

A Lexicographical Review on Collins COBUILD Advanced Dictionary of English

Jiangbo Wan

College of Foreign Languages and Literatures, Fudan University

Shanghai, China, 200433

E-mail: jbw@fudan.edu.cn

(Received: 17-5-14 / Accepted: 2-7-14)

Abstract

Collins COBUILD Advanced Dictionary of English (English-Chinese, abbreviated as CCADE in the below) is a new member of COBUILD corpus-based family. The employment of corpus has largely facilitated the procedure of data-gathering, entry-selection, definition-writing, the offering of citations, etc. Secondly, CCADE has carried forward its well-famed classroom-English defining principle, rather than the long-standing metalinguistic defining convention widely adopted by other dictionaries. However, in practice this principle is inevitably inconvenient to convey information and thus has a limited practicality. Thirdly, CCADE is exemplary in resorting to multimodality which corresponds to today's multiliteracy. Many of its measures are commendable – graphics, colors, information boxes, pictures, signals, diagrams, to name a few. In addition, CCADE is intended for a well-defined target readership – Chinese teachers and students, especially college students. Information on CET, TOEFL and TOEIC vocabulary, writing styles, etc, is laid out instructively to cater to their needs.

Keywords: CCADE, corpus-based, defining principle, metalanguage, multimodality.

1. Introduction

In 2009 and 2012, Higher Education Press (China) consecutively published two editions of *Collins COBUILD¹ Advanced Dictionary of English* (English-Chinese, abbreviated as CCADE in the below) in collaboration with Heinle Cengage Learning – respectively the full-color edition in 2009 and the bicolor edition in 2012. Claiming to be backed up with reliable and helpful corpus-based pedagogical resources, this latest semi-bilingual dictionary is tailored for Chinese teachers and students with a number of commendable user-friendly features. However, some efforts still lag.

¹ COBUILD is an acronym for Collins Birmingham University International Language Database, indicating partly the collaboration between Collins publishers and the University of Birmingham. It is a British research facility established at the University of Birmingham in 1980 and funded by Collins publishers.

2. A Brief Historical Review

ESL dictionaries, or English learner's dictionaries, became recognized as a new genre of dictionary ever since *The New Method English Dictionary* was published in 1935. Up to today, the development of ESL dictionaries has undergone two stages:

- 1) Pre-corpus ESL dictionaries, whose compilation was based on paper data as well as compilers' empirical linguistic knowledge;
- 2) Corpus-based ESL dictionaries, which profoundly rely on large electronic corpora of text for different tasks of dictionary-making: data-gathering, entry-selection, entry-preparation, definition-writing, the offering of citations and pragmatic information, etc.

The second stage was embarked upon with the début of *Collins COBUILD Dictionary of English* in 1987, which claimed to be based on the corpus of real examples of English, namely, "the Bank of English". In 1980s, Collins and the University of Birmingham led by John Sinclair, Professor of Modern English Language, developed the largest electronic corpus of the real English language that people speak and write every day. The corpus contains written material from websites, newspapers, magazines and books published around the world, and spoken material from TV, radio, and everyday conversations. New data is loaded into the corpus every month to help the Collins dictionary editors identify new words and meanings when they are first used. The birth of the lexical database revolutionized the approaches to dictionary-writing and brought forth a new generation of "corpus-driven dictionaries" for English language learners. (CCADE, xi) As a matter of fact, all COBUILD dictionaries are based on the information found in the Collins Corpus.

The advent of corpora in the late 1980s provoked fundamental changes in dictionary-making. Besides the Bank of English developed by COBUILD, there are a number of other large-scale corpora aiding the compilation of dictionaries (especially ESL dictionaries) in the past twenty years – the Longman-Lancaster Corpus, the Brown Corpus, the Longman Learners Corpus, the British National Corpus, etc. They have revolutionized the dictionary-making process and provided users with more useful and reliable information.

The past 27 years has witnessed a tremendous growth of corpus. When the first edition of *Collins COBUILD Dictionary of English* appeared in 1987, the size of the Bank of English was only 20 million words. However, up till now it contains a subset of 650 million words from a carefully chosen selection of sources to give a balanced and accurate reflection of contemporary English, and the full size of the Collins Corpus today reaches 4.5 billion words.²

3. Characteristics of the Corpus-based COBUILD Dictionaries

In traditional dictionary-making, besides a limited number of paper data and source dictionaries, lexicographers relied heavily on their personal knowledge and judgment. However, the process of making a corpus-based dictionary is very different in that computers play a crucial role during every step of dictionary-making. *The first step* is to assemble a large body of language data on computer from a variety of writings and speeches. *Then* the collected data can be presented on screen in a display for further study and analysis by means of a concordance program. Take the word 'agree' as an example (See Figure-1). The concordance data as such help lexicographers to discriminate between senses, identify recurrent collocations, and select suitable citations.

