

Developing Universal Dependencies for Mandarin Chinese

Herman Leung^{*} Rafaël Poiret[†] Tak-sum Wong^{*}
Xinying Chen[§] Kim Gerdes[¶] John Lee^{*}

^{*} City University of Hong Kong, [†] Zhejiang University,

[§] Xi'an Jiaotong University, [¶] Sorbonne Nouvelle, LPP (CNRS)

leung.hm@gmail.com, raphael_poiret@zju.edu.cn, tswong-c@my.cityu.edu.hk,
chenxinying@mail.xjtu.edu.cn, kim@gerdes.fr, jsylee@cityu.edu.hk

Abstract

This article proposes a Universal Dependency Annotation Scheme for Mandarin Chinese, including POS tags and dependency analysis. We identify cases of idiosyncrasy of Mandarin Chinese that are difficult to fit into the current schema which has mainly been based on the descriptions of various Indo-European languages. We discuss differences between our scheme and those of the Stanford Chinese Dependencies and the Chinese Dependency Treebank.

1 Introduction

At a time when dependency syntax is quasi-hegemonic in natural language processing (NLP), dependency treebank construction was until recently solely based on translating existing phrase-structure-based resources into dependencies (e.g., the Penn Treebank and the Stanford Dependency Parser; de Marneffe, MacCartney, & Manning, 2006), performed in the computer science departments, often in very applied perspectives. Only recently, the re-emergence of dependency-based linguistics put into question the syntactic principles underlying treebank construction. Although dependency annotation seems to be more consensual than based on phrase-structure analyses (possibly due to the lighter annotation without non-terminals or traces), different schools foster different annotation standards. For example the Prague Dependency Treebank (Böhmová, Hajič, Hajičová, & Hladká, 2003) is based on theoretical works of the Prague team (Sgall, Hajicová, & Panevová, 1986) and the Spanish MTT treebank (Mille, Vidal, Barga, & Wanner, 2009) is based on Mel'čuk's Meaning-Text Theory (1988). The annotation schemes differ mainly in the placement and number of different layers of annotation (semantics vs. deep-syntax vs. surface-syntax).

The Universal Dependencies (UD) project (de Marneffe et al., 2014; Nivre et al., 2016a) constitutes an important homogenization effort to synthesize ideas and experiences from different dependency treebanks in different languages, with the aim of facilitating multilingual research on syntax and parsing by proposing a unified annotation scheme for all languages. Up to the current version of UD (1.3)—which has released annotated data in 40 languages (Nivre et al., 2016b)—the proposed structure has been unique (no separate semantic or surface-syntactic annotation¹). The scheme has triggered some debate on the syntactic foundation of some choices that have been made (Osborne, 2015), in particular because UD does not rely on one theoretical framework and some of the proposed goals are necessarily contradictory: syntactic correctness, applicability of the schemes for NLP tools and purposes, and above all universality (similarity of structures across languages) cannot all be fulfilled at the same time (Gerdes & Kahane, 2016). Although no separate explicit annotation scheme exists for most UD treebanks, universality seems to outweigh other considerations.

This paper describes similar choices in our adoption of UD for Mandarin Chinese, but we will try to be explicit about the advantages and disadvantages of the choices we made. The gaps and problems we describe show more generally that syntactic category and function sets that were originally created for Indo-European languages need important changes and careful balancing of criteria to foster typologically different languages, so that the distinctions become truly universal as intended. Some of

¹ This may change in the upcoming version of the UD scheme (2.0).

these problems can be solved by a greater universality of the vocabulary to describe the syntactic distinctions. Some idiosyncrasies simply do not have a satisfying *and* universal description.

The article starts out with a brief overview of existing dependency annotation schemes for Mandarin Chinese and how they compare overall to the UD scheme. We describe a few of the Mandarin POS tag choices of our scheme in section 3. Section 4 is devoted to the important features of our dependency annotation scheme and the sub-types of dependency relations we introduce.

