Universal Dependencies for Japanese

Takaaki Tanaka*, Yusuke Miyao†, Masayuki Asahara‡, Sumire Uematsu§, Hiroshi Kanayama¶, Shinsuke Mori'||, Yuji Matsumoto‡

*NTT Communication Science Laboratories, †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-u.ac.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).

NINJAL1 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

1The National Institute for Japanese Language and Linguistics
scheme for the Japanese language was based on the lex-
emes 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 ver-
sions 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 segmenta-
tion 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
language. 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 in-
formation 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 Uni-
versal 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 coor-
dinating 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 co-
ordering conjunctive particle, these usages of と cannot be dis-
tinguished 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 こと “that” and と の こと “which,” and are classified as determiners DET, while others are tagged ADJ, e.g., 同じ こと same and 大きな こと 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 だ 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 sara “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).

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.

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. Dependance 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.
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” ataru 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 本を 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.

<table>
<thead>
<tr>
<th>Universal POS</th>
<th>UniDic POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>adjective_i</td>
</tr>
<tr>
<td></td>
<td>adnominal</td>
</tr>
<tr>
<td></td>
<td>noun(ajj)</td>
</tr>
<tr>
<td>ADP</td>
<td>particle(case)</td>
</tr>
<tr>
<td></td>
<td>particle(binding)</td>
</tr>
<tr>
<td>ADV</td>
<td>adverb</td>
</tr>
<tr>
<td></td>
<td>noun(adv)</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary verb</td>
</tr>
<tr>
<td></td>
<td>verb</td>
</tr>
<tr>
<td></td>
<td>verb</td>
</tr>
<tr>
<td></td>
<td>adjective_i</td>
</tr>
<tr>
<td>CONJ</td>
<td>particle(case)</td>
</tr>
<tr>
<td></td>
<td>particle(adv)</td>
</tr>
<tr>
<td></td>
<td>conjunction</td>
</tr>
<tr>
<td>DET</td>
<td>adnominal</td>
</tr>
<tr>
<td>INTJ</td>
<td>interjection</td>
</tr>
<tr>
<td>NOUN</td>
<td>noun</td>
</tr>
<tr>
<td></td>
<td>prefix</td>
</tr>
<tr>
<td></td>
<td>suffix</td>
</tr>
<tr>
<td>NUM</td>
<td>noun(num)</td>
</tr>
<tr>
<td>PART</td>
<td>particle</td>
</tr>
<tr>
<td></td>
<td>suffix</td>
</tr>
<tr>
<td>PRON</td>
<td>pronoun</td>
</tr>
<tr>
<td>PROPN</td>
<td>noun(proper)</td>
</tr>
<tr>
<td>PUNCT</td>
<td>supplementary</td>
</tr>
<tr>
<td>SCONJ</td>
<td>particle</td>
</tr>
<tr>
<td></td>
<td>particle</td>
</tr>
<tr>
<td>SYM</td>
<td>verb</td>
</tr>
<tr>
<td>VERB</td>
<td>verb</td>
</tr>
<tr>
<td>X</td>
<td>whitespace</td>
</tr>
</tbody>
</table>

