

# Identify Themes within a Text

*This recipe and exercise identify simple themes within a sample text*


## Ingredients

---

1. An electronic text to explore;
2. A List Words tool such as the [TAPoR List Words Tool](#);
3. A Concordance tool such as the [TAPoR Find Words - Concordance Tool](#);
4. A Collocation tool such as the [TAPoR Find Collocates Tool](#).

## Recipe Steps

---

1. **Take** an electronic text from a source such as [Project Gutenberg](#) (see discussion below);
2. **Prepare** text by removing any added infrastructure ;
3. **Generate** a word list (sorted by frequency) using the [TAPoR List Words Tool](#);
4. **Examine** list to see if anything unusual stands out;
5. **Refine** word list by applying a [stop list](#);
6. **Re-examine** list for particular words you expect or *don't* expect to see (see discussion below);
7. **Explore** keywords using [TAPoR Find Words - Concordance Tool](#) to discover their context;
8. **Identify** collocated words using [TAPoR Find Collocates Tool](#) to determine usage patterns.

## Discussion

---

### Finding a Text

Possible sources for electronic texts are listed in the [Electronic Texts Panel of TAPoR](#). When preparing text for analysis, you should be aware that academic infrastructure included in the text may obstruct reading the text for its original

construction. It may be useful to remove notes and other materials added by subsequent authors from the original work. You can use tools such [TAPoR Extract Text](#) to remove added material.

### Using a Word List

The word list can provide a first clue about the nature of the text. Questions which can be asked of the word list may include:

1. What are the basic preoccupations of this text?
2. What is unusual in the text?

### Exercise Steps

---

This tutorial uses Volume 2 of Thomas Macaulay's *History of England* which can be downloaded from [Project Gutenberg](#);

Run the [TAPoR List Words Tool](#) to generate a word list sorted by frequency.

Word		Count
The	-----	3591
Of	-----	2057
And	-----	1360
To	-----	1234
A	-----	850
Was	-----	848
In	-----	758
Had	-----	686
Been	-----	265
Be	-----	255
Not	-----	246
At	-----	240
On	-----	213
From	-----	212
Who	-----	201
They	-----	187
Their	-----	174
All	-----	153
King	-----	139

Examining the word list does not show an immediate pattern amongst the most common words in the text. The most frequent words were common function words such as 'The', 'A', etc. They don't appear to be particularly unique, so we decide to eliminate common function words;

Run the [TAPoR List Words Tool](#) again, *applying a list of words to exclude from the list.* One useful stop list, the [Glasgow stop words list](#);

Word		Count
King	-----	139
Great	-----	115
Parliament	-----	92
England	-----	86
House	-----	83
Men	-----	81
Time	-----	75
Government	-----	74
Charles	-----	73
Power	-----	68
Party	-----	66
Public	-----	59
Years	-----	57
France	-----	56
Long	-----	56
English	-----	55
Court	-----	54
Commons	-----	53
State	-----	52
Church	-----	51
New	-----	46
Man	-----	46
Country	-----	46

The list of frequent words is now more intriguing. Words such as : **King, Great, Parliament, England, House, Men, Time, Government, Charles, Power, Party, Public Years, Just** immediately stand out.

Let's see how these words are used within the context of the text.

Using the [TAPoR Find Words - Concordance Tool](#) will list places in the text that a particular word appears.

This search returns a list of these key words and the five words on either side of the target word. Several themes and paths for further exploration emerge from this process

the word *Time* occurs frequently. Does this suggest a focus on time passing or the importance of time to the story being related? Note the frequent use of the words 'long', 'years', 'old', 'good', 'passed', 'life', 'day', 'make', 'passed', 'did', 'soon' in relation to the word time.

Is there a theme of **Power** in this text? - Note the focus on titles, 'King', 'Parliament', 'Government', 'Charles', 'Power', 'State', 'Man', 'Nation', 'General', 'Crown', 'Duke', 'Royal', 'Head', 'Monarchy', 'Chief', 'High', 'Lord', 'Prince'.

How is power referred to in the text? Use the [TAPoR Find Collocates Tool](#) to explore the use of the word power.

**Power** is treated not a single entity, but is most often qualified: 'spiritual power', 'temporal power', 'coercive power', 'arbitrary power', 'uneasy power', 'power of the sword', 'political power'.

What is the author's attitude towards the common man? Use the [TAPoR Find Words - Concordance Tool](#) and input **Man** as the target.

Note that the word **Men** is generally disparaged: 'worst sort of men', 'worthless men', 'unhappy men', 'small men', 'men merited clemency', 'ambitious men', 'worst set of men of the world'.

What is the author's attitude towards the monarchy? Use [TAPoR Find Collocates Tool](#) to see how often the words 'King' and 'Charles' occur together.

Note that the title **the King** is commonly used, but not *King Charles* or *King James*. The proper name Charles is used frequently, but collocation of King and Charles are rare.

Parliament is superior. King is 'detested', 'disliked', 'impeached' - moreover, never used terms 'executed', killed.

The word **Court** is paired with a variety of disparaging terms, 'sycophants', 'concealed', 'quarreled with', '...abused', 'extravagance of the ...', 'the...excited the bitter indignation', 'vice and folly', 'disliked', 'faithlessness of the ...', 'seditious'

The **Commons** are collocated with positive terms: 'undoubted power', 'should be governed', 'elected', 'ancient and undoubted power', 'legally'.

Thus, using these simple tools easily identifies the themes of power, monarchy, the common man and time in Macaulay's *History of England*.

<statement of origin, caveats, etc>

V1.2 17 April 2007