2 Dependency schemes for Mandarin Chinese

Two widely used dependency schemes for Mandarin Chinese are Stanford Dependencies (SD) for Chinese (hereafter Stanford Chinese), developed by Huihsen Tseng and Pi-Chuan Chang (see Chang, 2009; Chang, Tseng, Jurafsky, & Manning, 2009), and the Chinese Dependency Treebank (CDT) developed by the Harbin Institute of Technology Research Center for Social Computing and Information Retrieval (see Che, Li, & Liu, 2012; HIT-SCIR, 2010; Zhang, Zhang, Che, & Liu, 2012). Stanford Chinese adopts its part-of-speech (POS) tagset directly from the Chinese Treebank (CTB) currently maintained at Brandeis University (Xue et al., 2013), also previously known as the Penn Chinese Treebank (hereafter Penn Chinese).

We have adapted the first version of Universal Dependencies (UD) for Mandarin Chinese (hereafter Mandarin UD) with reference to these two dependency schemes as well as the POS system of Penn Chinese. While we have taken many elements primarily from Stanford Chinese and CTB/Penn Chinese, due to their closer relation and structure to UD as well as existing SD-to-UD transformation tools, we have also made some choices that differ from some traditional Chinese linguistics analyses which Stanford Chinese and CDT follow. We will discuss these differences in detail in subsequent sections.

At the macroscopic level, our implementation of Mandarin UD differs from the other two dependency schemes in the division of labor between POS tags and relations, summarized in Table 1.

	Stanford Chinese	CDT (Harbin)	Mandarin UD
Total POS tags	33 (rich in verbs and function words)	26 (rich in nouns—8 total)	17
Total relations	45	15	57 (39 standard, 18 language-specific)

Table 1: Summary comparison of dependency schemes

The much smaller set of POS tags in Mandarin UD, albeit due to UD restriction, is compensated by a greater number of dependency relations, under the tenant of avoiding redundancy in annotating grammatical structures in both layers of labels. While Stanford Chinese has the highest amount of such redundancy among the three schemes, CDT takes the opposite approach of simplifying both the POS tags and dependency relations. Collapsing CDT’s 8 noun POS categories into just two (nouns and proper nouns), CDT would have only three more POS tags than Mandarin UD. However, CDT differs from SD and UD in its goal of being compatible with logical semantic representation (Zhang et al., 2014). On the other hand, UD was created as a successor to SD, which was initially created to represent English syntax only (de Marneffe et al., 2014). UD was crucially created for multilingual research on syntax and parsing (Nivre et al., 2016a), hence its emphasis on cross-linguistic compatibility and rules regarding how each language should adopt it.

In the next two sections, we discuss some of the more salient examples that illustrate the issues we encountered in developing Mandarin UD.

3 Parts-of-speech annotation

Mandarin UD uses all of UD’s 17 parts-of-speech (UDPOS) tags (Nivre et al., 2016a). We adopt heavily from the Penn Chinese Treebank POS system (Xia, 2000b) but differ from it in a few places, since UD’s tagset is smaller and does not correspond neatly to all of Penn Chinese’s tags. Since UD does not allow sub-typing of POS tags or language-specific tags, we adhere to this restriction. Below we discuss issues in adapting UDPOS for Mandarin with regard to predicate adjectives, localizers, and classifiers.

3.1 Predicate adjectives

In Chinese grammar what can be considered adjectives include two word classes, the non-predicate adjectives also known as 區別詞 *qūbiécí*, and the predicate adjectives also known as 形容詞 *xíngróngcí*. The non-predicate adjectives can only occur as prenominal modifiers. However, the predicate adjectives, despite its name, can occur both as a prenominal modifier (1a-b) and as a predicate (1c). When acting as a predicate, they are essentially intransitive stative verbs. When multisyllabic predicate adjectives act as prenominal modifiers, the particle 的 *de* is required (1b).

