Chinese Idioms: Stepping Into L2 Student’s Shoes

Mateja PETROVČIČ
University of Ljubljana, Slovenia
mateja.petrovcic@ff.uni-lj.si

Abstract

Idiomatic expressions play an important role in communication. They are widely used in both spoken and written discourse and are often one of the indicators of native-like proficiency. This paper discusses Chinese idioms chengyu in foreign language education, presents some observations regarding the scope of idioms predicted for the advanced level of Chinese, and introduces conceptual gaps between L1 and L2 as possible reasons for learners’ mistaken usages. This study argues that equipping students with skills and knowledge needed for reaching contextually appropriate use of idioms should be the main goal in the teaching/learning process and provides a set of practical suggestions as to which aspects of teaching idioms should be considered thoroughly.

Keywords: Chinese idioms; foreign language education; teaching and learning strategies; text corpora; language transfer

Povzetek

Idiomatski izrazi imajo v komunikaciji pomembno vlogo. Uporabljajo se v govornem in pisnem diskurzu, pogosto so tudi pokazatelj dobrega poznavanja jezika, ki meji na raven znanja naravnih govorcev. Članek obravnava kitajske pregovore chengyu pri poučevanju kitajščine kot tujega jezika in predstavi opažanja glede predvidenega izbora pregovorov, ki so glede na smernice poučevanja predvideni šele na najvišji jezikovni ravni, torej na ravni mojstrstva. Razprava o tem opozori na konceptualne vrzeli med L1 in L2 kot možnem izvoru nekaterih napak, ki jih tujci storimo. Članek poleg tega poudari pomen veščin in znanja, ki sta potrebna za doseganje kontekstualno ustrezne uporabe pregovorov, v sklepnih mislih pa ponudi še vrsto praktičnih predlogov o tem, čemu moramo pri poučevanju/učenju idiomatskih izrazov nameniti pozornost.

Ključne besede: kitajski pregovori; poučevanje tujih jezikov; strategije učenja in poučevanja; besedilni korpusi; jezikovni prenos
1 Introduction

Idiomatic expressions are a very fascinating part of a language. They make the language use more colorful, subtle, and precise. They are widely used in both spoken and written discourse and are considered as one of the hallmarks of native-like proficiency. Considering a wide range of situations in which idioms are used and the important role that they play in communication, there is little doubt that foreign language learners need to be familiar with their meaning and use (Vasiljevic, 2015).

The inclusion of idioms into the syllabus should therefore be an important part of teaching Chinese as a foreign language due to their constant presence in media, literary works, and daily conversations. As Wang and Luo claim (2021, p. 284), through learning idioms, students not only enrich their vocabulary, improve their language skills, and enhance the vividness of their language expression, but also better understand Chinese history and culture.

While in one’s native language idioms are typically acquired through exposure, they are difficult nuts to crack in a foreign language for several reasons. As Vasiljevic (2015) notes, even if the learners recognize the figurative use of expressions, they often lack the knowledge and the skills to disambiguate the phrase meaning in the way the native speakers may do due to their limited linguistic proficiency and vocabulary size. Moreover, many idioms are also culturally embedded and not motivated only by their lexical components, but also by the specific cultural and historical context in which they originated (Boers et al., 2004).

Even though learners of Chinese usually come across idioms at the intermediate and advanced level, this aspect of language use still causes several problems. Learners often make mistakes because they misunderstand the structural, semantical, or syntactical features of idioms (Shi, 2008).

Several previous studies have discussed the types of mistakes foreigners make and the reasons thereof (Shi, 2008; D. Xiao, 2016; Yang, 2011), some of them examined textbooks and learner’s dictionaries (Hong, 2012; S. Wang & Luo, 2021), or proposed improvements regarding teaching strategies and methods (Vasiljevic, 2015; Wang W. 王伟, 2016; D. Xiao, 2016; among others).

We agree with Vasiljevic (2015) that idiomatic language can and should be taught, whereby it is our duty as teachers to look for how the learning process can be optimized. Besides the published materials, online resources, and classroom idiom teaching activities, we hereby propose that teachers should equip students with the knowledge and skills of how to gain reliable information, encourage them to keep exploring language use to reinforce their language skills. The time available for contact hours is always limited, no matter how many classes there are scheduled in the syllabus.
To achieve this goal, we first need to understand how the present generations of students seek information, which resources they select, whether they evaluate their findings or not, and at which stage they get satisfied with answers. This study further points out the aspect which should be thoroughly considered at the advanced level of language teaching, namely the differences in the distribution of idioms across various domains.

Section 2 first reviews the terms related to idiomatic expressions in Chinese and closes with the decision to use English word *idiom* to denote Chinese *chengyu*. Section 3 focuses on idioms in foreign language education, points out some observations regarding the scope of idioms taught, and introduces some possible reasons for usage mistakes. Section 4 discusses the role of Chinese corpora and compares their strengths. Section 5 presents some major observations regarding the information-seeking process and students’ attempts to use given idioms on their own. Section 6 briefly touches on the fact that occurrences of idioms vary across domains. The paper closes with Section 7 which sums up this study’s findings.

