The Innovative PASS-Q-WORD
Robert Stock

In just a few months, the Pass-Q-Word computerized semi-bilingual dictionary for non-native learners of English will be released to the market. This dictionary is being developed as a joint project of Password Publishers Ltd and the Q Group PLC, combining the expertise of the two companies to produce a unique product.

Password Publishers, who have had international success with its series of semi-bilingual English learner's dictionaries, have provided the textual material and editorial work. The Q Group, a public company trading in the UK with a list of over 30 multimedia language learning courses, is the designer and developer. The partnership of these two firms in the production of the Pass-Q-Word dictionary insures that the product will have the highest content quality, while being an electronic-interactive dictionary that is both technically sound and fun to use.

Pass-Q-Word contains about 25,000 entries in English, including short definitions and examples of use, along with a translation into the user's native language. The contents include hundreds of new, up-to-date entries relevant to today's life, eg: affirmative action, AIDS, airbag, ATM, ATC, bar code, beeper, body language, carphone, cloning, CPU, conference call, couch potato, drunken driving, e-mail, endangered species, environmentalist, genetic engineering, global village, hacker, information superhighway, Internet, intranet, IOU, labour dispute, laptop, lazer printer, netiquette, personal watercraft, politically correct, on-line, optical scanner, serial killer, sexual harassment, smart card, spam, surrogate mother, touch-screen, up-to-the-minute, virtual reality, voice mail, Web site, win-win, WYSIWYG, yuppy, zap, etc.

The extremely attractive multimedia environment and simple user interface allow for ease of use, with options and paths clearly indicated and marked. The original drawings provide additional semantic clues for the user, which can be optionally viewed full-screen. In addition, the dictionary includes crossword puzzles, to give the learner the opportunity to practice the new words (and hear them at the same time).

The Pass-Q-Word dictionary utilizes the interactive nature of the computer to provide the ideal tool for students of English, combining the growing demand for more efficient ways to learn the language with the increased use of the computer in all aspects of life.

The main advantage of an electronic dictionary is the availability of features that are not possible with a printed dictionary, and which significantly reduce the time required to retrieve an item. In Pass-Q-Word, all the information available for any headword is immediately accessible and presented on the screen, along with the sound of the word, a drawing (where available) and a translation, eliminating the need to "turn" pages. Furthermore, there is a direct link between the dictionary and any text that the user is working on, as well as the ability to "cut and paste".
Some of the unique features of Pass-Q-Word are listed below:

- 25,000 English headwords including the latest entries
- sophisticated search routines to find a headword or idiom
- ability to access words through the native-language index as well as via the English headword
- presentation of different meanings associated with the headword along with its part of speech
- links between words and other information associated with that word
- human voice pronunciation (British and American) of every headword
- ability to record one's own pronunciation of the headword, and compare it with the default pronunciation
- ability to copy the headword, example, definition and translation, and to use it in other Windows applications such as Word
- graphic representation of the words previously looked at (history) so that it is easy to retrieve them
- original drawings to illustrate specific items and concepts
- separate index of special categories (e.g., parts of speech, punctuation)
- separate index of drawings
- separate index of the crossword puzzles
- context-sensitive help (in the native language)

Pass-Q-Word requires Windows 95. It comes on a CD-ROM, or as part of a network server. Its core can be loaded on the hard disk, to be accessed while the user is in a different Windows application.

[Windows and Word are trademarks of Microsoft Corporation.]
GLOSSER as a Practical Application of the Semi-Bilingual Dictionary Concept

Margit Langemets

Margit Langemets heads the Department of Lexicology at the Institute of the Estonian Language, and specializes in monolingual lexicography, dictionary typology and criticism, and computer applications. She graduated in Estonian Philology from Tartu University, then taught Computational Lexicography there. She initiated the publication of the semi-bilingual dictionary in Estonia, and was among its translators. Her current projects include completing the editing and computing of the Defining Dictionary of Literary Estonian, preparing a Lexicographic Text Corpus of Estonian, and computerizing several Estonian dictionaries.

