Phraseoculture in the construction of the corpus of the DiCoP: The treatment of the phraseographic microstructure

Lian CHEN

1 CY Cergy Paris University, France - LT2D laboratory (Jean Pruvost Centre)
INALCO, France - PLIDAM laboratory
loselychen@gmail.com

Abstract. This article illustrates the need to introduce phraseocultural information into specialized bilingual (or even multilingual) dictionaries, from the perspective of intercultural communication. In this sense, we present the Dictionary and Corpus of Phraseology (DiCoP) project, which is currently a computerized resource for French–Chinese and Chinese–French phraseology and will eventually be multilingual. This innovative project in the field of phraseomatics (informatics phraseology/phraseography) affects many linguistics and computer fields, such as contrastive phraseoculture, metaphraseography, phraseography, phraseotraductology, and natural language processing (NLP). In particular, we wish to demonstrate the importance of phraseoculture in computerized dictionary data and in phraseography.

This paper focuses on how the DiCoP electronic dictionary reflects the phraseocultural information of the phraseological units in their microstructure. It uses examples of such microstructure to illustrate their feasibility.

Keywords: digital phraseography, phraseoculture, corpus linguistics

1 Theoretical framework: Phraseoculturology and contrastive phraseology

A phraseological unit (PUs), or 熟语, is a “[…] polylexical sequence consisting of two or more categorically related, contiguous or non-contiguous graphic [lexies]” (Bolly 2011: 28). It is characterized linguistically by the following:

(i) a certain degree of syntactic fixidness (blocking of transformational properties and unalterable constituent order); and/or
(ii) a certain degree of semantic fixidness (at least partial non-compositionalcy); and/or
(iii) a certain degree of lexical fixidness (paradigmatic restriction); and/or
(iv) a constraint on use in a communication situation. (Bolly 2011: 28)

PUs are “the embodiment of a country’s culture.”¹ The study of this specific idiomatic cultural phenomenon in phraseology is called “linguo-culturology” (Teliya 1996;…

¹ 每一条熟语，都是民族文化的化身。(Yao 2013: 157).
Dobrovol’skij 1998; Szerszunowicz 2010, etc.), “culture bound” (Szerszunowicz 2009), or “shúyǔwénhuà熟语文化学” (Wang 2012). Inspired by R. Galisson’s (1988) “lexiculture” and to address this particular phenomenon of phraseology and comparative linguistics, we proposed the concept of “phraseoculturology” (L. Chen 2022b). This sub-discipline is characterized by a specific methodology and theory within the framework of the study of the national, cultural particularities of PUs. This new phraseocultural perspective necessitates a more attentive treatment of fixed expressions in bilingual dictionaries, which often neglect this aspect.

Applied contrastive phraseology responds to the interests and needs that arise when translating, learning a language, or developing a bi- or multilingual dictionary. The analysis and comparison of PUs from one language to another leads us to distinguish three types of relationships. The following examples concern French and Chinese idiomatic expressions (IEs)

1) “Perfect equivalence” signifies an identical nature, meaning, syntactic structure, and keyword(s) in two expressions from different languages. For example, the French IE *sans queue ni tête* (negation+tail+negation+head) has a perfect equivalent in the Chinese IE *(or chéngyǔ心高气傲)* mèitóu-méiwěi 没头没尾 (negation+head+negation+tail).

2) “Partial equivalence” occurs when the expressions have the same “fixed” nature (proverbs, IEs, etc.) in both or only one of the languages, without necessarily being IEs. For example, the expression *avoir la grosse tête* (have a big/swelled head) is the partial equivalent of *xīngāo-qiǎo* 心高气傲 (heart+high+air+proud); it is an IE in both languages, but the keywords differ. Similarly, the French proverb *bouche de miel, cœur de fiel* (mouth of honey, heart of gall) is the partial equivalent of the Chinese IE *fókǒu-shéxīn 佛口蛇心* (Buddha+mouth+snake+heart), “words of a Buddha and heart of a snake.”

