

# Chapter 14

## Exploring the unique method for encoding sinograms in the first known Chinese–Polish dictionary

Andrzej Swoboda

Jagiellonian University in Krakow


The aim of this chapter is to examine Doman Wieluch's Chinese character encoding method, introduced in his *Słownik chińsko-polski do czytania tekstów chińskich bez przygotowania* (Wieluch 1936) – the oldest (known) Chinese-Polish dictionary in history. In the first part of the chapter, the reference work in question is briefly described and discussed in the context of selected earlier lexicographic and linguistic publications. The macrostructure and microstructure of Wieluch's dictionary are outlined and compared to those of selected other monolingual and bilingual dictionaries from various periods. The second part of the chapter consists in a critical analysis of the author's encoding method: its general premise, specific rules (for isolating character components, replacing them with code words and arranging code words in sequences), selected examples of encoded headwords, as well as the peculiarities and limitations of the system. In the final part, Wieluch's ideas are discussed in the context of selected formerly and currently utilised methods for ordering, decomposing, encoding and inputting Chinese characters. It is hoped that this chapter will shed new light on this relatively unknown episode in the history of Chinese-Polish lexicography, which may be of interest to future dictionary authors.

### 1 Wieluch's Chinese-Polish dictionary

#### 1.1 Title(s) and author(s)

A relatively small number of Polish-Chinese and Chinese-Polish dictionaries have been published so far (cf. Sroka-Grądział 2018: 32). The earliest (surviving)



Andrzej Swoboda. 2025. Exploring the unique method for encoding sinograms in the first known Chinese–Polish dictionary. In Geoffrey Williams, Mathilde Le Meur & Andrés Echavarría Peláez (eds.), *West meets East: Papers in historical lexicography and lexicology from across the globe*, 271–295. Berlin: Language Science Press. DOI: 10.5281/zenodo.15394495 

reference work in the above-mentioned language pair is *Słownik chińsko-polski do czytania tekstów chińskich bez przygotowania* ('A Chinese-Polish Dictionary for Reading Chinese Texts Without Preparation'), also known by its Chinese title 切字中波辭典 (*Qiē Zì Zhōng-Bō Cídiǎn*<sup>1</sup> 'A Chinese-Polish Character-Splitting Dictionary'). It was compiled by Doman Wieluch, a Polish chemist, Sinologist and diplomat, who wrote the entries, the preface and the indexes to the dictionary. While Wieluch is credited as the sole author, additional contributions to his publication were made by three other individuals (Wieluch 1936: 15<sub>p</sub>)<sup>2</sup>. Yu Houo-Joei (虞和瑞 *Yú Héruì*), a Chinese diplomat and Polonist, helped to create the visual layout of the work and calligraphed the Chinese character components in the appendices (*Table I* and *Table II*). Witold Jabłoński, a renowned Polish Sinologist working at the University of Warsaw, provided Wieluch with source materials and advice. Konstanty Matuszek, Wieluch's colleague, was thanked by the author "for a great deal of painstaking and intelligent work put into the book".<sup>3</sup>

## 1.2 Characteristics and reception

*Słownik chińsko-polski...*, which contains approximately 6,000 entries on its 338 pages, was published by Wieluch himself in 1936 in Katowice. The reference work does not include a particularly large number of articles, which seems understandable, as the author himself perceived it as "merely the first step" for Chinese-Polish lexicography (Wieluch 1936: 14<sub>p</sub>) and explained that he did not have the financial means to publish a more comprehensive edition (Wieluch 1936: 3<sub>p</sub>). *Słownik...* remained the only bilingual general dictionary of its kind for several decades, but it was hardly ever used by Sinologists and translators and never received widespread acclaim. Its unpopularity may be demonstrated by the fact that it is actually absent from some of the publications which aim to give an account of the history of Polish-Chinese lexicography (cf. Sroka-Grądział 2018). To

---

<sup>1</sup>Unless otherwise stated, the pronunciation of Chinese words was indicated using the Hanyu Pinyin Romanisation system (漢語拼音 *Hànyǔ Pīnyīn*) throughout the chapter and all the spellings is given in traditional characters (繁體字 *fántǐzì* or 正體字 *zhèngtǐzì*). However, other transcriptions and simplified sinograms are preserved if they have been originally used in names or titles of publications.

<sup>2</sup>The three main sections of Wieluch's dictionary (excluding the *Tables I & II*) are numbered separately using Hindu-Arabic numerals, with no further distinctions. For the sake of clarity, the following letters were added to the relevant page numbers included in the references throughout this chapter: <sub>p</sub> (*Preface and Explanation*), <sub>d</sub> (*Dictionary*) and <sub>i</sub> (*Indexes*; see also section 1.4).

<sup>3</sup>Original quote: "za masę mrówczej i inteligentnej pracy, włożonej w niniejszą książkę" (Wieluch 1936: 15<sub>p</sub>).

some extent, the limited success of Wieluch's dictionary might be attributed to the lack of interest in Chinese-themed books among the general public or the relatively small print run of the work. In fact, only 200 numbered copies were originally created and some of the surviving volumes are incomplete.<sup>4</sup> On the other hand, it may also be suspected that one of the reasons for the relative obscurity of this reference work is the author's innovative and controversial approach to recording Chinese lemmata. Unlike most other lexicographers, Wieluch opted to encode the spelling of Chinese headwords in all of his dictionary entries instead of printing the relevant sinograms or writing them by hand, citing the complexity and inaccessibility of the Chinese script as the reason for this decision. The author explained his original character encoding system in the preface to the dictionary and promoted the system as a reliable and simple method of transcribing and printing Chinese texts, especially for foreign readers (Wieluch 1936: 1–3<sub>p</sub>). However, Wieluch's encoding system was never commonly accepted and did not appear in later bilingual dictionaries (see section 3.1).

### 1.3 Sources and inspirations

According to Wieluch, *Słownik chińsko-polski...* was compiled using the most recent editions of six previously published reference works (although no specific dates of publication were given). Four of these books were monolingual Chinese dictionaries: 說文解字 (*Shuōwén Jiězì*) by Xu Shen (許慎), 康熙字典 (*Kāngxī Zìdiǎn* '[the] Kangxi Dictionary') by Zhang Yushu (張玉書), Chen Tingjing (陳廷敬) et al., 攷正字彙 (*Kǎozhèng Zìhuì*) by Chen Haozi (陳淏子) and 通用新字彙 (*Tōngyòng Xīn Zìhuì*, author uncertain). In addition, Wieluch claimed to have consulted two bilingual dictionaries: one of them was the Chinese–English 漢英大辭典 (*Hàn-Yīng Dà Cídiǎn*) by O. Z. Tsang, also known as 張鵬雲 (*Zhāng Péngyún*), whereas the other was the Chinese–Russian 新中俄大字典 (*Xīn Zhōng-É Dà Zìdiǎn*) by 路大和 (*Lù Dàhé*); (Wieluch 1936: 14<sub>p</sub>).

Selected works of certain Western scholars, such as Angelo Zottoli, Séraphin Couvreur, Herbert Allen Giles and William Edward Soothill were reportedly also consulted by Wieluch while compiling his reference work, although he did not mention any specific titles. The author claimed to have relied on those publications in order to establish and describe the meaning of the Chinese lemmata

<sup>4</sup>For instance, during the preliminary research for this chapter it was revealed that the appendices (*Table I and Table II* – see section 1.4) were missing from the copy of *Słownik...* stored at the Jagiellonian Library in Kraków, which made that particular volume significantly more difficult to use.

included in his book, albeit to a limited extent (as his main sources were the aforementioned monolingual and bilingual Chinese dictionaries). He stressed that he had not limited himself to simply copying other Western authors and pointed out that the contents of *Słownik...* were in fact a result of his own deliberations (Wieluch 1936: 11<sub>p</sub>).

