation process.

Modules may also be defined according to levels of frequency and distribution of lexemes in the *ellexiko*-corpus. *Ellexiko* is now working on a module called “Lexikon zum öffentlichen Sprachgebrauch” (“Dictionary on Public Discourse”). It contains approximately 2,800 entries selected mainly by their (high) frequency in the *ellexiko*-corpus.<sup>1</sup>

The list of headwords<sup>2</sup> was taken exclusively from the *ellexiko*-corpus and was published on the Internet before starting the work on lexicographic modules. For each headword, sense-independent information generated automatically or semi-automatically from the underlying corpus is given. This concerns 300,000 single-word entries comprising details on spelling, spelling variation, and word division. Many entries also contain automatically chosen citations from the *ellexiko*-corpus (cf. Figure 6).

Lexicographically fully-described entries (as in the module “Dictionary on Public Discourse”) entail sense-independent information on morphology and word formation (“Lesartenübergreifende Angaben”) as well as a number of senses and their relationship. They also offer a large scope of sense-related information (“Lesartenbezogene Angaben”) in detail: meaning definition, collocations, syntagmatic patterns, sense-related terms, pragmatics, and grammar.

For the process of writing and presenting *ellexiko* on the Internet, the project uses numerous technologies and software tools, such as the corpus query and processing tool COSMAS II<sup>3</sup> developed at the IDS and its incorporated collocation

1 Examples given here are part of this module and may be viewed online.  
2 For the list of headwords in alphabetical or reverse order, see [http://www.owid.de/ellexiko/\\_Stichwortliste.html/](http://www.owid.de/ellexiko/_Stichwortliste.html/).  
3 For COSMAS II, see <http://www.ids-mannheim.de/cosmas2/>.

software package “Statistical collocation analysis and clustering.”<sup>1</sup> In addition, the corpus-linguistic research and development workbench CCDB<sup>2</sup> (also developed at IDS) is used. The corpus research software and the co-occurrence analysis are employed for numerous corpus-guided investigations within the practical working procedure.

**Sense disambiguation:** To disambiguate a highly frequent polysemous word in *lexiko*, we have “developed a disambiguation technique which is based on empirical and theoretical grounds. The lexicographic prerequisites of this disambiguation procedure are an elaborate theory, corpora, a data-processing software, and the linguistic competency of data interpreting.” (Storjohann 2003: 755). Employing COSMAS II and the co-occurrence analysis, the lexicographer disambiguates a polysemous word at three levels: the collocation level, the KWIC-level, and the text level and, thus, achieves a detailed semantic description (Storjohann 2003: 755). In a second step, the senses found by a strictly corpus-driven method are contrasted to the senses given in other dictionaries. If the lexicographer finds that in *lexiko* the range of senses seems yet to be incomplete, the corpus will be searched again. In this way, corpus-driven and corpus-based information on all senses of a lexical item is combined.

All senses are listed in *lexiko* in a frequency-based hierarchy: The most frequent sense in the *lexiko*-corpus comes first. Senses in *lexiko* belonging to each other such as the literal and the metaphoric sense are, thus, sometimes torn apart. At the same time, there may be word senses that are not related to the others at all. Accordingly, in the *lexiko*-entries, users find an explanation on how the senses are arranged and how they are connected. On the first screen opening after typing in the search word, there is a list of all senses, which are not numbered but labelled with words (cf. Figure 7); here ‘weltweit’ (‘worldwide’) and ‘pauschal’ (‘indiscriminate’). Under the rubric “Zum Zusammenhang der Lesarten” (“On the relationship between the senses”), their relationship is explained: The word sense ‘pauschal’ is an abstraction of the word sense ‘weltweit.’

**Sense-related information:** When clicking on one of the senses from the list shown in Figure 7, the user opens the complete range of sense-related information

(“Lesartenbezogene Angaben”) in *lexiko*: meaning definition (“Bedeutungserläuterung”), collocations (“Semantische Umgebung und lexikalische Mitspieler”), syntagmatic patterns (“Typische Verwendungen”), lexical relations (“Sinnverwandte Wörter”), pragmatics (“Besonderheiten des Gebrauchs”), and grammar (“Grammatik”). All information is extracted from the corpus, either in a corpus-driven or a corpus-based approach (cf. Haß 2005, Klosa 2007).

The information on collocations given in *lexiko* is an example of corpus-driven information: Typical, highly frequent words co-occurring with the headword are extracted from the corpus (using the software package “statistical collocation analysis and clustering”) and are arranged according to their semantic function by the lexicographer. The headword **global** in the sense ‘weltweit’ (‘worldwide’), for example, is used to specify actions or to characterize issues. The entry in *lexiko* gives statistically significant partner words as shown in Figure 8 (*Was macht man global? agieren, denken, ...* ‘What can be done globally? to act, to think, ...’; *Was ist global? Denken, Ebene, ...*, ‘What is global? thinking, level, ...’).

Besides characterizing syntagmatic semantic relations, *lexiko* also offers a new way of presenting paradigmatic sense relations in “a differentiated system of paradigmatic relations including synonymy, various subtypes of incompatibility (such as antonymy, complementarity, converseness, reversiveness, etc.), and vertical structures (such as hyponymy and meronymy)” (Storjohann 2005a: 1, cf. also Storjohann 2005b). For example, the entry **global** ‘weltweit’ lists as synonyms *international* (‘international’), *universell* (‘universal’), *weltumspannend* (‘global’), and *weltweit* (‘worldwide’). The complementary partner words *lokal* (‘local’), *provinziell* (‘provincial’), *regional* (‘regional’), and *national* (‘national’) are arranged into semantic groups as shown in Figure 9. Each of the partner words is illustrated in *lexiko* by a citation, and each one includes a hyperlink to the corresponding article. Other ways of presenting paradigmatic partners (e.g., in a semantic net) may be developed in the future.

In order to gather lexical information of this kind, the statistically significant co-selections of a word are analyzed with the corpus research software COSMAS II. Among the results of this analysis, the lexicographers will often find words related to the headword in a semantic way. This is why analyzing collocations automatically is the first step the *lexiko*-lexicographers take

#### Was ist global?

Denken  
Ebene  
Erwärmung  
Herausforderung  
Kapitalismus  
Klima  
Konkurrenz  
Krise  
Lösung  
Markt  
Maßnahme  
Netz  
Problem  
System  
Unternehmen  
Verantwortung  
Vernetzung  
Wettbewerb  
Zeitalter

#### Was macht man global?

agieren  
denken  
handeln

**Figure 8** Collocations for the entry **global** (‘weltweit’)

<sup>1</sup> For the software package “Statistical collocation analysis and clustering,” see <http://www.ids-mannheim.de/kl/projekte/methoden/ka.html/>.

<sup>2</sup> For CCDB, see <http://corpora.ids-mannheim.de/ccdb/> and Keibel and Belica (2007).

