

The Idiomatization and Constructionalization of ‘on thin ice’:  
A Diachronic Corpus Study . . . . . HIDETOSHI INOO 1

Phonetic Transcriptions in Free Online English-Japanese Dictionaries:  
A Comparative Analysis . . . . . HIROKO SAITO 15  
YOKO UCHIDA  
JUNKO SUGIMOTO  
MANAMI HIRAYAMA  
MARINA ARASHIRO

An Analysis of the *Collins COBUILD Advanced American English  
Dictionary*, Third Edition . . . . . TAKAHIRO KOKAWA 57  
RIKA AOKI  
RUMI TAKAHASHI  
KAZUO IKEDA

投稿規定 . . . . . 111  
編集後記 . . . . . 111

Iwasaki Linguistic Circle  
c/o Eigo-senko kyoiku junbishitsu  
Tokyo University of Foreign Studies  
3-11-1 Asahi-cho, Fuchu-shi  
Tokyo 183-8534  
Japan

Shoji Miyai, Chairman

2025年6月20日印刷 2025年6月29日発行

編集兼発行者 岩崎研究会 代表者 宮井捷二  
振替口座 00140-2-8737

〒183-8534 東京都府中市朝日町 3-11-1 東京外国語大学 英語専攻教育準備室  
email: iwaken.info@gmail.com

印刷所 有限会社紫藤印刷

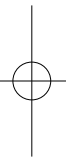


# The Idiomatization and Constructionalization of 'on thin ice': A Diachronic Corpus Study

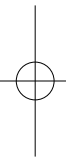
HIDETOSHI INOO

## 1. Introduction

This article examines the diachronic changes in the meaning and usage of the English phrase *on thin ice* through corpus data analysis, exploring how a compositional form evolved into a non-compositional idiomatic construction.

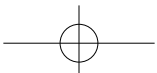


For this study, an idiom is defined as a linguistic expression comprising two or more lexical items whose meaning cannot be precisely predicted from its individual components. In other words, an idiom is a linguistic form with non-compositional meaning, such as *kick the bucket* 'die'. This definition aligns with Constructionist approaches that view construction as a pairing of form and function stored in the speaker's cognitive system. An idiom represents a specific type of construction characterized by its non-compositional meaning, requiring memorization to effectively communicate a particular conceptualization.



In everyday discourse, 'idiom' can refer to a broader range of expressions, including proverbs or customary collocations. In linguistic research terms, on the other hand, Taylor (2012: 69–99) proposed a definition to include expressions with syntactically divergent characteristics, for example, *him write a novel?* and *in fact*. However, for the purposes of this research, I will maintain a more focused definition centered on non-compositional meaning.

Traditionally, idioms have been marginalized in mainstream linguistic studies, particularly within the generative tradition, viewed as peripheral phenomena that do not cleanly fit into lexical or grammatical frameworks, due to their idiosyncrasies in meaning and/or syntactic behavior.



However, Fillmore, Kay, and O'Connor (1988: 534) challenged this perspective, arguing that idioms are ubiquitous and often highly productive. Dismissing them as exceptional would leave substantial aspects of language unexplained.

Idiomatization is used in this study to mean the change of a linguistic form consisting of two or more lexical items from compositional usage to idiomatic and figurative usage. Likewise, constructionalization means the change of a linguistic form from a non-construction to a construction. (The definition of a construction and constructionalization is discussed more in detail in Section 2.) Idiomatization and constructionalization take place simultaneously because they share the same mechanism of memorization or entrenchment of a certain combination of linguistic forms and its meaning in the speaker's mind. The difference is that idiomatization involves the non-compositional meaning of the linguistic form while constructionalization does not necessarily concern the meaning changes but formal changes.

Based upon these frameworks, this article examines the diachronic changes of *on thin ice*. In Section 2, the theoretical frameworks of this study are briefly reviewed, in Section 3 diachronic changes of *on thin ice* are analyzed through corpus data, in Section 4, possible motivations behind the changes are discussed, and in Section 5, the conclusion is presented.

## 2. Studies of Constructions and Idioms

### 2.1 Constructions and idioms

The definition of construction that most researchers in Constructionist approaches share is that by Goldberg (2006):

conventionalized pairings of form and function (Goldberg 2006: 3)

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg 2006: 5)

The definition of idioms in this present study is consistent with the definition of construction above; they are pairings of form and non-compositional meanings that need to be memorized by the speakers.

According to the usage-based model, which Constructionist approaches adopt as a theoretical background, language is acquired through repeated exposures to actual language uses. The same process that works during language acquisition plays an important role in the changes to an established adult language. Bybee (2015: xv) states that “language change takes place during language use, and the mechanisms that drive change are the psycholinguistic or cognitive process operating in everyday conversation and language use.”

In examining the evolution of the phrase *on thin ice* from a compositional expression to an idiom in this study, the Constructionist and the usage-based model as summarized above are assumed.

## 2.2 Diachronic changes in construction or constructionalization

Constructionalization is defined as the creation of new (combinations of) form-meaning signs by Traugott and Trousdale (2013: 22). The creation is not limited to a completely new form and new meaning; the transition from compositional form to idiom can be an instance of constructionalization if any new element exists in the combination that makes a sign. As stated in 2.1 above, a construction is a pairing of linguistic form and function stored in the mind of the speaker. Speakers use constructions as building blocks to say what they want to say. Each construction has a specific fixed function and is combined with other constructions to form a sentence. Such a sentence can be improvised online at the time of utterance. In contrast, a construction cannot be created online; it needs to be memorized by the speaker in advance. The transition from an impromptu speech to fixed and memorized knowledge may be called constructionalization. This process usually does not occur consciously. The usage-based model suggests that as the language user hears and utters some set expression repeatedly, the memory gradually and unconsciously becomes embedded.

### 3. Diachronic Changes of *on thin ice*

#### 3.1 Why *on thin ice*?

Before this study, approximately 2,000 idioms were collected from linguistic articles and dictionaries of idioms in both print and online sources. From this collection, I randomly selected idioms and traced their frequency distribution over time using the Corpus of Historical American English (COHA, Davies 2010-), aiming to identify the earliest stages of their idiomatic usages. COHA provides usage data of mainly American English, both written and spoken, from 1820 through 2019. Many of the traced idioms were established as fixed expressions prior to the 1820s, predating the earliest records in the corpus. The expression *on thin ice* proved to be an excellent example for examining how the meaning and usage of a certain form evolved over time because its earliest usage was within the time frame of COHA. The first usages of *on thin ice* on COHA appeared in the 1860s and remained through subsequent decades. The expression showed a remarkable increase towards the end of the 20th century, with usage remaining high since then (Table 1). A detailed analysis revealed that earlier examples were predominantly used in a literal sense (e.g. (1)), while usages from the 20th century onward were primarily in figurative and idiomatic meaning ‘in danger’ (e.g. (2)). This observed transition in usage will be discussed in detail in the following sections.

Table 1 Frequency of *on thin ice* in COHA (1820s-2010s)

decade	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
frequency	0	0	0	0	1	1	1	1	1	2	2	2	5	3	4	3	9	13	6	14

- (1) Talking with him was like skating **on thin ice**, and his companions had a constant mental vision of spots designated “dangerous.” (1870 fiction; emphasis added, likewise hereafter)
- (2) There was little in Job’s demeanor to reflect accurately whether he was secure or **on thin ice**, whether money was plentiful or scarce. (1950 fiction)

#### 3.2 Data

The sentences that include *on thin ice* were extracted from the COHA

full-text data using a custom Python script. To accurately distinguish between literal and figurative meanings, each extraction includes not only the sentence containing the phrase but also two preceding sentences. A total of sixty-eight usages were collected, spanning various genres and decades.

### 3.3 Method

The data have been tagged manually. The tag items are: (a) decade (1820s-2010s), (b) genre (academic, fiction, magazine, news, TV/movie), (c) meaning (literal, simile, figurative), (d) the verb of the clause or phrase where *on thin ice* occurs, (e) the subject (person, singular/plural, animate/inanimate) of the verb phrase, and (f) verb form (past/present/future, progressive/perfect, gerund, other).

Tag items (a) decade and (b) genre were adopted from the original file names of COHA full-text data, which reflect the categorization in COHA. Regarding (c) meaning, in simile, the phrase *on thin ice* itself is used in literal meaning and preceded by simile markers such as *like* or *as if*, as shown in (1). However, as we will see in Section 3.4, all the literal uses in the data were simile, and no usages in literal and non-simile were found, therefore the tag label simile (literal) was used in later sections. In (f) verb form, although English has no morphological future markers, to observe the semantic difference, a separate *future* tag was reserved for periphrastic future expressions such as *will* and *be going to*.

### 3.4 Results

The results of tallying the tag items are shown in Table 2. The data for the decades 1820–1850 have been suppressed in this table because there were no values (no usage examples) during this period.

In the 19th century, the expression was more often used in literal meaning as a simile, as in (1) shown above. In the 20th century, metaphorical usages gradually increased, and in the 21st century, only metaphorical meanings were found. In the following paragraphs (and corresponding sections of Table 2), each tag category (genre, verb, etc.) is divided by the meaning (simile or metaphor) first, and the characteristics of each meaning are analyzed, contrasting the simile (literal) and

Table 2 Usages of *on thin ice* in COHA

on thin ice (n=68)			1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	
		decade																	
		total count	1	1	1	1	1	2	2	2	5	3	4	3	9	13	6	14	
		simile (literal)		1	1	1		1			1					1			
		metaphor	1				1	1	2	2	4	3	4	3	9	12	6	14	
genre	simile (literal)	fiction		1	1	1		1			1					1			
		magazine																	
		news																	
		tv/movie																	
	metaphor	fiction	1					1	2	1	2	2	1	2	2	6	2	5	
		magazine					1				1	1					1	1	
		news									1				1	1			
		tv/movie										1		3	1	6	4	3	9
verb	simile (literal)	skate		1							1								
		walk				1	1												
		tread																	
		run						1											
		be																	
		other															1		
	metaphor	skate	1				1		1		3	1	1		2	2		1	
		walk														1		1	
subject	simile (person)	1SG																	
		2SG																	
		3SG			1	1	1		1			1					1		
		1PL																	
		2PL																	
		3PL																	
subject	metaphor	1SG									1	1	1		2	3	1	2	
		2SG								1	1		3	2	5	5	3	5	
		3SG	1				1	1	2		1	1			2	3	1	5	
		1PL													1				
		2PL																	
		3PL								1	1	1					1	1	2
subject	simile (animacy)	animate		1	1	1		1			1					1			
		inanimate																	
	metaphor	animate	1				1	1	2	2	4	3	4	3	9	11	5	11	
		inanimate															1	1	3
verb form	simile (literal)	present																	
		present progressive																	
		present perfect																	
		past															1		
		past progressive					1												
		past perfect																	
		gerund		1	1														
		other							1			1							
	metaphor	present									1	1	1	1	3	6	7	3	8
		present progressive										1		2		2	3	2	2
		present perfect																	1
		past							1				2			1	1		2
		past progressive										1					1	1	
		past perfect																	
metaphor	gerund									1									
	other	1				1	1	1			1		1					1	

figurative usages.

**Genre:** The phrase *on thin ice* was used more in fiction and TV/movie genres. It was less used in formal registers, such as magazine and news.

**Verb:** Toward the beginning of the 20th century, the verbs that collocate with the phrase were *skate*, *walk*, *tread*, and *run*, which describe the physical movement on the surface of thin ice (e.g. (3)). After that period, *be* and other verbs were increasingly used with the phrase such as in (4). This shows the transition of *on thin ice* from compositional expression coupled with surface movement verbs to independent, non-compositional expression with its own meaning that can work independently from the verb meanings.

- (3) He followed her into the house, trying to lessen his weight as if he were **walking** on thin ice; and the old house cracked its knuckles, but his foot-fall made not a sound. (1980 fiction)
- (4) Watch it, mister. You're on thin ice. (1960 TV/movie)

**Subject (person/animacy):** In earlier years up to the 1920s, the subjects that came with *on thin ice* were singular and animate, as can be observed in (3) above. Considering that the literal meaning of the usages was common in the corresponding period, more than one person being on thin ice would result in the immediate collapse of ice and their plunge into ice water, diminishing the effect of the phrase to express the risky situation of the subject in a subtle balance between her/his weight and the proof stress of the ice.

The transition from animate subjects to inanimate ones was observed in the 1990s and onward. This symbolizes the semantic change of *on thin ice* from compositional to figurative in another aspect. In literal meaning (and simile usages), the subject that moves on the surface of thin ice should be an animate entity that can maneuver its movement carefully. As the meaning shifts to the figurative, free from the danger of physically plunging into cold water, an inanimate subject can be used with the phrase to describe the risky situation whatever the subject may be. (5) is an example of such usage.

- (5) **These planes** are on thin ice. (2010 TV/movie)

**Verb form:** Change in tense and aspect is also noted. From the mid-20th century, when the transition to metaphoric uses occurred, present

tense and progressive aspect occurred more frequently. In literal and physical meaning, it is unlikely that somebody is on thin ice. But, in metaphorical usage, it is possible that somebody is at risk as in (6).

- (6) Right now, you **are walking** on thin ice. (2010 TV/movie)

#### 4. Discussion: Motivations for idiomatization

##### 4.1 Bodily experience and metaphorical mapping

The literal meaning of *on thin ice* can describe some object, probably a person, that is in contact with the upper surface of frozen water with apparently insufficient thickness to withstand her/his weight. This situation described in the literal sense will lead us to imagine that the person is prone to submerge into the ice-cold water beneath the ice because the ice might not be thick enough to support the person's weight. This imagery evoked by *on thin ice* may be based on our bodily experience of seeing, touching, or breaking thin ice in our lives. Most of us have not actually walked on thin ice even as a child, but normal cautions given by nearby adults in the locality of thin ice over a lake or pond or a depiction of such a scene on TV or other media combined with the bodily experience of handling a piece of thin ice would be sufficient to form a sense of danger in our mind.

This kind of encyclopedic knowledge based on our bodily experience about thin ice enables us to feel and understand the situation intuitively and easily by mapping the literal meaning of walking on thin ice onto other dangerous situations. This mapping from the risk of walking on thin ice to other risks is an example of metaphor, as theorized by Lakoff (1987: 276–278). According to Lakoff's theory, a "metaphor has a source domain, a target domain, and a source-to-target mapping". The source domain is a schema that is pervasive in experience, thus easy to understand, and with a simple and clearly distinguishable structure. In the example of *on thin ice*, the literal situation of walking on thin ice is the source domain. As mentioned before, this situation is familiar to most of us and easy to understand. The target domain is the risky situation described by the metaphorical reference to the phrase *on thin ice*, such as

a financial situation in (2), or aviation conditions in (5). There is a correlation between the source domain and the target domain, the riskiness of the situation, in the case of *on thin ice*.

A cross-linguistic comparison may be helpful to see the difference and significance of embodied encyclopedic knowledge behind the expression. In Japanese, there is a comparable phrase *hakuhyo o fumu* (薄氷を踏む) ‘step on thin ice’. This expression is synonymous with its English counterpart, meaning ‘in danger’, or ‘to take a risk’. This Japanese phrase appears to have grown on its own, because it can be traced back to as early as 1225 in *Hogen Monogatari* (*The Tale of Hogen*) (NINJAL 2024), far before Japan and Europe had direct contact, and thus should not be a translated expression from English or another European language. In contrast, Vietnamese or Spanish do not have similar expressions (private conversations with a native Vietnamese speaker-linguist, and a researcher of Spanish, respectively; no comparable expressions found in the *Cambridge Dictionary* (English-Vietnamese, English-Spanish)). In the warmer areas where these languages are spoken, ice does not occur naturally, therefore speakers of these languages would have little opportunity to experience the physical characteristics of thin ice on a frozen pond. Therefore, comparing a risky situation to walking on thin ice would not be easy to understand, or make sense, to the speakers of these languages.

#### 4.2 Motivations for speakers

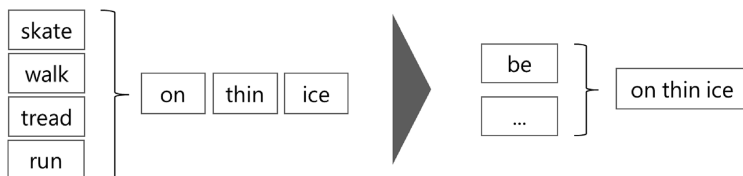
Thus, the figurative expression *on thin ice* is easy to understand and can communicate vivid sensations of a risky situation to English speakers (at least in the United States, where COHA data were collected). It was first used in simile, with such periphrastic markers as *like* or *as if*, explicitly indicating that being on thin ice is not factual but figurative during the 19th century. This expression seemed to be widely accepted, and repeated use led to its metaphorical usage without simile markers. Social change in the United States, such as industrialization, motorization, and financialization that took place in the 20th century (Wright 1990) might have played a role in making this phrase that describes a

risky situation, popular among American English speakers, who might have been increasingly aware of the risks involved in their changing lifestyles. The paring of the form *on thin ice* and its meaning of ‘in danger’ was understood and memorized by many speakers to the extent that using this phrase independently from such surface movement verbs as *skate*, *walk*, *tread*, or *run* was not inappropriate in mass media, including novels, magazines, and movies. By the time around the 1950s, it established itself as a full-fledged idiom, in other words, it was idiomatized, and at the same time, it was constructionalized. The possible motivations for the idiomatization and constructionalization of *on thin ice* could have been the sensory appeal of the physical characters of the thin ice, and social changes that made people aware of risks they face in everyday life. Whatever the driving force of the diffusion of metaphorical uses of *on thin ice*, it was accepted by a wide constituency of the American English speech community in the 20th century, through repeated use among such community members.

### 4.3 Constructionalization of *on thin ice*

The constructionalization of *on thin ice* is diagrammatically presented in Figure 1, where constructions are shown in boxes. On the left side, words (constructions) are combined to signify a meaning that is a sum of constituent word meanings. In this usage, surface movement verbs (*skate*, *walk*, etc.) collocate with the phrase *on thin ice* to describe the physical situation. In contrast, on the right side, constructionalized *on thin ice* has its own non-compositional meaning ‘in danger’, metaphorically extended from its literal meaning, which can collocate with the copula *be*, and other verbs not necessarily related to surface movement.

Figure 1 Diagrammatic representation of constructionalization of *on thin ice*



## 5. Conclusion

In the 19th century, the phrase *on thin ice* was used in literal meaning and simile with verbs indicating surface movement such as *skate*, *walk*, *tread*, and *run*. Since the beginning of the 20th century, it was increasingly used in metaphorical meaning, and the verbs used with it diversified into copula verb *be* and others with a broader sense from those described surface movement used in literal meaning. The usage with *be* can be analyzed as the phrase turned from a complement within a predicate verb phrase to a standalone adverbial. This change of usage is also reflected in the diversified genre, tense, subject person, and animacy. In the 21st century, simile (literal) usages were not found on COHA. This suggests that this expression has been entrenched in English speakers’ minds on its own, and constructionalization and idiomaticization are completed.

By describing the change of *on thin ice* from a compositional form to a non-compositional idiom with quantitative evidence, this study showcased a concrete example of idiomaticization and constructionalization.

There is a limitation to this study. COHA is the largest historical English language corpus that contains usages from 1820 to 2019 with more than 475 million words. Its data have been taken from materials available to the public, which means that such data as private conversations are not included in COHA. Considering that idioms tend to be used more frequently in informal, casual speech (Nunberg, Sag & Wasow 1994; Inoo 2025), it is possible that COHA data do not proportionately represent the true usage of idiomatic *on thin ice*. Also, because regional differences were not sorted out in COHA, any such potential differences could not be analyzed. Still, the methodology used here should be reasonably empirical, and if further data become available a similar method could be applied and more precise results would be obtained.

## ACKNOWLEDGEMENTS

The author’s thanks go to Mayu HORIE and NGUYEN Gia Thoai Du, both Graduate School of Global Studies, Tokyo University of Foreign Studies, for their valuable com-

ments. All remaining errors in this article are mine.

### WORKS CITED

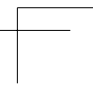
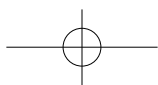
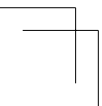
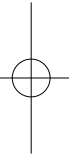
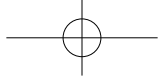
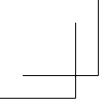
- Bybee, Joan. *Language Change*. Cambridge University Press, 2015.
- Fillmore, Charles J., Paul Kay and Catherine O'Connor. "Regularity and idiomaticity in grammatical constructions: The case of *LET ALONE*." *Language*, vol. 64, no. 3, 1988, pp. 501–538.
- Goldberg, Adele E. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, 2006.
- Inoo, Hidetoshi. "How speakers choose idiomatic constructions or compositional forms in English: A cognitive and functional analysis of synonymous pairs." *Gengo Chiiki Bunka Kenkyu (Language, Area and Culture Studies)*, vol. 31, 2025.
- Lakoff, George. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, 1987.
- Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow. "Idioms." *Language*, vol. 70, 1994, pp. 491–538.
- Taylor, John R. *The Mental Corpus: How Language is Represented in the Mind*. Oxford University Press, 2012.
- Traugott, Elizabeth Closs, and Graeme Trousdale. *Constructionalization and Constructional Changes*. Oxford University Press, 2013.
- Wright, Gavin. "The origins of American industrial success, 1879–1940". *The American Economic Review*, vol. 80, no. 4, 1990, pp. 651–668.

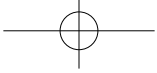
### DICTIONARIES and CORPORA

- Cambridge Dictionary*. Cambridge University Press & Assessment. <https://dictionary.cambridge.org/> (last visited: 20 September 2023)
- Davies, Mark. *The Corpus of Historical American English (COHA)*. <https://www.english-corpora.org/coha/>, 2010-.
- Farlex Inc. *The Free Dictionary*. <https://www.thefreedictionary.com/> (last visited: 20 September 2023)
- Grammarist*. <https://grammarist.com/> (last visited: 19 November 2023)
- Gulland, Daphne M., and David Hinds-Howell. *The Penguin Dictionary of English Idioms*, Second Edition. London: Penguin Books, 1994.
- Longman Dictionary of Contemporary English Online*. <https://www.ldoceonline.com/> (last visited: 4 December 2023)
- Merriam-Webster Dictionary*. Pearson Education Limited. <https://www.merriam-webster.com/> (last visited: 20 September 2023)
- National Institute for Japanese Language and Linguistics (NINJAL). *Nihongo Rekishi Kōpasu (Corpus of Historical Japanese, CHJ) Version 2024.3, Chūnagon Version 2.7.2*. <https://clrd.ninjal.ac.jp/chj/> (last visited: 16 December 2024)
- Oxford English Dictionary*. Oxford University Press. <https://www.oed.com/> (last visited: 19 November 2023)
- Oxford Advanced Learner's Dictionary*. Oxford University Press. <https://www.oxfordlearnersdictionaries.com/>

oxfordlearnersdictionaries.com/ (last visited: 4 December 2023)

*The Idioms: The Largest Idioms Dictionary*. <https://www.theidioms.com/> (author not indicated) (last visited: 25 May 2021)



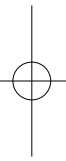


# Phonetic Transcriptions in Free Online English-Japanese Dictionaries: A Comparative Analysis

HIROKO SAITO  
JUNKO SUGIMOTO  
MARINA ARASHIRO

YOKO UCHIDA  
MANAMI HIRAYAMA

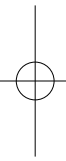
## 1. Introduction



Over the past few years, the authors of this paper have occasionally encountered “strange” phonetic transcriptions in homework answers submitted by students. When asked about the source, the students often replied that they had copied them from an online dictionary application. This issue came to our attention in a particularly striking way when we noticed the phonetic transcription /pɪ'gmɛɪljʌn/ (sic) for *Pygmalion* printed in the program for a play staged by a group of English-major university students. This prompted us to conduct a thorough investigation.

As Koyama and Nabei (2024)—along with many others involved in teaching English to college students in Japan—have observed, “the popular ‘dictionaries’ among recent L2 learners [...] have shifted from paper dictionaries to pocket electronic dictionaries. In addition, smartphone apps are becoming popular among young Japanese learners of English in the early 21st century” (p. 73).

Japanese learners of English have long been enthusiastic users of English-to-Japanese or English monolingual dictionaries. As a result, many high-quality learner’s dictionaries have been compiled in Japan since the country’s opening to the West. However, teachers today often lament that students no longer consult dictionaries when encountering unfamiliar words or phrases. This claim, however, is not entirely accu-



rate. Learners may not flip through the pages of printed dictionaries, but they frequently “flick” on smartphones or computer keyboards to search for unknown words. These mobile devices either carry built-in dictionary apps or allow instant access to free online dictionary sites via search engines, providing headwords, pronunciations, and meanings much like traditional printed dictionaries.

But is this truly the case? Do these two types of dictionaries—traditional and online—really provide the same kind and quality of information? And how did our students come to print /pɪ'gmeɪljən/ under the title of the play they were performing?

This is the question we aim to address by analyzing the phonetic symbols used in three popular online English-Japanese dictionaries in Japan. We will not examine offline pocket electronic dictionaries based on printed editions, whose contents are directly sourced from established publishers. Nor will we include subscription-based online dictionary sites provided by publishing houses, because they are rarely used by students.

In the following sections, we will first look at the differences between traditional printed dictionaries and digitized online dictionaries, followed by the methodology of our research and the results of our analyses. Specifically, we will closely examine how vowels and consonants are represented in the three most popular free online dictionaries in Japan: *Weblio*, *Eijiro*, and *goo*<sup>1)</sup>.

