

## CAPÍTULO 2

# TAKING A LOOK AT THE SUPPORT VERB CONSTRUCTION *V A LOOK*: A DEMONSTRATION OF METHODOLOGY

Bert Cappelle<sup>1</sup> (Université de Lille)

Pâmela Fagundes Travassos (Federal University of Rio de Janeiro)

### 2.1 INTRODUCTION: HOW WE EXPRESS VISUAL PERCEPTION

When we talk about acts of looking, we frequently adopt expressions and constructions that we also use to deal with a rather different sort of event, namely that of people or objects moving from one place to another (Cappelle 2020; Cifuentes-Férez 2014; Gruber 1967; Kawachi 2020; Matsumoto 2001, 2020; Matsumoto et al. 2021; Talmy 1996, 2000, 2018). In English, for example, someone can *gaze around the room* or *look out (of) the window*, just like they can *pace around the room* or *jump out (of) the window*. For both kinds of events – merely looking or actually moving – we can combine a verb – either a perception verb or a motion verb – and a prepositional phrase (PP) expressing the trajectory of the immaterial gaze or of the material person. Similarly, you can *throw a glance at your phone* just like you could, say, *throw a lance at a drone*, combining in both cases a verb of ballistic motion, a noun phrase, and a PP indicating the ‘target’. Vision thus finds parallels in self-agentive and caused motion.

---

1 University of Lille, Research Unit UMR 8163 STL, Univ. de Lille and CNRS, France.

You can also *get a glimpse* or *catch sight of something*, just like you can *get* or *catch* a ball that is thrown at you. In this alternative way of conceptualizing visual perception, the direction of motion is reversed, from the perceived object to the perceiver. When an object moves towards a person, it can change possession, as when a pop singer throws her jacket into the crowd and the person catching it can consider themselves the new owner. As can be expected, therefore, visual perception is often also expressed by means of expressions and constructions that involve notions of possession. Thus, apart from with the verb *get*, we also find visual perception encoded with *have* and *take* or by means of the ditransitive (double-object) construction, as in *give someone an angry look*, which, admittedly, is not just a matter of perceiving but also of making a facial expression. As is well known, the ditransitive construction is closely related to the so-called prepositional dative construction (as in *give something to someone*), which itself is structurally similar to the already-mentioned caused-motion construction, whereby the subject entity moves something to a certain goal or in a particular direction, as in *toss a startled glance towards the door*. In short, visual perception is frequently encoded in language in terms of the semantically interlinked concepts of motion, transfer, and possession.

At the same time, it should be clear that we can't take these expressions too literally. Consider the following authentic utterance, from the Corpus of Contemporary American English (Davies 2008-):

(1) Rita takes a forlorn glance towards the press area, where all of the other cameras are set up. (COCA)

This sentence mixes two different conceptualizations, that of the perceived object coming into the perceiver's possession (*take*) and that of the perceiver's gaze moving in the direction of (*towards*) the perceived object. Strictly speaking, these conceptualizations are incompatible. That (1) is not felt to be incomprehensible nonsense can only be seen as evidence that there is a great deal of semantic bleaching in the expressions used for visual perception: as language users, we treat the expressions, at least some of them, as somewhat dead metaphors. Moreover, while *giving* and *taking* are opposites when it comes to actual transfer, we find that different languages can use these opposites as default options for expressing visual perception. Consider English and Portuguese: English *take a look* (*at someone or something*) corresponds with (mainly Brazilian) Portuguese *dar uma olhada* (*em algo ou alguém*), literally 'give a look (at someone or something)'. Finally, in English itself, *take a look* and *have a look* practically mean the same, while outside of these (and other support verb) expressions, *take* generally refers to the gaining of possession while *have* is the result of this event.



The first aim of this chapter is to get a grasp of the *have/take a look* support verb construction in English. A support verb construction is one where a general-purpose verb (a 'light verb'), which is mainly restricted to carrying tense and aspect information, combines with a noun phrase whose head is semantically richer than the verb; together, the verb and the noun phrase express a certain action (see, e.g., Danlos 1992; Langer 2004; Namer 1998). Further examples in English are *have a drink*, *have a chat*,





*have a fight, take a bath, take a rest, give a scream, make a comment and do the washing-up.* Our second and no less important aim is to demonstrate how we can map a relatively small area of lexico-grammar like this one by using a widely available corpus tool: the corpus interface created by Mark Davies.<sup>2</sup> We thereby hope to give students and linguists new to corpus linguistics a glimpse of how much can be achieved with relatively little import of tool-external processing of the data – although we will also give a practical introduction to some of the many things the statistical tool R has to offer. Finally, we will attempt to integrate our empirical findings with some theoretical reflection about how lexico-grammatical knowledge is organized.



## 2.2 TAKE A LOOK AND HAVE A LOOK IN SOME DICTIONARIES OF ENGLISH





In the introduction, we mentioned in passing that *take a look* and *have a look* have practically identical meanings. Indeed, if we consult some of the major dictionaries of English, the impression that is raised is that *have* and *take* can be used interchangeably with *a look*. Consider for instance in Figure 1 some of the information (clipped here) in the *Longman Dictionary of Contemporary English* (LDOEC Online), s.v. *look*:

Figure 1: Fragment of LDOEC Online, s.v. *look*

**look<sup>2</sup>** ... **S1** **W1** noun  

**1 LOOK AT** [countable usually singular] an act of looking at something  
**have/take a look (at somebody/something)**  
 Let me have a look at that – I think it's mine.  
 Take a **good look** at the photo and see if you recognize anyone in it.  
 I took **one look** at the coat and decided it wasn't worth £50.  
**have/take a look around (also have/take a look round British English)** (=look at all the things in a particular place)  
 I have a special interest in old houses. Do you mind if I take a look around?

**2 EXPRESSION** [countable] an expression that you make with your eyes or face  
**give somebody a look**  
 Mike gave him such a severe look he didn't dare argue.  
 Why has Jake been giving me **dirty looks** (=unfriendly looks) all morning?

**3 CONSIDER** [singular] an act of examining something and thinking about it  
**have/take a look (at somebody/something)**  
 Have you had a chance to take a look at my proposal yet?  
 I asked the vet to have a **quick look** at the puppies as well.  
 It's time to take a **fresh look** at the old problem of low pay.  
 This month, take a **long hard look** (=examine very carefully) at where your money is going.

2 Another popular tool is SketchEngine, which allows you to search some very large corpora (<https://www.sketchengine.eu/>).

LDOEC Online provides several different meanings for the noun *look*. The first and the third meaning are expressed with either *have* or *take*; the second (which we also mentioned in the introduction) is combined with *give* and can combine with both the singular and the plural of *look*. The first meaning also manifests itself in the specific expression *have/take a look (a)round* (with *round* the preferred variant in British English), where again both support verbs are presented as alternatives. In the example sentences, some words are put in boldface, presumably because these are common combinations, known as ‘collocations’: *good look*, *took one look*, *dirty looks*, *quick look*, *fresh look* and *long hard look*. As for the latter, LDOEC Online actually also presents *take a (long) hard look at something/somebody* as a separate entry. This suggests that when *look* is modified by *hard* or *long hard*, the support verb will likely be *take*, not *have*. Something similar may be the case when the numeral *one* is used as a determiner: *take one look* could be more common than *have one look*.

As can be seen in Figure 2, somewhat similar information is provided by the *Oxford English Dictionary* (OED Online).

Figure 2: Fragment of OED Online, s.v. *look*, n.

**d. colloquial. to have (also take) a look:** to look at a person or thing, esp. (in later use) for the purpose of examination; to inspect or scrutinize someone or something. Also in extended use: to investigate or consider an idea, prospect, or other immaterial thing. Cf. *LOOK* *v.* 1a, *LOOK* *v.* 4a.

**(a)** With the object of one's gaze or consideration implied.

- 1673 J. DRYDEN *Assignment* III. i. 25 I must have a look before we go.  
 1777 J. WHITFIELD *Conjectures on Tyndaris of Horace* 21 Stept to Lloyd's, to take a look.  
 1830 J. P. COBBETT *Jrnl. Tour in Italy* 195 A young Roman lawyer...who is anxious, like myself, to have a look from the top of Mount Vesuvius.  
 1894 T. W. HYDE *Following Greek Cross* xxv. 153 About this time, Generals Meade and Warren came up on the rocks to take a look.  
 1920 P. GIBBS *Realities of War* IV. vii. 189 Come up and have a look, Jack.  
 1980 B. JACOBY *Girl in Mirror* II. 65 Doctor. How is she? Mother. No change. Doctor. Let's have a look, shall we?  
 2006 I. RANKIN *Naming of Dead* i. 13 I saw the sign yesterday... The one leading off the main drag. Decided I'd take a look.

(Hide quotations)

**(b)** With *at*. With the object of one's gaze or consideration specified. Cf. *Phrases* 2b(b).

- 1725 D. DEFOE *Tour Great Brit.* II. iii. 22 He would not go away without taking a look at the fine Gardens.  
 1769 G. S. CAREY *Shakespeare's Jubilee* 16 Let's have a look at you before you go.  
 1831 *New-Eng. Mag.* Dec. 494 A great Irishman was brought up into the operating theatre, the other morning, for the class to have a look at.  
 1885 M. COLLINS *Prettiest Woman in Warsaw* I. ii. 29 In the meantime I shall have a look at Warsaw.  
 1925 A. LOOS *Gentlemen prefer Blondes* iv. 109 So then Mons. Broussard stopped crying and put on his glasses and took a good look at us.  
 1960 J. IGGULDEN *Storms of Summer* 169 I'll get the quack at the Bush Hospital to have a look at it in the morning.  
 2012 *Action on Hearing Loss* Aug. 24 The government will soon replace Disability Living Allowance (DLA) with Personal Independence Payment (PIP) as part of its radical welfare reforms—we take a look at how this may affect you.

(Hide quotations)

**(c)** With adverbs or prepositional phrases, forming phrases corresponding to phrasal verbs at *LOOK* *v.* Esp. in **to have (also take) a look around (or round)**.

