

Unit 18 *HELP* or *HELP to*: what do corpora have to say? (Case study 2)

18.1 Introduction

This unit further explores, via a case study of the type of infinitive following *HELP*, variations in language varieties as discussed in units 10.5 and 14.4-14.5, and the recent language change as discussed in unit 15.5. This case study will use the tests of statistical significance introduced in unit 6.4.

Help is one of the most frequent words in the English language, ranking as the 245th most frequent word in the word frequency list of the British National Corpus (BNC). When we look at the most frequent verbs (lemmatized) in the BNC, *HELP* rises to the 72nd in the word frequency list, occurring 528.62 times per million words.

Furthermore, *HELP* is a unique verb in that it can control either a full infinitive or a bare infinitive, either with or without an intervening noun phrase (NP), as in the following examples, cited from the BNC:

- (a) The pattern *help to V*
Perhaps the book helped to prevent things from getting even worse.
- (b) The pattern *help NP to V*
I thought I could help him to forget.
- (c) The pattern *help V*
Savings can help finance other Community projects.
- (d) The pattern *help NP V*
We helped him get to his feet and into a chair.

While most English grammars and dictionaries take it for granted that the omission of *to* following *HELP* is arbitrary (e.g. Chalker 1984: 106; Eastwood 1992: 106; Murphy 1985: 110; Longman 1986, 1993; Collins 1995), others do take into account some factors that are possibly relevant to the omission of *to*. For example, Biber et al (1999: 73) observe that ‘AmE [American English] has an especially strong preference for the pattern *verb + bare infinitives* although the bare infinitive is more common than the *to*-infinitive in both varieties.’ This case study investigates some factors that may potentially influence a language user’s choice of a full or bare infinitive following *HELP*, namely, varieties of English, language change over the three decades from 1961 to 1991, as well as the syntactic conditions of an intervening noun phrase (NP), the infinitive marker *to* preceding *help*, and the passive construction. For a fuller account of this study, readers can refer to McEnery and Xiao (2005b), on which this case study is based.

In this case study, we will give you a step-by-step account of how to conduct the investigation using the corpus tool MonoConc Pro. We will also show you how to interpret the frequency data obtained from the concordancer by applying the statistics package SPSS for Windows.

18.2 Concordancing

As the first step in this investigation, we will concordance each of the four corpora of the Brown family to get the frequencies of patterns (a–d) introduced in unit 18.1. Note that the inflections of the controlling verb *HELP* (e.g. *help*, *helped*, *helps* and *helping*)

will also be counted. We suppose that the four corpora to be used in this case study are located in individual directories on your local drive:

LOB c:\My corpora\LOB
 FLOB c:\My corpora\FLOB
 Brown c:\My corpora\Brown
 Frown c:\My corpora\Frown

18.2.1 The pattern *HELP to V*

To get the frequency of the pattern *HELP to V* in LOB, do the following:

1. Start MonoConc Pro and you will see the interface of the concordancer as shown in Fig. 18.1.
2. Select *File* → *Load corpus* from the menu (Fig. 18.2).
3. You will see the *Select File(s) to Open* window (Fig. 18.3).
4. Locate and open the directory for LOB. You will find 15 files named LOB_A to LOB_R. Highlight the file named LOB_A by clicking on it with your mouse. Hold down the *SHIFT* key and the down arrow key on your keyboard until all of the 15 corpus files are selected, as shown in Fig. 18.4.
5. Press *Open*. In a few seconds, the corpus will be loaded into the concordancer. In the bottom left hand corner of the window, the program shows that there are 15 files currently loaded into the concordancer (Fig. 18.5).

