Annotation of Chinese Light Verb Constructions within UMR

Jingyi Li[1,2], Jin Zhao[2], Nianwen Xue[2], Shili Ge[1]
[1]Guangdong University of Foreign Studies
[2]Brandeis University
Corresponding author: lijingyi@mail.gdufs.edu.cn

Abstract

This paper discusses the challenges of annotating predicate-argument structures in Chinese light verb constructions (LVCs) within the Uniform Meaning Representation (UMR) framework, a cross-linguistic extension of Abstract Meaning Representation (AMR). A central challenge lies in reliably identifying LVCs in Chinese and determining their appropriate representation in UMR. We analyze the linguistic properties of Chinese LVCs, outline annotation difficulties for these structures and related constructions, and illustrate these issues through concrete examples. Our analysis focuses specifically on LVC.full types, where the light verb serves solely to convey morphological features and aspectual information. We exclude LVC.cause types, in which the light verb introduces an additional argument (e.g., a causal agent or source) to the event or state denoted by the nominal predicate. To address the practical challenge of semantic role assignment in Chinese LVCs, we propose a dual-path annotation approach: due to the compositional nature of these constructions, we recommend independently annotating the argument structure of the nominal predicate while systematically encoding the grammatical attributes and relations introduced by the light verb.

1 Introduction

The presentation of Light Verb Constructions (LVCs) continues to be a focal issue in both traditional linguistics and computational linguistics, garnering substantial attention over the years (Sag et al., 2002; Stevenson et al., 2004; Tu & Roth, 2011; Vincze et al., 2011; Nagy et al., 2020). LVCs are widely acknowledged as a universal linguistic phenomenon, composed of a verb—often referred to as "light"—paired with a single or compound predicative noun in its direct object position. The light verb makes only a minimal semantic contribution to the construction; instead, it primarily carries essential morphosyntactic properties such as person, number, tense, mood, and aspect (Savary et al., 2017; Bonn et al., 2023). Light verbs often exhibit unique and sometimes unpredictable behaviors across languages. Chinese light verbs, in particular, with their syntactic and semantic flexibility, combined with a distinctive distribution that sets them apart from regular verbs—which typically exhibit higher semantic content and more specific argument requirements—pose challenges for their identification and representation within various meaning representation frameworks. (Butt, 2010; Lin et al., 2014; Huang et al., 2014; Jiang et al., 2018; Bonn et al., 2023).

Uniform Meaning Representation (UMR), a recent graph-based framework designed to capture meaning across entire documents, provides promising opportunities for annotating LVCs, including those in Chinese, where compounding is a common word formation process (Bonn et al., 2023; Sun et al., 2023). Rooted in Abstract Meaning Representation (AMR), the fundamental components of a UMR graph are concepts and relations. At the sentence level, concepts typically map to words within a sentence, while at the document level, relations depict the semantic connections between these words (Bonn et al., 2024). At sentence level, UMR is flexible in allowing the use of both generic semantic roles, such as agent, theme, and patient, as well as predicate-specific roles, a practice widely adopted in the proposition bank approach to semantic role labeling (Xue and Palmer, 2009). Predicate-specific roles are defined in the PropBank Framesets, which provide entries for each predicate in a language (Xue and Palmer, 2005). Each predicate sense is assigned a unique set of core roles, labeled with numerical IDs prefixed by "Arg." For instance, the Chinese verb 认可 [renke, "accept"] has a first semantic frame, "认可-01," which defines the following roles:

Proceedings of the 23rd International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2025), pages 1–9

Arg0-agent: the entity described
Arg1-tar: the entity Arg0 accepts/ratifies

The Chinese verb 认可 [renke, “accept”] serves as an example where its defined roles can be applied to annotate occurrences of 认可 [renke, “accept”], even in contexts where some of its arguments are not explicitly stated. The LVC 获得 认可 [huode-renke, “get-accept”] in (1) can be annotated in UMR as follows:

(1) 这 一 方法 获得-认可 。
this one method get-accept .
“This method got accepted.”