The European Union is taking an increasing interest in possible ways of cooperating with Central and Eastern Europe, and so, even before these countries become members of the EU, the real cooperation has already begun. Estonian linguistics is not an exception to this.

The enormous information flow that reaches us every day via natural language makes one feel that it would be quite easy to "lose wisdom in a mass of knowledge, lose knowledge in a mass of information, lose information in a mass of data" (a quotation from Thomas Eliot). The Fourth Framework Programme COPERNICUS 1994 financed language engineering research, reusable language resources and several pilot applications. One of the aims of this kind of joint research project should certainly be bringing together academic scholars and people from industry to prove that "everything intended to be scientific need not necessarily be slow and unsaleable" (the words of a Hungarian colleague). For the first time, Estonia had a chance to participate in five such projects. One of them was GLOSSER, on which people from Bulgaria, Estonia, the Netherlands, France and Hungary worked together for a period of two years.

The result of this language technology project is a program called GLOSSER, designed to support the processes of reading and learning to read in a foreign language. This is the prototype of a system where a computer is used as a reference, not a language teacher, and assistance is offered to advanced learners who are not afraid of "machines" and who find it exciting and useful to use the computer application beside or instead of the tedious task of thumbing through a dictionary. GLOSSER differs from an ordinary dictionary lookup program in its analysing procedure that appears right on the screen. The starting point was a question arising while reading a text: How can one find the lemma and the right sense in the dictionary while meeting several kinds of word forms in the text? GLOSSER is meant as a tool for finding an answer to this question.

System architecture connects modules for morphological analysis and disambiguation, dictionary access and corpora search with an output model. Let us have a closer look at each of them.

Morphological Analysis and Part of Speech Disambiguation

Morphological analysis is necessary if one wishes to consult an on-line dictionary. GLOSSER was fortunate in having access to the Xerox POS Disambiguator for English language which, drawing on the Morphological Analyser, picks up the correct part of speech out of all possible morphological descriptions. The theoretical base of the disambiguator is English Constraint Grammar, a theory from the late 1980s, which determines the function of the word using special rules for morphological characteristics and context. These rules are the constraints.

Disambiguation means getting rid of such morphological descriptions that do not fit the specific context of the word, while semantics is not taken into account. The disambiguator makes its decision after looking through the whole sentence. Homonymy differs in type and extent in different languages. For example, English is noted for part of speech homographs, whereas Estonian is noted for homonymy of morphological forms. (In Estonian the number of word forms is very large: on average there are 33 different forms per word. About 40% of Estonian word forms are ambiguous.)

GLOSSER is a system that examines a sentence word by word, and there is only a word-based access to the dictionary. For example, in Xerox codes the input sentence The concert was nothing to write home about will be analysed word by word by the disambiguator (the+AT; concert+NN; be+BEDZ; nothing+PN; to+TO; write+VB; home+NN; about+IN ), although here is a multiword expression nothing to write home about with its dictionary definition at the end of the entry for home (placed after two parts of speech, several derivatives and several multiword expressions containing the headword). And how should a user know that this expression is located under home, but not under nothing? It should be the next stage of the research project to deal especially with multiword expressions and to enable an expression-based access to the dictionary. First, the system should check the possible belonging of a word in an expression, and if the answer is yes, then, secondly, display only this part of the dictionary entry (not the whole one)."}

The Dictionary

Reusability of lexicographic resources is a widespread trend in computational lexicography today. The only feasible option is to use an existing dictionary. For GLOSSER, the most suitable candidate was Password
mainly for its belonging to the semi-bilingual type of dictionary, but also for its appropriate size (25,000 headwords). The source language is represented by a headword, grammatical information, sense explanation and illustrative sentences. The target language is represented by brief translations for each meaning of the headword (a total of 37,000) or sub-headword (derivatives, multilingual expressions). The semi-bilingual dictionary is new and unique in Estonia, and GLOSSER was very happy to find this combination of monolingual and bilingual dictionaries in one volume.