- 1829 *Blackwood's Edinb. Mag.* July 77/2 You'd bether just take a bit of a look round on the worlt, an' bid it good by.  
 1898 H. G. WELLS *Jimmy Goggles* *God in Cosmopolitan* Nov. 109/1 I stopped once, and...had a look up.  
 1913 *Bull. Geogr. Soc. Philadelphia* 11 11 On approaching a strong rapid we usually took a look ahead to ascertain where the best line of advance lay.  
 1925 L. H. MYERS *Clio* (1945) vii. 49 I had a look through the window.  
 1932 J. BUCHAN *Gap in Curtain* ii. 97 He hoped, while in the country, to have a look round.  
 a1965 W. G. PURCELL *St. Croix Trail Country* (1967) 97 I stepped over to have a look down from the edge of the hill.  
 1982 P. REDMOND *Brookside* (Mersey TV shooting script) (O.E.D. Archive) Episode 4. 70 Just take a look around and see how much is going on the side.  
 2011 A. LEWIS *Everybody Jam* xiv. 99 I was having a quick look around.  
 2011 *Independent* 4 Aug. 13/1 Punk's not dead—it's just having...a look round a stately home.

(Hide quotations)

We can observe a few differences here with what the LDOEC presents. First, the OED makes a distinction between the use of *have/take a look* on its own (where the

target of the gaze remains implicit) and *have/take a look* combined with an *at*-PP. Second, the OED does not consider the third meaning listed by LDOEC, that of “examining something and thinking about it”, as a separate meaning but does mention this as a specific purpose of looking that arose as a later and extended use. Third, the expression *have/take a look (a)round* is given special prominence. The adverb *(a)round* is treated as identical to the particle in phrasal verbs with *look* as a verb (*to look (a)round*) and as essentially a very common choice alongside other adverbs or PPs that can occur (e.g. *take a look ahead*, *have a look through the window*, *take a look round a stately home*). Fourth, the expression *give somebody a look* is not mentioned separately. However, elsewhere in the entry of *look* as a noun (though this is not shown in Figure 2), the OED lists the meaning “[a]n expression of a specified thought or feeling by looking”. The verb *give* is not seen here as one that is privileged to express that meaning: the quotations contain also the verb *cast* or *flash*, as well as other grammatical environments (e.g. *He was regarding her with a tender, worried look*, an example from 1956). In its presentation of this use, the OED also mentions two collocations: *come-hither look* and, as in LDOEC Online, *dirty look*. Fifth, the OED presents *have* as the basic option: by mentioning “*also take*” between brackets, it suggests that *have* is more common.

Finally, let’s consider one more dictionary, *Collins English Dictionary* (Collins Online). In Figure 3, we show a part of the dictionary entry of *look*.

Figure 3: Fragment of Collins online, s.v. *look*

### 1. VERB

If you **look** in a particular direction, you direct your eyes in that direction, especially so that you can see what is there or see what something is like.

*I looked down the hallway to room number nine.* [VERB preposition/adverb]

*She turned to look at him.* [VERB preposition/adverb]

*He looked away, apparently enraged.* [VERB preposition/adverb]

*If you look, you'll see what was a lake.* [VERB]

**Synonyms:** see, view, consider, watch More Synonyms of **look**

**Look** is also a noun.

*Lucille took a last look in the mirror.*

*Assisi has a couple of churches that are worth a look if you have time.*

### 2. VERB

If you **look at** a book, newspaper, or magazine, you read it fairly quickly or read part of it.

*You've just got to look at the last bit of Act Three.* [VERB + at]

**Look** is also a noun.

*A quick look at Monday's British newspapers shows that there's plenty of interest in foreign news.* [+ at]

### 3. VERB

If someone, especially an expert, **looks** at something, they examine it, and then deal with it or say how it should be dealt with.

*Can you look at my back? I think something's wrong.* [VERB + at]

[Also VERB]

**Look** is also a noun.

*The car has not been running very well and a mechanic had to come over to have a look at it.* [+ at]

### 4. VERB

If you **look at** someone in a particular way, you look at them with your expression showing what you are feeling or thinking.

*She looked at him earnestly. 'You don't mind?'* [VERB + at]

**Look** is also a noun.

*He gave her a blank look, as if he had no idea who she was.*

*Sally spun round, a feigned look of surprise on her face.* [+ of]

Noticeable differences are that there are no separate senses mentioned for *look* as a noun. Instead, the meanings we have already encountered are mentioned under the verb uses, where we just read: “*Look* is also a noun”. This might appear as though the noun use is not given proper credit, but one could argue that this choice underscores the linguistic insight that *look* is really part then of a support verb construction, with the noun occurring with the same range of meanings that the related verb can express. As in LDOEC Online, the ‘examine’ meaning is treated separately (here in sense 3). In addition, there’s a special sense in which *look*, both as verb and as noun in the support verb construction, conveys the idea of viewing something quickly or partially (sense 2). The sense of looking in a particular way to reveal your thoughts or feelings is again mentioned separately (sense 4), and just like in OED Online, the verb *give* is not treated as one that is needed to trigger that interpretation, although we can see that the first example sentence does contain that verb.

We could look at more dictionaries and we would find further differences. For now, our perusal of the three dictionaries consulted allows us to distill a number of questions that should be possible to answer by means of corpus research:

Is *have a look* more frequent than *take a look*? Could there be regional preferences (e.g. British English versus American English)?

Are there specific collocations where one support verb is more common than another (e.g. *take a good look at something*)?

Related to the preceding: are there specific meanings or functions for which one support verb is more common than another?

For the meaning relating to a way of looking that is expressive of something (e.g. *a mournful look*), is the verb *give* the most typical environment?

The message we want to give here is that by just looking at information in dictionaries (the meanings as well as the examples), we can come up with a number of




empirical questions that we otherwise might not have thought of. In the next section, we will demonstrate how such questions can be tackled with the help of corpora.

## 2.3 GETTING A GRIP OF REAL DATA FOR *LOOK* WITH ENGLISH-CORPORA.ORG


### 2.3.1 GLOWBE: REGIONAL VARIATION AT A GLANCE











Can we find out whether different varieties of English – American English, British English, Australian English, etc. – have different preferences for *have a look* versus *take a look*? The answer is yes. One laborious way would be to consult several variety-specific corpora, such as the Corpus of Contemporary American English and the British National Corpus, or different corpora in the International Corpus of English (ICE) family, one after the other, and then compare the results obtained from each corpus. Fortunately, there is a much faster way, thanks to a corpus that provides data about these and many more regional varieties in one place: the Global Web-based English corpus, or GloWbE. This is housed at [www.english-corpora.org](http://www.english-corpora.org). Go to that website and you will get an overview of English corpora, as displayed in Figure 4. (For ‘monitor corpora’, which keep growing with new material up to the present date, the numbers you will find for the corpus size will be larger than in this figure.)

Figure 4: Overview of corpora housed at English-corpora.org



The most widely used online corpora [guided tour](#) overview, search types, variation, virtual corpora (quick overview), BYU<sup>†</sup>.

The links below are for the online interface. But you can also  download the corpora for use on your own computer.

Corpus (online access)	Download	# words	Dialect	Time period	Genre(s)
IWeb: The Intelligent Web-based Corpus		14 billion	6 countries	2017	Web
News on the Web (NOW)		12.9 billion+	20 countries	2010-yesterday	Web: News
Global Web-Based English (GloWbE)		1.9 billion	20 countries	2012-13	Web (incl blogs)
Wikipedia Corpus		1.9 billion	(Various)	2014	Wikipedia
Corpus of Contemporary American English (COCA)		1.0 billion	American	1990-2019	Balanced
Coronavirus Corpus		1074 million+	20 countries	Jan 2020-yesterday	Web: News
Corpus of Historical American English (COHA)		475 million	American	1820-2019	Balanced
The TV Corpus		325 million	6 countries	1950-2018	TV shows
The Movie Corpus		200 million	6 countries	1930-2018	Movies
Corpus of American Soap Operas		100 million	American	2001-2012	TV shows
Hansard Corpus		1.6 billion	British	1803-2005	Parliament
Early English Books Online		755 million	British	1470s-1690s	(Various)
Corpus of US Supreme Court Opinions		130 million	American	1790s-present	Legal opinions
TIME Magazine Corpus		100 million	American	1923-2006	Magazine
British National Corpus (BNC) *		100 million	British	1980s-1993	Balanced
Strathy Corpus (Canada)		50 million	Canadian	1970s-2000s	Balanced
CORE Corpus		50 million	6 countries	2014	Web
From Google Books n-grams (compare)					
American English		155 billion	American	1500s-2000s	(Various)
British English		34 billion	British	1500s-2000	(Various)

Most of these corpora have been developed and provided a uniform interface for by Mark Davies, a now-retired professor of linguistics at Illinois State University and later Brigham Young University. A few of the corpora, like the British National Corpus, pre-existed but are housed here so that they can be searched within the same interface. The website states that more than 130,000 people, in over 140 different countries, use these corpora on a monthly basis. If you are a first-time user, it is highly recommended that you take the guided tour (click on the link on top to open a pdf document). For now, we hope we can give you an idea of what the corpora and the online interface have to offer by showing you some concrete steps to be taken. Mark Davies also developed corpora for Spanish and for Portuguese (Davies 2016-a, b), for which he designed a similar interface (for documentation, see Davies 2008).



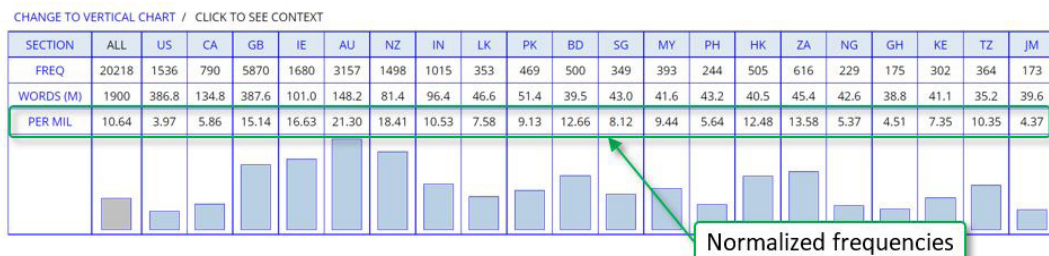
To access GloWbE, click on the name of the corpus in the list. In the search window, enter “HAVE a look” as a command (without any quotation marks), as shown in Figure 5.

