

Language Atlas of Japanese and Ryukyuan (LAJaR): A Linguistic Typology Database for Endangered Japonic Languages

Kanji KATO¹, So MIYAGAWA², and Natsuko NAKAGAWA³

EACL 2024 Workshop SIGTYP

¹ROIS-DS Center for Open Data in Humanities

²National Institute for Japanese Language and Linguistics

³Kyushu University

Overview

Aim

Building a WALS-like linguistic typology database, LAJaR

Target Languages

The Japonic language family

Case Study and Results

Elaborated description on the distribution of zero-copula

Overview

Aim

Building a WALS-like linguistic typology database, LAJaR

Target Languages

The Japonic language family

Case Study and Results

Elaborated description on the distribution of zero-copula

Overview

Aim

Building a WALS-like linguistic typology database, LAJaR

Target Languages

The Japonic language family

Case Study and Results

Elaborated description on the distribution of zero-copula

Overview

Aim

Building a WALS-like linguistic typology database, LAJaR

Target Languages

The Japonic language family

Case Study and Results

Elaborated description on the distribution of zero-copula

Outline

Introduction

- Overview
- Background
- Target Languages
- Previous Studies

Method

- Collection
- Publication
- Motivation

Case Study

Conclusion

- Conclusion
- Future Study

Introduction

- Overview
- Background
- Target Languages
- Previous Studies

Method

- Collection
- Publication
- Motivation

Case Study

Conclusion

- Conclusion
- Future Study

What is the Japonic Family?

The Japonic Language Family

- ▶ 2 groups
 - Japanese (and its areal varieties)
 - Ryukyuan (Amami, Okinawa, Miyako, Yaeyama, Yonaguni)
- ▶ Spoken in Japan
- ▶ Lack of mutual intelligibility
- ▶ Many varieties are endangered
- ▶ There is a large volume of descriptive studies.

Previous Studies: WALS Online, Grambank

- ▶ Large database of typological features of various languages of the world
- ▶ They enable scholars to observe the distributions of typological features on a map.
- ▶ Datasets are published under CC-BY 4.0 license.

Previous Studies: WALS Online, Grambank

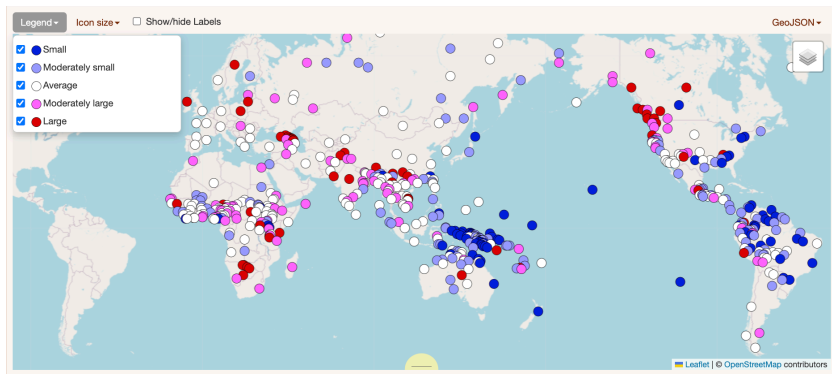


Figure 1: WALS 1A: 'Consonant Inventories' (Maddieson 2013)

Japanese languages in previous studies

Languages

Search a languoid by name. Matching names are formatted in bold font for **languages**, in italics for *genera* and underlined for families. This search does also take alternative names into account.

search for a languoid

Showing 1 to 4 of 4 entries (filtered from 2,662 total entries)

Name	WALS code	ISO 639-3	Genus	Family	Macroarea	Latitude	Longitude	Countries
<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>	--any--	<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Japan"/>
Ainu	ain	ain	Ainu	Ainu	Eurasia	43.00	143.00	Japan
Japanese	jpn	jpn	Japanese	Japanese	Eurasia	37.00	140.00	Japan
Nihon Shuwa (Japanese Sign Language)	nih	jsl	<u>Sign Languages</u>	<i>other</i>	Eurasia	36.00	139.00	Japan
Shuri	ryu	ryu	Japanese	Japanese	Eurasia	26.50	127.50	Japan

Showing 1 to 4 of 4 entries (filtered from 2,662 total entries)

Figure 2: Languages of Japan in WALS

Japanese Languages in Previous Studies

Japanese languages in previous studies

- ▶ WALS and Grambank do not cite any literature written in Japanese as sources of their datasets.
- ▶ This has led to a lack of languages and data incompleteness.

Our Goal

- ▶ Development of a typological dataset with a special focus on Japonic languages.

Introduction

- Overview
- Background
- Target Languages
- Previous Studies

Method

- Collection
- Publication
- Motivation

Case Study

Conclusion

- Conclusion
- Future Study

Method: Data Collection and Analysis

Data collection so far

- ▶ 200 features
- ▶ Some new features were added to WALS framework (e.g., Does the language have dual form for pronouns?).
- ▶ 30 varieties so far (7 varieties have been fully analyzed)
- ▶ Collected through fieldwork and literature research based on the WALS framework.

