UMR annotation of Chinese Verb compounds and related constructions

Haibo Sun[†], **Yifan Zhu**[*], **Jin Zhao**[◇], **Nianwen Xue**[‡]
Brandeis University
{ hsun[†],zhuyifan[*],jinzhao[◇],xuen[‡]}@brandeis.edu

Abstract

This paper discusses the challenges of annotating the predicate-argument structure of Chinese verb compounds in Uniform Meaning Representation (UMR), a recent meaning representation framework that extends Abstract Meaning Representation (AMR) to cross-linguistic settings. The key issue is to decide whether to annotate the argument structure of a verb compound as a whole, or to annotate the argument structure of their component verbs as well as the relations between them. We examine different types of Chinese verb compounds, and propose how to annotate them based on the principle of compositionality, level of grammaticalization, and productivity of component verbs. We propose a solution to the practical problem of having to define the semantic roles for Chinese verb compounds that are quite open-ended by separating compositional verb compounds from verb compounds that are non-compositional or have grammaticalized verb components. For compositional verb compounds, instead of annotating the argument structure of the verb compound as a whole, we annotate the argument structure of the component verbs as well as the semantic relations between them as creating an exhaustive list of such verb compounds is infeasible. Verb compounds with grammaticalized verb components also tend to be productive and we represent grammaticalized verb compounds as either attributes of the primary verb or as relations.

1 Introduction

Uniform Meaning Representation (UMR) (Gysel et al., 2021) is a meaning representation designed to annotate the semantic content of a text and it pairs a sentence-level representation with a document-level representation. Its sentence-level representation adopts the predicate-argument structure backbone of Abstract Meaning Representation (AMR) (Banarescu et al., 2013) but extends it to cross-linguistic settings by providing shared concepts and relations that can be applied cross-linguistically, particularly to morphologically complex low-resource languages. UMR also adds a document-level representation that captures linguistic phenomena such as coreference as well as temporal (Zhang and Xue, 2018; Yao et al., 2020) and modal dependencies (Vigus et al., 2019) that potentially go beyond sentence boundaries .

For the predicate-argument structure annotation, UMR is flexible in allowing the use of both generic semantic roles such as agent, theme, patient and predicate-specific roles, a practice popularized by the proposition bank approach to semantic role labeling (Palmer et al., 2005; Xue and Palmer, 2009). The predicate-specific roles in the propbanks are defined in *frame files* that have entries for each predicate in a language. For each sense of a predicate, a set of *core* roles are assigned unique numerical IDs that are prefixed by *Arg*. For instance, the non-polysemous English verb "sink" has the following roles:

Arg0: causer of sinking
Arg1: thing sinking
Arg2: extent
Arg3: start point
Arg4: end point, destination
Arg5: instrument

These roles can then be used to annotate instances of "sink", where not all arguments of sink may be realized:

(1) The enemy sank the ship.
(s / sink-01
:Arg0 (e / enemy)
:Arg1 (s2 / ship
:ref-number Singular)
:aspect Performance)

Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023), pages 75 - 84

For languages like Chinese where compounding is a robust word formation process (Packard, 2000), the predicate-specific approach of semantic role annotation in UMR provides both opportunities and challenges. For verb compounds that consist of verbs that each have their own argument structure, they can be represented in a straightforward manner in UMR, as shown in (2), where each component verb 哭 [ku, "cry"][1] and 湿 [shi, "wet"] has its own argument structure, and the semantic relation between them is one of *:cause-of*, indicating the former is the cause of the latter.

(2) 他 把 手帕 哭-湿 了
he ACC handkerchief cry-wet PF

"He cried so much that his handkerchief is wet."[2]

```
(x4a / 哭-01[ku, "cry"]
      :Arg0 (i / individual-person
            :ref-person 3rd
            :ref-number Singular)
      :cause-of (x4b / 湿-01[shi, "wet"]
                  :Arg0 (x3 / 手帕 [shou-pa,
                              "handkerchief"]))
      :aspect Performance)
```

The challenge, however, is that Chinese verb compounds involve various degrees of grammaticalization and idiomaticity, and it is not always appropriate to treat component verbs in a compound as separate predicates that each have their argument structures. In order for there to be consistent annotation, there needs to be a set of criteria that the annotator follows when determining which verb compounds should be treated holistically as having a single argument structure and which should have separate predicate-argument structures for their component verbs.

