

オックスフォード 上代日本語コーパス

The Oxford Corpus of Old Japanese

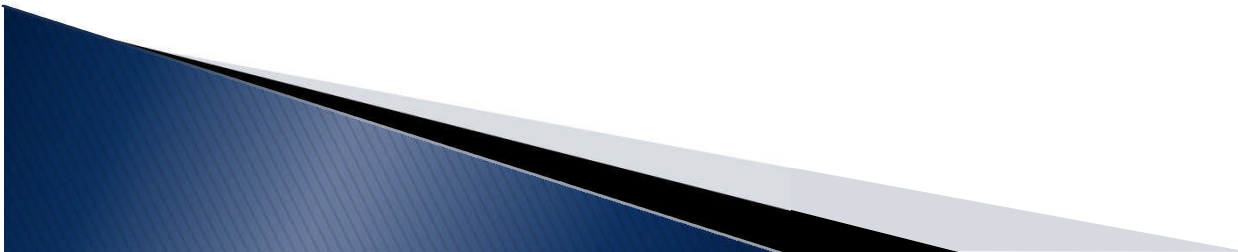
Bjarke FRELLESVIG

ビャーケ・フレレスビグ

- ▶ 日本語史における動詞の意味構造と項の具現化
Verb semantics and argument realization in pre-modern Japanese
- ▶ オックスフォード上代日本語コーパス
The Oxford Corpus of Old Japanese
 - 現在 Present
 - 将来 Future
- ▶ なぜローマ字化か
Why romanization

Verb semantics and argument realization in pre-modern Japanese VSARPJ

日本語史における
動詞の意味構造と項の具現化



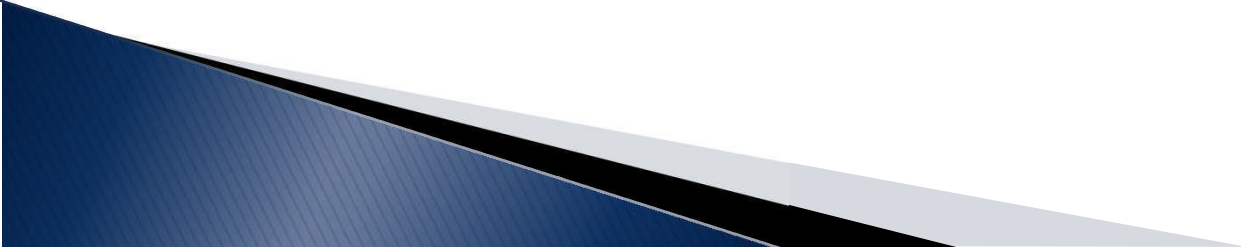
Verb semantics and argument realization in pre-modern Japanese VSARPJ

日本語史における動詞の意味構造と項の具現化

基金: Arts and Humanities Research Council, UK (イギリス人文科学研究振興
会) £1 million

期間: 2009年1月から2013年10月まで

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



VSARPJ 人事

- ▶ Bjarke Frellesvig (Oxford / NINJAL)
- ▶ Peter Sells (York University)
- ▶ Postdoctoral Researchers (PD研究員)
 - Stephen Wright Horn (Oxford)
 - Kerri L Russell (Oxford)
- ▶ Doctoral students (大学院生)
 - Daniel Trott (Oxford)
 - Zixi You (Oxford)

VSARPJ 人事

External members (外部メンバー)

- ▶ Anton ANTONOV (INALCO and CRLAO, Paris)
- ▶ Satoshi KINSUI (Osaka University)
- ▶ Tomohide KINUHATA (Fukuoka University)
- ▶ Yasuhiro KONDO (Aoyama Gakuin / NINJAL)
- ▶ Masayoshi SHIBATANI (Rice University / NINJAL)
- ▶ Akira WATANABE (University of Tokyo)
- ▶ John WHITMAN (NINJAL / Cornell University)
- ▶ Yuko YANAGIDA (Tsukuba University)

