

Gnome on the Range: Finding the Hypertextual Narratives in Ancient Wisdom Texts

K. Faith Lawrence
Department of Digital Humanities
King's College London
26-29 Drury Lane, London, UK
faith.lawrence@kcl.ac.uk

Anna Jordanous
Centre for e-Research
King's College London
26-29 Drury Lane, London, UK
anna.jordanous@kcl.ac.uk

ABSTRACT

In this paper we present the Sharing Ancient WisdomS (SAWS) project. Working with wisdom texts, or gnomologia, the project aims to produce an enhanced digital scholarly edition of the collected manuscripts which both makes the Greek, Arabic and Spanish texts available and demonstrates the hypertextual nature of these texts. By positioning the texts as collections of sayings, of which a given manuscript only shows one narrative path, we demonstrate how a hypertextual approach allows us to explore alternate narrative paths within and across the texts and support researchers as they study the context, significance and transmission of the wisdoms within these works.

Categories and Subject Descriptors

H.5 [HCI]: Hypertext/Hypermedia; J.5 [Arts and Humanities]: Classics

Keywords

Hypertext, Narrative, Linked Data, Digital Humanities, RDF, TEI XML, Manuscripts, Multilingual, Gnomologia

1. INTRODUCTION

This paper will present the work being done within the Sharing Ancient Wisdoms (SAWS) project to publish a selection of Gnomologia, or wisdom texts, and expose and visualise the relationships that exist within and between them. We argue that documents of this type represent collections of largely independent sections which exist as conceptual constructs outside the framework of any given edition. By regarding these textual nodes as separate entities which can be linked by the configuration of a published manuscript but also to each other across manuscripts through a variety of other relationships we can begin to discover the hypertextual network behind the works and gain new ways of looking at these textual elements and the narrative paths that can be drawn between them.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NHT'13, May 1, 2013, Paris, France.

Copyright 2013 ACM 978-1-4503-2005-4/13/05... \$15.00

This paper presents the technological, structural and presentational framework developed by the project and shows how this supports the hypertextual readings of the texts being used within the project.

1.1 Parallel Corpora and Hypertextual Links

The presentation of texts in multiple displays has a long history. Mimicking the common academic practice of having multiple books open at once it provides a familiar way of displaying a variety of manuscripts and other material on one screen. As a presentation tool for digital humanists it has enjoyed a popularity among researchers dealing with texts with multiple versions, either diachronic or synchronic variations or translations, or for the display of commentary and text together. For example, parallel text displays have been an especially useful technique within Bible studies(4) with tools such as the Bible Web App freely available online¹ for Biblical scholars, both inside and outside academia.

Outside Bible studies, frameworks such as the Versioning Machine(7) provide ways to visualise the multiple versions of a text in parallel, complete with critical commentary. In the case of multilingual texts, parallel corpora are commonly used to display translated texts². The Translation Array: Version Variation Visualization³ is using parallel displays to present multiple translations of Othello.

The question of alignment(12), that is matching up related areas of text at paragraph, line or word level when the documents are displayed in parallel, was one of the big challenges in the presentation of parallel corpora. The use of computer processing in this area means that while owing much to translation studies, today a significant amount of research in parallel corpora is driven by linguistic processing and is related to computation-driven applications such as machine/computer-aided translation and machine-aided learning(11). Within the SAWS project the decision was taken for the alignment to be driven by the manual annotation of relationships by our researchers due to both the expert knowledge required and the time needed to prepare the texts for computational analysis. As a result these annotations are seen within the project as scholarly claims within their own right, and with associated citation. Following the

¹<http://biblewebapp.com/v2/app/index.html>

²Texts maybe directly or indirectly translated. Indirect translations occur when the texts are linked via another text or texts

³<http://www.delightedbeauty.org/vvvclosed/>

model laid down in the SAWS ontology (see below) the annotated relationships lay down the groundwork for not only the alignment between given manuscripts but the hypertextual network and narrative paths within that network.

SAWS builds on and shares many of the methodologies of parallel corpora, comprising of a collection of related texts and translations. However the project differs in its focus on the use of linked data to model and expose the network of relationships that exist within and across the corpora, allowing us to see a saying, the building blocks of our texts, in not only its manuscript contexts but in the context of other related sayings from other manuscripts.

