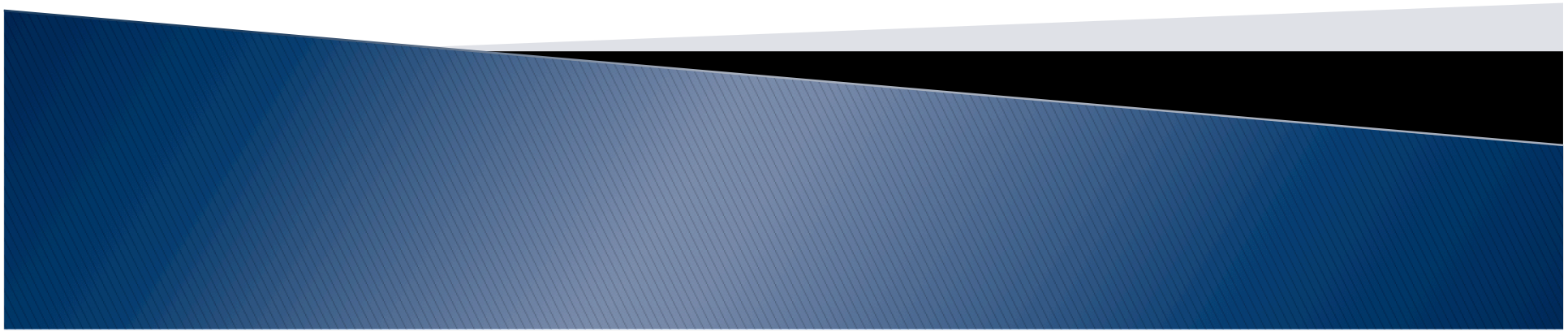
Research Centre for Japanese Language and Linguistics
University of Oxford

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/

Introducing the research project "Verb semantics and argument realization in pre-modern Japanese"

Bjarke Frellesvig, Stephen W Horn,
Kerri L Russell, Peter Sells



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

The Project (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 postaa-o haru
wall-DAT poster-ACC put up
'put posters up on the wall'

(Yuhki S. King. *Spray/Load verbs in Japanese: An Event Structure Perspective.* MA thesis, SOAS, 2008)

Argument realization

(9) penki-de kabe-o nuru
paint-with wall-ACC smear
'smear the wall with paint'

(10) *postaa-de kabe-o haru
poster-with wall-ACC put up

(11) kabe-ni postaa-o hari-tsukusu
wall-on poster-ACC hang-exhaust
'hang all the posters on the wall'

(12) kabe-o postaa-de hari-tsukusu
wall-ACC poster-with hang-exhaust
'hang the wall all over with posters'

The Project (VSARPJ)
Synchronic part

- A comprehensive account of the basic grammar of each of the stages of the Japanese language, from the beginning of its recorded history in the 8th century until the end of the 16th 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).

The Project (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.

VSARPJ: People

- Bjarke Frellesvig (Principal Investigator)
- Peter Sells (Co-Investigator, SOAS)
- Postdoctoral Researchers
  - Stephen Wright Horn (AHRC)
  - Kerri L Russell (AHRC)
  - Janick Wrona (Senior Research Associate)
- Doctoral students
  - Yuhki King (AHRC)
  - Zixi You
- Mary Dalrymple

VSARPJ: People

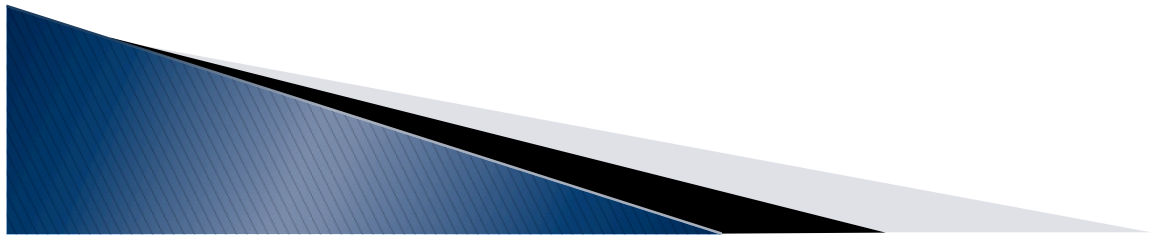
External members

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

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.



