

# Why romanize Japanese?

## Introducing The Oxford Corpus of Old Japanese

Bjarke FRELLESVIG

University of Oslo  
5 March 2014

# 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 8<sup>th</sup> century AD, Nara period)

# Funding and collaboration

## Funding

Arts and Humanities Research Council, UK

British Academy

JSPS/Tsukuba University

University of Oxford John Fell Fund

## Collaborating institutions

国立国語研究所

University of York

# OCOJ Content

# 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 **romanized, 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)
<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 and look-up in both directions possible.

# Annotation

# Annotation

is what makes a corpus a corpus  
and not simply a digital text

# 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.

# Texts

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

なぜローマ字化か

Why romanization?

Why transcription?



# *Man'yōshū* (万葉集) 15.3599

月余美能

比可里乎伎欲美

神鳴乃

伊素未乃宇良由

船出須和礼波

# *Man'yōshū* (万葉集) 15.3599

月余美能

比可里乎伎欲美

神鳴乃

伊素未乃宇良由

船出須和礼波

# *Man'yōshū* (万葉集) 15.3599

月余美能

月ヨミノ

比可里乎伎欲美

光ヲ清ミ

神嶋乃

神島ノ

伊素未乃宇良由

磯廻ノ浦ユ

船出須和礼波

船出ス我ハ

# *Man'yōshū* (万葉集) 15.3599

月余美能

月ヨミ ノ

比可里乎伎欲美

光 ヲ 清ミ

神嶋乃

神島 ノ

伊素未乃宇良由

磯廻 ノ 浦 ヲ

船出須和礼波

船出 ス 我 ハ

“The moonlight being bright, I will set out  
in my boat from the rocky bay of Kamishima”

# *Man'yōshū* (万葉集) 15.3599

月余美能

ツクヨミ ノ

比可里乎伎欲美

ヒカリ ヲ キヨミ

神嶋乃

カミシマ ノ

伊素未乃宇良由

イソミ ノ ウラ ユ

船出須和礼波

フナデ ス ワレ ハ

“The moonlight being bright, I will set out  
in my boat from the rocky bay of Kamishima”

# *Man'yōshū* (万葉集) 15.3599

月余美能

*tukuyomi no*

比可里乎伎欲美

*pikari wo kiywo-mi*

神嶋乃

*kamwisima no*

伊素未乃宇良由

*iswomwi no ura yu*

船出須和礼波

*puna-de su ware pa*

“The moonlight being bright, I will set out  
in my boat from the rocky bay of Kamishima”

# *Man'yōshū* (万葉集) 15.3599

月余美能

*tukuyomi no*

月ヨミ ノ

比可里乎伎欲美

*pikari wo kiywo-mi*

ヒカリ ヲ キヨミ

神嶋乃

*kamwisima no*

神島 ノ

伊素未乃宇良由

*iswomwi no ura yu*

イソミ ノ ウラ ユ

船出須和礼波

*puna-de su ware pa*

船出 ス ワレ ハ

“The moonlight being bright, I will set out  
in my boat from the rocky bay of Kamishima”

なぜローマ字化か

Why romanization?



なぜ音素表記か

Why phonemic transcription?

# Phonological representation

Phonological representation

Segmentation

Sound texture appreciation

# Phonological representation

当時の音韻の反映

# Phonological representation

## 当時の音韻の反映

Old Japanese /p/ (ハ行子音の音価)

Old Japanese Type A and B syllables  
(甲類・乙類音節の識別と解釈)

