CHAPTER 20

Constructional change and motion charts

What you will learn from this chapter:

This chapter introduces motion charts as a method for dynamic visualization of language change. More specifically, they enable one to detect and explore changes in the use of constructions by visualizing the relative frequencies of different lexemes that fill in the constructional slots. The method is illustrated with a case study that explores the changes in the use of future markers *will* and *be going to* by comparing the frequencies of infinitives that follow the markers.

20.1 The past and present of the future: Diachronic motion charts of *be going to* and *will*

20.1.1 Theoretical background and data

To perform this case study, you will need two add-on packages, which have to be installed and loaded:

```
> install.packages("googleVis")
> library(Rling); library(googleVis)
```

The landscape of future markers in English has been changing for centuries. On the one hand, the auxiliary *shall*, which used to be obligatory with the first person subjects, has given way to *will*. On the other hand, *be going to* has developed from a purposive motion construction to a full-fledged future marker. As a result, its frequency has been on the rise, as a screenshot of the Google Books Ngram viewer in Figure 20.1 demonstrates.

In this case study, we will zoom in on the 'division of labour' between *will* and *be going to*. How has their use changed since the beginning of the nineteenth century? More specifically, has *be going to* taken over some 'territory' of *will*?

We will use the data based on ngrams from the Corpus of Historical American English, or COHA (Davies 2011). In line with the distributional approach to semantics, we will investigate if there are any changes in the distribution of the infinitives that follow these markers. The infinitives were extracted from the ngram lists of *am/is/are going to + Inf* and *will + Inf*. The auxiliaries *be* and *have* were removed from the collexeme list. Only verbs that occur in all time periods (see below) were considered.

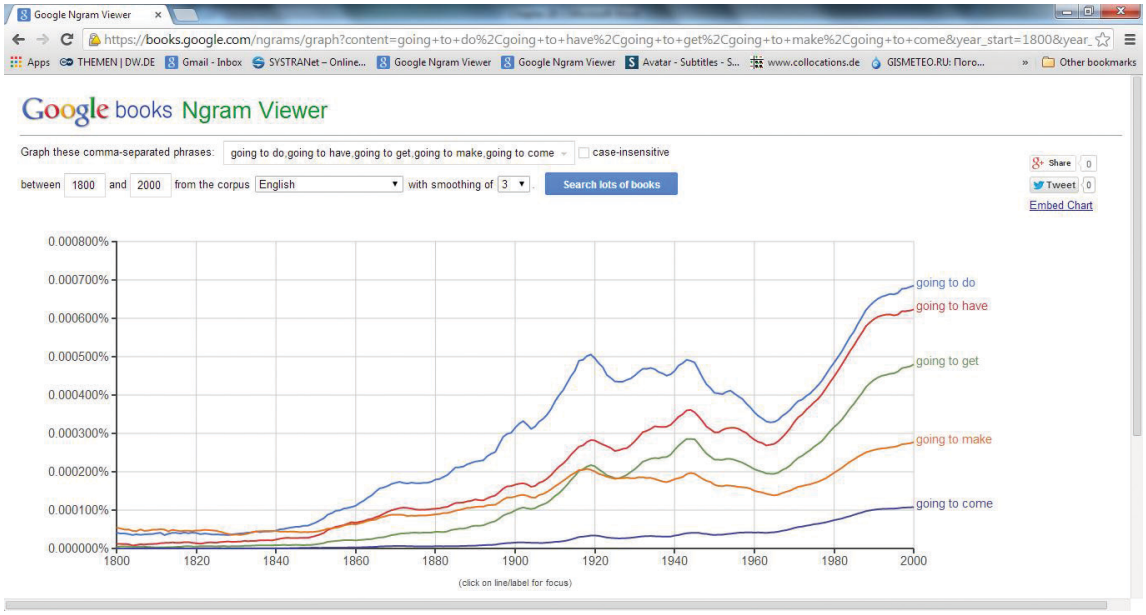

Figure 20.1. Google Ngram Viewer displaying frequencies of five ngrams with *going to +V* from 1800 to 2000

The dataset is called `fut`. It has the following structure:

```
> data(fut)
> str(fut)
'data.frame': 6973 obs. of 4 variables:
$ Verb : Factor w/ 367 levels "accept","accompany",..: 1 1 1 1 1 1
1 1 1 1 …
$ Period: num 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 …
$ will:     num   2.02  1.52   1.81   2.55   4.38 …
$ goingto: num   0     0      0      0      0 …
```

```
> head(fut)
Verb        Period   will        goingto
1 accept    1820     2.021075    0
2 accept    1830     1.524613    0
3 accept    1840     1.807208    0
4 accept    1850     2.546407    0
5 accept    1860     4.379536    0
6 accept    1870     3.385248    0
```

The number of unique verbs in the column `Verb` is 367. The periods examined are nineteen decades from the 1820s to the 2000s. In fact, COHA covers the period from the 1810s to the 2000s, but the subcorpus that represents the first decade is much smaller than the other subcorpora, so the 1810s were excluded. As a result, we have nineteen decades, which are shown in the column `Period`. The columns `will` and `goingto` contain the normalized frequencies (per million words) of the infinitives in the *will* and *be going to* construction, respectively.