**Figure 5:** Entering *have a look* as a string of lexical items, with *have* treated as a lemma

Capitals are used for the verb *have* to search for all inflected/contracted forms at once (*have, has, had, having, 've*). In other words, capitals force a lemma search, not a word form search. An alternative way of doing a lemma search is by enclosing the lexical item in square brackets: “[have] a look”. If you were to click “Find matching strings”, you would get to see the search results as a breakdown of the different forms of *have* combined with *a look*, with absolute frequencies in twenty different ‘World Englishes’. (You would then see that the string *'ve a look* has just one occurrence in British English and another single one in Honk Kong English.) As we’re mainly interested in discovering the relative frequencies of the expression *have a look*, whatever its inflected form, in different varieties of English, let us instead select the Chart function at the top (see Figure 6).

**Figure 6:** Retrieving a bar chart

If we now click “See frequency by section”, we obtain a bar chart of the ‘collapsed’ frequencies (i.e., with all the forms of *have* combined, as we did a lemma search) per variety of English (see Figure 7).

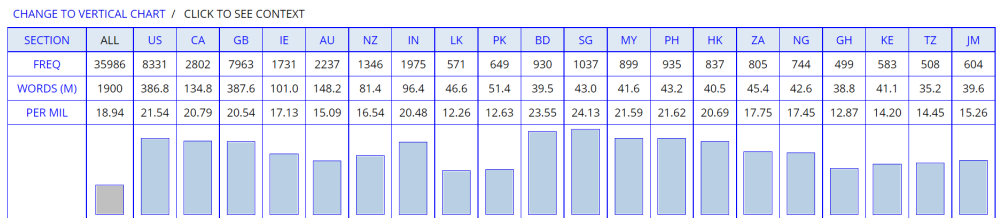
**Figure 7:** Bar chart of frequencies of *have a look* in twenty varieties of English worldwide

We can change the graph, if we so desire, to a so-called ‘vertical chart’, where the varieties are presented one under the other and the bars are then, in fact, horizontal. Looking at these results, we see that *have a look* is most frequent in Australian English (AU) and New Zealand English (NZ), as well as in Irish English (IE) and British English (GB). The string is also common in South African English (ZA), Bangladeshi English (BD) and Hong Kong English (HK). It will not have escaped the reader that these are all varieties spoken in countries with strong historical, cultural and economic ties, given Britain’s colonial past: countries that were part of the British Empire, which later developed into the Commonwealth of Nations, typically use the ‘British’ variant, in cases where American English and British English differ. In this case, indeed, it’s clear that we have relatively lower bars for American English (US) and Canadian English (CA), as well as for other varieties that are under ‘American’ influence, such as Philippine English (PH) and Jamaican English (JM).

There are three rows of numbers above the bars. The top row shows the absolute frequencies of the string in each regional variety. Thus, we can see that there are 5,870 tokens of this string in the British English component of GloWbE. The row underneath tells us how many words each component contains in total. For British English, this is more than 380 million words. As not all components are equally large – Tanzanian English ‘only’ has 35.2 million words – we need to know how often the string occurs as a proportion of the total size of the corpus. That is why the lowest row is most revealing: it shows the ‘normalized’ frequency of the string per million words in each corpus. It is these numbers that are represented by the bars. The first column, with the bar in grey, gives absolute and normalized frequencies of *have a look* in the entire corpus. The cell in the middle row also shows, as we could already see in Figure 1, that there are ‘1900 million’ words in the entire GloWbE, or indeed 1.9 billion words. Let’s remember that *have a look* has a frequency of 10.64 per million words in English worldwide.

We can now repeat the same procedure for *take a look*. To do this, first click on the “search” option which you’ll find above the results in the current window. This will bring you back to the screen with the search window we saw in Figures 5 and 6. Enter “TAKE a look”. The bar chart obtained looks different (see Figure 8).

Figure 8: Bar chart of frequencies of *take a look* in twenty varieties of English worldwide



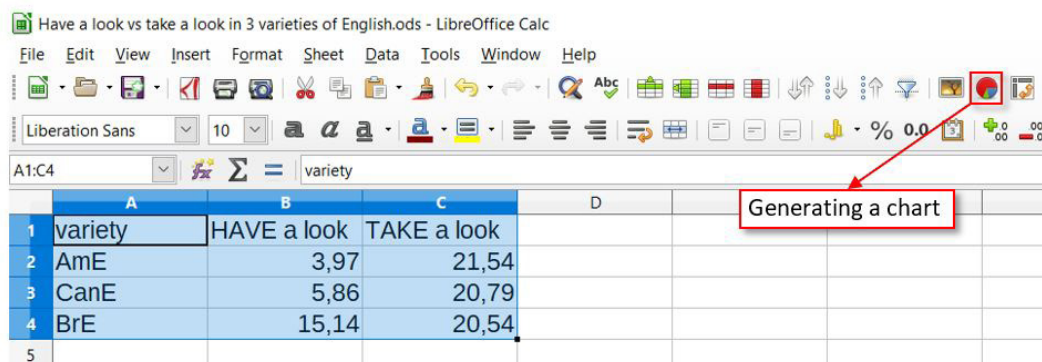
We can observe that the differences in relative frequencies across regional varieties are less pronounced for this expression. Crucially, the variant with *take* is overall more frequent than the variant with *have*: the string *take a look* occurs 18.94 times per million words, which approaches twice the frequency for the string *have a look*. This means that the OED has it backwards: it should more appropriately have written “*to take (also have) a look*”. In particular, in American English and Canadian English, where we saw that *have a look* is fairly infrequent, the frequencies are much higher for *take a look*. Even in Bangladeshi English, which we initially assumed followed British English, there is a clear preference for *take a look*.

If we want to present these results as part of a research report, we cannot just copy-paste screen grabs as we have done here. Instead, we should transfer the numbers of interest to a table in a word processor application. Numbers can also be entered in a spreadsheet application, such as Microsoft Excel or a free alternative such as LibreOffice Calc (see Figure 9).

Figure 9: Some selected corpus frequencies hand-entered in LibreOffice Calc

	A	B	C
1	variety	HAVE a look	TAKE a look
2	AmE	3,97	21,54
3	CanE	5,86	20,79
4	BrE	15,14	20,54

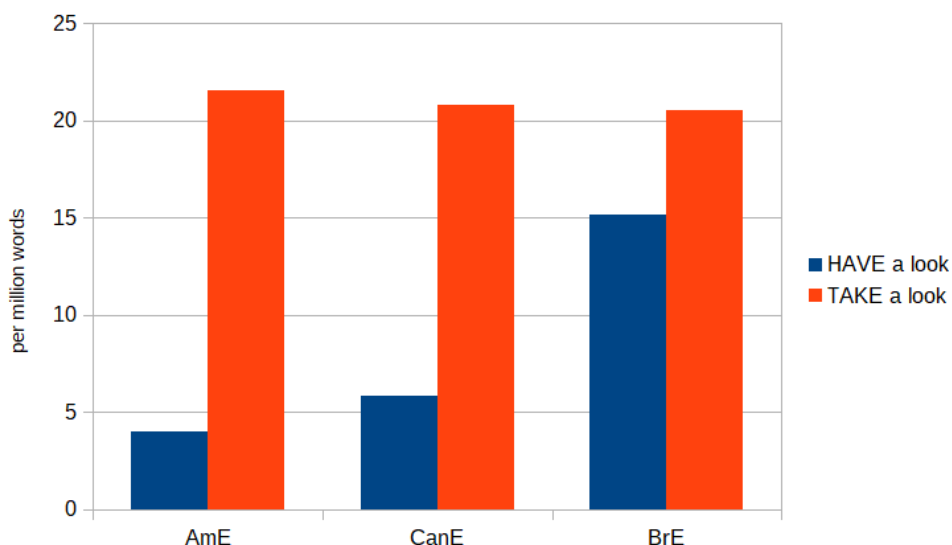
Selecting all the cells and then clicking the chart button (see Figure 10) will generate a graph.

**Figure 10:** Generating a bar chart and providing extra labels in LibreOffice Calc

If you are not happy with the standard color scheme in LibreOffice, you can change this by double-clicking anywhere on the chart, then right-clicking on the first series and selecting “Format Data Series...”. In the Area tab, a range of more palatable colors can be selected, including ten shades of gray. Each pair of bars can be changed this way. For our example, though, we have stuck with the standard colors.

The resulting table can then be selected and exported to your word processor application, as we have done here (see Figure 11). It might be preferable not to add any title within LibreOffice and instead add one in your word processor application, as this allows greater flexibility in terms of font size and the like. Moreover, it is not even necessary to use a separate spreadsheet application, as common word processor applications allow you to insert a chart directly into your document. Clicking the chart icon (which, depending on the application you use, looks like a pie chart icon or a bar chart icon) will then also allow you to make or adapt a data table, much like the one in Figure 9 above.

**Figure 11:** A simple bar chart generated via LibreOffice Calc



Now that we have this figure, two observations can be made very clearly. First of all, *take a look* is more frequent than *have a look* in all three varieties shown here. Secondly, this difference in frequency is much more pronounced in AmE and CanE than in BrE.

### 2.3.2 LOOKING UP COLLOCATIONAL PREFERENCES: WHICH ADJECTIVES ARE ATTRACTED TO WHICH VARIANT?

It will be remembered that *look* can be modified by one or more adjectives (*dirty, fresh, good, quick, long hard*, etc.). It would not come as a big surprise if the choice of *have* versus *make* was dependent on the occurrence of a particular adjective. How can we find out with the help of a corpus? This is possible with the “Compare” function in any of the corpora at English-corpora.org. Let’s use, for a change, the British National Corpus (BNC) (Davies 2004). In Figure 12, we show which information to enter.

**Figure 12:** Comparing *have a(n) ... look* and *take a(n) ... look* with respect to the adjectives that occur in them

List Chart Collocates **Compare** KWIC

HAVE a/an ADJ look Word1 [POS] ?

TAKE a/an ADJ look Word2 [POS]

ADJ Collocates [POS]

+ 4 3 2 1 0 0 1 2 3 4 +

Compare words Reset

☐ Sections Texts/Virtual Sort/Limit Options

Mastering the Collocates and the Compare functions in English-corpora.org is an acquired skill and takes quite a bit of trial and error. Here, we use capital letters for the verbs, as before, but not for the following article, even though we would like to retrieve both *a* and *an*. However, these two words are not treated in the corpus as forms of a single lemma. We solve this by using the vertical slash symbol; this way, the interface understands that we look for “*a* or *an*”. Next up is a word that should be an adjective, so we type ADJ. If you don’t know the part-of-speech abbreviation for a given word class, click on “[POS]?” next to the search and then again click “POS” or “Insert PoS” to obtain a menu of options. We end our string with the word *look*.