# Phonological representation

## 当時の音韻の反映

Old Japanese /p/

# *Man'yōshū* (万葉集) 15.3599

月余美能

*tukuyomi no*

月ヨミノ

比可里乎伎欲美

*pikari wo kiywo-mi*

光ヲ清ミ

神嶋乃

*kamwisima no*

神島ノ

伊素未乃宇良由

*iswomwi no ura yu*

磯廻ノ浦ユ

船出須和礼波

*puna-de su ware pa*

船出ス我ハ

# *Man'yōshū* (万葉集) 15.3599

月余美能

*tukuyomi no*

月ヨミノ

比可里乎伎欲美

*pikari wo kiywo-mi*

ピカリヲキヨミ

神嶋乃

*kamwisima no*

神島ノ

伊素未乃宇良由

*iswomwi no ura yu*

イソミノウラユ

船出須和礼波

*puna-de su ware pa*

プナ出スワレパ



# *Man'yōshū* (万葉集) 15.3599

月余美能

*tukuyomi no*

月ヨミノ

比可里乎伎欲美

*pikari wo kiywo-mi*

[hi] ヒカリ ヲ キヨミ

神嶋乃

*kamwisima no*

神島ノ

伊素未乃宇良由

*iswomwi no ura yu*

イソミノウラユ

船出須和礼波

*puna-de su ware pa*

[fu] フナ出 ス ワレ ハ [wa]

# Phonological representation

## 当時の音韻の反映

Type **A** and **B** syllables

Old Japanese

Early Middle J

Modern J

Type A      *ywo*

*yo*

*yo*

Type B      *yo*

# *Man'yōshū* (万葉集) 15.3599

月余美能

*tukuyomi no*

比可里乎伎欲美

*pikari wo kiywo-mi*

神嶋乃

*kamwisima no*

伊素未乃宇良由

*iswomwi no ura yu*

船出須和礼波

*puna-de su ware pa*

Japanese kana (hiragana and katakana)  
emerged in the Early Middle  
Japanese period and are therefore not  
designed to write or represent Old  
Japanese.

# *Man'yōshū* (万葉集) 15.3599

月余美能

ツクヨミノ

比可里乎伎欲美

ヒカリヲキヨミ

神嶋乃

カミシマノ

伊素未乃宇良由

イソミノウラユ

船出須和礼波

フナデスワレハ

# *Man'yōshū* (万葉集) 15.3599

月余美能

ツクヨ<sub>B</sub>ミノ

比可里乎伎欲美

ヒカリヲキヨ<sub>A</sub>ミ

神嶋乃

カミシマノ

伊素未乃宇良由

イソミノウラユ

船出須和礼波

フナデスワレハ

# *Man'yōshū* (万葉集) 15.3599

月余美能

*tukuyomi no*

比可里乎伎欲美

*pikari wo kiywo-mi*

神嶋乃

*kamwisima no*

伊素未乃宇良由

*iswomwi no ura yu*

船出須和礼波

*puna-de su ware pa*



Old Japanese

Early Middle J

Modern J

Type A      mi

mi

mi

Type B      mwi

# *Man'yōshū* (万葉集) 15.3599

月余美能

tukuyomi no

比可里乎伎欲美

pikari wo kiywo-mi

神嶋乃

kamwisima no

伊素未乃宇良由

iswomwi no ura yu

船出須和礼波

puna-de su ware pa

# *Man'yōshū* (万葉集) 15.3599

月余美能

ツクヨミノ

比可里乎伎欲美

ヒカリヲキヨミ

神嶋乃

カミシマノ

伊素未乃宇良由

イソミノウラユ

船出須和礼波

フナデスワレハ

# *Man'yōshū* (万葉集) 15.3599

月余美能

ツクヨ ミ<sub>A</sub>ノ

比可里乎伎欲美

ヒカリヲキヨ ミ<sub>A</sub>

神嶋乃

カ ミ<sub>B</sub>シマノ

伊素未乃宇良由

イソ ミ<sub>B</sub>ノウラユ

船出須和礼波

フナデスワレハ

# *Man'yōshū* (万葉集) 15.3599

月余美能

tukuyomi no

比可里乎伎欲美

pikari wo kiywo-mi

神嶋乃

kamwisima no

伊素未乃宇良由

iswomwi no ura yu

船出須和礼波

puna-de su ware pa

# Segmentation

## 形態素の分節

*nomu*

‘(I) drink; (I) who drink’

