



Feb. 27-Mar. 3.2019 ICLDC 6

Hawai'i University

# Raising language diversity awareness in Japan through web-based open access application

Kenan CELIK

Nobuko KIBE

National Institute for Japanese Language and Linguistics

大学共同利用機関法人 人間文化研究機構

国立国語研究所

National Institute for Japanese Language and Linguistics NINJAL

## Database of Endangered Languages of Japan (DELJ)

### the Database of Endangered Languages of Japan

Welcome to the Database of Endangered Languages of Japan!

This database is an open-access database that offers linguistic and audio materials of many endangered languages and dialects spoken in Japan. The basic vocabulary page ([here](#)) contains searchable basic vocabulary word lists of many dialects, each word being accompanied by its pronunciation by a native speaker. In the oral text page ([here](#)), you can enjoy many oral texts gathered in the field and provided with a full transcription and translation.



Database of Endangered Languages of Japan  
kikigengo.ninjal.ac.jp/en

# Language diversity in Japan

- How many languages spoken in Japan?
  - “Recent” languages (e.g. Korean, Chinese)
  - “Traditional” languages, spoken before 1853
- How many traditional languages spoken in Japan?

# Languages of Japan

## Japonic languages

## Ryukyuan languages

Japanese language(s)?

Northern Ryukyuan

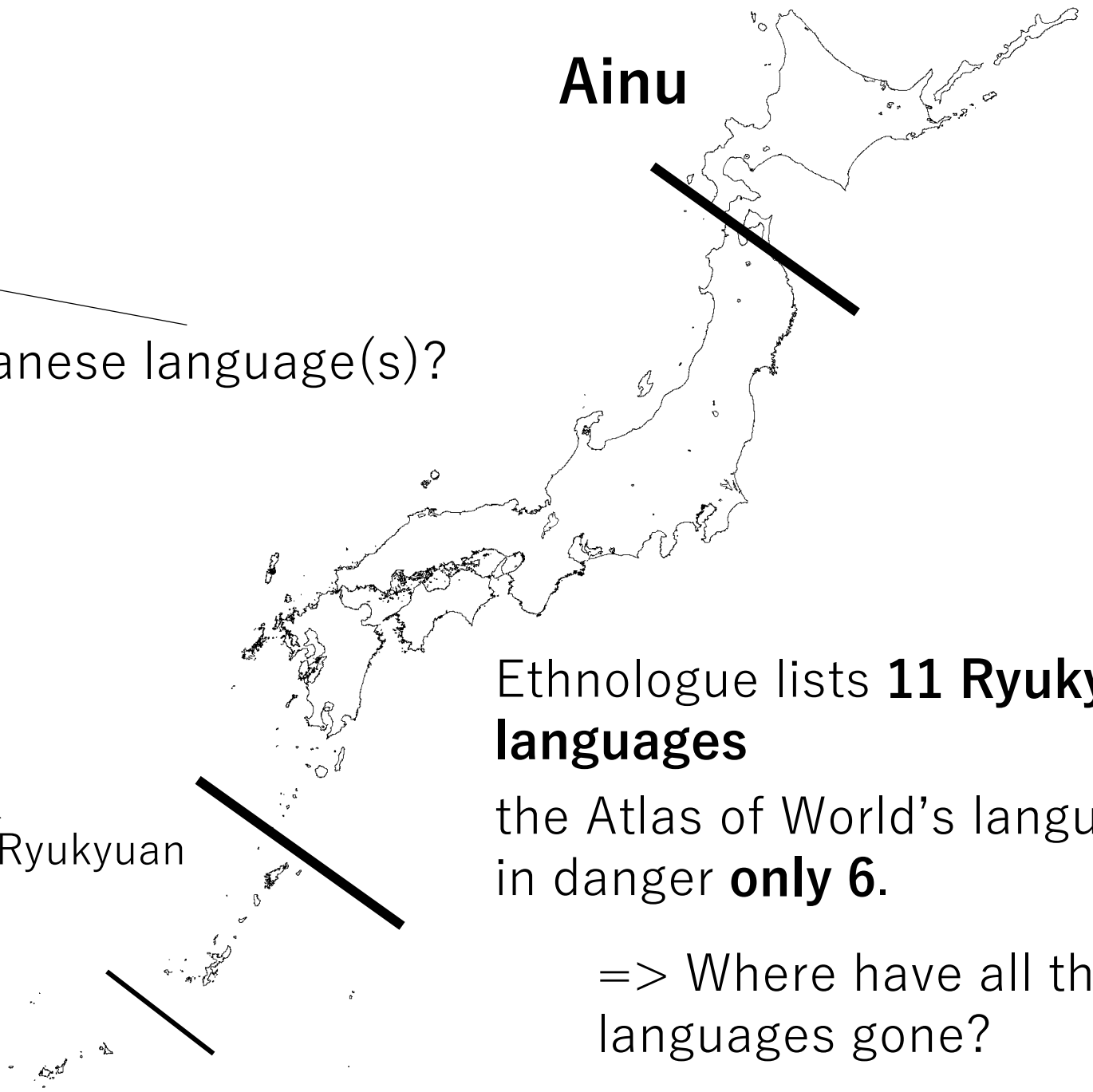
Southern Ryukyuan

**Ainu**

Ethnologue lists **11 Ryukyuan languages**

the Atlas of World's languages in danger **only 6.**

=> Where have all the languages gone?



