

---

## Using Genetic Data to Make Linguistic Arguments: opportunities and pitfalls

---

*Remo Nitschke*  
University of Arizona

In this talk I will show that using data from other fields, such as genetics may be very valuable to linguistics, but that we must take care to ensure that the data we use from other fields are suitable for the kinds of linguistic arguments we want to make. Huybregts (2017) brings forth a particular argument regarding the nature of click-phonemes and their role in human language evolution, based on genetic data on Southern Khoisan, Hadza and Sandawe language communities. I will show that the data they use is unsuitable for that specific argument. I will further provide different genetic evidence that calls theirs into question and provide some additional linguistic considerations that support a different view of the matter. Finally, I will propose an alternative story on the historical nature of click phonemes in these languages, suggesting that clicks might be a more recent linguistic innovation.

Huybregts (2017) provides a welcome example of how linguistics can utilize insights from other sciences, such as genetics. They use genetic evidence of the earliest known split of the modern human population, 121-128kya by the Khoisan group (Huybregts, 2017, 281), to date the latest possible emergence of a genetic language faculty, or I-language. Since geneticists are able to date the time frame during which the Khoisan group split from the main human population fairly consistently, we have to assume that I-language was present before said split, otherwise the modern descendants of said group would not be capable of language. They also argue that externalization of language must have occurred after the Khoisan split, as the descendants of said split are the only known language communities that have native clicks in their languages. While the former argument is sound and a highly valuable contribution, the latter is problematic for a variety of reasons.

Huybregts (2017) labels the modern descendants of the Khoisan divergence event as the communities that are encompassed by the Southern Khoisan (South Africa) language family and the isolates of the Hadza and Sandawe communities (East Africa), which constitute all of the language communities that have native click phonemes. They base this argument on evidence by Pickrell et al. (2012) who established distant shared ancestry between the Sandawe and Hadza groups of Eastern Africa and the Southern Khoisan, which are assumed to be the modern descendants of the first split in modern human populations. The Hadza and Sandawe languages were once considered to be part of the “Khoisan” languages, but they have since been re-analyzed as isolates of a larger “Khoisan” family, or perhaps not as related at all (cf. Güldemann and Sands, 2009, 205). Huybregts (2017) therefore needs to establish a genetic relationship between these groups and the Southern Khoisan language groups, to be able to make the argument that clicks developed only in that part of the human population that split 130kya. The descendant communities must therefore have retained clicks throughout their history up to this day. According to Huybregts (2017), language externalization happened after the Khoisan split from the general human population. This allowed the Khoisan group to develop click phonemes, which were then retained within the Southern Khoisan, the Hadza and the Sandawe for up to 110k years (Huybregts, 2017, 287), which explains why clicks are not natively found in any other language community. However, the genetic evidence is not as clear cut as the author presents it.

**Mismatches in the Genetic Evidence:** Pickrell et al. (2012, 4) attest that their analysis “provides strong evidence for a shared origin for the Khoisan-related genetic material in the Hadza and Sandawe”. However, Lachance et al. (2012), using whole genome sequencing of Hadza, Sandawe, and Cameroon Pygmy populations, come to a different result. They propose that ancient Sandawe and Hadza populations may have diverged from ancient Pygmy populations (Lachance et al., 2012, 458, F), which puts into question the idea that clicks developed in the ancient Khoisan divergence and have since been retained by the Hadza and Sandawe for 110k years (Huybregts, 2017, 287). Furthermore,

Shriner et al. (2018) suggest that “Hadza ancestry is closer to Omotic<sup>1</sup> ancestry than to Khoisan ancestry [...] [and] the core of the Sandawe sample is predominantly Omotic” (Shriner et al., 2018, 880). They suggest that the admixture found by Pickrell et al. (2012) is a result of a more recent migration event (cf. Shriner et al., 2018, 879).

In sum, the genetic data is not as clear cut as it might seem at a first glance, and it makes the story of clicks being a remnant of the first human diversification event unlikely.

**Linguistic Considerations:** Assuming that clicks have been part of the Khoisan branch of human ancestry for at least 100k years, is equally questionable from a linguistic point of view. We know that clicks are not uniquely acquirable by peoples with Khoisan ancestry, as many Bantu languages have extensively borrowed clicks from the Southern Khoisan families through the Bantu expansion (Huybregts, 2017). Clicks, like any other type of phoneme, are not restricted to genetic inheritance and may therefore be borrowed and transferred through language contact situations. The complete lack of clicks outside of Southern Khoisan, Bantu, Hadza and Sandawe, would therefore force us to assume, that, before the Bantu expansion none of the “Khoisan” languages encountered any language contact situations that would facilitate the transfer of clicks to another community. This does not only seem unlikely, but is also called into question through the same genetic studies cited by Huybregts (2017). For example, Pickrell et al. (2012) attest admixture in the Hadza ancestry with “a population most closely related to the Dinka”<sup>2</sup> (Pickrell et al., 2012, 4). We therefore know, that the Hadza have been in genetic contact (i.e. intermarriage) with other populations, which is a situation that facilitates strong language contact. If that is the case, then why are clicks not more widespread?