Both *Weblio* and *goo* are based on printed and pocket-size electronic dictionaries and so provide recorded pronunciations on their online platforms. However, since the audio pronunciation models are identical to those in the original electronic dictionaries, we will exclude them from our analysis. Instead, this study will focus solely on the phonetic symbols provided for the headwords in online dictionaries.

## 2. Printed vs. Digitized Dictionaries

Among the many differences—both positive and negative—between printed dictionaries and digitized dictionaries (including pocket electronic dictionaries and online versions), perhaps the most significant is

the issue of space. Printed dictionaries are constrained by the limited amount of space available for each entry. This restricts the number of headwords, definitions, grammatical notes, idioms, example sentences, and pronunciation variants that can be included per headword. In contrast, digitized and online dictionaries have a significant advantage in terms of space, allowing for far greater flexibility in the information they can provide.

Printed dictionaries typically offer phonetic transcriptions (often in IPA or a simplified phonetic format) for headwords. The system of transcribing pronunciation varies among dictionaries. For example, some use the small capital “i” for KIT words to highlight the quality difference from the FLEECE vowel, while others opt for a minimal set of symbols, distinguishing the two word sets with /i/ and /i/. This choice is up to the dictionary’s pronunciation editor, whose name appears on the title page of the dictionary. Once a system is selected, all reputable dictionaries adhere to the fundamental principle of assigning one symbol per vowel or consonant phoneme. If /i/ is chosen for “sit,” then all KIT vowels should be transcribed with this symbol—consistency within the dictionary is essential.

Due to space limitations, printed dictionaries must strike a balance between providing as much information as possible and conserving space. To achieve this, they employ complex abbreviation systems, such as using a bar “|” or incorporating italicization to indicate regional pronunciation variants (e.g., /æsk | ə:sk/, /kɑ:r/) and parentheses to denote optional elisions (e.g., /n(j)u:/). In order to understand these transcription rules and correctly pronounce the words, users must rely on pronunciation guides and symbol lists, which are usually located in the front and/or back matter, thus easy to find.

Digitized and online dictionaries, on the other hand, can display such information in full, without relying on abbreviation marks or complex systems. This allows for a clearer and more user-friendly presentation of pronunciation and other details. Needless to say, however, simply adding information from various sources in an unstructured manner is not sufficient. It is essential to integrate this information into a consis-

tent and well-organized system to ensure accuracy and usability.

One key feature of the online dictionaries we will analyze is that they are free of charge—an especially appealing aspect for our students. With a quick tap on their smartphone screens, they can access the meaning and pronunciation of an unfamiliar word almost instantly, at no cost. However, as with other freely available information on the internet, accuracy and credibility can be concerns. This is one important focus of our research.

(Sections 1 and 2 by Saito)

### 3. Methodology

In this study, transcription of pronunciation of three major online English-Japanese dictionaries is examined: *Weblio* (weblio 英和辞典, [1]), *Eijiro* (英辞郎 on the Web, [2]), and *goo* (goo 辞書, [3]). *Weblio* is an English-Japanese dictionary section on an integrated online encyclopedia site operated by the Tokyo-based GRAS Group Corporation [4, 5]. *Eijiro* is an English-Japanese/Japanese-English search service provided by ALC Inc., whose bilingual English-Japanese and Japanese-English data have been compiled by a group of translators and interpreters called Electronic Dictionary Project (EDP) [6, 7]. *goo* is an English-Japanese dictionary available in the search service provided by goo, a portal site operated by NTT DOCOMO, INC., which provides 13 dictionaries including a Japanese-English dictionary and a thesaurus [8, 9].

Among the three dictionaries, two have a printed dictionary as their main source of information for phonetic transcriptions. The pronunciation notation system for *Weblio* is based on *Shin Eiwa Chu Jiten*, 6th ed. (henceforth *SECJ6*) [10], and that for *goo* is based on *Progressive Eiwa Chu Jiten*, 5th ed. (henceforth *Progressive5*) [11]. Meanwhile, *Eijiro* is based on its own database compiled by the EDP [6, 12]. In all three dictionaries, the actual pronunciation of each entry can be heard by clicking the speaker icon next to it. All three are English-Japanese dictionaries that students can quickly and conveniently refer to when looking up a word online. Accordingly, although the subscription of free membership and/or paid service is available for *Weblio* and *Eijiro*, the target of

the present study is the phonetic transcriptions of free services for non-members.

Words in the Academic Word List (AWL; Coxhead 2000) were examined to shed light on issues related to pronunciation transcriptions in online dictionaries. The AWL is a list of words frequently used in academic publications but excludes those in the General Service List, which contains more basic words, including function words. Thus, most of the words in the AWL are polysyllabic and some are in inflected forms (e.g., *established*, *corresponding*, *outcomes*). The AWL contains 570 words, grouped into 10 sublists based on frequency of occurrence. All words in the AWL were first analyzed to gain insight into what and where the problems were. Then 300 words, that is, approximately half of the AWL (sublists 1, 3, 5, 7, and 9), were further analyzed in detail (henceforth AWL300). The results of this study were based mainly on an examination of these 300 words in three online dictionaries; additional headwords were included when necessary. Transcriptions of the words in the inflected forms were retrieved whenever possible. However, the availability differed from dictionary to dictionary (see §8). When inflected forms were absent, the pronunciation of the base forms was obtained.

The results of the analysis are presented in the following order: overview of the three online dictionaries (§4), transcription of stressed vowels (§5) and weak vowels (§6), description of American-British contrast (§7), and inflected forms (§8). Because the data collection and analysis were conducted from January 2023 to November 2024, any changes that might have been made since then on the side of the online dictionaries investigated are out of the scope of this paper.

#### 4. Overview

This section describes how the three online dictionaries (*Weblio*, *Eijiro*, and *goo*) provide information and the principles of phonetic transcription, and how the pronunciation of headwords is given in each dictionary. In addition, a set of phonetic symbols in these dictionaries will be introduced, along with some identified issues.

#### 4.1 Explanatory notes

In printed dictionaries, explanatory notes on pronunciation are usually found in the front and/or back matter. The notes introduce how pronunciations are presented in the dictionary and explain the usage of symbols and markers (e.g., lexical stress and American and British pronunciation differences). A list of phonetic symbols together with sample words is typically included to illustrate the phonetic value of each symbol.

Similar to printed dictionaries, *Weblio*, *Eijiro*, and *goo* also provide explanatory notes on how pronunciations are presented. However, unlike printed dictionaries, online dictionaries have no established customs. Thus, the three dictionaries differed in terms of where and how the notes were presented, making it difficult for users to access important information. Moreover, they differed significantly in terms of explanatory note content.

As *goo* is based on *Progressive5* (§3), it reproduced notes from the corresponding dictionary on its website (explanatory notes of *goo* [11]). The following five points were noted: (i) basic rules of phonetic transcription (e.g., the use of IPA symbols between slant brackets and the separation of American and British pronunciations by a vertical line), (ii) notation of primary and secondary word stress, (iii) words with different pronunciations depending on parts of speech and definitions, (iv) possible elision using italics, and (v) rules of stress transcription of compounds. Despite the explanation in the notes, slant brackets were missing in the actual phonetic transcriptions as in the examples below.

- (1) *goo*  
**research** 発音 risá:rtʃ, rí:sə:rtʃ  
**authority** 発音 əθó:rəti, əθár- | ɔ:θór-

We were unable to find a list of phonetic symbols with sample words on *goo*. Although *goo* uses symbols commonly employed in English-Japanese dictionaries, it would still be helpful for users to access both explanatory notes and a list of symbols.

Information about phonetic transcription in *Eijiro* is found in the page titled “explanatory notes” (explanatory notes of *Eijiro* [12]). Of a

list of 18 items in the notes, three are related to phonetic transcription. One is a general rule explaining that pronunciation is given next to the icon 発音 [= pronunciation] using special symbols, and when there are two or more variants, they are separated by a slash. A partial list of symbols is introduced under the heading “major phonetic symbols.” The other two observed features are unique to *Eijiro*. Upon the request of users who cannot understand IPA symbols, *Eijiro* provides *katakana* transcription as an aid with the icon カナ [= *kana*]. Also, for words the developers deemed necessary (e.g., ***ache***), the icon 発音 is highlighted in yellow to attract special attention to the pronunciation.

- (2) *Eijiro*  
***ache*** 発音\* éik, カナ エイク (\* The icon highlighted in yellow)  
***procedure*** 発音 prəsi:dʒər, カナ プロスイージャ(ア)

*Weblio* provides both a list of phonetic symbols and explanatory notes on pronunciation (explanatory notes of *Weblio* [10]). As explained in the following sections, substantial discrepancies between the explanatory notes and actual representations were identified in many entries. Possibly, these problems arose because *Weblio*'s explanatory notes are mere reproductions of the front and back matter in *SEJ6* and have not been formulated or revised for *Weblio*. Therefore, deviations were found in both technical aspects (e.g., phonetic symbols) and fundamental areas. For example, *Weblio* writes in the explanatory notes that it shows pronunciations right after its headword in slant brackets and separates American and British English differences by using a vertical line, /Am | Br/. However, *Weblio* actually shows the pronunciation after the definitions of a headword, by using the icon 発音記号・読み方 [= phonetic symbol, how to pronounce] and slant brackets. American-British differences are shown with the labels “(米国英語)” and “(英国英語)” [American English and British English, respectively], and are separated by a comma, rather than a vertical line. It also sporadically provides *katakana* transliteration marked with an icon フリガナ.

- (3) *Weblio*  
***lecture*** 発音記号・読み方 /léktʃə (米国英語), léktʃə (英国英語)/

**facilitate** 発音記号・読み方  
 /fə'sɪlə'teɪt (米国英語), fə'sɪlə'teɪt (英国英語)/  
フリガナ ファシリテイトウ

## 4.2 Basic rules of phonetic transcription

In all three online dictionaries, each entry had at least one pronunciation. However, the availability differed for inflected forms and derived words. The number of variants provided varied among the three dictionaries. An exploratory observation revealed that of the three, *goo* and *Weblio* are placed at two extremes, with *Eijiro* between them. Since *goo* follows the general description rules of a printed dictionary (*Progressive5*), its system is the most economical and consistent. On the other hand, *Weblio* provides full transcriptions for every entry, including forms with inflectional and derivational suffixes, and shows variations in pronunciation by constructing original transcriptions for entries that *SECJ6* does not fully show. This resulted in the copious inconsistencies observed in *Weblio*.

As for headwords with more than one pronunciation, both *goo* and *Weblio* explain in their notes that they removed the common part and employed a hyphen to indicate the omitted part. This is how *Progressive5* and *SECJ6* (*goo* and *Weblio*'s source dictionaries) describe words with multiple pronunciations. While *goo* follows this rule, *Weblio* generally provides a full transcription. As the redundant part is omitted with a hyphen in *SECJ6*, *Weblio* appears to generate its original transcriptions to provide the full form. While full-form representations have the merit of clarity from a user perspective, this part of the phonetic representations was found to be significantly inconsistent and erroneous (see §5, 6, 7).

## 4.3 Phonetic symbols

This section first introduces the set of phonetic symbols in the three dictionaries, and then discusses the anomalies and irregularities found.

All three dictionaries show pronunciation using IPA symbols. Table 4.1 summarizes the phonetic symbols as listed in the explanatory notes of

Table 4.1 Phonetic symbols in the explanatory notes of the three online dictionaries

## (i) Vowels

Keyword	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>	Keyword	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>
KIT	ɪ	i	i	PRICE	aɪ	ai	ai
DRESS	e	e	e	CHOICE	ɔɪ	ɔi	ɔi
TRAP	æ	æ	æ	MOUTH	aʊ	–	–
LOT	ɑ   ɔ	ɑ	ɑ   ɔ	NEAR	ɪə   ɪə	iər	iər
STRUT	ʌ	ʌ	ʌ	SQUARE	eə   eə	eər	eər
FOOT	ʊ	u	u	START	aɜ   a:	a:r	a:r
NURSE	ɝ:   ə:	ɝr	ɝ:r	NORTH	ɔɜ   ɔ:	ɔ:r	ɔ:r
FLEECE	i:	i:	i:	CURE	ʊə   ʊə	uər	uər
FACE	eɪ	eɪ	eɪ	HAPPY	i	–	–
PALM	ɑ:	ɑ:	ɑ:	MANUAL	ju	–	–
THOUGHT	ɔ:	ɔ:	ɔ:	LETTER	ɝ   ə	ər	ər
GOAT	ou   əʊ	ou	ou	HABIT	ɪ	–	–
GOOSE	u:	u:	u:	BONUS	ə	ə	ə

## (ii) Consonants

	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>
Stops	p, t, k; b, d, g (t for t-tapping)	–	p, t, k; b, d, g
Affricates	tʃ, dʒ	tʃ, dʒ	tʃ, dʒ
Fricatives	f, θ, s, ʃ, h; v, ð, z, ʒ	θ, ʃ; ð, ʒ	f, θ, s, ʃ, h; v, ð, z, ʒ
Nasals	m, n, ŋ	ŋ	m, n, ŋ
Laterals	l	–	l
Approximants	r, j, w	–	r, j, w

the three online dictionaries. Vowels are categorized using Wells's (1982) Standard Lexical Sets (with additional weak vowel categories), whereas consonant symbols are listed based on their manner of articulation. As already explained, *goo* does not provide a list of symbols; and thus, the symbols used in its source dictionary *Progressive5* are given.

Note that *Eijiro* does not provide an exhaustive list of symbols; it does not include weak vowel symbols, and it only provides American pronunciations for categories such as LOT and rhotic vowels. As for consonants, *Eijiro* only lists symbols that are not part of the Roman alphabet: /ʃ tʃ ʒ ɟ θ ð ŋ/. In addition, the MOUTH vowel is missing from the lists of both

*Eijiro* and *goo* (*Progressive5*).

### 4.3.1 Vowel symbols

The choice of vowel symbols varies considerably across dictionaries (see Table 4.1). Among the three dictionaries under discussion, *Eijiro* and *goo* choose /i/ and /u/ as short vowels at the expense of qualitative accuracy. For example, KIT and FLEECE are expressed as /i/ and /i:/, respectively. Meanwhile, *Weblio* adopts symbols such as /ɪ/ and /ʊ/, and therefore, the qualitative difference between KIT and FLEECE is expressed with different symbols /ɪ/ and /i:/. Although either representation is permissible, consistency within a dictionary is crucial (see also §2).

The fluctuation is most significant in the vowel categories containing /ɑ/. *Weblio* states that it employs the script /ɑ/ for American pronunciation of LOT, PRICE, and MOUTH. However, /ɑ/ (alpha symbol) and /a/ are also used interchangeably (e.g., *prime* /práim/ and *styles* /stáilz/). Similarly, while *Weblio* designates /ə/ for the weak vowel BONUS, /ʌ/ is frequently employed particularly in British pronunciation (e.g., *reaction* /ri:ækʃʌn/). For detailed discussions of phonological representations and the inconsistency of transcription, see Sections 5 (strong vowels) and 6 (weak vowels).

Lastly, in IPA notation, a length mark (: ) is used to mark long vowels. While both *goo* and *Eijiro* use (: ) consistently, *Weblio* uses not just an IPA length mark but also a colon (:) to indicate long vowels. Furthermore, a length mark is sometimes missing in a long vowel but used for weak vowels although it is unnecessary (e.g., American pronunciation of *removed* /ri'muvd/, British pronunciation of *capacity* /kʌ'pæsati:/).

Inconsistencies such as an alpha /ɑ/ in place of a script /a/ and a colon in place of a length mark are considered technical problems and will not be discussed in further sections.

### 4.3.2 Consonant symbols

Symbols for consonants are generally more straightforward than those for vowels. Furthermore, the three dictionaries employ the same set of consonant symbols as listed in Table 4.1. Along with regular symbols,

t-tapping is generally (though not systematically) shown with a dot (/t̚/) in *Weblio* (e.g., compare American pronunciations for **capacity** /kəp'æsəti/ and **generated** /'dʒɛnɜːɪtəd/) (see §7.2).

Some consonant symbols such as /θ/ are occasionally distorted in *Weblio*, shown in italicized two-byte character or in a different font: /θ/. However, the appearance seems to differ depending on the computer, web browser, or a combination of the two.

In *Weblio*, inconsistencies are also observed in the transcription of the r-sound: in addition to the symbol /r/ that is usually used in English phonemic transcription, the inverted /ɹ/ is found sporadically in both American and British pronunciations (e.g., **criteria** /kɹi'ɹiə(i).i.i.ə (米国英語), kraɹ'tɹi:ɹ (英国英語)/ and **paradigm** /p'ærədɔɹm (米国英語), 'pæɹ.ɹ.dəɹm (英国英語)/). As in this example, *Weblio* also sporadically uses a dot to indicate a syllable boundary.

#### 4.3.3 Possible elision of segments

Phonetic transcription occasionally involves segments that can be omitted. *Weblio* explains that such sounds are shown in parentheses, whereas *Eijiro* and *goo* show them in italics. However, we identified discrepancies between the explanations provided in the explanatory notes and the actual transcription. Discrepancies in the use of italics are most apparent in *goo*. In fact, none of the symbols expected to be in italics following the explanatory notes (and source dictionary *Progressive5*) are presented as such. For example, **source** is transcribed as /sɔːrs/ with non-italicized “r.” The lack of italicization may be a technical issue, but it results in a serious transcription problem such as the failure to distinguish between American and British pronunciations (see §7).

#### 4.3.4 Stress symbols

All three dictionaries state that acute accent marks are used for word stress: /ˈ/ for primary stress, and /ˌ/ for secondary stress. For *Eijiro*, stress marks are provided only for words with more than one syllable.

One problem noted is that *Weblio* also employs the IPA notations for some words: /ˈ/ for primary stress and /ˌ/ for secondary stress. In addi-

tion, both acute accent marks and vertical strokes are used incorrectly (see §5.1). According to IPA notation, a vertical stroke should be placed in front of a stressed syllable, but in *Weblio*, vertical strokes do not correspond to syllable boundaries.

- (4) *Weblio* (American pronunciations only)
- |                   |  |
|-------------------|--|
| Vertical stroke   | <b><i>factors</i></b> /'fæktəz/ (correct)                                  |
|                   | <b><i>draft</i></b> /dr'æft/; <b><i>implies</i></b> /ɪ'mplaɪz/ (incorrect) |
| Acute accent mark | <b><i>margin</i></b> /máɹədʒɪn/ (correct)                                  |
|                   | <b><i>academic</i></b> /ækədémɪk/ (incorrect)                              |

Sometimes, a vertical stroke is used for American pronunciation but an acute accent mark for British pronunciation, or vice versa (e.g., ***legal*** /lí:g(ə)l (米国英語), 'li:gəl (英国英語)/). Even more chaotic cases with both symbols used within a word were spotted (e.g., American pronunciation of ***external*** /'ɛkst'ɜ:nl/).

#### 4.4 Summary

As with the printed dictionaries, all three online dictionaries provided explanatory notes. Unfortunately, these are presented in ways that are difficult for users to access. Regarding phonetic symbols, there is no problem in having different dictionaries using different symbols as long as consistency is guaranteed and users understand the system and rules (§2). However, inconsistent and irregular use of symbols were found in the transcriptions of online dictionaries, especially *Weblio*. Some problems may be technical, but many seem to stem from the lack of a system, which is a problem peculiar to online dictionaries.

(Sections 3 and 4 by Arashiro)

### 5. Strong vowels

This section discusses the descriptions of strong vowels in the three online dictionaries under discussion (i.e., *Weblio*, *Eijiro*, and *goo*). Some methodological notes are in order first. The data in this section include vowels in syllables with primary stress in the database AWL300 (§3). When there are not enough words to examine for a vowel category, more words were added from outside of the database so that there were at

least ten words per vowel category. For example, the list had no words whose primarily stressed vowel is that in *boy*, so words such as *boy*, *noise* and *employ* were added to the dataset for the discussion in this section. When the stress position differed depending on the parts of speech of the word (e.g., *conflict*), only one category was considered. For example, for *conflict*, the pronunciation of the noun is considered, and the vowel in the initial syllable is analyzed. For inflected words, when the dictionary did not have the exact form in the headword, the bare form was used. Table 5.1 lists the data used in this section. Note that the rhotic vowels discussed in Section 7 are not included in the discussion in this section. Lexical variation here refers to variation stemming from the nature of the words and not variation between American and British varieties (see Section 7.5 for the discussion on the dialectal variation).

Table 5.1 Vowels and the number of words discussed in Section 5

Vowel	n	Vowel	n
FLEECE	14	GOOSE	12
KIT	42	FOOT	10
DRESS	45	STRUT	10
FACE	35	CHOICE	10
TRAP	21	PALM	10
PRICE	19	CLOTH	10
MOUTH	10	LOT	30
GOAT	13	THOUGHT	10
(Lexical variation)	8)		

Since dictionaries may use different symbols for the same vowel category, to compare the same vowel categories across dictionaries, the categories were indicated using the lexical sets proposed by Wells (1982). For example, the vowel as in *boy* is referred to as the CHOICE vowel in this section.

This section focuses on the following topics: stress marking (§5.1), vowel symbols in general (§5.2), some aspects of variation between American and British English (§5.3, §5.4, §5.5) and lexical variation (§5.6). Section 5.7 is a brief summary and a discussion of transcriptions spo-

radically found.

### 5.1 Stress marking

There are two issues in stress marking. One is the variation in the symbols used and the other is the presence or absence of stress marking. With respect to the variation, as discussed in the previous section (§4), dictionaries may use different symbols to mark stressed syllables or vowels. In the three dictionaries under discussion, *Eijiro* and *goo* consistently use an acute accent mark above the vowel symbol (e.g., /mæksəməm/ for **maximum**), while *Weblio* uses an acute accent mark for some words (e.g., /méθəd/ for **method**) and a vertical stroke for others (e.g., /m'æksəməm/ for **maximum**).

The variations in *Weblio* can fall into one of the following two cases. In the first case, the variation depends on whether the character is easily available with the acute accented mark. For example, the acute accent mark is used in /é/ because this accented character is readily available, whereas this is not the case with, for example, /æ/ and /ɔ/, which are rather special as a character, in which case the stress is indicated to the left of the vowel symbol (e.g., /m'æksəməm/; see §4.3.4 for this problem). The second case involves the availability of a transcription in the *SECF6* dictionary. Even characters for which the acute accented form is readily available (e.g., /í/, /ú/) sometimes have the vertical stroke in some words (e.g., /ri'muvd/ for **removed**) while the acute accent is used in other words (e.g., /əsú:m/ for **assume**). In these cases, the transcription with the vertical stroke usually involve inflected forms. Why is this? It appears to be the case that for many words investigated in this paper, *Weblio* uses the transcription available in *SECF6* for the American variety (and other kinds of transcriptions for the British variety). However, *SECF6* does not usually list inflected words as headwords. It seems, therefore, that where no pre-existing transcription is available, *Weblio* does not use the acute accent, the *SECF6* system, and instead uses the vertical stroke for the inflected forms.

Another issue regarding stress marking is the absence or presence of the marking in the first place. *Eijiro* sometimes lacks markings, but this

is usually not the case in the other two dictionaries. It appears that many of these cases involve monosyllabic words, in which case it makes sense not to indicate the stress position because there is only one syllable that can be stressed; thus, the stress marking is redundant. However, the transcription is not consistent; not all monosyllabic words are treated alike, and many of them are marked for stress.

## 5.2 Vowel symbols

All three dictionaries provide transcriptions for most words, although *goo* occasionally lacks transcriptions of certain words (e.g., *overlap*). The remainder of this subsection discusses the vowel symbols used in the three dictionaries.

Table 5.2 lists the symbols used in the dictionaries. This table provides the symbols used for the great majority of words in each vowel

Table 5.2 Vowel symbols at a glance

	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>
TRAP	'æ   'æ	æ	æ
PRICE	'ai, ái   'ai, ái	ái	ái
MOUTH	'au, áu   *	áu	áu
CLOTH (before r)	(see text in Section 5.4)		
CLOTH (elsewhere)	'ɔ:   'ɔ	ɔ, a   ɔ	ɔ́:   ɔ́
LOT	á, ('a)   'ɔ, ('a:)	á   ɔ́	á   ɔ́
THOUGHT	'ɔ:   ɔ: **	ɔ   ɔ:	ɔ́:
PALM	á:   a: **	a   a:	á:
DRESS	'ɛ, é   é, 'e, ('ɛ)	é	é
FACE	éi, ('e)   éi, 'ei	éi	éi
FLEECE	í:, ('i)   'i:, (i:) **	í:	í:
KIT	í, ('i)   'i, (i)	í	í
GOAT	óu   'əu	óu (UK áu)	óu
GOOSE	ú:, ('u)   ú:, 'u:	ú:	ú:
STRUT	'ʌ   'ʌ	ʌ	ʌ
CHOICE	'ɔi   ***	ɔi (ói)	ói
FOOT	'ʊ   ***	ú	ú

\* No transcription is provided. See text.

\*\* The length mark varies between a colon (:), and IPA mark (ː).

\*\*\* Transcription is provided only for a couple of words, with one word with stress mark ['ɔi] ['u] and the other without it [ɔi] [u].

category, but does not include those that are sporadically used (see Section 5.7 for those miscellaneous transcriptions). Thus, if there is only one symbol in a cell, it does not always mean that it is the only symbol used for the vowel category, and there may be a sporadic deviation in certain words that is judged as a miscellaneous transcription. In addition, the symbols in parentheses appear repeatedly but not most frequently in each category. Dictionaries may indicate American and British pronunciations separately (see Section 7 for more on this). When they are transcribed separately, the symbols are separated with a vertical bar (|) between them in Table 5.2: the American variety is given to the left of the bar while the British variety is given to the right of it. If there is more than one symbol within a variety, the symbols are separated by a comma (,). As shown in the Table, overall, *goo* provides the most principled and consistent transcription, *Eijiro* occasionally shows inconsistency (e.g., in stress marking), and *Weblio* shows the most fluctuating transcription.

