

Annotate Chinese Aspect with UMR—A Case Study on *The Little Prince*

Sijia Ge, Zilong Li, Alvin Po Chun Chen, Guanchao Wang

University of Colorado Boulder

{sijia.ge, zilong.li, alvin.chen, mike.wang}@colorado.edu

Abstract

Aspect is a valuable tool for determining the perspective from which an event is observed, allowing for viewing both at the situation and viewpoint level. Uniform Meaning Representation (UMR) seeks to provide a standard, typologically-informed representation of aspects across languages. It employs an aspectual lattice to adapt to different languages and design values that encompass both viewpoint aspect and situation aspects. During annotation, this can lead to ambiguous situations, especially considering that Chinese contains a large number of particles representing grammatical aspect. In the context of annotating the Chinese version of *The Little Prince*, we paid particular attention to the interactions between UMR aspect values and aspect markers, and we also want to know the annotation effectiveness and challenges under the UMR aspectual lattice. The factors contributing to disagreement among annotators including the lexical semantics, implications, and the influence of aspectual markers. We proposed an alternative split-way to annotate situation type and viewpoint and compare the agreement with the UMR annotations. Our work indicates UMR aspect guideline has a stronger emphasis on the situation type and sheds light on the challenges of aspect annotation in Chinese.

Keywords: Aspect, Uniform Meaning Representation, Chinese Language

1. Introduction

Uniform Meaning Representation (UMR) is a graph-based cross-linguistically applicable semantic representation. It encodes the meaning of natural language sentences and documents in a structured, human-machine-readable format. UMR is designed to facilitate interpretable natural language applications requiring deep semantic analysis (Gysel et al., 2021; Bonn et al., 2023).

UMR extends Abstract Meaning Representation (AMR) (Banarescu et al., 2013), improving it in two ways. Firstly, UMR extends to cover more linguistic categories. It adds categories such as aspect at the sentence level (Donatelli et al., 2018; Van Gysel et al., 2019) and coreference (O’Gorman et al., 2018) in document-level annotation. Secondly, UMR has been adapted for cross-lingual applicability, even for low-resource ones like Sacoya, Kukama, Sanapana, etc (Vigus et al., 2020; Van Gysel et al., 2021; Bonn et al., 2023).

This paper specifically focuses on aspect annotation in Mandarin Chinese. Aspectual meaning pertains to how the internal temporal structure of situations is presented (Friedrich et al., 2023). The concept of aspect can be approached from various angles, with one popular view being Smith (1991b)’s two-component theory. This theory identifies the **lexical aspect** (also known as the **aktionsart** or **situation types**) and **grammatical aspect** (also known as the **viewpoint**). The lexical aspect refers to the information inherently packaged in the lexicon. Vendler (1967) uses punctuality, telicity, and stativity to classify the lexical as-

pect into *State*, *Activity*, *Achievement*, *Accomplishment*, and *Semelfactive*. The details is shown as Table A1. Grammatical aspect, realized through the grammatical approach, pertains to how an event is perceived, either as a bounded whole (*Perfective*) or from within (*Imperfective*) (Velupillai, 2012; Friedrich et al., 2023; Chen, 2008).

In UMR, the guideline for annotating aspect is primarily inspired by the work of Croft (2012) and Donatelli et al. (2018, 2019). It relies on a lattice that spans a range from coarse-grained to fine-grained aspectual values (Van Gysel et al., 2019). The lattice is depicted in Figure 1. This enables UMR graphs to be annotated at the level of granularity most suitable for the language. The values designated in the lattice aim to capture the semantics of events from both qualitative and temporal perspectives (Croft, 2012; Chen et al., 2021). Temporal considerations encompass the boundedness of time scales, akin to the grammatical aspect, while qualitative considerations focus on the internal structure of events, resembling the lexical aspect.

Chinese is a language with a diverse range of aspect markers (particles) that are used to express the viewpoint (Li and Thompson, 1981). For instance, “过 guo” and “了 le” correspond to the *Perfective* aspect while “正在 zhengzai” and “着 zhe” correspond to the *Imperfective* aspect, particularly, *Progressive* aspect. The Chinese AMR corpus for *The Little Prince* has been annotated with aspect markers (Li et al., 2016, 2019). Bonn et al. (2023) highlighted that the aspect markers can be somewhat helpful in determining the UMR

aspect values. Yet there is a potential misalignment between the grammatical aspect that aspect markers correspond to and the aspectual lattice system in UMR. The latter's values encompass both viewpoint and situation aspects and have a stronger emphasis on the situation types.

In this paper, we investigate the above potential misalignment through practically annotating *The Little Prince* with UMR aspectual values. Our work seeks to address two main questions: 1) Are aspect markers genuinely beneficial for UMR aspectual annotation, particularly when focusing on different aspects? (§5) 2) How much would the agreement and effectiveness of annotation be with the UMR system, as opposed to the traditional separate discussions from the viewpoint and situation angles? Is it possible to categorize and analyze the complex cases, that may cause the disagreement of annotations, to offer insights for the development of refined annotation guidelines? (§6)

Our work revealed that there is not a one-to-one correspondence between aspect markers and aspect values. There is a moderate inter-annotation agreement for Chinese aspect annotation which arises from factors such as the interpretation of lexical semantics, implications, etc. We also provided the annotation result when proposing the annotation two aspects separately.

2. Related Work

2.1. Study on Aspect System of Chinese

Pioneering works on English by Vendler (1967), Dahl (1985), and Smith (1991a) significantly influenced indigenous studies on the Chinese aspect (Chen, 2008). The early investigation primarily revolved around classifying and understanding the grammaticalization process of particles that denote aspect, distinguishing between the external viewpoint to an event and the internal viewpoint, as influenced by Comrie (1976).

