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/

The Oxford Corpus of Old Japanese Lexicon

Kerri L. Russell & Zixi You
kerri.russell@orinst.ox.ac.uk & zyou@uclan.ac.uk
Digital Humanities at Oxford Summer School 2014

Introduction

- The Oxford Corpus of Old Japanese (OCOJ)
  - The VSARPJ project
  - Other research based on the OCOJ
- Annotation
- The Lexicon

Introduction: The OCOJ

- The OCOJ is an annotated digital corpus of all extant texts in Old Japanese (7th and 8th century CE).
- http://vsarpj.orinst.ox.ac.uk/corpus/
- Around 111,000 words.
- The OCOJ includes:
  - All extant texts
  - Annotation (TEI)
  - Translations
  - Dictionary (The Lexicon)

The Oxford Corpus of Old Japanese
オックスフォード上代日本語コーパス

- 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.

Introduction: The OCOJ

▸ People & collaborations
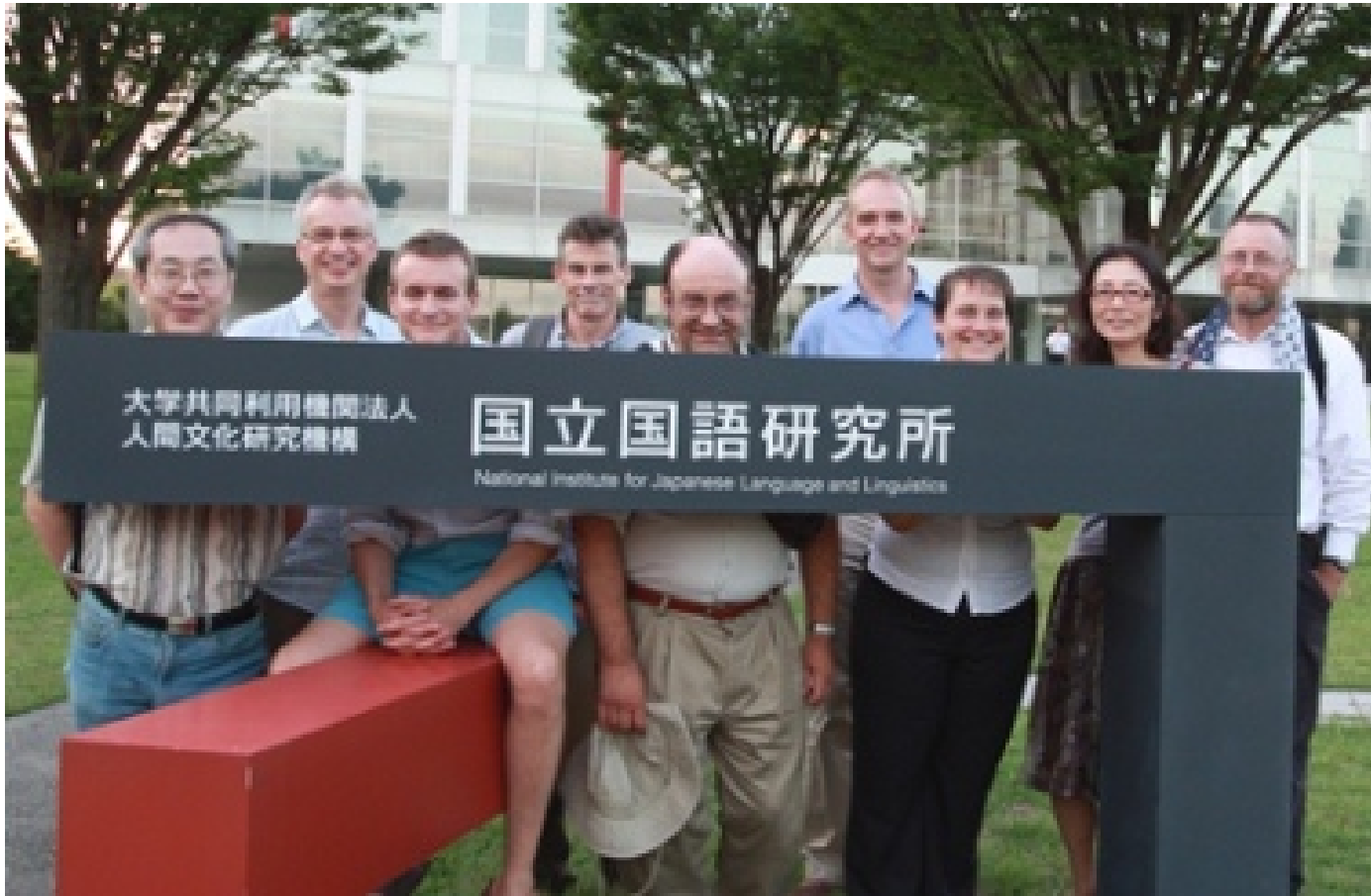
Introduction: The OCOJ

