Transcribe Bentham

Transcription Guidelines
Transcription Guidelines

Introduction
This page offers Guidelines for users of the Transcribe Bentham Transcription Desk. Here, you can find specific directions about how best to transcribe Bentham's writings, and how to encode specific features of the manuscripts.

Some of this may seem daunting, but do not be afraid to have a go at transcription - it is impossible to break anything, and any errors you might make can easily be reverted!

We recommend opening the below pages in new browser tabs or windows, so you can refer to the Guidelines while you are transcribing.

The Guidelines are divided into four sections:

- The Getting Started guide gives an overview of the transcription process, from start to finish (http://transcribe-bentham.ucl.ac.uk/td/Getting_Started).

- Basic Principles explain the basic principles of transcription and encoding (http://transcribe-bentham.ucl.ac.uk/td/Help:Transcription_Guidelines#Basic_Principles).

- Core Guidelines describe the manuscript features that users will encounter most frequently, and how to deal with them. Such features include additions, deletions, and notes (http://transcribe-bentham.ucl.ac.uk/td/Help:Transcription_Guidelines#Core_Guidelines).

- Supplementary Guidelines discuss the treatment of less-frequently occurring features of the manuscripts, such as ligatures, symbols, and foreign-language words (http://transcribe-bentham.ucl.ac.uk/td/Help:Transcription_Guidelines#Supplementary_Guidelines).

We are very grateful to our transcribers for their hard work! Check out our Credits page (http://blogs.ucl.ac.uk/transcribe-bentham/credits/) to find out how our volunteers are acknowledged for the work they do.

Need more information? Check out our Help pages (http://transcribe-bentham.ucl.ac.uk/td/Help:Contents), or email us (transcribe.bentham@ucl.ac.uk) — we're happy to help!
Basic Principles

Transcribing

Transcription refers to the text that the user reads from the Bentham manuscript and then copies into the Transcription Box.

When transcribing the text, your aim should be to produce a transcription which represents the text of the manuscript as accurately as possible (known as a diplomatic transcript).

Reproduce Bentham’s spellings, capitalisation and punctuation exactly as it appears on the manuscript, even if it seems incorrect to you. For example, Bentham and his scribes frequently get accents on foreign letters wrong or miss them out altogether. These mistakes should not be corrected. Do not expand any contracted words (Mr./mister) or write symbols as words (&/and).

Corrections and changes may be appropriate in a critical text, but the primary purpose of Transcribe Bentham is not to produce critical texts: it is to represent, in typographic form, the textual inscriptions of Bentham's manuscripts.

These transcriptions are intended as an important resource for future research and as a basis for work on the scholarly edition of *The Collected Works of Jeremy Bentham* ([http://www.ucl.ac.uk/bentham-project/publications/collected-works](http://www.ucl.ac.uk/bentham-project/publications/collected-works)), which is the definitive point of reference for all scholars interested in Bentham's ideas.

Do your best to transcribe as accurately as possible, but don’t worry too much if you cannot transcribe everything on a page. The Transcribe Bentham Editors may be able to help improve your transcript, as might other volunteers. Scholars who use any transcript available on Transcribe Bentham will also carefully verify the accuracy of the transcription before including it in their research.

Once you have transcribed a page, a final proof-read often makes it easier to spot any errors. Thinking about the sense of the words on the page can help too - although bear in mind that some of Bentham's papers do not make a lot of sense right away!

For help on deciphering Bentham's handwriting, have a look at our Palaeography Skills page ([http://transcribe-bentham.ucl.ac.uk/td/Help:Palaeography_Skills](http://transcribe-bentham.ucl.ac.uk/td/Help:Palaeography_Skills)).

If you are unsure about how to transcribe something, do what you think is best or send us an email and we'll try to help ([transcribe.bentham@ucl.ac.uk](mailto:transcribe.bentham@ucl.ac.uk)).
Encoding

We ask volunteers to encode their transcripts in Text Encoding Initiative (TEI) compliant XML. TEI is a de-facto standard for encoding electronic texts. Encoding can be done relatively simply by clicking the buttons on your transcription toolbar.