Wieluch's character encoding system was reportedly partially inspired by the earlier publications of John Chalmers and Léon Wieger (Wieluch 1936: 1<sub>p</sub>). Although, once again, no specific titles were mentioned by the author, it may be suspected that the works consulted by Wieluch were *An Account of the Structure of Chinese Characters Under 300 Primary Forms After the Shwoh-Wan, 100, A. D., and The Phonetic Shwoh-Wan, 1833* (Chalmers 1882), and *Chinese Characters: Their Origin, Etymology, History, Classification and Signification. A Thorough Study from Chinese Documents* (Wieger 1965, originally published in French in 1915).

#### 1.4 Contents

The first part of Wieluch's dictionary, *Wstęp i wyjaśnienie* ('Preface & Explanation') contains a general introduction to the reference work and to the author's original character encoding method. The latter was not only theoretically presented, but also practically demonstrated by including a sample Chinese text on page 12<sub>p</sub> and its encoded version on page 13<sub>p</sub>. The second part, called *Słownik* ('Dictionary'), is a single Chinese-Polish wordlist, which constitutes the main body of the publication. Its entries were arranged alphabetically based on the encoded spellings of the Chinese lemmata. The wordlist is followed by four indexes: the chronologically arranged index of Chinese dynasties and sovereigns, as well as three other indexes arranged alphabetically by character encodings: the index of Chinese era names, the index of Chinese provinces, prefectures, cities and towns, and the index of reversed character encodings. In addition, *Słownik polsko-chiński...* contains two appendices, printed on separate pieces of paper: *Table I* (an index of character components arranged alphabetically by code words) and *Table II* (an index of character components arranged by the number of brush strokes).

Most of the above-mentioned elements, such as prefaces, wordlists or indexes (including those containing typically extralinguistic information on history, geography or other related topics) appear quite frequently in various Chinese-Polish dictionaries (cf. Xu Lin (许琳), Yao Xishuang (姚喜双), et al. 2010, Kocyba-Grych & KołECKA 2012, etc.). On the other hand, indexes of reversed character encodings are exceedingly rare (if not completely unheard-of outside of Wieluch's

publication). According to the author, this special index was designed for deciphering unclear sinograms. Wieluch perceived it as a new, groundbreaking tool, which could enable future Sinologists (and other users) to make sense of partly erased or misprinted glyphs in the analysed Chinese texts (Wieluch 1936: 14<sub>p</sub>).

### 1.5 Macrostructure and microstructure

The entries in Wieluch's dictionary were arranged alphabetically by character encodings. This specific type of macrostructure is highly unusual for both monolingual and bilingual Chinese dictionaries. However, it does share certain characteristics with the traditional method of arranging Chinese-language lemmata used in *Shuowen Jiezi*, the *Kangxi Dictionary* and numerous other monolingual reference works (see 2.2, 3.1). In both of these types of macrostructure the order of entries is essentially determined by the structure of sinograms which appear in the spelling of the relevant headwords.

On the other hand, it might also be argued that the macrostructure of *Słownik...* is somewhat similar to the more contemporary and commonly used alphabetic order of entries based on the Hanyu Pinyin Romanisation system (cf. Xia Zhengnong (夏征农) & Chen Zhili (陈至立) 2009, Xu Lin (许琳), Yao Xishuang (姚喜双), et al. 2010, Kocyba-Grych & Kolečka 2012, Achinger & Radziwiłł 2015, etc.). In both of these types of macrostructure the order of headwords on the wordlist is determined by the conventional order of letters in the Latin alphabet, even if the methods of assigning sequences of letters to the relevant Chinese lemmata are anything but the same.

Either way, the macrostructure of *Słownik...* does appear radically dissimilar from the thematic order of entries which can be found in certain other Chinese–Polish dictionaries (cf. Wu Yuemei (吴月梅) 2010: Yin Xiangfeng (殷相峰) 2013: etc.).

Each entry in Wieluch's dictionary contains the encoding of one or more characters representing the relevant Chinese headword. In addition, every article includes one or more Polish equivalents of the lemma, one or more of its definitions in Polish or both of the aforementioned elements. Furthermore, each entry contains the information on the part (or parts) of speech to which the headword in question belongs. According to the author, this simple grammatical information (expressed by means of the commonly used Latin abbreviations, such as *n.*, *v.* or *adj.*) had been gleaned from Tsang's Chinese–English dictionary (漢英大辭典 *Hàn-Yīng Dà Cídiǎn* – cf. 1.3). As Wieluch himself pointed out, these part-of-speech symbols may occasionally seem to be at odds with the Polish equivalents or definitions provided in the entry, since – for the sake of conciseness – only

selected meanings of certain polysemic and multifunctional headwords had been considered (Wieluch 1936: 11<sub>p</sub>).

The pronunciation transcription of each headword, written in upper case, was indicated using a modified version of the Wade-Giles Romanisation. According to Wieluch, this particular transliteration system had been adapted to the spelling habits of Polish readers, mostly following prof. Jabłoński's suggestions (Wieluch 1936: 11<sub>p</sub>). The two most noticeable modifications were: the use of doubled letters instead of single letters with apostrophes<sup>5</sup> to denote aspirated consonants and the use of digraphs inspired by Polish orthography (with each letter doubled separately to represent an aspirated sound). For example, the pronunciation of the word 財 (*cái* 'wealth/property'), which typically would be rendered as *ts'ai*<sup>2</sup> in the Wade-Giles Romanisation system, was actually written as *TTSSAI* in *Słownik*.... Moreover, as can be seen in this example, the syllable tones were not indicated in any way in the transcription used in Wieluch's dictionary, as a result of which the information on pronunciation is incomplete for any given headword. Although it may be surprising, the above-mentioned imperfections appear to have been intended (or at least ignored) by the author. In fact, Wieluch admitted to "have treated the phonetic aspect of [his] dictionary with reprehensible negligence, as [he] wanted to emphasise that understanding Chinese texts [did] not depend on [knowing] their Chinese pronunciation."<sup>6</sup>

The phonetic transcription included in the entries may also serve a different purpose, as it sometimes provides additional information on the usage of certain headwords: the Romanisation is underlined if the lemma in question may function as a Chinese surname.

Some of the articles include other types of data, such as: information on antonyms, synonyms or other related words, information on spelling variants, full forms of abbreviated headwords, etc. Many compound words, collocations and fixed expressions appear in subentries, rather than as headwords in separate entries on the wordlist. In each of these subentries, the headword character encoding was replaced by the repetition symbol (./ or ./., depending on the number of repeated characters). For selected examples of microstructural elements in the entries of Wieluch's dictionary, see Figure 1.<sup>7</sup>

<sup>5</sup>Although the diacritic in question is often referred to as an *apostrophe* (i.e. ' ; cf. (Wieluch 1936: 11<sub>p</sub>), it may actually be the rough breathing character (i.e. ' or ', also known as the *spiritus asper*), as originally used by Thomas Wade.

<sup>6</sup>Original quote: "Stronę dźwiękową naszego słownika traktowałem z karygodnem [*sic*] niedbalstwem, chcąc uwydatnić niezależność rozumienia tekstów chińskich od ich chińskiego dźwięku." (Wieluch 1936: 11<sub>p</sub>).

<sup>7</sup>All figures included in this chapter were created using the photographs of the copy number 65 of *Słownik*..., which is currently owned by the Józef Lompa Voivodeship Pedagogical Library in Katowice (Pedagogiczna Biblioteka Wojewódzka im. Józefa Lompy w Katowicach).

In certain respects, the microstructure of *Ślownik...* is quite similar to that of many later Chinese-Polish dictionaries: the entries vary in length, but are usually relatively short; they contain some basic information on the key aspects of the headwords in question, such as their spelling, meaning, pronunciation or grammatical category (cf. Wu Yuemei (吴月梅) 2010, Xu Lin (许琳), Yao Xishuang (姚喜双), et al. 2010, Kocyba-Grych & Kolečka 2012, Yin Xiangfeng (殷相峰) 2013, etc.). However, the entries in Wieluch's dictionary are somewhat unique because of certain design choices made by the author. His novel Romanisation system is now completely obsolete and Hanyu Pinyin is almost invariably used instead; however, it should be noted that the latter system did not exist at the time of *Ślownik...*'s creation and the idea of adopting a universal, global Chinese Romanisation system for speakers of all languages was less commonly accepted. The other unprecedented element of Wieluch's dictionary articles – the character encodings – will be more thoroughly discussed in the next part of this chapter.