### 2 Idiomatic expressions in Chinese

A common overarching term for idiomatic expressions in Chinese is *xiyu* 习语 or *shuyu* 熟语. In a broader sense, the term *shuyu* 熟语 is used to denote a category, and in a narrower one, it is used to refer to individual items in the category (X. Zhang, 2012, p. 20). Generally speaking, it serves as an umbrella term which includes various subcategories, i.e. common phrases *guanyongyu* 惯用语, idioms *chengyu* 成语, proverbs *yanyu* 谚语 and two-part allegorical sayings *xiehouyu* 歇后语. While some authors consider just these four groups as *shuyu* 熟语, the others include other subcategories, as well. Feng (2011) or Wang and Wang (2013), for example, also count in common sayings *suyu* 俗语, aphorisms *geyan* 格言, maxims *zhennyu* 箴言, epigrams *jingju* 警句, slang *liyu* 俚, and many other specific forms of expressions. In Xiao (2020), two of the categories are named according to the length of idiomatic expressions, namely 3-characters common expressions *sanziyu* 三字语 and 4-characters common expressions *siziyu* 四字语 as separate groups. Categories of idiomatic expressions are overlapping in many aspects, and it is impossible to draw a clear dividing line between them.

As far as the term *chengyu* 成语 is concerned, it can be understood in a broader or a narrower sense as well. In the broader sense, it is a category of set phrases that includes *suyu* 俗语 and *yanyu* 谚语, but in the narrower sense, it refers to four-character set phrases that can be traced back to written records in ancient fables, histories, and literary works (X. Zhang, 2012, p. 26). *Chengyu*, therefore, form a special part of the Chinese lexicon that has a fixed structure, conveys concise and comprehensive meanings, contains rich cultural connotations, and has profound
historical and cultural value. Nowadays, *chengyu* are still widely used in written and spoken Chinese and learners come across them at the intermediate and advanced level (Shi, 2008).¹

Most of the publications translate *chengyu* as *idioms* but other expressions are used for them as well. For example, Zhao (2015) uses the word *proverb* as the English equivalent of *chengyu*. On the other hand, even the English term *idiom* is not understood uniformly.

In the academic papers, the term *idiom* is understood in various ways, stretching from the traditional definition as a fixed multi-word phrase whose meaning cannot be predicted from the literal meanings of individual words that constitute this phrase, to more profound definitions from modern lexicographers, such as Glucksberg (2001), Grant and Bauer (2004), Moon (1998), and other linguists (X. Zhang, 2012). In this paper, we will not discuss the definitions and various interpretations of the terms *idiomatic expression*, *quadra-syllabic idiomatic expression*, *idiom*, *fixed expression*, *proverb*, and others in detail, but will use the English word *idiom* to denote Chinese *chengyu*, conventionally used set phrases, which are historically allusive in origin, often highly fixed in structure (the four-character mould), usually opaque in meaning and typically archaic in style (R. Xiao & Hu, 2016, p. 107).

### 3 Idioms in foreign language education

As Jiao et al. (2011) note, contextually appropriate use of idioms tends to impress hearers, especially if an aptly used idiom is coming from the mouth of a foreigner. Since idioms are frequently used in formal speech and higher-level written materials, they serve as a marker of proficiency level.

For a learner of Chinese as a foreign language, mastering idioms is very intriguing. This is mostly due to their meaning and figurative usage, but even their syntactical behavior is a hard nut to crack (D. Xiao, 2016).

As Xiao (2016) noted, idioms are an important part of Chinese language learning and not enough emphasis is put on them. For their complexity, idioms are usually taught no sooner than at the upper-intermediate or advanced level. This is also evident from the HSK 1-6 vocabulary lists, where almost all of the slightly more than one hundred idioms are predicted for level 6.² In other words, idioms are listed among the advanced words, ranking from #2500 to #5000. Furthermore, as provided in Chinese

---

¹ The author does not explicitly state what is meant by "intermediate and advanced level" (*zhong-gaoji jieduan* 中高级阶段), but since the results of this study are based on the data from the HSK Dynamic Composition Corpus of Beijing Language and Culture University, we assume that the approximate level of learners in question was around HSK 5 or HSK 6.

Proficiency Grading Standards for International Chinese Language Education (Center for Language Education and Cooperation, 2021), idioms are supposed to be learned even later, ranking from #5457 to #11092, which is a negatively surprising trend in standard setting.

Nevertheless, no matter how late the idioms are scheduled in the international language education, students of sinology or Chinese studies as their major should achieve the highest possible level of language proficiency during their studies, but due to the limited time and scope of the BA/MA programs, it is impossible to achieve an advanced level of language proficiency just within three/five years. Therefore, rather than omitting idioms from the curriculum or just integrating selected idioms into the courses, it is more appropriate to equip students with skills and knowledge that are needed to reach contextually appropriate use and necessary for their lifelong learning.

Zhang (1999) argues that certain syntactical errors are caused due to the lack of understanding of idioms’ internal structure and could be clarified with a better understanding of their morphological features. Shi (2008) comes to similar observations and points out that foreigners make some mistakes because they are not aware that predicative idioms\(^3\) tend to be used in certain syntactical functions, whereas nominative idioms\(^4\) tend to be used in other syntactical functions. From this point of view, explaining idioms’ structure and each character’s meaning seems to have a positive impact on learners’ understanding.

From the perspective of a non-native Chinese teacher, we need to mention another conceptual gap that arises when students encounter unknown idioms. Several studies have pointed to syntactical errors in general, regardless of the learners’ L1 and problems that the language transfer might cause.

Although the intermediate learners should already have a sufficient understanding of Chinese notions of syntactical functions, the discrepancy in grammar and categories between L1 and L2 may lead to additional problems. For example, while the phrase on the desk in example (1) is considered a subject in Chinese, it is perceived as an adverbial phrase of place in Slovene.

