

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/

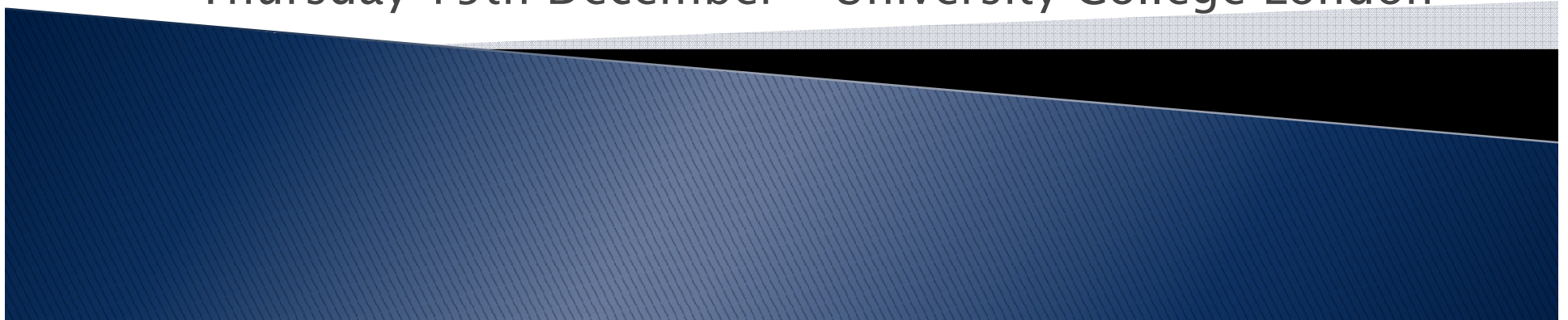
NIHU Scholarship at NINJAL

Stephen Wright Horn

Stephen.horn@orinst.ox.ac.jp

AHRC Mobility Showcase Events

Thursday 15th December – University College London



NIHU Scholarship scheme

Status: Postdoctoral Researcher at the
Research Centre for Japanese Language
and Linguistics, Oriental Institute,
University of Oxford

Placement: National Institute for Japanese
Language and Linguistics (NINJAL)

Term: 01/11/2010—31/01/2011

Conditions: Submit a report (approximately
2000 words) within 3 months of the end
of the scholarship



Application Process

Connection at Oxford: Fiona Groenhout
(Humanities Division Research Facilitator)

Previous connections at NINJAL: Kondo,
Yasuhiro (Corpus of Pre-modern Japanese)
Ogiso, Toshinobu (Statistics and
machine-learning)

Research Proposal: "Tagging noun phrases for
grammatical role in a pre-modern
Japanese language corpus."

Scholarship: Split between Horn and Russell



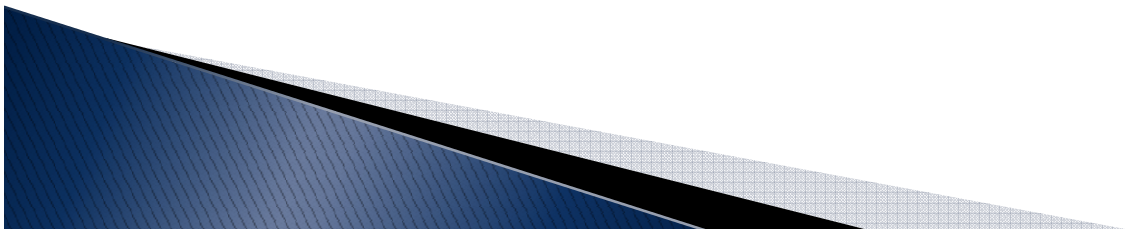
Award

AHRC Programmes Coordinator: Samuel
Lamshead

Contacts at NINJAL: Supervisor: Tanaka Makiro;
Administrative liason: Takayama Kazuo

