

# A Dependency Treebank for Telugu

**Taraka Rama**

Department of Informatics  
University of Oslo, Norway  
tarakark@ifi.uio.no

**Sowmya Vajjala**

Applied Linguistics and Technology Program  
Iowa State University, USA  
sowmya@iastate.edu

## Abstract

In this paper, we describe the annotation and development of Telugu treebank following the Universal Dependencies framework. We manually annotated 1328 sentences from a Telugu grammar textbook and the treebank is freely available from Universal Dependencies version 2.1.<sup>1</sup> In this paper, we discuss some language specific annotation issues and decisions; and report preliminary experiments with POS tagging and dependency parsing. To the best of our knowledge, this is the first freely accessible and open dependency treebank for Telugu.

## 1 Introduction

An annotated treebank is a pre-requisite for developing computational tools that support deeper language processing for any language. Treebanks are typically created with texts collected from specific genre such as news, fiction, Wikipedia, blogs, and Bible. There also exist treebanks for non-canonical text such as learner data (Berzak et al., 2016; Lee et al., 2017). While these treebanks have been used for the development of natural language parsers and other tools, they may not cover infrequent grammatical structures that do not occur in the specific domain of the training data. Grammar books provide an excellent set of examples for annotated sentences that cover a wide range of syntactic structures as these sentences are chosen to illustrate the interesting and unique features in a language. Additionally, such grammar book based treebanks can also be used to test the coverage of statistical parsers trained with a large amount of data from a specific domain or genre. Further, a typical grammar book features short sentences and allows rapid development of a treebank. Hence, they serve as a good starting point for developing a broad coverage treebank (Cöltekin, 2015).

Telugu is a Dravidian Language native to India with 74 million native speakers with a long history of written and oral literature. Despite some published research on development of part-of-speech taggers (PVS and Karthik, 2007) and a treebank (Vempaty et al., 2010), neither of the resources are publicly available. In this paper, we describe our efforts in developing a publicly available treebank for Telugu that covers a range of syntactic constructions and morphological phenomena. We manually annotated 1328 sentences from the Telugu grammar book by Krishnamurti and Gwynn (1985) with (universal) part-of-speech tags and dependency relations. We followed the Universal Dependencies (Nivre et al., 2016) framework for annotation, as it supports the development of treebanks for new languages through extensive documentation. We also report preliminary POS tagging and dependency parsing results using the treebank data and UDPipe (Straka et al., 2016).

The rest of the paper is organized as follows. We describe the related work in section 2. We provide a brief description of linguistic properties of Telugu in section 3. Then, we describe the corpus and annotation environment in section 4. We describe the annotation decisions during the annotation of POS tags in section 5. Section 6 briefly introduces Telugu morphology and is followed by section 7 that discusses dependency relations that are specific to Telugu. We present the results of our POS tagging and parsing experiments in section 8. Finally, we conclude the paper and discuss some directions for future work.

<sup>1</sup><http://universaldependencies.org/treebanks/te/>

## 2 Related Work

Treebanks for some South Asian languages were developed following the Paninian framework for dependency annotation (Begum et al., 2008) and some of them are publicly available.<sup>2</sup> There were some early efforts towards Telugu dependency treebank development following the Paninian framework (Vempaty et al., 2010). A Telugu treebank was also a part of an Indian language dependency tools contest (Husain et al., 2010). However, none of these resources are publicly available to the best of our knowledge. There have been efforts to convert some of the Indian language treebanks into Universal Dependencies (UD) framework (Tandon et al., 2016) and there is a reasonably large UD treebank for Hindi (Palmer et al., 2009; Bhat et al., 2017). However, except for Tamil (Ramasamy and Žabokrtský, 2012), converted from Prague dependency style (Hajič et al., 2017) to UD, there is no UD treebank for any other language from the Dravidian language family. In this scenario, the availability of free and open UD Telugu treebank would be a good starting point for the future of computational infrastructure support for Telugu and the Dravidian language family.

