

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/](http://www.orinst.ox.ac.uk/research/jap-ling/)

# Introduction of the research project 'Verb semantics and argument realization in pre-modern Japanese' and the Oxford Corpus of Old Japanese

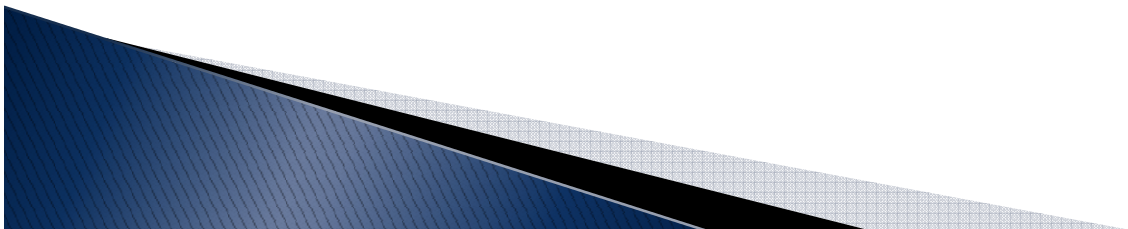
DiaSynCorp  
Workshop  
5–6 November 2011

Bjarke Frellesvig



# Verb semantics and argument realization in pre-modern Japanese

The Oxford Corpus of Old Japanese



# Verb semantics and argument realization in pre-modern Japanese:

## A comprehensive study of the basic syntax of pre-modern Japanese

Funding: The Arts and Humanities Research Council,  
UK, approximately £1 million

Period: 1 January 2009 – 30 September 2013

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

# People

- ▶ Bjarke Frellesvig (Principal Investigator)
- ▶ Peter Sells (Co-Investigator; York)
- ▶ Mary Dalrymple
- ▶ Postdoctoral Researchers
  - Stephen Wright Horn
  - Kerri L Russell
- ▶ Doctoral students
  - Zixi You (Split intransitivity in Old Japanese)
  - Daniel Trott (Aspect in Old Japanese)

# People

## External members

- ▶ Anton Antonov (INALCO and CRLAO, Paris)
- ▶ Satoshi Kinsui (Osaka University)
- ▶ Tomohide Kinuhata (Fukuoka University)
- ▶ Yasuhiro Kondo (Aoyama Gakuin/NINJAL)
- ▶ Masayoshi Shibatani (Rice University)
- ▶ Akira Watanabe (University of Tokyo)
- ▶ John Whitman (Cornell University/NINJAL)
- ▶ Yuko Yanagida (Tsukuba University)

Collaboration with the  
National Institute for Japanese  
Language and Linguistics

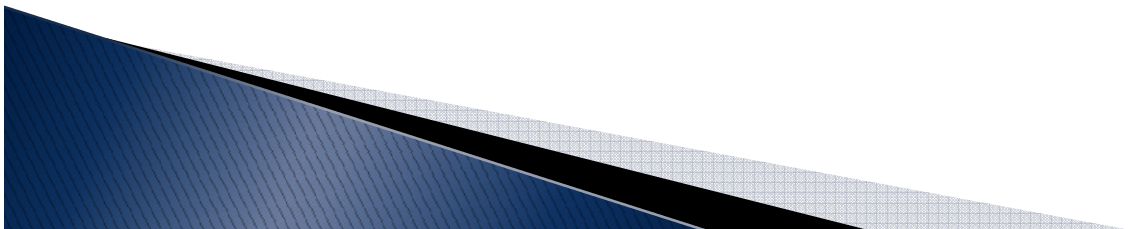
NINJAL

国立国語研究所

# VSARPJ

The research project investigates *argument realization* in detail for pre-modern Japanese.

Argument realization is a fundamentally important aspect of the syntax of a language which concerns the way in which verb meaning determines the number of arguments and their morpho-syntactic and semantic properties.



