SINO-PLATONIC PAPERS

Number 134

May, 2004

Two Steps Toward Digraphia in China

by Xieyan Hincha

Victor H. Mair, Editor
Sino-Platonic Papers

Department of East Asian Languages and Civilizations
University of Pennsylvania
Philadelphia, PA 19104-6305 USA
vmair@sas.upenn.edu
www.sino-platonic.org

SINO-PLATONIC PAPERS

FOUNDED 1986

Editor-in-Chief Victor H. Mair

Associate Editors
PAULA ROBERTS MARK SWOFFORD

ISSN

2157-9679 (print)

2157-9687 (online)

SINO-PLATONIC PAPERS is an occasional series dedicated to making available to specialists and the interested public the results of research that, because of its unconventional or controversial nature, might otherwise go unpublished. The editor-in-chief actively encourages younger, not yet well established, scholars and independent authors to submit manuscripts for consideration. Contributions in any of the major scholarly languages of the world, including romanized modern standard Mandarin (MSM) and Japanese, are acceptable. In special circumstances, papers written in one of the Sinitic topolects (*fangyan*) may be considered for publication.

Although the chief focus of *Sino-Platonic Papers* is on the intercultural relations of China with other peoples, challenging and creative studies on a wide variety of philological subjects will be entertained. This series is **not** the place for safe, sober, and stodgy presentations. *Sino- Platonic Papers* prefers lively work that, while taking reasonable risks to advance the field, capitalizes on brilliant new insights into the development of civilization.

Submissions are regularly sent out to be refereed, and extensive editorial suggestions for revision may be offered.

Sino-Platonic Papers emphasizes substance over form. We do, however, strongly recommend that prospective authors consult our style guidelines at www.sino-platonic.org/stylesheet.doc. Manuscripts should be submitted as electronic files, preferably in Microsoft Word format. You may wish to use our sample document template, available here: www.sino-platonic.org/spp.dot.

Beginning with issue no. 171, *Sino-Platonic Papers* has been published electronically on the Web at www.sino-platonic.org. Issues 1–170, however, will continue to be sold as paper copies until our stock runs out, after which they too will be made available on the Web.

Please note: When the editor goes on an expedition or research trip, all operations (including filling orders) may temporarily cease for up to three months at a time. In such circumstances, those who wish to purchase various issues of *SPP* are requested to wait patiently until he returns. If issues are urgently needed while the editor is away, they may be requested through Interlibrary Loan. You should also check our Web site at www.sino-platonic.org, as back issues are regularly rereleased for free as PDF editions.

Sino-Platonic Papers is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/2.5/ or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Abstract

- (a) Pinyin (Hanyu Pinyin Fang'an) is a reasonable and workable phonographic script for encoding modern standard Chinese (Putonghua). It is neither an "auxiliary instrument for transcribing Chinese characters" nor a "phonetic system for Chinese", as ISO and the PRC government put it.
- (b) The allegedly overwhelming homophony of monosyllabic morphemes and the resulting ambiguity (rendering the application of a Latin-based writing system impossible) do not exist. Whatever a blind Chinese is able to understand, can be encoded and reproduced, using a phonographic system of writing.
- (c) Homophonous morphemes are distinguished by the fact that they belong to different word-classes or other functional categories, or they are bound morphemes. Within a given word-class, they are not true homophones, but near-homophones, differentiated by tone change, rhotacization, the neutral tone, and other supra-segmental features. Homophony in modern Chinese is as peripheral as in other languages.
- (d) The means to differentiate near-homophonous morphemes are invisible with traditional Chinese characters, but they are (or could be) clearly indicated in Pinyin. Therefore, Pinyin is superior to Chinese characters in writing modern Chinese.

•

Two Steps Toward Digraphia in China

By Xieyan Hincha¹

From the outset, that is from 1955 onward, what was meant by Hànyũ Pīnyīn was unclear: a new script, a transcription system, or something similar. Ye Laishi, Lu Zhiwei, and Zhou Youguang jointly devised the "Hànyǔ Pīnyīn Wénzì Fāng'an Chūgăor" (First draft of a phonographic script for the Chinese language; abbreviated Chūgăor). That same year, the PRC Committee for Script Reform (Zhōngguó Wénzì Găigé Wĕiyuánhuì 中国文字改革委 员会; from Dec. 1985 on renamed: Guójiā Yǔyán Wénzì Gōngzuò Wĕiyuánhuì 国家语言 文字工作委员会 National Working Committee for Language and Script) deleted the key term Wénzi (script) from the Chūgǎor (first draft). It is worth noting that this was not done for stylistic reasons (cf. Zhou, Youguang 1999:197). There seems to have been no argument within the committee. The message was clear: Hànyũ Pīnyīn was not meant to be a script or to become a script. People began unanimously to say Hànyũ Pīnyīn Fāng'àn bú shì pīnyīn wénzì (Hanyu Pinyin is not a phonographic script), or even Hànyŭ Pīnyīn bú shì wénzì (Hanyu Pinyin is not a script). Presumably, that would have been the last word on the matter, had it not been for the momentous open-door policy that followed the Cultural Revolution. The People's Republic of China (PRC) had begun to participate in international standardization efforts as early as 1975, and proposed that the Hànyũ Pīnyīn Fāng'àn be employed instead of Wade-Giles. Before the arrival of ISO 7098:1982, the French representative, backed by other delegations, proposed that an orthographic regulation be added. The Chinese representative promised that a new orthography would subsequently be added (cf. Zhou, Youguang 1992:273). This reopened the question of whether Hànyũ Pīnyīn is a script. The question remains unanswered to this day.

Is Hànyǔ Pīnyīn a script?

Before answering this question, I would like briefly to discuss two important documents.

¹ Translated from the German by Paul Frank.

1. Law on Language and Script in the People's Republic of China

On October 31, 2000, the 中华人民共和国国家通用语言文字法 Zhōnghuá Rénmin Gònghéguó Guójiā Tōngyòng Yǔyán Wénzìfă (Law on the common national language and script of the People's Republic of China) was promulgated. Paragraph 18 concerns the Hànyǔ Pīnyīn Fāng'àn. The first sentence reads:

国家通用语言文字以《汉语拼音方案》作为拼写和注音工具 Guójiā tōngyòng yǔyán wénzì yǐ "Hànyǔ Pīnyīn Fāng'àn" zuòwéi pīnxiĕ hé zhùyīn gōngjù. (The common national language and script employ the Hanyu Pinyin Fang'an as an instrument for writing with letters and to indicate the pronunciation.)