Encoding and Markup are terms which may be used interchangeably. They refer to tags that are included in the transcription in order to identify features of the manuscript (such as deletions or marginal notes) in a manner that allows them to be processed by a computer.

By encoding your transcripts, you are helping to create a richer resource: researchers and students interested in Bentham's writing process, his deletions and revisions, will be afforded the opportunity to pursue this, owing to your work. Encoded transcripts also allow for more powerful and refined searching: rather than search for every occurrence of 'panopticon', it will be possible to see where 'panopticon' occurs only when deleted, for example.

Tags are used to identify parts of the transcription, and usually come in pairs, known as "opening" and "closing" tags (i.e. "<" and ">"). If, for instance, a user wished to note that the word "utility" was deleted from a manuscript they were transcribing, it would be tagged thus: <del>utility</del>.

To apply a set of tags, highlight a word or passage and click the appropriate button in the toolbar. Opening and closing tags will appear around your highlighted text. A closing tag can be identified by a slash after the "<". Sometimes only single tags are needed and to apply these, simply click the relevant button in the toolbar.

To see how the markup has been applied to your transcript, click the 'Preview' tab in the transcription interface.

Encoding may seem challenging but we hope that you will soon get the hang of it! As you become more confident, you may prefer to type the markup on your keyboard rather than using the transcription toolbar.

For more information, take a look at our Encoding page. If you are interested in learning more about TEI, we recommend visiting the excellent TEI by Example website (http://teibyexample.org/), which contains a number of tutorials and exercises.
Core Guidelines

Headings
If the page you are transcribing includes a title or heading, you may identify this feature by highlighting the transcribed text of the heading and clicking the button on the toolbar.

This will surround the heading with tags. Bentham occasionally provides more than one heading: in this instance, simply apply separate tags to each heading.

Footers
Page numbers and other details which sometimes appear in footers on Bentham's manuscripts should be marked-up in the same manner as headings.

Paragraphs
Once a paragraph from a manuscript has been transcribed, it may be identified by highlighting the text of the paragraph with the cursor and clicking the button on the toolbar. This surrounds the text with tags.

Any text not included in heading or note tags should be enclosed by tags, even if a discrete paragraph is not physically represented on the manuscript image (i.e. even if the text is a continuation of a paragraph from the previous manuscript page).

If the word at the beginning of a paragraph has been indented, this does not need to be reproduced in the transcript. The first word of the paragraph can be typed right next to the opening tag.
Line Breaks

In order to preserve the lineation of the manuscripts, a line break should be inserted directly after the final word or punctuation mark of each line. In order to do this, click the button on the toolbar: this inserts an `<lb/>` tag.

It is important to note that the `<lb/>` tag does not have opening and closing tags, as it is a milestone element, which marks a place in a text and does not have any content.

Once you have added an `<lb/>` tag, you can press enter and begin the next line of the transcript on a new line in the Transcript Box. This should make it easier for you to follow the correspondence between the image and the text you are transcribing.

Some users have noted an error occurring when using line break tags. If the tag is written as `<lb>` rather than `<lb/>`, then all text following the incorrect `<lb>` tag will not be displayed by the Transcription Desk when the transcript is saved. To correct this problem, simply find the incorrect line break tag and add the '/' to it.

Line-end Hyphenation

When a hyphenated word appears at the end of a line, transcribe the word without the hyphen along with any punctuation that immediately follows it. Then insert the `<lb/>`, by clicking.

In the example below, the word ‘circumstance’ is hyphenated at the end of the first line. The transcription should read as follows:

```
customs, religion of the inhabitants, every circumstance<lb/>
in which a difference in the point<lb/>
```

By transcribing hyphenated words in this way, you are making it easier for them to be picked up in keyword searches.
Page Breaks

Like line breaks, a page break is indicated in markup with a milestone element: `<pb/>`. When transcribing a folio that contains a double page (JB/027/124/001, for example), a page break should be inserted to mark the point at which one page ends and another begins.