An important issue that arises from Table 5.2 is the systematicity within a particular dictionary regarding the transcription of phonemic contrasts. Assuming that the transcriptions provided in these dictionaries are basically phonemic rather than phonetic (just as in the hard copy dictionaries (§2)), the same phoneme/vowel category would be transcribed using the same symbol or symbol combination. In other words, different symbols are expected to represent different phonemes/vowel categories.

This principle is easily found intact in *Eijiro* and *goo*, but not necessarily in *Weblio*. In *Eijiro* and *goo*, different symbols are used to represent different categories. In other words, all the vowels in the same lexical set/category are basically transcribed with one and the same symbol (although they sporadically show inconsistency). For example, all FLEECE vowels are transcribed as /i:/ and all KIT vowels as /i/: they are distinctly transcribed as expected from the principle of the way the symbols are used. (The overlapping use of symbols in some categories is due to dialectal variation as discussed in the next section.) However, in *Weblio*, there are cases where different symbols are used within the

same vowel category (e.g., PRICE, DRESS, LOT and MOUTH vowels). This goes against the principle of “one symbol for one category.” For example, the DRESS vowel is transcribed /e/ for some words but /ɛ/ for others, which would wrongly give the impression to the dictionary users that these two vowels are contrastive. Similarly, the same symbol is sometimes used across categories. For example, the FLEECE vowel is transcribed using the KIT vowel symbol for the second syllable in *techniques* /te'kniks/, and the FACE vowel is transcribed as if it were a DRESS vowel for the second syllable in *attained* /ʌ'tend/. If we take a closer look, this transcription, which appears to go against the principle, may follow the principle of the use of stress marking (i.e., an acute accent mark or vertical stroke) considered together with the vowel symbol, since the problematic cases seem to be commonly found in inflected forms for American pronunciation, where the vertical stroke is used instead of an acute accent mark to mark stress (as discussed in Section 5.1). However, it would be difficult for casual dictionary users to use this rather unusual rule (that combines the vowel symbol with the stress marking system) to understand the category to which the vowel belongs, unless it is clearly indicated along with the transcription.

### 5.3 American/British transcription 1

This and the next few subsections discuss aspects of the variation between American and British English (see Sections 7 and 8 also for other topics). The first issue is the fact that *Weblio* often specifies the transcription with the label “米国英語” ‘American pronunciation’ when the transcription could also apply to British English (labelled as 英国英語 elsewhere in *Weblio*). The lack of transcriptions for the British variety is particularly evident in the following categories: TRAP, KIT, PRICE, STRUT, MOUTH, CHOICE, and FOOT. These are categories that do not systemically vary between the American and British systems, but it is not certain whether this can account for the lack of British transcriptions for these categories, because entries for the GOOSE vowel, a similar category in that respect, consistently provide both the American and British transcriptions.

#### 5.4 American/British transcription 2: PALM, LOT, CLOTH, and THOUGHT vowels

One of the largest differences in the vowel systems between General American (GA) and Received Pronunciation (RP) is the distribution of low back vowels. The relevant lexical sets include PALM, LOT, CLOTH, and THOUGHT. Wells (1982) describes the distribution as summarized in Table 5.3: GA has two vowels whereas RP has three.

Table 5.3 Low back vowel contrasts in GA and RP in Wells (1982)

	PALM	LOT	CLOTH	THOUGHT
GA	ɑ		ɔ	
RP	ɑ:	ɒ	ɔ:	

The comparison can also be summarized as in (1).

- (1) a. GA: PALM = LOT vs. CLOTH = THOUGHT  
 b. RP: PALM vs. LOT = CLOTH vs. THOUGHT

The three dictionaries under discussion provide transcriptions for these four lexical sets as listed in Table 5.4. The CLOTH category includes only non-prevocalic vowels; we will treat the prevocalic vowels below. The three dictionaries exhibit slightly different patterns.

Table 5.4 Transcriptions for low back vowels in *Weblio*, *Eijiro*, and *goo*

	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>
PALM	ɑ̃:   ɑ:	ɑ   ɑ:	ɑ̃:
LOT	ɑ̃, (ɑ̃)   ɔ, (ɑ̃)	ɑ̃   ɔ̃	ɑ̃   ɔ̃
CLOTH	ɔ̃:   ɔ̃	ɔ̃, ɑ̃   ɔ̃	ɔ̃:   ɔ̃
THOUGHT	ɔ̃:   ɔ:	ɔ̃   ɔ:	ɔ̃:

First, *goo* describes the same system as Wells (1982) for UK, while it gives a slightly different system from it for US, with a three-way contrast instead of a two-way contrast, as summarized in (2). PALM and LOT vowels are transcribed differently for US, with the former with the length mark and the latter without it. This is a system found in the literature (e.g., Takebayashi and Saito 2008), with the merger often being noted between the two sets (i.e., PALM and LOT) (e.g., *ibid.*, p. 32).

- (2) *goo*
- a. US: PALM (ɑː) vs. LOT (ɑ) vs. CLOTH = THOUGHT (ɔː)
  - b. UK: PALM (ɑː) vs. LOT = CLOTH (ɔ) vs. THOUGHT (ɔː)

Second, in *Eijiro*, the PALM vowel in the US variety is transcribed mostly as /ɑ/ but only **father** and **aha** are transcribed as /ɑː/. The US pronunciation for **pa** shows variation between /ɑ/ and /ɔ/. If we ignore these cases, *Eijiro* provides the same system as Wells (1982), as in (3), but the transcription indicates a variation in the CLOTH vowel in the US, reflecting an ongoing sound change where CLOTH and LOT vowels are merged for many speakers.

- (3) *Eijiro*
- a. US: PALM = LOT (ɑ) vs. THOUGHT (ɔ), CLOTH (ɔ~ɑ)
  - b. UK: PALM (ɑː) vs. LOT = CLOTH (ɔ) vs. THOUGHT (ɔː)

Third, *Weblio*'s American system is the same as *goo*. For the British transcription, *Weblio* basically gives the same system as Wells (1982) (and the other two dictionaries) with the exception that some of the LOT words are wrongly transcribed with the unrounded /ɑː/. This error of the unrounded /ɑː/ for the LOT vowel is found in (i) inflected forms (e.g., **comments**) and (ii) **philosophy, technology, psychology, ideology, symbolic**, and **economic** (without the length mark for **economic**).

Turning to CLOTH vowels before /r/, the AWL300 list contains only three words: **minorities, authority, and priority**. All three dictionaries show inconsistencies in these words. *Eijiro* and *goo* transcribe the variation between the US and UK pronunciations, but how they vary is different. Compare this with *Longman Pronunciation Dictionary* (Wells 2008), which provides a consistent variation for all of them. *Weblio*'s transcription is confusing: the US/UK transcriptions may be reversed for **minorities**; unrounded /ɑ/ is wrongly used for the British pronunciation for **authority**; and the rounded /ɒ/, used for the British pronunciation for **priority**, is not indicated at all in the British pronunciation of other words except in the initial syllable of **objective**.

### 5.5 American/British transcription 3: GOAT vowel

The GOAT vowel is one for which some dictionaries (e.g., *SECF6*) indicate phonetic differences between American and British English. *Eijiro* and *goo* basically do not mark this variation (although *Eijiro* marks it sporadically, i.e., for **global** and **protocol** in the dataset). However, *Weblio* mostly indicates the variation with /óʊ/ and /'əʊ/ for the American and British transcriptions, respectively.

### 5.6 Lexical variation

There are words having more than one pronunciation. There are two types of such lexical variation. The first involves variation in the position of stressed syllables. These include **decades**, **conversely**, **research**, and **adults** in the dataset. The other type of variation involves words in which the stressed syllables do not vary but there is more than one way of pronouncing the stressed vowel. These are **process**, **data**, **theory**, and **route** in the dataset. Table 5.5 shows whether the variation is indicated in a particular dictionary. As shown in the Table, only *goo* indicates the variation for all words. The other two dictionaries do so for approximately half of the words. In other words, if the dictionaries are compared regarding whether they indicate the variation for particular words, for some words (e.g., **process**) they are consistent, whereas for others (e.g., **route**) they are not.

Table 5.5 Presence and absence of lexical variation

	<i>Eijiro</i>	<i>goo</i>	<i>Weblio</i>
process	√	√	√
data	√	√	√
theory		√	√
route	√	√	
decades	√	√	
conversely		√	√
research	√	√	
adults		√	

### 5.7 Summary

This section discussed the transcription of vowels in syllables with

primary stress in the dataset. The principle of one symbol (combination) for one (vowel) category was most consistently applied in *goo*; there are some inconsistencies in *Eijiro*; and the most confusing transcriptions were identified in *Weblio*.

We will end this section by identifying a few miscellaneous transcriptions (Section 5.2), which none of the three dictionaries is exempt from. For example, even *goo* has several words whose pronunciation is not provided in its entirety. All three dictionaries may have inconsistent transcriptions within the same vowel category: for example, a variation may or may not be indicated only for certain words within a category (e.g., only **boss** has a variation for American variety in CLOTH in *goo*; only **cough** does not have a variation for the American variety in CLOTH in *Eijiro*). *Weblio* may at times have somewhat unexpected transcriptions, including typos (e.g., only the suffix is given for **assessment** in the American transcription, and part of the transcription is omitted in the British transcriptions of **authority** and **temporary**). This kind of miscellaneous transcriptions is usually not reported in reviews of dictionaries published in hard copy. It is unknown why these inconsistent transcriptions are found in online dictionaries under discussion.

(Section 5 by Hirayama)

## 6. Weak vowels

This section describes the transcriptions of weak vowels in *Weblio*, *Eijiro*, and *goo*. Table 6.1 lists the weak vowel symbols used in the three online dictionaries, along with those used in *SECJ6* and *Progressive5*. Following Wells's (1982) Standard Lexical Sets for strong vowels, the five weak vowels are referred to as BONUS, HAPPY, MANUAL, HABIT, and

Table 6.1 Weak vowel symbols

Vowel	<i>SECJ6</i>	<i>Progressive5</i>	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>
H <u>APPY</u>	ɪ	ɪ	ɪ, i:   ɪ, i:	ɪ	ɪ
M <u>ANUAL</u>	ʊ	ʊ	ʊ, ʊ   u:, ʊ, ʌ	ʊ	ʊ
L <u>ETTER</u>	ɚ   ə	əɾ	ɚ, ɜ   ə, ɜ:	əɾ	əɾ
H <u>ABIT</u>	ɪ	ɪ	ɪ, ʌ	ɪ	ɪ
B <u>ONUS</u>	ə	ə	ə, ʌ	ə	ə

LETTER. As in the case of strong vowels, *Weblio* does not follow the principle “one symbol for one category” and uses multiple symbols for one vowel category.

### 6.1 HAPPY vowel

HAPPY vowels occur in prevocalic and word-final positions. Thirty-seven words were extracted from AWL300 for analysis: 24 in word-final and 13 in prevocalic positions. *Weblio* transcribes this vowel in three ways. One is a typical notation using /i/ for both American and British pronunciations as in **theory** /θi:əri | θəri/ and **scenario** /səné(ə)riðu | sənéná:ri`əu/ (13 words). The second group uses /i/ for American but /i:/ for British pronunciation as in **technology** /teknólédzi | te'kna:lédzi:/ and **create** /kriét | kri:'eit/ (23 words). The third group uses /i: (i:)/ for both American and British pronunciations as in **criteria** /krai'tiə(ɪ).i:ə | kraɪ'tiri:ʌ/ (1 word).

For the HAPPY vowel, both *Eijiro* and *goo* invariably use /i/ for American and British pronunciations (Table 6.1).

### 6.2 MANUAL vowel

Since there are only eight words with the MANUAL vowel in AWL300, additional sample words were added from Takebayashi and Saito (2008) so that the total number for analysis would be 10 each for prevocalic and preconsonantal positions. *Weblio* uses three patterns of transcription for prevocalic positions: (i) /u/ for American pronunciation and a British variant not provided (2 words; **individual**, **actual** /'æktʃuəl/), (ii) /u/ for American and /u:/ for British pronunciation (7 words; e.g., **continuous**, **mutual** /mju:tʃuəl | 'mju:tʃu:ʌl/), and (iii) /u/ for American and /ʌ/ for British pronunciation (1 word; **usual** /'ju:zuəl | 'ju:zʌwʌl/). As for preconsonantal positions, two patterns are identified: (i) /u/ for both American and British pronunciations (8 words; e.g., **document**, **volume** /vóljʊm | v'óljʊm/), and (ii) /u/ for American and /ʌ/ for British pronunciation (2 words; **calculate**, **simulation** /simjʊléɪʃən | simjʌ'leɪʃʌn/). The use of /ʌ/ as a weak vowel will be discussed in Section 6.6 below.

The MANUAL vowel is invariably transcribed with the symbol /u/ for

both American and British pronunciations in *Eijiro* and *goo* (Table 6.1).

### 6.3 LETTER vowel

Twenty-three words that include LETTER vowels are found in AWL300. Additionally, three words that include a sequence of V/r/V across weak syllables were examined.

First, LETTER vowels are transcribed in four different ways in *Weblio* (Table 6.2). There appears to be no clear rule, but the third group /ɜ̃ | ɜ:/ is used only for inflected forms: **factors**, **registered**, and **required**. The fourth one /ɚ | ɚ/ may simply be a typo.

Table 6.2 LETTER vowels in *Weblio*

Groups	US	UK	Example	n. of LETTER Vs
1	ɚ	ə	<b>energy</b> /énədʒi   énədʒi/	13
2	ɚ	ɜ:	<b>alter</b> /'ɔ:ltɚ   'ɔ:ltɜ:/	6
3	ɜ̃	ɜ:	<b>fact<u>ors</u></b> /'fæktɜ̃z   'fæktɜ:z/	3
4	ɚ	ɚ	<b>expos<u>ure</u></b> /'ɪkspóʊɜ̃   'ɪkspóʊɜ:/	1

Compared to *Weblio*, the notation of the LETTER vowel in *Eijiro* is fairly consistent using /ə̃r/ (Table 6.1). Using an italic “r” is efficient in indicating the lack of a postvocalic /r/ in a non-rhotic accent: **percent** /pə̃rsént/. When American-British difference is observed in other positions within a word, the full transcription is provided separately: **partnership** /([US] pá̃rtnə̃rʃɪp | ([UK] pá̃rtnə̃ʃɪp), **pursue** /([US] pə̃rsú: | ([UK] pə̃sjú:/. Transcription mistakes are sometimes observed, especially in inflected forms. For example, **registered** /rédʒɪstəd/ only gives the pronunciation without /r/, although its base form **register** /rédʒɪstər/ is transcribed with /r/.

Finally, the transcription of LETTER vowels in *goo* is consistent (Table 6.1). The only problem is that italicization is missing, which leads to failure in distinguishing American and British pronunciations (see also §7). Compare **sector** /séktər/ in *goo* and /séktə̃r/ in *Progressive5*, for example.

For three words with the V/r/V sequence across weak syllables (**considerable**, **liberal**, **monitoring**), a serious notation problem for British pronunciation is found in *Weblio*. As illustrated in the following two

words, an intervocalic /r/ is missing in the British pronunciation: **considerable** /kənsɪdərəbl̩ | kʌˈnsɪdɜːəbəl/, **liberal** /lɪb(ə)rəl | ˈlɪ,bɜːəl/. The same problem is observed in the notation of V/r/V that involves a strong vowel, which also fails in correctly transcribing British pronunciation (see §7.1).

#### 6.4 HABIT vowel

From AWL300, words that include /ɪ/ in weak syllables (i.e., syllables without primary or secondary stress) were extracted. Words in which American and British pronunciations differ in terms of stress and/or vowel quality are excluded. Furthermore, the word **prohibited** is added to the list because the third syllable is expected to have the HABIT vowel. Consequently, 58 target words with 62 instances of HABIT vowels are examined and transcribed in four ways in *Weblio*.

Table 6.3 HABIT vowels in *Weblio*

Group	US	UK	Example	n. of HABIT Vs
1	ɪ	ɪ	<b>coincide</b> /kəʊɪnsáɪd   kʰəʊɪnsáɪd/	48
2	ɪ	i:	<b>relaxed</b> /rɪl'ækst   ri:'lækst/	4
3	ɪ	ʌ	<b>devoted</b> /dɪvóʃtɪd   dɪ'vəʊtəd/	9
4	ʌ	ʌ	<b>prohibited</b> /pro'hɪbətəd   prəʊ'hɪbətəd/	1

Group 1 is a typical transcription of the HABIT vowel with the same /ɪ/ symbol for both American and British pronunciations. In Group 2, the long vowel /i:/ is used in British pronunciation, but only four instances are found. The third group, in which /ɪ/ is used for American but /ʌ/ for British, is unusual transcription. The nine instances in Group 3 are mostly derived forms with an inflectional or derivational suffix: **devoted**, **illustrated**, **isolated**, **hierarchical**, **justification**, **legislation**, **medical**. However, two underived forms (**philosophy**, **indicate** /ɪndɪkət̩ | ɪndʌkət̩/) are also included. Group 4 exhibits an interesting case in which the weak vowel in the third syllable of **prohibited** /pro'hɪbətəd | prəʊ'hɪbətəd/ is transcribed as /ʌ/ for both American and British pronunciations. The same vowel is written as /ɪ/ in the base form **prohibit** but is replaced by /ʌ/ in the *-ed* form.

*Eijiro* and *goo* both show consistency in using /i/ for both American and British pronunciations for the 62 instances of HABIT vowels in the 58 target words. Note that the two dictionaries use the same symbol /i/ for both HABIT and HAPPY vowels, along with the KIT vowel. This means that neither *Eijiro* nor *goo* indicate a qualitative difference of these three vowels, but this is a common transcription in English dictionaries published in Japan (cf. *Progressive5*) (§4.3.1).

### 6.5 BONUS vowel

The BONUS vowel is the most frequent English vowel, and dictionaries that adopt IPA symbols almost invariably transcribe it with the symbol /ə/. However, the transcription of this BONUS vowel in *Weblio* presents a complicated and confusing picture. Because the BONUS vowel is found in many words, the target for analysis is narrowed down to words for which *Weblio* provides both American and British pronunciations. Through an analysis of 148 target words, including 175 BONUS vowels, four different patterns are identified in *Weblio*.

Table 6.4 BONUS vowels in *Weblio*

Group	US	UK	Example	n. of BONUS Vs
1	ə	ə	<b>benefit</b> /bénəfit   bənəfit/	55
2	ə	ɪ	<b>intermediate</b> /ɪntə'mi:diət   ,ɪntɜ:'mi:di:ɪt/	9
3	ə	ʌ	<b>innovation</b> /ɪnə'veɪʃən   ,ɪnʌ'veɪʃʌn/	89
4	ʌ	ʌ	<b>specified</b> /'spesʌ,faid   'spesʌ,faid/	22

Group 1 is a typical transcription of BONUS vowels having the same /ə/ symbol for both American and British pronunciations. Group 2 may be simply reflecting distributional difference between the two varieties, that is, American pronunciation having /ə/ while British having /ɪ/ (Takebayashi 1996). An interesting transcription is spotted in Groups 3 and 4: /ə/ for American but /ʌ/ for British in the former, and /ʌ/ for both in the latter.

All 22 instances in Group 4 are derived forms: **anticipated**, **attained**, **circumstances**, **commenced**, **components**, **confirmed**, **constraints**, **converted**, **excluded**, **generated**, **identified**, **minorities**, **monitoring**, **prohibited**, **specified**, **ultimately**, **unified**, **variables**.

One point these words have in common is that *SECJ6* does not provide full transcription because they are not treated as headwords. Unlike Group 4, words in Group 3 are not restricted to derived forms, and the source of using /ʌ/ for British pronunciation is unclear.

*Eijiro* uses /ə/ for BONUS vowels consistently. Distributional difference between American and British pronunciations is observed using /ə/ for [US] and /i/ for [UK] in 16 words (17 instances): ***authority***, ***priority***, ***dominant***, ***facilitate***, ***policy***, ***specific***, ***integral***, ***comprehensive***, ***preliminary***, ***interpretation***, ***constitutional***, ***qualitative***, ***phenomenon***, ***challenge***, ***military***, ***financial***. The opposite pattern, namely /i/ for [US] and /ə/ for [UK], is found in words such as ***distorted*** / [US] *distórtid* | [UK] *distórtəd*/. It is uncertain whether this reflects a distributional difference or if it is simply inconsistent. For some words, *Eijiro* uses /ə/ for American but italicized /ə/ for British (see the next section for more detail).

Lastly, for all the target words, *goo* uses /ə/ for BONUS vowels following the transcription of *Progressive5*. Again, a problem concerning *goo* is a non-italic /ə/ (§6.3). Words such as ***coherence*** and ***environment***, for example, are transcribed with an italic /ə/ in *Progressive5* (/kouhíərəns/, /inváiərənmənt/) but as /ə/ in *goo* (/kouhíərəns/, /inváiərənmənt/).

## 6.6 The use of /ʌ/ as a weak vowel

A conundrum remains as to the use of /ʌ/ in place of the BONUS, HABIT, and MANUAL vowels in *Weblio*. As for /ə/, Pullum and Ladusaw (1996) explained that “In many American dialects there is no phonological distinction between [ʌ] and [ə], and American linguists often treat these two symbols as if they were interchangeable” (p. 18). This is reflected in the notation of American dictionaries such as *Merriam-Webster Online*, in which the word ***humble***, for example, is transcribed as /'həmbəl/. The stressed /ə/ corresponds to the STRUT vowel, whereas the unstressed /ə/ represents the BONUS vowel. Accordingly, it is theoretically possible to use /ʌ/ for both strong and weak vowels and distinguish the two by the presence/absence of a stress symbol. Since the choice of symbols is left to dictionary editors/makers, it is understandable if all instances of /ə/ in *SECJ6* are systematically converted to /ʌ/ in *Weblio*, but this is not the

case. Furthermore, *Weblio* uses /ʌ/ as a weak vowel for many British variants, but to the best of our knowledge, such transcription is not found in British dictionaries. The fact that /ʌ/ is also used for the **HABIT** and **MANUAL** vowels is unexpected and makes this matter even more confusing.

### 6.7 Transcription of suffixes

This section focuses on the transcription of suffixes to further verify consistency. From AWL300, derived forms with five suffixes (*-tion*, *-ment*, *-ical*, *-ity*, *-ed*) were chosen for analysis because there are five or more samples with phonetic transcriptions for both American and British pronunciations in *Weblio*.

(i) *-tion*: There are 36 words with the suffix *-tion*. Both *Eijiro* and *goo* consistently transcribe *-tion* as /-ʃən/ for both American and British pronunciations. *Weblio* shows inconsistency: twelve words are transcribed as /-ʃən/ but 24 words as /-ʃʌn/.

(ii) *-ment*: There are five words ending with *-ment*. Both *Eijiro* and *goo* transcribe the suffix consistently as /-mənt/. *Weblio* again shows inconsistency: two words with /-mənt/ (**document**, **environment**) but three with /-mʌnt/ (**enforcement**, **assessment**, **amendment**).

(iii) *-ical*: There are seven words that end with *-ical* (**hierarchical**, **technical**, **medical**, **empirical**, **chemical**, **identical**, **ethical**). The most consistent is *goo*, which provides only one variant with /-kəl/. However, note that an italicized /ə/ in *Progressive5* is not reflected in *goo*, which means it fails to indicate the possibility of pronouncing *-ical* with a syllabic /l/. *Eijiro* shows slight inconsistency. Of seven words, three words only give the American pronunciation /-kəl/. The remaining four words provide transcriptions for the two varieties: /-kəl/ for American and /-kəl/ for British. This is misleading since it appears as if the presence/absence of /ə/ is a difference between the two varieties. Lastly, as for *Weblio*, the notation itself is consistent: all seven words have /-ɪk(ə)l/ for American and /-ɪkʌl/ for British. However, the reason for using /ʌ/ for British pronunciation is unknown.

(iv) *-ed*: Regarding *-ed* ending, 12 words in which *-ed* is pronounced

with an extra vowel were analyzed: *suspended*, *converted*, *anticipated*, *rejected*, *submitted*, *excluded*, *generated*, *prohibited*, *illustrated*, *isolated*, *devoted*, *distorted*. First, *goo* does not provide phonetic transcriptions for most words, but when it does, the transcription is /-id/ (4 words). Similarly, *Eijiro* also does not give transcriptions for all words, but when it does, the transcription is /-əd/ (5 words). One exception is *distorted* with /-id | -əd/. *Weblio* again shows inconsistency. Of twelve words, four words have /-id | -id/, five words have /-id | -ʌd/, and three words have /-ʌd | -ʌd/.

(v) *-ity*: The last case is seven words with *-ity* ending. Of the three dictionaries, *goo* is the most consistent; it gives /-əti/ for American in all seven words and the British pronunciation is not provided. *Eijiro* gives /-əti/ for American and /-iti/ for British for two words (*authority* /[[US] əθɔːrəti | əθáɾəti | [UK] ɔːθɔːrɪti/, *priority* /[[US] praɪɔːrəti | praɪáɾəti | [UK] praɪɔːrɪti/). This is possibly because these words involve vowel differences in a stressed syllable, and thus require full transcription. Four words only give American pronunciation, and one word does not have a phonetic transcription (*minorities*). *Weblio* provides a confusing picture because *-ity* involves two weak vowels. A total of five patterns are identified: *authority* /-əti | -əti/, *priority* /-əti | -iti/, *validity* *entities*, and *stability* /-əti(z) | -iti:(z)/, *capacity* /-əti | -ʌti/, *minorities* /-ʌtiz | -ʌti:z/.

