

Alternative e-dictionaries: uncovering dark practices

Nesi, H.

Proof copy deposited in CURVE March 2015

Original citation & hyperlink:

Nesi, H. (2012) 'Alternative e-dictionaries: uncovering dark practices' in Sylviane Granger and Magali Paquot (Ed). Electronic Lexicography. Oxford, UK: Oxford University Press.

<http://ukcatalogue.oup.com/product/9780199654864.do>

Publisher statement: This material was originally published in Electronic Lexicography edited by S. Granger and M. Paquot, and has been reproduced by permission of Oxford University Press <http://ukcatalogue.oup.com/product/9780199654864.do>. For permission to reuse this material, please visit <http://www.oup.co.uk/academic/rights/permissions>.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the proof copy, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

CURVE is the Institutional Repository for Coventry University

<http://curve.coventry.ac.uk/open>

17

Alternative e-dictionaries: Uncovering dark practices

HILARY NESI

17.1 Introduction

Most published e-dictionary research involves studies of the way e-dictionaries are used, surveys of the habits and attitudes of e-dictionary users, or accounts of the development of new experimental e-dictionaries. The e-dictionaries that are described in detail in this research tend to be the prestigious varieties emanating from university centres or established publishing houses, and although surveys suggest that less prestigious types of e-dictionaries are far more popular with the general user, they are also far less likely to be specified individually in the academic literature, or evaluated in terms of their lexicographical content. We will call these less prestigious e-dictionaries ‘alternative e-dictionaries’, or ‘AEDs’. This chapter aims to redress the balance somewhat by considering AED resources, and in so doing will draw attention to the dangers of over-reliance on AED information.

17.2 Background

Perhaps understandably, a great deal of scholarly attention is paid to e-dictionaries that have been developed for research and educational purposes rather than for commercial gain. Such dictionaries tend to be small-scale, and are carefully designed to enable their developers to explore metalexicographical issues and concentrate on the needs of specific user groups, as advocated by Gelpi (2004) and Tarp (2009a). The *Louvain EAP Dictionary* (Granger and Paquot 2010b, and this volume), developed at the Université catholique de Louvain, for example, offers access routes not only via the lexeme or the user’s first language but also via communicative function, and the *Base lexicale du français*, developed at the Katholieke Universiteit Leuven, has a task- and problem-oriented interface which allows users to begin their search by choosing

from six “user-driven situations” (Verlinde 2010, Verlinde and Binon 2010, Verlinde and Peeters, this volume). Examples of other experimental e-dictionaries include *Lexin*, developed at Göteborg University to meet the needs of the immigrant community in Sweden (Hult 2008), OWID, an online portal for German developed at the Institute of German Language in Mannheim (Müller-Spitzer and Möhrs 2008), and the *Deutsch-Italienisches Fachwörterbuch der Linguistik*, a specialized linguistics dictionary developed at the University of Pisa (Flinz 2010).

In the case of experimental e-dictionaries such as these, the metalexigraphers are also part of the dictionary development team, and thus have the means to record in the form of log files the transactions between the dictionary database and the user’s computer. Nowadays e-dictionary developers can gather log-file data on every aspect of the consultation process, including search routes and time taken, but such details are potentially sensitive and so most published log-file studies, such as those of De Schryver and Joffe (2004), Bergenholtz and Johnsen (2007), Hult (2008), Prószyński and Kis (2002), and Verlinde (2010), continue to relate to the use of experimental or specialist e-dictionaries rather than commercial products.

The commercial e-dictionaries that receive the most metalexigraphic attention are those produced by prestigious publishing houses (cf. Dziemiński, this volume). Monolingual learners’ dictionaries on CD-ROM have been a major object of study by researchers not directly involved in e-dictionary development, such as Nesi (1996, 1999), Rizo-Rodríguez (2004), and Xu (2008). These dictionaries share at least some of their lexicographical content with familiar print editions, and this may make them seem more permanent and stable (Gelpi 2004: 9). They also usually include some explanatory documentation equivalent to the frontmatter in a printed dictionary. This is important for the ‘lexicographical approach’ to dictionary reviewing described by Nielsen (2009), which entails consideration of the interrelationship between dictionary components. Nielsen discusses the lexicographical approach to reviewing with reference to two online dictionaries, the *Oxford English Dictionary Online* (2008) and the *Merriam-Webster Online Dictionary* (2008); both are prestigious varieties with prefaces and user guides.

