Universal Dependencies for Japanese

Takaaki Tanaka[*], **Yusuke Miyao**[†], **Masayuki Asahara**[◇],
Sumire Uematsu[†], **Hiroshi Kanayama**[♠], **Shinsuke Mori**[♣], **Yuji Matsumoto**[‡]

[*]NTT Communication Science Labolatories, [†]National Institute of Informatics,
[◇]National Institute for Japanese Language and Linguistics,
[♠]IBM Research - Tokyo, [♣]Kyoto University, [‡]Nara Institute of Science and Technology
tanaka.takaaki@lab.ntt.co.jp, yusuke@nii.ac.jp, masayu-a@ninjal.ac.jp,
uematsu@nii.ac.jp, hkana@jp.ibm.com, forest@i.kyoto-uac.jp, matsu@is.naist.jp

Abstract

We present an attempt to port the international syntactic annotation scheme, Universal Dependencies, to the Japanese language in this paper. Since the Japanese syntactic structure is usually annotated on the basis of unique chunk-based dependencies, we first introduce word-based dependencies by using a word unit called the Short Unit Word, which usually corresponds to an entry in the lexicon UniDic. Porting is done by mapping the part-of-speech tagset in UniDic to the universal part-of-speech tagset, and converting a constituent-based treebank to a typed dependency tree. The conversion is not straightforward, and we discuss the problems that arose in the conversion and the current solutions. A treebank consisting of 10,000 sentences was built by converting the existent resources and currently released to the public.

Keywords: typed dependencies, Short Unit Word, multiword expression, UniDic

1. Introduction

The Universal Dependencies (UD) project has been developing cross-linguistically consistent treebank annotation for various languages in recent years. The goal of the project is to facilitate multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective (Nivre, 2015). The annotation scheme is based on (universal) Stanford dependencies (de Marneffe and Manning, 2008; de Marneffe et al., 2014) and Google universal part-of-speech (POS) tags (**UPOS**) (Petrov et al., 2012).

In our research, we attempt to port the UD annotation scheme to the Japanese language. The traditional annotation schemes for the Japanese language have been uniquely developed and are markedly different from other schemes, such as Penn Treebank-style annotation. Japanese syntactic parsing trees are usually represented as unlabeled dependency structures between bunsetsu chunks (base phrase units), as found in the Kyoto University Text Corpus (Kurohashi and Nagao, 2003) and the outputs of syntactic parsers (Kudo and Matsumoto, 2002; Kawahara and Kurohashi, 2006). Therefore, we must devise a method to construct word-based dependency structures that match the characteristics of the Japanese language (Uchimoto and Den, 2008; Mori et al., 2014; Tanaka and Nagata, 2015) and are able to derive the syntactic information required to assign relation types to dependencies.

We describe the conversion from the Japanese POS tagset to the UPOS tagset, the adaptation of the UD annotation for Japanese syntax, and the attempt to build a UD corpus by converting the existing resources. We also address the remaining issues that may emerge when applying the UD scheme to other languages.

2. Word unit

The definition of a word unit is indispensable in UD annotation, which is not a trivial question for Japanese, since a sentence is not segmented into words or morphemes by white space in its orthography. Thus, we have several word unit standards that can be found in corpus annotation schemata or in the outputs of morphological analyzers (Kudo et al., 2004; Neubig et al., 2011).

NINJAL[1] proposed several word unit standards for Japanese corpus linguistics, such as the minimum word unit (Maekawa et al., 2000). Since 2002, the Institute has maintained a morphological information annotated lexicon, UniDic (Den et al., 2008), and has proposed three types of word unit standards:

Short Unit Word (SUW): SUW is a minimal language unit that has a morphological function. SUW almost always corresponds to an entry in traditional Japanese dictionaries.

Middle Unit Word (MUW): MUW is based on the right-branching compound word construction and on phonological constructions, such as an accent phrase and/or sequential voicing.

Long Unit Word (LUW): LUW refers to the composition of bunsetsu units. An LUW has nearly the same content as functional words bounded by bunsetsu boundaries.