I will suggest different approach, in which clicks may be considered a more recent linguistic innovation, that possibly occurred in two unrelated places (Southern Africa and Eastern Africa) at possibly different points in the past. A similar suggestion has been entertained in the past by Güldemann and Sands (2009) and others. There is no *a priori* reason that should lead us to believe that clicks are a particularly ancient linguistic phenomenon, or that the “Khoisan” speech communities are particularly conservative. There is equally plausible evidence that suggests that clicks may be a recent innovation, and may have been innovated in two separate areas, or may have been brought to Eastern Africa by far more recent migration events.

## References

- Güldemann, T. and Sands, B. (2009). What click languages can and can’t tell us about language origins. In Botha, R. and Knight, C., editors, *The Cradle of Language*, pages 204–218. OUP, Oxford.
- Huybregts, M. A. C. (2017). Phonemic clicks and the mapping asymmetry: How language emerged and speech developed. *Neuroscience & Biobehavioral Reviews*, 81:279 – 294.
- Lachance, J., Vernot, B., Elbers, C. C., Ferwerda, B., Froment, A., Bodo, J.-M., Lema, G., Fu, W., Nyambo, T. B., Rebbeck, T. R., Zhang, K., Akey, J. M., , and Tishkoff, S. A. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse african hunter-gatherers. *Cell*, 150(3):457–469.
- Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S. W., Nakagawa, H., Naumann, C., Lipson, M., Loh, P.-R., Lachance, J., Mountain, J., Bustamante, C. D., Berger, B., Tishkoff, S. A., Henn, B. M., Stoneking, M., Reich, D., and Pakendorf, B. (2012). The genetic prehistory of southern africa. *Nature Communications*, 3(1143).
- Shriner, D., Tekola-Ayele, F., Adeyemo, A., and Rotimi, C. N. (2018). Genetic ancestry of hadza and sandawe peoples reveals ancient population structure in africa. *Genome Biol. Evol.*, 10(3):875–882.

---

<sup>1</sup>:modern speakers of Afro-Asiatic languages

<sup>2</sup>Unfortunately, they provide no numbers or further data on this ancestry. Attempts to contact the authors have yielded no results.

## **‘School’ versus ‘Home’: California-based Korean Americans’ Context-dependent Production of /u/ and /oo/ -- Andrew Cheng (UC Berkeley)**

California English is strongly represented in stereotypes by the fronting of back vowels /oo/ (as in *home*) and /u/ (as in *school*), one of the sound changes of the California Vowel Shift (CVS) [1, 2, 6]. However, participation in this shift is variable across the population of native California English speakers, according to age, sexuality, gender, ethnicity, and even a social orientation toward urban versus rural identity [3, 4, 5, 7].

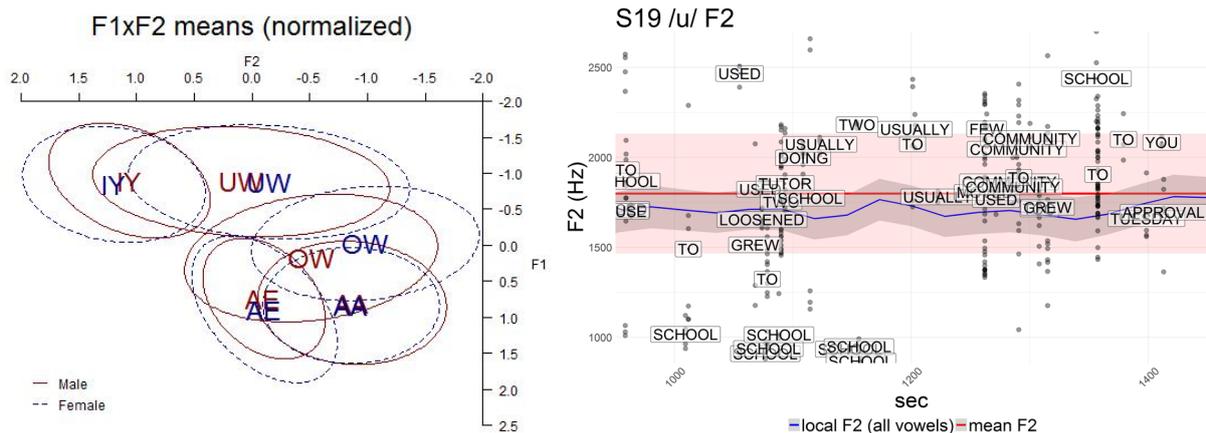
Korean Americans, an ethnic minority in California (at about 500,000, or 1.2% of the California population), mostly reside in metropolitan regions such as Los Angeles. Thirty percent are native born Americans, known as “second generation” Korean Americans (KAs). As native English speakers, dominant in English rather than Korean, the prevailing assumption is that second generation KAs will acquire the patterns of their local dialect, in this case California English. Though it has been argued that minority ethnic communities do not participate in sound changes of the majority group, in the case of the Mexican American community in California (including Chicano English speakers), /u/-fronting can be found [8].

