

From the editors

For the third time in a short while, *Moderna Språk* appears here as a special issue. On this occasion, it will be on the *Lexicon*, the term used both in its 'concrete' (i.e. the subject of dictionary studies = lexicography) and 'abstract' (i.e. the subject of vocabulary studies = lexicology) meaning. Our original idea was to publish a special issue for both the areas of *Translation* and the *Lexicon* together. However, to our pleasant surprise, we found that we received so many written contributions that a division into two separate numbers became necessary.

Of the articles in this issue particular mention should be made in the English section of a presentation by Paul Procter, director of the Cambridge Language Services, of the *Cambridge International Dictionary of English*. The French section is introduced by an article by one of France's foremost linguists, Professor Robert Martin, on *Le Trésor de la Langue Française* (the most comprehensive dictionary of the French language now in existence). In the Spanish section, there is an article by Pedro Benítez Pérez, a lecturer at the University of Alcalá de Henares, which contains a presentation of lexical frequency investigations in Spanish and poses the question of how word selections should be made. The article section is concluded by a short, historical study of translation by the Vice-Chancellor of Uppsala University, Professor Stig Strömholm.

The idea of having special issues has been well received by the readers of *Moderna Språk*. We intend, therefore, to have more special issues in the coming years. Themes under discussion, although still not decided upon, include *Language Acquisition* and *Feminist Studies of Language and Literature*. The two (possibly three) forthcoming numbers of the journal will, however, contain the usual 'mixed' selection of articles. Nevertheless, the editorial board appeals to readers for suggestions for further themes of future special issues.

Since the last number, Ulf Dantanus, one of our editors, has left the editorial board. We wish to take this opportunity to thank him for the great effort he has put into working with *Moderna Språk* over the past six years. His place will be taken by Ronald Paul, lecturer in English at the University of Göteborg, whom we warmly welcome onto the editorial board.

PAUL PROCTER

The Making of a Modern Dictionary

Paul Procter är f.n. verkställande direktör för Cambridge Language Services Ltd. Han har lång erfarenhet av lexikonproduktion inom den engelska förlagsvärlden. Bl. a. var han chefredaktör för första upplagan av *Longman Dictionary of Contemporary English*. Under de senaste åren har han som chefredaktör lett arbetet med den nya *Cambridge International Dictionary of English*, som han här presenterar. En recension av det nya lexikonet återfinns på annan plats i detta nummer.

The *Cambridge International Dictionary of English* (CIDE), published earlier this year, is the first major monolingual dictionary to be published by Cambridge University Press. Intended for use by foreign learners of English, it is based on a new and different approach to dictionary production, resulting in a new and different kind of learner's dictionary. In this article, I shall outline first what is new about the dictionary, and secondly what is new about its production.

The first stage in the preparation of CIDE was a basic, but very important one – that of asking potential users what they most wanted to see in a learner's dictionary. CUP therefore carried out extensive research among teachers and learners of English around the world. This research showed that the key features that the users wanted were *wide coverage*, *user-friendliness* and *clarity*. In other words, dictionary users want to be able to find the word or phrase they are looking for, to find it easily, and to be able to understand it when they get there. It was following these principles, and using state of the art technology, that CIDE was created.

In terms of giving wide coverage, CIDE contains a larger number of entries – 100,000 words and phrases – than other comparable dictionaries currently available. This was achieved in part by the use of a carefully designed page, free from the clutter that is often found in dictionaries, which allowed more textual information to be given. These words and phrases include many that are new, a feature which CUP's research showed was extremely important to users. Because CIDE was produced on a computer system using innovative software, much of which was developed in-house, the team of lexicographers writing the dictionary were able to produce it in just four years. This enabled the team to keep pace with linguistic changes, thus ensuring that CIDE's coverage is absolutely current. The new words and phrases, and new meanings of words, that appear in a learner's dictionary for the first time in CIDE include: in the area of politics – *anti-choice*, *depoliticize*, *feelgood*; in media and the arts – *infomercial*, *prequel*, *ragga*; in fashion and food – *bumbag*, *isotonic*, *oenophile*; technical and medical words – *genotype*, *mini-system*, *SAD*; words relating to sex – *homoerotic*,

femidom; in sport – *mogul*, *paintball*; and slang words – *diss*, *oik*, *scumbag*.