To do this, position the cursor at the relevant point in the transcription and click the button on the toolbar: this will insert a `<pb/>` tag.

A `<pb/>` tag does not need to be inserted at the end of a single page.

In JB/027/124/001, the page break would be recorded thus:

```html
<p>...we who are not of the Profession of the Law, cannot</p>
<p>positively assert</p>
<p><pb/></p>
<p><head>C</head></p>
<p><head>Prefat.</head></p>
<p>England has long been regarded</p>
```

Bentham occasionally quartered or divided large sheets into sections by drawing lines across the page. To identify when Bentham begins a new section, users should insert a page break by clicking `<pb/>`.

If a section has a footer, mark this up as a heading before inserting the page break.
Lines drawn across single sheets (e.g. JB/071/049/002) should not be considered as page breaks.

Additions

In its simplest form, the button in the toolbar is used to mark a part of the text that was added to the manuscript after the surrounding text was written. The exception to this is marginal additions, which are described below. This method may be used to mark additions, whether they are added above or (much less frequently) below the line. Highlight the addition and click the button to surround it with \(<\textit{add}>\)\(</\textit{add}>\) tags, as in the example below:

```
whatever <add>just</add> remark may
```

Deletions

Where a word or a sequence of words has been deleted in the manuscript, highlight the relevant text and click the button in the toolbar. This will surround the text with \(<\textit{del}>\)\(</\textit{del}>\) tags.

```
artificial: <del>tables of it's population:</del> tables of the
```

Just do what you think is best when deciding on the extent of deletions. Where the strikethrough does not physically cancel a punctuation mark that is apparently part of the deletion, you may assume that it forms part of the deletion. If in doubt about a particular example, you may send an email to the Transcribe Bentham editors (transcribe.bentham@ucl.ac.uk).
In some instances, entire pages or paragraphs are crossed out (e.g. JB/027/029/003), which indicate where Bentham or his scribes have used a particular passage when putting together a work. Text which is struck through in this manner should not be enclosed in deletion tags.

**Complex Additions and Deletions**

Transcribers will quickly become aware of instances of more complex intervention in the manuscripts, often where there is a combination of added and deleted text. One such example might be called 'substitution', where text added above the line is intended to replace text that is deleted with a strikethrough.

**Substitution**

The TEI provides guidelines about encoding such phenomena with the `<subst>` element, but for the purposes of this project, simply identifying text that is added and text that is deleted will suffice.

For the sake of consistency, transcribers are advised that when ordering substitutions like this, the deleted text should be transcribed first, followed by the added text, following the implicit order in which the respective parts originally appeared in the manuscript.

For example, once the relevant parts of text from the example above have been tagged, the transcription will look like this:

```
<del>[To bring]</del><add>I will reduce</add> the question at once
```

**Catchwords**

A catchword is the first word of the following page inserted at the right-hand lower corner of a manuscript folio, below the last line. They appear quite frequently in Bentham’s writings, and should be encoded in the same fashion as an addition, as in the example below:
...in the <add>act</add> can not <add>be</add></p>

Illegible Text

In the course of transcribing, you may encounter text that is illegible, either because Bentham's handwriting is difficult to read, or because it has been obscured by a strikethrough. There are slightly different ways to deal with each instance.

Undeleted

If a word or sequence of words on the manuscript is illegible, but has not been deleted, it may be identified by clicking the button in the toolbar. This inserts a <gap/> tag.

Note that <gap/>, like <lb/>, is a milestone element, and does not have any content.

Insert one <gap/> tag for each illegible word, if it is possible to distinguish the number of illegible words in a sequence.

Deleted

If the word or sequence of words is illegible because it has been deleted or struck through on the manuscript, you should use the <gap/> tag in conjunction with <del> tags to indicate the reason for illegibility.

But of that which remained, <del><gap/></del> as not

Questionable Reading

Where you have provided a transcription that you are not entirely certain about, this uncertainty may be registered by highlighting the word or sequence of words in question, and clicking the button on the toolbar.