Since KAs are known to attribute great import to maintenance of the Korean language for the purpose of cultural identity [9, 10, 11] and are almost always exposed to the Korean language prior to or simultaneous to any exposure to English, they will have acquired Korean phonology to some degree [12, 13]. For bilingual second generation KAs, then, the production of the backed Korean /u/ and /o/ may interfere with their participation in English back vowel fronting. (A similar argument is given for bilingual Mexican Americans.) However, as Fought (1999) found with Chicano English speakers, I argue that back vowel fronting is indeed found in the speech of young KAs. In addition, I propose that KAs do not uniformly demonstrate /u/- and /oo/-fronting, but are influenced by at least three contexts: first, the phonological rules triggered by adjacent segments; second, an individual speaker’s ties to Korean culture and KA peers may inhibit fronting [7]; third, considering the prevailing theory of interactional identity production [14], a speaker’s ever-shifting stance toward Korean culture and identity may be reflected in moment-by-moment fluctuations in fronting.

The data are drawn from twenty casual bilingual interviews held with second generation KAs. Each interview lasted between thirty minutes to one hour and included questions about the consultant’s language background and attitudes toward Korean language and identity. Interviews were transcribed in Praat [15], force-aligned [16], and analysed with IFC formant tracker [17]. Results show that male and female second generation KAs do demonstrate significant /u/-fronting, as calculated by distance between F2 of /u/ and the anchor vowel /i/, while male speakers demonstrate much more /oo/ fronting than females (Fig. 1). Gender was shown to affect the amount of fronting for /u/ ( $\chi^2(1)=10.99$ ,  $p<0.001$ ) and /oo/ ( $\chi^2(1)=16.896$ ,  $p<0.001$ ). In addition, a speaker’s self-identification as being ethnically “Korean American” (as opposed to simply “Korean”) correlated with fronted /oo/ ( $\chi^2(1)=4.29$ ,  $p=0.038$ ).

A closer look into individual speakers reveals that frontedness also varies widely over the course of an interview. For example, in Figure 2, Subject 19 maintains lower-than-average F2 for most tokens of the word *school*, to be expected given its pre-lateral phonological context. When he utters the word *school* with a particularly fronted /u/ (at around 1350sec), he is beginning to talk about how he did not associate with other Korean Americans in high school and knew only KAs who, like him, spoke much more English than Korean. The subject’s brief discussion of his social circle precedes a global rise in F2 for all vowels, of which the fronted “Californian” *school* is just the beginning.

These findings contribute to the broadening understanding of ethnic identity and sociolinguistic variation, while also promoting future studies of topic-based intra-individual linguistic variation to develop models of situational identity formation.



**Figure 1 (left).** Mean normalized F1 and F2 vowel plots for all second generation Korean American consultants, by gender. Both groups show a fronted /u/ (“UW”) token, and males showed a considerably fronted /oo/ (“OW”) token. **Figure 2 (right).** Raw F2 values for /u/ tokens for one subject, a second generation Korean American male, plotted over time, reveal that while the [+back] phonological environment results in a lower-than-average F2 for /u/ in most instances of the word *school*, there are tokens of *school* with much higher F2, which can be accounted for by looking at the topic and the speaker’s stance during each utterance.

[1] Teresa Pratt and Annette D’Onofrio. Jaw setting and the California Vowel Shift in parodic performance. *Language in Society*, 46(3):283-312, 2017.

[2] Leanne Hinton, Birch Moonwomon, Sue Bremner, Herb Luthin, Mary Van Clay, Jean Lerner, and Hazel Corcoran. It’s not just the Valley Girls: A study of California English. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 13:117-128, 1987.

[3] Robert J. Podesva. The California Vowel Shift and gay identity. *American Speech*, 86(1):32-51, 2011.

[4] Robert J. Podesva, Annette D’Onofrio, Janneke Von Hofwegen, and Seung Kyung Kim. Country ideology and the California Vowel Shift. *Language Variation and Change* 27(2):157-186, 2015.