Alternative e-dictionaries, on the other hand, are often difficult to describe in conventional lexicographic terms. A distinguishing feature of such products is that they combine diverse resources, some of which are up-to-date dictionaries from prestigious publishing houses, some of which are local dictionaries, often of old or uncertain provenance, and some of which are not produced by lexicographers at all. AED content is available via dictionary web portals and pocket electronic dictionaries (PEDs), or it can be downloaded to a personal computer or mobile device from the Internet, a USB stick, a CD-ROM, or a memory card. It is combined and re-combined for different markets, and on- and offline products with the same titles, produced by the same company, are not necessarily identical. Moreover, AED content is bewilderingly subject to change. Online dictionaries are always “under

construction”, as Gelpi (2004: 7) points out, and electronics firms such as Casio, Franklin, Seiko, and Sharp continually produce new PED models, each with different capabilities and features. AED companies have no interest in keeping historical records, or in referring users to earlier versions of their resources, and this makes it particularly hard for metalexigraphers to keep track of the content that was available in a particular product at a particular time.

Keeping up-to-date with AED resources can also be expensive, and as commercial developers are not in the habit of offering review copies or discounts on class sets, AED studies are likely to be hampered by the cost. Shizuka (2003) and Koyama and Takeuchi (2004b, 2007) acquired pocket electronic dictionaries from the Casio computer company for use in experiments in Japan, but this kind of support from manufacturers seems to be unusual, and Chen (2010), for example, complains of lack of funding to obtain PEDs for research.

Thus, despite the prevalence of AED use, AED material receives very little scholarly attention, at least in the West. For example, the Jin Shan Ci Ba rarely features in reviews, user studies, or dictionary skills training programmes, yet it is possibly the most widely used dictionary package in the world, well known to all mainland Chinese learners of other languages. Its developers claim that it is used in more than 50,000 educational institutes, businesses, and government agencies, by more than twenty million people.

Bypassing issues of AED content, researchers have tended to concentrate instead on the advantages and disadvantages of the consultation medium. Numerous surveys have compared electronic and print dictionary use (cf. Dziemianko, this volume), particularly by Japanese learners (Weschler and Pitts 2000, Perry 2003, Bower and McMillan 2007, Kobayashi 2008) and Chinese learners (Taylor and Chan 1994, Tang 1997, Deng 2005, Chen 2010), without actually drawing attention to differences in dictionary content that exist between one AED package and another. Recently some attempts have been made to study the way learners use free bilingual dictionary portals. Tseng (2009) tested English language learners’ interpretations of definitions from the Taiwanese Yahoo! portal, and Chon (2009) gathered think-aloud data on the productive use of the Korean Yahoo! and Naver portals. In both studies, however, the primary aim was to examine students’ dictionary skills, and so consultation problems tended to be associated with learners’ lack of dictionary strategies (“dictionary-based errors”, in Chon’s terms) or the unsuccessful use of these strategies (Chon’s “dictionary-based problems”), rather than with the design and content of the portals themselves.

Boonmoh and Nesi (2008) conducted a questionnaire survey of attitudes to pocket electronic dictionaries at a university in Thailand, and concluded that English language lecturers often blamed PEDs for consultation problems that could just as easily arise with print dictionaries. The following are typical lecturer responses to

questions about their students’ PED use; some indicate bad practice or lack of skill on the part of the students rather than defects in PED content.

- *When using a pocket electronic dictionary, they just type a Thai word and pick up one English word from the list provided. They don’t know how to compose a sentence by using the word.*
- *Students will look up words and use words without looking at the context.*
- *They tend to select the wrong word because of their lack of knowledge in part of speech.*
- *It is too convenient, so it is not challenging.*

Nevertheless, the lexicographic quality of AEDs is also a common cause of teachers’ complaints. Tang (1997), Koren (1997), Deng (2005), Stirling (2003), and Boonmoh and Nesi (2008) all found that teachers considered electronic dictionaries to lack important elements such as grammar codes, example sentences, and collocations, although these teachers also revealed themselves to be quite ignorant of e-dictionary features (Midlane 2005, Boonmoh and Nesi 2008), and did not usually differentiate between the prestigious and less-prestigious lexicographical content of AEDs, assuming that all AED dictionaries shared the same faults.

Table 17.1 shows the distribution of dictionary ownership amongst the thirty English language lecturers who responded to Boonmoh and Nesi’s questionnaire survey (2008). The older lecturers only used print dictionaries, although some of the younger lecturers sometimes referred to online dictionaries and prestigious monolingual dictionaries on CD-ROM (mostly those published by Longman, COBUILD, and Macmillan).