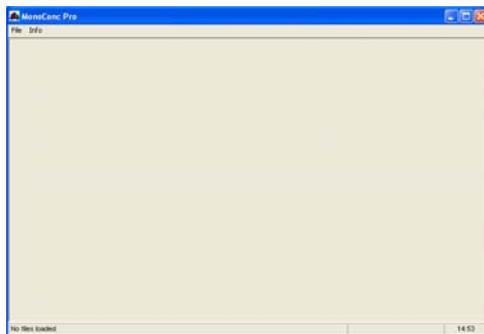


Fig. 18.1 MonoConc Pro interface

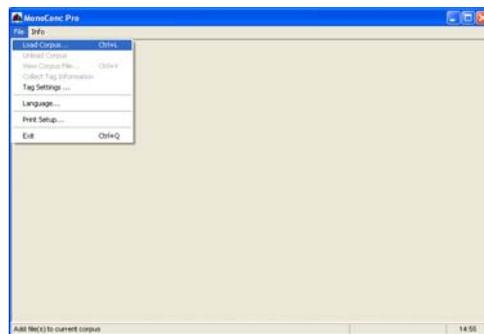


Fig. 18.2 Loading corpus

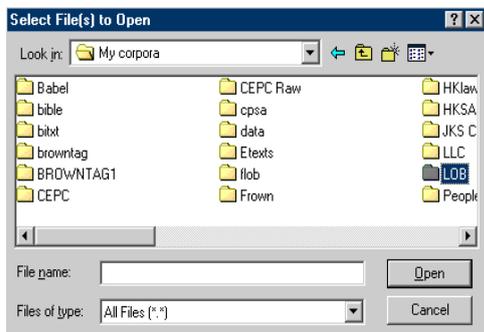


Fig. 18.3 Selecting file(s) to open

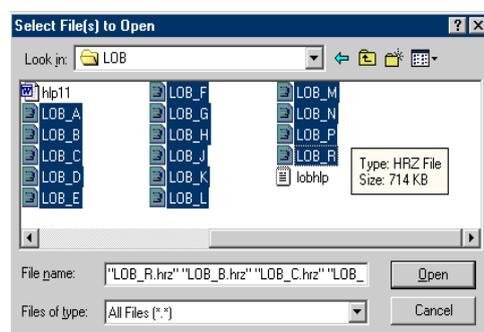


Fig. 18.4 Selecting corpus files

Unit 18 *Help* or *help* to: what do corpora have to say? 3

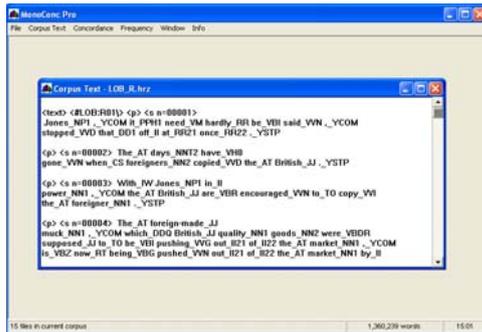


Fig. 18.5 Fifteen files loaded

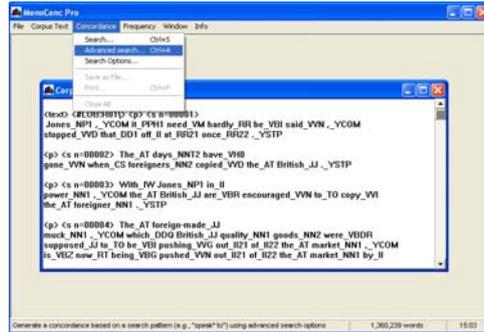


Fig. 18.6 Activating the concordancer

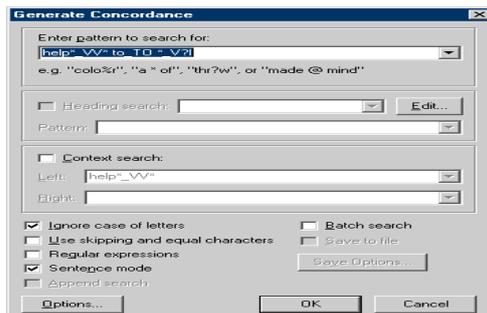


Fig. 18.7 Entering the search string

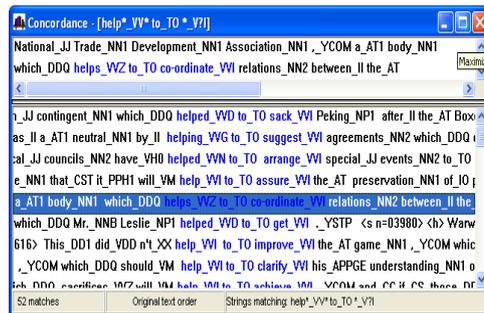


Fig. 18.8 'help to V' pattern

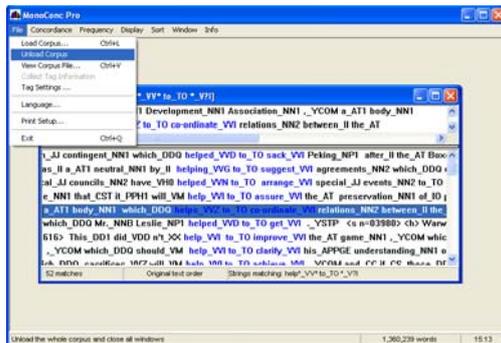


Fig. 18.9 Unloading the corpus

6. Select Concordance → Advanced Search, as shown Fig. 18.6.

7. When a new window appears, check the box preceding *Sentence mode* to ensure that the search string occurs in the same sentence. The box preceding *Ignore case of letters* is checked by default. Enter your search pattern *help* _VV* to _TO *_V?I* exactly as shown in Fig. 18.7 and press the OK button.

Here, the special character * allows the concordancer to match any number of characters so that the inflected forms of *HELP* are included. The special character ? stands for any single character. Therefore, the string *V?I* matches the infinitive form of any verb including *HAVE* and *BE*.

8. In a few seconds, you will see the concordancing result as shown in Fig. 18.8.

There are 52 matches of the search string. The lower part of the window shows the matched concordance lines in the KWIC (key-word-in-context) format, while the upper part of the window shows more context of a particular concordance line. If you

want to have a better view of the concordance lines and their contexts, press the *Maximise* button in the upper right hand corner of the concordance window.

9. Record the frequency of the pattern *HELP to V* in LOB. Select *File* → *Unload corpus* to initialize the concordancer, as shown in Fig. 18.9.