Subsequently, the Chinese aspect system evolved, particularly following Smith (1991a)'s introduction of a "two-component theory" that divided the aspect into situation types and viewpoint, building upon the ideas of Vendler (1967) and Comrie (1976). This framework separated the aspect system into two levels: the lexical level, focusing on categorizing the predicate based on the telic/atelic, and dynamic/punctual distinction, and the sentence level, concerned with the viewpoint aspect. The latter in Chinese involves aspect markers (particles) such as “在 zai” (expressing Progressive aspect), “了 le” (indicating Perfective, specifically a completion of a situation), “过 guo” (emphasizing the experience of having been through a situation), among others (Friedrich et al., 2023). Li and

Thompson (1981) classified Chinese aspects into 4 categories. Perfective (aspect marker is “了 le”), Durative aspect (same as Progressive, aspect markers such as “在 zai”, “着 zhe”), Experiential aspect (aspect marker such as “过 guo”) and Delimitative aspect, which is realized by repeating the verb, indicating the actualization of states and activities to some extent. Xiao and McEnergy (2004) proposed there are 4 Perfective aspects (Actual aspect, Experiential aspect, Delimitative aspect, and Completive aspect) and 4 Imperfective aspects (Durative aspect, Progressive aspect, Inchoative aspect, and Successive aspect), the latter two are classified into Phasal aspect by other researchers such as Binnick (1991)), expressing the different phases of a situation. Building upon the two-component theory, Chen (2008) proposed examining situations from different phases and further subdividing viewpoint based on the degree of grammaticalization, formulating a four-dimension frame.

2.2. Aspect Annotation Schemas and Data

In the realm of English aspect annotation, Mathew (2009) labeled habitual or episodic events for 1052 sentences from Penn Treebank (Marcus et al., 1993). Friedrich et al. (2016) categorized situation Entity types at the clause level including State, Event, Report, Generic Sentence, Generalizing Sentence, Question, and Imperative.

Donatelli et al. (2018, 2019) attached aspect annotation to AMR, covering multiple dimensions like :stative +/-, :ongoing +/-/?, :complete +/-, :habitual +/- and :completable +/- . This scheme serves as the cornerstone for the UMR aspects scheme.

Alikhani et al. (2022) constructed a corpus with image caption lexical aspects annotation across multiple languages, suggesting the consistency of expressing image semantics in terms of aspect and doing the automatical prediction experiments.

For Chinese, Li et al. (2016) and Song et al. (2020) used *The Little Prince* and CTB 8.0 as the raw data to annotate aspect markers. However, their annotations lack detailed specifications of aspect types, resulting in a shallow annotation.

3. Fundamental Aspect Values in UMR

UMR lattice encompasses 5 fundamental aspect values that are applicable across numerous languages: Habitual, State, Activity, Perform-

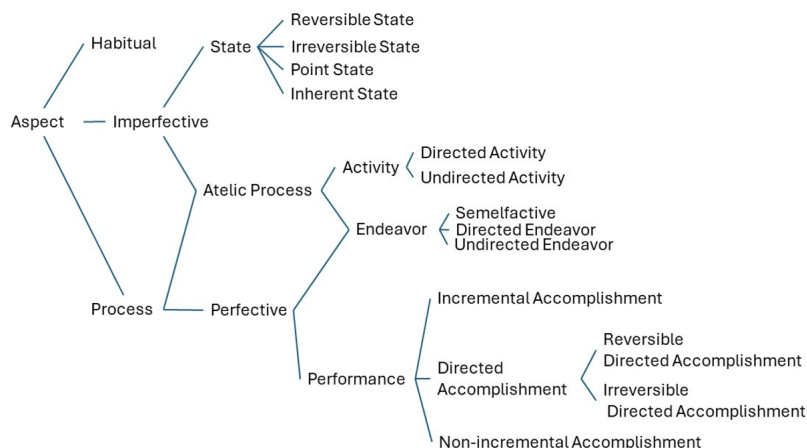


Figure 1: UMR aspectual lattice

mance, and Endeavor. The UMR guideline¹ employs a decision-tree-like methodology to guide annotators in selecting labels systematically.

- **Habitual** denotes events typically occurring or habitual in nature, akin to **Habitual** for viewpoint. In English, it often manifests in simple present constructions, with “used to” indicating habitual events in the past.
- **State** characterizes events where no changes occur during the event and are durative. Additionally, UMR extends to non-verbal clauses (e.g., “The doctor is tall”), modal verbs (e.g., “He wants to travel to Albuquerque”), and events within the scope of ability modals (e.g., “She is able to sing that aria”). A notable subcategory of **State** encompasses “inactive actions”, often involving postures (e.g., sit, lie), perception (e.g., see/look at), sensation (e.g., ache), and mental activities (e.g., think, understand). This one is the same as **State** in the situation types.
- **Activity** refers to ongoing events that have not reached a conclusion, contrasting with **Performance** and **Endeavor**. It aligns with the **Progressive** aspect under the viewpoint.
- **Performance** and **Endeavor** signify events that have concluded. The key distinction lies in whether the event attains a “natural result endpoint”, distinguishing between telic and atelic events. **Performance** implies a resulting state, often marked by completive aspectual markers (e.g., “finish” in English) or container adverbials, akin to **Achievement** and **Accomplishment** in **Vendler (1967)**’s system. **Endeavor** is annotated when explicit markers such as terminative aspect marking (e.g.,

“stop”), durative adverbials (e.g., “He ran for 5 minutes”), or non-result paths (e.g., “They walked along the river”) are present. This one is like **Activity** in the situation types.

