

Research Centre for Japanese Language and Linguistics
University of Oxford

オックスフォード大学 日本語研究センター
www.orinst.ox.ac.uk/research/jap-ling/

Digital Humanities in Oriental Studies

Kerri L. Russell
kerri.russell@orinst.ox.ac.uk
OI Methodology Seminar 21 February 2014

Introduction

- ▶ Digital Humanities
- ▶ The Oxford Corpus of Old Japanese (OCOJ)
- ▶ Research
 - The Lexicon

Digital Humanities

► Digital Humanities

- What is DH?
 - the use of technology for research in the humanities
 - creation and dissemination of research tools

► Digital Humanities @ Oxford

- Oxford is well known for DH research.
- Text Encoding Initiative (TEI)
- <http://digital.humanities.ox.ac.uk>



Home

News & Events

People & Projects

ICT Methods

Divisions & Units

Training

Support

About

You are here: [Home](#) >> People & Projects

People & Projects

Search people and projects at Oxford

Enter names or terms in the box below to search:

- People's surnames
- People's digital humanities interests
- Project titles
- Project descriptions

Leave the box empty to list all people and projects.

Search people and projects by Subject Areas and/or by ICT Methods

Subject Areas:

ICT Methods:

You can use the **Divisions & Units tab above** for a full list of units and the administrative divisions they are in, and to display the people and projects that fall under each one. The **ICT Methods tab above** gives detailed information on ICT Methods as well as listing people and projects under each broad methods category.

People & Projects

Find a full list of our [Past Featured Projects!](#)

Latest Updates: People

- [Kerri L Russell](#)
- [Neil Jefferies](#)
- [Milena Zeidler](#)

Latest Updates: Projects

- [Early Printing in Europe](#)
- [Integrating Broadside Ballad Resources Online](#)
- [Stationers' Register Online \(SRO\)](#)

Enter, confirm or update details

Members of the University should use the [Oxford login](#) to enter, confirm or update details.

People & Projects Search Results: Oriental Studies , All Methods

People (26)	Projects (17)
A Web-Based Teaching Course on Islamic Ceramics	Ashmolean Museum Jonathan Moffett
Digital Ashmolean Cuneiform	Ashmolean Museum Jacob Dahl
Iraq is the Navel of the World	Ashmolean Museum Eleanor Robson
Islamic Manuscripts Project	Bodleian Libraries Research & Learning Services Neil Jefferies
Ancient Arabian Languages and Cultures	Oriental Studies Faculty Daniel Burt
Book of Curiosities: An early 11th-century Arabic cosmography	Oriental Studies Faculty Emilie Savage-Smith
Cuneiform Digital Library Initiative	Oriental Studies Faculty Jacob Dahl
Diachronic Corpus of Sumerian Literature	Oriental Studies Faculty Jacob Dahl
Egyptian Mirage. A database of 19th-century "studio photographs"	Oriental Studies Faculty Jaromir Malek
Electronic Text Corpus of Sumerian Literature	Oriental Studies Faculty Jacob Dahl
Griffith Institute Archive	Oriental Studies Faculty Elizabeth Frood
Online Egyptological Bibliography	Oriental Studies Faculty John Baines
Scene details in Ancient Egyptian monuments: Oxford Expedition's electronic database and publications project (c.2960 - 2040 BC).	Oriental Studies Faculty Yvonne Harpur
The Arabic Documents of Norman Sicily	Oriental Studies Faculty Daniel Burt
Topographical Bibliography of Ancient Egyptian Hieroglyphic Texts, Statues, Reliefs, and Paintings	Oriental Studies Faculty Vincent Razanajao
UCLA Encyclopedia of Egyptology	Oriental Studies Faculty John Baines
Watercolours of Egypt made by Hector Horreau in 1838, now in the Archive of the Griffith Institute	Oriental Studies Faculty Jaromir Malek

Project Details

[Return to last search](#)

Project name: Griffith Institute Archive

Principal Investigator/ Director: John Baines

Oxford participants: • John Baines • Elizabeth Frood (Main Contact) • Vincent Razanajao

Other participants:

Project web page: <http://www.griffith.ox.ac.uk/griffitharchive.html>

Division: Humanities

Unit: Oriental Studies Faculty

Sub-unit: Griffith Institute

Start date: 1939

End date:

Partner organizations (inside or outside Oxford):

Funder: Check

Subject Area: Oriental Studies

Project description: The Griffith Institute has the largest specialized Egyptological archive in the world (<http://griffith.ashmus.ox.ac.uk/gri/4.html>). The material in the archive includes manuscripts, photographs, drawings, sketches, watercolours, papers, correspondence, notebooks, indexes and plans of Egyptian monuments. Howard Carter's records made during the excavation of the tomb of Tutankhamun are among the most interesting items in the Archive and are systematically being made available, as Tutankhamun: The Anatomy of an Excavation (<http://griffith.ashmus.ox.ac.uk/gri/4tut.html>), on the Institute's web pages. The database provides a simply search function.

Methods Used (Click on methods categories for explanations, or go to the **ICT Methods** tab; **NB:** not all users have registered categories, sub-headings or details.):

Category	Sub-Headings	Details
Data analysis	Searching/Linking --	Content-based image retrieval Searching and querying
Data Capture	Textual Input	Manual input and transcription
Data publishing and dissemination	Data and resource sharing	Resource sharing publishing
Data structuring and enhancement	Classifying and linking	Cataloguing and indexing
Practice-led research	Scanning, Photography and Images Image manipulation	

Other projects the participants have been involved in:

[Online Egyptological Bibliography](#)

[Topographical Bibliography of Ancient Egyptian Hieroglyphic Texts, Statues, Reliefs, and Paintings](#)

[UCLA Encyclopedia of Egyptology](#)

Last updated: 10/09/2012 11:54:51

Updated by: Erin Snyder

The OCOJ

- ▶ The Oxford Corpus of Old Japanese
 - <http://vsarpj.orinst.ox.ac.uk/corpus/>

Home

Texts

Display conventions

Tagging conventions

Searching the OCOJ

Work in progress

Events and news

People and Institutions

Funding bodies

Links

Contact



Arts & Humanities
Research Council



BRITISH
ACADEMY

Japan Office

オックスフォード大学日本事務所

Welcome to the Oxford Corpus of Old Japanese!

The Oxford Corpus of Old Japanese (abbreviated OCOJ) is a long-term research project which aims to develop a comprehensive annotated digital corpus of all extant texts in Japanese from the Old Japanese period. Old Japanese is the earliest attested stage of the Japanese language, largely the Japanese language of the Asuka and Nara periods of Japanese history (7th and 8th century AD). This is the formative literate period upon which the development of Japanese civilization is based, and these texts are of paramount importance for the study and understanding of the origins and development of civilization of Japan, including language, writing, literature, religion, history, and culture.

