

# Abstract Meaning Representation for MWE: A study of the mapping of aspectuality based on Mandarin light verb *jiayi*

**Lu Lu**

Hong Kong Polytechnic University  
lu-cbs.lu@polyu.edu.hk

**Nianwen Xue**

Brandeis University  
xuen@brandeis.edu

**Chu-Ren Huang**

Hong Kong Polytechnic University  
churen.huang@polyu.edu.hk

## Abstract

Multiword expression (MWE) refers to various types of linguistic units that are made up of more than one word. Light verb constructions (LVCs), as one of the least explored areas in MWEs, have idiosyncratic features that are difficult to capture in computational linguistics. In this paper, we addressed the aspectual differences between LVCs and their corresponding regular verb constructions with corpus data. Specifically, the *jiayi*-LVC in Mandarin Chinese was investigated as a case study, where idiosyncratic aspectual information in the LVC was proposed. This feature was not yet previously represented in abstract meaning representation (AMR), in which LVCs and its regular verb counterparts shared the same AMR. Given the semantic difference in *jiayi*-LVCs, we expand AMR by introducing aspect as the root node, while maintaining the core predicative component in meaning representation.

## 1 Introduction

Light verb constructions (LVCs) are a crucial type of multiword expression (MWE) that can be found across different languages (Butt 2010). One of the prominent semantic features in LVCs (e.g. *have a*

*bath* in English and *jinxing yanjiu* ‘carry out a study’ in Mandarin Chinese) is that the light verb (LV; i.e. *have* and *jinxing*) bears little semantic content and the actions are largely described by the eventive nominals (i.e. *bath* and *yanjiu* ‘research’), encoding the same semantic information as their regular verb counterparts. Since the meaning of LVCs differ from usual predicative structures or the direct aggregation of its semantic components, LVCs, as one of the least explored areas of MWEs in computational linguistics, pose a number of challenges in computational grammar, such as automatic word alignment, annotation and semantic representation. In the model of Abstract Meaning Representation (AMR), it is often assumed that the LVCs and its corresponding regular verb construction (RVC) share the same representation (Banarescu et al. 2012; Flanigan et al. 2014, Bu et al. 2016). AMR is a semantic framework addressing the predicate-argument relation of the whole sentence (to be more fully described in Section 4).

However, corpus data suggest that LVCs and RVCs have slightly different semantic meaning. In Urdu, for example, LVs play a central role in the meaning and morphosyntactic choices of the whole construction in (1). Although the LVs *par* ‘fall’ and *daal* ‘put’ both occur with *ciik<sup>h</sup>* ‘scream’ in Urdu, the LV in (1a), which involves an involuntary action, is preceded by an unmarked nominative subject, whereas the LV *daal* ‘put’ in (1b), denoting a conscious control over the action, requires the marked ergative case on the subject

argument. Based on this observation, Butt (1995) argues that it is the particular LV *daal* ‘put’ that contributes to the meaning of conscious choices of a given action, which further assigns a marked ergative case to the subject in (3b). In this analysis, the so-called ‘LV’ is not completely ‘bleached’, and has important role to play in AMR.

- (1) a. vo                          ciikh<sup>h</sup>      paŋ-aa  
 pron=NOM                  scream      fall-PRF.M.SG  
 ‘He began screaming suddenly (despite himself).’
- b. us=ne                        ciikh<sup>h</sup>      daal-aa  
 pron=ERG                  scream      put-PRF.M.SG  
 ‘He screamed violently (on purpose).’  
 (Butt 1995: 110)

Xu et al. (2020) focussed on another set of near synonyms like properties of LVCs in Mandarin Chinese. They first show that a subset of LVs is interchangeable in some contexts, which support the hypothesis of their lack of contribution to the meaning of the sentences. However, it is also shown that in other contexts, only a certain LVs can be used. For example, the LV *jiayi* tends to take accomplishment events in Mainland Mandarin Chinese, whereas other LVs do not show the same inclination. This observation underlines the fact that even though LVCs share largely similar grammatical patterns with their corresponding RVCs or other LVCs, each of them still contains some lexical meaning that can interact with the context to generate new meaning. Xu et al. (2020) used these distributional grammatical features to classify and predict different LVs. Our current study tackles the challenge of how to systematically represent such meaning differences, especially within the Universal Meaning Representation (UMR) initiative (Xue et al. 2019). Given the variations in meaning that an LV contributes to, a simple AMR graph may not be adequate. In this regard, a deeper understanding of the semantic content in LVCs contributes to our understanding of the meaning representation in natural language processing, revealing parameters of variations and commonalities between different structures and modelling systems. We also expect that the results can shed light on how to represent MWEs in AMR/UMR.