## 6.8 Summary

In summary, we have identified two main issues related to the transcription of weak vowels. One concerns the use of an italic /ə/ in both *Eijiro* and *goo*. *Eijiro* shows inconsistency in distinguishing American and British pronunciations in some cases, and *goo* fails to use an italic /ə/ entirely, resulting in missing out subtle pronunciation information such as the use of syllabic consonants.

The second issue is inconsistency in the choice of weak vowel symbols in *Weblio*, especially the inexplicable use of /ʌ/, which is the most prominent in the BONUS vowel but also found in the HABIT and MANUAL vowels. This is conspicuous in the transcriptions of derived forms and British pronunciations. As in the case of strong vowels, one reason for

this may be that their full transcription is not available in *SECJ6*. However, it should be noted that this notation is sometimes observed in American pronunciation and underived forms.

(Section 6 by Sugimoto)

## 7. Transcription of American and British variants

In this section, we delve deeper into how American and British pronunciation variants are described in the three online dictionaries.

Printed English-Japanese dictionaries usually show variations in the order of American English and British English, separating the two by a vertical bar (|). When more than one variant exists within a dialect, the variants are separated by a comma (see Table 7.1).

Table 7.1 Notation of American and British pronunciation variants

<i>SECJ6</i>	<i>Progressive5</i>	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>
/ Am   Br/	/ Am   Br/	/Am (米国英語), Br (英国英語)/	[US] Am   [UK] Br	Am   Br
/ Am 1, Am 2   Br 1, Br 2/	/ Am 1, Am 2   Br 1, Br 2/	—	[US] Am 1   Am 2   [UK] Br 1   Br 2	Am 1, Am 2   Br 1, Br 2

In contrast, *Weblio* uses the labels “(米国英語)” and “(英国英語)” instead of the vertical bar to indicate American and British variants. While the variants shown are supposed to correspond to the description of *SECJ6*, that is not always the case. (Note that the pronunciation descriptions taken from *SCEJ6* in this section all follow the print version.)

Deviation 1:

**ensure** *SECJ6*: /ɪnʃʊə, en- | -ʃʊ:, -ʃʊə/  
*Weblio*: /ɪnʃʊə (米国英語), ɪnʃʊ: (英国英語)/

**qualitative** *SECJ6*: /kwələtətɪv | kwəl(ə)tətɪv, -tətɪv/  
*Weblio*: /kwələtətɪv (米国英語), kw'əl(ə)tətɪv (英国英語)/

Deviation 2:

**advocate** *SECJ6*: /ædvəkət, -kət/  
*Weblio*: /'ædvəkət (米国英語), 'ædvʌkət (英国英語)/

**extract** *SECJ6*: /ɪkstrækt, eks-/  
*Weblio*: /ɪkstr'ækt (米国英語), 'e,kstrækt (英国英語)/

Deviation 3:

**emphasis**    *SECJ6*: /émfə́sɪs/  
*Weblio*: /émfə́sɪs (米国英語), 'emfə́sɪs (英国英語)/

As seen from the examples in Deviation 1, the rule of thumb for *Weblio* appears to adopt only the first variant found in *SECJ6*. Deviations 2 and 3 pose significant problems. Regarding Deviation 2, the variants that can in fact be heard in both dialects are incorrectly classified into American and British variants. Moreover, an unnecessary secondary stress is applied to the first syllable of the British variant. In the case of Deviation 3, the difference of the second syllable in the example **emphasis**, /fə/ for American and /fʌ/ for British pronunciation, is inexplicable (§6.5).

*Eijiro* has the labels [US] and [UK], and variants between dialects and within a dialect are separated by a vertical bar (|). *goo* employs the same system as its source dictionary, *Progressive5*, but without slashes (/ /).

## 7.1 Rhoticity

Among the differences in American and British pronunciation, the presence or absence of rhoticity is one of the most characteristic features. Accordingly, any English-Japanese dictionary systematically indicates r-fulness in American pronunciation and r-lessness in British pronunciation. *SECJ6* employs hooked-schwa /ə̃/ to express rhoticity, whereas *Progressive5* utilizes italicized “r” to indicate that vowels with it can be pronounced with or without rhoticity, depending on the dialects (Table 7.2).

Table 7.2 Transcription of rhoticity in American English

Vowel	n	<i>SECJ6</i>	<i>Weblio</i> *	<i>Progressive5</i>	<i>goo</i>	<i>Eijiro</i>
NEAR	3 (+7)	ɪə̃   ɪə	ɪə̃   ɪə	ɪə̃r	ɪə̃r	ɪə̃r
SQUARE	3 (+7)	eə̃   eə	eə̃   eə	eə̃r	eə̃r	eə̃r
START	4 (+6)	ɑ̃   ɑ	ɑ̃   ɑ	ɑ̃r	ɑ̃r	ɑ̃   ɑ
NORTH	14	ɔ̃   ɔ	ɔ̃   ɔ	ɔ̃r	ɔ̃r	ɔ̃   ɔ
CURE	3 (+7)	juə̃   juə	juə̃   juə	juə̃r	juə̃r	juə̃   juə
NURSE	11	ɜ̃   ɜ	ɜ̃   ɜ	ɜ̃r	ɜ̃r	ɜ̃r   ɜ

\* *Weblio* unsystematically/interchangeably uses the symbols (ɑ̃ for ɑ, and : for :).

Turning to the three dictionaries in question, *Weblio* adopts the hooked-schwa system, whereas *Eijiro* and *goo* use “r” to express rhoticity. By limiting our scope to strong vowels, the AWL300 dataset contained 38 words with rhotic vowels. For vowel types with an insufficient number of sample words, words outside the dataset (e.g., *peer*, *sphere* for NEAR vowels) were added so that at least 10 words were included in the analysis per vowel type.

Inconsistent and unsystematic descriptions are common in *Weblio* as in other vowels (§5, §6). Use of different symbols for inflected forms and derived forms are observed: e.g., *hierarchical* /ɑːr | ɑ:r/ (cf. *hierarchy* /ɑː | ɑ:/); *subordinate* /ɔːr | ɔ:r/ (cf. *ordinate* /ɔː | ɔ:/); *circumstances*, *confirmed*, *converted*, *inferred* /ɜːr | ɜ:r/ (cf. *circumstance* /ɜː | ɜ:/; *confirm*, *convert*, *infer* /ɜː | ɜ:/). Using different symbols for an identical vowel category can confuse dictionary users.

The serious notation problem regarding the sequence V/r/V across weak syllables (§6.3) is also found with V/r/V sequences in which the first syllable is strong and the second syllable is weak. Among four words that include such a sequence (*coherence*, *concurrent*, *assurance*, *inherent*), the British pronunciation of *concurrent* /kənˈkʊːrənt | kʌˈnkɜːrənt/ is missing /r/; *assurance* /əʃˈu(ə)rəns | əʃˈɔːrəns/ has an excessive /r/ in the British pronunciation. Further, *inherent* /ɪnhɪ(ə)rənt | ɪˈnhɜːrənt/ does not reflect the transcription of *SECJ6* /ɪnhɪ(ə)rənt, -hɜːr/—the two variants can in fact be heard both in American and British pronunciation.

*Eijiro* also had sporadic mistranscriptions (e.g., *confirmed* /ɹ/, *converted* /əː/ [cf. *convert* /əːr | əː/]). Use of /ɹ/ for the American pronunciation of START, which is different from the one used in the explanatory notes (/ɑːr/; see Table 4.1), was found, too. Moreover, inconsistency in the transcription was observed; among the other nine NURSE vowels in the dataset, *insert* was transcribed as /əːr/ instead of /əːr | əː/. An analysis of an additional 20 words randomly chosen by the author revealed a third type of transcription /əːr | əː/.

/əːr | əː/ *vertical, mercy, nurse, allergic, curse, nurture,*

*burden, term, firm, certainty, girl, purpose, church,*  
*thirsty, purse*  
 /ə:r/ *certain, pearl, perverse*  
 /ər | ə:/ *learn, world*

Transcribing the same vowel in three different ways can be confusing for dictionary users.

Additionally, an interesting observation was made regarding British spellings. While *labor*, *color*, and *theater* with American spelling have /ə:r/, their British counterparts *labour*, *colour*, and *theatre* have /ə/. This is misleading because there are varieties, such as Canadian English, in which rhoticity is maintained, although British spelling is adopted.

In the case of *goo*, the italic “r” is not used at all (e.g., *target* /ɑ:r/, *controversy* /ə:r/), resulting in the transcription of rhotic pronunciation only. This is not in accordance with the symbols used by *Progressive5*, which *goo* is based on. Italicized “r” has likely been modified to non-italicized “r” in the process of adopting symbols from *Progressive5*. The same can be observed with yod-dropping (§7.3). Nevertheless, the transcription of *goo* is the most consistent among the three.

## 7.2 T-tapping

T-tapping refers to the pronunciation of /t/ and /d/, which is typical of North American English, in which the tongue briefly taps the alveolar ridge. *SECF6* recognizes three environments where /t/ is realized as tapped-t: (1) between a strong vowel and a weak vowel; (2) between weak vowels; (3) between /n/ and a weak vowel (Wells 1982: 248). Furthermore, the symbol /t̬/ is used, which is reflected in *Weblio* (e.g., *negative* /négət̬ɪv | 'nɛg.ə.tɪv/).

Among the AWL300 dataset, 30 words with /t/ appeared in the aforementioned environments. However, its description in *Weblio* is not consistent, and five patterns can be observed, as shown in Table 7.3. While only Pattern 1 is permissible, Patterns 2 and 5 overlook the potential t-tapping of American pronunciation, and Patterns 3 and 4 may lead to the misunderstanding that t-tapping can also be heard in British pronunciation.

Table 7.3 Transcription of t-tapping in *Weblio*

Pattern	Am	Br	Example	n. of words
1	ɾ	t	<b>interaction</b> /ɪntə'ækʃən   ɪntɜ:'ækʃən/	19
2	t	t	<b>identified</b> /aɪ'dentɪ'faɪd   aɪ'dentɪ'faɪd/	7
3	ɾ	ɾ	<b>authority</b> /əθɔ'ɹəʃi   ɔ:ɹəʃi/	2
4	ɾ	–	<b>incompatible</b> /ɪnkəmp'ætəbl/	1
5	t	–	<b>submitted</b> /sə'bɪmɪtɪd/	1

This issue of inconsistency extends to the transcription of inflected forms. While **identified** /aɪ'dentɪ'faɪd | aɪ'dentɪ'faɪd/ shows Pattern 2, the base form **identify** /aɪdɛntɪ'faɪ | aɪ'dentɪ'faɪ/ correctly shows Pattern 1. The same is true with **minorities** /maɪ'nɔ:rɪtɪz | maɪ'nɔ:rɪtɪ:z/ — **minority** /mɪn'ɔ:rɪʃi | maɪ'nɔ:.ɹɪ.ti/, **entities** /'entɪtɪz | 'entɪtɪ:z/ — **entity** /'ɛntɪʃi | 'entɪtɪ:/, and **monitoring** /'mɔnɪtɜ:ɪŋ | 'mɔ:nɪtɜ:ɪŋ/ — **monitor** /mɔ'nɪtɜ | m'ɔnɪtɜ/. Although the reason for this discrepancy is unclear, the correct use of t-tapping /ɾ/ with base forms likely reflects the source of transcription (*SECJ6*), which provides full transcription only for base forms. In other words, where transcription was readily available from *SECJ6*, its system was applied; otherwise, transcription from another system was adopted, leading to the coexistence of two notation systems. The same argument is presented in Section 5.1 for stress marking. Although providing pronunciation for all inflected forms is a significant advantage of online dictionaries and is undoubtedly useful for the users (as discussed in §8), the inconsistency noted is likely to cause considerable confusion.

However, this does not explain why there are cases where t-tapping /ɾ/ is used in American and British variants (Pattern 3). The other word from the samples with this pattern is **data** /dɛɪtə | d'æʃə/. Both **authority** and **data** are headwords in *SECJ6*, and transcription in full form is available: /əθɔ'ɹəʃi, ɔ:-, -θɔ:r- | ɔ:θɔ:r-, ə-/ and /dɛɪtə, dɛʃtə | dɛɪ-, dɔ:-/ , respectively. With more than one variant recorded for American and British pronunciations, some confusion may have occurred, and the first variant and the second variant of American pronunciation have been distributed to American and British variant respectively. On a side note, such an erroneous distribution to American and British pronunciation is observ-

able in the transcription of other phonetic features, such as the weak form (弱形) and the strong form (強形). For example, *her* / (弱形) (h)ə̃ | (h)ə; (強形) hɜː | hɜː/ in *SECJ6* is transcribed as / (弱形) (h)ə̃ | (h)ə; (強形) h'ɜː/ in *Weblio*. The information of *SECJ6* was apparently transferred to *Weblio* without taking into account the meanings of vertical bar and weak/strong forms.

In the case of Patterns 4 and 5, the reason why the transcription for the British variety is missing is unclear (see Section 5.3). The best speculations for Pattern 4 are that there were no differences in the transcription of American and British varieties other than /t/, and that the American/British difference of /t/ was disregarded. Regarding Pattern 5, it can be speculated that the same process of Pattern 2 took place first, and only one variant remained with an inappropriate label [米国英語].

### 7.3 Yod-dropping

Collins, Mees, and Carley (2019: 75, 213) define yod-dropping as a frequent loss of /j/ after the dental and alveolar consonants /θ t d s z n l/ in most Americans. A total of 15 words, eight from the AWL300 dataset (*assume, constitutional, duration, mature, pursue, revenue, revolution, substitution*) and seven from Svartvik and Leech (2016: 164) (*due, enthusiastic, new, resume, suit, Tuesday, tune*), were included in the analysis. For each dictionary, several patterns observed are summarized in Table 7.4.

*Weblio*'s description of yod-dropping generally corresponds to that in *SECJ6*. However, Pattern 4 *substitution* is misrepresented: in *SECJ6*, it is transcribed as /sʌbstət(j)úːʃən | -tjúː-/. Additionally, Pattern 3 *revolution*, transcribed as /rɛvə'lúːʃən/ in *SECJ6*, should be represented in the same way in *Weblio*.

*Eijiro* exhibits five patterns. The differences between Patterns 2 and 3 are not clear.

Three patterns were found in the way the yod was missing in *goo*. Pattern 3 demonstrates a significant issue. Both the American and British pronunciations are shown with yod, yielding identical pronunciations. As pointed out in 7.1, all italicized symbols appear to have been changed

Table 7.4 Yod-dropping in *Weblio*, *Eijiro*, and *goo*

<i>Weblio</i>	
Pattern	Transcriptions
1: / (j)   j/	<b><i>due</i></b> /d(j)ú:   djú:/, <b><i>tune, new, enthusiastic, mature, revenue, duration, Tuesday, constitutional</i></b>
2: /-   (j)/	<b><i>assume</i></b> /əsúm   əs(j)ú:m/, <b><i>pursue, suit, resume</i></b>
3: /-   -/	<b><i>revolution</i></b> /rèvəlú:ʃən   ,rɛvʌˈlu:ʃʌn/
4: / (j)   -/	<b><i>substitution</i></b> /s`ʌbstət(j)ú:ʃən   ,sʌbstɪˈtu:ʃʌn/
<i>Eijiro</i>	
Pattern	Transcriptions
1: /j/	<b><i>mature</i></b> /mætúəɾ/, <b><i>substitution</i></b>
2: /j   j/	<b><i>duration</i></b> /djuréiʃən   djuréiʃən/, <b><i>enthusiastic</i></b>
3: /-, j   j/	<b><i>Tuesday</i></b> /tú:zdei, tjú:zdei   tjú:zdei/, <b><i>constitutional</i></b>
4: /-   j/	<b><i>assume</i></b> /əsúm   əsjú:m/, <b><i>pursue, revolution, revenue, due, tune, new, resume</i></b>
5: /-/	<b><i>suit</i></b> /sú:t/
<i>goo</i>	
Pattern	Transcriptions
1: /j/	<b><i>revolution</i></b> /rèvəljú:ʃən/
2: /-   j/	<b><i>assume</i></b> /əsúm   əsjú:m/, <b><i>pursue, suit, resume, enthusiastic</i></b>
3: /j   j/	<b><i>due</i></b> /djú:   djú:/, <b><i>duration, constitutional, substitution, revenue, mature, Tuesday, tune, new</i></b>

to non-italicized symbols in the process of adopting the symbols from *Progressive5* because if yod is italicized in the American variants, the variation is appropriately transcribed.

#### 7.4 BATH words

British speakers use PALM vowels for a set of words where American speakers use TRAP vowels. Collins et al. (2019: 94–95) explain that such BATH words have orthographic <a> followed by either a nasal (+ consonant) or a fricative (+ consonant). As only three words could be categorized into BATH words (***task, draft, behalf***) in the AWL300 dataset, an additional 17 words (***after, answer, ask, aunt, bath, calf, castle, command, dance, example, glass, grant, master, past, path, rather, sample***), considered typical examples of BATH words, were selected from Collins et al. (2019), and altogether 20 words were put

Table 7.5 Words that the British variant is recorded as a bath word

Pattern	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>	Words	n
1	√		√	<i>answer, aunt, calf, castle, draft, example, glass, grant, master, past, rather, task</i>	12
2	√	√	√	<i>after, ask, bath, dance, path</i>	5
3			√	<i>behalf, command, sample</i>	3

into analysis.

Table 7.5 shows that all 20 BATH words were recorded as such in *goo* (e.g., *calf* /kæf | ká:f/, *example* /igzæmpl | -zá:m-/).

Conversely, *Eijiro* did not have BATH vowels incorporated into most of the words and only had the TRAP variant recorded (e.g., *calf* /kæf/, *example* /igzæmpl/). It is unclear on what basis the BATH variant was selected with the five words in Pattern 2. Four of them (except for *after*) had BATH vowels as an American and British variant (e.g., *ask* /æsk, ásk | ú:sk/).

Turning to *Weblio*, the transcription is generally in accordance with *SECJ6*. However, with three words of Pattern 3 (*behalf, command, sample*), it is not apparent why the BATH vowel was not recorded, although the British variant is clearly indicated in *SECJ6* (e.g., *behalf* /bihæf, bæ- | -há:f/). Moreover, there are unexpected transcriptions, possibly typos, in *scenario* /səné(ə)riðu | sənéná:ri`əu/ and *aunt* /'ænt | ə:nt/. Such inconsistencies, which also appear sporadically in other parts of the transcription, may present challenges for casual dictionary users.

## 7.5 Non-systematic differences between American and British English pronunciations

In this section, we compare the transcriptions of words that exhibit non-systematic differences between American and British pronunciations. In addition to *data* in the dataset, 14 words were selected from Svartvik and Leech (2016: 165–166) and Trudgill and Hannah (2017: 56–57). Table 7.6 presents the transcriptions of the 14 words in *SECJ6* and the three dictionaries.

In *Weblio*, inconsistency (or possible errata) is observed, as has been

Table 7.6 Phonetic transcriptions of 14 words that have non-systematic differences

	<i>SECJ6</i>	<i>Weblio</i>	<i>Eijiro</i>	<i>goo</i>
<b>buoy</b>	bú:i, bó:i   bó:i	bú:i   'bu:i:	bú:i, bó:i   bó:i	bú:i, bó:i   bó:i
<b>clerk</b>	klɔ:k   klá:k	kl'ɔ:k   klá:k	klɔ:k   klɔ:k	klɔ:rk   klá:k
<b>data</b>	déi:tə, dá:tə   déi-, dá:-	déi:tə   d'æ:tə	déi:tə, dá:tə   déi:tə, dá:tə	déi:tə, dá:tə   déi:tə
<b>garage</b>	gərá:ɜ, -rá:dɜ   gɛ́ra:ɜ, -ra:dɜ	gərá:ɜ   g'æra:ɜ	gərá:ɜ, gərá:dɜ   gɛ́ra:ɜ	gərá:dɜ   gɛ́ra:dɜ
<b>herb</b>	(h)ɔ:b   há:b	(h)'ɔ:b   h'ɔ:b	ɔ:rb   hæ:b	á:rb   há:b
<b>leisure</b>	lí:zə, lé:zə   lé:zə	lí:zə   lé:zə	lí:zə   lé:zə	lí:zə   lé:zə
<b>lever</b>	lé:və, lí:-   lí:və	lé:və   lí:ə	lé:və   lí:və	lé:və   lí:və
<b>privacy</b>	práivəsi   prív-, práiv-	práivəsi   'praivəsi:	práivəsi	práivəsi
<b>process</b>	prá:ses, pró:-   prá:-	prá:ses   pr'ə:ses	prá:ses   pró:ses	prá:ses   pró:-
<b>route</b>	rú:t, ráut   rú:t	rú:t   ru:t	rú:t, ráut   rú:t	rú:t, ráut
<b>schedule</b>	skédʒu:l, -dʒu:l   fé:dju:l	skédʒu:l   fé:dju:l	skédʒu:l   fé:dʒu:l	skédʒu:l   fé:dju:l
<b>shone</b>	ʃóun   ʃón	ʃóun   ʃ'ɔn	ʃóun	ʃóun   ʃón
<b>tomato</b>	təméi:təu   -má:təu	təméi:təu   təmá:təu	təméi:təu   təmá:təu	təméi:təu   -má:-
<b>vase</b>	véis, véiz   vá:z	véis   vá:z	véis   vá:z	véis, véiz   vá:z
<b>zebra</b>	zí:brə   zéb-, zíb-	zí:brə   'zebɹə	zí:brə	zí:brə   zéb-

pointed out with other vowels (§7.1). In addition to the British variant of **data** with a tapped-t (already mentioned in Section 7.2), /v/ is missing in the British variant of **lever**, and a stress mark is missing in the British variant of **route**.

More importantly, with several words, the essential information in its source dictionary, *SECJ6*, is completely missing in *Weblio*. It does not tell that **buoy** may sound identical to “boy,” nor does it show how variants of **data** (/d'æ:tə/ for American variant, /dá:-/ for British variant), **privacy** (the British variant /prív-/), and **route** (the American variant /ráut/) are distributed in American and British pronunciation. Furthermore, the variants missing are those unexpected to Japanese users who are likely to be familiar only with the *katakana* counterparts “データ” (/de.e.ta/), “プライベート” (/pu.ra.i.ba.si.i/), and “ルート” (/ru.u.to/). This information is too precious to be omitted, and they should have provided guidelines for variant selection if they had to limit the inclusion of the variants.

For *Eijiro* and *goo*, the essential variants of the three words (**buoy**,

**data**, **route**, but **privacy**) mentioned above are included. *Eijiro* does not offer British variants for **shone** and **zebra** because of inexplicable reasons.

## 7.6 Summary

Analyses of American and British pronunciation variants highlighted three problems. The first is the inconsistency in the use of symbols supposed to show systematic differences between American and British pronunciations. Users may be confused by the use of different symbols to show rhoticity in *Weblio* and *Eijiro*. A similar issue arises with *Weblio*'s erroneous distribution of non-regional variants from the source dictionary, *SECJ6*, to American and British pronunciations. The second issue is the lack of italicization in *goo*, which is devastating in showing systematic differences in rhoticity and yod-dropping in American and British pronunciations. Finally, *Weblio*'s policy of including only one variant per word needs to be reconsidered because it can lead to missing critical information for understanding lexical differences in American and British pronunciation. (Section 7 by Uchida)

## 8. Inflected forms

Unlike dictionaries in hard copies, online dictionaries, given the unlimited space on the website, are free from page limitation. This can mean that they are not confined to the number of headwords to record. This advantage is in fact well taken in online dictionaries under discussion, especially in *Weblio*, which provides many inflected forms (e.g., **implies**), in addition to their bare forms, on separate webpages. This is not usually the case in dictionaries published in hard copies. Therefore, this section discusses inflected forms in particular as headwords. Among the dataset of 300 words (i.e., AWL300), there are 55 words that have inflectional suffixes attached to them, which are the focus of this section. The inflectional suffixes include the third-person singular *-s* on verbs, the *-ing* on verbs, plural suffixes on nouns and the *-ed* on verbs. There are not so many words in the first two cases: two forms with the third-person singular suffix (**implies** and **ceases**) and two with the *-ing*

suffix (*corresponding* and *monitoring*), but there are plenty of examples of the plural forms of nouns (20) and forms with the *-ed* suffix (31).

Overall, the accuracy of the transcription being put aside, *Weblio* provides the headwords and transcriptions for all 55 words, whereas *Eijiro* and *goo* may or may not provide them depending on the word. With respect to the forms with the third-person singular suffix *-s* on verbs (both of the two words are regular verbs) and forms with the regular plural suffix *-s* on nouns, *Weblio* has headword pages with their transcriptions for all of them, but the other two dictionaries provide headword pages for few and transcriptions for none, although for the irregular plural form (*criteria*), the transcription is available in all three dictionaries. Regarding the forms with suffixes *-ing* and *-ed*, *Weblio* provides the headword pages and transcriptions for all forms. The other two dictionaries provide them only when the form is tagged as an adjective (e.g., *corresponding*, *established* and *relaxed*); the other forms do not have independent pages.

