

Humanities Research Data Alliance Meeting  
The Text Problem: Scale, Heterogeneity, Language, Culture  
May 28, 2015  
Issues and Use Cases  
Gregory Crane

All sources and activities in the Humanities are mediated by language and culture. Even if we focus upon music or art, we express our opinions and refine our judgments through linguistic exchanges that reflect the languages that we have studied and the cultural backgrounds to which we have been exposed. We may in practice assume that all relevant voices will express themselves in the Developed World Scholarly Communication Culture (formal lectures and seminars, peer reviewed articles and monographs) and in a handful of languages (primarily English, but with some French, German, Italian, and Spanish -- with Mandarin and other non-Western languages already emerging). But even if we exclude everyone else from authoritative discussion, our primary sources include anything from any cultural system and in drawing on any language from the human record. This problem focuses upon the textual problem as the starting point for discussion.

This document focuses upon a relatively constrained, but strategic, problem: supporting work on a digitized subset of the textual record, a record that covers hundreds (rather than thousands) of languages and for which a substantial body of scanned print documents lends itself to optical character recognition (OCR) and optical symbol recognition (e.g., for music and math).

Strategic issues:

### **Scale:**

We need to integrate curated collections (from a single document to thousands) with mass collections that defy short term manual curation within the documents themselves (e.g., the 300,000 carefully digitized books produced in Germany through 1800 or the 5.1 million HathiTrust volumes that are in the public domain in the US) -- i.e., if a scholar creates a carefully marked up version of a poem by Herder, that markup should be linked to raw OCR output from scanned books and should be visible to anyone viewing any scanned book that contains that poem. In effect, this requires a synthesis of Digital Libraries (DL) with the Digital Humanities (DH). For DH, such a fusion is essential. Without integration with the DL world, DH publications become isolated and incommensurable.

Without integration with the DL world, DH projects cannot address problems at either very large or very deep scales.

- A very large project might trace the development of nationalism and the decline of religion as an organizing principle in European culture from 1450 to 1870. Such scholarship must extract as much as possible from analysis of as much digitized content as possible, drawing upon a synthesis of so-called distant reading and close reading applied to samples of the whole selected according to documented principles. The fact that such inquiry is not yet common, that established scholars are content with the practices to which they are accustomed and, indeed, that research across millions of books has become practicable in only very limited ways (e.g., Google N-Grams) are primarily important because they illustrate the work before us.
- Every digital edition represents -- or should represent -- a deep project, one in which each variant and each annotation (e.g., minimally, upper and lower case letters, punctuation, etc. in editions of European sources) associated with any word or combination of words in any version of any source must be a citable object.

## Heterogeneity:

Even if we focus upon the written record, our digital infrastructure must focus not only upon the physical written objects (books, inscriptions, papyri, etc.) but the textual content that those objects preserve. We need to address any subset of any set of words in any version of any text in the human record. Thus standards such as METS, EAD, TEI-HEADER, LIDO, PREMIS and DCCD (acronyms from a Darjah-DE internal report in German) may be necessary but they are by no means sufficient and no digital infrastructure should assume that focusing on such models will provide the functionality upon which new scholarship and teaching depend.

Humanities scholarship MUST have functionality comparable to the Functional Requirements for Bibliographic Records (FRBR) data model so that we can (1) organize our metadata also around logical works (e.g., the *Agamemnon* of Aeschylus or the many versions of the old French *Roman de la Rose*) and (2) distinguish between different versions of a work. We may assume that there was a single source text now lost to us (e.g., a single version of the *Agamemnon*) or that we have multiple different versions of a text, each of which represents a distinct entity and may have been the only version available to a particular community (e.g., different versions of a work by Bach scored differently as personnel at Leipzig changed or Christopher Marlowe's *Dr. Faustus*, which survives in two very different versions).

The Europeana Data Model integrates the FRBRoo data model. The Canonical Text Services/CITE architecture builds on the FRBR data model to provide functionality sufficient for texts that appear in multiple versions. Where we are able to organize texts

to conform with the CTS/CITE architecture, we vastly simplify, if not solve, the underlying problem of representing stable annotation. Implementation issues remain (e.g., what happens when we may have (for example) 3 OCR engines applied to 12 scans of 6 printed copies of an early modern book? We need to be able to treat each OCR-run as a distinct version even if human editors would completely agree on what the transcription should be).

We also need to be able to relate different scans of a written object to one another. Different scans will shoot the image at different angles. The same printed copy may be opened more fully in one scan vs. another. We may have the scanned photographs and 3D scans of inscriptions etc. We need to be able to relate a region of interest applied to one object across many different versions. (This applies also to iconography on Greek vases, which exist in 3D space).

## **Language:**

We need an infrastructure that supports researchers and students alike -- if only because researchers must always be students if they are expanding their intellectual range. In particular, there are so many languages to deal with that we need comprehensive and extensive reading support so that all users can do as much as possible with as many different sources in as many different languages as expertly as possible. European partners have a particular obligation/opportunity to develop infrastructures that make their national languages and their national textual cultures as accessible as possible.

The language problem falls into two overlapping parts, primary and secondary sources.

**Primary sources** can be in any human language, including the c. 6,800 languages still spoken in the early 21st century and the c. 300 languages that show up in very large library collections. Primary sources include the full human record -- from the earliest examples of writing in Sumer, Egypt, and China through newsfeeds, tweets, and other sources that accumulate in real time. For reasons of practicality, both technical and legal, we may focus initially upon textual sources through the early 20th century and especially upon those that are available as scans of printed pages but we need also to consider the challenges of managing less tractable sources (e.g., manuscripts, inscriptions) and real-time data in modern media (e.g., the reception of Ancient Greek Culture or of Goethe over the WWW and social media).

**Secondary sources** include, for the purposes of this discussion, academic articles and monographs. Critical editions mix primary sources with prose introductions and other secondary materials but constitute a relatively small percentage of published output (5%?). Reference works such as lexica, grammars, encyclopedias, curated indices and bibliographies also represent a small but strategic separate case.

More than half of all academic publication now takes place in languages other than English. Given the fact that all(?) of the HathiTrust member institutions are anglophone universities, their aggregate collection surely over-represents English language publications. The share of English-language books has declined from 57.6% (public domain books, which represent the world as of c. 1923) to 50.19%. Thus, even the HathiTrust libraries have, over the past c. 92 years, collected more books that are not in English than in English.

HathiTrust Full Collection (13.4m vols, 6.8m titles)			Public Domain Collection (5.1m vols, c. 38% of the total)		
Language	Count	Percent	Language	Count	Percent
<a href="#">English</a>	3,698,357	50.19	<a href="#">English</a>	1,398,228	57.61
<a href="#">German</a>	686,631	9.32	<a href="#">German</a>	296,839	12.23
<a href="#">French</a>	542,598	7.36	<a href="#">French</a>	245,670	10.12
<a href="#">Spanish</a>	321,179	4.36	<a href="#">Spanish</a>	87,550	3.61
<a href="#">Russian</a>	260,396	3.53	<a href="#">Latin</a>	79,410	3.27
<a href="#">Chinese</a>	254,652	3.46	<a href="#">Italian</a>	71,095	2.93
<a href="#">Japanese</a>	228,123	3.10	<a href="#">Japanese</a>	43,691	1.80
<a href="#">Italian</a>	193,841	2.63	<a href="#">Russian</a>	32,412	1.34
<a href="#">Arabic</a>	124,745	1.69	<a href="#">Dutch</a>	19,072	0.79
<a href="#">Latin</a>	100,943	1.37	<a href="#">Chinese</a>	17,837	0.73
<a href="#">Portuguese</a>	64,840	0.88	<a href="#">Portuguese</a>	14,078	0.58
<a href="#">Polish</a>	61,487	0.83	<a href="#">Swedish</a>	11,157	0.46
<a href="http://www.hathitrust.org/visualizations_languages">http://www.hathitrust.org/visualizations_languages</a>			<a href="http://www.hathitrust.org/visualizations_languages_pd">http://www.hathitrust.org/visualizations_languages_pd</a>		

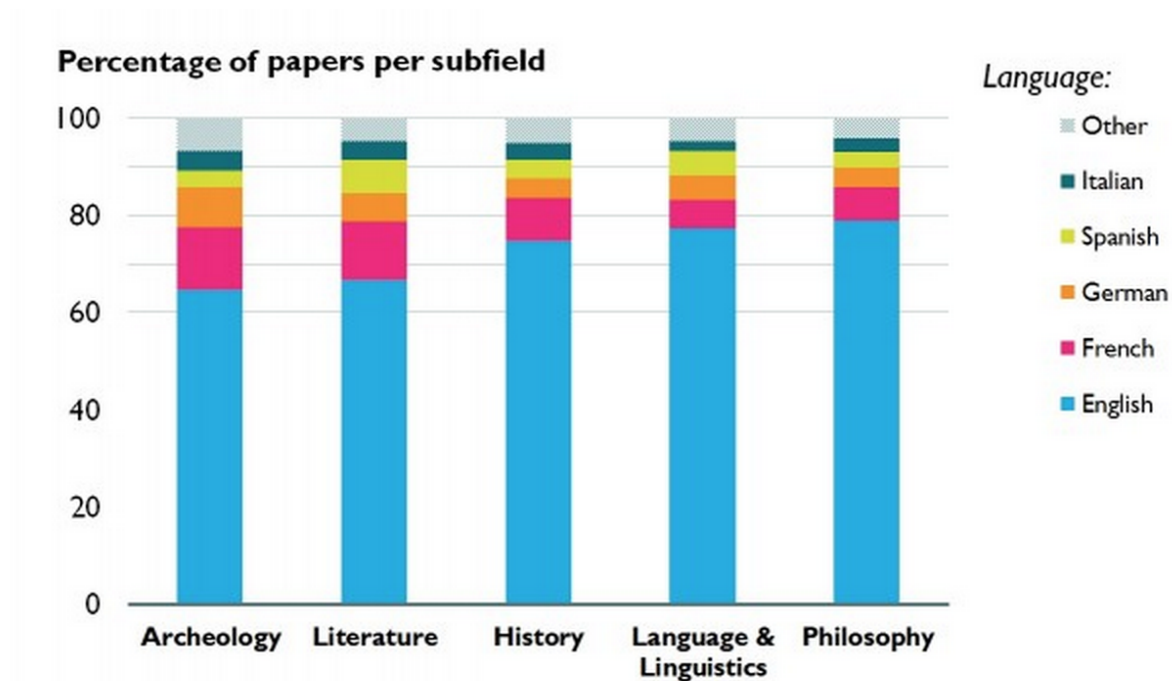
Europeana?

But if more than half of all monographs now collected by HathiTrust institutions are non-English, more than three-quarters (77%) of all indexed journals (according to one dataset) are in English, while 95% of all publications are in English, French, German, Italian or Spanish. Only French publications include a majority of articles (57%) written in the French language, while only 45% of the documents in Spanish publications are written in the Spanish language. Roughly three-quarters of all materials in French and German publications are now in English. Chinese, Russian and Portuguese are included altogether for a subset of the “other” category.

<http://www.researchtrends.com/issue-32-march-2013/publication-languages-in-the-arts-humanities-2/> (data from <http://www.scopus.com/>)

Country	Article count	Language (%)					
		English	French	German	Italian	Spanish	Other
United Kingdom	27400	98.0	0.7	0.3	0.1	0.6	0.3
United States	67815	97.3	0.8	0.2	0.1	1.4	0.2
The Netherlands	4985	89.8	1.4	1.3	0.0	0.5	7
Russia	1015	84.5	2.7	1.3	0.0	0.5	11
China	4231	78.2	0.4	0.4	0.1	0.2	20.7
Portugal	942	76.3	4.0	0.7	1.3	4.8	12.8
Germany	9824	66.9	2.1	28.3	0.4	1.2	1.1
Italy	5718	66.0	5.1	1.4	23.3	2.9	1.3
Spain	7975	48.2	3.0	0.5	0.4	45.5	2.4
France	10900	39.4	56.2	1.0	0.6	1.7	1.1
Overall	252443	77.0	7.1	4.2	2.4	4.1	5.2

**Table 1:** Overview of the percentage of Arts & Humanities papers published in English versus other languages per country (in 2008 – 2012), ordered by percentage of English use from most to least. Source: [Scopus](#)



**Figure 2:** Overview of the percentage of papers published in the top five languages per subfield of the Humanities (in 2008 – 2012), ordered by percentage of English use from least (left) to most (right). Source: [Scopus](#)

For the figure below, see: “Citation Analysis for Collection Development: A Comparative Study of Eight Humanities Fields”, Jennifer E. Knievel and Charlene Kellsey, *The Library Quarterly: Information, Community, Policy*, Vol. 75, No. 2 (April 2005), pp. 142-168 Published by: [The University of Chicago Press](#) Article DOI: 10.1086/431331  
Article Stable URL: <http://www.jstor.org/stable/10.1086/431331>

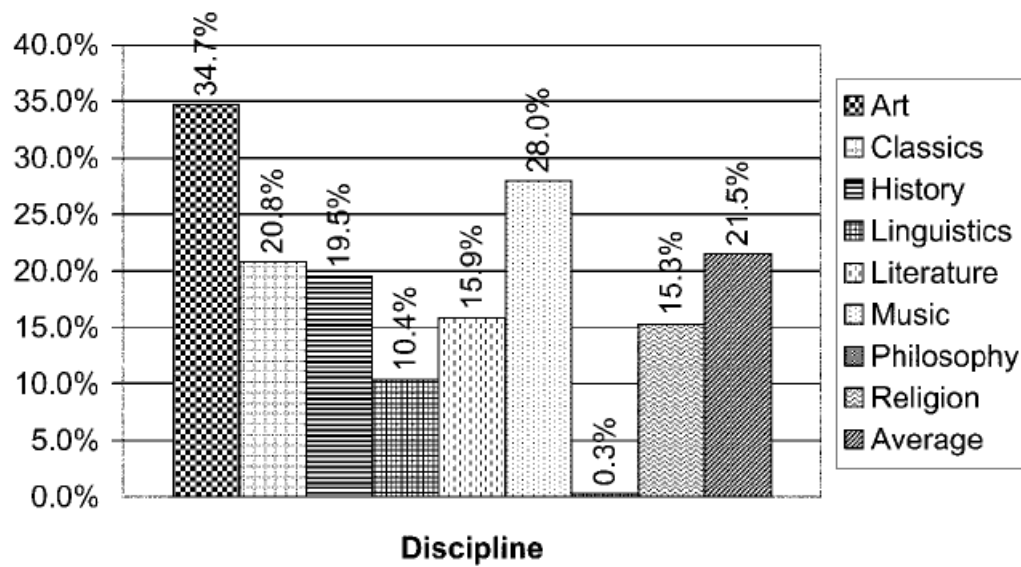


FIG. 1.—Citations to foreign language sources. Total number of citations counted was 9,131. Percentages are of total citations counted in all languages and source types for each discipline and are rounded to the nearest tenth.

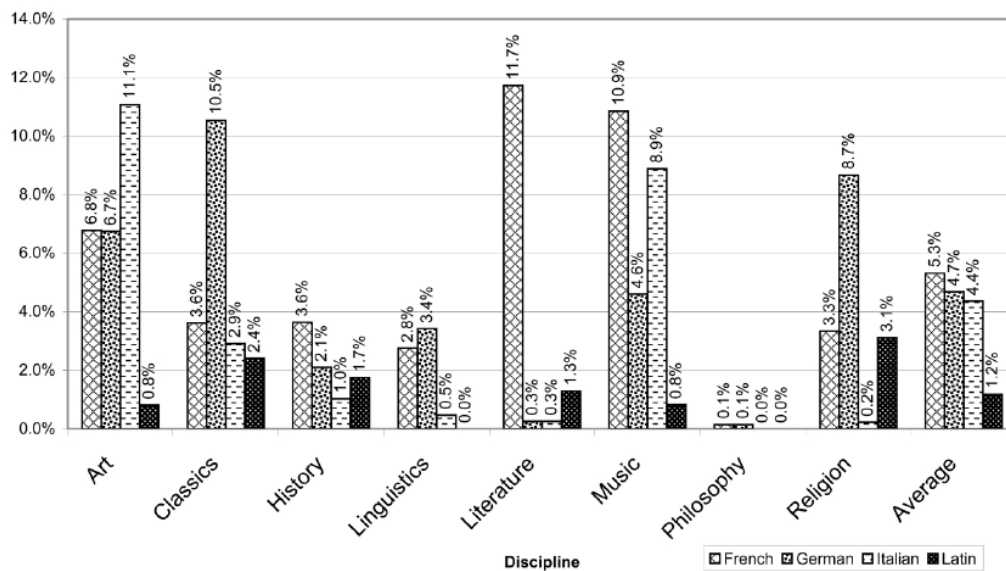


FIG. 2.—Distribution of foreign language citations. Total number of citations counted was 9,131. Languages counted but not displayed were Spanish, Portuguese, and "other." All were too small to be usefully displayed.

From <http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/18598/15201>

Table 4  
Language of Citations

TAPA Volume & Year	Language								Total
	English	German	French	Italian	Latin	Greek	Spanish	Other	
87 (1956) & 88 (1957)	604	413	150	102	44	9	1	4	1327
116 (1986)	955	340	89	35	1	1	0	0	1421
136 (2006)	1583	160	63	58	18	15	5	0	1902
Total	3142	913	302	195	63	25	6	4	4650

“The utilization of foreign language materials has greater implications for collection development and maintenance. In 1956 and 1957, English language materials accounted for only about half of the citations. By 2006, English materials represented 83.2% of all citations. During this time frame, the use of foreign language materials declined precipitously. For example, German language materials declined from 31.1% of all citations to 8.4% and French declined from 11.3% to 3.3%. These changes, however, may be deceiving, since many materials, especially books, may have been translated into English from the other languages in more recent years. Still, these changes do show that scholars in the field of classical philology rely heavily on materials in English. These results mirror those found by Kellsey and Knieval (2004) and Knieval and Kellsey (2005), although the *American Journal of Philology* cited a higher proportion of foreign language materials than did *TAPA*. As a result of such evidence, for many libraries collection development in the field of classical philology should focus primarily on English-language materials, although the evidence also reiterates the need for access to a wide variety of materials in other languages which may be provided through interlibrary loan or databases of foreign-language journals.”

## Culture

Most human engagement with, and discourse about, the human cultural record takes place outside of first world academic channels. Fatwahs from centers of learning such as Al-Azhar University in Egypt and the holy city of Qom in Iran may be the most important scholarly readings of the past but these circulate outside of the Western academic system. China is aggressively expanding its higher education system<sup>1</sup> and humanists must anticipate a new space of intellectual discourse, influenced by, but independent, of the European space in which English predominates and languages such as French, German, Italian and Spanish play a secondary role. The dominant languages of 21st scholarship may become, at least for US Humanities Scholarship, Arabic, Chinese, English, Spanish and Hindi, with languages such as Persian and Japanese emerging as comparable in influence (as they are in numbers of speakers) to German, French, and Italian. In each case, the linguistic challenges must be understood along with often unspoken cultural assumptions about scholarship, authority, civility, and identity.

<sup>1</sup> E.g.,  
[http://www.nytimes.com/2013/01/17/business/chinas-ambitious-goal-for-boom-in-college-graduates.html?\\_r=0](http://www.nytimes.com/2013/01/17/business/chinas-ambitious-goal-for-boom-in-college-graduates.html?_r=0)



The Research Data Alliance can provide a space whereby the fragmented efforts within the US and Europe can connect with each other and with nascent efforts from beyond Europe and North America.

## International Efforts

It is very difficult to form an overall picture of activity in the United States. The Digital Public Library of America (DPLA) and the HathiTrust provide two efforts that engage multiple institutions and are national in scale. The National Science Foundation (NSF) has had profoundly collaborative modes (as in the Digital Libraries Initiative Phase II) but, with its core funding under attack, has now concentrated upon its immediate national and (largely quantitative) social science, with little apparent interest in engaging with humanists. The National Endowment for the Humanities (NEH) struggles with very limited funding, providing short-term (2-3 year) grants that rarely exceed \$350,000 (of which at least \$100,000 usually disappears as overhead) and that cannot normally be extended. The Institute for Museum and Library Sciences (IMLS) has reorganized its funding but it remains limited in what it can do. The Mellon Foundation is the largest player but it is a private foundation, and the strategic impact and directions of its overall efforts can be difficult to track. At the so-called Digital Humanities Summit (in fact, a German-language all-projects meeting for DH efforts funded by the German BMBF), held in February 2015 in Berlin, a presentation on the state of the Digital Humanities internationally summarized US investments with a single slide of NEH funding, using this to illustrate the marginal role of DH development in the US.

We look to **Clarín** for a workflow for integrating many different knowledge sources, applying a comprehensive suite of automated services to each text and thus populating the baseline reading support environment.

We look to **Dariah** as the space in which individuals augment the results of automatically generated annotations, with contributions ranging from micro-publications to hybrid monograph scale works which integrate narrative prose with machine actionable data.

We draw upon mass digitization efforts available from **Big Collections**, such as the **HathiTrust**, **Archive.org**, **Europeana** and the **German Digital Library** as raw materials and as collections that are sufficiently comprehensive as to bring us above the level of isolated, curated docu-islands that may include one, hundreds, or even thousands of books. We need to deal with hundreds of thousands of documents as a minimum -- even as we recognize that millions of documents are available for download and computation (8 million from Archive.org).

In practice, data flows from the Big Collections to a Clarín workflow and then to individual users, who, in turn, collaboratively produce/review new data that flows back into the Big Collection space. NB: This assumes that the Big Collection space can manage annotations

that reference any given word or words and any given regions of interest within any document that they include.

## Use Cases

### Use Case #1 -- Deep Reading a particular source text

A reader wishes to make use of a source text in an unfamiliar language. Readers include professional scholars working on their specialty, native speakers reading an archaic text where the syntax and vocabulary are difficult (English speakers reading Chaucer or, even Shakespeare, examples from German? French?), L2 readers with varying amounts of background (again, an academic may be sufficiently proficient to lecture on Modern History in German but have difficulty with the vocabulary and syntax from essays by Wilhelm von Humboldt. Difficulty includes passages that are simply not understandable but also the laborious task of working through the language).

Some -- but by no means all -- of the reading methods are available in Perseus and/or Alpheios.net. Cross-language (and cross-cultural) reading support tools are essential to any intellectual infrastructure for a global society. The same infrastructure should be designed to support all readers because all readers are novices in most of the world's languages and ignorant of most cultural contexts, while all engaged readers should be assumed to be experts, or potential experts, in domains for which they interest themselves. Nothing is too advanced for the general public. Nothing is too basic and elementary for the advanced researcher -- at least, if the advanced researcher is ambitious and addressing large intellectual problems (as opposed to staying within domains sufficiently narrow and parochial to admit manually acquired expertise).

Example A: Reading 18th or 19th century German (i.e., German that differs from what is taught in modern German classes). Subgoal: Increase the number of those who read \*some\* Goethe and other formative works of German culture in German. [How many students of German in the English speaking world have read any of Faust? There are certainly still some but the numbers are also certainly not very big -- surely not big enough.]

Example B: Reading post-Classical Latin. The DFG has digitized c. 300,000 books produced in Germany through 1800 and more than half of these are reportedly in Latin. The Latin is often going to be very idiosyncratic, whether for reasons of literary genre or because the language is technical. Almost no print reference works exist to help readers with this.

Example C: Reading short passages in historical languages that we have not formally studied. A Grand RDA Challenge: Sheldon Pollock edited a series of essays illustrating World Philology. The Essays are all in English and contain small amounts of transliterated text

illustrating key words and even some short phrases in Sanskrit, Persian, Greek, Latin, Classical Chinese. Goal: Replace and increase the amount of source texts available in these essays, with the knowledge that interested readers can push beyond the opaque surface of the transliterated and un-queryable samples.

We can think of this as automating the production of reading support such as the Leipzig Gloss Rules, developed for linguists who may work with thousands of languages over a career and who prepare corpora knowing that virtually none of their users will be able to study the language that they are documenting. This has been applied to Ancient Egyptian and that work provides a general model for what we need, as well as an existence proof/proof of concept, for historical and literary languages.

## Examples

The following examples exemplify the suggestions in the 2009 article by Daniel A. Werning and Camilla Di Biase-Dyson. Personal preferences of the authors may show the variety of possible adaptations.

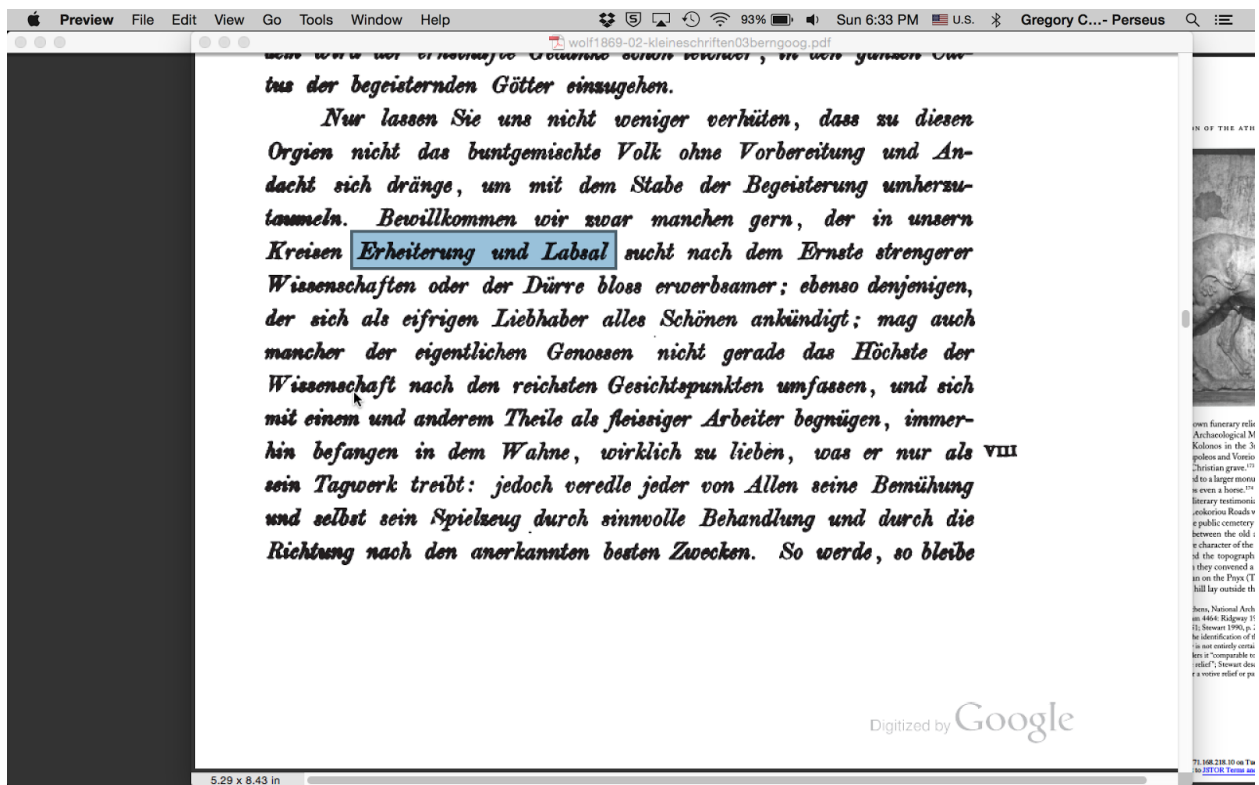
**New Kingdom Middle Egyptian** (Daniel A. Werning, 2009)

Hieroglyphic text witness			
Transcription	$^2m=tn$ $=wj$ $^c q<=j>$ $m=$ $t^2=$ $jmn-t$	$šzp$ $=wj$ $^c(w)-wj=tn$ $r=j$	$m=tn$ $=wj$ $<r>h\{r\}-kw$ $s-t=tn$ $hnt-(i)t$ $dw^2-t$
Glossing	<b>ATTN=2PL=1SG go_in:IPFV[=1SG] in= land(M.SG).STC= west-F.SG</b>	<b>recieve:IMP=1SG arm-M:DU=2PL to=1SG</b>	<b>ATTN=2PL=1SG get_to_know:RES-1SG seat-F.SG=2PL in_front-ADJZ.F underworld-F.SG</b>
Translation	<i>I am entering the land of the west.</i>	<i>Receive me! (Give) me your hands!</i>	<i>I know your place in the underworld.</i>

## Methods:

- Clarin: Standard operations, already available in one form or another, but here tuned to reading support.
  - Historically sensitive dictionary lookups: no one wants to slog through big on-line lexica. Identify archaic words (easy) and, as much as possible, word senses (harder). Point to glosses and to parallel passages. Simply using the German data from Google Books picks up a surprising number of archaic terms and suggests interesting histories for them: e.g., the phrase *Erheiterung und*

*Labsal* in an dedication to Goethe written in 1807 by Friedrich Wolf (1869 edition):



Readers should see IMMEDIATELY where there are obviously old-fashioned terms. The bump in frequency for both these words suggests that they were picked up as archaizing by National Socialists, who emphasized older language.

[https://books.google.com/ngrams/graph?content=Erheiterung%2CLabsal&year\\_start=1800&year\\_end=2000&corpus=20&smoothing=3&share=&direct\\_url=t1%3B%2CErheiterung%3B%2Cc0%3B.t1%3B%2CLabsal%3B%2Cc0](https://books.google.com/ngrams/graph?content=Erheiterung%2CLabsal&year_start=1800&year_end=2000&corpus=20&smoothing=3&share=&direct_url=t1%3B%2CErheiterung%3B%2Cc0%3B.t1%3B%2CLabsal%3B%2Cc0)



We want links to the Grimm lexicon but we want automated data also -- always, and always easy to spot.

- As many aligned Modern Language Translations as possible: Here we draw upon automated alignments at the word/phrase level as a starting point.
- Syntactic analyses: Here we draw upon automated syntactic analysis, providing potentially multiple interpretations from multiple parsers, some perhaps focused on specific issues (e.g., what usages are likely for verb X). Again, anyone ploughing through complex German, especially anyone who is not a native speaker reading text from the eighteenth or early nineteenth century, will have to spend a lot of time parsing out the syntax and will benefit from any clues about how sentences are structured. This sort of support must not be underestimated -- at least if there is a desire for more people to do more work with foreign languages (esp. those with moderately complex syntax).
- Version alignment: readers should always be conscious of how many versions of a work are available and should be able to see, as clearly as possible, how often, where, and (if desired) how different versions differ from each other.
- Text Reuse: any quotations from one text to another should be identified and linked to the source texts, e.g. from the same text,

und mit eigenen Schätzen bereichern mögen. So kam er denn in den Fall, die Worte Quintilians auf sich anwenden zu dürfen: „*Quantum notando consequi potuerant, interceptum boni iuvenes, sed nimium amantes mei, temerario 8 editionis honore vulgaverant.*“

Quintilian is, at least, named, but we are not given a citation. This quote should be linked to the Latin text of Quintilian -- AND THEN TO DIFFERENT MODERN LANGUAGE XLATIONS and to the reading support tools for Quintilian so that readers can puzzle through the point that Wolf is making.

- Automated Named entity support: for any given name, industry services such as Google now allow for fairly efficient disambiguation and readers can often hunt down info about people and places -- but readers need to visualize our best automatic analyses of the people, places, dates, etc. that show up in docs as a whole, as well as the relationships between them. This goes a bit beyond 4D geotemporal relationship browsers.

## Use Case #2: Editing/extending the results/services in Use Case #1

Users need to be able to refine the results from automated processes, generating new training data and, ideally, creating new services.

## Use Case #3: Tracing the Circulation of Ideas/Texts across time, space, and language

The starting point for all work must be generalizing the methods developed for, and demonstrated at, <http://books.cs.umass.edu/mellon/>:

**Language Detection:** a fundamental technology.

<http://books.cs.umass.edu/mellon/langid.html> -- far better data than library metadata:

[http://books.cs.umass.edu/mellon/langid\\_diff.html](http://books.cs.umass.edu/mellon/langid_diff.html)

**Duplicate Detection:** <http://books.cs.umass.edu/mellon/duplicates.html> -- Vergil is still more important (at least more frequently reprinted) than Shakespeare. Again, this is a fundamental function and cannot be accomplished with book-level library metadata.

**Duplicate Alignment:** <http://books.cs.umass.edu/mellon/alignment.html> -- This is essential for comparing versions, seeing what part of a work get anthologized etc. If you need to think about textual history or how a text appeared at a given time, you need this service.

**Quotation Detection:** <http://books.cs.umass.edu/mellon/quotes.html> -- What parts of particular texts get reused and where? This needs to be combined with translation alignment (e.g., we want to see what part of the Latin vulgate vs. Luther's Bible vs. the King James are cited in Protestant/Catholic work over time)

**Named Entity Recognition** (at scale and for noisy data):  
<http://books.cs.umass.edu/mellon/ner.html>

**Viral texts:** <http://viraltxts.org/> -- here there are no canonical texts. We compare everything against everything to find what does, in fact, get reprinted. Readers should be able to see what parts of a document/collection are most often quoted. Here the problem is harder (and the results will not be as exhaustive) because we are not comparing one text to N other texts (as when we look for passages that one texts shares with other texts), but N texts to N-1 texts.

**Multilingual topic models:** These are absolutely fundamental tools for scholarship and must be supported as broadly as possible. The following figure illustrates, for example, topics automatically detected in English and German in two Classics journals -- they immediately demonstrate differences in coverage between the journals and national communities of scholarship.

VII: Aligned topics in English (AJP), in dark gray extending upwards, and German (Hermes), in light gray extending downwards. Lines are at 1920, 1940, 1960, 1980 and 2000.

case languages language example nominative english genitive means cases object dative subject grammatical accusative article finnish like noun feminine

sanskrit self vedic mind buddhist three nature buddhism hindu type means through buddha reality meditation practice veda seeing compound

linguistics context words use specific type example particular meanings linguistic same terms different lexical analysis sound function literal within

rhetoric demosthenes cicero against speeches isocrates speech orator others style public along should teacher important private orators ten funeral law legal court case criminal laws under right rule principle will act crime civil cases decision parties judge courts

she daughter wife married mother sister marriage husband child children became queen gave bore woman named father birth herself

women woman men man female young male considered famous seen than sexual feminine beautiful role make reports sex parallel

love sexual desire beauty well because aphrodite character relationship eros another where describes twins part

sprachen sprache kasus beispiel fall akkusativ haus genitiv grammatik deutschen dativ latein beispiele kennt deutsch person ausdrücken präposition regel

sanskrit drei vier buddhismus indischen lehre bedeutet prakriti purusha dukkha wesen hinduismus zustand buddhistischen buddha meditation yoga

bedeutung zwischen beispiel verschiedene linguistik also beispielsweise wort verschiedenen zeichen lassen wörtern kontext sprachwissenschaft englisch gegenstand siehe unterscheidung spricht

cicero rede rhetorik reden redner allem ersten demosthenes ciceros bedeutung tod prozess seinem letzte wobei damit rhetor auseinandersetzung meidias recht gesetz wenn also lat law keine ohne wegen liegt tat non sog strafe gemäß deutschen grundsatz vertrag deutschland

tochter mutter ihr frau schwester vater ihrem ihre ihrer ihres heiratete ehe gattin kind verheiratet ihren kinder ehefrau geboren

frauen männer gehört mädchen männlichen frau anderen personen jungen junge ihre weiblichen mann männern bestand tritt bringen weiblicher männliche

liebe gilt menschen beziehung sexuelle zurück immer eher sehen geprägt personifikation hauptsächlich zuneigung