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

Present 現在

Oxford Corpus of Old Japanese

Comprises all poetic texts from the Old
Japanese period

上代日本語の韻文のテキスト

approximately 90,000 words
約 9 万語

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

Texts in the OCOJ

<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)

OCOJ

Original text 原文

Phonemic transcription ローマ字化

Annotation アノテーション

OCOJ

Original text 原文

Phonemic transcription ローマ字化

ONLINE DIGITAL TEXT

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



FACULTY OF ORIENTAL STUDIES UNIVERSITY OF OXFORD



The Oxford Corpus of Old Japanese

Home

Online corpus

Pre-modern
Japanese syntax
research project

Research Centre
for Japanese
Language and
Linguistics

Contact

This is a brief introduction, under continuing development, of the Oxford Corpus of Old Japanese. Old Japanese is mainly the language of the 8th century. The Oxford Corpus of Old Japanese is being constructed as part of the [Verb semantics and argument realization in pre-modern Japanese](#) project.

The corpus consists of Old Japanese [texts](#) which are supplied with a range of information in the form of xml tags, following the conventions of the [Text Encoding Initiative](#). The texts are romanized in a phonemic transcription and supplied with information about the original orthography; original script is also retained. Lexemes and morphemes are given unique identifiers and words are part-of-speech tagged. Inflecting words are supplied with information about morphology. Finally, information about syntactic constituency is encoded.

We are posting a very simple [online version](#) of the full Old Japanese poetic corpus which gives the original script and a phonemic transcription of the texts. The grammatical

古之

inisipye no

七賢

nana no sakasi-ki

人等毛

pito-domo *mo*

欲為物者

pori se-si mono pa

酒西有良師

sake *ni si* aru *rasi***MYS.3.341**

賢跡

sakasi-mito

物言從者

mono-ipu yworì pa

酒飲而

sake nomite

醉哭為師

wepi-naki suru *si*

益有良之

masari-taru *rasi***MYS.3.342**

將言為便

ipa-mu subye

將為便不知

se-mu subye sira-zu

極

kipamarite

貴物者

taputwo-ki mono pa

酒西有良之

sake *ni si* aru *rasi*

Man'yōshū (万葉集) 3.341

賢跡	sakasi-mito
物言從者	mono-ipu ywori pa
酒飲而	sake nomite
醉哭為師	wepi-naki suru si
益有良之	masari-taru <i>rasi</i>

“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

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

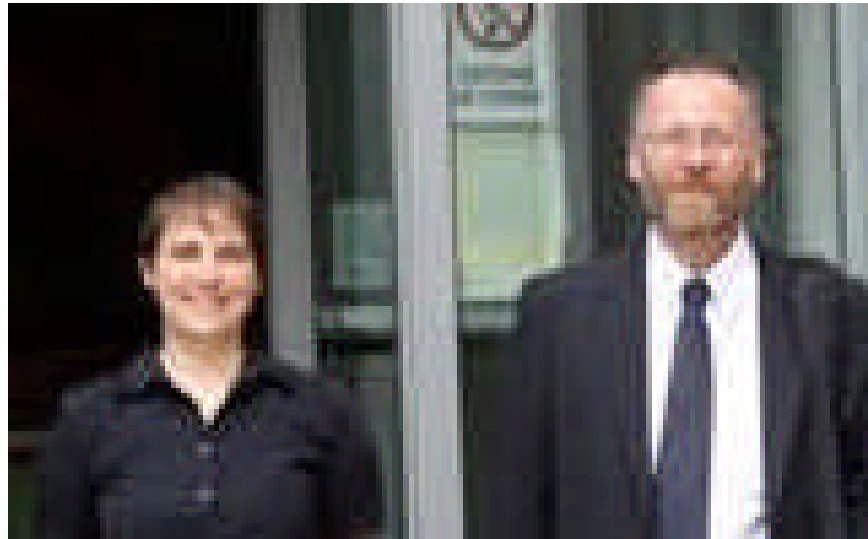
OCOJのアノテーション Annotation

TEIの基準に従ったXMLマークアップ

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

Manual mark-up

Kerri Russell and Stephen Horn



xml tag set in the OCOJ

<c> ... </c>	character
<m> ... </m>	morpheme
<w> ... </w>	word
<phr> ... </phr>	phrase
<cl> ... </cl>	clause
<s> ... </s>	sentence

