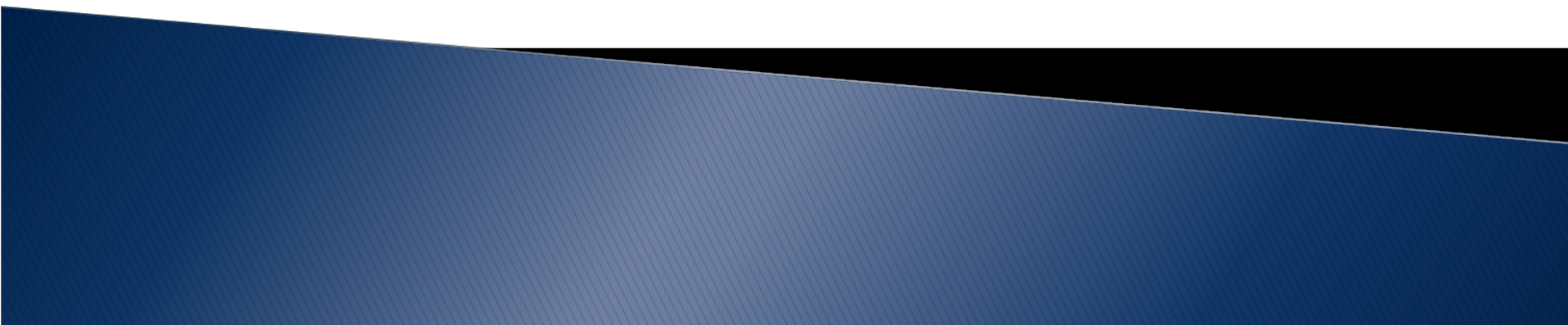


Argument structure in Old Japanese:

A corpus based study

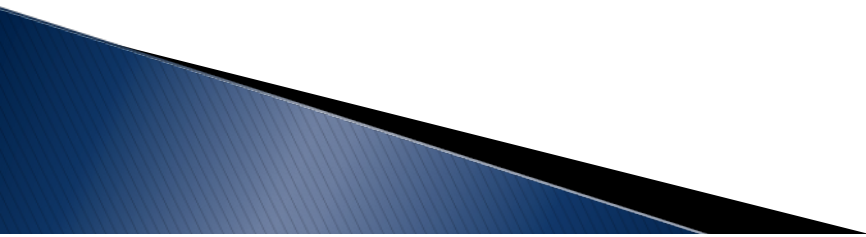
Kerri L Russell

Japanese Times Now Past (11-12 June 2010)



The VSARPJ Project

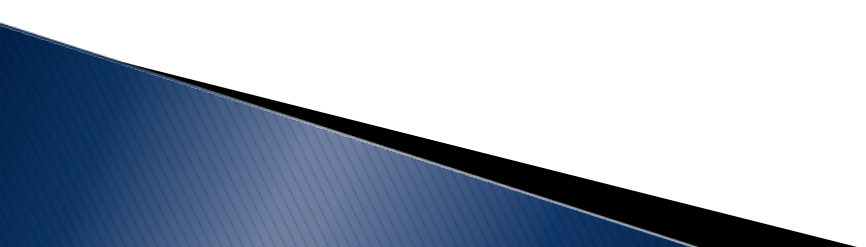
- Background
 - Goals
 - Create a tool for linguistic analyses
 - Argument structure
 - Texts
 - Funding
 - Arts & Humanities Research Council (AHRC)



The VSARPJ Project

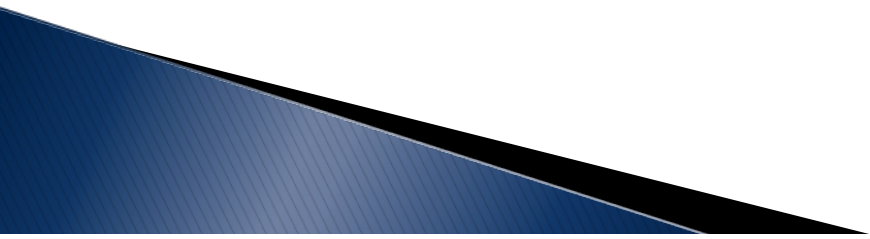
- People

- Bjarke Frellesvig – Principal Investigator
- Peter Sells (SOAS) – Co-Investigator
- Stephen Wright Horn & Kerri L Russell – Postdoctoral researchers
- Yuhki King & Zixi You – DPhil students
- External Members