Accomodation: Akaike Guest House

Disbursement: Pre-approved award account
through the Faculty of Oriental Studies



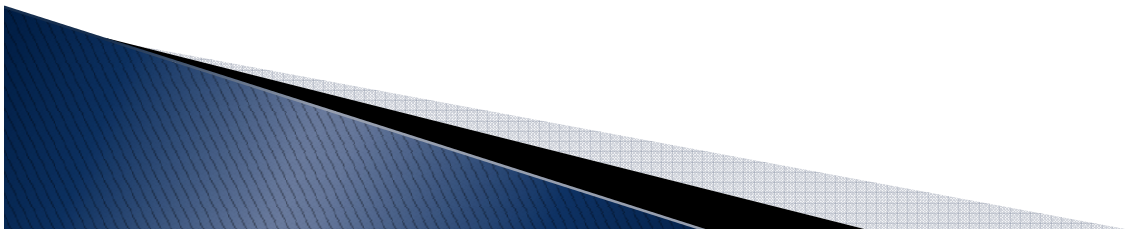
Akaike Guest House

(統計数理研究所 The Institute of Statistical Mathematics)



NINJAL facilities

- ▶ Visiting scholar's research office (work station, full access to network)
- ▶ Corpus development room (work station, full access to network, e-mail address)
- ▶ NINJAL databases, research library, National Institute of Japanese Literature
- ▶ Software for machine parsing and hand-annotation
- ▶ Seminars, research reports, colloquia



Dynagon annotator

Dynagon: プロジェクト - core_data (Access 2002 - 2003 ファイル形式) - Microsoft Access

検索 全文検索 | ファイル名検索 | 高度な検索 | モード切り替え | エクスポート | 一時的機能

分音結合モード: 全てOn | 全てOff | 同一属性全てOn

Hit件数: 55件

検索対象コーパス: CSJ

検索文字列: 日本語

並べ替え: 1: | 2: | 3: | 4: | 7/A

検	ファイル名	オーダー	新文脈	キー	後文脈	語彙素読み	語彙素	品詞	解析活用
<input checked="" type="checkbox"/>	0C02_01066.c	650	れるところかカーソルが...	日本	語が入らず、英数字しか...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	0C12_01186.c	750	が、英文はなかなか見つ...	日本	語と同じで、必ずしもこ...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	0C14_03063.c	1020	しも、やはり「冷やかしの...	日本	語が使えない人がここを...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW2X_00000.h	46010	在留地域、学んだ言語、現...	日本	語能力において極めて高...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	12150	・機軸の審議に際し、学...	日本	語教育の復興(1)日本語...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	12230	・学術に参画する予定です...	日本	語学習者の現状 近年、...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	12530	外国人の増加や他国との...	日本	語学習者の増加しており...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	13160	・一月文化庁調べによっ...	日本	語教育の支援 我が国に...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	13960	っています。文化庁では...	日本	語の学習を推進するため...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	14040	・こうした外国人に対する...	日本	語の学習を推進するため...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	14710	・こうした「学校の余剰教室...	日本	語の学習を推進するため...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	14910	・「留学」事業を全国十八...	日本	語を第二言語として学ぶ...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00000.c	15150	・て学ぶ人の中でも特に...	日本	語や生活習慣等を学ぶこ...	ニッポン	日本	名詞-固有名称-地名-国	
<input checked="" type="checkbox"/>	OW6X_00045.c	48180	・する外国人受入れ者の...	日本	語の理解が若しくは表現...	ニッポン	日本	名詞-固有名称-地名-国	