(s1x / 认可-01
:Arg1 (s1x2 / 方法 [fangfa, “method”]
:mod (s1x3 / 这 [zhe, “this”]))
:Aspect Performance
:MODSTR FullAff

The absence of clear morphological distinctions between certain Chinese nouns and their verbal counterparts, such as 认 可 [renke, “accept”], allows these lexical items to serve both nominal and verbal roles. In example (1), we identify 认可 as the main predicate rather than 获得 [huode, “get”], annotating 方法 [fangfa, “method”] as the Arg1 of 认可 [renke, “accept”]. This annotation choice reflects that the verb 获得 [huode, “get”] acts solely as a grammatical marker indicating successful completion of the event, without adding significant semantic content.

Such light verbs have received less scholarly attention as they mostly function as regular verbs and require clear linguistic features for accurate identification. Although comparative studies have examined LVCs in English and Chinese and explored variations across major Chinese-speaking regions, such as Hong Kong, Taiwan, and Beijing (Lin et al., 2014; Huang et al., 2014; Jiang et al., 2018; Tsou and Yip, 2020; Lu, 2016), research has primarily concentrated on specific verb groups, notably “do” (做 [zuo], 干 [gan], 搞 [gao]) and “give” (加 以 [jiayi], 予 以 [yuyi]). Therefore, further investigation into other less commonly studied LVC types is needed to enrich both linguistic analysis and computational modeling.

The absence of clear morphological distinctions between certain Chinese nouns and their verbal counterparts, along with the intricate modifiers and argument structures of nominal complements, makes annotating Chinese LVCs particularly challenging (Wang et al., 2023). These complexities highlight the need to strike a balance that ensures consistency across different types of Chinese LVCs—a task that is both essential and demanding. We adopt broad criteria for LVC annotation from the PARSEME guidelines, a European project aimed at processing multiword expressions, including LVCs (Savary et al., 2017). Jiang et al. (2018) applied these guidelines to the automatic tagging of Chinese light verbs and introduced valuable adaptations. Nevertheless, their research mainly focuses on tagging a restricted set of light verbs in the corpus and lacks detailed representations of LVCs within specific linguistic contexts.

The rest of the paper is structured as follows. Section 2 examines linguistics properties of Chinese LVCs within the UMR framework. Section 3 introduces refined criteria for systematically identifying these constructions. Section 4 highlights the distinctions between Chinese LVCs and causative constructions. Finally, Section 5 concludes the paper by summarizing the key findings and suggesting directions for future research. These contributions aim to improve UMR annotation practices and deepen the understanding of Chinese LVCs in semantic representation frameworks.

2 Linguistic Properties of LVCs

In this section, we set aside highly grammaticalized light verbs, such as the “do” and “give” groups, to focus on syntactic and semantic structure of vague cut cases of LVCs in Chinese and examine their diverse patterns.

2.1 Argument Structure

In UMR, light verbs are treated as having zero arguments, similar to auxiliary verbs, which also lack an argument structure (Xue and Palmer, 2005). The primary function of light verbs is to provide grammatical or aspectual support to the nominal predicate, which holds the core semantic content and carries the associated arguments (Bonn et al., 2023). The argument structure of an LVC thus depends entirely on the nominal predicate, which can have zero, one, or multiple arguments.

In Chinese, some nominal predicates naturally occur without requiring main arguments. For

example, the nominal predicate 爆炸 [baozha, "explode"] in sentence (2) can appear alone or together with the verb 发生 [fasheng, "occur"]. When 发生 is used, it explicitly indicates that the explosion event took place, allowing the introduction of specific details such as the time and location of the event. However, adding or omitting 发生 [fasheng, "occur"] does not change the fundamental meaning of the sentence: either way, the proposition remains that an explosion took place at the concert. Thus, 发生 [fasheng, "occur"] is considered a "light verb," as it does not contribute substantial new propositional content beyond signaling the occurrence of an event.

(2) 演唱会 于 22 时 33 分 发生-爆炸 。
concert at 22:33 occur-explode .
"An explosion occurred at the concert at 22:33."

(s2x / 爆炸-01 [baozha, "explode"]
:place (s2x2 / 演唱会 [yanchanghui, "concert"])
:temporal (s2d / date-entity
:time "h22m33")
:Aspect Performance
:MODSTR FullAff

Certain nominal predicates involve exactly one semantic argument. Example (3) illustrates this clearly:

(3) 该 团队 率先 取得-胜利 。
the team take the lead get-success .
"The team was the first to achieve victory."