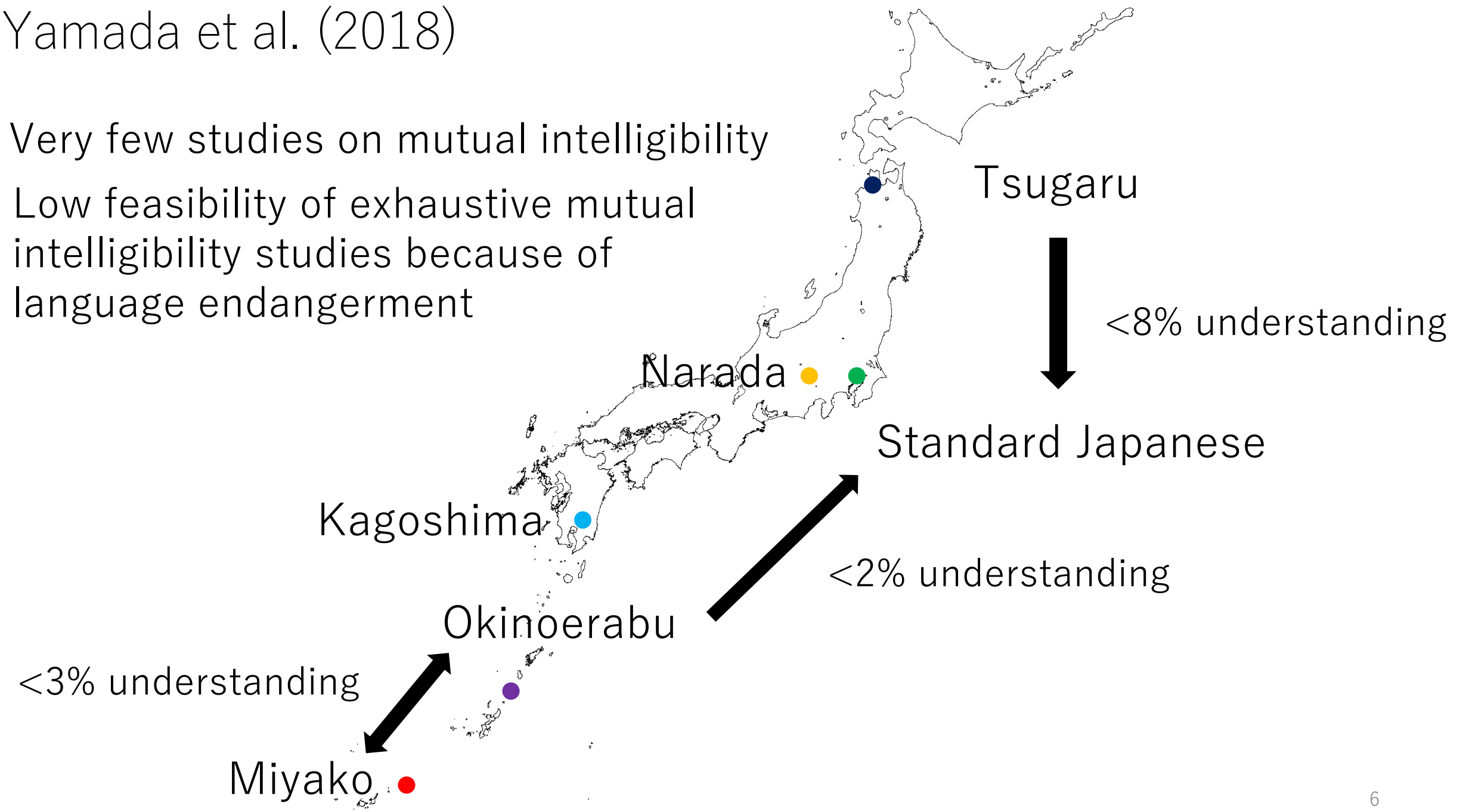
# Language diversity in Japan

- How to define language? => Criteria of mutual intelligibility (Hockett 1958)
- How many (mutually unintelligible) traditional languages spoken in Japan?
- Study testing Okinoerabu/Ikema Miyako and Tsugaru dialect (Yamada et al. 2018)
  - No mutual intelligibility between Okinoerabu and Ikema Miyako
  - No intelligibility of Tsugaru dialect by Standard Japanese speakers

Yamada et al. (2018)

