

# Fossil classifiers in Chinese chengyu and proverbs

Llinos Evans

University of Essex

---

## Abstract

*Chengyu and proverbs, being archaisms from the days of Classical Chinese, have preserved numerous traits from that period in Chinese history that cannot be replicated in modern Chinese varieties. This paper squarely focuses on the cases of 一介书生 yī jièshūshēng, 一抔黄土 yī póuhuángtǔ, and 一言既出, 驷马难追 yī yán jìchū, sì mǎ nán zhuī, and how their respective classifiers: 介 jiè, 抔 pōu, and 驷 sì, are all examples of lexical fossils when used as classifiers in these phrases. It further proposes evaluating the definition of lexical fossils within a Sinological context.*

**Keywords:** Chinese, Mandarin, Classical Chinese, wenyawen, wenyaw, classifiers, measure words, liangci, proverb, idiom, chengyu, historical linguistics, lexical fossils, fossil words, sinology, grammaticalisation

**Date of Submission:** 05.05.2025

**Date of Acceptance:** 19.09.2025

## Introduction

This paper aims to apply the study of lexical fossils to Mandarin Chinese, broaching the topic within East Asian languages. I decided to focus on *Chengyu* (成语), sometimes referred to as idioms in English. *Chengyu* are a system of set phrases, usually consisting of ~4 Han characters that have become fossilised in many varieties of Chinese, such as Mandarin and Cantonese (Huang C., 2017). For the purposes of this paper, I will be strictly analysing them from a Mandarin Chinese lens. *Chengyu* are often rooted in Classical and/or Literary Chinese, the former being the language spoken during the time of the Five Classics and Four Books (五经四书 *sǔjīng sìshū*, c.500 B.C.) and the latter the language mimicking its structures after the fact (Lu, 2016; Priestley and Shou-jung, 1962). Because many of these idioms are so old, they survive numerous grammatical, semantic, and structural changes that are seen outside of their original varieties, even having characters no longer in common use. They are, therefore, sometimes difficult to derive meaning from, with dictionaries dedicated to them existing (Huang C., 2017, pp. 2–3). While *Chengyu* are generally well-documented in western literature, scant attention has been paid to how classifiers (量词 *liàngcí*),

also known as “measure words,” are positioned within this linguistic phenomenon, or in Classical Chinese in general.<sup>1</sup>

Chinese classifiers, colloquially called “measure words,” are “indicators of prominent features that can be attached to a particular set or class of nouns” (Yip and Rimmington, 2016, p. 41). They must be applied to all nouns when occurring with a number or demonstrative, regardless of whether one needs to “measure” something at all (Yip and Rimmington, 2016, pp. 24–58). Therefore, a classifier, in a distributional sense, cannot be removed, only substituted by a limited selection of applicable classifiers, if any. In the following examples, we will adhere to the Leipzig Glossing Rules (Max Planck Institute, 2015).

一本书

*yī běn shū*

One CLF book

Without this classifier, this sentence would be considered ungrammatical. Without 本 *běn*, this sentence would be seen as strange:

\*一张书

\**yī zhāng shū*

\*One CLF book

This is because Chinese classifiers exist to assign semantic properties to their respective nouns (Del Gobbo, 2014; Yip and Rimmington, 2016, pp. 36–58). In this case, one could see 一本书 *yī běn shū* as a “roll” of a book, referring to the times of Chinese books being scrolls of bamboo slips. Classifiers furthermore resolve homophone issues. Observe:

一书

*yī shū*

To a Mandarin speaker, without context, this means nothing. One cannot determine if this is one uncle (叔), a pretty woman (姝), or the aforementioned book. 本 clarifies which “*shū*” one would be referring to, though like any language, it is not perfect.

Classifiers have their origins around the period of Old and Middle Chinese (Del Gobbo, 2014). Wang (1994) proposes that this began with Oracle Bone Script’s (甲骨文 *Jiǎgǔwén*) 丰 *fēng* (later 个 *gè*) and 介 *jiè*, before fully maturing during the Song dynasty. At this time, they were considered measure words, like “a cup of water,” rather than outright classifiers.