(s3x / 胜利-01 [shengli, "success"]
:ARG0 (s3x2 / 团队 [tuandui, "team"]
:mod (s3x3 / 该 [gai, "the"]))
:mod (s3x4 / 率先 [shuaixian, "take the lead"])
:Aspect Performance
:MODSTR FullAff)

The nominal predicate 胜利 [shengli, "success"] inherently involves one argument—the entity experiencing or achieving success (the team). The accompanying light verb 取得 [qude, "get"] does not introduce any additional arguments; it merely serves as a grammatical connector that enhances fluency. The UMR clearly annotates the team as ARG0, underscoring that the nominal predicate's single argument structure is preserved while the light verb remains semantically redundant.

Other nominal predicates can take multiple semantic arguments. Consider example (4):

(4) 科学家 对 遗骸 进行-检查。
scientists towards remains proceed-exam
"Scientists conducted an examination of the remains."

(s4x / 检查-01 [jiancha, "exam"]
:ARG0 (s4x3 / 科学家 [kexuejia, "scientists"]
:ARG1 (s4x5 / 遗骸 [yihai, "remains"]
:Aspect Performance
:MODSTR FullAff)

In this example, the nominal predicate 检查 [jiancha, "exam"] requires two semantic arguments: the agent performing the action (科学家 [kexuejia, "scientists"]) and the object of the action (遗骸 [yihai, "remains"]). The light verb 进行 [jinxing, "proceed"] contributes no independent semantic argument structure.

2.2 Adverbial and Attributive Modification

The incorporation of modifiers into Chinese LVCs significantly increases their structural complexity. Because Chinese adjectives can serve either attributively or adverbially, two distinct modification patterns emerge within LVCs. A modifier may directly describe the nominal predicate alone (attributive modification), or it may adverbially modify the entire LVC, thereby altering the interpretation of the entire event.

However, the possibility for a modifier to extend its scope beyond the nominal predicate should not be considered a definitive criterion in determining whether a construction qualifies as an LVC. Consider Example (5):

(5) 他们 展开 了 激烈的 争吵 。
They engage in PF intense dispute .
"They engaged in an intense dispute."

(s5x / 争吵-01 [zhengchao, "dispute"]
:ARG0 (s5p / person
:refer-person 3rd
:refer-number Plural)
:manner (s5x5 / 激烈 [jilie, "intense"])
:Aspect Performance
:MODSTR FullAff)

The adjective 激烈 [jilie, “intense”] allows two possible interpretations: it either describes only the nominal predicate 争吵 [zhengchao, “dispute”], resulting in “intense dispute,” or it modifies the entire event described by 展开争吵 [zhankai-zhengchao, “engage in dispute”], producing the reading “engaged intensely in a dispute).” Regardless of this ambiguity in interpretation, the argument structure remains stable, governed solely by the nominal predicate 争吵 [zhengchao, “dispute”], while the light verb 展开 [zhankai, “engage”] does not introduce any additional arguments.

Modifier placement further complicates the syntax of LVCs. Modifiers need not remain adjacent to their modified element; rather, they can freely occur either before, within, or after the construction. Example (6) illustrates a temporal modifier placed at the end of an LVC:

(6) 该团队 对其 进行-研究 长达 十年。
the team on it conduct research long ten years
“The team conducted research on it for as long as ten years.”

(s6x / 研究-01 [yanjiu, “research”]
:ARG0 (s6x2 / 团队 [tuandui, “team”]
:mod (s6x3 / 该 [gai, “the”]))
:ARG1 (s6x4 / 其 [qi, “it”]
:duration (s6t / temporal-quantity
:mod (s6x6 / 长 [chang, “long”])
:quant 10
:unit (s6x7 / 年 [nian, “year”]))
:Aspect Process
:MODSTR FullAff)

The temporal modifier 十年 [shinian, “ten years”] appears at the end of the construction yet semantically specifies the duration of the nominal predicate 研究 [yanjiu, “research”]. Despite this non-adjacent surface placement, the UMR annotation maintains consistency by explicitly linking this temporal element directly to the nominal predicate, highlighting the event’s duration rather than completion.

