An Investigation of Sentiment Polarity in Chinese VO Idioms Based on NLP Models

Xueyi Wen
School of Humanities
Nanyang Technological Univ.
XUEYI001@e.ntu.edu.sg

Hongzhi Xu
Inst. of Language Sciences
Shanghai Intl. Studies Univ.
hxu@shisu.edu.cn

Lili Yang
Sch. of Chinese Lang. and Lit.
Soochow University
llyang@suda.edu.cn

Abstract

Construction expressions pose significant challenges to Natural Language Processing (NLP) tasks such as natural language understanding and sentiment analysis due to their semantic opacity, where the overall meaning cannot be directly inferred from the literal meanings of their components. Chinese Verb-Object (VO) constructions are full of such special constructions and idioms, presenting specific challenges in sentiment analysis. This paper employs NLP and statistical methods to investigate the sentiment polarity of Chinese VO idioms and the correlations between sentiment polarity and literal and contextual meaning. The result of our analysis shows that Chinese VO idioms in general exhibit a negative sentiment bias, and their polarity is more closely tied to contextual factors than literal meaning. This finding aligns with the contextualized nature and semantic opacity of constructions.

Keywords: Chinese VO idioms; NLP; Sentiment Analysis

1 Introduction

Sentiment analysis, a key task in Natural Language Processing (NLP), aims to extract opinions, sentiments, evaluations, attitudes, and emotions from text (Liu, 2017). One major challenge of sentimental analysis is semantic opacity, particularly in expressions such as sarcasm, metaphor, and constructions, where traditional sentiment analysis struggles to capture the underlying meaning and emotional tendency.

Chinese, being a complex language and a major world language, displays frequent semantic opacity phenomena. The study of Chinese constructions' sentimental tendency is of significant importance for improving the performance of Chinese sentiment analysis systems. Attitudes and scalar are important factors impacting semantic opacity. Constructions, as non-recursive, non-trivial phrasal structures (Zhan, 2017), carry both attitudinal (sentiment polarity) and scalar meanings. Specifically, the attitudinal meaning, also termed emotional polarity, includes positive, negative, and neutral. However, due to the non-compositionality of form and meaning (Goldberg, 1995), it's often hard to induce construction meanings, which makes automatic parsing hard for them. Research shows that many Chinese constructions bear sentiment polarity (Zhan and Wang, 2020), with a tendency towards negative attitudes (Fang, 2017). This study focuses on Chinese Verb-Object (VO) idiomatic constructions, using large language models and statistical methods to explore sentiment polarity and its correlations behind its semantic opacity.

We address the following research questions:

1. Compared to literal meaning, do Chinese VO idioms tend to exhibit negative sentiment polarity?

2. Is the negative sentiment polarity of VO idioms related to their literal meaning?

3. Is the negative sentiment polarity of VO idioms related to context?

2 Related Work

Idiomatic expressions can, to some extent, reflect the sentiment polarity of their users. Many studies have used the sentiment-bearing properties of idioms to construct sentiment classification models. Xie and Wang (2014) utilized Chinese idiom resources, primarily four-character idioms, to build a novel unsupervised framework for training general-purpose sentiment classifiers. Williams et al. (2015) demonstrated that using English idioms as features can improve the performance of traditional sentiment analysis models. Wu and

Li (2022) constructed a Chinese construction corpus for sentiment analysis and found that although constructions are not the primary semantic units, they still carry a certain proportion of attitudinal semantic information. Similarly, Tahayna et al. (2022) showed that English idiom corpora annotated with sentiment polarity can enhance the performance of sentiment classifiers. These findings suggest a correlation between idioms and sentiment polarity, and several scholars have conducted research in this area. However, studies focusing specifically on Chinese VO idioms remain scarce, and there is a lack of established sentiment-annotated corpora for this category.

In Modern Chinese, there exists a large number of idiomatic constructions whose emotional meanings have been conventionalized through pragmatic usage. The classification and definition of constructions are diverse. Previous research on the sentiment of constructions has mostly been limited to individual cases or small categories. For example, Du (2005) investigated the "V/A 个 P" (*V/A gè P*, 'V/A one P') and "Q 才 VP" (*Q cái VP*, 'Q only then VP') structures in Chinese idioms, arguing that both function as exclamatory sentences expressing negation. Deng and Huang (2002) and Gan (2008) examined the "不 A 不 B" construction (*bù A bù B*, 'not A not B') from the perspectives of internal structural relations, constraints on negation, and processes of constructionalization. Li (2008) studied the pragmatic marker "问题是" (*wèntí shì*, 'the problem is') and its negative evaluative function. Li (2011, 2014) explored the constructions "X 真是(的)" (*X zhēnshì(de)*, 'X really is / what X really is') and "好你个 + X" (*hǎo nǐ ge X*, 'what a X / how X!') in terms of negative evaluation and underlying causes.