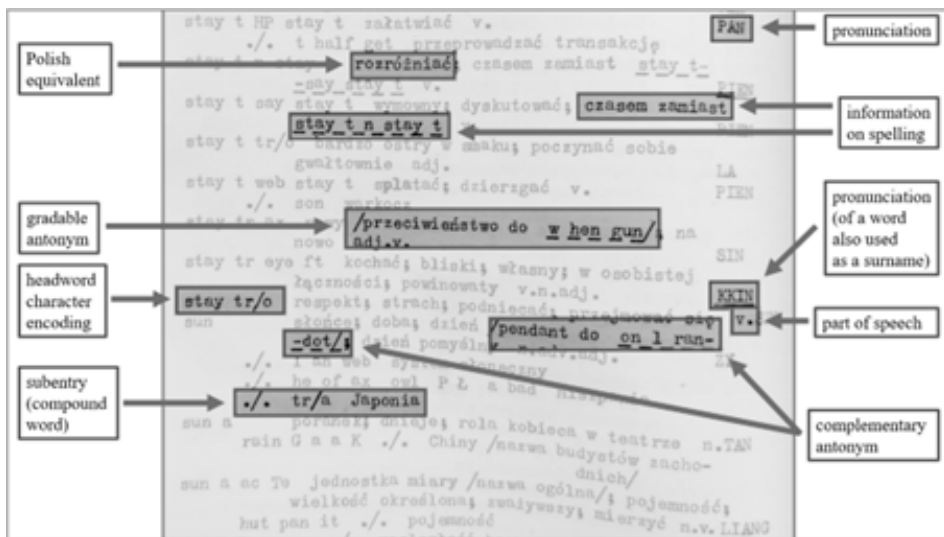


Figure 1: The microstructure of “Słownik...” (page 204<sub>d</sub>) – selected types of data (CC BY-NC-ND Doman Wieluch)

## 2 Wieluch's character encoding system

## 2.1 Introduction

As mentioned earlier, sinograms were used quite sparingly in Wieluch's dictionary: they only appear in the appendices, on the front cover and on page 12

(which contains a sample text in the regular form, next to its encoded version on page 13). All other Chinese characters were replaced throughout the dictionary by the author's original character encodings.

Every sinogram in Wieluch's system was essentially represented as a set of its parts, i.e. as a set of (*Chinese*) *character components*. Each of these character components was replaced by one of the 296 *code words*, which were supposed to be as simple and short as possible (Wieluch 1936: 1<sub>p</sub>). Every code word was either a specific letter of the Latin alphabet or a specific sequence thereof. While encoding entire Chinese characters, such code words were written in a particular order to represent the structure of the relevant sinograms. When two or more characters were being encoded, the corresponding sets of code words were written in rows, from left to right, top to bottom, so as to reflect the entire original text in a manner familiar to Western readers.

## 2.2 Selecting code words

The original criteria for selecting code words were explained by Wieluch in the preface to his dictionary (Wieluch 1936: 4-9<sub>p</sub>). A revised, simplified version of these criteria, based on the author's explanations and a new typological analysis of the character components used in his encoding system, will be presented in this part of the chapter. While the discrepancies between this new classification and the one originally included in the preface to Wieluch's dictionary are not particularly significant, they will be briefly discussed later on (see 2.3).

1. Some of the code words in Wieluch's system were chosen because of the similarity in shape between the relevant character components and certain letters of the Latin alphabet or groups thereof. For instance, the component 尸 was replaced by the code word *P*, the component 十 was substituted by the code word *t* and the component 乚 was paired with the code word *Ls* (Wieluch 1936: 4<sub>p</sub>, *Tables I & II*).
2. Some other code words were chosen because of the similarity in shape between the relevant character components and certain letters of the Greek alphabet. For instance, the component 厶 was assigned the code word *D*, as it resembles the Greek letter Δ; the component 冂 was replaced by the code word *G*, since it resembles the Greek letter Γ; the component 厶 was paired with the code word *s*, as it resembles the Greek letter Σ (Wieluch 1936: 4<sub>p</sub>, *Tables I & II*).



3. A number of other code words were chosen because of the similarity in shape between the relevant character components and the real-world objects denoted by the code words in question. For instance, the component 丶 was assigned the code word **dot**, the component 亾 was replaced by the code word **hill** and the component 𠂇 was paired with the code word **owl** (as it vaguely resembles the outline of a perched owl, (Wieluch 1936: 4<sub>p</sub>, *Tables I & II*).
4. Some other code words were chosen because of the similarity in shape between the relevant character components and some *elements* of certain real-world objects (or sets thereof). For instance, the component 冂 was assigned the code word **cot** [*sic* – see 2.3], as it vaguely resembles the outline of the main part of a coat (with missing sleeves); similarly, the component 𠂇 was replaced by the code word **pan**, as it is somewhat similar in shape to a pan hanging from a spit (with the bottom part of the pan missing; (Wieluch 1936: 4-5<sub>p</sub>, *Tables I & II*)).
5. Some of the code words were chosen because they denote the same objects which are referred to by the relevant character components (when these components are used as individual sinograms representing specific Chinese words). For instance, the component 身 (*shēn* ‘body’) was assigned the code word **body**; the component 犬 (*quǎn* ‘dog’) was replaced by the code word **dog**; the component 雨 (*yǔ* ‘rain’) was substituted by the code word **rain** and the component 羽 (*yǔ* ‘feather/wing’) was paired with the code word **wing** (Wieluch 1936: 5-7<sub>p</sub>, *Tables I & II*).
6. In certain cases a similar rule was followed, but the meaning of the code word in question was only close – and not identical – to that of the relevant Chinese character component. For instance, the component 虫 (*chóng* ‘insect/worm’) was assigned the code word **bee**; the component 鬼 (*guǐ* ‘ghost/demon’) was paired with the code word **elf**; the component 女 (*nǚ* ‘woman’) was replaced by the code word **her**, whereas the component 束 (*shù* ‘bunch, bouquet’) was substituted by the code word **rose** (Wieluch 1936: 7-8<sub>p</sub>, *Tables I & II*).
7. In some other cases a similar rule was applied, but the meaning of the code word was (almost) identical to that of a certain *more complex sinogram* containing the relevant Chinese character component. For instance, the component 亼 was assigned the code word **all**, as the component in question appears in the character 共 (*gòng* ‘common/total/altogether’); the

component 丁 was replaced by the code word **can**, as the component in question is part of the character 可 (*kě* ‘can, be able to’); the component 曲 was substituted by the code word **law**, as the component in question appears in the character 典 (*diǎn* ‘canon/law/ceremony’; Wieluch 1936: 8<sub>p</sub>, *Tables I & II*).

8. One of the code words, **hu**, was selected because of the similarity between its suggested pronunciation and the reading of the corresponding character component used as a separate sinogram: 互 *hù* (Wieluch 1936: 9<sub>p</sub>, *Tables I & II*).
9. The code word **a** was selected to represent the component 一 because of the perceived similarity between the traditional sequence of letters in the Latin alphabet and the conventional order of *Chinese character radicals* (i.e. principal character components, which often determine the order of articles in dictionaries – see section 3.1). While *a* is the first letter of the Latin alphabet, 一 (*yī*) is the first Chinese character radical (Wieluch 1936: 4<sub>p</sub>, *Tables I & II*).
10. Interestingly, Wieluch’s system is actually partially self-referential: some of the code words were selected because of the similarity in shape between the corresponding character components and certain other character components, to which certain other code words had already been assigned. For instance, as the component 一 had already been replaced by the code word **a**, and the component 尸 had already been replaced by the code word **P**, the component 尸 (which may be perceived as a combination of 一 and 尸) was substituted by the code word **a P**. Since the component 羊 had already been assigned the code word **ram**, the simpler component containing one fewer horizontal brush stroke at the bottom was paired with the code word **kid**. Similarly, as the component 車 had already been given the code word **car**, the component 𠂔 (whose bottom part is slightly different in shape) was replaced by the code word **van** (Wieluch 1936: 4<sub>p</sub>, 9<sub>p</sub>, *Tables I & II*).