# Argument realization

‘Argument realization: the study of the possible syntactic expressions of the arguments of a verb’

(Beth Levin and Malka Rappaport Hovav, 2005, *Argument realization*, p. 1)

- (1) The boy broke the window with a ball
- (2) The boy hit the window with a ball
- (3) The window broke
- (4) \*The window hit
- (5) I broke his leg / \*I broke him on the leg
- (6) I hit his leg / I hit him on the leg



# Argument realization

- (7) kabe-ni            penki-o            nuru  
    wall-DAT        paint-ACC        smear  
    'smear paint on the wall'
- (8) kabe-ni            posutaa-o        haru  
    wall-DAT        poster-ACC      put up  
    'put posters up on the wall'

# Argument realization

- (9) penki-de            kabe-o            nuru  
paint-with          wall-ACC          smear  
'smear the wall with paint'
- (10) \*posutaa-de      kabe-o            haru  
poster-with        wall-ACC          put up
- (11) kabe-ni            posutaa-o        hari-tsukusu  
wall-on            poster-ACC        hang-exhaust  
'hang all the posters on the wall'
- (12) kabe-o            posutaa-de        hari-tsukusu  
wall-ACC          poster-with        hang-exhaust  
'hang the wall all over with posters'

# VSARPJ

## Synchronic part

A comprehensive account of the basic grammar of the Old Japanese language (approximately the 8th century).

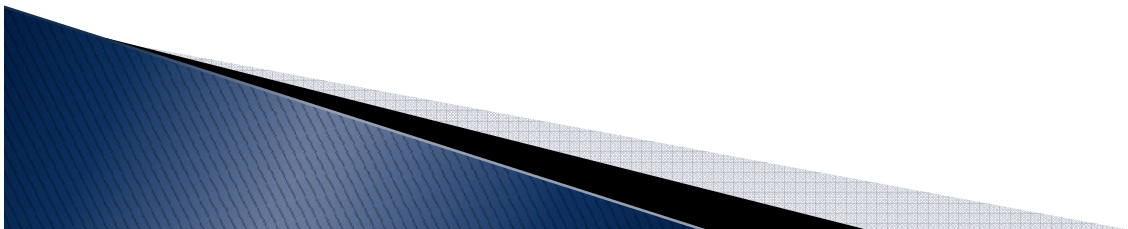
- The range of argument realization for individual verbs and for (semantically based) verb classes.
- Lexical semantic features relevant in the classification of verbs (for example, semantic roles, lexical aspect, telicity, type of motion, causation, unaccusativity/unergativity)?
- Patterns of argument realization.
- Other grammatical phenomena than argument realization which may be explained by verb semantics (for example, aspect, auxiliary selection, ellipsis, or specification of adjuncts or non-subcategorized arguments).

# VSARPJ

## Diachronic part

- ▶ An account of the development through time of verb semantics and argument realization, based on the results of the synchronic investigation.
- ▶ Inventory of changes through the history of Japanese in argument realization for individual verbs and for verb classes.
- ▶ General patterns of change, including possible development pathways for verb meanings and argument realization.

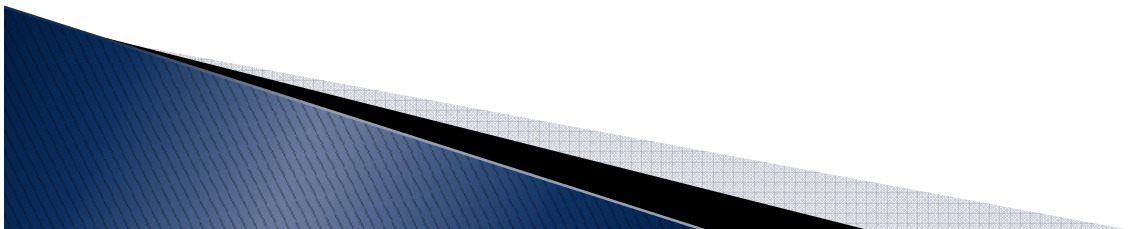
# The Oxford Corpus of Old Japanese



# Historical linguistics is corpus linguistics

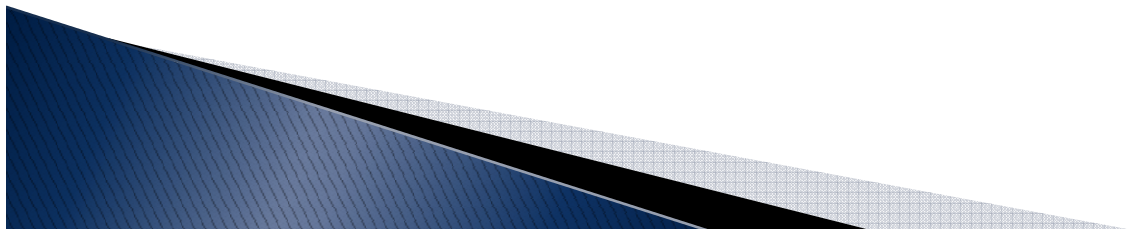
No recourse to native speaker intuition  
or directed testing with informants.