Nonetheless, research on the VO idiom category is limited, and large-scale statistical analyses beyond individual cases are lacking. Therefore, the present study focuses on the major category of VO idioms in Chinese, applying NLP methods to automatically assign sentiment values and employing statistical analysis to investigate sentiment tendencies within this relatively fixed category.

3 Construction of Chinese VO idiom sentimental corpus

The data for this study were drawn from the Chinese portion of the OpenSubtitles corpus in the Open Parallel Corpus (OPUS) (Tiedemann, 2012), the Chinese Linguistics Corpus (CCL) (Center for Chinese Linguistics, Peking University, 2003), and the *Xinhua Dictionary* (Chinese Language Editorial Committee, 2016). OPUS is a comprehensive database of parallel corpora widely used in machine translation research. It is compiled from aligned online translation texts collected from the internet. The Chinese portion of OpenSubtitles contains over 8 million lines and approximately 150 million characters.

The CCL Chinese corpus is developed by the Center for Chinese Linguistics at Peking University, comprising texts from the 11th century BCE to the present. It contains over 500 million characters of Modern Chinese and more than 200 million characters of Classical Chinese, with a total exceeding 700 million characters. In this study, the CCL corpus served as an additional large-scale source for retrieving sentences containing VO idioms, ensuring broader temporal coverage and linguistic variety.

We first summarized 540 Chinese VO-structured idioms trisyllabic form based on literature and dictionaries. The *Xinhua Dictionary* was used to provide both the literal meaning explanation and the idiomatic meaning explanation for each idiom, enabling analysis of the relationship between sentiment polarity and literal semantics.

Using regular expression matching, we retrieved sentences containing these VO idioms from the OPUS and CCL corpora. Considering the factors of idiom *usage frequency* and *familiarity*, idioms with fewer than 50 example sentences were excluded, resulting in 168 VO idioms. From these, we randomly sampled 25 idioms, totaling approximately 210,000 characters, for manual annotation. Each example sentence was assessed for semantic completeness and fluency to determine its validity. For valid sentences, we further annotated whether the VO instance was used in its literal sense or idiomatic sense. Accordingly, the annotation categories were: *invalid data*, *literal-meaning data*, and *idiomatic-meaning data*. The resulting manually annotated VO idiom corpus contains approximately 208,000 characters of valid data. The data information is shown in Appendix A.

4 Sentimental polarity analysis of Chinese VO idioms

4.1 Calculation and correlation model used

Our analysis addresses three related questions: whether a VO construction used idiomatically exhibits a systematically different sentiment polarity from its literal usage, and whether idiomatic polarity is predictable from the surrounding sentential context or its constituents. To answer these questions we use the sentence-level sentiment score as the fundamental observational unit and compute, for each annotated instance, up to three scalar values: the sentiment of the sentence with the VO used idiomatically ("idiom"), the sentiment of a sentence in which the same VO is used literally ("literal"), and the sentiment of the surrounding context obtained by masking the VO ("context"). The overall workflow of the annotation and sentiment scoring pipeline is illustrated in Figure 1. The values for each sentence are then aggregated within constructions to permit paired within-construction comparisons and pooled across constructions to assess overall tendencies.

As for statistical test, for each comparison, we first evaluate distributional assumptions using the Shapiro–Wilk test (α =0.05) together with graphical diagnostics. Where the paired differences or residuals approximate normality, we employ the paired Student's t-test for difference-in-means and Pearson's r for linear association; where normality or linearity is violated, we replace these with the Wilcoxon signed-rank test and Spearman's ρ, respectively. To mitigate issues arising from multiple per-construction tests, we control the false discovery rate via the Benjamini–Hochberg procedure and report effect sizes with confidence intervals. Figure 2 illustrates the logic of our first set of comparisons, which test whether idiomatic and literal sentiment values differ systematically both within and across VO constructions.

The second analysis examines whether idiomatic sentiment polarity is correlated with its surrounding context. For each idiomatic sentence, we remove the VO and compute a sentiment score for the remaining context. We then assess whether these contextual values covary with idiomatic sentiment, which would suggest that the local usage environment contributes to idiom polarity. Figure 3 depicts this procedure schematically. Finally, to account for unbalanced instance counts and construction-specific baselines, we also fit mixed-effects models with idiomatic sentiment as the dependent variable, contextual sentiment as a fixed effect, and random intercepts for construction.