2. SHARING ANCIENT WISDOMS

The SAWS project was funded under the 2009 HERA Cultural Dynamics: Inheritance and Identity call to promote the study and understanding of Greek and Arabic wisdom texts. Bringing together experts from the University of Vienna, University of Stockholm and King's College London, SAWS works with Arabic(2), Greek(5; 6), and Spanish texts, some previously unpublished, along with modern English translations. The initial texts were selected by the experts at the partner institutions who would be leading the document markup and preparation (see below). As well as making the texts more accessible to scholars and researchers, the SAWS project aims to expose the links and connections that exist between the sections of the documents beyond the immediate textual structure, allowing the hypertextual links and narratives of the works to be navigated and explored.

At the current time the SAWS corpus includes selections from Codex Bodleianus Digby, Codex Parisinus Graecus, Codex Vaticanus Graecus, Codex Vindobonensis, Graecus Theologicus, Proclus's *The Elements of Theology* (ed. Dodds), Hacı Mahmud Efendi, *Bocados de Oro*, Muhtar al-hikam wa mahasin al-kalim, *Miscellany Ayasofya*, and excerpts from Ali ibn Rabban al-Tabari's *Firdaws al-hikma*. It is hoped that the corpus will be expanded in the future, either directly or through partnerships with similar projects.

2.1 Gnomologia

Gnomologia or wisdom texts, historical collections of gnomes or aphorisms, proverbs and good advice, are of significant interest to classicists, byzantinists, medieval historians and philosophers as they formed a crucial route by which concepts fundamental to philosophy and ideas of behaviour and conduct were disseminated across the classical and Byzantine world and beyond. One key point of interest is the overlap between the texts, with some sayings appearing in some form in multiple texts showing transmission and translation across the manuscripts. The terms aphorism and saying are used interchangeably in this paper to refer to the distinct subsection of text that makes up that textual node. In most cases this includes a short narrative introduction to the aphorism or saying and then the aphorism or saying itself.

2.2 Difficulties of Working With Ancient Texts

The history of the proverbs collected within our sources contains many unknowns, with pivotal parts of the literary corpus lost. While there is evidence that the extant source

documents contain aphorisms which were themselves taken from earlier works, in some cases undergoing translation in the process, there is little direct evidence of the transmission chain beyond the similarities in text. In many cases the evidence points not to a direct connection but to a hypothesised intermediary manuscript.

Even where the texts themselves exist the bibliographic metadata related to them may be missing or in doubt. Details such as the author, location and time of creation may be unknown. Additionally the text as a conceptual work may have been pieced together from multiple, incomplete and/or damaged examples. Since these works had to be copied by hand, different versions of the same manuscripts often contain errors, editorial amendments and areas of illegibility which the scholar must evaluate to divine what they believe to be the intended output.

2.3 Narrative Paths

If we regard the source texts not as a selection of manuscripts but as a collection of aphorisms which may be structurally located in one or more places across the body of material, then we see they not only have a narrative path dictated by their physical structure, as decided by the collector of a given selection, but can legitimately be understood and studied through alternative narrative paths. We argue that the chain of links between the text nodes as created by the aphorisms represent a narrative pathway which can be explored to tell the story of the saying over and through different facets and relationships. Example paths of interest for the scholars within the SAWS project include the linking of sayings, linearly and in parallel, with and across collections by topic, conceptual reference, collection author, ascribed author or speaker of the aphorism and variations of an aphorisms.

The order in which individual sayings or themed groups of sayings appear in different manuscripts holds particular value for some researchers. Continuing our projection of these manuscripts as representing selected paths through the collection of aphorisms, we can regard this as an example narrative and treat it in the same way as an instance of a reader's interaction with a work of hypertext which may be recorded, replayed and viewed.

For others, the slight variations in aphorisms between versions are of great scholarly interest. This is especially true for the texts selected for the SAWS project as they include source texts from different languages and cultures but which contain the same sayings. The diachronic journey of a given aphorism as it passes between different authors and cultures represents a narrative path that would have been very interesting to expose and present. Unfortunately this was not possible to implement in our particular case due to the lack of existing information. While approximate dates for source texts are estimated, there is not enough evidence for scholars to date when the saying moved across cultures, such as Greek to Arabic or vice versa, or even, necessarily, the direction of transmission as the extant manuscripts cannot be conclusively said to represent the first instance of use, just an instance that has survived through to the modern day.