Underneath the windows for the two ‘words’ – in our case, these are word combinations – we have to specify which collocates we want to retrieve. Collocates are words that often occur nearby another word or phrase. In the Collocates search window, we again type ADJ, as that is the kind of word we want to look for inside the string. In fact, though, since we already used ADJ in the string itself, we do not strictly have to fill in this window. If we leave it empty, “\*” will appear automatically before we click “Compare words”. The asterisk is the ‘wildcard’ that stands for any lexical item of any word class. The bar with numbers allows us to specify how far to the left and/or to the right we want our collocates to be searched. Here, we don’t want it to be searched many words before or after the search strings. In fact, we want it to be searched *within* the search strings! Somewhat unpredictably – hence our insistence on trying things out a few times – it is the word *look* in each of the two search strings that is treated as the main context words. That is why we select the first spot to the left of our “Word 1” and “Word 2” as our position of interest. Note that for other types of searches, we can select not just a single position but a whole range, even up to nine words to the left and/or to the right of the central word. (For a span of more than four words, click once or twice on the plus sign.)

The results we obtain, after clicking “Compare words”, are displayed in Figure 13. They are quite revealing.



**Figure 13:** Two output tables of adjectives that are (or aren't) good collocates of *have a(n) ... look* compared to *take a(n) ... look*, and vice versa

SEE CONTEXT: CLICK ON NUMBERS (WORD 1 OR 2)  
 SORTED BY RATIO: CHANGE TO FREQUENCY  
 WORD 1 (W1): HAVE A[AN] \* LOOK (0.91) [HELP...]

	WORD	W1	W2	W1/W2	SCORE
1	LITTLE	25	0	50.0	55.1
2	WEE	12	0	24.0	26.5
3	PROPER	7	1	7.0	7.7
4	BETTER	9	3	3.0	3.3
5	QUICK	55	23	2.4	2.6
6	GOOD	69	41	1.7	1.9
7	CLOSE	11	25	0.4	0.5
8	CLOSER	21	65	0.3	0.4

WORD 2 (W2): TAKE A[AN] ADJ LOOK (1.10)

	WORD	W2	W1	W2/W1	SCORE
1	FRESH	26	2	13.0	11.8
2	CRITICAL	12	1	12.0	10.9
3	LONG	6	0	12.0	10.9
4	CAREFUL	11	1	11.0	10.0
5	HARD	11	1	11.0	10.0
6	HONEST	5	0	10.0	9.1
7	DETAILED	5	0	10.0	9.1
8	CLOSER	65	21	3.1	2.8
9	BRIEF	11	4	2.8	2.5
10	CLOSE	25	11	2.3	2.1
11	GOOD	41	69	0.6	0.5
12	QUICK	23	55	0.4	0.4

The output of our search contains two tables. The table on the left contains a rank-ordered list of words that go together very well (towards the top, in green) or not very well (towards the bottom, in a reddish hue) with *have a(n) ADJ look*, compared to the competing environment *take a(n) ADJ look*. Remember that we set up the search in such a way that these words are adjectives appearing in the slot before the noun *look*. By the way, at the top of the table on the left, “ADJ” in Word 1 is shown as “J\*”, which is just a quirk of the underlying corpus machinery – and another reason why, for research reporting purposes, it is best to produce your own table. The table on the right then shows how strongly adjectives are preferred by *take a(n) ADJ look* over the alternative with *have*.

Let’s take a more detailed (!) look at how we ought to read these tables. The number between brackets after our ‘word 1’ (W1) and ‘word 2’ (W2) – which, again, in our case are really more like lexical environment than single words – are the overall ratios of these two words. Here, there are 0.91 occurrences of *have a(n) ADJ look* in the British National Corpus for every occurrence of *take a(n) ADJ look*; conversely, there are 1.1 occurrences of the latter phrase for every occurrence of the former. Put simply, the variant with *have* is a little less frequent in British English than the one with *take*, confirming what we saw in Section 3.1 with data from GloWbE. The columns headed by W1 and W2 show the corpus frequencies for the words involved. W1/W2 is the ratio of the first environment to the second for each adjective in it. Thus, *quick* occurs 2.4 times as frequently with the *have* variant as with the *take* variant. As Mark Davies explains in the notes when you click on “[HELP...]”, when a competing variant occurs zero times, division by 0 is avoided by giving that variant a frequency of 0.5 instead of 0. That is how we get 50.0 as the W1/W2 ratio for *little*, which occurs 25 with *have* and 0 times, artificially set to 0.5 times, with *take*. The score in the right-most column of each table, finally, takes the overall relative frequency of the environment into account to either boost or attenuate the W1/W2 ratio or W2/W1 ratio. Thus, given that the *have* variant is less frequent than the *take* variant (see again its 0.91 overall ratio), if an adjective nonetheless prefers *have*, this is all the more surprising. In our case, the difference in frequency of the two competing environments is quite small, but as you

can imagine, for other cases of variation the score for a collocate might be wildly different from just its W1/W2 ratio. The score is calculated by dividing the W1/W2 ratio by the overall ratio of the host environment.

What the corpus allows us to do here is similar in spirit to a ‘distinctive collexeme analysis’, a method popular among corpus linguists to measure how strongly words are attracted to, or repulsed by, one particular construction compared to another one with a roughly equivalent meaning (Gries and Stefanowitsch 2004). However, distinctive collexeme analysis doesn’t just look for each collocate how often it occurs in one environment rather than another, competing environment, but also how often it occurs compared to all other collocates. Ultimately, it is a slightly more laborious kind of analysis than what Mark Davies’s corpus interface allows us to see.

So, what is it that we can see? The collocating adjectives reveal that the two support verb constructional alternatives are far from perfectly synonymous. Apparently, with *have* as a support verb, *look* is more often modified by ‘diminutive’ adjectives such as *little*, *wee* and *quick*, suggesting that *having a look at something* seems to be a perfunctory, fleeting action. With *take*, we find adjectives that suggest a more concentrated and considered, or a more specific way of looking: *fresh*, *critical*, *long*, *careful*, *hard*, *honest*, *detailed*, etc. Clearly, *taking a look at something* typically describes a cognitively more demanding action. It seems as though some of the active meaning of the lexical verb *take* – its dynamic and agentive semantics – is preserved in how we use it as a support verb. However, this could well be some wishful linguistic thinking. After all, there’s nothing very dynamic about *take a nap*. In any case, with *take*, the noun *look* is often synonymous to *perspective*, *view* or *approach*. This shows again that the meaning of *take a look* is not purely a matter of physically directing your eyes towards something.

The following two examples are quite representative examples of each variant of this support verb construction, in context:

(2) What are we doing? Well let’s let’s just **have a quick look** at this one. What’s this describing here? V equals large V or V nought or whatever you like, [unclear – cough] E to the minus R T over L. (BNC)

(3) The economic recession has caused European governments to **take a fresh look** at the extent to which a degree of interdependence in the past has been translated into a heavy dependence of the poorer areas on the richer north. (BNC)

Sentence (2) was taken from a tutorial lesson in the corpus. Looking at the context, the tutor *does* deal with contents for which one needs to have one’s wits together. It could be argued that the tutor tries to put the student at ease by merely *presenting* the action as something that doesn’t involve a lot of effort, as if to say: “This one will be easy for you to understand.”

## 2.3.3 'CODING' OR 'ANNOTATING' A SAMPLE OF CORPUS EXAMPLES BY HAND AND THEN PROCESSING YOUR FINDINGS WITH R

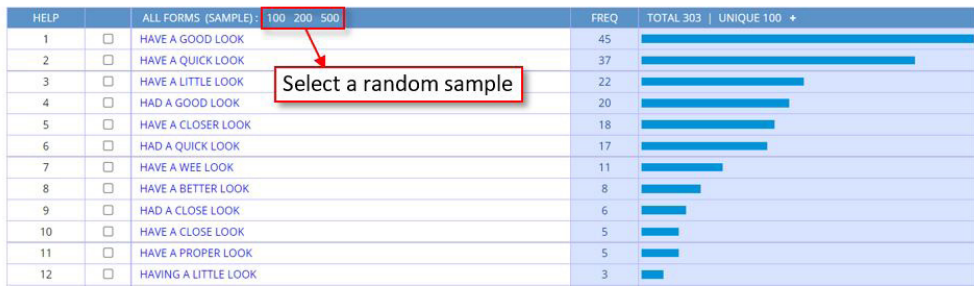
Grasping the full meaning of an expression goes beyond looking at the collocating lexical items. A fuller study should involve the following three steps:

1. getting one or more samples of sentences
2. analyzing the sentences with respect to one or several variables
3. processing the results with a tool for statistical analysis

### 2.3.3.1 Retrieving a sample of authentic sentences

So, first, you should retrieve a sufficiently large sample of authentic utterances from the corpus. To do this, use the List function, which is selected by default, and type in "HAVE a|an ADJ look" in the search window. In the results window, you can then select 100, 200 or 500 examples (see Figure 14).

**Figure 14:** Asking for a random sample of one of the competing forms



The result of this action is what's known as a set of 'concordance lines' (see Figure 15).