Repeat steps 1–9 for the other three corpora. The resulting frequencies should match those in Table 18.1. Note that the search patterns in this case study rely on part-of-speech (POS) tags. The concordance lines were not validated manually. Therefore, *You_PPY need_VV0 his_APPGE help_VVI to_TO cope_VVI ,_YCOM darling_NNI ._YSTP* (<s n=00490>) is included as an instance for the pattern *help to V*. Readers are advised to refer to McEnery and Xiao (2005b) for a more accurate and complete account of factors contributing to the choice of a full or bare infinitive.

Table 18.1 Frequencies of the pattern '*HELP to V*'

Corpus	Frequency
LOB	52
FLOB	48
Brown	38
Frown	32

18.2.2 The pattern of *HELP V*

To get the frequency of the pattern *HELP V* in LOB, repeat steps 1–9 in unit 18.2.1, replacing the search pattern in step 6 with *help*_VV* *_V?I*. Do the same for the other three corpora. The resulting frequencies should match those in Table 18.2.

Table 18.2 Frequencies of the pattern '*HELP V*'

Corpus	Frequency
LOB	11
FLOB	63
Brown	47
Frown	64

18.2.3 The pattern of *HELP NP to V*

An intervening noun phrase (NP) may vary in length. To simplify the concordancing process, we will only consider the case of a one-word intervening NP typically composed of a single noun or a pronoun. To get the frequency of the pattern *HELP NP to V* in LOB, do the following:

1. Repeat steps 1–6 from unit 18.2.1.
2. Press the *Advanced* button. In the new window, select *Batch search*, as shown in Fig. 18.10.
3. Press the *Edit search patterns* button. Enter the search patterns (*help*_VV* *_N* to_TO *_V?I* and *help*_VV* *_P* to_TO *_V?I*) on two separate lines as shown in Fig. 18.11. Press the *OK* button to return to the previous window.
4. Press the *OK* button and in a few seconds you will see the resulting concordance window, as shown in Fig. 18.12. There are 21 matches of the search strings. If you want to have a better view of the concordance lines and their contexts, press the *Maximise* button in the upper right hand corner of the concordance window.

5. Record the frequency of the pattern *HELP NP to V* in LOB. Select *File* → *Unload corpus* to reinitialize the concordancer, as shown in Fig. 18.9.

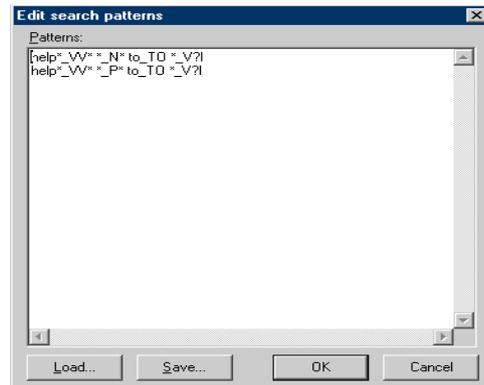
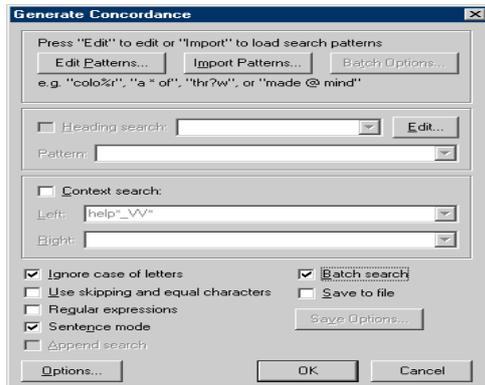


Fig. 18.10 The Advanced search window Fig. 18.11 The Edit batch search patterns



Fig. 18.12 The ‘*HELP NP to V*’ pattern

Repeat steps 1–5 above for the other three corpora. The resulting frequencies are shown in Table 18.3.

Table 18.3 Frequencies of the pattern ‘*HELP NP to V*’

Corpus	Frequency
LOB	21
FLOB	21
Brown	12
Frown	10

18.2.4 The pattern of *HELP NP V*

To get the frequency of the pattern *HELP NP V* in LOB, repeat steps 1–5 in unit 18.2.3, replacing the search pattern in step 3 with *help*_VV* *_N* *_VVI* and *help*_VV* *_P* *_VVI*. Do the same for the other three corpora. The resulting frequencies are shown in Table 18.4.

Having obtained the necessary frequency data from the corpora, we are now ready to examine the possible influence of language variety, language change, and an intervening NP on a language user’s choice of infinitive variants. The frequencies themselves do not tell you whether a particular factor affects the choice. You must interpret the frequencies by conducting statistical tests to determine whether the difference in frequencies obtained from the relevant corpora is statistically significant (see unit 6.4). If it is, you can conclude that the factor does influence the choice of a full or bare infinitive. Note, however, that frequency data must be interpreted with

caution. As sample size may affect the level of statistical significance, frequencies obtained from corpora of different sizes must be normalized to a common base (usually the average size of the corpora used, see unit 6.2). Since the four corpora used in this case study are of equal size, i.e., each containing roughly one million words, normalization is not necessary.