Automated sentiment scoring and preprocessing are performed with standard Chinese NLP tools: Jieba is used for segmentation and SnowNLP for sentence-level sentiment estimation (Feng, 2012; Zhang, 2012). To ensure our conclusions are not artifacts of a particular scorer or small-sample idiosyncrasies, we report robustness checks including confidence intervals for key statistics, repetition of principal tests with corresponding nonparametric measures, and a manual sanity check on a held-out subset.
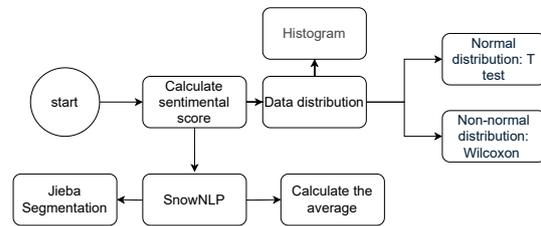

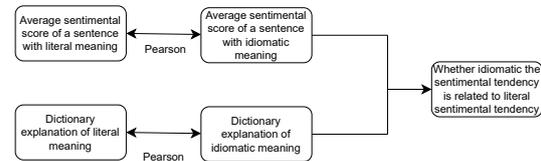Figure 1: Research methodology workflow.

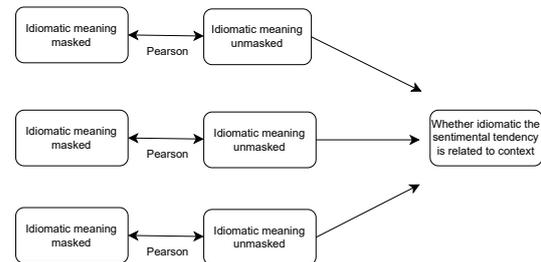Figure 2: Correlation between idiomatic and literal sentiment polarity.

Figure 3: Correlation between idiomatic sentiment polarity and context.

4.2 Data distribution

We used Python's visualization tools to plot normal probability plots for all literal and idiomatic sentences to examine the data distribution, as shown in Figure 4. It can be observed that the

data roughly follow a normal distribution. Therefore, a Student's t-test was applied to assess correlations across the overall dataset. However, due to variations in the frequency of VO construction usage, familiarity, and degree of idiomatization in Chinese, some literal usage examples for certain idioms were limited in the corpus. In cases where the sentence distribution of individual VO constructions deviated from normality, the Wilcoxon signed-rank test was employed for correlation analysis.
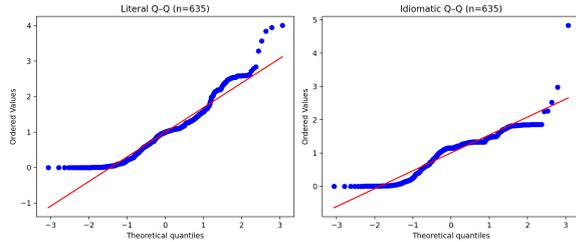

Figure 4: Overall data distribution: normal probability plots

4.3 Negative sentimental polarity of Chinese VO construction

As shown in Table 1, for most Chinese VO constructions, the average sentiment values of idiomatic usages are lower than those of their literal counterparts. The overall mean sentiment values across all VO constructions are 0.547 for literal meanings and 0.530 for idiomatic meanings, indicating a general tendency for idiomatic usages to convey slightly more negative sentiment.

To evaluate the statistical significance of these differences, paired t-tests were conducted for each VO construction as well as for the overall dataset; results for all constructions are provided in Appendix B1. The overall p-value is less than 0.01, confirming that the difference is statistically significant.

Table 1 lists several representative examples (a mix of idioms showing idiomatic negativity and idioms with positive idiomatic senses). The full set of 25 constructions and the complete significance test results are available in Appendix B2. In addition, the full table of all 25 constructions in Appendix C1, which reports sentiment scores based on dictionary definitions, also reveals a marked tendency for idiomatic meanings to be more negative than their literal counterparts (0.3414 vs. 0.63), further corroborating this finding.

ing.

Idiom	Literal Avg	Idiom Avg
拖下水 tuō xiàshuǐ 'drag someone into trouble'	0.506	0.468
有一手 yǒu yī shǒu 'be skillful / have a trick'	0.531	0.545
露头角 lù tóu jiǎo 'begin to show one's talent'	0.530	0.553
戴帽子 dài màozi 'be cuckolded'	0.592	0.588
吃闲饭 chī xiánfàn 'do nothing productive'	0.606	0.533
开绿灯 kāi lǜ dēng 'give the green light / permit'	0.489	0.510
Overall	**0.547**	**0.530**

Table 1: Representative examples of average sentiment values (literal vs. idiomatic) for selected VO constructions. Full table of all 25 constructions and full significance test results are in Appendix B2.

Overall, Chinese VO idiomatic constructions tend to convey more negative sentiment compared to their literal meanings. However, in some cases, the literal meaning may exhibit more positive sentiment than the idiomatic meaning. This phenomenon can be partly explained by certain idioms such as "露头角" (*lù tóu jiǎo*, 'begin to show one's talent') and "有一手" (*yǒu yī shǒu*, 'be skillful / have a trick'), whose idiomatic meanings are inherently positive and carry high sentiment values. In other cases, discrepancies may be due to limitations of the SnowNLP model in capturing idiomatic sentiment. For example, in the corpus, the sentence "她给他戴绿帽子" (*tā gěi tā dài lǜ màozi*) is automatically assigned a sentiment score of 0.662 (positive) by SnowNLP, whereas its actual idiomatic meaning, referring to infidelity in a romantic relationship, is negative.

