

Nunn Center Recommendations for Preparing Transcripts for OHMS

Draft, Last Updated 2/26/14

OHMS is not a word processing program. When you upload a transcript for synchronization and eventual viewing in OHMS, you must go through a few steps to ensure that all incompatible characters and formatting are addressed appropriately. Microsoft Word and other word processing applications often create invisible formatting that create proprietary dependencies that will cause anomalies in web-based and archival environments. For example Microsoft Word often will replace plain quotations with "smart" or curved quotations. Characters such as these must be converted to be more compatible with plain text, UTF-8 encoded text in a web environment. Many of the Nunn Center's procedures for formatting transcripts optimize presentation in OHMS. For additional guidelines on general transcript formatting, see the *Nunn Center Transcript Formatting Procedures*.

Many of the steps needed for OHMS compatibility can be automated using the Microsoft Word "macro" function. Essentially, in order to best utilize a transcript in OHMS, we recommend the following:

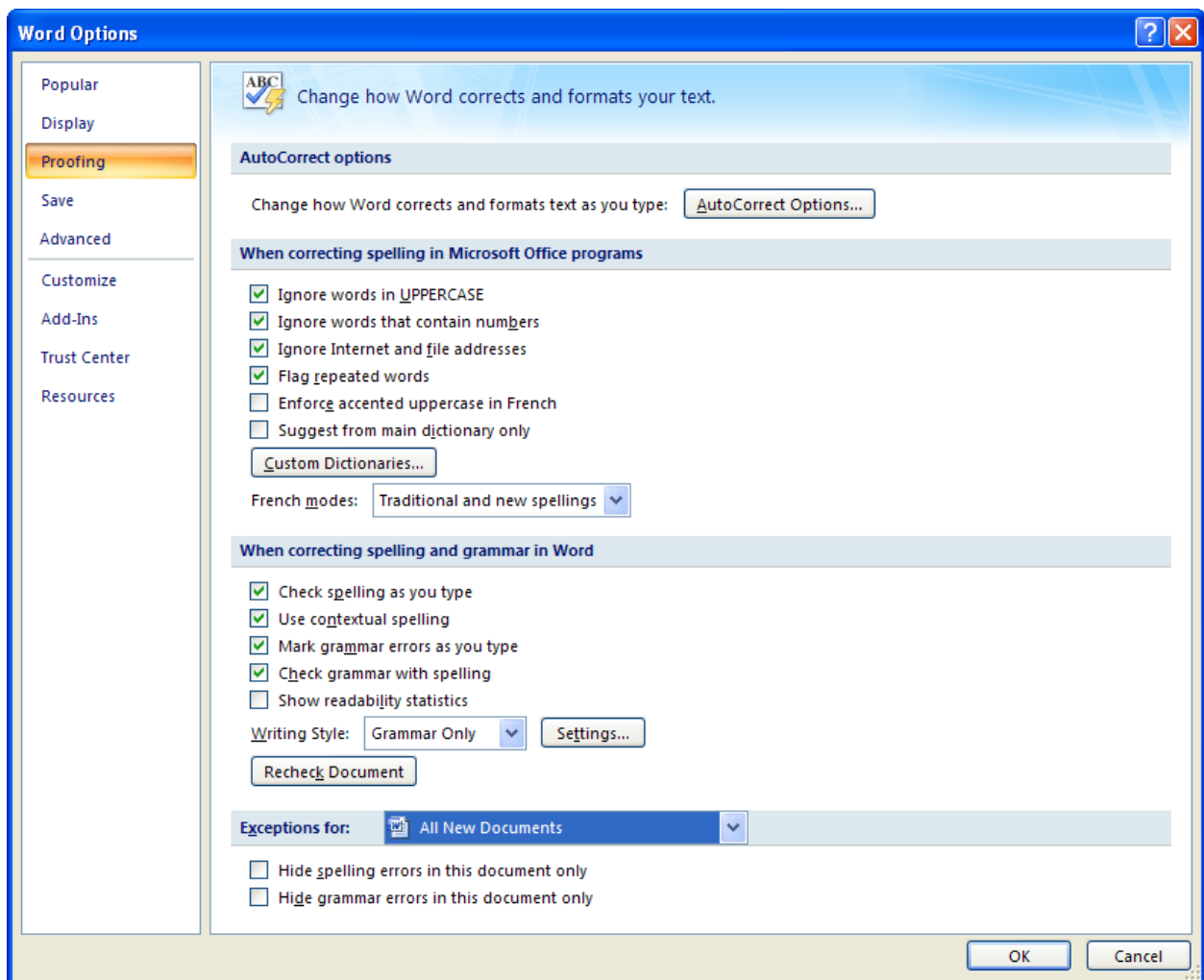
- Remove front matter and non spoken text.
- Remove Headers and Footers
- Clear formatting, section breaks
- Remove tabs, Remove blank space
- Make sure there is a colon after each Name Identifier in a transcript
- Replace soft returns with hard returns
- Every speaker must be on a separate line, requiring a hard return after each speaker.
- Change formatted text such as an em dash or en dash with double dashes
- Must be saved as a Unicode 6 (UTF-8) or Unicode 6.1 (UTF-8), plain text file. (.txt)

The following are the steps the Nunn Center goes through in order to prepare a transcript for OHMS. These recommendations are optimized for Microsoft Word 2007 (PC). These are recommendations based on our experience working with transcripts in OHMS. Sometimes Microsoft Word retains invisible formatting (such as encoded line breaks) that we have found

easier to remove in text editors such as Text Wrangler (Mac) or EditPad Pro (Windows). We will often perform finalization in one of these text editor programs before uploading to OHMS.