This study will use the corpus data of *jiayi*, with the abstract literal meaning of *give* in Mandarin Chinese, as a case study. It investigates the semantic features that are idiosyncratic to *jiayi* with data from Chinese Gigaword and presents the preliminary thoughts on how AMR can be refined to represent the semantic differences between LVCs and RVCs.

## 2 Corpus observation

In this study, the most recent edition of the newswire corpus—Chinese Gigaword (fifth edition)<sup>1</sup>—was consulted. To limit the influence of the varieties of Chinese language on the research results, only the corpora of Mainland Mandarin Chinese were employed at this initial stage of research. That is to say, texts from the corpora of Guangming Daily, People’s Daily, People’s Liberation Army Daily and Xinhua News Agency were selected for this paper. Queries of the LV *jiayi* were entered into AntConc and 55 instances of qualified *jiayi* were returned.

A closer look at the concordances of *jiayi*-LVCs show its distinctive aspectual feature. In the Chinese Gigaword corpus, *jiayi*-LVCs cannot take any aspect markers, such as *-le* and *-guo*, immediately after the LV, see (2). However, other semantically similar LVs, such as *geiyu*, can go with an aspect marker, see (3). Such properties have also been observed in Hu and Fan (1995), Diao (2004) and Kuo (2011).

- (2) You                          ren                  suiyi                  jiang  
 have                          people randomly      with
- zhexie                          lishi                  geming                  gequ  
 these                          history revolution      song
- cuangai                          jiayi                  xiedu.  
 falsify                          LV                  blaspheme  
 ‘Someone randomly falsified these revolutionary songs and blasphemed against them.’  
 (Chinese Gigaword)

- (3) Yingguo                  Bomingham                  Daxue

<sup>1</sup> The corpus can be downloaded from the Linguistic Data Consortium catalogue.

UK	Birmingham	University		pingjia.
zhuanmen	wei	zhekuan	che	evaluation
specifically	for	this	car	‘The University of Birmingham carried out an examination specifically for the car and spoke highly of it.’
jinxing-le	jiance,	jiyu-le	jigaode	(Chinese Gigaword)
LV-ASP	examine	LV-ASP	excellent	

Table 1. Tokens of different syntactic structures in the Chinese Gigaword corpus

	a. RV+OBJ1+O BJ2	b.OBL+RV+ OBJ2	c. RV+OBJ2	d. LV+OBJ1+A N	e. OBL+LV+AN	f. LV+AN	Total
<i>jiayi</i>	0	0	0	0	9	46	55
<i>geiyu</i>	29	19	148	307	89	371	963

[Examples:

- a. Ta **geiyu**-le [wo]<sub>OBJ1</sub> [rensheng zhong gengwei baogui-de jingyan]<sub>OBJ2</sub>  
 he give-PRF me life in more precious experience  
 ‘He gave me more precious life experience.’
- b. Guojia [dui huojiang qiye]<sub>OBL</sub> **geiyu** [300wan jiangli]<sub>OBJ2</sub>.  
 nation to award-winning company give 3.million reward  
 ‘The country rewarded 3 million yuan to award-winning companies.’
- c. Guojia ji yao xiang minzu diqu touru gengduode zijin,  
 nation not.only need to ethnic area invest more money  
 you yao **geiyu** [geng youhuide zhengce]<sub>OBJ2</sub>  
 but.also need give more preferential policy  
 ‘The country not only needs to make more investment in ethnic areas, but also needs to give more preferential policies (to them).’
- d. Ta zai shenghuoshang **geiyu** [ta]<sub>OBJ1</sub> [wuweibuzhide zhaogu]<sub>AN</sub>.  
 he on life LV her meticulous care  
 ‘He cared for her meticulously in life.’
- e. Ta [dui zhe yi zuofa]<sub>OBL</sub> **geiyu** [chongfen kending]<sub>AN</sub>.  
 he to this one practice LV adequate confirmation  
 ‘He highly confirmed this practice.’
- f. Ri mei ye **geiyu**-le [zugou guanzhu]<sub>AN</sub>.  
 Japanese media also LV-PRF enough attention  
 ‘Japanese media also paid enough attention (to this).’ ]

In what follows, we will explain and represent the aspectual properties of the *jiayi*-LVC observed in and generalised from corpus data: none of the aspect markers can be found in *jiayi*-LVCs.