Although typically only one of the above-mentioned rules was followed while choosing the code word for a specific character component, in some cases there may have actually been two applicable selection criteria. For example, the component 世 (*shì* ‘generation/world’) was paired with the code word **gen** because of the similarity in meaning between the words 世 and *gen[eration]*. Nevertheless, the same component was also presented by the author as a combination of three **t** (十) components put together. As he explained, such a description was

suggested by the above-mentioned similarity in shape, but also by the fact that a *generation* may be understood as a *period of thirty* (i.e. three times ten) *years* and 十 (*shí*) means ‘ten’ as a separate Chinese word (Wieluch 1936: 5<sub>p</sub>). Similarly, while the code word assigned to the previously mentioned component 一 (*a*) was selected because of the identical placement of the words *a* and 一 in monolingual dictionaries, Wieluch additionally cited the similarity in meaning between the English indefinite article *a* and the Chinese word 一 (*yī* ‘one’) as another reason for this pairing (Wieluch 1936: 4<sub>p</sub>).

While the above classification of code words within Wieluch’s system does give a general idea about the author’s intentions, it is by no means exhaustive. As Wieluch himself explained in the preface to his dictionary, some of the code words were in fact chosen “somewhat freely”, i.e. with hardly any regard for the above-mentioned criteria (Wieluch 1936: 9<sub>p</sub>). This issue, along with some other peculiar characteristics of the code word selection process within Wieluch’s system, will be elaborated on in the next part of the chapter.

### 2.3 Peculiarities of the code word selection

While considering the code word selection in Wieluch’s character encoding system, it might be useful to discuss some of the potentially controversial characteristics of this process. Although the task of dissecting and methodically recreating all Chinese characters is admittedly daunting and likely impossible to complete in a perfect way, it may be argued that some of the peculiarities of Wieluch’s code word selection process create certain practical limitations to his system and render it more challenging for the prospective users.

As previously indicated, some of the code words in Wieluch’s system were selected quite arbitrarily, without any explicitly presented reason, which the author himself admitted. For example, the component 又 was assigned the code word *it*, which seems in no way related to its shape, its reading (*yòu*) or its meaning as a separate sinogram (‘again/also’). Similarly, the component 甫 (*fǔ* ‘just, barely/[one’s courtesy name]’) was paired with the code word *my* (Wieluch 1936: 9<sub>p</sub>).

In some other cases there actually was a specific reason for choosing a certain code word, but it was not particularly obvious and did not clearly correspond to any of the previously established selection criteria. For instance, the component 齊 (*qí* ‘neat/simultaneous/[name]’) was assigned the code word *Poland* (i.e. the name of Wieluch’s home country), which the author himself explained with this somewhat fatalistic remark: “[齊] represents grain on a field, it denotes a fertile plain, and also a certain ancient state – for this patriotic gesture I am going

to have difficulties with my system in international education[/academia] and Polish Sinology is not going to support me [on this].”<sup>8</sup> In other words, in certain cases the previously delineated criteria for selecting code words were treated rather as general guidelines than as strict, unbreakable rules.

As may be understood from the quotation included in the previous paragraph, Wieluch hoped that his method for encoding sinograms could one day be used all over the world. To that end, he set out to create a truly international system for encoding Chinese characters, which would be useful not only to his compatriots, but also to speakers of various other languages. In order to accomplish that goal, he chose English as the source language for the majority of his code words, as evidenced by the examples included in the previous parts of this chapter; Wieluch 1936: 1-2<sub>p</sub>.

While Wieluch mostly abided by the above-mentioned rule, it should be noted that exceptions were made in certain cases. Some of the code words were actually derived from certain French or Latin lexemes and not their English counterparts. For instance, the French code words *bon* (‘good’) and *pas* (‘step’) were used instead of their English equivalents, as were the Latin code words *cor* (‘heart’), *et* (‘and’) and *mus* (‘mouse’; Wieluch 1936: 5-6<sub>p</sub>, 8<sub>p</sub>).<sup>9</sup> Although it might be argued that both French and Latin enjoyed the status of international languages at a certain point in history (or even that they continue to do so now), the use of such code words may have introduced additional (and possibly unnecessary) complexity to Wieluch’s character encoding system.

Furthermore, it should be noted that some of the aforementioned code words derived from French or Latin are quite similar to certain other code words derived from English, even though they obviously do not correspond to the same character components. For instance, while the code word *et* (‘and’) was selected for the component 而 (*ér* ‘and/and so/yet’), the code word *and* was chosen for the component 且 (*qiě* ‘and/moreover/yet’). Selecting each of these code words appears justified at first glance, as each of them (rather accurately) reflects the meaning of the relevant component as a separate Chinese character. On the other hand, it might be argued that the inclusion of both *et* (‘and’) and *and* as code words could be confusing to the potential users of the system, given that these

---

<sup>8</sup>Original quote: ”[齊] wyobraża łąn zboża, oznacza żyzną równinę, a także pewne starożytne państwo – za ten gest patryjotyczny [sic] będę miał trudności z moim systemem w nauce między-narodowej, a polska sinologia [sic] mnie nie poprze” (Wieluch 1936: 8<sub>p</sub>).

<sup>9</sup>The code words *cor*, *et* and *mus* can be identified as Latin (and not French or English) based on the comments and translations provided by Wieluch; for the same reason, it is certain that the code word *pas* is undoubtedly of French origin (Wieluch 1936: 5-6<sub>p</sub>, 8<sub>p</sub>).

words are identical in meaning in their respective source languages (Wieluch 1936: *Tables I & II*).

Another possible issue for the prospective foreign users of Wieluch's encoding system is the inclusion of the Polish letter *Ł* (also known as *L with stroke*, lower-case form: *ł*; Wieluch 1936: *Tables I & II*). As the letter in question only exists in the spelling systems of about a dozen languages (most notably Polish, but also Kashubian or Navajo, among others) and was actually even less common at the time of the creation of Wieluch's dictionary, it could be argued that its presence in a supposedly universal encoding system may have rendered it slightly more difficult for many potential users.

Since the code words in Wieluch's system were meant to be short, it is perhaps not surprising that some of them were coined by abbreviating certain relatively long source-language words. For example, **HP** was used instead of *horsepower*, **Me** was chosen instead of *metal*, **rt** was selected instead of *right* and **sec** appeared instead of *second*. Similarly, the code word **M** was used instead of the French *Monsieur* ('Mister') and the code word **Te** appeared instead of the Latin *Tellus* ('Earth'); (Wieluch 1936: 4–8<sub>p</sub>).

For the sake of brevity, non-standard spellings were also used as code words in certain cases. *Coat* was respelled as **cot** (cf. section 2.2), *door* was rendered as **dor**, *hand* was replaced by **had** and *plough* was written as **log** (Wieluch 1936: 4–5 :<sub>p</sub>). The analysis of these examples points to another possible problem with Wieluch's code word selection: while the above-mentioned words are intended to be understood as the unusual variant spellings of the words *coat*, *door*, *hand* and *plough*, they might be erroneously recognised as the regular spellings of the words *cot* ('a small or portable bed'), *dor* ('trick, joke/[a kind of insect]'), *had* ('[past simple or past participle form of *have*']') and *log* ('a thick piece of wood/a written record'), thus making the system more confusing.