Reflexive modifiers introduce additional layers of interpretation complexity. In Examples (7) and (8), the reflexive 他们自己 [tamen-ziji, “themselves”] demonstrates ambiguity contingent upon syntactic placement. When the reflexive modifier follows the nominal predicate, as in (7), it clearly signals possession:

(7) 工人们 取得 了 他们自己的 胜利。
workers achieve PF their-own victory
“The workers achieved their own victory.”

(s7x / 胜利-01 [shengli, “victory”]
:ARG0 (s7x2 / 工人 [gongren, “workers”]
:refer-number Plural)
:poss-of (s7x3 / 他 们 自 己 “tamen-ziji, themselves”)
:Aspect Performance
:MODSTR FullAff)

However, placing the reflexive before the entire LVC, as in (8), conveys collective agency rather than possession.

(8) 工人们 他们自己 取得 了 胜利。
workers themselves achieve PF victory
“The workers themselves achieved victory.”

(s8x / 胜利-01 [shengli, “victory”]
:ARG0 (s8x2 / 工人 [gongren, “workers”]
:refer-number Plural)
:Aspect Performance
:MODSTR FullAff)

2.3 Inherent Aspect

Events expressed through nominal constructions often pose significant challenges for aspectual annotation, primarily because they lack the explicit morphological or syntactic markers that typically signal aspectual distinctions. In LVCs, light verbs frequently combine with nominal predicates, thereby clarifying or altering the aspectual interpretation. While the default assumption might be that the aspect of a light verb aligns seamlessly with the nominal event it accompanies, discrepancies can occur and warrant careful analysis. For instance, in many instances, the light verb and the nominal share the same aspectual value, as exemplified by (10), where both 给予 [jiyu, “offer] and 帮助 [bangzhu, “assist”] converge on a process aspect indicating ongoing activity of offering help. Such compatibility between the lexical aspect of the light verbs and the nominal predicate typically simplifies the

annotation process because it provides a clear indication that the event has a definite end point.

(9) 慈善机构 承诺 给予-帮助。
charity organization promise offer-assist
"The charity organization promised to offer assistance."

(s10x / 承诺-01 [chengnuo, "promise"]
:ARG0 (s10x2 / 机构 [jigou, "organization"]
:mod (s10x4 / 慈善 [cishan, "charity"]))
:ARG1 (s10x3 / 帮助-01 [bangzhu, "assist"]
:Aspect Process
:MODSTR FullAff)
:Aspect Performance
:MODSTR FullAff)

However, more nuanced cases arise when the inherent aspect of the light verb diverges from that of the nominal event. Example (10) illustrates this situation: the nominal event 会谈 [huitan, "meet"] is intrinsically durational, unfolding over a three-hour span, thus suggesting an ongoing process. By contrast, the light verb 举行 [juxing, "hold"] tends to convey a sense of a discrete and complete occurrence—what can be classified as a performance aspect in UMR. When these two aspectual profiles come together in an LVC, the annotation must account for the fact that the event unfolds over a span of time but also concludes definitively once the meeting has taken place.

(10) 双方 举行 三小时 会谈。
both sides hold 3 hours meet
"Both sides held a three-hour meeting."

(s11x / 会谈-01 [huitan, "meet"]
:ARG0 (s11x2 / 双方 [shuangfang, "both sides"]
:temporal (s11t / temporal-quantity
:quant 3
:unit (s11x3 / 小时 [xiaoshi, "hour"]))
:Aspect Performance
:MODSTR FullAff)

2.4 Existential and Passive Oriented

In Chinese, there are two special sentence patterns that closely relate to LVCs: those oriented toward existence and those oriented toward passivity. The first type includes examples with the verbs 有 [you, "have"] or 存在 [cunzai, "exist"], both of which can function as light verbs in specific contexts. It is important to distinguish these uses from the typical Chinese existential you-construction, which parallels the English there-construction and expresses the existence, appearance, or disappearance of entities at a particular place or time. Consider the example in (11): although 存在 [cunzai, "exist"] typically means 'exist,' it does not convey its usual existential meaning but instead serves as a light verb. In this usage, it indicates a static relational state between the arguments rather than literal existence. The lexical meaning of 存在 [cunzai, "exist"] is bleached, and it instead operates as a grammatical marker that highlights the aspectual or stative nature of the nominal predicate.

