

Faculty of Arts
Master's Thesis
Digital Text Analysis

God among family, friends, and business partners.

Appendix A: Annotation guidelines

Jeroen Van Sweeveldt

Supervisor: Peter Petré

Co-supervisor: Sara Budts

Assessor: Jeroen Puttevils

University of Antwerp

Academic year 2023–2024

1. Introduction

This document serves as a guideline for the annotations to be used with the letter collections of Jeake and Marescoe-David in the context of the thesis *God among family, friends, and business partners. Appealing to the divine in merchant letter writing based on the correspondents' connection, 1640-1730*. The workflow of this thesis included digitising the corpora of the letter collections using Tesseract's optical character recognition engine. The digitised letters then need to be processed into data sets, which are then searched for the correspondent's use of divine appeals, by querying for words referring to the divine, that is, God or Jesus. A divine appeal is an instance where the correspondent calls upon a higher power, that is, God or Jesus, for assistance within a certain domain of the correspondent's life. A divine appeal thus implies an act from God, and the manner in which the appeal is formulated provides information on the degree in which the correspondent themselves acts along to achieve the desired outcome (divine approval), or, at the other end of the spectrum, remains passive, putting the course of events entirely in God's hands (resignation).

The processing of these digitised letters into data sets involves two different annotation procedures. The guidelines laid out in Section 2 are to be used as a logical layout analysis to assign layout element tags to every individual line put out by Tesseract's optical character recognition engine. These annotations form the groundwork of the metadata of every letter in the collections used to construct the final data sets.

The annotation guidelines from Section 3 are laid out by the research project '*Back to the Future: Future expectations and actions in late medieval and early modern Europe, c.1400–c.1830*'. The intention of this scheme is to annotate all instances of divine appeals to investigate how it governs the subject's future thinking. As a result, the scheme can be divided in three parts: where throughout the letter the divine was called upon, in which domain of the correspondent's life the divine was called upon, and the type of formula with which the divine was called upon. The type of formula is divided into three classes. The first implies future actions, the second refers to events that have already passed, or that the future is considered an accomplished fact. In the third class, the divine is not called upon as an acting force, but merely mentioned. The goal of these annotations is to investigate whether the divine appeals testify of a specific way of future thinking.¹

Finally, we would like to make a note on ambiguity. Typically, the context in which the divine appeals are made can cover topics spanning different domains. One example is Robert Oursel's comment on an outbreak of the plague in France:

'Comme vous dittes le change a diminué la semaine passée quoy que a mon advis c'est sans beaucoup de fondement, car le peu qu'il y a de contagion en cette ville fait grand peur a tous noz voisins quy ont interdit nostre commerce [...]. Dieu veuille nous en delivrer par sa grace' (*trade & finances, prayer*).²

This particular appeal is on an interesting intersection between *health & sickness, politics & war*, and *trade & finances*. I opted to go with *trade & finances*, because either way, the divine appeal is ultimately concerned with trading. Ideally, these ambiguities could be dealt with through inter-annotator agreements.

1 University of Antwerp, Budts and Heijmans (2023).

2 Roseveare (1987, pp. 87, 241).

2. Logical layout analysis

Logical layout analysis is the detection of page elements such as headers, footers, and body text.³ After programmatically processing Tesseract's `.hocr` files⁴, we annotate every individual line from both Jeake and Marescoe-David collections to assign them a page layout element. This action is necessary to generate the metadata for the final data sets.

Since both collections adhere to different layout styles, they obviously have layout elements unique to each collection. Hence, we will go over every page element for each collection separately.

2.1 Jeake

ADDRESSEE_RAW

This element contains the name of the recipient as written by the sender, that is, this name is not featured in captions or titles that are not part of the letter itself.

Examples: Goodman Coulton
Mr Sa: Jeake Junior

ADDRESSLINE

The name and address of the recipient as written by the sender. This line does not belong to captions or titles that are not part of the letter itself.

Examples: These To Mrs Frances Hartridge in Marden private
For Mr Samuell Jeake Junior at Rye in Sussex, these

BILL

The bill of lading that is included with the letter. The bill of lading is a shipping document that describes the shipped goods, their quantities, and prices.

Examples: One small barrell of mealle for the accoumpt of Master Jeake
Nett 81=3=08 att 36 per Cent 47

BODY

This is the letter's content, without the salutation, sign-off, address line, bill of lading, or any other element that is not part of the letter's topic of discussion.

Examples: You may thinke me too bold in presuming to present these confused
lines
thou marriedst twice against thy will why not a third time to thy poor
friend?

³ Smith (2009, p. 241).