**Figure 15:** Part of a random sample of 100 retrieved concordance lines

CLICK FOR MORE CONTEXT			EXPLORE NEW FEATURES	SAVE	TRANSLATE	ANALYZE
1	KBC	S_conv	need it. (SP:P51AG) Cos I just, I just want to <b>have a quick look</b> through it before I gave it (SP:P51AD) I'll bring it over after, as			
2	KBF	S_conv	couldn't get over (pause) you know, I say, I <b>had a quick look</b> the other at Marks (pause) you know, to even do the Christmas serviettes (pause)			
3	KBW	S_conv	(SP:P5087) (reading) The magician was fast asleep. Tim leaned over to <b>have a good look</b> at him, but his foot slipped and he fell in. He landed right			
4	KCY	S_conv	I think there might be yeah (SP:P50H9) Yeah (SP:P50H9) I ain't <b>had a proper look</b> I just put it on and said, work a few routes out you know			
5	KDS	S_conv	, still (unclear) wet! (SP:P50K6) (unclear) (SP:P50JX) (laugh) (pause) (laughing) <b>Had a worried look</b> on her face then! (SP:P50K6) Oh darling! (SP:P50JY) Sorry (ur			
6	KD8	S_conv	really nice Orion, a J reg. Anyway when they (pause) <b>had a good look</b> at it the front half was an Escort (pause) and the back half was an			
7	KDL	S_conv	yes (unclear)2. (SP:P50NY) Some more over there, do you want to <b>have a quick look?</b> (unclear) (pause) (SP:P50NX) Trying to think what Penny would like, but			
8	KDV	S_conv	round to the toys. (SP:P50SB) Go on then. You can <b>have a quick look</b> at the toys. (SP:P50SC) With you. (SP:P50SB) Can't afford to buy anything			
9	KE0	S_conv	the one (unclear) your dinner? (SP:P50SX) Oh ay! He was <b>having a good look</b> then I'd say! (SP:P50SY) (unclear) (SP:P50SX) Didn't look at as though			
10	KE6	S_conv	it's my turn to buy the (pause) draw prizes. I <b>had a quick look</b> round yesterday but I ain't got a clue! (SP:P50XB) And what what (SP:P50XA)			
11	KPA	S_conv	? (SP:P553C) (----) (SP:P553H) (----) no (pause) (unclear) and if we <b>have a close look</b> hey there's nobody in (SP:P553H) (laugh) (pause) we appear to			
12	JP3	S_meeting	it up on the machine to change again quickly but I've <b>had a quick look</b> at the situation with regard to the five per cent increase. (SP:P54GX) Yeah.			

Inspecting the sentences is important. Line 5 in the sample shows that there can be ‘noise’ in the data. This is not an example of the construction we’re interested: *a worried look* here refers to someone’s appearance. When we analyze a manageable amount of data in more detail, we have to sift out such irrelevant examples. For very large amounts of data, like those we obtained from GloWbE in Section 3.1, eliminating noise may not be practically feasible and we may just have to live with a certain degree of imprecision. Yet, we could always additionally take a small sample – say, 100 examples – and find out how precise the search is, that is, what proportion of correctly retrieved instances we have for our corpus query. We can then report on this precision rate to give the reader an idea of how reliable the graphs are. This solution, too, would present further practical problems, as the precision rate can vary across regional varieties, which would mean we have to examine samples of all varieties of interest. If there is just too much noise in the data, this may require you to think of a different search query. Ultimately, coming up with a satisfactory one is a matter of striking the right balance between precision – you want to have as little noise as possible – and so-called ‘recall’ – you want to have as many examples that really correspond to the pattern, so as not to miss any major portion. As for this latter notion, recall, you will understand that the more specific our search string is, for instance by adding *at* in our example, the higher the risk becomes that some examples that fall under the construction you are interested in will not be retrieved (e.g. a relevant example like *have a look round the new school* would not be included).

### 2.3.3.2 Annotating the data

Our next step, after taking a sample, is to copy the concordance lines to a spreadsheet and analyze them one by one with respect one or more ‘variables’ (also called ‘parameters’ or ‘factors’). Deciding which variables you want to focus on will depend on your specific research question(s). It is not unheard of that corpus examples are ‘coded’ or ‘annotated’ for twenty or more different variables, each of which may have several different values (or ‘factor levels’). Variables could pertain to morphological, syntactic, semantic, pragmatic, discursive and/or (when these data are available) sociolinguistic aspects of utterances and/or their producers.

In our case, we might be interested in, for instance, the pragmatic value it carries in its context, that is, whether it is used as part of a statement (most concordance lines in the fragment), a request (line 1, in a way), an invitation (lines 7 and 8), and so on. When the same exercise is done for a similarly sized sample of occurrences with “TAKE a|an ADJ look”, we might find that there are different proportions of pragmatic functions for the two competing patterns. Indeed, that is the case in our example, as we’ll see. But first, here are some details on the values we distinguished for our single variable ‘pragmatic function’:

### 1. Assertives

This category was assigned to the largest group of sentences, most of which are ordinary statements, e.g. *I had a furtive look round*. Note that the label was used as a shorthand for the most ordinary occurrences of the two competing patterns, including for sentences realizing speech acts that aren't assertives in a technical sense. For instance, in *Will you be all right with him while I take a better look at the other bloke?*, the proposition in which our expression occurs is not claimed by the speaker to be true (nor is it claimed to be false), as it occurs in a temporal clause that itself appears in an interrogative sentence. The *while*-clause here probably expresses a kind of intention. Yet, in order not to proliferate functional categories, we used *assertive* as a catch-all term for any sentence that wasn't of either of the next two categories.

### 2. Invitations and requests

These two pragmatic functions were treated as one single category, as they are closely linked: just like a request involves an action that the speaker would like to perform, an invitation refers to an action that the speaker thinks the hearer would like to perform. Note that sentences of the type in (2), with *let's...*, were also considered a kind of invitation, even if the action might not be one that the addressee is eager to perform prior to the suggestion. No further distinction was made between *let's*-utterances used to suggest a joint action by speaker and hearer and those used by the speaker to announce an examining action, as when a doctor says, *Let's have a wee look at this lump*. The latter could be argued to be a kind of request (with the permission assumed to be immediately granted) presented as an invitation, which is one more reason for why we lumped together (no pun intended) these two kinds of speech act.

### 3. Instructions

These come close to invitations, but this category is reserved for unmistakable orders and obligations. Instructions can also be directed at a group including the speaker, as in *It is time we take an honest look at what we do*.

We present the results of our small-scale corpus study in Table 1, which is a so-called 'contingency table':

**Table 1:** Pragmatic functions of the *HAVE a(n) ADJ look* and *TAKE a(n) ADJ look* support verb constructions in the BNC

	<i>Assertives</i>	<i>Invitations and requests</i>	<i>Instructions</i>	<i>Total</i>
HAVE a(n) ADJ look	50	24	6	80
TAKE a(n) ADJ look	72	9	19	100
Total	122	33	25	180

As you can see from the column with totals, twenty utterances out of a sample of a hundred retrieved tokens of *HAVE a(n) Adj look* have disappeared from the table. These were removed because they were irrelevant. All of them used *look* in the sense of appearance (e.g. *She had a seraphic look on her face*), as in line (5) in the concordance

in Figure 15. This gives us an idea of the precision rate of our corpus query: 80%. So, one in five retrieved utterances was noise. The precision rate for *TAKE a(n) ADJ look* was 100%. It also looks as though the *have* variant is more often used for invitations and requests than the *take* variant, while the latter is more often used for instructions than the former.

Before we go on, a quick word about how to organize a table. In Table 1, the two variants of the construction are presented one under the other in the column to the left. That is because they are treated as possibilities making up the ‘independent variable’. For each of these options, we are interested in finding out which pragmatic functions it can have, and these therefore make up the ‘dependent variable’. In a contingency table, the values of the dependent variable are usually presented at the top of the table, from left to right. By contrast, note that in Figure 9, ‘English variety’ (AmE, CanE, Br) was treated as the independent variable and the dependent variable consisted then of our two forms of the construction, *HAVE a look* and *TAKE a look*.

### 2.3.3.3 Processing the data statistically

The third step, after retrieving a sample and annotating the sentences, is to process the observed differences statistically. Are the differences in frequencies significant or are they likely to be merely coincidental variations you can get in a relatively small sample? After all, when you flip a coin just ten times, you may throw heads seven or eight times, without this having to force you to conclude that your coin is rigged and that you have a higher chance of getting heads as you do for getting tails. So, how can we then test whether or not there is any real association between the *have* vs. *take* variants of this support verb construction and the pragmatic functions realized by them?

Many linguists use the free software environment R (R Core Team 2018) to deal with the kind of statistics question we’re facing here. R is a programming language, which means you have to type in lines of code yourself to tell the program what to do. This may be off-putting if you’re used to clicking on icons and selecting options from scroll-down menus in a fully designed interface. Fortunately, there is a vibrant community of people from all over the world offering help to one another. An excellent introduction to how to use R for linguistic research is the book by Levshina (2015). What follows is based on Chapter 9 of Levshina’s book, about “Measuring associations between two categorical variables”. Other useful textbooks on using R in quantitative linguistic approaches are those by Desagulier (2017) and Gries (2021). For beginners, there is also Jeroen Claes’s (2019) crash course for linguists ([http://www.jeroenclaes.be/r\\_crashcourse\\_for\\_linguists/](http://www.jeroenclaes.be/r_crashcourse_for_linguists/)).

First you need to install the program. We suggest you also install RStudio IDE (integrated development environment), which makes your work in R a lot more user-friendly. You can download a free licence here:

<https://www.rstudio.com/products/rstudio/download/>



Once you have installed RStudio on your local computer and opened the application, you may first need to install a number of add-on ‘packages’. At the command prompt (the ‘>’ sign where the cursor is blinking in the console pane), you need to type in the following:

```
> install.packages("vcd")
```

This will automatically install a package that will allow you to visualize categorical data as well as compute the effect size of a variable.<sup>3</sup> You then also need to load this package, as follows:

```
> library(vcd)
```

To produce something like Table 1 in R, without the marginal frequencies (i.e. the totals), use the function `rbind()`, which connects vectors (i.e., number sequences) as rows forming a table:

```
> look <- rbind(c(50, 24, 6), c(72, 9, 19))
```

To check whether this has the desired effect of producing a table, simply type the name of the contingency table (which we have just named ‘look’, but we could also have named it ‘table’ or ‘data’):

```
> look
```

This is what you then see:

```
[,1] [,2] [,3]
[1,] 50  24   6
[2,] 72   9  19
```

To add more insightful labels to the rows and columns, you can use commands like these:

```
> rownames(look) <- c("HAVE a(n) ADJ look", "TAKE a(n) ADJ look")
> colnames(look) <- c("assert", "invit/req", "instruct")
```

This produces the following result:

```
> look

      assert invit/req instruct
HAVE a(n) ADJ look    50     24     6
TAKE a(n) ADJ look   72      9    19
```

<sup>3</sup> Another package is Quanteda, which you can use to conduct textual analysis, such as comparing speeches by different politicians and thereby creating word clouds that illustrate which words are typical for each of them, among many other features (<https://quanteda.io/index.html>). A package to carry out collexeme analysis, which we mentioned in passing in Section 3.2, has been developed by Susanne Flach and can be downloaded from this site: <https://sfla.ch/collostructions/>. It can then be installed and loaded in R (see Flach 2021).

It's also possible to calculate the proportions of every cell in a column, such that their sum adds up to 1.

```
> prop.table(look, 2)
```

```
assert invit/req instruct
```

```
HAVE a(n) ADJ look 0.4098361 0.7272727 0.24
```

```
TAKE a(n) ADJ look 0.5901639 0.2727273 0.76
```

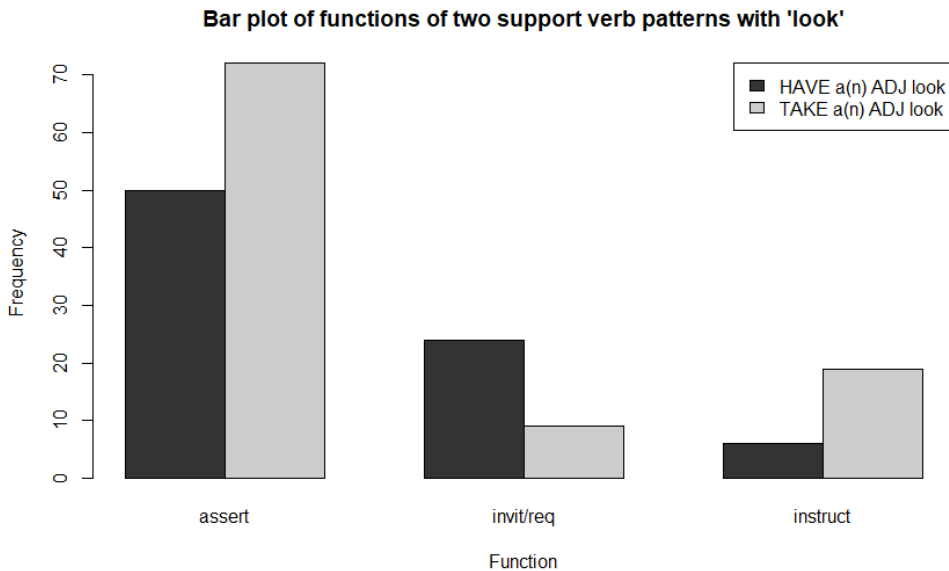
If you wanted to compute the proportions of cells making up a row, you would have to add “, 1” after the name of the table instead of adding “, 2”. (Not adding anything gives you the proportions of all cells in the table.)

You can generate a bar chart as follows:

```
> barplot(look, beside = TRUE, col = c("grey20", "grey80"), main =  
"Bar plot of functions of two support verb patterns with 'look'",  
xlab = "Function", ylab = "Frequency")  
> legend("topright", fill = c("grey20", "grey80"), c("HAVE a(n) ADJ look",  
"TAKE a(n) ADJ look"))
```

To produce a graph about something else, with another title, other labels and other variable names in the legend, simply replace what's rendered here in grey by descriptors that apply to your data. The code produces the chart shown in Figure 16:

**Figure 16:** Bar chart of pragmatic functions for two support verb alternatives, generated in R



So far, all of this is not anything that we couldn't have done more easily with the in-built chart function of a spreadsheet application like Microsoft Excel or LibreOffice Calc. Where R comes into its own is the possibility it offers to carry out statistical calculations. Presenting a table or a graph is just a matter of doing what is known as descriptive statistics. When we also want to test a hypothesis, we use so-called inferential statistics. This is the sort of statistics that allows us to assess whether we can generalize beyond our small sample the observation that, for instance, *have a look* appears to be more common for invitations and requests than *take a look*. Note from the proportions table above that the difference in proportions seems strongest for the value 'instructions'. But this category also represents the lowest share of the total sentences in the sample, so perhaps not that much can be inferred from these observations.

Let's start by comparing the observed frequencies in the table with the expected values, that is, the ones we would get if there were no relation whatsoever between choice of support verb and pragmatic use. We can type in the following two commands, each producing a table:

```
> look
```

```
      assert invit/req instruct
HAVE a(n) ADJ look   50    24    6
TAKE a(n) ADJ look   72     9   19
```

```
> chisq.test(look)$expected
```

```
      assert invit/req instruct
HAVE a(n) ADJ look 54.22222 14.66667 11.11111
TAKE a(n) ADJ look 67.77778 18.33333 13.88889
```

The first table is the one with the real observations – it could also be produced by using the command `chisq.test(look)$observed` – and the second one reproduces the table in such a way that all bumps and dips in the proportions are smoothed out. Note that the total observations for *have a look* is 80 and for *take a look* is 100 in our sample, which means that for each function, we could expect a four-to-five ratio under 'chance' conditions. So, per four tokens of *have a look*, there should then be five tokens for *take a look*, whatever the pragmatic function. Indeed, if you divide any expected frequency for *have a look* by the expected frequency for *take a look* just below that cell, you will get as a result 0.8 (or something very close to it, like 0.799999856). To get an idea of how far the observed frequencies are removed from the expected ones, we can ask for 'Pearson residuals' to be computed:

```
> chisq.test(look)$residuals
```

```
      assert invit/req instruct
HAVE a(n) ADJ look -0.5733931 2.437087 -1.533333
TAKE a(n) ADJ look 0.5128583 -2.179797 1.371455
```

These values are the differences between the observed and the expected values divided by the squared root of the expected value. A negative value is obtained when the observed frequency is smaller than the expected frequency; a positive value when it is greater. The higher the absolute value of a Pearson residual, the larger the difference between observed and expected frequency. Are the differences we see here sufficiently large to reject the ‘null hypothesis’, which says that there is no association between choice of verb and function? Of course, even if the null hypothesis were true, the observed frequencies are always likely to deviate somewhat from the ‘expected’ ones, which therefore, paradoxically, are not really *that* expected. The point is, though, that the null hypothesis becomes increasingly implausible the wider the gap is between the observed proportions and the ‘ideal’ ones (in our case, 4/5). To find out whether the difference is sufficiently large to be statistically significant, we can apply the ‘chi square’ ( $\chi^2$ ) test. This yields a p-value, which tells you how probable it is that the null hypothesis can still be true if you have a difference that is at least as big as that between the observed and the expected values. If the p-value is lower than 0.05, this means that there is less than 5% probability that the observed values can deviate as much from the expected values as they do under the null hypothesis being true. In other words, that’s rather too low a probability to still maintain that null hypothesis. Especially if the p-value dives under 0.01, let alone under 0.001, there remains hardly any reason at all to hold on to the assumption of no association between variables. This is how we can calculate chi square and the corresponding p-value in R:

```
> chisq.test(look)
```

Pearson’s Chi-squared test

data: look

X-squared = 15.515, df = 2, p-value = 0.0004276

What is meant by ‘df’ is degrees of freedom, which we always have to report, along with the chi square value and the p-value. For a two-by-three contingency table, the number of degrees of freedom is two. Why? Because there are only two cells that could be freely filled in. Once, say, two row values are randomly chosen, then we know what the third value is, as in a kind of Sudoku puzzle, based on that row’s marginal total. And then, once the entire first row is filled in, we have no choices left for the second row at all, because the marginal column totals tell us what these values should be. Importantly, here, the p-value is lower than 0.001, so we can report this result as follows: “We have found a significant association between verb choice in the English *have/take a(n) ADJ look* pattern and pragmatic function:  $\chi^2(2) = 15.52$ ,  $p < 0.001$ .”

Many linguists would leave it at that, but apart from knowing now that choosing one verb rather the other has an effect on which pragmatic value is likely to be realized,

we would also like to get an idea of the effect *size*. For this, we can use the function **assocstats()** (which stands for ‘association statistics’):

```
> assocstats(look)

X^2 df  P(> X^2)
Likelihood Ratio 15.940  2 0.00034561
Pearson          15.515  2 0.00042759

Phi-Coefficient  : NA
Contingency Coeff.: 0.282
Cramer's V       : 0.294
```

The function returns a value between 0 and 1 for three measures of effect size; however, with tables larger than two-by-two, the  $\phi$ -coefficient is unreliable and therefore not returned (‘NA’ standing for ‘not applicable’). It is recommended that for a two-by-three table, Cramér’s *V* is reported. In our case, the effect size is moderate (between 0.21 and 0.35), which means that, while the result of the chi-square test is significant, the two categorical variables (verb and function) are only moderately associated – the effect could be weaker but it could also be stronger. (Between 0 and 0.07, the association can be considered negligible; between 0.07 and 0.21 weak; between 0.21 and 0.35, as we said, moderate; over 0.35 strong. For a two-by-two contingency table, however, the association is weak for a value between 0.1 and 0.3, moderate for a value between 0.3 and 0.5, and strong only for a value over 0.5.

To get an idea of the direction(s) of association, we can calculate the odds ratios (ORs) for pairwise comparisons between the different pragmatic uses. To do so, we divide the odds for one use by the odds for another. First, though, let’s again reproduce the table so that we have direct access at the numbers:

```
> look

      assert invit/req instruct
HAVE a(n) ADJ look   50    24    6
TAKE a(n) ADJ look   72     9   19
```

Odds are ratios already in themselves, so an odds ratio is a ratio (i.e. a division) of two ratios (i.e. two divisions). It doesn’t matter which odds you use as numerator for each division, but it may be easiest to report your results if you take the largest of the two:

```
> (24/9)/(50/72)

[1] 3.84
```

```
> (24/9)/(6/19)
```

```
[1] 8.444444
```

```
> (50/72)/(6/19)
```

```
[1] 2.199074
```

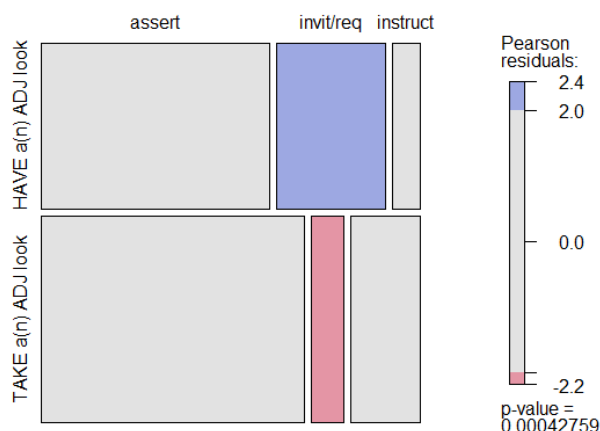
We can formulate these results as follows: “The odds that *V a(n) ADJ look* is used as an invitation or request rather than as an assertive is almost four times higher with *have* than with *take*; the odds that this construction is used as an invitation or request rather than as an instruction is more than eight times higher with *have* than with *take*; finally, the odds that it is used as an assertive rather than as an instruction is about twice as high with *have* as with *take*; Cramér’s  $V = 0.294$ ”. Note that it is customary to add to such a description of the odds ratio(s) the value of an effect size measure, which we obtained above.