We adopted SUW from the two types of word units, SUW and LUW, used for building the Balanced Corpus of Contemporary Written Japanese (BCCWJ) (Maekawa et al., 2014). SUWs correspond to an entry conveying morphological information in the UniDic. In this way, the UD

---

[1]The National Institute for Japanese Language and Linguistics

		魚フライを食べたかもしれないペルシャ猫 *"the Persian cat that may have eaten fried fish"*									
SUW	魚 NOUN *fish*	フライ NOUN *fry*	を ADP -ACC	食べ VERB *eat*	た AUX -PAST	か PART	も ADP	しれ VERB *know*	ない AUX -NEG	ペルシャ PROPN *Persia*	猫 NOUN *cat*
LUW	魚フライ NOUN *fried fish*		を ADP -ACC	食べ VERB *eat*	た AUX -PAST	かもしれない AUX *may*				ペルシャ猫 NOUN *Persian cat*	

Figure 1: Japanese word unit: Short unit word (SUW) and Long unit word (LUW).
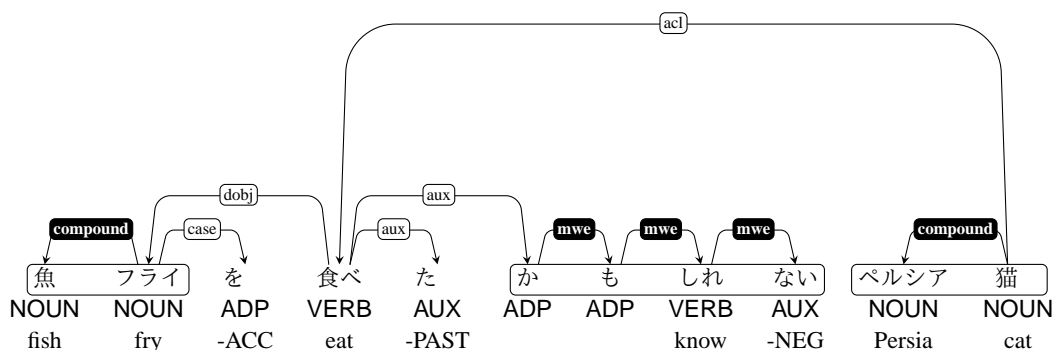

Figure 2: Dependencies for compounding SUWs into LUWs.

scheme for the Japanese language was based on the lexemes and the POS tagset defined in the UniDic. This was done because the UniDic guidelines are fully established and widely used in Japanese NLP. The UniDic has been maintained diachronically, and NINJAL has published versions of UniDic for several eras.

SUWs are sometimes too short for assigning the syntactic relation types for a pair of SUWs, while LUWs are more suitable for representing syntactic structure and function. Figure 1 is an example of Japanese unit word segmentation with SUWs and LUWs. The sentence contains some multiword expressions (multi SUWs) composing a single LUW, such as a compound noun and function phrase. For a compound noun LUW, e.g., 魚/フライ "fried fish", the internal relation is tagged with compound. An LUW that behaves like a single function word is annotated by a flat and head-initial structure. LUW internal dependencies are tagged with mwe in conformity to the UD scheme. In the example, the phrase か/も/しれ/ない *kamosirenai*, functioning like a single auxiliary verb, is annotated by a flat structure using mwe relations.

3. Part-of-speech annotation

UD employs UPOS tags, based on the Google universal tagset. The Japanese POSs are defined as a mapping from **UniDic POS** tags to UPOS tags. The POSs in Japanese corpora can be understood in two ways: lexicon-based and usage-based approaches. The lexicon-based approach involves extracting all possible categories for one word as labels. For example, the UniDic POS tag `noun(common.verbal.adjectival)` [2] means that a word can be a common noun, verbal noun, or adjective.

Usage-based labeling is determined by the contextual information in sentence.

We assume that the UPOS tagset is usage-based, though this is not clearly defined, and map the UniDic POS tagset to the UPOS tagset by disambiguating the lexicon-based UniDic POS using available context. For example, we must determine whether nominal verbs are tagged with VERB or NOUN depending on the context, as described in the rest of this section.

Table 1 shows a mapping from UniDic POS tags to Universal POS tags. The mapping between two tagsets is not a one-to-one correspondence, and thus, conversion is not straightforward. Issues that arise during the actual mapping for individual POS tags are described below.

