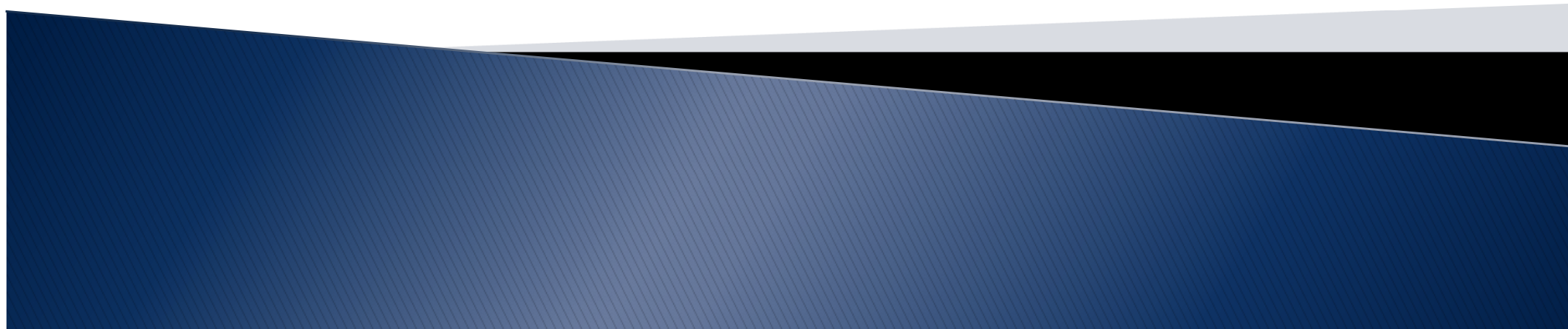


NINJAL 18 May 2014

The Oxford Corpus of Old Japanese



Oxford Corpus of Old Japanese

A comprehensive digital, annotated corpus of all extant texts in Japanese from the Old Japanese period (mainly 8th century AD, Nara period)

Funding and collaboration

Funding

Arts and Humanities Research Council, UK

British Academy

JSPS/Tsukuba University

John Fell Fund

Collaborating institutions

国立国語研究所

University of York

OCOJ Content

Texts

Translations

Dictionary

Annotation

Texts

The corpus will contain all extant texts in Japanese from the Old Japanese period, presented in **original script** and a **phonemic transcription**.

Texts in the OCOJ

Poetic texts

<i>Kojiki kayō</i> (古事記歌謡; 712)	(112 poems; 2527 words)
<i>Nihon shoki kayō</i> (日本書紀歌謡; 720)	(133 poems; 2444 words)
<i>Fudoki kayō</i> (風土記歌謡; 730s)	(20 poems; 271 words)
<i>Bussukoseki-ka</i> (仏足石歌; after 753)	(21 poems; 337 words)
<i>Man'yōshū</i> (万葉集; after 759)	(4685 poems; 83706 words)
<i>Shoku nihongi kayō</i> (続日本紀歌謡; 797)	(8 poems; 134 words)
<i>Jōgū shōtoku hōō teisetsu</i> (上宮聖徳法王帝説)	(4 poems; 60 words)

‘Prose’ texts

<i>Shoku nihongi Senmyō</i> (続日本紀宣命; 797)	(approx. 14,000 words)
<i>Engishiki Norito</i> (延喜式祝詞)	(approx. 6,500 words)

Texts in the OCOJ

Poetic texts

<i>Kojiki kayō</i> (古事記歌謡; 712)	(112 poems; 2527 words)
<i>Nihon shoki kayō</i> (日本書紀歌謡; 720)	(133 poems; 2444 words)
<i>Fudoki kayō</i> (風土記歌謡; 730s)	(20 poems; 271 words)
<i>Bussukoseki-ka</i> (仏足石歌; after 753)	(21 poems; 337 words)
<i>Man'yōshū</i> (万葉集; after 759)	(4685 poems; 83706 words)
<i>Shoku nihongi kayō</i> (続日本紀歌謡; 797)	(8 poems; 134 words)
<i>Jōgū shōtoku hōō teisetsu</i> (上宮聖徳法王帝説)	(4 poems; 60 words)

‘Prose’ texts

<i>Shoku nihongi Senmyō</i> (続日本紀宣命; 797)	(approx. 14,000 words)
<i>Engishiki Norito</i> (延喜式祝詞)	(approx. 6,500 words)
??? <i>Kojiki</i> (古事記; 712)	???

Translations

The texts are being supplied with **translations** into English.

Dictionary

A bilingual Old Japanese – English dictionary is being developed alongside and as an integrated part of the corpus. The dictionary part of the OCOJ is linked to the texts, making cross-reference in both directions possible.