² According to the information retrieved at <http://www.collins.co.uk/page/The+Collins+Corpus> on 2014/06/17.

different nationalities broadly AGREE over landscape quality, select
 book (i.e. stop because it cannot AGREE any communication with the envir
 represent the country, I cannot AGREE with this, he stated. Despite th
 ed process (which clearly cannot AGREE to any communication). In this c
 icted: "If your candidates can't AGREE among themselves we're not going
 I don't agree with it, I didn't AGREE with it in the '80s, I was a min
 unity or dey will get ya I don't AGREE wid all a dem cause some are ver
 country". He replied: " I don't AGREE with it, I didn't agree with it
 the shortcomings of law. Experts AGREE the US breached international co
 Small groups, say three or four, AGREE among themselves that they will

Figure-1: The word “agree” displayed in the concordance program in BNC Written Corpus³

During the third step, the lexicographers’ work, which has been “guided by standardized entry formats on screen that prescribed typography and layout for each category of information”, is loaded back into the computer in the form of a well-structured database. (Cowie, 2002: 120)

Compared with pre-corpus ESL dictionaries, corpus-based dictionaries are recognized with the following characteristics: 1) being authentic; 2) being descriptive; 3) being both quantitative and qualitative.

3.1 Being Authentic

With frequency counts information and other observed facts from the corpus, lexicographers are enabled to judge whether a word is a polysemy or a monosign, whether a word or a sense is preferred in written or spoken form, whether a verb is transitive or intransitive, whether an adjective has a comparative or superlative form, and the like. Furthermore, the most remarkable function of the corpus is its sufficient examples, which in turn rightly cater to ESL learners’ demand for authentic citations, so that lexicographers can avoid devising examples that are “silly, stilted, or clearly artificial”. (Landau, 2001: 308) As a result, lexicographers have almost all that takes to offer an authentic description of the entry word. Central to the revolutionary progress of corpus-based lexicography is the pursuit of proximity to the truth of language.

3.2 Being Descriptive

The traditional learner’s dictionaries had long been more prescriptive than descriptive. The classical works of prescriptive learner’s dictionaries are *A Learner’s Dictionary of Current English* (1948), *The Advanced Learner’s Dictionary of Current English* (1963), *Oxford Advanced Learner’s Dictionary of Current English* (abbreviated as OALDCE, 1974), *Longman Dictionary of Contemporary English* (abbreviated as LDOCE in the below, 1978), etc. Many of the compilers were English educators themselves – like H.E. Palmer and A.S. Hornby, who intended to integrate the prescriptive grammar rules and collocation restrictions into their lexical reference books. In order to achieve the goal of “correctness”, they recommended some forms and banned others. Their suggestions and recommendations were mainly based on normative codifying rules which made full use of such studies as etymology, lexicology, syntax, pragmatics, etc.

However, corpus-based dictionaries – like COBUILD dictionaries – have shown more descriptivist attitude towards the language they deal with. With the availability of database materials, the collection of evidence on actual usage is no longer as “difficult, expensive and time-consuming as in the past”. (Hartmann, 2000: 37) Consequently, CCADE corpus-based lexicographers make great efforts to present the English language as it is, rather than as it should be.

³ Cited from the webpage <http://conc.lex Tutor.ca/concordancers/wwwassocwords.pl> retrieved on 2014/06/17.

3.3 Being both Quantitative and Qualitative

The traditional compilation of ESL dictionaries was overwhelmingly qualitative. One case in point was LDOCE1, ie the first edition of *Longman Dictionary of Contemporary English* (1978). Consider the following example randomly chosen from the dictionary:

epic¹ /'epik/ *adj* [Wa5] *usu. apprec* (of stories, events, etc.) full of brave action and excitement, like an EPIC²: *the epic fight of one small ship against 6 enemy ships* —
ally *adv* [Wa4]

In this entry, the lexicographical treatment is highly informative – the pronunciation, the definition, the citation, and various labels for grammar and pragmatics. Here “[Wa5]” suggests that the word “epic” is an adjective that does not usually form a comparative or superlative; “[Wa4]” means that in this type of adverb the letters “-ally” are pronounced /-li/; and the label “*usu. apprec*” is a short form for “usually appreciative” indicating pragmatic context; the capitalized “EPIC²” suggests a cross-reference for the lemma “epic²”. Such densely qualitative information constitutes an important feature of LDOCE1 and many other ESL dictionaries, in different styles though.

