

A preliminary survey on linguistic areas in East Asia based on phonological features

Ian Joo and Yu-Yin Hsu

The Hong Kong Polytechnic University

In this presentation, we will demonstrate the preliminary findings of our project aimed at identifying linguistic areas in East Asia. A linguistic area is here defined as geographically close languages sharing a high proportion of linguistic features not due to genealogical relatedness but due to historical contact. East Asia is defined as the area consisting of China, Japan, Korea, and Mongolia.

Based on 19 phonological features as binary parameters in 52 East Asian languages that we sampled, we calculated the Simple Matching Coefficient of the features between geographically close languages. The 19 phonological features, seven of them from the World Atlas of Linguistic Structures (WALS, Dryer and Haspelmath 2013), are listed in Table 1, with a column indicating whether each feature is borrowed from WALS. We drew a line between two languages if their geographical coordinates are within 1,500km distance and their 19 binary features show a Simple Matching Coefficient higher than 0.7.

Table 1: List of phonological features

Feature	WALS
Consonant Clusters	No
Consonant Inventories	Yes
Coronal Sonorants	Yes
Falling Diphthongs	No
Front Rounded vowels	Yes
Glottal Stop	No
Labiodental Fricatives	No
Long Vowels	No
Palatal Nasal	No
Plosive Coda	No
Retroflex Consonants	No
Tone	Yes
Uvular Consonants	Yes
Velar Fricatives	No
Velar Nasal Onset	No
Voiced Plosives	No
Voiceless Glottal Fricative	No
Vowel Nasalization	Yes
Vowel Quality Inventories	Yes

Figure 1 shows the preliminary results, suggesting some interesting patterns:

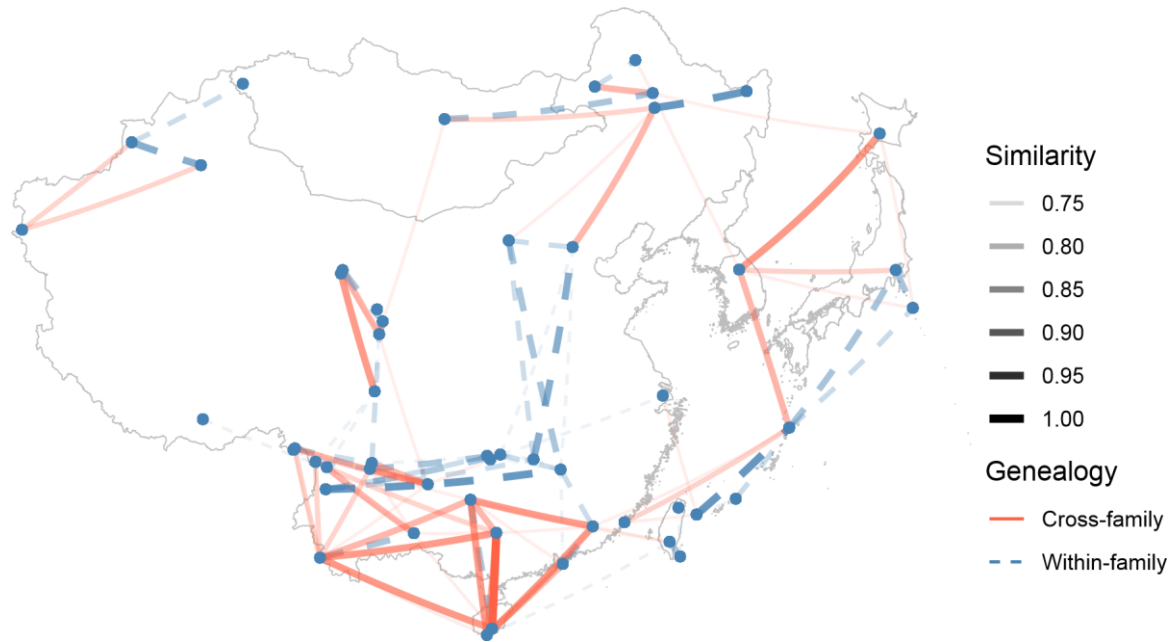


Figure 1: Connections representing phonological similarity between geographically close languages, across or within families

- The languages spoken in the Chinese provinces of Qinghai and Gansu share strong cross-family connections, as predicted by previous studies on the Qinghai-Gansu linguistic area (Xu 2017, Ch. 1, cf.).
- Languages in southwestern China are generally densely connected to each other, supporting the previous theories of the Mainland Southeast Asian linguistic area (Enfield 2018, cf.).
- Formosan languages show no similarity to Ryukyuan languages, despite their geographical proximity, in line with a genetic study demonstrating no genetic similarity between Taiwanese aboriginals and Ryukyuan islanders (Matsukusa et al. 2010).
- Manchu is connected to Mongolian and northern Sinitic languages (Mandarin and Hohhot Jin), in line with the historical contact between Manchu and these languages.
- Korean is most strongly similar to Ainu, and less so to Japonic languages and Dagur (Mongolic).
- Sarikoli, an Indo-European language spoken in northwestern China, show some connection with Turkic languages (Kazakh and Uyghur) spoken nearby.

Even though these observable patterns must be approached with caution given the preliminary stage of the data, they offer a promising outlook to our ongoing project and lead us to believe that with more features (other than phonological) examined, we will have a clearer view on linguistic areas within East Asia.

References

- Dryer, Matthew S. and Martin Haspelmath, eds. (2013). *WALS Online*. URL: <https://wals.info/>.
- Enfield, N. J. (2018). *Mainland Southeast Asian Languages: A Concise Typological Introduction*. Cambridge, UK ; New York, NY: Cambridge University Press.
- Matsukusa, Hirotaka et al. (2010). “A genetic analysis of the Sakishima islanders reveals no relationship with Taiwan aborigines but shared ancestry with Ainu and main-island Japanese”. In: *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 142.2, pp. 211–223.
- Xu, Dan (2017). *The Tangwang language: An interdisciplinary case study in Northwest China*. Cham: Springer.

Ian Joo: ian.joo@connect.polyu.hk

Yu-Yin Hsu: yu-yin.hsu@polyu.edu.hk