English is an international language, and CIDE is an international dictionary, providing coverage of British, American and Australian English, including spelling, pronunciation, grammar patterns and usage of all varieties. Furthermore, the dictionary's cultural content is, as far as possible, international, reflecting the fact that English is often used today as the only common tongue between groups of speakers of other languages. Every entry in CIDE was checked by an American and an Australian editor to ensure that the content of the entry was valid from the point of view of a native speaker of American or Australian English. This means that all learners of English, whether they find themselves in the streets of London, New York or Sydney, can use CIDE to help them understand and use the appropriate language and to avoid the misunderstandings that can arise from confusables. By referring to CIDE, for example, the user will be able to discover that if he or she is asked "Would you like some squash?" in Britain or Australia, it is a soft drink that is being offered, while in the US it is a vegetable, or that if a businessperson in the US "tables" an item at a meeting, they are leaving it for later discussion, while in Britain and Australia, they are suggesting it for discussion now.

One innovative feature in CIDE's coverage is the inclusion of information on false friends. These are words that can cause confusion to learners of English because of their apparent similarity to a word in the learner's own language. The English word 'sensible' for example, is a false friend to speakers of French and Spanish because in those languages it has the meaning 'sensitive'. The word 'gift' in English means 'present', while in Swedish it means 'poison'. A group of language researchers collected instances of false friends across a wide range of the world's languages, and a selection of these is contained in CIDE for 16 of these languages: Czech, Danish, Dutch, French, German, Greek, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish and Thai. The false friends tables list the words in the relevant languages alongside their false friends in English and provide the English words that should be used instead. Warning symbols that a particular word is a false friend for a particular language group are given at the dictionary entry for that word.

A final aspect of CIDE's depth of coverage is that because all entries in the dictionary were coded according to the subject to which they are related, it was possible to extract all those entries which came within a particular subject area, and send them to subject specialists for checking. These specialists, in a wide range of fields from aerospace to zoology, advised on the accuracy of definitions and examples relating to their particular subjects, as well as on the suitability of words in their specialism for inclusion in the dictionary.

In order to make it as easy of access as possible, CIDE has two main new features. The first is the structuring of the entries. Most dictionaries or-

der the definitions of homographs numerically, often with numerous different meanings being given in the same entry. In CIDE, each entry constitutes a single meaning, and homographs are distinguished by the use of guide-words, printed in boxes next to the headword. The word 'mug', for example, has four separate entries in CIDE, with the guidewords 'container', 'stupid person', 'face', and 'attack'. This enables the user to go straight to the definition that they are looking for, rather than wade through a long list of possible meanings. An extension of this ordering by sense is that morphologically related forms of the headword – usually related parts of speech – are all grouped together as run-on entries under the headword. So under 'mug (attack)', the user will also find 'mugger' and 'mugging'.

The second innovative feature of CIDE that facilitates the user's access to the dictionary is the phrase index. This is an index at the back of the book that directs the user to the page and line number of each of the 17,000 compounds and idioms, including phrasal verbs, that appear in the dictionary. When confronted with a multi-word unit, such as 'burn the midnight oil', it is difficult for the user to know where to look it up in the dictionary. The CIDE phrase index lists the phrase under each of its major components – 'burn', 'midnight' and 'oil' – with a reference at each to where the phrase can be found. Furthermore, because it is not always easy for a user to tell whether a particular verb-particle combination is a phrasal verb, or is being used in its literal sense, the phrase index also includes all literal verb-particle combinations that mimic phrasal verbs, to help the user distinguish between them.

To maximize clarity, the definitions in CIDE have been written using a carefully researched controlled defining vocabulary of 2000 core words. This defining vocabulary was built from a list of words that are frequently occurring and useful for explaining other words, and that are at a level of difficulty which is accessible for the average user of CIDE. Words that are often confused with other words in English, or with foreign words, were excluded from the defining vocabulary. Words that have different meanings in British and American English were also generally excluded. Where a word has several different senses, those senses which are within the defining vocabulary are specified. Although this use of a controlled vocabulary is a modern development in lexicography, the principle underlying it follows Dr Johnson, who said in his Preface to the *Dictionary of the English Language* in 1755, "to explain requires the use of terms less abstruse than that which is to be explained". CIDE's defining vocabulary ensures that the definition of a word in CIDE does not contain language that is more difficult than the word itself.

The definitions in CIDE are supplemented by over 100,000 example sentences and phrases which illustrate in clear natural English how the particular headwords are used in context. These examples also allow additional nuances of meaning, that cannot easily be given in a definition, to be

shown.

As further illustration of the natural use of English, CIDE also includes 2000 well-known quotations and references from literature, film, songs and TV. These include the kinds of quotations that are frequently alluded to in newspaper headlines and advertisements, and that native speakers often use in their discourse. Among the quotations given in CIDE are "Come up and see me some time", "To be or not to be" and "I've started so I'll finish".