Table 18.4 Frequencies of the pattern ‘*HELP NP V*’

Corpus	Frequency
LOB	6
FLOB	21
Brown	36
Frown	59

18.3 Language variety

In this section, we will examine the two major varieties of English represented in the corpora studied, namely, American English (AmE) and British English (BrE). Onions (1965), Lind (1983: 264) and Biber et al (1999: 73) observe that bare infinitives are more frequent in AmE than in BrE. To verify the validity of this observation, we will first rearrange the frequencies as shown in Table 18.5. As we are only interested in a comparison of language varieties in this section, the data gathered on intervening NPs will not be used; rather, the combined counts of infinitives with and without an intervening NP are used. We will use the frequencies in the column *Total* in Table 18.5 to compare LOB and Brown on the one hand, and FLOB and Frown on the other hand. The log-likelihood (LL) score is a reliable test for this purpose.

Table 18.5 Full and bare infinitives in AmE and BrE

Variety	Corpus	Inf-type	No NP	With NP	Total
BrE	LOB	Full	52	21	73
		Bare	11	6	17
	FLOB	Full	48	21	69
		Bare	63	21	84
AmE	Brown	Full	38	12	50
		Bare	47	36	83
	Frown	Full	30	10	40
		Bare	64	59	123

A convenient way to calculate the LL score is to use the statistics package SPSS for Windows by following the steps given below (assuming that SPSS for Windows Release 10.1):

1. Select *Start* → *Programs* → *SPSS for Windows* → *SPSS 10.1 for Windows*. You will see the interface for SPSS. Select *Type in data* and press the *OK* button as shown Fig. 18.13.
2. Select *Variable view* to define variables. In this case, we have three variables: corpus name (*corpus*), infinitive type (*type*), and frequency (*frequent*). The first two are string variables while the third is a numeric variable. As we are dealing with integers, the numerical is defined with no decimals (Fig. 18.14).
3. Select *Data view* to type in data as shown in Fig. 18.15.

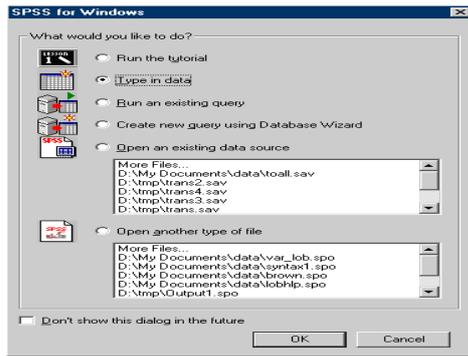


Fig. 18.13 SPSS for Windows

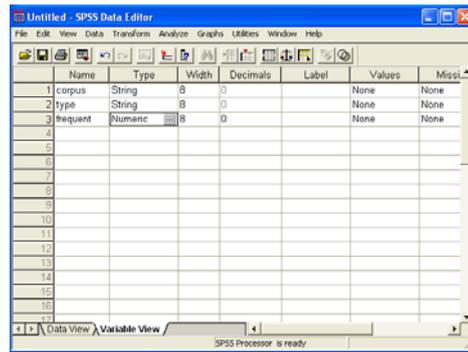


Fig. 18.14 Defining variables

	corpus	type	frequent
1	LOB	full-inf	73
2	LOB	bare-inf	17
3	Brown	full-inf	50
4	Brown	bare-inf	83