In order to understand this paragraph, one must first read and understand Paragraph 2, which reads:

本法所称的国家通用语言文字是普通话和规范汉字 Bén fǎ suǒchēng de guójiā tōngyòng yǔyán wénzì shì Pǔtōnghuà hé guīfàn Hànzì. (The common national language and script as defined by this law are Putonghua and standard Chinese characters.)

Summarized, these two sentences say that $P\bar{u}t\bar{o}nghu\dot{a}$ and standard Chinese characters employ the $H\dot{a}ny\bar{u}$ $P\bar{i}ny\bar{i}n$ $F\bar{a}ng'\dot{a}n$ as an instrument for writing (拼写 $p\bar{i}nxi\check{e}$) or to indicate the pronunciation (注音 $zh\dot{u}y\bar{i}n$).

The 现代汉语词典 Xiàndài Hànyǔ Cidiǎn (Dictionary of modern Chinese, 1996 edition. abbreviated Xiàn-Hàn) defines the words pīnxiĕ, pīnyīn, and zhùyīn as follows:

拼写: 用拼音字母按照拼音规则书写 pīnxiě: yòng pīnyīn zìmǔr ànzhao pīnyīn guīzé shūxiě (writing with pinyin letters in accordance with the rules of pinyin) (p. 974).

拼音: 把两个或两个以上的音素结合起来成为一个复合的音, 如 b 和 iao 拼成 biao (标) pīnyīn: bá liǎng ge huò liǎng ge yǐshàng de yīnsù jiéhé qǐlai chéngwéi yí gè fùhé de yīn, rú b hé iao pīnchéng biao (biāo) (combining two or more phonemes into a single sound complex, e.g., b and iao produce biao [biao]) (p. 974).

注音: 用符号表明文字的读音 zhùyīn: yòng fúhàor biǎomíng wénzì de dúyīn (clearly indicating the pronunciation of Chinese characters by means of symbols) (p.1646).

Based on Xiàn-Hàn, Paragraph 18, Subparagraph 1, must or can be understood to mean the following:

Both Pătonghuà and standard Chinese characters employ Hànyũ Pīnyīn Fāng'àn as an

"instrument" for writing with $P\bar{\imath}ny\bar{\imath}n$ letters according to the rules of $P\bar{\imath}ny\bar{\imath}n$, as well as to indicate the pronunciation of Chinese characters. This sentence is unclear in some respects. In this subparagraph, the legislators made at least three points:

- 1. Hànyǔ Pīnyīn Fāng'àn is an "instrument" (工具 gōngjù);
- 2. it is used to write *Pŭtōnghuà* with *Pīnyīn* letters in accordance with *Pīnyīn* rules (拼写 pīnxiĕ); and
- 3. in addition, it is used to indicate the pronunciation of standard Chinese characters.

The generic term "instrument" to denote Hànyŭ Pīnyīn Fāng'àn does not exist in linguistics. We may therefore forget it. The correct generic term is "graphic system.".

The legislators meant to say that $P\bar{\imath}ny\bar{\imath}n$ $F\bar{\imath}ng'\dot{\imath}n$ is a phonographic writing system for $P\bar{\imath}ut\bar{\imath}nghu\dot{\imath}a$. A phonographic script represents very specific sound units of a language system: either its sounds or phonemes or syllables, and so forth. But the formulation of this law contains several contradictions. Here I identify one of them:

It is no secret that the Chinese leadership cannot stand the thought that Chinese characters might be driven out by $P\bar{\imath}ny\bar{\imath}n$. That is why an attempt has been made to interpret the $P\bar{\imath}ny\bar{\imath}n$ system as a phonetic system and to subordinate it to Chinese characters, in order to defuse the discussion about the future of Chinese characters. That is how the "indication of the pronunciation of Chinese characters" formulation came about. Thus, the creation of a Latin script for Chinese was initially confounded with the problems of Chinese characters and concerns about their future, in order to then solve all these questions in one fell swoop. Let me state this more precisely:

The legislators are of course fully aware that thousands of Chinese characters are available as a functioning writing system for modern Chinese, but also that this system poses numerous problems to everyone who employs the Chinese language and script. Let me cite just one of these problems: Chinese characters contain little information about pronunciation. Some people go so far as to say that Chinese characters provide no indication as to pronunciation. This view is clearly mistaken. We are all familiar with the phonetic elements of Chinese characters. Formerly, it was claimed that approximately 90% of Chinese characters contained a phonetic element. Today it is estimated that approximately 57% of the 7,000 generally used modern characters, the so-called 通用字 Tōngyòngzì (generally used characters), contain a phonetic element. Of these, only approximately 37% contain sufficiently precise indications of pronunciation. (On this topic, see for example: 1. Su, Peicheng 2001a:372-386; 2. Article by Kang, Jiashen 1993, in Su,

Peicheng [ed.] 2001b:126-140; 3. Article by Li, Yan and Kang, Jiashen 1993, in Su, Peicheng [ed.] 2001b:141-154; and 4. Menzel, Cornelia 2000:49-60.)

For this reason Chinese leaders hope to complement the phonetic information lacking in Chinese characters by means of Pīnyīn, in order to create and gain acceptance for a uniform pronunciation of the modern Chinese language (普通话 Pǔtōnghuà). When you consult contemporary dictionaries, you find a Pīnyīn spelling behind every character, which is why people think that Chinese characters are the main script and Pīnyīn a transcription for it. The legislators have adjusted to this view. They have thus invalidated the pronunciation conveyed by Chinese characters and denied that they are a full-fledged script. Nowhere in the world is there a script that needs a transcription to function. Transcription systems are used only for spoken language, not for a script. What we find in dictionaries are lemmas (lexical entries). Every lexical item is given in two writing systems—in principle of equal value—that represent a unit of the spoken language.

Put another way: if we want to transcribe Chinese characters with $P\bar{\imath}ny\bar{\imath}n$, we are assuming that they are mute. But there is no fully developed mute script in the world. Is the Chinese script after 3,500 years of development still an early form of writing or perhaps a picture puzzle? How can the only surviving script from the Bronze Age, which is the pride of most Chinese people, be put in this category?

