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

Kojiki kayō (古事記歌謡; 712) (112 poems; 2527 words)

Nihon shoki kayō (日本書紀歌謡; 720) (133 poems; 2444 words)

Fudoki kayō (風土記歌謡; 730s) (20 poems; 271 words)

Bussukoseki-ka (仏足石歌; after 753) (21 poems; 337 words)

Man'yōshū (万葉集; after 759) (4685 poems; 83706 words)

Shoku nihongi kayō (続日本紀歌謡; 797) (8 poems; 134 words)

Jōgū shōtoku hōō teisetsu (上宮聖徳法王帝説) (4 poems; 60 words)

'Prose' texts

Shoku nihongi Senmyō (続日本紀宣命; 797)

(approx. 14,000 words)

Engishiki Norito (延喜式祝詞)

(approx. 6,500 words)

Texts in the OCOJ

Poetic texts

Kojiki kayō (古事記歌謡; 712) (112 poems; 2527 words)

Nihon shoki kayō (日本書紀歌謡; 720) (133 poems; 2444 words)

Fudoki kayō (風土記歌謡; 730s) (20 poems; 271 words)

Bussukoseki-ka (仏足石歌; after 753) (21 poems; 337 words)

Man'yōshū (万葉集; after 759) (4685 poems; 83706 words)

Shoku nihongi kayō (続日本紀歌謡; 797) (8 poems; 134 words)

Jōgū shōtoku hōō teisetsu (上宮聖徳法王帝説) (4 poems; 60 words)

'Prose' texts

Shoku nihongi Senmyō (続日本紀宣命; 797)

Engishiki Norito (延喜式祝詞)

???Kojiki (古事記; 712)

(approx. 14,000 words)

(approx. 6,500 words)

???

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.

Annotaation

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 Text Encoding Initiative (TEI)

Manual mark-up

Annotation

Linguistic information

Annotaation

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**.

賢跡 物言従者 酒飲而 酔哭為 益有良之

賢<mark>跡</mark> 物言従者

酒飲而

酔哭為師

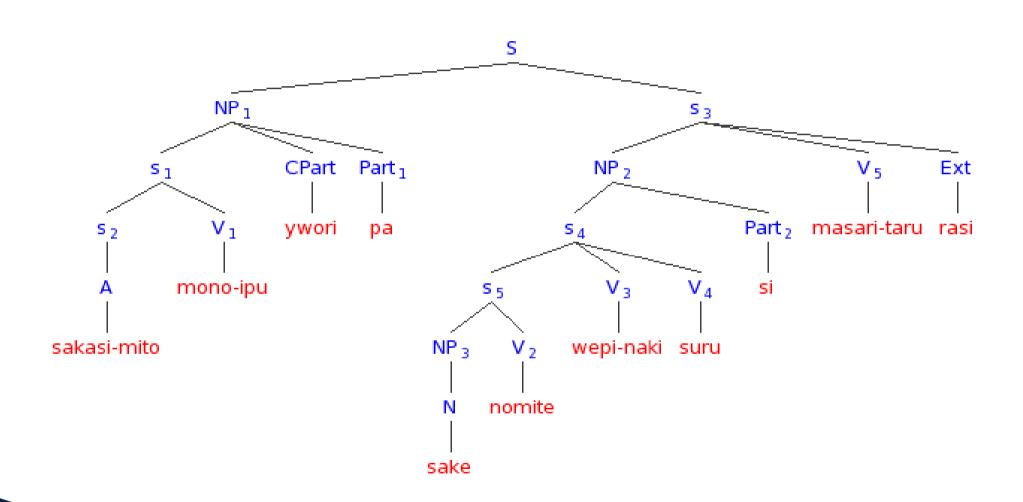
益有良之

sakasi-mi*to* サカシミト
mono-ipu ywori pa モノイフヨリハ
sake nomite サケノミテ
wepi-naki suru *si* エヒナキスルシ
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"

```
\leqs>
 <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="ab1" 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"/>
```

```
sakasi<sub>(adjective)</sub>-mito<sub>(adjcop, gerund)</sub>
mono<sub>(noun)</sub>-ipu<sub>(verb, adnconc)</sub> ywori<sub>(abl case particle)</sub> pa<sub>(top particle)</sub>
sake<sub>(noun)</sub> nomite<sub>(verb, gerund)</sub>
wepi<sub>(verb, stem)</sub>-naki<sub>(verb, infinitive)</sub> suru<sub>(verb, adnominal)</sub> si<sub>(res particle)</sub>
masari<sub>(verb, stem)</sub>-taru<sub>(auxiliary, stative, adnominal)</sub> rasi<sub>(extension, conclusive)</sub>
```