### 3 Aspectuality in *jiayi*-LVCs

As noted earlier, an LV is assumed to be form identical to its corresponding regular verb (RV) in a language. In Table 1, we summarised the

syntactic structures of *jiayi*-LVCs evidenced in the Chinese Gigaword corpus, and further, for a better understanding, contrasted it with another LV *geiyu* ‘give’ (bearing the same abstract meaning as *jiayi*), which can be used as an RV in a sentence (see the examples from a-c right after Table 1).

In Table 1, structures (a), (b) and (c), occurring in RVCs, can only be found in the contemporary *geiyu*, whereas the three structures were readily available in Classical Chinese for

both verbs. We can also see that a mixture of RVCs and LVCs is widely manifested in the Contemporary *geiyu*-LVCs, but such mixture is by no means occurs in *jiayi*-LVCs. As LVCs arguably enter the grammaticalisation cline developing from RVs to grammatical morphemes, it implies that *jiayi* is much closer to the grammatical end of the cline, compared to *geiyu*.

Additionally, from the perspective of semantics, the transition from (a) to (f) implies the generalisation in meaning, whereby new context entails more general meaning. This is, as observed in Heine and Kuteva (2007), one of the important factors responsible for grammaticalisation. However, as for *jiayi*, it has lost lexical content to a great extent that their contemporary syntactic context mostly favours the last two structures listed in Table 1. This supports the above claims that *jiayi* is at a later stage of grammaticalisation, compared to *geiyu*. In other words, *jiayi* is more grammaticalised than *geiyu*.

The percentage in Figure 1, based on the frequencies of RVCs and LVCs in Table 1 lends further weight to the claim that *geiyu* is at the earlier stage of grammaticalisation compared with *jiayi*.

Figure 1. The percentage concerning the distribution of LVCs and RVCs

	RVC	LVC
Lexeme	Grammatical morpheme	
	—————→	
<i>geiyu</i>	20.4%	79.6%
<i>jiayi</i>	-	100%

The earlier/latter stages of grammaticalisation are, interestingly, compatible with the realisation/non-realisation of the perfective aspect markers. It is assumed that the perfective aspectual information is co-provided by the verb and the construction (Michaelis 2004 and Goldberg and Jackendoff 2004) in the *jiayi*-LVCs. Given the usual grammaticalisation path is ‘independent lexical verb>grammatical morpheme (i.e. aspect marker in this case)’, the more grammaticalised an LV is, the more enriched the aspectual meaning will be (grammaticalised from an RV), and less likely it will resort to the verb to provide aspect information. As for *jiayi*, the verb is bounded and telic in its lexical meaning. This is especially evident in its lexical use, where *jia*, the obsolete

form of *jiayi* in Classical Chinese, was used. For example, in *jia bing wo* add military-force I ‘add more military force to me’, the event encoded by the verb *jia* has a specific end point, thus, when interacting with the context, implying the event can be viewed as a completed whole unit.

Given the above illustration, since *jiayi* is close to the end of grammaticalisation (in contrast to *geiyu*, as shown in Figure 1), we believe that the construction of *jiayi* is sufficient to embody the perfective aspect on its own right. Therefore, the fixed aspectual value internally conveyed in the *jiayi*-LVC makes it incompatible with any perfective aspect markers. This results in the non-realisation of the perfective aspect marker in the *geiyu*-LVCs.

#### 4 Mapping aspectuality into AMR: A preliminary thought

As generalised from the above linguistic observations, while the core predicative-argument relation remains the same in *jiayi*-LVCs and their corresponding RVCs, aspectual information is additionally encoded in LVCs. Therefore, other than the predicate relation and its arguments, which are the core semantic relation in the current AMR, we propose to refine the AMR modelling to represent aspectual information that set a particular LVC apart from its possible syntactic alternatives.

In AMR annotation, canonical meaning of a sentence is represented as a single-rooted, directed, acyclic graph with nodes labelled with concepts and edges labelled with relations. In AMR, the predicate argument structure is the core component. The predicate and its arguments are represented as nodes and the edges represent the relation between the predicate and each of its arguments in the AMR graph. As an illustration, the AMR notation and graph representation of sentence (4) can be found in (5).

