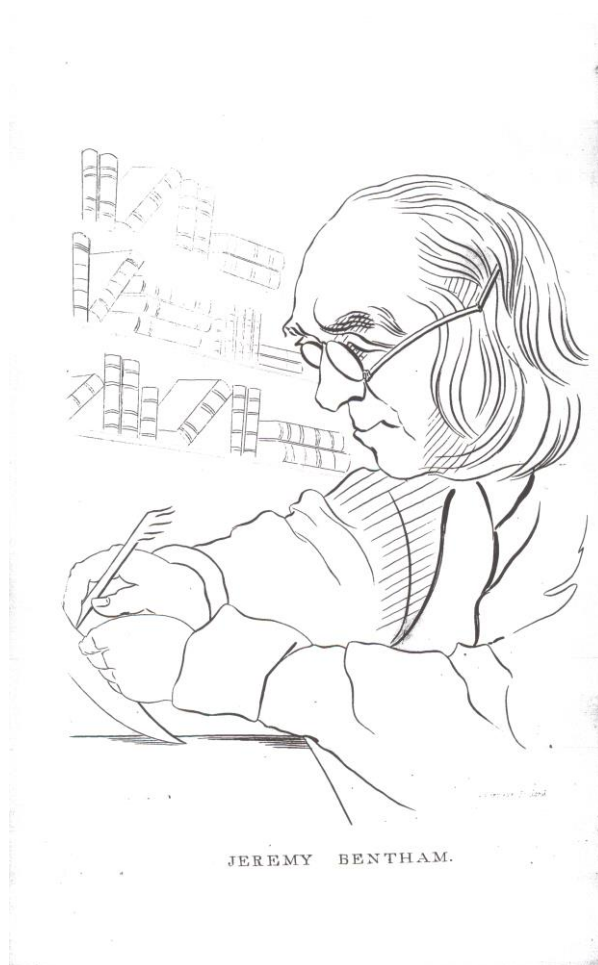




Transcribe Bentham Transcription Guidelines



Transcription Guidelines

Introduction

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

Some of the information below may seem daunting at first, but do not be afraid to have a go—it is impossible to break anything, and any errors you might make can easily be reverted.

The Guidelines are divided into four sections:

- The [Getting Started](#) guide gives an overview of the transcription process, from start to finish.
- The [Basic Principles](#) guide explains the essentials of transcription and encoding.
- The [Core Guidelines](#) describe the manuscript features that users will encounter most frequently, and how to deal with them.
- The [Supplementary Guidelines](#) discuss the treatment of less frequently occurring features of the manuscripts, such as ligatures, symbols, and foreign language words.

We are very grateful to our transcribers for their hard work. Visit our [Credits](#) page to find out how our volunteers are acknowledged for the work that they do.

Need more information? Check out our [Help](#) pages, or email us at transcribe.bentham@ucl.ac.uk—we are always 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 text, your aim should be to produce a transcription which represents the text of the manuscript as accurately as possible.

Reproduce Bentham's spellings, capitalisation, and punctuation exactly as they appear on the manuscript, even if they seem incorrect. For instance, Bentham and his scribes frequently got accents on foreign letters incorrect or omitted them altogether. These mistakes should not be corrected, nor should any contracted words be expanded (e.g. 'M' to 'Mister') or any symbols be rendered as words (e.g. '&' to 'and').

Do your best to transcribe as accurately as possible, but do not worry too much if you cannot transcribe everything on a page. The TB Editor may be able to help improve your transcript, as may other volunteers.

Once you have transcribed a page, a final proofread often makes it easier to spot any errors. Thinking about the sense of the words on the page can help too—although please bear in mind that some of Bentham's papers do not make a lot of sense right away, especially when taken out of context.

For help in deciphering Bentham's handwriting, please take a look at our [Palaeography Skills](#) page.

If you are unsure about how to transcribe something, do what you think is correct or send us an email at transcribe.bentham@ucl.ac.uk and we will try our best to help you as soon as possible.

Encoding

We ask volunteers to encode their transcripts in Text Encoding Initiative (TEI)-compliant XML format, an industry standard method for encoding electronic texts. Encoding can be done simply by clicking the buttons on your transcription toolbar.

By encoding your transcripts, you are helping to create a richer learning resource—owing to your efforts, researchers who are interested in Bentham's writing process, deletions and revisions, will gain valuable insight they might not otherwise have had. Encoded transcripts also allow for more powerful and refined searching.