(1) 桌子上 放着 一本书。

zhuōzi shàng fàngzhe yī běn shū

‘There is a book on the table.’ (Liu et al., 2004, p. 452)

---

\(^3\) Chinese: weicixing chengyu 谓词性成语

\(^4\) Chinese: ticixing chengyu 体词性成语
Similarly, it is not easily comprehensible for a Slovene learner of Chinese that 308 in example (2) below is an object and not a predicate complement.

(2) 他们住的房间号 是 308。

tāmen zhù de fángjiān hào shì sān líng bā

The number of their apartment is 308. (Liu et al., 2004, p. 463)

Seen from a different perspective, information that an idiom is used for example as a subject or object might not be as informative as expected.

Moreover, it is difficult to imagine how an idiom, explained in a form of a sentence, is used in a specific syntactical function. For example, an idiom meaning the losses outweigh the gains can be used as a predicate, a modifier of a noun, or a verbal complement, as shown in examples (3)-(5).

(3) 学习要循序渐进。当前的课业没有学好, 却把主要精力放在钻研那些高精尖的学问上, 真是得不偿失。（谓语）

Xuéxí yào xúnxùjiànjìn. Dāngqián de kèyè méiyǒu xuéhǎo, què bǎ zhǔyào jīnglì fàng zài zuānyǎn nàxiē gāo jīng jiān de xuéwèn shàng, zhēnshì débùchángshī. (wèiyǔ)

Learning should be done in a gradual manner. Not mastering the current coursework, yet devoting energy to the study of high-level knowledge, really brings more losses than gains. (predicate) (Pan, 2002, p. 166)

(4) 她正为干了这件得不偿失的事而懊悔。（定语）

Tā zhèng wéi gān le zhè jiàn débùchángshī de shì ér àohuǐ. (dìngyǔ)

She is regretting that she has done something that is not worth the loss. (modifier) (Pan, 2002, p. 166)

(5) 虽然得了一些实惠却失去了人心, 真是干得得不偿失啊! （补语）

Suīrán déle yīxiē shíhuì què shìqù rénxīn, zhēnshì gàn dé débùchángshī a! (bǔyǔ)

Although I got some benefits, I lost my heart. Doing so was not worth the loss. (complement) (Pan, 2002, p. 166)

In the era of smartphones, online resources, and one-click solutions, it is expected that the learners first seize for online dictionaries or google the answers, so let us
further look at the definition and top five examples provided by the Chinese-English online dictionary LINE Dict (https://dict.naver.com/linedict/), as one of the popular resources among students.  

(6) 得不偿失 débùchángshī
Definition: not be worth the effort
Usage examples:

得不偿失。
The gain is outweighed by the loss.
勿做得不偿失的事。
Do not pay too dear for your whistle.
那些发现得不偿失。
Those discoveries are not worth the candle.
北半球秋季变暖得不偿失
Losing more than we gain from autumn warming in the north.
寓意就是犯罪得不偿失。
And the moral is that crime doesn’t pay.

LINE Dict provides information that this idiom is classified as a HSK6 expression, with an English translation and several usage examples, whereby it does not mark idioms’ syntactic functions. Readers are supposed to be able to understand and independently create the appropriate usage without explicit instructions.

Wang and Luo (2021) note that displaying example sentences with the targeted idiom is considered effective, whereby the selection of examples plays a demonstrative role in the usage of idioms. Compared to the printed materials, online resources have the advantage of having unlimited space for example sentences, and can therefore provide better insight into the language use.

Several authors have already stressed that teaching idioms should give priority to high-frequency idioms, as well as their high-frequency meanings and grammatical functions (Hong, 2012; Jiao et al., 2011; S. Wang & Luo, 2021; among others). To retrieve frequency data—not to mention other information—language corpora are indispensable tools. The following section, therefore, discusses Chinese corpora in relation to idioms.

---

5 Before 2014, this dictionary was known as Nciku.
4 Idioms in Chinese corpora

As far as idioms are concerned, not all Chinese corpora are equally appropriate to use. Moreover, to make full use of the materials and to ensure a successful application of corpora in classroom instruction, specialized training in corpus analysis for teachers and students would also be necessary first (Ištvánová, 2021, p. 132). The following subsections present the pros and cons of the BCC corpus, the CCL corpus, corpora in Sketch Engine, and the Hanku corpus.

4.1 Idioms in the BCC corpus

The BCC corpus (http://bcc.blcu.edu.cn/) is the ideal choice for the initial analysis of idioms because it has got a special tag just for them (Xun et al., 2016, p. 107). To get a list of the most frequently used idioms, we ran a simple query "i" in the balanced corpus domain and then sorted the results by frequency. The pop-up window provides information about how many different idioms there are in the selected section, and what are the absolute frequencies of each idiom, as shown in Figure 1.

![Figure 1: List of the most frequent idioms in the multi-domain part of the BCC corpus with their absolute frequencies](image)

After extracting the top 1000 idioms with the highest frequency, we compared this frequency list with the HSK idioms. The results in Figure 2 below show that idioms aimed at foreign language education do not fully overlap with the upper tenth of the most frequently used idioms but are rather distributed as follows. Among the top hundred idioms, 37 items are listed as HSK vocabulary, whereas 63 of them are not on part of the HSK wordlists. Among idioms ranked between #101 and #200, 29 are scheduled for the HSK level 6, whereas 71 of them are not considered relevant, and so on.
An alternative option for idiom extraction would be the CCL corpus of Peking University (Zhan et al., 2019). In the CCL corpus, "i" stands for idioms as well, but the search is limited to the moshi chaxun 模式查询 (pattern query), which requires some specific context. For example, it is possible to search for a string of zhen 真 (really) + idiom “真(i,=4)”, but not for a list of idioms without any context.