In this paper, we will examine different types of verb compounds and propose how we will annotate them in UMR. When deciding if a verb compound needs to have the argument structure of their component verbs annotated, we consider idiomaticity (or non-compositionality), levels of grammaticalization, and productivity. For instance, when a verb compound becomes highly idiomatic and its meaning as a whole cannot be predicted from their component verbs, it no longer makes sense to annotate the predicate-argument structure of the component verbs. Similarly, when a component verb in a verb compound is highly grammaticalized and its lexical content is semantically "bleached", there is less value in annotating the predicate-argument structure of this component verb, and it is more appropriate to treat them as attributes for the primary predicates or as relations between the primary predicates and one of its arguments.

When examining these verb compounds, we will classify them into broad categories based on syntactic and semantic relations between their component verbs, as they have been standardly done in linguistic annotation work (Xue et al., 2005). Here we focus on verb compounds that consist of two verbs, in the form of V_1 and V_2. They include resultative (VR) and directional verb (VD) compounds, subordinating compounds (VSB) in which the first verb modifies the second verb, coordinating compounds (VCD) in which the verbs either happen sequentially or have an equal status, and verb compounds that have the second verb as a copula verb (VCP). We will also examine the UMR annotation of light verb constructions that are similar in form but not content to verb compounds.

The rest of the paper is organized as follows. In Section 2, we examine different types of verb compounds and show how they are treated in UMR. In Section 3, we discuss how to annotate related verb constructions in UMR. We discuss related work in Section 4, and conclude in Section 5.

2 Types of verb compounds

In this section we examine different types of verb compounds, and show how we plan to annotate them in UMR.

2.1 Resultative verb compounds

Resultative verb compounds are a type of verb compounds that have been discussed extensively in linguistic literature (Thompson, 1973; Lu, 1977; Li, 1990; Packard, 2000). In general, resultative verb compounds are in the form of V_1 and V_2 where V_2 is broadly considered to be the result of V_1. As we will show, however, the semantic relation between the two component verbs tends to be quite diverse and is not always strictly

[1]Throughout the paper, the pinyin and translation in square brackets are not part of the UMR annotation and are merely provided for readability

[2]The glossing abbreviations used in this paper are: PF: perfective aspect, PRG: progressive aspect, EXP: experiential aspect, CL: classifier, ACC: accusative case marker

one of cause and result. Similarly, there is also quite a bit of variability in the argument structure of each component verb. In addition, the meaning of some resultative verb compounds cannot be predicted from their component verbs, and they are thus non-compositional, in which case the argument structure as a whole should be annotated. In other cases, one of their component verbs are grammaticalized to a certain degree. When this happens, it also makes sense to annotate the argument structure of the verb compound as a whole rather than the argument structure of each component verb.

2.1.1 Compositional Resultative Compounds

When resultative verb compounds are compositional, the argument structure of the component verbs is annotated. This is illustrated in (3), where the athlete's running lead to her shoes being broken. V_1 跑 [pao, "running"] is the cause, and V_2 坏 [huai, "break"] is the result. In UMR, this relation is labeled as **:cause-of** indicating that the first predicate is the cause of the second predicate, or conversely, the second predicate is the result of the first predicate.

(3) 运动员 跑坏 了 鞋 。
Athlete run-break PF shoe .

"The athlete broke (her/his) shoes because of running."

```
(x2b / 跑-01 [pao, "running"]
      :Arg0 (x1 / 运动员 [ yundongyuan,
                  "athlete"]
      :cause-of (x2a / 坏-01[huai, "break"]
                  :Arg0 (x4 / 鞋 [xie, "shoe"])))
```

By annotating the argument structure of the component verbs, we obviate the need to create a frame file for the verb compound as a whole, in addition to those for the component verbs. However, the annotator needs a reliable compositionality test to determine if this verb compound is compositional. We can test the compositionality of this verb compound by paraphrasing this sentence as ''运动员 ["athlete"] 跑 ["run"]" , and ''鞋子 ["shoe"] 坏 ["break"] 了 [PF]". If the component verbs ''跑 ["run"]" and ''坏 ["break"]" have the same meaning in the paraphrase as they do in the verb compound, then we know this verb compound is compositional. Otherwise it is not.

