

論文

# Hyphenation in Present-day English Complex Words: Analysis of Style Manuals and Corpora

NISHIBU Mayumi

## 要 旨

本稿では、英語の合成語 (complex words) にハイフンが挿入される条件について、主要なスタイルマニュアル (=出版物などにおいて統一した言葉遣いや表記方法を規定する手引き) の記述を比較し、さらに英・米語のコーパスと *Oxford English Dictionary Online* (OED-Online) を用いて使用の実態を分析した。三種類のスタイルマニュアルを分析した結果、どれも複合形容詞におけるハイフンの使用に関する項目が最多で共通点が多いが、それぞれには特徴的な記述があった。また、特に米語に於いてハイフンをなくして一語にする、あるいはスペースを入れる傾向が強まっていることが分かった。また、コーパスを使用してハイフンで結ばれた合成語を分析した結果、形容詞に次いで名詞の機能を持つものが圧倒的に多いことと、イギリス英語ではより頻繁にハイフンを施す傾向があること、発話中の間投詞やフィラーの音を繋ぐためにハイフンが使用されていることが明らかになった。

キーワード：ハイフン、スタイルマニュアル、コーパス、合成語、  
複合語、接辞

## 1. Hyphenation

The usage of a hyphen can be categorized into the following three types: soft

hyphens, hard hyphens, and phonetic hyphens.

A soft hyphen is inserted to break a word at the end of a line on word processing software. Phonetic hyphens are used to show stuttering (e.g., “Y-yes”) or to exaggerate slow and careful pronunciations in utterances (e.g., “No-o-o,” “P-l-e-a-s-e!”).

Hard hyphens, which this paper discusses, are used to join two or more components in complex words. “Complex word” is the upper category, subsuming “compounds” and “derivatives.” A compound consists of two or more bases, whereas a derivative consists of at least one base combined with affixes.

Most compounds consist of two bases (e.g., *ready-made*, *drop-out*), while a smaller number of them contain more than two (e.g., *up-to-date*, *mother-in-law*). Compounds can be represented in either one or sometimes more orthographical patterns among the following three patterns: open (spaced, separated), hyphenated, and solid (closed, juxtaposed). For example, either form can be seen in some compounds in the forms of *workplace* (solid), *work place* (open), and *work-place* (hyphenated).

It is often stated that compounds emerge in open form at the first stage, become hyphenated when they become more established afterward, and gradually change into solid form (Quirk et al., 1985; Huddleston & Pullum, 2002).

When the sequence of two or more words functions as an attributive modifier in a nominal, hyphenation is set to default (e.g., *a London-based company*, *out-of-town shopping*, *a never-to-be-forgotten experience*). Huddleston & Pullum (2002, p. 1762) term these as “syntactic hyphens” and state that they include attributive modifiers ending with past and present participles (e.g., *a well-argued reply*, *a hard-drinking man*), those ending with nouns (e.g., *a four-point plan*, *a fast-food outlet*) and those consisting of a phrase (e.g., *a creamier-than-average taste*, *the Hobart-to-Sydney classic*). Some stylebooks call them “temporal compounds” (APA, 2020, p. 162; OUP, 2016, p. 133).

Huddleston & Pullum (2002, p. 1761) distinguish “lexical hyphens” from syntactic hyphens. Lexical hyphens are found in established complex words formed morphologically by the lexical word-formation process; in other words, they are lexicalized combinations of bases of a compound or those of the affix and base of a derivative.

In the case of compounds, lexical hyphens are frequently found in the following ten cases.

