The negative and question forms of *used to* in present-day English

Kazuko Fujimoto, Soka University

Abstract

Which negative and question forms of used to *should be taught is a long-discussed question. I have found a variety of views in grammar books and a diversity of negative and question forms in major English learners' dictionaries. The main focus of their discussion is on the acceptability of the use of* used to *after* did. *This study examined the frequency and the spoken/written distribution of each negative and question form in British/American English corpora. My corpus findings indicate that the dictionaries do not fully reflect the actual usage of* used to — *as a whole they seem to reflect grammarians' and dictionary editors' attitudes towards it. It may be suggested that the forms prioritized in teaching should be reconsidered based on the frequency of the negative and question forms of* used to.

1   Introduction

Used to, which refers to a habitual action or state in the past, is called a 'quasi-auxiliary verb' (Peters 2004: 560).[1] Since it is used as both an auxiliary and a lexical verb, opinions have long been divided among grammarians as to which negative and question forms of *used to* are acceptable. The variation in the forms is also observed in learners' dictionaries. This diversity of the forms in the learners' dictionaries and grammar books motivated my corpus-based study of *used to*. In this paper, I examine the frequency and the spoken/written distribution of the negative and question forms of *used to* in British English (BrE) and American English (AmE) corpora so that classroom instruction of these forms may be informed by patterns of actual use.

2   Material

First, I will compare examples and usage notes in the latest edition of eight major advanced learners' dictionaries: *Cambridge advanced learner's dictionary* (third edition, 2008) (*CALD*3), *Collins COBUILD advanced dictionary of English* (seventh edition, 2012) (*CCADE*7), *Longman dictionary of contemporary English* (fifth edition, 2009) (*LDOCE*5), *Macmillan English dictionary for advanced learners* (second edition, 2007) (*MED*2), *Oxford advanced learner's dictionary* (eighth edition, 2010) (*OALD*8), *Collins COBUILD advanced dictionary of American English* (2007) (*CollinsAM*), *Longman advanced American dictionary* (second edition, 2007) (*LAAD*2), and *Merriam-Webster's advanced learner's English dictionary* (2008) (*MWALED*).[2] The first five dictionaries are (basically) for learners of BrE, and the last three are (basically) for those of AmE. I will also refer to English grammar and usage books when it is necessary to compare some more information especially for the forms about which diversity is observed among the dictionaries. Next, the frequency and distribution of the negative and questions forms will be examined in three types of corpora: the Brown family of corpora (BFC),[3] the British National Corpus (BNC),[4] and the Corpus of Contemporary American English (COCA). BFC is a seven million word collection of six comparable corpora of written BrE and written AmE from 1931 to 2006: the BLOB Corpus (BrE, 1931) (BLOB), the LOB Corpus (BrE, 1961) (LOB), the Brown Corpus (AmE, 1961) (Brown), the FLOB Corpus (BrE, 1991) (FLOB), the Frown Corpus (AmE, 1992) (Frown), and the BE2006 Corpus (BrE, 2006) (BE06). BNC is a 100 million word collection of BrE from 1970s to 1993, consisting of 90 per cent written and ten per cent transcribed spoken sub-corpora. COCA is an AmE corpus of more than 450 million words, composed of 80 per cent written and 20 per cent spoken sub-corpora from 1990 to 2012 (as of August 2012).[5]

I will examine the following negative and question forms presented in the learners' dictionaries: *did not/didn't use to*, *did not/didn't used to*, *used not to*, *never used to*, *use to not*, *usen't to*, and *usedn't to* for the negative forms, and *did ... use to ... ?*, *did ... used to ... ?*, and *used ... to ... ?* for the question forms.

3   Results

First the results of the learners' dictionary comparison will be presented, and then the corpus search results will be shown.

