

Plain Text Guidelines: Mixed Matter

Overview

The purpose of this document is to establish guidelines for the .txt file used in preparing an LDLT edition of a mixed matter text. These files should be plain text (i.e. .txt, not .rtf) and should be saved in UTF-8 encoding.

Structure

For purposes of the LDLT model, mixed matter texts are separated into sections, each of which can contain one or more “chunks”. A “chunk” is either a prose paragraph, a poem or quote from a poem, or a quote from a dramatic text. Chunks should be separated by one blank line. ### Numbering #### Section numbers Sections should be numbered sequentially at the beginning of the first paragraph of the section, with numerals only and no additional punctuation. #### Paragraphs and sentences Paragraphs do not need to be individually numbered – the script handles this automatically. Within each prose paragraph, sentences should be numbered sequentially with numbers in parentheses, e.g.: > (1) this is a sentence. (2) this is another sentence.

Poems

Poems do not need to be individually numbered. However, each poem, quote from a poem, or chunk of dramatic text should be marked “POEM” on the line directly above the poem itself. If the poem is the beginning of a new section, the section number should go before the word POEM, e.g.: > 4 POEM
> this is the first line of a poem

Line numbers

Lines should be broken up with line breaks, i.e. the newline character (\n). Our scripts can automatically number lines of verse and can handle transposed lines. If the poem starts at line 1, you do not need to mark this. If it does not start at line 1, you should include the line number at the end of the first line, separated from the text by one space and with no additional punctuation, e.g.: > this is the fourth line 4

For transposed or discontinuous line numbers, the script will number sequentially until it sees a numbered line, then jump to that number, then number sequentially until it sees another number. So, if the editor thinks that line 3 should be between line 1 and line 2, you should encode this: > "This is a line
> here's a third line 3
> here's the second line 2

> here's the fourth line 4
> and the fifth line

Speakers

The mixed matter script can handle speakers in poetry chunks only. Speakers should be encoded at the beginning of their speech in parentheses, like so: > (Speaker) these are some words

You do not need to include the speaker on every line, only on the first line of a new speaker, e.g.: > (Speaker1) this is a line
> here's another line by Speaker1
> (Speaker2) now Speaker2 is talking
> still Speaker2

Speakers can also change in the middle of lines, like this: > (Speaker1) here someone is talking (Speaker2) now someone else is

For information on how to handle uncertain speakers, see the CSV guidelines.

Greek quotations

Many texts contain quotations in Greek, and our scripts can handle them. In cases where Greek is used, it is especially important to make sure that the file is encoded in UTF-8. You should also check the output to ensure the characters remained encoded correctly during the transformation to XML. No special markup is needed.

Editorial Markup

Editorial additions

Editorial additions should be enclosed in angle brackets <>, e.g.: > “these are some <added> words.”

Lacunae

Lacunae (i.e. places where the editor believes there is text missing) should be marked with three asterisks, e.g.: > there is probably some *** here

For multiline lacunae, you should include three asterisks *** on each missing line.

Crux

Sections which the editor believes are corrupted but chooses not to emend should be enclosed in daggers † †, e.g.: > here is some †confusing text corrupted maybe†

Editorial deletions

Editorial deletions should be enclosed in curly braces {} e.g.: > these are some {deleted} words