The only data are closed sets of texts in  
dead languages.



# Electronic texts and corpora

- ▶ Access and searchability
- ▶ Encoding of linguistic information

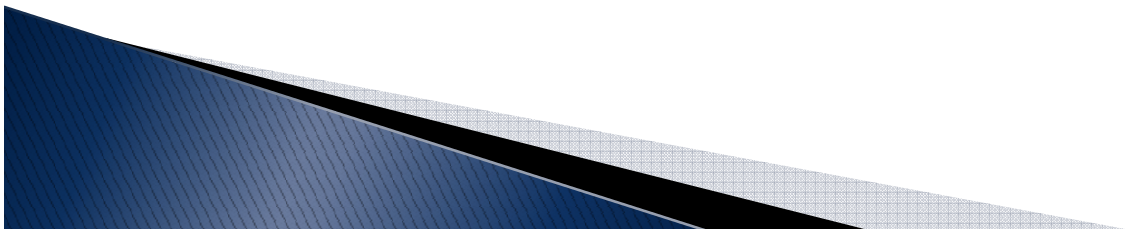


# Oxford Corpus of Old Japanese

Comprises all poetic texts from the Old Japanese period

approximately 90,000 words

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





# Old Japanese texts

## IN the Oxford Corpus of OJ

- ▶ *Kojiki kayō* 古事記歌謡 712
- ▶ *Nihon shoki kayō* 日本書紀歌謡 720
- ▶ *Fudoki kayō* 風土記歌謡 730s
- ▶ *Bussokuseki-ka* 仏足石歌 after 753
- ▶ *Man'yōshū* 万葉集 after 759
- ▶ *Shoku nihongi kayō* 続日本紀歌謡 797

## NOT in the Oxford Corpus of OJ

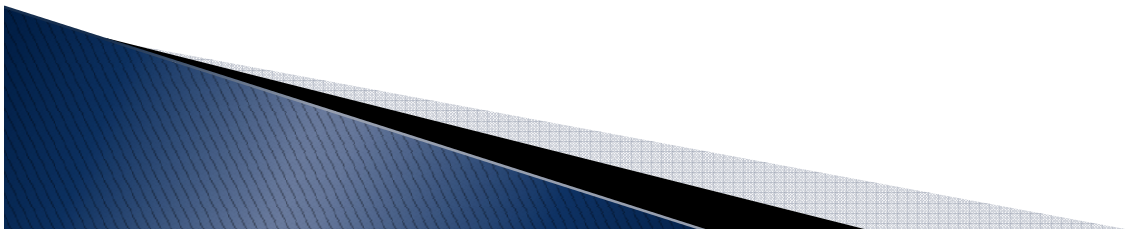
- ▶ *Shoku nihongi Senmyō* 宣命 697–791
- ▶ *Engishiki Norito* 祝詞 (compiled 927)

# Oxford Corpus of Old Japanese

Manual tagging

XML mark-up following the standards of the  
Text Encoding Initiative (TEI)

Phonemic transcription and original script,  
noting logographic and phonographic  
writing



# Oxford Corpus of Old Japanese

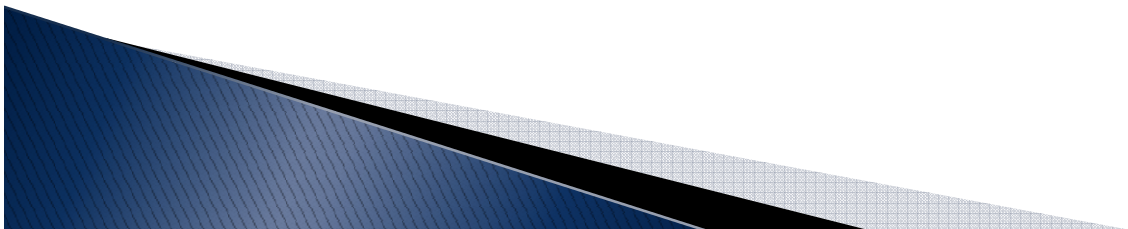
## Annotation

Lexeme and morpheme ID

Lexicon

Part of speech

Morphology



# Oxford Corpus of Old Japanese

## Annotation

Syntactic constituency:

noun phrases

clauses

topics

right dislocated elements

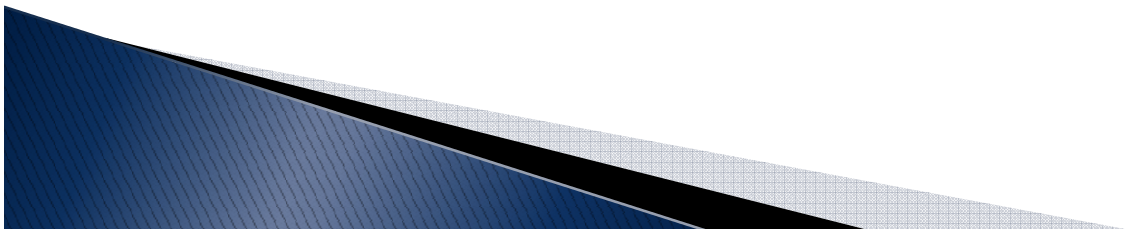
(Almost) no empty elements

Argumenthood

# Oxford Corpus of Old Japanese

## Annotation

Semantic roles (discontinued)



# Text example

## *Man'yōshū* 8.1606

君待跡

吾戀居者

我屋戸乃

簾令動

秋之風吹

君待跡

kimi      matu *to*  
my.lord   wait   as

吾戀居者

wa ga kwopwi woreba  
I GEN love      PROG.as

我屋戸乃

wa ga yadwo *no*  
I GEN hut      GEN

簾令動

sudare ugwokasi  
curtain move

秋之風吹

aki      no      kaze puku  
autumn GEN wind blow

# Writing

- ▶ Phonographic writing

- 我屋戸乃      wa ga yadwo *no*

- ▶ Logographic writing

- 我屋戸乃      wa ga yadwo *no*

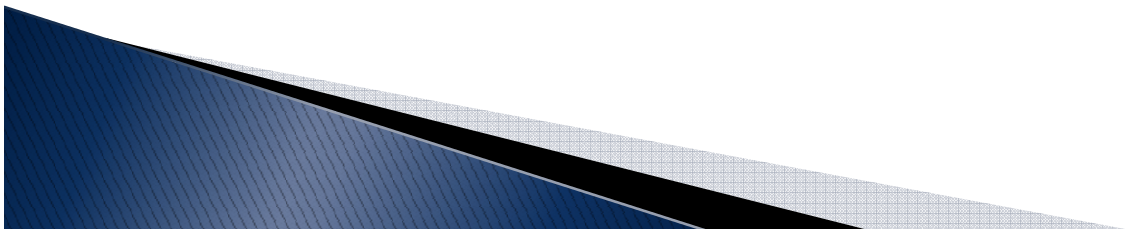
- No writing

- 我屋戸乃      wa ga yadwo *no*



# The problem with logographic writing: Identification

為者



# The problem with logographic writing: Identification

為者

*sure-ba*

provisional  
'as, when'

*se-ba*

conditional  
'if'

*suru pa*

adnominal + TOP

# Morphology

**kimi**<sub>(noun)</sub> **matu**<sub>(verb, adnconc)</sub> **to**<sub>(particle, conj)</sub>

**wa**<sub>(pronoun)</sub> **ga**<sub>(particle, case, gen)</sub> **kwopwi**<sub>(verb, infinitive)</sub> **woreba**<sub>(verb, provisional)</sub>

**wa**<sub>(pronoun)</sub> **ga**<sub>(particle, case, gen)</sub> **yadwo**<sub>(noun)</sub> **no**<sub>(particle, case, gen)</sub>

**sudare**<sub>(noun)</sub> **ugwokasi**<sub>(verb, infinitive)</sub>

**aki**<sub>(noun)</sub> **no**<sub>(particle, case, gen)</sub> **kaze**<sub>(noun)</sub> **puku**<sub>(verb, adnconc)</sub>