Another compositional verb compound example is provided in (4):

(4) 他 买-亏 了 这 衣服
he buy-loss PF this clothes

"He bought this dress at a loss."

```
(x2a / 买-01[mai, "buy"]
      :Arg0 (i / individual-person
                  :ref-person 3rd
                  :ref-number Singular)
      :Arg1 (x5 / 衣服 [yifu, "clothes"]
                  :mod (x4 / 这 [zhe, "this"]))
      :cause-of (x2b / 亏-01[kui, "at a loss"]
                  :Arg0 i )
      :aspect Performance)
```

This example means that the person bought clothes at a time when the price of the clothes was high, and he thus suffered a loss in the sense that he could have bought them when the price was lower. In this case, the buying event is straightly speaking not the "cause" of the loss. It is the timing of the buying event that caused the loss. The net consequence is that he suffered the loss. The UMR does not make such fine-grained distinctions, and **:cause-of** is still used to annotate the relation between the two events.

2.1.2 Non-compositional Resultative Compounds

While the theoretical linguistics work focuses on compositional verb compounds, in practical UMR annotation there is a need to consistently distinguish them from non-compositional verb compounds. In non-compositional verb compounds, while both component verbs can function as standalone verbs, the meaning of the verb compound is no longer predictable from their component verbs. An example is provided in (5):

(5) 该 产业 能 带动 经济
This industry can drag-move economy
发展。
development.

"This industry can spur economic development. "

```
(x4 / 带动-01 [daidong,"drag"+"move"
                  = "spur"]
      :Arg0 (x2 / 产业 [chanye, "industry"]
            :mod (x1 / 该 [gai, "this"]))
      :Arg1 (x6 / 发展-01 [fazhan,
            "development"]
            :Arg1 (x5 / 经济 [ jingji,
                  "economy"]))
      :modstr NeutAff)
```

In (5), the meaning of verb compound 带动 [daidong, “spur”] has diverged from the meaning of their component verbs, 带 [dai, “drag”] and 动 [dong, “dong”]. While the meaning of the verb compound 带动 is abstract, the meanings of their component verbs are concrete.

In yet another type of resultative verb compounds the result verb V_2 is grammaticalized and has largely been reduced to some aspectual meaning. Yet they are not fully grammaticalized as Chinese aspectual markers 着 [zhe, PRG], 了 [le, PF], and 过 [guo, EXP]. One sign of their grammaticalization is that they tend to be productive, and can co-occur with a wide range of V_1s. Since V_2 in the verb compound is grammaticalized, its meaning in the verb compound also diverges from its meaning if it is used in isolation. In this sense, it is also non-compositional. For example, in (6), 掉 [diao] originally means "to drop" as a stand-alone verb, but when it forms a verb compound with another verb as V_2, it means finishing up something by means of V_1. 吃掉 [chidiao] in (6) thus means “eat up”. Since 掉 [diao] is grammaticalized and does not alter the predicate-argument structure of the verb compound in any way, we do not annotate the argument structure of this verb. Since it is partially grammaticalized, it contributes to the aspectual value of V_1, which is *Performance* in this case. Since it is not fully grammaticalized, we still use the entire verb compound as the predicate rather than just the first verb, which would be the case if it is fully grammaticalized as the aspect markers.

(6) 小孩吃-掉 了 糖果。
Kid eat-drop PF candy.

“The kid ate up the candy.”

(x2 / 吃掉-01 [chidiao, “eat up”]
:Arg0 (x1 / 小孩 [xiaohai, “kid”])
:Arg1 (42 / 糖果 [tangguo, “candy”])
:aspect Performance)

Another such example is 完 [wan, “finish”], which forms “phase resultative verb compounds” with V_1 (Li and Thompson, 1981; Woo, 2021). It indicates the completion of the event denoted by V_1 and is also partially grammaticalized and does not contribute to the argument structure of the verb compound. In (7), 写完 [xiewan] means “finish writing”, with 完 [wan] contributes to the completion reading of the verb compound, and this is reflected in the aspectual value *Performance* for the event.

(7) 小孩写完 了 作业。
kid write-finish PF homework.

“The kid finished doing his homework.”

(x2 / 写完 [xie, “write”]
:Arg0 (x1 / 小孩 [xiaohai, “kid”])
:Arg1 (s2 / 作业 [zuoye, “homework”])
:aspect Performance)

Verbs like 写完 [xiewan, “finish writing”], 听惯 [tingguan, “get used to listening”] contribute to the aspectual meaning of V_1. Since they are not fully grammaticalized, we use the verb compound as a whole as the UMR concept to avoid loss of meaning. Since V_2 is partially grammaticalized and is productive, creating separate frame file entries for these verb compounds is impractical as there is potentially a long list of such verb compounds. Since such verbs do not contribute to the argument structure of the verb compound, this means the argument structure of the verb compound as a whole is the same as the argument structure of V_1. We could exploit this property and link the argument structure of the verb compounds ending with such verbs to the argument structure of V_1 as aliases, a practice that is similar to how phrasal verbs in English like “eat up” is annotated in the Propbank (Palmer et al., 2005).