4.4 The difference of Chinese VO construction sentimental polarity between idiomatic and literal meaning

Based on idiomatic and literal meanings extracted from the *Xinhua Dictionary* and scored using SnowNLP, the comparison is shown in Table 2. Note that for the idioms "讨生活" (*tǎo shēnghuó*) and "吃闲饭" (*chī xiánfàn*), literal explanations were not listed in the dictionary and are therefore excluded from the comparison. Idioms such as "露头角" (*lù tóu jiǎo*) and "有一手" (*yǒu yī shǒu*) have positive sentiment values in their idiomatic meanings (0.619 and 0.894, respectively). For the remaining VO idioms, the sentiment val-

ues of the idiomatic meanings are lower than those of the literal meanings, consistent with our expectations. For instance, the literal meaning of "乱弹琴" (*luàn tán qín*) is "to play music in a disorderly manner without clear melody or rhythm, sounding disharmonious," which has a sentiment score of 0.891. Its idiomatic meaning, "to act recklessly or talk nonsense," has a sentiment score of 0.017. Both human intuition and the machine-assigned sentiment score clearly indicate a negative sentiment tendency when this VO construction is used idiomatically.

Idiom	Literal Sentiment	Idiom Sentiment
乱弹琴 luàn tán qín 'act recklessly / talk nonsense'	0.891	0.017
有一手 yǒu yī shǒu 'be skillful / have a trick'	0.999	0.894
露头角 lù tóu jiǎo 'begin to show one's talent'	0.797	0.719
拖下水 tuō xiàshuǐ 'drag someone into trouble'	0.070	0.216
Overall	0.634	0.341

Table 2: Representative examples of dictionary-based sentiment values for selected VO constructions. Full table of all 25 constructions is in Appendix C1.

The Pearson correlation coefficients for the two comparison methods are presented in Table 3. Both coefficients are close to 0, indicating that there is almost no linear relationship between the idiomatic sentiment and its literal meaning. Once an expression becomes idiomatic, its usage often diverges substantially from the original literal meaning. Some expressions carry negative connotations idiomatically, but their literal meanings may not convey strong negative sentiment, or may even exhibit opposite sentiment polarity.

Method	p-value	Pearson r
Sentence examples	0.929	-0.002
Dictionary explanations	0.567	0.121

Table 3: Correlation test between literal and idiomatic sentiment values.

4.5 Chinese VO construction sentimental polarity correlation with context

In this study, we assessed the sentiment of idiomatic usages within context by masking the idiom in sentences using the [SEP] token in Python. For example, the sentence:

Eg. 4.5 你可以跳楼，但你会把我拖下水，那就让你成为杀人凶手。
nǑ1d0   kǑ113yǑ131   tiǑ0e0olǑ0f3u, dǑ0e0n nǑ1d0 huǑ0ec bǑ101 wǑ1d0 [SEP][SEP], nǑ0e0 jiǑ0f9 rǑ0e0ng nǑ1d0 chǑ0e9ngwǑ0e9i shǑ101rǑ0e9n xiǑ101ngshǑ1d0.
'You may jump off the building, but you will [drag me into trouble], and then you will become a murderer.'

was processed as:

你可以跳楼，但你会把我[SEP][SEP]，那就让你成为杀人凶手。
nǐ kěyǐ tiàolóu, dàn nǐ huì bǎ wǒ [SEP][SEP], nà jiù ràng nǐ chéngwéi shārén xiōngshǒu.
'You may jump off the building, but if you [SEP][SEP], then you will become a murderer.'

The masked sentence was then tokenized and analyzed for sentiment. Differences between the idiomatic sentiment of VO constructions and the corresponding contextual sentiment were evaluated for statistical significance. Representative results are shown in Table 4. The full table of all constructions is provided in Appendix C2. For clarity and data hygiene, we removed two constructions ("生活" and "吃") since their literal usages are rare. We also calculated the overall effect sizes to quantify their practical magnitude. Aggregating over all 25 idiomatic sentences (N = 2489) yields Pearson $r = 0.634$ ($p < 0.001$), which corresponds to $r^2 \approx 0.402$. This shows that contextual embeddings explain roughly 40.2% of the variance in idiomatic sentiment scores. In regression terms, this implies Cohen's $f^2 = r^2/(1 - r^2) \approx 0.672$, a large effect by conventional benchmarks.

5 Discussion on sentiment tendencies and correlations of VO constructions

From the above data analysis, the feasibility and results of using NLP large models to investigate sentiment tendencies and correlations of Chinese constructions can be observed. This provides a foundation for future research, offering data-driven analysis and empirical support.