3.1 Dictionaries

Table 1 shows the negative and question forms in the entries and usage boxes of *used to* in the eight learners' dictionaries:

Table 1: Negative and question forms in the entries and the usage boxes of *used to*

Negatives	CALD3	CCADE7	LDOCE5	MED2	OALD8	CollinsAM	LAAD2	MWALED
did not/didn't use to	—	+	♦+	♦+	♦+	♦+	♦+	+
did not/didn't used to	—	+	—	—	—	—	—	—
used not to	♦+	♦+	+	+	+	+	—	+
never used to	—	—	+	—	—	—	—	♦+
used to not	—	—	—	—	—	+	—	—
usen't to	—	—	—	+	—	—	—	—
usedn't to	—	—	—	+	+	—	—	—
Questions	CALD3	CCADE7	LDOCE5	MED2	OALD8	CollinsAM	LAAD2	MWALED
did ... use to ...?	—	—	♦+	♦+	♦+	—	♦+	♦+
did ... used to ...?	—	—	—	—	—	—	—	—
used ... to ...?	—	—	—	—	+	—	—	—

'+' indicates the presence of the negative/question form, and '−' its absence.

'♦' indicates that it is the first or only negative/question form in the entries.

The upper part of Table 1 presents results on the negative forms. All the dictionaries except *CALD*3 have *did not/didn't use to*, and this is the first negative form in the five dictionaries. Only *CCADE*7 has *did not/didn't used to*. All the dictionaries except *LAAD*2 give *used not to*. As for *never used to*, *LDOCE*5 and *MWALED* have this form. Only *CollinsAM* contains *used to not*. As regards the contracted forms *usen't to* and *usedn't to*, *MED*2 has both, and *OALD*8 has the latter form only. Compared with the negative forms, there is less variety of question forms in the dictionaries. *CALD*3, *CCADE*7, and *CollinsAM* provide no information about the question forms. It is notable that five of the eight dictionaries give *did ... use to ...?* for the first or only question form. None of the dictionaries gives *did ... used to ... ?*

Examining the dictionary usage notes, I can see that there is a difference of opinion about the use of *did ... use to* and *did ... used to* for the negative and question forms. *MWALED* notes that when *did* is used, *use to* is usually used. According to *MED*2, *did ... used to* is sometimes used, but many people con-

sider this to be wrong. *OALD*8 does not accept the spelling *used to* with *did*. Though *CCADE*7 has *did not/didn't used to*, it does not mention the difference in acceptability between *did ... use to* and *did ... used to*. Quirk *et al*. (1985: 140) state that *did ... used to* "is often regarded as nonstandard". *Webster's dictionary of English usage* (1989: 933–934) remarks that the correct form after *did* is *use to* in AmE. In contrast, Garner (2009: xxxv, 836) defends the form *didn't used to* as being "a universally accepted informal form", claiming that the negative written form should not be *didn't use to* (Garner (ibid.) does not give any comments about the question forms of *used to*). Swan (2005: 595) comments that *did ... used to* is often used instead of *did ... use to*. Carter and McCarthy (2006: 661) accept both, and they seem to have a positive attitude towards *did ... used to* because they give examples not with *did ... use to* but with *did ... used to*. As for *never used to*, only *MWALED* gives this for its first negative form. Let us look at some more information in grammar and usage books. Swan (2005: 595) states that *never used to* is "the most common negative". Garner (2009: 836) considers this form "the standard form that can save you headaches". This means that *never used to* is useful in avoiding the controversy about the use of the double past tense marker (i.e. the use of *used* with *didn't*). As for *used to not,* only *CollinsAM* has this form with no usage information. Information about this form cannot be found in the grammar and usage books examined in this study either. With regard to *used not to* and *used ... to?* (where *used to* is treated as an auxiliary), seven of the eight dictionaries have *used not to*, but only one dictionary has *used ... to?*

Judging from the usage notes in the dictionaries, it may roughly be said that the negative and question forms with *did* are regarded as more informal or typical of spoken English, and those without *did* more formal or typical of written English, used in BrE. It should be added that according to *LDOCE*5, *never used to* (not with *did*) is spoken English.