In terms of transcription accuracy, forms with the suffixes *-ed* and *-s* will be discussed. The allomorphs that consist only of consonants (e.g., /t/, /z/) are accurately transcribed. However, the allomorphs that have an inserted vowel (e.g., *devoted*) may vary within a dictionary in terms of the vowel symbol, especially in *Eijiro* and *Weblio*. For example, *Eijiro* transcribes the suffix in *devoted* as /əd/ but that in *distorted* as /id/ for the American pronunciation. *Weblio*'s transcription includes the unexpected /ʌ/ for the inserted vowel in many cases, which makes up part of the general problem in its transcriptions of reduced vowels discussed in Section 6. *Weblio* also unexpectedly gives /-tɪd/ for the entire word form for *isolated*, *illustrated* and *distorted* (for American pronunciation) presumably because it takes its transcription from *SEJ6*, which gives only /-tɪd/ as it is shown on *Weblio*'s page. (Section 8 by Hirayama)

## 9. Conclusion

This study examined the phonetic transcriptions of the three online English-Japanese dictionaries that are available free of charge: *Weblio*, *Eijiro*, and *goo*. Key issues identified include discrepancies from the

source dictionaries that likely took place in the transplanting process, and inconsistencies in the notation system (e.g., use of multiple phonetic symbols for one phonemic category, absence of italics, and inconsistent stress marking), with causes that can only be speculated (e.g., concatenating different sources without unifying the symbols used in the original).

So now we know where the students got their *Pygmalion* pronunciation from!

Providing consistent transcriptions is fundamental to any dictionary. Also, the capacity of online dictionaries to hold more information than printed ones should be fully utilized, for example, by providing full phonetic transcriptions for all words and offering easily accessible explanatory notes. The role of online dictionaries is becoming increasingly significant, as university students often rely on them as a convenient tool for checking unfamiliar English words and phrases. It is expected that the accuracy and quality be guaranteed while making the most of the advantages of online dictionaries.

#### NOTES

- 1) On 14 May, 2025, NTT DOCOMO, INC. announced that the *goo* dictionary service will be terminated at 13:00, 25 June. The authors came across this news at the time of proofreading.

#### DICTIONARIES and their ABBREVIATIONS

<i>Weblio</i>	<i>Weblio Eiwa-Jiten Waei-Jiten</i>
<i>Eijiro</i>	<i>Eijiro on the WEB</i>
<i>goo</i>	<i>goo Jisho</i>
<i>SECJ6</i>	<i>Shin Eiwa Chu Jiten [Kenkyusha's New College English-Japanese Dictionary]</i> (6th ed.), 1994.
<i>Progressive5</i>	<i>Progressive Eiwa Chu Jiten [Shogakukan Progressive English-Japanese Dictionary]</i> (5th ed.), 2012.

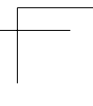
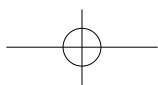
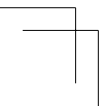
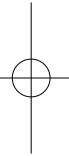
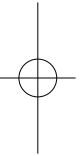
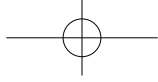
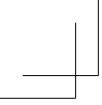
#### WEBPAGES

- [ 1 ] Weblio Eiwa-Jiten Waei-Jiten <<https://ejje.weblio.jp/>> Accessed 31 Jan. 2025.
- [ 2 ] Eijiro on the WEB <<https://eow.alc.co.jp/>> Accessed 31 Jan. 2025.
- [ 3 ] goo Jisho <<https://dictionary.goo.ne.jp/en/>> Accessed 31 Jan. 2025.

- [ 4 ] Weblio, Copyrights and Trademarks <[https://help.weblio.jp/ejje/company/c\\_info/h0023](https://help.weblio.jp/ejje/company/c_info/h0023)> Accessed 31 Jan. 2025.
- [ 5 ] GRAS Group Corporation <<https://gras-group.co.jp>> Accessed 31 Jan. 2025.
- [ 6 ] What is “Eijiro”? <<https://eow-support.alc.co.jp/hc/ja/articles/31894248675993>> Accessed 31 Jan. 2025.
- [ 7 ] Electronic Dictionary Project (EDP) <<https://eow-support.alc.co.jp/hc/ja/>> Accessed 31 Jan. 2025.
- [ 8 ] goo <<https://www.goo.ne.jp>> Accessed 31 Jan. 2025.
- [ 9 ] goo sitemap <<https://dictionary.goo.ne.jp/sitemap/>> Accessed 31 Jan. 2025.
- [10] Explanatory notes of Weblio <<https://ejje.weblio.jp/explanatory-notes/#1>> Accessed 31 Jan. 2025.
- [11] Explanatory notes of goo <<https://dictionary.goo.ne.jp/help/ej/>> Accessed 31 Jan. 2025.
- [12] Explanatory notes of Eijiro <<https://eow-support.alc.co.jp/hc/ja/articles/33903242775193>> Accessed 31 Jan. 2025.

## REFERENCES

- Collins, Beverley, Inger M. Mees, and Paul Carley. 2019. *Practical English Phonetics and Phonology: A Resource Book for Students*. 4th ed. London: Routledge.
- Coxhead, Averil. 2000. *The Academic Wordlist*. <[https://www.wgtn.ac.nz/lals/resources/academic\\_wordlist](https://www.wgtn.ac.nz/lals/resources/academic_wordlist)>
- Koyama, Toshiko, and Toshiyo Nabei. 2024. “A qualitative analysis of how look-up behavior differs in two conditions.” *Lexicon* 54, 73–101.
- Merriam-Webster Online* <<https://www.merriam-webster.com/>>
- Pullum, Geoffrey K., and William A. Ladusaw. 1996. *Phonetic Symbol Guide*. 2nd ed. Chicago: University of Chicago Press.
- Svartvik, Jan, and Geoffrey Leech. 2016. *English: One Tongue, Many Voices*. 2nd ed. New York: Palgrave Macmillan.
- Takebayashi, Shigeru. 1996. *Eigo Onseigaku* [English Phonetics]. Tokyo: Kenkyusha.
- Takebayashi, Shigeru, and Hiroko Saito. 2008. *Shinso-ban Eigo Onseigaku Nyumon* [Introduction to English Phonetics: New Edition]. Tokyo: Taishukan Shoten.
- Trudgill, Peter, and Jean Hannah. 2017. *International English: A Guide to Varieties of English Around the World*. 6th ed. London: Routledge.
- Wells, J. C. 1982. *Accents of English: An Introduction*. Cambridge: Cambridge University Press.
- Wells, J. C. 2008. *Longman Pronunciation Dictionary*. 3rd ed. Harlow: Pearson Education.



# An Analysis of the *Collins COBUILD Advanced American English Dictionary*, Third Edition

TAKAHIRO KOKAWA      RIKA AOKI  
RUMI TAKAHASHI      KAZUO IKEDA

## 1. Introduction

The year 2023 saw the publications of two major revised editions of EFL dictionaries from a major British publishing company, HarperCollins Publishers Limited (henceforth Collins). These are, the *Collins COBUILD Advanced Learner's Dictionary of English*, 10th edition (henceforth *COB10*) and the *Collins COBUILD Advanced American English Dictionary*, Third Edition (henceforth *COBAm3*), in this order. Though the former does not have the word 'British' in its title, the two books may be considered twin dictionaries with the same concept, design and size, featuring two major varieties of World Englishes, namely, British and American English respectively. *COBAm3* represents a significant update over its predecessor, the *Collins COBUILD Advanced American English Dictionary*, Second Edition published in 2016 (henceforth *COBAm2*). As we can see, there are seven years' span between the publications of *COBAm2* (2016) and *COBAm3* (2023) as opposed to the five years between the ninth edition of the COBUILD (2018, henceforth *COB9*) and *COB10* (2023). Still, considering that the first edition of the *Collins COBUILD Advanced Dictionary of American English* henceforth *COBAm1* was released in 2007, the speed of revision from its second edition to the third edition was faster than the first major update of the American version of the dictionary. Our team has made an analysis on *COB10* in Kokawa et al. (2024). In this paper, we would like to review *COBAm3*, mainly in comparison with *COBAm2* and its British counterpart, *COB10*.

At the time of writing this paper, only hardcover copies were available in the Japanese (and probably the international) market for *COBAm3* (ISBN: 978-0-00-860778-4) and *COB10* (ISBN: 978-0-00-844490-7), while for *COBAm2*, paperback editions were available (ISBN: 978-0-00-813577-5). We have used these respective editions for our survey here.

The front cover designs of *COBAm3* and *COB10* are identical in layout, with their titles (and in the case of *COB10*, its edition number) at the top:

Collins   COBUILD	Collins   COBUILD
ADVANCED	ADVANCED
AMERICAN	LEARNER'S
ENGLISH	DICTIONARY
DICTIONARY ( <i>COBAm3</i> )	10 <sup>th</sup> EDITION ( <i>COB10</i> )

Below the titles, we find yellow upper-case and lower-case letters C c, and at the bottom is the same front cover text, 'For upper-intermediate and advanced learners of English'. The design is the same in *COB9*, while the front cover had a simpler composition in *COBAm2*, with only the title and the edition number (Collins COBUILD Advanced American English Dictionary, NEW 2nd edition) shown, and a simpler text 'THE SOURCE OF AUTHENTIC ENGLISH' at the bottom. The covers for the American editions (*COBAm2* and *COBAm3*) are deep red, while the British versions (*COB9* and *COB10*) are deep blue.

On the other hand, the back cover of *COBAm2* was more informative, with 15 lines of description on what the dictionary could offer learners, including the claims 'Be confident in English', 'Expand your vocabulary' and 'Improve your fluency'. It also mentioned that the dictionary was 'ideal for students preparing for TOEFL<sup>®</sup>, IELTS, and Cambridge English: First, Advanced, and Proficiency' (tests) (back cover of *COBAm2*). Only the targeted CEFR level (namely, CEFR B2+) is mentioned in *COB10*, *COBAm3* (and *COB9*).

All four dictionaries reference the Collins Corpus on which the Collins COBUILD dictionaries are based. The text on the back covers of

each edition goes as follows:

You can trust Collins COBUILD -----  
All of our explanations, examples, and special features are based on our  
**4.5-billion**-word database of the English language, the Collins Cor-  
pus, which means that you can trust COBUILD to help you speak  
and write accurate and up-to-date English. The corpus is updated  
every month and has been at the heart of Collins COBUILD for over  
**25 years**.

-----  
(*COBAm2*, 2016, emphasis ours)

You can trust Collins COBUILD -----  
All of our explanations, examples, and special features are based on our  
**4.5-billion**-word database of the English language, the Collins Cor-  
pus, which means that you can trust COBUILD to help you speak  
and write accurate and up-to-date English. The corpus is updated  
every month and has been at the heart of Collins COBUILD for over  
**30 years**.

-----  
(*COB9*, 2018, emphasis ours)

You can trust Collins COBUILD -----  
All of our explanations, examples, and special features are based on our  
**4.5-billion**-word database of the English language, the Collins Cor-  
pus, which means that you can trust COBUILD to help you speak  
and write accurate and up-to-date English. The corpus is updated  
every month and has been at the heart of Collins COBUILD for over  
**30 years**.

-----  
(*COB10*, 2023, emphasis ours)

You can trust Collins COBUILD -----  
All of our explanations, examples, and special features are based on our  
**20-billion**-word database of the English language, the Collins Cor-  
pus, which means that you can trust COBUILD to help you speak  
and write accurate and up-to-date English. The corpus is updated  
every month and has been at the heart of Collins COBUILD for over  
**30 years**.

-----  
(*COBAm3*, 2023, emphasis ours)

What seems rather strange is that the size of the Collins Corpus suddenly jumps from 4.5 billion (*COBAm2* (2016), *COB9* (2018) and *COB10* (2023)) to 20 billion (*COBAm3* (2023)) in the latest version of its American edition, when all four dictionaries boast of monthly update of its corpus. It must have constantly grown during this period (between 2016 and 2023).

As mentioned in ‘About COBUILD dictionaries’, the first COBUILD dictionary was published in 1987 (p. v in the front matter of *COBAm3*—this text represents the fundamental concept of the COBUILD corpus and dictionary project, and is identical with those found in *COBAm2*, *COB9* and *COB10*). Thus, the COBUILD corpus project must have started earlier. Therefore, Collins could have updated the history length of COBUILD to 35 years in the latter two editions, but the number 35 may not have sounded well-rounded enough in English-speaking countries compared to 25 and 30 years.

There is also reference made to the integrated Collins reference site, ‘collinsdictionary.com’, on the back cover of both *COBAm3* and *COB10*, which goes:

Use the COBUILD Learner’s dictionary at  
collinsdictionary.com

The reference to Collins dictionary site used to be as follows on the *COBAm2* back cover:

Use the COBUILD Learner’s Dictionary at  
**www.collinsdictionary.com/COBUILD**

This URL is no longer valid, however.

The ‘collinsdictionary site is now titled ‘Free online dictionary, thesaurus, and reference materials’ and includes a whole host of reference materials regarding language and language learning. However, the most useful reference may be integrated multi-language dictionary and thesaurus. (cf. Chapter 8 of Kokawa et al. (2024:67–69)).

When we looked up a word entered in *COBAm3* but not in *COB10* (or

in *COBAm2*), such as **penmanship**, in the Collins online dictionary, the source dictionary is indicated as follows:

*Collins COBUILD Advanced Learner's Dictionary*. Copyright ©  
*HarperCollins Publishers*

This is instead of the entry being referred to as taking its information from the *Collins COBUILD Advanced American English Dictionary*. Thus, it may be surmised that at least with regard to the Collins online dictionary, the two editions, *COBAm3* and *COB10*, are considered as one EFL dictionary.

We found the following information in the front and back matter of *COBAm3*.

Front matter (pp. i–xxxv)

(title page) (p. i)

(Copyright page) (p. ii)

(Table of) Contents (p. iii)

About COBUILD dictionaries (p. v)

Acknowledgements (pp. vi and vii)

Guide to dictionary entries (pp. viii and ix)

Guide to dictionary features (Vocabulary in Context, Synonyms, Collocations, Usage notes, Prefixes and suffixes, Word histories, Visual Dictionary) (pp. x–xii)

Introduction (p. xiii)

Guide to definitions (pp. xiv–xvi)

Pronunciation (pp. xvii–xix)

List of grammatical notations (Word classes, Words and abbreviations used in patterns)

(pp. xx–xxi)

Explanation of grammatical terms (pp. xxii–xxxiii)

Irregular verbs (pp. xxxiv–xxxv)

A-Z dictionary text (pp. 1–1499)

Color ‘Visual Dictionary’ pages paged independently 1–16 inserted

between dictionary text pp. 774–775

**APPENDICES** (pp. 1501–1627)

Style and Usage (Writing style guide pp. 1505–1519, Special information pp. 1520–1523, Language in use (pp. 1524–1545)

Grammar (pp. 1549–1566)

Business English grammar (pp. 1567–1576)

Academic English grammar (pp. 1577–1591)

Glossary of Grammatical Terms (pp. 1595–1603)

Word Lists—Frequent words (pp. 1607–1619), Academic Word List (pp. 1620–1623)

Credits (for illustrations in Visual dictionary supplement and photographs all through the dictionary) (p. 1627)

With regard to the front matter in *COBAm3*, the majority of the contents represents an exact copy of, or a slight update over those in *COBAm2*. We would only like to compare the Acknowledgements pages in *COBAm3* (pp. vi and vii) and one in *COB10* (p. iv) and note the composition of the staff in both dictionaries. *COBAm3* first acknowledges the major contributors for the original COBUILD concept, namely, John Sinclair, Patrick Hanks, Gwyneth Fox, Richard Thomas, and 25 other lexicographers followed by 12 contributors to the first edition. The dictionary also appreciates the permission given for copyright materials offered by publishers and other sources for the Collins corpus. Then, it lists Paul Nation as Consultant, four people as ‘Editors’, nine names as reviewers—international (eight from Japan and one from Taiwan) and 22 people as ‘Reviewers—USA’.

*COB10* Acknowledgements are simpler. First it goes:

The publishers would like to acknowledge the following for their invaluable contribution to the original COBUILD concept and for their work on previous editions of this dictionary. (*COB10* iv)

Then the dictionary cites John Sinclair as ‘Founding editor-in-chief’, followed by the names of 76 editors and lexicographers in alphabetical order of their surnames, without any distinction or classification. Some

names of seemingly senior lexicographers listed in the first part of *COBAm3* Acknowledgements are found also in the *COB10* list, but the majority in *COB10* and *COBAm3* are different. Also, none of the names found in the Reviewers (both international and USA) sections in *COBAm3* seem to be listed in *COB10*.

It appears that *COBAm3* was revised by a different team from *COB10*, with supervision from senior staff common to *COB10* and *COBAm3*.

The A-Z dictionary text of *COBAm3* includes 1,499 pages, as opposed to 1,539 in *COBAm2* and 1,736 in *COB10*. However, our count showed that *COBAm3*'s one page is made up of two columns with 77 maximum lines each, while *COB10* and *COBAm3* also contain two columns but with 74 maximum lines. Thus, the amount of text information may not be so different in the three editions.

In Kokawa et al. (2024:24–25) we have discussed the striking change in presentation from *COB9* to *COB10*, namely, from A-Z text pages in (full) color (*COB9*) to black-and-white (*COB10*). The same regrettable policy alteration was applied to the revision from *COBAm2* to *COBAm3*. Also, the so-called 'Picture Dictionaries' featured throughout the A-Z dictionary in full color in *COBAm2* (and in *COB9*) were relocated to the 16 color pages between L and M entries in *COBAm3*, just as in *COB10*. We will not repeat our earlier comments here, but do note that the reduction of visibility, user-friendliness and the amount of information brought about by this alteration may not be negligible.

Still, through our limited research we found a steady series of reasonable updates from the second to the third editions of the *COBUILD Advanced American English Dictionary*. We would like to review them in different categories one by one below.

## **2. Headwords and Related Information**

### **2.1. Count of headwords and related information**

#### **2.1.1. Scope of the survey and entry count in the sample portions**

We counted the independent headwords (or entries) in one out of

every 50 pages starting from page 2 up to 1452 in *COBAm3*, and counted the headwords in the same (entry-based) portions in *COBAm2* and *COB10*. Run-ons and run-ins were not counted in our survey here. The sample parts constitute approximately 1.93% of the total *COBAm3* pages. In the sample sections, we found 636 headwords in *COBAm3*. The numbers of headwords in *COBAm2* and *COB10* in the corresponding portions are 628 and 657 respectively.

### 2.1.2. Entries characteristic to one or two dictionaries

Now we would like to have a look at those headwords only found in one or two dictionaries in the *COBAm2*, *COBAm3* and *COB10* group. Here differences due to the edition numbers (*COBAm2* versus *COBAm3*) and those derived from American/British editions (i.e. *COBAm2* and *COBAm3* versus *COB10*) are jointly reviewed. The results are tabulated below.

Table 2.1. Headwords characteristic in one or two editions

<i>COBAm2</i>	<i>COBAm3</i>	<i>COB10</i>	Items found in our sample portions
✓			-ability, -able, e-
✓		✓	nationalise
	✓		penmanship
	✓	✓	bar mitzvah, batswoman, flipped, gig economy, gimlet, jeggings, Jerusalem artichoke, pension scheme, signalwoman, signboard (10 items)
		✓	AOB, bather, bathing costume, batman, baton charge, breakfast television, champers, complaisant, complementation, cross-Channel, cross-party, each way, flipping, giant-killer, giant-killing, gilt-edged, ginger group, head girl, head lamp, medium wave, Open University, penn'orth, penny-farthing, pen-pusher, pensionable, pension book, PPS, practise, practised, praetorian guard, school dinner, school leaver, Square Mile, suss, sussed, washing line, washing powder, washing-up, washing-up liquid, waste disposal (40 items)
✓	✓		bath salts, crossbeam, crosstown, crossways, gift certificate, head-butt, national holiday, national security, PPO, racquetball, schmuck, school district, undiplo-matic, undiscriminating (14 items)

First of all, affixes (prefixes and suffixes) that were treated as independent headwords in *COBAm2* have now come to be entered as prefix or suffix columns in *COBAm3* and *COB10* (in our sample portions, **-ability**, **-able** and **e-**). Also, a British variant in spelling, namely **nationalise** as a headword in *COBAm3* was deprived of headword status and demoted to a note under **nationalize** (In BRIT, also use **nationalise**). This seems to be due to the systematic change in editorial policy rather than for linguistic reasons. No other deleted items were found in our sample portion in the update from the second to the third editions.

On the other hand, several items are found added in the revision from *COBAm2* to *COBAm3*. The word **penmanship** is uniquely found in *COBAm3*. We found 10 new items in the two editions published in 2023, namely *COBAm3* and *COB10*. These are **bar mitzvah**, **bats-woman**, **flipped**, **gig economy**, **gimlet**, **jeggings**, **Jerusalem artichoke**, **pension scheme**, **signalwoman**, **signboard**. Such entries are apparently not always new words or words that have trended recently. In the Collins online dictionary mentioned in Section 1 above, a ‘Recorded Usage’ chart derived from the Google Books Ngram Viewer is shown for some lexical items, and the chart for **signboard** shows it has been in constant use at least since the 17th and 18th centuries. The same is true for **pension scheme**. On the other hand, **flipped** (as in ‘flipped learning’), **gig economy**, **gimlet** may reflect recent trends in language. The lexical item **signalwoman** as opposed to **signalman** may have been added for gender equality consideration (cf. Kokawa et al. 2024:27 and 4.2.3 below), and reflect the recent trend in society as well.

It is notable that in our sample portions we found nearly three times more headwords unique to the British *COB10* (40 items), as opposed to headwords found only in its American counterparts (14 items). Naturally, it is unsurprising that items designating British institutions and localities (**cross-Channel**, **Open University**, **Square Mile**, and so forth) should have found their way into the British edition, but *COB10* also appears to contain much more interesting ‘lexical’, rather than so-called ‘encyclopedic’ items. Although smaller in number, items only found in American editions are also interesting. Incidentally, it is strange

that the item **national security** found its way only in the American editions, when it is a common (not to say global) concern for many countries all the time. Reference to the British expression ‘bank holiday’ in *COBAm3* for the entry **national holiday** may have also aroused learners’ interest. For the addition of new items outside our present scope of sampling, see 4.2.1 below.

We have found six cases of main headword (i.e. the first headword cited in an entry) difference between American (*COBAm2* and *COBAm3*) and British (*COB10*) editions. These all come from contrast between American and British spelling conventions. Most of the practices relate to whether the words are spelled solid, hyphenated or open, and the customs are not at all categorical, varying within the countries on both sides of the Atlantic. In many cases, the three dictionaries cite the other spelling possibilities after the main headwords as in: in BRIT, use **centre** (s.v. **center** in *COBAm3*) and **under way** also **underway** (s.v. **under way** in *COB10*). Six instances found are as follows:

Table 2.2. Differences in main headwords in *COBAm2*, *COBAm3* and *COB10* (1)

<i>COBAm2</i> , <i>COBAm3</i>	<i>COB10</i>	<i>COBAm2</i> , <i>COBAm3</i>	<i>COB10</i>
flipchart	flip chart	underway	under way
headfirst	head-first	washed-up	washed up
time-honored	time-honoured	washrag	wash-rag

Also, the following items differed in main headwords in terms of plurality and regional varieties within the headwords.

Table 2.3. Differences in main headwords in *COBAm2*, *COBAm3* and *COB10* (2)

<i>COBAm2</i>	<i>COBAm3</i>	<i>COB10</i>
suspender	suspenders	suspender
penny stock	penny stocks	penny shares

Incidentally, the headword **trousers** was already presented in *COBAm2* as **trousers** instead of **trouser**, so the update from **suspender** to **suspenders** in the *COBAm* revision may be part of the publisher’s policy for more systematic headword presentation.

### 2.1.3. Frequency band designation in *COBAm2*, *COBAm3* and *COB10*

As discussed in Kokawa et al. (2020:47–48) and Kokawa et al. (2024: 28), the fourth and later editions of COBUILD employ a three-diamond, five-level system to designate the frequency of lexical items, entry by entry. The more the number of diamonds applied to the entry, the higher the level of frequency the dictionary deems the lexical item has. The system is the same in *COBAm2* and *COBAm3*, and the application of diamonds, or the actual designations of frequency bands have not changed from *COBAm2* to *COBAm3* as far as our sample survey portions are concerned.

The following table shows the difference of the number of diamonds among the three dictionaries as found in our sample survey (the sample portions are the same as that for headword sampling).

Table 2.4. Difference in frequency band designations in *COBAm2*, *COBAm3* and *COB10*

headwords	<i>COBAm2</i>	<i>COBAm3</i>	<i>COB10</i>	headwords	<i>COBAm2</i>	<i>COBAm3</i>	<i>COB10</i>
anymore	◆◆	◆◆	0	IMF	0	0	◆
didn't	◆◆◆	◆◆◆	0	PR	0	0	◆
heading	0	0	◆	practice	◆◆◆	◆◆◆	◆◆

The non-application of lozenges to **anymore** in *COB10* is reasonable, as **anymore** is mainly an American spelling. The choice of zero for **heading** and **PR** may reflect actual language use. *COB10* defines **PR** as the abbreviation for proportional representation as well as public relations, while the two American versions mention only its being the abbreviated form for the latter. It may be not fair for *COBAm2* and *COBAm3* to apply no diamonds to **IMF**, while, in 2024 the United States is by far the largest contributor in terms of quota in the institution. This may therefore only reflect the level of interest amongst speakers of American English. The difference of rating for **practice** may reflect the fact that British English has a spelling variant **practise** for its verb use.

(Section 2 by Kokawa)

### 3. Pronunciation

This section examines how *COBAm3* represents the pronunciation of American English (GenAm), particularly in terms of the distinction between GenAm and British English (RP). By comparing the phonetic transcriptions in *COBAm3* with those in *COB10* and *COBAm1*, this section aims to determine whether *COBAm3* maintains consistency.