This will surround the relevant text with <unclear></unclear> tags.
Ampersands

Bentham uses the ampersand sign (&) quite frequently in his manuscripts. When it occurs in a manuscript you are transcribing, you should click the button on the toolbar: this will add a piece of code (& amp;) which will render the ampersand correctly in the saved transcription.

The reason you cannot simply type a ‘&’ character on your keyboard is that in markup, ‘&’ is an escape character which invokes an alternative interpretation on subsequent characters in a character sequence.

Marginal Notes & Summaries

Bentham wrote in the margins of a manuscript for two main purposes: to add text to a portion of the manuscript that was already written, or to provide a summary of the text opposite.

In the first of these instances, Bentham often used a symbol in the main text of the manuscript to identify the point of attachment of the note: the symbol would then be reproduced at the text of the note in the margin. When this occurs, transcribe the text of the marginal note at the relevant point of attachment in the main text of the manuscript. Then, in order to identify it as a marginal note, highlight the text, and click the button.

This will surround the text with <note></note> tags.

You can include <lb/> tags inside the <note></note> tags to indicate if the note includes several lines of text.

When a symbol is not provided for the note at the point of attachment, you should encode the note at the point in the main text at which you think it is relevant.

Marginal notes

a former chapter be true <del><add>just</add></del>, that <note>even in a civilised<lb/> life</note> the whole<lb/>
complement of punishment that is judged
The <note> tags will generally be nested within <p> tags. In rare circumstances, a note will apply to a heading, and will then appear nested within <head> tags.

Marginal Summaries
Marginal summaries are intended to provide a brief summary of adjacent text in Bentham's manuscripts. They are usually written in pencil and can be difficult to transcribe. This marginalia does not need to be transcribed. But if you would like to, it should be transcribed and encoded in the same fashion as marginal notes.

Underlined Text
When a word has been underlined in the manuscript, you may identify it by highlighting the relevant text and clicking the button on the toolbar: this will surround the text with tags containing an attribute: <hi rend="underline"></hi>

You may occasionally encounter pieces of text that have double or multiple underlinings. You may simply tag these in the same fashion as single-underlined text.

Superscript
Text in superscript is distinct from additions, where text has been added to the manuscript after the surrounding text has been written. A common example of superscript is seen in ordinal numbers, where the letters often appear above the line (3rd).

To encode an instance of superscript, highlight the relevant text and hit the button on the transcription toolbar. This will surround the text with this piece of code: <hi rend='superscript'></hi>, as in the examples below:
Unusual Spellings and Abbreviations

There are occasional instances in the manuscripts where Bentham employs an unusual spelling for a familiar word: these may include previously-acceptable spellings which are no longer in use, or idiosyncratic misspellings. Bentham also uses abbreviations or archaic contractions in words such as employ'd or suppos'd.

The Latin word ‘sic’ indicates that the word has been transcribed exactly as it appears in the manuscript. Where unusual words or abbreviations occur, they may be encoded by highlighting the relevant word and clicking the button on the toolbar. This will result in tags surrounding the word.

If you encounter a word that appears to have an unfamiliar spelling, you may refer to this list of unusual spellings (http://transcribe-bentham.ucl.ac.uk/td/Alternative_spellings) to see whether it is one that Bentham used frequently. You may also add words to this list to benefit other transcribers.

The tags should not be used for familiar contractions like it's, don't, they're, and so on.

<sic>compleat</sic> code of laws
Supplementary Guidelines

User Comments

In the event that you encounter something in the course of your transcription that is not covered by these Guidelines, you should insert a comment in the transcription to alert Transcribe Bentham editors and other transcribers to the issue. In order to do this, you should click the button on the toolbar. This will generate these characters: <!-- -->. You should type your comment between the dashes. The text of your comment will not appear in the saved transcription but will remain present in the Transcription Box.

<!-- There is an unusual feature at this point in the manuscript -->

If you have questions about an unusual feature in a manuscript, send an email to the Transcribe Bentham Editors (transcribe.bentham@ucl.ac.uk) with the name of the manuscript and information about the nature of your discovery.