3.2   Corpus data

I attempted to make a synchronic and diachronic analysis with BFC so that BrE and AmE would be compared by tracing the change of the usage of *used to* from 1931 to 2006. However, as a result of the corpus-family search, I obtained only five negative forms and no question forms of *used to*: one occurrence of *never used to* in each of LOB (BrE, 1961) and Brown (AmE, 1961), one occurrence of *didn't use to* in each of FLOB (BrE, 1991) and Frown (AmE, 1991), and one occurrence of *didn't used to* in Frown (AmE, 1991). The frequency per million words of each case is between 0.87 and 0.88. The small number of occurrences prevents a meaningful comparison between BrE and AmE regarding changes in

the use of negative and question forms of *used to*. Since the size of BFC is about one-fourteenth of BNC and about one-sixty-seventh of COCA, the small number of occurrences might be due to corpus size. Or this result from the BFC search might show the extremely low frequency of the negative and question forms of *used to* in written language from 1931 to 2006. In this section the analysis will mainly be based on BNC and COCA, with which written and spoken language in BrE and AmE can be compared, though a diachronic comparative analysis cannot be made.

Before moving to the main data analysis, I will examine the frequency of *used to* in the written and spoken sub-corpora of BNC and COCA. In BNC, the frequency per million words of *used to* (including negative and question forms) is 86.36 in the written sub-corpus (BNC_W) and 738.82 in the spoken sub-corpus (BNC_S); in COCA, 97.14 in the written sub-corpus (COCA_W) and 113.49 in the spoken sub-corpus (COCA_S). It is clear that *used to* is more frequent in the spoken sub-corpora of BNC and COCA than in their written sub-corpora. The frequency of *used to* is much higher in their spoken sub-corpora than in their written sub-corpora (and the difference is statistically significant at the level of *p<0.0001*).

Tables 2 and 3 show the results of the BNC and COCA search for the negative and question forms of *used to* in the written and spoken sub-corpora, respectively. Column 6 (%DIFF) shows the '% difference' of the normalized frequencies (per million words – PMW) in BNC_W and COCA_W as compared to BNC_S and COCA_S (shown in columns 3 and 5 respectively). In the PMW column of the reference corpus in Table 3, instances of zero frequency were substituted with '1E-19'[6] to avoid the problem of dividing by zero, as proposed in Gabrielatos and Marchi (2011), who also outline the steps for calculating %DIFF.

Table 2: Spoken vs. written in BNC

	Written (BNC_W)		Spoken (BNC_S)				
Negatives	RF	PMW	RF	PMW	%DIFF	LL	*p*-value
did not use to	0	0.000	0	0.000		0.00	
didn't use to	4	0.046	12	1.153	-96.05	36.79	< 0.0001
did not used to	0	0.000	0	0.000		0.00	
didn't used to	4	0.046	19	1.825	-97.51	64.97	< 0.0001
used not to	7	0.080	2	0.192	-58.55	1.01	
never used to	48	0.546	87	8.357	-93.47	225.72	< 0.0001
used to not	0	0.000	0	0.000		0.00	
usen't to	0	0.000	0	0.000		0.00	
usedn't to	0	0.000	0	0.000		0.00	
Questions	RF	PMW	RF	PMW	%DIFF	LL	*p*-value
did ... use to ... ?	3	0.034	10	0.961	-96.45	31.53	< 0.0001
did ... used to ... ?	3	0.034	68	6.532	-99.48	281.19	< 0.0001
used ... to ... ?	0	0.000	0	0.000		0.00	

RF= raw frequency; LL=log likelihood values. Negative values of %DIFF indicate that the form is more frequent in BNC_S than in BNC_W.