Ever since 1987, corpus-based COBUILD dictionaries have carried on the comprehensive qualitative treatment while they incorporate quantitative information. Look at the following example quoted from CCADE:

sauce ◆◆◆ (sauces) N-MASS A sauce is a thick liquid which is served with other food.
 酱; 调味汁 □ ...*pasta cooked in a sauce of garlic, tomatoes, and cheese.* ...用大蒜、
 番茄和奶酪调味汁煮的意大利面

In this entry, the qualitative tradition of pre-corpus ESL dictionaries is juxtaposed with the quantitative influence of corpus-based methodology. In addition to all the information concerning pronunciation, inflections, grammar, citations, Chinese equivalents, etc, we also see icons like “◆◆◆” indicating the frequency banding. In CCADE, the most frequent words are signaled with three diamonds (e.g. “better”, “international”, and “tax”), the next most frequent two (e.g. “protection”, “reserve”, and “truth”), and the ones which are less frequent have one diamond (e.g. “wedding”, “bet”, and “dish”). Words which occur less frequently but still deserve an entry in the dictionary do not have any diamonds. With the help of this hierarchy system, users will have a more concrete idea as to which words are most fundamental to be learned.

As a matter of fact, the frequency banding is only explicit quantitative information. Some quantitative information might be implicit. Compared with previous COBUILD dictionaries, some entry words or senses might be added or deleted in CCADE. The rule to include or exclude an entry word or sense is determined by the quantitative data offered by the corpus. For instance, new words like “bird flu”, “multitasking”, and “reality TV” have entered the CCADE entry list according to the index of corpus quantitative information.

The COBUILD corpus allows lexicographers to quickly search a target word or collocation from the ocean of language resources, locate instances of a target word or phrase, sort examples systematically, and, therefore, offer more reasonable discrimination of senses, more accurate definitions, more authentic citations, and much more of other useful and reliable information. The dictionary thus made reflects more facts of the English language as a whole. Just as Landau (2001: 305) pointed out, “the corpus is a tool that has breathed new life into the art of lexicography. ..., for the first time lexicographers at least have a sound basis for making decisions and can no longer plead ignorance”.

4. Definitions with Classroom English

Besides being corpus-based, the other revolutionary feature of COBUILD dictionaries is their persistent principle of writing definitions with classroom full-sentence English, which has challenged the type of metalanguage institutionalized for centuries.

4.1 A Confrontation with Conventional Defining Metalanguage

According to Hartmann (2000: 93), “metalanguage” as a lexicography term “includes such conventions as grammatical codes, labeling of usage and the formulation of definitions”. The conventional definition is usually an interpretation of the entry word in equivalent structures. For instance, a verb is interpreted with verbs or verbal phases, and an adjective is interpreted with adjectives or adjective structures.

However, COBUILD dictionaries initiated their special way of explaining meanings in 1987. They use full sentences in a “direct and informal style of the explanations”, “similar to the way in which teachers in a classroom would explain what a word means”. (CCLD, 1996: v) This defining technique is intended for both decoding and encoding learning activities, offering information of not only receptive learning activities but also the special collocations to contribute to their productive learning activities. Please compare the following examples quoted from *Oxford Advanced Learner’s Dictionary* (2000, abbreviated as OALD5) and CCADE.

OALD5:

twinkle v 1. to shine with a light that changes constantly from bright to faint. 2. ...

patient n (a) a person who is receiving medical treatment, esp in a hospital. (b) a person who is registered with a doctor, dentist, etc and is treated by her or him when necessary.

CCADE:

twinkle □ V-1 If a star or a light **twinkles**, it shines with an unsteady light which rapidly and constantly changes from bright to faint. ...

patient ◆◆◇ (patients) □ N-COUNT A **patient** is a person who is receiving medical treatment from a doctor or hospital. A **patient** is also someone who is taken care of by a particular doctor. ...

By using structures like “If ..., it ...” and “A ... is a person who ...”, CCADE attempts to tell its users in natural English which would be used by their English teachers in the classroom. Systematically, CCADE inserts the headword into a syntagmatic sequence that represents its typical environment and explains the meaning with vocabulary and grammatical structures that occur naturally with the headword.

COBUILD dictionaries vary their defining patterns with the nature of the headword. For instances:

Nouns:

- 1) (a/an/0/the) □ is (a/an/0/the/someone)⁴
- 2) (If you/someone/something) + V + □ (+ prep) (sb/sth), you ...
- 3) If you are a □, ...
- 4) Your □ is ...
- 5) If a man has □, ...

⁴ □ stands for the headword or the expression that is being defined.

Verbs:

- 1) To \bar{X} (someone/something) means **to** ...
- 2) (when/if) you \bar{X} (someone/something), you ...
- 3) When (someone/something) is \bar{X} ed, ...
- 4) (someone/something) (who/that/which) \bar{X} s ...