(1a)	好花 <i>hǎo huā</i> good flower 'good flower'	(1b)	美麗的花 <i>měilì de huā</i> beautiful DE flower 'beautiful flower'	(1c)	那花很好 <i>nà huā hěn hǎo</i> that flower very good 'that flower is good'
------	--	------	--	------	---

While Penn Chinese treats predicate adjectives as a type of verb ('VA') separate from the nominal modifier ('JJ'), we group them together as CDT does with the non-predicate adjectives as 'ADJ', and treat the particle 的 *de* as an adjectival marker in this case (such as in (1b)). The advantage of shuffling this subclass of verbs is that we are able to separate the intransitive stative uses of the predicate verbs from other verbs, since UD does not allow one to create subcategorical POS tags. Additionally, the modifier and predicate uses of predicate verbs are easily differentiated from each other simply by looking at their head in the dependency representation. Since the decision to tag predicate adjectives as ADJ is also supported in other languages such as in the Japanese implementation of UD (Tanaka et al., 2016), we consider our categorization to be more advantageous for cross-linguistic comparison.

3.2 Localizers

This class of words is known in Chinese linguistic literature as 方位詞 *fāngwèicí*. They come after a noun and primarily indicate spatial information in relation to the noun (with grammaticized uses for temporal and other abstract concepts of location), and are often additionally paired with the preposition 在 *zài*. Examples include 上 *shàng* 'above', 中 *zhōng* 'middle', 外 *wài* 'outside', 前 *qián* 'front', 旁 *páng* 'side', among others. Both Penn Chinese and CDT give localizers a unique tag—'LC' for 'localizer' (2a) and 'nd' for 'direction noun' (2b), respectively.

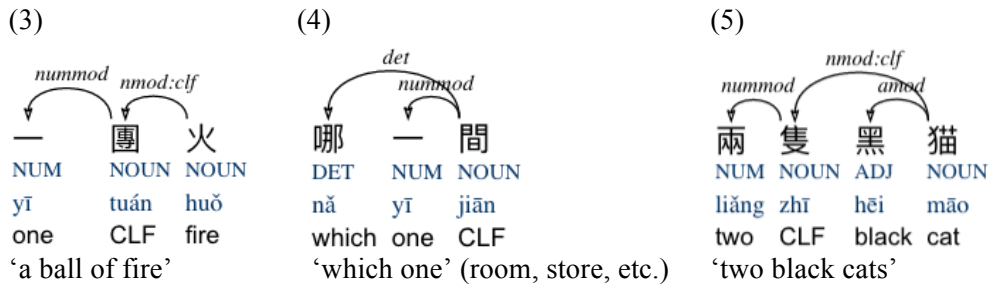
(2a) <u>Stanford Chinese</u>	(2b) <u>CDT</u>	(2c) <u>Mandarin UD</u>
<p>在山腰間 P NN LC</p>	<p>在山腰間 p nl nd</p>	<p>在山腰間 ADP NOUN ADP</p>
zài shānyāo at mountain.waist 'on the mountainside'	jiān between	

Although localizers historically derive from nouns, they no longer have the same syntactic distribution of regular nouns and are rather limited in their usage in providing spatial information to another noun. While CDT categorizes them as nouns and Penn Chinese appears semi-agnostic in giving them a unique tag, we treat them as postpositions after early research (Peyraube, 1980; Ernst, 1988) as well as recent research (Djamouri, Waltraud, & Whitman, 2013; Waltraud, 2015), the latter of which take into account historical Chinese data as well as cross-linguistic observations, notably German which also has both prepositions and postpositions. Therefore we tag them as ADP, as adpositions (2c).

3.3 Classifiers

Classifiers are an indispensable lexical category in Mandarin as well as many East Asian and Southeast Asian languages. In Mandarin, they are often obligatorily present with a numeral modifying a

noun (3)². Often they are also the head of a nominal phrase when a regular noun is not present (4). They differ from nouns in that classifiers can be preceded by numerals in all syntactic contexts, but nouns can only be preceded by numerals without an intervening classifier in exceptional cases. Additionally, attributive adjectives can never immediately precede or modify a classifier, but they do so with nouns, so that a noun phrase involving all three must have the order *classifier-adjective-noun* (5). It is likely due to the unique syntactic distribution of classifiers that both Penn Chinese and CDT give them unique POS tags—‘M’ for ‘measure word’ and ‘q’ for ‘quantity’, respectively.