2.1.3 Variants of resultative verb compounds

One test for resultative verb compounds that have been recognized very early on is that resultative verb compounds can have an infix between V_1 and V_2 to indicate “potential”. The infix can either be 得 [de, “able”] or 不 [bu, “not able”], and this is illustrated in (8):

(8) a. 柜子 打-得-开
cabinet open-ABL-open

“The cabinet can be opened.”

(x2a / 打开-01 [dakai, “open”]
:Arg1 (x1 / 柜子 [guizi, “cabinet”])
:aspect State
:MODSTR NeutAff)

b. 柜子 打-不-开
cabinet open-NEG-open

“The cabinet cannot be opened.”

(x2a / 打开-01 [dakai, “open”]
:Arg1 (x1 / 柜子 [guizi, “cabinet”])
:aspect State
:MODSTR FullNeg)

In both (8a) and (8b), the resultative verb compound is 打开 [dakai, "open"]. The compound is non-compositional in that 打 has a different meaning when used as a standalone verb than it is in the compound. In (8a), the infix adds a modal meaning to the verb compound so that it means "can be opened", while in (8b), it adds the infix 不 to mean that "cannot be opened". The modal meaning in UMR is annotated as modal strength (:MODSTR) with values FullNeg (fully negative) and NeutAFF (neutral affirmative).

2.1.4 Pseudo-resultative compounds

Some V_1 V_2 constructions look like resultative compounds in form, but upon closer examination they are not. In this section, we discuss a few such examples. In UMR annotation, it is important to separate such cases from resultative verb compounds as the semantic relation between these two verbs is not one of cause and result. One example is (9), where V_1 is an argument of V_2. In this example V_1 研制 [yanzhi, "develop"] is actually an argument of the V_2 成功 [chenggong, "succeed"], as what is successful is the research and development activity denoted by V_1. The fact that this endeavor succeeded implies the completion of the event denoted by V_1, as indicated by the aspecutal value of *Performance*.

(9) 新 药 研制 成功
new medicine develop succeed
"New medicine has been successfully developed. "

(x3 / 研制-01 [yanzhi, "develop"]
:Arg1 (x2 / 药 [yao, "medicine"]
:mod (x1 /新 [xin, "new"]))
:Arg0-of (x4 / 成功-01 [chenggong, "succeed"])
:aspect Performance)

Some verb compounds closely resemble resultative verb compounds but they are in fact object-oriented depictives. 买-贵 [mai-gui, "buy-expensive"] in (10) is such an example. It differs from 买-亏 [mai-kui, "buy-loss"] in (4) by one character, but has a very different interpretation. The semantic relation between V_1 and V_2 is one of temporal co-occurence (as indicated by the *:temporal* role), meaning the clothes were bought at a time when they were expensive, not that the buying event made the clothes more expensive, as would be case if there is a cause-result interpretation.

(10) 这 件 衣服 小王 买-贵 了
This CL clothes Xiaowang buy-costly PF.
"Xiaowang bought this piece of clothes at a high price."

(x2a / 买-01[mai, "buy"]
:Arg0 (x1 / 小王 [Xiaowang, (name)])
:Arg1 (x5 / 衣服 [yifu, "clothes"]
:mod (x4 / 这 [zhe, "this"]))
:temporal (x2b / 贵-01 [gui, "costly"]
:Arg0 x5)
:aspect Performance)

Another example is 挖-浅 [wa-qian] in (11), where V_2 indicates a deviation from the expected result from V_1 (Li, 2007) rather than the result. This is annotated with with the UMR abstract concept *but-91*, which captures the semantic relation between the events denoted by the two verbs.

(11) 这 口 井 小张 挖-浅 了
This CL well Xiaozhang dig-shallow PF.
"Xiaowang dug this well but it was too shallow."

