

The Johnson Letters, 1542-1552

Named Entity Annotation Guidelines

Version 1.3 May 2023

1. Introduction

These guidelines for named entity annotation are intended for use with the Johnson Letters in the context of the European Research Council-funded project 'Back to the Future: Future expectations and actions in late medieval and early modern Europe, c.1400-c.1830' at the University of Antwerp. The project aims to gain insight into "how people in the past thought about the future and how this affected their actions."¹

The Johnson Letters are comprised of the extant handwritten correspondence collected by the 16th century English merchant John Johnson, head of the family firm Johnson & Company. Barbara Winchester compiled an edition of the letters in 1953 by transcribing the original letters.² Winchester's edition was then digitised from February 2020 for use on the project's bespoke Back2theFuture annotation platform. The digitised letters constitute the corpus.

The corpus of 881 letters is primarily Early Modern English, for which orthography has not yet stabilised. There are 10 Early Modern French letters, and a single letter is a mix of both Early Modern French and Early Modern English. However, many Early Modern English letters contain words or expressions in Early Modern French and/or Early Modern Dutch. Furthermore, the small pool of corresponding authors are mutually inflecting each others' language and idiosyncratic spelling preferences. The scope for individual spelling variation is thus wide, which can make interpretation, and thereby annotation, difficult. These guidelines therefore aim to present examples of the entities to be annotated in their breadth of variation.

Named entities are definite noun phrases that reference specific types of real-world objects, such as individuals, organisations, and locations. Named Entity Recognition (NER) systems

¹ University of Antwerp. (2020, February 1). *Case studies*. <https://www.uantwerpen.be/en/projects/back-to-the-future/case-studies/>

² Winchester, B. (1953) *The Johnson Letters, 1542-1552: Edited with an introduction by Barbara Winchester* [Doctoral dissertation, Birkbeck]. British Library.

typically include temporal and numerical expressions as named entities, too. The entities annotated in this corpus include named entities associated with its formal definition and extend the NER systems approach with a number of additional kinds of annotations. "MARKET", for example, designates an intersection of temporal and geographic content specific to this corpus.

2. General instructions

2.1 Back2theFuture annotation platform

For a detailed account of navigating and using the Back2theFuture annotation platform, see the Back2theFuture annotation platform manual.

2.2 Nesting

Although the annotation platform permits nesting one entity inside another entity, as well as allowing for partially overlapping entities, these guidelines do not permit nested entities.

2.3 Punctuation

As per the functioning of the annotation platform, punctuation cannot be treated distinctly. For example, take "*And your lettre which was brought me by Thomas Fleeton!*". The exclamation mark will be retained if selecting "*Thomas Fleeton*" for annotation.

2.4 Winchester's transcription

Winchester's amendments are always inside square brackets or surrounded by forward slashes, and where text is missing there are ellipses in place. As per section 2.3, punctuation cannot be treated independently, thus when a selected span of text includes square brackets or forward slashes they will remain in the annotated span.

2.5 Ambiguities and interpretation

Ambiguities typically arise as a result of particularly cryptic spelling variants. Nearly as often it is simply because a word or name has fallen out of usage. Two methods have proven appropriate for negotiating these problems. Firstly, obey a phonetic pronunciation of the ambiguous case and see what you divine. Secondly, and more properly, search the Oxford English Dictionary for historical and/or obsolete forms. A simple Google search is also particularly helpful.

There is no method for tagging ambiguous cases. Ideally, inter-annotator agreements would resolve ambiguities.

2.6 Tagging schema

Following Jurafsky and Martin, a BIO tagging system is used:

In BIO tagging we label any token that begins a span of interest with the label B, tokens that occur inside a span are tagged with an I, and any tokens outside of any span of interest are labeled O. While there is only one O tag, we'll have distinct B and I tags for each named entity class.³

3. Entities

3.1 "NAME"

When the entity refers to an individual person using a proper name or names.

Where an individual is referred to with a single forename or surname along with titular information, the name and titular information are annotated, including where titular information refers to a geographical location (see section 3.2).

Kinship terms directly preceding or following full names are not annotated. However, kinship terms directly preceding or following a single forename or surname are annotated along with possessive pronouns that immediately precede the kinship term.

Honorific titles accompanied by proper names are annotated. Entities that refer to members of royal families or politicians by their honorific title alone are annotated inclusive of articles. Note that entities that refer to religious figures should not be annotated (see section 3.8.).

The personal name element of company names are annotated when they refer to an individual, however a company's full trading name does not form part of the annotated entity.

Examples of "NAME" entities:

"Otwell Johnson"

"Mastar Jonsson"

"Lord Gray"

"my cossyn Cave"

³ Jurafsky, Daniel and James H. Martin. 2023. *Speech and Language Processing*. Draft version, Jan. 7, 2023. Available from <https://web.stanford.edu/jurafsky/slp3/>.

"Jackeson the grocer"

"the Quene"

3.2 "LOCATION"

When the entity refers to a real-world location.