There are some differences in the meaning of the term when compared to the system proposed by **Vendler (1967)** and **Smith (1991a)** as these values combine certain viewpoint with the lexical aspects.

4. Annotation Aspect in Chinese Little Prince

Due to the absence of an existing aspect annotation dataset in Chinese with UMR guidelines, we conducted manual annotations. This allowed us to explore the relationship between aspect markers and UMR aspect values, as well as to assess the annotation quality under the UMR guidelines.

4.1. Data Source

Our data source is Chinese AMR corpus for *The Little Prince*², which comprises 1562 sentences along with fully annotated AMR graphs (Li et al., 2016). A total of 517 aspect markers were labeled.

It’s worth noting that there are parallel “*The Little Prince*” AMR corpora available in multiple languages. This opens up opportunities for cross-lingual comparison and analysis and contributes to our choice for selecting the data.

4.2. Annotation Labels

For our annotation, we have chosen 5 fundamental aspectual values (**Habitual**, **State**, **Activity**, **Endeavor**, **Performance**). Unlike low-resource languages, where initial annotation often

¹<https://github.com/umr4nlp/umr-guidelines>

²https://www.cs.brandeis.edu/~clp/camr/res/littleprince12.1%28sl_new%29.txt

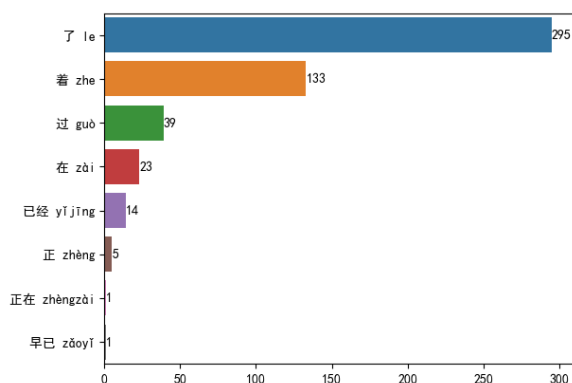


Figure 2: The distribution of aspect markers

employs coarse-grained label sets due to limited linguistic analysis, the Chinese aspect system has already undergone extensive research. Thus we opted for this fundamental label set. A finer-grained label set is deemed unnecessary, as it could increase the annotators' workload and chance of disagreement. This fundamental label set has been consistently applied in multiple languages, ensuring comparability in annotation and analysis.

4.3. Procedure and Inter-Annotation Agreement Analysis

In this study, we enlisted 4 volunteers to perform annotation. All volunteers are graduate students with expertise in linguistics and NLP, and they are native Mandarin Chinese speakers. Due to time constraints, we chose to focus our annotation efforts on sentences containing aspect markers. This approach aligns with our goal of observing the role of these markers in the selection of aspect values during the annotation process. Different from other aspect particles, the usage of “了 le” in Chinese is rather complex. Its functions can be divided into aspect particle (“了 le¹” in literature) and modal particle (“了 le²” in literature). “了 le” often appears at the end of a sentence or a word, but there is still considerable controversy over its exact functions in different positions. We annotate all “了 le” except for following cases: either “了 le” is part of an insertion or combined with “要 yao” to form “要....了”, indicating future tense (can be translated into “be going to”). Only 6 cases with “了 le” are excluded and take up 1.1% of all original markers. 511 aspect markers are left, which span across 403 sentences, and take up 26% of 1562 sentences in total. The aspect markers distribution is illustrated in Figure 2. We followed the aspect annotation part in the UMR guideline for annotation. Before commencing the annotation, all annotators diligently studied the UMR guidelines and received training on a small set of data from *The Little Prince*. For the interest of research, volunteers were asked to

mark challenging cases. For instance, in situations where multiple labels are applicable to a single event, annotators were permitted to assign multiple labels but were instructed to order them based on their own preference and pick up the first one as their choice to compute agreement, etc.

Initially, we began with 2 volunteers for annotation and observed a low inter-annotation agreement (IAA) based on the Kappa value (Cohen, 1960). we decided to engage more volunteers to avoid individual bias, bringing the total to 4 annotators. The Kappa values remained in the range of 0.47 to 0.58 (moderate agreement) with the expanded team when calculating pairwise consistency. We also computed Fleiss's Kappa, which is used to evaluate annotation consistency when multiple annotators are involved (Fleiss, 1971). The resulting Fleiss's Kappa value was 0.53 (moderate agreement) and 238 annotations are agreed by all 4 annotations. There are 6 annotations, and each annotator chose a completely different label.

In the following section, we initially conducted a quantitative analysis to examine the alignment between aspect markers and aspect values, addressing the first research question. This step was undertaken to determine the extent to which aspect markers influence aspect value annotation. Subsequently, to address the second research question, we carefully reviewed the annotations, aiming to comprehend the factors responsible for the lower-than-anticipated IAA. We gathered feedback from the annotators and engaged in discussions to categorize disagreed annotations. Furthermore, we selected representative examples from each category for empirical analysis.

5. RQ1: Relationship Between Aspect Markers and Aspect Labels

We initially tackled the question of whether aspect markers have a discernible influence on the annotation of aspect values. To investigate this, we established a ground truth by selecting the majority label among the 4 annotators for each annotation case. This means that if at least 3 annotators agree on one label, we take that label as the final label. In instances where a majority label could not be determined, all 4 annotators deliberated collectively until a consensus was reached, resulting in 151 annotations in this case. We then calculated the correlated distribution between the aspect marker and the aspect value. The results are presented in Table 1. The table reveals that there is no such case where the “marker” and “value” correspond one-to-one. It's important to note that aspect markers primarily focus on the viewpoint aspect, while the UMR lattice encompasses both viewpoint and situation types.