(x5a / 挖-01 [wa, "dig"]
:aspect Performance
:Arg0 (i / individual-person
:name (n / name
:op ‘‘小张" [Xiaozhang]))
:Arg1 (x3 / 井 [jing, "well"]
:mod (x1 / 这)
:unit (x2 / 口 [kou, CL]))
:Arg1-of (b / but-91
:Arg2 (x5b / 浅 [qian,
"shallow"]
:Arg0 x3)))

What we have presented above are just a few examples of apparent resultative verb compounds that have other semantic relations. They are unlikely to be exhaustive and further research is needed to uncover more such examples.

2.2 Subordinating Compounds

Syntactically subordinating compounds in Chinese are compounds where V_1 is a modifier to V_2. An example is provided in (12), where V_1 describes the manner of V_2, represented in UMR as a *:manner* relation. That is, the student bikes to school rather than by any other means.

(12) 这 个 学生 骑行 前往 学校。
this CL student cycle head-to school.

"The student bikes to school."

(x1 / 前往 [qianwang, "head to"]
:Arg0 (x2 / 学生 [xuesheng, "student"])
:Arg1 (x3 / 学校 [xuexiao, "school"])
:manner (x4 / 骑行 [qixing, "cycle"])
:aspect Habitual)

Not all V_1 indicates the manner of V_2, and some subordinating verb compounds are depictives. This is illustrated in , where 活捉 [huo-zhuo, "catch alive"] is an object-oriented depictive that means when V_2 happens, the tiger is in the state of V_1. That is, the tiger was captured while it was alive. This is indicated by the *:temporal* relation between $V1$ and V_2.

(13) 猎人 活-捉 了 这 只 老虎。
hunter alive-catch PF this CL tiger .

"The hunter caught this tiger alive."

(x1 / 捉 [zhuo, "catch"]
:Arg0 (x2 / 猎人 [lieren, "hunter"])
:Arg1 (x3 / 老虎 [laohu, "tigerl"]
:mod (x4 / 这 [zhe, "this"])
:unit (x5 / 只 [zhi, CL]))
:temporal (x6 / 活 [huo, "alive")
:aspect Performance)

Examples like 活 捉 are compositional, but subordinating conjunctions can also be non-compositional. The literal meaning of (14), 三思 is "think three times", but the verb compound actually just means "think carefully". 三思 should thus not decomposed in UMR annotation and treated as a single concept.

(14) 购买前 要 三思 。
buy before should three-think

"(You) need to think carefully before (you) buy (it)"

(x4 / 三思 [sansi, "think carefully"]
:temporal (x2 / 前 [qian, "before"]
:op (x1 / 购买[goumai,
buy])))
:aspect Process)

2.3 Coordinating compounds

Coordinating verb compounds are compounds in which V_1 and V_2 are viewed as equals in their importance, and they also be tend to have similar argument structures. In UMR, the two verbs in the verb compound are typically annotated as arguments to an abstract concept *and* that indicates a discourse relation, and they typically share arguments. This is illustrated in (15). In this sentence, 开发 [kaifa, "develop"] and 建设 [jianshe, "build"] share the same argument 港口 [gangkou, "port"].

(15) 开发 建设 港口.
Develop build port.

"To develop and build the port"

(s1a / and
:op1(x1 / 开发-01 [kaifa, "develop"]
:Arg1 (x3 / 港口 [gangkou,
"port"]))
:op2 (x2 / 建设-01 [jianshe, "build"]
:Arg1 x3))

Compositional coordinating verb compounds like 开发-建设 should be distinguished from non-compositional verb compounds like 褒贬 [baobian, "pass judgment on"], where the meaning of the verb compound as a whole cannot be systematically predicted from the individual verbs, although it is clear they are still related. In this case, the verb compound should be treated as a single concept, as in (16):

(16) 他 喜欢褒贬 人.
he like praise-criticize others.

"He likes to pass judgment on others"

(x1 / 喜欢-01 [xihuan, "like"]
:Arg0 (x3 / individual-person
:ref-person 3rd
:ref-number Singular)
:Arg1 (x2 / 褒贬-01 [baobian,
"pass judgment on"])
:Arg0 x3
:Arg1 (x4/人 [ren, "people"])))

2.4 Verb compounds that have a copula

Chinese copula include 是 [shi, "be"], 为 [wei, "be"] and 成 [cheng, "become"], and they can form a verb compound as V_2 with another verb.

Since when used in a verb compound the sole purpose of 是 and 为 is to introduce another argument to V_1, in UMR annotation, they will not be represented as a separate concept. This is illustrated as (17), where 是 simply introduces *:Arg2* of V_1:

(17) 小王 被 看作-是好 人
xiaowang PAS see-is good person
"Xiaowang is viewed as a good person."