By way of preliminary conclusion, one remark is in order: the legislators did not clarify what $H any \tilde{u} P \bar{v} ny \bar{v} n F ang' an$ is, and thus failed to solve the problem. A new development, however, is that a law has for the first time disqualified Chinese characters as a script.

2. ISO Standard 7098

An international standard was published in 1982: ISO Standard 7098. The only valid edition is the second one of 1991, but in China a translation of the 1982 text is still used, albeit with the title and date of the second edition (cf. Guojia Yuwei Biaozhunhua Gongzuo Weiyuanhui Bangongshi 1999:498). Many people claim that Hanyu Pinyin Fang'an is the same as this ISO standard. This shows that they have not read the text carefully (or at all). The ISO standard deals with three issues:

2.1. A "romanization of the Chinese language"

Paragraph 1, entitled "Scope," reads:

This International Standard explains the principles of romanization of Modern Chinese or putonghuà, the official language of the People's Republic of China as defined in the Directives for the promotion of putonghuà, promulgated in 1956-02-06 by the State Council.

Oddly enough, this paragraph does not speak of a script but a "Westernization", or, to be more precise, a romanization of the Chinese language, Pŭtōnghuà.

That is not what the drafters meant, but the formulation shows that they did not know what their task was. From paragraph 2 onward, they no longer speak of *Pŭtōnghuà*, but rather about converting one writing system into another. How is this accomplished? ISO lists three methods: transliteration, transcription, and a combination of both methods. See paragraph 2.5:

To carry out romanization (the conversion of non-Latin writing systems to the Latin alphabet) it is possible to use either transliteration or transcription or a combination of these two methods, according to the nature of the converted system.

One has to wonder what such a combination might look like. No wonder that the text contains no description of such a mixed method. All that is left are transliteration and transcription.

2. 2. Transliteration

By transliteration is meant the letter-by-letter conversion of a text written in an alphabet into another alphabetical script, if necessary using diacritical marks, in such a way that the text can be correctly converted back into the original text by means of a transliteration table. Even GB drafters sometimes misunderstand the term "transliteration." For example, in the 中文书刊名称汉语拼音拼写法 Transliterating rules of Chinese phonetic Alphabet on titles for books and periodicals in Chinese (Guojia Jishu Jianduju 1992), the writing of book and periodical titles is regarded as a transliteration.

In the case of Chinese characters, ISO has established that a transliteration between Chinese characters and $P\bar{\imath}ny\bar{\imath}n$ is impossible: the supposedly more than 40,000 ("ideophonographic") characters cannot be represented by the 26 letters of the Latin alphabet. There is no doubt about that. This clearly shows that $H\grave{a}ny\check{u}$ $P\bar{\imath}ny\bar{\imath}n$ $F\bar{a}ng'\grave{a}n$ is not a transliteration system, because it does not fulfill all the criteria of a transliteration system.

Here I would like to add a remark: The ISO identified Chinese characters as "ideo-

phonographic" characters (known in Chinese as 形声字 xingshēngzì). The Chinese title of this standard is: 中文的罗马化 Zhōngwén de Luómǎhuà (Romanization of Chinese). Yet not one of these six characters is "ideophonographic." The ISO disregarded the five other traditional categories of 六书 liùshū (the six ways in which Chinese characters are formed), and the figure of more than 40,000 Chinese characters is flat wrong (cf. Hincha, Xieyan 2003:117-129). But the ISO found an apparent way out to somehow convert characters into Latin letters: transcription. Paragraph 3.1 reads:

The structure of ideophonographic scripts, where conveyance of meaning is of greater importance than that of pronunciation, entails the existence of a large number of characters (more than 40 000 in the case of Chinese), thus making sign by sign transliteration impossible and resulting in the need to devise a system of transcription.

2. 3. Transcription

The central point of the ISO standard is the desire to employ Hànyũ Pīnyīn Fāng'àn to transcribe Chinese characters:

Each character must therefore be transcribed by one or more Latin letters standing for the pronunciation or pronunciations of the character in question. This means that the transcriber must be familiar with the reading or readings of the text to be transcribed (§ 3.1).

Therefore, the Latin letters of $H any \bar{u} P \bar{u} ny \bar{v} n F \bar{u} ng' an$ are used to transcribe the character pronunciation(s). The same mistake found in the Chinese law is repeated here. This time Chinese characters are disqualified at an international level as a script.

The claim that $P\bar{i}ny\bar{i}n$ can transcribe Chinese characters is based on a popular view that says that characters represent meaning while an alphabetic script represents phonemes. But in fact both Chinese characters and a phoneme script represent pronunciation and meaning (cf. Zhao, Yuanren 1999:145-149). A Chinese character does not have all that many disadvantages in representing pronunciation, nor that many advantages in representing meaning. A character represents a morpheme; consequently it also represents the pronunciation (cf. Zhao, Yuanren 1999:229).

A page farther down, paragraph 4 sets up a contradiction:

Hànyũ pĩnyĩn făng'àn (Chinese phonetic system) or pĩnyĩn, which was officially

adopted on 1958-02-11 by the National Assembly of the People's Republic of China, is used to transcribe Chinese. The transcriber writes down the pronunciation of the characters according to their readings in pătōnghuà.

The first sentence talks about the transcription of the Chinese language, while the second goes back to the pronunciation of characters.

It is time to ask what exactly is a transcription system. It is a graphic system whose elements unambiguously represent the sounds of a spoken language. The transcription can be narrow or broad: in both cases one graphic symbol represents in principle precisely **one** single sound. If necessary, digraphs are used: $\langle ng \rangle$, $\langle sh \rangle$ and suchlike. We need to examine whether $H any \bar{u} P \bar{u} n y \bar{u} n F \bar{u} n g' an really is a transcription system.$

3. Is Hànyũ Pīnyīn Fāng'àn a phonetic transcription system for the Chinese language?

If we consult the birth certificate of Hànyũ Pĩnyĩn Fāng'àn, we find rules that are common knowledge.