⁴ Tesseract's `.hocr` file extension is shorthand for `HTML OCR`. These are Extensible Hypertext Markup Language files, containing metadata for the OCR output, including coordinates for word and line bounding boxes. We extract each line from these files along with their location on the page that is calculated with these coordinates, and save them to a `.csv` file.

CHAPTER

The title of the current chapter in the collection. This is not part of a letter.

Examples: Religious conflicts in Rye
The persecution of dissenters, 1684

DATE

The date of writing mentioned in the letter.

Examples: March the 24 1653
July the 16

DATELINE

The date and place of writing mentioned in the letter. Other elements like sender name can be included as well.

Examples: London, 10 November 1668
Loving Bro & Cosin Pembroey September 29 1676

FOOTNOTE

Supplementary text provided by the collection's curator to clarify the content of a letter, or any other written passage in the collection.

Examples: *The most foolish of bipeds.
John Allin, former vicar of Rye.

HEADER

The page header. Contains the title of the overarching part of the collection, as well as the page number.

Examples: Conflicts and Courtships, 1640-c.1662 23
Adolescence and Starting Out, 1668-1669 75.

ID

The unique document or letter identifier.

Examples: ESRO FRE 4842
RMA, Selmes Manuscripts, RYEYT N39.59.1

LATIN

Untranslated Latin text written in the correspondence.

Examples: Filius tuus obedientis[simul]s:
audiendi cupidus maneo, scribebam Raptissime

LOCATION

Place of writing, as stated by the sender in the letter.

Examples: Southwarke
{from the] Gatehouse Westminster.

MULTI

This line contains multiple layout elements. In the case of the Jeake collection, it can contain both the letter's title and unique identifier, the sign-off and dateline, the date and sender_raw information, or the sign-off and sender_raw information.

Examples: with heading. ESRO FRE 4223 fos. 102-3
Rye January 28" Your very lo: father

NOISE

This can either be unintelligible data or data without apparent meaning. It can also refer to data that is useless to our analysis. An example is the companion text in the Jeake collection written by Murphy to contextualise the selection of letter presented in the collection.

Examples: He was apparently well-intentioned but acting on limited
experience and received
aoe G d

POSTSCRIPT

A supplement that the sender chose to add to the letter after it had already been signed off.

Examples: I have forgotten to tell you that your letter dated of the
19" March I did
Excuse my evil writing but I slept but half an hour last night
[damaged

SALUTATION

The sender's greeting to the recipient at the opening of the letter.

Examples: My deare and loveing Cosin
Ah my dear (but unkind) Heart

SENDER_RAW

The author mentioning his own name within the letter, usually at the conclusion of the letter.

Examples: Ch: Blackwood
Aunt Elizabeth Dighton

SIGN-OFF

The sender's concluding notes before ending the letter.

Examples: Farewell, desirous of your health,
Remember me to all friends, yours here remember to you.

TITLE

The letter's title, given by the curator of the collection.

Examples: [5] Samuel Jeake senior to Frances Hartridge, 1 July 1650.
Copy letter.
[221] Nathaniel Hartshorne to Elizabeth Jeake, undated. Original letter,

2.2 Marescoe-David**BILL**

The bill of lading that is included with the letter. The bill of lading is a shipping document that describes the shipped goods, their quantities, and prices.

Examples: Ardasse 16 to 17s [per lb.]
From which Mr Simons has drawn £1.300

BODY

This is the letter's content, without the salutation, sign-off, addressline, bill of lading, or any other element that is not part of the letter's topic of discussion.

Examples: 'Wee writt you, Mr Lethicullier, the 6th instant in answer to yours of
the 2nd past when
Concerning the Tar Company's bill of exchange, the matter will be
settled as soon as Mr

CHAPTER

The title of the current chapter in the collection. This is not part of a letter.

Examples: B. LETTERS TO LEONORA MARESCOE & PETER JOYE
C. LETTERS TO LEONORA MARESCOE & JACOB DAVID

DATE

The date of writing. In this collection, this information is standardised by the curator of this collection at the beginning of each letter.

Examples: (ii)⁵22 FEBRUARY 1668
July the 1617 JANUARY 1670

⁵ Often, letters were copied. This Latin number indicates that the author of the collection chose to include a copy of the letter, because the original was damaged (Roseveare 1987, p. 209).

DATE_OF_ARRIVAL

The date when the addressee received the letter. This information is standardised by the curator of this collection at the beginning of each letter.

Examples: 3 Feb.
16 Mar.

DATE_OF_REPLY

The date when the addressee replied to the letter. This information is standardised by the curator of this collection at the beginning of each letter.

Examples: 2 Mar.
3 Apr.