Very few studies on mutual intelligibility

Low feasibility of exhaustive mutual intelligibility studies because of language endangerment



# Language diversity in Japan

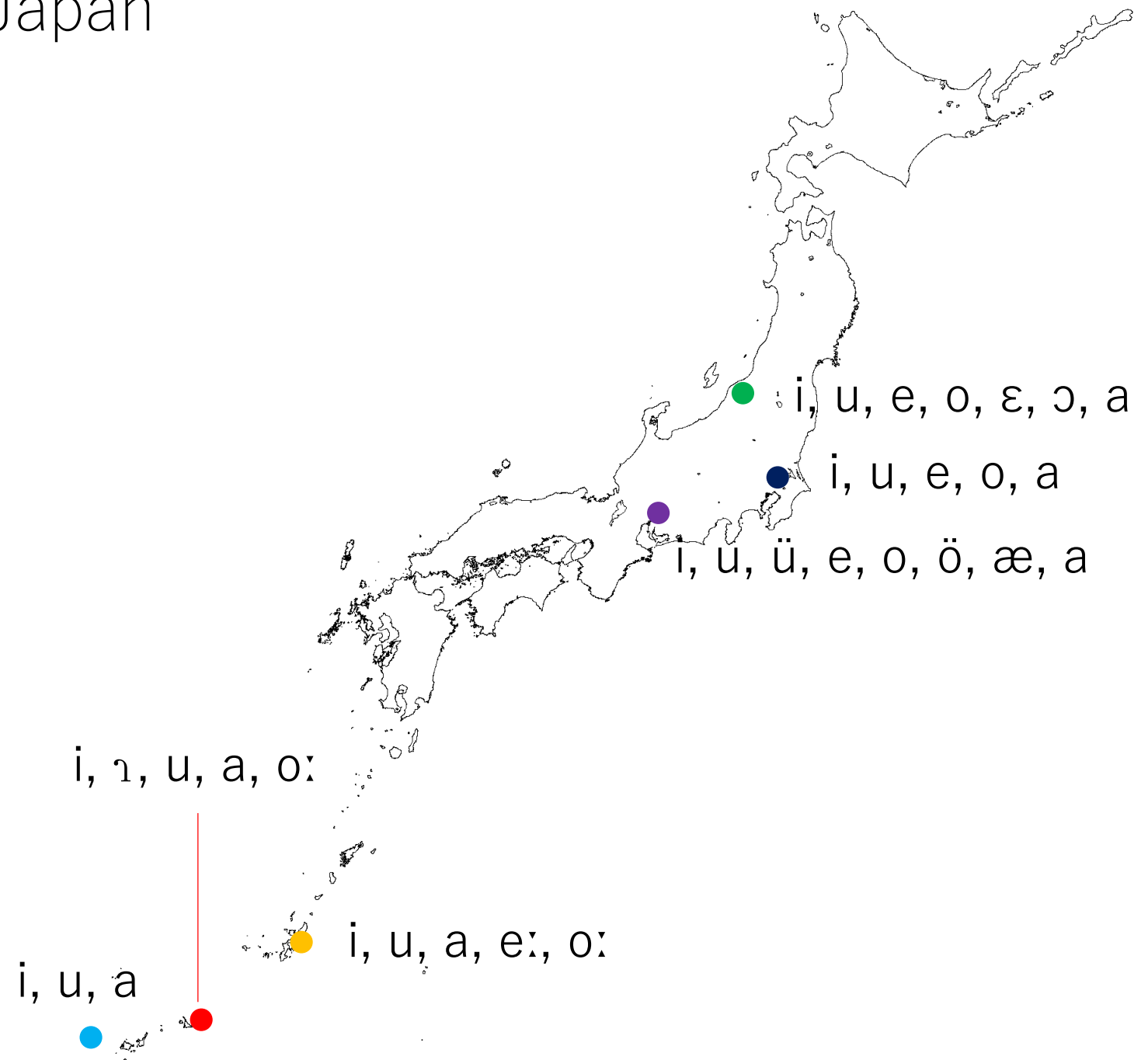
- How many (mutually unintelligible) traditional languages spoken in Japan
  - => Well... we don't really know (and that is the result)
- We **do know** that we do not know how many mutually unintelligible traditional languages there are in Japan