- Funding bodies
  - VSARPJ project: nearly £1 million from the AHRC (1 January 2009 – 30 June 2014)
  - Named a British Academy Research Project (from 13 August 2012)
  - Japan Society for the Promotion of Science/University of Tsukuba
  - John Fell Fund (1 July 2014 – 30 June 2015)

Introduction: The VSARPJ project

▸ The OCOJ started as a research tool for the "Verb semantics and argument realization in pre-modern Japanese" (VSARPJ) project.

◦ http://vsarpj.orinst.ox.ac.uk/

◦ The OCOJ was not initially a corpus building project; it was a corpus designed for a specific research project.

◦ This informed many of our choices as to how to annotate the corpus.

- Lexeme identification (POS, inflection, function)
- Syntactic markup

Introduction: Research

- We have around 20 publications so far (published or soon to be published) based on research using the OCOJ.
- Frellesvig, Bjarke, Stephen W. Horn, Kerri L. Russell, and Peter Sells. 2011-Present. *The Oxford Corpus of Old Japanese*.
  - Others can be found on the VSARPJ website. Topics include:
    - Verb-verb compounds
    - Noun incorporation
    - Null pronouns in various clause types

Introduction: Research

- Participation/collaboration
  - NINJAL Diachronic Corpus project
  - Study of the History of the Japanese Language Using Statistics and Machine-Learning
  - The Japanese Lexicon: A Rendaku Encyclopedia
  - A Diachronic Contrastive Study of Japanese Interrogatives
  - *Senmyō* Project (Tsukuba)
  - HiCor Network (Oxford)

Introduction: Research

- DPhil Dissertations
  - Trott, Daniel. *Tense and Aspect in Old Japanese*. Completed 2014.
  - You, Zixi. *Split Intransitivity in Old Japanese*. Submitted 2013.
  - Mann, Laurence. *New Directions in* Engishiki Norito *Studies* (working title). In progress.

Introduction: Research

- **MA Dissertations**
  - Kitahara, Chiho. *Complex Predicates in Historical Japanese – Analysis of Motion Verbs come/go in Old Japanese*. Completed 2011. (SOAS)
  - Mæhlum, Petter. *Plant names in Old Japanese*. In progress, University of Oslo.
- **BA Dissertations**
  - Millichip, Daniel. *Senmyō: A translation and commentary*. Completed 2012.

Annotation

▸ The OCOJ is marked up with XML tags following the guidelines of the Text Encoding Initiative.

▸ XML tags include:
  ◦ <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

Annotation

▸ Here's a sample poem (MYS.8.1606)

君待跡

吾戀居者

我屋戸乃

簾令動

秋之風吹

Annotation

- Romanization

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

Annotation

▸ Orthography

```
<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>
```

Annotation

▸ 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>

Annotation

▸ Syntax: phrases, clauses, and sentences

```xml
<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>
```

Annotation

▸ Bibliographic information

```xml
  <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>
```

Annotation

▸ People and place names

```xml
<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>
```

The Lexicon

- The Lexicon refers to two things:
  - lexicon.xml
  - A collection of files consisting of pre-defined searches, generated with xslt run on lexicon.xml and the corpus
- Lexicon.xml was initially created as a place to store UIDs.
  - Based on TEI Dictionaries module
- It eventually developed into a bidirectional dictionary, making cross-referencing from the OCOJ to Lexicon possible.
- It is the only bilingual OJ-English dictionary in existence.

The Lexicon: lexicon.xml

```xml
<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: lexicon.xml

```xml
<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: lexicon.xml

```xml
<superEntry xml:id="per1-main">
    <entry xml:id="per1a">
        <form>
            <orth stage="I">tenno kooken</orth>
        </form>
        <sense n="1">
            <def>Empress Kooken</def>
        </sense>
    </entry>
    <entry xml:id="per1b">
        <form>
            <orth stage="I">daijoutenno kooken</orth>
        </form>
        <sense n="1">
            <def>Retired Empress Kooken</def>
        </sense>
    </entry>
    <entry xml:id="per1c">
        <form>
            <orth stage="I">tenno syootoku</orth>
        </form>
        <sense n="1">
            <def>Empress Syootoku</def>
        </sense>
    </entry>
</superEntry>
```

The Lexicon

▸ We have two ways of searching the corpus, both are linked to the Lexicon:

- ◦ A simple search engine
- ◦ The Lexicon (stored searches)

Lexical Item: oti   lemma search

oti "" L030356b

Orthography   Constituent boundary
○ phonogra[...]   oti- "fall" L030356a   ☑ sentence
○ logograph[...]   otiir- "fall in place;fit" L030355a   ☑ clause
◉ all[default]   otiiri "" L030355b   ☑ phrase   [search]

Lexical Item

Orthography
○ phonographically attested
○ logographically attested
◉ all[default]

Lemma search L030356a

Source
○ cOJ
○ EOJ
◉ all[default]

Constituent boundary
☑ sentence
☑ clause
☑ phrase

search

Lexical Item Information
Search lemma   L030356a
Lexical class verb
Stem form   *oti-*
Inflection type UB
Definition   fall

Search Result Information
Source   all
Orthography   all
Number of attestations 49
Constituent boundaries sentence
                        clause
                        phrase

italic phonologically attested
plain   logographically attested
bold verb
♠   link to tree images
«,»   sentence boundary
{,}   clause boundary
more

⇕No.	Source	⇕Before	⇕Search lemma	⇕After	[Updated:2014-
1	KK.5♠g T	a}{{*uti-***mwiru**}sima no saki-zaki]{{*kaki-***mwiru**}iswo no saki]	*oti*	-*zu*}{{[waka-kusa]no}tuma]**mota**-se rame}»«{[a pa mo yo]{	
2	KK.82♠g T	«{{[miya-pito no ayupi no kwo-**suzu**]	*oti*	-ni-ki to}[miya-pito]**toyomu**}»«[satwo-bito mo][yume]»	
3	KK.100♠g T	[pina wo]**op**-yeri}}{{po tu ye no ye no ura-ba pa}{[naka tu ye ni]	*oti*	-**purabape**}}{{[naka tu ye no ye no ura-ba pa}{[simwo tu ye ni]	
4	KK.100♠g T	*ti-***purabape**}}{{[naka tu ye no ye no ura-ba pa}{[simwo tu ye ni]	*oti*	-**purabape**}}[si-du ye no ye no ura-ba pa}{{{{[ari-kinu no]m	
5	KK.100♠g T	[mi-pye no kwo ga]**sasagas**-eru}midu-tama-uki ni][**uki**-si}abura]	*oti*	-**nadusapi**}{[mina][koworo]koworo-ni}}[ko si mo]{{aya-ni ka	
6	MYS.1.6♠g T	«{{[yama-kwosi no kaze wo]tokizi-mi}{[{**nuru**}ywo]	*oti*	-*zu*}{{[ipye n]**aru**}imo wo]{**kakete**}*sinwopi-tu*}»	
7	MYS.1.25♠g T	oki]na-ki ga}goto}{{[so no ame no]ma-na-ki ga}goto}{[kuma mo]{	*oti*	-*zu*}}{**omopitutu** so}**ko**-si [so no yama-miti wo]»	
8	MYS.1.26♠g T	tokizi-ki ga}goto}{{[so no ame no]ma-na-ki ga}goto}{[kuma mo]{	*oti*	-*zu*}}{**omopitutu** so}**kuru** [so no yama miti wo]»	
9	MYS.1.79♠g T	ne]**ukete**}{{{[{[wa ga]**yuku**}kapa no]kapa kuma no *yaswokuma*]	*oti*	-*zu*}[yorodu-tabi]**kapyeri-mi situtu**}{{[[tama-poko]no}miti]**yuk**	
10	MYS.2.137a♠g T	««[{[aki-yama *ni*]	**oturu**	}momiti-ba]»{simasi-ku pa}{na-**tiri-magapi**-*so*}»«{[imo ga atari]	
11	MYS.2.137b♠g T	««[{[aki-yama *ni*]	**oturu**	}momiti-ba]»{simasi-ku pa}{{tiri}na-**magapi**-*so*}»«{[imo ga atari	
12	MYS.4.514♠g T	«{{{[[[wa ga se-kwo ga]**kyes**-eru}koromo no]pari-me]	*oti*	-*zu*}**iri**-*ni-kye rasi* mo} [wa ga kokoro sape]»	
13	MYS.4.711♠g T	«{[[[[{[[[kamo-tori no]**aswobu**}ko no ike *ni*] [ko no pa]	*otite*	}**ukab**-yeru}kokoro] [wa ga]**omopa**-na*ku*}]*ni*}»	
14	MYS.6.1010♠g T	-kwi no pa]**sinwogi**}**puru**}yuki no}puri pa}{**masu** tomo}} [tuti *ni*]	*oti*	-*me ya mo*}»	
15	MYS.6.1053♠g T	utagi *no* miya pa}{[momo-kwi]mo-ku}} [yama pa]{ko-daka-*si*}»«[{[	*oti*	-*tagitu*}se *no*]oto mo} {kiywo-*si*}»«[{[ugupisu *no*]**ki-naku**}paru-	
16	MYS.6.909♠g T	«[[{{[yama]taka-*mi*}{[sira-yupu-bana]*ni*]	*oti*	-*tagitu*}taki no]kaputi pa}{{[**miredo**]**aka**-nu *kamo*}»	

The Lexicon

▸ The Lexicon uses markup in the corpus to display data based on various search conditions.