Nonetheless, due to the classifiers’ partially similar syntactic distribution (when acting as the head of a noun phrase) as well as close relation to nouns—with analogy to measure words in languages such as English (e.g., *a head of cattle*) and Danish (e.g., *en kop kaffe* ‘a cup of coffee’)—we decided to place them under the tag NOUN, since UDPOS prohibits language-specific tags. We also considered PART (‘particle’), but this would conflate classifiers with function words that cannot function as heads.

The distinction between regular nouns and classifiers is currently preserved with the dependency relation *nmod:clf*, which is used to label a classifier when it precedes a noun (3, 5). However, the distinction is lost when the classifier is the head of a noun phrase (4). We are not currently implementing features in our Mandarin UD, but a *NounType* feature may be a future consideration. We are also in discussion with the UD community in order to include this widely neglected category in the upcoming UD 2.0 specifications.

4 Syntax annotation

Our adoption of UD for Mandarin Chinese has presented a number of syntactic challenges, some of which are due to particular constructions whose analyses are controversial or under-researched, and some are due to what might be insufficiencies in the UD design itself. We discuss some of these issues in the subsections below.

We use 39 of the 40 dependency relations available in UD as laid out in Nivre et al. (2016), leaving out *expl* since expletives do not exist in Mandarin, and additionally propose 18 language-specific dependency relations as permitted by UD, shown in Table 2.

Label	Description	Label	Description
acl:irrealis	‘irrealis descriptive clause’	discourse:sp	‘sentence particle’
acl:realis	‘realis descriptive clause’	dojb:light	‘direct object of light verb’
advcl:purpose	‘adverbial purpose clause’	mark:dec	‘adjectival/complementizer/nominalizer DE’
advmod:df	‘duration & frequency adverbial modifiers’	mark:dev	‘manner adverbializer DE’
aux:aspect	‘aspect marker’	nmod:agent	‘agent in long BEI phrase’
case:loc	‘localizer’	nmod:clf	‘classifier modifier’
compound:dir	‘directional verb compound’	nmod:dojb	‘direct object in BA phrase’
compound:der	‘descriptive/extent DE compound’	nmod:poss	‘possessive nominal modifier’
compound:ov	‘other verb compound’	nmod:tmod	‘temporal nominal modifier’

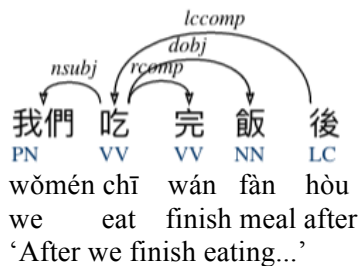
Table 2: Proposed language-specific relations in Mandarin UD

2 All tree diagrams in this paper illustrate Mandarin UD annotation unless explicitly labeled otherwise.

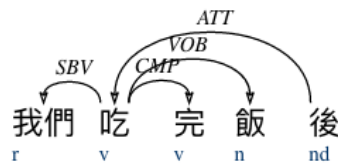
4.1 Adpositional phrases

One major systematic difference between UD and the previous SD (Stanford Dependencies) as well as CDT is in the treatment of adpositional phrases. Conforming to the SD system, Stanford Chinese treats all prepositions as well as postpositions as the head of adpositional phrases, with the nouns they introduce as their dependents. CDT employs the same treatment, as seen earlier in (2a-c) under the discussion of localizers in section 3.2. Since many of these adpositions also have grammaticized functions which introduce clauses instead of noun phrases, some clauses are also treated as dependents of these functions words in these schemes, illustrated in (6a-b) with a temporal adverbial clause marked by the clause-final 後 *hòu* ‘after’ (grammaticized from the postposition 後 *hòu* ‘behind’).

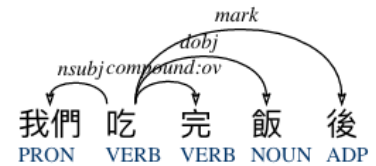
(6a) Stanford Chinese



(6b) CDT



(6c) Mandarin UD

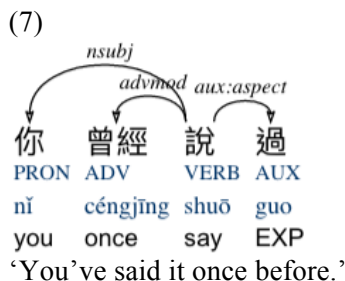


This approach of treating these function words as heads is abandoned in UD in favor of assigning heads to content words instead of function words (Nivre et al., 2016a). This not only means that the direction of the dependency relations for the above-mentioned function words are reversed in our implementation of Mandarin UD, as illustrated in (6a-c), but some of the head-child relations are also shifted, as illustrated in the earlier examples of localizers in (2a-c).