# Language diversity in Japan

- How many traditional language *varieties* spoken in Japan?
  - That we (kind of) know: *at least* **several hundreds**
- Not really surprising, historical depth comparable to Romance languages (>1500 years Pellard 2015) in an island environment



# Linguistic diversity in Japan



# Language diversity in Japan

- This wealth of language diversity in Japan is not properly acknowledged, especially among non-linguists
- Negative attitudes towards (one's own) “dialect” and linguistic diversity (Not being able to speak one's own traditional dialect considered as social status)

How to raise awareness of linguistic diversity in Japan?  
How to make it known? How to appeal to it?

# Paradigm shift in Japanese linguistics

- Since 2000s, radical paradigm shift in Jap. linguistics: proper acknowledgement of linguistic diversity
- Numerous publications of very good reference grammar describing particular language varieties

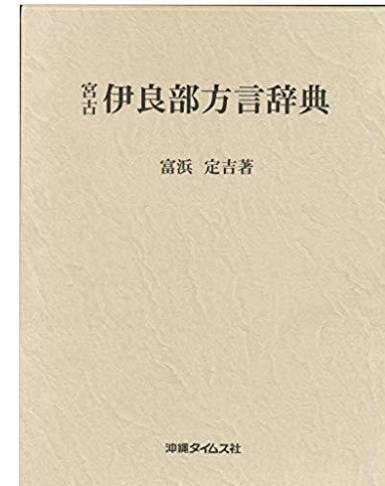
Source	Language branch	Language variety	Written in
Miyara 1995	Ryukyuan	Ishigaki	Japanese
Shimoji 2008	Ryukyuan	Miyako Irabu	English
Pellard 2009	Ryukyuan	Miyako Ogami	French
Niinaga 2014	Ryukyuan	Amami Yuwan	English
Kubozono et al. 2015	Japanese	Koshikijima	Japanese
Yokoyama 2017	Ryukyuan	Okinoerabu	Japanese

# Some issues

- These research raise several issues:
  - Highly atomistic, each descriptive research focusing on one language variety spoken in one village
  - Highly technical, not accessible to non-linguists.
  - No standardization of data (e.g. different traditions)
  - Little internationalization (= access of data to non-Japanese speakers)
  - Do not always (=always never) publish the linguistic primary data (audio files/texts etc.), Great description, poor documentation
- They are not sufficient to get the big picture of language diversity in Japan and convey it to non-linguists.