レコード: 1 / 55

ファイル名	オーダー	出現書字形	語彙素読み	語彙素	品詞	解析活用型	活用形	語義	出現発音形	ruby	id	originalTr
0C02_01066.c	610	なぜ	ナゼ	何故	助詞-副助詞				ナゼ		765288443086670	なぜ
0C02_01066.c	620	か	カ	か	助詞-副助詞				カ		153052680949298	か
0C02_01066.c	630	そこ	ソコ	某処	代名詞				ソコ		584666170432788	そこ
0C02_01066.c	640	に	ニ	に	助詞-格助詞				ニ		774551828549653	に
0C02_01066.c	650	日本	ニッポン	日本	名詞-固有名称				ニッポン		782165849927476	日本
0C02_01066.c	660	語	ゴ	語	名詞-普通名詞一般				ゴ		36652003482009	語
0C02_01066.c	670	が	ガ	が	助詞-格助詞				ガ		216952043151001	が
0C02_01066.c	680	入ら	ハイル	入る	動詞一般	五段-ラ行一般	未然形一般		ハイル		810615813296195	入ら
0C02_01066.c	690	ず	ズ	ず	助動詞	連用形一般			ズ		536405077687564	ず

レコード: 1 / 20

ファイル名	オーダー	出現書字形	語彙素読み	語彙素	品詞	解析活用型	活用形	語義	出現発音形	ruby	id	ori
0C02_01066.c	610	なぜ	ナゼ	何故	助詞-副助詞				ナゼ		765288443086670	なぜ
0C02_01066.c	620	か	カ	か	助詞-副助詞				カ		153052680949298	か
0C02_01066.c	630	そこ	ソコ	某処	代名詞				ソコ		584666170432788	そこ
0C02_01066.c	640	に	ニ	に	助詞-格助詞				ニ		774551828549653	に
0C02_01066.c	650	日本	ニッポン	日本	名詞-固有名称-地名-国				ニッポン		782165849927476	日本
0C02_01066.c	660	語	ゴ	語	名詞-普通名詞一般				ゴ		36652003482009	語
0C02_01066.c	670	が	ガ	が	助詞-格助詞				ガ		216952043151001	が
0C02_01066.c	680	入ら	ハイル	入る	動詞一般	五段-ラ行一般	未然形一般		ハイル		810615813296195	入ら
0C02_01066.c	690	ず	ズ	ず	助動詞	連用形一般			ズ		536405077687564	ず

レコード: 1 / 20

出現書字形	結合	語彙素読み	語彙素	品詞	解析活用型	活用形	語義	出現発音形	CD	lid
日本	Lex	未	ニッポン	日本	名詞-固有名称-地名-国			ニッポン	1	782
語	Lex	未	ゴ	語	名詞-普通名詞一般			ゴ	2	366
が	Lex	未	ガ	助詞-格助詞				ガ	3	216

レコード: 1 / 3

実行

検索対象コーパス: [OC_core][OW_core][OW_hand][PB_core][PN_core][PN_hand]

UniDic electronic lexicon

The screenshot displays the UniDic Explorer application window. The interface is divided into several sections:

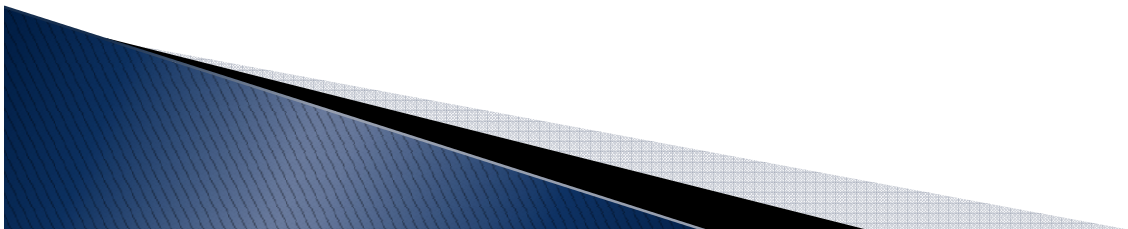
- Search Area:** Includes a search input field with "ミル" entered, search options (e.g., "完全一致", "前方一致"), and a "検索" button.
- UniDic ツリー (Tree View):** A hierarchical tree on the left showing morphological data. The selected node is "36920見る (用)", which is expanded to show its constituent morphemes like "1181441ミル (動詞-非自立可能)" and "1181442ミレル (動詞-非自立可能)".
- 語彙素 (Morpheme) Details:** A central panel showing detailed information for the selected morpheme. It includes fields for:
 - 語彙素ID: 36920 (用例: c2702r1014w641b720n)
 - 語彙素読み: ミル
 - 語彙素: 見る
 - 分類語彙表番号: 出典 [Macp]
 - 語形ID: 1181442 (用例: c15r2y28-D14-N2-B1)
 - 語形: ミレル (活用型: 下一段-ラ行)
 - 書字形ID: 302449156 (用例: y2B4Y126(138))
 - 書字形: 観れる
 - 発音形ID: 302449153 (アクセント型: 2)
 - 発音形: ミレル (アクセント結合型: C1)
- ツリーの操作 (子孫ノードも同時に処理します):** A control panel at the bottom right for managing the tree nodes, with buttons for "クリア", "コピー", "移動", and "削除".
- データベース:** A SQL query window at the bottom left showing the query: `SELECT * FROM 短単位語彙素 WHERE 語彙素読み like 'N'ミル' ORDER BY 短単位語彙素, 語彙素ID`.