Tags are used to identify parts of the transcription and usually come in pairs, known as ‘opening’ and ‘closing’ tags. If, for instance, a user wished to note that the word ‘utility’ was deleted from a manuscript they were transcribing, it would be tagged as `utility`.

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 forward slash after the ‘<’. Sometimes only single tags are needed (such as for the ‘`<gap/>`’ and ‘`<lb/>`’ tags)—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 at first but 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, please take a look at our [Encoding](#) page. If you are interested in learning more about TEI, we recommend visiting the [TEI by Example](#) website, which contains a number of tutorials and exercises.

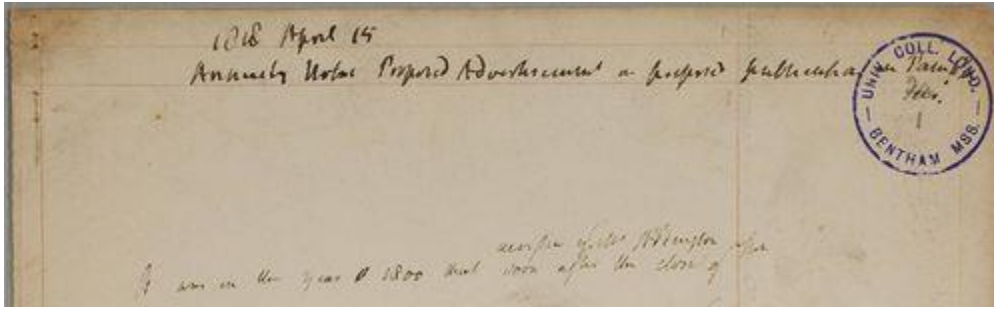
Core Guidelines

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 `<head></head>` tags. Bentham occasionally provides headings that span multiple lines—in this instance, simply add a `<lb/>` tag to mark the line break/s in the heading.

Using multiple heading tags on each manuscript may create problems for the TB Editor when converting the text to XML and may slow down the checking process, so should be avoided wherever possible.



<head>1818 April 15</lb/> Annuity Notes Proposed Advertisement on proposed publication on painless fees</head>

Footers

Page numbers and other details which sometimes appear in footers on Bentham's manuscripts should be marked-up in the same manner as notes (i.e. using <note></note> tags).


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 [[File:Jb-button-paragraph.png]] button on the toolbar. This surrounds the text with <p></p> tags.

All text that is not included in heading or note tags should be enclosed by <p></p> tags.

If the word at the beginning of a paragraph has been indented, the indentation does not need to be reproduced in the transcript. The first word of the paragraph can be typed right next to the opening <p> 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 what is called a 'milestone element', which marks a place in a text and does not have any content.

Once you have added an <lb/> tag, please press return and begin the next line of the transcript on a new line in the Transcription Box. This will make it easier for you to follow the text you are transcribing, and for the TB Editor to check your work quickly.

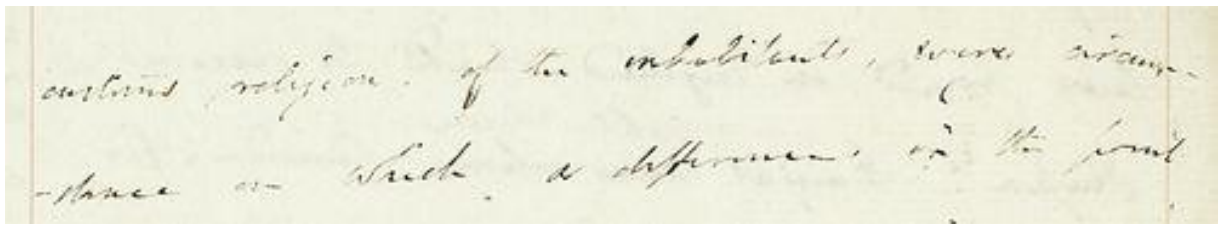
If the tag is written as `<lb>` rather than `<lb/>`, then all text following the incorrect `<lb>` tag will not be displayed by the Transcription Desk when the transcript is saved. To correct this problem, simply find the incorrect line break tag and add the `'/'` to it.

Line break tags are not required at the end of a paragraph or at the end of a manuscript—simple `</p>` or `</note>` tags are fine.