OCOJのアノテーション

Writing 原文の表記法

xml tags: <c>

Part-of-speech 品詞

Lexeme and morpheme UID 語彙素と形態素のID番号

Morphology, inflection 活用

xml tags: <w> <m>

Syntax シンタクス (要素・構造)

xml tags: <s> <cl> <phr>

Writing 原文の表記法

xml tag: <c>

表音文字 (万葉がな)

Phonographic writing

表語文字

Logographic writing

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

Man'yōshū (万葉集) 3.341

表語文字 表音文字

賢跡

sakasi-mi**to**

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

Man'yōshū (万葉集) 3.341

表語文字 (plain)

表音文字 (*italics*)

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

Man'yōshū (万葉集) 15.3599

月余美能

比可里乎伎欲美

神嶋乃

伊素未乃宇良由

船出須和礼波

Man'yōshū (万葉集) 15.3599

表語文字 表音文字

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

表語文字 表音文字

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 ュ

船出須和礼波

puna-de su ware pa

船出 ス 我 ハ

Man'yōshū (万葉集) 15.3599

表語文字 (plain)

表音文字 (*italics*)

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

語彙素と形態素の情報

Word and Morpheme Information

xml tags: <w>, <m>

Part-of-speech 品詞

Lexeme and morpheme UID [Unique Identifier]

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)

xml tags:

<s> Sentence 文

<cl> Clause 節

<phr> Noun phrase 名詞句
argument (項)

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

```

<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"/>

```



```

<cl>
  <phr type="arg">
    <cl>
      <cl>
        <phr type="arg">
          <w>
            <c type="logo">sake</c>
          </w>
        </phr>
        <w inflection="gerund" lemma="L031386a" lemmaRef="29226" type="verb">
          <c type="logo">nomite</c>
        </w>
      </cl>
    <lb corresp="#MYS.3.341-orig_3" xml:id="MYS.3.341-trans_3"/>
    <w>
      <w inflection="stem" lemma="L031930a" lemmaRef="39012" type="verb">
        <c type="logo">wepi</c>
      </w>
      <w inflection="infinitive" lemma="L031267a" type="verb">
        <c type="logo">naki</c>
      </w>
    </w>
    <w inflection="adnominal" lemma="L030919a" lemmaRef="19537" type="verb">
      <c type="logo">suru</c>
    </w>
  </cl>
  <w lemma="L000526a" lemmaRef="14922" subtype="res" type="particle">
    <c type="phon">si</c>
  </w>
</phr>
<lb corresp="#MYS.3.341-orig_4" xml:id="MYS.3.341-trans_4"/>

```



```
<w>
  <w inflection="stem" lemma="L031630a" lemmaRef="35662" type="verb">
    <c type="logo">masari</c>
  </w>
  <m function="stative" inflection="adnominal" lemma="L000022" lemmaRef="94822" type="auxiliary">
    <c type="logo">taru</c>
  </m>
</w>
<w inflection="conclusive" lemma="L000806" lemmaRef="39702" type="extension">
  <c type="phon">rasi</c>
</w>
</cl>
</s>
```

Man'yōshū (万葉集) 3.341

sakasi_(adjective)-mito_(adjcop, gerund)