(x3 / 看作-01 [kanzuo, "viewed as"]
:Arg1 (i / individual-person
:name (n / name
:op ‘‘小王" [Xiaowang]))
:Arg2 (x5 / 人 [ren, "person"]
:mod (x4 / 好 [hao, "good"]))
:aspect Performance)

As a V_2 in a verb compound, the copula 成 [cheng, "become"] also introduces an argument to V_1, but it indicates change and has a meaning of its own. For this reason, we treat it as a separate predicate taking its own arguments, but it is also an argument itself to V_1. This is illustrated in (18):

(18) 气旋 增强-成 风暴
Cyclone intensify-become storm.
The cyclone intensifies into a storm.

(x2a / 增强-01 [zengqiang, "intensify"]
:Arg1 (x1 气旋 [qixuan, "storm"])
:Arg2 (x2b / 成 [cheng, "become"]
:Arg0 s1x3
:Arg1 (s1x4 / 风暴 [fengbao,
"storm"]))

In the example above, 增强 [zengqiang, "intensify"] as a verb has two arguments: Arg0 is the agent/cause and Arg1 is the thing strengthened. This verb implies a transition. Thus, we propose a new core argument Arg2 to indicate the end state of intensification. Thus, if we want to keep the copula, 成 [cheng, "become"], it would become part of Arg2.

2.5 Directional verb compounds

Modern Chinese has a closed list of direction verbs (Lu, 1977; Packard, 2000) that can serve V_2 in a verb compound, forming that has been described in literature as a directional verb compound. Here we focus on two main types of such verb compounds, compositional and partially grammaticalized verb compounds.

2.5.1 Compositional Directional Verb Compounds

In compositional directional verb compounds, both V_1 and V_2 have their own argument structures, with V_2 serving as a direction or goal of V_1. This is illustrated in (19):

(19) 老师 走-进 学校。
teacher walk-enter school.
"The teacher walked into the school."

(x2a / 走 [zou, "walk"]
:Arg0 (x1 / 老师 [laoshi, "teacher"])
:goal (x2b / 进 [jin, "enter"]
:Arg0 x4
:Arg1 (x3/ 学校 [xuexiao,
"school"]))
:aspect Performance)

2.5.2 Non-compositional Directional Verb Compounds

The direction verb in directional verbs are frequently partially grammaticalized in the sense that they no longer have their own argument structure, and only indicate a direction for V_1, sometimes literally and other times metaphorically. The examples in (20) show that while the meaning of V_1 递 in the directional verb compound 递-过来 does not change when it is in a compound (20a,b), V_2 过来 cannot be used in isolation (20c).

(20) a. 他 递-过来 一 杯 水
he hand-come one cup water.
"He handed over a glass of water."

b. 他 递 一 杯 水
he hand one cup water.
"He handed over a glass of water."

c. * 一杯 水 过来
one cup water come.
"A glass of water came."

Partially grammaticalized verbs tend to be productive in the sense that the direction verb can co-occur with a wide range of other verbs to form verb compounds. We approach such verb compounds similarly as we do with grammaticalized resultative verb compounds by treating the verb compounds as a single UMR concept, but link the argument structure of the verb compound to that of V_1 so that we do not have create separate frame file entries for such compounds.

(21) 他 递-过来 了 一 杯 水
he hand-come PF one cup water

"He handed over a glass of water."

(x2a / 递-过来-01 [di, "hand over"]
:Arg0 (i / individual-person
:ref-person 3rd
:ref-number Singular)
:Arg1 (x6 / 水 [shui, "water"]
: quant 1
: unit 杯 [bei, "cup"])
:aspect Performance)

2.6 Ambiguity between direction and resultative verb compounds

Some verbs in Chinese are ambiguous between a resultative reading and a directional reading, and it is not always possible to say the verb compound is resultative or directional without a specific context.

For example, the verb 开 as V_2 in (22a) means "away", while in (22b) V_2 means "open". This means that 踢开 [tikai, "kick open / kick away"] in (22) is a directional verb compound while in (22b) it is a resultative verb compound. When it is a resultative verb compound as in (22b), the verb is actually compositional, and are decomposed into two separate concepts in UMR annotation. When it is a directional verb compound, it is non-compositional as there is not an "away" sense when 开 [kai, "open"] is used as a standalone verb. It is also partially grammaticalized in the sense that it can form a verb directional verb compound with a wide range of other verbs. As such we will treat the verb compound as a whole as a UMR concept, but linking its argument structure to that of its V_1, adopting a similar approach to other directional verb compounds. What this example suggests that compositionality is tied to specific senses of a word in a particular context rather than the word as a whole.

