**What makes a classic? Popularity of works of two great English novelists in relation to readability**

Nicole de Groot (11419156), Lenka Bajčetić (11429488) and Linda Boeckhout (11011246)

**Introduction**

Many classic English novelists are still widely read today. It is striking that certain works enjoy a wider contemporary readership than other works of the same author. The object of this research is to ascertain whether there is a correlation between readability of a certain work of a popular author and its contemporary popularity. As an indicative factor in determining popularity we take the ratings from the website Goodreads.com.

**Method**

**Choice of datasets**

For this research we have selected the works of Charles Dickens and Jane Austen. We have chosen these authors because their work is readily available on project Gutenberg, and rated on the website Goodreads. Both authors still enjoy a wide contemporary readership and their oeuvres are large enough to compare popular and less popular books.

**Selected works of Austen and Dickens**

Only fully published novels will be part of the dataset. This means that short stories and novellas will be excluded. In the case of Jane Austen, this means that the novels that were unfinished at the time of her death will be excluded (*Lady Susan* and *Sanditon*), as will the *Juvenilia*, the short stories she wrote before she rose to fame. In the case of Charles Dickens, we will exclude his short stories and novellas like *A Christmas Carol* as they are usually published in ever-varying collections. This restriction is made because we want to make an accurate comparison between the readability of a novel and the popularity, and the popularity of short stories, individual novellas and compilations is hard to ascertain. Picture books and abridged versions are also excluded: the popularity of certain editions (like picture books) are beyond the scope of this research and abridged versions would not give an accurate representation of the contemporary readability of the original work.

**Selection of popularity parameters**

The parameters for popularity will be taken from the website goodreads.com. We will take the following factors into account: 1. Overall rating (scale of 1-5) 2. Number of ratings 3. Number of editions 4. Number of reviews

It is important to realise that the popularity on goodreads.com does not indicate the historical and cultural relevancy of a certain work, nor does it indicate, for instance, whether a book is still widely read and studied in schools and universities. It does, however, indicate the attraction a certain work still holds for contemporary readers, which is relevant to our research.

**Retrieval of data**

The oeuvres of Charles Dickens and Jane Austen will be retrieved by using the project Gutenberg API. The popularity data is not readily available via the Goodreads API, but is listed with the entry of the book. We will gather this data manually and determine the popularity by the available parameters on the website.

**Processing the data**

We will use Cygwin/Unix to retrieve the novels from project Gutenberg. Once the data has been retrieved, we will use the koRpus package and Tree Tagger in Rstudio to assess the chosen texts with established readability formulas in the field. (Flesh-Kincaid). The popularity data from goodreads will be established by a formula that takes all four parameters into account.

**Discipline**

The answers of this research are relevant to the field of English literature studies. If readability turns out to be related to the popularity of certain works, it might explain why certain works of acclaimed authors enjoy a wider contemporary popularity than other works. If there is no correlation at all, alternative explanations have to be sought, for instance in the subject matter of the book.