Method: How to Publish

Data will be

- ▶ Visualized with CLLD format.
- ▶ Distributed under CC-BY 4.0 license.

Why New Database?

- ▶ To include features that are not adopted in the previous studies, but are studied well in the context of the Japonic family.
- ▶ To make it accessible to the local communities.

Introduction

- Overview
- Background
- Target Languages
- Previous Studies

Method

- Collection
- Publication
- Motivation

Case Study

Conclusion

- Conclusion
- Future Study

Case Study: Zero Copula

(1) a. I am a teacher.

b. *I a teacher.

(2) *ya uchitel*

I teacher

'I am a teacher.' (Russian)

- ▶ Languages vary as to whether they allow zero copula.

Case Study: Zero Copula

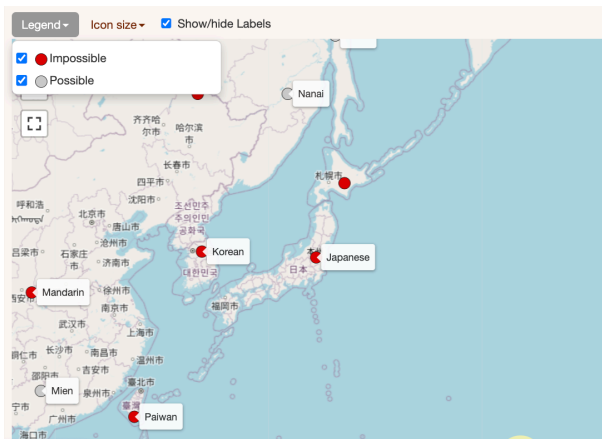


Figure 3: WALS 120A 'Zero Copula for Predicate Nominals'

Case Study: Zero Copula

- ▶ According to WALS 120A, there is no language that allows zero copula in Japan.

Case Study: Zero Copula

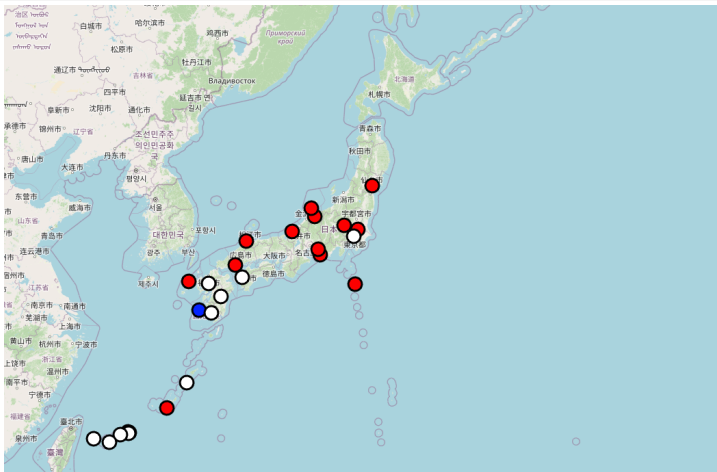


Figure 4: LAJaR 120A

Case Study: Zero Copula

- (3) a. *wan=ja sinsii*
 I=TOP teacher
 ‘I am a teacher.’ (Amami, Kato 2022: p.38)
- b. *an hita sensee=zjad=do*
 that person.TOP teacher=COP=SFP
 ‘That person is a teacher.’
 (Kagoshima-Japanese, Hiratsuka 2018: p.115)

Result

- ▶ There is a areal variation within the Japonic family regarding the feature 120A, that has not been reflected on the previous datasets.

Introduction

- Overview
- Background
- Target Languages
- Previous Studies

Method

- Collection
- Publication
- Motivation

Case Study

Conclusion

- Conclusion
- Future Study

Conclusion

- ▶ We have been developed a typological dataset on the Japonic languages with more detail.

Future Study

- ▶ Data visualization
- ▶ Data expansion
- ▶ We have just "zoomed in".

References

- Yusuke Hiratsuka. 2018. Kagoshimaken Kagoshimashi hôgen. In *Zenkoku Hôgen Bumpô Jiten Shiryôshû 4: Katsuyô Taikê 3*, pages 107–116. Hôgen Bumpô Kenkyûkai. [The dialect of Kagoshima-city, Kagoshima Prefecture].
- Kanji Kato. 2022. Tokunoshima (Kagoshima, Northern Ryukyuan). In Michinori Shimoji, editor, *An introduction to the Japonic languages : grammatical sketches of Japanese dialects and Ryukyuan languages*, chapter 2, pages 25–57. Brill, Leiden ; Boston.
- Ian Maddieson. 2013. Consonant inventories (v2020.3). In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Zenodo.

Acknowledgement

This work was supported by ROIS-DS-JOINT (#048RP2022) to S. Miyagawa.