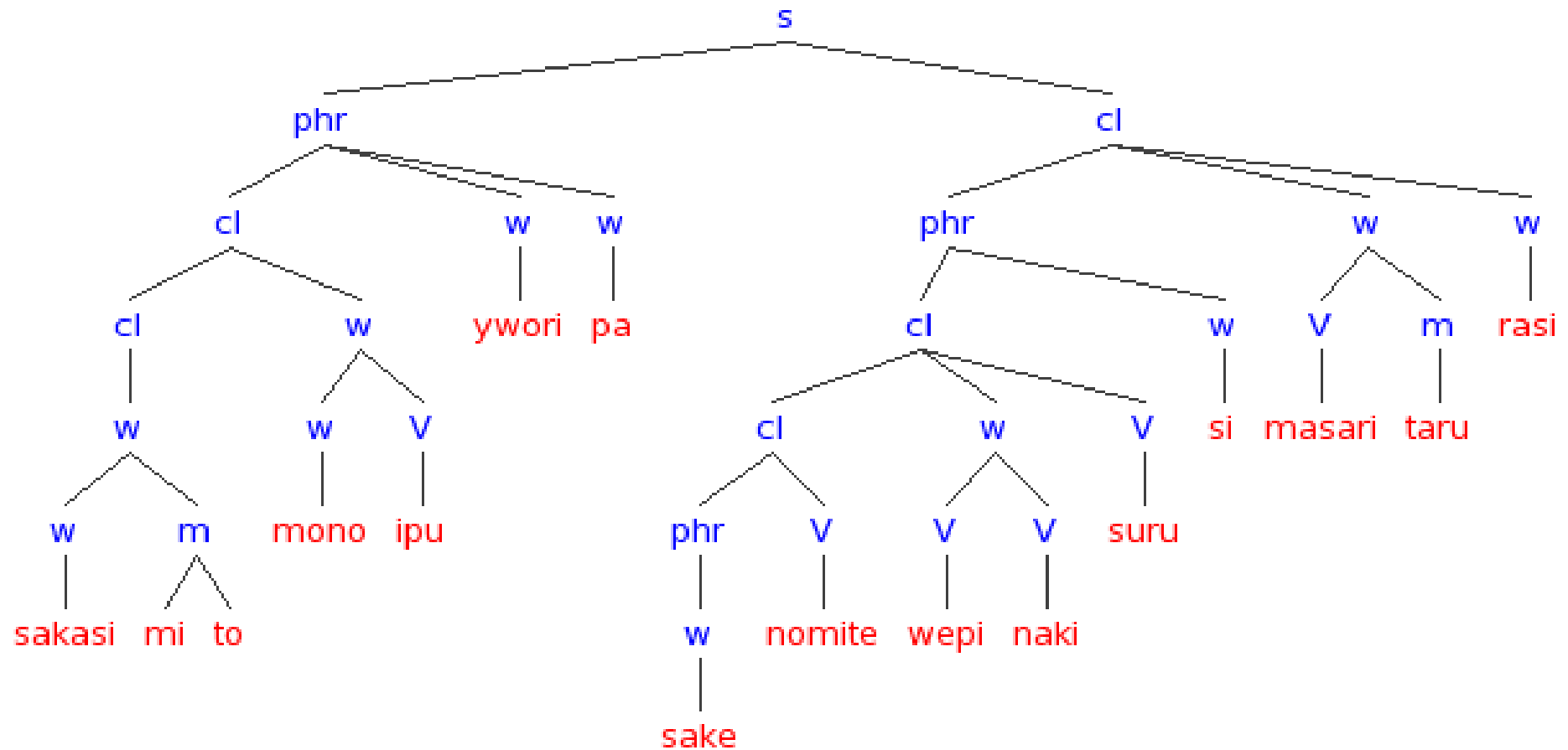
mono_(noun)-ipu_(verb, adnconc) ywori_(abl case particle) pa_(top particle)

sake_(noun) nomite_(verb, gerund)