Fig. 18.15 Typing in the data

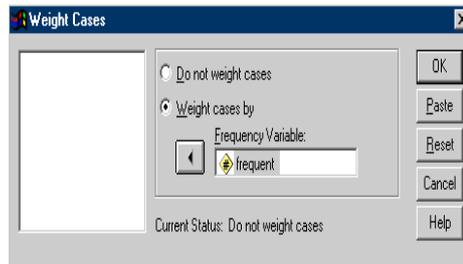


Fig. 18.16 Weighting the cases

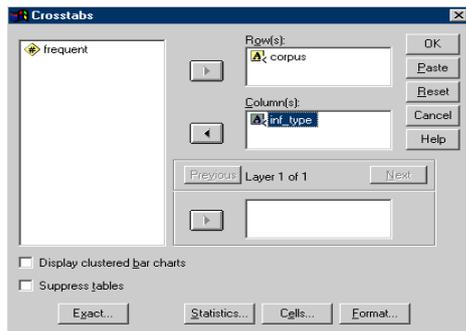


Fig. 18.17 Defining row and column

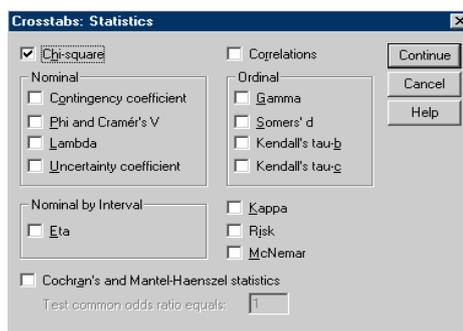


Fig. 18.18 Selecting the statistical test

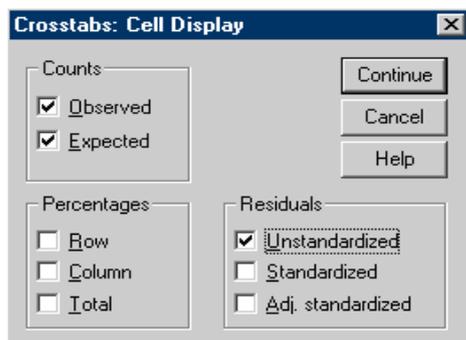


Fig. 18.19 Selecting the cells

4. As we type in the numerical value for the variable *frequent* directly, the value needs to be weighted. Select *Data* → *Weight cases* from the menu. You will see a window *Weight cases*. Select *Weight cases by* and highlight the variable *frequent* in the left panel, then click on the right arrow and press the *OK* button (Fig. 18.16).

5. You will then return to the previous window. Select *Analyze* → *Descriptive statistics* → *Crosstab* from the menu and you will come to the *Crosstab* window. Define the *Row(s)* with the variable *corpus* and the *Column(s)* with the variable *type* (Fig. 18.17).
6. Press the *Statistics* button, select *Chi-square*, and then press *Continue* (Fig. 18.18).
7. Press the *Cells* button, and select both *Observed* and *Expected* under the label of *Counts*. As the frequencies used are not normalized, also select *Unstandardized* under the label of *Residuals*. Press *Continue* (Fig. 18.19).
8. Now you return to the *Crosstab* window. Press *OK* and in a few seconds you will be taken to the output window. You can ignore the first two tables and come to the table labelled *Chi-Square Tests*.

Note that the calculated log-likelihood ratio is 43.435 for 1 degree of freedom (d.f.), and the 2-sided significance level (0.000) is less than 0.001. If you consult an appendix table labelled the chi-square distribution in a textbook or reference book for statistics (e.g. Oakes 1998: 266), you will find the critical value for statistical significance at $p < 0.001$ is 10.83 with 1 d.f. The calculated LL score is considerably greater than this critical value. Therefore, we can more than 99.9 percent confident that the difference in the frequencies of full and bare infinitives in BrE and AmE in the 1960s is statistically significant.

Before exiting SPSS, clear the data in step 3 (Fig. 18.15) and enter the frequencies from FLOB and Frown to compare BrE and AmE in the 1990s. This time, you can skip steps 4, 6 and 7 and simply repeat steps 5 and 8. If you have exited SPSS and restarted the package, however, you will have to repeat the whole process. You will find the LL score calculated on the basis of frequencies from FLOB and Frown is 14.750 for 1 d.f., which is greater than the critical value 10.83. The 2-sided significance level (0.000) is still less than 0.001. This means that the difference between BrE and AmE in the early 1990s is also statistically significant.

It is interesting to note that the contrast between BrE and AmE in the 1990s is not so marked as in the 1960s, as reflected by the much smaller LL score for the data in the 1990s. For the moment we will simply note this difference, though we will return to it in unit 18.4.

HELP (NP) do is prevalent in AmE simply because this construction is of American provenance, though it has penetrated rapidly into BrE (cf. Onions 1965; Lind 1983: 264). As such, Zandvoort (1966) classifies this construction as an Americanism and claims that ‘except in American English, however, *to help* usually takes an infinitive with *to*’ (cf. Lind 1983: 264). However, if we take language change into account, which we will do in unit 18.4, we find that Zandvoort’s claim no longer holds.

18.4 Language change

Language change over time has affected the choice of a full or bare infinitive following *HELP* in both AmE and BrE. The bare infinitive after *HELP* is not mentioned in the *Concise Oxford Dictionary* (COD) but is now considered to be dialectal or vulgar in the *Oxford English Dictionary* (OED). The *Supplement to the OED* removes this label and judges it as being ‘a common colloq. form’ (cf. Kjellmer 1985: 264). Vallins (1951: 56) claims that ‘the construction is not seriously questioned now (as it might have been twenty years ago) even in normal literary writing.’ This section examines recent data to demonstrate the possible effect of language change on a

language user's choice of alternative infinitive variants. We will use the frequency data in Table 18.5 to compare BrE in the 1960s and in the 1990s on the one hand, and AmE in the corresponding periods on the other hand.

Let us first compare LOB and FLOB using the same procedure as in unit 18.3. Clear the data in step 3 (Fig. 18.15) and enter the frequencies from LOB and FLOB. Repeat steps 5 and 8. You will find the LL score calculated on the basis of frequencies from the two corpora is 32.059 for 1 d.f., which is greater than the critical value 10.83. The 2-sided significance level (0.000) is still less than 0.001.

Do the same with the frequencies from Brown and Frown. You will find the LL score calculated on the basis of frequencies from the two corpora is 5.884 for 1 d.f., which is greater than the critical value 3.84 for statistical significance at $p < 0.05$. The 2-sided exact significance level is 0.016.