TABLE 17.1. Thai English language lecturers’ ownership of dictionaries

Monolingual dictionaries in book form	29
Monolingual dictionaries on CD-ROM	22
Bilingual dictionaries online	11
Bilingual dictionaries in book form	10
Monolingual dictionaries online	9
Pocket electronic dictionaries	4

Most lecturers rejected the idea of PEDs, and expressed little faith in their lexicographical content. Typically, PEDs were characterized as inaccurate and lacking in usage examples:

- *It doesn’t give the correct meaning and doesn’t tell students clearly how to use the word in different contexts.*
- *It always gives the wrong usage.*
- *limited vocabulary and doesn’t provide sentence examples, too expensive.*

- *it is not as detailed as a dictionary in book form.*
- *A PED does not provide usage, examples. And as a consequence, students are likely to make mistakes.*

The only lecturer who preferred the PED format used it in combination with other sources:

- *I will use it only when I cannot think of English vocabulary. However, I will have to check how that word is used from a monolingual either in book form or on CD-ROM.*

All these respondents were apparently ignorant of the fact that the most popular PED model at their university contained the *Oxford Advanced Learner's Dictionary*. The majority (18 out of 30) admitted that they had no idea what dictionaries PEDs were likely to contain.

Those who are not teachers seem on the whole to be less worried about the accuracy and detail of dictionary entries. As Kilgarriff (2005b) pointed out, “a dictionary is a dictionary is a dictionary, good or bad”. Studies of students’ attitudes towards e-dictionary use suggest that they like AED bilingual dictionaries, and tend to value them for their advanced technology and accessibility rather than for their lexicographic content. The 1,211 Thai student respondents in Boonmoh and Nesi’s (2008) survey, for example, had been recommended by their teachers to buy a *Longman Active Study Dictionary (LASD)* package, complete with CD-ROM, but although 82 per cent claimed to own a monolingual print dictionary (probably *LASD*), only 28 per cent claimed to own a monolingual dictionary on CD-ROM, suggesting that many had not even bothered to open the CD-ROM envelope inside their *LASD* copies. AEDs were used far more frequently. Nesi (2010: 214–16) found that students who were given copies of a Macmillan learners’ dictionary on CD-ROM rated its quality more highly than the AEDs they normally consulted, acknowledging that it had “better explanations” and “more information”. However, most of these students quickly reverted to AED use (in the form of PEDs and/or software on their home computers), because they liked the many additional features AEDs offered, for example ready-made wordlists, language tests, games, and personal organizer functions. They also found it easier to understand bilingual dictionary information, whether on CD-ROM, the computer hard drive, or in a more portable format; PEDs were particularly convenient for use in class or when travelling. In Nielsen’s (2009) terms, AEDs had lower “lexicographic information costs” and required less effort to consult. The distinction between “usability” and “usefulness” (Laufer and Kimmel 1997: 362) is relevant here. Laufer and Kimmel define “dictionary usability” as “the willingness on the part of the consumer to use the dictionary in question, and his/her satisfaction from it”, while “dictionary usefulness” is “the extent to which a dictionary is helpful in providing the necessary information to its user”. AEDs have high usability ratings, despite the doubts expressed regarding their usefulness.

Some surveys report teachers' suspicions that e-dictionary consultation is too fast and too easy for deep learning to take place (Taylor and Chan 1994, Sharpe 1995, Zhang 2004, Stirling 2003, Boonmoh and Nesi 2008). In experimental studies where subjects have consulted identical lexicographical content on a computer screen or in print, however, task performance has either not differed significantly (Nesi 2000b, Koyama and Takeuchi 2004b, 2007) or has been significantly better for e-dictionary users (Shizuka 2003, Dziemianko 2010) (cf. Dziemianko, this volume). Thus, although the restricted size of PED screens doubtless impedes consultation of longer entries, and there is certainly scope for further investigation into the pedagogical effects of e-learning, the evidence so far suggests that teachers' reported dislike of e-dictionaries is largely due to the low quality of some AED resources, and their ignorance of better-quality AED resources, rather than the electronic medium itself.

Gelpi (2004) offers some guidelines for the evaluation of online bilingual dictionaries, presenting a list of quality indicators, including the need for a "real and public author", and explicit indications of when and how often the source is updated. AEDs often fail to meet both of these criteria, and in the next part of this chapter I will demonstrate how, with some notable exceptions, the world's most popular e-dictionaries are unreliable sources of linguistic information.