In the case of (2a-c), our implementation unfortunately loses hierarchical information between the preposition and postposition because we treat them both as direct dependents of the noun (2c), in contrast to Stanford Chinese (2a) and CDT’s (2b) approaches that indicate the postpositional/localizer phrase is embedded in the prepositional phrase. However, our approach is necessary to meet the demands of the UD rubric in treating content words as governors of function words.

4.2 Aspect particles

There are three basic aspect particles in Mandarin: perfective 了 *le*, durative 着 *zhe*, and experiential 過 *guo*. Although they are written as individual characters, they can be considered suffixes since they attach immediately after verbs and have lost their original tones, making them prosodically (as well as morphosyntactically) dependent. Etymologically they likely grammaticized from verbs but no longer retain any verbal characteristics (aside from conveying aspect). We propose the language-specific `aux:aspect` to link these particles to the verb (7).



These aspect markers fit well as a subcategory of the `aux` relation because that is the only relation whose associated tag AUX has a definitional mention of verbal marking including aspect (besides tense, mood, as well as person and number marking). However, UD specifies that the `aux` relation as well as the AUX tag are used only for auxiliary *verbs*, which seems to be problematic because verbal markers

are not all verbs or verbal in origin cross-linguistically. Mandarin aspect particles/suffixes are neither auxiliary verbs nor verbs, but unlike inflectional elements in morphologically rich languages, we treat them as separate tokens. Nonetheless, not all verbal markers in all languages are affixes (for example, see Bickel and Nichols, 2013), so in those languages the unbound verbal markers should be treated as individual tokens and should not be marked as part of the word token. We therefore believe there is a gap in the UD schema and that the UD definition for `aux` (and `AUX`) may benefit from revision in order to take non-affixal, non-verb verbal markers into account.

We have also considered using `compound` to link the aspect markers to the verb, but since this relation is used only for derivational or compositional structures rather than inflectional markers, it is not ideal, either. We tentatively annotate the aspect markers with `aux:aspect`, despite violation of the current UD definitions.

4.3 Sentence-final particles

Similar to classifiers, sentence-final particles are an areal feature across many East and Southeast Asian languages. They occur at the end of a sentence or utterance (8), and may have a wide variety of non-referential meanings and functions that modify the entire sentence, including modality, speech register, “speaker/hearer orientation”, and other discourse and pragmatic uses. In Mandarin, these particles include 嗎 *ma* (interrogative), 了 *le* (new situation), 吧 *ba* (command/suggestion), 的 *de* (certainty), among others. They are integrated into the sentence they attach to as part of its prosodic contour with no pause in between.



Although not all sentence-final particles have discourse functions, and more importantly they are very different from interjections, they seem to fit best as a sub-relation of `discourse` despite the fact that this relation is currently classified as a nominal dependent (of the main governor of a clause). We have also considered `advmod`, but the function and especially syntactic distribution of these particles are quite different from adverbs given that adverbs in Mandarin are usually preverbal and do not have a fixed syntactic position. We believe the lack of an obvious and natural space for these particles may be another possible gap in the UD schema. We tentatively propose `discourse:sp` for “sentence particle” to keep the name applicable cross-linguistically (for example, see Dryer, 2008, for varying positions of question particles across languages).