UD treebanks were developed from scratch for several low-resource languages such as Kazakh (Makazhanov et al.) and Buryat (Badmaeva and Tyers, 2017) in the recent years. Several existing treebanks are also being converted into UD. While there is no Dravidian language in UD other than Tamil, there exist treebanks for other agglutinative languages such as Finnish (Pyysalo et al., 2015), Hungarian (Vincze et al., 2017), Turkish (Cöltekin, 2015; Sulubacak et al., 2016), and Estonian (Muischnek et al., 2016) which provided us with useful insights in dealing with language-specific morphological and syntactic phenomena for Telugu.

## 3 Telugu

Telugu is one of the 22 languages with official status in India. Telugu belongs to the South-Central subgroup of the Dravidian language family<sup>3</sup> and is mainly spoken in Southern India. The Dravidian language family was the subject of both historical and comparative linguistic research in the later half of twentieth century (Krishnamurti, 2003). Telugu is an agglutinative language like other Dravidian languages such as Tamil or Malayalam. The dominant word order in Telugu is Subject-Object-Verb (SOV) with inclination towards pro-drop. Telugu verbs inflect for gender, number and person. The “be” (*yun-*, “existential”) verb in Telugu shows agreement with the subject for gender, number, and person. The existential verb has a negative counterpart “not to be” (*le:-*) which can participate in both light and serial verb constructions and also act as the main verb.<sup>4</sup> Telugu does not have a dominant overt coordination as in English or Hindi. Telugu forms subordinate clauses through verbal nouns, verbal adjective, and converbs. Control constructions marked by  $\times_{\text{COMP}}$  relation are rare (less than ten instances in our treebank) in Telugu.

## 4 Corpus and Annotation

The Telugu treebank currently consists of 1328 sentences and 6465 tokens. The sentences were manually typed in Telugu script (derived from Brahmi script) from the examples in chapters 7–29 in (Krishnamurti and Gwynn, 1985). Many sentences in this book are collected from contemporary Telugu fiction of that time (1960s-80s). Both the authors manually annotated all the sentences, and disagreements were adjudicated after discussion. Annotation was done using the Brat (Stenetorp et al., 2012) tool and the conversion to CONLL format was done using a Python script. We annotated the sentences with UD POS tags and dependency relations (annotation guidelines available on the UD website<sup>5</sup>). The whole process of annotation and correction process took 4 months. The following sections outline our annotation decisions with examples.

<sup>2</sup>Available at <http://kcis.iiit.ac.in/>

<sup>3</sup>The Dravidian language family is one of the four language families spoken in India (the others being Indo-European, Austro-Asiatic, and Sino-Tibetan).

<sup>4</sup>The closest parallel of a negative verb is the negative auxiliary verb in Finnish <http://wals.info/chapter/112> and in Kurmanji (*tune*).

<sup>5</sup><http://universaldependencies.org/guidelines.html>

## 5 Part-of-speech annotation

UD specification defines 17 POS tags, of which we used 14 tags in Telugu. We did not use the tags - X, SYM and AUX. X was not used as there were no instances of unanalyzable foreign words that are not loan words in the corpus. SYM was not used as there were no symbols in the sentences. Words that resemble auxiliary verbs in Telugu also function as main verbs in the sentence. Hence, we did not use the AUX tag in our annotations. While we largely followed UD guidelines for POS tagging, we also made accommodations for some language specific phenomena. Among the open-class words, while Nouns, Proper Nouns and Verbs are relatively straightforward to tag, we made specific annotation decisions for adjectives and adverbs. Verbs functioning as Nouns or Adjectives were tagged Verbs but annotated with appropriate dependency relation (e.g., *acl* for verbal adjectives) to the head. We mark the morphological feature `VerbForm` with `Vnoun` and `Part` respectively.