(4) wo xihuan dangao.

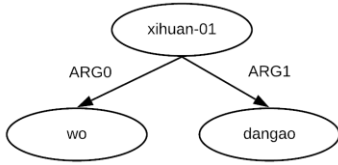
I like cake  
‘I like cakes.’

(5) a. (x0/xihuan-01

:ARG0 (x1/wo

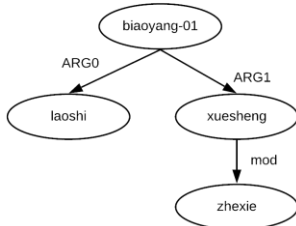
:ARG1 (x2/dangao))

b.



As can be seen, the advantage of representing predicative core elements lies in the central positioning of the predicate argument structure. However, AMR does not represent aspectual information (Banarescu et al. 2012). While Bu et al. (2016) add aspect as a non-core semantic relation particularly designed for Chinese AMR, it does not specify how aspect is embedded into Chinese LVCs. As regards to LVCs, Bonial and Palmer (2016) further argue the approach that LVCs and its corresponding RVCs share the same AMR may be adequate for English LVCs, but it needs to be evaluated for other languages, as cross-linguistically there is some semantic space that cannot be covered by their corresponding RVCs. Given the corpus observation and aspectual justification in Sections 2 and 3, we argue that LVCs and its corresponding RVCs encode different aspectual information in Chinese, and argue that the current AMR does not yet properly handle aspectual encodings with different syntactic realisations: for example, the AMR-graph in (6) can represent both LVC and the corresponding RVC in (7).

(6)



(7) a. LVC

Laoshi dui zhexie xuesheng  
 teacher to these student

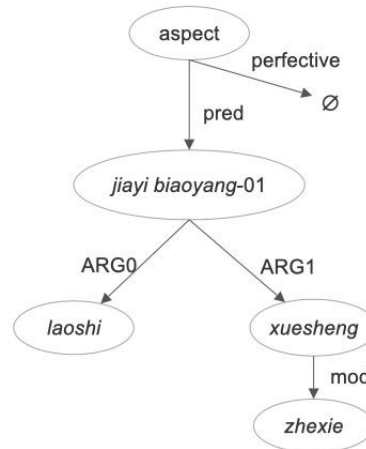
jiayi biaoyang  
 LV praise  
 ‘The teacher made a praise to these students.’

b. RVC counterpart

Laoshi biaoyang-le zhexie xuesheng.  
 teacher praise-ASP these student  
 ‘The teacher praised these students.’

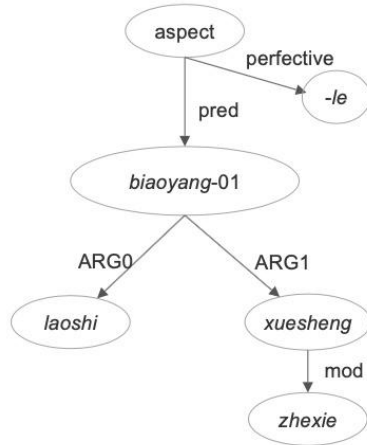
In our research, instead of treating aspect as a modifier, we propose to consider it as a root assigning aspectual value and acting over the predicative node. Consider the AMR graph as shown in (8). The aspect acts as the root node taking aspectual value over the predicate. In example (8), aspect is represented as a property over the predicate (i.e. *jiayi* and the nominal complement), demonstrating the aspectuality in *jiayi*-LVC (i.e. the predicate). In this case, the perfective aspectual value is left empty in AMR, as it is inherently contained in the *jiayi*-LVC. This parallels with the justification of the aspectual features in *jiayi*-LVCs in Section 3. Additionally, from the perspective of lexical semantics, the literal sense of *jiayi*, which is ‘add’, encodes the telic property on its own, and thus can be viewed as a simple whole. Therefore, the perfective aspect of the *jiayi*-LVC in (7a) is not realised in the node of aspect.

(8) AMR graph of example (7a)



Compared to its corresponding RVC (where the nominal complement in LVC is used as the main predicate), the aspect value, represented in the root node, is shared between the verbal predicate and the aspect marker in RVC. As represented in (9), the aspectual value pertaining to the RV *biaoyang* is specified and realised in the aspect node.