Example A in tabular form:

Syllables without consonantal initial sound	Orthographic rules	Spelling
i	y + i	yi
ia	$i \rightarrow y + a$	ya
ie	$i \rightarrow y + e$	ye
iao	$i \rightarrow y + ao$	yao
iou	$i \rightarrow y + ou$	you
ian	$i \rightarrow y + an$	yan
in	y + in	yin
iang	$i \rightarrow y + ang$	yang
ing	y + ing	ying
iong	$i \rightarrow y + ong$	yong

If Hànyũ Pīnyīn Fāng'àn were a transcription system, this table would contain three state-prescribed violations of the transcription principle, namely: y+i, y+in, and y+ing. In all three of these cases, two letters represent one sound. The same is true when writing y+u and w+u. This rule does not concern phonetic transcription; rather, it is an orthographic

rule: in these cases $\langle y \rangle$ and $\langle w \rangle$ are artificial and arbitrary initial symbols. But phonetically these are not consonants. Consequently, in this respect $H any \tilde{u} P ny \tilde{u} n$ F ang 'an is not a transcription system.

Example B:

Instead of [jiou] one has to write jiu. Strangely enough, the main vowel [o] is omitted, although it is the more important vowel, which also carries the tone. What has happened to phonetics? Similarly, instead of [shuei] one writes shui. This is another instance of an orthographic rule rather than phonetic writing.

Example C:

For every letter of the alphabet, there is an upper-case letter (capital letter) that has no additional phonetic function. The use of upper-case letters is regulated in the national $P\bar{\imath}ny\bar{\imath}n$ orthography standard (Guojia Jishu Jianduju 1996). This standard also regulates word boundaries, punctuation marks, and other matters. None of these are phonetic units.

Example D:

The sound [ü] is only written with the letter $\langle u \rangle$ after $\langle n \rangle$ and $\langle l \rangle$. After the letters $\langle j \rangle$, $\langle q \rangle$, $\langle x \rangle$, and $\langle y \rangle$ the diaeresis is deleted. This is not a *phonetic* rule either, but rather an *orthographic* one.

4. Is Hànyŭ Pīnyīn a script?

Next I would like to add a few facts to the preceding four arguments that $H any \bar{u} P \bar{u} ny \bar{u}$

- 1. In 1976, the State Bureau of Surveying and Mapping and the Committee for Script Reform issued a joint directive entitled 少数民族语地名汉语拼音字母音译转写法 Shǎoshùr Minzúyǔ Dìmíngr Hànyǔ Pīnyīn Zìmǔr Yīnyì Zhuánxiéfǔ (loosely translated: phonetic transliteration between letters in minority-language place names and Hanyu Pinyin letters). This standard contains three transliteration tables that stipulate how Uighur, Mongolian, and Tibetan letters are to be transliterated into Hànyǔ Pīnyīn letters (Guojia Cehui Zongju / Wenzi Gaige Weiyuanhui 1976). Since a transliteration between writing systems takes place, Hànyǔ Pīnyīn can only be a script (i.e., a writing system).
 - 2. In April 1958, the Chinese Ministry of Education, the Ministry of the Interior, the

Committee for Script Reform, and the Chinese Association of the Blind and the Deaf jointly founded a Braille Committee. Shortly thereafter, the committee made two drafts based on Hànyǔ Pīnyīn Fāng'àn. The first is 汉语拼音盲字方案 Hànyǔ Pīnyīn Mángzìr Fāng'àn (loosely translated: draft braille script based on Hanyu Pinyin). The other is 词 儿连写规则 Cir Liánxiĕ Guīzé (rules for writing as one word) (cf. Teng, Weimin, Li, Weihong 1997:44). After a long period of hesitation, in 1995 the national standard 中国 盲文 Chinese Braille was published (Guojia Jishu Jianduju 1995). This braille system is in principle a conversion of Hànyǔ Pīnyīn into braille letters, that is to say a transliteration. If braille is a script, which no one doubts, especially not blind people, then Hànyǔ Pīnyīn must also be a script.

Incidentally, the Law on Language and Script contains not one word about braille. The legislators simply overlooked blind people. It is also very rare to find a reference to braille in Chinese specialist graphological literature (文字学 wénzìxué). Yet we should not forget that braille is also a script for the Chinese language. Blind people are also Chinese!

As is well known, the Chinese leadership refuses to recognize Hànyũ Pīnyīn as a script and to permit digraphia. But scientific facts demonstrate that Hànyũ Pīnyīn Fāng'àn, including its orthography, is a writing system for Chinese. As far as I am concerned, the answer to the question posed above is a categorical yes.

Is the problem of homophony unsolvable?

Even those who agree with my arguments and conclusions, doubt that it would be possible to introduce $P\bar{\imath}ny\bar{\imath}n$ as a commonly used standard script and thus establish digraphia in China. The main reason for their doubts has long been known, and it is a serious one: an extremely large number of syllables in Chinese are homophonous and would, so the doubters' argument goes, be indistinguishable in the Latin alphabet. Put in linguistic terms: phonographic scripts are either phonemic scripts or syllabic scripts (syllabaries). $P\bar{\imath}ny\bar{\imath}n$ is a phonemic script. In principle it can represent all Chinese phonemes, but it is not capable of differentiating between homophonous syllables. Even a syllabic script like the Japanese kana would be incapable of this. If Chinese characters had not existed for millennia—argue the doubters—they would have to be invented. If this argument were true, $P\bar{\imath}ny\bar{\imath}n$ ought to be resolutely given up once and for all. According to the doubters' argument, $P\bar{\imath}ny\bar{\imath}n$ is an almost perfect script, but unusable for Chinese.

This claim is quite possibly utterly wrong, because it is based on an incorrect analysis of unsuitable data. If I were able to scientifically prove this with facts, $P\bar{\imath}ny\bar{\imath}n$ would clearly have to be accepted as a script.

In the specialist literature one often finds linguists picking a syllable and counting how many characters are connected to it. A tried-and-tested syllable is yi. The Xian-Han lists no less than 91 lexical items (lemmas) under yi. Based on this, linguists reach the categorical conclusion that the Latin alphabet is unsuited to differentiate between 91 "words" with the yi pronunciation. As far as many people are concerned, the case is closed. What they fail to take into consideration is that even a dictionary such as Xian-Han, which supposedly represents the modern language, also contains—for sound reasons—many Chinese characters from the classical literary language. Many lexical items are not words, but rather bound morphemes. Others are yiti characters (variant forms of a Chinese character, also known as allographs). Moreover, modern dictionaries do not necessarily document language as it is spoken and written today. As is well known, all dictionaries are out-of-date the day they are published.