“Assigning Semantic Roles in a Corpus of Old Japanese”

Presentation: “上代日本語のコーパスにおける意味役割の付与”

Venue: National Institute for Japanese Language and Linguistics (国立国語研究所),
Tachikawa City, Tokyo, Japan, 20 January 2011

<http://vsarpj.orinst.ox.ac.uk/files/horn.ninjal.pdf>



“Queries for a Corpus of the Man'yōshū”

Venue: Department of Education: Special Lecture, Hiroshima University, 25 January 2011

<http://vsarpj.orinst.ox.ac.uk/files/horn.hiroshima.pdf>

【特別講義】
万葉集のコーパスから
見たいもの

万葉のことばの世界からどんな文法が見えてくるのでしょうか。

スティーブ・ホーン氏
オックスフォード大学東洋学部
日本語研究センター

日時: 2011年1月25日(火)
12:50~14:20
場所: 教育学研究科 第二会議室(2階)
(使用言語は日本語です!)

小学館 「新日本古典文学全集」

Electronic files: *Kojiki, Nihon shoki, Fudoki, Man'yôshû, Taketori monogatari, Ise monogatari, Tosa niki, Makura no sôshi, Genji monogatari, Konjaku monogatari, etc.*

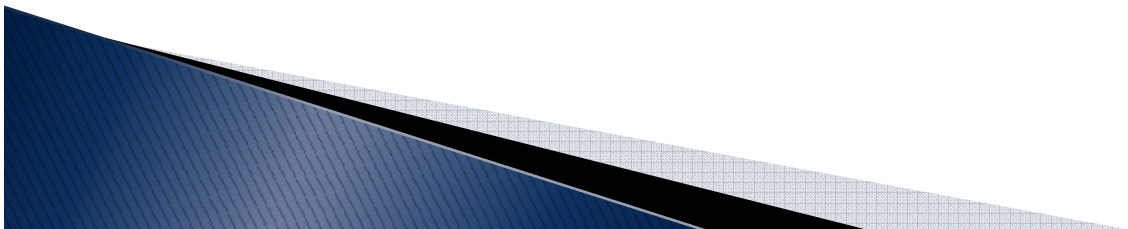
Morphological analysis by NINJAL: *Ise monogatari, Tosa niki*

Syntactic parsing by VSARPJ: *Ise monogatari, Tosa niki*



Follow-up

- ▶ Sharing resources
- ▶ Presenting and publishing
- ▶ Future collaboration



Ise Monogatari (fully parsed)