Texts
Old Japanese (– 800)

- *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 Senmyō* 宣命 697–791
- *Shoku nihongi kayō* 続日本紀歌謡 797
- *Engishiki Norito* 祝詞 (compiled 927)

Texts
Early Middle Japanese (800–1200)

- *Kokin wakashū preface* 古今和歌集仮名序 914
- *Ise monogatari* 伊勢物語 early 10th c.
- *Tosa nikki* 土佐日記 935
- *Taketori monogatari* 竹取物語 mid 10th c.
- *Kagerō nikki* 蜻蛉日記 2nd half 10th c.
- *Ochikubo monogatari* 落窪物語 late 10th c.
- *Makura no sōshi* 枕草子 c. 1000
- *Genji monogatari* 源氏物語 1001–1010
- *Sarashina nikki* 更級日記 1059–1060
- *Konjaku monogatari-shū* 今昔物語集 1120

Texts
Late Middle Japanese (1200–1600)

- *Hōjōki* 方丈記 1212

- *Esopo no fabulas* 1593
- *Feiqe monogatari* 1593

Electronic texts and corpora

- Access and searchability
- Encoding of linguistic information

Features of the The Oxford Corpus of pre-modern Japanese

- Phonemic transcription
- XML mark-up following the internationally recognized standards of the Text Encoding Initiative (TEI)

- Orthography
- Part of speech
- Morphology
- Lexeme and morpheme ID; Lexicon
- Syntax

Phonemic transcription

- 恋 'love'
  - NJ *koi*
  - historical kana spelling: こひ

Phonemic transcription

< 800		800		950		1000		1100
OJ *kwopwi*	>	EMJ *kwopi*	>	*kopi*	>	*kowi*	>	*koi*

Orthography

- 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*

Parts of speech

- Main lexical and grammatical word classes.
  - Nouns, adverbs, verbs, adjectives, adjectival nouns, copula, numerals.
  - Auxiliary verbs, particles, auxiliaries, extensions, prefixes, suffixes, counters.

Morphology
Inflectional categories

- **Finite**

Conclusive	*sinu*
Adnominal	*sinuru*
Exclamatory	*sinure*
Imperative	*sine*
Neg. conject.	*sinazi*
Optative	*sinana*
Prohibitive	*na sini so*

Morphology
Inflectional categories

- **Non-finite**

- Infinitive *sini*
- Gerund *sinite*
- Continuative *sinitutu*
- Conditional *sinaba*
- Provisional *sinureba*
- Concessive *sinuredo*

- **Nominal** *sinuraku*

Morphology
Inflectional categories

- Additional categories in the mark-up
  - Adn/conc *yuku*
  - Inf/conc *ari*

Lexicon

- Unique ID numbers for lexical and grammatical elements.
- Lexical information.

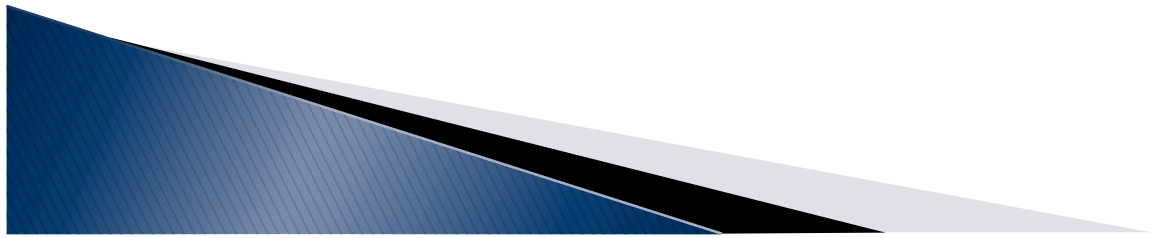
Lexical entry

From the 8th century: **kwopwi-** >
From 800: **kwopi-** >
From before 950: **kopi-** >
From c. 950–1000: **kowi-** >
From c. 1100: **koi-**

Part of speech: **verb**
Conjugation class: **upper bigrade**
Gloss: **love**

Syntax

- Syntactic categories
  - Clause
  - Phrase
- Syntactic relation types
  - Argument
  - Adjunct



Syntax

- Topics
- High and low topics
- Right dislocated elements
- Focused elements
- Nominalized clauses
- Clause types (main, subordinate, adnominal)
- *Kakari-musubi*
- Noun phrase heads
- Particle scope

Syntax

- Clause predicates
- Analytic predicates
- Zero marked arguments
- Topicalized arguments.
- Order of case marked and zero marked arguments, including ordering relative to focused elements.