Line-end hyphenation

When a hyphenated word appears at the end of a line, transcribe the full word without the hyphen, along with any punctuation that immediately follows it, and then insert the `<lb/>` tag 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

in which a difference in the point

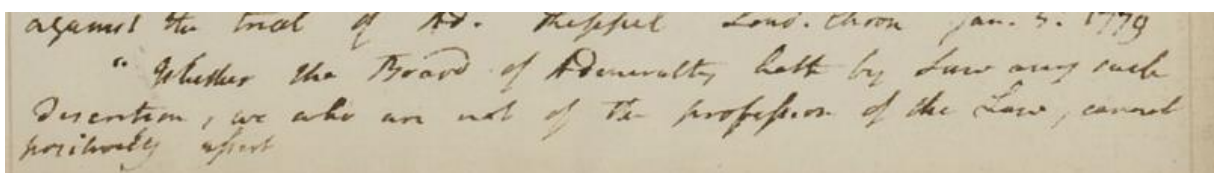
By transcribing hyphenated words in this way, you are making it easier for them to be picked up in keyword searches.

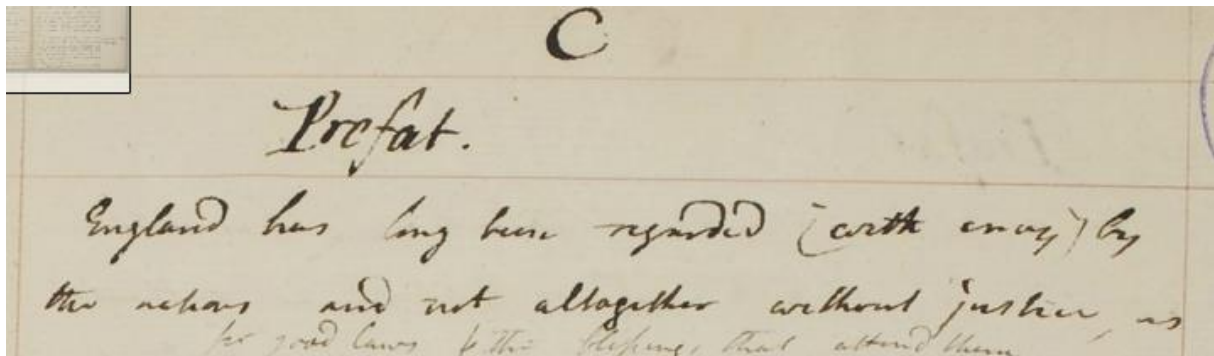
Page breaks

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.

In order 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.

In [JB/027/124/001](#), the page break would be recorded as follows:





<p>...we who are not of the Profession of the Law, cannot

positively assert</p>

<pb/>

<head>C

Prefat.</head>

<p>England has long been regarded [with envy] by</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**.

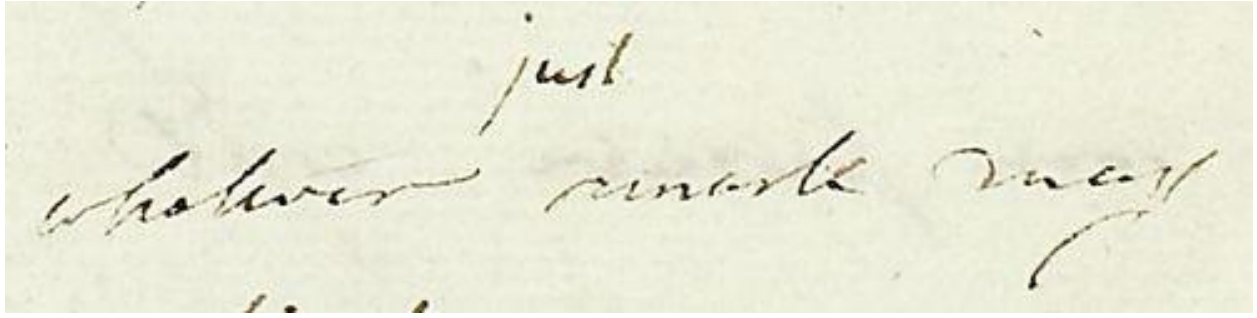
If a section has a footer, mark this up as a heading before inserting the page break.

Lines drawn across single sheets (e.g. JB/071/049/002) should not be considered as page breaks.

Additions

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. This method may be used to mark additions, whether they are added above or (very rarely) below the line.

Highlight the addition and click the **[+]** button to surround it with <add></add> tags, as in the example below:




whatever <add>just</add> remark may

Please avoid using additions within additions. Where Bentham has added text above a particular line, and then added further text above that, the best practice is to distinguish the extra addition using forward slashes.

For instances where Bentham has used superscript letters (i.e. in '1st' or '2^d'), please use the superscript tags, as described below.