For a contingency table, especially one of a higher than two-by-two dimension, we can nicely capture the conspicuous overuse and underuse of certain fields in the table in a so-called ‘mosaic plot’:

```
> mosaic(look, shade = TRUE, varnames = FALSE)
```

This generates the visualisation shown in Figure 17.

**Figure 17:** Mosaic plot of pragmatic uses of *HAVE a(n) ADJ look* versus *TAKE a(n) ADJ look*





Blue tiles represent positive residuals (light blue: between 2 and 4). Reddish tiles represent negative residuals (pink: between -2 and -4). There are only moderate cases of over- and underuse in our data. (Values above 4 would show in dark blue and values below 4 in magenta.) The plot shows us immediately that invitations and requests are conspicuously common with *have* and conspicuously uncommon with *take*.

Finally, we may want to inspect which residuals make a significant contribution to this distribution, something which standardized residuals (i.e., residuals divided by their standard error) are used for. The following code can be entered for this:

```
> chisq.test(look)$stdres
```

This produces the table below:

	assert	invit/req	instruct
HAVE a(n) ADJ look	-1.355224	3.618136	-2.216885
TAKE a(n) ADJ look	1.355224	-3.618136	2.216885

Cells with standardized residuals greater than 1.96 or smaller than -1.96 contain values that deviate significantly from expected values at the 0.05 level of significance. With values exceeding  $\pm 2.58$ , the deviation reaches the 0.01 level of significance, and with values outside  $\pm 3.29$ , there is significance at the 0.001 level. The table we have obtained allows us to see that cells in the assertives column fail to make any significant contribution, that cells in the invitations and requests column make a statistically extremely significant contribution at the 0.001 level, and that the cells in the instructions column also contribute significantly to the obtained  $\chi^2$ -statistic value<sup>4</sup>, but only at the 0.05 level. This means that we can confidently conclude that instructions and especially invitations/requests are pragmatic functions for which speakers make a clearly different selection of *have a(n) ADJ look* versus *take a(n) ADJ look*.

### 2.3.4 GIVING SOMEONE A CERTAIN LOOK: FINDING ALTERNATIVES TO THE VERB

Remember that there is also pattern in which more than just looking is involved (cf. meaning 2 in Figure 1), namely where someone gives or sends – again metaphorically, of course – a particular look to another person (or, less likely, an

4 For a prepared spreadsheet document that processes the results in a table and produces an interpretation that you can copy and paste into your paper, go to: [https://www.researchgate.net/publication/361115565\\_Chi\\_square\\_effect\\_size\\_and\\_odds\\_ratio\\_calculator\\_for\\_corpus\\_linguistics](https://www.researchgate.net/publication/361115565_Chi_square_effect_size_and_odds_ratio_calculator_for_corpus_linguistics)

object). An added aspect of meaning then is that the looker reveals their emotion while looking at someone. To find out which verbs are used in the pattern, apart from *give* or *send*, we could try to retrieve instances of the so-called ditransitive pattern, with the recipient of the look realized as a personal pronoun (*me, you, him, her, us, them*). Such a search query is shown in Figure 18. The results the interface produces are displayed in Figure 19. The search was again carried out in the British National Corpus (BNC, Davies 2004).

**Figure 18:** Search query designed to retrieve instances corresponding to the expression *give someone a(n) ADJ look*<sup>5</sup>

ListChartCollocatesCompareKWIC

PRON a|an ADJ look

Word/phrase [POS] ?

VERB

Collocates [POS]

+ 4 3 2 1 0 0 1 2 3 4 +

Find collocates

Reset

☐ Sections

Texts/Virtual

Sort/Limit

Options

# HITS

100

# KWIC

200

GROUP BY

LEMMAS

DISPLAY

RAW FREQ

SAVE LISTS

NO

**Figure 19:** Results retrieved by the query: verb lemmas used in this pattern, ordered by frequency

HELP			FREQ
1	<input type="checkbox"/>	[GIVE]	224
2	<input type="checkbox"/>	[SHOOT]	47
3	<input type="checkbox"/>	[THROW]	24
4	<input type="checkbox"/>	[SEND]	13
5	<input type="checkbox"/>	[CAST]	9
6	<input type="checkbox"/>	[FLING]	5
7	<input type="checkbox"/>	[SLIDE]	4
8	<input type="checkbox"/>	[DART]	3
9	<input type="checkbox"/>	[FLICK]	2
10	<input type="checkbox"/>	[FLASH]	2
11	<input type="checkbox"/>	[LEND]	2
12	<input type="checkbox"/>	[TOSS]	1
13	<input type="checkbox"/>	[SLANT]	1
14	<input type="checkbox"/>	[OFFER]	1
15	<input type="checkbox"/>	[EARN]	1
16	<input type="checkbox"/>	[DELIVER]	1
		TOTAL	340

5 In addition to looking for this structure, with an adjective before the noun *look*, we could formulate a query that retrieves instances with *look* followed by *of ADJ NOUN*, such as *give someone a look of {utter contempt / pure disgust / mild surprise}*.

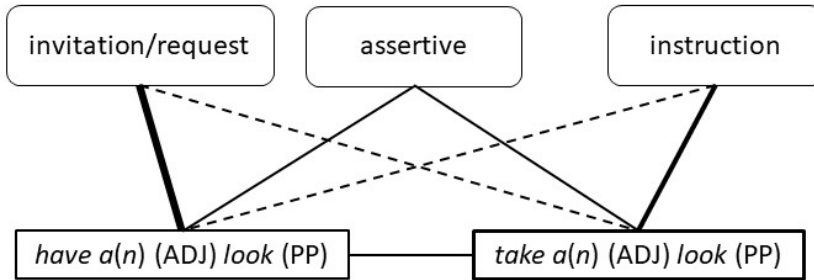
Grouping the results by lemmas, which we did in the Options tab of the search window, ensures that different inflectional forms for a single verb are not scattered in the output.

*Give* is clearly the most common verb, although we can't fully take the numbers at face value. If we click on [GIVE] in the result table and inspect the actual examples, we can see that a few are not of the desired type, such as *His thick moustache, neatly trimmed, gave him a handsome look*. Even when such examples are weeded out, *give* is still by far the most common verb. Note that apart from *give*, the verbs *shoot* and *throw* are very frequent, even more so than *send*. These and a few other verbs (*cast*, *fling*, *dart*, *flick*, *toss*) belong to what Levin (1993) calls 'verbs of throwing', that is, verbs describing "instantaneously causing ballistic motion" by imparting a force (Gropen et al. 1989).

## 2.4 IMPLICATIONS FOR THE ORGANIZATION OF LEXICAL AND GRAMMATICAL KNOWLEDGE

What we have found in the course of answering the questions raised at the end of Section 2 is that we can posit constructions of visual perception and visual expression with *look* at various levels of generality. At close to the highest level, we may find the [V *a(n)* (Adj) *look* (Adv/PP)] construction with a fairly general meaning having to do with directing one's eyes somewhere. This general construction has a few 'daughter' constructions, two of which are the by now well-known support verb constructions *have a look* and *take a look*. These are close synonyms, but each of them has nonetheless its own preferred range of uses. How can we represent their different functional profiles? In a theoretical framework such as Construction Grammar (cf., e.g., Goldberg 2003 for a quick outline of the basic tenets), idiomatic expressions are seen as constructions defined as cognitively 'entrenched' form-function units. Yet, general pragmatic functions such as asserting something, inviting someone to do something, ordering someone to do something and the like ought to be recognized as having an independent existence outside of constructions. After all, unlike the conceptual content of lexical items, such as the meaning of *cat* or *sneeze*, such a pragmatic function is not realized by a single, highly privileged linguistic form – rather, there are very many possible forms for it. We can therefore capture the relations between the forms and the pragmatic functions as in Figure 20.

**Figure 20:** Network of pragmatic functions associated with two roughly synonymous support verb constructions



This simple network contains the constructional idioms *have a(n) (Adj) look (PP)* and *take a(n) (Adj) look (PP)*, which are shown to be linked in a ‘horizontal’ relationship, as these are two sisters that have much the same meaning. They can be seen as ‘allostructions’ (Cappelle 2006), that is, two alternative formal manifestations of a slightly more general construction – but less general than the *[V a(n) (Adj) look (Adv/PP)]* construction mentioned above – in which the choice of verb (i.e., whether it is *have* or *take*) is left unspecified. It may be observed that *[take a(n) (Adj) look (PP)]* has been enclosed in a box with a somewhat thicker line, indicating it has higher frequency and is thus cognitively more strongly engraved or entrenched in the mind. We know from a sample of authentic data retrieved and analysed (cf. Section 3.3) that requests and invitations are more strongly associated with *have a look* than with *take a look* and this is indicated in the network with a thicker connecting line for the former and a thin dashed line for the latter. For instructions, the opposite pattern holds, but somewhat less pronouncedly so.

By representing the pragmatic values as external to each constructional alternative, the partial network in Figure 20 reflects a way of looking at the ‘construct-i-con’ (i.e., the repository of constructions in the minds of speakers) that contains more links between nodes, which are themselves less crammed with information (cf. Diessel 2019, Hilpert and Flach in preparation). The whole of a language user’s grammar and lexicon might thus be conceived of as a network of nodes that are linked with several other nodes. When one node is accessed (because it is selected by the speaker or processed by the hearer) its activation spreads through the network. This is compatible with the idea that pragmatics can be ‘part of’ a construction (Cappelle 2017).