There is a tool that makes the search for homophonous words easier. The Hanyu Pinyin Cihui (1989, abbreviated Cihui) lists more than 60,400 of the most important words, based on texts that were printed in Chinese characters. These Chinese characters were then converted into Pinyin in accordance with the ziyin (character pronunciation), i.e., according to the established pronunciation of isolated characters. This was meant to establish the way Pinyin is written. I have looked for homophones in this Cihui and counted 2,874 groups of two to eight monosyllabic and polysyllabic homophonous lemmas. I found a total of 6,435 lemmas that are not unambiguous. This is an alarming result: 10.65% of 60,400 written items are not identifiable. This would disqualify Pinyin as a commonly used standard script.

But this verdict is rash and unfair. I found the following examples in the Cihui:

- 1. a. bùjiàn 不见
 - b. bùjiàn 部件

This entry is incorrect. According to the tone sandhi rules, 1.a. is pronounced bújiàn. Thus these two lemmas are not phonetically identical.

- 2. a. bǎn 板
 - b. băn 版

In the spoken language, 2.a. is only pronounced banr.

- 3. a. jiàoshì 教士
 - b. jiàoshì 教室
- 3.a. jiàoshì exists only in dictionaries. In the spoken language the word for "priest" is jiàoshi.
- 3.b. Nor is *jiàoshì*, meaning "classroom," rendered entirely correctly: in everyday language it is often pronounced *jiàoshī*. A comparable word is *bàngōngshī* 办公室 (office) (cf. Kupfer, Peter 2003:93).
 - 4. a. báihuà 白桦
 - b. báihuà 白话
- 4. a. The word for "white birch" is báihuá. huà is in fact only the pronunciation of the single character 桦 in dictionaries.
 - 4. b. Three words are hidden behind the characters 白话:

báihuà: empty talk;

báihuàr²: simple, clear formulation, e.g.: shuō dà báihuàr;

báihuà: Baihua (the spoken vernacular as opposed to the classical written language Wenyan).

This word occurs only in combinations, e.g.: báihuàwénr 白话文儿, báihuà xiǎoshuōr 白话小说儿.

- 5. a. bǐjià 比价
 - b. bǐjià 笔架
- 5.a. Two meanings are concealed by the characters 比价: bǐjià meaning "rate of exchange" and bǐ jià meaning "to compare prices." It would be worth considering whether a blank (space character) ought to be inserted: bǐ jià. This would remove this "word" from the word list.
 - 5.b. The word 笔架, meaning "pen rack" or "brush holder," is pronounced bǐjiàr.
 - 6.a. bănhuà 板话
 - b. bănhuà 版画

² Not listed in the Xiandai Hanyu Cidian.

6.a. I do not know what the "word" bănhuà 板话 means, nor can I find it in dictionaries such as Xiàn-Hàn and Hànyŭ Dà Cidiăn (Xia, Zhengnong et al. 1997).³

6.b.: 版画 meaning "graphic reproduction" is only pronounced bănhuàr.

These six pairs are just examples that could easily be multiplied. What is going on here? Chinese characters have been converted into *Pīnyīn* according to syllables. But the following factors have been disregarded:

- Tone sandhi;
- Rhotacization (r-coloring; érhuà);
- Neutral or fifth tone (qīngshēngr);
- Tonal variations in the spoken language;
- Semantic differences, including word-class differences;
- Word boundaries.

In addition there are artificial accidental errors due to sloppy work.

Pseudo-homophonous written forms were created in this way. They must of course be eliminated. But that is not enough. Our goal cannot be to reduce the proportion of homophones from 10.65% to some smaller number. Even if one were painstakingly to arrive at a lower percentage, it would not be of much use, because no one can determine the percentage of homophones that would be acceptable for the Chinese language written in the $P\bar{\imath}ny\bar{\imath}n$ script. I therefore had to break off my work with the Cihui and to reorient my thinking.

The eminent linguist Zhao Yuanren pointed the way. He once wrote—I am giving the general gist of his statement—that there is no country and no place where people speak a language whose purpose it is not to be understood. Although compared to other Chinese dialects the Beijing language has the fewest syllables, homophony poses no obstacles to communication in it. This indicates that there is a self-healing process in the language (cf. Zhao, Yuanren 2002:42-43). Zhao added that a phonographic script such as Guóyũ Luómãzì can therefore be employed for the Chinese language. There appear to be technical problems, but the difficulties reside not in the mechanics of the script but rather in the psychology of society, including that of individuals and the government (cf. Zhao, Yuanren 1999:151).

³ Translator's note: The ABC Chinese-English Comprehensive Dictionary defines bǎnhuà 板话 as "rhythmic comic talk."

I take a step further in this direction: When you converse with illiterates or, to be more precise, with people who have no knowledge of written language (the script-blind, as they are known in Chinese), you have to conclude that there is no misunderstanding, no asking what a word meant, and no breakdown in communication because of homophonous Chinese words. Consequently, there is ample reason for hoping that the use of an effective phonographic script for Chinese will succeed. I find Zhao's reference to a self-healing process in the language to be very helpful. He cited four pieces of circumstantial evidence (Zhao, Yuanren 2002:42), which I do not need to repeat here. To bolster Zhao's claim, I simply add my own observations, which I have not seen in the linguistic literature to which I have had access.

Example 1: 方便

1. fāngbiàn qúnzhòng de shēnghuó 方便群众的生活 (Alleviating the life of the masses)

In connection with the word shēnghuó 生活, fāngbiàn 方便 is a verb.

- 2. shàng wăng hĕn fāngbian 上网很方便 (going online/getting on the Internet is easy)
 In this case fāngbian 方便 is an adjective.
 - 3. bă fāngbian rànggei biéren 把方便让给别人 (letting someone else enjoy the comfort)

Here fangbian 方便 is used as a noun.

4. fāngbian yixiàr 方便一下儿 (to go to the toilet)

fāngbian 方便 is a verb in this case.

These examples shown in tabular form:

Homophones	Parts of speech
2. fāngbian	adjective
3. fāngbian	noun
4. fāngbian	verb
1. fāngbiàn	verb

These examples show instances of a general rule:

Phonetically identical forms belong to different parts of speech. Within one part of speech (verb in this case), they are *not* phonetically identical.

Example 2: 被,背,辈,倍

1. shū bèi sī pò le 书被撕破了 (the book was torn)

bèi 被 is a function word in this case.