We obtained the electronic version of the dictionary text in layout format, the so-called typographic view, which is concerned with the two-dimensional printed page. These layout codes had to be stripped and converted into a suitable format. Our task was to analyse the typographic view (the raw text format) fully, to be able to transform the text into the lexical view, i.e. lexical data as those might appear in a database, without concern for their exact textual form.

The list of headwords was sent to Xerox for testing. Prof. L. Karttunen made a network from the list and checked them against the English transducer Xerox supplied for GLOSSER. About 400 headwords were not recognized by the analyzer and needed to be added to the system. For example a) British spelling for words that are in the morphological analyzer only with American spelling (apologise, ardour, etc.); b) French words in their original French orthography (café, courtège, etc.); and c) words that are not found in the American Heritage Dictionary (casuaria, dhoti, kampung, rambutan, etc.). The latter words originate from several other local editions. The Estonian version was supplemented with 'kroon', 'sprotsid', etc.

Usually, the microstructure of a dictionary is hierarchic and, depending on type, rather complex. The conservative form of a traditional printed dictionary, because of its implicit information, is satisfactory for users, but not for various computer systems, which require information types to be set out explicitly.

For encoding the text of the printed dictionary TEI Guidelines were consulted. The encoding format has to adhere to the rigorous principles of traditional dictionaries and present them in such a way as to facilitate dictionary reusability and automatic processing. The Guidelines use the Standard Generalised Markup Language (SGML) to define their encoding scheme. It provides for a formal definition in terms of elements and attributes, and rules governing their appearance in a text. A dictionary is seen as a linear text stream interspersed with markup. The tags provide an indication of the content of the fields they delimit. Each of the information fields has an opening marker </> and an end marker </>. One field can embed another.

• Entry
  <entry> contains a reasonably well-structured dictionary entry. <hom> and <sub> mark the sub-
division of entries into part-of-speech homographs and sub-headwords (derived words, multiform expressions, idioms), respectively. The tags serve to group information relating to each component. The attribute (type=xref) marks a cross-reference. Entries comprise several constituent parts (form, sense, usage, etc), each providing a different type of information about the word treated. Exceptional cases are still characteristic of lexical data. Information of the same kind can appear at different levels in the same entry.

• Form
  <form> is the first item in an entry. <en> gives the print-form of the headword. Orthography and stress of a single element lexeme are not separated. <pr> contains the pronunciation(s) of the word. Other information, such as variant or alternative, abbreviated and full forms, inflected forms illustrating the inflectional pattern of the headword, orthographic form(s) for displaying gender contrasts, negative word or use, and comparison, is presented by additional attributes (type=var / abbr / full / infl / gen / neg / comp). Collocations (T-shirt under T) of the headword to show multiple-word lexical items are also marked.

• Grammatical Information
  For encoding, the tag <gr> is used to group all grammatical information about a lexical item. Usually it consists of <pos> for indicating the part(s) of speech. But in addition there might be <colloc-type=prep> for prepositions and <nr> for the grammatical number associated with a form. The 'plural' specification may apply either to a) the inflected forms provided (passers-by), or b) the headword itself (noun plural).

• Sense Information
  <sense> stands for for semantic description. It groups information (forms, grammatical information, usage, translation(s), etc) about the given sense of a word. Attribute (n=’...’) indicates the sense number. <df> contains the text of the definition. Definitions describe the meaning of some lexical item, most often of the headword of the entry, while in some cases they describe examples. <ee> contains the Estonian translation text. Definitions and translations are usually accompanied by examples. <ex> contains an example text with at least one occurrence of the word form, used in the sense being described, in it. Examples may still contain other elements, eg. <pr>, <usg>, <pos>, etc.