While the majority of code words refer specifically to single character components (as the general rules would suggest), this is actually not always the case. To some extent, this multireferentiality of certain code words mirrors one of the characteristics of the Chinese script: the same character component may be written in a slightly different way (i.e. have a somewhat different shape) when it appears as a component in different complex sinograms. For instance, the character component 心 (*xīn* 'heart/mind/centre') is written as 心 in the character 想 (*xiǎng* 'to think/to want/to miss'), but as 忄 in the character 怕 (*pà* 'to fear, to dread') and as 忄 in the character 恭 (*gōng* 'respectful, polite'); however, all of these variant forms are replaced by the same code word (**cor** 'heart') in Wieluch's encoding system (see Figure 2). Since both the origin and the meaning of all such variant forms are essentially identical, the author's decision to group all these forms

under the same code word appears justified, even if it might increase the complexity of his system. Nevertheless, in certain rare cases the decision to assign the same code word to more than one component seems to have been rather arbitrary, as no etymological reason may be cited to support it. For instance, while the code word **tow** corresponds to the mutually related components 申 and 电, it also refers to the similarly-shaped, but unrelated component 𠂇 (Wieluch 1936: *Tables I & II*).

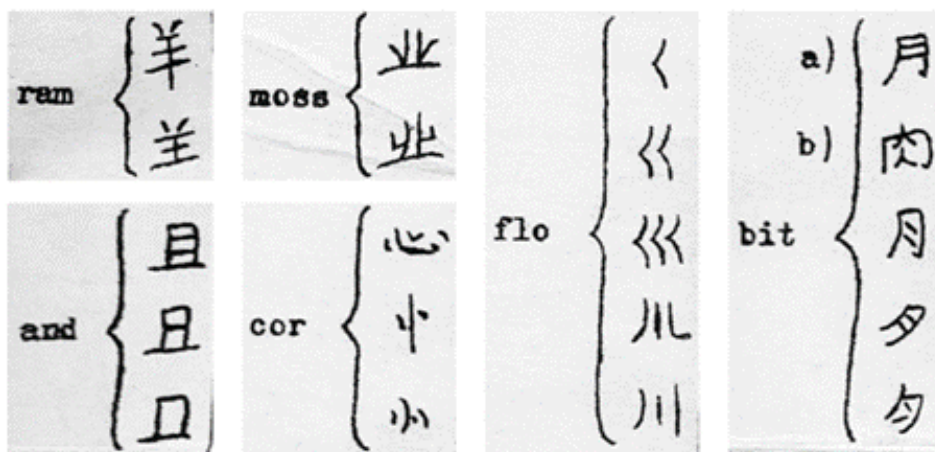


Figure 2: Multireferentiality of code words in Wieluch's system (CC BY-NC-ND Doman Wieluch)

Another possible issue with Wieluch's code word selection process is the fact that both upper-case and lower-case letters were used as single-letter code words. In certain cases the relevant corresponding elements are markedly different: for example, the code word **b** (derived from the English word *born*) corresponds to the component 生 (*shēng* 'to be born/to give birth') and the code word **B** (selected because of the similarity in shape between the letter and the relevant component) corresponds to the component 𠂇, which makes these two code words relatively easy to distinguish. On the other hand, some pairs of single-letter code words (containing the same letter in upper case and in lower case) were chosen solely because of the similarities in shape between the letters and the relevant components, which makes selecting and identifying such code words significantly more challenging. These difficulties are sometimes exacerbated by the multireferentiality of such single-letter code words and the similarities between the letters themselves. This issue is perhaps most noticeable when we consider the code words **I**, **i**, **L** and **l** (all of them being distinct and multireferential), which together correspond to a total of eleven different character components (see Figure 3; Wieluch

1936: *Tables I & II*). The code words **o** (口) and **O** (口) may be considered a similar, but somewhat particular example, as they are additionally differentiated by the combination properties of their corresponding components. While the **o** component never contains other components, the **O** component invariably does. These properties may be observed in the encoding of the character 固 (gù ‘solid, hard’), rendered as **O t o** (Wieluch 1936: 4<sub>p</sub>, 163<sub>d</sub>; see also 2.4).

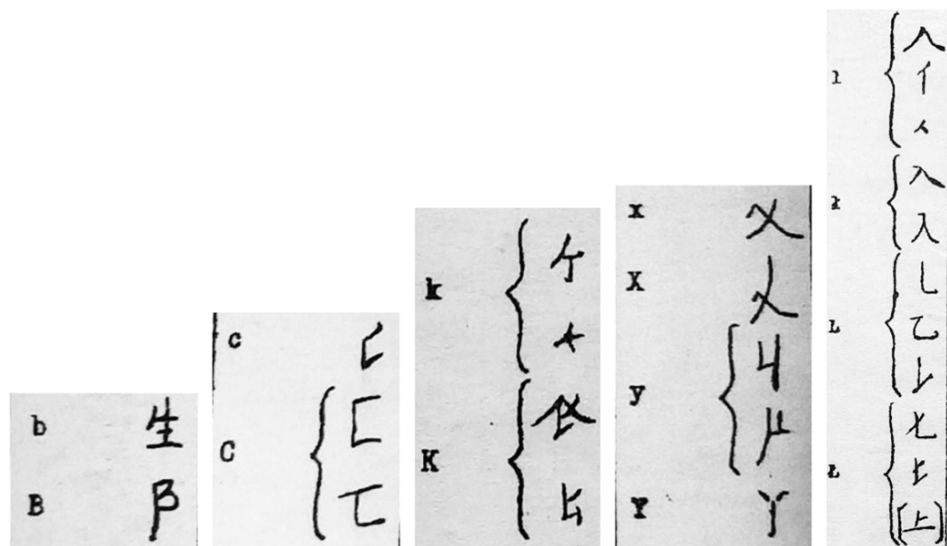


Figure 3: Upper-case and lower-case letters as code words in Wieluch's system (CC BY-NC-ND Doman Wieluch)

The above-mentioned multireferentiality of certain code words is additionally complicated by the fact that some character components may occur as variant forms which are identical to certain other components. As a result, some forms are associated with more than one code word (see Figure 4 – the code words associated with variant forms are given in square brackets). For example, while the character component 上 was primarily paired with the monoreferential code word **up**, it was also described as the third variant form associated with the code word **L** (cf. Figure 3. Wieluch 1936: *Tables I & II*).

Finally, as Wieluch himself admitted in the preface to his dictionary, his encoding system is actually not completely exhaustive. While the code words listed in his publication are perfectly sufficient to encode numerous sinograms, some Chinese character components were not assigned any specific code words of their own. Thus, if one of those components appears in a certain Chinese character (and cannot be broken down into two or more encodable components), it has

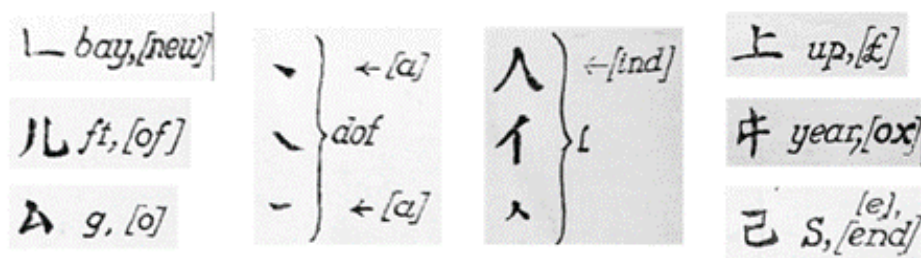


Figure 4: Multireferentiality of character components in Wieluch's system (CC BY-NC-ND Doman Wieluch)

to be encoded in the same way as we would normally encode *the most similar* component explicitly included in the system (Wieluch 1936: 9<sub>p</sub>).

As the various types of code words and the peculiarities of the process of code word selection in Wieluch's system have already been discussed, it might be useful to end this part of the chapter by briefly summarising the differences between the previously presented classification of code words and the one originally included in the preface to *Słownik...*:

1. In the original classification, the code word **ah**, which had been derived from the code word **a** because of the similarity in shape between the relevant corresponding components (see Figure 5), was not described as yet another example of a code word based on another, previously selected code word, but instead was listed in a separate category (Wieluch 1936: 4<sub>p</sub>).
2. In the original classification, the code words based on other, already selected code words, such as **a P** (戸), based on **a** (一) and **P** (尸) or **van** (夷), based on **car** (車), were divided into two different groups: combined components (such as **a P**) and modified components (such as **van**; Wieluch 1936: 4-9<sub>p</sub>).
3. In the original classification, there was a small separate group for code words based on mnemonics, even though these code words arguably could have been placed in other categories. One of these code words was **bud** (朱), selected because of the similarity in shape between its corresponding component and the component paired with the code word **tr** (木 'tree'; Wieluch 1936: 8-9<sub>p</sub>).
4. In the original classification, the arbitrarily assigned code words, such as **it** (又) or **my** (甫) were grouped in a separate category (Wieluch 1936: 9<sub>p</sub>).