Statistical tests show that in both BrE and in AmE, language change over the three decades has indeed exerted influence over the choice between infinitive variants. Fig. 18.20 illustrates this point well. As can be seen in the figure, the proportions of bare infinitives in both BrE and AmE increased considerably from the 1960s to the 1990s. The contrast in BrE is even more marked than in AmE, as reflected by the much greater LL ratio and higher significance level for the BrE data. The reason for this apparent difference is that by 1961 AmE was already much more tolerant of bare infinitives than BrE (see Fig. 18.20). Consequently a greater shift towards the use of the bare infinitives in the period 1961–1991 was possible for BrE, resulting in a more marked change.

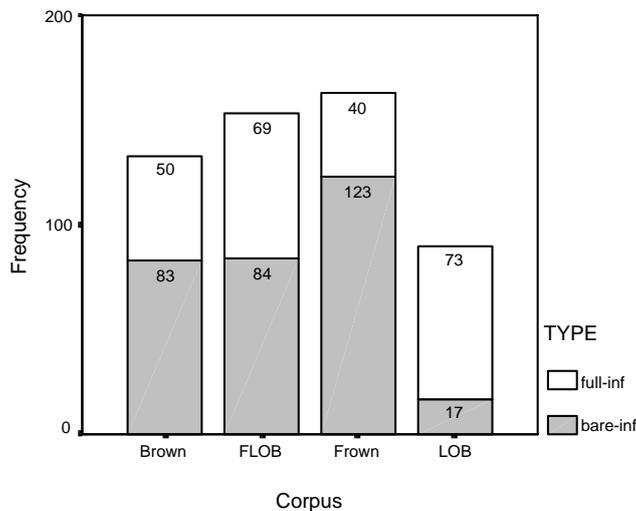


Fig. 18.20 Frequencies of infinitives in the 4 corpora

18.5 An intervening NP

Biber et al (1999: 73), Lind (1983: 269) and Kjellmer (1985: 158) claim that bare infinitives occur more frequently after *HELP* with an intervening NP than where there is no intervening NP. To determine whether this argument is viable, we will first rearrange the frequencies as shown in Table 18.6. It can be seen from the table that an intervening NP typically contributes an increase of 18 percent to the proportion of bare infinitives in the AmE data. In the BrE data, however, the effect of an intervening NP is unpredictable.

Table 18.6 The effect of an intervening NP (proportions)

Corpus	Inf-type	No NP		With NP		± %
		Fre.	%	Fre.	%	
LOB	Full	52	82.54	21	77.78	-4.76
	Bare	11	17.46	6	22.22	+4.76
FLOB	Full	48	43.24	21	50.00	+6.76
	Bare	63	56.76	21	50.00	-6.76
Brown	Full	38	44.71	12	25.00	-19.71
	Bare	47	55.29	36	75.00	+19.71
Frown	Full	30	31.91	10	14.49	-17.42
	Bare	64	68.09	59	85.51	+17.42

This finding is further supported by the LL tests, as shown in Table 18.7. For a difference to be statistically significant, the calculated LL score must be greater than the critical value for significance at $p < 0.05$ (1 d.f.), 3.84. As can be seen from the table, in the BrE data, the increase or decrease in the proportion of bare infinitives contributed to by an intervening NP is not statistically significant, whereas in the AmE data, such increases are significant in both corpora. This provides further evidence that AmE behaves differently from BrE (cf. unit 18.3). It is also of interest to note that the increase in the proportion of bare infinitives in AmE in the 1990s (significance level $p = 0.009$) is more significant than that in the 1960s (significance level $p = 0.022$). This finding is in line with our conclusion in unit 18.4.

Table 18.7 The effect of an intervening NP (LL tests)

Corpus	Inf-type	No NP	With NP	LL ratio	Sig. level
LOB	Full	52	21	0.273	0.601
	Bare	11	6		
FLOB	Full	48	21	0.560	0.454
	Bare	63	21		
Brown	Full	38	12	5.239	0.022
	Bare	47	36		
Frown	Full	30	10	6.819	0.009
	Bare	64	59		

18.6 The infinitive marker preceding *HELP*

This section tests the claim, made by Biber et al (1999: 737), Lind (1983: 269) and Kjellmer (1985: 159), that the infinitive marker *to* preceding the controlling verb *HELP* is a decisive syntactic condition encouraging the omission of *to* following *HELP*. As an intervening NP may influence the choice of a full or bare infinitive in AmE, we will exclude this factor. Considering that the pattern *to help* occurs only rarely in the individual corpora used, we will take our four corpora as a whole. MonoConc Pro allows one to load different corpora consecutively. In this section, we will also show you how to search corpora using regular expressions and how to sort concordance lines in a certain order. To obtain the relevant frequency data, do the following:

1. Activate MonoConc Pro and load the four corpora one at a time using the procedures in unit 18.2.1. You will see, in the bottom left hand corner of the window, 60 files are selected.
2. Select *Concordance* → *Search* from the menu, a window will appear for you to enter a search pattern.