#### 3.1. Overview

The transcription system of *COBAm3* remains unchanged compared with that of *COBAm1*. According to the front matter of *COBAm3*, it retains its original approach to pronunciation representation, emphasizing intelligibility across English varieties, as in both *COBAm1* and *COB10*. The model pronunciations ‘are the result of a program of monitoring spoken English and consulting leading reference works’ (xvii), which is the same wording found in *COB10*.

A notable feature of *COBAm1*’s transcription system, which differentiates it from other dictionaries, is its use of underlining to mark stress and the symbol /y/ to represent the /j/ sound in the International Phonetic Alphabet (IPA). This convention remained unchanged in *COBAm3*. *COBAm1* states that ‘the symbols used in the dictionary are adapted from those of the International Phonetic Alphabet (IPA), as standardized in the English Pronouncing Dictionary by Daniel Jones (14th Edition, revised by A.C. Gimson and S.M. Ramsaran 1988)’ (*EPD14* hereafter). However, *COBAm3* does not explicitly explain why it deviates from IPA and *EPD14* for these two aspects. From a pedagogical perspective, it can be assumed that underlining makes stress placement easier for learners to recognize, and that the use of /y/ instead of /j/ is more intuitive, as it aligns with phonics or pronunciation respelling conventions. However, this raises the following question: if /y/ is used to represent /j/ for the sake of intuitiveness, should *COBAm3* also use /j/ instead of /dʒ/ to transcribe the initial sound in **joy**? Future research on user preferences could evaluate the effectiveness of this approach.

Another distinctive feature of the *COBAm1* transcription system is its use of /v/ in place of /a/ in IPA transcriptions of GenAm. As GenAm

does not have the /ɒ/ vowel, most dictionaries use /ɑ(:)/ to represent the LOT vowel in IPA notation. While *COBAm1* included a note in its ‘Guide’ (xxiii) explaining that /ɒ/ and /ɑ/ are the same in GenAm, this note is absent in *COBAm3*. Consequently, users may incorrectly assume that GenAm and RP share the same vowel sound. To clarify this discrepancy, *COBAm3* should either reinstate the explanatory note or adopt a different symbol, such as /ɑ/.

*COBAm3*’s stated policy is to provide only one pronunciation for most words unless the differences between RP and GenAm are particularly noticeable. In such cases, the RP variant is transcribed after the label BRIT, or multiple pronunciations are included if they are common in either RP or GenAm. While *COBAm3* claims to systematically indicate major distinctions between RP and GenAm, closer examination reveals inconsistencies and ambiguities in its representations. The following sections briefly examine selected GenAm features, including rhoticity (3.2.), *ask* words (3.3.), open back vowels (3.4.), yod-dropping (3.5.), vowels before the intervocalic /r/ (3.6.), and lexical items (3.7.). In 3.8., the comparison with the online dictionary is covered.

### 3.2. Rhoticity

As in previous editions, *COBAm3* transcribes rhotic vowels using a mixed system, as described by Masuda et al. (2008). In this system, /ə/ appears before /r/ following vowels that are not low back or central, while low back or central vowels are transcribed without /ə/. Masuda et al. suggested two possible explanations, which, according to them, are unlikely: either /ə/ is syllabic, contradicting stress-marking conventions, or it represents an inglide before /r/, which would be inconsistent with *COBAm1*’s phonetic detail (63–64).

Another possible explanation is that this system simplifies the transcription process by adapting the existing RP transcription framework. That is, rather than transcribing GenAm pronunciation from scratch, *COBAm3* may use RP transcriptions and add GenAm-specific sounds, namely /r/, where necessary. This approach accounts for its biletal transcription system in showing rhoticity.

### 3.3. *Ask* words

In GenAm, *ask* words are pronounced with /æ/, whereas in RP, they are pronounced with /ɑ/. This distinction is noted in the front matter of *COBAm3*. Based on *COBAm1*'s single-pronunciation policy, one would expect *ask* words to be transcribed only with /æ/. However, while most *ask* words are indeed represented with /æ/, some headwords show variations: for example, **Slav** is transcribed with /ɑ/, **ask**, **advantage**, **bask**, **basket**, **glass**, **hasp**, and **lather** with /ɑ, æ/, and **aunt** and **bath** with /æ, ɑ/, all without the BRIT label. This distribution mirrors Masuda et al.'s (2008) findings. Thus, although *COBAm3* reflects GenAm variants for *ask* words, inconsistencies remain in its transcription pattern.

### 3.4. Open back vowels

In GenAm, two vowel mergers involve the open back vowels: the LOT-PALM merger and the LOT-THOUGHT merger. In the former, the LOT vowel /ɒ/ in RP is pronounced as /ɑ/ in GenAm, making it equivalent to the PALM vowel. Most dictionaries, including *LDOCE6* and *OALD10*, reflect this merger, but *COBAm3* maintains a distinction, transcribing **lot**, **dog**, and **common** with /ɒ/ and **palm**, **father**, and **calm** with /ɑ/. This decision, although unique, may misrepresent real-world GenAm.

For the LOT-THOUGHT merger, many GenAm speakers no longer distinguish between the LOT and THOUGHT vowels. Studies such as Dinkin (2011) indicate that this merger is spreading, although it remains an ongoing shift. *COBAm3*, similar to most learners' dictionaries, maintains the distinction: THOUGHT words such as **caught**, **dawn**, and **walk** are transcribed with /ɔ/, while LOT words are transcribed with /ɒ/. Considering the transcription of these two mergers, as noted by Masuda et al. (2008) for *COBAm1*, *COBAm3* also employs a unique three-way contrast for open back vowels, which is not observed in other learners' dictionaries. This system may not be fully pedagogic because of a lack of faithful reflection of ongoing vowel mergers in GenAm.

### 3.5. Yod-dropping

Yod-dropping, the deletion of /j/ from /ju/ or /jʊ/, is 'nearing comple-

tion in many parts of the English-speaking world, including middle-class England, the northern U.S., and Canada' (Chambers, 'Social Embedding of Changes in Progress', 17). In GenAm, /j/ is typically elided after coronals, /t d n s z θ l/, leading to pronunciations such as 'toon' for **tune**, 'doo' for **dew**, and 'nooz' for **news** (Chambers, 'Social Embedding of Changes in Progress' 17).

*COBAm3* retains *COBAm1*'s yod-dropping transcriptions: /j/ is omitted in primary-stressed syllables, as in **tune** /tun/, **new** /nu/, and **pursuit** /pərsut/, while it is retained in unstressed syllables, as in **annual** /ænyuəl/, **failure** /feɪlyər/, **educate** /edʒuket/, **issue** /ɪʃu/, and **situate** /sɪtʃueɪ/. The latter three reflect yod coalescence, as noted by Masuda et al. (2008). However, for words with the /j/ sound on secondary stressed syllables, transcriptions are inconsistent: **avenue** and **residue** list both forms (/ævɪnyu, -nu/ and /rɛzɪdu, -dyu/), whereas **attitude** has only the yod-less form. This inconsistency may reflect North American trends in retaining yod in final unstressed syllables (Chambers, 'Yod-Dropping in an English Accent' 9), although no explicit rationale has been provided.

In summary, the transcription of yod-dropping remains unchanged from that in *COBAm1*, with some inconsistencies in the transcription rules remaining.

### 3.6. Vowels before intervocalic /r/

In GenAm, vowel contrast is lost when vowels are followed by post-vocalic /r/ (Wells, 1982). Specifically, the distinction between mid and low front vowels before /r/ is often neutralized. This phonological change, known as the *Mary-marry-merry* merger, eliminates the contrast between /ɛə(eə)/, /æɹ/, and /ɛ(e)/. *COBAm1* employs a three-way distinction for the three lexical sets, and *COBAm3* maintains this policy. Thus, *COBAm3* can be considered conservative in terms of preserving this distinction.

### 3.7. Lexical incidence

This section focuses on the lexical items that exhibit unsystematic variations between GenAm and RP. To examine whether the transcrip-

tion of *COBAm3* reflects regional lexical differences and whether it has changed from *COBAm1*, the transcriptions of the example words identified by Wells (1982) as differing between GenAm and RP were analyzed, following Masuda et al. (2008). The results showed that among the 41 words available as headwords in *COBAm3*, 28 words were transcribed only in their GenAm form. These words included **advertisement**, **ate**, **ballet**, **beta**, **borough**, **clerk**, **depot**, **docile**, **erase**, **figure**, **from** (strong form), **herb**, **iodine**, **laboratory**, **lieutenant**, **massage**, **of** (strong form), **process** (noun), **progress**, **quinine**, **schedule**, **shone**, **suggest**, **thorough**, **tomato**, **vermouth**, **wrath**, and **Z**.

Additionally, even when *COBAm3* provides both GenAm and RP forms, as in **vase** /vɛɪs, vɑːz/, the RP form is not marked with the code BRIT, which contradicts the transcription policy stated in the front matter (xvii). Regarding **address** (noun), *COBAm3* includes a pronunciation note under its headword.

The noun is pronounced /ədres/ or /ædres/. The verb is pronounced /ədres/.

Here, if prioritizing GenAm, listing the second variant (/ædres/) first would be more appropriate, as it is more prevalent pronunciation in GenAm.

In conclusion, *COBAm3* primarily focuses on and prioritizes GenAm, which benefits users by aligning with their needs and reducing their cognitive load. However, this approach contradicts the dictionary's stated policy, which claims to provide alternative pronunciations when multiple variants are common in either American or British English.

### 3.8. Online dictionary

As noted by Kokawa et al. (2024), Collins provides an online dictionary (Collins English Dictionary), that is centered on British English but also includes American English. In terms of pronunciation, users can access audio recordings of headwords in both RP and GenAm, including newly added headwords from *COBAm3*. However, some low-frequency headwords have only RP pronunciations because the dictionary is British-centered. These features are highly beneficial as they provide access to pronunciation information and its variations, which

are not fully available in print editions. Additionally, for 21,093 commonly used words, the dictionary offers links to pronunciation videos on YouTube.

The pronunciation transcription system used in *COBAm3* and that used in the online dictionary are not identical. This discrepancy arises from the online dictionary's British-centered approach as well as differences in notation, such as the representation of the /j/ sound in IPA, as mentioned in Section 3.1. That is, although the online dictionary follows a transcription system used in widely recognized dictionaries, representing the initial sound in **yellow** as /j/, *COBAm3* maintains its unique transcription (i.e., /y/). This inconsistency can confuse learners, particularly beginners who may struggle with multiple transcription systems.

Although this variation is unlikely to significantly hinder comprehension, maintaining consistency across formats would enhance the user experience, as it would ensure a standardized approach to pronunciation. These discrepancies likely stem from differences in how print and digital platforms prioritize accessibility and user needs; however, it remains unclear why the transcription systems have not been fully standardized across all formats.

### 3.9. Conclusion

In summary, the pronunciation transcription system for *COBAm3* has not changed since the first edition of the dictionary, with some inconsistencies and deviations from its policy remaining. Although the actual sound files are available online, further updates and a clearer distinction between RP and GenAm would be helpful for users.

(Section 3 by Aoki)

## 4. Definitions

### 4.1. Overview

The explanation of definitions found in the front matter of *COBAm3* is similar to that in the front matter of *COBAm2*.

First, the 'Guide to dictionary entries' in the front matter of

*COBAm3* resembles that of *COBAm2* but it is more similar to that of *COB10*. For example, the style of definitions is explained using exactly the same words in both *COBAm3* and *COB10*, as follows: ‘Full-sentence definitions show the ways in which the headword is used’ (*COBAm3* viii; *COB10* vi). By contrast, the characterization is a little different in the guide in *COBAm2*: ‘Full-sentence definitions show *how* the headword is used’ (viii; italics added). This former similarity may be because *COBAm3* was published in the same year as *COB10*.

The ‘Guide to dictionary entries’ in *COBAm3* was likely adopted from *COB10*. This supposition is supported by the fact that the two guides are similar to each other to the extent that an explanation of CEFR levels is found in the guide in *COBAm3* although CEFR levels are not indicated in the entries in *COBAm3*.

Second, the ‘Guide to definitions’ in the front matter of *COBAm2* and that in the front matter of *COBAm3* are also similar to each other. For example, there is an explanation of definitions in the ‘Guide to definitions’ in the front matter of *COBAm2* and that of *COBAm3*, and the explanation is exactly the same across the two editions:

One of the features of the *Collins COBUILD Advanced American English Dictionary* is that the definitions are written in full sentences, using vocabulary and grammatical structures that occur naturally with the word being explained. These definitions give you much more than the meaning of the word you are looking up, and also contain information on usage, register, typical context, and syntax. Whenever possible, words are explained using simpler and more common words. This gives us a natural defining vocabulary, with most words in our definitions being among the 2,500 most common words in English (xiv).

This text has been revised and updated as follows for *COB10*:

One of the most striking features of the *Collins COBUILD Advanced Learner’s Dictionary* is that the definitions (or explanations) are written in full sentences, using vocabulary and grammar that occur naturally with the word being explained. This enables us

to give a lot of information about the way a word or meaning is used by speakers of the language. Whenever possible, words are explained using simpler and more common words. This gives us a natural defining vocabulary, with most words in our definitions being amongst the 2,500 most common words in English (xii).

The underlined sentence in the *COB10* text differs from that in the parallel texts in *COBAm2* and *COBAm3*. However, the remainder of the text is the same as that found in *COBAm2* and *COBAm3*. It is interesting to note that, although *COBAm3* was published in the same year *COB10* was published, the discussion of the definitions in *COBAm3* remained the same as the one in *COBAm2* and was not the text revised in *COB10*.

The only difference between the 'Guide to definitions' in *COBAm2* and that of *COBAm3* is the difference in the examples of geographical labels. *COBAm2* gives AUSTRALIAN and SCOTTISH as example labels in addition to the labels AM and BRIT, while *COBAm3* gives only SCOTTISH as an example. The reason for this reduction is not clear. It is interesting to note that *COB10*, which was published in the same year as *COBAm3*, gives more examples than *COBAm3*. The examples of geographical labels in *COB10* are as follows: AUSTRALIAN, IRISH, NORTHERN ENGLISH, and SCOTTISH.

Although they are similar, another difference appears between the 'Guide to definitions' in *COBAm2* and *COBAm3* and that in *COB10*. The examples given in the section titled 'Other kinds of definition' are the same in *COBAm2* and *COBAm3*, but they differ from those in *COB10*. Compare:

Lastly, some definitions are expressed as if they are cross-references. For example:

**rd.** is a written abbreviation for **road**.

**e-commerce** is the same as **e-business**. (*COBAm2*, *COBAm3*)

**hr.** is a written abbreviation for **hour**.

A **banker's draft** is the same as a **bank draft**. (*COB10*)

Other than this, the text in the ‘Guide to definitions’ in both *COBAm2* and *COBAm3* are the same as that in *COB10*.

Third, the variety of style labels and pragmatic labels in the ‘Guide to definitions’, along with their meanings, are similar across *COBAm3* and *COBAm2*. However, there is only one difference between the examples in *COBAm3* and those in *COBAm2*: the example given for the style label HUMOROUS in *COBAm2* is **gents**, which is the same as the example given for the same label in *COB10*, whereas it is replaced with **mustachioed**, which means ‘having a mustache’, in *COBAm3*. It is not clear why **gents** was replaced with **mustachioed** in *COBAm3*, although it could have been considered old-fashioned.

Incidentally, the style labels in use in *COBAm3* are different from those in *COB10*. The style labels that are common between *COBAm3* and *COB10* are BUSINESS, COMPUTING, DIALECT, FORMAL, HUMOUROUS, INFORMAL, JOURNALISM, LEGAL, LITERARY, MEDICAL, MILITARY, OFFENSIVE, OLD-FASHIONED, SPOKEN, TECHNICAL, TRADEMARK, and WRITTEN. Two additional labels, RUDE and VERY OFFENSIVE, are found in *COB10*, whereas the label VULGAR, common to *COBAm2* and *COBAm3*, is replaced with VERY RUDE in *COB10*.

The pragmatic labels in *COBAm3* are shared with *COBAm2* and *COB10*. However, the labels are characterized differently, and the examples in *COB10* are different from those in *COBAm3*. The pragmatic labels are characterized at greater length in *COB10* than in *COBAm3*, largely because the former uses simpler words, while each label is explained using the key word itself in the latter. Compare:

APPROVAL: The label APPROVAL indicates that you use the word or expression to show that you like or admire the person or thing you are talking about. An example of a word with this label is *broad-minded*. (*COB10*, xv; an underline added)

APPROVAL: used to show that you approve of the person or thing you are talking about, e.g., **angelic**. (*COBAm3*, xvi; an underline added)

In this example, instead of the key phrase ‘approve of’ used in *COBAm3*, the paraphrase ‘like or admire’ is given in *COB10*. In addition, the example given for the label in *COB10* is different from that in *COBAm3*. As noted, the example for the label APPROVAL in *COB10* is ‘broad-minded’, whereas the example in *COBAm3* is ‘angelic’.

Thus, the ‘Guide to dictionary entries’ in *COBAm3* seems to have been adopted from *COB10*, while ‘Guide to definitions’ was inherited from *COBAm2* to *COBAm3* without much revision. The latter guide in *COB10*, retitled ‘Definitions’, is also similar to that in the two American versions.

## 4.2. Features of the new definitions

### 4.2.1 Effects of recent social changes

We found some word senses not present in *COBAm2* that were newly added in *COBAm3* in the selected sample pages. These newly added word senses in *COBAm3* are similar to or are the same as those in *COB10*. This could be expected because *COBAm3* was published the same year as *COB10*.

We selected 2 sample pages for every 100 pages of *COBAm3*, beginning with page 1, or approximately 2% of the dictionary. This approach is the same as the one that will be used in Chapter 5 of this article. We also referred to other pages besides the sample pages, as needed.

The introduction to *COBAm3* emphasizes that the entries in the dictionary are systematically compiled from corpus data. It asserts that the reliance on corpus research has enabled lexicographers to construct definitions with greater confidence and precision. The inclusion of novel lexical meanings in *COBAm3* that are absent in *COBAm2* can be attributed to its development being anchored in the Collins Corpus, a constantly updated linguistic database that contains 20 billion words. One example of new lexical meanings in *COBAm3* is the second sense given for the headword **eagle**:

#### **eagle**

[2] N-COUNT In golf, if you get an **eagle**, you get the golf ball

into a hole in two strokes fewer than the number of strokes which has been set as the standard for a good player.

Another example is the second sense of **medium-term**:

**medium-term**

2 N-SING When you talk about what happens in the **medium term**, you are talking about what happens over a few months or years, either in the future or after a particular event. □ *She risks losing her salary in the medium-term and damaging her long-term career prospects.*

In these examples, a new meaning was added to an existing headword from *COBAm2* and incorporated into *COBAm3*. In addition, *COBAm3* also includes entirely new words that are absent in *COBAm2*. Naturally, the definitions of these new words are exclusive to *COBAm3*. In the sampled pages, the following new headwords were identified in *COBAm3*: **bat mitzvah**, **batswoman**, **dyspraxia**, **flexitarian**, **flight risk**, **Jerusalem artichoke**, **open access (open-access)**, **pox**, **PPE**, **signalwoman**, and **signboard**.

The newly included headwords in *COBAm3* reflect recent changes in society, such as technological advancements, lifestyle changes, and, most notably, the impact of the COVID-19 pandemic.

Among the newly coined terms that have emerged with recent technological advancements and are included in *COBAm3* are **AI**, **block-chain**, **smartphone**, and **wearable**. We observed some differences between the definitions of these newly added words in *COBAm3* and those in *COB10*, such as in the definitions of **artificial intelligence**:

**Artificial intelligence** is a type of computer technology concerned with making machines work in an intelligent way, similar to the way that the human mind works. (*COBAm3*)

**Artificial intelligence** is a type of computer technology concerned with making machines work in an intelligent way, similar to the way that the human mind works. The abbreviation **AI** is also used. (*COB10*, an underline added)

The definition in *COB10* is nearly the same as that in *COBAm3*, but the sentence underlined in the extract is an addition to the definition in *COB10*. As this example illustrates, not all definitions in *COBAm3* match those in *COB10*. However, the definitions of **smartphone**, **blockchain**, and **wearable** in *COBAm3* were all the same as those in *COB10*.

One newly coined term that has emerged alongside recent changes in society is **cryptocurrency**. With the definition of this term, unlike that of **AI**, an additional sentence has been appended at the end of the definition in *COBAm3* that was not found in the definition of *COB10*. Compare:

**Cryptocurrency** is a kind of digital currency that is created and exchanged using cryptography. It is often shortened to **crypto**. (*COBAm3*, an underline added)

**Cryptocurrency** is a kind of digital currency that is created and exchanged using cryptography. (*COB10*)

However, in many neologisms that are reflective of recent changes in society, numerous instances, including **cancel culture**, **carbon footprint**, **deepfake**, **FOMO**, **global village**, **influencer**, and **meme**, show identical definitions between *COBAm3* and *COB10*.

Another category of novel words reflects the influence of the COVID-19 pandemic. While some terms were newly coined, others were originally limited to technical uses among medical professionals but became widely used and generalized within the context of COVID-19. The following are certain examples found in *COBAm3* and *COB10*: **Contact tracing**, **isolation**, **pandemic**, **PPE**, **quarantine**, **social distance** (*COB10*)/**distancing** (*COBAm3*), and **super-spreader**. However, it is not certain that these terms were included in *COBAm3* and *COB10* as a result of the influence of COVID-19, in that neither their definitions nor the examples provided for them explicitly reference COVID-19.

The definitions of **hybrid** and **remote** in *hybrid working* and *remote working* reflect changes in work practices brought about by the COVID-

19 pandemic. However, these new meanings are included in *COB10*, not in *COBAm3*. These examples also indicate that *COBAm3* did not simply adopt its definitions from *COB10*.

#### 4.2.2. Effects of DEI

Textbooks, academic papers, and other resources are often intended to incorporate a range of diverse voices and equitable perspectives to promote inclusive education. Dictionaries that are tailored for EFL learners are often included in these educational materials. They are among the teaching materials playing a key role in promoting diversity, equity, and inclusion (DEI) within education, catering to learners from diverse linguistic and cultural backgrounds. The definitions of some headwords in *COBAm3* likely have had their definitions revised in consideration of DEI principles. The revision of the definition for **flesh-colored** could be an example:

Something that is **flesh-colored** is the color of any shade of a person's skin. (*COBAm3*)

**Flesh-colored** is yellowish pink in color. (*COBAm2*)

This could be an example of the evolution of language that reflects the values of inclusivity. By defining **flesh-colored** as encompassing all skin tones rather than limiting it to a narrow interpretation, *COBAm3* acknowledges the diversity of humanity and challenges outdated perspectives. This shift could help foster representation and reinforce the idea that no single skin color is default. These thoughtful changes of definition can help make language a tool for inclusion, not exclusion. This approach aligns with the principles of DEI, as it promotes understanding and respect across cultures and identities.

Although this change may seem small, it could be an example of how modern lexicography is evolving to uphold inclusivity and fairness. Language can bear within its implicit bias. Revisions of definitions that take DEI into consideration could help ensure that words be used to reflect respect and equality for all, regardless of their attributes or back-

grounds.

#### 4.2.3. Effects of gender identity

Social awareness of inclusiveness and gender identity, which are the key principles of DEI, are growing. The shift away from the use of ‘he’ as a default pronoun to more inclusive options such as ‘he or she’ or the singular ‘they’ acknowledges and respects the diversity of gender identities. In the example given below, the underlined sentence in the definition of the second sense of **he** in *COBAm2* was revised for *COBAm3* to show that the use of ‘he or she’ or the singular ‘they’ is more prevalent in present-day society. It is noteworthy that the subject of the sentence has been revised to ‘Many people’ in *COBAm3*.

##### **he** PRON-SING

[2] In written English, **he** is sometimes used to refer to a person without saying whether that person is a man or a woman. Some people dislike this use and prefer to use ‘he or she’ or ‘they.’ (*COBAm2*, an underline added)

[2] In written English, **he** is sometimes used to refer to a person without saying whether that person is a man or a woman. Many people prefer to use ‘he or she’ or ‘they’ as an alternative to ‘he.’ (*COBAm3*, an underline added)

The use of singular ‘they’ is explained as follows in *COBAm3*:

##### **They** PRON-PLURAL

[2] You use **they** instead of ‘he or she’ to refer to a person without saying whether that person is a man or a woman. □ *The teacher is not responsible for the student’s success or failure. They are only there to help the student learn.* (*COBAm3*)

In the example given above, the teacher is referred to as ‘they’. *COBAm2* states that some people consider this use of singular ‘they’ to be incorrect. However, since *COBAm3* does not include such a note, it can be inferred that this usage has become widely accepted. The revised definition of ‘he’ and the expanded use of ‘they’ could be considered as part of the societal movement toward greater inclusivity. Singular ‘they’, in

particular, has recently gained additional prominence in our society because it is gender-neutral and does not assume or impose any binary gender identity on the person that is being referred to. This seemingly minor linguistic adjustment can significantly enhance the sense of belonging that is experienced by nonbinary individuals and those who favor gender-neutral language.