3) Finally, the most difficult translation/transmission concerns “non-equivalence.” Fixed expressions (FEs) belong to this category of figures of speech that are rarely translated without loss, or that may remain misunderstood outside the language and culture of origin. Non-equivalence contains two aspects: a) “Semantic vacancy,”

---

2 An IE is a polylexical sequence, semantically non-compositional (its global meaning is not always deducible from the meaning of the elements that formally compose it), syntactically fixed or fixed by usage (it does not always submit to the combinatorial rules which govern free syntax), and loaded with cultural implicits (L. Chen 2021: 123).

3 Here, the term quasi-equivalence would be more appropriate. It is difficult to find two units identical in all respects, because units are “bound to differ, at least slightly, in some parameters” (Fiedler 2007: 118).

4 French IEs correspond to Chinese *chéngyǔ* (polylexical sequences, non-compositional, syntactic blocking); both are characterized by a high degree of fixedness. The *chéngyǔ* has a quadrisyllabic, fixed basic form (Sabban [1978] 1980; Henry 2016). It is often an expression from classical literature used in modern Chinese as a compound lexic. Its style is elegant and concise, and it frequently contains strong allusive content (Ma 1978; Shi 1979; Liu 1985; Wen 2006, etc.). Even if modern creations also exist, these expressions are generally issued from a tradition—often oral and popular in French, more often written and literary in Chinese—and evoke daily realities in a metaphorical, very pictorial, and picturesque way.

5 In our examples, we reproduce the pinyin transcription (Latin alphabet transcription of Mandarin Chinese pronunciation), the sinograms accompanied by their word-for-word translation, a literal translation, their figurative or implicit meaning, and an equivalent in French if possible.
which has a specific cultural dimension that cannot be transposed from one language to another, is explained by the specificities of each culture: customs, historical references (e.g., avoir les dents du bonheur; to have lucky teeth), toponyms (e.g., oies du Capitole; geese of the Capitol), literary sources (e.g., n’avoir pas froid aux yeux; to not be faint hearted), etc. b) “Lexical vacancy,” wherein the words are not translatable from one language to another or the concept does not exist: in one language, there is a lexical vacancy, while in the other, the word has a specific cultural connotation (e.g., dùzhànáotóu 独占鳌头 alone+occupy+Ao6+head; all alone on the Great Dragon Turtle’s head/to get first place in a competition/to be the champion/to top the list).

J. Chen (2004) has noted that the phenomenon of “cultural non-equivalence” (wénhuà bù děng zhí 文化不等值) should be considered in the translation of PUs and that it creates an obvious difficulty in the development of a bilingual dictionary. Without corresponding lexies in the target language, it is difficult for the lexicographer to respect or reflect the culture of the source language.

2 Phraseoculture, poorly treated in specialized French–Chinese and Chinese–French bilingual dictionaries: A metaphraseographic approach

According to Diaz (1981: 75), IEs should be approached from four points of view, which show the insufficiencies of the treatment of IEs in the dictionary: at the language level, at the level of use, at the level of discourse, and at the cultural level. Unfortunately, “the dictionary rarely provides information about the connotation that the linguistic community attributes to an IE and that the learner will not be able to establish on his or her own” (Ibid.). Some dictionaries pay special attention to “linguoculturological” (Teliya et al. 2006 [2014]; Bartmiński 1996, 1999, 2012) or “lexicultural” (Rey et Chantreau 2003) principles in the analysis of phraseologisms. However, these dictionaries with a partly cultural vocation are mainly monolingual projects and therefore do not allow the determination of cultural specificities.

We looked for FEs carrying strong allusions (history, fable, mythology, culture, etc.) in bilingual dictionaries of FEs (three specialized Chinese–French dictionaries and three French–Chinese dictionaries) and analyzed them with a metaphraseographic approach. We find that phraseoculture is not satisfactorily addressed. The example of a