(1) Lexical Hyphens in Compounds

- i. Compound adjective      *bone-dry, oil-rich, red-hot, snow-white*
- ii. Contains transitive prep      *free-for-all, sergeant-at-arms, sister-in-law*
- iii. Intransitive prep as 2<sup>nd</sup> base      *break-in, build-up, drop-out, phone-in, stand-off*
- iv. Coordinative compound      *Alsace-Lorraine, freeze-dry, murder-suicide*
- v. Nominal compound + -ed      *one-eyed, red-faced, three-bedroomed*
- vi. Numerals and fractions      *twenty-one, ninety-nine, five-eighths*
- vii. Dephrasal compound      *cold-shoulder (V), has-been (N), old-maidish*
- viii. Verb with noun as 1<sup>st</sup> base      *baby-sit, gift-wrap, hand-wash, tape-record*
- ix. 1<sup>st</sup> base is letter-name      *H-bomb, t-shirt, U-tern, V-sign*
- x. Rhyming-base compounds      *clap-trap, hoity-toity, teeny-weeny, walkie-talkie*

(Huddleston & Pullum, 2002, p. 1762)

The compounds in example (1) are so common and familiar that they could be regarded as lexicalized. However, a dichotomy between lexical and syntactic compounds seems to be difficult to set because the level of lexicalization or familiarity is comparative and gradable in nature. In particular, (i) compound adjectives and (iv) coordinative compounds can include a variety of combinations of component bases. Many of them are difficult to discern between lexical and syntactic ones. Given the definition of lexicalization is the state in which the combination of each element in a compound is established and each element is no more productive, (vi) numerals and fractions should be different from the others since the combinations of numerals and fractions are innumerable. In addition, (i) compound adjective in (1) apparently limits to compounds ending with an adjective. However, the term should have included (v) nominal compound + *-ed* and (vi) numerals and fractions since they can behave as adjectives modifying a noun. Let us discuss these problems in later sections.

On the other hand, when one or more affixes are attached to a base, the words are termed as derivatives. Most derivatives are represented in solid orthography (e.g., *understandable, childish*), while some are hyphenated between a base and an affix (e.g., *non-standard, co-operative, child-like*). Affixes are, by definition, bound morphemes and cannot appear independently. Thus, they appear in solid or hyphenated form. Except for word-like suffixes such as *-wise* and *-like*, most suffixes combine with a base without a hyphen. Several prefixes (e.g., *co-, de-, re-*) are hyphenated to avoid the double

occurrence of a vowel (e.g., *co-author*, *re-elect*, *de-ice*) and repetition (e.g., *re-release*). Derivatives that are established and those written shorter with a relatively smaller number of letters tend to be solid without a hyphen.

The orthographical choice is sometimes capricious and depends on individual tastes, dictionary houses, and the guidelines written in stylebooks. The following section examines the guidelines regarding hyphenation in major style manuals.

## 2. Style Manuals

We have so far reviewed the previous literature by linguists. This section investigates how hyphenation is prescribed in three major style manuals: *the New Oxford Style Manual (Oxford)* (2016), *the Associated Press Stylebook (AP)* (2022), and *the Publication Manual of the American Psychological Association (7<sup>th</sup> ed.) (APA)* (2019). The former two manuals are used worldwide as a guideline for writing in English by writers, while *APA* is for writing academic papers and books.

### 2.1 Overview of the three style manuals

All three manuals explain the usage of hyphens and give plenty of advice on when to hyphenate or not. This section overviews how each stylebook prescribes the usage of hyphens in compounds and derivatives. Table 1 summarizes hyphenation conditions in several types of compounds mentioned in the three style manuals.

The manuals explain compounds by grammatical function: adjectives, verbs, and nouns. All three books have allocated their largest spaces to explain the hyphenation in compound adjectives, with nine conditions in *Oxford* and *APA* and five in *AP*. There are eight conditions where a hyphen is necessary, whereas there are seven conditions where a hyphen should be avoided in Table 1.

For a compound adjective, its position, before the noun it modifies or after the noun, is vital. When a compound adjective is a premodifier in a nominal, a hyphen is indispensable in most cases. On the other hand, a hyphen is not needed when used in predicative use in a clause, except for adjective + a participle form of a verb (*Oxford*) and longer phrases, such as *state-of-the-art (AP)*.