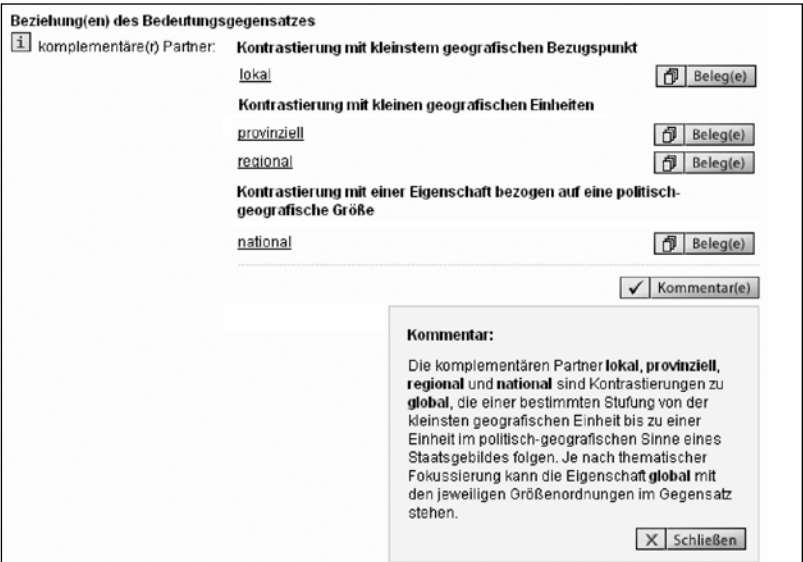


Figure 9 Groups of complementary partner words for **global** ‘weltweit’ (‘worldwide’)

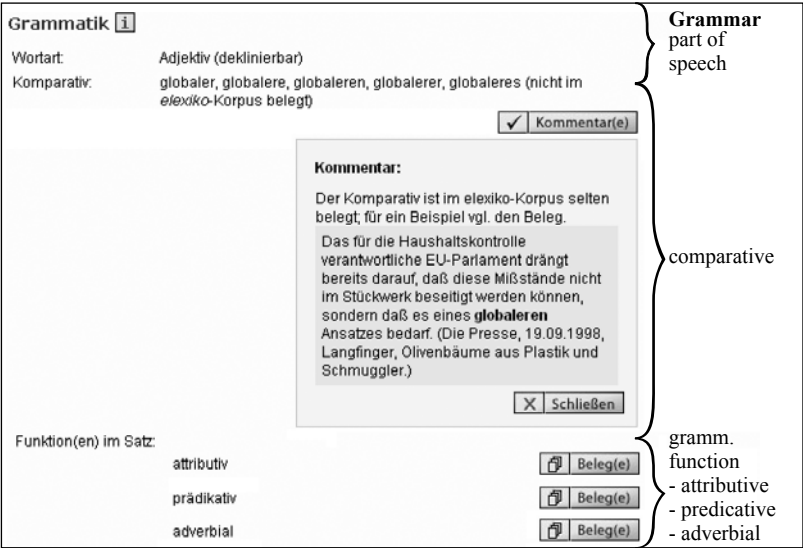


Figure 10 Grammatical information on **global** ‘pauschal’



Figure 11 Part of the XML file of the article **global** in *ellexiko*, representing the syntactic functions of the adjective

on their way to identifying paradigmatic relations - without relying on intuition or personal linguistic competence. In a second step, the lexicographers evaluate the results gathered automatically: They classify the sense relations found and search for citations from the corpus texts that exemplify the sense relation concerned. Since in some cases the corpus-tools cannot provide a comprehensive description of the sense relational patterns, the corpus is checked based on a comparison with other dictionaries in a third, supplementary procedure, particularly so as to extend or complete paradigmatic descriptions.

**Grammar in *ellexiko*:** Grammatical information in *ellexiko* is also based on the corpus, and it is given for each sense. For adjectives, besides naming the part of speech, users find information on comparison and on attributive, adverbial, or predicative use (see Figure 10). All uses are only given if they are found in the corpus. Corpus-based grammatical information in *ellexiko* is, thus, more detailed and more reliable than information given in many other dictionaries. On the other hand, a strictly corpus-based approach may cause problems: “When you analyze corpus data, you constantly find that inflectional forms given in the dictionary are in fact not attested in actual text productions, or the opposite situation of inflectional forms occurring that are not part of the official paradigm. The problem is primarily one of interpretation: is the absence of an inflectional form an indication that it does not exist, or is it an indication that the corpus is simply not large enough?” (Trap-Jensen 2002, cf. also Klosa and Müller-Spitzer 2007).

As can be seen from the information on comparison in Figure 10, each comparative form for the headword **global** is checked in the *ellexiko*-corpus, and, if not found there, this is recorded (“nicht im *ellexiko*-Korpus belegt”, i.e., “no evidence in the *ellexiko*-corpus”). In a commentary accompanying the comparative forms, the lexicographer may note the rare use in the corpus and give an example from the corpus texts.

**4 Data structures and administration**  
**Dynamic customizable microstructures:**

With respect to their contents, the individual participating projects and their compiled lexicographic resources in OWID are independent of each other. However, it has been obvious from the very beginning that the value of OWID would be increased if more common access structures for the different contents could be developed and if the lexicographic data would be interlinked more adequately. Above all, we



wanted to respect requirements of modern lexicography and dictionary research. For example, the dictionary user interface should be adaptable to specific dictionary consulting situations by creating dynamic customizable microstructures. “It is one thing to be able to store ever more data, but another thing entirely to present just the data users want in response to a particular look-up.” (De Schryver 2003: 178, cf. also Engelberg and Lemnitzer 2001, Storrer 2001, and on the modelling concept Müller-Spitzer 2007) So on the one hand, in order to create a basis for a common access structure to the content, consistent principles for modelling and structuring the contents were applied to all integrated products. On the other hand, OWID will also be kept open for the possible integration of externally developed lexicographic resources, namely, reference works that are written outside IDS and other lexicological resources.

**Data modelling:** The approach chosen here not only guarantees that different lexicographic products can be integrated under the management of OWID on the macro structure level, that is, the level of headwords, but that dictionaries can also be accessed on a more granular level. Therefore, the attempt was to harmonize modelling on the level of content structure, that is, the level of the individual lexicographic information.

**Technical architecture of OWID:** OWID uses a single modelling process for all projects: All lexicographic data are stored as XML files. For each individual resource, a special tailor-made XML-DTD/XML-schema was developed. In these DTDs, the microstructures of the individual dictionaries are defined. Focusing on the interconnectedness of the individual projects, a modular system was established where identical phenomena were modelled identically and only once. The dictionary entries are then written with an XML editor and stored in an Oracle database system. For the Internet presentation, the XML data are transformed by an XSLT stylesheet to HTML. To provide a uniform structure for lexicographic information of the same type contained in different dictionaries, a DTD library was created for OWID, where specific DTDs contain all entities, elements, or attributes that are shared by all entry structures. Due to this segmentation, the modelling level already shows what information is accessible across the different dictionaries. This procedure requires each individual information unit to be granularly tagged in all entry structures, but it also allows for automatic access to each content unit. Figure 11 provides an initial impression of

the overall tagging granularity. It shows a part of the XML file of the entry **global** from *lexiko* (corresponding to the bottom part of Figure 10), illustrating the tagging of information concerning the grammatical functions of adjectives.

**Editing system EDAS:** Our internal editing system EDAS (Electronic Dictionary Administration System) allows lexicographers to use the granular XML structures for very special search questions with XPath, an expression language used to access or refer to parts of an XML document. On the basis of XML structures as in Figure 11, it is, for example, possible to search for all adjectives that are used attributively in prenominal position (cf. Figure 12).

As a result, adjectives from *lexiko* and the neologism dictionary are displayed. Searches on the content of paraphrases are possible as well. Figure 13 shows the results of a query that requests all entries that have the word “Computer\*” in their paraphrase. It can be seen that all entries displayed in this search result belong to computer-related vocabulary.