# Some issues

- In the case of linguistics materials written by non-linguists, i.d. speakers from local communities: no quality control, resulting in anything from masterpieces (Tomihama 2013) to dubious
- What to do?



# Database of Endangered Languages of Japan

- Platform to address these issues, respecting:
  - fundamentally comparative
  - standardized data
  - quality control
  - easy to update (smooth flow from field to data publication)
  - free/open-access
  - appealing/user-friendly
  - bilingual (Japanese/English)
- => the Database of Endangered Languages of Japan, result of 2 projects at NINJAL:
  - *General Study for Research and Conservation of Endangered Dialects in Japan* (2010 and 2015)
  - *Endangered Languages and Dialects in Japan* (2016-2021)

## Database of Endangered Languages of Japan (DELJ)

### the Database of Endangered Languages of Japan

Welcome to the Database of Endangered Languages of Japan!

This database is an open-access database that offers linguistic and audio materials of many endangered languages and dialects spoken in Japan. The basic vocabulary page ([here](#)) contains searchable basic vocabulary word lists of many dialects, each word being accompanied by its pronunciation by a native speaker. In the oral text page ([here](#)), you can enjoy many oral texts gathered in the field and provided with a full transcription and translation.



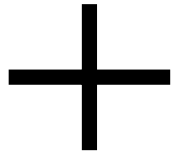
Database of Endangered Languages of Japan  
kikigengo.ninjal.ac.jp/en

# Database of Endangered Languages of Japan



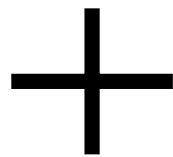
## Team of researchers

40 researchers working on 40 language varieties



## Database

Basic vocabulary component  
Oral text component



## Homepage ([kikigengo.ninjal.ac.jp](http://kikigengo.ninjal.ac.jp))

makes database searchable and realized with designer



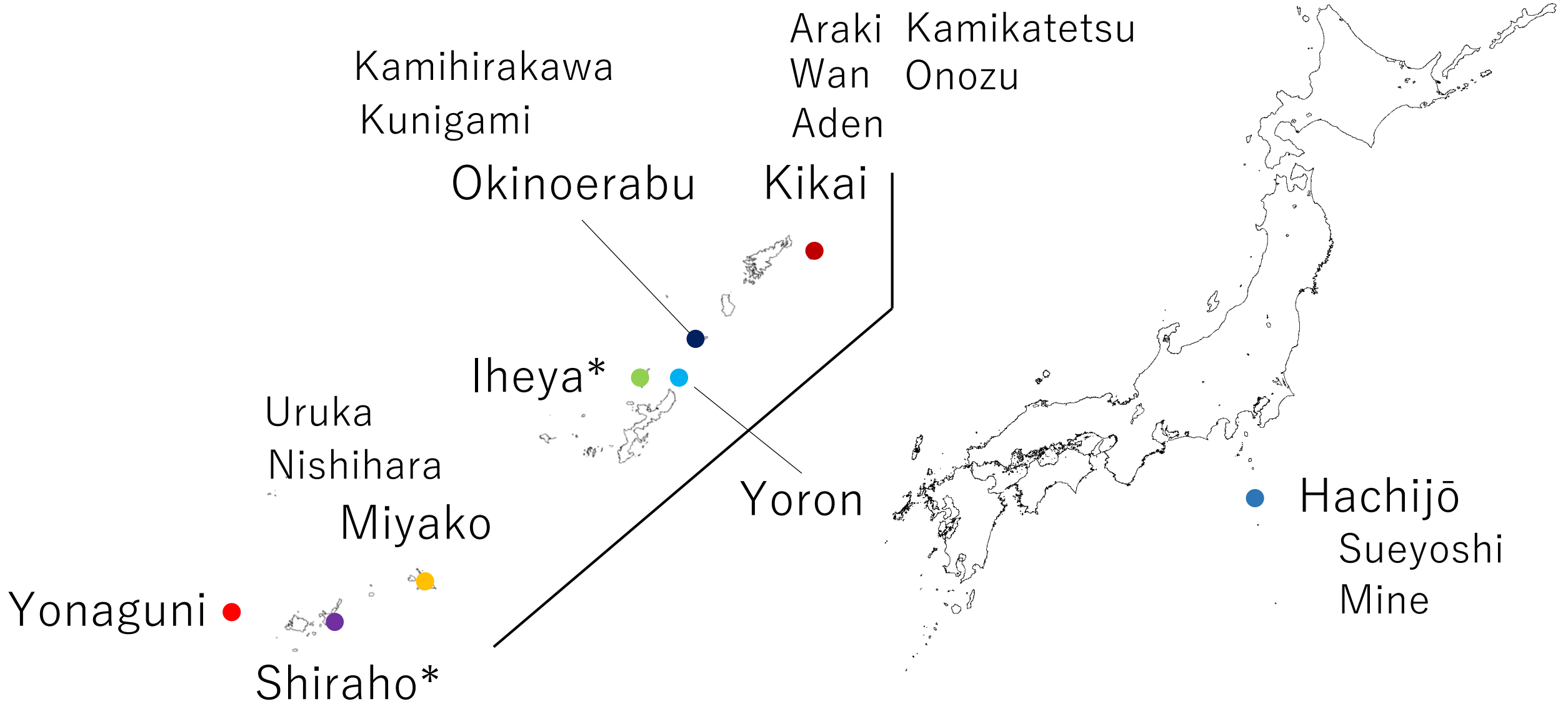
# Database of Endangered Languages of Japan (Basic Vocabulary)