5. In the original classification, some of the code words discussed in this chapter, such as *owl* (巴), belonged to different, but similar categories (Wieluch 1936: 4-9<sub>p</sub>).

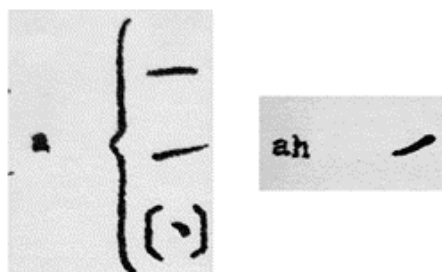


Figure 5: Components corresponding to the code words *a* and *ah* in Wieluch's system (CC BY-NC-ND Doman Wieluch)

## 2.4 Encoding entire sinograms

The process of replacing Chinese character components with code words, described in greater detail in the previous parts of the chapter, was a key element of Wieluch's encoding system. However, it should be noted that most sinograms are actually composed of two or more character components. For this reason, it was also necessary to devise a set of rules regarding the process of forming entire sequences of code words, which could represent more complex Chinese characters.

If a certain sinogram consisted of a single component, the relevant code word was simply written down instead of the character in question. For instance, the character 失 (*shī* 'to lose/to miss/to fail') was substituted by the code word *lose*, the character 山 (*shān* 'mountain') was replaced by the code word *Mt* and the character 言 (*yán* 'speech/to say') was substituted by the code word *say* (Wieluch 1936: 137<sub>d</sub>, 146<sub>d</sub>, 190<sub>d</sub>).

As indicated earlier, variant forms of the same character component were typically substituted by the same multireferential code word in Wieluch's system. While the author suggested the possibility of clearly identifying specific component forms occurring in complex sinograms by adding numerals to the relevant code words (Wieluch 1936: 9<sub>p</sub>), he decided against adopting this solution in his dictionary. However, if two variant forms of a certain component could exist as two different separate sinograms (representing two different words), the relevant

forms were additionally marked with the letters *a*) and *b*),<sup>10</sup> so as to avoid ambiguity. This solution was used for exactly two pairs of characters corresponding to two pairs of variant component forms. 月 (*yuè* ‘moon/month’) was assigned the sequence **bit a**), while 肉 (*ròu* ‘meat, flesh’) was replaced by the sequence **bit b**) (cf. Figure 2); 子 (*zǐ* ‘son/child’) was represented as **son a**), whereas 孑 (*jié* ‘lonely, alone’) was substituted by **son b**) (Wieluch 1936: 9<sub>p</sub>, 23<sub>d</sub>, 198<sub>d</sub>, *Tables I & II*).

If a certain sinogram included two or more character components, the consecutive code words corresponding to these components were written one after another with either single spaces or hyphens (-) between them. The specific order of these code words, reflecting the structure of the original Chinese glyph, was determined by the following rules:

1. If the character contained two or more vertical components placed next to one another, the relevant code words were written from the left-hand side to the right-hand side of the sinogram. For example, the character 袖 (*xiù* ‘sleeve’) was encoded as **robe by**, as it includes the component 衤 (corresponding to the code word **robe**) on the left-hand side and the component 由 (corresponding to the code word **by**) on the right-hand side. Similarly, the character 謝 (*xiè* ‘to thank/to wither/to decline’) was replaced by the sequence **say body cm**, as it includes the component 言 (corresponding to the code word **say**) on the left-hand side, the component 身 (corresponding to the code word **body**) in the centre and the component 寸 (corresponding to the code word **cm**) on the right-hand side (Wieluch 1936: 10<sub>p</sub>, 186<sub>d</sub>, 191<sub>d</sub>).
2. If the character contained two or more horizontal components placed next to one another, the relevant code words were written from the top part to the bottom part of the sinogram. For instance, the character 分 (*fēn* ‘to divide’) was encoded as **of n**, since it includes the component 八 (corresponding to the code word **of**) as its top part and the component 刀 (corresponding to the code word **n**) as its bottom part. Similarly, the character 夢 (*mèng* ‘dream/to dream’), which consists of the components 亼 (**w**), 𠂔

<sup>10</sup>In fact, these symbols were variously spelled as *a*) and *b*) (Wieluch 1936: *Tables I & II*), /*a*/ and /*b*/ (Wieluch 1936: 23<sub>d</sub>) or /*forma a*/ and /*forma b*/ (Wieluch 1936: 198<sub>d</sub>) in the dictionary. For the sake of clarity, only the first of these three pairs of symbols is consistently used in this chapter.

(*net*), 冫 (*m*) and 夕 (*eve*), was represented by the sequence **w net m eve** (Wieluch 1936: 10<sub>p</sub>, 165<sub>d</sub>, 238<sub>d</sub>).<sup>11</sup>

3. If the character contained two or more components arranged in a different manner, the relevant code words were written from the top left-hand corner to the bottom right-hand corner of the sinogram. For example, the character 但 (*dàn* ‘but/merely’) was encoded as **l sun a**, as it includes the component 亻 (*l*) on the left-hand side and the components 日 (*sun*) and 一 (*a*) on the right-hand side. The character 架 (*jià* ‘to support/frame’) was encoded as **HP o tr**, since it includes the components 力 (*HP*) and 口 (*o*) as its top part, and the component 木 (*tr*) as its bottom part. The character 能 (*néng* ‘can, to be able’) was encoded as **g bit L L**, as it includes the components 厶 (*g*) and 月 (*bit*) on the left-hand side and two 乚 (*L*) components on the right-hand side. The character 固 (*gù* ‘solid, hard’) was rendered as **O t o**, since it includes the components 十 (*t*) and 口 (*o*) placed inside the component 囗 (*O*; cf. 2.3; Wieluch 1936: 10<sub>p</sub>, 62<sub>d</sub>, 108<sub>d</sub>, 133<sub>d</sub>, 163<sub>d</sub>).

In some rare cases, applying the above-mentioned rules was insufficient to unequivocally represent a certain character, as the only difference between the glyph in question and a certain other sinogram was the placement of components, with identically encoded components placed on the left-hand and right-hand sides in one character, and in the top and bottom parts in the other. The encodings of such characters included additional symbols to indicate the vertical (||) or horizontal (=) relative placement of the relevant components. For instance, the character 叭 (*bā* ‘[sound of a sharp noise]’) was represented by the sequence **o of ||**, while the character 只 (*zhǐ* ‘only, merely’) was replaced by the sequence **o of =**. A similar solution was adopted to encode more complex sinograms, such as 暉 (*huī*, ‘sunshine/to shine upon’), paired with the sequence **sun m car |=** and 暈 (*yūn* ‘dizzy/to faint’ or *yùn* ‘halo/dizzy/to faint’), substituted by the sequence **sun m car =** (Wieluch 1936: 10<sub>p</sub>, 159<sub>d</sub>, 206<sub>d</sub>).