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

| Lexical Item: | | lemma search L031386a Lexical Item I | | formation | Search Result Information | italic phonologically attested | |
|---------------|---|--|---|---|--|--|--|
| © : | ography phonographicall logographically all[default] | ly attested Constituent boundary Value Va | Search lemma Lexical class Stem form Inflection type Definition | L031386a verb nom- QD drink | Source all Orthography all Number of attestations 22 Constituent boundaries sentence clause phrase | plain logographically attested bold verb link to tree images «,» sentence boundary clause boundary | |
| \$No | o. Source | ‡ Before | | Search lemma | ♦After | [Updated: 2012-Dec-0 | |
| 1 | MYS.3.338 | i] na-ki}mono WO] omopa-zupa} [{[pito-tuki] no} {nigo | OF-eru} sake WO] ▶ | nomu | be-ku aru <i>rasi</i> }» | | |
| 2 | MYS.3.341€ | «[{{sakasi-mito}mono-ipu}ywori | nomite | } wepi-naki suru } si j masari-taru rasi } » | | | |
| 3 | MYS.3.344 | «{[ana]miniku}»«{{[{{[sakasi-ra wo]su to}[sake] | | noma | -nu}pito woj {yo-ku}mireba}} [saru ni ka moj {ni-mu}» | | |
| 4 | MYS.3.346 ♠ | «{{{[{yworu}pikaru}tama]∅ to}ipu | nomite | }[{[kokoro wo]yaru}ni][ani]sika-me ya mo}» | | | |
| 5 | MYS.3.350♠ | «[{{[moda]worite}[sakasi-ra]suru | nomite | }wepi-naki suru}ni] [napo]sika-zu-kyeri}» | | | |
| 6 | MYS.4.555 ⊕ | «{[{[kimi ga tame] kami-si}mati-zake][yasu no nwo | noma | -mu){{[tomo]na-si}nisite}» | | | |
| 7 | MYS.5.821 ⊕ | «[{{{[[awo-yanagwi][ume to]]no pana wo | nomite | }no noti pa₁ { { tiri -nu tomo} yo-si}» | | | |
| 8 | MYS.5.833♠ | ru no₁ ki -taraba⟩ [kaku si koso] ([ume wo₁ kazasit e | noma | -me}» | | | |
| 9 | MYS.6.973 ☆ | ipu}»«{{uti-nade so}negwi-tamapu}»«{[{[{kapyeri-ko-mu}pi]a | | noma | -mu}ki]Ø so} [ko no toyo-mi-ki pa]» | | |
| 10 | MYS.6.995 | «{{[kaku]\$ | nomi | -koso}»«{[kusa-kwi sura]{[paru pa]{opwitutu}}}[aki pa]{tiri-yuk | | | |
| 11 | MYS.7.1128 ± | [[asibi]nasu]sakaye-si}kimi ga]pori-si}wi no ipa-w | nomedo | }aka-nu kamo}» | | | |
| 12 | MYS.7.1142€ | saki-ku) yo-kye-mu to [{ipabasiru}tarumi no midu wo] musubite | | nomi | -tu;» | | |
| 13 | MYS.7.1295 ⊕ | I) mikasa no yama ni] [{[tukwi] no} pune] idu} » « {[{[miyabwi-wo no]} | | nomu | }sakaduki ni] { [kage] ni} mi-yetutu}» | | |
| 14 | MYS.8.1656 ± | «[{{[saka-duki ni][ume no pana]ukabe}[| {omopu}doti] | nomite | } <u>no</u> noti pa}{{ tiri -nu tomo}yo-si}» | | |
| 15 | MYS.8.1657 ⊕ | «[tukasa ni mo] {yurusi-tamap-yeri}»«{ [{ [l | ko-yopi nomwi] | noma | -mu}sake]Ø kamo}»«{tiri-kosu na[yume]}» | | |
| 16 | MYS.12.2925 ± | «{{{[midori-kwo no tame koso]}[omo pa]{motom | nu <u>to}}ipe</u> }[ti] | nome | ya}»«{ [kimi ga] [omo] motomu ramu}» | | |
| 17 | MYS.13.3260 | ı-na-ku so}} [pito pa] {kumu to}} ipu}»«{{{{tokizi-ku | 1 SO}}[pito pa]{ | nomu | to}}ipu}»«{{{{[{kumu}pito]no}ma- | na-ki ga}goto}{{{[{ nomu }pito | |

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

Annotaation

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

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

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

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

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)