Although these two corpora are very convenient for some basic information on Chinese idioms, their functions are limited to relatively simple queries and do not support the corpus query language to create flexible queries.

### 4.2 Idioms in SkE corpora

For more refined queries with corpus query language (CQL), Sketch Engine is a better choice. This tool returns cleaner and more precise results, but it is not ideal to extract idioms as such in the first place.

In the part-of-speech (POS) tagged corpora, there are usually no special tags for idioms, although it would be useful and convenient for Chinese since they are recognized as special fixed expressions and phraseological units. On the other hand, since idioms may be used in various syntactical functions, they are often treated as one token and are assigned the corresponding tags. For example, in corpora that are POS annotated by the Stanford Log-linear Part-Of-Speech Tagger using Chinese Penn Treebank, the majority of idioms are tagged as verbs, as shown in Table 1.
Table 1: Structure of the 1,000 idioms from BCC corpus according to their POS tags in Chinese Web 2017 (zhTenTen17) Simplified

<table>
<thead>
<tr>
<th>Tag</th>
<th>Absolute frequency</th>
<th>Relative frequency</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>VV</td>
<td>8,380,349</td>
<td>505.05</td>
<td>83.80%</td>
</tr>
<tr>
<td>NN</td>
<td>930,931</td>
<td>56.10</td>
<td>9.31%</td>
</tr>
<tr>
<td>AD</td>
<td>405,127</td>
<td>24.42</td>
<td>4.05%</td>
</tr>
<tr>
<td>VA</td>
<td>204,299</td>
<td>12.31</td>
<td>2.04%</td>
</tr>
<tr>
<td>CD</td>
<td>56,622</td>
<td>3.41</td>
<td>0.57%</td>
</tr>
<tr>
<td>Others (JJ, NR, NT, M, OD)</td>
<td>22,672</td>
<td>1.37</td>
<td>0.23%</td>
</tr>
</tbody>
</table>

Some idioms are consistently tagged with only one POS, for example, qiánsuǒwèiyōu 前所未有 (hitherto unknown; such as never previously existed; unprecedented) as a verb; sìmiànbāfāng 四面八方 (from all sides; all around; all directions; all quarters) as a noun; yǔcǐtóngshí 与此同时 (at the same time; in the meantime; meanwhile) as an adjective, etc. Some idioms, on the other hand, have been assigned various POS tags. For example, the idiom shíshìqiúshì 实事求是 (seek the truth from facts; base on facts; be practical and realistic) is treated as a verb (VV), adverb (AD), predicative adjective (VA), or a noun (NN).

Even though 83.8% of all idioms are tagged as verbs, the query of four-character-verbs does not return only idioms. The proportion of noise increases if we expand the query to include the other tags from Table 1.

Apart from Chinese Web 2017, Sketch Engine offers several other Chinese corpora that are annotated with either Chinese Penn Treebank part-of-speech tagset, Chinese Symbols part-of-speech tagset or Chinese NEUCSP part-of-speech tagset.

Chinese Penn Treebank part-of-speech tagset,\(^6\) which is available in Chinese corpora annotated with Stanford taggers, does not have any special tag for idioms. This is also the case in Chinese Symbols part-of-speech tagset,\(^7\) which is available in Chinese corpora annotated with tagging tool developed by Institute of Information Science and CKIP group in Academia Sinica, for example, Chinese GigaWord 2 Corpus.

On the other hand, Chinese corpora annotated by the NEUCSP tagging tool developed by the Natural Language Processing Group at Northeastern University, China, use Chinese NEUCSP part-of-speech tagset,\(^8\) which is in many respects different

---


from the former two. In this tagset, idioms are labeled as ‘habitual language’ (*xiyongyu* 习用语). This tag is also used for phrases, such as *méi shénme* 没什么 (it’s nothing), *shì de* 是的 (that’s right), *shíshí shàng* 事实上 (in fact), *méi guanxi* 没关系 (it doesn’t matter; never mind), etc.

For the analysis of syntactical features of idioms, the Chinese Corpus Hanku could be used, as well, because it fully supports CQL (Gajdoš et al., 2016). Since it uses NoSketch Engine, i.e. an open-source version of the Sketch Engine corpus manager, it requires a more profound knowledge of the corpus query language to get the results that are otherwise available in the word sketch tool of Sketch Engine.

We thus suggest a combination of corpora to achieve optimal results. In our case this means the BCC corpus to extract idioms, then the Chinese Web 2017 for the syntactical analysis of the selected idioms, and the BCC corpus again for the part presented in Section 6 below.

## 5 Unveiling idiom’s meaning and usage

We agree with the view that learning is not committing a set of facts to memory, and teachers’ role is not only to transmit the required knowledge, but to guide students and give them the ability to use resources to find, evaluate, and apply information (Lujan & DiCarlo, 2006, p. 17). Developing life-long skills such as critical thinking, problem-solving, and communication is much more useful in the long term. The time available in the classes is always limited, no matter how many hours there are scheduled in the syllabus.

To achieve this goal, we first need to understand how the present generations of students are seeking information, which resources they select, whether they evaluate their findings or not, and at which stage they are satisfied with answers.

