The Challenges of Non-English Language Collections

Roberta Dougherty
Amber Kanner Clooney
Amelia Wilson
Improving Access to World Collections of Arabic-Script Materials

Roberta L. Dougherty
Librarian for Middle East Studies
Yale University Library
MELA 2020
word cloud!

http://mela.us
Arabic and Middle Eastern Electronic Library

AMEEL

AMEEL is a Web-based portal and a digital collection of information for the study of the Middle East, including its history, culture, development, and contemporary face. Within this portal, Yale University Library integrates existing scholarly digital content with newly digitized resources to make such materials easier to find and use efficiently and freely.

Currently, AMEEL holds approximately 350,000 pages of full text, indexed and searchable in the language of publication including Arabic and Western scripts. The full text in AMEEL has been extracted using Optical Character Recognition software (OCR) rather than re-keying all works. Our digitization team aimed for accuracy rates of higher than 90% when converting modern Arabic text.

What's in AMEEL?

- Journals: a regional selection of academic journals from Tunisia to Saudi Arabia
- Abstracts from Brill's Encyclopedia of Islam, THREE
- Dictionaries: Lane's Lexicon, among others
- Manuscripts: selections from Yale's Beinecke Rare Book Library and the Medical
Many hundreds of blog posts on OA materials
Particularly useful: alphabetical lists of historical newspapers and journals
Arabic Collections Online (ACO) is a publicly available digital library of public domain Arabic language content. ACO currently provides digital access to 15,531 volumes across 9,356 subjects drawn from rich Arabic collections of distinguished research libraries. Established with support from NYU Abu Dhabi, and currently supported by major grants from Arcadia, a charitable fund of Lisbet Rasing and Peter Baldwin, and Carnegie Corporation of New York, this mass digitization project aims to feature up to 23,000 volumes from the library collections of NYU and partner institutions. READ MORE...
A hazine or a khazina is a repository. It is where you put things. Thus, Hazine is an online resource that seeks to function as a repository of writing on repositories. It was started by two Ottoman historians, Nir Shafir and Chris Marklewicz, to provide archives and library guides for researchers working on the Middle East, North Africa, and Islamicate societies. Recognizing that research can be opaque and that the research process itself is often self-taught, Hazine was meant to help acquaint researchers, especially those stepping into the field for the first time, with archives and how to navigate them. Heather Hughes was brought onto the team in 2016 to provide a librarian perspective. Heather invited N. Al Mansour to join the team in 2018. Since ‘relaunching’ in 2018, Hazine has continued to provide the research community with archive reviews, but Hughes and Mansour have since reworked the website to be more broadly inclusive of cultural heritage workers and highlight their contributions to research, while also featuring research techniques that address some of the unique challenges to Islamic and Middle East Studies research. They hope to continue to build the site both as a resource and as an inclusive space where cooperation and learning about the research process is fostered.
Ottoman-era calendar page for 20 April 1911

- Arabic & Turkish w/ Islamic & Rumi (Ottoman) calendar
- Bulgarian (in Cyrillic)
- Greek (w/ Julian calendar)
- French (w/ Gregorian calendar)
- Armenian
- Ladino (in Hebrew script; w/ Jewish calendar)
THANK YOU
Yiddish Collections in an English Website

Amber Kanner Clooney
Web Developer and Digital Projects Librarian
Yiddish

- Yiddish is a Germanic language that is written with a modified Hebrew script.
- 30% of vocabulary comes from Hebrew
- Vocabulary also comes from Eastern European and Romance Languages
- Modern Yiddish includes a lot of English vocabulary.
- Around 600,000 Yiddish speakers in the world today.
- Still spoken at home in many Hasidic and Haredi Jewish communities.
Yiddish Book Center

- Founded in 1980 to rescue and distribute Yiddish books which were being discarded.
- Since its founding Yiddish Book Center has rescued more than 1 million books.
- In 1997, the Center began scanning and producing reprints of many books.
- The Center now has 12,000 scanned, freely available books in its Steven Spielberg Digital Yiddish Library.
Digital Collections at yiddishbookcenter.org

- More than 800 archival recordings of events and lectures
- 140 Audiobooks
- Nearly 12,000 digitized Yiddish books
- Over 650 digitized Yizkor books
- More than 800 born-digital oral history videos produced by our Wexler Oral History Project
- All collections are available with a single search on the organization’s main website.
- A single search will provide unique archival and rare books along with articles, podcasts, translations, and other curated content.
How to Incorporate Yiddish on an English Website?