Table 3: Spoken vs. written in COCA

	Written (COCA_W)		Spoken (COCA_S)				
Negatives	RF	PMW	RF	PMW	%DIFF	LL	*p*-value
did not use to	1	0.003	0	1E-19	+3E+18	0.46	
didn't use to	22	0.060	8	0.084	-28.84	0.64	
did not used to	0	0.000	1	0.010	-100.00	3.16	
didn't used to	84	0.228	32	0.335	-32.08	3.26	
used not to	7	0.019	2	0.021	-9.44	0.02	
never used to	124	0.336	41	0.430	-21.74	1.78	
used to not	6	0.016	5	0.052	-68.95	3.42	
usen't to	0	0.000	0	0.000		0.00	
usedn't to	0	0.000	0	0.000		0.00	
Questions	RF	PMW	RF	PMW	%DIFF	LL	*p*-value
did ... use to ... ?	5	0.014	3	0.031	-56.87	1.21	
did ... used to ... ?	43	0.117	27	0.283	-58.79	11.87	< 0.001
used ... to ... ?	0	0.000	0	0.000		0.00	

Positive values of %DIFF indicate that the form is more frequent in COCA_W than in COCA_S, and negative values of %DIFF indicate that the form is more frequent in COCA_S than in COCA_W.

It is noticed that the frequency of the negative and question forms of *used to* is extremely low in both the written and spoken sub-corpora in BNC and COCA. In most cases in BNC and all the cases in COCA, the frequency per million words is less than 1.00. As far as the negative and question forms drawn from BNC and COCA are concerned, all the forms except *did not use to* in COCA are more frequent in their spoken sub-corpora than in their written sub-corpora. %DIFF shows that the difference of frequency between BNC_W and BNC_S is larger than that between COCA_W and COCA_S. Very little can be safely inferred about the use of words with zero occurrences in both BNC and COCA. However, the absence of corpus instances is more important in BNC_W and COCA_W than in BNC_S and COCA_S, because the written sub-corpora are much larger than the spoken sub-corpora; that is, their larger size affords more opportunities for words to appear.

As is presented in Table 2, the negative forms *didn't used to*, *didn't use to*, and *never used to* (in decreasing % DIFF order) and the question forms *did ... used to ... ?* and *did ... use to ... ?* (in decreasing % DIFF order) are much more frequent in BNC_S than in BNC_W. The difference is statistically significant ($p<0.0001$). The negative form *used not to* is more frequent in BNC_S than in BNC_W as well, but the difference is not statistically significant. Table 3 shows that among the negative and question forms observed in COCA, only the non-contracted negative form *did not use to* is more frequent in COCA_W than in COCA_S. However, due to its very low frequency in each sub-corpus, the difference is not statistically significant. The only form with a significant difference between COCA_W and COCA_S is the question form *did ... used to ...?* This form is significantly more used in COCA_S ($p<0.001$).

The differences of frequency between BNC_W and COCA_W and between BNC_S and COCA_S were also examined. The negative form *used not to* is more frequent in BNC than in COCA. The difference between their written sub-corpora is significant at $p<0.01$, and the difference between their spoken sub-corpora is significant at $p<0.05$. *Never used to* is more frequent in BNC as well. The difference between their written sub-corpora is significant at $p<0.01$, and the difference between their spoken sub-corpora is significant at $p<0.0001$. It should be added that the negative form *used to not* is more frequent in COCA than in BNC, but the difference is not statistically significant.