(11) 他 与 袭击案 存在-关联 。
he with the attack case exist-connect .
"He is connected to the attack case."

(s13x / 关联-01 [guanlian, "connect"]
:ARG0 (s13p / person
:refer-person 3rd
:refer-number Singular)
:ARG1 (s13x2 / 袭击案 [xiji-an, "attack case"])
:Aspect State)

The second category involves passive-oriented syntactic patterns, where the grammatical structure shifts from an active to a passive voice while preserving the core semantic representation. Crucially, this syntactic alternation—exemplified by passive markers such as 受到 [shoudao, "undergo"], 遭到 [zaodao, "suffer"], and 被 [bei, "be"]—does not alter the thematic roles or event structure of the nominal predicate. Light verbs act as voice heads that license syntactic reconfiguration without modifying lexical-semantic content. In (12), while the passive construction elevates the patient 敌人 [diren, "enemy"] to subject position, the nominal predicate 攻击 [gongji, "attack"] remains the semantic core of the event. The light verb 受到 [shoudao, "undergo"] functions solely to signal passive voice and suppress roles.

(12) 敌人 受到 猛烈 攻击。
enemy undergo fierce attack
"The enemy was heavily attacked."

(s15x / 攻击-01 [gongji, "attack"]

:ARG1 (s15x2 / 敌人 [diren, “enemy”])
:manner (s15x3 / 猛烈 [menglie, “fierce”])
:Aspect Performance
:MODSTR FullAff)

3 Broad criteria for determining LVCs

Building on the tests developed by PARSEME, the UMR annotation guidelines for LVCs in Chinese introduce four specific tests to determine whether a verb with a predicative noun as complement qualifies as an LVC.

3.1 Test 1

Test 1 evaluates whether the complement of a light verb is a predicative noun. For example, in the phrase “make a contribution,” the noun “contribution” is predicative because it represents an event or action that corresponds to the verb “contribute.” Conversely, in “make a cake,” the noun “cake” is not predicative, as it does not have a verbal counterpart. Therefore, the former passes Test 1 and proceeds to the next stage, while the latter is excluded.

Notably, verbs are often mistaken for predicative nouns in Chinese, primarily because of the unmarked morphological status shared by predicative nouns and their verbal counterparts. For instance, in (13), the verb combination 引用报导 specifically conveys the meaning “to be cited and reported,” rather than suggesting that Chinese media cited reports or the reporting event created by other outlets. In the latter interpretation, 报导 [baodao, “report”] would act as a predicative noun. However, in this context, it does not meet the requirements of Test 1, as it functions as a verb.

(13) 中文 媒体 引用 报导 该 新闻。
Chinese media cite report the news
“The Chinese media cited and reported the news.”

(s16a / and
:op1 (s16x / 引用-01 [yinyong, “cite”]
:ARG0 (s16x2 / 媒体 [meiti, “media”]
:medium (s16x3 / 中文 [zhongwen, “Chinese”]))
:ARG1 (s16x4 / 新闻 [xinwen, “news”]
:mod (s16x6 / 该 [gai, “the”])))
:Aspect Performance
:MODSTR FullAff)
:op2 (s16x7 / 报导-01 [baodao, “report”]
:ARG0 s16x2
:ARG1 s16x4
:Aspect Performance
:MODSTR FullAff))

3.2 Test 2

Test 2 assesses whether the subject of a verb within the construction also functions as an argument of the nominal predicate. For instance, in the sentence “made a presentation to his boss,” the subject of the verb “make” serves as the agent of the nominal predicate “presentation,” thereby satisfying Test 2. Conversely, in “John’s boss interrupted his presentation,” the subject “John’s boss” does not hold a thematic role related to the nominal predicate “presentation,” resulting in a failure to meet Test 2.