▸ This makes it possible to examine properties of lexical items grouped by various features.

▸ The Lexicon is currently online on a password protected site.

The OCOJ

- The Lexicon
- The English - OJ Dictionary
- The OJ - English Dictionary
- Valency Dictionary of Selected Old Japanese Verbs [test mode - don't trust what you see here]
- The OCOJ
- The OCOJ (with translations)
- The OCOJ glossed (with translations)
- The OCOJ .pdf (7 July 2014)
- The Norito and Senmyo .pdf (15 November 2013)
- All non-poetic texts
  - Norito texts
  - Semmyo texts
- All non-poetic texts glossed
  - Norito texts glossed
  - Semmyo texts glossed
- SM Trees

Various lists

- All verbs
- All nouns
- All adjectives
- All adverbs

- Verbs and auxiliaries by conjugation class:
  - Irregular
  - K-Irregular
  - Lower Bigrade
  - N-Irregular
  - Quadrigrade
  - R-Irregular
  - S-Irregular
  - Upper Bigrade
  - Upper Monograde
  - unknown

- Adjectives by conjugation class (not complete):
  - -ku adjectives
  - -siku adjectives

The Lexicon

▸ User-friendly display of entries in The Lexicon

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

The Lexicon

▸ The online Lexicon allows us to view:
  ◦ all texts (plain, glossed, and tree views)

The Lexicon

▸ Plain text view

MYS.8.1606 [gloss](#) [tree](#)

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 Lexicon

- ▸ Glossed view

« { { { [**kimi** (L004266 noun lord)] matu (verb adnconc L031644a wait) *to* (L000531a [concessive conjunctional particle]) }

[**wa** (L042057 noun 1st person pronoun) ga (L000503 [genitive case particle])] kwopwi (verb stem L030731a love) -woreba (verb provisional progressive L031957a be sitting) }

{ [wa (L042057 noun 1st person pronoun) ga (L000503 [genitive case particle]) yadwo (L050030 noun dwelling) *no* (L000520 [genitive case particle])

sudare (L050031 noun curtains)] ugokasi (verb infinitive L030247a move t) }

[aki (L050032 noun fall, autumn) no (L000520 [genitive case particle]) **kaze** (L050008a noun wind)] puku (verb adnconc L031516a blow) } »