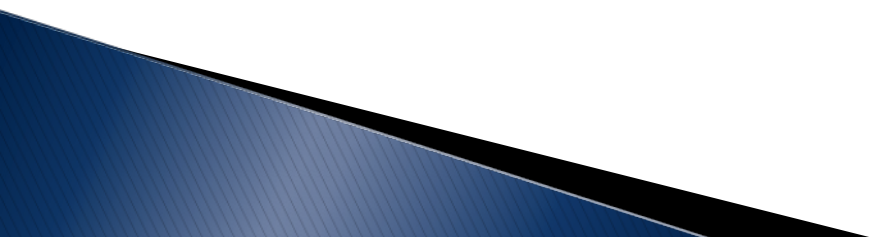
Markup

- TEI
- XML



Orthography

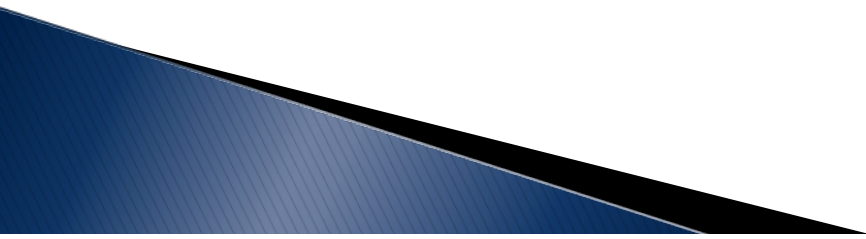
- Preserve information about orthography
 - Logographic
 - Phonographic
 - Not recorded in text
 - Large or small phonograms
- Character tags: `<c>word</c>`
- Character attributes: `@type="some value"`
`<c type="phon">word</c>`



Morphology

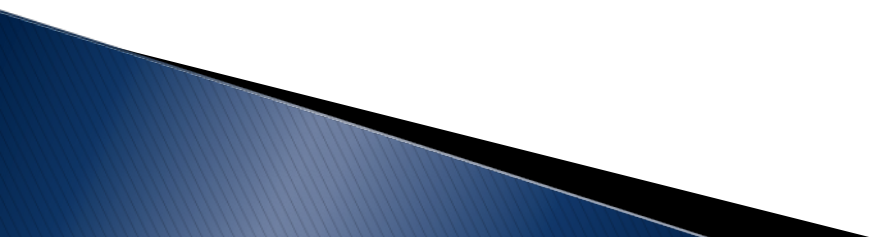
- The <w/> tag
- @type: POS
 - Verbs
 - @inflection
 - Statives and progressives: @function
 - Particles
 - @subtype
 - Case & final particles: @function

<w type="verb" inflection="infinitive">verb</w>



Morphology

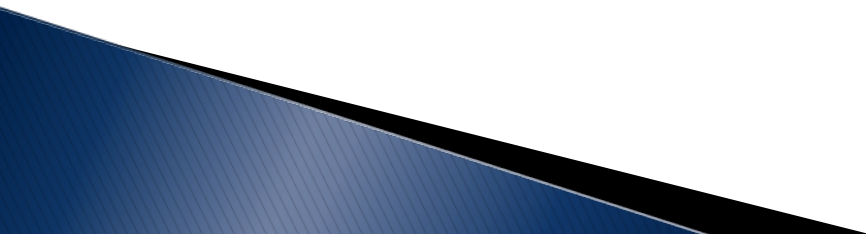
- The <m/> tag
- @type: auxiliaries, adjectival copulas, prefixes, suffixes, circumfix, counters, numerals
 - Auxiliaries
 - @function
 - @inflection
 - Adjectival copula
 - @inflection
 - Verbal prefixes
 - @function



The dictionary

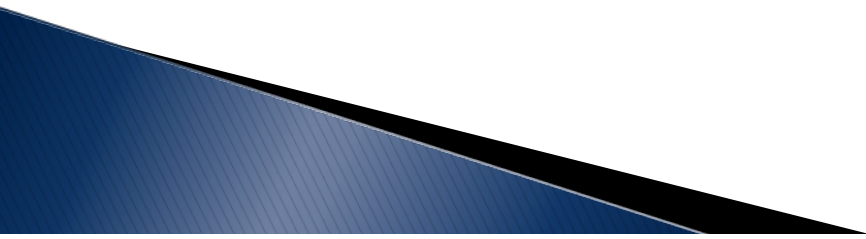
- Search words through time
- All `<w/>` and `<m/>` are assigned numbers
 - `<w type="noun" ana="#L004266">kimi</w>`
- This links to an entry in the dictionary which contains:
 - Unique identification number
 - Pronunciation changes
 - POS
 - Usage
 - Start/end
 - Dialect (EOJ)