This annotation is reserved for entities that retain only location-based significance. Entities that refer to major trade markets retain an additional temporal significance and will be annotated as "MARKET" (see section 3.5). All other markets are annotated for their location-based significance.

Territories with a geopolitical border are annotated. This includes cities and smaller units such as communes and boroughs; regions that refer to the internal divisions of a state up to the level of country, including traditional regions; and countries.

The built environment is annotated. This includes terrestrial locations such as streets, squares, ports, and waterways; and buildings such as courts or universities. Where the built environment is referred to generically with accompanying location information, the instance of the built environment and location are annotated.

The natural environment is annotated. This includes mountains and valleys, as well as waterways and seas.

Extraterrestrial locations are not annotated.

Examples of "LOCATION" entities:

"Andwarpe"

"Flandres"

"Le Pays basque"

"Engleter"

"Flete Strete"

"Temple Bare"

"the Courte at Calleis"

"Narrowe Sees"

3.3 "NATION"

When the entity refers to one or more people or objects by their nationality or origin.

A nation or origin is a territory with a geopolitical border. This includes regions that refer to the internal divisions of a state up to and including the level of country.

Annotations are inclusive of articles and demonstratives in a construction. Quantifiers and modifiers are not annotated.

Examples of "NATION" entities:

"Flemysse"

"Les Holandois"

"Ynglyshemen"

"les Allemans"

"Harlamers"

3.4 "MARKET"

When the entity refers to a significant trade market.

This special set of entities are used by merchants to signal geographic and temporal information. An example of each significant trade market is given in the example below.

Entities that refer to a significant trade market without naming the market itself are annotated. The greater context of the letter is especially critical in such cases.

Annotations are inclusive of articles and demonstratives in a construction.

Examples:

"Synxon marte"

"Payz Marte"

"Coventreye Fayre"

"Bams [m]arte"

"Colld marte"

"Cold"

"Paske marte"

"halffe Pasche"

"this marte"

3.5 "DATE"

When the entity refers to an absolute date or period.

A construction with at least one fundamental calendar unit (day, month, year) is annotated.⁴ Day segments (e.g. afternoon, evening) are not annotated, however, if they form part of an absolute date construction they are annotated. Dates in the Christian liturgical year are annotated.

⁴ I am loosely following the distinction made in Cohen, D & Schwer, S. R. (2011) Proximal Deixis with Calendar Terms: Cross-linguistic Patterns of Temporal Reference. Unpublished manuscript. Retrieved from <https://shs.hal.science/halshs-01059978>

Named days of the week (e.g. Monday, Tuesday) are not annotated, however, if they form part of an absolute date construction they are annotated. Seasons or other recurrent periods are not annotated. Constructions that signal indeterminate or relative temporalities are not annotated (see section 3.6).

Annotations are inclusive of articles and demonstratives. Entities expressed in words, numerals, or combinations of words and numerals are annotated.

Examples:

<i>"le 19"</i>	<i>"Marche"</i>
<i>"xviiijth(1) day of Aprylle nexte"</i>	<i>"Eester weke"</i>
<i>"13th of this present"</i>	<i>"12th in February, 1542"</i>
<i>"anno 1546, the xxiiijth in Aprill"</i>	<i>"Wytson tyd"</i>
<i>"Monday, the second of this monythe"</i>	
<i>"anno 1542, on Ashe Wedensdaye"</i>	

3.6 "TIME"

When the entity refers to an indeterminate date or period.

A construction that directly signals temporality without using fundamental calendar unit (day, month, year) information is annotated. Adverbs are annotated if the speaker intends to refer to a real if indeterminate point in time. A construction that refers to the present without any fundamental calendar unit information is annotated. Named days of the week (e.g. Monday, Tuesday) are annotated if they do not form part of an absolute date construction.

Annotations are inclusive of articles and demonstratives. Entities expressed in words, numerals, or combinations of words and numerals are annotated.

Examples of "TIME":

<i>"shertlye"</i>	<i>"wekelye"</i>
<i>"this yere"</i>	<i>"every day"</i>
<i>"aboute this daye"</i>	<i>"Frydaye next"</i>
<i>"withyn thys viij days"</i>	<i>"Yesternight"</i>

"3 dayes ago"

"hens but 10 or 12 before"

"by the daye"

" the [day] above-wrytyn"

3.7 "PRICE"

When the entity refers to monetary units.

All forms that express numerical quantities pertaining to monetary units are annotated. Currency names, including abbreviated names, and symbols are annotated along with numerical amounts.

Examples of "PRICE":

"7s 8d st."

"8s"

" xxvj li"

"forty powndes"

"70 li 11 s 8 d Fl."

"a bagge of shillinges"

"140 crownes and xxs Fl"

3.8 "GOD"

When the entity refers to the Christian God.

Annotations are inclusive of articles and modifiers.

Examples of "GOD":

"God"

"Our Lord"

"His Godhed"

"Nostre Seigneur"

"the Lord Almyghty"

"Living God"

"Deo,"

"Dieu"