Starting from the research questions and based on the data analysis, it can be seen that Chinese

Idiom	Pearson r	p-value
拖下水 tuō xiàshuǐ 'drag someone into trouble'	0.955	0.001
有一手 yǒu yī shǒu 'be skillful'	0.978	0.000
开绿灯 kāi lǜ dēng 'permit'	0.990	0.000
咬耳朵 yǎo ěrduo 'whisper in ears'	0.989	0.000
乱弹琴 luàn tán qín ' talk nonsense'	0.992	0.000
Overall	0.634	0.000

Table 4: Representative Pearson correlation between idiomatic sentiment and contextual embedding sentiment for selected VO constructions. Full results are in Appendix C2.

VO constructions exhibit a negative sentiment tendency compared to their literal meanings. This is consistent with the findings of Wu et al. (2017), who observed that negative-attitude meanings in Chinese idioms dominate both frequency and sentiment values in idiomatic corpora. This indicates that negative-attitude meanings in Chinese are largely conveyed through constructional meanings (Fang, 2017). At the same time, it reflects the significant asymmetry in the distribution of constructional meanings, the asymmetry of positive and negative attitudes at the constructional level in Chinese, and the manifestation of the negativity bias at the cognitive level (Rozin and Royzman, 2001).

Regarding the correlation analysis of idiomatic sentiment, two approaches were considered: dictionary-based sentiment correlations and literal-to-idiomatic sentiment correlations. The results show that the sentiment tendency of VO idioms is not correlated with their literal meanings. As a constructional structure, VO idioms cannot have their meaning or form predicted from their component parts once they have become a conventionalized expression (Goldberg, 1995). This is consistent with cognitive construction grammar perspectives. Meanwhile, the sentiment of the context surrounding idioms is correlated with the idiomatic sentiment but not with the literal meaning. This also highlights the implicit nature of idiomatic sentiment: negative evaluations in Chinese VO idioms are associated with contextual sentiment.

For example, consider the sentence processed in the study (see Example in 4.5):

Here, the idiom 拖下水 (tuō xiàshuǐ 'drag someone into trouble') was masked with [SEP] tokens for sentiment analysis. This illustrates that the contextual sentiment is correlated with the idiomatic sentiment, while it is not correlated with the literal meaning of the construction. This aligns with the notion of "constructionalization context" proposed by Traugott and Trousdale (2013), referring to the multifaceted linguistic environment influencing the formation of constructions, including discourse or textual context. The occurrence of a unit, partially or entirely, depends on its context, which can be described through various relations, including syntactic, morphological, phonological, semantic, and pragmatic functions.

6 Limitations

Despite the contributions of this study, several limitations should be acknowledged. First, the data scale could be further expanded. In the current study, only 25 commonly used VO idioms with at least 50 occurrences in the corpus were randomly selected. Therefore, the present study should be viewed as an exploratory case study illustrating the feasibility of combining NLP-based sentiment analysis with statistical testing. Future work will expand the coverage to a larger set of idiomatic expressions, allowing us to test whether the negative polarity tendency observed here generalizes across the lexicon. Second, the NLP model employed may have limitations in sentiment analysis accuracy, as automatic scoring may not fully capture nuanced idiomatic meanings. Third, the contextual information used in this study is limited to the single sentence containing the construction, which may omit discourse-level context and affect sentiment interpretation.

Future research directions include enlarging the database and data selection, collecting more examples of VO idioms, and improving the sentiment analysis models by incorporating pre-trained embeddings or additional sentiment features. Moreover, extending the context to include preceding and following sentences will better capture discourse-level sentiment effects for VO idioms.

7 Conclusion

This study investigated sentiment polarity and correlations of Chinese VO idioms using NLP-based sentiment analysis and statistical methods. The results show that VO idioms generally convey more

negative sentiment than their literal meanings, and that idiomatic sentiment is largely independent of the literal meaning but correlated with the contextual sentiment. These findings contribute to a better understanding of constructional semantics and sentiment in Chinese, providing a data-driven foundation for future research on idiomatic expressions. Furthermore, this study demonstrates the feasibility of leveraging NLP models for large-scale analysis of sentiment in constructions, offering empirical support for research in cognitive construction grammar and the pragmatics of idiomatic expressions.

References

Center for Chinese Linguistics, Peking University. 2003. Chinese linguistics corpus (ccl). http://ccl.pku.edu.cn/. Accessed: 2025-08-10.

Chinese Language Editorial Committee. 2016. *Xinhua Dictionary*, 12th edition. Commercial Press, Beijing.

Yingshu Deng and Gu Huang. 2002. On the negation meaning and constraints of "no a no b". *Chinese Language Learning*, (4):16–20.

Daoliu Du. 2005. A study on the structures "v/a个p" and "q才vp" in chinese idioms. *Journal of Chinese Language Learning*, (1):48–55.