Aspect Markers	Activity	Endeavor	Habitual	Performance	State
了 <i>le</i>	2	58	1	121	113
着 <i>zhe</i>	53	0	11	0	69
过 <i>guo</i>	0	11	0	10	18
在 <i>zai</i>	17	0	2	0	4
已经 <i>yi jing</i>	0	1	0	4	9
正 <i>zheng</i>	1	0	0	0	4
正在 <i>zheng zai</i>	1	0	0	0	0
早已 <i>zao yi</i>	0	0	0	0	1

Table 1: The confusion matrix between aspect markers and aspect values

From the table, we can see some aspect markers express its aspect features, such as “了 *le*” marks the *Perfective*, which involves the *Endeavor* and *Performance* values. “着 *zhe*” is a marker for *Progressive*, corresponding to *Activity* value in UMR lattice. As UMR aspectual lattice also involves situation types, many events with these markers may also be annotated as *State*. Relying on the aspect marker to determine the aspect value may not be conclusive regarding the varying distribution split from the table. Even so, the inheritance meaning of the label can help us exclude impossible labels and narrow down the range. For example, “着 *zhe*” is a *Progressive* marker that never occurs in the situations that should be labeled with *Performance* or *Endeavor*. Those events are ended events. Similarly, “过 *guo*” seems never to occur in the ongoing events as it is an *Perfective* marker.

6. RQ2: Patterns of Disagreement Cases

In this section, we delve into specific cases that exhibit ambiguity and disagreement among annotators. Our aim is to elucidate the fuzzy instances that contributed to the low inter-annotation agreements. We selected examples that, among the 4 annotators, were considered uncertain (assigned multiple labels for a single event) by at least 3 annotators. These represent commonly ambiguous cases. Additionally, we identified cases where none of the labels received support from three or more annotators, indicating a lack of consensus and leading to disagreement. We collected feedback from annotators and then categorized these examples together through discussion. The cases extracted through this approach represent 16.8% of the total annotations.

6.1. The Blurred Definition of “Inactive Actions”

There are a lot of cases with “着 *zhe*” that make it hard to choose a label between *State* and *Activity*. It roughly constitutes 30% of the total cases.

In the UMR guidelines, “inactive actions” (Croft, 2012) are considered as *State*. Some verbs deviate from prototypical “inactive actions”. Since “着 *zhe*” is a *Progressive* marker, combining with some verbs can be considered as *Activity* or *State*. e.g.³,

- (1) 但是，却有一种
but, CONT EXIST one CLF can.
说不出的东西在
not.express NPAR thing *zai*.IPFV
默默地放着光芒 ...
saliently VPAR emit *zhe*.IPFV light ...

Yet through the silence something throb, and gleams ... (id:1255, CONT=contrast, NPAR=structural particle for Noun, VPAR=structural particle for verb)

Here “在 *zai*” and “着 *zhe*” express the ongoing meaning. Emitting light certainly requires some mechanisms to radiate the light continuously, which involves a dynamic process rather than static. Therefore, it can be classified as both an “inactive action” and labeled as *State*, or labeled as *Activity* based on the context and interpretation.

- (2) 在法国，正夕阳
in.LOC France, *zheng*.PROG sun
西下，只要
down.to.west, only.COND
在一分钟内
within.PREP one minute under.PREP
赶到法国就可看到日落。
reach France then could see sunset .

If you could fly to France in one minute, you could go straight into the sunset, right from noon. (id:280 PREP=preposition)

This sentence indeed has multiple interpretations. “The sun rises in the east and sets to the

³I adopted the English AMR *The Little Prince* corpus as the corresponding English translation for reference as well as the sentence id, but ignore the graph for saving space, the glossing rule adopted from the Leipzig Glossing Rules: <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>

west every day” is a basic commonsense, suggesting that it should be considered as *Habitual* based on the step-by-step annotation procedure defined in the guideline. However the aspect marker “正 zheng” emphasizes the progressive and ongoing status, which can be considered as *Activity*. The third interpretation considers it as an “inactive action”, where it’s more like a continuous state. Similarly, “所有的星星上都好象开着花” (all the stars are a - bloom with flowers, id:1425). It’s unclear whether “开花” (bloom) is a kind of “inactive action” or not.

The point here is the aspect marker carries the ongoing meaning while the verb itself might be a stative verb. The UMR framework, which allows only one aspect label, struggles to handle these cases because they can involve values in both viewpoint and situation aspects separately.

6.2. Confusing Between Endeavor and Performance

One of the challenges arises from annotators struggling to reach an agreement on whether an event falls under the category of *Performance* or *Endeavor*. This accounts for approximately 13% to 20% of the total challenging cases. Distinguishing these two categories relies heavily on the telicity of the verb, which can be hard to determine sometimes.

This highly relies on how individual annotators define the specific natural outcome of a verb. Verbs that frequently lead to this confusion include words like “帮” (help), “讲” (tell) in phrases like “讲这个故事” (tell this story), among others. Annotators who favor *Endeavor* argue that “帮” (help) does not inherently result in a specific natural result. You can give assistance to someone you’ve helped, but it is not necessary to achieve a particular goal or bring benefits to the recipient. Similarly, in the case of “讲这个故事” (tell this story), there is no requirement to present the whole story comprehensively, maybe just mentioning an episode of the whole. Annotators who favor *Performance* consider it’s a default value defined in the guideline and there is no concrete marker in the context to indicate *Endeavor*.