• Metadata
• Romanization
• Keyboards and encoding
• Fonts and Diacritics
• Synonyms and spelling
Metadata

• All metadata in non-roman characters is Romanized.
• In some cases, we also include a translated field.
• Minimum of two title fields; sometimes more in different languages.
About this program:

*sholem aleykhem un amerike*

*Mit*: Chone Shmeruk

**Sholem Aleichem and America**

*Featuring: Chone Shmeruk*
Romanization Tables

● Most scholars of Yiddish use Romanization tables from YIVO Institute for Jewish Research. All of our metadata for our digital collections uses YIVO romanization...

● EXCEPT...

  – For our 12,000 MARC records, which use Library of Congress.
  – Because of the need to standardize MARC records across libraries, we use LoC for MARC and YIVO for everything else.
Romanization and Spelling

- ALA Romanization rules have changed over time. Prior to 1948, words from German retained the German spelling.
- Post-1948, words from German were transliterated consistent with other vocabulary.
How Do We Spell This?

- Standard modern Yiddish Romanized spelling is Gerekhtigkeyt

- Our metadata uses the title transcribed as on the resource, which is the German spelling of the word.
Keyboards

- Metadata is Romanized to provide results from a standard QWERTY keyboard.
- Characters have to be normalized for search in other ways:
  - Yiddish vs. Hebrew characters and diacritics may have differences in encoding
  - Encoding differences can lead to different search results
  - Yiddishbookcenter.org attempts to normalize characters for broader search results.
Keyboards

Hebrew Keyboard layout

Yiddish Pasekh Layout

In catalog searches double vov (װ) gives different results when typed on a Hebrew vs. Yiddish keyboard.
Fonts and Diacritics

• Diacritics for Hungarian titles

A Rákospalotai Zsidó Hitközség története
Tel-Aviv: [publisher not identified], 1978

Part of NYPL-Yiddish Book Center Yizkor Book Collection

A Tegnap városa: a nagyváradi zsidóság emlékkönyve
SCHÖN, DEZSŐ
Tel-Aviv: Kiadta a Nagyváradról Elszármazottak Egyesülete Izraelben, 1981

Part of NYPL-Yiddish Book Center Yizkor Book Collection
Fonts

● We use the Malaga font, but that font doesn’t have all the diacritics we use for the multiple languages on the site.

● We fall back to Noto for some diacritics.

● For Yiddish script, we had one option for a font (Botanika).
Spelling and Pronunciations

- Different pronunciations in Hebrew vs. Yiddish:
  - Shabbat vs shabbos
  - Bat mitzvah vs bas mitzvah

- In modern American usage, the Hebrew version is frequently used for the same word.

- Multiple spellings for different words as well.
Spelling and Pronunciation

● We handle some of this with the Solr synonyms table.

● Here are all the ways you can search for Passover on our website:
  - Passover, pessach, пъсах, пасха, pesah, pessah, pesach, pesaj, paskah, пасха, pasaka, pesahs, pascha, peszah, passower, pesax, pesakh, peysekh, peysakh

● We use a phonetic matching algorithm which allows certain romanizations to match phonetically with words written in Yiddish script.

● This will allow us to catch searches with some variant spellings.
Visit our website at https://www.yiddishbookcenter.org

Questions and comments can be sent to: digitalcollections@yiddishbookcenter.org

OCR site is at http://ocr.yiddishbookcenter.org
Navigating Access and Use of Indigenous Tlingit Archive Materials

Amelia Wilson
Executive Director
Huna Heritage Foundation, Hoonah, Alaska
Our Mission:
Enrich Our Way of Life in perpetuity

“WOOSH - JEE - EEN”

HUNA HERITAGE FOUNDATION
Hoonah, Alaska
Processing Physical Collection
Translation Challenges

Shayadayéik.