The forward slash symbol (/) was used between two code words in certain sequences to indicate that one character component was placed on top of another. For instance, the character 本 (*běn* ‘root/origin’) was replaced by the sequence **tr/a**, as it consists of the component 一 (*a*) placed on top of the bottom part of the component 木 (*tr*). The character 便 (*biàn* ‘suitable, convenient’ or *pián* ‘[bound form]’) was represented by the sequence **l a X/sun**, since it contains the

<sup>11</sup>The component 卩 in the character 夢 may be substituted by the component 卩<sup>+</sup> in certain fonts; nonetheless, both 卩 and 卩<sup>+</sup> are treated as variant forms of the same component in Wieluch’s system and are both replaced by the code word **w**.

component 亻 (*l*) on the left-hand side, the component 一 (*a*) in the top right-hand corner and the combination of the components 又 (*X*) and 日 (*sun*) in the bottom right-hand corner. Similarly, the character 費 (*fèi* ‘to cost/to spend/fee’) was replaced by the sequence **bow/f i d**, as it consists of the components 丩 (*f*) and 丨 (*i*) placed over the component 弓 (*bow*) in the top part of the sinogram and the component 貝 (*d*) in its bottom part (Wieluch 1936: 10<sub>p</sub>, 28<sub>d</sub>, 124<sub>d</sub>, 217<sub>d</sub>).

Numerals were added in front of certain code words in some of the sequences in order to avoid repetition. For example, the sinogram 桑 (*sāng* ‘mulberry tree’) was replaced by the sequence **3 it tr**, as it contains three 又 (*it*) components in the top part and the component 木 (*tr*) in the bottom part. Similarly, the character 饒 (*ráo* ‘abundant’) was assigned the sequence **eat 3 Te a ft**, as it contains the component 饣 (*eat*) on the left-hand side, three 土 (*Te*) components in the top right-hand corner and the components 一 (*a*) and 儿 (*ft*) in the bottom right-hand corner (Wieluch 1936: 10<sub>p</sub>, 54<sub>d</sub>, 116<sub>d</sub>).

However, some other sequences did include repeated code words, since no numerals were used in the same way as in the above-mentioned examples. For instance, the character 鑰 (*yuè/yào* ‘key’), which contains the component 金 (*Me*) on the left-hand side, the component 厶 (*D*) in the top right-hand corner, three 口 (*o*) components below it and the component encoded as **vol** in the bottom right-hand corner, was substituted by the sequence **Me D o o o vol**. Similarly, the sinogram 傘 (*sǎn* ‘umbrella’), including five 人 (*l*) components and one 十 (*t*) component, was replaced by the sequence **l l l t l l** (Wieluch 1936: 10<sub>p</sub>, 129<sub>d</sub>, 141<sub>d</sub>).

In some cases it was possible to dissect the same character in different ways, thus generating more than one corresponding set of components and, by extension, more than one possible sequence of code words. To avoid any confusion on the part of the reader, only one of the possible sequences was consistently used to represent such sinograms. For example, the character 辛 (*xīn* ‘pungent/laborious/eighth in order’) was assigned the sequence **stay t** (立十), and not **\*on kid**. Such sequences were also consistently used in more complex sinograms: for instance, 辣 (*là* ‘spicy’) was represented as **stay t tr/o** (立十木/口), and not as **\*on kid tr/o** (Wieluch 1936: 10<sub>p</sub>, 204<sub>d</sub>).

*Słownik chińsko-polski...* is essentially a *character dictionary* (or 字典 *zìdiǎn*; cf. Huang Chu-Ren et al. (2016: 545-551), as the vast majority of its headwords are monosyllabic lexemes, originally represented by single sinograms. Compound words, just like collocations and fixed expressions, typically appear in subentries; they contain repetition symbols (./.) instead of the headword characters and they are written with double spaces between the encodings of separate glyphs. For example, the word 日本 (*Rìběn* ‘Japan’) was listed as ./ **tr/a** under the headword 日 (*rì* ‘sun/day’ **sun**), while the word 義大利 (*Yìdàlì* ‘Italy’) was listed as ./ **big gr**

**n** under the headword 義 (yì ‘justice, righteousness’ **ram I**; Wieluch 1936: 183<sub>d</sub>, 204<sub>d</sub>; cf. 1.5, Figure 1). However, if a certain sinogram is used exclusively in the spelling of bound forms, i.e. parts of disyllabic, two-character words, these disyllabic lexemes appear as headwords, and not as compound words in subentries. Examples of such headwords include 骷髏 (*kūlóu* ‘human skull/human skeleton’), rendered as **os bit t o os bit lot**, 霹靂 (*pīlì* ‘thunderclap/a bolt [from the blue]’), replaced by **rain P o stay t rain G gr gr leg** and 邂逅 (*xièhòu* ‘to meet by chance’), substituted by **way n use n ox way G a o** (Wieluch 1936: 172<sub>d</sub>, 182<sub>d</sub>, 245<sub>d</sub>).

### 3 Wieluch’s method and its applications in the context of other systems

#### 3.1 Ordering dictionary entries

Wieluch’s method of encoding sinograms may be compared to a number of other systems designed to categorise Chinese characters by their structure, i.e. by the shape and number of their components. Such systems have been used for centuries for the purpose of establishing the order of sinograms (and, by extension, Chinese headwords) on the wordlists of Chinese dictionaries (cf. 1.5, 2.2). Examples of such systems include the system of 540 *character radicals* introduced in 說文解字 (*Shuowen Jiezi*) by Xu Shen (許慎), the system of 214 radicals (or 部首 *bùshǒu* in Chinese), introduced in 字彙 (*Zihui*) by Mei Yingzuo (梅膺祚) and featured in the 康熙字典 (*Kangxi Dictionary*) by Zhang Yushu (張玉書), Chen Tingjing (陳廷敬) et al. (cf. 1.3), as well as the systems of approximately 200 (e.g. 188, 189, 200 or 201) components in later simplified-character dictionaries (cf. Cheung Kam-Siu 2016: 531–539, Huang Chu-Ren et al. 2016: 545–551).

Many character components are identical in the 540-radical system, the 214-radical system and Wieluch’s 296-code word system (disregarding the differences in font styles). For instance, 火 (*huǒ* ‘fire’) is known as the 382<sup>nd</sup> radical in *Shuowen Jiezi*, the 86<sup>th</sup> radical in the *Kangxi Dictionary* and a component paired with the code word **hot** in *Śłownik*.... Similarly, 言 (*yán* ‘speech/to say’) is the 56<sup>th</sup> radical in *Shuowen Jiezi*, the 149<sup>th</sup> radical in the *Kangxi Dictionary* and a component represented in *Śłownik*... by the code word **say**. Such correspondences often extend to the variant forms of components: for example, 灬 is classified as a variant form of 火 (**hot**) in all of these systems. In certain cases, the code word proposed by Wieluch is analogous to a single *Kangxi* (*Zihui*) radical, but to two separate *Shuowen* radicals. For instance, although the components 乙 and 乚 are both denoted by Wieluch’s code word **L** and classified as the *Kangxi* radical number 5, in *Shuowen Jiezi* 乙 is the 514<sup>th</sup> radical and 乚 is the 456<sup>th</sup> radical.

Nevertheless, Wieluch's system is actually unique in some respects. For example, despite the fact that the component 曰 (*yuē* 'to speak, to say'; not to be confused with 日 *rì* 'sun/day') is described as either the 152<sup>nd</sup> *Shuowen* radical or the 73<sup>rd</sup> *Kangxi* radical, it is a complex sinogram represented as **O a** (口一) in *Słownik...* Similarly, while the component 竹/<sup>木</sup> (*zhú* 'bamboo') is also a character radical (number 143 or 118, respectively), in Wieluch's system it is encoded as **k** **k**, since each of its two parts (cf. Figure 3) is classified as a separate component (cf. *HanDian* (漢典) 2023, Wieluch 1936: *Tables I & II*, 117–122<sub>d</sub>, 162<sub>d</sub>).

Although the system of 214 radicals and its slightly altered versions are still commonly used in the creation of wordlists, indexes and entries of both monolingual and bilingual Chinese dictionaries (cf. Xia Zhengnong (夏征农) & Chen Zhili (陈至立) 2009, Xu Lin (许琳), Yao Xishuang (姚喜双), et al. 2010, Kocyba-Grych & Kolečka 2012, Achinger et al. 2023, *HanDian* (漢典) 2023, *MDBG Chinese Dictionary* 2023 etc.), Xu Shen's system has long fallen out of use and Wieluch's system never became popular, as evidenced by the fact that it has not been featured in any of the later Chinese-Polish dictionaries (cf. Sroka-Grądział 2018). This is perhaps not particularly surprising, given that the advancements made since the publication of *Słownik...* in the fields of linguistics (e.g. the invention of the Hanyu Pinyin Romanisation system), computer science (e.g. the development of the Internet or the Unicode standard) and electronics (e.g. the introduction of touchscreen devices) have largely contributed to rendering Wieluch's method replaceable and obsolete.