3. Press the *Advanced* button and check the boxes labelled *Regular expressions* and *Sentence mode*. The box labelled *Ignore case of letters* is checked by default. Enter the search string (*to_TO help_VVI .*_V.I*) exactly as shown in Fig. 18.21. There are 69 matches for this search pattern. Record the number of matches.

It should be noted that the special characters * and ? are used differently as wildcards and in regular expressions. Wildcards are common in a number of programs, including word processors and concordancers. Wildcards allow one to search words or expressions where one character is unspecified (with the unspecified character represented by a question mark) or many characters are unspecified (represented by an asterisk). Regular expressions also allow one to search for words or expressions where characters are unspecified. However, the symbols used in regular expressions, while at times similar to wildcards, have different meanings. The wildcard ? is equivalent to the regular expression . whereas the wildcard * is equivalent to the regular expression .* . When the box preceding *Regular expression* is unchecked (as in step 7 in unit 18.2.1, see Fig. 18.7), the asterisk matches any number (zero or more) of character(s). Note that when you use WordSmith tools, the special character * is used as a wildcard, as in step 7. When the option of regular expressions is selected, however, the syntax of regular expression must be followed, where the special character . stands for any single character. The asterisk is only a quantifier. The asterisk means 'equal to or more than zero' occurrence(s) of the preceding character. Therefore, if you forget to include the special character ., the concordancer will find no match at all, because there is no instance of *to help* followed by zero or more white spaces and an underscore (e.g. *to_TO help_VVI _VVI*). When a quantifier is absent, its default value is 1. As such, *V.I* matches *VVI*, *VHI*, *VDI* and *VBI*.

4. Repeat step 3 and enter the search string *to_TO help_VVI to_TO .*_V.?I*. There are only 3 matches. If you highlight the concordance lines, you will find that all of them occur in BrE (1 in LOB and 2 in FLOB).

5. Repeat step 3 and enter the search string *help_VV[OI] .*_V.?I*. There are 115 matches for this search pattern. As this search string also matches *to help V*, we need to subtract the count of *to help V* from the total of 115. Select *Sort* → *1 left* → *No Second Sort* from the menu, the concordance lines will be sorted alphabetically according to the 1st word on the left of the search term in ascending order (Fig. 18.22).

6. Locate the first occurrence of *to help V* and move down to count until the last occurrence (Fig. 18.23). You will find 69 such instances. The count is exactly what we find in step 3. Subtract 69 from 115, the result 46 is the frequency of the pattern *help V*.

7. Repeat step 3 and enter the search string *help_VV[OI] to_TO .*_V.?I*. The string *VV[OI]* matches both *VV0* and *VVI*. We need to include the latter because *help* can be preceded by a modal or auxiliary verb such as *can* and *did*. There are 52 matches for the search string. As 3 of these are of the pattern *to help to V*, there are 49 instances of the *help to V* type.

Unit 18 *Help* or *help to*: what do corpora help to say? 12

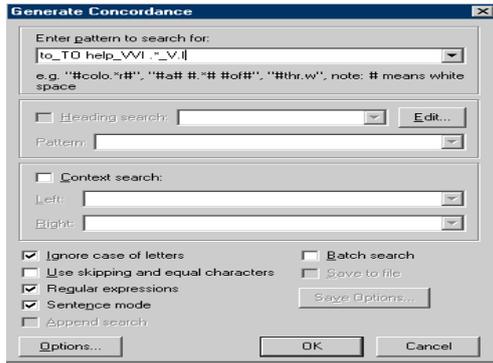


Fig. 18.21 Using regular expression

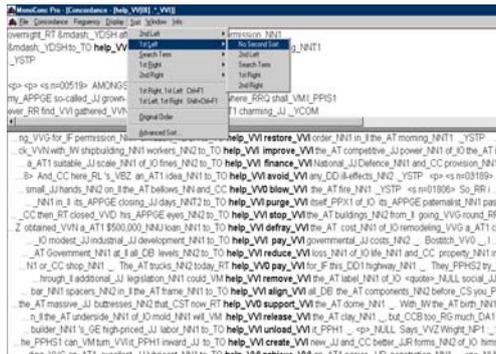


Fig. 18.22 Sorting the concordance lines

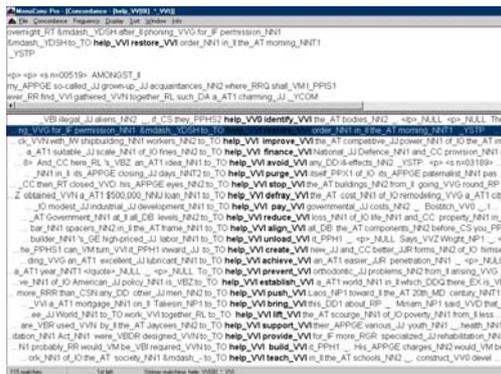


Fig. 18.23 Sorted according to 1st left

The frequency data obtained above is rearranged as shown in Table 18.8 and Fig. 18.24.