Adjectives:

- 1) (a man/someone/something) (who/that) is \bar{X} + **V**
- 2) (a/an/0/the) \bar{X} + **N** is/relates to, etc.
- 3) If (something/someone) is \bar{X} , it is ...⁵

Interestingly, the choice of “you” or “someone”, “a man”, etc is mysterious. Though not stated, the compilers seem to avoid disagreeable or uncomfortable associations with the pronoun “you”. For such words or expressions with derogatory or embarrassing indications as “bastard”, “bribe”, “cheat”, “kidnap”, “have an affair”, or “betray one’s country”, the compilers choose the third person so as to observe the principle of politeness.

4.2 A Controversial Defining Principle

Such a defining method has been adopted ever since COBUILD1 was published in 1987 and has become a controversial defining principle in the field of ESL dictionary making.

Béjoint refers to this defining technique as “the most notable step forward in the direction of the lexicographical treatment of idiomaticity” (1994: 222). Landau also gives his similar positive comment, “I applaud this innovation in ESL lexicography; I believe it has improved the quality of definition in many cases, especially in the treatment of idioms” (2001: 164).

While some lexicographers buy this defining technique, others don’t. Fillmore (1969: 59) compares the different treatment of “rhyme” in COBUILD1 and in LDOCE2 as follows:

LDOCE2: rhyme (with) (of words or lines of poetry) to end with the same sound, including a vowel.

COBUILD1: rhyme if one word rhymes with another, or if two words rhyme, they have a very similar sound.

He believes that LDOCE2 obviously offers more information, though COBUILD1 is wordier. In order to convey the same amount of information, COBUILD1 will have to use more cumbersome formations like “if one word or line of poetry rhymes with another or if two words or two lines of poetry rhyme, etc”. If a definition attempts to pack too much information into a single sentence, it soon becomes unmanageable. (Béjoint, 1994: 224) Therefore, to define with classroom full-sentence English is not as efficient as to define with the traditional metalanguage.

As an ESL learner, teacher and lexicographer, I, for one, do not see much overwhelming necessity in adopting this defining technique. Instead, I have found more inconveniences and liabilities related to this principle. (Wan, 2006: 12)

First of all, the redundant “full-sentence” definition makes it difficult for users to search for the most crucial part of its interpretation. Take, for example again, the definition of “patient” as a noun – “A *patient* is a person who is receiving medical treatment from a doctor or hospital. A *patient* is also someone who is taken care of by a particular doctor”. Here, the italicized part is nothing but a mere repetition of the headword, carrying no new information

⁵ Based on Béjoint’s *Tradition and Innovation in Modern English Dictionaries* (1994: 223).

at all. Besides, structures like “When you ...”, “If you call someone a ..., you think that ...”, and “If you say that ..., you mean that ...” have taken up much of the precious space which could well have been utilized to convey more substantial information. The core of the definition is consequently buried under the heavily formalistic wording, bringing extra burdens to the user.

Secondly, by so wording, the definition usually cannot convey as much information as traditional metalinguistic method. Due to the syntactical restrictions, “full-sentence” definitions can hardly allow too many points to be stated in a single sentence. As a counter policy, a lexicographer would either use more sentences or more complicated sentence structures or omit something that may sound less important. In effect, COBUILD lexicographers prefer the latter. Since they aim to show how typical collocations work, some less typical collocations are left out from the definition, which inevitably reduces the entropy of information.

Last but not least, compared with the traditional definition style, the “full-sentence” definition style has a limited substitutability. Since the traditional style uses equivalent structures to define the headword, it allows the definition wording to fit into different real contexts and therefore enjoys a more flexible substitutability. However, with the prescriptive contextual definition, more often than not, COBUILD users find it hard to substitute the target word in a real situation.

I stand firmly on the footing that dictionaries are above all a tool instead of a classroom. Admittedly, COBUILD dictionaries have initiated the defining method in the treatment of idiomaticity, but it may also have undermined the overall efficiency of the dictionary if users’ need is taken into consideration. What users expect from a dictionary should always be laid prior to what the dictionary expects its users to learn.

5. Multimodality

Recent technological developments have far-reaching impacts on the way people communicate with each other. Nowadays, texts have been enormously enriched with various semiotic modes like emoticons, pictures, diagrams, sounds, videos, etc.