Regarding noun + noun compounds as a modifier, *Oxford* and *AP* suggest omitting a hyphen, particularly when they are lexicalized compound nouns. *AP* even encourages not

**Table 1. Guidelines of hyphenation in compounds**

Compound features		Oxford	AP	APA	
<b>Adjectives</b>	Prenominal modifiers	○ present-, past-participle form as 2 <sup>nd</sup> base	✓		<i>decision-making behavior, water-deprived animals</i>
		○ phrases	✓	<i>black-and-white photograph, know-it-all attitude</i>	<i>trial-by-trial analysis, one-on-one interview</i>
		○ adj + noun	<i>first-class seat, low-level radioactive waste</i>		<i>high-anxiety group, middle-class families</i>
		○ number as 1 <sup>st</sup> element			<i>six-trial problem, 12<sup>th</sup>-grade students, 16-min interval</i>
		○ fractions			<i>two-thirds majority</i>
		○ vowel or consonant collision	<i>drip-proof, part-time</i>		
		○ others		<i>well-</i>	
		○ × coordinative dual heritage	<i>Greek-American wife</i> ○	<i>African American, Mexican American</i> ×	
		× lexicalized noun + noun compounds	<i>labour market liberalization</i>	<i>climate change report, emergency room visit</i>	
		× capitalized elements	<i>British Museum staff, Latin American studies</i>		
		× comparative or superlative			<i>better written paper, highest scoring students</i>
		× Latin phrases (adj. or adv.)	✓		<i>a posteriori test, post hoc comparisons</i>

	×	letter/number as 2 <sup>nd</sup> element			<i>Group B participants, Type II error</i>
	×	-ly in 1 <sup>st</sup> element	<i>happily married couple</i>	<i>easily remembered rule</i>	<i>widely used test</i>
Predicative	○	adj. + verb participle	<i>good-looking, double-breasted</i>		
	○	longer, uncommon		<i>state-of-the-art, fill-in-the-blank</i>	
	×	All the others	✓	✓	✓
Verbs	○	derived from compound nouns	<i>machine-gun</i>		
	○	lexicalized		<i>speed-walk, spoon-feed</i>	
	×	phrasal verbs		<i>back up, set out</i>	
	×	All the others	✓	✓	✓
Nouns	○	derived from phrasal verbs	<i>back-up</i>		
	○	confusion avoidance	✓	<i>sister-in-law, hand-me-downs, merry-go-round</i>	✓
	×	fractions			<i>one third</i>
	×	coordinative dual heritage		<i>African American, Mexican American</i>	
	×	lexicalized	✓		✓
	×	scientific terms	<i>liquid crystal display, sodium chloride solution</i>		✓

**Notes:** ○ indicates “hyphenate” and × “do not hyphenate.” ✓ indicates the same description is provided without examples. Italicized items are examples quoted from each style manual.

to insert a hyphen in lexicalized adjective + noun compound modifiers, while the other two publishers regard it as necessary.

To summarize the hyphenation in compound adjectives, a hyphen is required when a compound adjective has the second or last element of a participial form of a verb, a phrasal construction of more than two elements, an adjective + noun construction, a numeral as

the first element, and a fraction. In addition, some conventions with phonological reasons can be found in *drip-proof* and *part-time*.

On the other hand, compound verbs rarely need a hyphen except for ones converted from another compound and specially lexicalized ones, such as *machine-gun* (verb), *speed-walk*, and *spoon-feed*.

The hyphenation in compound nouns is controversial, particularly for the ones with dual heritage meanings. As can be seen in the following sections, *American Mexican* was combined with a hyphen in the past, but the recent change prefers the open form in the US. Scientific terms are not hyphenated. Compound nouns converted from phrasal verbs need a hyphen, according to *Oxford*, although more and more compound nouns have been written in solid form (OUP, 2016).

Table 2 summarizes the descriptions of hyphenation in derivatives.