*nomi*

‘drinking’

*nomu*

ノム

‘(I) drink; (I) who drink’

*nomi*

ノミ

‘drinking’



*nom + u*

ノ ム

‘(I) drink; (I) who drink’

*nom + i*

ノ ミ

‘drinking’

万5.904

三枝之

中爾乎祢牟登

愛久

志我可多良倍婆

# 万5.904

三枝之  
三枝ノ  
*sakikusa no*

中爾乎祢牟登  
中ニヲ寝ムト  
*naka ni wo ne-mu to*

愛久  
愛シク  
*utokusiku*

志我可多良倍婆  
シガ語ラヘバ  
*si ga katarapeba*

“When he adorably said that he wanted to sleep between us, like the *sakikusa* (three blade/stalk grass)”

# 万5.904

三枝	之	中	爾	乎	衿牟	登
三枝	ノ	中	ニ	ヲ	寢ム	ト
<i>sakikusa no</i>		<i>naka ni</i>		<i>wo</i>	<i>ne-mu</i>	<i>to</i>
<i>sakikusa</i> like		middle DAT		ACC	sleep-CONJ	that

愛久	志 我	可多良倍婆
愛シク	シ ガ	語ラヘバ
<i>utukusiku</i>	<i>si ga</i>	<i>katarapeba</i>
adorably	he GEN	say.PROV

“When he adorably said that he wanted to sleep between us, like the *sakikusa* (three blade/stalk grass)”

# 万5.904

三枝	之	中	爾	乎	衿牟	登
三枝	ノ	中	ニ	ヲ	寢ム	ト
sakikusa	no	naka	ni	wo	ne-mu	to
sakikusa	like	middle	DAT	ACC	sleep-CONJ	that

愛久	志	我	可	多	良	倍	婆
愛シク	シ	ガ	語	ラ	ヘ	バ	
utokusiku	si	ga	katarapeba				
adorably	he	GEN	say	PROV			

“When he adorably said that he wanted to sleep between us, like the *sakikusa* (three blade/stalk grass)”

# 万5.904

三枝	之	中	爾	乎	衿牟	登
三枝	ノ	中	ニ	ヲ	寢ム	ト
sakikusa	no	naka	ni	wo	ne-mu	to
sakikusa	like	middle	DAT	ACC	sleep-CONJ	that

愛久	志	我	可多良倍婆
愛シク	シ	ガ	語ラヘバ
utokusiku	si	ga	katarapeba
adorably	he	GEN	say.PROV

“When he adorably said that he wanted to sleep between us, like the *sakikusa* (three blade/stalk grass)”

# 万5.904

三枝	之	中	爾	乎	衿牟	登
三枝	ノ	中	ニ	ヲ	寢ム	ト
sakikusa	no	naka	ni	wo	ne-mu	to
sakikusa	like	middle		ACC	sleep-CONJ	that

愛久	志 我	可多良倍婆
愛シク	シ ガ	語ラヘバ
utokusiku	si ga	katarapeba
adorably	he GEN	say.PROV

“When he adorably said that he wanted to sleep between us, like the *sakikusa* (three blade/stalk grass)”

