

Chapter 7: Exercises – answers

1) Interpret the following three graphs.

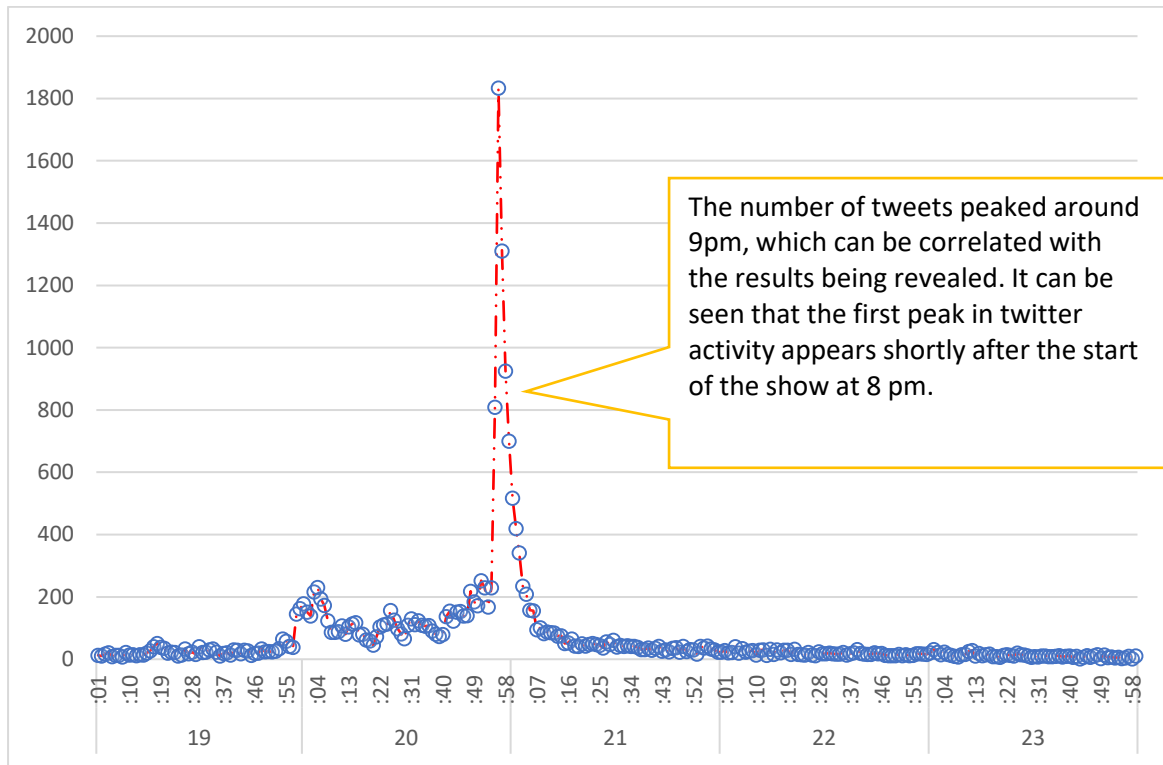


Figure 7.20 Number of tweets related to an episode of the UK X-factor: 16/11/2014, 7-11pm