---

<sup>1</sup> This is likely a consequence of Classifiers mainly appearing in later iterations of Classical Chinese, which are rarely studied themselves. However, this aspect falls outside the scope of this paper.

Lexical fossils, known as “fossil words” in older literature, are defined by the Oxford English Dictionary (Oxford University Press, 2025) as “a word or other linguistic form which has become obsolete except in isolated regions or in set phrases, idioms, or collocations.” From an English perspective, one can easily point to phrases such as “to and fro,” “biding one’s time,” and “taking umbrage” (Coffey, 2013). Extending this to linguistic forms, Colchester is a compound of the River Colne and *ceaster*, the Old English name for a Roman town. Ceaster occurs in many placenames (e.g. Leicester, Cirencester, Manchester), but is not functional in modern English and has thus fossilised; it went from a word to a suffix, before becoming a non-productive lexical fossil. However, this definition is somewhat murky, and proposals to revise and formalise a working definition within historical linguistics has been proposed by Coffey (2013), which could assist with applying the term to non-European languages.

These forms are, therefore, fossils: They have no life of their own, they do not occur anywhere outside of their fossilised areas. Should their last remaining source fall into disuse, they will go the way of words like *heora* and *hlisful*. These are not simple archaisms like the use of *thou*, as they see no use in any other sense: They only appear in their “fossilised” domains.

## Methodology

When selecting *Chengyu* to analyse, several criteria were chosen, with the main aim being to outline a conservative definition for these fossil classifiers:

- The *Chengyu* must appear in dictionaries.
- The proposed classifier must not be in active use outside of *Chengyu*. While the 斛 *hú* in 源泉万斛 *yuánquánwànhú* is obsolete, it is still a formal unit of measurement (Institute of Linguistics CASS, 2016, pp. 1786–1796). Similarly, the 枚 *méi* in 不胜枚举 *bùshèngméijǔ*, while not the modern sense of the word, is still used for small objects and bombs, and experiencing neologistic usage for “cute” people (Jiao, 2002; Shi and Jing-Schmidt, 2020). Therefore, they do not meet this criterion.
- Classically aligned or otherwise highly literary phrases (such as 万乘之国) are excluded, as they are learned borrowings.
- Out of preference, the *Chengyu* should have a traceable etymology.

Based on these criteria, I selected the following *Chengyu* (citations denote dictionary sources):

一介书生

*yījièshūshēng*

One-CLF-book-born

“A (mere) scholar”

Sources: Huang (2017); Institute of Linguistics CASS (2016); Taiwan Ministry of Education (1994)

一抔黄土

*yīpóuhuángtǔ*

One-CLF-yellow-dirt

“One loess”

Sources: Huang (2017); Institute of Linguistics CASS (2016); Taiwan Ministry of Education, (1994)

一言既出，驷马难追

*yī yán jì chū, sì mǎ nán zhuī*

One-word-since-release, CLF-horse-hard-chase

*“With one word, a horse will flee and be difficult to catch”*

Source: Taiwan Ministry of Education (1994)

I will argue that each of these fit the distributional qualities necessary for them to be classifiers in the Classical Chinese definition. Analysis will therefore be conducted with the following tests, which were chosen based on the syntactic qualities of classifiers themselves:

- Distribution: Does the word appear before or after the noun?
- Replacement: Does replacing the word change the meaning of the text or otherwise make it ungrammatical?
- Elision: Can the word be removed?

## Analysis

I will first take the following two *Chengyu*, for they both follow the same One-CLF-noun structure, which is the most common way classifiers manifest in modern Mandarin.

一介书生

*yījièshūshēng*

One-CLF-book-born

“A (mere) scholar”

一抔黄土

*yīpóuhuángtǔ*

One-CLF-yellow-dirt

“One loess”

介, as aforementioned, was a common classifier in the time of Classical Chinese. This *Chengyu* is seen in the following passage from 滕王阁序 *Téngwáng Gé Xù* by Tang dynasty poet 王勃 Wáng Bó (664):

勃，三尺微命，一介书生。无路请缨，等终军之弱冠。

*Bó, sān chǐ wēi mìng, yījiè shūshēng. Wú lù qǐngyīng, děng zhōng jūn zhī ruòguān.*

“I, Wang Bo, of humble life, am **a mere scholar**. I have no way to serve like Zhong Jun, despite being as young as he was.”