2. bèi kèwénr 背课文儿 (learning the text off by heart)

bèi 背 is a verb.

3. ěrduo bèi 耳朵背 (to be hard of hearing)

bèi 背 is an adjective.

4. yǐzi bèir 椅子背儿 (chair-back)

bèir 背儿 is a noun.

5. bèi téng 背疼 (the back aches)

bèi 背 is also a noun in this case.

6. yí bèir 一辈儿 (a generation)

bèir 辈儿 is a measure word.

7. (yí) bèi (一)倍 (single/(one) time)

· bèi 倍 is also a measure word.

In tabular form:

Homophones	Parts of speech	Chinese	Homophones	Parts of speech	Chinese
		characters			characters
1. bèi	function word	被	4. bèir	noun	背儿
2. bèi	verb	背	6. bèir	measure word	辈儿
3. bèi	adjective	背			
5. <i>bèi</i>	noun	背			
7. bèi	measure word	倍			

Non-homophones	Parts of speech	Chinese characters
4. bèir	noun	背儿
5. bèi	noun	背
6. bèir	measure word	辈儿
7. bèi	measure word	倍

The same result:

The homophones do not belong to a single part of speech, and within the same part of speech the forms are not homophonous.

Example 3: 办, 扮, 拌, 半, 瓣, 伴, 绊

- 1. bàn shìr 办事儿 (dealing with a task/getting something done): verb
- 2. bàn zhuōr 办桌儿 (organizing a banquet): verb
- 3. bàn zhàor 办照儿 (having a permit drawn up): verb
- 4. bàn guiliănr 扮鬼脸儿 (making faces; grimace): verb
- 5. bàn xiànr 拌馅儿 (mixing and stirring a filling): verb
- 6. bàn xiǎoshi 半小时 (a half hour): numeral
- 7. yi bànr suàn 一瓣儿蒜 (clove of garlic): measure word
- 8. jié bànr 结伴儿 (keeping company; "going with"): noun
- 9. xià bànr 下绊儿 (setting up an obstacle to get someone to stumble/bring about sb.'s downfall): noun

Homophones	Chinese characters	Parts of speech
1. bàn + shìr	办事儿	verb
2. $ban + zhu\bar{o}r$	办桌儿	verb
3. bàn + zhàor	办照儿	verb
4. bàn + guiliănr	扮鬼脸儿	verb
5. bàn + xiànr	拌馅儿	verb
6. bàn (xiǎoshí)	半(小时)	numeral

Homophones	Chinese characters	Parts of speech
7. bànr	瓣儿	measure word
8. jié + bànr	结伴儿	noun
9. xià + bànr	下绊儿	noun

In examples 1 to 5, the forms "bàn" are homophonous, and in 8 and 9 "bànr" are also homophonous, but their differences are defined by the additional component. Number 6 belongs to another word class. 7, 8, and 9 are homophonous, but 7 belongs to a different word class. Thus, in principle we get the same result.

To summarize:

Homophony is avoided, but with the most economical means. Economy is not a foreign word in linguistics. How the Chinese language deals with homophony is a question that ought to be investigated more thoroughly. Many orthographic problems would quite possibly disappear automatically if this were done.

Next I discuss the question of demarcating⁴ words and word groups.

Too much emphasis has been placed on homophony. It is therefore not surprising that other important matters have been overlooked.

In the spoken language there are words that cannot be differentiated by means of pronunciation, word class, and characters. But in principle, they could be differentiated in writing by means of $P\bar{\imath}ny\bar{\imath}n$. Next I give a few examples.

Example 1: 拼音

According to the explanation found in Xiàn-Hàn, which I cited above, 拼音 is not one word but two: 拼 $p\bar{\imath}n$ (V) and 音 $y\bar{\imath}n$ (N). The dictionary is thus wrong to go against its own definition and write " $p\bar{\imath}ny\bar{\imath}n$ " instead of " $p\bar{\imath}n$ $y\bar{\imath}n$."

There is a second meaning hidden behind the term $p\bar{\imath}ny\bar{\imath}n$ 拼音. This meaning becomes apparent when one combines $p\bar{\imath}ny\bar{\imath}n$ 拼音 and wénzì 文字. $p\bar{\imath}ny\bar{\imath}n$ 拼音 means "phonographic" and is a single word, an adjective to be precise.

There is also a third meaning. When we hear "xué pīnyīn" and "xiẽ pīnyīn" we know what this means: learning and writing Chinese with letters according to the rules of Hànyũ

⁴ Translator's note: Abgrenzung

Pīnyīn Fang'an. In this case, "pīnyīn" is an abbreviation for Hànyŭ Pīnyīn or Hànyŭ Pīnyīn Fāng'àn (cf. Zhou, Youguang 1992:227). Since in this case "pīnyīn" is used as an abbreviation of the technical term Hànyũ Pīnyīn Fāng'àn, it ought to be capitalized. The following cases could be differentiated:

Syllables	Pinyin script	Chinese characters
pīnyīn	pīn yīn (V+N)	拼音
pīnyīn	pīnyīn (Adj)	拼音
pīnyīn	Pīnyīn (N)	拼音

This example shows that $P\bar{\imath}ny\bar{\imath}n$ can do something which Chinese characters cannot do: differentiating homophonous words.

Example 2: 拼音不是拼音文字

The sentence 拼音不是拼音文字 has three meanings:

First: Pīnyīn is no script. The noun wénzì 文字 is being negated.

In this case the word pīnyīn 拼音 ought to have been omitted, because the adjective is irrelevant to the negation. This sentence ought to read: Pīnyīn bú shì wénzì 拼音不是文字.

Second, the attribute $p\bar{\imath}ny\bar{\imath}n$ 拼音 can be negated, which results in the following meaning: $P\bar{\imath}ny\bar{\imath}n$ is a script, but not a phoneme-, sound-, letter-, phonetic, or alphabetic script. To what type of a script $P\bar{\imath}ny\bar{\imath}n$ belongs is a question that is left unanswered. The Chinese leadership has classified $P\bar{\imath}ny\bar{\imath}n$ as an "auxiliary instrument" (see above), whatever that may mean.

Third: One can negate "pīnyīn wénzì 拼音文字" and say: Pīnyīn is not a phonographic script, that is to say, it is not a writing system for the phonographic representation of Chinese. Instead, it is something altogether different, such as a "transcription system for Chinese characters."