# {Clauses} and <Phrases> arguments and adjuncts

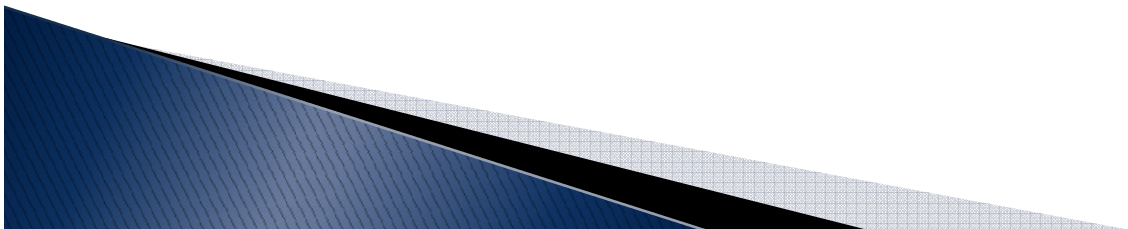
{ { { < kimi<sub>(noun)</sub> > matu<sub>(verb, adnconc)</sub> to<sub>(particle, conj)</sub> }  
< wa<sub>(pronoun)</sub> ga<sub>(particle, case, gen)</sub> > kwopwi<sub>(verb, infinitive)</sub> woreba<sub>(verb, provisional)</sub> }  
{ < wa<sub>(pronoun)</sub> ga<sub>(particle, case, gen)</sub> yadwo<sub>(noun)</sub> no<sub>(particle, case, gen)</sub>  
sudare<sub>(noun)</sub> > ugwokasi<sub>(verb, infinitive)</sub> }  
< aki<sub>(noun)</sub> no<sub>(particle, case, gen)</sub> kaze<sub>(noun)</sub> > puku<sub>(verb, adnconc)</sub> }

# The Oxford Corpus of OJ provides

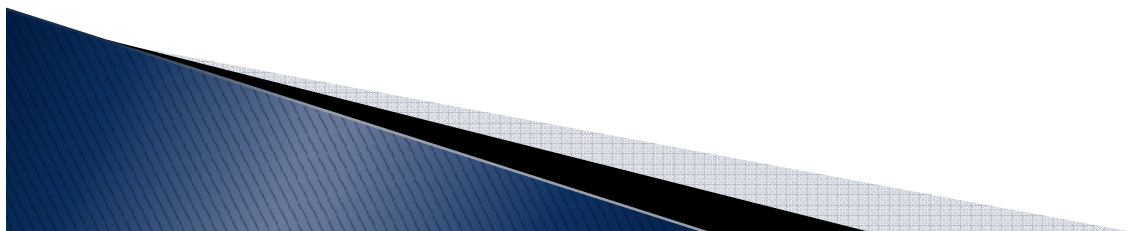
- ▶ Easy access to the texts
- ▶ Searchability of individual words, especially verbs, regardless of their inflected form; or of sets of words
- ▶ Searchability of syntactic frames for individual verbs
- ▶ Searchability of structural patterns, e.g. case marking and word order, very generally, or limited to (sets of) specific verbs

# Further perspectives for the OCOJ

- ▶ Flexibility
- ▶ Expandability
- ▶ Other applications
  - Linguistics
  - Literature
  - History



# XML mark-up



```
<text> <body> <div> <ab xml:id="MYS.8.1606"> <s> <cl> <cl> <phr> <w
type="noun" ana="#L004266"> <c type="logo">kimi</c> </w> </phr> <w
type="verb" inflection="adnconc" ana="#L031644"> <c type="logo">matu</c> </w>
<w type="particle" subtype="conj" ana="#L000541"> <c type="phon">to</c> </w>
</cl> <lb/> <cl> <phr type="arg"> <w type="pronoun" ana="#L042057"> <c
type="logo">wa</c> </w> <w type="particle" subtype="case" function="gen"
ana="#L000503"> <c type="noLogo">ga</c> </w> </phr> <w type="verb"
inflection="infinitive" ana="#L030731"> <c type="logo">kwopwi</c> </w> <w
type="verb" inflection="provisional" ana="#L031957" function="progressive">
<c type="logo">woreba</c> </w> </cl> <lb/> <cl type="adjunct"> <phr
type="arg"> <w type="pronoun" ana="#L042057"><c type="logo">wa</c> </w> <w
type="particle" subtype="case" function="gen" ana="#L000503"> <c
type="noLogo">ga</c> </w> <w type="noun" ana="#L050030"> <c
type="logo">yadwo</c> </w> <w type="particle" subtype="case" function="gen"
ana="#L000520"> <c type="phon">no</c> </w> <lb/> <w type="noun"
ana="#L050031"> <c type="logo">sudare</c> </w> </phr> <w type="verb"
inflection="infinitive" ana="#L030247"> <c type="logo">ugwokasi</c> </w>
```



ご清聴ありがとうございました

## Websites

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

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