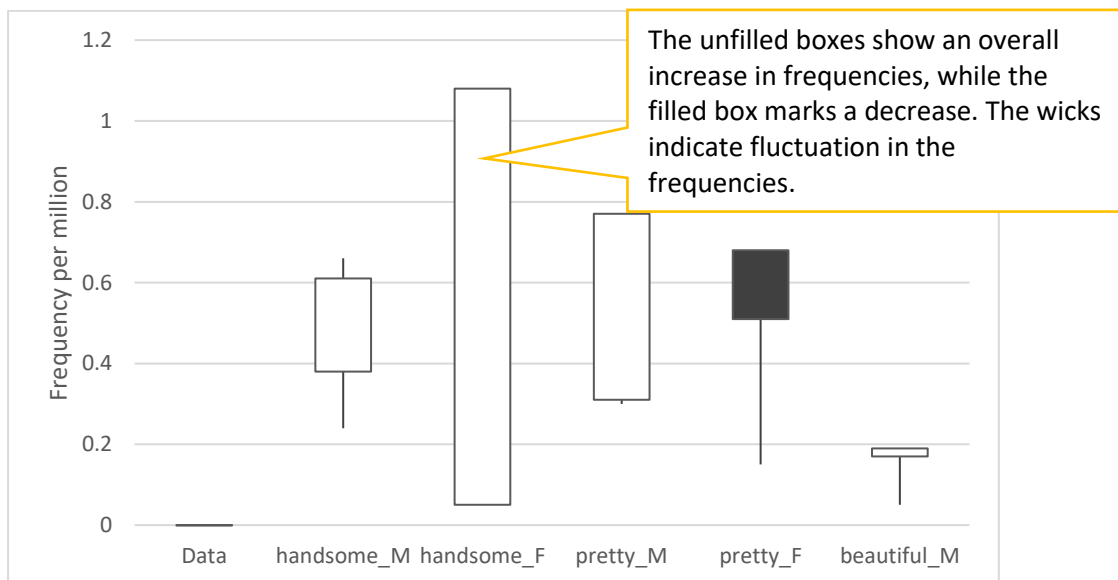


Figure 7.21 Development of frequencies of *handsome*, *pretty* and *beautiful* followed by a male (M) or female (F) person in the 17th century.

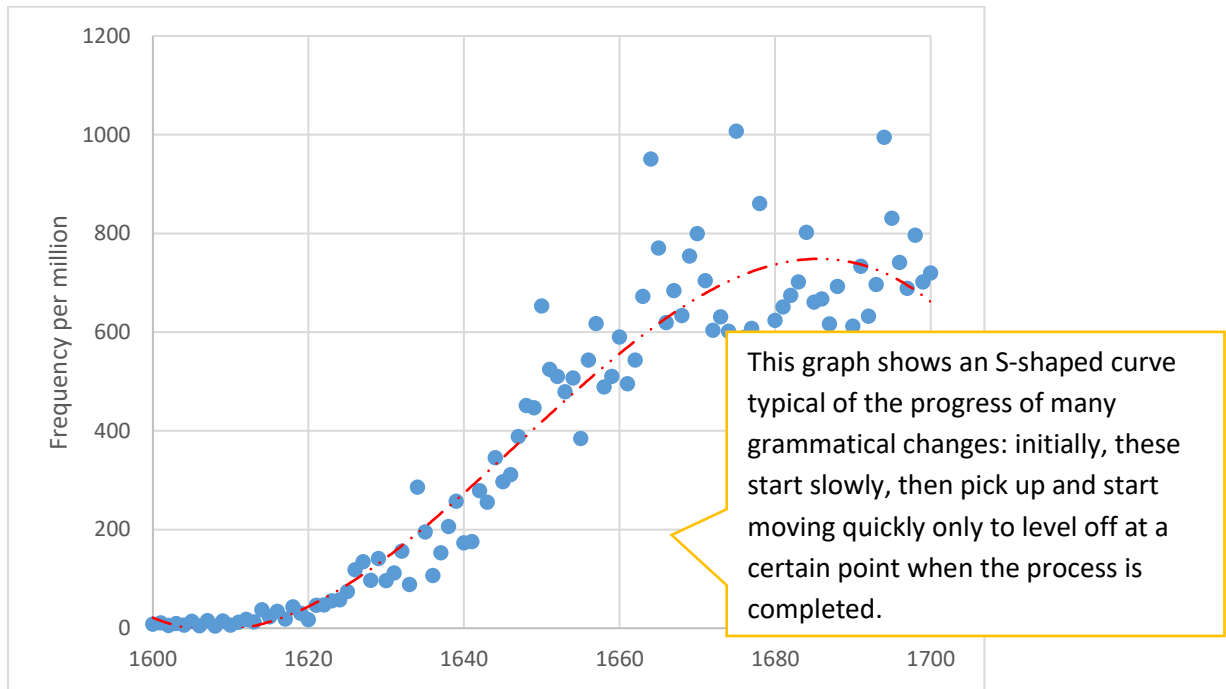


Figure 7.22 Development of frequencies of the possessive pronoun *its* in the 17th century.

2) Fill in the blanks in the descriptions below.

Over the course of the 20th century, the frequencies of the modal *shall* (), *should* (), *may* (), *might* (), *must* () and *will* () **decreased**, while the frequencies of *can* () and *could* () **increased**.

In the 17th century, the adjective *handsome* used with a female person () **increased** , while *pretty* () **fluctuated** in this context; *beautiful* () used with a female person **stayed approximately the same with a dip in the middle of the period**

3) Look critically at the trends in the four figures below. Which of these represents the largest change? **It is important to critically compare the scales on the y-axes of the line graphs. This means looking at the origin (minimum value) as well as the maximum value. It is also important to assess whether the scale is linear or not. So, while a) and b) appear to display the same trend (the slope is the same), they**

represent two different realities. a) represents a steady increase from 10,000 to 40,000, whereas b) represents a much smaller increase from 10,000 to 10,003. In fact, b) indicates the smallest increase of all the four images (by 3 points only); the other graphs a), c) and d) show the same increase (10k – 40k) on different scales. Note that the scale on the y-axis of graph d) is not linear but logarithmic (with the base of 2).