17.3 Alternative electronic dictionary content

In order to critique AEDs, this section will examine their various components, their lexicographical content, and the potential of new collaborative procedures to influence AED content development.

17.3.1 AED components

Alongside local dictionaries and non-lexicographical resources, AEDs typically contain some monolingual, bilingual, and/or bilingualized dictionaries licensed from highly regarded international publishing houses. For example, the PEDs used in Nesi and Boonmoh's (2009) study contained the monolingual *Concise American Heritage Dictionary*, the *Oxford Advanced Learner's Dictionary*, and the *Longman Dictionary of American English*, and Chen (2010) reports that her students' PEDs were installed with bilingualized learner's dictionaries from Longman, Oxford, and COBUILD. Lingoos, a popular Chinese dictionary portal, offers access to the *Merriam Webster's Collegiate* and the *American Heritage Dictionary*, together with all the major monolingual English advanced learners' dictionaries: the *Cambridge Advanced Learner's Dictionary (CALD)*, the *Collins COBUILD English Dictionary (COBUILD)*, the *Longman Dictionary of Contemporary English (LDOCE)*, the *Macmillan English Dictionary (MED)*, and the *Oxford Advanced Learner's Dictionary (OALD)*. The Babylon website offers dictionaries published by Oxford, Duden, Larousse, and

Merriam Webster, and the Jin Shan Ci Ba CD-ROM and Internet download contains the *American Heritage Chinese-English Dictionary* and a bilingualized version of the *Oxford Advanced Learner's Dictionary*, as well as many other dictionaries and thesauruses of more varying quality.

Licences from famous brands are useful for AED developers because they add prestige to the dictionary package as a whole. Some AEDs such as Jin Shan Ci Ba offer free web access to unlicensed dictionary material, but require users to purchase the package containing licensed dictionaries such as the *Oxford Advanced Learner's Dictionary*. This practice is comparable to that of many publishers who offer a simple web interface in order to boost sales of their dictionaries in print or on CD-ROM. On the AED sites, however, specific information about their prestigious dictionaries is sometimes left deliberately vague, perhaps through ignorance on the part of the site developers, perhaps to allow for last-minute alterations to AED content, or perhaps to imply more extensive prestige content than is actually the case. On some portal sites various publications from Oxford University Press are referred to by the blanket term "Oxford Dictionary" or "Oxford English Dictionary", for example. This is the case with the popular e-dictionary portal Babylon, as is evident in the extract in Figure 17.1, from an online synchronous conversation between myself and 'Michelle', a Babylon representative.

You are now chatting with "Michelle"

Michelle: Hey

Hilary: Can you give me full details of Oxford Dictionary mentioned as part of the Lifetime License package? Is it available by monthly subscription?

Michelle: nope

Hilary: So you don't know anything about the Oxford Dictionary?

Michelle: the monthly version is a light version paid for every month

Michelle: Yes i do

Michelle: the full version of Babylon is for lifetime use

Michelle: with the Oxford

Hilary: But what about the Oxford Dictionary? What is its full title?

Michelle: Oxford dictionary is the most professional English dictionary we work with. It adds an extensive content in English, technical terms, example of a usage of the work you are searching for in a sentence, and a professional thesaurus.

Hilary: Is it published by Oxford University Press?

Michelle: yes, it's the same famous old Oxford dictionary book Published by Oxford, all its content goes into your Babylon.

FIGURE 17.1 A conversation with a Babylon representative.

I had expressed interest in subscribing to Babylon’s ‘Lifetime License’ package, an upgrade from a cheaper monthly subscription package.

When pressed, Michelle did provide two Oxford titles, but still failed to provide explicit information about the updating process, one of Gelpi’s quality indicators (2004) (see Figure 17.2).

Hilary: There are many (possibly hundreds) of dictionaries published by Oxford University Press. Which one is it?

Michelle: The Concise Oxford Dictionary

Michelle: The Concise Oxford Thesaurus

Hilary: That’s helpful, thanks. Publication date? Or edition?

Michelle: it’s updated all the time.

FIGURE 17.2 Further conversation with a Babylon representative.

It is perhaps not surprising that information about Oxford products is so vague, given that, according to their website, syndicated material from the Oxford Reference Online can be selected and tailored according to a company’s wishes, with all kinds of modifications to its “functionality, content and design”.