By presenting two or more modes at the same time during an interaction; multimodality may be viewed as a ‘communicational ensemble’ of modes. (Kress and van Leeuwen, 2001: 111) It becomes a common phenomenon in many fields of people’s lives today – email, text messages, books, newspapers, internet, classroom teaching, conferences, so on and so forth. People used to conduct a type of linear reading mainly during their reading activities, whereas today people have to read graphically and audiovisually. Consequently, there arises a new form of literacy – multiliteracy. (Hu, 2007: 7) In order to cope with the advancement of communication, today’s citizens are supposed to be equipped with the ability to understand and produce information in multimodality. In linguistics, multimodality is also known as multi-semiotic. In the field of lexicography, multimodality refers to the integration of different semiotic codes, e.g. texts, graphics, colors, pictures, signals, boxes of special functions, diagrams, attached CDs, front matter and back matter, and the like.

Compared with all the previous ESL dictionaries – including previous COBUILD dictionaries, CCADE is nonetheless exemplary in resorting to multimodality. Many of its measures are commendable.

5.1 Colors

CCADE (2009) is a full-color dictionary in its real sense. Not only are the pictures printed in color, but also all the text has taken advantage of full-color typography. For example, headwords and their frequency bandings are printed in either blue or magenta. (For more information, please see Part 5 “Adaptation to China’s Market”.) All the sense numbers blue. All the headword numbers magenta. All the derivatives blue. All the guidewords blue. All the

phrasal verbs blue, led by a rightward triangle in blue. All the information boxes are specially colored. This colored typography effectively enhances readability and helps its users locate required information with much less efforts.

CCADE (2012), the bicolor edition, is an economic choice for both the publisher and the users. Although visually less vigorous than the full-color edition, it has exploited to the fullest measure of the two colors, black and blue – black lean type, black bold type, dark backgrounds for the headwords of special purpose boxes, blue lean type, blue bold type, blue signals, light and dark blue backgrounds, etc.

5.2 Information Boxes

5.2.1 Outline Boxes

Many details of CCADE reveal an earnest user-friendly attitude of the lexicographers. For longer entries, they especially offer a colored box of outlining information, which offers an instant visual guidance to the major meanings of the headword so that the user can immediately locate the meaning, collocation or citation he or she is looking for.

5.2.2 Usage Notes

CCADE lexicographers offer tips on language variety, pronunciation, and grammar by highlighting and explaining shades of meaning, clarifying cultural references, and highlighting important grammatical information. They understand that usually ESL learners have a lot of difficulties in differentiating British English, American English, and other native varieties of English, in pronouncing a certain word correctly in different senses or contexts, and in applying some subtle grammatical rules. To resolve these uncertainties for ESL learners, they make full use of information boxes.

As is seen in Figure 2, the boxed concise information of language variety, pronunciation, and grammar is extremely clear and helpful, thus enhancing its readability to a great extent.

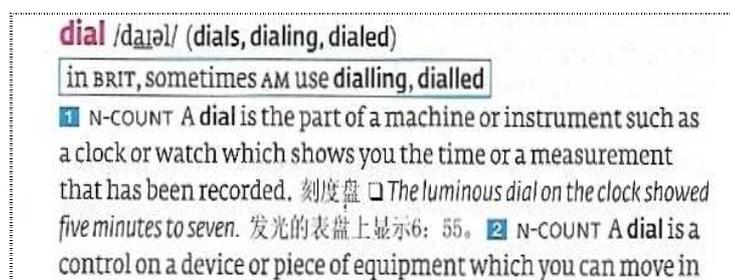


Figure-2: Example of “Usage Notes” from CCADE

5.2.3 Demarcations between Semantically Related Words

Some semantically related words, for example “expect”, “wait for”, and “look forward to”, are very close in meaning and confusing to ESL learners. By using the mode of information boxes, CCADE informs its users with extra demarcation interpretations and strengthens the pedagogical function of the dictionary.

5.2.4 Thesaurus

Thesaurus can effectively help upsize users’ word power. The thesaurus boxes of both synonyms and antonyms for high frequency words have transformed CCADE into a mini-thesaurus at once.

5.2.5 Collocations

As aforementioned, CCADE's full-sentence defining technique can hardly include all the important collocations into a single definition. Therefore, the information boxes of collocations, namely "Word Partnership", compensate for this disadvantage to some extent.

5.2.6 Morphology

CCADE increases users' language awareness also by showing how words are built in English, which is useful for learners in all areas of academic work as well as in daily communication. Focusing on prefixes, suffixes, and word roots, each morphology box – namely "Word Link" – provides a simple definition of the morpheme and then gives three examples of it used in a word, encouraging users to look up these words to further solidify their understanding (see Figure-3).