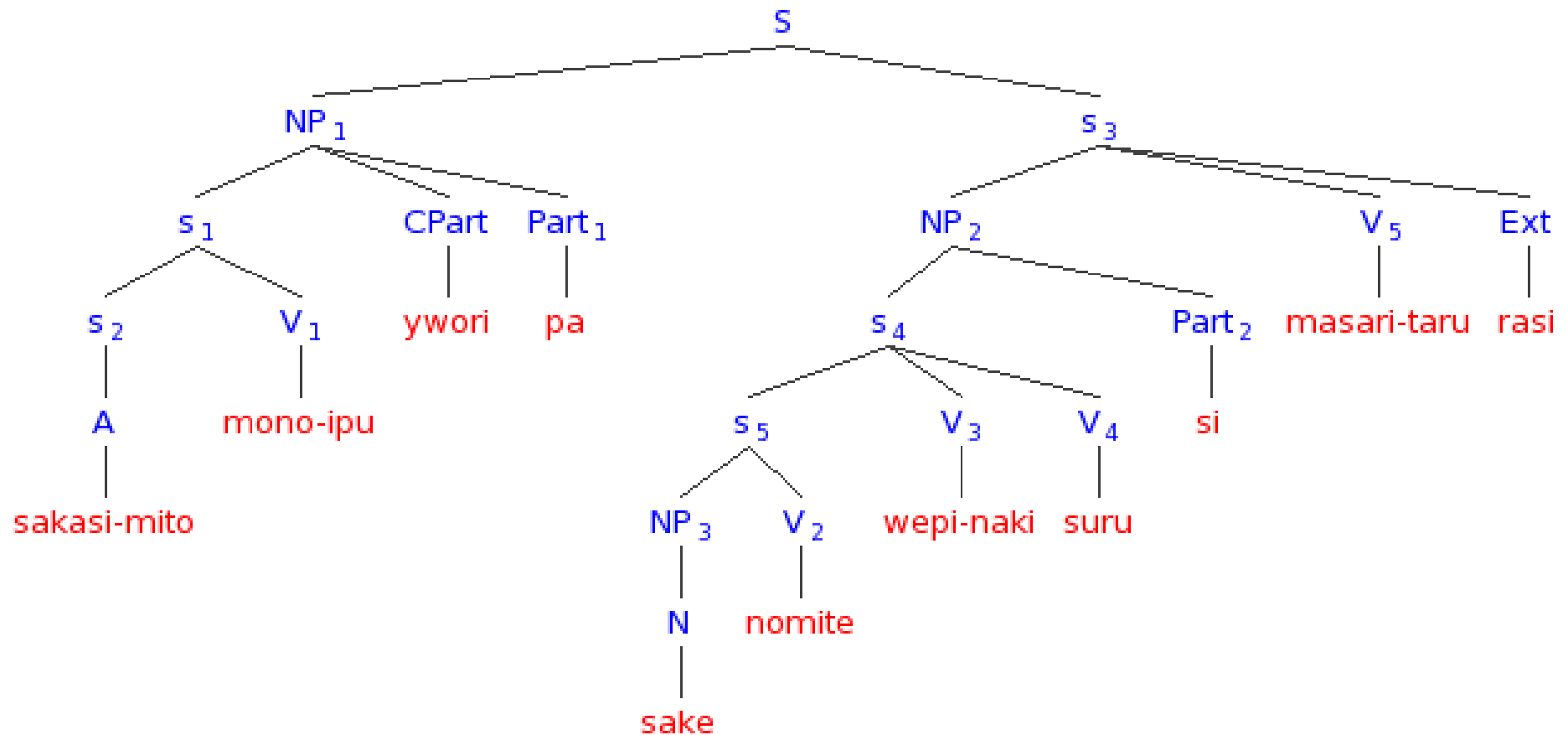
wepi_(verb, stem)-naki_(verb, infinitive) suru_(verb, adnominal) si_(res particle)