As part of their undergraduate studies, students of the sinology program at the University of Ljubljana are asked to retrieve information about given idioms and write a report describing their ‘discovery path’. This is an open-ended survey conducted on a yearly basis that provides respondents with the opportunity to explore information retrieval in their personal way. Every student selects a package of five different idioms and is asked to do whatever they want to understand the meaning and usage of the given ‘items’. In the concluding part of these mini-inquiries, students are asked to present a few examples that they consider useful, explain their decision, sketch a situation to be suitable for that idiom, and form a sentence with it.

---

9 The provided idioms are those with the highest frequency of occurrences, regardless of whether they are part of the HSK vocabulary or not.
The observations retrieved from these assessments provide valuable insights into the habits and preferences of current generations regarding information retrieval. This paper does not aim to discuss all the findings in detail, but rather highlights some noteworthy characteristics and common features.

First, although having easily determined that the ‘task items’ were idiomatic expressions and being fully aware that figurative meanings are not a simple sum of the meanings of the component words, all students devoted their time and energy to decompose idioms and tried to understand the meaning of individual characters. This supports the views expressed by Matlock and Heredia (2002) that non-experienced foreign language learners first establish direct connections between literal and nonliteral meanings of figurative expressions. Moreover, when processing an L2 idiom, L2 learners first attempt to make sense of it by translating it literally into L1.

Trying to understand the individual characters and relations among them seems to be a desirable thing in Chinese. Namely, the syntactical features of idioms are said to be closely related to the idiom’s internal structure. According to Shi (2008), predicative idioms tend to be used as predicates, attributives, adverbial adjuncts, and complements. On the other hand, nominative idioms would normally be used as subjects or objects.

Second, students needed fewer efforts for the so-called decomposable or compositional idioms, where the figurative meanings were relatively closely related to the literal meanings of their constituent words (e.g. qiánsuǒwèiyǒu 前所未有 unprecedented, lit. before-not.existing). The results have shown that even at the pre-intermediate level, students do not encounter great difficulties understanding them. Based on these observations, we claim that certain high-frequency compositional idioms may already be taught at the pre-intermediate level and do not need to be postponed to the advanced level.

Similarly, students devoted more time to non-decomposable or non-compositional idioms, probably since the meanings of individual characters seem to be unrelated to their compositional analysis (e.g. tuōyǐng’érchū 脱颖而出 escape from-an.awn.of.wheat-in.such.manner-go.out; literal meaning refers to the fibers that often stick through a sack of wheat; fig. talent being fully exposed, talent revealing itself, distinguish oneself, rise above others). If they found a consistent equivalent in L1, they still managed to express the correct idea, whereas the resulting sentences often deviated from the expected usage. See example (9) below.

On the other hand, if the idiom resulted in a series of various translations, such as for example (6) above, mistakes were very frequent even if the idiom itself was a compositional one.

Third, the results have also shown that students did not give priority to the dictionaries/webpages with English translations when searching for usage examples. If
a translation was provided, they often re-checked it and compared it with other possible translations at various translation services. Surprisingly, approximately two-thirds of students seized AI-powered translation and dictionary tools, such as ReversoContext (https://context.reverso.net/) or Linguee (https://www.linguee.com/).

Forth, while focusing on the meaning of idioms themselves, students often neglected the provided context. This confirms the previous observations that learners often lack the skills to take advantage of contextual clues (Vasiljevic, 2015, p. 3). Although the selected examples shared some obvious features, students failed to notice them and to use them accordingly in their own production. For example, returning back to the idiom tuōyǐng’érchū 脱颖而出 ‘distinguish oneself’, numerous examples point to the collocations with the subject rěncái 人才 ‘talent’, prepositional phrase zài/cóng … zhōng 在/从……中 ‘among …’, presence of modal verb néng 能 ‘be able to’, etc., but none of these clues were applied in the created example.10 Speaking of collocates, none of the students came to the idea to make queries in language corpora. A follow-up discussion revealed that they were not even aware of language corpora in their L1.

Finally, the analysis has shown that idioms with similar syntactical behavior in L1 and L2 are easier to learn and use properly than those that differ in this respect. This aspect of idiom research has been ignored in scholarly papers.

For example, the idiom zhòngsuǒzhōuzhī 众所周知 is explained as widely known; as everyone knows; as is known to all, and it is usually used independently at the beginning of a sentence. As such, it is in our case similar to the corresponding expressions in students’ native language.11 It is therefore not very difficult to create well-formed sentences, as shown in examples (7)-(8) below.

(7) 众所周知，美国人口大部分居住于东半部。

Zhòngsuǒzhōuzhī, Měiguó rénkǒu dà bùfèn jūzhù yú dōng bàn bù.
‘As everyone knows, the majority of the American population lives in the eastern half’ (A phrase used at the beginning of a sentence)

(8) 吸烟可能致癌是个众所周知的事实。

Xīyān kěnéng zhì’ái shì gè zhòngsuǒzhōuzhī de shìshí.
‘It is a well-known fact that smoking may cause cancer.’ (modifier of a noun)

10 It should be noted that there were two opposite approaches in forming sentences. The safe-players tended to make just minor adjustments to the existing examples, which resulted in correct use, while the risk-takers tried to be creative and applied idioms to their own contexts.

11 Since there are no Chinese-Slovene dictionaries of idioms available, and students are used to using Chinese-English materials, we neglect the Slovene translation here. Nevertheless, some other idioms may have better Chinese-Slovene equivalents.
The reasons for such smooth application might be attributed to the compositional nature of this idiom, however, we claim that the influence of L1 grammatical features and discrepancy in grammatical categories play an important role. Even the examples with compositional idioms reflected this issue. For instance, sequences in examples (9) and (10) below show at which part of the ‘discovery path’ the usage deviated from expected.

