



TRACING SIGNIFICANT CHANGE IN 17TH-CENTURY ENGLISH LEXIS: THE CIVIL-WAR EFFECT

Tanja Säily
Department of Modern Languages
University of Helsinki
Terttu Nevalainen
Department of Modern Languages
University of Helsinki
Jefrey Lijffijt
Department of Information and Computer Science
Aalto University

INTRODUCTION

Michel et al. (2010) plotted frequencies of **culturally loaded words and phrases** over time in the Google Books database, a massive corpus of five million digitized books. They found that cultural change could be detected through a quantitative analysis of lexical change.

Could we use similar methods to investigate the lexical impact of the **English Civil War (1642–1651)** (cf. Raumolin-Brunberg 1998)?



Figure 1. Oliver Cromwell leading the New Model Army at the Battle of Naseby during the English Civil War. From Photos.com.

PROBLEMS WITH GOOGLE BOOKS

The Google Books database reveals that the use of the word *war* increases in British English books during the 17th century and that some of the peaks in the frequency coincide with the Civil War (Figure 2). There is a great deal of noise, however, and it is unclear how many of the occurrences actually refer to that particular war. Furthermore, the peaks could be accidental, as the valleys could be due to **sparseness of data** for some of the years.

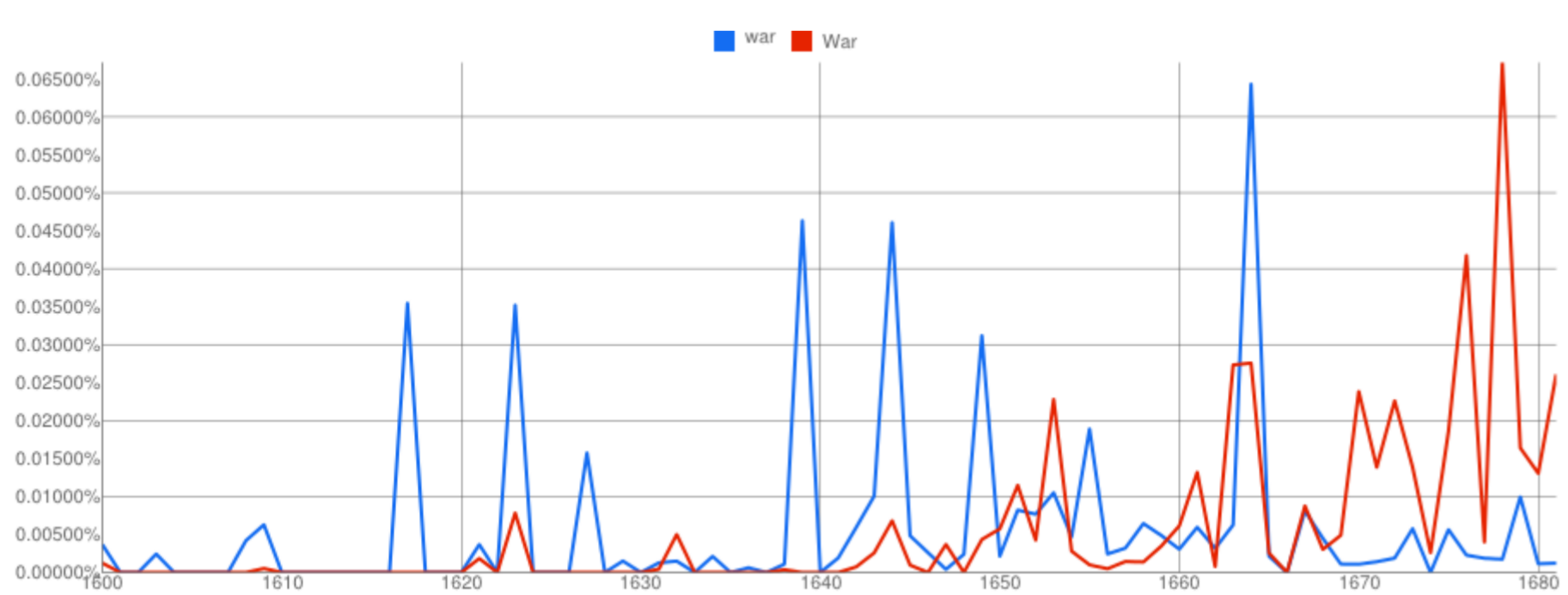


Figure 2. Normalized frequencies of the word *war* (blue = lowercase, red = capitalized) in the British English section of the Google Books Ngram Viewer, 1600–1681.

One way to improve the reliability of the Google Books data is to analyze the actual n-

gram frequencies aggregated over longer **time periods**. It turns out that the first 40-year period is very poorly represented in the Google Books database. The second period is more comprehensive, and does mark the overall increase in the use of *war* (Figure 3).