Particle In traditional Japanese grammar, particles are classified into several subcategories; UniDic has six particle subcategories. Some particles can be mapped to UPOS tags using the subcategories, while some are split and assigned different UPOS tags. Case particles (`particle(case)`) and binding particles (`particle(binding)`) correspond to ADP when they are attached to a noun phrase as case markers. Note that the case particle と (`particle(case)`) [3] is tagged with CONJ when it marks a conjunct in a coordinating conjunction. Phrase final postpositional particles (`particle(phrase_final)`) are classified into PART. Conjunctive particles (`particle(conjunctive)`), which introduce subordinate clauses, and nominal particles (`particle(nominal)`), which introduce noun phrases as complementizers, are mapped to SCONJ.

---

[2]Typewriter font is used for UniDic POS tags in this paper.

[3]Since the UniDic POS tagset does not have a tag for the coordinating conjunctive particle, these usages of と cannot be distinguished only by POS.

Adnominal There is a small group of adnominal words tagged with `adnominal` in UniDic, which are similar to attributive adjectives but not conjugative words. Some words in the class correspond to demonstrative and possessive pronouns, e.g., あの *ano* "that" and どの *dono* "which," and are classified as determiners DET, while others are tagged ADJ, e.g., 同じ *onaji* "same" and 大きな *ookina* "big." This is because, in other language, the former functions as a determiner. However, the Japanese language does not have articles and the traditional Japanese grammar does not have the determiner word class.

Auxiliary verb In some cases it is difficult to distinguish between main and auxiliary verbs.

- 走っ た *hashit ta* "ran"
- 走って いる *hashit te iru* "running"
- 走って ほしい *hashit te hoshii* "want (you) to run"
- 走り 始める *hashiri hajimeru* "begin to run"

In our first example, we clearly have an auxiliary verb, because た does not appear independently. The other cases, however, are unclear, because verbs like いる, ほしい and 始める can also be used as main verbs. In the above examples, the usual meanings of these verbs are replaced (similar to a light verb) and auxiliary meanings are added to the preceding verbs.

These verbs are defined as 非自立 verb(bound) in UniDic, and we define this type of verb preceded by another verb as an auxiliary verb. If these verbs appear independently, they are regarded as the main verb.

Nominal verb and nominal adjective Words in this category noun(common.verbal_suru) are basically nouns and function as verbs when followed by an auxiliary verb, e.g., する *suru* "do." The stems of nominal verbs, e.g., 報告 *houkoku* "report," are tagged with VERB as heads when they are used as verb (1). They are still tagged with NOUN when used as nouns (2).

The noun(common.adjectival) words are similarly tagged with NOUN or ADJ depending on context. That is, the stems of the nominal adjectives, e.g., 自由 *jiyuu* "free," are tagged ADJ as heads when used as adjectives, and tagged NOUN when used as nouns, as shown in (3) and (4).
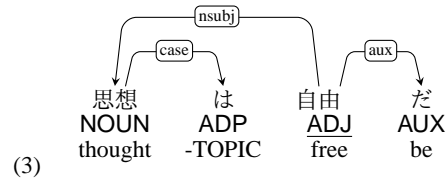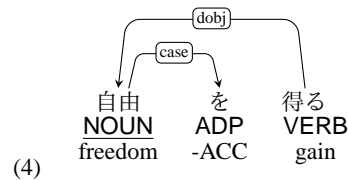ADJ as
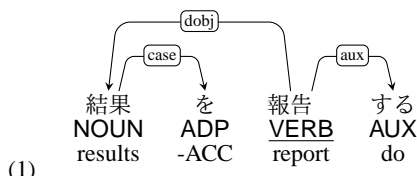
(1) '(Someone) reports the results.'
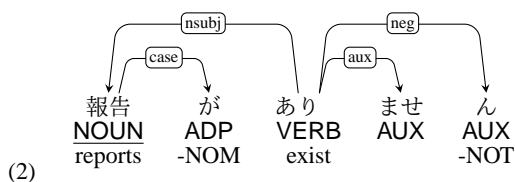
(2) 'There is no report.'

(3) 'Thought is free.'

(4) '(Someone) gains freedom.'