In most cases, constructions that pass Test 1 also pass Test 2. However, exceptions do exist. For instance, in (14), the verb 支持 [zhichi, “support”] in the expression 支持反恐 [zhichi-fankong, “support counter-terrorism”] presents a counterexample. In this case, the subject of 支持 [zhichi, “support”] is not inherently an argument of the nominal predicate 反恐 [fankong, “counter-terrorism”], as it does not directly engage in anti-terrorism actions. Rather, the subject expresses an external stance of approval or endorsement, without participating in the actual event, thus failing Test 2.

(14) 清真寺 曾 就 支持 反恐
Mosque once on support counter-terrorism
和 生命尊严 布告。
and life dignity issue a statement
“The mosque once issued a statement supporting counter-terrorism and the dignity of life.”

(s17x / 布告-00 [bugao, “issue a statement”]
:ARG0 (s17x2 / 清真寺 [qingzhensi, “Mosque”])
:ARG1 (s17a / and
:op1 (s17x4 / 支持-01 [zhichi, “support”]
:ARG1 (s17x5 / 反-01 [fan, “counter”]
:ARG1 (s17x6 / 恐 [kong, “terrorism”])))
:op2 (s17x7 / 尊严 [zunyan, “dignity”]

:mod (s17x8 / 生命 [shengming, "life"])))
:mod (s17x3 / 曾 [ceng, "once"])
:Aspect Performance
:MODSTR FullAff)

3.3 Test 3

Test 3 is designed to determine whether a given verb introduces substantial semantic content beyond merely hosting morphological features such as tense, mood, and aspect, or contributing syntactic structure for the nominal predicate. In essence, this test seeks to distinguish genuinely "light" verbs from those that add meaningful lexical semantics. If a verb simply facilitates the nominal predicate's argument structure or supplies grammatical inflections without introducing new propositional content, it can be considered light.

In UMR, applying Test 3 is relatively straightforward. If removing the verb does not alter the core propositional meaning, the verb can be classified as light. However, if the omission leads to a loss or shift in essential semantic content, the verb is not considered light. For example, in (15), the verb 引起 [yinqi, "draw"] contributes more than just grammatical support—it introduces the causative meaning of bringing about attention. This is semantically richer than the role of a typical light verb, which would merely provide aspectual or syntactic support to the nominal predicate 注意 [zhuyi, "attention"] without adding new event content. Thus, 引起 [yinqi, "draw"] fails the test for lightness, as it adds distinct lexical meaning to the clause.

(15) 他 已 引起 情报部门
he already draw the intelligence agency
注意 。
attention.
"He has drawn the attention of the intelligence department."

(s19x / 引起-01 [yinqi, "draw"]
:ARG0 (s19p / person
:refer-person 3rd
:refer-number Singular)
:ARG1 (s19x2 / 注意-01 [zhuyi, "attention"]
:ARG0 (s19x3 / 部门 [bumen, "agency"]
:mod (s19x4 / 情报 [qingbao, "intelligence"]))
:Aspect State
:MODSTR FullAff)
:mod (s19x5 / 已 [yi, "already"])
:Aspect Performance
:MODSTR FullAff)

4 Causative Constructions

Certain verb constructions in Chinese resemble LVCs in form or function, particularly those expressing causative relations, but they do not fully satisfy the core definitional criteria of LVCs. While not the primary focus of this study, these constructions merit careful consideration, as their syntactic and semantic characteristics can easily be mistaken for genuine LVCs.

In certain complex transitive verb constructions that can be interpreted as externally caused events, the process of causativization consistently appears to be feasible. Basciano (2013), for instance, highlights verbs such as 弄醒 [nongxing, "wake"), 弄哭 [nongku, "make cry"], 搞丢 [gaodiu, "lose"], and 搞坏 [gaohuai, "destroy"), all of which demonstrate this pattern. Similarly, Chung (2006) investigates verbs containing the root 加 [jia, "add"], including 加宽 [jiakuan, "widen"], 加高 [jiagao, "heighten"], and 加强 [jiaqiang, "to strengthen"], observing that the 加 [jia, "add"] facilitates the formation of transitive counterparts of change-of-state verbs derived from open-scale adjectives that denote incremental increases.