The screenshot displays the XML Editor interface for the file IM.xml. The XML content is as follows:

```
1 <?xml-stylesheet type="text/css" href=" ../sinzensyuu.css"?>
2 <sample sampleID="1202_伊勢物語" no="1202" title="伊勢物語" fileName="1202伊勢物語_100728">
3   <div id="00000001">
4     <pb no="118"/>
5     <s>
6       <pc orthToken="" IForm="" lemma="" lemmaID="23" kana="" pos="空白" Form="" pronToken="" wType="記号" start="10" end="20" orderID="10" BOS="True"/>
7       <c/>
8       <phr>
9         <w orthToken="むかし" IForm="ムカシ" lemma="昔" lemmaID="37012" kana="ムカシ" pos="名詞-普通名詞-副詞可能" Form="ムカシ" pronToken="ムカシ" wType="和" start="20" end="50"
10          orderID="20">むかし</w>
11       </phr>
12       <pc orthToken="、" IForm="" lemma="、" lemmaID="24" kana="" pos="補助記号-読点" Form="" pronToken="" wType="記号" start="50" end="60" orderID="30"></pc>
13       <a id="0"/>
14       <phr>
15         <w orthToken="男" IForm="オトコ" lemma="男" lemmaID="5063" kana="オトコ" pos="名詞-普通名詞-一般" Form="オトコ" pronToken="オトコ" wType="和" start="60" end="70"
16          orderID="40">男</w>
17       </phr>
18       <pc orthToken="、" IForm="" lemma="、" lemmaID="24" kana="" pos="補助記号-読点" Form="" pronToken="" wType="記号" start="70" end="80" orderID="50"></pc>
19       <a id="1"/>
20       <c/>
21       <note org="1" text="1"/>
22       <w orthToken="初冠" IForm="ウイコウブリ" lemma="初冠" lemmaID="196417" kana="ウイコウブリ" pos="名詞-普通名詞-サ変可能" Form="ウイコウブリ" pronToken="ウイコウブリ" wType="和" start="80"
23          end="100" orderID="60">初冠</w>
24       <w>
25         <w orthToken="し" IForm="スル" lemma="為る" lemmaID="19537" kana="シ" pos="動詞-非自立可能" Form="ス" cType="文語サ行変格" cForm="連用形-一般" pronToken="シ" wType="和"
26          start="100" end="110" orderID="70">し</w>
27         <m orthToken="て" IForm="テ" lemma="て" lemmaID="24874" kana="テ" pos="助詞-接続助詞" Form="テ" pronToken="テ" wType="和" start="110" end="120" orderID="80"
28          bound="suffix">て</m>
29       </w>
30     </s>
31   </div>
32 </sample>
```

The interface includes a menu bar (File, Edit, Find, Project, Options, Tools, Document, Window, Help), a toolbar with various editing tools, and a search bar at the bottom. The taskbar at the bottom shows the Windows Start button, several application icons, and system tray information including the date and time (22:50, 2011/12/06).

142nd Meeting of the Linguistic Society of Japan



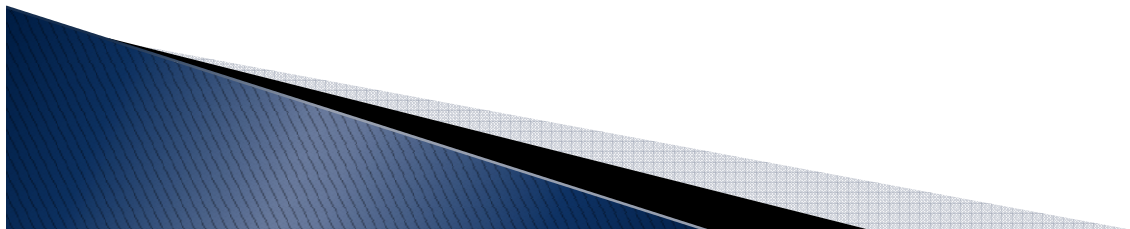
Open Symposium: Event and Property Predications in Language

Presentation: “The so-called ‘subject raising to object’ construction in Japanese.”

Venue: Nihon University, College of Humanities and Sciences, June 18–19, 2011

Publication (to appear):

“日本語のいわゆるく主語から目的語への繰り上げ構文。” Chapter in 影山太郎(編). 属性叙述の世界. くろしお出版, 2012.