• Usage Information
  Most dictionaries provide restrictive labels and phrases indicating the usage (<usg>) of given words or particular senses. Attributes help to define usage more precisely. A distinction is made between a definition and an additional descriptive phrase (type=hint; eg of horses; in football, hockey, etc). Geographic area, national or regional use (type=geo) is marked in some cases (whisky... Irish and American whiskey). Not much is told about regional style (type=reg), but there are style labels, such as 'formal', 'informal', 'offensive', 'rare', 'unkind', etc.
• Cross References
These refer the reader to additional information elsewhere in the dictionary. They may be free-standing within an entry. The metalanguage remains untagged (labels 'same as', 'see', 'see below', etc). <xr> groups the information relating to a cross-reference (a phrase or sentence). <ptr target=...> defines a pointer to another location. Cross reference-like cases occurring occasionally in definition texts remain untagged.

• Notes
Notes about usage, grammar, etc, may be placed within an entry. <note> contains a note or annotation. Notes may give extra information about form (as part of a word; with capital) or grammar (in questions, negative sentences, etc; placed after a noun).

• Related Entries
These are included in many dictionaries for direct derivatives or inflected forms of the headword, or for compounds, phrases, collocations, and idioms. <re> contains a degenerate entry embedded inside a larger entry. It is often of reduced form, consisting mainly of nouns without any sense information.

Text Corpus
The results of morphological analysis also serve as input to corpus search. Lexeme-based search looks for further occurrences of the same string. Up to now GLOSSER has relied on another EC project for bilingual corpora, which involved work in (re)aligning the texts. The corpus should be big enough to cover the 10,000 most frequent words, i.e. ca 5MB.

User Interface
The system is supported by Unix and Windows environments. The main window consists of four child windows: the Input Window, the Disambiguator Window, the Dictionary Window and the Corpus Window. Input may be typed in or read in from elsewhere in the computer memory. After marking the text in the input window, one can either analyse it or look it up in the dictionary, or get different examples of aligned sentences with their translations of the word.

Figure Main Window:

• Notes
1 In the dictionary project COMPASS (finished in 1996) all multiword expressions were coded.
3 C. M. Sperberg-McQueen, L. Bumard (eds), Guidelines for Electronic Text Encoding and Interchange (TEI.P3). Chicago, Oxford, 1994

• References
Dictionaries in Asia and ASIALEX

Amy Chi

Amy Chi graduated in Applied Linguistics from Exeter University, and works as Instructor at the Language Centre of Hong Kong University of Science and Technology. Her most recent publication is a critical user-oriented review of the Collins Cobuild Dictionary. She specialises in pedagogical lexicography, dictionary use, and teacher and learner training, and is currently reading for her doctorate at Macquarie University, with a thesis on pedagogical lexicography.

It was the wish of Dr. Gregory James, Director of the Language Centre of Hong Kong University of Science and Technology, to hold a conference of lexicography in Asia. The dream was realised with a grant from the Japanese government and full support from scholars in the field. A conference entitled "Dictionaries in Asia: research and pedagogical implications", which was organised by the Centre, was held in Hong Kong from March 26 to 29, 1997.

Papers were arranged according to four themes, one for each day:
- The Dictionary-Making Process
- Using Dictionaries in Language Teaching
- Bilingual Lexicography
- Lexicography in Asia.

The aims of the Conference were threefold. First, it aimed to bring together prominent speakers in the field of lexicography to share their expertise and help promote the development of dictionaries in Asian regions. Second, it was to be a rendezvous for lexicographers, publishers and language teachers (English and Chinese) to discuss the making of various dictionaries and their use in language learning. Third, the Conference was also being held with a view to forming an Asian Association for Lexicography (ASIALEX).

To achieve the first goal, we started by inviting distinguished scholars in the field from various Asian countries to be plenary speakers. Since no connection of any kind had existed, we went through quite a difficult stage to locate the whereabouts of potential speakers. After much search, we got hold of their addresses and started writing to them. Since we were not sure whether the information obtained was up-to-date, and in some parts of Asia the postal service is still not very well developed, it was a long and anxious wait. Every reply was received with excitement. Altogether we invited eight prominent figures from countries in the region: Jordan, PR China, Korea, Japan, Singapore, Indonesia, and India.

Since the Conference was targeted at promoting lexicography in the region, besides featuring well-established speakers in the field, we attempted to include potential 'stars' of the future. We were very lucky to receive a total of HK$168,000 from the Japan Foundation Asia Center and Siemens (HK), to provide scholarships for twelve postgraduate students from various Asian countries to present papers at the Conference. The sponsorship was awarded purely on merit and by open competition.