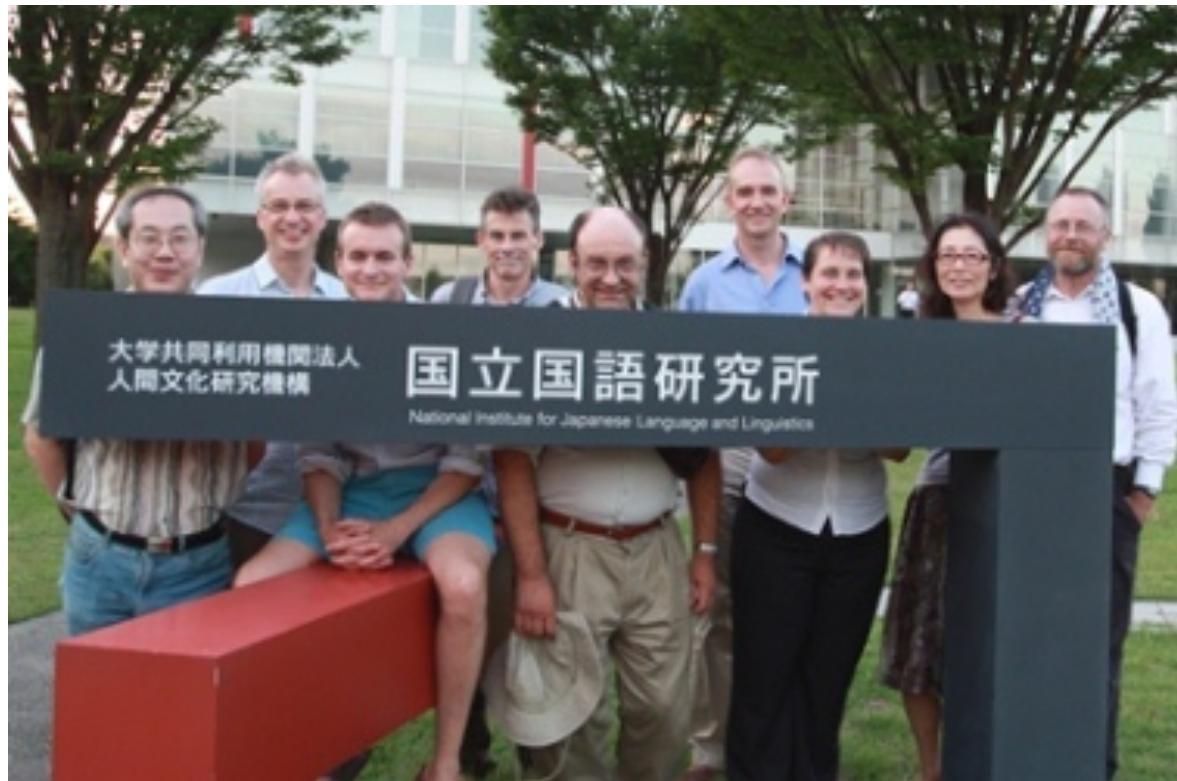
The OCOJ will contain

1. **Texts:** The corpus will contain all extant texts in Japanese from the Old Japanese period, presented in original script and phonemic transcription. See [here](#) for the texts and [here](#) for display conventions.
2. **Annotation:** A large amount of information about the texts will be encoded and made searchable. This will include linguistic information (orthographic, phonological, morphological, syntactic, semantic, and lexical information), as well as literary, biographical, historical, geographical and other information. The digital format makes it possible to add information of any kind continuously. See [here](#) for tagging conventions.
3. **Translations:** The texts will be supplied with translations into English. The translations will be linked to the texts.
4. **Dictionary:** A bilingual Old Japanese – English dictionary will be developed alongside and as an integrated part of the corpus. The dictionary part of the OCOJ will be linked to the texts, making cross-reference in both directions possible.

The OCOJ will provide a research and reference resource of value to specialist scholars and students of early Japan, but it will also provide wide and easy general access to a large body of important texts and materials for anyone interested in Japanese language, history and culture. The OCOJ will be published online and continuously updated, and it will be made easily accessible and searchable through a web-based interface. At the moment we post a simple digital text which gives the original script and a phonemic transcription of the texts.

The OCOJ

► People



The OCOJ

- ▶ Verb semantics and argument realization
in pre-modern Japanese (VSARPJ)
 - <http://vsarpj.orinst.ox.ac.uk>

The OCOJ

- ▶ The OCOJ is marked up with XML tags following the guidelines of the Text Encoding Initiative.
- ▶ XML tags:
 - <s>...</s> for sentences
 - <cl>...</cl> for clauses
 - <phr>...</phr> for phrases
 - <w>...</w> for word-like things
 - <m>...</m> for morphemes
 - <c>...</c> for characters and orthography

The OCOJ

- ▶ A poem (MYS.8.1606)

君待跡

吾戀居者

我屋戶乃

簾令動

秋之風吹

The OCOJ: Romanization

► A poem (MYS.8.1606)

君待跡

kimi matu *to*

吾戀居者

wa ga kwopwi-woreba

我屋戸乃

wa ga yadwo *no*

簾令動

sudare ugokasi

秋之風吹

aki no kaze puku

The OCOJ: Orthography

- ▶ When tagging the corpus, we encoded information about how a word was written: logographically, phonographically, or not written in text.

```
<c type="logo">kimi</c>
```

```
<c type="logo">matu</c>
```

```
<c type="phon">to</c>
```

```
<c type="logo">wa</c>
```

```
<c type="noLogo">ga</c>
```

```
<c type="logo">kwopwi</c>
```

```
<c type="logo">woreba</c>
```