Syntax

- Syntactic frames for individual verbs, within individual stages of the language and across different stages.

Text example

君待跡 / 吾戀居者 / 我屋戸乃 / 簾令動 / 秋之風吹
きみまつと / わがこひをれば / わがやどの /
すだれうごかし / あきのかぜふく　(MYS 8.1606)

君之行 / 氣長成奴 / 山多都祢 / 迎加将行 / 待尓可将待
きみがゆき / けながくなりぬ / やまたづね /
むかへかゆかむ / まちにかまたむ　(MYS 2.85)

如此許 / 戀乍不有者 / 高山之 / 磐根四巻手 / 死奈麻死物呼
かくばかり / こひつつあらずは / たかやまの /
いはねしまきて / しなましものを　(MYS 2.86)

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

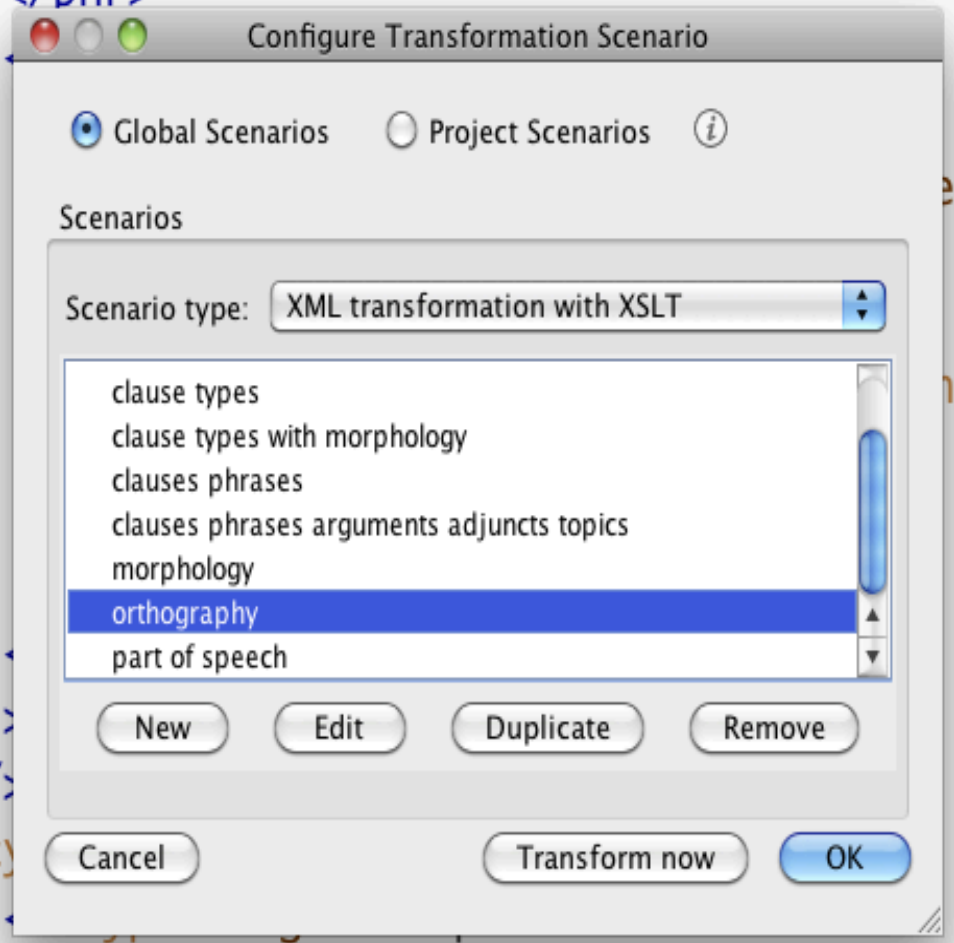
秋之風吹　　aki　no　kaze puku
autumn GEN wind blow