35

The Lexicon

▸ Tree view
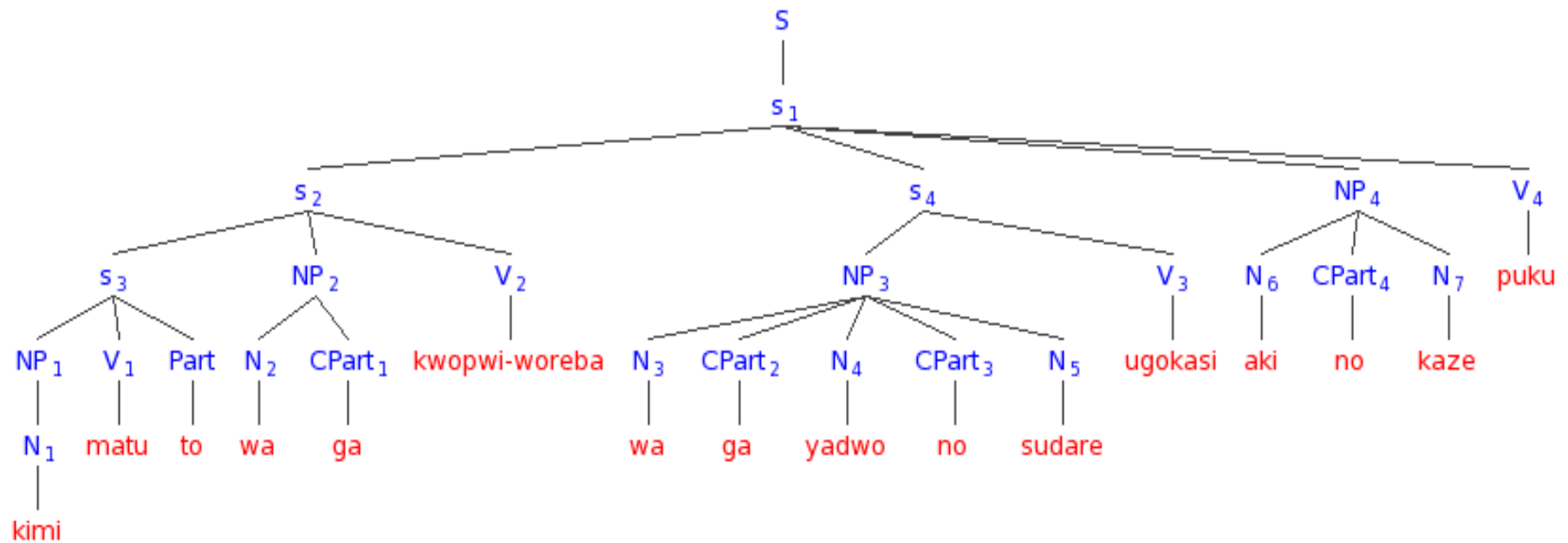
The Lexicon

- The online Lexicon allows us to view:
  - English – OJ and OJ – English dictionaries

English	OJ	attestations ▲	(logo/phono/no text)	Statistics	Attestations	Clause Stru
[genitive case particle]	no	7123	(3543 / 3580 / 0)	statistics	attestations	
[dative case particle]	ni n	3810	(695 / 3115 / 0)	statistics	attestations	
copula	no ni	3768	(1336 / 2432 / 0)	statistics	attestations	clause struct
[topic particle]	mo	2549	(214 / 2335 / 0)	statistics	attestations	
[topic particle]	pa ba	2479	(1528 / 951 / 0)	statistics	attestations	
[genitive case particle]	ga go	2377	(1577 / 800 / 0)	statistics	attestations	
[conjectural]	-m-	2172	(1080 / 1092 / 0)	statistics	attestations	
[accusative case particle]	wo	2168	(166 / 2002 / 0)	statistics	attestations	
[negative]	-n- -zu	1748	(1204 / 544 / 0)	statistics	attestations	
see	mi-	1311	(1132 / 177 / 2)	statistics	attestations	clause struct
[complementizer]	to t	1179	(173 / 1006 / 0)	statistics	attestations	
1st person pronoun	wa	1115	(875 / 240 / 0)	statistics	attestations	
[simple past]	-ki- -si-	1076	(209 / 867 / 0)	statistics	attestations	
exist	ar-	1019	(710 / 309 / 0)	statistics	attestations	clause struct
mountain	yama	992	(815 / 177 / 0)	statistics	attestations	
[perfective]	-n-	990	(299 / 691 / 0)	statistics	attestations	
lord	kimi gi	967	(770 / 197 / 0)	statistics	attestations	
my lord	kimi gi	967	(770 / 197 / 0)	statistics	attestations	
think	omop- mop-	850	(654 / 196 / 0)	statistics	attestations	clause struct
come	ko-	813	(538 / 275 / 0)	statistics	attestations	clause struct
person, people	pito	753	(591 / 162 / 0)	statistics	attestations	
woman, women	pito	753	(591 / 162 / 0)	statistics	attestations	
[restrictive particle]	si	719	(125 / 594 / 0)	statistics	attestations	
[focus particle]	ka	704	(110 / 594 / 0)	statistics	attestations	
[stative]	-yer-	695	(453 / 242 / 0)	statistics	attestations	

The Lexicon

▸ The online Lexicon allows us to view:
  ◦ Lists:
    • all verbs, nouns, adjectives, adverbs
    • verbs, adjectives, and auxiliaries by conjugation class, by inflected forms, and by clause types, verbs by their semantic classification, etc.

▸ For today, let's look at verbs of motion

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

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
stem	40	1
adnominal	5	0
conclusive	2	0
gerund	1	0
nominal	0	0
semblative	0	0

L030356a oti- fall

Simple Forms	Text (no gloss)	Text (with gloss)	Trees
otite	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:	Text (no gloss)	Text (with gloss)	
oti-zu	KK.5	KK.5	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-tagitu	MYS.9.1736	MYS.9.1736	MYS.9.1736

[L030356a oti-](#) fall

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

The Lexicon

- 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.

The Lexicon

▸ We have plans to make parts of the OCOJ Lexicon public. If you're interested, please get in touch and we'll let you know when it's online.

Questions and Comments Welcome

Kerri L Russell & Zixi You
Kerri.russell@orinst.ox.ac.uk & zyou@uclan.ac.uk