The OCOJ: Morphology

- ▶ Part of speech, lexeme, and morpheme identification

<w lemma="L004266">**kimi**</w>

<w type="verb" inflection="adnconc"
lemma="L031644a">**matu** </w>

<w type="particle" subtype="conj"
lemma="L000531a">**to**</w>

The OCOJ: Syntax

- ▶ Markup of phrases, clauses, and sentences

```
<cl>
  <cl>
    <phr type="arg">
      <w lemma="L004266">
        <c type="logo">kimi</c>
      </w>
    </phr>
    <w type="verb" inflection="adnconc" lemma="L031644a" lemmaRef="35830">
      <c type="logo">matu</c>
    </w>
    <w type="particle" subtype="conj" lemma="L000531a">
      <c type="phon">to</c>
    </w>
  </cl>
  <lb xml:id="MYS.8.1606-trans_1" corresp="#MYS.8.1606-orig_1"/>
  <phr type="arg">
    <w lemma="L042057" lemmaRef="41100">
      <c type="logo">wa</c>
    </w>
    <w type="particle" subtype="case" function="gen" lemma="L000503" lemmaRef="7889">
      <c type="noLogo">ga</c>
    </w>
  </phr>
  <w>
    <w type="verb" inflection="stem" lemma="L030731a" lemmaRef="52566">
      <c type="logo">kwopwi</c>
    </w>
    <w type="verb" inflection="provisional" function="progressive" lemma="L031957a" lemmaRef="5360">
      <c type="logo">woreba</c>
    </w>
  </w>
</cl>
```

The OCOJ: Historical information

- ▶ Bibliographic information

```
<sourceDesc>
  <xi:include href="texts/Manyoshu.xml" parse="xml" xmlns:xi="http://www.w3.org/2001/XInclude"/>
</sourceDesc> _____
</fileDesc>
<profileDesc>
  <creation>
    <date/>
  </creation>
  <langUsage>
    <language ident="ojp">Old Japanese</language>
  </langUsage>
  <textClass>
    <catRef target="#manyoshu"/>
  </textClass>
  . . .

```

The OCOJ: Historical information

- ▶ People and place names

```
<phr type="arg">
  <cl ana="mk1">
    <phr type="arg">
      <w>
        <w type="adjective" inflection="stem">
          <c type="phon">awo</c>
        </w>
        <w>
          <c type="phon">ni</c>
        </w>
      </w>
    </phr>
    <w>
      <w type="adjective">
        <c type="phon">yo</c>
      </w>
      <m type="adjcop" lemma="L000033" inflection="conclusive">
        <c type="phon">si</c>
      </m>
    </w>
  </cl>
  <lb xml:id="KK.58-trans_5" corresp="#KK.58-orig_5"/>
  <w lemma="pln1">
    <c type="phon">nara</c>
  </w>
  <w type="particle" subtype="case" function="acc" lemma="L000534" lemmaRef="41407">
    <c type="phon">wo</c>
  </w>
</phr>
```

Displaying data

► Plain text view

[MYS.8.1606 gloss tree](#)

君待跡	kimi matu <i>to</i>
吾戀居者	wa <u>ga</u> kwopwi-woreba
我屋戸乃	wa <u>ga</u> yadwo <i>no</i>
簾令動	sudare ugokasi
秋之風吹	aki no kaze puku