君之行	kimi ga yuki my.lord GEN go
氣長成奴	*ke* nagaku nari-*nu* day long become-PERF
山多都祢	yama *tadune* mountain visit
迎加将行	mukape *ka* yuka-mu meet FOC go-CONJ
待尓可将待	mati *ni* *ka* mata-mu wait DAT FOC wait-CONJ

如此許	kaku bakari this.way only
戀乍不有者	kwopwitutu ara-zupa loving be(STAT)-NEG.if
高山之	takayama no Takayama GEN
磐根四巻手	ipane *si* maki*te* rock EMPH rolling
死奈麻死物呼	si*na-masi* monowo die-OPT EMPH

```
<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>
```

Display options



Orthography

kimi matu *to*
wa ga kwopwi woreba
wa ga yadwo *no*
sudare ugwokasi
aki no kaze puku

kimi ga yuki
ke nagaku nari*nu*
yama *tadune*
mukape *ka* yukamu
mati *ni ka* matamu

Part of Speech

kimi ga yuki
ke nagaku nari*nu*
yama *tadune*
mukape *ka* yukamu
mati *ni ka* matamu

kaku bakari
kwopwitutu arazupa
taka yama no
ipane *si* maki*te*
si*namasi* monowo

Morphology

kimi(noun) matu(verb, adnconc) *to*(particle, conj)

wa(pronoun) ga(particle, case, gen) kwopwi(verb, infinitive) woreba(verb, provisional)

wa(pronoun) ga(particle, case, gen) yadwo(noun) *no*(particle, case, gen)

sudare(noun) ugwokasi(verb, infinitive)

aki(noun) no(particle, case, gen) kaze(noun) puku(verb, adnconc)

Morphology

kimi(noun) ga(particle, case, gen) yuki(verb, infinitive)

ke(noun) naga(adjective)-ku(adjcop, infinitive) nari(verb, stem)-*nu*(auxiliary, conclusive)

yama(noun) *tadune*(verb, infinitive)

mukape(verb, infinitive) *ka*(particle, foc) yuka(verb, stem)-mu(auxiliary, adnconc)

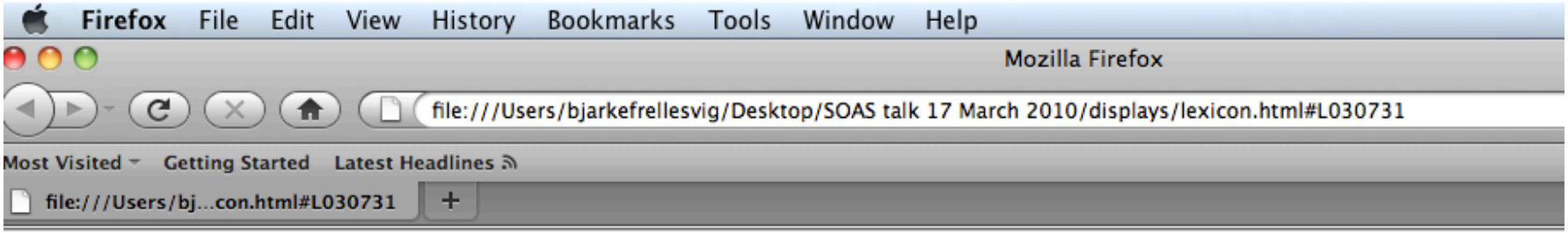
mati(verb, stem) *ni*(particle, case, dat) *ka*(particle, foc) mata(verb, stem)-mu(auxiliary, adnconc)

Morphology

kaku (adverb) bakari (particle, res)
kwopwitutu (verb, continuative) ara (verb, stem) -zupa (auxiliary, conditional)
takataka (adjective, stem) yama (noun) no (particle, case, gen)
ipane (noun) *si* (particle, res) maki*te* (verb, gerund)
si*na* (verb, stem) -*masi* (auxiliary, adnconc) monowo (particle, conj)

Lexicon

kimi(noun) matu(verb, adnconc) *to*(particle, conj)
wa(pronoun) ga(particle, case, gen) kwopwi(verb, infinitive) woreba(verb, provisional)
wa(pronoun) ga(particle, case, gen) yadwo(noun) *no*(particle, case, gen)
sudare(noun) ugwokasi(verb, infinitive)
aki(noun) no(particle, case, gen) kaze(noun) puku(verb, adnconc)