In the second and third case it is not possible to make a differentiation by means of Chinese characters. In the third case $P\bar{\imath}ny\bar{\imath}n$ could have inserted a hyphen between $p\bar{\imath}ny\bar{\imath}n$ and

wénzì, i.e., pīnyīn-wénzì. According to the current spelling rules, this is unfortunately not possible. But the hyphen would have clarified the meaning.

Example 3: 中华人民共和国国家通用语言文字法

国家通用语言文字法 Guójiā Tōngyòng Yũyán Wénzìfã is ambiguous. Since the Chinese leadership did not dare introduce word boundaries in the Chinese-character script, this formulation is open to at least the following interpretations:

- a. a written body of rules for the language, applicable throughout the country;
- b. law on language and script, applicable throughout the country;
- c. law on the common national language and script.

Only from the content of the law can one conclude that meaning c. is the one that is meant. A sensible orthography would have made this meaning immediately clear: Guójiā Tōng-yòng-Yŭyán/Wénzì Fă.

Example 4: 一个中国

On the international stage, $-\uparrow + \equiv yi \ ge \ Zh\bar{o}nggu\acute{o}$ is known as "ein China," "one China," "odin Kitaj," and so forth. No one knows the identity of the person who was so intelligent or cryptic as to coin this expression, which causes much confusion and unnecessary conflict. According to several dictionaries, the word $Zh\bar{o}nggu\acute{o} + \equiv already$ appears in the $Zh\bar{o}ngy\bar{o}ng$, part of the Liji, compiled more than 2000 years ago, with the meaning "Middle Kingdom" (see, for example, Ciyuan 1981). In the centuries that followed, it was used with varying meanings. Since the middle of the nineteenth century, $Zh\bar{o}nggu\acute{o} + \equiv has$ been the established name of China's territory (see, for example, Cihai 1999). Chinese people have grown accustomed to calling their country by this name (cf. He, Rong 1974 and Wang, Yunwu 1968). It is difficult to determine when an additional, largely ignored, meaning cropped up in mainland Chinese dictionaries:

"中华人民共和国的简称" (abbreviation of Zhōnghuá Rénmín Gònghéguó, People's Republic of China). For example, one finds this definition in the latest edition of Cíhǎi and in the Xiàn-Hàn, but not in the Hànyǔ Dà Cídiǎn and Hànyǔ Cídiǎn (1957). At any rate, in mainland China two terms have arisen which are homophonous and homographic. Nowadays, the PRC government insists that there is "yí ge Zhōngguó" and demands that

the representatives of Taiwan and the rest of the world's governments accept this usage with the stress on the numeral. It is generally thought that this is a fairly transparent term that is meant to counter certain separatist tendencies in Taiwan. In fact, this interpretation just scratches the surface. For it remains open whether the old or the new meaning is meant, or whether those who employ the term are alternately thinking of one or the other meaning. Such ambiguities could easily be avoided with the *Pinyin* script. For example:

Zhōngguó stands for "China";

Zhōng Guó or Zhōng-Guó is used as an abbreviation of Zhōnghuá Rénmín Gònghéguó, just as Yin, Binyong et al. have suggested Tái Méng 台盟 or Tái-Méng as abbreviations of Táiwān Mínzhǔ Zìzhì Tóngméng 台湾民主自治同盟 (Taiwan Democratic Self-Government League) (see Shangwu Yinshuguan Cishu Yanjiu Zhongxin 2002:127).

Example 5: 文化大革命

What is 文化大革命? It is China's world-famous "Cultural Revolution" (1966-1977). Many people have probably asked themselves what "Cultural Revolution" means. Only recently, the host of a cultural-affairs television show in Germany said, "Cultural Revolution? It's difficult to understand what is meant by this. It sounds harmless enough, but the reality was brutal." It is not just people in the West who grapple with the incomprehensibility of the term "Cultural Revolution"; so do many Chinese people.

Before the word wénhuà 文化, meaning "culture," was imported from Japan to China, there was already a concept of wénhuà 文化. It is as old as the word Zhōngguó 中国. Four words are concealed behind 文化, namely 文 wén (civil); 治 zhì (administration); 教 jiào (to educate); and 化 huà (to persuade). 文 wén is an abbreviation of 文治 wén zhì (civil administration); 化 huà is an abbreviation of 教化 jiào huà (to persuade through education). Thus 文化 is a short way of writing "civil administration and persuasion through education." This refers to the peaceful, Confucian-based moral education of the people (cf. Gernet, Jacques 1983:85, 295).

The current word "culture," which is also written 文化, has a different origin. To translate the Western concept of culture, the Japanese coined the word *bunka*, which is written 文化 (see Liu, Zhengtan et al. 1984, s.v. wenhua). The Chinese imported this character combination from Japan and pronounced it according to the rules of their own

language: wénhuà. In this way, the modern term has been superimposed on the Chinese "civil administration and persuasion through education." The original Chinese meaning is largely unknown nowadays.

In today's Chinese, 文化 is understood to mean "knowledge," "education," and "culture," including learning Chinese characters. For example: xué wénhuà 学文化 (acquiring knowledge), yŏu wénhuà 有文化 (well educated; cultured), or méi wénhuà 没文化 (uneducated; uncultured) (On this issue, see for example: He, Jiuying 1999:31—42).

The story of gé mìng or gémìng 革命 is similar to that of 文化.

- a. traditionally gé mìng 革命 meant simply change of dynasty. gé 革 means weapons of war (Gao, Shufan 1989:2031); mìng 命 means "Mandate of Heaven." This meant the violent overthrow and replacement of the "Son of Heaven" (the emperor), who had lost heaven's mandate.
- b. Since ming 命 also means "life" (e.g., yì tiáo rénming 一条人命: a human life), a popular meaning evolved: "taking or destroying a life." gé ming 革命 is semantically closely related to shā tóu 杀头 "to behead" or "decapitate."
- c. A third meaning of géming 革命 is "revolution," borrowed from the Japanese ka-kumei 革命 (see Liu, Zhengtan et al., 1984, s.v. geming). The Xiàn-Hàn tries to explain this word in Marxist terms. Ordinary Chinese people associate it with the use of violence.