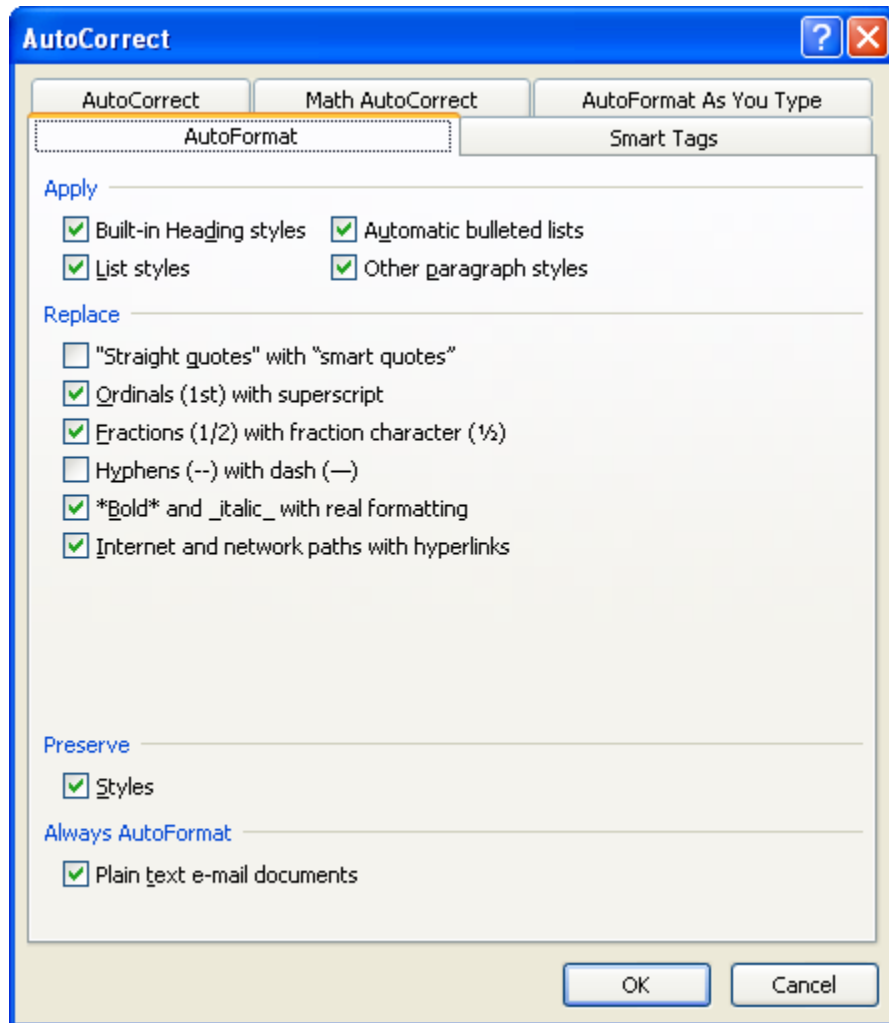
1. Change AutoCorrect Functions: apostrophes and parentheses from smart quotes to straight quotes

- Choose "Word Options"
- Choose "Proofing"
- Choose "Autocorrect Options"



- In "Autoformat As You Type" tab under "Replace" de-select "Straight quotes with smart quotes" and "Hyphens (--) with dash (--).

- In "Autoformat" tab under "Replace" de-select "Straight quotes with smart quotes" and "Hyphens (--) with dash (--).

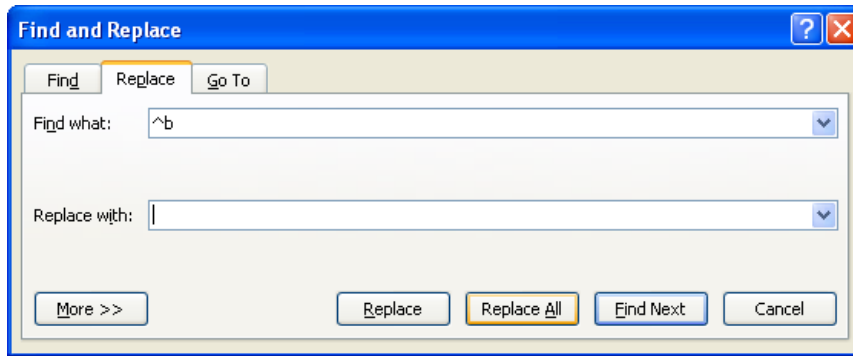


2. Delete any extra formatting (e.g., Letterhead, footer, page numbers)

3. Delete section breaks

In "Find and Replace" (Ctrl+f) or ("Home" menu, "Editing", "Replace")