20.1.2 Motion charts

To find out if there are any changes in the division of labour between these future constructions, we will use a visualization technique called motion charts. It was introduced to diachronic linguistic analysis by Martin Hilpert (2011).[1] Motion charts represent linguistic changes in time dynamically as changes in the position of some objects (for example, collexemes), with regard to two dimensions (typically, two near-synonymous constructions). This allows one to see how the associations between a word and the constructions change over time, which may be interpreted as evidence of semantic change. To construct motion charts, one can use the package `googleVis`. The function `gvisMotionChart()` creates an object referring to the Google Visualization API (Application Programming Interface). To view the output one needs a browser Flash and an Internet connection. In order to capture a motion chart for off-line presentation, e.g. in a PPT slideshow, one additionally needs screen-capturing software.

There are only three obligatory arguments of the function: the data frame itself, `idvar`, which specifies the column in the data frame that contains the identities of the points, and `timevar`, which specifies the column that corresponds to the time dimension. The other options are described on the help page of the function.

```
> mch <- gvisMotionChart(fut, idvar = "Verb", timevar = "Period")
> plot(mch)
```

The resulting plot, which is shown in Figure 20.2, opens in a web browser. The horizontal axis shows the normalized frequencies of the verbs after *will*. The vertical axis demonstrates the frequencies after *be going to*. The bubbles correspond to the individual verbs. They can be coloured in different colours if you choose this option from the drop-down menu *Color* in the top right corner. One can also change the size of the bubbles to represent their frequencies with either construction with the help of the drop-down menu *Size*. The frequencies are shown on the linear scale, although one can represent them on the logarithmic scale (see the buttons at the ends of the axes). In the latter case, the scale will be stretched at the lower range and shrunk for high frequencies.

One can see that the normalized frequencies of the verbs in the 1820s are much higher in the *will* construction. To see how they change over time, click on the *Play* button (the black triangle) at the bottom left corner of the plot. The circles will begin to move. The algorithm interpolates the frequencies for every year between the decades, so that the motion appears continuous. Slow motion corresponds to small changes in the frequencies, whereas quick motion shows that the changes are big. To stop the motion, one can click on the *Pause* button.

Figure 20.3 shows the verb frequencies in the 1900s. One can see that the frequencies of verbs in *be going to* future have increased. Still, the normalized frequencies of *will* have not decreased, either. For a few most frequent verbs, they have even slightly increased.

---

1. See http://members.unine.ch/martin.hilpert/motion.html, where you can find video tutorials and a number of interesting examples (last access 11.06.2015).

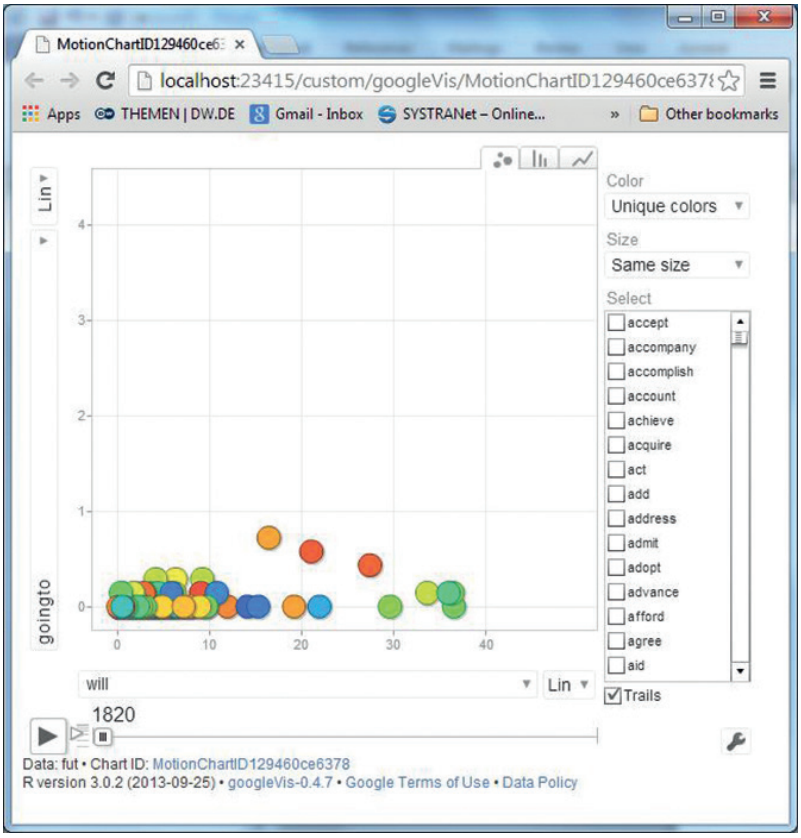

Figure 20.2. A screenshot of a motion chart of the verbs used with *be going to* and *will*: the 1820s

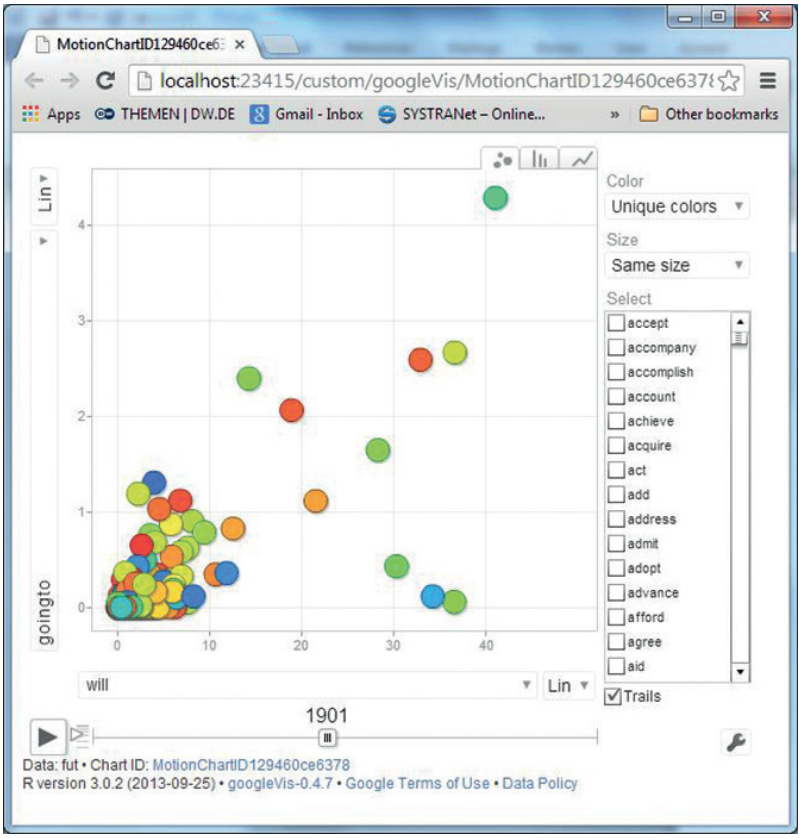

Figure 20.3. A screenshot of a motion chart of the verbs used with *be going to* and *will*: the 1900s

Figure 20.4 displays the frequencies in the 2000s, the last decade captured by the corpus. It is apparent that many verbs are now used more frequently with *be going to*. At the same time, the use of *will* has slightly decreased. Still, it would be wrong to say that *be going to* is the predominant marker because the weighted frequency of *will* is still greater than that of *be going to*. This will become obvious when you compare the units of the horizontal and vertical axes.

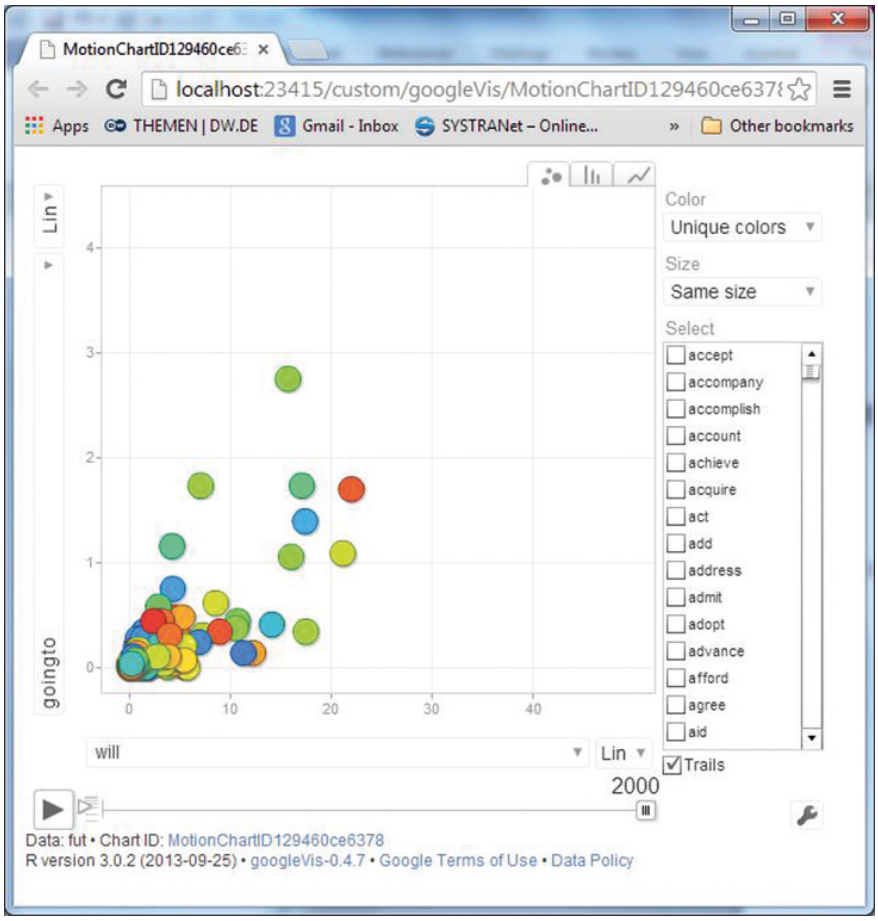

Figure 20.4. A screenshot of a motion chart of the verbs used with *be going to* and *will*: the 2000s

In addition to this general picture, one can also trace the changes in the patterns of use of individual verbs. One can click on the circles, or check the boxes corresponding to the individual verbs in the scrollable list on the right. One can also trace the trajectories of changes of a word by checking the *Trail* box in the bottom right corner (normally it should be checked by default). For instance, Figure 20.5 displays the trajectory of *take* over the entire period. It shows that the constructional preferences of *take* have somewhat shifted towards the future construction with *be going to*. However, this development was not linear. There was an intermediate period, when the verb had increasing frequencies with both *will* and *be going to*. Still, after this period the frequency of the verb with *will* has slightly decreased.

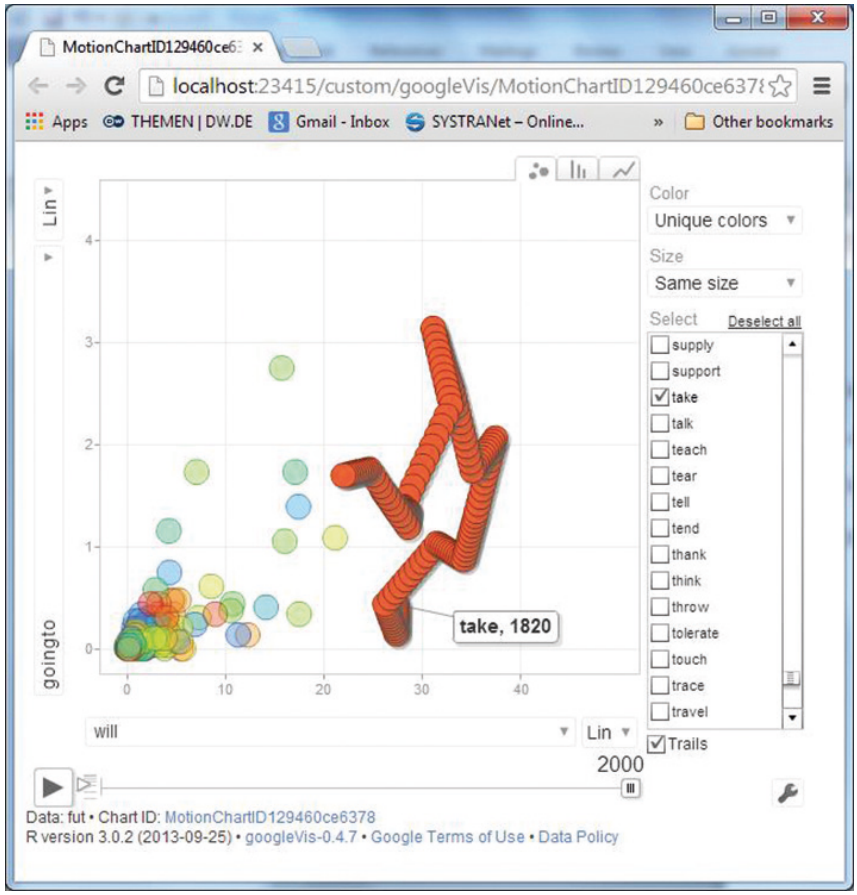

Figure 20.5. The trace of *take* that shows the changing frequencies of the verb in two future constructions