Table 18.8 Frequencies of full and bare infinitives after (*to*) *help*

Inf-type	<i>help</i>	<i>to help</i>
Full	49	3
Bare	46	69

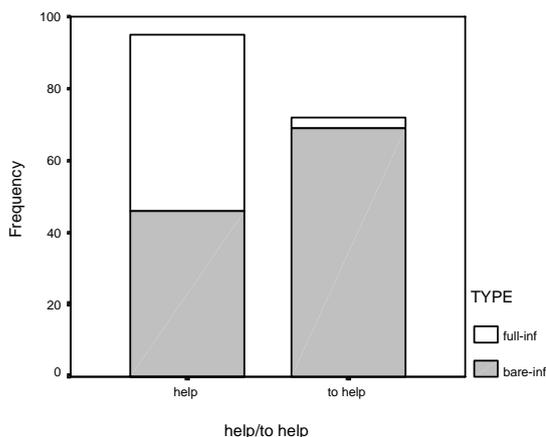


Fig. 18.24 A contrast of *help* and *to help*

As can be seen in Fig. 18.24, there is a marked contrast between *HELP* and *to help*. While bare infinitives account for less than 50 percent of the total number of infinitives following *help*, they make up more than 95 percent when an infinitive

marker precedes *help*. In the AmE data, the pattern *to help to V* is simply non-existent. The log-likelihood (LL) test shows that the difference as illustrated in Fig. 18.24 is statistically significant. The calculated LL score is 50.601, which is considerably greater than the critical value 10.83 for statistical significance at $p < 0.001$. Therefore, we are more than 99.9 percent confident that *to* preceding *help* encourages a language user to choose a bare infinitive after *HELP*.

18.7 The passive construction

Palmer (1965: 169) claims that the ‘passive occurs [...] only with *to*: They were helped to do it.’ To test this claim, we will first find the frequencies of passive and non-passive constructions in our corpora. As the passive is an infrequent linguistic feature, we will not contrast the use of the passive synchronically or diachronically. Rather we will study all of the examples from the four corpora together. This section will show you how to interpret the frequency data for an infrequent linguistic feature. Using the same concordancing procedure as in unit 18.6, enter the search strings as shown in Table 18.9. The frequencies you gather will match those in the table. Search string No. 1 matches the pattern *BE helped to V* while No. 2 matches *BE helped V*. In search string Nos. 3 and 4, *help.*_VV.** matches both the non-inflected form and all inflections of the controlling verb *HELP*, including passive constructions. Therefore, the counts of passive constructions should be subtracted from the frequencies for Nos. 3 and 4 to obtain the frequencies of non-passive constructions, as shown in Table 18.10.

Table 18.9 Search strings for (non-)passive constructions

No.	Search string	Frequency
1	.*_VB.* helped_VVN to_TO .*_V.?I	6
2	.*_VB.* helped_VVN .*_V.?I	0
3	help.*_VV.* to_TO .*_V.?I	170
4	help.*_VV.* .*_V.?I	185

Table 18.10 Frequencies of infinitives for (non-)passives

Pattern	Frequency
<i>BE helped to V</i>	6
<i>BE helped V</i>	0
<i>help to V</i>	164
<i>help V</i>	185

To determine whether the difference between the passive and non-passive constructions is statistically significant, we will not use the log-likelihood test as at least one of four cells in the 2x2 contingency table has an expected value less than 5. Under such conditions the log-likelihood test is unreliable (cf. Howitt and Cramer 2001: 121–123). Rather, we will use Fisher’s exact test to determine the exact significance level as it is more reliable in these circumstances.

The Fisher’s exact significance test is automatically calculated by SPSS if at least one of the cells of the contingency table has an expected value less than 5 when you select the chi-square test. Alternatively you may press the *Exact* button and then check the bullet point preceding *Exact*. You can find the 2-sided exact significance level in the table labelled *Chi-Square Tests* in the output window. In this case, Fisher’s exact test shows an exact significance level of 0.012, meaning we can be 98.8 percent sure that the passive prefers a full infinitive following *HELP*.

18.8 Unit summary and suggestions for further study

This case study used four *prêt-a-porter* corpora (LOB, Brown, FLOB, and Frown) to explore the potential factors that may influence a language user's choice of a full or bare infinitive after *HELP*. Our findings are summarized as follows:

AmE shows a stronger preference for bare infinitives after *HELP* than BrE. Language change over the three decades from 1961 to 1991 has produced a bias in favour of bare infinitives after *HELP* in both AmE and BrE. An intervening NP may increase the proportion of bare infinitives after *HELP* in AmE whereas in BrE, the effect of an intervening NP is unpredictable. The infinitive marker *to* preceding *help* encourages a language user to choose a bare infinitive. The passive construction exclusively selects bare infinitives.

In this case study, we gave a step-by-step demonstration of how to explore a corpus using MonoConc Pro, using its advanced features such as searching a corpus using regular expressions, sorting concordance lines, and the batch search function. We also showed how to interpret frequency data with the SPSS statistics package.

In unit 18.2.3, we only considered a single noun or pronoun as an intervening NP.

Now use the batch search function of MonoConc Pro to include the cases of an article or determiner followed by a noun, e.g., *a/the boy* and *those students*. Do the above findings still hold?