Wieluch himself made the following remark: "[T]here does not – and cannot – exist a dictionary [...] which would *unequivocally* lead [the reader] from the sound [of a certain Chinese word] to the glyph which it denotes."<sup>12</sup> While this statement technically still holds true today for the majority of monosyllabic Chinese words, modern users of electronic Chinese dictionaries (cf. Achinger et al. 2023, *HanDian* (漢典) 2023, *MDBG Chinese Dictionary* 2023 etc.) have mostly become accustomed to searching headwords by rewriting the unfamiliar characters on the screen or using pronunciation-based queries. While the latter method usually does not yield accurate results (with certain exceptions: for example, the queries *shei*, *shei2* etc. unequivocally lead to the word 誰 *shéi* 'who?'), it is relatively easy to browse a number of (nearly) homophonic lemmata, which can be generated by the dictionary in a matter of split seconds. Similarly, the modern paper-based monolingual and bilingual Chinese dictionaries usually also have a Romanisation-based (typically Hanyu Pinyin-based) alphabetic macrostructure

<sup>12</sup>Original quote: "[N]ie istnieje i istnieć nie może słownik [...], któryby [sic] od tego dźwięku jednoznacznie prowadził do wyrażonego przezeń charakteru [sic]." (Wieluch 1936: 3<sub>p</sub>).

(cf. Xia Zhengnong (夏征农) & Chen Zhili (陈至立) 2009, Xu Lin (许琳), Yao Xishuang (姚喜双), et al. 2010, Kocyba-Grych & Kolečka 2012, Achinger & Radziwiłł 2015, etc.; cf. 1.5). While this significant shift in the common lexicographic practice needs to be acknowledged from the modern perspective, it may be argued that predicting such a drastic change was rather difficult (if not completely impossible) in Wieluch's times.

### 3.2 Encoding and inputting sinograms

Wieluch's method, designed to uniquely represent every sinogram as a sequence of code words denoting character components, may also be compared to its modern counterparts, such as the Intelligent Chinese Character Encoding System. For example, the character 峰 (*fēng* 'peak, summit') is rendered as **Mt ran ff** (山久丰) in Wieluch's system and as 山久丰 in ICCES (Wieluch 1936: 148<sub>d</sub>, Huang Chu-Ren et al. 2016: 546–547). Unlike the method presented in *Słownik...*, the modern encoding systems are not designed to transcribe entire Chinese texts (which nowadays are rather easy to type), but may be used in electronic dictionaries to facilitate the analysis and acquisition of unfamiliar characters or to introduce cross-references to variant forms of sinograms (cf. Achinger et al. 2023, *HanDian* (漢典) 2023, *MDBG Chinese Dictionary* 2023, Huang Chu-Ren et al. 2016: 546–547).

As a tool for writing in Chinese, Wieluch's system is based on roughly the same principles as the modern shape-based Chinese character input methods, such as *Cangjie* (倉頡), *Zhengma* (鄭碼) or *Wubi* (五筆). For example, just like the word 這 (*zhè* 'this') could be rendered as **way say** (讠言) in Wieluch's system, it may also be encoded as **YYMR** (卜卜一口) in *Cangjie* (cf. *HanDian* (漢典) 2023, Wieluch 1936: 246<sub>d</sub>). However, compared to the above-mentioned methods, Wieluch's system contains significantly more basic components and (at least currently) cannot be used to input Chinese characters on electronic devices. While it may be hypothesised that Wieluch's system could be adapted for this purpose, it probably would not be more efficient than the already existing methods.

### 3.3 Concluding remarks

Wieluch's encoding system was certainly a unique and interesting proposition, which aimed at solving the actual problems of contemporary non-native speakers and learners of Chinese. Although it was inspired by some earlier works of Western scholarship and the traditional classifications of character radicals, it was the very first method of analysing and transcribing sinograms to be practically applied in the creation of a Chinese–Polish dictionary. The groundbreaking

character of Wieluch's ideas is clearly demonstrated by the fact that the reference work in question was written at a time when most likely no such publications existed.

Today, this original Chinese character encoding system and the unprecedented dictionary compiled to showcase its abilities are both largely forgotten, not unlike Doman Wieluch himself. It may be argued that the solutions which he proposed almost ninety years ago were too complicated or too arbitrary, at least in certain respects. We might also suspect that the system was not sufficiently popularised at the time or demonstrate that it was effectively rendered obsolete by the advancement of technology and science. Nevertheless, appreciating this ambitious, even if ultimately unsuccessful endeavour is an essential prerequisite to fully understanding the history of Chinese-Polish lexicography. Although *Słownik chińsko-polski do czytania tekstów chińskich bez przygotowania* is not likely to receive widespread acclaim in the modern era, studying Wieluch's magnum opus could be a source of inspiration for future dictionary authors.

## References

- Achinger, Krzysztof et al. 2023. *Słownik chińsko-polski i polsko-chiński – Bohan*. <http://www.bohan.pl> (23 April, 2023).
- Achinger, Krzysztof & Antoni Radziwiłł. 2015. *Polsko-chiński i chińsko-polski słownik biznesowy* 波汉-汉波商务词典. Kęty: Wydawnictwo Bohan.
- Chalmers, John. 1882. *An account of the structure of Chinese characters under 300 primary forms after the shwoh-Wan, 100, a. D., and the phonetic shwoh-Wan, 1833*. London: Trübner & Co.
- Cheung Kam-Siu. 2016. Lexicography. In Chan Sin-wai et al. (eds.), *The Routledge encyclopedia of the Chinese language*, 215–226. London/New York: Routledge.
- HanDian (漢典). 2023. HanDian. <https://www.zdic.net> (23 April, 2023).
- Huang Chu-Ren, Li Lan & Su Xinchun. 2016. Lexicography in the contemporary period. In Chan Sin-wai et al. (eds.), *The Routledge encyclopedia of the Chinese language*, 545–562. London/New York: Routledge.
- Kocyba-Grych, Katarzyna & Zuzanna Kolečka. 2012. *Słownik chińsko-polski*. Warsaw: Wydawnictwo Edgard.
- MDBG Chinese Dictionary. 2023. MDBG. <https://www.mdbg.net> (23 April, 2023).
- Sroka-Grądziel, Anna. 2018. Z zagadnień leksykografii polsko-chińskiej. *Gdańskie Studia Azji Wschodniej* 13. 30–42.



- Wieger, Léon. 1965. *Chinese characters: Their origin, etymology, history, classification and signification. A thorough study from Chinese documents*. Trans. by Leo Davrout. 2nd edition, enlarged and revised according to the 4th French edition. New York: Paragon Book Reprint Corp./Dover Publications, Inc.
- Wieluch, Doman. 1936. *Słownik chińsko-polski do czytania tekstów chińskich bez przygotowania* 切字中波辭典. Katowice: (self-published).
- Wu Yuemei (吴月梅). 2010. *Obrazkowy słownik języka chińskiego* 汉语图解词典 (波兰语版). Beijing: 商务印书馆 (The Commercial Press).
- Xia Zhengnong (夏征农) & Chen Zhili (陈至立). 2009. *辞海*. 6th edition. Shanghai: 上海辞书出版社 (Shanghai Lexicographical Publishing House).
- Xu Lin (许琳), Yao Xishuang (姚喜双), et al. 2010. *Podręczny słownik języka chińskiego* 汉语800字. Trans. by Justyna Szpakowska. Beijing: 外语教学与研究出版社 (Foreign Language Teaching & Research Press).
- Yin Xiangfeng (殷相峰) (ed.). 2013. *汉语波兰语分类词典* *chińsko-polski słownik tematyczny*. Beijing: 商务印书馆 (The Commercial Press).