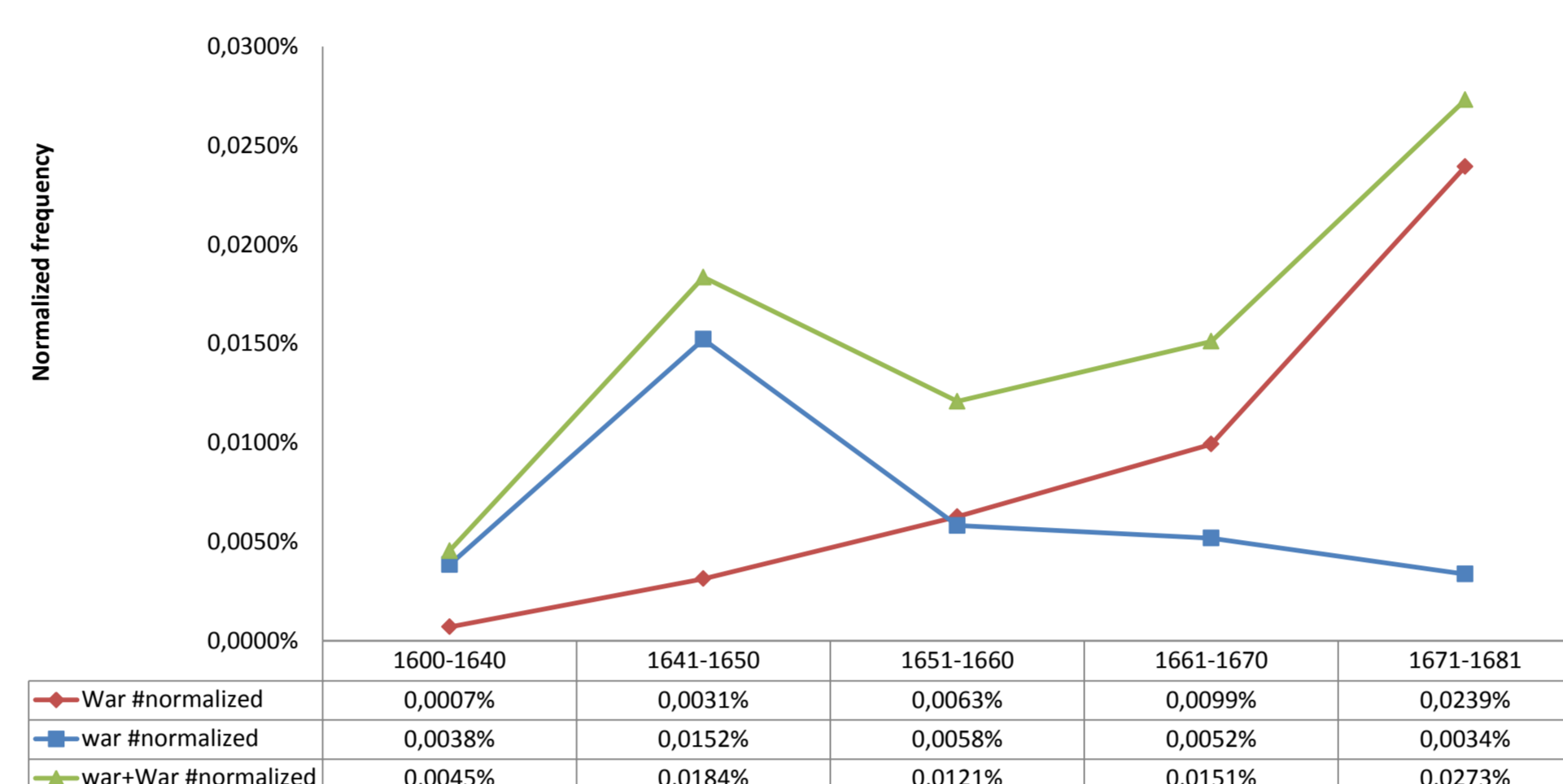


Figure 3. Normalized frequencies of the word *war* (blue = lowercase, red = capitalized, green = combined) in the British English section of the Google Books database, 1600–1681, divided into five time periods.

How about the **reactions of individuals** who experienced the war? We looked for answers in the *Corpus of Early English Correspondence* by comparing the decades during and after the War with the preceding 40-year period.

CORPUS OF EARLY ENGLISH CORRESPONDENCE

Compiled at the University of Helsinki, the 2.6-million-word *Corpus of Early English Correspondence* (CEEC) consists of 5,961 letters written by 778 people between the years 1410 and 1681. It includes **sociolinguistic metadata** on letters, writers and recipients. We used the 17th-century part of the corpus divided into two subcorpora, 1600–1639 and 1640–1681, and compared the frequencies of each word in them.

METHOD

There has been extensive debate on which statistical test is most appropriate when studying the similarity of two corpora (e.g., Kilgarriff 2001, Paquot & Bestgen 2009). We reviewed the effects of using particular statistical tests and their parameter choices for identifying significant changes. We found that **the choice of statistical test matters** (see Table 1 in the handout): tests that take into account only the number of word occurrences, such as the log-likelihood ratio test (Dunning 1993), can lead to identifying false positive

differences. This warrants the use of tests that also take into account the dispersion of words, such as the **bootstrap test** (Lijffijt et al. forthcoming).

We manually went through the 1,292 words that were significantly overused in either period according to both tests, and checked promising items against the Society > Armed Hostility section of the *Historical Thesaurus of the Oxford English Dictionary* (HT). Outside the HT, some proper nouns and the word *restoration* were also included. This resulted in a list of 70 possibly war-related words.

RESULTS

We studied the actual instances in context and determined that 35 of the 70 words had something to do with the war in at least some of the cases. Out of these words, 34 were overused during the latter period (1640–81) and only one in the former period (1600–39).

A further examination of the 34 words confirmed that at least some of the instances of each word were related to the Civil War or its aftermath. The clearest examples of **war-related overuse** in the latter period were the words *armies*, *colonel*, *militia*, *officers*, *ordnance*, *regiment*, *restoration* and *war*.

While most of the overuse was due to **men belonging to the professional rank**, the war also featured in letters written by women, such as Dorothy Osborne (Figure 4, Example 1).



Figure 4. Dorothy, Lady Temple, by Gaspar Netscher, 1671. From Wikimedia Commons.

(1) if I am not mistaken that Monk has a brother lives in Cornwall, an honest Gentleman I have heard, and one that was a great acquaintance of a Brother of mine who was *killed* there during the *War* ... (Dorothy Osborne to her future husband, William Temple, 1653)

REFERENCES

See handout.