Suffix UniDic has the `suffix` (and `prefix`) categories as independent word classes. A typical suffix adds meaning to the preceding noun and forms a new noun phrase. For instance, the suffix 達 *tachi* is a type of plural suffix and makes a plural form as in 学生/達 "students". These are tagged with NOUN.

Another type of suffix alters the POS of the preceding word. For example, the suffix さ *sa* changes an adjective into a noun as in (5), and っぽい *ppoi* changes a noun into an adjective as in (6). These words are tagged with PART and are dependent on the preceding words. Unlike nominal verbs and nominal adjectives, the POS of the preceding word remains the same as the original one in the annotation. This is because the preceding words do not have the syntactic properties of the altered POS without suffixes, while nominal verbs and nominal adjectives have the properties of verbs and adjectives by themselves. This kind of construction is very generative, and it is considered to be morphological (a word formation), rather than a syntactic relation.

(5) 'Cuteness'
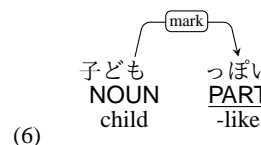
(6) 'Child-like'

4. Syntax annotation

Syntactic dependency types in Japanese are defined in order to be as in conformance with the principles of UD as possible. However, the definition of Japanese syntax under UD involves several issues that should be discussed. For example, the definition of "clause" is not clear. Dependency types rely on the definition of a "clause," such as the distinction between nsubj and csubj. Thus, we need to define a clause from the viewpoint of UD annotation. The dependency types expl and xcomp are not used since no corresponding Japanese constructions exist.

1653

Universal POS	UniDic POS	
ADJ	`adjective_i`	赤い "red"
	`adnominal`	いろんな "various"
	`noun(adjectival)`	自由/だ "free"
ADP	`particle(case)`	が (nominative case), を (accusative case)
	`particle(binding)`	は (binding)
ADV	`adverb`	とても "very", 必ず "always"
	`noun(adverbial)`	完全/に "absolutely"
AUX	`auxiliary_verb`	食べ/た "ate" (past tense)
	`verb(bound)`	食べ/て/いる "be eating" (progressive)
	`verb(bound)`	勉強/する "study" (verbification)
	`adjective_i`	食べ/ない "*not* study" (negation)
CONJ	`particle(case)`	コーヒー/と/牛乳 "coffee *and* milk"
	`particle(adverbial)`	コーヒー/か/牛乳 "coffee *or* milk"
	`conjunction`	しかし "but"
DET	`adnominal`	この "this", その "that" (demonstrative)
INTJ	`interjection`	ああ "oh", えっと "well"
NOUN	`noun`	猫 "cat", 質問 "question"
	`prefix`	副/社長 "*vice* president"
	`suffix`	学生/達 "student*s*" (plural), 付属/品 "accessories (*lit.*) supplementary *parts*"
NUM	`noun(numeral)`	十 "ten"
PART	`particle(phrase_final)`	良い/ね "good, *isn't it*"
	`suffix(adjectival_noun)`	衝撃/的/だ "(something) is shocking"
PRON	`pronoun`	私 "I", 彼女 "she", いつ "when"
PROPN	`noun(proper.*.*)`	京都 "Kyoto", 鈴木 "Suzuki"
PUNCT	`supplementary_symbol`	. (period), 「」 (parentheses)
SCONJ	`particle(conjunctive)`	食べ/て/寝る "eat, *then* sleep"
	`particle(nominal)`	食べる/の/が/好き "(I) like *to* eat" (nominal particle)
SYM	`supplementary_symbol`	+. −, <, >
VERB	`verb`	遊ぶ "play"
	`noun(common.verbal_suru)`	勉強/する "study"
X	`whitespace`	white space

Table 1: Mapping from UniDic POS to Universal POS. The symbols '/' denote the borders of SUWs.

In the following paragraphs, we show the annotation scheme for some basic constructions.

Core dependent of the predicate Core dependents of the predicate can be either subject, direct object or indirect object in the context of UD. It is difficult to strictly define verb valency in Japanese; generally, a postpositional phrase that is a dependent of the predicate is assigned with subject, direct object or indirect object depending on its case particle, が *ga*, を *o* or に *ni*. Even if the case particle is hidden by a topical marker は *wa* in a topicalized phrase and the adverbial particle も *mo*, the relation is assigned on the basis of the original particle.