DATELINE

The date and location of writing. This information is standardised by the curator of this collection at the beginning of each letter.

Examples: SEVILLE 29 NOVEMBER 1667
YARMOUTH 20 JULY 1668

EXCHANGE_RATE

The sender's update on the current exchange rates in relevant locations.

Examples: Antwerp: 23st
London—34s 2g 2 usances.

FOOTNOTE

Supplementary text provided by the collection's curator to clarify the content of a letter, translations, or any other written passage in the collection.

Examples: * News had reached the West of England of large catches of fish in Newfoundland, and the season later proved
2 'Sy syn luyden van grote concepten en veel omslach.'

FRENCH

Untranslated French text written in the correspondence.

Examples: Plusicurs navires sont arrivés au bas de nostre riviere entre
lesquels sont Thomas Frost et
Puisque la Triple Alliance s'est faite entre l' Angleterre, Hollande
et Suede l'on croit que

HEADER

The page header. Contains the title of the part of the collection, as well as the page number.

Examples: Markets and Merchants of the Late Seventeenth Century 227
Appendix A 573

ID

The unique document or letter identifier.

Examples: 13
[A5]

LANGUAGE

A tag denoting the original language of the translated letter.

Examples: [D]
[D] & [F]

LOCATION

The location of writing. This information is standardised by the curator of this collection at the beginning of each letter.

Examples: STOCKHOLM (i)
VENICE

MULTI

This line contains multiple layout elements. In the case of the Marescoe-David collection, it can contain both the letter's title and unique identifier, the language tag and part of the letter's body, or multiple languages on a single line, where the writer switches from one language to the other.

Examples: 124 J. REENSTIERNA to CM
as was intended, for it is now too late in the season. Au reste
Mademoiselle je ne puis finir celle

NOISE

This can either be unintelligible data or data without apparent meaning. It can also refer to data that is useless to our analysis.

Examples: [unendorsed]
14 eT es

POSTSCRIPT

A supplement that the sender chose to add to the letter after it had already been signed off.

Examples: [P.S.] Advise how insurance runeth to and from Vergine and Barbads.'
[P.S.] I intreat you to send me the remnant of your promise. My man
hath a generall

SALUTATION

The sender's greeting to the recipient at the opening of the letter. There is only one instance of this in the Marescoe-David collection.

Examples: Sir,

SENDER_RAW

The author mentioning his own name within the letter, usually at the conclusion of the letter.

Examples: [H. Potter].
Eduard Walker.

SIGN-OFF

The sender's concluding notes before ending the letter.

Examples: Yours att command, W.S.'
Bless you, and with heartiest regards I remain, &c.

TITLE

The letter's title, given by the curator of the collection.

Examples: ABRAHAM MOMMA to CM
TAR COMPANY DIRECTORS to CM

YEAR

The year during which the letter was received and answered. This information is standardised by the curator of this collection at the beginning of each letter.

Examples: 1668
1669

3. Corpus annotation

3.1 Letter structure

The letter structure was broken down into four chunks labelled *opening*, *body*, *bridging*, and *closing*, based on the methodology laid out by Rutten and Van der Wal.⁶ Examples of opening formulae are:

'Lo Cousin After my kind love with my wives to youer good selfe and wife hoping youer health as blessed be the Lord we are att present.'⁷

'I last wrote you on the 9th of October, from Hamburg. Since then I have arrived here in good health and have found all my people likewise well, praise be to God who protects us.'⁸

Closing formulae mostly consist of commendation formulae, which serve both to place the addressee under divine protection, as well as to construe the writer's identity as religious:⁹

⁶ See Rutten and Van der Wal (2014, pp. 85–114).

⁷ Murphy (2018, p. 208)

⁸ Roseveare (1987, p. 258)

⁹ Rutten and Van der Wal (2014, pp. 83–5).

'I Commit you to the Lord & rest Sir Yours to Command.'¹⁰

'Shall weare them for your sakes, so praying God to accumulate you with his blessings we kindly salute you and rest, honoured Cosin your most obliged servant and kinsman, &c.'¹¹

Rutten and Van der Wal (2014, pp. 111–2) In our data set, these have been annotated as bridging formulae provide a final piece of intersubjective information:

'I desire you to excuse my suddaine departure when I was last at Rye which was for diverse causes necessitated, but future opportunity god willing may oceaion a longer tariance.'¹²

'For some days I have been lying in a fever so I have been unable to write myself. However, with God's help I hope to be well soon, although I am still very weak.'¹³

3.2 Divine appeal formula types

We classify the divine appeal hits according to how they were used in discourse. More specifically, we annotate for instances in which a correspondent calls upon divine assistance, that is, an action of God to aid them. This classification is of particular importance to the Back to the Future project because these annotations provide information about the degree to whether the subject considered themselves in charge of their actions, or if they perceived God was in control, and that these divine appeals testified of a specific way of future thinking and could factor as one possible representation of secularisation.¹⁴

Class 1: Future-oriented. These hits are oriented towards the future and future actions.