Deletions

Where a word or a sequence of words has been struck-through in the manuscript, highlight the relevant text and click the  button in the toolbar. This will surround the text with tags.



artificial: tables of its population: 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, please email the Transcribe Bentham editors at transcribe.bentham@ucl.ac.uk.

In some instances, entire pages or paragraphs are crossed out (e.g. [JB/027/029/003](#)). For simplicity, text which is struck through in this manner does not need to be enclosed in deletion tags.

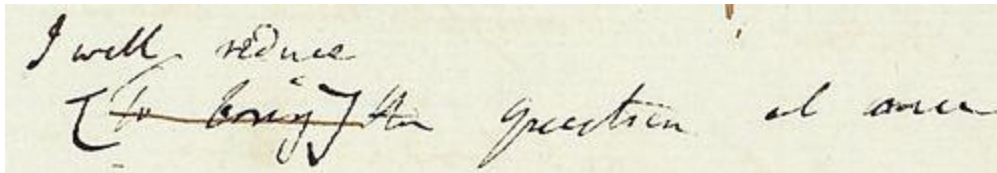
Just as it is best to avoid additions within additions, please also avoid using deletions within deletions—both of these practices prevent the transcription preview from displaying correctly and cause issues when we save your transcripts as .xml files.

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 is called 'substitution', where text added above the line is intended to replace text that is deleted with a strikethrough.

In marking substitutions, simply identifying text that has been added and text that has been deleted in its proper sequence will suffice.



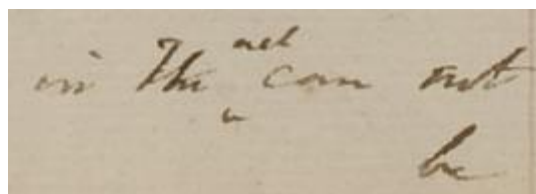
Transcribers are advised that when ordering substitutions like this, the deleted text should be transcribed first, followed by the added text, following the implicit order in which the respective parts originally appeared in the manuscript.

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

`[To bring] <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:




`<p>in the act can not</p>`

`<add>be</add></p>`

Illegible text

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

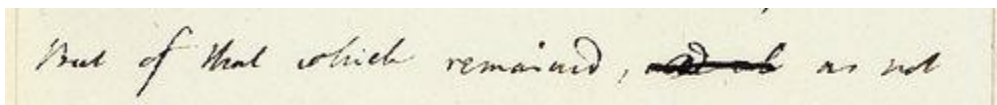
If a word or sequence of words on the manuscript is illegible, but has not been deleted, it may be identified by clicking the  button in the toolbar. This inserts a `<gap/>` tag.

Note that <gap/>, like <lb/>, is a milestone element, and does not have any content—as such it does not have opening and closing segments.

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


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 tags to indicate the reason for illegibility.

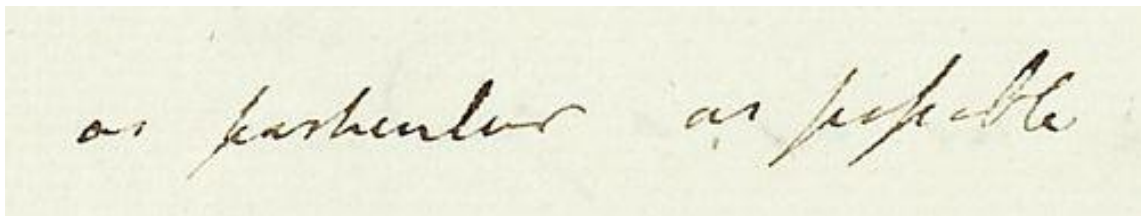


But of that which remained, <gap/> as not

Questionable readings


Where you have provided a transcription that you are not entirely certain about, this uncertainty may be noted 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.



as <unclear>particular</unclear> as <unclear>possible</unclear>


Ampersands

Bentham uses the ampersand sign (i.e. '&') quite frequently in his manuscripts. When it occurs in a manuscript that you are transcribing, click the  button on the toolbar—this will add a piece of code ('&#amp;#38;') which will render the ampersand correctly in the saved transcription.

The reason you cannot simply type a '&' character on your keyboard is that an ampersand is what is called an 'escape character' in TEI encoding, which applies an alternative interpretation to any subsequent characters in a given sequence. Escape characters are not used within Transcribe Bentham.

