

Keyword Analysis

Simone Rebora

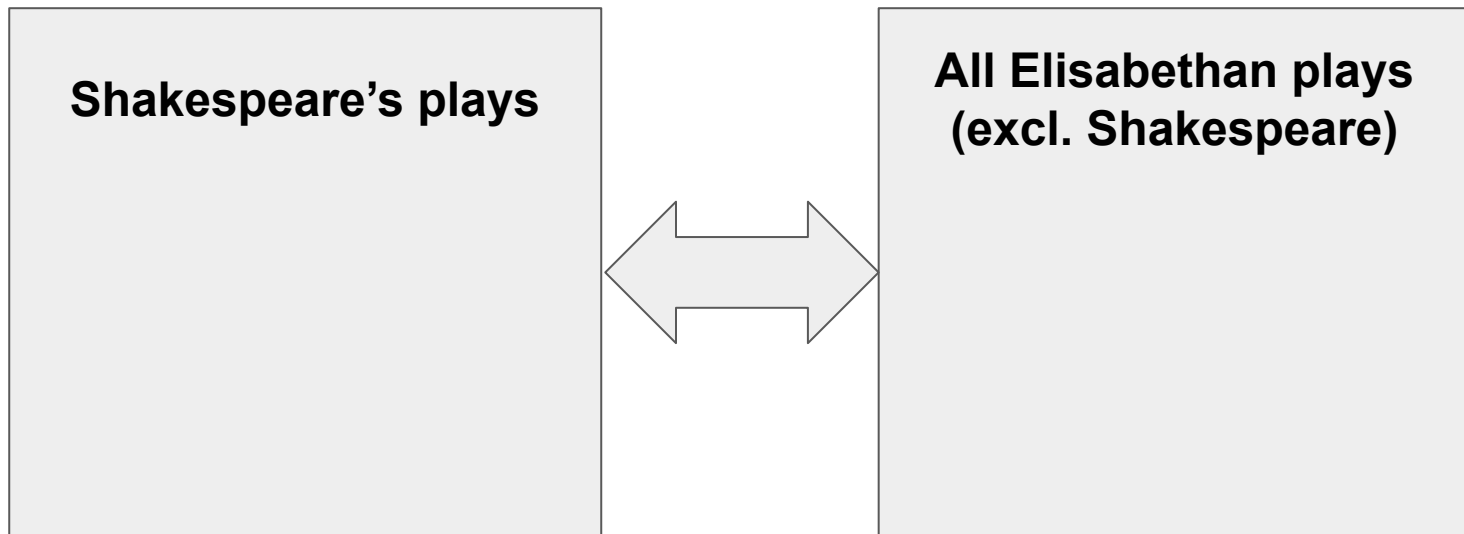
The Idea

Comparing the use of words in one document
(or a group of documents)

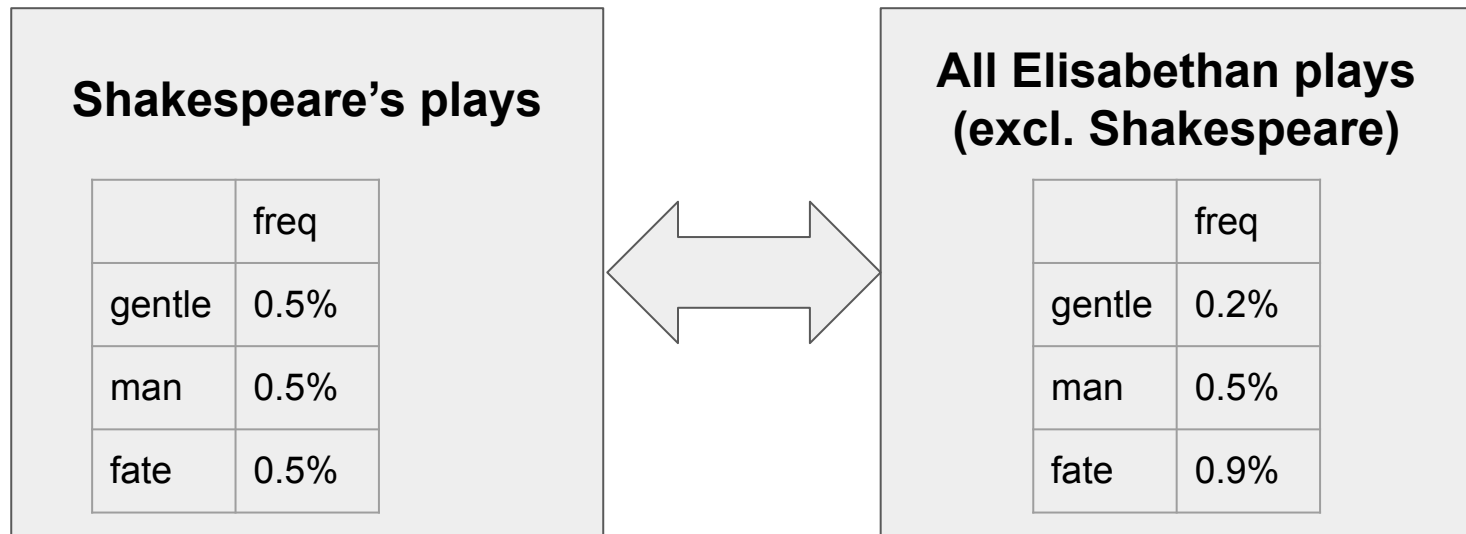
...with their expected use

Case study: see the differences between the linguistic choices of one author and common word usage