Conducting a movement test with 介 *jiè*, the meaning instantly falls apart:

\*一书介生

*\*yī shū jiè shēng*

It can be seen that “一书” looks as though it needs 本 before it: Indeed, Chinese is highly reliant on word order to make sense. Disregarding that makes it sound as though there is a book introduction, as 介 *jiè* in modern Mandarin is commonly seen in the word 介绍 *jièshào*, but it nevertheless remains nonsensical.

Likewise, 生 cannot be moved either:

\*生一介书

*\*yī shū jiè shēng*

This would sound as though one is giving birth to a book. Obviously, this is not adequate.

However, a substitution test, swapping 介 *jiè*, works in a limited capacity:

一个书生

*yī gè shūshēng*

一位书生

*yī wèi shūshēng*

\*一本书生

*\*yī běn shū shēng*

The use of 个 and 位 is grammatical, but they lose their classical semantic meaning, as the *Chengyu* is broken because it is no longer referencing the humbleness of Wang Bo. 个 works simply because 书生 itself is a valid word for a scholar, and the structure becomes less formal as a result, and 位, the respectful form for a learned individual, also works for modern Mandarin. However, they lose their classical semantic meaning: They are no longer referencing the humbleness of Wang Bo. 本, on the other hand, does not work at all, as it implies there is “a roll of a scholar”. The ungrammatical nature of 本, therefore, implies that 介 is assigning semantic

value to 书生 in some capacity, be it in the *Chengyu*'s reference or in the classifier sense. This therefore adds further criteria to whether or not something is a classifier.

Elision, while grammatical in Classical Chinese, is ungrammatical in Mandarin Chinese, the language actually spoken, which supports the claim that this is a classifier.

一书生

\**yī shūshēng*

Therefore, the 介 *jiè* in 一介书生 *yījiè shūshēng* fits all the definitions to be a fossilised classifier, as this is the only phrase in modern Mandarin to use it, and 介 *jiè* is thus no longer functional.

The fossilisation of 抔 *póu* in 一抔黄土 *yīpóuhuángtǔ* can also easily be deduced following the same logic. One example of usage can be found in 史记 *shǐjì*, known as Records of the Grand Historian in English, within the biography of Zhang Shizhi (Sima, 1975):

“假令愚民取长陵一抔土，陛下何以加其法乎？”

*Jiǎ lìng yúmín qǔ cháng líng yī póu tǔ, bìxià héyǐ jiā qí fǎ hū?*

"If a fool took a **handful of soil** from Chang Ling, why would Your Majesty punish him?"

As with 一介书生 *yījiè shūshēng*, we can deduce that 抔 *póu* is a classifier using the same syntax tests as before.

A movement test shows the meaning falling apart:

\*一黄抔土

\**yī huángpóutǔ*

A substitution test works, but only in a limited capacity with the generic 个 *gè*:

一个黄土

*yī gè huángtǔ*

\*一本黄土

\**yī běn huángtǔ*

An elision test fails in the context of Mandarin grammar:

\*一黄土

\**yī huángtǔ*

In all respects, it follows the exact same syntax as 一介书生 *yījiè shūshēng* and is absent from any usage outside of these scenarios. Therefore, 介 *jiè* and 杯 *póu* can be called lexical fossils when referring to their classifier usages.

Next, we will analyse 驷 *sì*, which is used in many idioms (Huang, 2017, e.g. 高车驷马; 驷马高门, 驷不及舌), and in many cases, it appears as a lexical fossil. It originally meant a chariot or carriage drawn by four horses, but later became a classifier for horses themselves (Lu, 2016). It was then supplanted by 匹 *pǐ* as early as the Ming dynasty.<sup>2</sup>

Thus, we can conduct a substitution test to showcase this evolution:

一匹马

*yī pǐ mǎ*

One CLF horse

\*一驷马

\**yī sì mǎ*

驷 *sì* is simply not an operational classifier in modern Mandarin Chinese, and thus using it to refer to horses is ungrammatical, as the semantic meaning associated with the classifier (while itself grammaticalised) is lost.