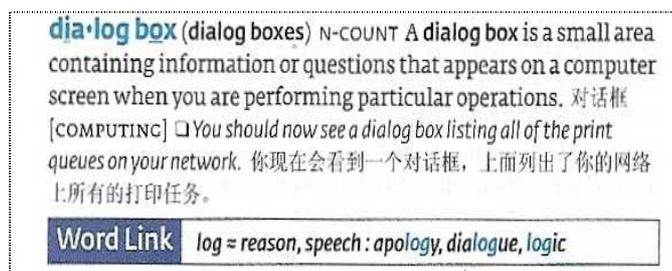


Figure-3: The "Word Link" box illustrating the meaning of the root "log"

5.3 Pictures

Compared with full texts, pictures are more pleasant to eyes and minds. We tend to remember the information illustrated with pictures rather than if it is written in full texts. In some cases, pictures are more effective than word pictures, e.g., pictures of vegetables and fruits function far better than verbal descriptions of those. Just like recent editions of LDOCE and OALD, CCADE also uses a large number of pictures to interest and cater to its user.

In CCADE, pictures are printed also in the form of information box. They mainly fall into two categories: Word Web and Picture Dictionary.

5.3.1 Word Web

Word Web boxes in CCADE present topic-related vocabulary through encyclopedia-like readings combined with brilliant pictures, creating opportunities for deeper understanding of the language and concepts and encouraging users' language exploration. Take, for example, the Word-Web box entitled "laugh". Here all key words ("laughter", "humor", "bursts out laughing", "howling with laughter", and "giggles") in bold are defined in the dictionary. Upon looking up one word, learners discover other related words that draw them further into the dictionary and the language. By weaving related words into a systematic web, CCADE is more like an activator than other dictionaries.

5.3.2 Picture Dictionary

Picture Dictionary boxes are usually used to illustrate a process or concept, especially that benefits from a visual presentation. In CCADE, pictures like "color", "core", "jewelry", "age", "egg", "dessert", "cut", "body", "basketball", "barn", etc are both informative and entertaining.

But for the outline boxes and usage notes, there are estimated over 3,000 pedagogical features in CCADE, mainly in the form of information boxes and pictures. All these modalities have effectively increased the readability and information entropy on the whole.

5.4 Icons

CCADE uses a set of visual icons to enhance information and clarify the layout. For example, the aforementioned diamond patterns like “◆◇◇” are used to indicate frequency banding; the square pattern (□) is used to introduce a citation; the magenta star (★) is used to signal CET-6 words; the magenta triangle (▲) is used to signal CET-6 plus words; the simple arrow (→) suggests cross-reference; the wide arrow (⇒) introduces compilers’ suggestions for using the dictionary; the underlines as in “good evening” suggest the correct stresses in pronouncing the phrases; the blue numbered squares (e.g., ⑧) signal different senses; the black dot (•) introduces a derivative. In short, it makes full use of varied icons to concisely guide the user.

5.5 The Front Matter and the Back Matter

The front matter and the back matter are two components in a dictionary’s macrostructure. The component parts preceding the central word-list section are the front matter, and the component parts “located between the central word-list section and the end of the work” are the back matter. (Hartmann, 2000: 12)

CCADE has been very original in utilizing the front matter and the back matter. Besides the usual components, CCADE also has an “Activity Guide” in the front matter, which actually is a series of educational exercises for its users. Users may learn to recognize different types of information and optimize the benefits rendered by the lexicographers.

In the back matter, there are also a number of ingenious modes of information that other dictionaries would envy: “Index”, “Brief Grammar Reference”, “Brief Writer’s Handbook”, “Brief Speaker’s Handbook”, “Words That Frequently Appear on TOEFL and TOEIC”, “Text Messaging and Emoticons”, “Academic Word List”, “USA State Names and Capitals”, “Geographical Places and Nationalities,” and “Words to Remember”.

The 100-page “Index” is a very original attempt. As a matter of fact, this is an alphabetically-ordered index of the Chinese equivalents found in the dictionary proper. Accordingly, this index directs the user to the relevant English entry through the medium of Chinese, which adds to the dictionary’s accessibility in a potential way. Though the CCADE claims this index “is not a dictionary as such”, it almost serves as a quick Chinese-English dictionary for the user.

Besides, “Brief Writer’s Handbook” is a very usual manual to help ESL learners and teachers fulfill different tasks – writing essays, dissertations, letters, and resumes, and marking students’ papers. It includes such important parts as Punctuation, Capitalization, Italicization, Spelling, Grammar, Documentation, Letter Format, Resumes, and Proofreading Marks.

The most interesting part of CCADE is “Words to Remember”, a section open to its users. They are like an unfinished address list for the users to fill in their own new vocabulary, definitions, and sentences. This dictionary is interactive, considering that the users also participate in dictionary writing. And, this dictionary is growing, since it unfolds infinite possibilities of personal gains and discoveries.