In sentence (7), the relationship between the predicate 与える "give" *ataeru* and the postpositional phrases 私/が "I"-NOM, 彼/に "he"-DAT, and 本/を "book"-ACC are tagged nsubj, dobj, and iobj, respectively. Even if the phrase 本/を "book"-ACC is topicalized, i.e., 私が彼に本 は 与えた. "As for the book, I gave it to him. (I may not give anything else.)," the dependency between the argument 本 and the verb 与える is still tagged dobj.
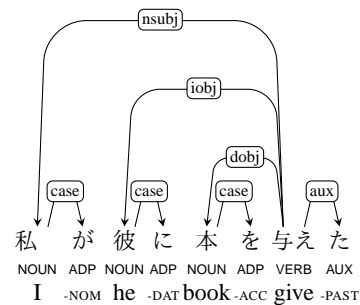
(7)

'I gave him a book'

Nominal subject and clausal subject Concerning the distinction between nominal subject nsubj and clausal subject csubj, we have the following gradation.

- 食べる の が 大切 だ "Eating is important"
- 食べる こと が 大切 だ "Eating is important"
- 食べる ところ が 大切 だ "The place where (we) eat is important"

The first one is the csubj case because の *no* is a complementizer (SCONJ), which does not appear as a content word independently. However, the following examples are unclear cases.

1654

relation	definition	typical construction in Japanese
nsubj	nominal subject	predicate → a postpositional phrase with the case marker が
dobj	direct object	predicate → a postpositional phrase with the case marker を
iobj	indirect object	predicate → a postpositional phrase with the case marker に
csubj	clausal subject	predicate → a postpositional phrase with the case marker が
dislocated	dislocated elements	predicate → a postpositional phrase with the topic marker は
nmod	nominal modifier	noun phrase with the genitive marker の → noun phrase
amod	adjective modifier	noun phrase → ADJ
advmod	adverbial modifier	predicate → ADV
case	case marking	noun phrase → ADP
mark	marker	subordinating clause → complementizer と
aux	auxiliary	predicate → AUX
cop	copula	noun phrase →the copular auxiliary だ
neg	negation	predicate →the negative auxiliary ない
auxpass	passive auxiliary	predicate → the passive auxiliaries れる / られる
acl	clausal modifier of noun	head noun → relative clause
advcl	adverbial clause modifier	head verb → adverbial clause
ccomp	clausal complement	predicate of main clause → complement clause
xcomp	open clausal complement	not used
expl	expletive	not used

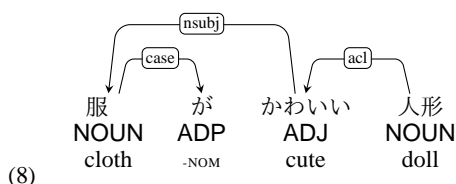Table 2: Mapping to syntactic annotation in UD

Formal nouns こと *koto* "fact" and ところ *tokoro* "place" (NOUN) can have clausal complements and form noun phrases denoting the action expressed by the clause. This occurs when the expressions are used as nouns having content, but in these examples these words have light meanings.

In the current definition, we define only the first case, i.e., a phrase introduced by の, as a clausal subject, while the other cases are regarded as noun phrases. Here, the second example has almost the same meaning as the first.
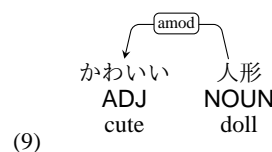
Adnominal clause and adjective There are two types of noun modifying clauses: the clausal complement of a noun and a relative clause. In the UD scheme, these two types are not distinguished and are tagged acl [4].

The dependency between formal noun こと *koto* "fact" and the clausal complement is termed the clausal modification of a noun. 食べる *taberu* "eat" こと means eating (or the fact that someone eats) in the example above, and the relation between them is also tagged acl.