The lack of temporal or other definitive hierarchy means that any given group of related sayings can be interrogated from any starting point and with no restrictions on order so long as direct links are followed. One of the main aims of the project is to allow researchers to navigate the textual content in the manner, forefronting the interconnected and often parallel existence of the aphorisms, going beyond what is currently found in standard implementations of digital editions of historical texts.

3. BEYOND DIGITAL EDITIONS

In creating the digital edition of the texts it was seen as important to not only present the text in a traditional scholarly edition but to offer researchers ways to visualise and explore the paths between the text sections and then see the text within those contexts.

The source texts were all XML-encoded by researchers at partner institutions, each taking responsibility for documents in their specific areas, and languages, of expertise. To allow direct reference to any given section such as an aphorism, a script was then run on the XML documents to assign unique identifiers to each tagged area. Initially, structured but otherwise meaningless identifiers were used; however this was seen as a less than ideal solution(3). The project is in the process of moving to the Canonical Text Services protocol, an emerging citation standard within digital humanities, for our identifiers(8).

The project uses the TEI XML standard(9) for text encoding. One aspect was to develop a customised TEI schema to define more specifically how the parts of the texts should be described across the partner institutions. Further, an ontology was developed to lay out the conceptual model of the relationships between the text sections within their related contexts and provide a taxonomy of terms for use by the scholars involved in the markup process.

The SAWS ontology was developed as an extension of FRBRoo(1) with significant input from the domain experts as to what relationships could be stated to exist with acceptable scholarly rigour and therefore what narratives could be created across the texts. These relationships were also encoded in the XML through the use of the TEI *relation* entity(10).

Kiln, created by researchers at King's College London, is a platform for deploying websites driven by XSL transformations of XML encoded source files. In conjunction with Solr, this framework offers the potential to create indexed and browsable scholarly editions. Additionally XSLT was used to extract the relationships as RDF triples which could then be used to power the visualisation and navigation of the generated narrative paths (see Figure 1).

To allow more dynamic views of the textual sections as controlled by the user, the libraries behind the Versioning Machine framework(7) were expanded and repurposed. Using the mechanism for displaying parallel versions we converted the system to take a collection of aphorisms, initially the collections defined by the historical texts, to populate the parallel views. Options were then given to the user to highlight sections which had a link to another section of text, automatically lining up linked sections horizontally across

the parallel views. Highlighting was used to emphasis the cross-view path with the different types of relationships being colour-coded. (see Figure 2).

The framework is still under development (see below). As it is extended further and fully integrated within the digital edition, the option to define the collection of sayings which will populate the vertical views will be given to the user. This will allow thematic or non-structurally related groups, or even individual sayings, to make up the vertical axis in addition to collections defined by a given document, further breaking the aphorisms out from their linear textual structure.

4. FUTURE WORK

The framework described above has been used to create a static demonstrator to assist humanities scholars working with the project to gain a better understanding of how such a treatment of the manuscripts would support their needs. Following these discussions a more specific set of visualisation options was agreed upon and those options, as part of the dynamic version of the extended digital edition, are under development. As part of this ongoing development we hope to carry out user interaction testing to improve the usability of the edition for humanities researchers and others with an interest in this area.

We are also in discussion with researchers working with other corpora with similar compositions and complex interconnections with a view to both expanding the manuscripts that are included in our application and exporting our methodology to other projects.

5. CONCLUSIONS

Scholars working with ancient wisdom texts are interested in the interconnections between and transmissions of aphorisms collected within their manuscripts. Through the use of multiple display techniques we are creating ways to allow researchers to interact with this source material as hypertextual narrative. By positioning the texts in this way we can expose the narrative paths that exist within and across the manuscripts and allow scholars to follow and recontextualise the wisdoms as significant objects in themselves, facilitating their study beyond a proscribed pathway as presented in a given manuscript.

In this paper we have laid out the reasons why we have chosen this methodology for the manuscripts used in the SAWS project and how we have implemented this in a way to support scholars in their research through the exposure of the narrative paths, both historic and conceptual, that exist within and between those documents.

6. ACKNOWLEDGMENTS

The SAWS project is funded by HERA (Humanities in the European Research Area) as part of a programme to investigate cultural dynamics in Europe. It is composed of teams at the Department of Digital Humanities and the Centre for e-Research at King's College London, University of Stockholm and the University of Vienna.