**Adjectives (ADJ):** Adjectives in Telugu are indeclinable. Oblique nouns functioning as adjectives are tagged as NOUN with the relation `nmod:poss` to the head noun. Adjectives with a pronominal suffix (e.g., *manci=va:ɖu* good-3-SG-M. suffix, meaning: good one) are tagged as PRON and not adjective, as they refer to entities. Figure 1a illustrates a sentence with pronominalized adjective in predicate position with words transcribed in IPA. Adjectives denoting dimensions such as tall (*poɖugu*) or short (*poɖṭṭi*) do not need a pronominal suffix to function as the root of a sentence non-verbal construction. In such cases, we treat such as adjectives as abstract nouns and mark the POS tag as NOUN.

**Adverbs (ADV):** Krishnamurti and Gwynn (1985) and Krishnamurti (2003) note that adverbs of time and place behave as nouns (can inflect with case markers) in Telugu. We adopt the judgment into our treebank and mark all adverbs of time and place as nouns. We annotate an inflected time or location noun as NOUN and annotate it as the dependent of the dependency relations `obl` or `obl:tmod`. This is shown in Figures 1b and 1c.

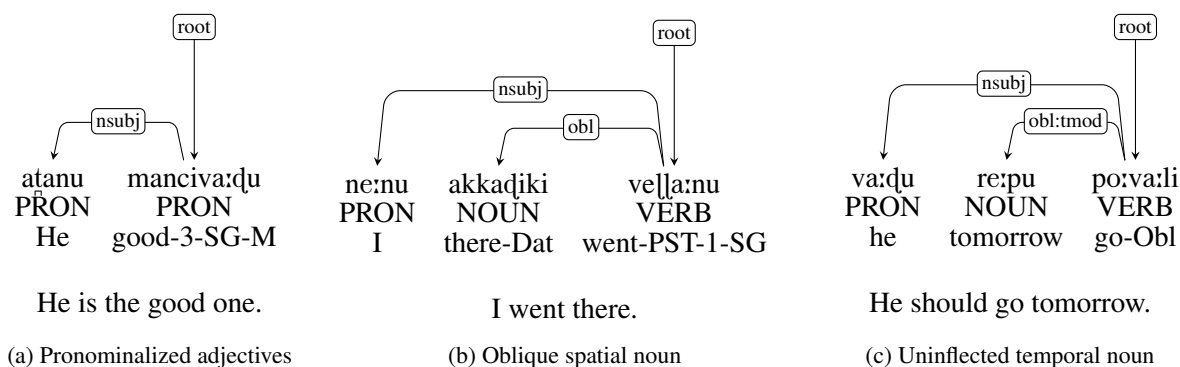


Figure 1: Part-of-speech tag examples

In the case of closed class words – Adpositions, Determiners, Numerals – we made the following language specific decisions:

**Adpositions (ADP):** Telugu uses postpositions and suffixes to denote cases. Postpositions are tagged as ADP and are dependent of nominals through `case` relation. Some adverbs indicating temporal or location information that appear as nominal modifiers are also tagged ADP.

**Determiners (DET):** UD guidelines distinguish 6 kinds of determiners. Of those, Telugu does not have articles and possessive determiners. We mark distal/proximal demonstratives and interrogative determiners that precede a nominal as DET. Telugu does not have relative pronouns and forms relative clauses through nominalization or verbal adjectives.

**Numerals (NUM):** Following UD guidelines, we tagged all numbers, fractions and multi-word numeric expressions with the NUM tag. However, numbers can also function as adjective, adverb or noun in Telugu, and can be inflected. Inflected numbers which do not appear in a multi-word numeric expression

are marked according to their syntactic function. UD guidelines also describe the tagging of non-cardinal numbers according to their syntactic function.<sup>6</sup>

## 6 Morphology

Telugu verbs show agreement with the agent in number, gender and person. Telugu has two genders: masculine and non-masculine and we will annotate with the same categories. Telugu has two numbers: singular and plural. Telugu nominals show highly inflected case system with nominative, dative, instrumental, genitive, commitative, ablative, and locative cases. Postpositions also function as adessive (*mi:da* “on”), purposive and comparative cases. Complex cases are formed through a combination of base case markers and postpositions.