**Table 2. Guidelines of hyphenation in derivatives**

Affix features		Oxford	AP	APA
<b>Prefixes</b>	○ vowel or consonant collision	<i>anti-intellectual, non-negotiable, semi-invalid, pre-eminent</i>	✓	(double a, i, o) <i>meta-analysis, anti-intellectual, co-occur</i>
	○ repetition	<i>re-release, sub-subcategory</i>	<i>sub-subparagraph</i>	
	○ before a capitalized name, a date, a numeral	<i>anti-Darwinism, mid-Victorian</i>	✓	<i>pro-Freudian, Stroop-like, post-1977</i>
	○ more than one elements			<i>non-achievement-oriented students</i>
	○ ambiguity avoidance	<i>pro-life, re-form, re-cover</i>	<i>re-creation</i>	<i>re-pair, re-form, un-ionized</i>
	○ others	<i>ex-, mid-</i>	<i>self-, all-, ex-, half-</i>	<i>self-</i>
	× double- <i>e</i> ( <i>re-, pre-</i> )		<i>preempted, reelected</i>	<i>reexamine, preexisting</i>
	× double- <i>o</i> ( <i>co-</i> )		<i>cooperate, coordinate</i>	
<b>Suffixes</b>	○ two <i>ls</i>	<i>bell-less, shell-like</i>		
	○ others	<i>-proof, -scape, -wide</i>	<i>-free, (-based, -elect)</i>	<i>-self</i>
	× All the others	✓	✓	✓

**Notes:** ○ indicates “hyphenate” and × “do not hyphenate.” ✓ indicates the same description is provided without examples. Italicized items are examples quoted from each style manual. (*-based, -elect* are not suffixes in its strict sense.)

All three manuals agree that a hyphen is unnecessary when an affix is attached to a base. However, hyphenation is preferred where awkward pronunciation or semantic ambiguity may happen or where prefixes or suffixes look like a single word (e.g., *all-*, *ex-*, *half-*, *self-*, *-self*, *-like*, *-free*). However, the difference can be found in the hyphenation in double-*e* and double-*o*. The US guidelines demand solid forms, while the British one suggests hyphenated forms for these patterns.

All the manuals state that writers should refer to specific dictionaries when they are unsure whether a hyphen is necessary. However, it is noteworthy that each manual recommends different dictionaries, as shown in Table 3.

**Table 3. Dictionaries for reference**

Oxford (2016)	AP (2022)	APA (2019)
<i>New Oxford Dictionary for Writers and Editors</i>	<i>Webster's New World College Dictionary</i>	<i>Merriam-Webster.com Dictionary</i>
<i>Oxford Dictionary of English</i>		<i>APA Dictionary of Psychology</i>
<i>Oxford English Dictionary Online</i>		

The two American style manuals recommend Merriam-Webster's dictionaries, whereas *Oxford* does Oxford's dictionaries. It is worth noting that the online dictionaries, *the Oxford English Dictionary Online (OED-Online)* and *Merriam-Webster.com Dictionary*, are included in their recommendation.

This section has overviewed the descriptions of the three style manuals. The following section explores the distinctive features of each stylebook.

## **2.2 The New Oxford Style Manual (2016)**

This manual represents the British way of hyphenation and the revised edition of *the Oxford Style Manual* (2003). This new book provides the whole usage of a hyphen, not limited to compounds and derivatives. For example, it states that a hyphen is used for numbers (e.g., to spell-out numbers from 21 to 99, *twenty-three*, *thirty-fourth*), compass points (e.g., *south-east*, *south-south-east*), phonological representations of stammering, paused, and intermittent speech (e.g., *P-p-perhaps*, *Uh-oh*), and combining forms (e.g., *quasi-scientific*). The new edition sometimes mentions different orthography in the US practice, where the hyphen is less used. It also describes a recent tendency to avoid



hyphenation in an adjectival noun + noun combination (OUP, 2016, p. 58).