Many more details could be added to this network representation, which is intentionally a simplification of the rather complex linguistic reality. For instance,

different versions of each idiom with a specific adjective (e.g. *have a quick look*, *take a good look*) are likely to be part of a competent speaker's lexicon of conventionalized sequences. The ties that these more specific idioms have with the pragmatic functions may be stronger or weaker than is the case for the two versions without a modifier shown in Figure 20. That there can be two modifiers (e.g. *take a good hard look*) suggests that the network should include even slightly more general schemas in which *look* appears as the head of an NP and in which the preceding (still optional) position is just specified to be semantically a modifier, which can then be syntactically realized by means of more than one adjective, or by an adjective that is itself modified (e.g. *take a real(ly) good look*). The presence of general schemas in the network, however, does not obviate the need for also storing specific instantiations with high usage frequency. It is Langacker (1987) who perhaps most eloquently rejected what he referred to as the 'rule/list fallacy', according to which there would be no reason to separately memorize any individual sequences that are rule-governed. While that would perhaps lead to a more economical, parsimonious representation of the speaker's construct-i-con, this may not be in line with cognition. It is worth quoting Langacker (1987: 42) here in full:

*"[S]pecific forms like beads, shoes, toes, and walls can be omitted from the grammar of English if one provides a general rule of plural formation that accounts for them. Listing regular forms thus implies lost generalizations and failure to achieve an optimal analysis.*

This line of thought illustrates the rule/list fallacy. It is fallacious because it assumes that one is forced to choose between rules and lists: the options are posed as rules alone vs. lists alone. If these are the only two options, it can be argued that the rules must be chosen, for lists by themselves do not express generalizations. There is in reality a third choice, however, namely both rules and lists. It is plausible to suppose that speakers of English master many regular plural forms as fixed units in addition to learning the general rule of plural formation. Similarly, it is clear that speakers learn as fixed units a large number of conventional expressions that are nevertheless fully analyzable and regular in formation (absolutely incredible, cheap imitation, great idea, turn the page, let the cat out, and so forth). General statements and particular statements can perfectly well coexist in the cognitive representation of linguistic phenomena, just as we learn certain products by rote in addition to mastering general procedures for multiplication. To the extent that this is so, an accurate linguistic description claiming psychological reality must contain both rules expressing generalizations and specific forms learned as fixed units, even if the specific forms accord fully with the rules." (Langacker 1987: 42)

Applied to our constructions, just because *have a quick look* or *take a long hard look* are sequences in full accordance with 'rules' (i.e., schematic patterns) allowing for one or more optional adjectives, this does not mean that these specific sequences cannot be entered in the constructional network as conventional, routinized units.

## 2.5 IDEAS FOR FURTHER DISCUSSION AND PRACTICE ACTIVITIES

Our demonstration has shown how corpora can be fruitfully used to test predictions based on what we find in dictionaries. Sometimes our intuitions also tell us something about what is probable, and such intuitions, too, can be confirmed or disconfirmed with the help of corpora. Below are some ideas for further corpus exploration that you might find useful to try out if you're a beginning corpus user. Teachers could also use these as class or homework activities. The first two exercises are quite easy; a more challenging task, involving some statistical processing of data, is given at the end.

### 2.5.1 HAVE GOT A LOOK

Do you think *have got a look* can be used as an alternative of *have* (or *take*) *a look*? Use a corpus to find out whether and how often *have got a look* occurs with the meaning of 'have/take a look'. What can you conclude from these quantitative results?

### 2.5.2 THE DOCUMENTARY HAS/TAKES A LOOK AT...

Based on what you now know about the semantic differences of *have a look* and *take a look*, which of these do you think is more likely to occur: *The documentary has a(n) ADJ look at...* or *The documentary takes a(n) ADJ look at...*? Why?

Do you know the name of the linguistic phenomenon whereby a person (in this case a documentary maker) is referred to indirectly, via the term of a concept closely associated with it (here, the thing made)? Can you find further such examples by means of a corpus?

### 2.5.3 HAVE/TAKE A LOOK AT GENRES

Which of the two alternatives, *have a look* or *take a look*, do you think is most common in spoken language? Present some data obtained from a corpus in a table and/or a chart that you produce yourself, without copying-and-pasting from a screen. (Hint: in the same way that GloWbE has different language varieties as 'sections', mono-variety corpora such as the BNC and COCA are composed of data from different genres such as spoken language, fiction, magazines, newspapers, academic books and papers, etc. With the Chart function, you can see the frequency of a search string by genre.)

Use R to carry out a statistical analysis of the difference you find between spoken and non-spoken English and the choice of expression.



## REFERENCES

- Cappelle, B. 2006. Particle placement and the case for ‘allostructions’. *Constructions* SV1-7/2006, 1–28. [www.constructions-online.de](http://www.constructions-online.de), urn:nbn:de:0009-4-6839, ISSN 1860-2010. Last accessed 1 January 2020.
- Cappelle, B. 2017. What’s pragmatics doing outside constructions? In I. Depraetere, & R. Salkie (Eds.), *Semantics and pragmatics. Drawing a line* (115–151). Cham: Springer.
- Cappelle, B. 2020. Looking into visual motion expressions in Dutch, English and French: How languages stick to well-trodden typological paths. In Y. Matsumoto, & K. Kawachi (Eds.), *Broader Perspectives on Motion Event Descriptions* (235–279). Amsterdam & Philadelphia: John Benjamins.
- Cifuentes-Férez, P. 2014. A closer look at Paths of vision, Manner of vision and their translation from English into Spanish. *Languages in Contrast*, 14(2), 214–250.
- Claes, Jeroen. 2019. R Crashcourse for linguists. Available online at [http://www.jeroenclaes.be/r\\_crashcourse\\_for\\_linguists/](http://www.jeroenclaes.be/r_crashcourse_for_linguists/).
- Danlos, L. 1992. Support verb constructions: Linguistic properties, representation, translation. *Journal of French Language Studies*, 2(1), 1–32. doi:10.1017/S0959269500001137
- Davies, M. 2004. *British National Corpus* (from Oxford University Press). Available online at <https://www.english-corpora.org/bnc/>.
- Davies, M. 2008. New directions in Spanish and Portuguese corpus linguistics. *Studies in Hispanic and Lusophone Linguistics*, 1(1), 149–186.
- Davies, M. 2008-. *The Corpus of Contemporary American English* (COCA). Available online at <https://www.english-corpora.org/coca/>.
- Davies, M. 2009. Creating useful historical corpora: A comparison of Corde, The Corpus del Español, and the Corpus do Português. In A. Enrique-Arias (Ed.), *Diacronía de las lenguas iberorrománicas Nuevas aportaciones desde la lingüística de corpus* (137–166). Madrid: Iberoamericana/Vervuert.
- Davies, M. 2013. *Corpus of Global Web-Based English*. Available online at <https://www.english-corpora.org/glowbe/>.
- Davies, M. 2016-a. *Corpus del Español: Web/Dialects*. Available online at <http://www.corpusdelespanol.org/web-dial/>.
- Davies, M. 2016-b. *Corpus do Português: Web/Dialects*. Available online at <http://www.corpusdoportugues.org/web-dial/>.
- Desagulier, G. 2017. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. New York: Springer.
- Diesel, H. 2019. *The Grammar Network: How Linguistic Structure Is Shaped by Language*

- Use. Cambridge: Cambridge University Press.
- Flach, S. 2021. *Collostructions: An R implementation for the family of collostructional methods*. Package version v.0.2.0. <https://sfla.ch/collostructions/>.
- Goldberg, A. E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224.
- Gries, S. Th. 2021. Statistics for Linguistics with R: A Practical Introduction (3rd revised ed.). Berlin: Walter de Gruyter.
- Gries, S. Th. & A. Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspectives on ‘alternations’. *International Journal of Corpus Linguistics*, 9(1), 97–129.
- Gropen, J., S. Pinker, M. Hollander, R. Goldberg and R. Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language*, 65(2), 203–257.
- Gruber, J. S. 1967. Look and see. *Language*, 43(4), 937–947.
- Hilpert, M. & S. Flach. In preparation. Modals in the network model of Construction Grammar. In I. Depraetere, B. Cappelle, & M. Hilpert (Eds.), *Models of Modals: From Pragmatics and Corpus Linguistics to Machine Learning*. Berlin: Mouton de Gruyter.
- Kawachi, K. 2020. Should Talmy’s motion typology be expanded to visual motion?: An investigation into expressions of motion, agentive motion, and visual motion in Sidaama (Sidamo). In Y. Matsumoto, & K. Kawachi (Eds.), *Broader perspectives on motion event descriptions* (205–234). Amsterdam & Philadelphia: John Benjamins.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar. Theoretical Prerequisites* (Volume I). Stanford: Stanford University Press.
- Langer, S. 2004. A linguistic test battery for support verb constructions. *Linguisticae Investigationes*, 27(2), 171–184.
- Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago & London: University of Chicago Press.
- look. *LCOED Online*, Longman, June 2021, <https://www.ldoceonline.com/dictionary/look>. Accessed 27 June 2021.
- look, n. *OED Online*, Oxford University Press, June 2021, [www.oed.com/view/Entry/110129](http://www.oed.com/view/Entry/110129). Accessed 27 June 2021.
- Matsumoto, Y. 2001. Lexicalization patterns and caused and fictive motion: The case of typological split. Handout for a lecture at SUNY Buffalo, NY.
- Matsumoto, Y., et al. 2021. Linguistic representations of visual motion: A crosslinguistic experimental study.
- Namer, F. 1998. Support verb constructions. In F. van Eynde, and P. Schmidt (Eds.),

*Linguistic specifications for typed feature structure formalisms* (315–343). Luxembourg: Office for Official Publications of the European Communities.

R Core Team. 2018. R version 3.5.0. R: A language and environment for statistical computing. R Foundation for Statistical Computing Platform, Vienna, Austria. <https://www.R-project.org/>.

Talmy, L. 1996. Fictive motion in language and ‘ception’. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (211–276). Cambridge, MA & London: The MIT Press.

Talmy, L. 2000. *Toward a cognitive semantics* (2 Volumes). Cambridge, MA: MIT Press.

Talmy, L. 2018. *Ten lectures on cognitive semantics*. Leiden, The Netherlands: Koninklijke Brill.