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

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 (CONJ), which does not appear as a content word independently. However, the following examples are unclear cases.
<table>
<thead>
<tr>
<th>relation</th>
<th>definition</th>
<th>typical construction in Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj</td>
<td>nominal subject</td>
<td>predicate → a postpositional phrase with the case marker が</td>
</tr>
<tr>
<td>dobj</td>
<td>direct object</td>
<td>predicate → a postpositional phrase with the case marker を</td>
</tr>
<tr>
<td>iobj</td>
<td>indirect object</td>
<td>predicate → a postpositional phrase with the case marker を</td>
</tr>
<tr>
<td>csubj</td>
<td>clausal subject</td>
<td>predicate → a postpositional phrase with the case marker に</td>
</tr>
<tr>
<td>dislocated</td>
<td>dislocated elements</td>
<td>predicate → a postpositional phrase with the topic marker は</td>
</tr>
<tr>
<td>nmod</td>
<td>nominal modifier</td>
<td>noun phrase with the genitive marker の → noun phrase</td>
</tr>
<tr>
<td>amod</td>
<td>adjective modifier</td>
<td>noun phrase → ADJ</td>
</tr>
<tr>
<td>advmod</td>
<td>adverbial modifier</td>
<td>predicate → ADV</td>
</tr>
<tr>
<td>case</td>
<td>case marking</td>
<td>noun phrase → ADP</td>
</tr>
<tr>
<td>mark</td>
<td>marker</td>
<td>subordinating clause → complementizer と</td>
</tr>
<tr>
<td>aux</td>
<td>auxiliary</td>
<td>predicate → AUX</td>
</tr>
<tr>
<td>cop</td>
<td>copula</td>
<td>noun phrase → the copular auxiliary だ</td>
</tr>
<tr>
<td>neg</td>
<td>negation</td>
<td>predicate → the negative auxiliary ない</td>
</tr>
<tr>
<td>auxpass</td>
<td>passive auxiliary</td>
<td>predicate → the passive auxiliaries れる / るるる</td>
</tr>
<tr>
<td>acl</td>
<td>clausal modifier of noun</td>
<td>head noun → relative clause</td>
</tr>
<tr>
<td>advcl</td>
<td>adverbial clause modifier</td>
<td>head verb → adverbial clause</td>
</tr>
<tr>
<td>ccomp</td>
<td>clausal complement</td>
<td>predicate of main clause → complement clause</td>
</tr>
<tr>
<td>xcomp</td>
<td>open clausal complement</td>
<td>not used</td>
</tr>
<tr>
<td>expl</td>
<td>expletive</td>
<td>not used</td>
</tr>
</tbody>
</table>

(a) doll whose clothes are cute'  
kaibei ADJ NOUN  
ADJ cute  
NOUN doll  

(9)  
'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)  
'Taro is a student.'

5. Corpus

It is reasonable to obtain Japanese UD corpora by converting existing 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 "eat," is distinguished from the subject of (12).

```
(11)
'Taro ate an apple. (active)'

Taro NOUN
apple NOUN
eat VERB

(12)
'An apple was eaten by Taro. (passive)'

りんご NOUN
case ADP
taro NOUN
eat VERB

(13)
'Hanako makes Taro eat an apple. (causative)'

Hanako NOUN
taro NOUN
apple NOUN
eat VERB
aux AUX

(14)
'Hanako asks Taro to eat an apple. (benefactive)'

Hanako NOUN
taro NOUN
apple NOUN
eat VERB
bene AUX
```

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 "eat"/"食べ"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べる"/"食べた


\(^6\)Voice information can be included as a morphological feature; however it is not clearly represented as a syntactic structure.

**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.
Section 4.

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.”

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.

For elephants, trunks are long.

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.

7

8.

Bibliographical References


Parallel Universal Dependencies. Contribute to UniversalDependencies/UD_Japanese-PUD development by creating an account on GitHub. GitHub is where the world builds software. Millions of developers and companies build, ship, and maintain their software on GitHub â€“ the largest and most advanced development platform in the world. Sign up for free. Dismiss. master. 2018. Universal Dependencies Version 2 for Japanese. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pages 1824-1831, Miyazaki, Japan. BCCWJ-DepPara: A Syntactic Annotation Treebank on the 'Balanced Corpus of Contemporary Written Japanese.Â 2016. Universal Dependencies for Japanese. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), pages 1651-1658. Constructing a practical constituent parser from a Japanese treebank with function labels. â€™ IBM Research - Tokyo, â€™Kyoto University, â€¦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-u.ac.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 rst introduce word-based dependencies by using a word unit called the Short Unit Word, which usually cor...