---

6 Æo: Legendary turtle with dragon body.
7 The learner who does not master the entire system may have doubts and risk not using the IE wisely. These productions deviating from generally accepted usage constitute errors of language.
8 Learners seek to know if expressions are usual and frequent. The indications they can find in the dictionary (pop., fam., etc.) are too vague and imprecise and only very relatively define the levels of language and the variations of register.
9 The absence of context makes it impossible for learners to foresee which situations they will be able to use an IE in.
10 Quoted by Pamies (2017).
11 (Doan and Weng 1999; Beijing Publishing House 1980; Sun 2012)
12 (Yue and Xiao 2000; Sun 2010; Cai 2014)
Chinese FE from a famous historical reference on page 747 of Sun’s Chinese–French dictionary (2012) demonstrates this:

臥薪嘗膽 (wò xīn cháng dǎn) Coucher sur de la paille et goûter du fiel/couvrir sa vengeance à travers de dures épreuves/boire le calice jusqu’à la lie pour se venger au moment venu/« L’homme a souffert! Il a bu tous les calices » (J. Anouilh)

The meaning of this Chinese expression is as follows:

[to sleep on brushwood and taste gall in order to recall one’s humiliations/to maintain one’s resolve for revenge]

The current Chinese–French bilingual dictionaries include simplified characters, traditional characters, and Pinyin transcriptions, and they offer a literal translation, a free translation, and when possible, an idiomatic equivalent in French. However, sources and phraseocultural explanations are missing. The absence of these prevents users from gaining a sufficient understanding from the word-for-word (臥薪嘗膽 wò xīn cháng dǎn 臥薪尝胆; sleep+firewood+taste+gallbladder) and global translations. A short historical addition (see below for the DiCoP) for expressions with strong allusive content would facilitate the mastery of these expressions and reinforce the role of the dictionary in terms of cultural transmission.

Similarly, French–Chinese dictionaries only offer an idiomatic, if it exists, and an example(s) of use. We can take the obvious example of the grue (crane), whose image is very positive in China, in contrast to in the French culture, where this bird is very negatively connoted.

Su’s dictionary (2010), on page 589, presents the following:

grue n.f. faire le pied de grue loc. v. 鶴立, 久等
① Et d’indiquer que (sic) des yeux les dizaines de Noires et de mulâtresses qui font, tout près (sic) le pied de grue en se pavanant lascivement. (Le Point, 18-03-1995) 还要指出，几十个黑女人或黑白混血女人就在近旁猥亵作态地等客呢。 (Sun 2010: 589)

[cane (noun). cool one’s heels
Example : ① And to indicate that the eyes of the dozens of black women and mulatto girls who are waiting nearby, strutting lasciviously.]

As this example demonstrates, these works are not free from approximations, gaps, or even errors and inaccuracies in their translations (“sic”). In these dictionaries, the presence of 謂 is limited to the compilation of glossaries, and specialized dictionaries only provide meanings in the form of paraphrases and some examples, with few etymologies in Chinese. Bilingual Chinese–French dictionaries still require much development.

Cai’s (2014) dictionary does offer specific cultural explanations. The author provides fairly thorough explanations for each FE, especially for those with strong allusive content. It provides examples of the use of these expressions and evokes their origins and metaphorical meanings. However, many FEs that are rich in phraseoculture are not included; for example, être dans les bras de Morpheus (in the arms of Morpheus) or avoir des yeux de lynx (have eyes like a hawk/lynx-eyed), etc.

Moreover, their particular phraseocultural dimension is not always associated with an appropriate translation. Take the example of avoir la grosse tête (have a big or swollen head/be full of oneself) in Cai (2014) on page 36. In the absence of a more precise comparative study of the two cultures, the translator does not propose a satisfactory correspondent in Chinese, and on the other hand does not give any clarification on the
connotation of the word “cœur” (heart) in Chinese, often synonymous with the French word “tête” (head): xīngáo-qiào 心高气傲 (heart+high+air+proud). Indeed, there is a notable difference between the “holistic” conception of the heart in Chinese, and the Western dichotomy between “cardiocentrism” and “cerebrocentrism” (Yu 2009; L. Chen 2022a).

It will probably be stated that a paper dictionary cannot explain everything, unless it is thousands of pages long. This is the pitfall of paper dictionaries, which are subject to physical volume constraints and must therefore make choices to be able to remain easily manipulated. Electronic dictionaries, with a system of links, can circumvent this problem, which represents a major advantage that it would be a pity not to exploit.