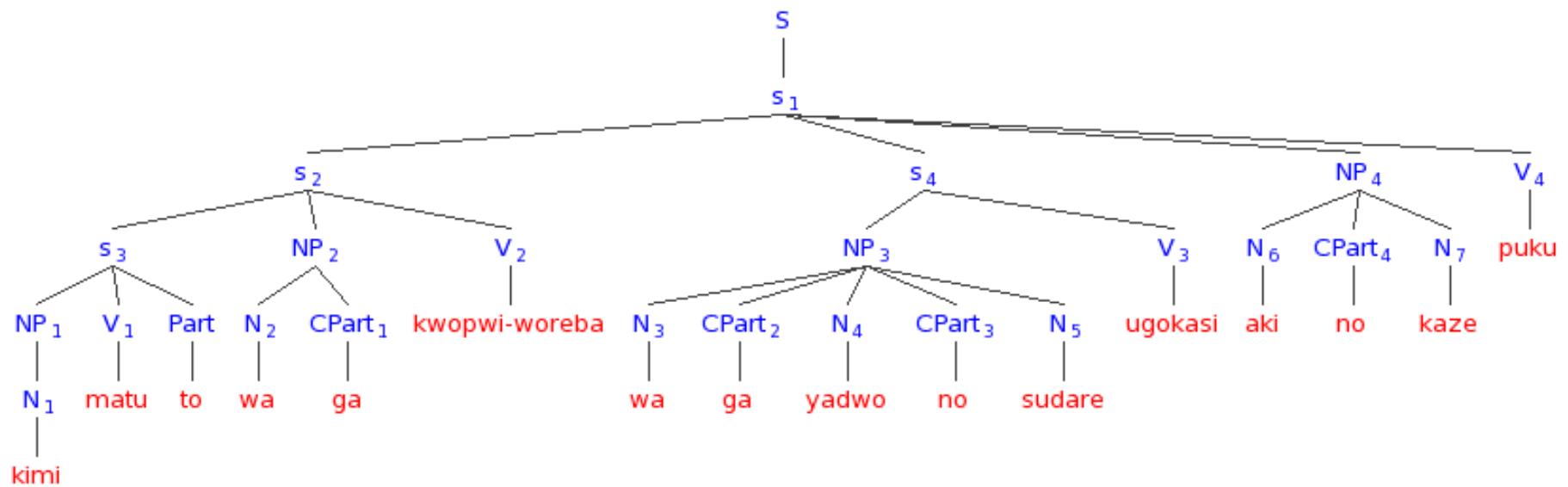
Displaying data

► Glossed view

```
<< { { { [ kimi(L004266 lord) ] matu(verb adnconc L031644a 35830 wait) to() } }  
[ wa(L042057 41100 1st person pronoun) ga(L000503 [genitive case particle]) ] kwopwi(verb stem L030731a 52566 love)-woreba(verb provisional progressive L031957a 5360 be sitting) }  
{ [ wa(L042057 41100 1st person pronoun) ga(L000503 [genitive case particle]) yadwo no(L000520 [genitive case particle])  
sudare ] ugokasi(verb infinitive L030247a 3094 move t) }  
[ aki no(L000520 [genitive case particle]) kaze ] puku(verb adnconc L031516a 32591 blow) } »
```

Displaying data

► Tree view



The Lexicon

- ▶ The Lexicon has two main functions:
 - It stores information about items in the corpus.
 - It allows us to retrieve information.

The Lexicon

► An entry:

```
- - -  
<superEntry xml:id="L031840-main">  
  <entry xml:id="L031840a">  
    <form type="stem">  
      <orth stage="I">yuk-</orth>  
      <gramGrp>  
        <pos>verb</pos>  
        <iType type="QD"/>  
        <gram type="class">motion</gram>  
      </gramGrp>  
    </form>  
    <sense n="1">  
      <def>go</def>  
    </sense>  
    <re type="related">  
      <form>  
        <orth>  
          <ref target="L030145a">ik-</ref>  
        </orth>  
      </form>  
    </re>  
  </entry>  
  <entry xml:id="L031840b">  
    <form type="noun">  
      <orth stage="I">yuki</orth>  
      <gramGrp>  
        <pos>noun</pos>  
      </gramGrp>
```

The Lexicon

► Place names:

```
<superEntry xml:id="pln1-main">
  <entry xml:id="pln1">
    <form>
      <orth stage="I">nara</orth>
    </form>
    <sense n="1">
      <def>Nara</def>
    </sense>
  </entry>
</superEntry>
```

The Lexicon

► People:

```
<superEntry xml:id="perl-main">
  <entry xml:id="perla">
    <form>
      <orth stage="I">tenno kooken</orth>
    </form>
    <sense n="1">
      <def>Empress Kooken</def>
    </sense>
  </entry>
  <entry xml:id="perl1b">
    <form>
      <orth stage="I">daijoutenno kooken</orth>
    </form>
    <sense n="1">
      <def>Retired Empress Kooken</def>
    </sense>
  </entry>
  <entry xml:id="perl1c">
    <form>
      <orth stage="I">tenno syootoku</orth>
    </form>
    <sense n="1">
      <def>Empress Syootoku</def>
    </sense>
  </entry>
</superEntry>
```

Displaying data

► Lexicon entry

yuk-
L031840a

Part of speech: **verb**
Conjugation class: **quadrigrade**
Verb classification: **motion**