### **2.3 *The Associated Press Stylebook (2022)***

The AP's consistent primary policy is to “use a hyphen whenever ambiguity would result if it were omitted” and “no hyphen is needed when the meaning is clear and unambiguous without the hyphen”(AP, 2022, pp. 232–3). Notably, it reiterates that we should not use a long hyphenated phrasal modifier and instead rewrite it using a post-nominal modification (e.g., not *a how-to-use-hyphens-wisely guide*, but *a guide about how to use hyphens wisely*) (p. 232). The guideline made changes in 2019 not to hyphenate dual-heritage compounds (e.g., *African American*) and double-*e* (e.g., *reelected*, *preempted*). It even proposes to use the open form for compound modifiers, such as *third grade teacher*, *special effects embellishment*, and *public land management*, because their meanings are clear without a hyphen (p. 233). These aspects indicate that the US practice tries to reduce hyphenation.

### **2.4 *The Publication Manual of the American Psychological Association (7<sup>th</sup> ed.) (2019)***

Since this manual aims at academics, the directions are more specific and clear-cut than the other two. In particular, as seen in Table 1, many more conditions where a hyphen should not be used are carefully written, including comparative and superlative forms of compound adjectives. It also specifies hyphenation in compounds with numbers and (capital) letters in detail.

## **3. Corpus-based Observation: *BNC* and *COCA***

In any case, hyphens show that components of a complex word originate from morpho-semantically separable elements. This section analyzes the actual data in two large, balanced corpora.

The corpora used in the analysis are *the British National Corpus (BNC)* and *the Corpus of Contemporary American English (COCA)* on [www.english-corpora.org](http://www.english-corpora.org). *BNC* comprises approximately 10 million spoken and 90 million written words of British English. A small portion (8%) of the data was collected between 1960 and 1984, and the majority (92%) was collected between 1985 and 1993. *COCA* contains more than 1 billion

words collected annually from 1990 to 2019 in American contexts, 13% of which contain transcripts of unscripted conversations (Davies, 2022).

### 3.1 Method

The analysis retrieved the 100 most frequent hyphenated complex words in the two corpora. It examined to what extent compounds and derivatives could be observed in the corpus data. As noted in the previous section, the ten categories of lexical compounds in (1) needed some revision to analyze actual data. In the analysis, the distinction between “lexical” and “syntactic” compounds was not made, and samples were categorized by the formal patterns of elements into 17 groups of compounds along with derivatives. Derivatives were designated according to the list of affixes retrieved from *OED-Online*.

The focus was also placed on the word class of frequent hyphenated words. The word class classification of samples was based on *OED-Online*. In addition, some differences between British and US English were expected to be found when using the two corpora.

### 3.2 Results and Discussion

The 100 frequent words retrieved in each corpus were examined in terms of their elements’ formal patterns. The results are shown in Table 4.

In the compounds section in Table 4, numerals and fractions accounted for the most significant ratio in *BNC* and the second largest in *COCA* (*BNC* 19%, *COCA* 23%). The adjective-noun combination was also conspicuous in both corpora (*BNC* 16%, *COCA* 24%), and its ratio was the largest in *COCA*.

The results were similar in both corpora except for four features. First, interjections and fillers appeared only in *COCA*, such as *good-bye*, *mm-hm*, and *uh-huh*. The interjections and fillers do not represent rhyming sounds or stammering mentioned in the former sections. Second, the words beginning with *e-*, shortening of *electronic*, such as *e-mail(s)*, appeared only in *COCA*. They have been so accepted recently that the hyphen after *e* increasingly drops. A plausible reason is that *COCA* contains more recent texts and many more spoken utterances than *BNC*.

Third, the words referring to compass points, *north-east*, *north-west*, *south-east*, and *south-west*, appeared only in the UK English corpus, which follows the Oxford guideline.