In bilingual dictionaries, if the lexicographer is content to communicate the meaning, ignoring the social and cultural dimension, then the translation and the transfer function are incomplete. It is in this context that we propose the Dictionary and Corpus of Phraseology (DiCoP) project.

3 Innovation of phraseographic microstructure in the DiCoP: The treatment of the contrastive phraseoculture of fixed expressions

3.1 Presentation of the DiCoP project

In line with the current electronic era, the DiCoP project aims at developing an electronic dictionary of multilingual phraseology (currently bilingual), based on a corpus of phraseological units (collocations, proverbs, IEs, puns, defrosting, etc.) and associated databases to determine their frequency of use (in newspapers, literary works, manuals, etc.) in practice and thus their vitality, to improve their automatic translation, and with the aim of giving easier access to the phraseological units.

Within the context of our thesis (L. Chen 2021) on French and Chinese IEs and their respective translations/equivalences, we have built a corpus of 2,400 entries related to the human body and animals. We are continuing this work and enriching the lexicographical corpus with fields such as gastronomy, numbers, and plants. In the digital era, we believe it crucial to integrate phraseoculture into this DiCoP, which is inspired by the lexicographical tradition of Le Robert. The DiCoP is aimed at students, teachers, translators, and professionals from all disciplines.

3.2 Computerized dictionary data and the design of an innovative microstructure in the DiCoP: The addition of contrastive phraseoculture

The DiCoP is currently constituted on the basis of two corpora, Chinese–French and French–Chinese. It pays particular attention to phraseoculture, especially for PUs of partial equivalence or non-equivalence. Take the example of grue (crane) in French–Chinese in the DiCoP.

13 Site: phraseologia.com (development in progress).
Grue n.f. Faire le pied de grue « attendre debout » [exp.] v. 鹄立, 久等

**词源:** 该表达式来源于17世纪的固定表达faire la jambe de grue和faire de la grue (和16世纪的faire la grue)。在Faire de la grue中，grue这个名词有动词“等待”的含义 (如在诗人Maurice Scève作品中)，但是Bonaventure des Périers认为，在16世纪的语境中faire de la grue中的grue一词也含有鸟类：鹤的比喻义。

**比喻义:** 在16世纪，鹤的比喻用途通常是贬义的（être grue 傻蠢, suivre la multitude comme les grues [Calvin] s'en aller comme des grues [ibid.]), 像鹤一样跟随众人，没有自己的想法。以及Le Roux在1752年给出的变体: être planté comme une grue. 与être planté comme un sot同义：一动不动地站着，等待很久，久等。最后，“妓女”的意思来自faire le pied de grue, 人行道上妓女等待客人的形象比喻。

**实例:** ①Et d'indiquer des yeux les dizaines de Noires et de mulattresses qui font, tout près, le pied de grue en se pavanant lascivement. (Le Point, 18-03-1995) 并使眼色表明附近几十个黑人妇女和黑白混血女人正在作出淫荡姿态等客上门呢。

**Example:** ① And to indicate that the eyes of the dozens of black women and mulatto girls who are waiting nearby, strutting lasciviously.

**Source:** this expression succeeds in the 17th century to “faire la jambe de grue” and to the form “faire de la grue” (“faire de la grue”, 16th century). In this last expression, “grue” (crane) appears to be used as a verbal noun, derived from the verb "gruer" (to wait) in Maurice Scève. However, the context in Bonaventure des Périers shows that it also metaphorically refers to a bird.

**Metaphor:** The figurative uses of crane in the 16th century are generally pejorative (“to be a crane” means to be a “fool” and to follow the multitude like the cranes [Calvin]. In addition, the variant given by Le Roux in 1752, “to be planted like a crane,” also meant “to be planted like a fool.” Finally, the meaning of “prostitute” comes from “faire le pied de grue,” that is, “to wait for the customer on the sidewalk”).

**Example:** ① And to indicate that the eyes of the dozens of black women and mulatto girls who are waiting nearby, strutting lasciviously.

For these language-specific expressions, a phraseocultural annotation is essential: the crane in French has a meaning of prostitute.