Annotation

A large amount of information about the texts is being encoded and made searchable.

This includes **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 and to **search** for any combination of information in the corpus.

Annotation

XML mark-up following the internationally recognized standards of the **T**ext **E**ncoding **I**nitiative (TEI)

Manual mark-up

Annotation

Linguistic information

Annotation

Writing 原文の表記法

Part-of-speech 品詞

Lexeme and morpheme UID 語彙素と形態素のID番号

Morphology, inflection 活用

Syntax シンタクス (要素・構造)

Syntactic analysis marks and delimits **sentences, clauses,**
and **phrases.**

Man'yōshū (万葉集) 3.341

賢跡
物言從者
酒飲而
醉哭為師
益有良之

Man'yōshū (万葉集) 3.341

賢跡

sakasi-mi*to* サカシミト

物言従者

mono-ipu ywori pa モノイフヨリハ

酒飲而

sake nomite サケノミテ

酔哭為師

wepi-naki suru *si* エヒナキスルシ

益有良之

masari-taru *rasi* マサリタルラシ

“It should be better to drink saké and weep drunkenly
than talking in a clever fashion”

```

<s>
  <phr>
    <cl>
      <cl>
        <w>
          <w type="adjective">
            <c type="logo">sakasi</c>
          </w>
          <m inflection="gerund" lemma="L000034" lemmaRef="36226" type="adjcop">
            <c type="logo">mi</c>
            <c type="phon">to</c>
          </m>
        </w>
      </cl>
      <lb corresp="#MYS.3.341-orig_1" xml:id="MYS.3.341-trans_1"/>
      <w lemma="L031771a">
        <w lemma="L050042">
          <c type="logo">mono</c>
        </w>
        <w inflection="adnconc" lemma="L030199a" lemmaRef="1571" type="verb">
          <c type="logo">ipu</c>
        </w>
      </w>
    </cl>
    <w function="abl" lemma="L000540" subtype="case" type="particle">
      <c type="logo">ywori</c>
    </w>
    <w lemma="L000522" lemmaRef="29321" subtype="top" type="particle">
      <c type="logo">pa</c>
    </w>
  </phr>
  <lb corresp="#MYS.3.341-orig_2" xml:id="MYS.3.341-trans_2"/>

```


Man'yōshū (万葉集) 3.341

sakasi_(adjective)-mito_(adjcop, gerund)

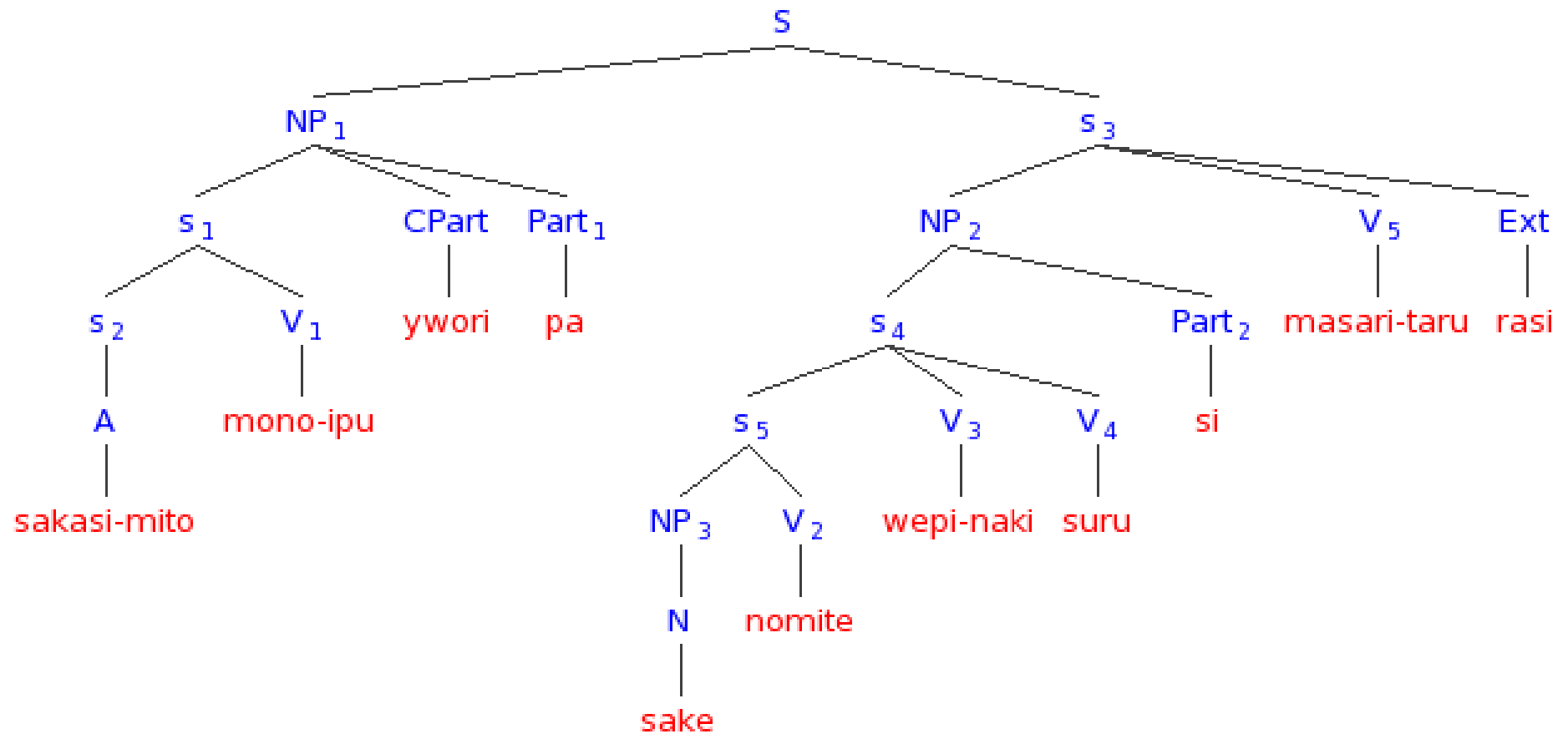
mono_(noun)-ipu_(verb, adnconc) ywori_(abl case particle) pa_(top particle)