(9) AMR graph of example (7b)



We believe the advantages of the approach can be seen in three folds. It differentiates two syntactically and semantically similar structures, while maintaining both the predicative core elements of the sentence. Further, since the aspect node has the same representation in *jiayi*-LVCs and RVCs (except that the grammatical aspect is realised or not), it lowers the cognitive load for annotation and processing, especially for those who are not familiar with the grammatical system of the language. Lastly, this approach has the potential to generalise into other LVs, leading to a universal representation of aspects and its interaction with the predicate.

## 5 Conclusion

In this study, generalising from corpus observation, we argue that perfective aspectual meaning is internally encoded in the *jiayi*-LVCs in Mandarin Chinese, thus highlighting the semantic differences between LVCs and corresponding RVCs. Given this, we refined the AMR to capture the aspectual encoding and its interaction with the predicate and the aspect marker. We proposed a root node aspectual feature in Chinese AMR, while maintaining the predicative core element in the original AMR graph. This preliminary work, drawing on the two roughly equivalent

constructions of LVCs and their RVC counterparts, enriches the representation of AMRs with the feature of aspect. In the next, we will expand the representation with more corpus data and experiment with small-scale annotation and testing to work on its feasibility regarding universal representation.

## References

- Banarescu, L. Bonial, C., Cai, S. Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K. Koehn, P., Palmer, M. and Schneider, N. 2013. ‘Abstract meaning representation for sembanking’. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 178–186.
- Bonial, C., and Palmer, M. (2016). ‘Comprehensive and consistent PropBank light verb annotation’. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 3980–3985.
- Bu, L., Y. Wen, B. Li, N. Xue. 2016, ‘Zhongwen chouxiang yuyi biaoshi biao zhu guifan (V1.2)’ [The annotation guidelines of Chinese Abstract Meaning Representation]. [https://www.cs.brandeis.edu/~clp/camr/res/CAMR\\_GL\\_v1.2.pdf](https://www.cs.brandeis.edu/~clp/camr/res/CAMR_GL_v1.2.pdf) (accessed on July 10, 2020)
- Butt, M. 1995. *The structure of complex predicates in Urdu*. Stanford: CSLI.
- Butt, M. (2010). ‘The light verb jungle: Still hacking away’. *Complex predicates: Cross-linguistic perspectives on event structure*. 48–78. Cambridge: Cambridge University Press.
- Diao, Y. 2004. *Xiandai hanyu yuyi dongci yanjiu* [The research on weak verbs in contemporary Chinese]. Dalian: Liaoning Normal University Press.
- Flanigan, J., Thomson, S., Carbonell, J. G., Dyer, C., and Smith, N. A. 2014. ‘A discriminative graph-based parser for the abstract meaning representation’. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Volume 1*:1426–1436).

- Goldberg, A.E. and Jackendoff, R., 2004. 'The English resultative as a family of constructions'. *Language* 80 (3): 532–568.
- Heine, B., and Kuteva, T. 2007. *The genesis of grammar: A reconstruction*. Oxford: Oxford University Press.
- Hu, Y. and Fan, X. 1995. *Dongci yanjiu* [The research on verbs]. Kaifeng: Henan University Press.
- Kuo, P. J. A. 2011. 'Case study of obligatory object fronting in Mandarin Chinese'. In *Online proceedings of glow in Asia workshop for young scholars*. Accessed [http://faculty.human.mie-u.ac.jp/~glow\\_mie/Workshop\\_Proceedings/11Kuo.pdf](http://faculty.human.mie-u.ac.jp/~glow_mie/Workshop_Proceedings/11Kuo.pdf) (Accessed on 15 Aug, 2016)
- Michaelis, L. A. 2004. 'Type shifting in construction grammar: An integrated approach to aspectual coercion'. *Cognitive Linguistics* 15 (1): 1–68.
- Xu, H., Meng, J., Lin, J. and Huang, C.-R. 2020. 'Light verb variations and varieties of Mandarin Chinese: Comparable corpus driven approaches to grammatical variations'. *Corpus Linguistics and Linguistic Theories*. [ahead of print]
- Xue, N., Croft, W., Hajic, J., Huang, C.-R., Oepen, S., Palmer, M., and Pustejovsky, J. 2019. Proceedings of the First International Workshop on Designing Meaning Representations. In *Proceedings of the First International Workshop on Designing Meaning Representations*.