Here let us consider the frequency ranking of the negative and question forms in BNC and COCA (see Tables 2 and 3). First, I will examine the three top negative forms and then the ranking of the question forms (three question forms were examined in this study). In BNC_W, the most frequent negative form is *never used to*, the second is *used not to*, and the third *didn't use to* and *didn't used to* (these two forms have the same frequency). In BNC_S, COCA_W, and COCA_S, *never used to* ranks first, *didn't used to* second, and *didn't use to* third. As for the question forms, in BNC_W, *did ... use to ... ?* and *did ... used to ... ?* rank first with the same frequency and *used ... to ... ?* third with no instances. BNC_S, COCA_W, and COCA_S have the same frequency ranking: *did ... used to ... ?* ranks first, *did ... use to ... ?* second, and *used ... to ... ?* third with no instances. It is worthwhile mentioning that the most frequent negative form is *never used to* in all the four sub-corpora: BNC_W, BNC_S, COCA_W, and COCA_S. The frequency of *did ... used to* is the same as that of *did ... use to* in BNC_W and higher than that of *did ... use to* in BNC_S, COCA_W, and COCA_S. However, it should be noted that there is a transcription problem because in spoken English it is difficult to make a distinction between *use to* and *used to* (Peters op. cit.: 560).[7]

Finally, the ratio between the negative forms and the question forms of *used to* is examined. In both BNC and COCA, the frequency of the negative forms is higher than that of the question forms (see Tables 2 and 3). In BNC_W the negative forms are 10.5 times more frequent than the question forms; in BNC_S 1.5 times; in COCA_W 5.1 times; in COCA_S 3.0 times.

4   Discussion

Comparing the dictionaries and grammar and usage books, I can conclude that there is a clear difference of opinion about the use of the following forms: *did ... use to* and *did ... used to*. If the double past tense *did ... used to* is thought to be incorrect, the forms *didn't use to* and *did ... use to ... ?* are likely to be chosen. The fact that seven of the eight dictionaries have *didn't use to*, and that five of the eight dictionaries have *did ... use to ... ?* might indicate that these forms are recommended for teaching. On the other hand, if *used to* is regarded as an invariant idiomatic phrase, *used to* can be used after *did*. My corpus findings show that for both the negative and question forms, after *did*, *used to* is more frequent than *use to* in BNC_S, COCA_W, and COCA_S (in BNC_W, they have equal frequency), though as regards the spoken sub-corpora it might be necessary to take into account the transcription problem of *use to* and *used to* as mentioned above. Only one of the eight dictionaries gives *didn't used to*, and none of them has *did ... used to ...?* The corpus findings also reveal that *never used to* is the most frequent of the negative forms in all the sub-corpora examined. However, this negative form can be found in only two of the eight learners' dictionaries. The corpus search has also shown that the frequency of the negative and question forms is extremely low, and that the frequency of the question forms is lower than that of the negative forms. The reason why three of the eight learners' dictionaries do not give any question forms of *used to* at all might be their very low frequency.

Judging from the frequency ranking and the controversy discussed in this paper, *never used to* can be used for the negative form. However, when the extremely low frequency of the negative and question forms of *used to* is considered, different constructions may be used for both forms. *MWALED* paraphrases "He never *used to* smoke" as "[H]e never smoked in the past". For the question construction, Leech and Svartvik (2002: 257) give an example, stating, "[A] different construction is often a more natural choice, for example: *Did he smoke when he first knew him?*"

5   Conclusion

The learners' dictionaries in general provide different results from my corpus findings. The dictionary editors seem to hesitate to approve *did ... used to* because this form is felt to be less acceptable, though my findings show that it is more common than *did ... use to*. They also seem to pay little attention to the negative form *never used to*, which is the most frequent negative form in my corpus search. Leech (2011: 13–14, 18, 27) emphasizes the importance of frequency information for "language learning and teaching purposes" (the principle "more frequent = more important to learn"). I would like to say that more frequent negative and question forms of *used to* should be prioritized in teaching or when we consider their low frequency, different constructions which express similar meanings should be taught.