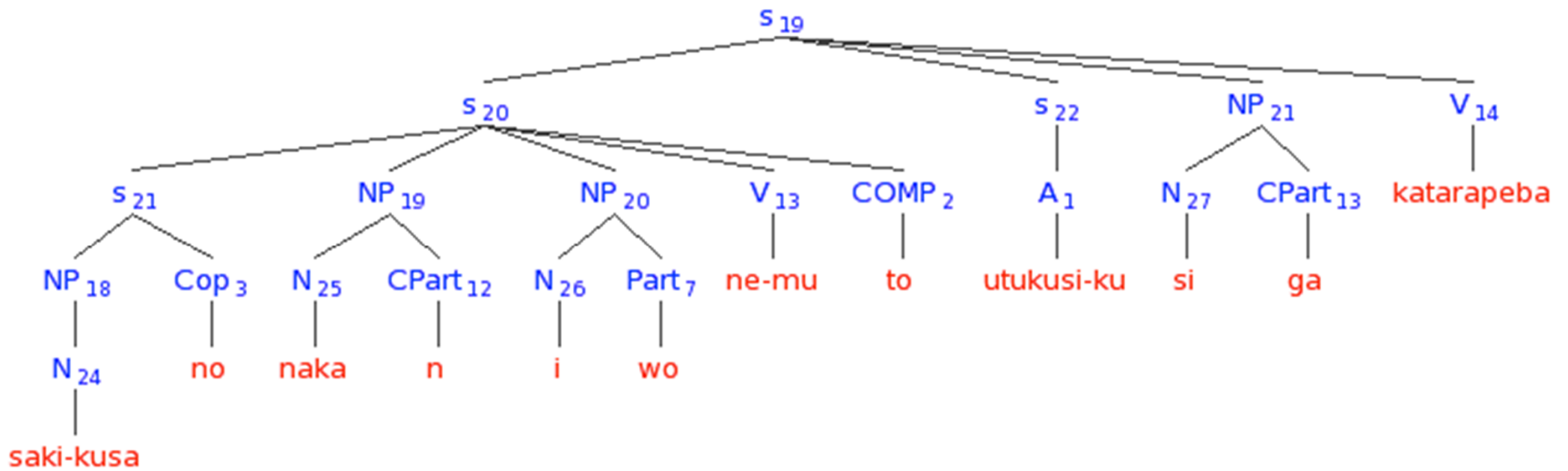
# 万5.904

三枝之	中	爾	乎	衿牟	登
三枝ノ	中	ニ	ヲ	寝ム	ト
sakikusa no	naka n	i	wo	ne-mu	to
sakikusa like	middle DAT	sleep	ACC	sleep-CONJ	that

愛久	志 我	可多良倍婆
愛シク	シ ガ	語ラヘバ
utokusiku	si ga	katarapeba
adorably	he GEN	say.PROV

“When he adorably said that he wanted to sleep between us, like the *sakikusa* (three blade/stalk grass)”





# Sound texture appreciation

## サウンドテクスチャの鑑賞

# *Man'yōshū* (万葉集) 3.341

賢跡

*sakasi-mito*

賢シミト

物言従者

*mono-ipu ywori pa*

物言フ ヨリ ハ

酒飲而

*sake nomite*

酒 飲ミテ

酔哭為師

*wepi-naki suru si*

酔ヒ泣キ スル シ

益有良之

*masari-taru rasi*

勝リタル ラシ

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

# *Man'yōshū* (万葉集) 3.341

賢跡

sakasi-mito

sakashimito

物言従者

mono-ipu ywori pa

mono yuu yori wa

酒飲而

sake nomite

sake nomite

醉哭為師

wepi-naki suru si

einaki suru shi

益有良之

masari-taru rasi

masaritaru shi

# *Man'yōshū* (万葉集) 3.341

賢跡

sakasi-mito

**sakashimito**

物言従者

mono-ipu ywori pa

**mono-yuu yori wa**

酒飲而

sake nomite

**sake nomite**

醉哭為師

wepi-naki suru *si*

**einaki suru shi**

益有良之

masari-taru *rasi*

**masaritaru shi**

# *Man'yōshū* (万葉集) 3.341

賢跡

*sakasi-mito*

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

# *Man'yōshū* (万葉集) 3.341

賢跡

*sakasi-mito*

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

# *Man'yōshū* (万葉集) 3.341

賢跡

*sakasi-mito*

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



# Why phonemic transcription?

# Why phonemic transcription?

Phonological representation

# Why phonemic transcription?

Phonological representation

Segmentation

Sound texture appreciation

ご清聴ありがとうございました