Telugu verbs show tense, aspect, and mood. Verbs are typically active voice and passive constructions are not common. Causative constructions (*VOICE=Cau*) are formed by adding *-inc* to the transitive verb. Telugu also has a reflexive suffix *-kon* that is added to causative and transitive verb stems to denote that the agent is also the patient. We mark such a reflexive verb with *REFLEX=Yes*. There are two tenses: past and non-past. Telugu does not have a negative particle and shows negation through *-a-* marker that occurs before index markers. Verb can show aspect: habitual (*HAB*), progressive (*PROG*), perfect (*PERF*), prospective (*PROSP*) which are available in UD. The mood features are imperative (*IMP*), conditional (*CND*), potential (*POT*), necessitative (*NEC*), inceptive (*INC*), hortative (*HOR*).<sup>7</sup> Morphological annotation is not a part of the UD 2.1 release for Telugu and is a part of future work.

## 7 Universal Dependency Relations

Our treebank has 42 dependency relations, of which 11 are language specific. They are listed below in Table 1. Relations that are not seen in other language UD treebanks are marked with \*.

Relation	Description
<i>acl:relcl</i>	Relative clause
<i>advcl:cond</i>	Conditional Adverbial clause
<i>compound:lvc</i>	Light verb construction
<i>compound:redup</i>	Reduplicative construction
<i>compound:svc</i>	Serial verb construction
<i>nmod:cmp*</i>	Nominal comparative modifier
<i>nmod:poss</i>	Nominal possessive modifier
<i>nmod:tmod</i>	Nominal temporal modifier
<i>nsubj:nc</i>	non-canonical subjects (e.g., dative subjects)
<i>obl:tmod</i>	Oblique case-temporal
<i>obl:cau*</i>	Oblique case-causative (Section 6)

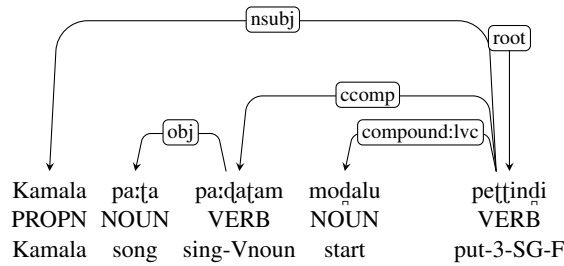
Table 1: Language Specific Dependency Relations for Telugu

While some relations such as *acl:relcl* and *nmod:poss* exist in several other language treebanks, other relations are not very common. Some of them are discussed below.

**Light verb constructions:** Light verbs are noun-verb constructions where the semantic content is in the noun even if the syntactic head is the verb. These constructions are wide spread in Hindi-Urdu (Butt, 2010; Vaidya et al., 2016). However, the Hindi UD treebank (converted from Paninian dependency treebank) does not seem to tag this construction specifically though it is annotated with *POF* (part-of) relation in the original Paninian treebank. We tag this construction explicitly using *compound:lvc*. In UD 2.1, this construction is explicitly marked in Farsi, Kazakh, Kurmanji, Marathi, Turkish and Uyghur along with Telugu. Recent work in Hungarian (Vincze et al., 2017) described these constructions using the label *dobj:lvc*. Light verb construction is illustrated in figure 2 where a noun (start) followed by a verb is used as a light verb compound.