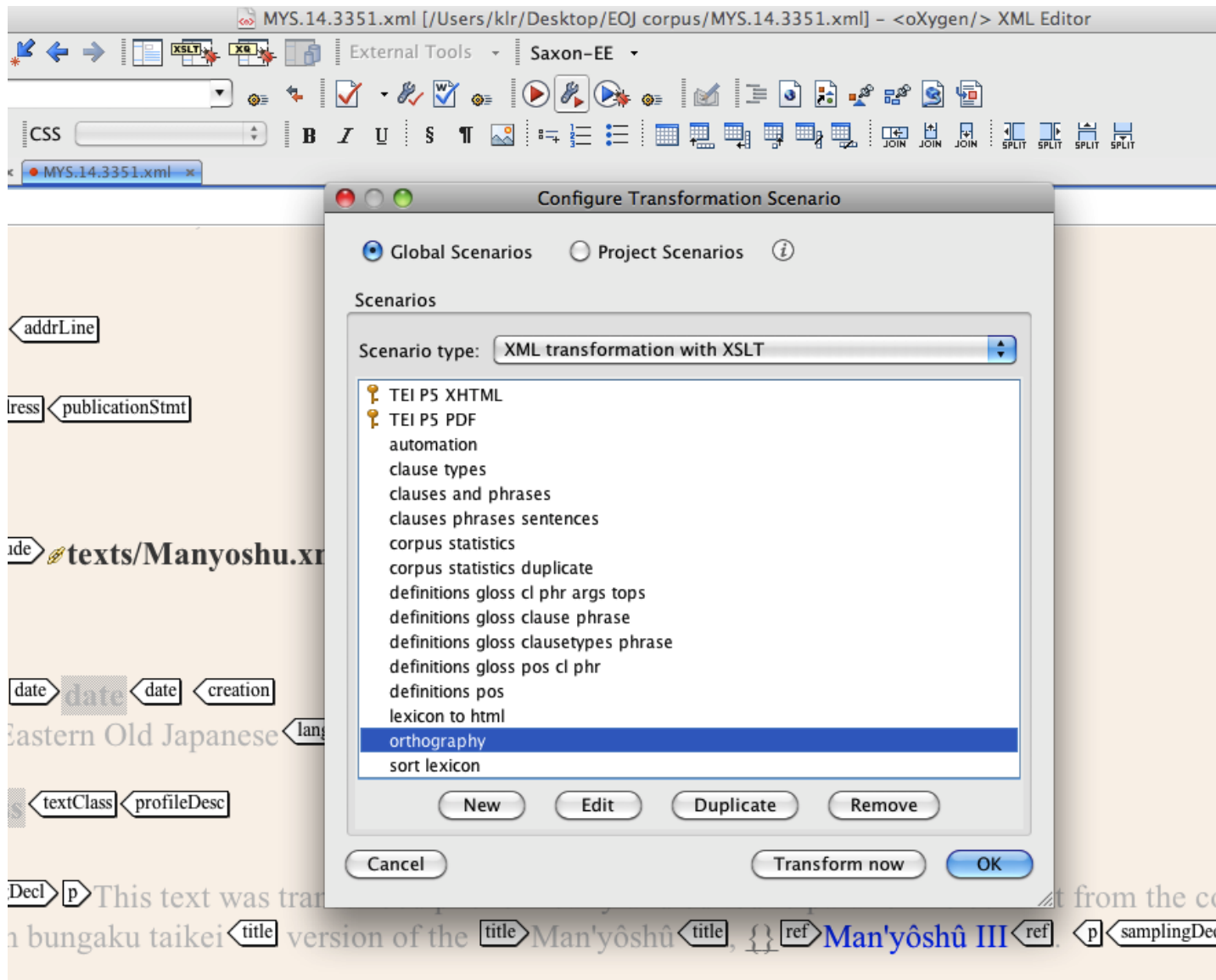
Syntax

- Syntactic Markup
 - `<phr/>`
 - `<cl/>`
 - `<s/>`
 - Arguments and adjuncts (@type)
 - Clauses can be main, subordinate, etc. (not marked)
 - Topics (not marked)
 - Right dislocated elements (not marked)
- 

Eastern Old Japanese

- EOJ poems found in *Man'yōshū* Books 14 and 20
- Each poem is identified by dialect area
- Ability to conduct various searches





Eastern Old Japanese

- Orthography

筑波祢爾

由伎可母布良留

伊奈乎可母

加奈思吉兒呂我

爾努保佐流可母

tukupane ni

yuki ka mo puraru

ina wo ka mo

kanasiki *kworo* ga

ninwo posaru kamo

MYS.14.3351

Eastern Old Japanese

- Part of speech

筑波祢爾

由伎可母布良留

伊奈乎可母

加奈思吉兒呂我

爾努保佐流可母

tukupane ni

yuki ka mo puraru

ina wo ka mo

kanasiki kwo ro ga

ninwo posaru kamo

MYS.14.3351

Eastern Old Japanese

- Clauses and phrases

筑波祢爾

由伎可母布良留

伊奈乎可母

加奈思吉兒呂我

爾努保佐流可母

[{ < tukupane ni > }

< yuki ka mo > { puraru }]

[< ina wo ka mo >

{ < { kanasiki } *kworo* ga >

< ninwo > posaru kamo }]