**Search options:** These search options can only be used internally by the lexicographers. For the dictionary user, the search potential hidden in these data structures has only been partly revealed so far. However, the search option “Erweiterte Suche” (Advanced Search) within each dictionary already facilitates detailed searches for specific information. It is, for instance, possible

- to search in *lexiko* for all nouns with an old spelling variant that are compounds;

**Erweiterte Suche**

enthält    
 [Suchkriterium hinzufügen]

1. Hartz-IV-sicher
2. abgezockt
3. achtziger
4. afrikanisch
5. aktuell
6. alarmistisch
7. alt
8. amerikanisch
9. angefasst
10. arbeitslos
11. atheistisch
12. aufgestellt
13. bandförmig
14. bayerisch
15. behindert
16. beruflich
17. billig
18. bisherig

**Figure 12** XPath search in EDAS (extract), displaying the list of adjectives corresponding to the restriction expressed as a search term

**Erweiterte Suche**

enthält  mit Inhalt   
 [Suchkriterium hinzufügen]

1. Banking
2. Barcode
3. Beamer
4. Bildschirmschoner
5. Bluetooth
6. Browser
7. Computerwurm
8. Cookie
9. Cybersex
10. DVD
11. Dateiformat
12. Datenautobahn
13. Datenhandschuh
14. Datenhighway
15. Direktbank
16. Direktbanking
17. Doppelklick
18. Download

**Figure 13** XPath search in EDAS (extract), displaying the list of headwords that include “Computer\*” in their paraphrase

- to search in the dictionary of neologisms for all new lexemes (Neologismtyp=“Neulexem”) that entered the German language in the early 1990s (Aufkommen=“Anfang der 90er Jahre”). Search results are words like *wegzappen* ‘to zap to another channel’, *Neufünfland* ‘New Laender’, or *abspacen* (partly calque of ‘to space out’);
- to search in the dictionary of neologisms for all verbs (Wortart=“Verb”) that gained a new sense (Neologismtyp=“Neubedeutung”) in the 1990s. (Results are verbs like *blicken* ‘to understand something’) or *surfen* (new sense ‘browsing through the Internet’).

These examples show only some of many possibilities. Similar searches can be defined for all approximately 450 possible elements and their additional attributes available within the OWID modular entry structures.

**OWID and the user:** It is our goal to provide a maximum of flexibility for the user interface of OWID. Therefore, the data are modelled solely with respect to the content; aspects of presentation and dictionary use are kept apart from considerations about data modelling. Thus, the same data can be displayed differently for numerous user types and look-up situations with no need to transform it. However, until now for electronic lexicography, it has not been systematically investigated in large empirical studies which functionalities are useful for particular user groups and situations. Therefore, we are running an academic project focused on user research and data

relations within and between lexicographic resources. For OWID, today’s challenge is - besides the continuous extension and enhancement of the individual dictionaries and integration of new dictionaries - to provide the user with an increasing range of more flexible display possibilities of this machine-readable puzzle, which can lead to new forms of using lexicographic information. In this manner, OWID is a living system that is modified and expanded continuously.

## 5 Competence and workflow

**Technological and lexicographic competences:** In order to compile Internet dictionaries like *ellexiko* and create dictionary portals like OWID, a wide range of competences are required. The tasks that make up our Internet-lexicographic enterprise fall into four major domains: (I) corpus linguistics, (II) lexicography proper, (III) text-technology and computer lexicography, and (IV) metalexicographic research.

**(I) Corpus linguistics:** All our lexicographic projects are based on corpora. This requires the constant acquisition and maintenance of large corpora of present-day German. Corpus research software (“COSMAS II”) is developed within the institute as well as methods of corpus analysis, in particular methods based on co-occurrence analyses. Most of these developments are made available to the public via Internet. Other methods have been developed mainly for internal use, for example, the semi-automatic compilation of frequency-based lemma lists, methods for extracting and representing frequency changes in word use, and the automatic insertion of lexicographic examples from corpora. These activities require competences in the domain of statistics, text technology, data processing, etc.

Most of the work within the area of corpus linguistics is done in projects outside the department *Lexik* (reference corpora, co-occurrence analysis, corpus research software), in particular in the project group “Corpus Linguistics”<sup>1</sup> and the data processing division. The development of methods for the automatic extraction of lexicographic information from corpora and the acquisition of historic corpora are activities located within the department. These will be intensified in the future to meet our particular lexicographic needs.

**(II) Lexicography proper:** Despite the indispensable corpus linguistic, text technological, and computer lexicographic

<sup>1</sup> See <http://www.ids-mannheim.de/kl/projekte/methoden/>.

activities, most of the work falls into the domain of lexicography proper. As described in Section 2, this involves the application of corpus linguistic methods, the analysis of the results, and the compilation of lexicographic articles. It also involves a large amount of conceptual work and accompanying activities such as the maintenance of an elaborate editorial manual, as in the project *ellexiko*. The work requires experienced lexicographers and linguists and the ability for sophisticated use of corpus analysis methods. Since there is still a lack of institutionalized training and study programs for lexicographers<sup>1</sup>, we invite students for short periods of practical training, employ student researchers, and try to integrate lexicographers in the making into our lexicographic projects.

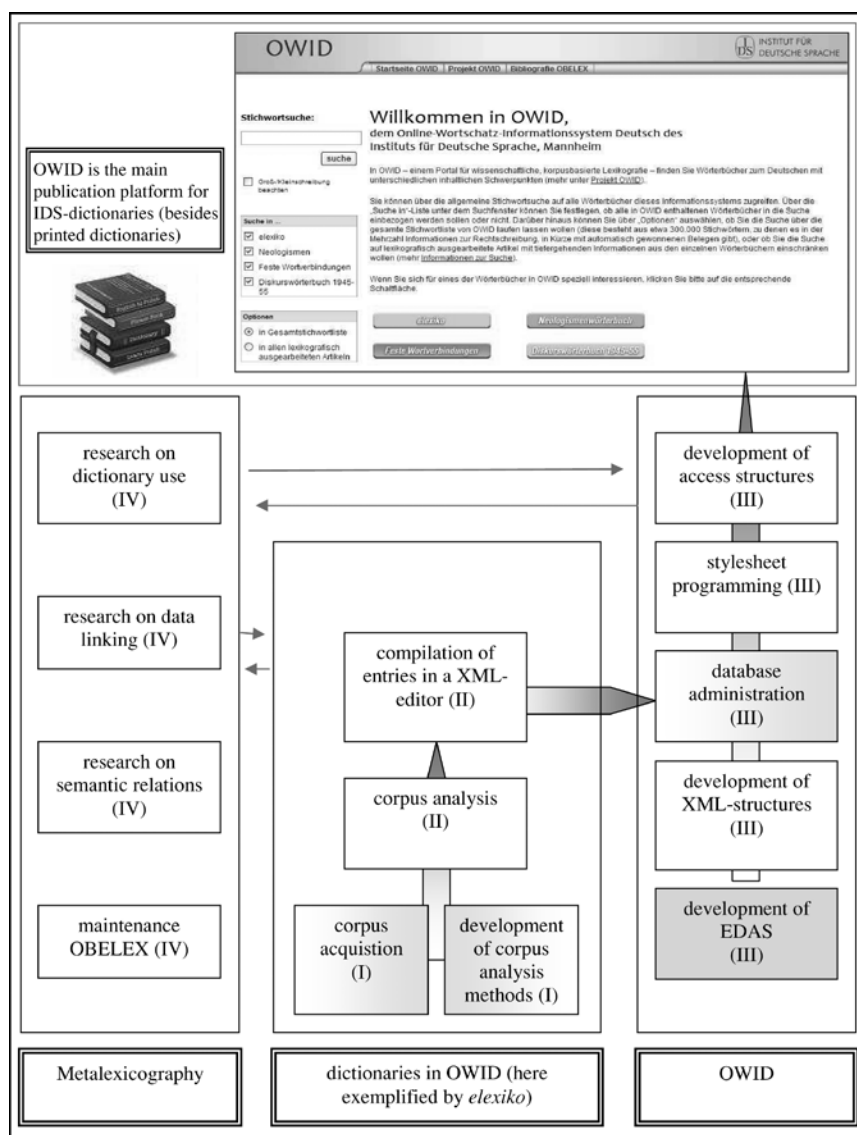
**(III) Text technology and computer lexicography:** As described in Section 4, the lexicographic data are created in XML form, stored in an Oracle database, and then transformed by XSLT stylesheets into the HTML surface representations visible to the user. In addition, web-based methods of research on dictionary use are developed. Thus, the competences required range from text technology over database management to web technology.