Mei Fang. 2017. Conventionalization of negative evaluation expressions. *Zhongguo Yuwen*, (2):131–147.

Shuo Feng. 2012. Jieba: Chinese word segmentation. Accessed: 2025-08-10.

Lihao Gan. 2008. The construction meaning and negative tendency of "no a no b": A cognitive and pragmatic analysis. *Rhetoric Learning*, (2):56–60.

Adele E. Goldberg. 1995. *A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Xiaojun Li. 2011. The negative evaluative function of the construction "x真是(的)". *Contemporary Rhetoric*, (4):23–30.

Xiaojun Li. 2014. A study on the negative evaluation of the construction "好你个+x". *Journal of Pragmatics Research*, (3):45–53.

Zongjiang Li. 2008. Negative evaluation in the pragmatic marker "the problem is". *Modern Rhetoric*, (2):34–39.

Bing Liu. 2017. Sentiment analysis: Mining opinions, sentiments, and emotions.

Paul Rozin and Edward B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4):296–320.

B. M. Tahayna, R. K. Ayyasamy, and R. Akbar. 2022. Automatic sentiment annotation of idiomatic expressions for sentiment analysis task. In *IEEE Access*, volume 10, pages 122234–122242.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214–2218. European Language Resources Association (ELRA).

Elizabeth C. Traugott and Graeme Trousdale. 2013. *Constructionalization and Constructional Change*. Oxford University Press, Oxford.

L. Williams, C. Bannister, M. Arribas-Ayllon, A. Preece, and I. Spasić. 2015. The role of idioms in sentiment analysis. *Expert Systems with Applications*, 42(21):7375–7385.

Yinqing Wu and Dejun Li. 2022. A study of chinese construction corpus compilation and application for sentiment analysis. In *Proceedings of the 21st Chinese National Conference on Computational Linguistics*, pages 475–484. Chinese Information Processing Society of China.

Yinqing Wu and 1 others. 2017. Sentiment distribution in chinese idioms. *Journal of Chinese Linguistics*, 45(2):123–145.

Songxian Xie and Ting Wang. 2014. Construction of unsupervised sentiment classifier on idioms resources. *Journal of Central South University*, 21(4):1376–1384.

Weidong Zhan. 2017. From phrase to construction: Theoretical issues in building a construction knowledge base. *Journal of Chinese Information Processing*, 31(1):230–238.

Weidong Zhan and Jiajun Wang. 2020. Representation and annotation of constructions in modern chinese. In *Proceedings of the 21st International Workshop on Chinese Lexical Semantics*.

Shuo Zhang. 2012. Snownlp: A python library for chinese text processing. Accessed: 2025-08-14.

A Data distribution

B Average sentiment values and significance tests for 25 Chinese VO constructions and their idiomatic meanings

C Full dictionary- and context-based sentiment tables

Table A1: Frequencies of literal and idiomatic uses of 25 Chinese VO idioms in the annotated corpus.

Idiom	Literal sentences	Idiomatic sentences
擦屁股 (cā pìgu) — wipe someone's butt	132	137
戴帽子 (dài màozi) — wear a hat	677	56
照镜子 (zhào jìngzi) — look in the mirror	180	59
回老家 (huí lǎojiā) — return to old home	206	45
拖下水 (tuō xiàshuǐ) — drag into water	7	186
开后门 (kāi hòumén) — open back door	54	15
说梦话 (shuō mènghuà) — talk in sleep	48	25
啃骨头 (kěn gǔtou) — gnaw bone	16	12
做文章 (zuò wénzhāng) — write article / exploit for gain	3	70
开绿灯 (kāi lǜ dēng) — give the green light / permit	7	31
剃光头 (tì guāngtóu) — shave bald head	38	3
浇冷水 (jiāo lěngshuǐ) — pour cold water	30	19
有一手 (yǒu yī shǒu) — be skillful / have a trick	7	168
踢皮球 (tī píqiú) — kick the ball / pass responsibility	6	53
咬耳朵 (yǎo ěrduo) — whisper in ears	9	74
摇尾巴 (yáo wěibā) — wag tail / fawn on someone	66	8
讨生活 (tǎo shēnghuó) — make a living	0	94
乱弹琴 (luàn tánqín) — act recklessly / talk nonsense	8	47
露头角 (lù tóujiǎo) — begin to show one's talent	0	491
露马脚 (lù mǎjiǎo) — reveal one's hidden flaw	0	48
开夜车 (kāi yèchē) — work/study late at night	5	60
打哈哈 (dǎ hāhā) — laugh off / pretend to agree	1	89
出风头 (chū fēngtóu) — show off / attract attention	1	496
吃闲饭 (chī xiánfàn) — live idly / do nothing productive	0	86
睡大觉 (shuì dàjiào) — sleep heavily / nap	111	117
Total	**1612**	**2489**

Table B1: Average sentiment values for 25 Chinese VO constructions and their idiomatic meanings.

