Stephen Palmstrom April 12, 2013

# Introduction

After documents have been converted from legacy fonts to Unicode, it is useful to compare the legacy and Unicode documents to see how successful the conversion has been. This can be done by looking at the documents side-by-side, but this can become difficult, particularly as there may be subtle differences between the fonts that may change the pagination. It is therefore useful to be able to see the two documents in one place, line-by-line. This program does this.

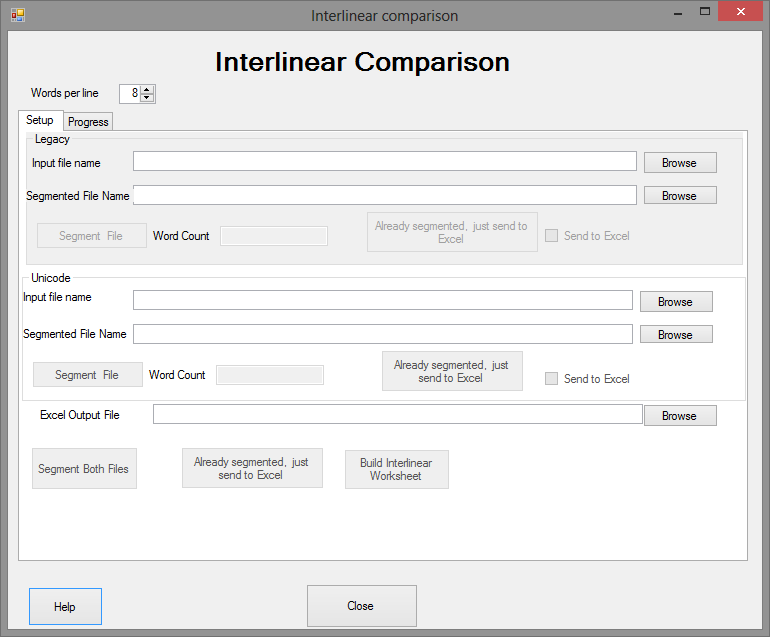
This document describes how to use Interlinear, a program that takes two Microsoft Word documents, generates segmented files (i.e. files with a fixed number of words per line) and copies them into an Excel spreadsheet that gives a line-by-line comparison of the two documents.

# Installation

Open the Interlinear folder in the zip file and execute setup.exe. Until the code can be digitally signed, Windows 8 may throw up security warnings that you will have to ignore.

# Execution

After installation, you will see this:



Click the Browse buttons to insert the file names for legacy, Unicode and the Excel spreadsheet. The buttons allowing you to segment the files and create the Excel output will be enabled as you add the file names.

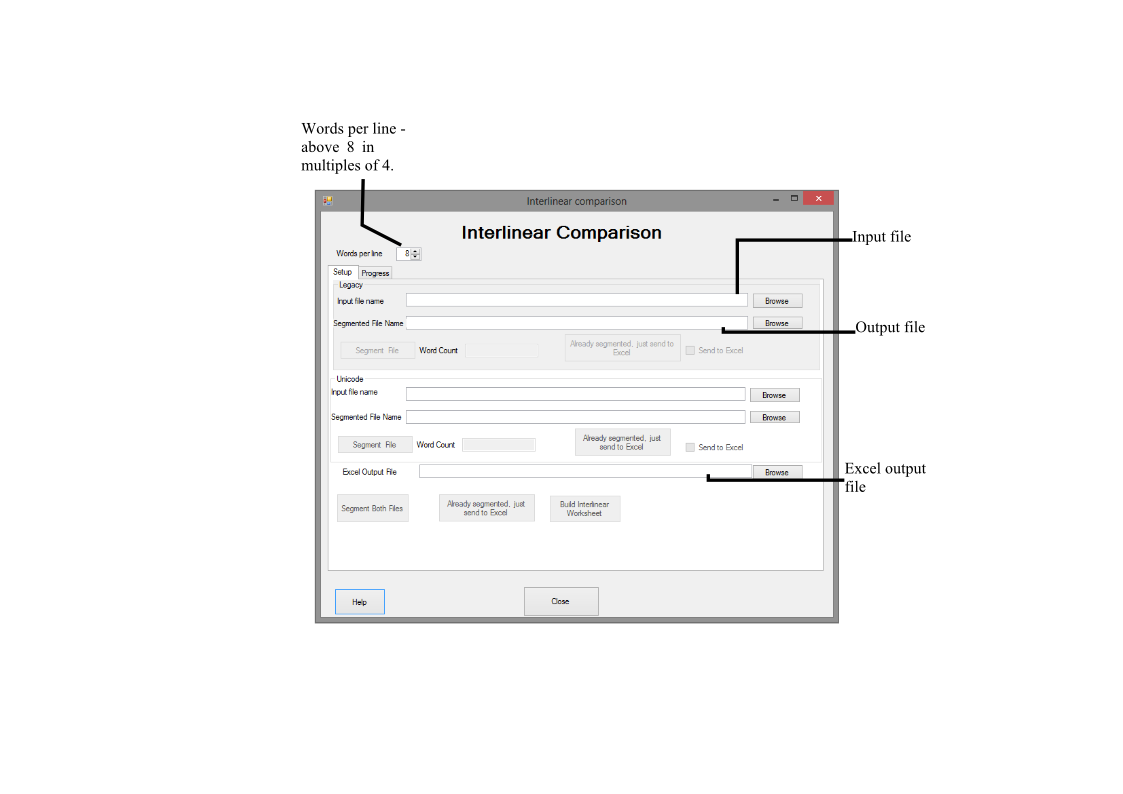


Figure 1 Explaining input and output files.

Once you have entered the file names you can click on the Segment buttons. If the Excel file name has been entered, the default is that the segmented data are sent to Excel. If you have all the file names completed, the Segment Both button will do segment and create the Excel file in one operation. The progress tab reports progress after you have started the process.

# The Excel File

The Excel file consistes of three worksheets.

1. Interlinear, where legacy and Unicode texts are in odd and even rows respectively. The fonts are those first encountered in each document.
2. Legacy – the legacy lines in odd rows
3. Unicode – the Unicode lines in even rows

Note that the last two are pasted from the segmented Word documents, and the first one has cells equated to the other two depending on whether the row is odd even. Note that this means that the Interlinear worksheet will not show changes to the font in mid-line, unlike the other worksheets which were populated by pasting from Word. This means that you should check the equivalent lines in Legacy and Unicode if there are apparent discrepancies in the Interlinear worksheet.