These core competences are covered within the department. Additional support comes from the grammar department that is experienced in Oracle database application. It is responsible for the development of the editorial system EDAS and the generic bibliography system on which OBELEX is based.

**(IV) Metalexicographic research:** Finally, the department is engaged in numerous types of metalexicographic research: Its main focus is currently research on dictionary use, on the linking of lexicographic data, and on the investigation and lexicographic representation of lexical semantic relations, in particular with respect to *ellexiko* and OWID. Apart from the core lexicographic, linguistic, and corpuslinguistic knowledge, the research on dictionary use in particular requires competences about research methods within the domain of empirical social sciences and Internet-based user research. Besides these activities, the OBELEX bibliography is developed in OWID.

**Personnel and workflow:** Figure 14 gives an overview on the different activities that are involved in creating our Internet portal and dictionaries. It also illustrates which core activities are carried out within the

department and which support is necessary from other departments within the Institute. Read from bottom to top, the chart also gives a rough description of the workflow. Many employees are involved in various tasks and projects, but most of them allocate only part of their time to activities closely related to our Internet lexicographic projects. Furthermore, most of them are involved in more than one of the above-mentioned four task domains. Altogether, the equivalent of 10-12 full-time positions is allocated to Internet-lexicographic activities within the department. Since a number of researchers have part-time positions, up to 18 lexicographers, linguists, text technologists, etc. are involved in our Internet-lexicographic projects. This does not include the support given by approximately 8 student researcher positions. The distribution of



**Figure 14** Task domains with regard to Internet lexicography (I = corpus linguistics, II = lexicography proper, III = text technology and computer lexicography, IV metalexicography) and an indication of the work flow. The shaded areas characterize the work done outside of the department Lexik

<sup>1</sup> However, a "European Master in Lexicography" is currently being developed.

Coordination and conception of Internet lexicography	10 %
Task domain I: Corpus linguistics (in particular, automatic extraction of lexicographic information)	5 %
Task domain II: Lexicographic work proper (half of it in <i>elexiko</i> , the other half in the other dictionary projects mentioned in Section 2); including other activities like teaching classes, workshops on corpus-based lexicography, and Internet lexicography, etc.	50 %
Task domain III: Text technology and computer lexicography (DTDs, stylesheets, Oracle)	20 %
Task domain IV: Metalexicographic research (as described above)	15 %

**Figure 15** The approximate distribution of positions within the department *Lexik*

positions within the department *Lexik* among our main Internet-lexicographic task domains appears in Figure 15.

Most of these positions are financed by the regular budget of the Institute, some of them - in particular, those allocated to metalexicographic research, the automatic extraction of lexicographic data from corpora, and the compilation of the dictionary of proverbs - are financed also with funds acquired from third-party donors.

### 6 Outlook

Internet lexicography will remain a main task for the department *Lexik*. The challenges posed by a thorough corpus orientation will continue to shape our lexicographic work in the future. This also holds for the conceptual demands that accompany the step from print to Internet lexicography, as is evident in projects like *elexiko*.

OWID will be the main platform for our lexicographic work. We are currently preparing the integration of several other dictionaries into OWID. However, for those dictionaries that were not modelled as fine-grained XML structures from the beginning, the conversion into proper XML structures is still tedious work. With the experience we are gaining, these tasks are becoming easier to carry out. Also, new dictionary projects are more likely to begin with adequately modelled XML structures from the start. New OWID dictionaries also pose new questions with respect to how to link the data and present them in a user-friendly and linguistically interesting format. Thus, besides broadening the lexicographic basis by integrating new dictionaries, the main and most interesting task will be to integrate the data in a way that goes beyond the creation of a common external access structure.

### Bibliography

**De Schryver, G.-M. 2003.** 'Lexicographer's Dreams in the Electronic-Dictionary Age.' *International Journal of Lexicography* 16.2: 143-199.

**Diskurswörterbuch 1945-55. 2007.** *OWID – Online Wortschatz-Informationssystem Deutsch*. Mannheim, Institut für Deutsche Sprache. [www.owid.de/Diskurs1945-55/index.html/](http://www.owid.de/Diskurs1945-55/index.html/).

**elexiko. 2003ff.** *OWID – Online Wortschatz-Informationssystem Deutsch*. Mannheim, Institut für Deutsche Sprache. [www.owid.de/elexiko/\\_index.html/](http://www.owid.de/elexiko/_index.html/).

**Engelberg, S., and Lemnitzer, L. 2001.** *Lexikographie und Wörterbuchbenutzung* (= Stauffenburg Einführungen, Band 14). Tübingen: Narr.

**Engelberg, S., and Müller-Spitzer, C. Forthcoming.** 'Dictionary portal' in R.H. Gouws, U. Heid, W. Schweickhard, and H.E. Wiegand (eds.), *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography*. Berlin / New York: de Gruyter.

**eValBu. 2009ff.** 'elektronisches Valenzwörterbuch deutscher Verben' in *OWID – Online Wortschatz-Informationssystem Deutsch*. Mannheim: Institut für Deutsche Sprache.

**Feste Wortverbindungen. 2007ff.** *OWID – Online Wortschatz-Informationssystem Deutsch*. Mannheim: Institut für Deutsche Sprache. [www.owid.de/Wortverbindungen/index.html/](http://www.owid.de/Wortverbindungen/index.html/).

**Harras, G., Winkler, E., Erb, S., and Proost, K. 2004.** *Handbuch deutscher Kommunikationsverben. Teil 1: Wörterbuch* (= Schriften des Instituts für Deutsche Sprache 10.1). Berlin / New York: de Gruyter.

**Harras, G., Proost, K., and Winkler, E. 2007.** *Handbuch deutscher Kommunikationsverben. Teil 2: Lexikalische Strukturen* (= Schriften des Instituts für Deutsche Sprache 10.2). Berlin / New York: de Gruyter.

**Haß, U. (ed.) 2005.** *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz*. (Schriften des Instituts für Deutsche Sprache), Berlin / New York: de Gruyter.

**Herberg, D., Kinne, M., and Steffens, D. 2004.** *Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. Unter Mitarbeit von E. Tellenbach und D. al-Wadi* (Schriften des Instituts für Deutsche Sprache 11). Berlin / New York: de Gruyter.

**Kämper, H. 2005.** *Der Schulddiskurs in der frühen Nachkriegszeit. Ein Beitrag zur Geschichte des sprachlichen Umbruchs nach 1945* (Studia Linguistica Germanica 78). Berlin / New York: de Gruyter.