The fourth point is that hyphenation in *sitting-room* and *dining-room* was observable

Hyphenation in Present-day English Complex Words

**Table 4. Common 100 hyphenated complex words in *BNC* and *COCA***

		Compounds			
Formal patterns		BNC	COCA		
Numerals and fractions	numbers	<i>2-1, forty-five, twenty-one, twenty-two, twenty-four, twenty-five</i>	6	<i>1-2, 2-3, 6-0, 6-1, 6-2, 6-3, 6-4, twenty-five, twenty-four</i>	9
	fractions	<i>one-third, two-thirds</i>	2	<i>one-third, two-thirds</i>	2
	num - noun /adj	<i>one-day, second-hand, eighteenth-century, first-class, two-year, three-year, five-year, nineteenth-century, twentieth-century, three-quarters, three-dimensional</i>	11	<i>10-year, one-year, two-year, three-year, four-year, five-year, one-time, one-way, nineteenth-century, first-time, third-party, 3-d</i>	12
Transitive preposition		<i>day-to-day, up-to-date</i>	2	<i>day-to-day, face-to-face, one-on-one</i>	3
Verb - particle (phrasal verbs)		<i>back-up, break-up, built-in, build-up, follow-up, line-up, make-up, set-up</i>	8	<i>built-in, follow-up</i>	2
Preposition - noun		<i>in-house, in-service</i>	2	<i>in-depth, on-line</i>	2
Dephrasal conversion		<i>would-be</i>	1	<i>would-be</i>	1
Nominal compound -ed		<i>middle-aged, old-fashioned</i>	2	<i>middle-aged, old-fashioned</i>	2
Verb-ing ending	noun-Ving	<i>decision-making, policy-making, time-consuming</i>	3	<i>award-winning, decision-making</i>	2
	adv-Ving	<i>far-reaching, long-standing, wide-ranging</i>	3	<i>best-selling, long-standing</i>	2
	well-Ving	<i>well-being</i>	1	<i>well-being</i>	1
Verb-pp ending	noun-Vpp	<i>object-oriented</i>	1		0
	adv-Vpp	<i>short-lived, so-called</i>	2	<i>so-called</i>	1
	well-Vpp	<i>well-known</i>	1	<i>well-known</i>	1
Verb-ing - noun/adj		<i>dining-room, sitting-room, working-class</i>	3	<i>working-class</i>	1
Adj - noun	adj-noun	<i>full-time, half-time, half-way, inner-city, large-scale, long-term, middle-class, no-one, part-time, present-day, short-term, small-scale</i>	12	<i>all-star, all-time, full-time, half-hour, high-end, high-level, high-profile, high-quality, high-risk, high-speed, high-tech, large-scale, long-distance, long-term, low-cost, low-income, middle-class, part-time, real-life, real-time, real-world, same-sex, short-term</i>	23
	adj /noun /adv-noun	<i>left-hand, left-wing, right-hand, right-wing</i>	4	<i>right-wing</i>	1

Letter names and capital letters	<i>t-shirt, X-ray</i>	2	<i>e-mail, e-mails, t-shirt, t-shirts, X-ray</i>	5
Compass points	<i>north-east, north-west, south-east, south-west</i>	4		0
Coordinative dual heritage	<i>Anglo-Saxon</i>	1	<i>African-American, African-Americans</i>	2
Interjections and fillers		0	<i>bye-bye, good-bye, mm-hm, mm-hmm, uh-huh, uh-oh</i>	6
Proper nouns	<i>Hewlett-Packard, Jean-Claude, MS-DOS, Sally-Anne</i>	4	<i>al-Qaeda, Coca-Cola, hip-hop, Wal-Mart</i>	4
Combining forms	<i>by-election, socio-economic</i>	2		0
Others	<i>attorney-general, one-off, secretary-general, world-wide</i>	4	<i>brand-new, cease-fire, hands-on, health-care, president-elect, Wi-Fi, year-round</i>	7
<b>Derivatives</b>				
<i>co-</i>	<i>co-op, co-operate, co-operation, co-operative, co-ordinate, co-ordinated, co-ordinator, co-ordinating, co-ordination</i>	9	<i>co-founder, co-host, co-workers</i>	3
<i>self-</i>	<i>self-employed, self-esteem</i>	2	<i>self-defense, self-esteem</i>	2
Others	<i>non-existent, pre-tax, pre-war, post-war, semi-final, semi-finals, sub-committee, vice-president</i>	8	<i>ex-wife, non-profit, pro-life, re-elected, re-election</i>	5