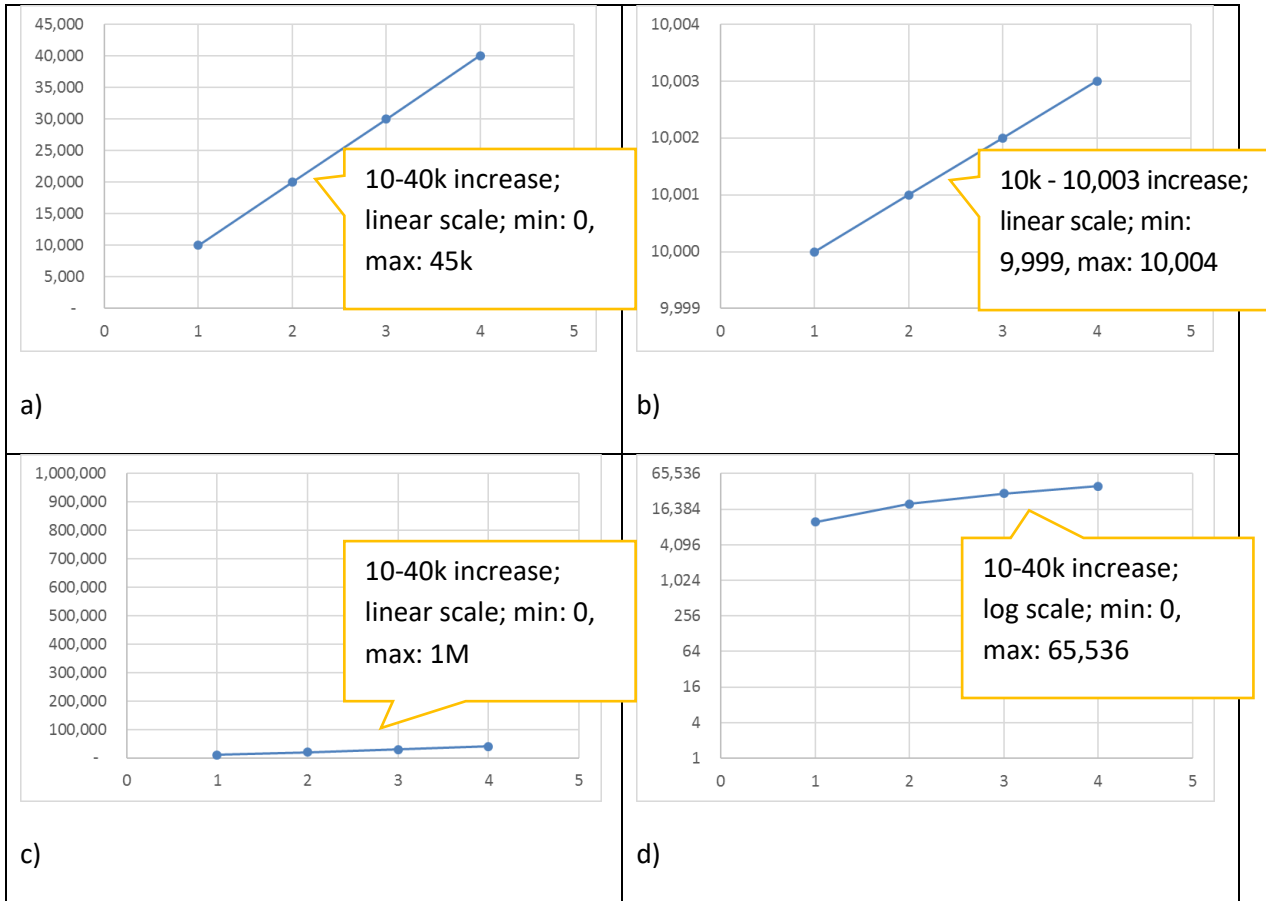


Figure 7.23 Four frequency change scenarios

- 4) Interpret the following peaks and troughs graphs showing the development of *handsome* and *pretty* in the 17th century.

Handsome shows a nice steady increase without much fluctuation with a small dip towards the end of the century. All the data points are within 99% CI limits (grey area). *Pretty*, on the other hand, shows much more fluctuation in terms of the relative frequencies per million over the course of the 17th century with the largest dip around the middle of the century. This pattern, however, is typical of lexical usage changes. Note that all data points are within 95% CI limits.