For a relative clause, the head of the dependency is the noun modified by the clause; the dependent is the main predicate of the clause as shown in (8). However, it is difficult to distinguish between a clause and a non-clause because there is no difference between a simple adjective-noun construction as in (9) and a relative clause construction. This is because relative clauses are not accompanied by a relativizer. Our current solution is to define an adjective without any arguments, e.g., nsubj, and auxiliary verbs, e.g., た *ta*, as a non-clause amod or otherwise as a clause acl.
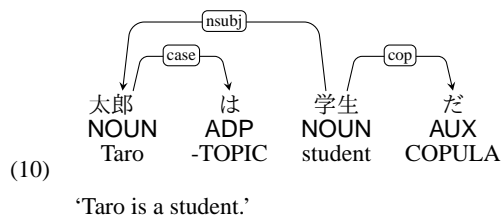
(8)
服 NOUN cloth / が ADP -NOM / かわいい ADJ cute / 人形 NOUN doll
(nsubj, case, acl)

----

[4]In UD for English, the relative clause is currently classified into a subclass acl:relcl.

'a doll whose clothes are cute'

(9)
かわいい ADJ cute / 人形 NOUN doll
(amod)

'a cute doll'

Copula The dependency type cop is reserved for the copular auxiliary だ *da*. This auxiliary typically follows a noun phrase to form a copular clause. A postpositional phrase with a nominative case is commonly needed to complete sentence as in (10). Note that we treat the auxiliary だ after nominal adjectives noun(common.adjectival) as an auxiliary instead of a copula as shown in (3).

(10)
太郎 NOUN Taro / は ADP -TOPIC / 学生 NOUN student / だ AUX COPULA
(nsubj, case, cop)

'Taro is a student.'

5. Corpus

It is reasonable to obtain Japanese UD corpora by converting existent linguistic resources; however, a direct conversion from the major Japanese corpora such as the Kyoto University Text Corpus (Kurohashi and Nagao, 2003) is not simple since they lack syntactic information (unlabeled) and the structure is not suitable to recover constituents (bunsetsu chunk-based dependency trees). Therefore, we first constructed conversion rules for use with Japanese constituent treebank (Tanaka and Nagata, 2013)
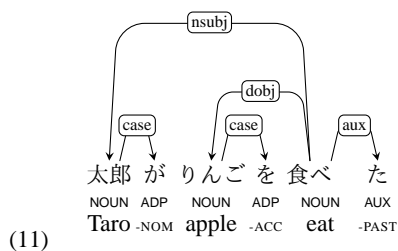
for the Mainichi Shimbun Newspaper. The treebank was initially built by converting the Kyoto University Text Corpus and was manually annotated. The treebank has clause level annotations with syntactic function labels, e.g., syntactic role and clause type, and coordination construction, which are required for UD annotation. The treebank is composed of complete binary trees, and can be easily converted to dependency tree by adapting the head percolation rules and dependency type rules for each partial tree. The UD corpus is composed of 10,000 sentences, and it contains 267,631 tokens. The data is available on the UD website [5]. Moreover, we attempted to construct conversion rules for BCCWJ with third-party annotations in order to build UD resources covering a wide variety of genres including books, magazines, blogs, etc.
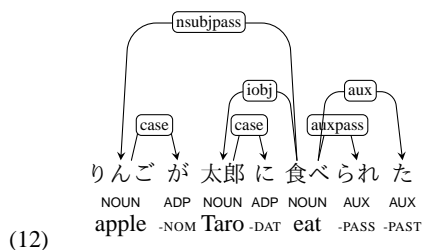
6. Discussion

Several issues remain before the Japanese syntactic structure can be fully covered in the UD scheme.

Voice In UD, the passive voice is marked with special dependency types, such as nsubjpass and auxpass. This is useful in recognizing semantic dependencies. Sentence (12) is the passivized sentence of sentence (11). The subject of (11), which is an agent of 食べる "eat," is distinguished from the subject of (12).