Idiom	Literal Avg	Idiom Avg
擦屁股 (cā pìgu) — wipe someone's butt	0.481	0.487
戴帽子 (dài màozi) — wear a hat	0.592	0.588
照镜子 (zhào jìngzi) — look in the mirror	0.528	0.518
回老家 (huí lǎojiā) — return to hometown	0.537	0.528
拖下水 (tuō xiàshuǐ) — drag someone into trouble	0.506	0.468
开后门 (kāi hòumén) — use backdoor	0.465	0.488
说梦话 (shuō mènghuà) — talk in one's sleep	0.499	0.498
啃骨头 (kěn gǔtou) — gnaw a bone	0.539	0.547
做文章 (zuò wénzhāng) — make an issue / exploit for gain	0.552	0.533
开绿灯 (kāi lǜ dēng) — give the green light / permit	0.489	0.510
剃光头 (tì guāngtóu) — shave head	0.581	0.568
浇冷水 (jiāo lěngshuǐ) — pour cold water	0.480	0.483
有一手 (yǒu yī shǒu) — be skillful / have a trick	0.531	0.545
咬耳朵 (yǎo ěrduo) — whisper in ears	0.530	0.522
摇尾巴 (yáo wěibā) — wag tail / fawn on someone	0.514	0.505
踢皮球 (tī píqiú) — pass the buck / pass responsibility	0.520	0.498
讨生活 (tǎo shēnghuó) — make a living	0.558	0.545
乱弹琴 (luàn tánqín) — act recklessly / talk nonsense	0.541	0.535
露头角 (lù tóujiǎo) — begin to show one's talent	0.530	0.553
露马脚 (lù mǎjiǎo) — reveal one's hidden flaw	0.504	0.520
开夜车 (kāi yèchē) — work late / study late at night	0.505	0.523
打哈哈 (dǎ hāhā) — joke around / laugh off	0.511	0.528
出风头 (chū fēngtou) — show off / be in limelight	0.543	0.542
吃闲饭 (chī xiánfàn) — be idle / live idly	0.606	0.533
睡大觉 (shuì dàjiào) — sleep heavily / nap	0.516	0.521
Overall	0.547	0.530

Table B2: Significance tests for 25 Chinese VO constructions (paired t-test and Wilcoxon signed-rank test).

Idiom	t-statistic	t-p-value	w-statistic	w-p-value
擦屁股 (cā pìgu) — wipe someone's butt	-1.424	0.156	4470.000	0.333
戴帽子 (dài màozi) — wear a hat	0.480	0.631	740.000	0.371
照镜子 (zhào jìngzi) — look in the mirror	1.962	0.051	992.000	0.101
回老家 (huí lǎojiā) — return to hometown	1.628	0.105	472.000	0.615
拖下水 (tuō xiàshuǐ) — drag someone into trouble	2.638	0.009	4.000	0.109
开后门 (kāi hòumén) — use backdoor	-1.784	0.079	60.000	0.284
说梦话 (shuō mènghuà) — talk in one's sleep	0.087	0.931	134.000	0.443
啃骨头 (kěn gǔtou) — gnaw a bone	-0.468	0.644	29.000	0.470
做文章 (zuò wénzhāng) — make an issue / exploit for gain	0.797	0.428	0.000	0.250
开绿灯 (kāi lǜ dēng) — give the green light / permit	-1.029	0.310	11.000	0.688
剃光头 (tì guāngtóu) — shave head	0.561	0.578	4.000	0.875
浇冷水 (jiāo lěngshuǐ) — pour cold water	-0.232	0.817	77.000	0.490
有一手 (yǒu yī shǒu) — be skillful / have a trick	-2.741	0.007	131.000	0.021
咬耳朵 (yǎo ěrduo) — whisper in ears	0.902	0.370	7.000	0.297
摇尾巴 (yáo wěibā) — wag tail / fawn on someone	0.853	0.397	10.000	0.578
踢皮球 (tī píqiú) — pass the buck	1.952	0.056	3.000	0.313
讨生活 (tǎo shēnghuó) — make a living	0.731	0.466	0.000	0.500
乱弹琴 (luàn tánqín) — act recklessly / talk nonsense	0.576	0.567	9.000	0.844
露头角 (lù tóujiǎo) — begin to show one's talent	-1.143	0.253	0.000	0.500
露马脚 (lù mǎjiǎo) — reveal one's hidden flaw	-0.785	0.436	1.000	1.000
开夜车 (kāi yèchē) — work late / study late at night	-1.471	0.146	3.000	0.156
打哈哈 (dǎ hāhā) — joke around / laugh off	-1.098	0.275	0.000	0.500
出风头 (chū fēngtou) — show off / be in limelight	0.093	0.926	1.000	1.000
吃闲饭 (chī xiánfàn) — be idle / live idly	4.272	0.000	0.000	0.500
睡大觉 (shuì dàjiào) — sleep heavily / nap	-1.295	0.196	2745.000	0.286
Overall	10.046	0.000	512790.000	0.000

Table C1: Full dictionary-based sentiment values for all 25 Chinese VO constructions. '–' indicates that a literal explanation was not listed in the dictionary.