However, although arrangements with prestige publishing houses such as Oxford must help to sell AED products, in fact the dictionaries from these publishers may not be the ones that are usually consulted. The default dictionary offered in dictionary portals and PEDs is often a local one, and unless the settings are changed, or the search word is not in the local dictionary database, the entries for this local dictionary are the ones that automatically appear. Thus although the various versions of the Jin Shan Ci Ba (‘Kingsoft Powerword’) contain international resources, the default setting seems to be the Jian Ming Ying Han Ci Dian, a somewhat unreliable, undated, general Chinese-English dictionary. The Wikipedia entry for the dictionary package advises users to bypass this default and “pick good dictionaries” instead, although it also acknowledges that it “is not clear how to do this”. Nesi and Boonmoh (2009) found that the prestigious monolingual dictionaries in Thai students’ PEDs were never consulted, as local English-Thai and Thai-English dictionaries were offered as the default choice.

17.3.2 AED quality

E-dictionaries are particularly popular in East Asia, and for some East Asian languages there are no bilingualized dictionaries, and the only available bilingual dictionaries are locally produced without the benefit of corpus resources. Some of these dictionaries are simply basic translation tools, with little or no information

about grammar and usage. Chon (2009), for example, notes the lack of grammar information on the Korean Naver website. Some of the dictionaries also contain serious mistranslations. For example, Nesi and Boonmoh (2009) found PEDs listing *bulb* as a verb, supposedly equivalent to the archaic Thai form /ngok hua/ ('to appear') and Boonmoh *et al.* (2005) found PEDs translating the Thai noun /bot/ ('chapter' or 'unit') as "a foot; a stanza of verse; the words of a song or a play" (see also Hanks, this volume, on bad definitions). Moreover, some local dictionaries are digitalized versions of older print dictionaries, and errors seem to have crept in during the digitization process. The misprints in Thai PEDs reported by Boonmoh *et al.* (2005), for example "hehind" for *behind*, and "lear" for *learning*, would appear to be scanning errors.

The poor quality of many local dictionaries is exacerbated by the AED tendency to aggregate large numbers of dictionaries compiled at different times and for different purposes, offering a 'jump' facility between sources. The current Jin Shan Ci Ba download offers "more than 140 practical dictionaries", and "98 Academic and Professional Directions" [sic] which appear to be thesauruses of various kinds, for example for "Soil Science", "Ship Engineering", and "Railway Science". Thus when the dictionary package is used productively, typing in just a few letters can lead, via the jump facility, to a display of archaic, technical, or misprinted word forms each of which may originate in just one source.

This problem is compounded by the tendency for local dictionary compilers to over-extend lexical derivation rules, adding affixes wherever they might potentially occur. For example, Boonmoh *et al.* (2005) found the following words beginning with *suppo-* in the PED CyberDict 3 Advance: *support*, *supportable*, *supporter*, *supportive*, and *supposable*. *Supportable* has very low frequency in the British National Corpus (BNC), and there are no recorded instances of *supposable* in the BNC. Similarly, Nesi (2010) found the sample of headwords in Figure 17.3 offered in a Jin Shan Ci Ba download.

Words marked * in Figure 17.3 do not appear in the 56-million-word Collins COBUILD Wordbank, and *board* is the only collocate of *examination* to be counted amongst the 100 most statistically significant.

Some AEDs also include figurative expressions that are no longer current, perhaps originating in out-of-date sources or added to enlarge the advertised lemma stock. The Jin Shan Ci Ba translation *nigger brown* has now been removed, but the Korean Naver site and the Japanese and Korean versions of Yahoo! online all include the biblical expression *not worth a jew's eye*, apparently originally listed in *Brewer's Dictionary of Phrase and Fable* (undated). Similar entries with the potential to offend can be found in many AEDs, often without any kind of taboo warning. Other rather quaint but less offensive expressions listed in AEDs include *not worth a leek*, *thick as mutton*, and *thick as herrings*. It could be argued that such entries might one day be useful to those who use AEDs for receptive purposes, but AEDs are also very widely

employed as writing aids, and the *nigger brown* translation caused an outcry when it was used in 2007 by a Chinese furniture company exporting goods to the West. This incident features in the Wikipedia entry for PowerWord¹ (see also Baemans 2007 and many other online blogs citing the CNN news report ‘Chinese translation error blamed for slur on sofa label’).

exam
examen*
EXAMETNET*
examinable*
examinant*
examine*
examination
examination board
examination class
examination finding
examination in chief
examination of scheme
examination of auditor
examination of budget
examinationism*
examinationist*
examination procedure
examinator*
examinatorial*

FIGURE 17.3 A sample of headwords from Jin Shan Ci Ba.