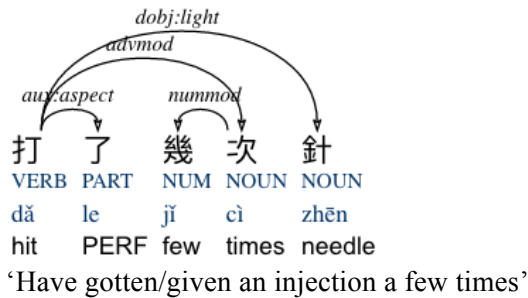
4.4 Light verb constructions

A number of verbs in Mandarin have semantically weakened and are used in combination with nouns to express new verbal concepts. A salient example is the verb 打 *dǎ*, which as a main verb by itself has the meaning “to hit” or “to strike.” However, this meaning is no longer present, or only a vestige of it is retained, when the verb combines with nouns like 針 *zhēn* ‘needle’, 球 *qiú* ‘ball’, 電話 *diànhuà* ‘telephone’: 打針 *dǎzhēn* ‘to get/give an injection’, 打球 *dǎqiú* ‘to play (a) ball (game)’, 打電話 *dǎdiànhuà* ‘to make a phone call’.

Ordinarily, UD includes light verb constructions under `compound`, listing Persian and Japanese as examples of languages with these constructions. However, in Mandarin the verb-noun “compound” is not a tight unit. The aspect markers still attach directly after the verb before the noun, as do duration and frequency adverbial phrases (9). The verb can further compound itself (10). For this reason we

propose *dojb:light* to link the noun to the verb since the noun still behaves like a direct object, and the dependency relation will enable a distinction between the light and full verb usages of the same Chinese character.

(9)



(10)



4.5 Descriptive clauses

Among the most difficult structures we encountered in implementing UD for Mandarin are descriptive clauses. Li & Thompson (1981) describe two related constructions of the sequence [NP1] + V1 + [NP2] + V2 in which the second verb V2 (or the clause it heads) describes or comments on the second noun phrase NP2, and NP2 also serves as an object of the first verb V1. They refer to them as “realis” and “irrealis descriptive clauses” (pp. 611-620). For the *realis* kind, the clause headed by V2 describes something that is “in the here and now of the ‘real world’ ” (p. 612)—such as 我很喜歡吃 *wǒ hén xǐhuān chī* ‘I very much enjoyed eating’ in (11). For the *irrealis* kind, the V2 clause describes “an unrealized event” which is “hypothetical or projected information” (p. 612), as shown by 吃 *chī* ‘eat’ in (12).

- (11) [他] 炒了 [一個菜] (我很喜歡吃) (cf. Li & Thompson, 1981: 612)
 tā chǎo le yí gè cài wǒ hén xǐhuān chī
 he fry PERF one CLF dish I very like eat
 ‘He cooked a dish (that I very much enjoyed eating).’ *realis descriptive clause*

- (12) [我] 找 [東西] (吃)
 wǒ zhǎo dōngxǐ chī
 I seek thing eat
 ‘I’m looking for stuff (to eat).’ *irrealis descriptive clause*

The fact that the clause headed by V2 is entirely optional in (11-12) (shown in parentheses), and that NP2 is not necessarily the subject of V2 (but an object of V2 in (12) and of a verb in an embedded clausal argument of V2 in (11)), sets these descriptive clauses apart from “pivotal constructions” (as described in Chao, 1968, and Li & Thompson, 1981). Pivotal constructions, as a type of control structure (specifically, object control), obligatorily requires the V2 clause as a clausal complement of V1, and the “pivot” NP2 must also be the subject of V2. For this reason, *x_{COMP}*, the relation appropriate for pivotal constructions, is a bad fit for descriptive clauses.

One possible analysis of these descriptive clauses is that they are clausal modifiers of NP2, given that they are descriptive of NP2, always involve a predicate, and are optional. This fits the definition of the relation *ac1* (‘clausal modifier of noun’), in which case V2 would be an *ac1* dependent of NP2 (13). However, this analysis goes against the fact that Chinese is typologically regarded as a language that is strictly head-final with regards to nouns. That is, in noun phrases the noun is always last and any modifiers of that noun must precede it.

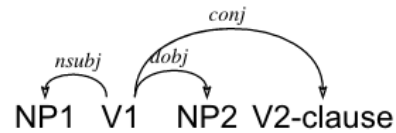
Li & Thompson suggests that the *realis* descriptive construction such as in (11) is biclausal, where V2 heads a separate clause that has a dropped argument (an object coreferential with NP2 in these cases of (11-12)). However, this structure has received very little attention in the literature (Peng, 2016), and no study known to us to date has proposed a detailed syntactic analysis which refutes or agrees with Li & Thompson’s hypothesis. Nonetheless, another option for us is to follow Li & Thompson in treating

them as coordinated clauses with the `conj` relation linking V1 and V2 (14). Simplified diagrams in (13-14) illustrate the annotation differences between our two choices of `acl` and `conj` for realis descriptive clauses.