- Divine approval: the writer asks for God's assistance to carry out his own intentions. Although it is a joint agency between the subject and God, the subject has control over these intentions.

'...as for your shute and other things that you writ for **I intend to send god willing** the next weeke the Taylor.'¹⁵

'...and, **if God permits, I shall** at once inform you of their arrival.'¹⁶

- Confidence: the writer is confident that God will take care of things.

'...as **God shall** save my soule which is of far greater worth than a watch this is true.'¹⁷

10 Murphy (2018, p. 200).

11 Roseveare (1987, p. 389).

12 Murphy (2018, p. 47)

13 Roseveare (1987, p. 251)

14 Budts and Heijmans (2023).

15 Murphy (2018, p. 144).

16 Roseveare (1987, p. 450).

17 Murphy (2018, p. 52).

'I most heartily thank you and pray to God to preserve you, your wife and children in good health and prosper you in every way he can in this and many coming years, which I **trust the Lord will do**.'¹⁸

- Prayer: the writer prays to God for something beneficial to happen. These hits are in reference to events or circumstances beyond the subject's own sphere of control.

'Oh that the **lord would** give me a spirit to bear these troubles.'¹⁹

'**May God** bring him and also Thomas Master in safety!'²⁰

- Preservation: similar to prayer, except here the writer appeals to God for the addressee's well-being.

'...the good **Lord keep you & direct you**.'²¹

'I wish you much happiness, **God's blessing** and all prosperity.'²²

- Resignation: the writer asserts that, ultimately, God is in control of events.

'I yet am resolved to rest on providence & **let the Lord do with me as seemeth good in his sight**.'²³

'...but we must bear patiently with **the will of the Lord**, against which there is nothing we can do.'²⁴

Class 2: These hits refer to divine actions that have happened in the past, or when no modality is possible anymore²⁵, because the author perceives the future as an accomplished fact.

- Presupposed or past divine actions: divine actions in if-clauses, or actions that have happened in the past:

'...for the **Lord seemed** to hide himselfe from her.'²⁶

'...**if the Lord is pleased** to bless me again.'²⁷

- Gratitude: the writer thanks God for an intervention advantageous to him:

18 Roseveare (1987, p. 527).

19 Murphy (2018, p. 181).

20 Roseveare (1987, p. 543).

21 Murphy (2018, p. 191).

22 Roseveare (1987, p. 445).

23 Murphy (2018, p. 170).

24 Roseveare (1987, p. 314).

25 According to Portner (2009, p. 1), modality is "the linguistic phenomenon whereby grammar allows one to say things about, or on the basis of, situations which need not be real".

26 Murphy (2018, p. 23).

27 Roseveare (1987, p. 308).

'**Thanks be to the lord** wee are all pretty well at present.'²⁸

'I see that Capt. Nicolas Payton has arrived safely at Harwich, **God be praised!**'²⁹

Class 3: Hits in this class differentiate from the other classes in the sense that they do not relate to direct action performed by the divine. Rather, these hits are found in formulaic constructions, or God is referred to as an object instead of an acting agent.

- Fixed phrase: these are conventional formulaic uses containing the word God or one of its near-synonyms. A non-exhaustive list of examples are dates:

'...only is it expected a great many will be excommunicate for not receiving the sacraments at Easter or the next **Lords day** according to the order of the B. of London.'³⁰

Another example are standard forms used in letter formatting:

'This is all by way of reply; **may God bless you, &c.**'³¹

- God as object: instances in which God is mentioned, but not called upon.

'...I may walke close with **God** in this crooked and perverse generation among whome I live.'³²

'...but the idea that any monopoly should arise from that is odious and accursed of **God**'³³

The Jeake correspondence, however, is a special case in this regard. While God as object occurs only five times in the Marescoe-David corpus, it does so 97 times in the Jeake corpus. I therefore created five subcategories of this formula, which are as follows.

- Allegiance to God: the writer refers to the religion professed by the subject.

'...she **professed Jesus Christ**, and also she was a sister by Baptisme.'³⁴

'And about 10 the last night when her speech was not articulate being asked by another whether **she had any sense of the love of God upon her heart.**'³⁵

- Duty to God: the writer argues that something is a divine rule to which they must abide. This rule is usually dictated by holy scripture.