References

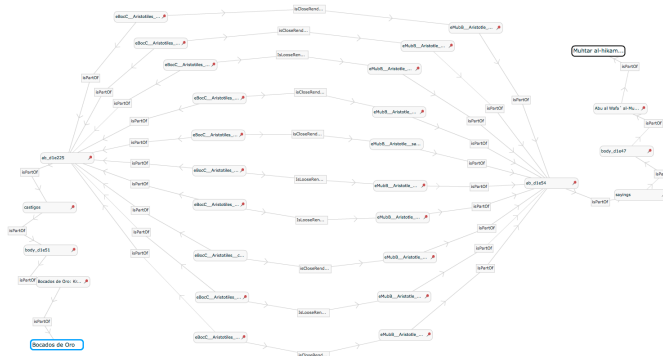


Figure 1: Screenshot of the relationship paths between aphorisms in example Arabic and Spanish texts

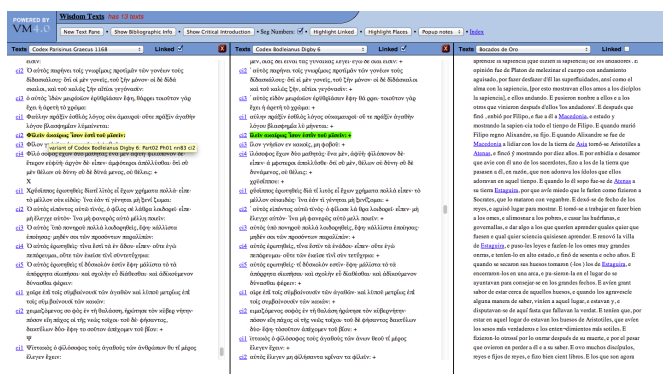


Figure 2: Screenshot of the texts displayed in parallel with the linking texts highlighted

[1] M. Doerr and P. LeBoeuf. Modelling intellectual processes: The frbr - crm harmonization. In C. Thanos, F. Borri, and L. Candela, editors, *Digital Libraries: Research and Development*, volume 4877 of *Lecture Notes in Computer Science*, pages 114–123. Springer, Berlin / Heidelberg, 2007.

[2] D. Gutas. Classical Arabic wisdom literature: Nature and scope. *Journal of the American Oriental Society*, 101(1):49–86, 1981.

[3] A. Jordanous and K. F. Lawrence. Exploring manuscripts: Sharing ancient wisdoms across the semantic web. In *Proceedings of WIMS '12: International Conference on Web Intelligence, Mining and Semantics*, Craiova, Romania, 2012.

[4] P. Resnik, M. B. Olsen, and M. Diab. The Bible as a Parallel Corpus: Annotating the 'Book of 2000 Tongues'. *Computers and the Humanities*, 33(1–2):129–153, April 1999.

[5] C. Roueché. Sharing Ancient Wisdoms. In *Aliento: Corpus anciens et bases de données*, pages 297–299, 2002.

[6] D. Searby. Sharing Ancient Wisdoms: On Preparing a Digital Hyperedition of Greek Gnostic Collections. In *Ars Edendi Workshop Proceedings 2010*, Centre for Medieval Studies, University of Toronto, 2012.

[7] M. N. Smith. Mith's lean, mean, versification machine. In *Proceedings of ALL/ACH 2002: New Directions in Humanities Computing*, pages 122–7, University of Tübingen, 2002.

[8] N. Smith. Citation in classical studies. *Digital Humanities Quarterly*, 3(1), 2009.

[9] TEI. TEI: Text Encoding Initiative. <http://www.tei-c.org>, 2011. Last accessed 31st October 2011.

[10] TEI. Text Encoding Initiative / Tracker / Feature Requests / Encoding RDF relationships in TEI - ID: 3309894 (Discussion). <http://sourceforge.net/tracker/>, 2011. Last accessed 31st October 2011.

[11] J. Tiedemann. *Bitext Alignment*, volume 14 of *Synthesis Lectures on Human Language Technologies Series*. Morgan & Claypool Publishers, 2011.

[12] J. Véronis. *Parallel Text Processing: Alignment and Use of Translation Corpora*, chapter Evaluation of parallel text alignment systems: The Arcade Project, pages P. 369–388. Springer, 2000.