Extra-Linguistic Data for Understanding Dialect Mutual Intelligibility

Chin-Chuan Cheng
City University of Hong Kong
University of Illinois at Urbana-Champaign

We do dialect surveys all the time. We look at some linguistic phenomena and attempt to make a generalization from the collected data about a certain aspect of language. We might revisit the phenomena and collect more data. Another person might study the same subject and might come up with a different set of data. Thus, there is no pure data to speak of. What we collect are the matters that are filtered by our approach, framework or theory. When we ask a slightly different but entirely relevant question, our data become useless in the argumentation. I will illustrate this point by using our own study of Chinese dialect mutual intelligibility. I hope to conclude that extra-linguistic matters such as population movement, geographical barriers, etc. will help us understand how speech communities were formed in the past.

We approach the question of mutual intelligibility by first collecting the sound correspondence patterns between a pair of dialects. For example, the nasal n initial consonant in Beijing corresponds to zero or null initial, l, n, S, and s on the basis of the Chinese dialect database (Beijing University 1962, Cheng 1994a). In our database there were 53 cognate items. The average for the 5 patterns was 10.6. If the number of cognates in a pattern was smaller than the average, the pattern was considered as interfering noise. Otherwise it was taken as communication-enhancing signal. The corresponding sounds may be different as indicated below:

|							
a.	n	zero	3	10.6	noise	different	zero occurs elsewhere in Beijing
b.	n	l	1	10.6	noise	different	l occurs elsewhere in Beijing
c.	n	n	27	10.6	signal	same	
d.	n	S	21	10.6	signal	different	S does not occur in Beijing
e.	n	s	1	10.6	noise	different	s occurs elsewhere in Beijing

A weighting scale was established for the patterns. We calculated the values of all the sound correspondence patterns between Beijing and Jinan to derive an index of mutual intelligibility. We actually calculated the mutual intelligibility of all the pairs of Beijing, Jinan, Xi'an, Taiyuan, Hankou, Chengdu, Yangzhou, Suzhou, Wenzhou, Changsha, Shuangfeng, Nanchang, Meixian, Guangzhou, Xiamen, Chaozhou, and Fuzhou (Cheng 1994b, 1996). The mutual intelligibility indices were used in a cluster analysis to establish the affinity tree given in Figure 1. In the figure, the higher the degree at which point two dialects or groups of dialects are connected the

higher the mutual intelligibility.

Now that we have established the dialect mutual intelligibility, we should be able to explain a few things. We may ask why Chaozhou in Guangdong and Shuangfeng in Hunan are so low in mutual intelligibility. Our linguistic answer is that the sound correspondence patterns between the pair of dialects yielded a low sum. Thus we answer the question in terms of our own internal data. If we further ask why they yielded a low sum, then we need to explain how the dialects were formed historically and how they evolved without extensive interaction. But we have little information on such matters. By looking at some maps of China we see that mountains separate the two regions. Perhaps the mountains constituted barriers for communication and thus caused the low mutual intelligibility. On the other hand, we see that Hankou in Hubei and Chengdu in Sichuan have the highest degree of mutual intelligibility. They belong to Southwestern Mandarin. And yet they are separated by even more towering mountains and the treacherous Three Gorges. So geography does not seem to help here. Historians will tell us that population movement centuries ago caused the spread of the Southwestern Mandarin in that region.

The movements of the Kejia people and the formation of their speech islands have been discussed often. Individual attempts have also been made to explain certain dialect situations in term of culture (Zhou and You 1986). The Chinese language has a continuous written history of over 3,000 years. It covers a large portion of Asia. It would seem that data are abundant for our use. But data are derived from the first approximation of analysis. As far as I know, at present extra-linguistic data in large amount for the Chinese language in time and space are not readily available. We hope that the use of a standard geographical information system will allow us to obtain useful data from other disciplines for integration with linguistic data so as to answer many relevant linguistic and other questions.