1 Gloss: **go**

Related forms:
See also: [ik-'go'](#)

[statistics](#) [attestations](#) [clause structure](#)

yuki
L031840b

Part of speech: **noun**
[statistics](#) [attestations](#)

Research

- ▶ We can use the information in the OCOJ to look for patterns.
- ▶ The online Lexicon allows us to view:
 - all texts (plain and glossed views)
 - English – OJ and OJ – English dictionaries
 - Lists:
 - all verbs and number of attestations
 - verbs, adjectives, and auxiliaries by conjugation class, by inflected forms, and by clause types, verbs by their semantic classification, etc.

Research

- ▶ Verbs of motion
 - sorted by attestation

form	attestations ▲ (logøphonøno text)	gloss	statistics	attestations	claus struc
<u>ko-</u>	813	(538 / 275 / 0) come	statistics	attestations	claus struc
<u>yuk-</u>	692	(490 / 202 / 0) go	statistics	attestations	claus struc
<u>tat-</u>	501	(318 / 182 / 1) stand; depart	statistics	attestations	claus struc
<u>ap-</u>	496	(391 / 103 / 2) meet	statistics	attestations	claus struc
<u>ide-</u>	304	(219 / 83 / 2) go out	statistics	attestations	claus struc
<u>imas-</u>	274	(156 / 118 / 0) be; come; go [respect]	statistics	attestations	claus struc
<u>watar-</u>	254	(179 / 75 / 0) go across	statistics	attestations	claus struc
<u>tir-</u>	180	(138 / 42 / 0) scatter i	statistics	attestations	claus struc
<u>sugwi-</u>	167	(119 / 48 / 0) pass time	statistics	attestations	claus struc
<u>kog-</u>	143	(76 / 67 / 0) row	statistics	attestations	claus struc
<u>sar-</u>	139	(98 / 41 / 0) move on; leave	statistics	attestations	claus struc
<u>yor-</u>	138	(93 / 45 / 0) approach	statistics	attestations	claus struc
<u>kwoye-</u>	134	(99 / 35 / 0) pass over	statistics	attestations	claus struc
<u>kapyer-</u>	131	(89 / 42 / 0) return	statistics	attestations	claus struc

oti-

L030356a

Part of speech: **verb**

Conjugation class: [**upper bigrade**](#)

Verb classification: [**motion**](#)

1 Gloss: **fall**

Related forms:

Derived from: [**-i-**](#) '[theme]'

Transitivity relationship with: [**otos-**](#) 'drop t' [**otor-**](#) 'be inferior'

[**statistics**](#) [**attestations**](#) [**clause structure**](#)

oti

L030356b

Part of speech: **noun**

[L030356a](#) oti- 'fall'

Number of attestations in COJ: 48

Number of attestations written phonographically: 16

Number of attestations written phono-logographically: 0

Number of attestations written logographically: 32

Number of attestations written logo-phonographically: 0

Number of attestations with no orthographic representation in the text: 0

Number of attestations in kake/jo structures and other forms with no text: 0

Number of attestations in EOJ: 1

Number of attestations written phonographically: 1

Number of attestations written phono-logographically: 0

Number of attestations written logographically: 0

Number of attestations written logo-phonographically: 0

Number of attestations with no orthographic representation in the text: 0

Number of attestations in kake/jo structures and other forms with no text: 0

Inflected forms:

inflection

	count	count
	cOJ	EOJ
<u>stem</u>	40	1
<u>adnominal</u>	5	0
<u>conclusive</u>	2	0
<u>gerund</u>	1	0
nominal	0	0
semblative	0	0

L030356a oti- fall

Simple Forms

otite

Text (no gloss)

Text (with gloss)

Trees

oturu

[MYS.4.711](#)

[MYS.4.711](#)

[MYS.4.711](#)

oturu

[MYS.2.137a](#)

[MYS.2.137a](#)

[MYS.2.137a](#)

oturu

[MYS.2.137b](#)

[MYS.2.137b](#)

[MYS.2.137b](#)

oturu

[MYS.8.1617](#)

[MYS.8.1617](#)

[MYS.8.1617](#)

oturu

[MYS.11.2676](#)

[MYS.11.2676](#)