(9) **Stage 1:** Literal meaning of 脱颖而出 is confusing but its figurative meaning sounds understandable: *talent revealing itself, to rise above others, to distinguish oneself*

**Stage 2:** mapping English equivalents to L1:
izkazati se, izstopati

**Stage 3:** example sentences in Chinese can be perfectly translated to Slovene with the verb izstopati

**Stage 4:** production of example in Slovene:
*On vedno bolj izstopa.* ‘He is increasingly rising above others.’ (perfectly acceptable)

**Stage 5:** translation of *he is increasingly + rising above others* in Chinese:
他越来越 + 脱颖而出

Comments: We cannot claim that this sentence is wrong, because such instances are available in corpora but are extremely rare (relative frequency less than 0.01). The reason is that the expression yuelaiyue normally requires an adjective and not a verb(al phrase).\(^\text{12}\)

(10) **Stage 1:** 兴致勃勃 means:
*high-spirited (adjective) (source: https://en.bab.la/)*
alternative information: full of zest

**Stage 2:** mapping English equivalents to L1:
poln žara, srečen, dobro razpoložen

**Stage 3:** example sentences in Chinese can be perfectly translated to Slovene with these adjectival phrases

**Stage 4:** production of example in Slovene:
*Mi smo vsi srečni/polni žara, ker bo kmalu konec leta.*
‘We’ll finish classes soon, so we’re full of zest.’

\(^{12}\) Collocation details reveal that yuelaiyue+VA appears with RF 132.55, whereas yuelaiyue+VV appears with RF 24.76.
Stage 5: translation of *we + full of zest* in Chinese with the following explanation: this idiom is used as an adjective. When adjectives are used as predicates, they usually require *hen* (*N+hen+Adj*).

快要停课了，我们 + 很 + 兴致勃勃。
Kuàiyào tíngkè le, wǒmen + hěn + xìngzhìbóbó.

Comments: This student was not mistaken in her conclusions that adjectives can function as predicates. Moreover, based on the explicit information from the selected online dictionary and informal confirmation in L1 that this is an adjectival expression, her way of thinking was correct. However, placing more focus on the idiom’s internal structure and near context would raise some doubts in the suggested example.

As the next step, idioms that are explained in the form of a longer phrase or sentence, are more difficult to be further used in a specific syntactical function. For example, consider the idiom *yìngyùn’érshēng* 应运而生 with the explanation to *arise or emerge as times demand; to arise according to objective demands; emerge as the times require; arise at an opportune time*. While forming sentences with this idiom as a predicate is still feasible due to the verb arise/emerge as the anchoring headword, it is more difficult to use it as a modifier of a noun. Consider examples (11)–(12) below.

(11) 随着网络的出现，网络语言也应运而生。
Suízhe wǎngluò de chūxiàn, wǎngluò yǔyán yě yìngyùn’érshēng.
‘With the advent of the Internet, network language also emerged as the times require. >> With the advent of the Internet, network language also came into being.’ (predicate)

(12) 慕课作为网络自主时代里应运而生的产物，是一种新型的教学资源。
Mùkè zuòwéi wǎngluò zìzhǔ shídài lǐ yìngyùn’érshēng de chǎnwù, shì yī zhǒng xīnxíng de jiàoxué zīyuán.
‘As a new product (that emerged as the times require) in the era of network autonomy, MOOCs are a new type of teaching resource.’ (modifier of a noun)

The most difficult idioms to grasp have proven to belong to the group of idioms that are culturally embedded. These are idioms whose meanings are not motivated only by their lexical components but also by the specific cultural and historical context in which they originated (Boers, Demecheleer & Eyckmans 2004 in Vasiljevic, 2015). For example, consider the explanation and user guidelines for the idiom *fēngshēng-hèlì* 风声鹤唳 in *A Chinese-English Dictionary of Chinese Idioms* (CEDCI).
In 383, Fu Jian's army of 800,000 strong were badly beaten by the troops of the Eastern Jin, totaling 100,000. On their way back home, the fleeing army mistook the sound of the wind and the cry of cranes as the coming of the pursuing Eastern Jin's army. (Met) To be full of suspicion and fear. Derogatory: mainly describes someone who is extremely frightened and scared. Used as 1) predicate; 2) complement (often cooperates with verbs like 闹 and 搞); 3) attributive (often modifies 程度, 地步 and 情况 as well as nouns referring to people); 4) adverbial adjunct; and 5) object (often used in cooperation with the predicate verbs 感到 and 听). May be used together with 草木皆兵 (Pan, 2002, p. 219).