Therefore, the character combination 文化大革命 has at least two meanings:

- a. a radical transformation of the moral education of the people, such as the replacement of traditional school education by education through manual labor;
- b. the violent overthrow of the governing elite, understood as the dynasty, and its replacement by a new leadership (Chairman Mao Zedong 毛泽东 [1893-1976] instead of head of state Liu Shaoqi 刘少奇 [1898-1969]).

In addition, there is the mechanical word-for-word translation "Cultural Revolution," which has understandably remained incomprehensible to the Western world.

⁵ Translator's note: Bildung can mean education or culture.

Neither Chinese characters nor the pronunciation can differentiate between the three meanings. The $P\bar{\imath}ny\bar{\imath}n$ script, on the other hand, makes it possible. The following arrangement could be agreed upon:

First: wén huà is an abbreviation of wén zhì jiào huà (see above);

wénhuà is the modern word "culture," comprising "knowledge" and "education."

Second: gé ming stands for "violent overthrow of something";

géming is the modern word "revolution."

What people in the West call the "Cultural Revolution" was wénhuà gé mìng to the Chinese people. But to the man who coined the term, Mao Zedong, who was very knowledgeable about the Chinese classical tradition, it quite possibly meant wén huà gé mìng.

Much more thorough investigation is required before we can formulate laws governing examples such as the ones I have cited. The employment of a good phonography in a commonly used standard script can help clarify ambiguities that are obscured by Chinese characters. $P\bar{\imath}ny\bar{\imath}n$ is the more precise writing system. This is an additional reason for introducing the $P\bar{\imath}ny\bar{\imath}n$ script, initially in the form of digraphia.

List of references

Cihai 1999. Cihai. Shanghai: Shanghai Cishu Chubanshe.

Ciyuan 1981. Ciyuan. Beijing: Shangwu Yinshuguan.

ISO/TC46/SC2 1991. Information and documentation - Romanization of Chinese (ISO 7098:1991 (E)).

Gao, Shufan 1989. Zhongwen Xing-Yin-Yi Zonghe Da Zidian. Beijing: Zhonghua Shuju Gernet, Jacques 1983. Die chinesische Welt. Second German edition. Frankfurt am Main: Insel Verlag

Guojia Cehui Zongju / Wenzi Gaige Weiyuanhui 1976. Shaoshur Minzuyu Dimingr Hanyu Pinyin Zimur Yinyi Zhuanxiefa.

Guojia Yuwei Biaozhunhua Gongzuo Weiyuanhui Bangongshi 1999. Guojia Yuyan Wenzi Guifan he Biaozhun Xuanbian. Beijing: Zhongguo Biaozhun Chubanshe.

Guojia Jishu Jianduju 1996. Hanyu Pinyin Zhengcifa Jiben Guize Basic rules for Hanyu Pinyin Orthography (GB/T 16159-1996).

Guojia Jishu Jianduju 1992. Zhongwen Shukan Mingcheng Hanyu Pinyin Pinxiefa

Transliterating rules of Chinese phonetic Alphabet on titles for books and periodicals in Chinese (GB 3259-92).

Guojia Jishu Jianduju 1995. Zhongguo Mangwen Chinese braille (GB/T 15720-1995).

Hanyu Cidian 1957, reprint of Guoyu Cidian 1947. *Hanyu Cidian*. Beijing: Shangwu Yinshuguan.

"Hanyu Pinyin Cihui" Bianxiezu 1991. Hanyu Pinyin Cihui. Beijing: Yuwen Chubanshe.

He, Jiuying 1999. Hanzi Wenhuaxue. Shenyang: Liaoning Renmin Chubanshe

He, Rong 1974. Guoyu Ribao Cidian. Taipei: Guoyu Ribao She.

Hincha, Xieyan 2003. Die chinesische Graphemik. In CHUN 18:117-129.

Kang, Jiashen 1993. Xiandai Hanyu Xingshengzi Xingfu Yanjiu. In Su, Peicheng (ed.), 2001b:126-140

Kupfer, Peter 2003. Fragen der Grammatikalität im neueren Sprachgebrauch. In CHUN 18:93

Li, Yan / Kang, Jiashen 1993. Xiandai Hanyu Xingshengzi Shengfu Yanjiu. In Su, Peicheng (ed.), 2001b:141-154.

Liu, Zhengtan et al., 1984. *Hanyu Wailaici Cidian*. Shanghai: Shanghai Cishu Chubanshe.

Menzel, Cornelia 2000. Zur Phonetizität der chinesischen Schrift. In CHUN 16:49-60

Shangwu Yinshuguan Cishu Yanjiu Zhongxin 2002. Xinhua Pinxie Cidian. Beijing: Shangwu Yinshuguan.

Su, Peicheng 2001a. Ershi Shiji de Xiandai Hanzi Yanjiu. Taiyuan: Shuhai Chubanshe.

Su, Peicheng (ed.) 2001b. Xiandai Hanzixue Cankao Ziliao. Beijing: Beijing Daxue Chubanshe.

Teng, Weimin / Li, Weihong u. a. 1997. Zhongguo Mangwen. Beijing: Huaxia Chubanshe.

Xia, Zhengnong u. a. 1997. Hanyu Da Cidian. Shanghai: Hanyu Da Cidian Chubanshe.

Xiandai Hanyu Cidian 1996. Xiandai Hanyu Cidian. Beijing: Shangwu Yinshuguan.

Wang, Yunwu 1968. Wang Yunwu Da Cidian. Hong Kong: Huitong Shudian.

Zhao, Yuanren 2002. Zhao Yuanren Yuyanxue Lunwenji. Beijing: Shangwu Yinshuguan.

Zhao, Yuanren 1999. Yuyan Wenti. Beijing: Shangwu Yinshuguan.

Zhonghua Renmin Gongheguo Guojia Tongyong Yuyan Wenzifa 2001. Beijing: Zhongguo Fazhi Chubanshe.

Zhou, Youguang 1999. Xin Shidai de Xin Yuwen. Beijing: Sanlian Shudian.

Zhou, Youguang 1992. Xin Yuwen de Jianshe. Beijing: Yuwen Chubanshe.

Since June 2006, all new issues of *Sino-Platonic Papers* have been published electronically on the Web and are accessible to readers at no charge. Back issues are also being released periodically in e-editions, also free. For a complete catalog of *Sino-Platonic Papers*, with links to free issues, visit the *SPP* Web site.

www.sino-platonic.org