The multimodality of CCADE allows linear reading and graphic reading to compensate each other, thus enhancing the efficiency of dictionary consultation. Besides the modes aforementioned, multimodality is also reflected in other modes like different types of fonts, colored thumb index, the attached CD, etc. The guideline of multimodality is embodied in numerous microscopic details as well as its macroscopic planning.

6. Adaptation to China's Market

CCADE reaches out to China's users not only through its semi-bilingual orientation, but also through its consideration of China's potential users. It explores China's college English teaching syllabus and recognizes CET vocabulary in its layout.⁶ In the dictionary, CET-4 vocabulary words are printed in the color of magenta. CET-6 words are magenta and are marked with magenta star (★). CET-6+ words are magenta and are marked with magenta triangle (▲). All such devices are very user-friendly for Chinese young people, especially to college students who are expecting to enter the future competitive job markets with a presentable English qualification.

CCADE is a semibilingual dictionary, which offers translations for all definitions, senses, and examples. Those translations complement the English monolingual material and provide additional support to the users when they encounter a difficult word or expression.

In addition, CCADE still takes into consideration the practical needs of Chinese college students for further overseas studies or future employment. In its back matter, an alphabetical list of "Words That Frequently Appear on TOEFL and TOEIC" makes the dictionary also a handy reference book for TOEFL and TOEIC preparation, which would definitely caters to a large number of its college customers.

Unlike previous COBUILD dictionaries, CCADE stresses more on American English than British English, with consideration of the special needs of China's market where learners are eager to learn American English.⁷ For the first time, a COBUILD dictionary marks only the American pronunciation. For grammar and collocations, CCADE also gives as much information as possible about differences between American and British Englishes. In the "Brief Writer's Handbook" from the back matter, it introduces mainly American writing styles – Chicago Style and MLA Style. Also in the back matter, CCADE even offers a list of "USA State Names and Capitals". No wonder Professor Hu Zhuanglin (2009) said jocularly or regretfully that this dictionary should have been entitled as "Collins COBUILD Advanced Dictionary of American English" to ensure a better marketing in China.

7. Enlightenment to Chinese Lexicographers

CCADE enjoys many notable features: corpus-based compilation, natural-English defining principle, multimodality, China-oriented, etc. Though some efforts might not prove successful, it still sets a good example for bilingual dictionary compiling, especially for China's bilingual lexicographers.

As aforementioned, corpus-based compilation can ensure the reliability of data-gathering, entry-selection, definition-writing, citations, and pragmatic information. But in China's lexicographical circle, corpus-based compilation is still far from a desirable level. Though some dictionaries have claimed to be based on bilingual corpus, the outcome is not comparable to that of English monolingual dictionaries. On the one hand, parallel corpora in China are still undersized. As is known, a corpus will only boost dictionary compilation when it reaches a considerable size (100 million bytes at least) and covers a wide range of language styles and sources. However, there are only few parallel corpora in China, to name a few, CONULEXID built jointly by Nanjing University and the Commercial Press, HKUST built by the Hong Kong University of Science and Technology, and some other small-sized corpora built by different universities for linguistic and other purposes. (Li, 2006: 44) Another problem has to do with the cost. As is known, to establish, process and maintain a large-size corpus requires highly on human and financial resources, and it usually takes tens of years to

⁶ CET is an abbreviation for College English Test of China Mainland. In the below, CET Band Four is abbreviated as CET-4, CET Band Six as CET-6.

⁷ Another reason could be that CCADE is jointly published with Cengage Learning, an American publisher.

collect and process sufficient raw data. However, Chinese lexicographers or publishers usually cannot afford such a high cost of resources and wait that long for a product to come out. To tell the truth, as bilingual dictionary compilers, we have to optimize www.baidu.com, www.google.com.hk, and other searching engines as our personal corpus, which prove at least better than nothing. Therefore, this paper hereby would call on more constructive cooperation between the few corpora in China and between lexicographers and corpora scientists. Only by striving together can we combat the problem and actually build up a high quality corpus-based bilingual dictionary.

Even when marching into the 21st century, with all the significant infiltration of Internet technology to homes, schools, and offices via various access methods, and with all the changes in people's way of reading, China's lexicographers have not reacted promptly enough to cater to the changed users or the changed market. Most of our dictionary products are text-exclusive and non-interactive, unfolding an unfriendly posture, whether consciously or unconsciously, to the changing readership. Admittedly, multimodality makes dictionary-making more taxing to compile, but there is no denying that compared to contemporary magazines, newspapers, or even textbooks, our dictionaries have lagged far behind. Therefore, there is an urgent need that today's dictionary compiling resort to the new technology of communications, as well as the reading habits of present-day users. The colors, multipurpose information boxes, information icons, pictures, visual designs, and even sounds and videos form the new components of today's media and writings. A dictionary aims to be informative, as much as possible, and multimodality should be a necessary consideration for all lexicographers nowadays. So long as it helps enhance readability and information delivery, it should be called a reason to change. So far, multimodality has not been widely recognized in China's dictionary compiling. We should neither ignore the trend nor neglect the need of the younger generation. Otherwise, we may lose our future readers.