Causative constructions inherently involve distinct semantic roles for both the causing event and the resultant state (Tham, 2015; Sun et al., 2023). Therefore, the constituent verbs in such constructions function as independent predicates, each maintaining its own argument structure. This property distinguishes causative verb compounds clearly from true LVCs, even if one verb appears semantically "lighter" than the other. Specifically, the so-called "light" verb in causative compounds still contributes a distinct argument structure, disqualifying it from classification as a genuine light verb. For example, in (16), the verb compound 打破 [dapo, "break"] encodes two separate events: an action event 打 [da, "beat"] and a resultant state 破 [po, "break"]. Thus, 打 [da, "beat"] denotes a causing action, and 破 [po, "break"] expresses the resulting event. This forms a compositional resultative construction rather than an LVC.

(16) 他 打-破 了 桌上的 花瓶。
he beat-break PF on the table vase
"He broke the vase on the table."

(s21x / 打-015 [da, "beat"]
:ARG0 (s21p / person
:refer-person 3rd
:refer-number Singular)
:Cause-of (s21x2 / 破-04 [po, "break"]
:ARG0 (s21x3 / 花瓶 [heaping, "vase"]
:place (s21x4 / 桌 上 [zhuozi, "table"))
:Aspect State
:MODSTR FullAff)
:Aspect Performance
:MODSTR FullAff)

From a semantic perspective, metaphorization involves extending a verb's literal, physical meaning into a more abstract domain. In Chinese, certain verbs display such metaphorization, making their identification and annotation more challenging. The same verb compound 打破 [dapo, "break"] can function in both a literal sense, as seen in (16), and a metaphorical sense in (17), where 打破 [dapo, "break"] is best interpreted as expressing the disruption or alteration of an abstract state. In UMR annotation, the first frame of 打破 [dapo, "break"] treats the bird sound (ARG0) as the agent and the stillness (ARG1) as the theme.

(17) 鸟声 打破 了 清晨的宁静。
the sounds of birds break PF tranquility of the early morning
"The sound of birds broke the tranquility of the early morning."

(s22x / 打破-01 [dapo, "break"]
:ARG0 (s22x2 / 鸟声 [niaosheng, "the sound of birds"])
:ARG1 (s22x3 / 宁静 [ningjing, "tranquility"]
:temporal (s22x4 / 清晨 [qingchen, "the early morning"]))
:Aspect Performance
:MODSTR FullAff)
:quant (s26x4 / 多)))
:ARG1 (s26x5 / 慰问)
:Aspect Performance
:MODSTR FullAff)

5 Conclusion

In this paper, we have explored the challenges inherent in annotating Chinese light verb constructions within the Uniform Meaning Representation framework. Through an analysis of their structural and semantic characteristics, we illustrated how the syntactic flexibility of Chinese light verbs complicates accurate annotation. We addressed key issues such as identifying LVCs, annotating argument structures, and distinguishing these constructions from similar forms. Our findings reinforce prior research (Savary et al., 2017; Lin et al., 2014), confirming that Chinese light verbs primarily fulfill grammatical roles rather than contributing substantive semantic meaning. Nevertheless, their distinctive syntactic versatility calls for refined annotation guidelines to mitigate potential misclassifications. To address this, we proposed a dual-path annotation method, separately encoding the argument structure of nominal predicates and the grammatical properties of light verbs. This methodology sets the stage for future studies to investigate the intricate syntactic, semantic, and contextual dimensions of LVCs. Ultimately, our work aims to enhance both linguistic research and computational modeling of Chinese and other languages exhibiting similar complexities.

Acknowledgments

This work was supported by a grant from the CNS Division of the National Science Foundation (Award No. 2213804), titled "Building a Broad Infrastructure for Uniform Meaning Representations." Additional financial support was provided by the Guangdong Provincial Philosophy and Social Science Planning General Project (Grant No. GD23CWY10) and a research grant (No. LAI202302) from the Laboratory of Language and Artificial Intelligence, Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies. The views expressed in this paper are those of the authors and do not necessarily reflect those of the National Science Foundation, the Guangdong Provincial Philosophy and Social Science Planning Office, or the Laboratory of Language and Artificial Intelligence. We are grateful to William Croft, James Pustejovsky, Martha Palmer, Andrew Cowell, and other collaborators on the UMR project for their insightful discussions. We also

thank the annotators of the Chinese UMR team, particularly Haibo Sun, Hsiao-Jung Cheng, and Dan Xing for raising important issues related to the annotation of Chinese Light Verb Constructions. All errors and omissions remain the responsibility of the authors.