Marginal Notes and Summaries

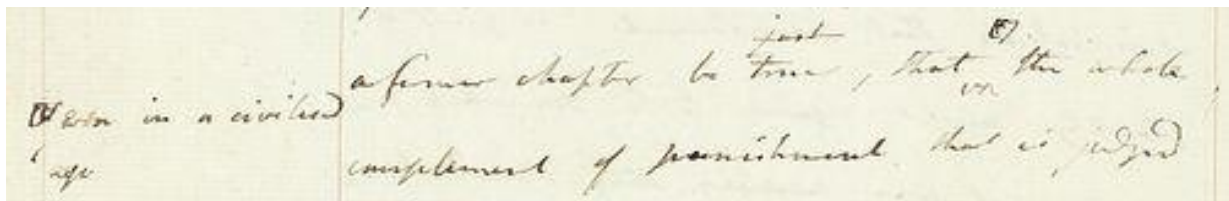
Bentham often wrote in the margins of a manuscript for two main purposes: either to add text to a portion of the manuscript that was already written, or to provide a summary of the text adjacent to it for the purpose of structuring his work.

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. If in doubt, it is best to place the `<note>` at the end of the paragraph which it appears alongside.




a former chapter be true ~~<add>just</add>~~, that `<note>`even
in a civilised`<lb/>` life`</note>` the whole`<lb/>`
`<note>`complement of punishment that is judged

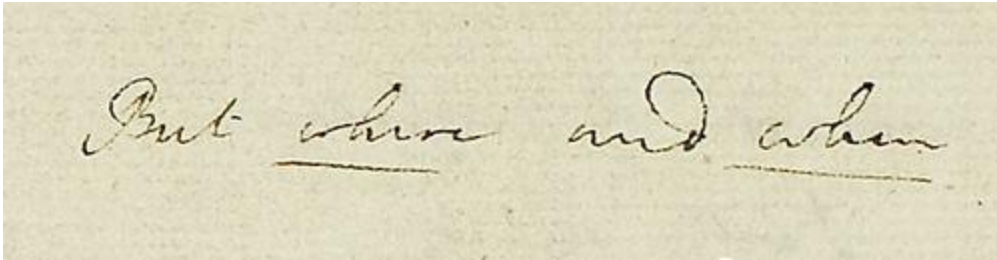
The `<note>` tags will generally be nested within `<p>` tags. In rare circumstances, a note will apply to a heading, and will then appear nested within `<head>` tags.

Marginal summaries 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. 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 and placed before the paragraph to which it corresponds.

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 the following tags: `<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.




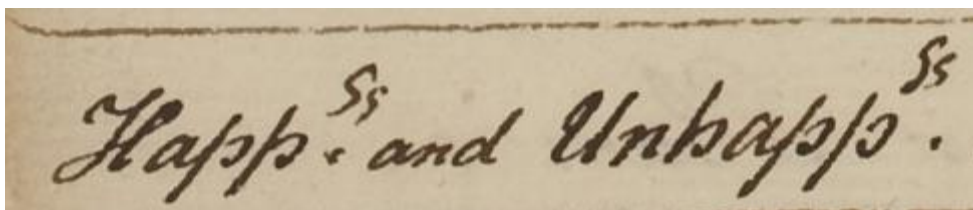
But `<hi rend="underline">where</hi>` and `<hi rend="underline">when</hi>`

Superscript

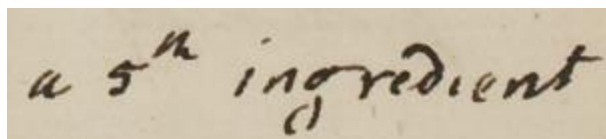
Text in superscript is distinct from additions, where, as described above, alternative text has been added to the manuscript after the surrounding text has been written.

A common example of superscript is seen in ordinal numbers, where the letters often appear above the line (e.g. '3rd').

To encode an instance of superscript, highlight the relevant text and click the  button on the transcription toolbar. This will surround the text with a piece of code (`<hi rend='superscript'></hi>`), as in the following examples:




Happ.`<hi rend='superscript'>ss</hi>` and Unhapp.`<hi rend='superscript'>ss</hi>`



a `<hi rend='superscript'>th</hi>` ingredient

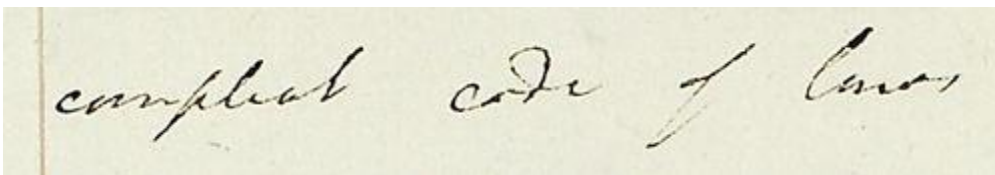
Unusual Spellings and Abbreviations

There are 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.