sake_(noun) nomite_(verb, gerund)

wepi_(verb, stem)-naki_(verb, infinitive) suru_(verb, adnominal) si_(res particle)

masari_(verb, stem)-taru_(auxiliary, stative, adnominal) rasi_(extension, conclusive)

Man'yōshū (万葉集) 3.341



検索 Searching

The key to using an annotated text corpus is searchability.

検索 Searching

Lemmatization allows us to search the corpus for words regardless of their inflected form by doing a simple KWIC (KeyWord In Context) search:

nomu	drink.conclusive/adnominal ‘(I) drink; (I) who drink’ 飲.終止連体形
nomi	drink.infinitive ‘drinking’ 飲.連用形
noma-nu	drink-negative.adnominal ‘(I) who don’t drinking’ 飲-否定.連体形

[illegible]

Searching

More importantly,
any combination of information
contained in the corpus can be
searched.

Searching

More importantly,
any combination of information
contained in the corpus can be
searched.

At the moment, such ‘advanced’
searches have to be programmed

Annotation

‘Other’ information

Annotation

The annotation is being expanded to include amongst other information such as the following, both *in* and *about* the texts

- literary
- biographical
- historical
- geographical
- other information

Access and conditions of use

Conditions of use

Citation

Cite the corpus if you use it for any kind of publication

Frellesvig, Bjarke, Stephen Wright Horn, Kerri L. Russell, and Peter, Sells.
The Oxford Corpus of Old Japanese.

Please include the retrieval date as the date of publication

Conditions of use

Citation

Cite the corpus if you use it for any kind of publication

Frellesvig, Bjarke, Stephen Wright Horn, Kerri L. Russell, and Peter, Sells.

The Oxford Corpus of Old Japanese.

Please include the retrieval date as the date of publication

Don't share access

Conditions of use

Citation

Cite the corpus if you use it for any kind of publication

Frellesvig, Bjarke, Stephen Wright Horn, Kerri L. Russell, and Peter, Sells.

The Oxford Corpus of Old Japanese.

Please include the retrieval date as the date of publication

Don't share access

Tell us

Conditions of use

Citation

Cite the corpus if you use it for any kind of publication

Frellesvig, Bjarke, Stephen Wright Horn, Kerri L. Russell, and Peter, Sells.
The Oxford Corpus of Old Japanese.

Please include the retrieval date as the date of publication

Don't share access

Tell us

- about mistakes (so we can correct them)
- about your results

Access

Corpus webpage:

<http://vsarpj.orinst.ox.ac.uk/corpus/>

Access

Corpus webpage:

<http://vsarpj.orinst.ox.ac.uk/corpus/>

Lexicon:

<http://vsarpj.orinst.ox.ac.uk/lexicon/>

KWIC:

<http://rcjll.orient.ox.ac.uk/cse/cse02/start.html>

Username: bruce

Password: elvis42

Grammatical roles

Subject, Object, OBLIQUE

Semantic roles of arguments

Agent, Theme; Goal (including recipient, beneficiary, etc.)

Path, Source (verbs of motion)

Location (verbs of location / possession)

Experiencer (verbs of experience)

Causee (morphological causatives)

Empty (copula subjects; subjects raised to objects)