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 ciikh ‘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 *ḍaal* ‘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.

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)—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).

1. 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.’

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.’

3. Yingguo Bominghan Daxue

---

1. The corpus can be downloaded from the Linguistic Data Consortium catalogue.
The University of Birmingham carried out an examination specifically for the car and spoke highly of it. (Chinese Gigaword)

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.

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))
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 aspe ctual 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 aspe ctual justification in Sections 2 and 3, we argue that LVCs and its corresponding RVCs encode different aspe ctual information in Chinese, and argue that the current AMR does not yet properly handle aspe ctual 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 aspe ctual value and acting over the predicative node. Consider the AMR graph as shown in (8). The aspect acts as the root node taking aspe ctual 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 aspe ctuality in jiayi-LVC (i.e. the predicate). In this case, the perfective aspe ctual value is left empty in AMR, as it is inherently contained in the jiayi-LVC. This parallels with the justification of the aspe ctual features in jiayi-LVC 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 aspe ctual value pertaining to the RV biaoyang is specified and realised in the aspect node.
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 biaozhu guifan (V1.2)’ [The annotation guidelines of Chinese Abstract Meaning Representation]. 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 perpectives 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, Č., 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 (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*. 