Thus, the motion charts suggest growing preferences for *be going to* and dispreferences for *will*-future, although the latter still remains the predominant marker. Let us now have a look at two verbs of motion, *come* and *go*. Figure 20.6 displays their traces in time. One can see that their patterns have changed in the direction from *will* to *be going to* in a similar fashion. The possibility of *be going to* followed by a verb of motion, especially *go*, indicates that the initial meaning of directed motion is bleached. This indicates that the grammaticalization of *be going to* was still going on in the time span covered by the motion chart.

These findings are in line with previous research. For instance, Hilpert (2013: Section 2.4.4) compared the relative frequencies of *be going to* and *will* in two periods, the 1960–1980s and 1990s. The proportion of *be going to* has increased slightly with time. However, he also found an interaction of register and time. Namely, *be going to* increased its relative frequency in informal registers, whereas *will* came to be used more often in formal contexts. Informalization of speech genres may be a possible explanation of the gradual shift towards *be going to*, but one needs a multivariate analysis, such as logistic regression (see Chapter 12), to disentangle different factors that may come into play.

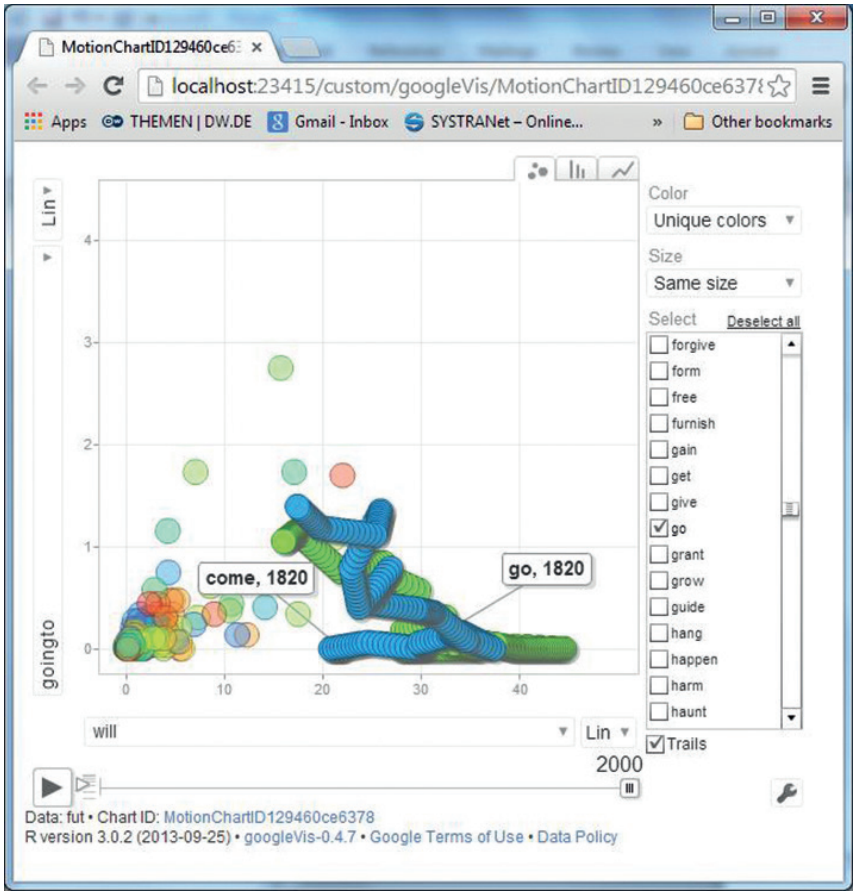

Figure 20.6. The traces of *come* and *go*

20.2 Summary

This chapter has presented motion charts, a method that has been used, among other things, for the visualization of changes in the division of labour between two constructions in time. A comparison of only two constructions may sometimes be all that is needed, but clearly human language is more complex. For example, it might be useful to include *gonna* + Infinitive in the analysis, as well as *shall* + Infinitive as a future marker, in addition to *going to* and *will*, in order to see a fuller picture. Hilpert (2011: Section 5; 2013: Section 2.6) discusses ways in which motion charts can be used to represent changes in more complex, mutivariate datasets. Still, even pairwise comparisons can help us formulate new research hypotheses, which can be later tested with the help of multivariate statistical methods, including those described in this book.