Foreign Language

While transcribing Bentham's manuscripts, you may encounter languages other than English: this may occur in isolated words, brief passages, or longer sections of writing. You may encode such instances by highlighting the relevant non-English text, and clicking the button on the toolbar. This will surround the text with <foreign></foreign> tags.

Where non-English words include diacritics such as accents (é) or circumflexes (ô), these should be transcribed. You can produce such characters in Microsoft Word (or a similar programme) using keyboard shortcuts or the 'Symbols' menu. You can then copy and paste the character into the Transcription Box. Alternatively, you could copy and paste the character from another website.

<d'une fantaisie contrariée>

<!--d'une fantaisie contrariée-->
Dashes

Occasionally, you will encounter dashes of varying lengths in the manuscripts. In general, these may correspond to the en-dash and the em-dash. In printing houses, pieces of type that held the letters ‘n’ and ‘m’ were used as units for measuring and estimating the amount of printed matter in a line or page. Thus, a dash the width of a letter ‘n’ became known as an en-dash, while a longer dash, the width of a letter ‘m’, was called an em-dash.

Use your discretion to judge whether a Bentham dash is best represented by an en- or em-dash. There is no need to worry too much about which is which - as long as some form of dash is included in the transcript.

For an en-dash, you may simply type a hyphen (−) into the transcription box; if inserting an em-dash, you should click the button on the toolbar. This will insert a Unicode character code (& #x2014;) which will enable the representation of the em-dash in web browsers.

Pencil markings

Most manuscript pages were imprinted with a University College London stamp in the process of being catalogued. Two numbers are usually written in pencil inside this stamp to indicate the box and folio number of that particular page. Both the stamp and the pencil numbers should not be transcribed.

Any other pencil markings which appear on a page do not need to be transcribed. These pencil scribblings include marginal summaries, headings and corrections.

But if you can read and would like to transcribe text written in pencil, you are free to do so. Please add a User Comment before any text written in pencil: !-- text written in pencil -->

Ligatures

A ligature is character where two or more letters are joined together (such as æ). Bentham occasionally uses ligatures in his writing. Should you encounter one, you should simply transcribe the individual letters of the ligature (‘æ’ rather than ‘æ’), and insert a User Comment containing the word 'ligature' directly afterwards.

oeconomy!-- ligature -->
Symbols

Bentham uses a number of symbols (e.g. section sign: §), with varying regularity. If it possible to reproduce a symbol from the keys on your keyboard or by copying and pasting from another website, you should do so. Otherwise, you should simply register the presence of a symbol with a User Comment: <!-- symbol -->

Tables

Bentham sometimes presented information in tabular form, in rows and columns.

It is difficult to replicate the format of a table in a simple transcription, so you should concentrate on making sure that the text from the table is reproduced accurately in your transcription.

Depending on the shape of the table, it may make more sense to transcribe the text row-by-row or column-by-column. You should note the presence of tabular text in the transcript by inserting a User Comment before the table: <!-- the following text appears in a table -->.

Printed text

The Bentham collection contains a significant number of printed texts including Parliamentary Bills and contemporary pamphlets. Sometimes pages contains a mixture of printed and handwritten text. Printed texts can be transcribed according to the same guidelines as handwritten manuscripts. You can include a User Comment to note when printed text appears on a page.
Italics

If you are transcribing a printed text, italicised passages can appear. You can make a note of this by inserting a User comment before the italicised text: <!-- next paragraph appears in italics -->.

Brackets

In your transcription, you may represent the various types of brackets used in the manuscripts, including parentheses ( ), square brackets [ ], or braces { }. Take care not to use angle brackets < >, as these are used only for markup elements.

This document was first written in May 2010 and was last updated in February 2018. The Guidelines have evolved slightly over time in line with editorial discussions about transcription and encoding.

If anything in these Guidelines is unclear, please send an email to the Transcribe Bentham Editors (transcribe.bentham@ucl.ac.uk).