

# A proto-Ryukyuan Database: an aggregating model of dialectal lexical data

CELIK Kenan (National Institute for Japanese Language and Linguistics)

NAKAZAWA Kohei (Shinshu University)

ASO Reiko (Meio University)

In the investigation into the history of the Japonic languages, the reconstruction of proto-Ryukyuan (pRk), i.e. the most recent common ancestor of all the languages belonging to the Ryukyuan branch of the Japonic languages, is a necessary step. Detailed reconstructions of Proto-Ryukyuan have already been proposed (Hattori 1978-79, Thorpe 1986), but they predate the publication of major lexical works on several Ryukyuan dialects, among others Maezawa (2011), Tomihama (2013), Kajiku (2020), Tokuyama, Celik (2020). More recent research on pRk like Pellard (2009, 2015) does integrate more up-to-date data, but even since the publication of this research, the availability of lexical data on Ryukyuan dialects has grown at a dizzying pace.

Needless to say, any proposed reconstruction of pRk should be tested against new data and be updated accordingly. The unprecedented growth in lexical data on Ryukyuan dialects opens up the possibility of a robust testing of the models proposed so far, but, at the same time, it raises the non-trivial problem of how to integrate this uninterrupted flow of new dialectal data into the reconstruction. A solution to this problem would be a data management system that is able to continuously integrate the latest data into the existing model of pRk.

A pioneering answer to this problem is Igarashi's "Japonic classified vocabulary" (Igarashi 2016), a periodically updated database of reconstructed pRk Forms, containing in its latest version (v.7 Igarashi 2019) more than 1800 pRk words. However, this database suffers from several flaws. First, it does not make public the dialectal data on which the reconstructions are based. Another major flaw lies in the approach adopted, which could be described as "extractive data-building". Lexical data is extracted from each source into the database without keeping a co-reference with the data in the source. This design, in which the link with the primary source is lost, not only limits research possibilities, but, more importantly, it also cannot cope with dynamic lexical data building, in which the primary source is regularly updated.

In the face of this, we constructed a proto-Ryukyuan database (approximately 6000 cognate sets) adopting a radically different approach, which could be described as "aggregate data-building". Namely, we devised an ID system enabling us to link cognates across different lexical data sets, whatever their data structure. Since the link with all primary sources is kept, not only does this design efficiently cope with dynamic lexical building, but it also vouches for more flexible data building. For example, since we link cognates to a common cognate set and not to a proto-form, there is no need to decide whether a particular cognate set goes back to the proto-language or not. This approach also

ensures transparency in the reconstruction as we can refer to all the dialectal data with a unique ID. Lastly, this design also opens up many other research possibilities than the reconstruction of pRk, like for instance the mapping of isoglosses or the exploration of the geographic factors of word variation.

#### References

- Hattori, Shirō. 1978–1979. Nihon sogo ni tsuite 1–22. *Gengo* 7(1)–7(3), 7(6)–8(12).
- Igarashi, Yōsuke. 2016. Akusento-gatano taiōni motoduite nichiryū-sogowo saiken suru tameno goi-risuto "nichiryūgo-ruibetu-goi". *Nihongo-gakkai 2016 nendo shunki taikai yokōshū*, 233-238.
- Igarashi, Yōsuke. 2019. Japonic Classified Vocabulary (v.7, 2019/5/17) electronic data (last accessed 2021/8/18).
- Kajiku, Shin'ichi. 2020. *Hatoma hōgen jiten*. Tokyo: The National Institute for Japanese Language and Linguistics.
- Maeara, Tōru. 2011. *Taketomi hōgen jiten*. Ishigaki: Nanzansha.
- Pellard, Thomas. *Ōgami: Éléments de description d'un parler du Sud des Ryūkyū*. Linguistique. Ecole des Hautes Etudes en Sciences Sociales (EHESS) PhD thesis.
- Pellard, Thomas. 2015. The linguistic archeology of the Ryukyu Islands. In *Handbook of the Ryukyuan languages: History, structure, and use*, eds., Patrick Heinrich, Shinsho Miyara, Michinori Shimoji, 13–37. Berlin: De Gruyter Mouton.
- Thorpe, Maner L. 1983. *Ryūkyūan language history*. University of Southern California PhD thesis.
- Tokuyama, Shun'ei, Celik, Kenan. 2020. *Minami Ryūkyū Miyakogo Tarama hōgen jiten*. Tokyo: The National Institute for Japanese Language and Linguistics.
- Tomihama, Sadakichi. 2013. *Miyako Irabu hōgen jiten*. Okinawa: The Okinawa Times.