

AOR Phase 1 Transcription Data Summary and Analysis (Final Report)

The transcription and checking work undertaken in the period after the Interim Report (12/1/2015 – 31/9/2016) consisted of the transcription and subsequent checking of all the annotations in Thomas Tusser’s *Fiue hundred pointes of good husbandrie*, the new addition to the AOR1 corpus, as well as checking the remaining 969 transcriptions, as identified in the [previous version](#) of this document (which was an appendix to the AOR Interim Report). This work has been conducted by the AOR post-doctoral fellow, Jaap Geraerts, JHU research assistant Christ Geekie, and UCL research assistants Amanda Brunton and Kristof Smeyers. Moreover, these transcriptions as well as transcriptions made in the earliest stages of the project have been checked yet another time (even though these transcriptions already had been through a process of checking). Overall, this extra round of checking has further enhanced the quality of the AOR data set which, at the end of the grant period of AOR Phase one, consisted of 2,694 XML transcriptions. As the AOR corpus consists of a total of 5,658 digital images, this means that 47.6 per cent of the images contain reader interventions.

Table 1.1 shows how these transcription files are distributed among the 13 volumes in the AOR Phase 1 corpus.

Table 1.1 Number of transcription files per book in corpus

Title book	Number of files
Buchanan, <i>Ane detectioun of the duinges of Marie Quene of Scottes</i>	51
Buchanan, <i>De Maria Scotorum regina</i>	18
Castiglione, <i>Il libro del cortegiano</i>	92
Castiglione, <i>The covrtyer of Covnt Baldessar Castilio</i>	390
Domenichi, <i>Facetie, motti, et burle & Guicciardini, Detti et fatti piacevoli et gravi</i>	422
Freigius, <i>Paratitla</i>	60
Frontinus, <i>The strategems</i>	228
Livy, <i>Ab urbe condita</i>	652
Machiavelli, <i>Art of warre</i>	239
Melanchton, <i>Selectarum declamationum</i>	133
Olaus, <i>Historia de gentibus septentrionalibus</i>	122
Smith, <i>De recta & emendata linguæ Anglicæ</i>	105
Tusser, <i>Fiue hundred pointes of good husbandrie</i>	182
Total	2694

The XML files comprise several “elements”, each element corresponding to a type of annotation: marginalia (a verbally written note); underlining (words and passages underlined or highlighted in some way); and a non-verbal mark or symbol. Additional tagging protocols were developed, as well, to deal with more exceptional cases of annotations such as visual drawings; numerals (passages marked up, for example, “1, 2, 3”); and errata (where the annotator appears to have corrected or specifically altered the printed text). In addition, we created the ability to quantify the specific number of words written by Harvey as marginalia, and the number of printed words underlined by Harvey within each book and across the full corpus of books.

As detailed in Table 1.2, below, the 2,694 transcription files contain a total of 109,417 individual elements. More specifically, they contain a total of 73,247 words of manuscript annotations and 227,024 underscored words, comprising a total body of annotated materials amounting to 300,721 words.

Table 1.2 General overview of corpus

Title book	Marginalia	Underline	Mark	Symbol	Drawing	Numeral	Errata	Total	Marginalia_ words	Underline words	Total
Buchanan, <i>Ane detectioun...</i>	13	635	221	4	0	0	5	878	176	2070	2246
Buchanan, <i>De Maria Scotorum regina</i>	8	48	3	5	0	0	2	61	39	156	195
Castiglione, <i>Il libro del cortegiano</i>	92	116	48	0	0	0	6	262	672	368	1040
Castiglione, <i>The covrtyer of Covnt Baldessar Castilio</i>	399	12090	5262	644	2	28	18	18443	2371	44568	46939
Domenichi, <i>Facetie, motti.. & Guicciardini, Detti et fatti..</i>	3087	10906	5905	697	1	13	13	20622	36256	32899	69155
Freigius, <i>Paratitla</i>	50	468	331	31	0	13	2	895	1489	975	2464
Frontinus, <i>The strategems</i>	875	3747	2637	147	0	18	3	7427	9227	11915	21142
Livy, <i>Ab urbe condita</i>	854	26110	10744	1145	1	66	60	38980	18632	84351	102983
Machiavelli, <i>Art of warre</i>	215	7288	4082	92	0	51	76	11804	2444	27102	29546
Melanchton, <i>Selectarum declamationum</i>	45	1129	327	13	1	8	9	1532	349	2968	3318
Olaus, <i>Historia de gentibus septentrionalibus</i>	53	1552	293	18	0	20	3	1939	669	4538	5207
Smith, <i>De recta & emendata linguæ Anglicæ</i>	42	687	224	6	0	3	3	965	323	1966	2289
Tusser, <i>Fiue hundred pointes of good husbandrie</i>	94	3954	1495	59	1	3	3	5609	600	13147	13747
Total	5827	68730	31572	2861	6	223	203	109417	73247	227023	300270

Thomas Tusser's *Fiue hundred pointes of good husbandrie* has been a significant addition to the corpus, both in terms of its quantitative (although it is a rather small book, it is heavily annotated throughout) and qualitative aspects (it is the only book of the corpus written in verse). This qualitative aspect of Tusser's book has gotten ample attention in a [blog post](#) by Kristof Smeyers and [one](#) by Jaap Geraerts. The authors of both blogs have stressed that in general this qualitative difference between Tusser's book and the rest of the corpus was not reflected in quantitative terms – the way in which Tusser's *Fiue hundred pointes of good husbandrie* has been annotated mirrors the annotation patterns found in the other books in the AOR corpus. The fact that the majority of the marginal notes that are to be found in Tusser's book were written in English and not in Latin (the language used for the vast majority of marginal notes in the AOR corpus), might be indicative of Harvey's interest in the linguistic qualities of the book. However, in the other English language books within the AOR corpus we also see a greater number of annotations made in English, and it might be that Harvey tended to annotate the book in the language in which it was written. Still, a significant percentage of the marginal annotations made in the English-language books in the corpus are in Latin (in Frontinus' *Strategemes* the marginal annotations written in English and Latin are almost equal), which raises questions whether readers had a preference to address certain topics in a particular language (relating to their mastery of a particular vocabulary in a given language). We have not pursued this line of inquiry further, but these and related questions show how useful it can be to subject the dataset to a relatively straight-forward statistical analysis.

One of the tasks we set ourselves in the Interim Report from the remainder of the grant period was to deal with several remaining questions pertaining to the humanistic content of the transcriptions that remained unsolved. Although we have been able to solve a number of these outstanding issues, we have not been able to solve them all. These remaining issues and uncertainties have been flagged in the XML transcriptions and are also included in a list which will be uploaded to the AOR site later this week. These efforts indicate our intention to communicate the uncertainties, which are immanent to any research project undertaken in the humanities, to our user base. Over the course of AOR2, as our understanding of early modern reading practices matures and perhaps based on the suggestions we receive from our users, we will revisit these remaining issues.