**Kämper, H. 2007.** *Opfer - Täter - Nichttäter. Ein Wörterbuch zum Schulddiskurs*

- 1945-1955. Berlin / New York: de Gruyter.
- Keibel, H., and Belica, C. 2007.** 'CCDB: A Corpus-Linguistic Research and Development Workbench' in *Proceedings of Corpus Linguistics 2007, Birmingham*. ([http://www.corpus.bham.ac.uk/corplingproceedings07/paper/134\\_Paper.pdf](http://www.corpus.bham.ac.uk/corplingproceedings07/paper/134_Paper.pdf)).
- Klosa, A. 2007.** 'Korpusgestützte Lexikographie: besser, schneller, umfangreicher?' in Kallmeyer, W., and Zifonun, G. (eds.), *Sprachkorpora - Datenmengen und Erkenntnisfortschritt* (Jahrbuch des Instituts für Deutsche Sprache 2006). Berlin / New York: de Gruyter, 105-122.
- Klosa, A., and Müller-Spitzer, C. 2007.** 'Grammatische Angaben in *elexico* und ihre Modellierung' in Gottlieb, H. and Mogensen, E. (eds.), *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004*. Amsterdam: Benjamins, 13-37.
- Klosa, A., Schnörch, U., and Storjohann, P. 2006.** 'ELEXIKO – A lexical and lexicological, corpus-based hypertext information system at the Institut für Deutsche Sprache, Mannheim' in Marengo, C. et al. (eds.), *Proceedings of the 12th EURALEX International Congress (Atti del XII Congresso Internazionale di Lessicografia), EURALEX 2006, Turin, Italy, September 6th – 9th, 2006. Vol. 1*. Turin: Edizioni dell'Orso Alessandria, 425-430.
- Kommunikationsverben online. 2009ff.** OWID – Online Wortschatz-Informationssystem Deutsch. Mannheim: Institut für Deutsche Sprache.
- Müller-Spitzer, C. 2007.** *Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis* (Studien zur deutschen Sprache 42). Tübingen: Narr.
- Neologismenwörterbuch. 2005ff.** OWID – Online Wortschatz-Informationssystem Deutsch. Mannheim: Institut für Deutsche Sprache. [www.owid.de/Neologismen/index.html/](http://www.owid.de/Neologismen/index.html/).
- OWID – Online-Wortschatz-Informationssystem Deutsch. 2008ff.** Mannheim: Institut für Deutsche Sprache. [www.owid.de/](http://www.owid.de/).
- Schumacher, H., Kubczak, J., and Schmidt, R. 2004.** *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen: Narr.
- Storjohann, P. 2003.** 'The lexicographic use of corpora and computational tools for disambiguation' in Archer, D., Rayson, P., Wilson, A., and McEnery, T. (eds.), *Proceedings of Corpus Linguistics 2003 Conference, UCREL technical paper number 16*. UCREL, Lancaster University.
- Storjohann, P. 2005a.** 'Corpus-driven vs. corpus-based approach to the study of relational patterns' in *Proceedings of the Corpus Linguistics Conference 2005 in Birmingham. Vol. 1, no. 1*. ([www.corpus.bham.ac.uk/pclc/](http://www.corpus.bham.ac.uk/pclc/)).
- Storjohann, P. 2005b.** '*elexico* – A Corpus-Based Monolingual German Dictionary.' *Hermes, Journal of Linguistics* 34. Århus, 55-83.
- Storrer, A. 2001.** 'Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie' in Lemberg, I., Schröder, B., and Storrer, A. (eds.), *Chancen und Perspektiven computergestützter Lexikographie* (Lexicographica Series Maior 107). Tübingen: Niemeyer, 53-69.
- Trap-Jensen, L. 2002.** 'Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases' in Braasch, A., and Povlsen, C. (eds.), *Proceedings of the Tenth EURALEX International Congress*. Copenhagen: CST, 503-508.

### The Institute for German Language (Institut für Deutsche Sprache)

is the central non-university research institution for the study and documentation of the structure, the contemporary usage, and the recent history of the German language. The IDS was founded in 1964 in Mannheim and is still located there. The research is carried out in three departments, the Department of Grammar, the Department of Pragmatics, and the Department of Lexical Studies. Project clusters on corpus linguistics and research infrastructure, as well as the Central Data Processing Section and the Public Relations Section, complement the work of the institute.

In the Department of Grammar, the grammatical structures of German are identified and described, also including their comparison to other languages. The Department of Pragmatics carries out research on spoken German and linguistic behaviour in conversation. Within the Department of Lexical Studies (Lexik), three main lines of research are currently pursued, (i) lexicology and lexicography from a cultural-historical perspective, (ii) lexical theory with respect to syntagmatic aspects of lexical items, and (iii) corpus-based Internet-lexicography as described in the article at hand.

[www.ids.mannheim.de](http://www.ids.mannheim.de)





## KirzeN, the international dictionary

Balázs Sipócz



**Balázs Sipócz** is COO of KirzeN. He heads the production of the site and defines strategies to achieve the allotted goals. He has been working on Web projects with Docler Holding, including some of the largest Websites on the Internet, since 2004.  
sipocz.balazs@kirzen.hu

KirzeN is a business partner of K Dictionaries.

KirzeN Hungary Ltd. is the youngest division of Docler Holding, one of the most successful and continuously growing companies in Hungary. It was founded and is owned by György Gattyán.  
<http://kirzen.com>  
<http://docler.com>

Photo of the new Docler building in Budapest.



KirzeN is an innovative online dictionary created by an ambitious company in Hungary that dared to dream big. The idea of the Website stemmed from an urge to create something that could become unique, useful and positively exciting. After a long period of studying the current market leaders in online translation and the main dictionary sites, it was concluded that the overwhelming majority use either a multilingual dictionary core or language pairs that can be selected at the main input field to allow two-way searches between two languages. Furthermore, we noted the meagre availability of “minor” languages and the lack of functions for non-verbal information. Thus, we decided to develop an online dictionary that includes a wider range of languages, actually aiming to cover as many living and extinct languages as possible, and that provides users with an easy-to-handle interface in different languages—for free.

KirzeN is not only a tool for looking up words and phrases in diverse languages; it is based on a brand-new concept that combines high-quality and well-structured lexicographic data together with the advantages of social networking and file sharing. The site is now available on the Internet under several domains (.com, .net, .eu, .us, .in, .cn, etc.) in 16 interface languages. It is possible to search for translations and definitions in more than a million items in over 40 languages, including most of the European languages and the major Asian ones. Each word and phrase has its own sub-page that is easily accessible by clicking on the relevant record. The page contains all of the available lexicographic content as well as the uploaded multimedia content for that word or phrase. After the registration process, users can fill out their own profile and contact other members of the community. Registered members gain access to the KirzeN Workshop, where the action takes place. There they need to choose a mentor to cooperate with. They

can upload new translations, definitions, pronunciation samples, pictures or videos for any entry. Since we are committed to allowing only quality content to appear on the site, there is a strict reviewing system. Each uploaded item has to gain the approval of two reviewers before being registered in the records. The reviewer, who should be an active member of the community and who contributes often to editing the entries, may become a mentor. The mentor makes decisions about questionable content and helps newcomers to become acquainted with the rules. Members gain rewards for quality work, in the form of points that are redeemable for valuable gifts in the KirzeN Shop. Points are given for approved content according to the difficulty level and language competence. In addition to the community functions, users can communicate through a private messaging system, and soon we will launch our own instant messaging system as well. In view of the success of Wikipedia and YouTube, we strongly believe in the power of the community and that our members will share and gain knowledge through such a dictionary.

After nearly one year of preliminary work developing an appropriate database structure and the interfaces for searching, uploading and editing content, most of the planned functions are up and running. At present the site's main goals are to attract mentors, extend the number of the languages in the database, and bring to perfection the grammatical component of the entries in the dictionary. Obviously, KirzeN is not only for surfers who need just a quick translation; it has more to offer to users worldwide who are generally interested in a language, wish to discuss their experiences and ideas, and are ready to elaborate the dictionary by contributing to or editing new entries. Moreover, professional lexicographers, experts on “minor” languages and students are all welcome to make their comments on KirzeN and have their work recognized and appreciated worldwide.

We realize it is becoming more and more difficult to find an online dictionary that suits your needs, and we hope and believe that with KirzeN such needs will be fulfilled. Admittedly, we cannot forget that developing a site requires continuous hard work, so we do our best to enhance the present features and attain the planned ones. Hopefully, the presence of KirzeN, the international dictionary, on the World Wide Web will satisfy many users.

## TheFreeDictionary.com story

Michal Goldman

When Steve Boymel and Nick Simonov, co-founders of Farlex, Inc., launched TheFreeDictionary.com in 2003, they made it their mission to create the world's most comprehensive online dictionary. TheFreeDictionary (TFD) debuted as a simple, free online dictionary catering to English-speaking users, but it has since evolved into one of the most authoritative and comprehensive reference sites on the Web and now serves a global network of users. A prime online tool for active information seekers, TFD has been accessed by more than 1.6 billion visitors.