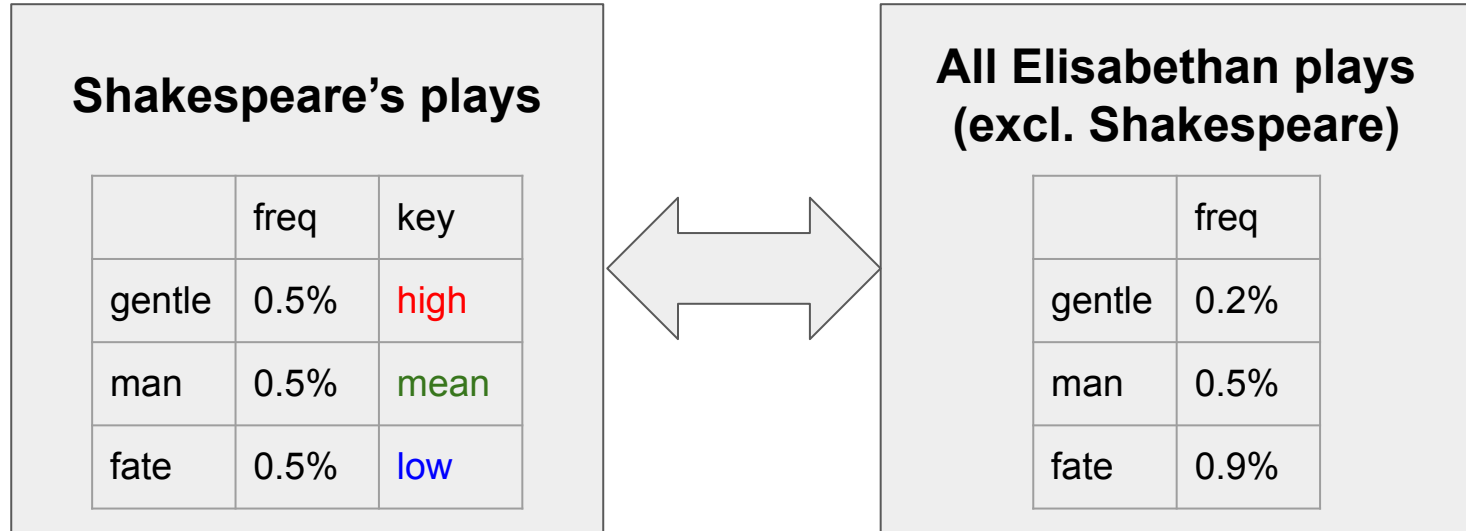
Log-likelihood



Log-likelihood



Log-likelihood



Graphical Interfaces

AntConc

<https://www.laurenceanthony.net/software/antconc/>

The screenshot displays the AntConc graphical user interface. The 'Target Corpus' is 'AmE06_Learned' with 80 files and 161469 tokens. The 'Reference Corpus' is 'AmE06' with 500 files and 1017879 tokens. A search for the word 'word' has been performed, resulting in 42 hits. The interface shows a list of 12 hits, each with a file name, left context, the hit word, and right context. The hit word is highlighted in blue, and the right context contains color-coded grammatical information (e.g., 'is a noun or adjective', 'is a specific', 'is a verb'). The 'Search Query' is 'word', and the 'Results Set' is 'All hits'. The 'Context Size' is set to '10 token(s)'. The 'Sort Options' are 'Sort to right', 'Sort 1 1R', 'Sort 2 2R', 'Sort 3 3R', and 'Order by freq'. The progress bar at the bottom indicates 100% completion. The time taken for creating the KWIC display is 0.0043 ms.

	File	Left Context	Hit	Right Context
1	AmE06_J32.txt	ing word is a verb, but not when the -ing	word	is a noun or adjective. So the grammar of
2	AmE06_J47.txt	object, action or quality. Learning the category boundaries for each	word	is a specific "problem of induction." Children are placed
3	AmE06_J32.txt	a-hunting dog. The a- is possible when the -ing	word	is a verb, but not when the -ing word
4	AmE06_J34.txt	which coincides with word-initial position, and that the entire	word	is dominated by a single syllable. The gesture-calculations
5	AmE06_J59.txt	as Marjorie Perloff puts it, or perhaps that they put	word	and image in a mutually interpretive context, as the
6	AmE06_J59.txt	Stein's verbal portraits of these painters "attempt to fuse	word	and image," as Marjorie Perloff puts it, or perhaps
7	AmE06_J60.txt	an ancient purity and directness, a pre-Babelic unity of	word	and thing. In so doing, though, photography was just
8	AmE06_J47.txt	the word, or if the child understands and says the	word.	The second component is a list of 63 communicative, social,
9	AmE06_J47.txt	to indicate if they have heard their child say the	word.	The second major component utilizes an innovative sentence pair
10	AmE06_J47.txt	Consider what would appear to be the simplest condition of	word	learning. A fluent speaker of the language, such as
11	AmE06_J47.txt	MCDI and laboratory measures of comprehension, production, and	word	learning. The second is the Twins' Early Development Study (
12	AmE06_J47.txt	Each type of reference is a plausible use of the	word.	And even if we were correct in believing that "

With R

Quanteda package

<https://quanteda.io/>