- Basic vocabulary items with sound files

so far ...

- **13 language varieties** covered
- **8080 data points** (appr. 620 words per variety)

# Basic vocabulary component



\*in preparation

# Database of Endangered Languages of Japan (Basic Vocabulary)

- Fieldwork (data gathering) based on standardized questionnaire (1200 basic vocab. items)
- Data processing
  - transcription (IPA + Japanese orthography)
  - prosodic information
  - comments
  - example sentences
  - processing of sound files
- Data is centrally integrated into the DB
- Publication

# Database of Endangered Languages of Japan (Oral texts)

so far...

- 9 language varieties
- 22 oral texts
  
- Transcription (IPA + Kana)
- Gloss
- Free translation (English/Japanese)

# Database of Endangered Languages of Japan

- (semi-)Centralized organization enables:
  - Standardization of data => better comprehension of diversity
  - Quality control => data usable by linguists & enjoyable by non-linguists
  - Balanced choice of areas (not yet quite there)
- If you want to know about linguistic diversity in Japan, you have to visit [kikigengo.ninjal.ac.jp](http://kikigengo.ninjal.ac.jp)!

# Database of Endangered Languages of Japan: possibilities for the future

- Database is perpetually expanding (hope to get the attention of more and more people)
- technical perspective: tool for publishing online talking dictionaries (or lexical data)
- data perspective: resource for research, but also for other purposes like education, revitalization etc.

# Database of Endangered Languages of Japan

- If you want to know about linguistic diversity in Japan, you have to visit **[kikigengo.ninjal.ac.jp/en](http://kikigengo.ninjal.ac.jp/en)**!
- Thank you for listening!

# References

- Hockett, C. (1958). A course in modern linguistics. New York: Macmillan.
- Pellard, T. (2015). The linguistic archeology of the Ryukyu Islands. Heinrich, Patrick; Miyara, Shinsho; Shimoji, Michinori. Handbook of the Ryukyuan languages: History, structure, and use, De Gruyter Mouton: 13–37.
- Tomohama, S. (2013). Miyako Irabu hogen jiten [A dictionary of Miyako-Irabu]. Okinawa Times.
- Yamada, M., Takubo, Y., Iwasaki, S., Celik, K., Harada, S., Kibe, N., Lau, T., Nakagawa, N., Niinaga, Y., Otsuki, T., Sato, M., Shirata, R., van der Lubbe, G., Yokoyama, A. (2018). Experimental Study of Inter-language and Inter-generational Intelligibility: Methodology and Case Studies of Ryukyuan Languages. 26<sup>th</sup> Japanese-Korean Conference, UCLA.