Simonov and Boymel founded TFD as a basic dictionary of everyday terms, and their seamless partnership has allowed the site to grow by leaps and bounds. It now hosts scores of dictionaries in dozens of languages, several encyclopedias, a literature reference library, and a search engine.

While Simonov oversees the architecture and technology that powers it, Boymel focuses on the site's day-to-day operations as well as the company's business dealings and partnerships. The two lead a dedicated core team of six individuals who meticulously maintain the site's content and create daily articles for the homepage. They, in turn, are supported by an extensive network of about 50 contractors.

Farlex has been able to turn TFD into a preeminent reference destination by partnering with the elite among publishers of reference material, including the Houghton Mifflin Company, Saunders, Cambridge University Press, Columbia University Press, Reed Elsevier, Encyclopædia Britannica, HarperCollins, K Dictionaries, Larousse Editorial, Research Machines, The Computer Language Company, Gale (Cengage), and the McGraw-Hill Companies. The wealth of educational content from these and other publishers is organized into a number of easy-to-use sections: Dictionary/Thesaurus; Medical Dictionary; Legal Dictionary; Financial Dictionary; Acronyms and Abbreviations; Idioms; Encyclopedia; Wikipedia; and Hutchinson Encyclopedia. Convenient, color-coded tabs located at the top of each search page allow users to quickly navigate between these different sections and gain a broader understanding of researched topics.

Each day, the site's English-language homepage features a number of educational items such as "Word of the Day," "News of

the Day," "Article of the Day," "This Day in History," "Today's Birthday," and "Quote of the Day." These items are exclusively created by the site's editorial team and link to relevant dictionary and encyclopedia entries within the site. Much of this content is also available through the popular "Word of the Day" email, a free service provided to more than 25,000 subscribers.

The user-friendly interface available at TFD appeals to a wide range of Internet-savvy visitors seeking a personal window to the Web. The fully customizable format—which allows users to simply add, remove, drag, and drop content windows to create personalized homepages—can be accessed from every modern browser without any installation, download, or additional software. The assortment of customizable offerings includes a collection of vocabulary games designed to entertain and educate, an innovative mail feature that allows users to view messages in their email inboxes, and a picture option that lets them upload their favorite photos for display directly on their homepage. In addition, the syndicated news directory offers hundreds of RSS feeds from dozens of reputable publications, websites, and news sources. Because TFD generates revenue through advertising, all of the site's services and tools are available to users free of charge.

TFD's English dictionary has many unique features that allow users to better investigate and understand searched terms. In addition to the standard part-of-speech/definition/etymology information, the site provides users with audio files containing both American English and UK English word pronunciations, a thorough list of synonyms and antonyms, and usage examples from both classic literature and recently published articles. Users are also afforded the option of translating words into a host of different languages and can do so by simply selecting the desired language from a drop down menu.

At TFD, the foremost priority is the users—as many as 2 million visitors search for information on the site every day—and Farlex's editors pride themselves on the personal attention they give to each and every email they receive. Over the years, TFD's staff has reviewed tens of thousands of questions and comments, and many of the site's unique features have come as a direct response to user requests. One such example is the "My Word List" tool, which allows users to create a personalized list of



**Michal T. Goldman** graduated summa cum laude from the University of Pennsylvania in 2005 with a degree in Visual Studies. She has been an editor for Farlex since 2007.  
michal.goldman@farlex.com

Farlex, Inc. is an independent, privately-held provider of online reference products based in Huntingdon Valley, PA. Founded in 2004, Farlex provides innovative, easy-to-use reference and learning tools. Its flagship websites, TheFreeDictionary.com and TheFreeLibrary.com, contain over five million pages of books, periodicals, and reference information from the world's most respected sources.  
www.farlex.com

words and terms on the site and easily access their definitions and encyclopedia entries. The feature has become an invaluable learning tool for people from all walks of life, including students, professionals, and people learning a new language.

Over the years, TFD's users have repeatedly expressed interest in a feature that would allow them to provide the site with their own definitions, and, in 2007, Farlex addressed their requests by launching a second dictionary site—Definition-Of (www.Definition-Of.com). The site is built around user-contributed definitions and allows visitors to rank the quality of existing entries.

In May 2008, Farlex announced the largest single addition of reference material to TFD, including, for the first time, foreign language resources. The overwhelmingly positive user response to those four initial foreign language dictionaries encouraged the development team to further enhance the site's multilingual offerings, and, just months later, Farlex released TFD's newly expanded translations section, which allows users to translate dictionary terms from English into more than 40 different languages. At the same time, the site's developers launched the homepage in 10 new languages: Chinese, Dutch, French, German, Greek, Italian, Norwegian, Portuguese, Russian,

and Spanish. By expanding the scope of its resources to include titles outside of the English language, TFD addresses the multilingual needs of its ever-expanding user base.

While TFD does not track user information, IP address data indicate that about half of the site's users are based in the US. International Web surfers compose the remainder of its clientele. Many users are not native English speakers and use the site to improve their English skills or even teach themselves the language. The unique multilingual translator feature, which allows users to easily translate words from one language to another and just as easily switch to a third, has become an invaluable tool for such users.

TFD reached an exciting milestone in 2008, welcoming its billionth visitor in just five years. Since that time, the number of reference titles published by the site has expanded exponentially. The Farlex team remains committed to continuously enhancing TFD and is always developing new features that keep it on the cutting edge of online reference site technology. This ceaseless dedication to growing and evolving the site distinguishes TFD from other reference websites and makes it a singular destination for online information seekers.

Farlex is a business partner of K Dictionaries.

## The BLDS Italian dictionary core

Palma Gallana

### Introduction: a dictionary streaming to perfection

The BLDS is an extraordinary instrument. This is what I thought the first time I came across it and what I have become more and more convinced of during my work on it. It is not a common dictionary but a mono-bilingual one, "a dictionary streaming to perfection". It is user-friendly and made to measure for students and speakers of a foreign language at intermediate level, who can find in it all they need to understand the meaning of a word and use it correctly.

The dictionary is bilingual, built upon a monolingual base. Definitions and examples are in the source language, but are adapted in order to be easily understood by someone who is studying and needs to be coached to understand the meaning and use of words. Far from being banal or inaccurate, the definitions are formulated with simple terms, belonging to the dictionary core

itself. Examples are actual, practical, useful and not academic.

Providing this monolingual base is extraordinary for two reasons:

- a propaedeutic purpose, since the user is oriented to a soft monolingual approach before eventually getting to pure monolingual dictionaries.
- a linguistic purpose, since the user tastes the real savour of a foreign language, confronting definitions, examples, expressions, etc. in it.

BLDS is a rationally well-organised, balanced, complete monolingual core, predisposed to become a bilingual dictionary, where headwords are correlated by pronunciation and, together with the examples and phrases, each sense is translated into L2. This satisfies the need of finding punctual correspondences in L2, besides leading to a complete understanding of the word.

Traditional tools such as indicators of sense and use, synonyms and antonyms, etc, complete the work. The result is a fully bi-directional dictionary, interesting and useful for both L1 and L2 users.

### The Italian core: two dictionaries in one

“How fantastic it would be to have an Italian BLDS core”, I thought, and the day the adventure began I was very excited to have the privilege to edit it myself. It was different from the traditional dictionaries I had worked on, and I found myself with the opportunity to finally make the dictionary I had lacked during my years of study and work. Though far from perfect, the actual Italian core is the result of great efforts, with its merits, qualities and limits.