From the 8th century: **kwopwi-** > From 800: **kwopi-** > From before 950: **kopi-** > From c. 950-1000: **kowi-** > From c. 1100: **koi-**
Part of speech: **verb**
Conjugation class: **upper bigrade**
Gloss: **love**

From the 8th century: **kwopwi** > From 800: **kwopi** > From before 950: **kopi** > From c. 950-1000: **kowi** > From c. 1100: **koi**
Part of speech: **noun**

From the 8th century: **kwopwiamar-** > From 800: **kwopiamar-** > From before 950: **kopiamar-** > From c. 950-1000: **kowiamar-** > From c. 11
Part of speech: **verb**
Conjugation class: **quadrigrade**
Gloss: **be infatuated**

From the 8th century: **kwopwiamari** > From 800: **kwopiamari** > From before 950: **kopiamari** > From c. 950-1000: **kowiamari** > From c. 11
Part of speech: **noun**

From the 8th century: **kwopwisin-** > From 800: **kwopisin-** > From before 950: **kopisin-** > From c. 950-1000: **kowisin-** > From c. 1100: **koisin**
Part of speech: **verb**
Conjugation class: **quadrigrade**
Gloss: **die of love**

From the 8th century: **kwopwisini** > From 800: **kwopisini** > From before 950: **kopisini** > From c. 950-1000: **kowisini** > From c. 1100: **koisini**
Part of speech: **noun**

Syntax
{Clauses} and <Phrases>

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

{Clauses} and <Phrases> arguments and adjuncts

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

{Clauses} and <Phrases> arguments and adjuncts

{ { < kimi$_{(noun)}$ ga$_{(particle, case, gen)}$ > yuki$_{(verb, infinitive)}$ }
{ < *ke*$_{(noun)}$ > naga$_{(adjective)}$- ku$_{(adjcop, infinitive)}$ } nari$_{(verb, stem)}$- *nu*$_{(auxiliary, conclusive)}$ }
{ < { { < yama$_{(noun)}$ > *tadune*$_{(verb, infinitive)}$ }
mukape$_{(verb, infinitive)}$ } *ka*$_{(particle, foc)}$ > yuka$_{(verb, stem)}$- mu$_{(auxiliary, adnconc)}$ }
{ < { mati$_{(verb, stem)}$ } *ni*$_{(particle, case, dat)}$ *ka*$_{(particle, foc)}$ > mata$_{(verb, stem)}$- mu$_{(auxiliary, adnconc)}$ }

{Clauses} and <Phrases> arguments and adjuncts

{ { < kaku$_{(\text{adverb})}$ bakari$_{(\text{particle, res})}$ >
kwopwitutu$_{(\text{verb, continuative})}$ ara$_{(\text{verb, stem})}$- zupa$_{(\text{auxiliary, conditional})}$ }
{ < takataka$_{(\text{adjective, stem})}$ yama$_{(\text{noun})}$ no$_{(\text{particle, case, gen})}$
ipane$_{(\text{noun})}$ *si*$_{(\text{particle, res})}$ > maki*te*$_{(\text{verb, gerund})}$ }
si*na*$_{(\text{verb, stem})}$- *masi*$_{(\text{auxiliary, adnconc})}$ monowo$_{(\text{particle, conj})}$ }

{Clauses} and <Phrases>
arguments adjuncts topics
right dislocated elements

{ { < yama$_{(noun)}$ no$_{(particle, case, gen)}$ *pa*$_{(noun)}$ *ni*$_{(particle, case, dat)}$ >
< *adimura*$_{(noun)}$ > { sawaki$_{(verb, infinitive)}$ }
yuku$_{(verb, adnconc)}$ *naredo*$_{(extension, concessive)}$ }
< ware$_{(pronoun)}$ pa$_{(particle, top)}$ > { *sabusi*$_{(adjective)}$- $_{(adjcop, conclusive)}$ *we*$_{(particle, intj)}$ }
{ { < kimi$_{(noun)}$ > *ni*$_{(copula, infinitive)}$ *si*$_{(particle, res)}$ } ara$_{(verb, stem)}$ - ne*ba*$_{(auxiliary, provisional)}$ } }