AED headword lists are vastly increased by aggregating large numbers of dictionaries, and by including rare multi-word expressions and potentially acceptable but actually unattested derived forms. This may benefit an AED by raising its status, especially in the eyes of more unsophisticated dictionary users who may be impressed by the sheer size of the package, and by the fact that there are translations for almost all the words they will ever encounter receptively. All commercial publishers emphasize the extent of their dictionary’s coverage in order to attract customers, and Nielsen (2009: 27) points out that reviewers tend to evaluate dictionaries in terms of the size of their lemma stock, placing too much emphasis on linguistic categories rather than significant lexicographical features. The lack of restrictive labels and usage information in the entries provided by local and specialist AED dictionaries is a problem for writers and for language learners, however, especially as many AEDs seem to be intended as learning tools, given their facilities for word-list creation and vocabulary testing.

The unreliability of AEDs stems partly from their reliance on poor-quality dictionaries, but also partly from their use of unedited non-lexicographical data. Many

¹ <http://en.wikipedia.org/wiki/PowerWord> (last accessed 29 April 2011).

AEDs supplement their dictionary resources with examples taken from online encyclopedias and media websites, as is the case with Lingoes and the Korean Doosan Dong-A Prime on the Daum portal. Some, like the online Jin Shan Ci Ba, offer web examples as a free resource while reserving their more prestigious dictionaries for paying customers to download. Some AEDs also make use of automatic translators. Lingoes, for example, draws on the Jukuu search engine for sentence translation in Chinese, English, and Japanese, and the Jin Shan Ci Ba works with the machine translation device Jinshan Kuaiyi.

The results of these approaches are often disastrous. Mair (2007) blames a combination of the Jin Shan Ce Ba and the Jinshan Kuaiyi for the production of “absurdly crude English mistranslations in bizarrely inappropriate contexts” in China, and the same seems to be true for the *Doosan Dong-a Prime* English dictionary, online on the Daum South Korean web portal. On this site lexicographical content similar to that provided in the *Doosan Dong-a Prime* print dictionaries is supplemented with bar charts showing quantitative collocation information based on web data, and illustrative examples taken from online sources. The amount of additional illustrative material on the website may give Korean English language learners the impression that the compilers are concerned with contextual appropriacy, but in fact the illustrations are unedited and highly misleading. Examples (1) to (3) provide contexts for ‘dark coffee’ (available at the time of writing by typing ‘dark coffee’ in the search box on the homepage at <http://engdic.daum.net/>). These sentences appear to be automatically generated translations of Korean postings to USENET, the Internet discussion system.

- (1) What did Francis arrive the cup before the dark coffee?
- (2) It should change the dark coffee and arrive it through its monolith.
- (3) She’d rather kick furiously than call with George’s dark coffee.

17.3.3 Collaborative approaches to AED content development

Collaborative or “bottom-up” lexicographical methods (Carr 1997) have the potential to correct mistranslations and absurd examples, and generally improve online AED resources, although quality control remains an issue (Docherty 2000, De Schryver 2003). The bottom-up approach has been facilitated by the invention of Wiki software, which became available in the early 2000s as an open source tool (Nesi 2009). The Wiktionary, which describes itself as a “wiki-based Open Content dictionary”, appeared in 2002, and the Wiktionary system now covers more than 400 languages (cf. Meyer and Gurevych, this volume). Although many apparent neologisms and local slang expressions are recorded, Wiktionary contributors rarely seem to have taken the trouble to provide source details or usage information, and some entries over-extend the English lexical derivation rules in order to include derived

forms which are only potentially acceptable, such as *examinable*, *supportable*, and *supposable*, all listed without restrictive labels.

Many other AEDs are now starting to invite user contributions through wikis, or via postings on blogs or fora. For example, the German *bab.la*, a dictionary and language learning portal which claims about 30,568 unique visitors per day, operates on a wiki-style model, allowing users to contribute content and give feedback. Similarly, CC-CEDICT invites contributions to a wiki and encourages users to submit new entries to its ‘editor’ website. These contributions are then used to inform the development of the MDBG, a Chinese-English online dictionary which claims 400,000 unique visitors per month.

The Leo Dictionary, on the other hand, hosts fora in order to discuss difficult translations and questions concerning the dictionary. The collaborative nature of Leo is critiqued in Wikipedia (3 March 2011), where it is argued that the absence of an editorial hand results in “duplication and a lack of additional lexical information such as gender, inflected and irregular forms, plurals, and other contextual indicators”. Nevertheless the Leo fora are very active and quick to identify problems with the dictionary content, as demonstrated in the exchange in Figure 17.4, where participants JTB and CM2DD discuss (bilingually) the use of the word *unthink*, translated as *umdenken* in Leo’s English-German dictionary. *Unthink* is a typical potentially acceptable derivation for which there are translations in a number of AEDs, for example ET House, Doosan Dong-a Prime, Wikipedia, and the Yahoo! sites for Japan, Korea, and Hong Kong.