(22) a. 小孩踢开 了 皮球
Kid kick-away PF ball

"The Kid kicked the ball away."

(x2 / 踢开-01 [ti, "kick away"]
:Arg0 (x1 小孩 [xiaohai, "kid"])
:Arg1 (x4 / 皮球 [piqiu, "ball"])
:aspect Performance)

b. 小孩踢-开 了 门
Kid kick-open PF ball

"The Kid kicked the door open."

(x2a / 踢-01 [ti, "kick"]
:Arg0 (x1 小孩 [xiaohai, "kid"])
:Arg1 (x4 / 门 [men, "door"])
:Arg0-of (x2b / 开[kai,
"open"])
:aspect Performance)

3 Related verb constructions

Light verb constructions are also worth discussion here. When determining the type of semantic relations that hold between V_1 and V_2 in a verb compound in UMR annotation, it is important to first determine whether it is a verb compound in the first place. One construction that is similar to verb compounds in appearance is the light verb construction, which also has a verb followed by a deverbal noun that is identical in form to verbs as Chinese verbs can function as a noun without having to have a derivational suffix. This is illustrated in (23), where 进行 [jinxing "hold"] 讨论 [taolun, "discussion"] is a light verb construction in which

(23) 小王 用 图表 进行-讨论
xiaowang use graph operate-discuss

"Xiaowang use graph to discuss. "

(x4a / 讨论-01[taolun, "discuss"]
:instrument (x2 / 图表[tubiao,
"graph"])
:Arg0 (i / individual-person
:name (n / name
:op ''小王" [Xiaowang]))
:aspect Performance
:MODSTR FullAff)

进行 [jinxing, "process"] is a light verb in this sentence, thus it is not annotated in the graph. 讨论 [taolun, "discuss"] is treated as the predicate.

4 Related work

Theoretical discussion on V-V compounds . Chinese V-V compounds, particularly resultative verb compounds have received a lot of discussion in theoretical linguistics literature. Most of the discussion centers on the issue of whether such compounds are formed in lexicon or in syntax. (Li,

1990; Gu, 1992; Thompson, 1973; Li, 2007) generally hold that V-V compounds are produced in lexicon and it is theta identification that restricts the possible V-V constructions. However, (Lu, 1977; James Huang, 1992) hold the view that V-V compounds are generated via syntactic operations. Some other works (Cheng, 1997) charted a course in the middle, arguing that verb compounds are generated in both lexicon and syntax. There are also discussions (Paul, 2022; Lu, 1973) specifically on directional verb compounds. Most of the discussions have implicitly assumed these compounds are compositional without providing a set of criteria for how to distinguish compositional from non-compositional verb compounds. For UMR annotation, however, out of necessity we first to determine whether they are compositional or not as we have to decide what concepts to propose. The question of whether they are generated in syntax or the lexicon is of secondary importance. Here we provide a classification of different types of verb compounds in Chinese and show how compositional verb compounds can be distinguished from non-compositional verb compounds and how each type of verb compounds can be annotated in UMR.

Semantic role annotation in the Chinese Propbank, Chinese AMR, and UMR The practice of defining predicate-specific semantic roles started with the Proposition Bank (Palmer et al., 2005) and this practice has been adopted in the construction of the Chinese Propbank (Xue and Palmer, 2009). Before the argument structure of a predicate can be annotated, a frame file that defines the semantic roles for each sense of that predicate has to be created. For a language like English in which verb compounds are uncommon, the list of verbal and nominal predicates is relatively small[3]. However, for a language like Chinese where verb compounding is a common process, as we have discussed, the number of frame files can be quite large[4] if we consider each verb compound as a new predicate that needs a frame file. This issue is inherited by the Chinese AMR Project (Li et al., 2019, 2016) and the UMR project (Gysel et al., 2021) as they adopt the same approach to predicate-argument structure annotation. We propose a solution in which the argument structure of the component verbs are annotated together with the relation between them if the verb compound is compositional. This way we do not to create a new frame files every time we see a new verb compound as long as the frame files for the individual verbs are already available.