[MYS.11.2676](#)

oturu

[MYS.13.3232](#)

[MYS.13.3232](#)

[MYS.13.3232](#)

oturu

[MYS.13.3233](#)

[MYS.13.3233](#)

[MYS.13.3233](#)

oturu

[MYS.14.3392](#)

[MYS.14.3392](#)

[MYS.14.3392](#)

Use with auxiliaries or suffixes:

oti-zu

Text (no gloss)

Text (with gloss)

[KK.5](#)

oti-ni-ki

[KK.82](#)

[KK.82](#)

[KK.82](#)

oti-ni-ki

[NSK.73](#)

[NSK.73](#)

[NSK.73](#)

oti-zu

[MYS.1.6](#)

[MYS.1.6](#)

[MYS.1.6](#)

oti-zu

[MYS.1.25](#)

[MYS.1.25](#)

[MYS.1.25](#)

oti-zu

[MYS.1.26](#)

[MYS.1.26](#)

[MYS.1.26](#)

oti-zu

[MYS.1.79](#)

[MYS.1.79](#)

[MYS.1.79](#)

oti-zu

[MYS.4.514](#)

[MYS.4.514](#)

[MYS.4.514](#)

oti-me

[MYS.6.1010](#)

[MYS.6.1010](#)

[MYS.6.1010](#)

oti-me

[MYS.8.1653](#)

[MYS.8.1653](#)

[MYS.8.1653](#)

First member of a compound:

	Text (no gloss)	Text (with gloss)	
oti-purabape	KK.100	KK.100	KK.100
oti-purabape	KK.100	KK.100	KK.100
oti-nadusapi	KK.100	KK.100	KK.100
oti-tagitu	MYS.6.909	MYS.6.909	MYS.6.909
oti-tagitu	MYS.6.920	MYS.6.920	MYS.6.920
oti-tagitu	MYS.6.1053	MYS.6.1053	MYS.6.1053
oti-tagitu	MYS.7.1107	MYS.7.1107	MYS.7.1107
oti-tagitu	MYS.7.1127	MYS.7.1127	MYS.7.1127
oti-tagiti	MYS.9.1714	MYS.9.1714	MYS.9.1714
oti-taegitu	MYS.9.1736	MYS.9.1736	MYS.9.1736

Clause Structure ▾	Verb	Following Verb	Text
	oturu		MY
	oti-tagitu		MY
	oti-tagiti-taru		MY
	oti-tagitu		MY
CL-arg	oti-tagitu		MY
CL-arg	oti-tagitu		MY
NP	oti-ni-ki		KK.
NP	oti-nadusapi		KK.
NP	oti-ni-ki		NSK
NP	oti-zu		MY
NP	oti-zu		MY
NP	oti-zu		MY
NP	oti-zu		MY
NP NP	oti-zu		KK.
NP NP-ni	oti-me		MY
NP NP-ni	oti-mo		MY
NP NP-no NP-ni	oti-me		MY
NP-ni	oturu		MY
NP-ni	oturu		MY
NP-ni	oti-purabape		KK.
NP-ni	oti-purabape		KK.
NP-ni	oti-me		MY
NP-ni	oti-tagitu		MY
NP-ni NP	otite		MY
NP-ni-ya	oti-mu		MY
NP-no NP-ni	oti-mu		MY
NP-wo	oti-zu		MY
NP-zo NP-ni	oti-ni-kyeru		MY

Research

- ▶ Displaying data in this way gives us another way to study data to hopefully find new patterns.
- ▶ As we enhance our annotation, it will be possible to look at different kinds of data.
 - mapping of place names, so we can use a map to view all texts referencing a given place
 - common topics/words used by people of a certain age, rank, sex, etc.
- ▶ This is ongoing. We can add information for a variety of research topics.

Networks

- ▶ Oxford has several DH related networks:
 - digitalhumanities@maillist.ox.ac.uk
 - digital-editions@maillist.ox.ac.uk
 - hicor@maillist.ox.ac.uk

Research Centre for Japanese Language and Linguistics
University of Oxford

オックスフォード大学 日本語研究センター
www.orinst.ox.ac.uk/research/jap-ling/

Questions and Comments Welcome

Kerri L. Russell
kerri.russell@orinst.ox.ac.uk