Examples representing such syntactical functions are:

(13) 中国对的攻势非常猛烈, 阿根廷队门前风声鹤唳。
Zhōngguó duì de gōngshì fēicháng měngliè, Āgēntíng duì mén qián fēngshēnghèlì.
‘The Chinese team’s offensive was very fierce, and Argentina in front of the goal was frightened.’
(predicate of subject Argentina)

(14) 整个东北被他弄得风声鹤唳。
Zhěnggè dōngběi bèi tā nòng dé fēngshēnghèlì.
‘The whole northeast was shaken by him.’
(complement of the verb to make, lit. has been made shaken)

(15) 资本市场有点风声鹤唳的感觉，股市、债市都在跌。
Zīběn shìchǎng yǒudiǎn fēngshēnghèlì de gǎnjué, gǔshì, zhài shì dōu zài diē.
‘The capital market feels a bit jittery. The stock market and bond market are both falling.’
(attributive/modifier of a noun feeling)

(16) 如果你总是风声鹤唳地看过去，生活只好让你四面楚歌。
Rúguǒ nǐ zǒngshì fēngshēnghèlì de kàn guòqù, shēnghuò zhǐhǎo ràng nǐ sìmiànxūngē.
‘If you always look jittery at the past, life will treat you surrounded by enemies on all sides.’
(adverbial adjunct to the predicate to look)

(17) 犯了罪的人对每一个声音都感到风声鹤唳。
Fànle zuì de rén duì měi yīgè shēngyīn dōu gǎndào fēngshēnghèlì.
‘Those who have committed a crime feel very nervous about every voice.’
(object of verb to feel)
 Needless to say, for culturally embedded idioms, it is very difficult to determine appropriate contexts, even though the dictionary tells us that this idiom mainly describes someone who is extremely frightened. As observed from examples (13)-(17) retrieved from corpora, this expression can describe *fearful atmosphere* in sports, market economy, among a certain population, etc. Without detailed queries L2 learners cannot understand and use such idioms properly.

At this point, we must remind ourselves that contemporary generations first seize for online resources that return results in no time. Therefore, providing guidelines of how to refine their searches would be of great importance for their life-long learning. On the other hand, students should also be aware that specialized materials such as dictionaries of Chinese idioms capture the essence of idioms and are important supplementary reference works.

We agree with Wang and Luo (2021) that it is still a challenge for Chinese learners and teachers to answer the questions of how to effectively master the meaning and usage of idioms, however, we disagree with their opinion that there is a lack of idiom learning dictionaries for CFL learners (S. Wang & Luo, 2021, p. 285). Although compiled and published back in 2002, Pan’s *A Chinese-English Dictionary of Chinese Idioms* is still a relatively good publication for learning Chinese as a foreign language. This book represents idioms relatively accurate, explains their literal and figurative meanings, which are accompanied with short but rich syntactical information, followed by up to four examples of use in various syntactical functions. In our supplementary research, we have evaluated the reliability of provided information and given examples concerning frequency of occurrences in Chinese corpora. The results were quite satisfying. Another reference work worth advising is Jiao, Kubler and Zhang’s *500 Common Chinese Idioms: An Annotated Frequency Dictionary* (Jiao et al., 2011). This book is grounded in corpus research, focused on learner usage, accompanied by not numerous yet well chosen example sentences, explanations, and usage notes.

A selection of various online and printed materials in combination with the corpora would therefore be an invaluable aid to both comprehension and production.

6 Further notes on idioms across various domains

As Feng (2011) states, idioms often appear in many types of texts, from literary texts, political texts, or scientific texts, to people's daily conversations.

Therefore, we conducted further follow-up queries across various domains in the BCC corpus. Apart from the multi-domain (*duōlǐnyù 多领域*), which was designed as

---

13 One respondent in the generation 2020/2021 frankly admitted that her first intention was to dig into a comprehensive dictionary in her private library, however, she lost her nerves because of the slow pace of progress.
a balanced corpus, the BCC corpus is further composed of the following four sub-
corpora of contemporary Chinese: literature (wénxué 文学), press (bàokān 报刊),
Weibo (Wēibó 微博) and science and technology (kē-ji 科技). Weibo represents
colloquial Chinese, whereas the section on science and technology reflects the
academic writing style. All five parts are independent and do not overlap.

The analysis of idiom distribution over the five domains has shown notable
differences across the domains. As demonstrated in Table 2, the most prominent
difference exists between literary and academic texts. There is just a 37.5% overlap in
the list of 1,000 the most frequent idioms per field. In other words, 62.5% of idioms
that are frequently used in literary texts are not among the top thousand idioms in
academic texts and vice versa. Table 2 below further demonstrates the ratio of idioms
that are used in all registers compared.

Table 2: Ratio of common idioms across fields (range: top 1,000 most frequent idioms per field)

<table>
<thead>
<tr>
<th>Field</th>
<th>Literature</th>
<th>Press</th>
<th>Colloquial Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Press</td>
<td>401 (40.1%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Colloquial Chinese</td>
<td>518 (51.8%)</td>
<td>449 (44.9%)</td>
<td>-</td>
</tr>
<tr>
<td>Academic Chinese</td>
<td>375 (37.5%)</td>
<td>676 (67.6%)</td>
<td>455 (45.5%)</td>
</tr>
</tbody>
</table>

For example, the following idioms are expected to be found in the literary texts,
but not very likely in the academic papers: yīběn-zhèngjīng 一本正经 (be poker-faced;
humorless; in all seriousness), yīyán-bùfā 一言不发 (not to utter a single word; be as
silent as the grave), xīnmǎn-yìzú 心满意足 (to one’s heart’s content; be fully satisfied and content) and
luànqī-bāzāo 乱七八糟 (be out of order; all a hideous mess; be in wild disorder), among
others. Vice versa, the following idioms are very likely to appear in the academic papers,
but not in the literary works: xíngzhīyǒuxiào 行之有效 (effective), jiānkǔfèndòu 艰苦
奋斗 (work hard and perseveringly), yīndìzhìyí 因地制宜 (suit one’s measures to local
conditions), xiāngfǔxiāngchéng 相辅相成 (be inseparably interconnected) and
tuōyǐng’érchū 脱颖而出 (to stand out), among others.