6.3. The Event Involves the Inception or Maintenance of a State

A significant portion of the challenges arises from distinguishing between *Performance/Endeavor* and *State*, accounting for roughly 25% of the challenging cases. This confusion is related to the *Perfective* marker “了 le” and “过 guo”, “过 guo” signals that the event has been experienced at least once at some indefinite time which is usually in the indefinite

past (Li and Thompson, 1981), it can be taken as providing an existential quantification over times that are earlier than the “过 guo” sentence’s reference time (Mangione and Li, 1993). It can be translated into “ever” in English. On the other hand, “了 le” presents a situation in its entirety, without reference to its internal structure (Li and Thompson, 1981; Li, 1990; Shi, 1990; Smith, 1991a). It can represent the completion or the termination depending on whether the verb is telic or atelic (Chen, 2008). Sometimes the aspect marker “了 le” combined with events could express a kind of state meaning, e.g., the negation marker with “过 guo” emphasizes the subject does not possess a kind of experience.

Below is an example:

- (3) 他 从来 没 闻 过 一
3SG.M never NEG smell *guo*.PFV one
朵 花.
CLF flower.

He has never smelled a flower. (id:333)

The predicate in this sentence is “闻” (smell), it should be annotated as *Endeavor* if we consider it is not an “inactive action”. The action has ended and there is no natural endpoint for the action “smell”. However, it emphasizes the negation maker and expresses the missing of such an experience, which is more like a *State* rather than the *Endeavor*. The ambiguity in this case arises from the interplay among the inherent meaning of the verb, the *Perfective* marker “过 guo”, and the negation marker. It highlights the challenge of annotating aspectual values based on context and the implications carried by various linguistic elements.

Certain verbs can be inherently ambiguous as they can refer to both a specific event and the result carried by that event, which often corresponds to a new status or state. When combined with some aspect markers such as “过 guo”, and “了 le”, the interpretation of the sentence can be tricky sometimes, this happens in both Chinese and English. One example out of *The Little Prince* is

- (4) 它 睡着 了 ...
3SG fall.asleep *le*.PFV ...
He has gone to sleep...

This sentence suggests that “He” has fallen asleep in the past and remains in that sleeping status until speech time. Such a sentence should be annotated as *Performance* as the process “睡着” (fall asleep) expresses a kind of resultant meaning and cannot be terminated, similar to “build” and “bake” in English, we won’t say “停止睡着” (terminated falling asleep) in Chinese. But this event also refers to the ongoing state. Similar verbs like “认识” (make acquaintance), and “结婚” (marry) present a similar duality. If I say “就这样，我认识了小王

子” (and that is how I made the acquaintance of the little prince, id:102), it can mean both specific action “结识” (made the acquaintance), such as introducing each other and the durative status of knowing each other. Likewise, saying “我结婚了” (I have already gotten married) implies the specific action for being married (held the wedding or return the marriage license to the county clerk) and still in marital status. On the other hand, “我结过婚” (I have ever married) implies that I experienced the specific action while I am not in a marital status right now. The example is even more salient when combined with negation marker as mentioned, “我没结过婚” (I have never gotten married) implies the speaker is single right now, which is a kind of state, rather than indicating the absence of “marry” action. However, the current UMR guidelines classify such cases as negation in terms of polarity and *Performance* in terms of aspect, rather than considering them as a whole state (Vigus et al., 2020). Similar events in the corpus like “我的朋友带着他的小羊已经离去六年了” (six years have already passed since my friend went away from me, with his sheep). Here “went away” is *Performance* if we consider it as a simple action. However, the speaker emphasizes the ongoing passage of time from the past up to the speech time.

Clearly, the challenge here lies in whether we should strictly adhere to different grammatical categories or focus on the meaning as a bundled unit, as well as whether we should consider the implication in the annotation. Undoubtedly, implication plays a role in understanding aspectual meaning, which is composed of the verb’s inherent meaning, tense, arguments, and the context that triggers softer implications and hard inferences (Friedrich et al., 2023). Moens and Steedman (1988)’s approach is an alternative way to resolve the issue, it incorporates the consequent state into consideration. This is similar to including the phases in the annotation. Chao (1968) argued that “了 le” can indicate the inception or inchoative of a situation.

6.4. “着 zhe” Used in Comitative Events

This involves multiple events occurring simultaneously, accounting for approximately 10% of all cases. The aspect marker “着 zhe” indicates the durative, resultive state, often serving as background information (Friedrich et al., 2023). It can be used with the “V+ 着+V” construction (see Klein et al. 2000) and involves multiple predicates, as illustrated below:

- (5) 那会儿 我 正 忙
 at-that-time 1SG zheng.PROG busy
 着 从 发动机 上
 zhe.IPFV from.PROP engine LOC

卸下 一颗 拧 得
 unscrew one CLF fasten ADJP
 太 紧 的 螺 丝 。
 too.DRG tight NPAR screw .

At that moment I was very busy unscrewing a bolt that had got stuck in my engine. (id:280, PROP=prepositio,ADJP=sturctual particle for adverb and adjectives, DRG=degree,NPAR=participle for nouns)

In this case, two predicates are involved: “忙” (busy) and “卸下” (unscrew). “忙” (busy) is *State*, while “卸下” (unscrewing) is an ongoing dynamic process. The question arises whether we should consider this sentence as one event or two. If we consider it as one event, we face challenges in selecting the appropriate label since these two labels are different. If we consider them separately, it might seem contradictory to the language sense, as “着 zhe” stresses simultaneity. Such a situation is cross-lingual, akin to the “be busy doing” construction in English.