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 `<sic></sic>` tags being added around the word or words.

If you encounter a word that appears to have an unfamiliar spelling, you may refer to this [list of unusual spellings](#) to see whether it is one that Bentham used frequently.


The `<sic></sic>` tags should not be used for familiar contractions like ‘it’s’, ‘don’t’, ‘they’re’, and so on. Bentham also uses abbreviations or contractions in words such as ‘employ’d’ or ‘suppos’d’—these should generally be tagged as unusual spellings, but abbreviations such as ‘Ch.’ for ‘Chapter’, for instance, do not need to be tagged as such.



`<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.

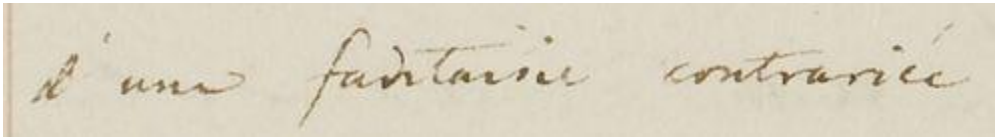
`<!-- There is an unusual feature at this point in the manuscript -->`

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 wherever possible. 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.




<foreign>d'une fantaisie contrariée</foreign>

Dashes

You will often encounter dashes of varying lengths in the manuscripts—either hyphens (-), en-dashes (–), or em-dashes (—).

Use your discretion to determine whether a Bentham dash is best represented by a hyphen, an en-dash or an 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 a hyphen or en-dash, you may simply type a hyphen (-) into the transcription box. For an em-dash, you should click the  button on the toolbar. This will insert a Unicode character code ('—') which will enable the representation of the em-dash in your web browser.

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—which may include marginal summaries, headings, and corrections—do not need to be transcribed. However, if you can read and

would like to transcribe any text written in pencil, you are free to do so. Please add a User Comment before any text written in pencil, like so: '<!-- text written in pencil -->'.

Ligatures or Diphthongs

A ligature or diphthong is a 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 ('oe' rather than 'œ'), and insert a User Comment containing the word 'ligature' directly afterwards.

oeconomy<!-- ligature -->

Wherever an 'æ' diphthong appears, this may be inserted using the Special Characters drop-down menu on the toolbar.

Symbols

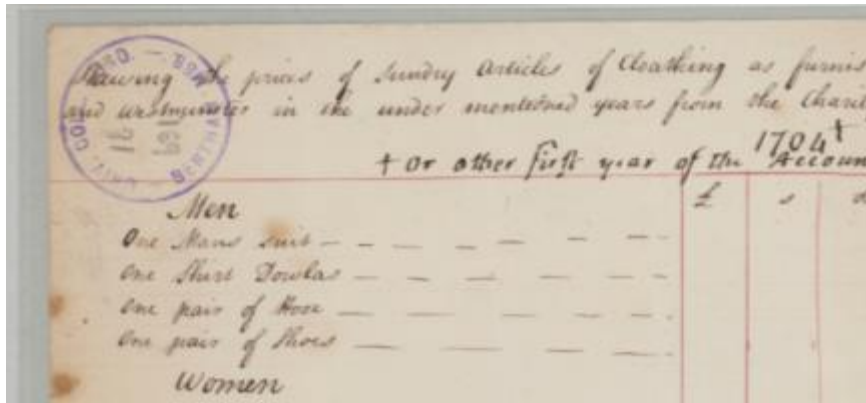
Bentham used a number of symbols in his writings, including section symbols (§) and manicules (☞). These may be added using the Special Characters drop-down menu on the toolbar. If it is possible to reproduce others symbol from the keys on your keyboard or by copying and pasting from another website, 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 Transcribe Bentham, 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, for which you can include a User Comment to note when printed text appears on a page.

Printed text may be transcribed according to the same guidelines as handwritten manuscripts. Transcribing printed text is a good way to get used to TEI encoding, but please be aware that, unlike Bentham's own text sheets, printed texts within the Bentham Papers will not be included in *The Collected Works of Jeremy Bentham*.

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 January 2020. 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 TB Editor at transcribe.bentham@ucl.ac.uk.