To support labelling it as a classifier, and thus lexical fossil status, we will review 一言既出，驷马难追 *yīyánjìchū, sì mǎ nánzhuī*. While more of a proverb (Taiwan Ministry of Education, 1994), it can be seen in 邓析子 Deng Xizi's 转辞 Zhuanci chapter, written during the Spring & Autumn period (Sturgeon, 2025):

“一言而非，驷马不能追；一言而急，驷马不能及。”

*yī yán ér fēi, sì mǎ bùnéng zhuī; yī yán ér jí, sì mǎ bùnéng jí.*

one-word-and-not CLF-horse-NEG-can-run; one-word-and-urgent, CLF-horse-NEG-can-reach

“If a word is wrong, four horses cannot catch up; if a word is urgent, four horses cannot catch up.”

---

<sup>2</sup> For an example of this intermittent stage, see 騙經 (The Book of Swindles) by 張應俞 Zhang Yingyu:  
<https://zh.wikisource.org/wiki/%E9%A8%99%E7%B6%93/%E8%84%AB%E5%89%9D%E9%A8%99>

Of course, the phrase has changed, removing 不 *bù* to create the modern four-character structure, but what is important here is 驷 *sì*, the classifier in this structure.

In the case of this classifier, its usage in Mandarin Chinese may not immediately be obvious. This is because this *Chengyu* uses a Classical Chinese convention of not necessarily requiring a numeral; this will be proven in the movement test, which also proves it is grammatical in Mandarin Chinese.

We can conduct a movement test by moving 驷 *sì* to the right of 马 *mǎ*, wherein it breaks Mandarin Chinese grammatical conventions, as a classifier cannot succeed a noun. However, it is grammatical to do this in Classical Chinese, so we know it was a classifier at that point in time. This is important, as if we are going to evaluate a fossil classifier in an idiom, we must understand the original grammatical status of the classifier to re-evaluate from a modern perspective.

\*一言既出，马驷难追

*\*yī yán jì chū, mǎ sì nán zhuī*

one-word-already-go out horse-CLF-hard-chase

In the case of both Mandarin and Classical Chinese, moving the classifier beyond these boundaries breaks the meaning completely.

\*一言既出，马难驷追

*\*yī yán jì chū, mǎ nán sì zhuī*

one-word-already-go out horse-hard-\*CLF-chase

The classifier can be replaced, but only by specific classifiers and by breaking the *Chengyu* in the process:

一言既出，头马难追

*yī yán jì chū, tóu mǎ nán zhuī*

\*一言既出，本马难追

*\*yī yán jì chū, běn mǎ nán zhuī*

头 can be used to refer to livestock in general: It is possible for it to work with horses, but one may believe the speaker were referring to a 马骡 (*mǎ luó*, mule).



Therefore, given that this classifier usage is unseen outside of idioms, 驷 *sì* as a classifier is indeed a lexical fossil. Furthermore, the Han character<sup>3</sup> is not used outside of Literary Chinese terms, such as 良驷 *liángsì* “a fine steed”, which means an argument for the character itself being a lexical fossil in Mandarin Chinese is a potential area of research.

## Conclusion

This paper aimed to illustrate examples of lexical fossils in Mandarin Chinese, bringing their study to East Asian languages, something that is rarely conducted. Ergo, it only scratches the surface of the topic, focusing on classifiers within Mandarin Chinese *Chengyu*. Only 介 *jiè*, 怀 *póu*, and 驷 *sì* have been argued for out of potentially far more. I believe that this sort of research could be used to broach new topics in teaching *Chengyu* to learners of Chinese varieties, as well as showcase the language’s deep history to the world at large. Formalising a concrete definition of “lexical fossil,” especially when evaluating other Chinese varieties and their use of classifiers, could assist with helping the working definition proposed by Coffey (2013) to flourish outside of European languages. Further research on the syntax of *Chengyu* as Classical Chinese sentences could also shed further light on the potential for discovering more lexical fossils.