(13) clausal modifier of noun (`acl`)



(14) coordinated clauses (`conj`)



While Peng’s (2016) comprehensive corpus study has brought *realis* descriptive clauses to the fore in recent years, much less research has been done on the *irrealis* ones. Li & Thompson themselves offer only several descriptive sentences on this structure, and the biclausal analysis for the *realis* constructions doesn’t fit the *irrealis* ones. Sentence (15) illustrates that the subject of V2 is not coreferential with any of the noun phrases within the sentence:

(15) [我] 有 [衣服] 要 洗

wǒ yǒu yīfǔ yào xǐ

I have clothes need wash

‘I have clothes that need to be washed.’

irrealis descriptive clause

In (15), the subject of the second verb is necessarily *unspecified*; whether the person who would hypothetically wash the clothes is intended to be the speaker herself or the addressee or some third party is completely arbitrary. Since subjects are only dropped in Mandarin if it can be understood pragmatically from previous context (Chao, 1968), the hypothesis that the second verb is part of an independent clause with a dropped subject is unlikely, and thus the `conj` relation would be inappropriate.

The two similar constructions above, *realis* and *irrealis* descriptive clauses, occur very frequently in Chinese, as illustrated in Peng’s (2016) study of the *realis* descriptive constructions in particular, yet we have found no explicit mention in either Stanford Chinese or CDT on how to treat these structures.

Since these two types of clauses share the function of providing additional descriptive information on NP2, we propose to keep them under the same category of relations and use `acl:realis` and `acl:irrealis`, which is preferable to splitting them apart between `conj` and `acl`. Creating the two language-specific relations for these structures will also allow for easier automatic conversion in the future for either only one of the clause types or both, should new research reveal that they should be analyzed differently.

5 Conclusion and future work

We have presented our attempt to adopt Universal Dependencies (UD) to Mandarin Chinese with consideration and reference to two other dependency schemes previously created for this language, illustrating some of the challenges and solutions we have encountered and made with regard to the morphosyntactic properties of Mandarin. Through these discussions we identified possible gaps in the current UD design, especially with regards to verbal markers and sentence particles that lie beyond the purview of adverbial modifiers and discourse markers. We also identified two common structures in Mandarin, the *realis* and *irrealis* descriptive clauses, that may have eluded analysis and explicit treatment in other Chinese treebank schemes.

We are in the process of applying our proposed annotation scheme to the Mandarin Chinese text in a Cantonese-Mandarin Parallel Corpus (Lee, 2011) of over 8000 lines of text. We plan to develop Universal Dependencies for Cantonese as well, to enable comparative studies on the grammars of the two Chinese languages. Once the treebanks for Mandarin and Cantonese are finalized, we hope to release them as part of the UD project, to be made publicly available through its website (<http://universaldependencies.org>).

Acknowledgements

This work was supported by a grant from the PROCORE-France/Hong Kong Joint Research Scheme sponsored by the Research Grants Council and the Consulate General of France in Hong Kong (Reference No.: F-CityU107/15 and N° 35322RG); and by a Strategic Research Grant (Project No. 7004494) from City University of Hong Kong.