20th International Conference on Historical Linguistics (ICHLXX)



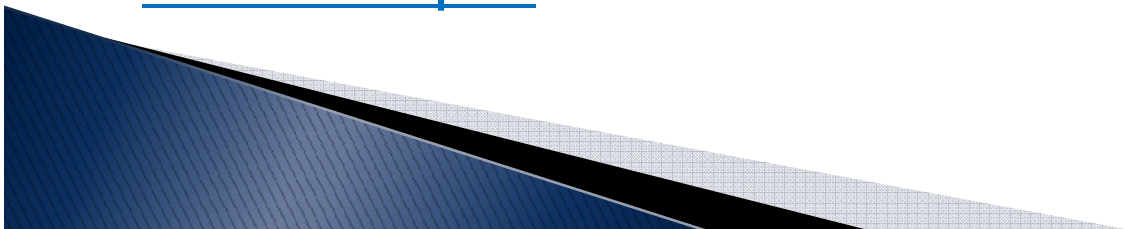
"Null Arguments in Pre-modern Japanese"

Authors: Bjarke Frellesvig, Stephen Wright
Horn, and Kerri L Russell.

Venue: National Museum of Ethnology, Osaka,
Japan, 25–30 July 2011

Publication (to appear) : "Null Arguments in
Pre-modern Japanese" *Acta Linguistica
Hafniensia* 45.1, (January 2013)

<http://vsarpj.orinst.ox.ac.uk/files/FrellesvigHornRussell.ICHL.pdf>



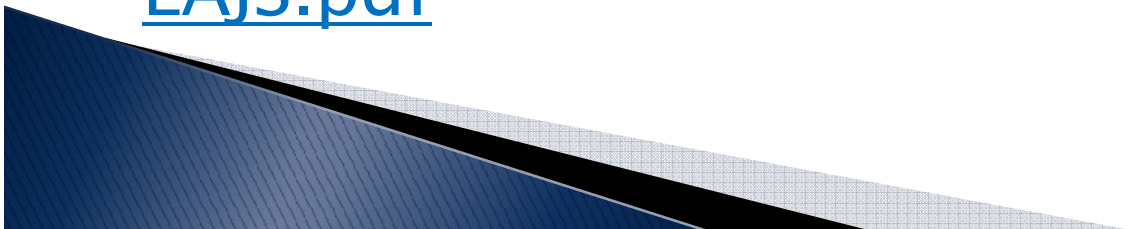
The 13th International Conference of EAJS

Presentation: Stephen Wright Horn and Kerri L Russell. "Semantic roles and case marking in Old Japanese"

Venue: Tallinn University, Tallinn, Estonia

Date: 24–27 August 2011

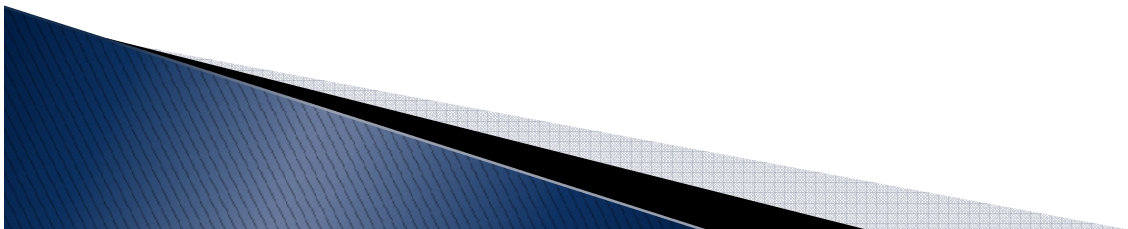
<http://vsarpj.orinst.ox.ac.uk/files/HornRussell.EAJS.pdf>



Membership on NINJAL project teams

Design of a diachronic corpus (Project leader: Kondo, Yasuhiro)

Study of the history of the Japanese language using statistics and machine-learning (Project leader: Ogiso, Toshinobu)



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/

Questions and Comments Welcome

Stephen Wright Horn
stephen.horn@orinst.ox.ac.jp