[5] Lauren Hall-Lew. The completion of a sound change in California English. In *Proceedings of the International Congress of Phonetic Sciences*, 17:807-810, 2011.

[6] Robert Kennedy and James Grama. Chain Shifting and Centralization in California Vowels: An Acoustic Analysis. *American Speech*, 87(1):39-56, 2012.

[7] Penelope Eckert. Where do ethnolects stop? *International Journal of Bilingualism*, 12(1-2):25-42, 2008.

[8] Carmen Fought. A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics*, 3(1):5-23, 1999.

[9] Grace Cho, Kyung-Sook Cho, and Lucy Tse. Why ethnic minorities want to develop their heritage language: The case of Korean-Americans. *Language, Culture, and Curriculum*, 10(2):106-112, 1997

[10] Jin Sook Lee. The Korean language in America: The role of cultural identity in heritage language learning. *Language, Culture, and Curriculum*, 15(2):117-133, 2002.

[11] Sarah J. Shin. *Developing in Two Languages: Korean Children in America*. Multilingual Matters, Clevedon, UK, 2005.

[12] James E. Flege. Language contact in Bilingualism: Phonetic systems interactions. In Cole & Hualde (eds.), *Laboratory Phonology 9*. Berlin; Mouton de Gruyter, 2006.

[13] Johanne Paradis. Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism* 5(1):19-38, 2001.

[14] Mary Bucholtz and Kira Hall. Identity and interaction: A sociocultural linguistic approach. *Discourse Studies* 7(4-5):585-614, 2005.

[15] Paul Boersma and David Weenink. Praat: Doing phonetics by computer [Computer program]. Version 6.0.40, 2018.

[16] Ingrid Rosenfelder, Josef Fruehwald, Keelan Evanini, and Jiahong Yuan. FAVE (Forced Alignment and Vowel Extraction) Program Suite [Computer program]. 2011.

[17] Ronald L. Sprouse and Keith Johnson. The Berkeley Phonetics Machine. In *INTERSPEECH* 2016:1623-1626. 2016.

# **From Kanauji to Kanpuria: Language Contact and Change in Kanpur**

**Pankaj Dwivedi**

**Indian Institute of Technology Ropar**

**Central Institute of Indian Languages**

## **Abstract**

India is a multilingual country and so are its all states, districts, cities, and towns. It cherishes its great diversity in terms of languages, cultures, beliefs, value systems, food habits, and religions. Kanpur is a major district in the state of Uttar Pradesh, India. Due to Kanpur being a prominent educational and industrial hub of the state, every year thousands of students and workers migrate to the district in the search of better prospects in career and professional life. And, these people also significantly add to its linguistic repertoire and code-matrix.

Kanauji is spoken by about 6 million people in total across six districts of Uttar Pradesh including Kanpur. It is surrounded by four other major dialects of Hindi. Kanauji has little recognition in academia and administration. Additionally, pressure from Standard Hindi, English and presence of other prominent languages and dialects are changing its structure on all linguistic levels, especially in the urban areas of Kanpur. As a result, in the last decade, there has been emergence of new linguistic variety, popularly known as Kanpuria among youths, which primarily is a mix of Standard Hindi, Kanauji and English (Chaturvedi, 2015). Dwivedi and Kar (2016, 2018) also provide examples of language contact situation in Kanpur.

This paper sheds light on the types and degrees of bilingualism among Kanauji speakers across different age groups and gender, giving detailed examples from code-mixing and code-switching instances. Language attitude of Kanauji speaking people towards Kanpuria, Kanauji and Hindi will also be discussed.

The data for this study has been collected during a 15-day linguistic fieldwork and stored in both text and speech media. The database is annotated using Praat software. The limitation of the study lies in the fact that Kanpur is a large district surrounded by neighboring varieties viz. Bundeli, Awadhi, Braj, and Bagheli on different sides of borders. Therefore, the morphological and phonological structure of Kanpuria may slightly differ on these borders areas.

**Keywords:** Western Hindi, Kanpuria, Kanauji, Language contact, Language Change

**Reference:-**

- Chaturvedi, Sujata. 2015. A Sociolinguistic study of linguistic variation and code matrix in Kanpur. *Procedia-Social and Behavioral Sciences*, 192, 107-115.
- Dwivedi, Pankaj, and Kar, Somdev. 2018. Phonology of Kanauji. In G. Sharma (Ed), *Advances in Hindi Language Teaching and Applied Linguistics* (pp. 189-220). Munich: Lincom Europa
- Dwivedi, Pankaj, and Kar, Somdev. 2016. Kannauji of Kanpur: A brief overview. *Acta Linguistica Asiatica*, 6(1): 101-119.