One of the innovative features of the Conference was its focus on the pedagogical implications of dictionaries. I believe there is a triangular relationship among lexicographers, teachers and language students. In this 'lexicographic triangle', with lexicographers and students in two corners, language teachers play an important role in bridging the gap between the sophisticated scholars and the unskilled learners. Hence, it was only natural to make the conference a meeting-place for all those concerned.

On the second day of the Conference, presenters from all over the world discussed research data and experience in using various dictionaries to teach language. The kinds of dictionaries involved ranged from monolingual English or Chinese printed copies to electronic bilingual ones.

Moreover, there were dictionary workshops given on that day by Oxford University Press, Longman Asia and Commercial Press (HK), with the theme "How to use dictionaries to learn reading and writing". Local secondary school teachers of English and Chinese were invited to join workshops as well as the Conference. Over 250 teachers came.

The third goal of the Conference was to form an Asian Association of Lexicography (ASIALEX), which will act as a catalyst for future cooperative research. In Europe, Australia, Africa and North America, pan-continental associations of lexicography already exist - EURALEX, AUSTRALEX, AFRILEX and DSNA respectively. It was hoped that ASIALEX would act as a focus for lexicographic development in Asia, as well as a clearing-house for research and development information and regional cooperation, and with the prospect of attracting grant funding for disbursement.

A Preparatory Committee of ASIALEX was set up four months before the Conference to make all the necessary arrangements. Its members included scholars from Hong Kong and PR China. Moreover, representatives from AFRILEX, AUSTRALEX and EURALEX, namely, Prof. D. Prinsloo, Dr. C. Yallop and Dr. R.R.K. Hartmann respectively, were invited as consultants for both the Conference and the Preparatory Committee.

On the last day of the conference the first AGM of ASIALEX was held. Over seventy participants joined in the proceedings. The
AGM was chaired by Dr. A. Taylor, representative of the Preparatory Committee. It began with a plenary forum with representatives from the various LEXES describing their experience in setting up their associations, followed by Dr. T. McArthur providing an overview of the need for - and benefits of - global links and cooperation between workers in the field of lexicography. An Executive Committee was then elected and ASIALEX was inaugurated in the presence of all the participants.

The ASIALEX Executive Committee (1997-1999) is made up of the following members:

- President: Prof Huang Jianhua Guangdong University of Foreign Studies, P.R. China
- Vice-President: Prof Sangsup LEE Yonsei University, Korea
- Honorary-Secretary: Ms Amy CHI HKUST, Hong Kong
- Honorary Treasurer: Dr Steven LUK The Commercial Press (HK) Ltd, Hong Kong

Committee Members:

- Dr Turki DIAB University of Jordan, Jordan
- Dr LU Gusun Fudan University, P.R. China
- Mr TONO Yukio Tokyo Gakugei University, Japan

The Conference ended with the formal inauguration of ASIALEX. I was very fortunate to have the opportunity to be the convener of the Conference, and I would like to express my deepest gratitude to those who helped in making the event a success. From what I have observed - following a Chinese saying “there are certainly a lot of pearls buried under the hay” - much has been done in various countries in Asia but most, if not all, has not been properly documented and brought to light. In order to disseminate to other parts of the world work that has been carried out in the region, the Language Centre will work on the publication of selected papers presented at the Conference and ensure a wide distribution. In the long run, however, it will be the responsibility of ASIALEX to remedy the situation.

Towards PEOPLEPLEX
Ilan Kernerman

I was thrilled to take part in the Dictionaries in Asia Conference and the inauguration of ASIALEX. The need for a forum of this kind has long been felt, and the event lived up to expectations.

It might seem strange no such framework existed so far, since Asia was the cradle for language and dictionary-making thousands of years ago, and its lexicographic tradition flourished through the ages to modern times. The 20th century’s prominent milestones in pedagogical lexicography stem from the work of M West in India and AS Homby in Japan. Some of the world’s finest dictionaries are made in Japan and its neighbours, as well as valuable research, but they are little known of elsewhere.

