

The Oxford Corpus of Old Japanese

- ▶ Annotated digital corpus of all extant texts in Old Japanese (7th and 8th century CE)
- ▶ Remaining texts from the Old Japanese Period
 - ▶ Around 111, 000 words
- ▶ Contains:
 - ▶ Online digital texts: original script and phonemic transcription
 - ▶ Annotations
 - ▶ Translations (into English)
 - ▶ Dictionary (Lexicon)
- ▶ The Lexicon: <http://rcjll.orient.ox.ac.uk/lexicon.html>
- ▶ Text-string search: <http://vsarpj.orinst.ox.ac.uk/corpus/ojcorpus.html>
<http://vsarpj.orinst.ox.ac.uk/corpus/NScorpus.html>
- ▶ In collaboration with National Institute for Japanese Language and Linguistics (NINJAL) and the University of York
- ▶ Supported by the British Academy and the Arts and Humanities Research Council, UK.

Creation

- ▶ Began as a research project: “Verb semantics and argument realization in pre-modern Japanese” (VSARPJ) project
- ▶ Designed to support the project
 - ▶ 4983 Poetic texts
 - ▶ 480 Poems from Eastern Old Japanese
 - ▶ 27 Norito (祝詞) and 62 Senmyō (宣命)

Annotation

- ▶ Lexeme identification (POS, inflection, function)
- ▶ Syntactic markup
- ▶ Wrapped in XML tags
 - ▶ Romanization
君待跡 kimi matu to
 - ▶ Orthography
<c type="logo">kimi</c>
<c type="logo">matu</c>
<c type="phon">to</c>
 - ▶ Morphology: POS, lexeme, and morpheme identification
<w lemma="L004266">kimi</w>
<w type="verb" inflection="adnconc"
lemma="L031644a">matu </w>
<w type="particle" subtype="conj"
lemma="L000531a">to</w>
 - ▶ Syntax: phrases, clauses, sentences
 - ▶ Bibliographic Information
 - ▶ People and place names

Usage

- ▶ Used to examine reference-coding properties provisional and conditional clauses
- ▶ A provisional clause (*e-ba*) presents one event as being connected to another, roughly meaning ‘when S1, S2’ or ‘because S1, S2’

▶ Eg: 友達 が 行けば、わたし は イベント へ 行きます。

tomodachi ga ikeba watashi wa ibento he ikimasu

friend NOM go-PRO I TOP event DIR go

“If my friend is going I will go to the event”

- ▶ Question: In the provisional clause, can the overt subject of the S2 clause appear in front of the S1 clause (and thus S1 is embedded inside of S2)?
- ▶ Answer: YES
- ▶ “...based on around 900 examples. A Provisional S1 is embedded within an S2, though it is quite ‘high’ within S2.” (Russell et. al, 2013b, p1)

[S2 [S1 *masurawo-no ei ywobi-tate-sikaba*] *sa-wosika-noi muna-wake-yuka-mu*]

fine.man-GEN call-stand-PAST.PROV PFX-buck-GENi breast-divide-go-CONJECTURAL

“The buck must be pressing through (the brush) because the fine men flushed it out.” (MYS 20.4320)

References

- ▶ Frellesvig, B., Horn S.W., Russell K.L., and Sells P. (2013b.) Provisional and Conditional Clauses in Old Japanese. In Ozge, Umut (eds.), *MIT Working Papers in Linguistics*, 67. Paper presented at the Proceedings of the 8th Workshop on Altaic Formal Linguistics, (pp.65-78).
- ▶ Russell K.L. (2013). *The Oxford Corpus of Old Japanese* (Powerpoint slides). Retrieved 26 January, 2015: <http://vsarpj.orinst.ox.ac.uk/files/Russell.HiCor.May13.pdf>
- ▶ Russell K.L. and You Z. (2014). *The Oxford Corpus of Old Japanese Lexicon* (Powerpoint slides). Retrieved 26 January, 2015: <http://vsarpj.orinst.ox.ac.uk/files/RussellYou.DHOxSS.pdf>

ISLRN

Title	The Oxford Corpus of Old Japanese (OCOJ)
Resource Type	Extant Texts from Old Japanese Period
Source/URL	http://vsarpj.orinst.ox.ac.uk/corpus/index.html
Format/MIME Type	text/xml
Size/Duration	111,000 words
Access Medium	online
Description	<p>The Oxford Corpus of Old Japanese (abbreviated OCOJ) is a long-term research project which aims to develop a comprehensive annotated digital corpus of all extant texts in Japanese from the Old Japanese period. Old Japanese is the earliest attested stage of the Japanese language, largely the Japanese language of the Asuka and Nara periods of Japanese history (7th and 8th century AD). This is the formative literate period upon which the development of Japanese civilization is based, and these texts are of paramount importance for the study and understanding of the origins and development of civilization of Japan, including language, writing, literature, religion, history, and culture.</p>
Version	?
Media Type	Transcribed text
Language	Old Japanese used within the period of 7th and 8th Century AD
Resource Creators	Faculty of Oriental Studies, Oxford; National Institute for Japanese Language and Linguistics (NINJAL) and the University of York
Distributor	University of Oxford
Rights Holder	University of Oxford