It is also noteworthy that the use of ‘she’ or ‘her’ is consistently avoided in *COBAm3*’s definitions. The changes in attitudes toward the use of gendered terms such as ‘she’ or ‘her’ could reflect broader social changes, particularly with respect to gender inclusivity. Many at present prefer to use gender-neutral terms and pronouns to promote inclusivity and avoid stereotypes or assumptions about an individual’s gender identity. In line with this, the second sense of **she** is revised as follows in *COBAm3*. Compare:

**she** PRON-SING

[2] Some writers may use **she** to refer to a person who is not identified as either male or female. They do this because they wish to avoid using the pronoun ‘he’ all the time. Some people dislike this use and prefer to use ‘he or she’ or ‘they.’ (*COBAm2*, an underline added)

[2] Some writers may use **she** to refer to a person who is not identified as either male or female. Many people prefer to use ‘he or she’ or ‘they’ as an alternative to ‘she.’ (*COBAm3*, an underline added)

As was the case in the definition for **he**, the subject of the last sentence is revised to ‘Many people’ in *COBAm3*. This shows that the gender-neutral use of **she** had not attracted sufficient approval, and the use of ‘he or she’ or singular ‘they’ became more prevalent.

The same revision has been done to the definitions of **him** and **her** in *COBAm3*. See the following:

**him** PRON-SING

[2] In written English, **him** is sometimes used to refer to a person without saying whether that person is a man or a woman. Some people dislike this use and prefer to use ‘him or her’ or ‘them.’

(*COBAm2*, an underline added)

[2] In written English, **him** is sometimes used to refer to a person without saying whether that person is a man or a woman. Many people prefer to use ‘him or her’ or ‘them’ as an alternative to ‘him.’ (*COBAm3*, an underline added)

**her** PRON-SING

[2] In written English, **her** is sometimes used to refer to a person without saying whether that person is a man or a woman. Some people dislike this use and prefer to use ‘him or her’ or ‘them.’ (*COBAm2*, an underline added)

[2] In written English, **her** is sometimes used to refer to a person without saying whether that person is a man or a woman. Many people prefer to use ‘him or her’ or ‘them’ as an alternative to ‘her.’ (*COBAm3*, an underline added)

The definitions of **himself** and **herself** in *COBAm3* have been revised in the same way.

In response to this change in society, words such as ‘person’, and ‘individual’, also seem to be favored in place of gender-conscious pronouns such as ‘she’ or ‘her’ when the gender of the person who is referred to does not require specification. The following revision of the definition of **surrogacy** in *COBAm3* aligns with this trend:

**Surrogacy** is an arrangement by which a woman gives birth to a baby on behalf of a woman who is physically unable to have babies herself, and then gives the baby to her. (*COBAm2*, underlines added)

The underlined noun and pronouns were revised as follows in *COBAm3*:

**Surrogacy** is an arrangement by which a woman gives birth to a baby on behalf of someone who is physically unable to have babies themselves, and then gives the baby to that person. (*COBAm3*, underlines added)

We immediately notice that ‘a woman’ is replaced with ‘someone’, ‘herself’

is replaced with ‘themselves’, and ‘her’ is replaced with ‘that person’ in *COBAm3*.

The tendency to avoid any specification of a general person’s gender is not limited to any use of personal pronouns but can also be observed in nouns that were once historically differentiated based on gender. For example, for the definition of **understudy**, the noun ‘actress’ is eliminated with in *COBAm3*. Compare:

An actor’s **understudy** is the person who has learned their part in a play and can act the part if the actor or actress is ill. (*COBAm2*)

An actor’s **understudy** is the person who has learned their part in a play and can act the part if the actor is ill. (*COBAm3*)

The inclusive use of the noun ‘actor’ in *COBAm3* is explained in the definition of the headword **actress** in *COBAm3* as follows:

An **actress** is a woman whose job is acting in plays or films. Some women who act prefer to be called ‘actors’ rather than ‘actresses.’ (*COBAm3*, an underline added)

The underlined sentence was added to the definition in *COBAm3*, and the information it provides is reflected in the definition for **understudy**.

Thus, *COBAm3* has revised its word definitions in a way that reflects contemporary awareness of gender identity and the use of gender-neutral terms wherever possible to avoid specifying gender. Lexicographers’ efforts to reduce the use of binary gendered language in word definitions appear to align with DEI principles.

#### 4.3. Cross-references

The sample pages of *COBAm3* revealed that under some headwords, new cross-references are added that were not present in *COBAm2*. These cross-references are incorporated at the end of each entry. These refer to related sections, such as ‘Vocabulary in Context’ and ‘Visual Dictionaries’ (VDs). In *COBAm2*, it was clearly indicated in a cross-reference the key features (such as a ‘Word Web’ or a ‘Picture Dictionary’)

that the given headword referred to, increasing its user-friendliness. However, in *COBAm3*, the VD is generally referred to as the ‘color supplement’, and there is no specific indication of which headword’s VD should be referenced. Thus, users must locate the relevant VD on their own, a user-unfriendly requirement.

As an example of the unclarity regarding which VD is to be referred to, let us consider the headword **abdomen**. In *COBAm3*, a cross-reference not present in *COBAm2* is added following the definition of this noun. However, the cross-reference merely directs the user to ‘See also color supplement’, without specifying which VD should be consulted. From the meaning of **abdomen**, one might think to refer to the VDs that depict the human body, but those VDs only illustrate various human organs, having no diagram related to **the abdomen**.

As a matter of fact, the word **abdomen** appears in a VD titled INSECT, showing the abdomen of a grasshopper, as illustrated in the diagram. This VD is within a section entitled ‘PARTS OF’ in the middle matter, including VDs identifying various components of objects or living beings that are illustrated within them. To locate the appropriate VD for the **abdomen**, users must first exclude the ‘TYPES OF’ section (categorizing types of entities) and focus on the ‘PARTS OF’ section of the VDs. Then, they must further narrow their search, avoiding diagrams entitled BIRD or FISH (which one might expect could contain ‘abdomen’ in their diagrams) to select the diagram of the INSECT. This convoluted process leaves users to fall back on trial-and-error to locate intended VDs.

The VDs in the middle matter of *COBAm3* are identical to those in *COB10*, as are the cross-references. Thus, it might be thought that *COBAm3* would have adopted the cross-references in *COB10*, inheriting the search-related challenges accompanying them.

For an illustration of the issues with cross-references in *COBAm3*, another example can be found from the headword **bat**. For this entry, the final word sense is a cross-reference stating, ‘See also color supplement’. This cross-reference does not exist in *COBAm2*. This is because, in *COBAm2*, the Picture Dictionary (PD) and Word Web for **bat** per-

tain to the animal *bat*. The PD in *COBAm2*, for instance, is associated with the headword **cave**, and it depicts bats living in a cave. This PD is no longer present in the VDs in *COBAm3*.

In *COBAm3*, the PDs are referred to as Visual Dictionaries (VDs), in alignment with *COB10*, and their content is identical to those in *COB10*. In addition, some of the PDs do not have their counterpart VDs in *COBAm3*. For example, there is no VD associated with **bat** in the sense of the animal *bat*. The VD associated with **bat** in *COBAm3* pertains to the piece of sports equipment used to hit a ball, such as in baseball and cricket.

The VD linked to **the bat** in *COBAm3* is found in the ‘TYPES OF’ section under the title BATS, STICKS, AND RACKETS, where photographs of baseball and cricket bats are provided. To access this VD, users must first locate the ‘TYPES OF’ section among all the VDs and then identify the VD that is linked to the headword **bat**. This process is likely more challenging than one might expect because COBUILD dictionaries bring together multiple meanings of a polysemous word for a single headword. When attempting to locate the relevant VD using the cross-references at the end of the entries, users will initially be unable to discern whether the VD is in reference to the animal or the piece of sports equipment. This example demonstrates that the cross-references in *COBAm3* have a problematic approach to user-friendliness, making it more difficult for users to navigate the dictionary and identify intended references.

*COBAm3* contains another type of cross-reference, linked to the feature boxes. In the entry for the headword **champion** in *COBAm3*, the fourth word sense forms a cross-reference, stating, ‘See feature box at **competition**’. This feature box refers to a column titled ‘Vocabulary in Context’ (VC) located near the headword **competition**. The VC columns are part of the framework that is adopted from *COB10*. They are, in fact, identical to those found in *COB10*. Knowing that a feature box refers to a VC column makes it easier to locate the specific VC that is associated with a given headword, as it is clearly indicated in the cross-references. Thus, the references to VCs are superior in terms of

user-friendliness to references to VDs.

#### 4.4. Conclusion

*COBAm3* closely resembles *COBAm2* and *COB10* in its structure, but it incorporates a mixture of continuity and updates. While its ‘Guide to dictionary entries’ seems to be derived from *COB10*, the ‘Guide to definitions’ largely mirrors *COBAm2*, with minor changes. *COBAm3* introduces new definitions that reflect social changes, including technological advancements, changes in lifestyle, and the COVID-19 pandemic, which align with many definitions of *COB10*.

*COBAm3* likewise demonstrates a commitment to the principles of DEI, revising definitions, adopting gender-neutral language, and emphasizing inclusivity. For instance, **flesh-colored** is redefined to refer to all skin tones, and binary gender references are avoided in headwords such as **surrogacy**.

Cross-references in *COBAm3* present inconsistencies in user-friendliness. References to VDs often lack clarity, requiring users to rely on trial-and-error to locate the relevant diagrams they are searching for. Conversely, references to feature boxes such as VCs are clear and accessible, highlighting variations in usability across different tools in *COBAm3*.

In summary, the definitions and dictionary features of *COBAm3* are more similar to those in *COB10* than to those in *COBAm2*. This similarity may be influenced by the fact that *COBAm3* was published in the same year as *COB10*. This is also likely to be rooted in the utilization of the Collins Corpus, as in *COB10*. Given its resemblance to *COB10*, *COBAm3* exhibits similar issues to it, such as those related to cross-references, suggesting that there is room for improvement.

(Section 4 by Takahashi)

#### 5. Examples

In this section, we review examples in *COBAm3*. *COBAm2* used examples based on the Collins Corpus with 4.5 billion words, and *COBAm3* uses the Collins Corpus with over 20 billion words, which is about 4.5 times more. On the other hand, *COB7*, *COB8*, *COB9*, and

*COB10* used the Collins Corpus with 4.5 billion words. We compared mainly the examples used under the same range of headwords in *COBAm2* and *COBAm3* here.

### 5.1. Comparison of examples in *COBAm2*, *COBAm3*, *COB8*, and *COB10*

At first, we compared the numbers of examples used under the same range of headwords in *COBAm2* (2016) and *COBAm3* (2023), as well as *COB8* (2014) and *COB10* (2023), which were published about the same times. Table 5.1. shows the results of the survey of examples in these

Table 5.1. Comparison of the number of examples among *COBAm2*, *COBAm3*, *COB8*, and *COB10*

	page	entry words	<i>COBAm2</i>	<i>COBAm3</i>	<i>COB8</i>	<i>COB10</i>
1	1-2	A - abode	68	67	88	91
2	101-102	basket - battery	74	76	81	79
3	201-202	chained - chance	56	56	72	78
4	301-302	crony - crossword	75	75	97	88
5	401-402	dying - early	77	76	100	100
6	501-502	flesh-colored - flock	79	82	110	114
7	601-602	haven't - headline	73	73	87	87
8	701-702	jailhouse - jet	76	75	93	94
9	801-802	mechanize - meet	72	73	109	108
10	901-902	open - operative	115	119	148	158
11	1001-1002	power - praise	73	74	96	99
12	1101-1102	resplendent - restoration	74	74	72	72
13	1201-1202	siege - signposted	77	78	96	96
14	1301-1302	surreptitious - sustain	72	72	96	96
15	1401-1402	underpopulated - undiscriminating	83	84	112	113
total of samples			1,144	1,154	1,457	1,473
total text pages			1,539	1,499	1,828	1,736
sample pages/total text pages			1.95%	2.00%	1.64%	1.73%

editions. *COBAm2* has 1,539 text pages (A–Z) and *COBAm3* has 1,499 text pages. *COBAm3* has fewer text pages than *COBAm2*, but we noticed a little increase in the total number of examples from *COBAm2* to *COBAm3*. In the same way, *COB10* has fewer text pages than *COB8*, but we noticed a little increase in the total number of examples from *COB8* to *COB10*.<sup>1)</sup>

### 5.2. Changes to examples from *COBAm2* to *COBAm3*

We examined the changes from *COBAm2* to *COBAm3* in detail here. The sample pages are about 2% of the total text pages of *COBAm2* and *COBAm3*. Table 5.2. shows the results of the survey of examples con-

Table 5.2. Changes to the Number of Examples from *COBAm2* to *COBAm3*

	Pages	Headwords	<i>COBAm2</i>	Added	Deleted	<i>COBAm3</i>	Replaced
1	1 - 2	A - abode	68	0	-1	67	2
2	101 - 102	basket - battery	74	+3	-2	76	0
3	201 - 202	chained - chance	56	0	0	56	1
4	301 - 302	crony - crossword	75	+2	-2	75	1
5	401 - 402	dying - early	77	0	-1	76	1
6	501 - 502	flesh-colored - flock	79	+3	0	82	1
7	601 - 602	haven't - headline	73	+3	-1	74	2
8	701 - 702	jailhouse - jet	76	+2	-1	75	2
9	801 - 802	mechanize - meet	72	0	0	73	2
10	901 - 902	open - operative	115	+5	-1	119	4
11	1001 - 1002	power - praise	73	+1	0	74	1
12	1101 - 1102	resplendent - resto- ration	74	0	0	74	0
13	1201 - 1202	siege - signposted	77	+3	-2	78	1
14	1301 - 1302	surreptitious - sustain	72	0	0	72	3
15	1401 - 1402	underpopulated - indiscriminating	83	+1	0	84	1
Total samples in the sampling pages			1,144	+23	-11	1,154	22
Total text pages (A - Z)			1,539			1,499	

ducted on the same range of headwords in both editions. In about 30 sample pages out of the 1,499 text pages in *COBAm3*, 1,154 examples are found: 23 examples have been added, three deleted, and 30 replaced or partially modified in *COBAm3*.

### 5.3. Additions of examples

In the sampling pages, 15 examples have been added to *COBAm2*. Among these additions, there are 12 new examples added with new headwords, and three new examples to the old headwords in *COBAm3*.

#### 5.3.1. New examples with new headwords

The following 12 examples have been added to the 10 new headwords in *COBAm3*:

(1) **batswoman**

□ *She crafted a well-earned reputation as a batswoman and wicketkeeper during her heyday.*

(2) **flexitarian**

□ *I stopped eating meat as a teenager, but I now call myself a flexitarian.*

(3) **flight risk**

□ *US authorities are opposing his request for bail, saying that his is a flight risk.*

(4) **flipped**

□ *Stanforth said he was planning to take the flipped classroom several stages further.*

(5) **jeggings**

□ *Flats work fabulously with this season's jeggings and a long line vest.*

(6) **open access**

□ *There's 24-hour open access to computers in the library.*

□ *Its faculty buildings are well equipped, with banks of open-access computers.*

(7) **open mic, open mike**

□ *...an open mic night.*

□ *Are you brave enough to sing at the café's Thursday night open mike?*

(8) **PPE**

□ *Workers require guidance regarding what PPE can be removed, and*

*how, for a break.*

(9) **signboard**

☐ *The signboard outside the factory read, 'baby milk plant.'*

(10) **understaffing**

☐ *Understaffing and overcrowding led to an increase in prison violence.*

### 5.3.2. New examples added to the old headwords

There are three examples added to the old headwords including one to a new definition:

(1) **bath tub**

☐ *...a gigantic marble bathtub.*

(2) **head**

☐ *...the youngest, the most ambitious professor ever to head the department.*

(3) **medium term** 2 N-ING When you talk about what happens in **the medium term**, you are talking about what happens over a few months or years, either or after a particular event. ☐ *She risks losing her salary in the medium-term and damaging her long-term career prospects.*

### 5.4. Deleted examples

In the sampling pages, two examples were deleted from *COBAm2*.

(1) **bathing trunks**

☐ *He came down a few minutes later in his bathing trunks.*

(2) **dyslexic**

☐ *Many dyslexics have above average intelligence.*

### 5.5. Replacement of examples

There are 31 cases of replacement or partial modification of examples in *COBAm3*. Many illustrative examples have been replaced with shorter or simpler examples without proper nouns. Some have been replaced with more appropriate, less controversial or discriminatory ones. (Underlined parts in the examples, our modification, show the parts where changes occurred).

(1) **abeyance**

□ The threat is, at the least, in abeyance.

← □ The Russian threat is, at the least, in abeyance.

(2) **able-bodied**

□ The gym can be used by both able-bodied and disabled people.

← □ No doubt such robots would be very useful in the homes of the able-bodied, too.

(3) **abode**

□ 30 percent of the clinic's patients are of no fixed abode.

← □ 30 percent of psychiatric hospital beds are occupied by people of no fixed abode.

(4) **batter**

□ The gang battered him unconscious.

← □ A karate expert battered a man to death.

(5) **championship**

□ ...the world showjumping [sic] championship.

← □ ...the world chess championship.

(6) **crony**

□ He played a round of golf with his golf cronies.

← □ ...lunchtime drinking sessions with his business cronies.

(7) **cropped**

□ He wore a black t-shirt and cropped jeans.

← □ ...women athletes wear cropped tops and tight shorts.

(8) **crossover**

□ I told her the crossover from acting to singing is easier than singing to acting.

← □ I told her the crossover from actress to singer is easier than singer to actress.<sup>2)</sup>

(9) **dynamic**

□ South Asia was known as the most dynamic economic region in the world.

← □ South Asia continues to be the most dynamic economic region in the world.

(10) **flock**

□ The criticisms will not stop people flocking to see the movie.

← □ The criticisms will not stop people flocking to see the film.

(11) **he**

- ☐ *The teacher should encourage the child to proceed as far as he can.*  
← ☐ *The teacher should encourage the child to proceed as far as he can, and when he is stuck, ask for help.*

(12) **head**

- ☐ *These problems came to a head in September when five of the staff were fired.*  
← ☐ *These problems came to a head in September when five of the station's journalists were fired.*

(13) **headhunt**

- ☐ *She was soon headhunted by an investment bank.*  
← ☐ *He was headhunted by Barkers last October to build an advertising team.*

(14) **headline**

- ☐ *The front cover of your July 28 issue has the headline: "Time for action."*  
← ☐ *The Sydney Morning Herald carried the headline: "Sorry Ma'am, Most Australians Want a Republic."*

(15) **Jane Doe**

- ☐ *The new action, filed in the Manhattan federal court, is on behalf of a woman named only as Jane Doe.*  
← ☐ *The patient, referred to in the indictment as "Jane Doe," knew of Mr. Weill's secret plan.*

(16) **jeer**

- ☐ *Marchers jeered at passers-by, but there was no violence, nor any arrests.*  
← ☐ *Marchers jeered at white passers-by, but there was no violence, nor any arrests.*

(17) **medalist**

- ☐ *...the gold medalist.*  
← ☐ *...the Olympic gold medalists.*

(18) **medium**

- ☐ *He says he has been talking to his late father through a medium.*  
← ☐ *Bruce Willis says he has been talking to his dead brother through a medium.*

(19) ① **open**

- ☐ *...dressing informally in open shirt and cowboy boots.*  
← ☐ *The top can be worn buttoned up or open over a T-shirt.*

(20) ② **open**

- ☐ *The rebels have opened the road to the capital.*

- ← □ The rebels have opened the road from Monrovia to the Ivory Coast.
- (21) ④ **open**  
 □ He said he wanted to see how the country was opening up to the world.  
 ← □ He said he wanted to see how Albania was opening up to the world.
- (22) **operating system**  
 □ Microsoft's Windows operating system.  
 ← □ Microsoft's Windows NT operating system.
- (23) **operation**  
 □ The company's electronics operation employs around 5,000 people.  
 ← □ Thorn's electronics operation employs around 5,000 people.
- (24) **praise**  
 □ All the guests are full of praise for the staff and service they received.  
 ← □ All the ladies are full of praise for the staff and service they received.
- (25) **sign**  
 □ They signed a new quarterback.  
 ← □ The Minnesota Vikings signed Herschel Walker from the Dallas Cowboys.
- (26) **signer**  
 □ We provide signers for deaf people and readers for blind people.  
 ← □ We provide signers for deaf people and readers for the blind.
- (27) **significant**  
 □ A small, but significant number of 11-year-olds are illiterate.  
 ← □ Most 11-year-olds are not encouraged to develop reading skills; a small but significant number are illiterate.
- (28) **surrogate**  
 □ Martin had become his surrogate son.  
 ← □ Martin had become Howard Cosell's surrogate son.
- (29) **susceptible**  
 □ Young people are often seen as being the most susceptible to advertisements.  
 ← □ Young people are the most susceptible to advertisements.
- (30) **suspended animation**  
 □ She lay in a state of suspended animation, waiting for dawn, when she would rise.  
 ← □ She lay in a state of suspended animation, waiting for dawnlight.
- (31) **underworld**  
 □ ...an underworld of gangs, drugs and violence

← □...a Spanish Harlem underworld of gangs, drugs and violence

## 5.6. Conclusion

In this section we compared the examples used under the same range of headwords in *COBAm2* (2016) and *COBAm3* (2023), as well as *COB8* (2014) and *COB10* (2023), which were published about the same times. *COBAm3* has fewer text pages than *COBAm2*, but we noticed a little increase in the total number of examples from *COBAm2* to *COBAm3*. In the same way, *COB10* has fewer text pages than *COB8*, but we noticed a little increase in the total number of examples from *COB8* to *COB10*. In the sample pages in *COBAm3*, 1,154 examples are found: 15 examples have been added, three deleted from *COBAm2*, and 30 examples replaced or partially modified in *COBAm3*. Some have been replaced with less controversial or discriminatory ones. Many of these illustrative examples have been replaced with shorter or simpler examples without proper nouns. Shorter examples have enabled the new edition to have more examples in fewer pages in total, and examples without proper nouns have made it easier for many users to understand.

(Section 5 by Ikeda)

## 6. Featured Columns

### 6.1. Vocabulary in Context

Kokawa et al. (2024) highlighted four revisions made to ‘Vocabulary in Context’ in *COB10* to enhance visibility, ensure consistency with headwords, and reflect a more diverse worldview. These revisions include updates such as redefining **nuclear family** and recognizing the equal importance of paternity and maternity leave. However, these changes have not been incorporated into *COBAm3*, suggesting that *COBAm3* does not refer to *COB10* in terms of ‘Vocabulary in Context’. From an educational standpoint, the discrepancy between *COB10* and *COBAm3*, in which the former reflects these revisions, whereas the latter does not, raises concerns about fairness and consistency in representation.

In terms of content, each column in *COBAm3* accounts for differ-

ences between American and British vocabulary and cultural contexts, which are pedagogically beneficial for users. For example, in the column **DIY**, *COBAm3* uses ‘drop cloths’ instead of ‘dust sheets’, the latter being the term used in *COB10*. Similarly, in the column on **Poverty**, *COBAm3* refers to homelessness as ‘living in a shelter or as a squatter’, while *COB10* describes it as ‘living in a squat or a shelter’, reflecting the British-specific usage of ‘a squat’ to mean an empty building illegally occupied by people. From a cultural perspective, the **Football** column in *COBAm3* explains that ‘football’ in the United States refers to American football and also includes an explanation of soccer. In contrast, *COB10* only provides an explanation of football without mentioning American football.

Regarding additional online resources related to ‘Vocabulary in Context’, those available for *COB10* can also be accessed by *COBAm3* users, though they are not officially designed for *COBAm3*. Notably, there are no ‘Vocabulary in Context’ resources listed under the *COBAm3* category, which instead consists of Collocations, Frequency, Geographical Differences, Homographs, Prefixes, Suffixes, Verb Patterns, Visual Dictionary, Vocabulary, Word Families, and Word History. The unequal availability of resources may contribute to disparities in learning opportunities between the users of the two dictionaries.

Thus, while the latest revisions made in *COB10* are not available in *COBAm3*, it provides culturally appropriate and customized content, which is advantageous for its users. In future editions, ensuring equal access to resources across both dictionaries will help ensure fairness and consistency among users.

(Section 6.1. by Aoki)

## 6.2. Visual Dictionaries

### 6.2.1. Picture Dictionaries in *COBAm2* and Visual Dictionaries in *COB10* and *COBAm3*

In *COBAm2*, featured columns named ‘Picture Dictionary’ panels were distributed all throughout the A-Z dictionary text of the dictionary. There were two types of Picture Dictionary, namely, ‘variety’ types (see Kokawa et al. 2020:74–77), which the COBUILD dictionaries

refer to as ‘types of’ panels, and the other is ‘constituent’ type (*ibid.*), that the dictionaries call ‘parts of’ panels. (*ibid.*) They were all allocated near the entry of the panel title, as were all other featured columns in the dictionary.

There are 46 Picture Dictionary panels presented in *COBAm2*: 27 constituent types (**age, Arctic, area, barn, baseball, basketball, bed, body, cave, core, desert, face, family, foot, football, globe, golf, gymnastics, hand, house, landforms, mountain, river, sign language, skateboarding, soccer** and **tennis**) and 19 variety types (**brass, bread, clouds, color, cook, cut, dessert, dish, egg, hand, herb, jewelry, keyboard, percussion, shapes, solids, strings, volume** and **woodwinds**). We found eight sports panels of the constituent type (**baseball, basketball, football, golf, gymnastics, skateboarding, soccer** and **tennis**) and the panel introducing names of musical instruments were conspicuous in the variety type (**brass, keyboard, percussion, strings** and **woodwinds**).

In *COBAm3*, the general name for the panels presented as the full-color 16-page supplement between L and M (pp. 774 and 775) A-Z dictionary text is ‘Visual Dictionary’ as in the case of *COB10*. However, as the change of its name from Picture Dictionary to Visual Dictionary suggests, it’s not merely that Picture Dictionary panels are assembled as the center supplement. Though the total number of Picture/Visual Dictionary panels are nearly the same in *COBAm2* and *COBAm3* (about 45), the constituting panels in the two dictionaries are quite different. Visual Dictionary panels in *COBAm3* seem to be based on those in *COB9*, as noted below. Now, we will focus mainly on the comparison of Visual Dictionaries in *COBAm3* and *COB10*, and briefly discuss the merit of illustration presentation in the paper editions of a contemporary dictionary.