{Clauses} and <Phrases>
arguments adjuncts topics

< { < { < *wominapyesi*$_{(noun)}$ >
saku$_{(verb,\ adnconc)}$ } sapa$_{(noun)}$ *ni*$_{(particle,\ case,\ dat)}$ > opu*ru*$_{(verb,\ adnominal)}$ }
pana*katumi*$_{(noun)}$ >
{ < { < *katute*$_{(adverb)}$ *mo*$_{(particle,\ top)}$ > { sira$_{(verb,\ stem)}$ - nu$_{(auxiliary,\ adnominal)}$ } }
kwopwi$_{(noun)}$ *mo*$_{(particle,\ top)}$ > { *suru*$_{(verb,\ adnominal)}$ *kamo*$_{(particle,\ finl,\ emph)}$ } }

Clause types: main, subordinate1, subordinate2, adnominal

kimi matu *to*
wa ga kwopwi woreba
wa ga yadwo *no*
sudare ugwokasi
aki no kaze puku

Clause types: main, subordinate1, subordinate2, adnominal

kimi ga yuki
ke nagaku nari*nu*
yama *tadune*
mukape *ka* yukamu
mati *ni ka* matamu

Clause types: main, subordinate1, subordinate2, adnominal, nominal

*tukusi*bune

imada *mo* koneba

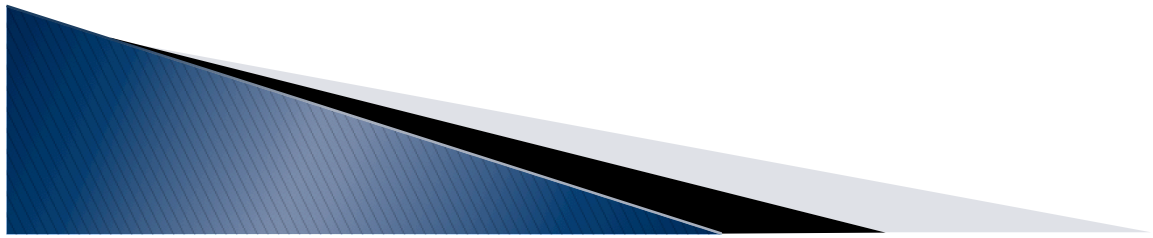
arakazime

araburu kimi *wo*

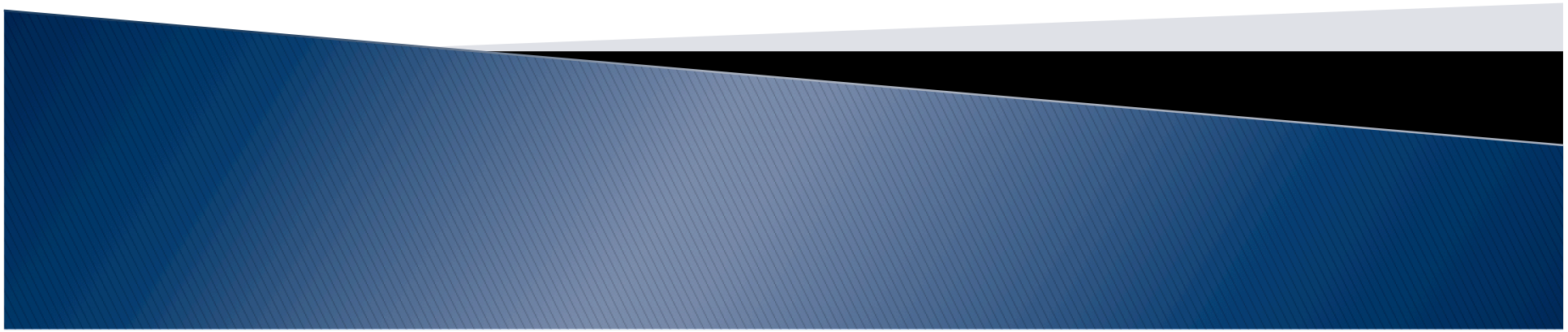
miru ga kanasi*sa*

Further perspectives for the corpus

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



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/

ご清聴ありがとうございました