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

Man'yōshū (万葉集) 3.341



Man'yōshū (万葉集) 3.341



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

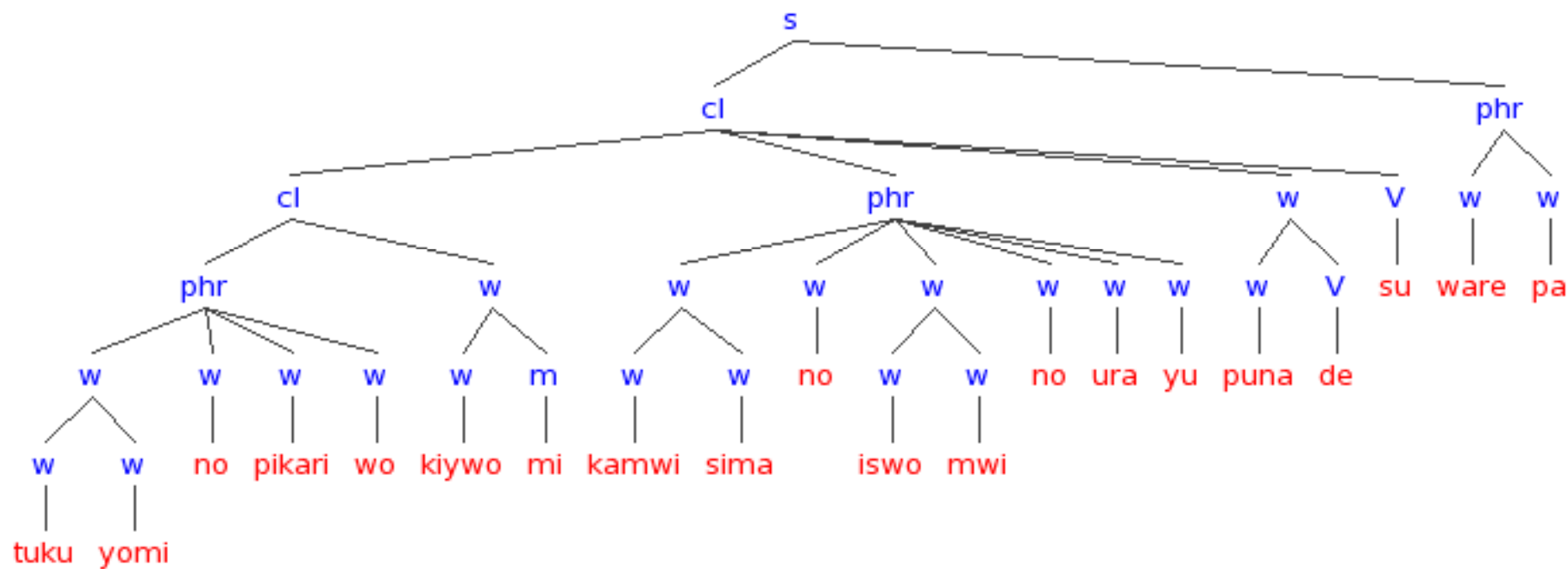
iswomwi no ura yu

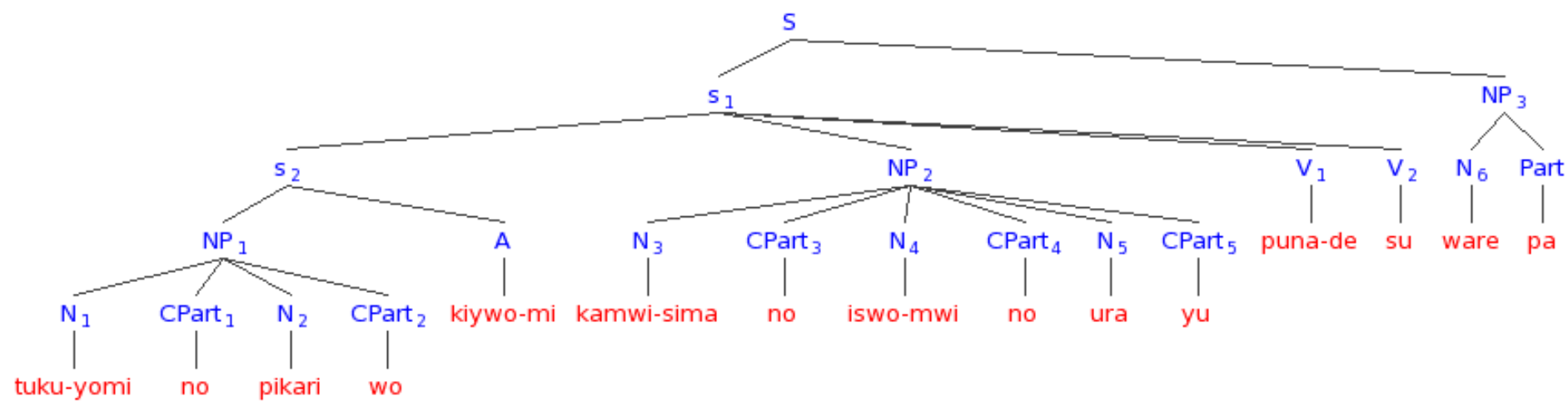
磯廻 ノ 浦 ヌ

船出須和礼波

puna-de su ware pa

船出 ス 我 ハ





検索

Searching

検索 Searching

語彙素と形態素のID番号 Lexeme and morpheme UID

=

Lemmatization

語彙素と形態素の表記と活用に関係なく検索可能

検索 Searching

単純検索 (一つの検索条件)

Simple search (a single search criterion)

KWIC (KeyWord In Context)

ある語彙素または形態素のすべての例

検索 Searching

高度検索 (検索条件の組み合わせ)

Advanced searches
(combining search criteria)

例 : all clauses (節)

whose predicate is in the conclusive form (終止形)

which contain a noun phrase (NP) that contains

a WH-word (疑問詞 有), but has

no focus particle (係助詞 無)

必要な情報

Necessary information

Writing 表記法

Part-of-speech 品詞

Morphology, inflection 活用

Clause 節

Noun phrase 名詞句

ID番号(Lemmatization)と辞書

UID (Lemmatization) and Lexicon

ID番号(Lemmatization)と辞書

UID (Lemmatization) and Lexicon

Another important function of the UIDs is to provide a link between the texts and the Lexicon associated with the corpus.

The Lexicon contains basic lexical information about each lexeme and morpheme, including POS, meaning or function, etc.

The Lexicon also contains information about the occurrence in the corpus of each lexeme and morpheme

Lexicon 辞書

「飲ム」の項目

nom-

L031386a

Part of speech: **verb**