only in *BNC*. In the US practice, these words are not hyphenated but are written in open form. According to Oxford style manuals, compound nouns consisting of a present participle and a noun formerly followed the “walking-stick rule.” If the *stick* itself was not walking, namely if the noun was not providing the action expressed by the verb, then the compound was hyphenated. On the other hand, if the noun provided the action, such as *walking wounded* and *walking delegate*, where the nouns were walking, the two elements were spaced. The samples in *BNC* seem to be the remaining of the rule. Nevertheless, this rule is no longer borne out, and they are now written as two words even in the UK (OUP, 2003; OUP, 2016).

Regarding derivatives, differences between the British and US English practices are evident. In *BNC*, many more prefixes were hyphenated than in *COCA*. In the UK English corpus, many words beginning with *co-*, and a few *pre-* and *semi-* were hyphenated.

Generally, English in the UK prefers to insert hyphens in complex words, while in the US, solid forms are more popular (Quirk et al., 1985; Huddleston & Pullum, 2002; OUP, 2016).

The results in Table 4 could reflect differences between English in the UK and the US, between the periods of collecting texts, and between spoken and written registers in each corpus.

The subsequent examination was conducted regarding the word classes of 100 common hyphenated samples. Word classes of multifunctional words are arranged in descending order of frequency in *OED-Online*. For example, when a word is labeled as “noun, adj, adv,” it functions mainly as a noun, and its secondary function is adjective, and the third one is adverb. Table 5 indicates the token and ratio of adjectives, adverbs, nouns, and verbs.

**Table 5. Word classes of common hyphenated complex words in *BNC* and *COCA***

	ADJ				ADV		NOUN				VERB	
	adj	adj adv	adj adv noun	adj noun	adv adj	adv adj noun	noun adj	noun adj adv	noun	noun verb	verb adj	verb
<i>BNC</i>	31	5	1	14	1	4	7	3	30	0	2	2
	51%				5%		40%				4%	
<i>COCA</i>	42	3	1	12	0	0	6	0	27	2	0	1
	58%				0%		35%				1%	

*Note:* *COCA* included six occurrences of interjections and fillers.

Except for the fact that *COCA* included 6 (6%) occurrences of interjections and fillers, the ratios were similar between the two corpora. The most significant number of the hyphenated complex words were adjectives (*BNC* 51%, *COCA* 58%), followed by nouns (*BNC* 40%, *COCA* 35%), while adverbs and verbs were rare. When all samples with adjective use were counted, the total ratio was over 60% (*BNC* 68%, *COCA* 64%). In the same way, the ratio of total nouns amounted to around 60% (*BNC* 59%, *COCA* 48%). These results conclude that hyphenated complex words are adjectives in the most significant number of cases and nouns in the second place.

## 4. Analysis of *OED-Online*

Similar to the previous section, word classes of common hyphenated complex words were examined using *the Oxford English Dictionary Online* (2023–). It is the third revised online edition of *the Oxford English Dictionary*. It updates every three months with 500–1000 new headwords, which makes it up-to-date evidence of the present-day English language.

The online platform has been transformed into the current one, allowing the users to download retrieved data from *OED-Online* in a CSV file. The platform provides a search engine, and the results can be sorted by frequency, alphabet, or the year of the first appearance.

The frequency in *OED-Online* ranges from 8 (very high-frequency words) to 1 (very low-frequency). Per million words, words in Band 8 appear more than 1000 times (0.02 % types in the total entries), and those in Band 7 appear from 100–1,000 times (0.18%). In the same way, Band 6 includes words appearing 10–100 times (1%), and Band 5 does those appearing 1–10 times (4%). Band 1 shows the lowest frequency, 0.01 or lower times (18% of the total entries).