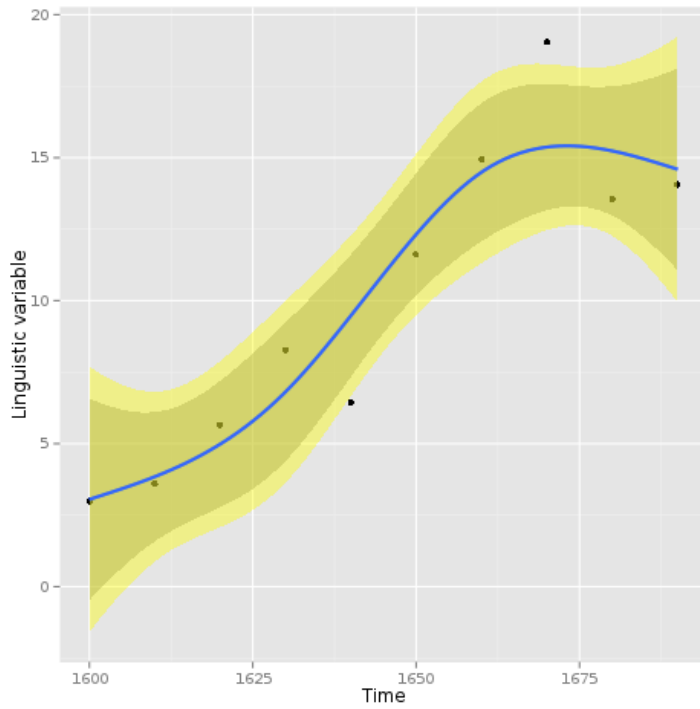


Figure 7.24 Handsome in the 17th century.

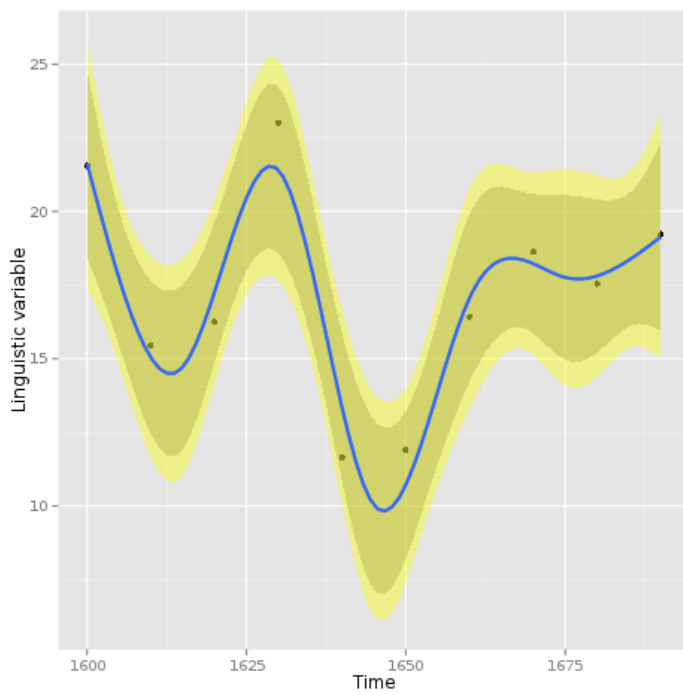
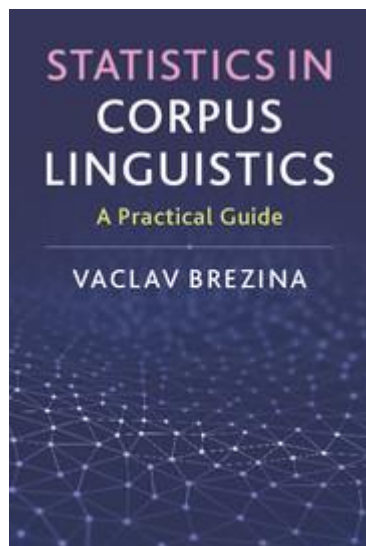


Figure 7.25 Pretty in the 17th century.



Brezina, V. (2018). [*Statistics in Corpus Linguistics: A Practical Guide*](#). Cambridge: Cambridge University Press.

Do you use language corpora in your research or study, but find that you struggle with statistics? This practical introduction will equip you to understand the key principles of statistical thinking and apply these concepts to your own research, without the need for prior statistical knowledge. The book gives step-by-step guidance through the process of statistical analysis and provides multiple examples of how statistical techniques can be used to analyse and visualise linguistic data. It also includes a useful selection of discussion questions and exercises which you can use to check your understanding.

The book comes with a Companion website, which provides additional materials (answers to exercises, datasets, advanced materials, teaching slides etc.) and [Lancaster Stats Tools online](#), a free click-and-analyse statistical tool for easy calculation of the statistical measures discussed in the book.