Lidzéech kwshé shuwushxeex kwsh

Aḵ aat shayadayéik áwe du ...

aḵ éesh du dlaak’ áwe kúnax.

Áwe kuu hú tsú áa na.éich.

Yá du kéidi gagí xaa.át

Sgóonde nagútch hú aḵ náx.

Aḵ éesh.

Áwe.

Frank O. Williams.

Tlé Lingít ɂ’éínáx ash wudlis’aak.

It is lowered.

It was the difficulty of it that killed her, maybe.

My aunt was lowed down, there, his ...

it was my father’s sister, really.

And she is the one who always sat there.

My father.

That is.
Spelling Challenges
Acknowledgments:

The Editor wishes to thank the many individuals and organizations who helped on this primer.

Materials by Jeff Leer, Constance Naish and Gillian Story provided the resources for these programmed instructions.

The first draft was developed as a project of the Alaska Native Language Center of the Center for Northern Educational Research of the University of Alaska during the fiscal year 1972-73.

The revised draft was field tested at the 4th annual Tlingit language workshop at Sheldon Jackson College, June 1974.

I wish to thank my colleagues and our students for their corrections and suggestions: Henry and Claribel Davis, Katherine Mills, Zero Duenhauer, William and Emma Davis were instructors and consultants, and our native speaking students were: Mitchell Martin, Eva Speaks, Margaret Stevens, Virginia Searl, Franklin James, Elizabeth Jack, John Marks, Peter Marks, Nellie Lord, Emily Williams. Non speakers were: Della Cheney, Doris Hutchinson, Marie Thalmeyer, Edith Bovce and Andy Hope.

The final manuscript was typed by Georgina Davis.

Cover design by Robert Davis
Proofreading by Jeff Leer

Published by Tlingit Readers, Inc.
in association with
Alaska Native Language Center

Printed at Sheldon Jackson College by Tlingit Readers, Inc.
June 1974 360 copies

Printers: Peter Marks (Juneau Indian Studies Program)
Andrew Hope III (Tlingit Readers, Inc.)

Technical assistance: Daryll Scepanwski, Sheldon Jackson College

Copyright 1974 by Tlingit Readers, Inc.

Introduction

This is a mini-primer for Tlingit literacy. Let's look at these words:

1. Literacy: This book is not for teaching the Tlingit language to non speakers. It is designed to help native speakers learn to read and write.

2. Primer: This book is for beginning readers.

3. Mini: This book is designed to be a minimal outline. In most cases, only one page is given to each letter.

A set of more detailed reading books is being composed by Jeff Leer and Constance Naish. Both texts will eventually be used: the mini-primer to give a fast introduction to the letters, and the Leer-Naish readers to give additional graded ready practice and self tests.

Also, an elementary alphabet book is being composed by Henry Davis.

"Spell it like it sounds": English and Tlingit

People ask, 'Why not spell it like it sounds? Why not spell the Tlingit word for 'money' like the girl's name 'Donna'?'

Let's look at Donna. Take the letter o and look at these words:

\begin{verbatim}
hot
ten
off
no
who
woman
women
\end{verbatim}

In English the one letter o can have at least seven different sounds. Only one word "hot"—sounds like "Donna" but that also sounds like "a" in father. Some words like Polish and polish are spelled alike but sound different.
Clan Owned Stories & Songs

Local Tlingit Protocols

Stewards of Cultural Records
Successes & Solutions

http://archives.hunaheritage.org/

- Translation of Language Recordings
  – New language learning resources
- Incorporation of Tlingit language in narrative descriptions
- Mukurtu CMS
  – Access parameters
  – Dictionary feature
- Clan leadership support, input & guidance
- Board of Trustees as authoritative body responsible for ensuring stewardship
- Policies that incorporate local Tlingit protocols
Gunalchéesh
Thank you

Amelia M. Wilson, Executive Director
Huna Heritage Foundation
P.O. Box 275, Hoonah, Alaska 99829
Email amelia.wilson@hunaheritage.com
Phone (907) 789-8582