6.5. Overlapping between Habitual, Activity, and State

There are other cases that challenge annotations to make a decision among *Habitual*, *State*, and *Activity*. For example, in the sentence “第二个行星上住着一个爱虚荣的人” (the second planet was inhabited by a conceited man, id:604), the interpretation can be in three ways: 1) A conceited man used to live on the second planet. This should be labeled as *Habitual*. 2) A conceited man was living on the second planet. This can be labeled as *Activity* due to “着 zhe” in the sentence. 3) more aligned with the English translation here, take “inhabit” as an “inactive action” and label it as *State*. 2) and 3) relate to the discussion in the section 6.1.

7. Discussion

Furthermore, we identified two categories of aspects in Chinese that are not well-covered and pose compatibility challenges with the current UMR guideline.

7.1. Iterative

There is a subtype of *Imperfective* aspect known as *Iterative*, denoting repetition (iteration) of an event on a single occasion. It indicates that the event occurred as a series of bounded wholes (Velupillai, 2012). For instance, “He practices the presentation over and over” and “He goes in and out”.

In Chinese, the meaning type of *Iterative* can be varied. It can be the repetition of a single action, like “走来走去” (to walk around), alternation of different actions like “走走停停” (start-stop), or the repetition of symmetry actions involving different subjects, like “一问一答” (one ask one answer). Or different subject engaged in different actions happening simultaneously, as in “进进出出” (in-out). The examples in *The Little Prince* are shown below:

- (6) 小 王 子 的 话 在
 little Prince POSS words LOC
 我 的 脑 海 中
 1SG.POSS NPAR brain PROP
 跳来跳去。
 hopping.back.and.forth

The little prince 's last words came reeling back into my memory (id:1239)

In this example, the author employs a metaphorical expression, “跳” (jump) in the speaker’s brain means I recall it over and over again. The construction “跳来跳去” (jump here and there back and forth) can be generalized into “V 来V 去”.

In UMR, there is no such specific value corresponding to the *Iterative*.

7.2. Duplication

Smith (1991b) introduced the concepts of “open situation” and “closed situation” based on the existence of a final point in events. In the UMR, there are two possible values for annotating ended events: *Endeavor* and *Performance*. In Chinese, there is a construction called *Duplication* expressing events that occur in a short time and lack a natural endpoint, even when the aktionsart of the verb is telic. While *Semelfactive* can apply to this situation, it often occurs as iterative actions that return to the initial state at the end. *Duplication* is a milder construction that emphasizes a short duration and a small number of occurrences, making the action unable to extend indefinitely. It is characterized by less volition and implies a sense of attempting arbitrariness, and casualness (Chen, 2008). As a result, the Chinese *Duplication* aspect cannot be completely equated with the *Semelfactive* aspect. An example is illustrated below:

- (7) 我 使 劲 地 揉 了
 1person.SG vigorously SP-verb rub le
 揉 眼 睛 ,
 rub eyes .

I rubbed my eyes hard. (id:48)

In this context, it means that I just rubbed my eyes briefly. The Chinese verb “揉” (rub) is typi-

cally considered a durative verb. However, even when combined with an atelic verb, *Duplication* in Chinese expresses the repetition of a simple verb several times within a short period in a casual manner. This construction is somewhat akin to the combination of an atelic verb with a durative time expression in English, such as “she coughed for 1 minute”. For *Duplication*, the emphasis lies on a shorter duration, a smaller number of repetitions, and a casual quality.

The existing UMR aspectual lattice lacks specific values to account for the aspect constructions mentioned above. To circumvent this issue, we can opt for a more general label during annotation. However, using a coarser label may result in a loss of the specific internal semantics and structure of these constructions.

7.3. Guideline Retrofitting Proposal

To enhance the UMR annotation process, we propose annotating the situation and viewpoint separately. We engaged annotators in this approach to compare their agreement with the UMR guideline. For the grammatical aspect, we selected labels including *Perfective*, *Progressive*, *Habitual*, and *Iterative*. The latter three are subcategories of *Imperfective*. For situation types, we chose 5 values from Table A1 as labels. We provided annotators with two tables as the annotation guideline and referred them to aspect-related chapters in Saeed (2011)’s and Velupillai (2012)’s works. These two books are utilized in the CU-Boulder linguistic department as textbooks. Two table notes Table A2&Table A3 can be found in the Appendices.

We calculated agreement when annotating the lexical aspect only, when annotating the grammatical aspect only, and when annotating both of them simultaneously. Additionally, we provided the pairwise Cohen’s agreement range when computing for lexical aspect only and grammatical aspect only. The detailed agreement results are presented in Table 2. Furthermore, we attempted to merge *Progressive*, *Habitual*, and *Iterative* into the *Imperfective* label to observe differences in agreement in terms of grammatical aspect, that is “grammatical aspect-binary” in the Table 2.

Overall, our analysis indicates that annotating lexical and grammatical aspects separately may not serve as a viable solution to the low agreement observed in UMR aspect annotation; in fact, the agreement tends to be lower than when annotating with the UMR guideline. This discrepancy is somewhat expected, given the multitude of labels—ten in total—resulting from the combination of lexical aspect and even if just two grammatical aspect labels (*Perfective* and *Imperfective*). Notably, while annotating grammatical aspect in isolation

Annotation type	Pair-wise cohen's kappa	Fleiss's kappa
Umr lattice	0.48~0.59	0.53
Lexical aspect	0.45~0.57	0.48
Grammatical aspect	0.57~0.69	0.65
Grammatical aspect-binary	0.6~0.73	0.69
Lexical aspect + grammatical aspect	N/A	0.41
Lexical aspect + grammatical aspect-binary	N/A	0.42

Table 2: The annotation result when annotating separately

yields high agreement (row 3 and row 4), this decreases significantly when combined with lexical aspect (row 5 and row 6), underscoring the challenges primarily stemming from lexical aspect annotation. Annotating lexical aspect poses significant challenges for annotators, who often encounter puzzles discussed in section 6, such as distinguishing between *State* and *Activity*, as well as *State* and *Accomplishment/Achievement*. This difficulty is exacerbated compared to UMR guidelines due to its step-by-step annotation process, wherein annotators must prioritize one label from multiple available options. The similar challenges coupled with the closer agreement score, affirm that the UMR aspect lattice leans more towards the lexical aspect angle.