Traditional monolingual dictionaries explain senses of words and eventually illustrate them with one or more examples. In medium-size dictionaries the definitions are often non-extensive, reduced to a couple of synonyms. Only bigger dictionaries usually add one or two extra synonyms in a special section, but that occurs only for few terms, not systematically, and synonyms are treated as secondary, information. The result is that if you need to find a synonym or an alternative term to avoid repetition you have to look it up in a specific synonym dictionary. This second dictionary might not have the same division of senses as the first one, so you have to start your query again.

Exhausted by this double check, when I started editing the Italian core, I had the idea to make two dictionaries in one: a monolingual core with extensive definitions and an additional synonym apparatus. This apparatus comes as the first element of the semantic section. The entry in fact is structured so as to introduce the sense of a word (after the **morphology** section) starting from the **context** it deals with/ belongs to and the **synonyms**. Then, an extensive **definition** explains the meaning of the headword and one or more **examples** and/or **idiomatic expressions** illustrate it. Moreover, the definition is preceded by a short **sense indicator**, meant as a substitute of the definition itself in the edition exclusively for native Italian speakers.

### morphology

If irregular, the feminine and plural forms are given for nouns and adjectives. For intransitive verbs, the auxiliary is specified, and accurate distinction is made between intransitive pronominal and reflexive verbs. Specificities concerning verb tense/mood required by syntactic words (prepositions, conjunctions, etc) or concerning elisions,

euphonic forms or combined forms are given in explanatory notes.

### context

A lot of effort has been put into the lexicographical structure, which provides specific **indicators of use and sense** throughout the dictionary. **Register labels**, which identify the levels of expression beyond standard language (familiar, vulgar, bureaucratic, euphemistic, etc); **sense qualifiers** or rhetorical indications, which denote particular expressive modalities (figurative, extensive, ironic, pejorative, etc); and **sector/field indicators**, which are labels for specialist terms (anatomy, economy, sport, etc) have been accurately tracked in order to guide and/or alert the reader during the consultation. The **range of application** of the word, that is to say the object, person or animal it refers to, is also given when useful or appropriate.

By contextualising the meanings with the multiple and abundant sense indicators, the dictionary helps the user to identify the most suitable and translation.

### synonyms

About ninety percent of the headwords are correlated by one or more **synonyms**. They appear right after the grammatical information and introduce the semantic section. Synonyms act as a kind of accordion that permits the semantic dimension of the headword to extend. They do this by proposing an equivalent to the headword that is on the same level and can replace it. Such a range of words that are close to the lemma gives the user an idea of the sense he/she is going to encounter.

This approach is very useful for multiple reasons:

- the user might already know one of the synonyms, so the headword becomes less unfamiliar.
- by setting the headword in the company of terms that mean the same, different channels with different possible contexts open in the mind of the reader, and thus he/she gets a wider panorama of the possible nuances that a headword may carry within a given sense.
- a set of synonyms can turn out to be a good didactic tool to learn new words and establish linguistic associations.
- during a search for a word in the electronic version of the dictionary, a term that is not registered as a headword may be found among the synonyms. In this case it is possible to get to the meaning of the synonym or approach it by means of association with the headword it is linked to, which has its proper definition, example and translation. Since the number



**Palma Gallana** graduated in 1993 in Russian language and literature from the Catholic University of Milan and has since been working in linguistics and lexicography as author, editor and translator. She masters five modern languages, is passionate in particular about grammar and lexicon, and has edited dozens of grammar books and dictionaries in various languages, especially oriental (Arabic, Chinese, Japanese, Hungarian). Dr Gallana was chief editor of the Russian/Italian dictionary published by A. Vallardi in 2005 and of two slang dictionaries (American and English) by the same publisher. She also compiled an economic-financial dictionary, several grammar manuals and exercise books of the Russian language for Italian natives and of Italian language for Russian natives. Her cooperation with K Dictionaries began with a French-Italian dictionary translation in 2007. She is the chief editor of KD's new Italian dictionary core.  
gallana.palma@hsr.it



## An extract from a BLDS Italian entry:

**andare**<sup>1</sup> [an'dare] *vi* (aus. *essere*) **1** =muoversi; {*senza meta*} spostarsi senza una meta precisa ◇ *andare a piedi* ◇ *andare in bicicletta* ◇ *andare piano* ◇ *andare a spasso* ◆ **andare a zonzo** passeggiare senza meta ◆ **andare d'accordo** *fig* trovarsi bene insieme ◆ **andare fiero** mostrarsi fiero ◆ **andare sul sicuro** non rischiare ◆ **andare a naso** *fig* seguire l'istinto **2** =dirigersi; =recarsi; =procedere; {*verso una direzione / meta*} spostarsi in una direzione o verso una meta ◇ *andare avanti* ◇ *andare a casa* ◇ *andare in vacanza* ◆ **andare in onda** essere trasmesso alla televisione o alla radio ◆ **andare in scena** *theat* essere rappresentato ◆ **andare a nozze** *fig* trovarsi molto bene in una data situazione **3** tempo: =passare; =trascorrere; =volare; indica lo scorrere del tempo ◇ *Due settimane se ne sono andate solo per scegliere la località.* **4** {*essere destinato*} essere destinato ◇ *Questi CD vanno sul primo scaffale.* ◇ *La forchetta va a sinistra, il coltello a destra.* **5** abbigliamento: =essere; {*stare*} stare in un certo modo ◇ *Le scarpe mi vanno strette.* ◇ *La giacca mi va alla perfezione.* ◆ **andare a pennello** calzare o vestire bene **6** attività: =procedere; =svolgersi; avere luogo o avanzare in un certo modo ◇ *Come vanno le cose sul lavoro?* ◇ *Va tutto bene.* ◆ **andare a gonfie vele** ottimamente ◆ **andare a buon fine** concludersi con successo **7** apparecchio: =funzionare; adempiere la propria funzione ◇ *La televisione non va più.* **8** =diventare; =modificarsi;

of headwords is limited, as is the scope of the dictionary, synonyms help to create a virtual extensibility.

### antonyms

**Antonyms** too occur, though less frequently than synonyms, and only in cases when it is useful to highlight the headword by means of semantic contraposition or when it is necessary to clarify the sense of a headword by saying what it is not about.

### sense indicator

The sense indicator is a word or brief phrase that immediately evokes the context for collocating a particular sense of the headword and offers an intuitive association for it. Since it is represented by a word or group of words (glosses) that frequently appear together with the headword, it forms typical combinations that a native speaker would consider as natural.

The sense indicator can specify:

- the **general context** by means of a complement (in literature, of an animal, etc). The sense indicator is sometimes represented by part of the definition, if general enough. Since the definition and sense indicator do not appear together in the same dictionary version, this won't generate redundancy.
- a **category** by means of a generic term, the **hypernym**, to whose field of application the headword belongs as one of its particular representatives ("animal" for 'horse', "profession" for 'lawyer', etc).

The sense indicator can also be represented by:

- a **generic synonym**, that is to say the most generic among synonyms, usually used in the formulation of the definition, that gives the sense of the headword and eventually can replace it.
- a **specific hyponym**, that is to say a particular representative of the headword, that plays in such a case the role of generic term.

Particular attention has been put into avoiding redundancy as concerns the "synonym-definition" section and "synonym-sense indicator" section. Since synonyms are visible always, and definition or sense indicator alternatively, a term that appears in the definition may occur as a sense indicator or one that appears as a sense indicator may appear in the definition, but neither will occur as a synonym (e.g. *pentimento* **1** =mortificazione (rimorso) rimorso per ciò che si è fatto **2** =rammarico (rimpianto) rimpianto per ciò che non si è fatto).

### definition

The definition is a brief (approximate) description in simple Italian, formulated

with words from the dictionary's headword list. Definitions are succinct and clear, as often as possible are autonomous and do not contain derivatives of the defined word.