### 6.2.2. ‘Types of’ type Visual Dictionary panels in *COBAm3* and *COB10*

There were altogether 23 ‘types of’ panels in *COBAm3* and 21 panels in *COB10*. The summary of ‘types of’ Visual Dictionary panels in

*COBAm3* and *COB10* are tabulated in Table 6.2.1.:

We should note that *COB9* already had the same format of Visual Dictionaries as *COB10* (and *COBAm3*), although panels in *COB9* were presented in the A-Z text of the dictionary, and individual Visual Dictionary panels in *COB10* are just an update of the *COB9* Visual Dictionaries presented in different parts in the A-Z text of the dictionary (see Kokawa et al. 2024:50–53). We should also pay attention to the fact that *COB9* had two different panels named **DIY toolkit** and **woodworking tools** containing 9 and 12 pictures respectively, as well as the panel **ships** with the same 12 pictures found in *COBAm3* **SHIPS** panel. What we can surmise is that the *COB10* and *COBAm3* teams developed their respective Visual Dictionary panels by updating *COB9* Visual Dictionary panels independently from each other. Each Visual Dictionary panel in *COB9* had basically the same number of pictures in those in *COB10*, so presumably the *COBAm3* team arranged the *COB9* panels for their American version by omitting some items that may not fit well in the new edition. For instance, an American lexicographer for *COBAm3* may have called what is designated by the picture **toilet bag** a toilet pouch by which logic it is not considered a ‘bag’ any more. Also, such trademark items as **Stanley knife**<sup>®</sup> and **Strimmer**<sup>®</sup> may not be popularly used in the United States.

That the *COBAm3* team retained the **SHIPS** panel found in *COBAm2* was a helpful judgement for EFL users, as the **BOATS** panels in *COBAm3* and *COB10* represent mainly small types of vessels.

### 6.2.3. ‘Parts of’ type Visual Dictionary panels in *COBAm3* and *COB10*

Listed below in Table 6.2.2. are the ‘Parts of’ type Visual Dictionary panels found in *COBAm3* and *COB10*.

Both *COB10* and *COBAm3* present 23 ‘Parts of’-type Visual Dictionary panels and, in all cases employed the same pictures as each other. In 20 panels out of 23 they applied the same captions as well. The panel title, BIKE in *COBAm3* could well have been something like BICYCLE/BIKE instead of just BIKE but that may not be a major concern.

Table 6.2.1. 'Types of' type Visual Dictionary Panels in *COB10* and *COBAm3*

Panel title	No. of pics in <i>COB10</i>	No. of pics in <i>COBAm3</i>	Difference of constituting pictures ( <i>COB10</i> — <i>COBAm3</i> respectively) (-) = found only in <i>COB10</i> , (+) = found only in <i>COBAm3</i>
ARCHITECTURE	12	12	no difference
BAGS	12	11	toilet bag (-) <b>plastic bag--carrier bag</b> different picture for <b>saddle bag</b>
BATS ( <i>COB10</i> ) BATS, STICKS AND RACKETS ( <i>COBAm3</i> )	8	8	no difference
BEDS	12	11	(-) couchette <b>camp bed -cot</b>
BIKES	9	8	(+) e-bike, (-) penny-farthing, (-) quad bike
BOATS	12	12	<b>inflatable—dingy</b>
CARS	12	12	<b>estate wagon—station wagon</b> <b>people carrier—minivan</b>
CHAIRS	12	11	<b>banquette—bench</b> <b>deckchair—deck chair</b> (-) <b>bar stool</b>
CLOCKS	7	7	the same constituent pictures presented in different order (layout)
CUPS AND GLASSES	9	9	no difference
DEVICES	12	12	no difference
DIY	21	8	8 pictures in DIY panel and 12 pictures in WOOD-WORKING TOOLS (see below) in <i>COBAm3</i> (-) <b>Stanley knife®</b>
FOOTWEAR	12	10	<b>bregues, courtshoes, trainers and wellingtons (wellies)</b> ( <i>COB10</i> ) <b>oxfords, pumps, rubber boots, sneakers</b> ( <i>COBAm3</i> ) respectively
GARDENING TOOLS	12	11	<b>secateurs</b> ( <i>COB10</i> ), <b>pruning shears</b> (-) <b>Strimmer®</b>
GEMSTONES	11	11	no difference
HATS	12	10	<b>cloth cap/flat cap</b> renamed to <b>cloth cap</b> <b>balaclava</b> renamed into <b>ski mask</b> (-) <b>fascinator</b> , (-) <b>trilby</b>
HOMES	11	11	<b>flat</b> renamed into <b>apartment</b> with picture replaced <b>mobile home</b> renamed into <b>manufactured home</b>
KITCHEN UTENSILS	12	11	<b>tin opener</b> renamed into <b>can opener</b> (-) <b>fish slice</b>
METALS	12	12	<b>aluminium</b> renamed into <b>aluminum</b>

NUTS	7	7	singular captions are all rendered into plural (e.g. <b>almond</b> (COB10)— <b>almonds</b> (COBAm3))
ROCKS	11	11	no difference
SHIPS	no panel	12	panel found only in COBAm3
WOODWORKING TOOLS	no panel	12	<b>wrench, spanner, spirit level, vice, washer</b> in the DIY panel in COB10 recaptioned into <b>monkey wrench, wrench, level, vise and washers</b> in COBAm3 respectively

One thing that is noticeable is the difference in the names of exterior car parts between the American and British varieties of English. It is certainly the case that *COBAm3* is an American dictionary, and therefore American expressions should be listed first. However, it is still worthwhile to present alternatives. It can be presumed that few learners of English would buy both dictionaries the size of *COBAm3* and *COB10*, and even if a learner chooses one of the major varieties of English as the one to learn, access to other usages is still of value. In *COBAm3* (and in *COB10* as well), the dictionary provides a wide variety of information regarding other varieties of English than they mainly discuss in the A-Z part of English, such as In BRIT, use .... Would not such provision of information be desirable in Visual Dictionary supplement (not only in the ‘Parts of’ panels but also in ‘Types of’) as well? It may make the panels a little more cluttered, but would provide learners with the opportunity to compare and learn expressions from the major varieties of English in this context.

The Collins online dictionary that we discussed in Section 1 above is evolving constantly, and now in many cases presents images of the word looked up if the word is a concrete noun. Thus, with regard to images of individual lexical items, Web search including Collins online may provide the reader with more than enough information if they have a device connected to the Internet. The advantage of a Visual Dictionary as found in *COBAm3* and *COB10* may be in its collective nature. The Visual Dictionaries in both are now somehow disconnected from the A-Z entries. They used to be near the entries to which they were related in the dictionary text, but now they are independent from the relevant entries and assembled in the center supplement. However, it is also true

Table 6.2.2. 'Parts of' type Visual Dictionary Panels in *COB10* and *COBAm3*

Panel title	No. of captions in <i>COB10</i>	No. of captions in <i>COBAm3</i>	Difference of captions ( <i>COB10</i> — <i>COBAm3</i> )
COMPUTER	4	4	<captions identical>
AEROPLANE ( <i>COB10</i> )/ AIRPLANE ( <i>COBAm3</i> )	9	9	<captions identical>
BICYCLE ( <i>COB10</i> )/ BIKE ( <i>COBAm3</i> )	12	12	<captions identical>
CAR	16 (exterior 11, interior 5)	16 (exterior 11, interior 5)	windscreen — windshield, windscreen wipers — windshield wipers, bonnet — hood, number plate — license plate wing — fender, wing-mirror — side-view mirror trunk — boot indicator only in <i>COB10</i> horn only in <i>COBAm3</i>
SHIP	9	9	<captions identical>
BIRD	8	8	<captions identical>
FISH	6	6	<captions identical>
FLOWER	5	5	<captions identical>
INSECT	7	7	<captions identical>
COMPASS POINTS	8	8	north-east – northeast, south-east – southeast, north-west – northwest, south-west -- southwest
CONTINENTS	7	7	<captions identical>
GREENHOUSE EFFECT	5	5	<captions identical>
VOLCANO	4	5	crater only in <i>COBAm3</i>
WATER CYCLE	4	4	<captions identical>
WIND TURBINE	4	4	<captions identical>
DIGESTIVE SYSTEM	7	7	<captions identical>
EAR	6	6	<captions identical>
EYE	6	6	<captions identical>
HEART	5	5	<captions identical>
ORGANS	6	6	<captions identical>
RESPIRATORY SYSTEM	9	9	<captions identical>
SKELETON	10	10	<captions identical>
TEETH	4	4	<captions identical>

that 16-page full-color collective visual panels are interesting and attractive, forming a picture book or in fact, a picture dictionary. We men-

tioned above that in *COBAm2* many Picture Dictionary panels regarding sports and musical instruments were presented. These panels were, unfortunately, discontinued in *COBAm3*. Many learners of English might find value in such images of their favorite pastimes in the Visual Dictionary supplement.

The inclusion of such panels as GREENHOUSE EFFECT and WATER CYCLE in the latest edition is welcome in terms of environment issue awareness as in the case of *COB10* (see Kokawa et al. 2024:53).

#### 6.2.4. Conclusion

The Visual Dictionary of *COBAm3* can be surmised to have been developed with *COB9* Visual Dictionary as a basis, rather than through the update of the Picture Dictionary panels in *COBAm2*. It might be considered regrettable that the links between the A-Z dictionary text entries and the Visual Dictionary section has been rendered less direct, but it has its own merits in terms of a collective supplement. The provisions of British English counterpart information would be desirable in the Visual Dictionary panels as well. (Section 6.2 by Kokawa)

#### 6.3. SYNONYMS

The 'SYNONYMS' column constitutes one of the feature boxes that have been adopted from *COB10* nearly verbatim. These feature boxes differ from the 'Thesaurus' boxes that are found in *COBAm2*. Within the 'SYNONYMS' boxes, synonyms for the headwords are listed, each of which is accompanied by illustrative usage examples. While the synonyms that are presented in the 'SYNONYMS' boxes remain consistent with those in *COB10*, slight alterations have been made to the usage examples to account for the differences between American and British English. For instance, in the case of the 'SYNONYMS' box for the headword **charge**, the synonyms that are listed in both *COB10* and *COBAm3* are identical; however, a comparison of the usage examples for the synonym **cost** shows that the currency unit denoted is 'pounds' in *COB10* and 'dollars' in *COBAm3*. A similar adjustment can be

observed in the case of ‘SYNONYMS’ box for the headword **earnings**. It is evident that, while *COBAm3* incorporates the ‘SYNONYMS’ boxes from *COB10*, it adapts their features in a way that caters to its audience, presumed to consist of American English speakers, by modifying certain details, such as replacing ‘pounds’ with ‘dollars’ to reflect differences in currency. (Section 6.3. by Takahashi)

#### 6.4. Collocations

‘Collocations’ columns were introduced into *COB9* in place of Word Partnership boxes in *COB8*. There is no change in the members of ‘Collocations’ columns between *COB9* and *COB10*. *COBAm3* inherited these columns. Columns of **flat** and **holiday** disappeared from *COB10*, and columns of **apartment** and **attention** appeared in *COBAm3*. Americanisms such as **behavior**, **color**, **defense**, and **humor** are also seen. In total, 453 columns of ‘Collocations’ are seen in *COBAm3*.

#### 6.5. Prefixes and Suffixes

There was no change in the contents of ‘Prefix’ boxes and ‘Suffix’ boxes between *COB9* and *COB10*. Here we compare the ‘Prefix’ boxes and the ‘Suffix’ boxes in *COB10* and those in *COBAm3*. *COBAm2* didn’t treat the prefixes and suffixes as columns but as headwords. From *COB10* only the ‘Prefix’ column of **part-** disappeared, and that of **nano-** appeared in *COBAm3*. The ‘Suffix’ columns of **-ous** and **-sion** disappeared from *COB10* though they could have been included in *COBAm3*. The ‘Suffix’ column of **-ise** appeared in *COBAm3*.

#### 6.6. Word History

*COB9* has introduced 46 ‘Word History’ columns to show the etymologies of words. They took place of 490 kinds of Word Link boxes in *COB8*. *COB10* inherited the same ‘Word History’ columns.

There is no change in the contents of ‘Word History’ columns between *COB10* and *COBAm3*, except the difference in spelling between **pyjamas** and **pajamas**. **Whiskey** is spelled **whisky** in the ‘Word History’ column. 46 columns are in *COBAm3*.

(Sections 6.4.–6.6. by Ikeda)

## 7. Information in the back matter

### 7.1. Frequent Words

*COBAm3* has ‘a list of 3,000 most frequent words in Collins Corpus’. Actually, the list has 3,185 words, which is the same as in *COBAm2*. *COBAm1* (2007) had a list of 3,224 words as ‘Defining Vocabulary’.

*COB10* has a list of 3,110 most frequent words from the Collins Corpus. We compared the words in the lists of both *COBAm3* and *COB10*. The following 188 words are in the list of *COBAm3* but not in the list of *COB10*.

abandoned	abortion	accord	acquisition	acre
activist	advanced	advocate	afterward	agricultural
ah	all right	allied	altogether	amendment
anymore	April	argue	Asian	assured
audio	August	auto	avenue	await
banker	basketball	bass	bearing	beating
behavior	bloody	bound	broadcasting	broker
brown	Caribbean	caring	center	charter
Christmas	colonel	color	colored	commentator
congressional	cooperate	‘cos	dancing	December
defense	didn’t	doesn’t	downtown	dressed
east	eastern	enforcement	English	European
false	February	fighter	fixed	flavor
Friday	given	god	gone	gray
harbor	high school	highway	honor	humor
inch	included	initially	Islam	Islamic
IT	January	Jesus	jew	July
June	kilometer	labor	Latin	lawsuit
let’s	made-up	maker	March	marked
May	meter <sup>3)</sup>	Middle East	Miss	mom
Monday	Mr	Mrs.	Ms.	navy
Nazi	neighbor	neighborhood	news agency	newscaster
north	northeast	northern	northwest	noted

November	October	of course	offense	oh
okey	Olympic	one's	opposed	organized
organizer	parliamentary	percent	photographer	pickup
playoff	point of view	policeman	pp.	publisher
publishing	purple	qualified	quite	railroad
rating	real estate	reporting	resignation	revolutionary
rider	Roman	rumor	runner	satisfied
Saturday	secretary-general	Secretary of State	Security Council	September
shaped	sir	south	southeast	southwest
speaker	speaking	State Department	Sunday	terror
terrorism	testing	theater	Third World	threatening
Thursday	till	toward	transportation	traveler
treasury	Tuesday	U.N.	united	United Nations
Wall street	Wednesday	west	western	White House
wireless	withdrawal	Wooden		

On the other hand, the following 115 words are in the list of *COB10* but not in the list of *COBAm3*.

academy	accommodation	ace	advertiser	afterwards
aged	allegedly	all-round	an	behaviour
bench	Bible	bonus	boyfriend	bulletin
cannot	carefully	celebrity	cent	centre
CEO	children	clearly	click	colour
commonwealth	completely	co-operate	courier	currently
CV	decent	defence	dine	DNA
documentary	East	Eastern	email	ending
episode	Euro	excited	fantastic	fitness
flavour	garage	GDP	girlfriend	GNP
God	GP	grey	harbour	heading
higher education	honour	humour	industrialized	infrastructure
inspire	investigation	kilometre	labour	left-wing
legend	lifestyle	metre	MP	neighbour
NHS	North	North-east	Northern	north-west
offence	offline	ok	pal	PC

per cent	petrol	PM	PR	programme
quit	railway	ranking	regularly	restoration
rumour	South	south-east	south-west	spokesperson
stunning	supermarket	teenage	theatre	Tory
totally	tourism	towards	trainer	training
UK	UN	unite	VAT	viewer
West	Western	worse	worst	zero

Comparing the list of *COBAm3* and the list of *COB10*, we find the following 33 pairs of words which have different spellings.

behavior - behaviour	center - centre	color - colour	cooperate - co-operate	defense - defence
east - East	eastern - Eastern	flavor - flavour	god - God	gray - grey
harbor - harbour	honor - honour	humor - humour	kilometer - kilometre	labor - labour
meter <sup>2)</sup> - metre	neighbor - neighbour	north - North	northeast - North-east	northern - Northern
northwest - north-west	offense - offence	okey - ok	percent - per cent	rumor - rumour
south - South	southeast - south-east	southwest - south-west	theater - theatre	toward - towards
U.N. - UN	west - West	western - Western		

## 7.2. Academic Word List

*COBAm3* has a list of 570 words of **Academic Word List** with their sublists from 1 to 10. The list appeared in *COBAm1* (2007), and remained unchanged in *COBAm2* and *COBAm3*. The list also appeared from *COB6* to *COB10*. The list has been available online since *COB8* (2014). (Sections 7.1.–7.2. by Ikeda)

## 8. Concluding remarks

Through our limited research here, we have seen a steady and well-thought-out update of information through the revisions from the second to the third editions of the *COBUILD Advanced Learner's Dictionary*, throughout the various information categories. It is now up-to-date not only in the linguistic sense, but also in terms of conformity with social trends such as gender equality, DEI and environmental awareness (See Sections 4 and 6.2 above).

On the other hand, the change from color printing to black and white

in the A-Z part of the dictionary means that the visibility, amount of information and user-friendliness have been somewhat reduced in *COBAm3*. This may have been caused by a concern for cost on the part of the publishers and is also true of the *COB9* to *COB10* revision but may be regarded overall as a regrettable policy change. Also, one more feature that may await future consideration may be that CEFR level markers (A1, A2, B1 and B2) applied to different senses of some entries in *COB10* were somehow not employed in the contemporarily-compiled *COBAm3*. The fact that the CEFR is a European standard may have affected this decision by the Collins team.

Still, *COBAm3*, a substantially renewed edition of the American version of one of the major EFL dictionaries from a major publisher is very welcome, providing EFL learners with another choice not only between the British and American versions of Collins EFL dictionaries of parallel design, concept and up-to-datedness, but also among other dictionaries for learners of English.

## NOTES

### Section 1

- 1) In *COB10*, **nationalise** is entered as a headword.
- 2) A note such as 'National holidays are normally called bank holidays in Britain' may be suggestable.

### Section 5

- 1) Prefixes **a-** and **e-**, and suffixes **-ability** and **-able** in *COBAm2* are treated in PREFIX boxes and SUFFIX boxes in *COBAm3*, so the examples of these affixes are not included here.
- 2) The definition of **actress** says as follows: ... Some women who act prefer to be called 'actors' rather than 'actresses'.
- 3) Here we treat metre in the original form in the list as meter because in the body the headword metre should be referred to the headword meter in *COBAm3*.

## REFERENCES

- Chambers, Jack K. "Social embedding of changes in progress." *Journal of English Linguistics* 26.1, 1998, pp. 5–36.
- Chambers, Jack K. "Yod-Dropping in an English Accent" *Journal of the Phonetic Society of Japan* 6.3, 2002, pp. 4–11.

- Dinkin, Aaron J. "Weakening resistance: Progress toward the low back merger in New York State." *Language Variation and Change* 23.3, 2011, pp. 315–345.
- Kokawa, Takahiro, et al. "An Analysis of *Collins COBUILD Advanced Dictionary of English*, Seventh Edition," *LEXICON* No. 45, 2015, pp. 74–137.
- Kokawa, Takahiro et al. "An analysis of *Collins COBUILD Advanced Learner's Dictionary*, Ninth edition," *LEXICON* No. 50, 2020, pp. 42–108.
- Kokawa, Takahiro et al. "An analysis of *Collins COBUILD Advanced Learner's Dictionary*, Tenth edition," *LEXICON* No. 54, 2024, pp. 21–72.
- Masuda, Hideo, et al. "An Analysis of *Collins COBUILD Advanced Dictionary of American English*," *LEXICON* No. 38, 2008, pp. 47–156.
- Takebayashi, Shigeru. *Eigo Onseigaku [English Phonetics]*. Tokyo: Kenkyusha, 1996.
- Wells, John C. *Accents of English 3. Beyond the British Isles*. Cambridge: Cambridge University Press, 1982.

## DICTIONARIES and their ABBREVIATIONS

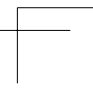
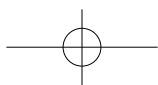
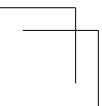
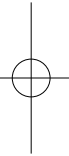
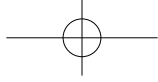
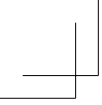
- COBAm1*: *Collins COBUILD Advanced Dictionary of American English*. First published. Ed. by J. Sinclair. Glasgow: Thomson Heinle. 2007.
- COBAm2*: *Collins COBUILD Advanced American English Dictionary*. Second edition. Ed. by J. Sinclair. Glasgow: Thomson Heinle. 2016.
- COBAm3*: *Collins COBUILD Advanced American English Dictionary*. Third edition. Ed. by J. Sinclair. Glasgow: Thomson Heinle. 2023.
- COB6*: *Collins COBUILD Advanced Dictionary of English*, Sixth edition. Ed. by J. Sinclair. Glasgow: HarperCollins Publishers 2008.
- COB7*: *Collins COBUILD Advanced Dictionary of English*, Seventh edition. Ed. by J. Sinclair. Boston: Heinle Cengage Learning. Glasgow: HarperCollins Publishers 2012.
- COB8*: *Collins COBUILD Advanced Learner's Dictionary*, Eighth edition. Ed. by J. Sinclair. Glasgow: HarperCollins Publishers 2014.
- COB9*: *Collins COBUILD Advanced Learner's Dictionary*, Ninth edition. Ed. by J. Sinclair. Glasgow: HarperCollins Publishers 2018.
- COB10*: *Collins COBUILD Advanced Learner's Dictionary*, Tenth edition. Ed. by J. Sinclair. Glasgow: HarperCollins Publishers 2023.
- EPD14*: *English Pronouncing Dictionary*. 14th edition. Daniel Jones (ed. by A. C. Gimson, revisions and supplement by S. M. Ramsaran). Cambridge: Cambridge University Press. 1988.
- LDOCE6*: *Longman Dictionary of Contemporary English*. 6th ed. Ed. by Laurence Delacroix. Harlow: Pearson Education, 2014.
- OALD10*: *Oxford Advanced Learner's Dictionary of Current English*, Tenth edition. Ed. by Diana Lea and Jennifer Bradbery. Oxford: Oxford University Press. 2020.

## WEBSITES

- Academic Word List: <https://www.victoria.ac.nz/recources/academicwordlist> (accessed on March 25, 2025)
- Collins Online Dictionary: <https://www.collinsdictionary.com/> (accessed on March 25,

2025)

ELT COBUILD Worksheets Resources: <https://collins.co.uk/pages/elt-cobuild-reference-cobuild-worksheets-resources/> (accessed on March 25, 2025)



## 投稿規定

(1) 投稿は岩崎研究会会員に限る。但し、非会員であっても論文審査委員から推薦のあった場合は特別に認める。(2) 論文の内容は未発表のものに限る。(3) 用語は英語に限り、原則として native check を受けたものとする。(4) 注 (note) は後注とし、章ごとに通し番号を付ける。(5) ギリシャ字、ロシア字以外の特殊文字はできるだけローマ字化してほしい。音声記号は国際音声学協会 (IPA) 所定のものを用いる。(6) 引用文献：書式は MLA Style に従う。(7) 枚数：論文はワープロ原稿で、1行はアルファベットの小文字で 70 字、450 行以内。(8) 原稿はすべて論文審査委員による審査の上採否を決定する。共同執筆論文を別として、論文の掲載は毎号 1 人 1 篇とする。(9) 都合により短縮を求めることがある。印刷上の体裁および論文の掲載年度については編集委員に一任する。(10) 抜刷は 20 部までを無料で、別に本誌 1 部を呈上する。(11) 原稿は随時受付ける。(12) なお、詳細は別に定める。

### 論文審査委員

赤須 薫	土肥 一夫
岩崎 春雄	馬場 彰
浦田 和幸	斎藤 弘子
増田 秀夫	清水あつ子
宮井 捷二	投野由紀夫
山田 茂	小室 夕里

編集後記 編集の担当が赤須さんから小室さんに引き継がれました。赤須さんにはお忙しい中、長い間編集の仕事をしていただきありがとうございました。その期間にはいろいろなことがありました。中でも Lexicon の原稿応募が非常に少なかった時期がありました。赤須さんの懸命な努力のおかげで Lexicon の一時休刊などの事態は避けられました。大学の教員の仕事は私の時代と違って極めて多忙になりました。その中での Lexicon の編集作業は大きな負担であったことと思います。これは小室さんにとっても同じことでしょうか、Lexicon をより充実した内容にするためにご努力されるようお願いいたします。

今回も優れた論文が 3 編掲載されました。

稲生さんの論文は on thin ice という idiom に絡めて先行研究やコーパスなどを参照し自らの主張を見事にまとめ上げたものです。日本語の「薄氷を踏む」は古くからある比喩なのですね。また学習英和辞書の on thin ice の記述も様々ですね。

音声学の共同論文は斎藤弘子教授の退官記念論文です。このような論文は斎藤さんをはじめとする我が岩研の音声学を専門とする人でないと成しえない業績です。

小川さんチームは昨年に引き続き COBUILD の分析にあたってくれました。Editions による違いは各項目にわたって辞書学者・編集者にも大変参考になります。社会的な変化による新しい見出し語、例えば COVID-19 pandemic 関連の語についての指摘などは注目されます。

(2025 年 5 月 31 日 S. M.)