MDBG and Leo are not typical AEDs because they originated in research projects and still seem to be motivated at least in part by academic rather than commercial interests. MDBG is powered by CC-CEDICT, an offshoot of Jim Breen’s EDICT Japanese dictionary project now managed by the Electronic Dictionary Research and Development Group at Monash University. It is funded by donations as well as advertising revenue. Leo started life as a research project at the Technische Universität München and is now operated commercially by a company formed by members of the original research team. Some of the other, more commercially-minded AEDs seem to manage collaborative input less successfully, perhaps because they do not want to encourage criticism of the content on offer, or because their users are less disposed to critique their perceived authority.

I searched for *unthink* in the free online version of the Jin Shan Ci Ba, which has no official dictionary content but refers users to its own blog and forum contributions, and to various local web sources such as the China Daily. In the online version of the Jin Shan Ci Ba the entry for *unthink* offers a translation into everyday Chinese, with no accompanying warning about the rarity of the English word (the Chinese translation roughly means ‘don’t want to’). However, the entry also contains a footnote signalling that the translation does not come from an authoritative source, and it offers users the opportunity to edit and amend the entry information.

Betritt	Beispiel für "to unthink" - "umdenken"	
Kommentar	<p>Hallo!</p> <p>Bei der Suche bin ich auf "to unthink" gestoßen, einmal übersetzt mit "aus dem Gedächtnis verbannen" und einmal mit "umdenken". Ich habe das noch nie gehört und mein Sprachgefühl meutert (was an sich nichts zu bedeuten hat) und wollte fragen, ob das wirklich das "wir müssen in dieser Sache umdenken" meint oder ein anderes "umdenken" und ob mir vielleicht jemand ein Kontextbeispiel geben könnte.</p> <p>Vielen Dank & viele Grüße,</p> <p>JTB</p>	
Autor:	JTB	02 Mar 11 09:41
Kommentar	<p>I don't think "unthink" and "umdenken" match at all. You need something more like "rethink" or "change our way of thinking", depending on the context.</p> <p><i>unthink</i> <i>Forget, remove from thought; cancel out or reverse by a mental effort.</i> New SOED</p>	
1	Autor	C M 2 D D (236324) 02 Mar 11 10:03
Kommentar	<p>Thank you!</p> <p>So my feeling that it sounded weird didn't betray me. Should I open a "Falscher Eintrag?" thread and may I cite your source for that?</p>	
2	Autor	jtb 02 Mar 11 10:08
Kommentar	<p>The Macquarie Dictionary does give a definition "to reverse or retract by thinking, as in changing one's mind" and I've found examples where it seems to be used kind of that way.</p> <p>"The ambition was to "unthink" and revive the intellectual tradition of holistic theorizing about large historical processes that was started by Marx and Weber" (Encyclopedia of Nationalism: Fundamental Themes.)</p> <p>Or here:</p> <p><i>in addition to rethinking, which is "normal," I believe we need to "unthink " nineteenth-century social science, because many of its presumptions - which, in my view, are misleading and constrictive - still have far too strong a hold on our mentalities</i> http://books.google.de/books?id=H8wnle1KwMUC&pg=PA1&dq=%22to+unthink%22...</p> <p>Looks as if it may be used to mean "reverse our way of thinking about sth" - seems to be transitive, perhaps. A scholarly use which people would probably understand if they read, but it isn't commonly known in non-scholarly English AFAIK.</p> <p>Maybe someone with the big OED can see if it's marked as rare or anything?</p>	
3	Autor	C M 2 D D (236324) 02 Mar 11 10:14

FIGURE 17.4 A discussion on the *Leo* site about the use of *unthink*.

Unfortunately, there is no real evidence that amendments to such entries actually take place. Contributions to the Jin Shan Ci Ba blogs and fora seem to consist almost entirely of didactic material. Contributors assume the role of authoritative teachers, and present advice to users on how to translate a Chinese word or phrase into English, or an English expression into Chinese. Subsequent messages then tend to thank the contributor and practise the vocabulary information that has been imparted. The posting in Figure 17.5 is typical.