While the aspect marker contributes to the relatively consistent annotation of grammatical aspect, there remain instances where annotators cannot make the decision solely through the aspect marker. For instance, in the sentence “我们唤醒了这口井，它现在唱起歌来了” (we have wakened the well, and it is singing, id:1290), the presence of “了 le” may intuitively suggest a *Perfective* aspect. However, it is arguably more appropriate to consider it as *Progressive*, as “了 le” and the preceding “起来” together imply the inchoative phase of the “唱” (sing) action, which is ongoing at the speech point rather than being completed (*Perfective*). Another example, “夜晚，当你望着天空的时候...” (at night, when you gaze at the sky...,id:1452), includes “着 zhe”, often indicates the *Progressive* aspect. However, the sentence conveys a constant conditional situation, suggesting a habitual rather than a specific ongoing action. Thus, we prefer to annotate it as *Habitual*.

8. Conclusion

Ongoing efforts within the UMR community are dedicated to developing aspectual representations that consider the topological aspects and work consistently across various languages. Data annotation is a particularly challenging aspect of this endeavor (Friedrich et al., 2023). This paper represents our attempt to apply the UMR aspect scheme to the Chinese *The Little Prince* AMR corpus, with a particular focus on the influence of aspect markers on labeling decisions.

The major contributions of our work can be summarized as follows: 1. We engaged four annotators with linguistic expertise and training to perform annotations. Despite this approach, we encountered a less substantial inter-annotator agreement score, highlighting the inherent challenges in this task. 2. Our analysis reveals that aspect markers in Chinese do not exhibit a one-to-one correspondence with the aspect values defined in the UMR lattice. This discrepancy can be attributed to the fact that aspect markers primarily emphasize the viewpoint, while the UMR framework combines both viewpoint and situation aspects in its considerations. 3. We systematically categorized and analyzed complex cases that led to inconsistencies among annotators. These cases often revolved around the interpretations of annotators regarding the lexical semantics of predicates, implications, and specific linguistic constructions. This underscores the UMR guidelines' inclination toward lexical aspects and the importance of lexical semantic understanding. Additionally, we identified language-specific aspects in Chinese, such as *Duplication* and *Iteration*, which are not adequately addressed by the current UMR guidelines. 4. We proposed a refined scheme to annotate the lexical aspect and grammatical aspect separately, such a way failed to improve the agreement because there are still many tricky cases when annotating situation types, similar to annotations with UMR, the similar puzzle and agreement score result also suggest the UMR aspect is more bias towards situation type. The agreement for the viewpoint annotation verifies the correlations between aspect markers and the viewpoint.

9. Acknowledgements

Thanks to Prof. Andrew Cowell and Jie Cao from the University of Colorado Boulder who provides the feedback for the paper and Prof. Xiuju Shi from Shanxi University who offers valuable comments on the tricky Chinese annotation cases.

10. Bibliographical References

- Malihe Alikhani, Thomas Kober, Bashar Alhafni, Yue Chen, Mert Inan, Elizabeth Nielsen, Shahab Raji, Mark Steedman, and Matthew Stone. 2022. [Zero-shot cross-linguistic learning of event semantics](#).
- 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, Sofia, Bulgaria. Association for Computational Linguistics.
- R.I. Binnick. 1991. *Time and the Verb: A Guide to Tense and Aspect*. Filologia y linguistica. Oxford University Press.
- Julia Bonn, Skatje Myers, Jens E. L. Van Gysel, Lukas Denk, Meagan Vigus, Jin Zhao, Andrew Cowell, William Croft, Jan Hajič, James H. Martin, Alexis Palmer, Martha Palmer, James Pustejovsky, Zdenka Urešová, Rosa Vallejos, and Nianwen Xue. 2023. [Mapping AMR to UMR: Resources for adapting existing corpora for cross-lingual compatibility](#). In *Proceedings of the 21st International Workshop on Treebanks and Linguistic Theories (TLT, GURT/SyntaxFest 2023)*, pages 74–95, Washington, D.C. Association for Computational Linguistics.
- Yuen Ren Chao. 1968. *A grammar of spoken Chinese / by Yuen Ren Chao*. University of California Press Berkeley.
- Daniel Chen, Martha Palmer, and Meagan Vigus. 2021. [AutoAspect: Automatic annotation of tense and aspect for uniform meaning representations](#). In *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 36–45, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Qianrui Chen. 2008. *The typological perspective of Chinese aspectual research*. The Commercial Press.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.
- Bernard Comrie. 1976. *Aspect : an introduction to the study of verbal aspect and related problems*. Cambridge University Press London, New York, [etc.].
- William Croft. 2012. *Verbs: Aspect and Causal Structure*. Oxford University Press.
- Osten. Dahl. 1985. *Tense and aspect systems*. B. Blackwell Oxford [Oxfordshire] ; New York, NY.
- L. Donatelli, Nathan Schneider, W. Bruce Croft, and Michael Regan. 2019. [Tense and aspect semantics for sentential amr](#).
- Lucia Donatelli, Michael Regan, William Croft, and Nathan Schneider. 2018. [Annotation of tense and aspect semantics for sentential AMR](#). In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 96–108, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- JL Fleiss. 1971. [Measuring nominal scale agreement among many raters](#). *Psychological bulletin*, 76(5):378—382.
- Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. 2016. [Situation entity types: automatic classification of clause-level aspect](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1757–1768, Berlin, Germany. Association for Computational Linguistics.
- Annemarie Friedrich, Nianwen Xue, and Alexis Palmer. 2023. [A kind introduction to lexical and grammatical aspect, with a survey of computational approaches](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 599–622, Dubrovnik, Croatia. Association for Computational Linguistics.
- 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](#). *KI - Künstliche Intelligenz*, 35:343 – 360.
- Wolfgang Klein, Ping Li, and Henriette Hendriks. 2000. [Aspect and assertion in mandarin chinese](#). *Natural Language Linguistic Theory*, 18(4):723–770.
- 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, Berlin, Germany. Association for Computational Linguistics.