Figure 1. Dialect Affinity Based on Mutual Intelligibility
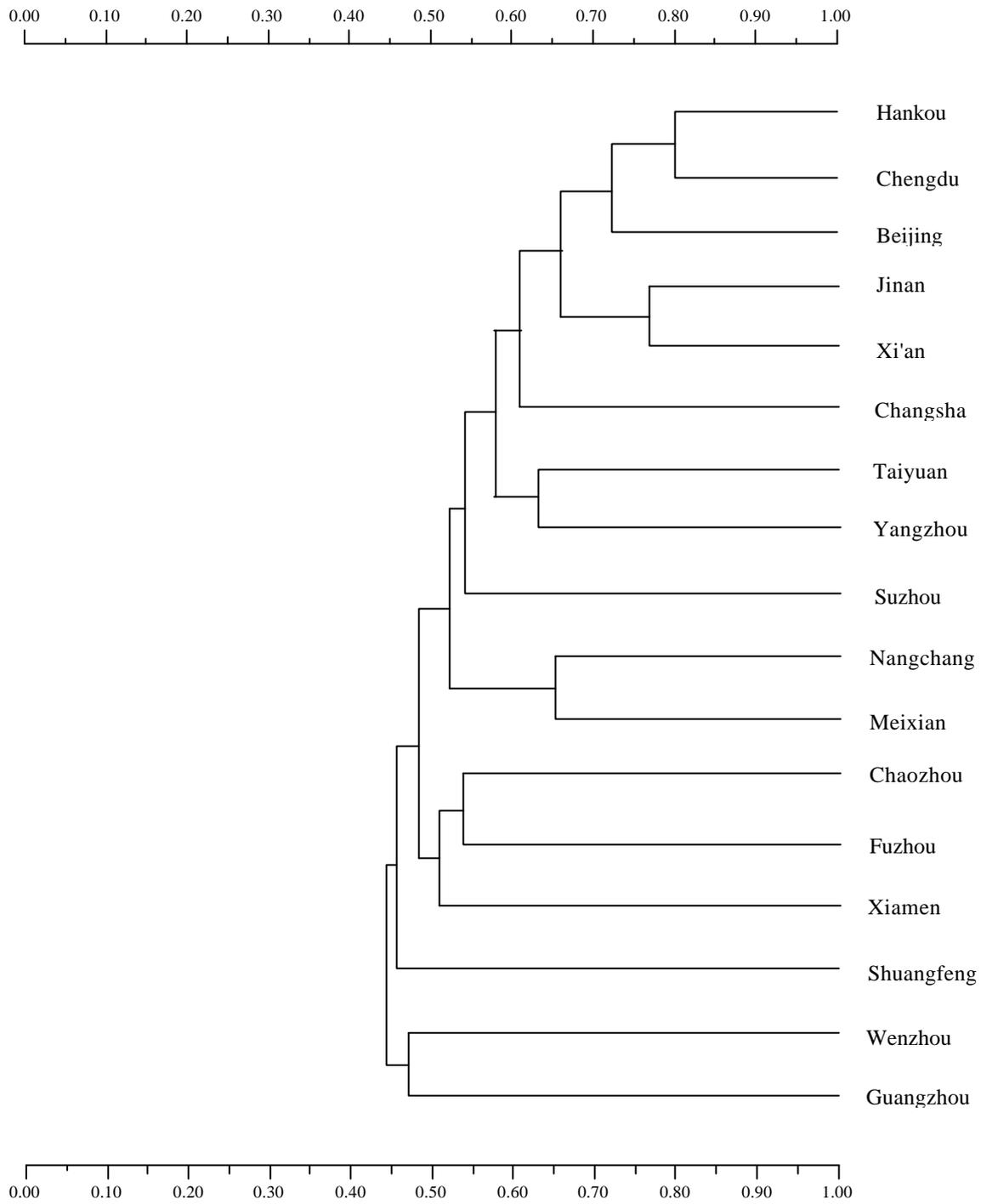
REFERENCES

Beijing University, *(Chinese Dialect Character Pronunciation List)*, 1962, Beijing: Wenzi Gaige Chubanshe.

Chen, Matthew Y. and Ovid J.L. Tzeng, (eds.), *In Honor of William S-Y. Wang: Interdisciplinary Studies on Language and Language Change*, 1994, Taipei: Pyramid Press.

Cheng, Chin-Chuan, "DOC: Its Birth and Life", in Matthew Y. Chen and Ovid J.L. Tzeng, (eds.), *In Honor of William S-Y. Wang: Interdisciplinary Studies on Language and Language Change*, 1994a, pp. 71-86.

Cheng, Chin-Chuan, " (Calculation of Chinese Dialect Mutual Intelligibility)", *(Zhongguo Yuwen)*, 238 1994b, pp. 35-43.

Cheng, Chin-Chuan, "Quantifying Dialect Mutual Intelligibility", in Huang and Li, (eds.), *New Horizons in Chinese Linguistics*, 1996, pp. 269-292.

Huang, James and Audrey Li, (eds.), *New Horizons in Chinese Linguistics*, 1996, Boston: Kluwer Academic Publishers.

Zhou, Zhenhe and Rujie You. *(Dialects and Chinese Culture)*, 1986, Shanghai: Shanghai Renmin Chubanshe.