今天的几个句子是有关预订的 1. I'll see if there are any flights.

我来看看有没有班机。 2. I don't want a night flight. 我不要夜航班机。

3. When am I supposed to check in?

我应该在什么时候翻办理呢 ?

FIGURE 17.5 A *Jin Shan Ci Ba* blog posting.

The contributor presents some language data with what appear to be his or her own translations, as no source is provided. The follow-on postings do not develop any discussion about the appropriacy of the translations in various contexts, but instead reproduce the English sentences, presumably as a means of memorizing them. In the process, two mistakes are introduced—one respondent writes “I don’t want a right fight” instead of “I don’t want a night flight”—but this is not corrected. My helpful Chinese informants suggested that those who registered to contribute to the *Jin Shan Ci Ba* sites might have the right to delete any critical comments relating to their postings. My informants were not eager to correct any of the mistakes themselves, because they thought that this might result in the withdrawal of their own posting rights. Thus these and many other errors remain, and are incorporated into the reference material on the site.

Like many AEDs, the *Jin Shan Ci Ba* seems keen to record idioms and new words, but of course postings on these topics are particularly problematic, because figurative language does not easily translate, and the new words are often nonce formations created for humorous effect in a specific media context. These additions to the dictionary seem to be treated with the same seriousness as the most essential vocabulary, although they are of little communicative use. Thus numerous unattested expressions are presented as English idioms in the *Jin Shan Ci Ba* fora, such as “Many girls want to marry into the purple” (where “purple” is intended to imply wealth). I also found definitions for English words that had no currency, or a very restricted one, for example: “The husbeen is the insignificant other, the husband that has given up all sense of individuality and independence to keep his wife happy.”

17.4 Conclusion

As has been noted, the digital content of AEDs is difficult to pin down, and it is scarcely surprising that they are rarely described in either the didactic or the metalexigraphic literature. However, we should take note of the fact that AEDs are increasingly the reference sources of choice for language learners around the world. Rather than ignoring them, we should be discussing them, drawing attention

to their defects, and also analysing their appeal. The more dictionary users learn to critique lexicographical content, the more likely it is that they will turn to electronic sources which provide accurate advice about lexical meaning and use. Published reviews of AEDs will help this to happen, as will classroom intervention and dictionary skills materials which direct users to the best AED components, even if these are packaged together with multiple lower-quality dictionaries. Producers of high-quality dictionaries may still be able to maintain a competitive edge, especially if they continue to develop those peripheral e-dictionary facilities such as audio and video files, word-list creation tools, language tests, and language games, all popular with users and unique to the electronic medium.

Perhaps in time user contributions will help AEDs to improve, but it seems clear that although collaboratively produced encyclopedias can be made worthy of comparison with eminent published brands, good modern dictionaries cannot be created without corpus-derived insights into grammar and usage, and collaboratively produced dictionaries will always require additional lexicographical input in order to be really useful reference tools.

Acknowledgments

I wish to thank Yeon-Kyung Bae, Liang Liao, and Wang Liyuan for their invaluable help with the translation and interpretation of Chinese and Korean e-dictionary material.

Dictionaries

Summers, Della (ed.) (2004). *Longman Active Study Dictionary*. Harlow: Pearson Education.

Alternative e-dictionary references

bab.la language portal: <http://bab.la/>.

Babylon online: <http://dictionary.babylon.com/>.

CC-CEDICT Wiki: <http://cc-cedict.org/wiki/>.

Doosan Dong-a Prime on the *Daum* portal: <http://engdic.daum.net/>.

ET-House Korean-English Dictionary: <http://www.et-house.com/>.

Jin Shan Ci Ba (Kingsoft Powerword) online portal: <http://www.iciba.com/>.

Jin Shan Ci Ba (Kingsoft Powerword) download site: <http://ciba.hp009.com/>.

Jin Shan Ci Ba (Kingsoft Powerword) discussion forum: <http://bbs.hp009.com/>.

Leo Dictionary: <http://www.leo.org/>.

Lingoes: <http://www.lingoes.cn/zh/dictionary/index.html>.

MDBG free online Chinese-English dictionary: www.mdbg.net.

Naver dictionary portal: <http://dic.naver.com/>.

Wiktionary (English language version): http://en.wiktionary.org/wiki/Wiktionary:Main_Page.

Yahoo! Hong Kong: <http://hk.dictionary.yahoo.com/>.

Yahoo! Japan: <http://dic.yahoo.co.jp/>.

Yahoo! Korea: <http://kr.dic.yahoo.com>.