Important Features

Metadata:

* Date
* Publication
* Publication type (weekly, daily, periodical)
* Publication audience
* Editor
* Author (if possible)
  + Personography
* Publication address?
* Source/Citation

Body:

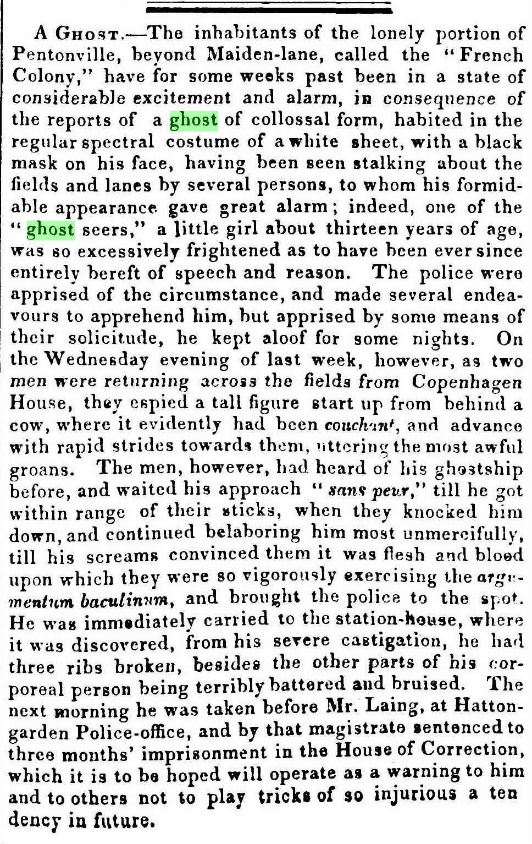
* Paragraph organization
* Place references
* People references
  + Direct references to personography
  + Indirect references to unidentified people
* Political/class references
  + This is a little bit difficult to parse?
  + What is a class reference? Reference to a person’s economic standing, so servant/master?
* Ghost references
* Framing devices

Deciding what goes into the collection:

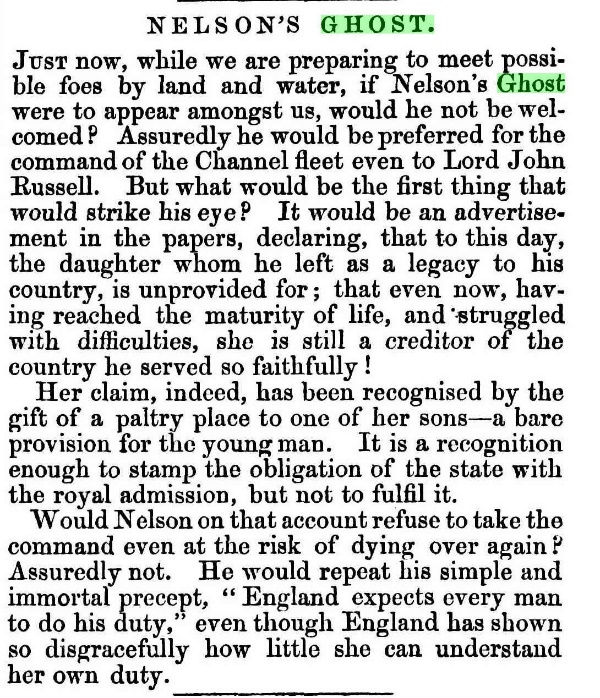
* “non-fiction”: news reports in a paper of record, or another newspaper that mostly publishes news stories and editorials
  + Report of hoax
  + Early report of haunting that turns out to be a hoax later
  + Report of reaction to hoax
* “ambiguous non-fiction”: possibly fiction, possibly non-fiction, but framed as non-fiction. Usually repeating a story told to the author, who expresses doubt as to the authenticity of the haunting
  + Alternatively, a true account which USES some ghost story as a tool to further some other idea, likely political
* “fictional reporting of ambiguity”: a most likely fictional account of a hoax or possible hoax, still framed as non-fiction reported to the author, the author expresses some doubt about the reliability of the ghost story

Examples:

Non-fiction



Ambiguous non-fiction



Fictional reporting of ambiguity

[C:\Users\gabikeane\Desktop\vict\_hoax\clippings\ghostcutghost\_chambers\_1889.pdf](file:///C:\Users\gabikeane\Desktop\vict_hoax\clippings\ghostcutghost_chambers_1889.pdf)

Something I don’t want (but will probably collect and then exclude anyway)

[C:\Users\gabikeane\Desktop\vict\_hoax\clippings\ghostofastory\_bentleys\_1847.pdf](file:///C:\Users\gabikeane\Desktop\vict_hoax\clippings\ghostofastory_bentleys_1847.pdf)

# Copyright

<https://www.copyrightaid.co.uk/forum/viewtopic.php?t=1431>

<https://copyright.cornell.edu/publicdomain>

According to Copyright Aid, I am within my rights to make full transcriptions of texts (including electronically, with my own edits) that are out of copyright. Gale and other resources own the rights to the scans and photographs (much as they own the physical rights to the newspapers themselves) and own the images. For this reason, I cannot republish the images (oops) but I can republish the texts I derive from those images (hooray). Of course, I will need to cite where those texts came from (just like when I visit a library and photograph things there), but I don’t need to worry about them asking me to remove my texts, as they cannot legally own the copyright to something that old that they didn’t purchase and renew rights to.

# Editorial Notes

Using ABBYY Finereader (FIGURE OUT WHICH VERSION) to create transcriptions of texts. There are a few issues with ABBYY that I need to either address or work around

1. Two column items sometimes are read the way a human would read them: left column and then right column. Other times, **columns are read from left to right as a human would read a one-column text**. This can happen in the same text, sort of an every-other-page deal. Fixing it consumes time, so this problem needs some attention.
2. Typographic long em dashes are longer than modern em dashes. For this reason, **long em dashes must be replaced with two modern em dashes**. This should be consistent across texts, so that I can easily find and replace them should I find a better typographical equivalent. For the purpose of initial editing work, I won’t do this.
3. Quotation marks, question marks, and other **punctuation should not be separated from letters with a space**. Typographically this was done to improve readability, but with computer text it is distracting
4. **Capital letters**, especially ALL CAPS, can often be misread. Using Notepad ++ or another plain text editor with spell check features is helpful for catching these errors
5. **Curly quotes** should be used. David Birnbaum is partially to blame for this, but also when I go tag quotations the Regex is a heck of a lot easier.