In the case of some words, a greater clarity of definition can be achieved by the use of pictures rather than words. Over 2000 vocabulary items are shown in CIDE in clear, distinctive black and white drawings and silhouettes. The illustrations include pictures of lexical groups, such as types of bags, musical instrument and herbs and spices, and pictures of real-life settings, such as rooms in a house, as well as illustrations of words with multiple meanings, such as one of 'bar' which shows a coffee bar, a bar of chocolate, a bar of soap, a bar of music, a bar on a window and a bar of an electric fire. These items were selected for illustration not only on the basis of how usefully their meaning could be shown pictorially, but also in terms of how visually interesting the pictures were.

Another aspect of CIDE where considerable attention was paid to clarity is that of grammar. A clear and easy to understand coding system is used to illustrate the syntactic behaviour of headwords. Specific grammar patterns are shown in example sentences, each with a code attached, so that it is possible for the user easily to identify which grammar pattern is being shown in which example.

Additional information about grammar is also provided in some of the 120 language portraits that appear in the dictionary. These have been written by ELT authors, teachers and editors, and in addition to grammar cover a wide range of topics, such as punctuation, phonetics, style and vocabulary. They are alphabetically arranged through the body of the dictionary, and have been designed for use in the classroom or for private study.

I should like to turn now to the way in which the dictionary was produced. CIDE is the first dictionary to have been written using an Integrated Language Database. This consists of a database system on which the dictionary information is recorded, intertwined with a corpus of 100,000,000 words of written and spoken English. This corpus was used in various ways. First, it was used by the lexicographers as a source of words and phrases for inclusion in the dictionary, although the lexicographers' judgement was also used in deciding whether or not a particular word should be included. For example, the corpus produced numerous occurrences of the word, 'loadsamoney', deriving from a British television programme in the late 1980's, but the team judged it, quite rightly as is now evident, as being too ephemeral for inclusion. Entries within the corpus are tagged according to their source, so the lexicographers were able to identify whether a word is changing in the frequency with which it is used, whether its meaning is

changing in any way, or whether it is used mainly in newspapers, fiction or spoken English.

Another use of the corpus that was made by the CIDE team was as a source of example material. In writing the examples, the lexicographers used material from the corpus, which they were able to cut out from the corpus and paste into the dictionary electronically. This was, however, usually edited first in order to remove any unnecessarily complicated vocabulary or syntax, thereby making it more learner-friendly. The material was also edited to remove any bias, in the areas of, for example, gender, race or religion.

The use of corpora for the production of dictionaries is not, in itself, of course, new. What is innovative about the Cambridge system is that the intertwining of the dictionary database with the corpus enables the lexicographers to carry out a much more detailed analysis of the way in which a particular word is used. For instance, once the lexicographer has entered onto the database particular linguistic information about a word, corpus searches on that word can be carried out using that linguistic information. One example of the way in which this works is the use of information about inflections. When a lexicographer carries out a corpus search for a particular word, the software first accesses the dictionary database to identify what the inflected forms of that word are, and then produces all the occurrences of all the inflections within the corpus. The lexicographers are further helped in this by having each word in the corpus tagged with its part of speech. It is therefore possible for the lexicographer to search for all instances of the noun 'bear', and be provided with all occurrences of 'bear' or 'bears', or of the verb 'bear' and be given 'bear', 'bears', 'bearing', 'bore'. The Cambridge system also uses information given in the dictionary database to sense tag the corpus, making it possible to search, for example, for all instances of the verb 'fire' meaning to dismiss someone from their job, or 'fire' meaning to shoot a gun.

The system also contains a highly advanced piece of collocation analysis software, which was developed by the CIDE programmers. This software searches the corpus for those words which occur within a specified range of the particular word being considered, and calculates the statistical significance of those words occurring together, thereby identifying those words which co-occur most commonly. This information can then be assessed by the lexicographer and included in the dictionary where appropriate. For example, a collocation search on the word 'view' reveals that the words which most significantly co-occur with it are 'point', 'of', 'panoramic', 'dim', 'from', 'shared', 'world', 'my', 'jaundiced', 'sanguine' and 'takes'. The lexicographer can use this information to show in the dictionary that 'point of view' and 'take a dim view' are frequently occurring units of language, that views are often taken or shared, and that what you see from the top of a mountain is a panoramic view, not a wide one or an

extensive one. Collocations are given in example sentences in CIDE, marked up in bold print, making them immediately apparent to the user, and thereby helping the user to develop that awareness of collocation that is essential for achieving fluency.