We cannot claim that certain idioms appear in just one kind of text, but there is a
stronger or weaker tendency to use them in various domains, as further indicated in
Table 3.
Table 3: Number of occurrences in literary texts versus academic texts

<table>
<thead>
<tr>
<th>Idioms frequently used in literature, but not among the first 1,000 most frequent idioms in the academic texts.</th>
<th>Literature</th>
<th>Academic Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>一本正经</td>
<td>1,386</td>
<td>401</td>
</tr>
<tr>
<td>一言不发</td>
<td>1,339</td>
<td>304</td>
</tr>
<tr>
<td>心不在焉</td>
<td>1,215</td>
<td>364</td>
</tr>
<tr>
<td>心满意足</td>
<td>1,209</td>
<td>491</td>
</tr>
<tr>
<td>乱七八糟</td>
<td>1,204</td>
<td>544</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Idioms frequently used in academic texts, but not among the first 1,000 most frequent idioms in the literary texts</th>
<th>Literature</th>
<th>Academic Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>行之有效</td>
<td>57</td>
<td>15,732</td>
</tr>
<tr>
<td>艰苦奋斗</td>
<td>36</td>
<td>15,202</td>
</tr>
<tr>
<td>因地制宜</td>
<td>16</td>
<td>12,965</td>
</tr>
<tr>
<td>相辅相成</td>
<td>56</td>
<td>9,876</td>
</tr>
<tr>
<td>脱颖而出</td>
<td>93</td>
<td>8,495</td>
</tr>
</tbody>
</table>

Certain idioms are therefore more typical for some domains than others. For example, data from BCC have shown that the idiom *jiānkūfèndòu* 艰苦奋斗 (to struggle arduously; work hard and perseveringly) appeared 21,840 times per billion words in the texts from the press, but only 12 times in the literary texts. Similarly, *zhòngsuǒzhōuzhī* 众所周知 (as everyone knows) was used 6,473 times in the academic texts, but only 136 times in the literary texts. The idiom *luànqībāzāo* 乱七八糟 (be out of order; all a hideous mess) tends to be used the most in spoken language with 5,734 occurrences and the least in the academic texts with only 181 occurrences per billion words.

Therefore, when providing information about language use and instructing upper-intermediate to advanced learners what to pay attention to, we suggest it is worth considering domains, as well.

7 Conclusions

Idiomatic expressions play an important role in communication. They are widely used in both spoken and written discourse and are often one of the hallmarks of native-like proficiency. Considering the role of idioms and their vivid use, it is obvious that they should have a special place in second language acquisition, as well.

The time available for studies is always limited, therefore it is important to make the best use of it. For a learner of Chinese as a foreign language, mastering idioms is very intriguing, and according to HSK, they are usually taught after reaching upper-intermediate or advanced level. Our study has shown that there is no need to associate them with insurmountable difficulties because even at the pre-intermediate level, students are capable of proper perception and partly accurate production.

From the above, we suggest familiarizing students with idioms no later than at the intermediate level, whereby it is important to keep an eye on the main goal. That is to
say that teachers should try to equip students with the skills and knowledge needed for proper language use and independent lifelong learning. Needless to say, introducing idioms just as amazing and amusing stories (chengyu gushi 成语故事) is not enough to understand how these idiomatic expressions are used in the context.

As several previous studies have already pointed out, idiom teaching should pay attention to several aspects, from the role of frequency in either idiom selection, the hierarchy of meanings, grammatical functions or collocations, to the importance of their internal structure. Besides that, Hong (2012) emphasized the importance of detailed definitions and exemplifications with typical use scenarios in the textbooks. We believe that online resources should pay attention to this feature, as well, and list the good dictionary examples with extended contexts as their top results. Online resources are, unlike printed materials, unlimited in space and therefore more flexible for adding numerous additional examples. AI-powered translation and dictionary tools extract the passages from real-life contexts and cover a wide range of registers of speech. A higher degree of exposure to various examples is beneficial for developing learners’ yuyangan (language feel/intuitive competence), and thanks to user-friendly display options, there is no need to limit results to a set of foreigners’ tailored vocabulary. This, however, does not demote the importance of printed reference works.

Apart from obtaining information from the above-mentioned resources, teachers should foster students to verify their assumptions and refine their new insights in corpora, as well. As our survey has shown, this aspect of information retrieval is still very weak. As Ištvánová (2021, p. 132) noted, the direct use of corpora enhances the process of second language acquisition, however, specialized training in corpus analysis for teachers and students is necessary to ensure its successful application.

Despite all the advances of modern technologies, learners should raise their awareness of potential conceptual gaps between L1 and L2 that might lead to usage mistakes. These pitfalls are inherently related to the specifics of L1, therefore more local studies are needed to examine learning difficulties in relation to different L1 backgrounds. Moreover, getting acquainted with Chinese through English first might cause additional misunderstandings because of this additional step.

Acknowledgments

The research for this paper was supported by the Slovenian Research Agency (ARRS) in the framework of the research core funding Asian Languages and Cultures (P6-0243).
References


Xiao, B. 肖冰. (2020). *Chengyu, yanyu, suyu, guanyongyu, xiehouyu, sanziyu, siziyu 成语, 谚语, 俗语, 惯用语, 故事语, 三字语, 四字语 (Idioms, proverbs, common sayings, common phrases, allegorical sayings, common expressions (3 characters), common expressions (4 characters)).* Xin Ya chubanshe.