The definition usually centres on the semantic nature of the term; however, especially in case of grammar terms, it can explain the grammatical-syntactic function and application.

### examples

The contextualization of terms is achieved by means of up-to-date natural examples, which reflect real contemporary Italian language. Examples are rather short and use simple words from the dictionary's list of headwords. They can be **syntagms** (usually) or **full sentences** (if they are the only or best way to show the commonest use of a term).

### idiomatic expressions

**Idioms** and **set phrases** (and relative examples) with their metaphorical evocative meaning, not only have a particular linguistic "taste" but also help impart an idea of Italian culture, history and life.

### selection criteria

Headwords, examples and expressions have been selected so as to offer a wide panorama of the actual, active and living language. The **headword selection** ranges from spoken/informal words to the terms of specialist sectors (e.g. finance, technology, law, etc) which occur most frequently in today's communication. The extremes, too specific and too slangy, have been excluded in order to collocate the dictionary on a well-balanced middle level.

Within the headwords and sub-entries the meanings are ordered according to frequency of use and logical consequence (literal before figurative or extensive, general before specific or sector, etc).

### Conclusion: work is always in progress

The conclusion is that now that the core is finished, it is time to start it again. At present there is a core, quite rich as far as lexicography is concerned. But there are many components that are missing and might be added, especially on a morphological level (informative notes on conjugations, or references to conjugation tables; degrees of comparison for adjectives; augmentative, diminutive and other altered forms; syllabication; etc). The headword list can be augmented and extended so as to include also prefixes, suffixes and acronyms. And many more integrations could be conceived, since "streaming to perfection" never ends. Work on the second phase of the Italian dictionary core is currently underway.



## BLDS sample pages

Page from the Japanese-French  
section of Dictionnaire Assimil  
Kernerman Japonais-Français  
Français-Japonais (below)

ko「re ㄣ kushon

256

japonais-français

ko「re ㄣ kushon コレクション *n* collection *f*◇ コレクションに加える *korekushon ni kuwaeru ajouter qqch à sa collection*ko「rema ㄣ de これまで *n*1 jusqu'à maintenant ◇ これまで誰も登ったことのない山 *koremade dare mo nobotta koto no nai yama montagne que personne n'a jamais gravie*2 fin *f* ◇ しんどい仕事もこれまでだ。 *Shindoi shigoto mo koremade da. J'en ai fini de la corvée.*ko「re ㄣ ra これら *pron* ceux-ci / celles-ciko「ri ㄣ ru こりる 懲りる *vi* agir en tenant compte de ses échecs ◇ 酒に懲りる *sake ni koriru échaudé par de mauvaises expériences, renoncer à l'alcool*ko「ritsu こりつ 孤立 *n* [+suru] isolement *m*◇ 世界から孤立する国 *sekai kara koritsu suru kuni pays qui s'isole du reste du monde*ko「ro ㄣ ro ころ 頃 *n*1 temps *m*, époque *f* ◇ 子どもの頃の思い出 *kodomo no koro no omoide souvenir du temps où j'étais enfant*2 moment *m* opportun ◇ 頃を見計らう *koro o mihakarau choisir le moment opportun*ko「robu ころぶ 転ぶ *vi* tomber ◇ 雪道で滑って転ぶ *yukimichi de subette korobu tomber en glissant sur le chemin enneigé*ko「rogaru ころがる 転がる *vi*1 rouler ◇ 石が坂を転がる *ishi ga saka o korogaru les pierres roulent dans la pente*2 tomber ◇ つまずいて地面に転がる *tsumazuite jimen ni korogaru tomber par terre en trébuchant*ko「rogasu ころがす 転がす *vt*1 faire rouler ◇ ボールを転がす *booru o korogasu faire rouler un ballon*2 renverser ◇ コップを転がす *koppu o korogasu renverser un verre*ko「rokke コロケ *n* croquette *f* ◇ コロケを揚げる *korokke o ageru faire frire des croquettes*ko「rokoro ころころ *adv*1 en roulant ◇ 石がころころ転がる *ishi ga korokoro korogaru la pierre roule*2 avec des formes arrondies ◇ ころころ太った子猫 *korokoro futotta koneko chaton rondelet*3 brusquement ◇ 話がころころ変わる *hanashi ga korokoro kawaru la conversation change brusquement*

عربية-فرنسية

234

سَمِعَ

beurrier *m*, crémier *m* *nm* [sam:'a:n] سَمَّانْNous avons *إِشْتَرَيْنَا سَمَّانًا مِنَ السَّمَانِ* ◇

acheté du beurre chez le crémier.

fertiliser *v* [sam:'ada] سَمَّدَIl a fertilisé le champ. *سَمَّدَ الْحَقْلَ* ◇empoisonner *v* [sasm:'ama] سَمَّمَIl a empoisonné le chien. *سَمَّمَ الْكَلْبَ* ◇corrompre les idées de qqn *سَمَّمَ أَفْكَارَهُ* ◇empoisonner les relations entre *سَمَّمَ الْعِلَاقَاتِ* ◇engraisser *v* [sam:'ana] سَمَّنَIl a engraisé le veau. *سَمَّنَ الْعِجْلَ لِيَذْخُلَ* ◇

entendre, ouïr

Il a entendu une voix.

écouter

Il a écouté la conférence.

obéir, céder

Il a écouté son père.

réputation *f*

bonne réputation

mauvaise réputation

s *m pl*

frais

volant

m

n

le poissons

ture *f*

pêche

m

grillé

frit

ur *f*

e est

centimètres d'épaisseur.

m

hé deux poissons.

sons

, poison *m*

lé le poison.

f *pl*

mortel

nf [sam:'a:'a] سَمَاعَةٌ

ur *m*, combiné *m*

ur du téléphone

ope *m*

n

raisonné la viande au sumac.

nm [sam:'a:k] سَمَّاك

nier *m*, marchand *m* de poissons

illait comme poissonnier.

س

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

سَمَّاك

Page from the Arabic-French  
section of Dictionnaire Assimil  
Kernerman Arabe-Français  
Français-Arabe (above).

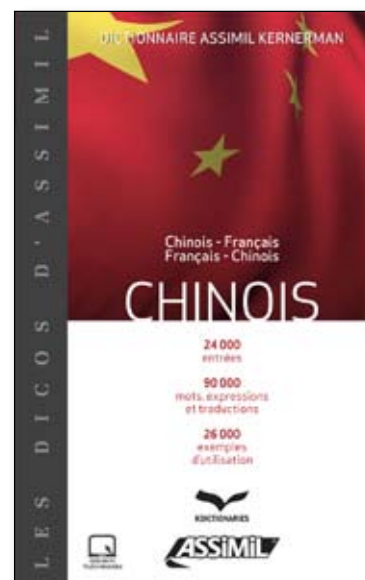
# Dictionnaires Assimil Kernerman, 2009

Published by Assimil, Paris.  
<http://assimil.com>

Electronic applications available  
from  
<http://lesdicosdassimil.com>



Arabe-Français / Français-Arabe  
ISBN 978-2-7005-0446-0 · 1111 pages



Chinois-Français / Français-Chinois  
ISBN 978-2-7005-0436-1 · 958 + 28 pages



Grec-Français / Français-Grec  
ISBN 978-2-7005-0448-4 · 1012 pages



Japonais-Français / Français-Japonais  
ISBN 978-2-7005-0445-3 · 1278 pages



Polonais-Français / Français-Polonais  
ISBN 978-2-7005-0449-1 · 1078 pages



Portugais-Français / Français-Portugais  
ISBN 978-2-7005-0438-5 · 952 pages



Russe-Français / Français-Russe  
ISBN 978-2-7005-0437-8 · 1211 pages



Turc-Français / Français-Turc  
ISBN 978-2-7005-0447-7 · 1078 pages