In addition to economic-political factors, this lack may be mainly due to Asia’s inherent diversity, not being a homogenous entity of any sort. Linguistically, unlike most European tongues which pertain to the Indo-European family - Asian languages share no common background, apart from being human. That natural human link is true just as well for the entire world. Asia can project a microcosmos of it and, thus, establishing ASIALEX is a significant step toward forming a global lexicographical association.

A future Globalex (or Unilex, in the words of Tom McArthur) concerns globalization and co-existence in multilingual societies. English is the international lingua franca, localized Englishes, effects on the mother tongues, etc, as well as repercussions from hi-tech and tele-com, on-line interactivity and automatic translation, ‘Dictionizers’ and ‘Quicktionaries’, and so on.

This forthcoming forum should not replace national or regional LEXes, but accommodate the varied issues. As such, geography is no sound base for its foundation, nor for the soon-to-come dictionaries that will hardly be what we imagine now.

Beyond countries and behind computers there are people. First of all, and after all. People are the most common denominator for lexicography all over the world.

People in Asia, left to right: Huang Jianhua; Ilan Kernerman; Gregory James; Anne Pakir (Singapore); Turki Diab; P R Subramanian (India)
Kernerman Semi-Bilingual Dictionaries

RECENT TITLES

PASSWORD ENGLISH
DICTIONARY FOR SPEAKERS OF RUSSIAN
Translated by Irina Laduseva, Tatjana Koltsowa, Sophie Troubilov and Larisa Kraskova.
Edited by Ants Pihlak.
TEA Language Centre Ltd.
Tallinn, Estonia 1996
388 pages.
Including a Russian-English Index.

BOOKMAN ENGLISH
DICTIONARY FOR SPEAKERS OF CHINESE
Edited by Jerome Su.
Bookman Books Ltd.
Taipei, Taiwan R.O.C. 1997
ISBN 957-586-668-1
931 pages.

ENSK-ENSK ORDABÓK MED ÍSLENSKUM LYKILORDUM
Íslensk þynging: Geir Svansson
Mál og Menning
Reykjavik, Iceland 1996
ISBN 9979-3-1436-2
632 pages.

ENSK-ENSK ORÐABÓK
Med Íslenskum Lykilorðum
Íslensk þynging: Geir Svansson
Mál og Menning
Reykjavik, Iceland 1996
ISBN 9979-3-1436-2
632 pages.

This semi-bilingual Traditional-Character version contains up-to-date entries, eg beeper, CD-ROM, E-mail, etc, as well as Taiwan and Taiwanese.

PASSWORD ANGLICKÝ
VÝKLADOVÝ SLOVNÍK SO SLOVENSKÝM EKVIVALENTEM 4/E
Translated by Andrea Cániková.
Revised by PhDr. A. Kubišová.
Media Trade - Slovenské Pedagogické Nakladatelstvo (SPN)
Bratislava, Slovak Republic 1997
ISBN 80-08-02489-5
860 pages.

The revised semi-bilingual Slovak version, including a Slovak-English Index, is the only dictionary approved by the Slovak Ministry of Education.

ENGLISH POLISH LEARNER’S
DICTIONARY
Translated by Jadwiga Fisiak, Michal Jankowsky and Tadeusz W. Lange.
Edited by Prof. Jacek Fisiak.
Wydawnictwo Naukowe PWN (Polish Scientific Publishers Ltd.)
Warszawa, Poland 1996
ISBN 83-01-12004-5
946 pages.

Including a Polish-English Index, as well as an innovative Phonetic Index - enabling users to locate an English entry according to its pronunciation, as exemplified in the extract above.

PASSWORD ENGLISH
DICTIONARY FOR SPEAKERS OF LATVIAN
Maija Trailona and Rasma Mosere.
Zvaigzne ABC Publishers Ltd.
Riga, Latvia 1996
ISBN 9984-04-385-1
822 pages.