- Find what: ^b
- Replace with: (leave blank)
- Click "Replace All"



4. Replace Soft Returns with Hard Returns

"Find and Replace" (Ctrl + f) or ("Home" menu, "Editing", "Replace")

- Find what: ^l
- Replace with: ^p
- Click "Replace All"

5. Run the Nunn Center OHMS transcript-preparation Macro (Microsoft Word)

The Macro developed by the Nunn Center automates the following steps:

- It replaces single paragraph mark with two paragraph marks.
- It replaces smart apostrophes with plain apostrophes.
- It replaces smart quotes with plain quotes.
- It replaces em dashes with double-dashes.
- It replaces en dashes with double-dashes.
- It replaces ellipses with double-dashes.
- It replaces tabs with a deletion.

In order to implement the Macro undergo the following steps:

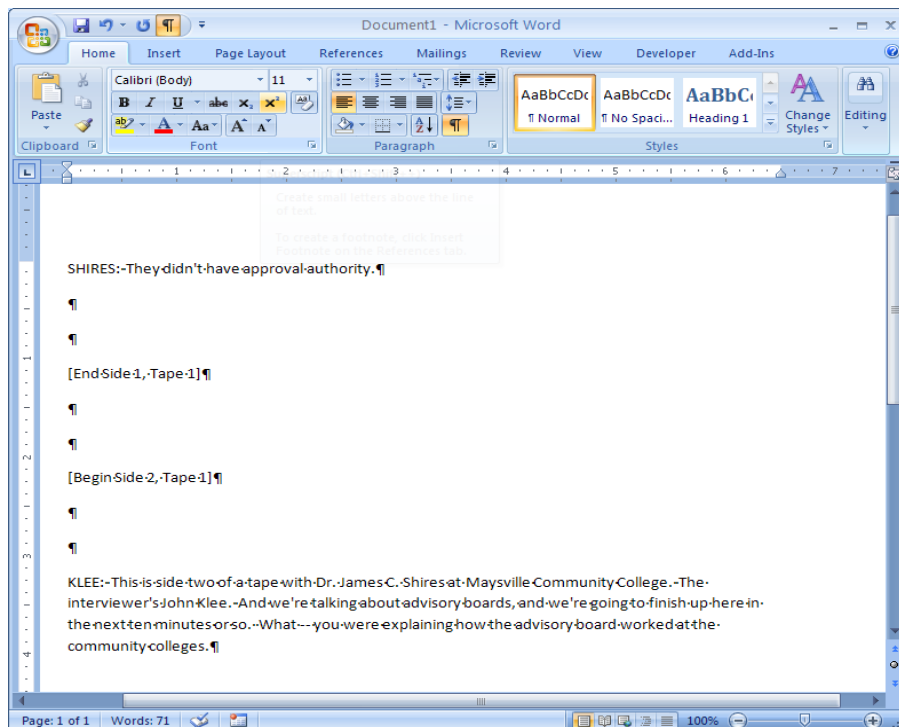
- Click the "Office" button (top left circle) and choose "Word Options"
- Click the "Popular" tab. Under "Top options for working with Word" select "Show Developer Tab in the Ribbon" and click "OK"
- In "Developer" tab, click on "Visual Basic"
- Right click the "Normal" heading, and choose "Import File".

- Locate and select desired macro and click "Open". (You may need to choose "All files" under the "Files of type" option depending on how your macro document was saved.)
- Close out of "Visual Basic" window and return to Word document by clicking x in top right hand corner
- In "Developer" tab, click on "Macros"
- Select desired macro (automatically named "Ohms_textprep_final_final") and click "Run"

6. Make sure there are three paragraph marks between every interview break, for example [End of tape 1].

In "Find and Replace" (Ctrl+f) or ("Home" menu, "Editing", "Replace")

- Find what: [
- Make sure there are three paragraph marks between every break



7. Save document as a plain text file (.txt)

Save the file as a Unicode 6 (UTF-8) or Unicode 6.1 (UTF-8), plain text file. (.txt)

- Choose "Save As"
- Choose "Other Formats"
- Type file name and choose save location

- In "Save as Type" scroll down to "Plain Text"
- Click "Save"
- Choose "Other encoding"
- Click to highlight "Unicode (UTF-8)" on a non-Mac or "Unicode 6.1 UTF-8" on a Mac
- Click "OK" to Save

8. Format Troubleshooting

If, after following these steps, you are still viewing anomalous line breaks in the transcript sync module or in the preview viewer, you may want to perform the following final steps on the .txt file and reupload:

- Open the .txt file in an alternative program like Text Wrangler (Mac) or EditPad Pro (Windows) for final processing.
- Some applications offer a function to "Remove Line Breaks" (Text Wrangler)
- Copy and Paste text into a new document and choose the "Copy Text Only" option
- Save as UTF 8 but when prompted for choices on line breaks, choose:

