User Interface Challenges
for the Study of Folklore and Literature

Timothy R. Tangherlini (UCLA)
Some key problems

• Design cycle
  o Many Humanities projects have a very short funding cycle, with very limited funds ($3000 is not an unusual amount)
  o Limits the type of iteration through design that could converge on user centered interfaces
    • Humanities scholar might be an excellent scholar but
    • Not an excellent designer and
    • Not an excellent data scientist

• How do we shift from analog interaction with highly curated data (e.g. reading a novel) to computationally mediated interaction with uncurated data (e.g. trying to find passages from millions of books in HathiTrust)?
Four Challenges in Folklore and Literature

• Collections and Archives
• Indexing and Classification
  o curation
• Visualization and Navigation
• Analysis
The Folklore Macroscope

WitchHunter, TrollFinder, TreasureX, ElfYelp

With Peter Broadwell (UCLA)
& James Abello (Rutgers), and David Mimno (Cornell)
Macroscopes

- Macroscopes provide a ‘vision of the whole,’ helping us ‘synthesize’ the related elements and detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, macroscopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend

-Katy Börner (2011)

Folklore Macrooscope Tools

The Danish Folklore Nexus (Tangherlini and Broadwell 2013)
A curated study environment, included with Danish Folktales, Legends, and Other Stories

ETKspace (Broadwell and Tangherlini 2014)
A faceted browsing interface to the full Tang Kristensen corpus

WitchHunter (Tangherlini and Broadwell 2015)
Devised for geo-semantic mapping and pattern discovery in geo-located folklore

GhostScope (Broadwell and Tangherlini 2015)
Conceptual mapping of storytellers’ mental geographies

ElfYelp (Tangherlini and Broadwell 2015)
A tool for geo-topic modeling and exploration
DFL 1.0: Classification and Navigation in a Curated Corpus

- Navigation must “add value”
- How do we leverage the huge amount of work that we have done in acquiring, organizing and providing dynamic labels for our research materials?
Danish Folklore Nexus

The oldest parish clerk that I know of in Svejstrup was Niels Ring. He was first a servant for the manor lord at Venge farm and he appointed him to the position. He was so harsh that once when he punished a young girl, he lifted up her clothes and but her bare bottom up on top of a glowing wood-burning stove for so long that her skin was half burned off when she came down. But then he got too old and the children teased him, they filled his chair with pins.

The next parish clerk was Søren Jul. He was a clever man, but strict, and when he got angry he'd throw all the boys together in a heap and then he'd walk all over them with his leather-soled dogs and then he'd yell, "You Goddamn wolf children!" But regardless how strict he was, they still played tricks, they did those boys. One of them was called Mads, he'd been born up at his great grandmother's house. So one day he'd gotten off to a bit of a late start, and when he got to school he wasn't even finished with his breakfast, since he arrived chewing on a cheese rind and a piece of bread. When he gets there, the parish clerk had just started the lessons. Then his desk mate, Niels Rebslæer, sees his chance and while Mads is sitting there chewing on the cheese rind, he reaches out and shoves it down his throat. Now Mads was sitting there and couldn't get the rind either up or down. The other children began to laugh and the parish clerk couldn't control them. "What is going on?" he asks them gruffly. Immediately, Niels Rebslæer jumps up and says, "It's Mads, he's ingenious!" Regardless of how strict he was, the parish clerk couldn't help laughing, but they both had to go and get spanked anyway.
Novel Navigation

Related stories based on multi-dimensional classification
James Abello’s *GraphView*
Drill-down
Benefits

Downloads

If you are unable to install the Danish Folklore Nexus from the disk included with your bound volume, you can download a copy of the disk [here](https://etkspace.scandinavian.ucla.edu/danishfolklore/updates/default.html).

The copy you download will be an iso file. To use this file, you will need to mount it on your system. Download it to your desktop.

On Mac, right click the iso file, DFL.iso, and open it with the DiskImageMouter application. The DFL disk will now appear as a disk on your desktop. Double click on the DFL disk and install the Danish Folklore Nexus.

On Windows 8, right click on the iso file, and select "mount". The DFL disk will now be mounted as a DVD on your system. Open it from Explorer, and install the Danish Folklore Nexus from there.

On Windows XP, Windows Vista, and Windows 7, you will need to download the Virtual CD-Rom Control Panel from the Microsoft Download center ([here](https://etkspace.scandinavian.ucla.edu/danishfolklore/updates/default.html)) and follow the instructions for mounting a virtual drive. Once mounted, install the Danish Folklore Nexus from the mounted disk.

Updates

Any updates for the Danish Folktales, Legends, and Other Stories Digital Content will be listed below as it becomes available.
Disadvantages

- Relatively fixed content
- Predetermined pathways through curated content
- Limited to curated content
(Un)curated Collections

- Archives (or libraries) are often curated at a fairly high level
  - Metadata
  - Standard classifications

- Often have limited means for discovery
  - Focus on precision

- Goal of scholar often not easily served by these methods to access collections
  - Scholar often wants to generate a (dynamic) research corpus
The foods served at this particular banquet would be considered unusual since they constitute everyday fare and not particularly exciting everyday fare at that. In this case, the poor food makes the celebration memorable. Kirsten Marie had previously told a story about another banquet that was memorable, in that case because the hosts should not have been able to afford an elaborate celebration (KMP 2.07). In both cases, she emphasizes the violation of expectations. In the published version, Tang Kristensen has separated these descriptions and, in so doing, erased the continuity of the descriptions and the clear sense of expectations for banquets.

**English publication text**: At one big birth celebration, they weren't served anything but (buckwheat) gruel and fish as far as I can remember.
Repurposing other software

• Recognition that there are certain interfaces that our users already understand and with which they are comfortable
• Recognition that we do not have support to design ground-up interface
• Recognition that a “one-off” solution does not scale well to other projects
• Challenge: Can we adopt other interfaces for Humanities challenges?
A sample search

Churches and their construction. Monasteries, holy springs, etc. (and) About the Devil and being in league with him (2318 results)

Export metadata as: XML, CSV, Plain Text
Drill-down and export
Advantages of DFL 2.0

• Selection of a subset of the larger corpus
  o Subset can be
    • subjected to other types of analysis (e.g. GIS)
    • packaged for use in DFL 1.0

• User defined path

• Addressable searches
  o Allows for rapid reproduction of searches

• Extensions
  o “features” added to each of the main categories
    • Eg. add gender, occupation, dob to “people” category
    • Information derived from other collections
  o Use map generated bounding boxes to create a search
Disadvantages

- Fitting a collection to the interface
- Faceted browsing presupposes that there is some fundamental unit (here the story) that is the goal of a search
Other Closely Related Approaches

• Dutch folklore archive and beta geo-focused interface
  - fast / Solr index
  - aesthetically pleasing
  - geography is foregrounded
  - not easy to generate portable research subset
  - not clear what is being mapped

• Sagnagrunnar, Icelandic legend database
  - fast
  - geography is foregrounded / Google maps
  - multiple visualizations
  - idiosyncratic classification
  - no way to generate research collection
Accessing Analytics

- Many current projects in folklore focus on access to archival materials or visualizations generated on the collection as a whole.
- Some of our work has focused on creating a series of tools to access analyses of aspects of the collection in whole or in part.
- “Operationalizing Distant Reading”…
ElfYelp

Information about the selected region

Mound dwellers, ghosts and Satan's influence (2.22%):
- Mound dwellers (Hidden folk): Cassation of the destruction of a mound (2.09%)
- Legends about farms and towns: Old named villages (2.02%)
- Mound dwellers (Hidden folk): Changelings, the old child (1.82%)
- Our forebears' way of thinking and spiritual life: Blad ministers (1.9%)  
- Heroes and their sport: Giants build churches (1.79%)
- Heroes and their sport: Giants throw stones at churches (1.69%)
- Lights and portents: Lights for buildings, etc. Moving (1.6%)
- Outdoor life: Anecdotes from named rakers' lives (1.57%)
- Manor lords, ladies and mistresses: Blood spots (1.55%)
- Manor lords, ladies and mistresses: Wailed in (1.36%)
- Female revenants: The murdered child, mother with the knife and washing it or the clothes (1.26%)
- Heroes and their sport: Dams (1.21%)
- About the Devil and being in league with him: Black dogs and the like show themselves (1.2%)
- Mound dwellers (Hidden folk): Smiths in mounds (1.17%)
- Mound dwellers (Hidden folk): Fording of the hidden folk (1.16%)

Things that happen in fields (1.24%):
- Heroes and their sport: Stones and earth thrown at churches, builders disagree (4.76%)
- From the time of village: Tenancy and purchase of farms (4.21%)
- Small kings and their feasts: Kings. Enemy invasions: Sweden north of Skanderborg and Silkeborg (3.59%)
- Diverse hauntings and conjuring of revenants: Black shapes (3.51%)
- Diverse hauntings and conjuring of revenants: Sheriffs, bailiffs, etc (3.17%)
- Our forebears' way of thinking and spiritual life: Confirmation preparations (3.07%)
- Female revenants: In with maidens (2.89%)
- Diverse place legends: Odin, and sailways (2.79%)
- Religious legends: The Syll's prophecy (2.75%)
- Cunning men and women and their activities: Wise Norwegians (2.68%)
- Traveling monsters: Odin's hunters go through a house or farm (2.47%)
- Outdoor life: Floods and sand movement (2.25%)
- Female revenants: Old hags (2.16%)
- Mound dwellers (Hidden folk): Hidden folks' silver and gold (2.11%)
- Diverse place legends: Sand movements (2.06%)

Liminal areas and people or boundaries (0.57%):
- Churches and their construction. Monasteries, holy springs, etc.: Virgins build churches (5.99%)
- Outdoor life: Road- and postal service (2.73%)
- Diverse place legends: Old forests, Viborg (2.23%)
- Elves: The dance of the elves (2.19%)
- Diverse place legends: Sounds, fjords, and sailways (2.18%)
- Social gatherings and parties: Birth and churching women, going to the altar, etc (2.18%)
- Witches and their sport: Witch burned on the pyre (2.04%)
- Diverse place legends: Sand movements (1.64%)
- Outdoor life: Various livelihoods (1.61%)
- Our forebears' way of thinking and spiritual life: Landscape descriptions (1.55%)

Normalized by global frequency:

1. brak: 0.85 (6 times in region, 12 times globally)
2. turbe: 0.5 (4 times in region, 8 times globally)
3. langa: 0.4 (2 times in region, 5 times globally)
4. den: 0.4 (2 times in region, 5 times globally)
5. sile: 0.39 (4 times in region, 11 times globally)
6. gje: 0.33 (1 times in region, 3 times globally)
7. sol: 0.31 (5 times in region, 16 times globally)
8. lase: 0.31 (8 times in region, 28 times globally)
9. snabe: 0.3 (5 times in region, 10 times globally)
10. kuskas: 0.29 (2 times in region, 7 times globally)
Teaching Folklore

Getting Students to Think Like a Folklorist through Interface Design
The Online Korean / Korean American Folklore Archive

- Conceptualized as a student driven project
- Intends to teach about
  - Fieldwork
  - Archiving
  - Human Subjects Protection
- While generating an anonymized, usable dynamic dataset for the study of Korean American traditional expressive forms
- Two Goals: Pedagogy and Research
Design drivers

• “Folklore is traditional expressive forms circulating in and across social networks embedded in time and place”
  o Implications:
    • Collectors
    • Informants (Consultants)
    • Contexts
    • Field data

• Study of folklore is based on the analysis of many variants
  o Implications
    • Classification
    • Historic-geographic search
Science

Results (Crawl by date of collection, most recent first)

Collector
Male, 28, College, student, english, korean

Consultant
Female, 25, College, student, English, Korean, Korean, LA, CA, USA

Context
January 1, 2012, afternoon, Sunny
Business, Los angeles, ca, usa

Data
Fieldnotes
She did not know where the story originated, however, her version of the story was more based on science rather than myth or superstition. During the deathly hot summer of her younger years in Korea, her mother had informed her about not leaving the fan on directly towards the face. If she had, she would die. The oxygen would be repelled away from the nose, thus the person would suffocate and die. When asked about leaving it on in another area of the room, she also stated that the door or window needed to be open to circulate oxygen.
The Folklore Collector
Dashboard

DASHBOARD

Collector Profile
Timothy Tangharlo
Email: tango@hurrel.uc.edu
Address: 3456, CA 90024 USA
DOB: 1993-10-20
Age: 417
Gender: Male
Marital Status: married
Occupation: used care salesman
Educational Level: No Schooling
Heritage: European
Languages Spoken: English, Danish, Korean, and all other languages

Contexts
TT Birthday Party 1/16/2014
Date: 2014-01-16
Time: after noon
Description: Tommy's birthday party took place in abandoned vel

Test
Date: 0000-00-00
Time: after noon
Description: one two three

Consultants
Tommy Tran
Jihoon Kim
Donald Trump

Data
Hardcore Gang Birthday Parties Description: TT: What did you think of the party? JH: Pretty rad. TT: I was scared, really, most of the time.
Filename: Screen Shot 2014-01-13 at 8.12.43 PM.png

Hardcore Gang Birthday Parties Description: Photograph of rickety stairs leading to party hall...
Filename: Screen Shot 2014-01-13 at 8.25.56 PM.png

Hardcore Gang Birthday Parties Description: blah blah
Filename:

Test Description: hello
Filename: 02_loc_final.pdf

Delete
Delete
Delete
Delete
Delete
Delete

Delete
Delete
Delete
Delete
Delete
Delete
Delete

Collectors

**COLLECTOR PROFILE**

<table>
<thead>
<tr>
<th>Last Name: *</th>
<th>First Name: *</th>
<th>Initial:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tangharini</td>
<td>Timothy</td>
<td>R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Email Address: *</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:teng@tumnet.utsa.edu">teng@tumnet.utsa.edu</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Street Address:</th>
<th>City:</th>
<th>State/Province:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roebing Ave</td>
<td>Los Angeles</td>
<td>CA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Zip/Postal Code:</th>
<th>Country:</th>
</tr>
</thead>
<tbody>
<tr>
<td>92024</td>
<td>USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DOB (yyyy/mm/dd): *</th>
<th>Age:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980-10-02</td>
<td>41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender: *</th>
<th>Marital status: *</th>
<th>Occupation: *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>married</td>
<td>used car salesman</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Educational Level:</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Schooling</td>
</tr>
<tr>
<td>High School</td>
</tr>
<tr>
<td>College</td>
</tr>
<tr>
<td>Middle School</td>
</tr>
<tr>
<td>Professional Degree</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Heritage:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
</tr>
<tr>
<td>Hindu</td>
</tr>
<tr>
<td>Taiwanese</td>
</tr>
<tr>
<td>Native American</td>
</tr>
<tr>
<td>South Asian</td>
</tr>
<tr>
<td>South African</td>
</tr>
<tr>
<td>South American</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Languages spoken: *</th>
</tr>
</thead>
<tbody>
<tr>
<td>English, Danish, Korean, and all other languages</td>
</tr>
</tbody>
</table>
Consultants  
(not informants)

<table>
<thead>
<tr>
<th>Last Name:</th>
<th>Tran</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Name:</td>
<td>Tommy</td>
</tr>
<tr>
<td>Initial:</td>
<td></td>
</tr>
<tr>
<td>Show Level:</td>
<td>None</td>
</tr>
<tr>
<td>Street Address:</td>
<td>1220 Levering Ave</td>
</tr>
<tr>
<td>City:</td>
<td>Los Angeles</td>
</tr>
<tr>
<td>State/Province:</td>
<td>CA</td>
</tr>
<tr>
<td>Zip/Postal Code:</td>
<td>90024</td>
</tr>
<tr>
<td>Country:</td>
<td>USA</td>
</tr>
<tr>
<td>Date of Birth (yyyy/mm/dd):</td>
<td>1997-02-04</td>
</tr>
<tr>
<td>Place of Birth:</td>
<td>Onju City</td>
</tr>
<tr>
<td>State/Province:</td>
<td>Korea</td>
</tr>
<tr>
<td>Country:</td>
<td>Korea</td>
</tr>
<tr>
<td>Gender:</td>
<td>Male</td>
</tr>
<tr>
<td>Marital status:</td>
<td>Single</td>
</tr>
<tr>
<td>Occupation:</td>
<td>Warehouse Wrangler</td>
</tr>
<tr>
<td>Educational Level:</td>
<td>Graduate Degree</td>
</tr>
<tr>
<td>Income Level:</td>
<td>0 - 10,000</td>
</tr>
<tr>
<td>Heritage:</td>
<td>Korean</td>
</tr>
<tr>
<td>Languages spoken:</td>
<td>English</td>
</tr>
<tr>
<td>Did you immigrate to the United States?</td>
<td>Yes</td>
</tr>
<tr>
<td>Date of Immigration (yyyy/mm/dd):</td>
<td>1990-12-09</td>
</tr>
</tbody>
</table>

Uploaded Consent Form:  
File Name: 20160331_090332.jpg  
File Type: jpeg  
File Size: 10924640 Bytes  
Submit  Cancel
This interview took place in a residence in the afternoon on an overcast day which smelled of imminent rain.

The residence was enveloped in a soft blanket of silence. Considering that it was a holiday, the streets were fairly silent, except for the appearance of a random ice cream truck blaring out promises of delicious sweets to lure children into buying mediocre ice cream. The occasional greetings of people visiting the neighbors' houses could be heard from the distance.
Collector: Can you explain what fan death is?  
M: If you keep the fan on while you sleep, you can die because the fan lowers your body temperature.

Collector: As a person who studied science, do you think this theory has any scientific basis?  
M: I'm not sure. I've heard stuff on the news but I personally don't know anyone who has been hurt from it.
Summary Challenges

- Folklore collections are often organized according to limiting indexing systems or idiosyncratic organizations.
- Goal is to be able to (a) navigate large collections based on researcher determined criteria and (b) discover patterns at scale that can lead to (c) the selection of resources for further examination.
- Intuitive interfaces that operate at human-interactive timescales.
- Making apparent the underlying organizational schema to help users infer limitations.
Sub-Corpus Topic Modeling

Challenges for Search and Discovery in the Study of Literature

With Peter Leonard (Yale)
General flow

Sub-Corpus Topic Modeling (STM)

1 Sub-corpus Selection

2 Chunking

3 Sub-corpus Modeling (LDA)

4 STM curation

5 STM based search

6 Expert input / refinement
STM 1.0 Dashboard:

Topic curation

Keyword word cloud
Topic labeling
n-gram word cloud

Ranked saturated chunks
Chunk source work
Chunk saturation
Identified chunk

Topic Modeling Evald Tang Kristensen's Danish Folklore

Exploring topics in detail

Word clouds
Chunks per year
Related topics
Forgotten Works:
Ingvor Bondesen
Extensions: Finding Thor

Topic Modeling Old Norse

Edda Snorra Sturlusonar

Source: 50 20
Stopped: Target: Heimskringla

Zero words:

porr jatti
porr hafra
hefir verit
segja stóð porr
rá porr reidr

Zero phrases:

porr jatti
porr hafra
hefir verit
segja stóð porr
rá porr reidr

Zero chunks:

Olaf's konungs byrjar skemtar:
Hraldur saga hins hafsaga Chunk 23
Hraldur saga hins hafsaga Chunk 73
Hraldur konungi graed 17
Hraldur saga hins hafsaga Chunk 6
Magnus konungs

Topic #1 words:

jarl laetr
of log
vísí gefr

Topic #1 phrases:

jarl laetr
of log
vísí gefr

Topic #1 chunks:

Magnus konungs berfts Chunk 30
Magnus binda Chunk 22
Hraldur saga hins hafsaga Chunk 18
Hraldur konungi graed Chunk 35

Jarlegnir Olaf's konungs. Sa atburður gerðisk í Niðarósi, at skrín Olaf's konungs var borit um stræti, at hofugt varð skrínit svá at eigi fengu menn borit frá or stæð; en síðan var skrínit niðr sett ok brott upp strætít ok leit, hvat þar var undir, ok fánzí þar barns lík, er myrt hafði verit ok fólgít þar; var þat þá á brot borít, en þeir aprt strætít, svá sem þær hafði verit, en borít skrín at vanda.
STM 2.0

• Our original STM machine was a proof of concept machine
  o Finely tuned to the target corpus
  o Preprocessing and modeling done offline and refined multiple times
  o It provided excellent and unexpected results:
    • Discovered discourse in the social science literature that reflected the growing importance of Social Darwinism in late 19th and early 20th century Denmark
    • Provided an interesting method to confirm the findings of Pil Dahlerup in her mid 1980s “Det moderne gennembruds kvinder”, a rejoinder to the Georg Brandes’s classic, “Det moderne gennembruds mænd”
Motivating question(s):

• How can we make this
• Broadly available,
• Flexible to the heterogeneous textual data with which Humanities researchers work,
• Provide results interpretable by people new to the field of computational Humanities
Multi-step process

- Upload
  - Corpus
  - Sub-corpus
- Preprocessing
- Modeling
- Curation
  - Labeling
  - Hiding topics
  - Multi-level modeling (differing levels of granularity)
- Exploration
  - Visualization
Upload

- Devise a corpus or subcorpus
- Upload it to the site
Corpus upload

STM, Scandinavian Section, UCLA

Manage Corpus  Topic Modeling

Metadata

- Corpus
- Subcorpus

niels_lyhne

Add Files (929 files in corpus)

- Drop files here to upload
- Already Preprocessed

Delete Corpus

Delete

Corpus name
Add files
Delete existing corpus
Topic browsing og curation

Valg af modellets granularitet
Preprocess

- Chunking (currently off-line)
  - By paragraph
    - Our original STM approach
    - See also Algee-Hewitt, Heuser and Moretti (2015)
  - By page
  - By chapter
  - User defined chunking
- Stop-words
- Tokenizing and/or stemming
Preprocess

Choice of regex tokenization

Choice of stopword files
Model the corpus

- Model the corpus
  - Choose level of granularity
Modeling
(currently LDA, plan to add other topic models)
Infer topics

• On the main corpus based on the subcorpus
• Standard topic modeling—
  o Make the subcorpus = main corpus
Build model

Choose corpus

Choose modeling
Curation

- Labeling and hiding topics
Multi-level topic curation

- **Topic granularity**
- **Show / hide topics**
Topic browsing and search

Top Phrases

falskhed forstillelse

- højt loftet
- sveget ideal haanlig
- elskov kjærlighed
- mandens ideal
- ædlere end elskede
- ydmyge beje kneæ
- elskede hele
- siden busten
- hvem mødt
- hele sjæl igjennem
- al glans al renheds
- muligt end
- forglemmelse selv stolthed ydmyghed
- egnes skjønhed
- derunder selv
- mod dens
- fremmed grund
- sorg langt
- siden gardinet

Related Topics

- alle, selv, samme, ...
of niels_lyhne_stopped (corpus) (50 topics)

- ret, ganske, gamle, ...
of niels_lyhne_stopped (corpus) (50 topics)

- laa, meget, graa, ...
of niels_lyhne_stopped (corpus) (50 topics)

- blev, saae, efter, ...
of niels_lyhne_stopped (corpus) (50 topics)

- haand, ord, hen, ...
of niels_lyhne_stopped (corpus) (50 topics)

- blev, mellem, tale, ...
of niels_lyhne_stopped (corpus) (50 topics)

- sige, vilde, maa, ...
of niels_lyhne_stopped (corpus) (50 topics)
Navigation and Discovery

• Topic saturation per chunk
• Discovering similar chunks
• Network representation
Han følte, at det i hendes Sjæl, som var Sjæl for det Yppige, Blussende, Sandseligt-bløde i hendes Skjenhed, vilde aldrig lade sig drage til ham, det vilde ingensinde tage ham i Favn med disse blændende Junoarme, al Evighed elskovssvagt til denne vellystaandende Nakke hans Kys i Vold. Han saae det godt, den unge Pige hende kunde han vinde, havde han vel vundet, og hun, den Yppige, det troed han forvist, hun havde følt, hvo den ærlie Skjenhed, der var den i hende, mystisk havde rørt sig i sin levende Grav, for at favne ham med slan Jomfruarme, møde ham med bange Jomfrulæber. Men hans Elskov var ikke den. Han elskede kun det, der i stod til at vinde, elskede just denne Nakke med dens varme Blomsterhvidhed og dens Skjær af duggende Grunde under det dunkle Hær. Han hulked i Eiskovsavn og vred sine Hænder i higende Afmagt, han slynged sine Ar under et Træ, læned sin Kind mod dets Bark og græd.
Networks of topics
Ongoing challenges

• Add user login and profiles and sharing so that STM can be a private / collaborative space
• Move the computation into the cloud, so that creating models is not as time intensive as it currently is
• Make large collections freely available for:
  o Subcorpus selection and creation
  o Corpus search
Beta version of Search/Selection

Danish Books 1860-1870

73 Books filtered from 1605 originally (Reset All Filters)

sorted by: labels; then by... • ○ grouped as sorted
1 • 2 • 3 • 4 Next »

Breve fra og til Johan Ludvig Heiberg
By Johan L. Heiberg
Reitzel, 1862. 240 pages

Breve fra og til Johan Ludvig Heiberg
By Johan Ludvig Heiberg and Thomasine Gyldenloung
C.A. Reitzels forlag, 1862. 240 pages

Johan Ludvig Heiberg's poetiske skrifter
By Johan Ludvig Heiberg
C.A. Reitzel, 1862. 496 pages

Publication Date

11

Type an Author's name...

Type a Publisher...

Type a Subject...

...or select from a list

60 Johan Ludvig Heiberg
12 Johan L. Heiberg
1 Fiske Icelandic Collection
1 Peder Andreas Christian Heiberg
1 Thomasine Gyldenloung

24 C. A. Reitzel
18 Reitzel
12 C.A. Reitzel.
5 C. A. Reitzel.
1 C.A. Reitzels forlag
1 P.G. Philipsen
1 Reitzels Forlag
1 W. Prior

...or select from a list

17 History / General
15 Danish literature
15 Drama
15 Literature
11 Drama / English, Irish, Scottish, Welsh
11 Drama / General
1 Diatoms
1 Mythology, Norse
1 Performing Arts / Theater / General
1 Performing Arts / Theater / Stagecraft
1 Poets, Danish
1 Theaters
Challenges in the representation of Old Norse-Icelandic

IceMorph
http://tango.bol.ucla.edu/andreas09/index.html

Work with Kryztof Urban (DataCrush), Aurelijus Vijunas (Natl Kaohsiung Univ), Peter M Broadwell
ICEMorph: Download

ICEMorph is a morphological analysis and look-up tool for the study of Old Norse-Icelandic, among the most morphologically complex of the Germanic languages.
The analysis tool uses a functional programming language, FM/Haskell, to tackle the problem of this complexity. Machine learning algorithms boost the performance of the system.
The look-up tool is based on two of the most important Old Icelandic-English dictionaries: Cleasby-Vigfusson An Icelandic-English Dictionary (1874), and Zoëga Old Icelandic Dictionary (1910).

Download

You can download the latest version of the tagged corpus below.
ICEMorph and the tagged corpus is covered by the Creative Commons Open Archives non-commercial "by" license. The current citation for ICEMorph and the tagged corpus is:

The tagged texts are based on:

Click on the link below to download the current tagged corpus.
Fornaldarsögur Norðurlanda
Why is this hard?

• Old Icelandic is *highly inflected*
  o Making search quite difficult
  o Making learning it quite difficult

• Old Icelandic corpus is normalized
  o To different normalization standards
  o Scholars often want to work with variant readings

• Challenge:
  o Can we devise tools to make accurate/sophisticate research a bit easier
  o Can we help students learn and give non-experts access
Vowel Changes: Umlaut and Breaking

A large part of the complexity of Old Icelandic (OIC) morphology is due to activity of the so-called “umlauts” and “breakings”. Whereas the previous term is merely a synonym of “vowel assimilation”, used in a more narrow sense, the interpretation of the latter term, “breaking”, has been a matter of debate, and the final solution has yet to be determined.

Dissimilation

Dissimilation is fairly rare in Icelandic, yet it has contributed to the development of several irregular paradigms, cf. below:

<table>
<thead>
<tr>
<th>maðr ‘man’ (sg.)</th>
<th>teðr/tennr ‘teeth’ (pl.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom. maðr -- *mann-r</td>
<td>teðr/tennr: teðr -- *tenn-r, whereas the by-form tennr is analogical</td>
</tr>
<tr>
<td>Acc. mann</td>
<td>teðr/tennr</td>
</tr>
<tr>
<td>Dat. manni</td>
<td>to. nnum</td>
</tr>
<tr>
<td>Gen. man(n)</td>
<td>tanna</td>
</tr>
</tbody>
</table>

Degemination

Degemination occurs commonly in consonant clusters of the type -C1C1C2(C3)-, cf. sékkva ‘sink’ (transitive v.) vs. sékpa ‘id.’ (1. p. sg. preterite), or spenna ‘to make tense’ vs. spenta ‘id.’ (1. p. sg. pret.), etc. In numerous instances geminates are restored contrary to what seems to be a phonological reduction rule, cf. sakkta, spennta, etc. The reason for the restoration of geminates is primarily analogy, to be discussed in 2. below. In certain instances of a different type, cf. leggja ‘lay’ (inf.) or ‘legs’ (gen. pl.), bekkja ‘benches’ (gen. pl.), etc., the spelling would seem to be calling for degemination. However, it does not happen, because the consonant j in such verbs only denotes the frontness of the preceding stop /g/ : /j/, and does not represent an individual phoneme /j/.

Devoicing in word final position

Devoicing in word-final position has led to certain irregularities as well. The consonants that were affected more frequently than others were d, g and y < ð. The consonant ð in absolute final position would become /h/ and disappear, engendering irregular development of the preceding mono- and diphthongs, cf. the development of the 1./3. p. sg. preterite of the verb stiga ‘step’ (*stiyan): Proto-Scandinavian *stály -- *stáyl -- *sté -- OIC. sté. Otherwise, the diphthong *al usually developed into OIC. eí, cf. *bálal ‘waited’ -- OIC. bálí, *skáll ‘shined’ -- OIC. skéinn, etc. It also developed into ei in those word-forms, where ð was not final, e.g. 2. p. sg. steig-t -- *stáyl-t. However, due to systemic pressure (“Systemzwang”) a new phonologically irregular by-form steig was formed both for the 1. and the 3. persons.

The disappearance of the word-final /h/ (originally, ð) in a similar way affected the diphthong au, irregularly changing it to ó (OIC. ó), cf. OIC. fló ‘flew’ -- *flauh -- *flauy instead of the expected au, cf. OIC. skau › ‘shot’ -- *skau, etc. In a similar fashion, the monophthong a was lengthened to á, cf. vá ‘hit (with a weapon)’ -- *wah -- *way.

As a result of devoicing in word-final position, the consonant /d/ often devoiced and subsequently assimilated preceding sonorants, cf. OIC. bátt ‘bound’ (1./3. p. sg. pret.) -- *bátt -- *bátt -- *band, etc. The consonant /g/ often behaved in the same way, devoicing and assimilating a preceding nasal, cf. sprakk ‘exploded’ (1./3. p. sg. pret.) -- *sprank -- *sprank -- *sprang. Since in those word-forms where the conditions for word-final devoicing were not met, devoicing did not occur, paradigms eventually became quite irregular, cf. singular sprakk vs. plural sprungum, or singular bått vs. plural bundum.
Sketching the process

1. Expert Knowledge
2. IceMorph Website
3. Update and add
4. Update
5. Bayesian classification
6. Base Tagged Corpus
Sketching the interface

It might make sense to have the prototype reflection appear here for nouns. (I’m not sure what to do about verbs) so you can swap back and forth. The prototype could also be labeled "ya-stem" or whatever depending on example.
Three Tools

• Morphological analyzer
  http://icemorph.scandinavian.ucla.edu/saga/landnama

• Inflected dictionary browser
  http://etkspace.scandinavian.ucla.edu/~monit/search.php

• Reading environment
  http://etkspace.scandinavian.ucla.edu/~monit/saga/
<table>
<thead>
<tr>
<th>Overview</th>
<th>Lexeme</th>
<th>Inflection</th>
<th>Dictionary Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) man (irrespective of sex), human being (guð skapaði síðarst menn tvá, er ættir eru frá komnar); síndi m manni, one showed it to another, it went from from hand to hand; fjöldi manns, a great number of people; múgmanns, crowd of people; (2) degree in kinship; vera at þriðja, frórðax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Zoega</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Edit**

<table>
<thead>
<tr>
<th>Lemma</th>
<th>maðr</th>
<th>Normalized</th>
<th>maðr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound</td>
<td>maðr</td>
<td>Normalized</td>
<td>maðr</td>
</tr>
</tbody>
</table>

**Translation**

m., nnr und ranni, Vellekla, (for the change of nn before r into ð see the introduction to letter N); gen. manns, dat. manni, acc. mann, plur. menn, qs. menn-r; with the article, menninir, so always in old writers, but in mod. menninir erroneously, as if from mennir: the plur. meðr, answering to the sing. maðr, occurs in old poets—meðr vituð öðling æðra, Fms. vii. 87 (in a verse); Norð-meðr róa naðri, vi. 309 (in a verse); meðr fengu mikit veðr, Edda 102; hirð-meðr, veðja, Rekst., all verses of the 11th and 12th centuries; er meðr Myrkvið kalla, Akv. 5: meðr hlutu sár. Fbr. 75 new Ed. (in a verse): æn. pl. manna. dat. mönnnum. acc. menn. In Ballads
leita

Edit Lexeme and Class

POSC: verb weak (1)

View Inflections

Translations

Edit Translation

Cleasby-Vigfusson:

að, prop. a causal from līta, [Ulf. wlaition = περιβλέπεσθαι, North. E. to lə at c-u, or absol., héraðsmenn leituðu hennar ok fundu hana eigi, Nj. 14; le Landn. 27, 32, Fms. i. 27, 71; leita e-m kvánfanga, Eg. 22, passim: metapl. þótt hans væri eigi vel leitað, Fb. ii. 73: — hón leitaði í eina hirzlu, Fms. iv manna-forrða, Nj. 149; leita at e-m, to seek for, Fms. ix. 218; ok leiti þær a Fms. i. 69: the phrase, leita sést staðar, to go on one's business, cæcar, Hr annarrar en mín ef þú þarft nokkurs við, Nj. 74; leita sér heilla, Landn. 33; leita ráða, to seek for advice, Nj. 75. 2. to enquire, examine, Sks. 638, Hor aggressive, Nj. 16, Ö. H. 222, (á-leitinn); leita á við e-n, to contest, call in leita? Nj. 3; en þat ræð ek, at þú leitir eigi optarr á hreysti mína, that thou after,’ seek for, passim; leita eptir máli, to follow a case, take it up, Nj. 75; enquire into, Eg. 536: leita upp, to seek out, Germ. anfsuchen, Fms. x. 71: 382; leita við

Edit Translation

Zoega:
Saga reader

maðr (n. r.) m., mn. und ranni, Veilekla. (for the change of m. before r into r see the introduction to letter N); gen. manns, dat. manni, acc. mann, plur. menn, q. menn-n; with the article, mennirr, so always in old writers, but in mod. mennirr. in erroneously, as it from mennirr. the plur. mennirr, answering to the sing. maðr. occurs in old poets—maðr víttu Ólög maðr. Fms. vll. 87 (in verse); Nord-maðr rás naðr, vi. 309 (in a verse); maðr fengu miði víðr, Edda 102; hníð-maðr, vegja. Rekšt., all verses of the 11th and 12th centuries; er maðr Myrkvi kalla, Akv. 5. meðr hlutur, Fbr. 75 new Ed. (in a verse); gen. pl. manna, dat. mönnum, acc. menn. In Ballads and Rimur after the 18th century, and hence in ecol. writers of later times, a nom. mann is now and then used, esp. in comps. influenced by Germ. and Engli., e.g. hræysti-mann, Skaða R. 56; or for the sake of rhyme, ætla þe ekki, aurr mann, al komast muni straffast hann, Pass. 14. 17: [Ulf. manna = ætlaumoc; in other Teut. languages speit, or better mann.] [translation from: Clausen-Vigfusson]

(noun m. 1) man (irrespective of sex), human being (gud skapaði síðarst menn tvá, æ ætli eru frá komnar); sýndi í manni, one showed it to another; it went from hand to hand; fjöldi menns, a great number of people; mercury, of a crowd, of people; (2) degree in kinship; vera at þröða, fröðax [translation from: Zoega]

I. Saga 1: Völsunga saga, Chapter 1

(nónn m. r.) 1. noun, masculine gender, r-stem

menn
Inflection: Singular Dative
POS: noun: masculine gender, r-stem

<table>
<thead>
<tr>
<th>Case</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>maðr</td>
<td>menn</td>
</tr>
<tr>
<td>Accusative</td>
<td>mann</td>
<td>menn</td>
</tr>
<tr>
<td>Genitive</td>
<td>manns</td>
<td>mannya</td>
</tr>
<tr>
<td>Dative</td>
<td>mann</td>
<td>mönnum</td>
</tr>
</tbody>
</table>

Chapter 19
Chapter 20
Chapter 21
Chapter 22
Chapter 23
Thank you

- UCLA Council on Research
- John Simon Guggenheim Foundation
- American Council of Learned Societies
- The Mellon Foundation
- Google Books
- NSF
- Institute for Pure and Applied Mathematics
- Dansk Folkemindesamling

- Collaborators: Peter M. Broadwell, Peter Leonard, James Abello, David Mimno, Mads Rosendahl Thomsen, Kryztof Urban, Aurelijus Vijunas