MYS.14.3351

Eastern Old Japanese

- Clause types

信濃奈流
須我能安良能爾
保登等芸須
奈久許惠伎氣婆
登伎須疑爾家里

sinano n aru
suga no arano ni
pototogisu
naku kowe kikeba
toki suginikyeri

MYS.14.3352

Eastern Old Japanese

信濃奈流

須我能安良能爾

保登等芸須

奈久許惠伎気婆

登伎須疑爾家里

[{ { < { < { < *sinano*_(noun) *n*_(particle, case, dat) > *aru*_(verb, adnominal) }
*suga*_(noun) *no*_(particle, case, gen) *arano*_(noun) *ni*_(particle, case, dat) >
< *pototogisu*_(noun) >
*naku*_(verb, adnconc) } *kowe*_(noun) > *kikeba*_(verb, provisional) }
< *toki*_(noun) > *sugi*_(verb, stem)- *ni*_(auxiliary, stem)- *kyeri*_(auxiliary, conclusive) }]

MYS.14.3352

Eastern Old Japanese

筑波祢爾

由伎可母布良留

伊奈乎可母

加奈思吉兒呂我

爾努保佐流可母

[{ < tukupane_(noun) ni_(particle, case, dat) > }
< yuki_(noun) ka_(particle, foc) mo_(particle, top) > { pur_(verb, stem)- aru_(auxiliary, adnominal) }]
[< ina wo_(particle, intj) ka_(particle, foc) mo_(particle, top) >
{ < { kanasi_(adjective)- ki_(adjcop, adnominal) } kwo_(noun) ro_(suffix) ga_(particle, case, gen) >
< ninwo_(noun) > pos_(verb, stem)- aru_(auxiliary, adnominal) kamo_(particle, finl, emph) }]

Eastern Old Japanese

筑波祢爾

由伎可母布良留

伊奈乎可母

加奈思吉兒呂我

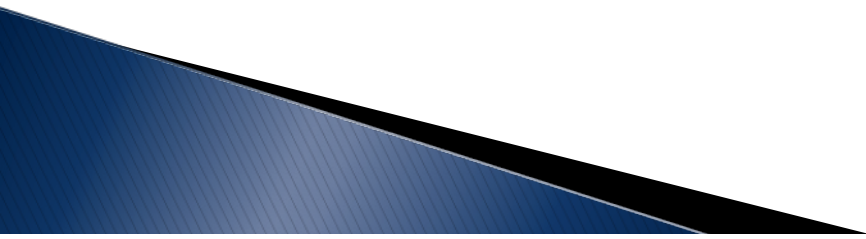
爾努保佐流可母

[{ < tukupane_(noun) ni_(particle, case, dat) > }
< yuki_(noun) ka_(particle, foc) mo_(particle, top) > { pur_(verb, stem)- aru_(auxiliary, adnominal) }]
[< ina wo_(particle, intj) ka_(particle, foc) mo_(particle, top) >
{ < { kanasi_(adjective)- ki_(adjcop, adnominal) } kwo_(noun) ro_(suffix) ga_(particle, case, gen) >
< ninwo_(noun) > pos_(verb, stem)- aru_(auxiliary, adnominal) kamo_(particle, finl, emph) }]

Eastern Old Japanese

- Definitions

From the 8th century: **ro**
Part of speech: **bound morpheme**
Gloss: [**diminutive suffix**]



Eastern Old Japanese

- Corpus statistics

poems:	239
words:	3414
phrases:	987
argument phrases:	359
argument phrases with no case particle:	156
adjunct phrases:	41
clauses:	1055
argument clauses:	29
etc.	

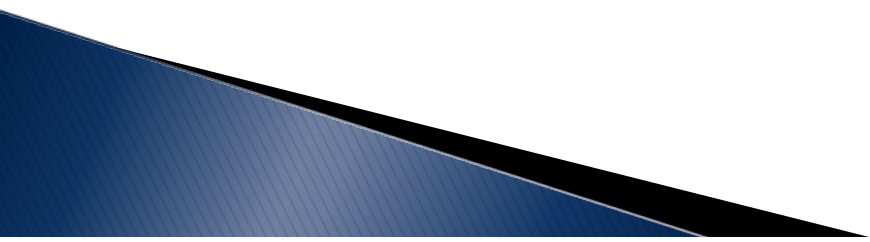
Eastern Old Japanese

- Hitachi (CEOJ) statistics

poems:	15
words:	234
phrases:	65
argument phrases:	27
argument phrases with no case particle:	11
adjunct phrases:	3
clauses:	77
argument clauses:	2
etc.	

Next...

- Analysis
- <http://vsarpj.orinst.ox.ac.uk/>



Questions and Comments Welcome

Kerri L Russell
kerri.russell@orinst.ox.ac.uk