Conjugation class: **quadrigrade**

1 Gloss: **drink**

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

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

The future 将来

The Oxford Corpus of Old Japanese
recognized as a
British Academy Research Project

2012年4月
大英学士院研究プロジェクトとして
認められた

追加するテキスト TEXTS to be added

Shoku nihongi Senmyō (続日本紀宣命; 797)

Engishiki Norito (延喜式祝詞)

Kojiki (古事記; 712)

追加する情報 INFORMATION to be added

- ▶ The syntactic information will be expanded. The next step is to add information about subject and object status of argument NPs.

シンタックス

項のNP : 主語と目的語の情報

追加する情報

INFORMATION to be added

- ▶ The annotation will be enriched beyond linguistic information to include literary, historical, geographical and other information.
文学的、歴史的、地理的、その他の情報
- ▶ The texts will be supplied with translations into English.
テキストの英訳

DICTIONARY

辞書

The dictionary associated with the OCOJ will be expanded to constitute a real bilingual

Old Japanese -- English dictionary

上代日本語・英語 辞書

The dictionary will be hyperlinked to the texts, making cross-reference in both directions possible.

COLLABORATION

Collaborating institutions

NINJAL

University of York

External members

Makiro TANAKA (NINJAL)

Toshinobu OGISO (NINJAL)

Alexander VOVIN (University of Hawai‘i)

Sven OSTERKAMP (Ruhr-Universität Bochum)

....

なぜローマ字化か

Why romanization?

なぜ音素表記か

Why phonemic transcription?