The main aims of the CIDE team were to produce a dictionary that was contemporary, comprehensive, clear and easily accessible. The new features that the dictionary contains are designed to realize those aims. The new production system created specifically for it was designed to achieve them.

M

CURRENT RESEARCH

Two new doctoral dissertations in English literature from Göteborg University:

Janina Nordius, *"I Am Myself Alone": Solitude and Transcendence in John Cowper Powys.*

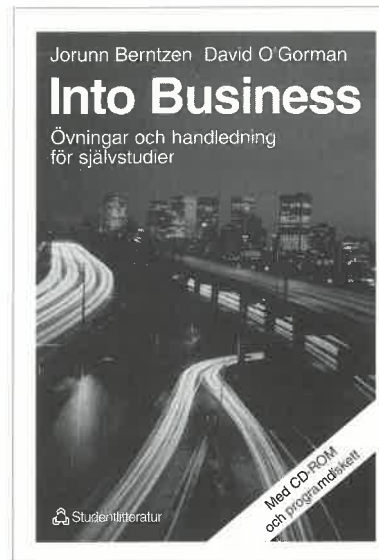
The aim of this study is to examine John Cowper Powys's fictional inquiry into *solitude* (a comprehensive term used here for *all* aspects of being, or feeling, alone) and to show how this inquiry constitutes an integral structure in all his major novels. The analysis is based on close readings of the novels *Wolf Solent* (1929), *A Glastonbury Romance* (1932), *Weymouth Sands* (1934), *Maiden Castle* (1936; 1990), *Owen Glendower* (1940), and *Porius* (1951; 1994), each of which is discussed in a separate chapter. Powys's philosophy of solitude—whose main tenet is that it is both possible and desirable to make solitude a source of personal happiness—is continually being tested against the fictional reality of his novels, a process which often entails severe criticism as well as further development of the philosophy.

Address: Janina Nordius, Engelska institutionen, Göteborgs universitet, S-412 98 Göteborg, Sweden.

Claudia Egerer, *Fictions of (In)Betweenness*

In this study, the author seeks to show how both fictional and theoretical texts 'worry the lines' between conceptions of home and exile. The author examines the ways in which these concepts are re-thought and engaged with in the novels by Louise Erdrich, J.M. Coetzee and David Malouf and in the writings of Edward Said, Jacques Derrida and Homi Bhabha. The texts discussed position themselves as texts of (in)betweenness in that they are all engaged in 'thinking the between.' That is, they are dissatisfied with 'home' as the privileged term and attempt to re-conceptualize the border between home and exile in terms of the unhomeliness of home and the homefulness of exile.

Address: Claudia Egerer, Engelska institutionen, Göteborgs universitet, S-412 98 Göteborg, Sweden.

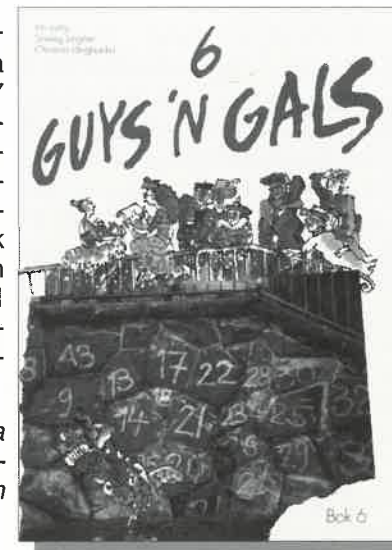


Affärsengelska – nu på CD-ROM

Vår nya CD-ROM bygger på kursmaterialet **Into Business** med textbok, övningsbok, lärarhandledning och kassetter. Med CDn kan man öva uttal och hörförståelse på ett helt nytt sätt. Dels arbetar man med diktamensövningar där man lyssnar till texten och skriver in det man hör med möjlighet att kontrollera rätt svar i facit. Dels tränar man uttalsövningar genom att spela in sin egen röst och jämföra uttalet med CDn. Materialet finns för självstudier och för skolor.

Guys 'n Gals heter vårt flexibla läromedel i engelska på gymnasienivå. Det består av 7 böcker som var och en innehåller text, övningar och grammatik, vilket gör materialet väl lämpat för kursindelade, individualiserade studier. Till varje bok finns ett kassettpaket och en omfattande lärarhandledning med kopieringsunderlag och stordia-underlag. Böckerna täcker engelska A och B.

Du är välkommen att beställa tjänsteexemplar om du funderar på att använda en bok som kurslitteratur.



Se hela vår utgivning på Internet <http://www.studli.se/>

 **Studentlitteratur**
Box 141, 221 00 Lund Tel 046-312000 Fax 046-304962