## References

- Coffey, S. J. (2013). Lexical Fossils in Present-Day English: Describing and Delimiting the Phenomenon. In R. W. McConchie, T. Juvonen, M. Kaunisto, M. Nevala, & J. Tyrkkö (Eds.), *Selected Proceedings of the 2012 Symposium on New Approaches in English Historical Lexis (HEL-LEX 3)* (pp. 47–53). Cascadilla Proceedings Project.
- Del Gobbo, F. (2014). Classifiers. In C. -T. J. Huang, Y. -H. A. Li, & A. Simpson (Eds.), *The Handbook of Chinese Linguistics* (1st ed., pp. 26–48). Wiley.  
<https://doi.org/10.1002/9781118584552.ch2>
- Huang C. (2017). 汉语成语词典 [*Hanyu Chengyu Cidian*] (New Revised Edition). 四川辞书出版社 [Sichuan Cishu Chubanshe].
- Institute of Linguistics CASS (Ed.). (2016). 现代汉语词典 [*Xiandai Hanyu Cidian*] (7th ed.). 商务印书馆 [The Commercial Press].
- Jiao, F. (ed.) (2002) 汉英量词词典 [*A Chinese-English Dictionary of Measure Words*]. 2nd ed. Beijing: Sinolingua.
- Lu Z. (with Wen Y., Su J., Bu Q., Gao J., & He Q.). (2016). 实用古代汉语词典 [*Shiyong Gudai Hanyu Cidian*] (2nd Ed.). 甘肃教育出版社 [Gansu Education Publishing House].

---

<sup>3</sup> That is, a 汉字 *Hànzì* Chinese character, such as 孔 *kǒng*. Han characters form the writing system used by Classical and Mandarin Chinese, among other languages.

Max Planck Institute. (2015, May 15). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. Max Planck Institute for Evolutionary Anthropology. <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>

Oxford University Press. (2025). *Oxford English Dictionary* [Dictionary]. Oxford English Dictionary. <http://www.oed.com>

Priestley, K. E., & Shou-jung, C. (1962). *China's Men of Letters, Yesterday and Today*. Dragonfly Books.

Shi, H.H. and Jing-Schmidt, Z. (2020) 'Little cutie one piece: An innovative human classifier and its social indexicality in Chinese digital culture', *Chinese Language and Discourse*, 11(1), pp. 31–54. Available at: <https://doi.org/10.1075/cld.00023.shi>.

Sima, Q. (1975). *史記 [Shǐjì]* (Y. Pei, Z. Sima, & S. Zhang, Eds.; 7th ed., Vol. 1). 中華書局出版 [Zhonghua Book Company].

Sturgeon, D. (Ed.). (2025). 鄧析子一卷 [Deng Xizi, One Volume]. In *意林 [Yilin]* (Digitized ed., Vol. 1). Chinese Text Project. <https://ctext.org/text.pl?node=566757&if=en&remap=gb>

Taiwan Ministry of Education. (1994). *重編國語辭典修定本 [Revised Mandarin Chinese Dictionary]*. Taiwan Ministry of Education.

Wang, B. (664). 秋日登洪府滕王閣餞別序 [Qiū rì dēng hóng fǔ téngwánggé jiànbíe xù]. <https://zh.wikisource.org/wiki/%E6%BB%95%E7%8E%8B%E9%96%A3%E5%BA%8F>

Wang, L. (1994) *Origin and development of classifiers in Chinese*. PhD dissertation. The Ohio State University. Available at: [https://web.archive.org/web/20190711012236/https://etd.ohiolink.edu/!etd.send\\_file%3Faccession%3Dosu1487856076411916%26disposition%3Dinline](https://web.archive.org/web/20190711012236/https://etd.ohiolink.edu/!etd.send_file%3Faccession%3Dosu1487856076411916%26disposition%3Dinline) (Accessed: 5 May 2025).

Yip, P.-C., & Rimmington, D. (2016). *Chinese: A comprehensive grammar*. 2nd edn. Routledge.

## Copyright Statement

© Llinos Evans. This article is licensed under a [Creative Commons Attribution 4.0 International Licence \(CC BY-ND\)](https://creativecommons.org/licenses/by-nd/4.0/).