References

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. 2002. Multiword expressions: A pain in the neck for NLP. In Computational Linguistics and Intelligent Text Processing: Third International Conference, CICLing 2002, Mexico City (pp. 1–15).

Stevenson, S., Fazly, A., & North, R. 2004. Statistical measures of the semi-productivity of light verb constructions. In Proceedings of the Workshop on Multiword Expressions: Integrating Processing (pp. 1–8).

Tu, Y., & Roth, D. 2011. Learning English light verb constructions: Contextual or statistical. In Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World (pp. 31–39).

Vincze, V., Nagy, I., & Berend, G. 2011. Detecting noun compounds and light verb constructions: A contrastive study. In Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World (pp. 116–121).

Nagy, I., Rácz, A., & Vincze, V. 2020. Detecting light verb constructions across languages. Natural Language Engineering, 26(3), 319–348.

Savary, A., Ramisch, C., Cordeiro, S. R., Sangati, F., Vincze, V., Qasemi Zadeh, B., Candito, M., Cap, F., Giouli, V., Stoyanova, I., & Doucet, A. 2017. The PARSEME shared task on automatic identification of verbal multiword expressions. In Proceedings of the 13th Workshop on Multiword Expressions at EACL (pp. 31–47).

Bonn, J., Cowell, A., Hajic, J., Palmer, A., Palmer, M., Pustejovsky, J., Sun, H., Uresova, Z., Wein, S., Xue, N., & Jin, Z. 2023. UMR annotation of multiword expressions. In Proceedings of the 4th International Workshop on Designing Meaning Representations.

Butt, M. 2010. The light verb jungle: Still hacking away. In Complex Predicates in Cross-Linguistic Perspective (pp. 48–78).

Lin, J., et al. 2014. Annotation and classification of light verbs and light verb variations in Mandarin Chinese. In Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing.

Huang, C. R., Lin, J., Jiang, M., & Xu, H. 2014. Corpus-based study and identification of Mandarin Chinese light verb variations. In Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects (pp. 1–10).

Jiang, M., Klyueva, N., Xu, H., & Huang, C. R. 2018. Annotating Chinese light verb constructions according to PARSEME guidelines.

Sun, H., Zhu, Y., Zhao, J., & Xue, N. 2023. UMR annotation of Chinese verb compounds and related constructions. In Proceedings of the First International Workshop on Construction Grammars and NLP (pp. 75–84).

Bonn, J., Buchholz, M. J., Chun, J., Cowell, A., Croft, W., Denk, L., Ge, S., Hajic, J., Lai, K., Martin, J. H., & Myers, S. 2024. Building a broad infrastructure for Uniform Meaning Representations. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 2537–2547).

Xue, N., & Palmer, M. 2005. Automatic semantic role labeling for Chinese verbs. In Proceedings of IJCAI (pp. 1160–1165).

Tsou, B. K., & Yip, K. F. 2020. A corpus-based comparative study of light verbs in three Chinese speech communities. In Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation (pp. 302–311).

Lu, L. 2016. Information structure in Mandarin Chinese light verb constructions. SOAS Working Papers in Linguistics, 18, 53–74.

Wang, D., Jiang, G., & Zheng, Y. 2023. Walking out of the light verb jungle: Exploring the translation strategies of light verb constructions in Chinese–English consecutive interpreting. Frontiers in Psychology, 14, 1113973.

Basciano, B. 2013. Causative light verbs in Mandarin Chinese (and beyond). In Morphology in Toulouse: Selected Proceedings of Décembrettes, 7(1), 57–89.

Chung, K. S. 2006. Mandarin compound verbs. Crane Publishing.

Tham, S. W. 2015. Resultative verb compounds in Mandarin. In The Oxford Handbook of Chinese Linguistics (pp. 306–321).