[3]See a list of frame files here: https://verbs.colorado.edu/verb-index/

[4]See a list of Chinese frame files here: https://chinese-propbank.herokuapp.com

5 Conclusion

In this paper we describe the challenge to annotate Chinese verb compounds in the Uniform Meaning Representation framework as compounding is a productive process in Chinese. We propose a solution that is based on treating different types of verb compounds differently based on compositionality, levels of grammaticalization, and productivity of these verb compounds. For compounds that are non-compositional, we annotate the argument structure of the verb compound as a whole, but for compositional verb compounds, we annotate the argument structure of their component verbs, obviating the need to create an additional frame file entry for the compound verb as a whole. For verb compounds that have highly grammaticalized verb components, we also annotate the argument structure of the verb compound as a whole, but link its argument structure to that of the primary verb in the verb compound so that there is no need to create a completely new frame file.

Acknowledgement

This work is supported in part by a grant from the CNS Division of National Science Foundation (Award No: 2213804) entitled "Building a Broad Infrastructure for Uniform Meaning Representations". All views expressed in this paper are those of the authors and do not necessarily represent the view of the National Science Foundation. We would like to thank William Croft, James Pustejovsky, Martha Palmer, James Andrew Cowell, and other collaborators of the UMR project for useful discussions. We would also like to thank annotators of the Chinese UMR team, particularly Ching-wen Chen and Alicia Tu for raising issues related to the UMR annotation of Chinese verb compounds. All shortcomings of the paper are those of the authors.

References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina

Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.

LL Cheng. 1997. Resultative compounds and lexical relational structures. *Chinese Languages and Linguistics III: Morphology and Lexicon*, pages 167–197.

Yang Gu. 1992. *The syntax of resultative and causative compounds in Chinese*. Cornell University.

Jens Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Timothy J. O'Gorman, Andrew Cowell, W. Bruce Croft, Chu-Ren Huang, Jan Hajic, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. Designing a uniform meaning representation for natural language processing. *Künstliche Intell.*, 35:343–360.

C-T James Huang. 1992. Complex predicates in control. *Control and grammar*, pages 109–147.

Bin Li, Yuan Wen, Weiguang Qu, Lijun Bu, and Nianwen Xue. 2016. Annotating the little prince with chinese amrs. In *Proceedings of the 10th Linguistic Annotation Workshop held in Conjunction with ACL 2016 (LAW-X 2016)*, pages 7–15.

Bin Li, Yuan Wen, Li Song, Weiguang Qu, and Nianwen Xue. 2019. Building a chinese amr bank with concept and relation alignments. *Linguistic Issues in Language Technology*, 18.

Chao Li. 2007. *Mandarin resultative verb compounds: Where syntax, semantics, and pragmatics meet*. Yale University.

Charles N Li and Sandra A Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Univ of California Press.

Yafei Li. 1990. On vv compounds in chinese. *Natural language & linguistic theory*, 8(2):177–207.

John HT Lu. 1973. The verb—verb construction with a directional complement in mandarin. *Journal of Chinese Linguistics*, pages 239–255.

John HT Lu. 1977. Resultative verb compounds vs. directional verb compounds in mandarin. *Journal of Chinese linguistics*, pages 276–313.

Jerome L Packard. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Waltraud Paul. 2022. Svcs in disguise. *New Explorations in Chinese Theoretical Syntax: Studies in honor of Yen-Hui Audrey Li*, 272:133.

Sandra Annear Thompson. 1973. Resultative verb compounds in mandarin chinese: A case for lexical rules. *Language*, pages 361–379.

Meagan Vigus, Jens EL Van Gysel, and William Croft. 2019. A dependency structure annotation for modality. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 182–198.

I-hao Woo. 2021. Pedagogical and theoretical issues around the resultative verb compound construction in mandarin chinese. *International Journal of Chinese Language Teaching*, 2(1):17–35.

Naiwen Xue, Fei Xia, Fu-Dong Chiou, and Marta Palmer. 2005. The penn chinese treebank: Phrase structure annotation of a large corpus. *Natural language engineering*, 11(2):207–238.

Nianwen Xue and Martha Palmer. 2009. Adding semantic roles to the chinese treebank. *Natural Language Engineering*, 15(1):143–172.

Jiarui Yao, Haoling Qiu, Bonan Min, and Nianwen Xue. 2020. Annotating temporal dependency graphs via crowdsourcing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Yuchen Zhang and Nianwen Xue. 2018. Structured interpretation of temporal relations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).