References

- Bickel, Balthasar and Johanna Nichols. 2013. Inflectional Synthesis of the Verb. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, chapter 22, <http://wals.info/chapter/22>.
- Böhmová, Alena, Jan Hajič, Eva Hajičová, and Barbora Hladká. 2003. The Prague dependency treebank. In *Treebanks*. Springer Netherlands, pages 103-127.
- Chang, Pi-Chuan. 2009. Improving Chinese-English Machine Translation through Better Source-side Linguistic Processing. Ph.D. Dissertation, Stanford University.
- Chang, Pi-Chuan, Huihsin Tseng, Dan Jurafsky, and Christopher D. Manning. 2009. Discriminative Reordering with Chinese Grammatical Relations Features. *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation*: 51-59.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. University of California Press.
- Che, Wanxiang, Zhenghua Li, and Ting Liu. 2012. Chinese Dependency Treebank 1.0 LDC2012T05. Linguistic Data Consortium, Philadelphia. <https://catalog.ldc.upenn.edu/LDC2012T05>
- Djamouri, Redouane, Paul Waltraud, and John Whitman. 2013. Postpositions vs prepositions in Mandarin Chinese: The articulation of disharmony. In Theresa Biberauer and Michelle Sheehan, editors, *Theoretical Approaches to Disharmonic Word Order*. Oxford University Press, pages 4-105.
- Dryer, Matthew S. 2008. Position of polar question particles. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, chapter 92, <http://wals.info/chapter/92>.
- Ernst, Thomas. 1988. Chinese postpositions—Again. *Journal of Chinese Linguistics* 16(2): 219-244.
- Gerdes, Kim, and Sylvain Kahane. 2016. Dependency Annotation Choices: Assessing Theoretical and Practical Issues of Universal Dependencies. *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*: 131.
- Harbin Institute of Technology Research Center for Social Computing and Information Retrieval (哈尔滨工业大学信息检索研究中心) [HIT-SCIR]. 2010. HIT-CIR Chinese Dependency Treebank Annotation Guideline (HIT-CIR 汉语依存树库标注规范).
- Lee, John. 2011. Toward a Parallel Corpus of Spoken Cantonese and Written Chinese. *Proceedings of the 5th International Joint Conference on Natural Language Processing*: 1462-1466.
- Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. University of California Press.
- de Marneffe, Marie-Catherine, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*: 449-454.
- de Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal Stanford Dependencies: A cross-linguistic typology. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*: 4584-4592.

- Mel'čuk, Igor. 1988. *Dependency syntax: Theory and Practice*. SUNY Press.
- Mille, Simon, Vanesa Vidal, Alicia Barga, and Leo Wanner. 2009. Creating an MTT Treebank of Spanish. *Proceedings of the Fourth International Conference on Meaning-Text Theory*.
- Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016a. Universal Dependencies v1: A Multilingual Treebank Collection. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*: 1659-1666.
- Nivre, Joakim, Željko Agić, Lars Ahrenberg, et al., 2016, *Universal Dependencies 1.3*, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University in Prague, <http://hdl.handle.net/11234/1-1699>.
- Osborne, Timothy. 2015. Diagnostics for Constituents: Dependency, Constituency, and the Status of Function Words. *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*: 251-260.
- Peng, Rui. 2016. Chinese Descriptive Pivotal Construction: Taxonomy and Prototypicality. *Language and Linguistics* 17(4): 529-573.
- Peyraube, Alain. 1980. *Les constructions locatives en chinois moderne*. Editions Langages croisés, Paris.
- Sgall, Petr, Eva Hajicová, and Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Kluwer, Dordrecht.
- Tanaka, Takaaki, Yusuke Miyao, Masayuki Asahara, Sumire Uematsu, Hiroshi Kanayama, Shinsuke Mori, and Yuji Matsumoto. 2016. Universal Dependencies for Japanese. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*: 1651-1658.
- Waltraud, Paul. 2015. *New Perspectives on Chinese Syntax*. De Gruyter, Berlin.
- Xia, Fei. 2000a. The Segmentation Guidelines for the Penn Chinese Treebank (3.0). University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-00-06, http://repository.upenn.edu/ircs_reports/37/
- Xia, Fei. 2000b. The Part-of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0). University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-00-07, http://repository.upenn.edu/ircs_reports/38/
- Xue, Nianwen, Xihong Zhang, Zixin Jiang, Martha Palmer, Fei Xia, Fu-Dong Chiou, and Meiyu Chang. 2013. Chinese Treebank 8.0 LDC2013T21. Linguistic Data Consortium, Philadelphia, <https://catalog.ldc.upenn.edu/ldc2013t21>
- Zhang, Meishan, Yue Zhang, Wanxiang Che, and Ting Liu. 2014. A Semantics Oriented Grammar for Chinese Treebanking. *CICLing 2014, Part I: LNCS 8403*: 366-378.