(11)

nsubj / dobj / case / case / aux
太郎 が りんご を 食べ た
NOUN ADP NOUN ADP NOUN AUX
Taro -NOM apple -ACC eat -PAST

'Taro ate an apple. (active)'

(12)

nsubjpass / iobj / aux / case / case / auxpass
りんご が 太郎 に 食べ られ た
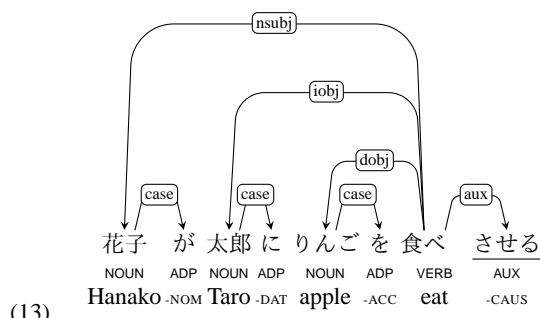NOUN ADP NOUN ADP NOUN AUX AUX
apple -NOM Taro -DAT eat -PASS -PAST
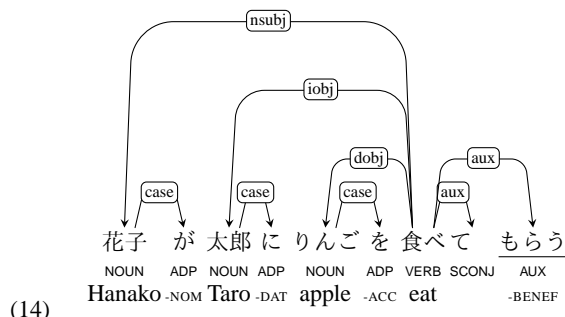
'An apple was eaten by Taro. (passive)'

Japanese syntax involves another voice that involves case alternations done by adding specific auxiliary verbs to predicates similar to passive voice construction, as shown in (13) and (14). The problem here is that the relationship between the verb and its subject in this voice is not distinguished from the relationship indicated by the original active voice [6]. The subject of the verb phrases 食べ/さ/

---

[6] Voice information can be included as a morphological feature; however it is not clearly represented as a syntactic structure.

せる "eat +CAUS" or 食べ/て/もらう "eat +BENEF", 花子 "Hanako", is an additional argument, which is a causer in sentence (13) and a benefactor in sentence (14). It is normally annotated with nsubj, following the current UD scheme, while the subject in sentence (12), りんご "apple", which is the patient of the verb, is annotated with the special relation nsubjpass. In addition, the proto-agent of the verb 食べ "eat", 太郎 "Taro", is tagged with iobj because it is marked with the case marker に "ni" in these constructions.

We do not have a method to indicate these case alternations in the current UD. Currently, we give dependency types on the basis of surface expressions, without any markings of case alternations.

(13)

nsubj / iobj / dobj / case / case / case / aux
花子 が 太郎 に りんご を 食べ させる
NOUN ADP NOUN ADP NOUN ADP VERB AUX
Hanako -NOM Taro -DAT apple -ACC eat -CAUS

'Hanako makes Taro eat an apple. (causative)'

(14)

nsubj / iobj / dobj / aux / case / case / case / aux
花子 が 太郎 に りんご を 食べて もらう
NOUN ADP NOUN ADP NOUN ADP VERB SCONJ AUX
Hanako -NOM Taro -DAT apple -ACC eat -BENEF

'Hanako asks Taro to eat an apple. (benefactive)'

Coordination We take the first conjunct as the head in the coordinating construction in the fashion of the UD scheme. However, because Japanese is a head final language, the last conjunct tends to be the head. Using first conjunct head construction for a head final language creates issues because different constructions share the one annotation. For example, the annotation (15) has two possible interpretations. かわいい犬と猫 "cute dogs and cats" could be interpreted as having the adjective modifying the first conjunct 犬 "dog" or the adjective modifying the whole construction 犬/と/猫 "dog and cat". It would be preferable to adapt a scheme for choosing the head of a coordinating construction depending on the properties of the target language.

1656

(15)
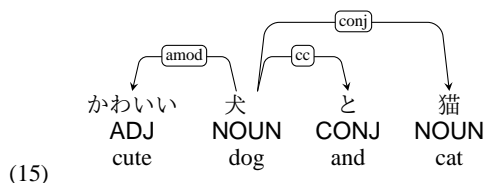
'cute dogs and cats'

Topic phrase The dependency type dislocated is assigned for topic phrases. A topic phrase introduces the topic of a sentence, and is typically a prepositional phrase with a topic marker は *wa*. One of the most famous examples is 象 は 鼻 が 長い 。 *zou wa hana ga nagai* "For elephants, trunks are long." [7]
However, this type is also used for fronted or postposed elements that do not fulfill the usual core grammatical relations; for example, the relation between "office" and "me" in the sentence "This is our office, me and Sam." It is possible to argue that these constructions share the same dependency type.