Acknowledgements

Notes

1. Quirk *et al*. (1985: 137–140) call this a 'marginal modal'; Biber *et al*. (1999: 490), a 'semi-modal'.
2. The following eight dictionaries are examined: *Cambridge advanced learner's dictionary* (third edition). 2008. Cambridge: Cambridge University Press, *Collins COBUILD advanced dictionary of American English*. 2007. Boston: Thomson Heinle, *Collins COBUILD advanced dictionary of English* (seventh edition). 2012. Boston: Heinle & Heinle Publishers, *Longman advanced American dictionary* (second edition). 2007. Harlow: Pearson Education Limited, *Longman dictionary of contemporary English* (fifth edition). 2009. Harlow: Pearson Education Limited, *Macmillan English dictionary for advanced learners* (second edition). 2007. Oxford: Macmillan Education, *Merriam-Webster's advanced learner's English dictionary*. 2008. Springfield, MA: Merriam-Webster, Inc., and *Oxford advanced learner's dictionary* (eighth edition). 2010. Oxford: Oxford University Press.
3. The Brown Family (extended): *powered by CQPweb* was used. Available at http://cqpweb.lancs.ac.uk/ (accessed 21 August 2012).

4. The CQP-Edition of *BNCweb* was used. Available at http://bncweb.lancs.ac.uk/ (accessed 22 August 2012).

5. Since COCA is a monitor corpus, it is updated regularly. As of 22 August 2012, the latest update was in June 2012. Available at http://corpus.byu.edu/coca/ (accessed 22 August 2012).

6. '1E-19' denotes the number 0.0000000000000000001. '3E+18' in the %DIFF column in COCA_S denotes 3,000,000,000,000,000,000. For details, see Gabrielatos and Marchi (2011).

7. See Hoffmann *et al*. (2008: 37–38) about transcription problems in BNC, and <http://corpus.byu.edu/coca/> for the information on the COCA spoken transcripts.

References

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.

Carter, Ronald and Michael McCarthy. 2006. *Cambridge grammar of English*. Cambridge: Cambridge University Press.

Gabrielatos, Costas and Anna Marchi. 2011. Keyness: Matching metrics to definitions. Paper presented at *Corpus Linguistics in the South. Theoretical-methodological challenges in corpus approaches to discourse studies – and some ways of addressing them*, University of Portsmouth, 5 November 2011. Available at http://eprints.lancs.ac.uk/51449/ (accessed 13 February 2012).

Gabrielatos, Costas and Anna Marchi. 2012. Keyness: Appropriate metrics and practical issues. Paper presented at *CADS International Conference 2012. Corpus-assisted discourse studies: More than the sum of discourse analysis and computing?*, University of Bologna, 14 September 2012. Available at http://repository.edgehill.ac.uk/4196/ (accessed 8 November 2012).

Garner, Bryan A. 2009. *Garner's modern American usage* (second edition). New York: Oxford University Press.

Gilman, Ward E. 1989. *Webster's dictionary of English usage*. MA: Merriam-Webster Inc.

Hands, Penny (ed.). 2011. *Collins CUBUILD English grammar* (third edition). Glasgow: HarperCollins Publishers.

Hoffmann, Sebastian, Stefan Evert, Nicholas Smith, David Lee and Ylva Berglund. 2008. *Corpus linguistics with BNCweb – a practical guide*. Frankfurt am Main: Peter Lang.

Leech, Geoffrey. 2011. Frequency, corpora and language learning. In F. Meunier, S. Cock, G. Gilquin and M. Paquot (eds.). *A taste for corpora*, 7–31. Amsterdam: John Benjamins Publishing Company.

Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

Leech, Geoffrey and Jan Svartvik. 2002. *A communicative grammar of English* (third edition). Harlow: Pearson Education Limited.

Mair, Christian and Geoffrey Leech. 2006. Current change in English syntax. In B. Aarts and A. MacMahon (eds.). *The handbook of English linguistics*, 318–342. Oxford: Blackwell.

Peters, Pam. 2004. *The Cambridge guide to English usage.* Cambridge: Cambridge University Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Swan, Michael. 2005. *Practical English usage* (third edition). Oxford: Oxford University Press.