- Bin Li, Yuan Wen, Li Song, Weiguang Qu, and Ni-anwen Xue. 2019. [Building a Chinese AMR bank with concept and relation alignments](#). *Linguistic Issues in Language Technology*, 18.
- Charles N. Li and Sandra A. Thompson. 1981. *Mandarin Chinese : a functional reference grammar / Charles N. Li and Sandra A. Thompson*. University of California Press Berkeley.
- P. Li. 1990. [Aspect and Aktionsart in Child Mandarin](#). Li Ping.
- L. Mangione and Dingzuan Li. 1993. [A compositional analysis of -guo and -le](#). *Journal of Chinese Linguistics*, 21(1):65–122.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, 19(2):313–330.
- Thomas A. Mathew. 2009. Supervised categorization for habitual versus episodic sentences.
- Marc Moens and Mark Steedman. 1988. [Temporal ontology and temporal reference](#). *Computational Linguistics*, 14(2):15–28.
- Tim O’Gorman, Michael Regan, Kira Griffitt, Ulf Her-mjakob, Kevin Knight, and Martha Palmer. 2018. [AMR beyond the sentence: the multi-sentence AMR corpus](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3693–3702, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- J.I. Saeed. 2011. [Semantics](#). Introducing Linguistics. Wiley.
- Ziqiang Shi. 1990. Decomposition of perfectivity and inchoativity and the meaning of the particle *le* in mandarin chinese. *Journal of Chinese linguistics*, pages 95–124.
- Carlota Smith. 1991a. [The Aspectual System of Mandarin Chinese](#), volume 43, pages 343–390.
- Carlota Smith. 1991b. [The parameter of aspect](#).
- Li Song, Yuan Wen, Sijia Ge, Bin Li, and Weiguang Qu. 2020. [An Easier and Efficient Framework to Annotate Semantic Roles: Evidence from the Chinese AMR Corpus](#), pages 474–485.
- Jens E. L. Van Gysel, Meagan Vigus, Lukas Denk, Andrew Cowell, Rosa Vallejos, Tim O’Gorman, and William Croft. 2021. [Theoretical and practical issues in the semantic annotation of four indigenous languages](#). In *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 12–22, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Jens E. L. Van Gysel, Meagan Vigus, Pavlina Kalm, Sook-kyung Lee, Michael Regan, and William Croft. 2019. [Cross-linguistic semantic annotation: Reconciling the language-specific and the universal](#). In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 1–14, Florence, Italy. Association for Computational Linguistics.
- V. Velupillai. 2012. [An Introduction to Linguistic Typology](#). John Benjamins Publishing Company.
- Zeno Vendler. 1967. [Linguistics in Philosophy](#). Cornell University Press, Ithaca, NY.
- Meagan Vigus, Jens E. L. Van Gysel, Tim O’Gorman, Andrew Cowell, Rosa Vallejos, and William Croft. 2020. [Cross-lingual annotation: a road map for low- and no-resource languages](#). In *Proceedings of the Second International Workshop on Designing Meaning Representations*, pages 30–40, Barcelona Spain (online). Association for Computational Linguistics.
- R. Xiao and T. McEnery. 2004. [Aspect in Mandarin Chinese: A Corpus-based Study](#). Studies in language companion series. J. Benjamins Pub.

11. Appendices

Appendix A. Situation types

situation types	static	durative	telic	example
State	+	+	-	He hated ice cream.
Activity	-	+	-	Your cats watched those birds.
Accomplishment	-	+	+	He learned Japanese.
Achievement	-	-	+	He reached the peak.
Semelfactive	-	-	-	The gate banged.

Table A1: Situation types (follows Smith (1991b))

Appendix B. The table notes for annotating situation types and viewpoint

Type	Durative	Dynamic	Telicity	Example	Explanation
State	+	-	-	know the answer, love Mary	can't say "I am knowing", because it's a state
Activity	+	+	-	laugh, swim,run	I run 10 mins, durative, dynamic, no result
Achievement	-	+	+	reach the top	reach the top is just a instant action, not durative,
Accomplishment	+	+	+	build a house, make a cake	build requires a duration, cannot finish if we break in the middle
Semelfactive	-	+	-	tap,knock,tap,wink	instant action

Table A2: The table notes for the situation types

type	example	explanation
progressive	I am eating	specifically denotes that the event is ongoing
perfective	I watched the movie	view the event as a whole from an outside perspective
habitual	used to	denotes that an event takes place regularly or is true for an extended period
iterative	He blew and blew	indicating that the event took place as a series of bounded wholes,repeative

Table A3: The table notes for the viewpoint