(16)
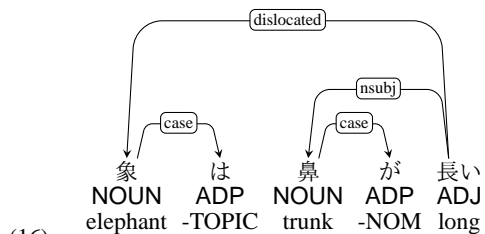
'For elephants, trunks are long. '

7. Conclusion

We have presented an attempt to apply the UD annotation scheme on Japanese language annotation and build a UD corpus. Porting the UD scheme to Japanese is not straightforward. We have enumerated the issues related to morphological and syntactic phenomena in Japanese and shown our current solutions. We believe that the remaining issues, including voice and coordination provide hints towards a better UD scheme in terms of the commonization of phenomena in various languages.

The first draft of the guidelines for UD Japanese was released on June 1st, 2015, and the first treebank is available through the UD website.

8. Bibliographical References

de Marneffe, M.-C. and Manning, C. D. (2008). The stanford typed dependencies representation. In *Proceedings of COLING 2008 Workshop on Cross-framework and Cross-domain Parser Evaluation*.

de Marneffe, M.-C., Silveira, N., Dozat, T., Haverinen, K., Ginter, F., Nivre, J., and Manning, C. D. (2014). Universal stanford dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, LREC 2014, pages 4584–4592.

Den, Y., Nakamura, J., Ogiso, T., and Ogura, H. (2008). A proper approach to Japanese morphological analysis: Dictionary, model and evaluation. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, LREC 2008, pages 1019–1024.

Kawahara, D. and Kurohashi, S. (2006). A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, HLT-NAACL 2006, pages 176–183.

Kudo, T. and Matsumoto, Y. (2002). Japanese dependency analysis using cascaded chunking. In *Proceedings of the 6th Conference on Natural Language Learning*, volume 20 of *CoNLL 2002*, pages 1–7.

Kudo, T., Yamamoto, K., and Matsumoto, Y. (2004). Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP 2004, pages 230–237.

Kurohashi, S. and Nagao, M., (2003). *Building a Japanese Parsed Corpus – while Improving the Parsing System*, chapter 14, pages 249–260. Kluwer Academic Publishers.

Maekawa, K., Koiso, H., Fukui, S., and Isahara, H. (2000). Spontaneous speech corpus of Japanese. In *Proceedings of the Second International Conference on Language Resources and Evaluation*, LREC 2000, pages 947–952.

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., and Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, 48(2):345–371.

Mori, S., Ogura, H., and Sasada, T. (2014). A Japanese word dependency corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, LREC 2014, pages 753–758.

Neubig, G., Nakata, Y., and Mori, S. (2011). Pointwise prediction for robust, adaptable Japanese morphological analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, ACL HLT 2011, pages 529–533.

Nivre, J. (2015). Towards a universal grammar for natural language processing. In *Proceedings of the 16th International Conference of Computational Linguistics and Intelligent Text Processing*, CICLing 2015, pages 3–16.

Petrov, S., Das, D., and McDonald, R. (2012). A universal part-of-speech tagset. In *Proceedings of the International Conference on Language Resources and Evaluation*, LREC 2012, pages 2089–2096.

Tanaka, T. and Nagata, M. (2013). Constructing a practical constituent parser from a Japanese treebank with function labels. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, SPMRL 2013, pages 108–118.

Tanaka, T. and Nagata, M. (2015). Word-based Japanese

---

[7]Note that the relation dislocated is not used as a topicalized phrase and also a core argument of a predicate, as described in Section 4.

typed dependency parsing with grammatical function analysis. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, volume 2 of *ACL 2015*, pages 237–242.

Uchimoto, K. and Den, Y. (2008). Word-level dependency-structure annotation to corpus of spontaneous Japanese and its application. In *Proceedings of the International Conference on Language Resources and Evaluation*, LREC 2008, pages 3118–3122.