### 4.1 Method

The method for an analysis using *OED-Online* was as follows. First, headwords consisting of hyphenated elements were retrieved and sorted by frequency in descending order. Then, the results were downloaded in a CSV file. It included the headwords, their word classes, their first appearance year, and the frequency band. Secondly, 235 headwords in 7–5 frequency Bands were collected, and other less frequent ones were eliminated. No entry was found in Band 8, *COVID-19* belonged to Band 7, and seven words (*long-term*, *in-person*, *so-called*, *year-old*, *short-term*, *well-being*, *well-known*) belonged to Band 6. Words of Band 5 amounted to 227 types, with many of them appearing alphabetically in the list. Therefore, the analysis included all words in Band 5. The following section shows the word-class proportions of the frequent hyphenated words.

### 4.2 Results and Discussion

The 235 frequent hyphenated words in *OED-Online* are summarized according to

their word classes in Table 6.

**Table 6. Word classes of common hyphenated complex words in *OED-Online***

ADJ				ADV		NOUN			VERB
adj	adj adv	adj adv noun	adj noun	adv adj	adv adj noun	noun adj	noun adj adv	noun	verb
81	6	2	32	3	5	18	1	79	8
34.5%	2.6%	0.9%	13.6%	1.3%	2.1%	7.7%	0.4%	33.6%	3.4%
51.5%				3.4%		41.7%			3.4%

Similar to the results shown in Table 5, Table 6 shows that more than half of frequent hyphenated complex words were adjectives (51.5%), and when nouns and adverbs with adjective use were added, the figure amounted to 63.0%. The ratio was followed by nouns (41.7%), while that of verbs was the lowest (3.4%). Among 79 nouns, 19 items (8.1%) were numerals and fractions, which can be prenominal adjectives. In summary, most hyphenated complex words can be considered adjectives and nouns.

## 5. Conclusion

This study has revealed the following. The style manuals heavily focus on hyphenation in compound adjectives or prenominal modifiers, although they mention the rare use of a hyphen in compound verbs. For compound nouns, *the New Oxford Style Manual* (2016) and *the Associated Press Stylebook* (2022) advised not to use a hyphen unless semantic ambiguity or pronunciation awkwardness occurs without a hyphen. In particular, coordinative compounds with dual heritage meaning (e.g., *American Mexican*) no longer need a hyphen in the US. Instead, open orthography is recommended for them.

The most frequent formal pattern of hyphenated compounds is adjective + noun, which is more conspicuous in the US English corpus. The significant differences between *BNC* and *COCA* can be found in UK English, which uses hyphens in compass points and prefixes (*co-*, *pre-*, *semi-*, *sub-*), and in US English, which uses hyphens in interjections and fillers.

Of course, the use of hyphens in the guidelines in style manuals and the actual orthography are influenced by each other and will change over time. In addition, the use

of spellcheckers in word processing software, such as MS Word Editor and Grammarly, must massively restrict our orthography towards a unified form now and in the future.

## References

- American Psychological Association. (2019). *The Publication Manual of the American Psychological Association* (7<sup>th</sup> ed.).
- Associated Press. (2022). *The Associated Press Stylebook: 2022–2024*. Basic Books.
- Davies, M. (2022). Texts. *The Corpus of Contemporary American English*. <https://www.english-corpora.org/COCA/>
- Huddleston, R. & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. CUP.
- Oxford University Press. (2003). *The Oxford Style Manual*.
- Oxford University Press. (2016). *The New Oxford Style Manual*.
- Oxford University Press. (2023–). *The Oxford English Dictionary Online*. <https://www.oed.com/>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.

## Corpora

- Davies, M. (2004). *The British National Corpus* (from Oxford University Press). Available online at <https://www.english-corpora.org/bnc/>
- Davies, M. (2008–). *The Corpus of Contemporary American English*. Available online at <https://www.english-corpora.org/coca/>