

The Discrepancy Between Spoken and Written Chinese— Methodological Notes on Linguistics

Ľuboš Gajdoš

The issue of choosing language data on which synchronous linguistic research is being done appears in many ways not only to be relevant to the goal of the research, but also to the validity of the research results. The problem which particularly concerns us here is the discrepancy between speech on the one hand and written language on the other. In this context, we have often encountered in the past a situation where the result of the research conducted on a variety of the Chinese language has been generalized to the entire synchronous state of the language, i.e. to all other varieties of the language, while ignoring the mentioned discrepancy between the spoken and written forms. The discrepancy between the spoken and written forms is likely to be present in any natural language with a written tradition, but the degree of difference between languages is uneven: e.g. compared to the Slovak language, it may be stated that the situation in Chinese is in this respect extraordinary.¹ Nevertheless, it is surprising that the quantitative (qualitative) research on discrepancies between different varieties of the language has not yet aroused the attention of Chinese linguistics to such an extent as would have been adequate for the unique situation of this natural language.

1 The degree of non-correspondence between written language and speech across languages is determined by many factors. One of them, certainly, is the unequal extent of a written tradition.

Kouyu versus shumianyu

Considering the above mentioned notes, modern Chinese is differentiated into spoken language, *kǒuyǔ* 口語, and written language, *shūmiànyǔ* 書面語. Both terms are defined very vaguely in Chinese linguistics, so let us use at least the definition of *kouyu* which is presented in a university textbook for students of Chinese language:² »*Kouyu* is the oral language used by people having a colloquial style.« *Shumianyu* is then »a language recorded in Chinese characters and constituted by *kouyu*. It can be easily re-cultivated until it becomes very concise, rigorous, compact, and thus has a different style than *kouyu*.«³ *Kouyu* is usually translated into foreign languages with an equivalent—‘spoken language’, occasionally ‘colloquial style’.⁴ *Shumianyu* is generally translated as ‘written language’, less commonly as ‘literary language’. *Shumianyu* is, *inter alia*, lexically characterized by an abundance of archaisms, idioms and sayings.⁵

Although the written Chinese of the 20th century underwent (especially after the May Fourth Movement) tremendous changes which brought it closer to the spoken language (oral language), the differences between them are still enormous. The *status quo* of the modern Chinese language is a result of these changes with its positive and negative aspects which both affect surveyed language data.⁶ For example, the markers *le* 了, *zhe* 著, *guo* 過 and complements are less frequent in *shumianyu*,⁷ and some grammatical phenomena do not even occur in *shumianyu* at all or only to a limited extent.⁸ Richard Xiao and Tony McEnery, who in their research on verbal aspect were also engaged with this

2 Huang Borong 黃伯榮 and Liao Xudong 廖序東, *Xiandai Hanyu* 現代漢語 [Modern Chinese] (Beijing: Gaodeng jiaoyu chubanshe, 1997), 1.

3 *Ibid.*, 1.

4 For more information, see William Hannas, *Asia's Orthographic Dilemma* (Honolulu: University of Hawaii Press, 1997), 248.

5 For more information, see Cao Wei 曹燁, »Xiandai Hanyu kouyu ci he shumianyu ci de chayi chutan« 現代漢語口語詞和書面語詞的差異初探 [A Preliminary Study on the Differences in Vocabulary between the Colloquial and the Literary in Modern Chinese], *Yuyan jiaoxue yu yanjiu* 語言教學與研究 6/2009, 39–44.

6 For more information, see Hannas, *Asia's Orthographic Dilemma*, 128–129, or John DeFrancis, *The Chinese Language. Fact and Fantasy* (Honolulu: University of Hawaii Press, 1984), 152–153.

7 Richard Xiao and Tony McEnery, *Aspect in Mandarin Chinese* (Amsterdam: John Benjamins, 2004).

8 Verb reduplication may illustrate this point.

subject, came to the conclusion that aspect markers are significantly more present in narrative texts.⁹ Likewise, Liu Yuehua and his colleagues claim that complements are more frequent in spoken than in written language.¹⁰

A decisive factor for the discrepancy, among other things, is the intellectualization of a language. For example, one of the essential sources that affects the current written language in lexis and syntax is the literary language *wenyanwen*. Feng Shengli, in relation to the syntax of written Chinese, has observed that »*shumianyu* differs from *wenyanwen* as well as from *kouyu*, and it is therefore quite possible to speak of a separate grammar system.«¹¹ Thus, a synchronous state of the written language is the result of this influence (undoubtedly, not the only one), in which the spoken language and the literary language *wenyanwen* are blended together. A consequence of these trends is the written language, which although based on the spoken language includes such »foreign« elements—the residue of the literary language *wenyanwen*—which in some cases may give rise to the situation that a percipient of a text (which is being read aloud) does not understand it.¹² Despite the great efforts of language reformers in the 20th century, this phenomenon could not be removed from the modern language. On the contrary, we are witnessing the increasing leakage of *wenyanwen* into the written language of *putonghua* 普通話. This fact is documented by Chen Ping's argument that the present share of *wenyanwen* in written *putonghua* is 30 to 50 per cent.¹³ Therefore, it may be concluded that the written language *shumianyu* in terms of its vocabulary and syntax is a sort of intermediate stage on the axis *wenyanwen* → *shumianyu* ↔ *kouyu*, while *shumianyu* and *kouyu* mutually interact.