8. Conclusion

As a new member of COBUILD family, CCADE has inherited its corpus-based tradition profoundly, which has largely facilitated the procedure of data-gathering, entry-selection, definition-writing, the offering of citations, etc. Secondly, CCADE has carried forward its well-famed classroom-English defining principle, rather than the long-standing metalinguistic defining convention widely adopted by other dictionaries. However, in practice this principle is inevitably inconvenient to convey information and thus has a limited practicality. On the other hand, CCADE is exemplary in resorting to multimodality which corresponds to today's multiliteracy. Many of its measures are commendable – graphics, colors, information boxes, pictures, signals, diagrams, to name a few. In addition, CCADE is intended for a well-defined target readership – Chinese teachers and students, especially college students. Information on CET, TOEFL and TOEIC vocabulary, writing styles, American English, etc is laid out instructively to satisfy their needs.

Honestly but unfortunately, the prospect of paper language dictionaries does not look very engaging, be it CCADE or LDOCE. Even if they boast so many merits, its genetic problem – paper medium – becomes the heel of Achilles. Today, electronic dictionaries, especially pocket e-dictionaries, constitute a sizable threat to traditional dictionaries. Last semester when we offered the latest LDOCE as a prize for student winners in a campus English competition, many of them politely declined the gift because it was too big to carry. The paper dictionary compilers, especially learner's dictionary compilers, are faced with a dilemma. Their appreciable efforts only generate an underappreciated product. Even if some dictionaries are accompanied with CDs, those CDs still require a computer or laptop to function.

On the other hand, pocket e-dictionaries, dictionary applications (eg those intended for Android devices, iPhone, iPad, Surface) are fast upgrading with much larger capacity – with tens of dictionaries in one – and with much more functions – offering audio information, assisting personal language learning, note-taking, entertaining, etc. Besides being equally reliable, they are faster, lighter, more informative, more multimodal, more interactive, and more multifunctional. If we take a look at how computers upgrade, how tape recorders

transform, and how telephones develop, we would definitely be aware of the leap that it takes for dictionaries to become a must in our life. The sooner, the better.

References

- [1] H. Béjoint, *Tradition and Innovation in Modern English Dictionaries*, (1994), New York: Oxford University Press.
- [2] A.P. Cowie, *English Dictionaries for Foreign Learners: A History*, (2002), Oxford University Press.
- [3] C.J. Fillmore, Review of COBUILD dictionary of the English language, *International Journal of Lexicography*, 2(1) (1969), 57-83.
- [4] R.R.K. Hartmann *et al.*, *Dictionary of Lexicography*, (2000), Beijing: Foreign Language Teaching and Research Press.
- [5] Z. Hu, Dictionary digitalization and foreign language teaching, Shanghai: *Seminar on Dictionary Digitalization and English Teaching*, (2009).
- [6] Z. Hu, Multimodalization in social semiotics, *Language Teaching and Research*, 122(1) (2007), 1-10.
- [7] G. Kress and T. van Leeuwen, *Multimodal Discourse: The Modes and Media of Contemporary Communication*, (2001), London: Hodder Education Publishers.
- [8] S.I. Landau, *Dictionaries: The Art and Craft of Lexicography*, (2001), Cambridge: Cambridge University Press.
- [9] D. Li, Use of parallel corpus in bilingual dictionary making, *Journal of PLA University of Foreign Languages*, 29(3) (2006), 41-44.
- [10] J. Wan, *A Study on Bilingual Dictionaries and Translation*, (2006), Shanghai: Fudan University Press.
- [11] CCADE (*Collins COBUILD Advanced Dictionary of English*), (2009, 2012), Beijing: Higher Education Press.
- [12] CCLD (*Collins COBUILD Learner's Dictionary*), (1996), London: Harper Collins Publishers.
- [13] LDOCE (*Longman Dictionary of Contemporary English*), (1978, 1987, 1995, 2003), London: Longman Group UK Ltd.
- [14] OALDCE (*Oxford Advanced Learner's Dictionary of Current English*), (1974), London: Oxford University Press.
- [15] OALD (*Oxford Advanced Learner's Dictionary*), (2000), Beijing: The Commercial Press, Oxford: Oxford University Press.
- [16] <http://www.collins.co.uk/page/The+Collins+Corpus>, Retrieved on 2014/06/17.
- [17] http://www.lex tutor.ca/concordancers/concord_e.html, Retrieved on 2014/06/17.