28 Murphy (2018, p. 97).

29 Roseveare (1987, p. 456).

30 Murphy (2018, p. 170).

31 Roseveare (1987, p. 456).

32 Murphy (2018, pp. 48-9).

33 Roseveare (1987, p. 399).

34 Murphy (2018, p. 26).

35 Murphy (2018, p. 204)

'It is both commanded and practised as a **duty in the word of God as a rule for us to walke by.**³⁶

'Remember to pay the resolutions you made to god in the time of your trouble.'³⁷

- Guidance: writer sends words of (spiritual) advice or encouragement.

'**Encourage your selfe in god** in all hard times.'³⁸

'let all the dispensations of providence **draw you nearer to God.**'³⁹

- Citation: writer directly cites holy scripture.

'...Gen 8 20:21 verses **And Noah builded an alter unto the lord...**'⁴⁰

'...for the meek shall inherit the earth Math 5:5 & **ther shall bee abundance of peace for as long as the sun & moon endure the lord...**'⁴¹

- Manipulation: the writer uses God to change the recipient's opinions or actions.

I know thou expectest **mercy at Gods hands** (which I pray God grant in the hour at death & day of judgement) **but how canst though looke for any & exercise so much cruelty and falseness?**'⁴²

'...but my Love there will be atime of searching all the secrets of our hearts, no covers nor pretences **can cloake us from the all seeing eye of God.**'⁴³

3.3 Divine appeal domains

These annotations refer to the context in which the correspondent appeals to the divine. Domains are collections of subdomains grouped by topic relevance.

36 Murphy (2018, p. 25).

37 Murphy (2018, p. 47).

38 Murphy (2018, p. 46).

39 Murphy (2018, p. 278).

40 Murphy (2018, p. 155).

41 Murphy (2018, p. 156).

42 Murphy (2018, pp. 53–4).

43 Murphy (2018, p. 54).

Domain Subdomains	Explanation
Health & sickness Death Health Sickness	This domain refers to all conversation topics concerning the health (or lack thereof) of senders, addressees, or third parties mentioned throughout the correspondence.
Politics & war Peace Politics Prosecution War	Referring to all topics discussing the political climate. These can range from policies concerning the sender or addressee, to the expression of concerns regarding peace or ongoing war, to legal troubles affecting the sender or addressee.
Social life Birth Communication Courtship Family Marriage Meeting Welfare	Informal topics discussing the social lives of the correspondents. Such topics include family matters, expectations towards communication style, meeting arrangements, or material welfare outside of trading life. Other subdomains include correspondents courting one another, the expansion of social circles through marriage, or the birth of a child.
Trade & finances Crisis Debt Economy Investment Loss Trade	Topics regarding commercial trade. Typically these include trading arrangements, but can also be about crises affecting the trade, such goods being destroyed by fire or bankruptcy, investments made by the sender or addressee, loss of either money or trading wares, the debts of the sender, addressee, or third parties discussed, or the economical situation locally or internationally.
Travel & transport Travel Transport	Topics referring to any type of movement of a person or ware from point a to b. It is annotated as <code>travel</code> when the movement concerns a person, <code>transport</code> when the movement concerns trading wares.
Other	Divine appeals that cannot be placed in a meaningful context. Usually these are highly formulaic instances such as sign-offs, where the writer commends the addressee into the hands of God.

Table 1

Domains, their subdomains, and what they mean to the topic of conversation.

References

- Budts, Sara and Elisabeth Heijmans. 2023. Secularization of future thinking in practice: a comparison between english and french merchant letters (16th–18th century). In *14th European Social Science History Conference*, Gothenburg.
- Murphy, Anne Louise. 2018. *The Worlds of the Jeake Family of Rye, 1640–1736*. Oxford University Press.
- Portner, Paul. 2009. *Modality*. Oxford University Press.
- Roseveare, Henry G. 1987. *Markets and merchants of the late seventeenth century: the Marescoe-David letters, 1668–1680*. Oxford University Press.
- Rutten, Gijsbert and Marijke J Van der Wal. 2014. Letters as loot. *A sociolinguistic approach to seventeenth-and eighteenth-century Dutch*. Amsterdam/Philadelphia: Benjamins (*Advances in Historical Sociolinguistics* 2).
- Smith, Ray. 2009. Hybrid page layout analysis via tab-stop detection. In *ICDAR '09: Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, pages 241–245, IEEE Computer Society, Washington, DC, USA.
- University of Antwerp. 2020. Back to the future: Future expectations and actions in late medieval and early modern europe, c.1400-c.1830. <https://www.uantwerpen.be/en/projects/back-to-the-future/>. Accessed: January 4 2024.