The problem of linguistic research on Chinese in the 20th century is the fact that for instance grammatical phenomena have been frequently studied on texts which actually did not truly represent/record spoken language (speech) at that

9 Xiao and McEnery, *Aspect in Mandarin Chinese*, 249.

10 Liu Yuehua 劉月華 and Pan Wenyu 潘文娛 and 故 (韋、華), *Shiyong xiandai Hanyu yufa* 實用現代漢語語法 [Practical Chinese Grammar] (Beijing: Waiyu jiaoxue yu yanjiu chubanshe, 2004), 534.

11 Feng Shengli 馮勝利 馮勝利, »Shumianyu yufa ji jiaoxue de xiangdui dulixing« 書面語語法及教學的相對獨立性 [author's translation »An Independent Grammar for Written Chinese in Second Language Teaching«], *Yuyan jiaoxue yu yanjiu* 2/2003, 53–63.

12 Incidentally, this is also one of the reasons why it is not satisfactorily possible to replace Chinese characters with a phonetic script (for example, based on Latin) at present. For more details see DeFrancis, »The Indispensability Myth«, ch. in *The Chinese Language*, 189–203.

13 Chen Ping, *Modern Chinese: History and Sociolinguistics* (Cambridge University Press, 1999).

time; quite often the examples were drawn from literary works of the first half of the 20th century (the so-called modern literature 現代文學),¹⁴ a period which constituted a *de facto* transition between *baihua*wen¹⁵ and *wenyan*wen,¹⁶ although from the point of view of literary theory, the works of this period already belong to *baihua*wen literature. Despite this evidence, research results were repeatedly generalized onto the entire synchronous state of the language, and in their application intervened in the field of language education as well.¹⁷

Current linguistic research must therefore be seen through the critical view of that mentioned above. Should we therefore reject all of this kind of research? Certainly not, for the results of research on any language data can help describe the general view of the synchronous state of Chinese language and can be looked upon as an integral part of the language mosaic. In this regard, is the use of a large general corpus (vast bulk of language data) which covers all varieties of language a solution? Yes and no. Yes, it may help to objectify the research and its applications. However, the difficulty in describing different varieties still persists. So what is the appropriate source of language data in terms of media, area and social status of native Chinese speakers, etc.?¹⁸ Our knowledge suggests

14 It became a manifestation of some kind of »good manners« in Chinese synchronous linguistics in the second half of the 20th century to cite only the so-called classics of modern literature. In this regard, it is worth noting what John Sinclair has observed about literary texts in English (in connection with corpus building). He stated that these texts are a small representation of normal everyday English and they tended to be innovative. For more information, see John Sinclair, *Corpus, Concordance, Collocation* (Oxford University Press, 1991), 17. Although this claim is valid for English, we suppose that to some extent it is universally applicable to literary texts of any natural language.

15 *Baihua*wen ('written vernacular Chinese') can be considered as the predecessor of the current standard language. For more information, see Chen Ping, *Modern Chinese: History and Sociolinguistics*.

16 On the other hand, this claim is not absolutely valid because among the classics of so-called modern literature can be found works with a language close to speech.

17 A good example, otherwise very inspiring, is provided in Fang Yuqing's 房玉清 grammar *Shiyong Hanyu yufa* 實用漢語語法 [Practical Chinese Grammar] (Beijing: Beijing yuyan wenhua daxue chubanshe, 1991).

18 There are some examples of spoken corpora (in this case Wenzhou dialect), such as *Wenzhou Spoken Corpus/Wenzhou kouyu yuyan ziliao ku* 溫州口語語言資料庫 <corpora.tapor.ualberta.ca/wenzhou/> (last retrieval Dec 20, 2009). Nevertheless, they are still insufficient in some crucial factors. For more information on this issue, see Gu Yueguo, »Sampling Situated Discourse for Spoken Chinese Corpus«, <ling.cass.cn/dangdai/gu_papers/sampling%20situated%20discourse.

such a corpus does not yet exist in a form that may satisfactorily meet these requirements.¹⁹

Conclusions

If there was a corpus of spoken language with adequate parameters, we may then draw outlines of different varieties of language. But not only that: it may serve as a fixed point of reference that reflects the current state of language as well as a basis to which future states of language may be compared in terms of lexical change and syntactical patterns. All these elements could also help to study the stratification of the Chinese language, a field of linguistics that deserves more attention as until now only a fraction of stratification has been examined.

Besides linguistic research, description is not the goal in itself, but rather a means for a better picture of *shumianyu* (supported by statistical data) as well. It should result in a grammar (or dictionary) of *shumianyu* which is needed in second language acquisition nowadays. Thus, students (primarily not native speakers) of Chinese would have a better opportunity to master both the written and spoken forms.

This contribution should be seen as a stimulus for further research on the spoken language rather than as an answer to questions concerning the discrepancy between *kouyu* and *shumianyu*.

Comenius University in Bratislava, Department of East Asian Studies

pdf> (May 10, 2011), or Chui Kawai and Lai Huei-ling, »The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min«, *Taiwan Journal of Linguistics* 6,2 (2008), 119–144.

19 M. A. K. Halliday and E. McDonald in their »Metafunctional Profile of the Grammar of Chinese«, in *Language Typology. A Functional Perspective*, ed. by A. Caffarel (Philadelphia: John Benjamins, 2004), 380–396, also noticed this fact.