In Chinese–French, we propose a more exhaustive microstructure based on simplified and traditional characters, Pinyin, the word-by-word translation of monosyllabic or disyllabic characters or lexies, and the figurative meanings, sources, and pragmatic uses of FEs according to the following model:

卧薪尝胆 [臥薪嘗膽] (coucher, bois de chauffage, goûter, vésicule biliaire)

**Pinyin:** wò xīn cháng dǎn

**Traduction Littérale:** coucher sur de la paille et goûter du fiel

**Signification Implicite:** couver sa vengeance à travers de dures épreuves/boire le calice jusqu'à la lie pour se venger au moment venu

**Source:** « 史记 Shiji » (Mémoires du Grand Historien ou Mémoires historiques) (109 - 91 av. J.-C.), Histoire: L’Épisode se déroule, en 494 av. J.-C. Vaincu par l'empereur de l'État de Wu (吴), l'empereur Gou-Jian de l'État de Yue (越) décide de prendre sa revanche. Afin de ne pas oublier l'opportunité qui couvre son pays déchu et de s'affermir dans sa résolution de se venger, il dort sur de la paille et goûte souvent la sérécition d'une vésicule biliaire suspendue au mur de sa chambre. Cet exercice de mortification le rend plus fort et il finit par vaincre l'État de Wu.

**Exemple d'emploi:** 一时失败何足畏,若有卧薪尝胆的壮志,一定能反败为胜。Il n'y a rien à craindre d'une défaite, car avec de l'ambition et en surmontant les épreuves, on peut la transformer en victoire.
This idiomatic expression originated from the following history:
The episode takes place in 494 BC. Defeated by the Emperor of the State of Wu (吴), Emperor Gou-Jian of the State of Yue (越) decides to take his revenge. In order to not forget the opprobrium that covers his fallen country and to strengthen his resolve to take revenge, he sleeps on firewood and often tastes the gallbladder hanging on the wall of his room. This exercise of mortification makes him stronger, and he eventually defeats the State of Wu.

4 Conclusion

Bilingual lexicographers (French–Chinese and Chinese–French) J. Huang and C. Chen (2003: 101) note that “dictionaries are the index of culture.” They are the key to opening the door to the knowledge and culture contained in semantics and constitute important reference works for learning and mastering foreign languages. For J. Pruvost (2006), the role of the dictionary as a “tool of a language and a culture” need no longer be demonstrated. As T. Szende (2003) notes, “Establishing identity relations between the terms of two languages within the framework of a bilingual dictionary is as much a linguistic operation as a cultural one.” (mentioned by M. Murano 2011: 60) The bilingual dictionary “is revealed today in its phase of transformation as a tool for reflection on culture” (Celotti 2002: 464), and “can also serve as a bridge between cultures, or (as cultures are not monolithic) between two sets of cultural understandings” (Rodger 2006: 572).

Closely linked to a socioculture that differs from one country to another, phraseoculturology is an object of attention in linguistics insofar as it is a matter of considering “a particular and fundamental dimension of words which, unfortunately, is lacking in lexicography as well as in dictionary” (Pruvost 2005: 16). As it represents a major source of difficulties, it should occupy a more substantial place in the dictionary, which is “the preferred medium for compiling all lexical units […] as well as specific idiomatic expressions” (L. Chen 2021: 278). This DiCoP project is part of a contrastive French–Chinese perspective and is therefore related to lexicology, phraseography, phraseotranslation (according to overall meaning, context, and related traits rather than word by word), and the development of natural language processing (NLP).

It could also be the subject of further phraseodidactic exploitation (such as the introduction of phraseological units in learning earlier and more regularly according to their opacity, frequency in daily life, etc.). Moreover, the advent and subsequent generalization of computer tools has considerably modified the modes of consulting and designing dictionaries, offering possibilities for enriching content, as well as the additional development models of contributory, collaborative, and participatory. DiCoP intends to take advantage of open annotation and commentary functions to allow for an intercommunication between readers and lexicographers in an interactive practice.

---

14 辞书是文化的索引。
References


24