Idiom	Literal Sentiment	Idiom Sentiment
擦屁股 (cā pìgu) — wipe someone's butt	0.116	0.031
戴帽子 (dài màozi) — wear a hat; (idiom) be cuckolded	0.950	0.451
照镜子 (zhào jìngzi) — look in the mirror	0.923	0.661
回老家 (huí lǎojiā) — return to hometown	0.975	0.433
拖下水 (tuō xiàshuǐ) — drag someone into trouble	0.070	0.216
开后门 (kāi hòumén) — use backdoor / bribe	0.012	0.190
说梦话 (shuō mènghuà) — talk in one's sleep	0.508	0.465
啃骨头 (kěn gǔtou) — gnaw a bone	0.367	0.355
做文章 (zuò wénzhāng) — make an article / exploit for gain	0.999	0.227
开绿灯 (kāi lǜ dēng) — give the green light / permit	0.084	0.008
剃光头 (tì guāngtóu) — shave head	0.406	0.541
浇冷水 (jiāo lěngshuǐ) — pour cold water	0.077	0.342
有一手 (yǒu yī shǒu) — be skillful / have a trick	0.999	0.894
咬耳朵 (yǎo ěrduo) — whisper in ears	0.626	0.545
摇尾巴 (yáo wěibā) — wag tail / fawn on someone	0.999	0.073
踢皮球 (tī píqiú) — kick the ball / pass responsibility	0.990	0.215
讨生活 (tǎo shēnghuó) — make a living	–	0.618
乱弹琴 (luàn tán qín) — act recklessly / talk nonsense	0.891	0.018
露头角 (lù tóu jiǎo) — begin to show one's talent	0.797	0.719
露马脚 (lù mǎjiǎo) — expose one's hidden problem	0.614	0.270
开夜车 (kāi yèchē) — work/study late at night	0.224	0.189
打哈哈 (dǎ hāhā) — laugh off / pretend to agree	0.940	0.400
出风头 (chū fēngtóu) — show off / attract attention	0.493	0.094
吃闲饭 (chī xiánfàn) — live idly / do nothing productive	–	0.157
睡大觉 (shuì dàjiào) — sleep heavily / nap	0.784	0.406
Overall	0.634	0.341

Table C2: Full Pearson correlation between idiomatic sentiment and contextual embedding sentiment for all Chinese VO constructions.

Idiom	Pearson r	p-value
擦屁股 (cā pìgu) — wipe someone's butt	0.842	0.000
戴帽子 (dài màozi) — wear a hat; (idiom) be cuckolded	0.034	0.800
照镜子 (zhào jìngzi) — look in the mirror	0.153	0.201
回老家 (huí lǎojiā) — return to hometown	-0.061	0.691
拖下水 (tuō xiàshuǐ) — drag someone into trouble	0.955	0.001
开后门 (kāi hòumén) — use backdoor / bribe	-0.081	0.748
说梦话 (shuō mènghuà) — talk in one's sleep	-0.470	0.015
啃骨头 (kěn gǔtou) — gnaw a bone	-0.474	0.119
做文章 (zuò wénzhāng) — make an article / exploit for gain	0.828	0.379
开绿灯 (kāi lǜ dēng) — give the green light / permit	0.990	0.000
剃光头 (tì guāngtóu) — shave head	0.381	0.619
浇冷水 (jiāo lěngshuǐ) — pour cold water	0.020	0.936
有一手 (yǒu yī shǒu) — be skillful / have a trick	0.978	0.000
咬耳朵 (yǎo ěrduo) — whisper in ears	0.989	0.000
摇尾巴 (yáo wěibā) — wag tail / fawn on someone	0.489	0.266
踢皮球 (tī píqiú) — kick the ball / pass responsibility	0.907	0.034
乱弹琴 (luàn tán qín) — act recklessly / talk nonsense	0.992	0.000
露头角 (lù tóu jiǎo) — begin to show one's talent	1.000	1.000
露马脚 (lù mǎjiǎo) — expose one's hidden problem	1.000	1.000
开夜车 (kāi yèchē) — work/study late at night	0.997	0.000
打哈哈 (dǎ hāhā) — laugh off / pretend to agree	1.000	1.000
出风头 (chū fēngtóu) — show off / attract attention	1.000	1.000
睡大觉 (shuì dàjiào) — sleep heavily / nap	0.993	0.000
Overall	0.689	0.000