**Appendix A: Systematic criteria for binning labels.**

Labels that were explicitly a language name were coded as separate categories from labels that talked only about the region, or regional accent of Mandarin (cf. Bucholtz et al. 2007 for similar treatment of language name labels vs. dialect labels of Californian participants). For the name “Shanghainese” (上海话) specifically, even though this is not an explicit reference to the language family of Wu languages (吴), we coded the entry *上海话* ‘Shanghainese’ as referring to the non-Mandarin language of Wu Shanghainese. The format of *[place name] + 话* ‘[place name] + speech’ is not always consistent in whether it refers to a variety of Mandarin or a non-Mandarin language. In the case of *Shanghai+speech* (Shanghainese), this term is used in China to refer to the non-Mandarin language. However, something like *Northeast+speech* (东北话 ‘Northeast dialect’) refers to the variety of Mandarin spoken in the Northeast. Thus, *上海话* ‘Shanghainese’ was coded as a separate language, whereas *上海* ‘Shanghai’ was coded as Shanghai-accented Mandarin, since it made reference only to the location, and not a separate language.

If a label had more than one place name in it, it was separated into two entries and given one count to each place name that was mentioned. Labels that explicitly mentioned cities were coded as separate from the province. For example, Guangzhou (a city in Guangdong) was coded separately from Guangdong (the province), because Guangzhou speech is not the same as speech everywhere else in Guangdong; there is much diversity across the province in the ways people speak (see, e.g., Wu et al. 2021; Yue-Hashimoto 1991). Labels that were not a region or language name, but rather a general description (e.g. “more Standard”), were coded as a separate entry as well. If the general description labels mentioned specific regions, an entry was added to each of those regions.

Labels that mentioned a variety of Mandarin spoken in many areas, such as Yilu Mandarin (spoken in many provinces), were coded as a separate entry but did not add an entry for each province that that variety is spoken in, since those labels referred to a specific type of Mandarin speech, rather than an area. However, labels that were a three-character province abbreviation code were coded as separate entries for each province mentioned because those kinds of labels make reference to specific locations. The three-character abbreviation code is a common way for people to refer to three neighboring provinces with a shorthand; this is also used in license plate abbreviations in China, e.g., *京津冀* *Jing-Jin-Ji*, ‘Beijing + Tianjin + Hebei’, a standard shorthand for referring to those three regions at once (Chung & Lam 2010).

In coding the names for Hokkien/Southern Min (non-Mandarin) varieties in the South, we employed systematic ways to separate Taiwanese Southern Min from Southern Min as a language family, which would also encompass the Hokkien varieties spoken in Fujian. Southern Min, or *闽南 Minnan* ‘lit. Min-South’, is a language family that encompasses Hokkien. The term “Taiwanese” refers to the variety of Southern Min spoken in Taiwan. The term “Hokkien” refers to the variety of Southern Min spoken in Fujian (Chiung 2015). The term *闽南话 Minnanhua* ‘Southern Min variety’ is often used to describe both the language spoken in Taiwan and in Fujian. The Taiwan and Fujian varieties are quite mutually intelligible in some regions of Fujian, while other regions of Fujian differ from the Taiwan variety more (Hsieh 2014). However, a clear term for Taiwan Southern Min is “Taiwanese”, so if a label said *台语* *Taiyu* ‘Taiwanese’ specifically, I coded it as separate from *闽南话* *Minnanhua* ‘Southern Min’, because it potentially marked a political distinction (cf., Chiung 2015).

Labels that just said “Taiwan” were coded as Taiwanese-accented Mandarin (or, Taiwan Mandarin). Labels that said “Fujian” were coded as Fujian-accented Mandarin. These labels make reference to the location and not the language that is spoken in that region.