Phonological representation

当時の音韻の反映

Segmentation

形態素の分節

Phonological representation

当時の音韻の反映

Phonological representation

当時の音韻の反映

ハ行子音の音価: /p/

甲類・乙類音節の識別と解釈

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 ヌ

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

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

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

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

ツクよミの

比可里乎伎欲美

ヒカリヲキヨミ

神嶋乃

カミシマの

伊素未乃宇良由

イソみのウラユ

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_乙ミ_甲ノ_乙

比可里乎伎欲美

ヒ_甲カリヲキ_甲ヨ_甲ミ_甲

神嶋乃

カミ_乙シマノ_乙

伊素未乃宇良由

イソ_甲ミ_乙ノ_乙ウラユ

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₂ミ₁ノ₂

比可里乎伎欲美

ヒ₁カリヲキ₁ヨ₁ミ₁

神嶋乃

カミ₂シマノ₂

伊素未乃宇良由

イソ₁ミ₂ノ₂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

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

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

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

Phonological representation

当時の音韻の反映

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

Man'yōshū (万葉集) 3.341

賢跡

sakasi-mito

賢シミト

物言従者

mono-ipu yworī 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

Man'yōshū (万葉集) 3.341

賢跡

sakasi-mito

サカシミト

物言従者

mono-ipu yworī 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

Segmentation

形態素の分節

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
utukusiku

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	祢牟	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
utukusiku	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	祢牟	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
utukusiku	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	祢牟	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
utukusiku	si ga	katarapeba
adorably	he GEN	say.PROV

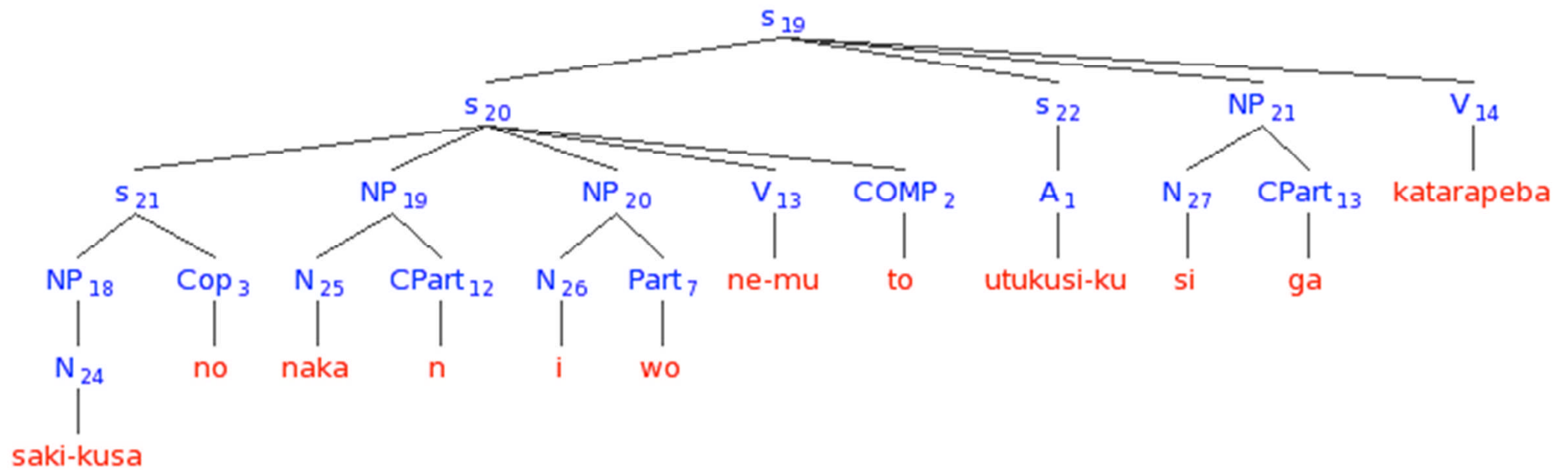
“When he adorably said that he wanted to sleep between us, like the *sakikusa* (three blade/stalk grass)”

万5.904

三枝之	中	爾	乎	祢牟	登	
三枝ノ	中	二	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
utukusiku	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)”



ご清聴ありがとうございました