<sup>6</sup><http://universaldependencies.org/u/pos/NUM.html>

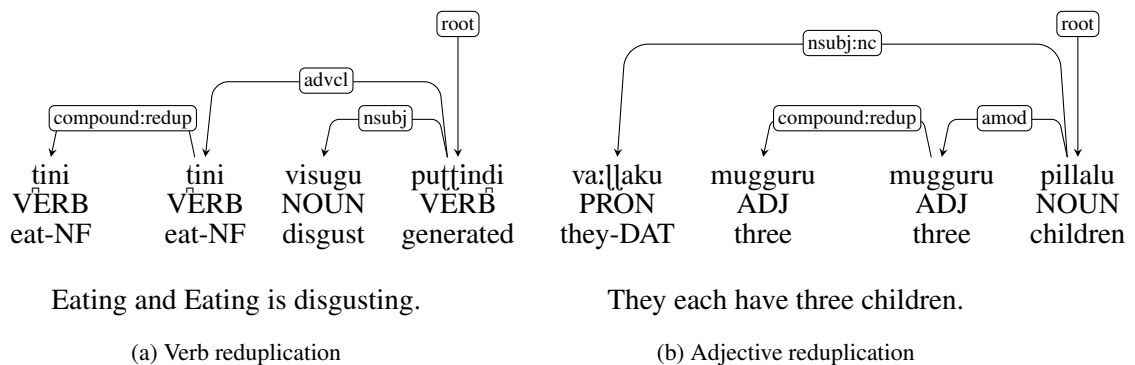
<sup>7</sup>Inceptive and Hortative moods are not available in UD.



Kamala started singing a song.

Figure 2: Light verb construction with a nominalized clausal complement.

**Reduplication:** Reduplication is the morphological process in which whole or parts of the word are repeated to denote a syntactic function. Reduplication (both partial and complete) is a common phenomenon in several languages although it is explicitly marked in only five other UD 2.1 languages – Hindi, Kurmanji, Marathi, Turkish and Uyghur – with the relation `compound:redup`. We mark all reduplicated words with this relation and treat the final word as the head. In Telugu, reduplication can occur across POS categories such as determiner, verbs, adjectives, nouns, and adverbs. We show examples of verb and adjective reduplication in figure 3.



Eating and Eating is disgusting.

They each have three children.

(a) Verb reduplication

(b) Adjective reduplication

Figure 3: Reduplication in verbs and adjectives.

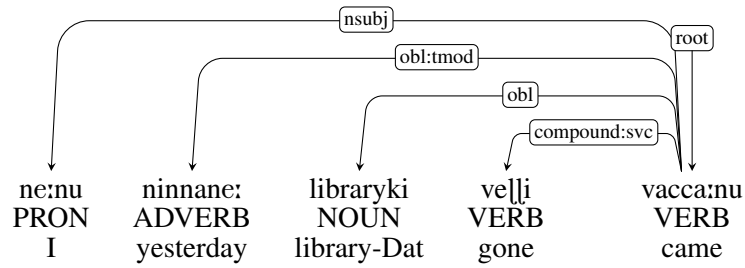
**Serial Verbs:** The Dravidian comparative literature defines serial verbs as a series of finite verbs which are present in Old Telugu but absent in Modern Telugu. There is no limit to the number of participating verbs in such a construction. We employ the definition of serial verbs from Velupillai (2012, 332) that a series of verbs referring to a single event is labeled as serial verb. These are different from other compound verbs such as V-V complex predicates in that they describe a sequence of actions. We mark such constructions with `compound:svc` (cf. Figure 4). The Dravidian comparative literature treats these constructions as adverbial clauses.

**Non-verbal predication:** In this paragraph, we present non-overt copula (cf. Figure 5a) and negative verb (cf. Figure 5b) which is specific to Telugu. Equative, attributive, possession, and benefaction constructions consist of NP+NP and lack an overt copula. Location construction (negation variant) shown in Figure 5b shows agreement and does not fall under the definition of non-verbal predication in UD.

**Genitives:** Genitive constructions can be formed through a preceding nominal dependent in nominative case, postposition (*yokka*), and oblique noun. We mark all these relations as `nmod:poss` (cf. figure 6).

**Comparatives:** Comparative constructions are formed through a special postposition *kanṭe* which is marked as a dependent of the second nominal through `case` relation (cf. figure 7).<sup>8</sup>

<sup>8</sup>At the time of submission of the paper, we marked the relation between the two nominals using `nmod:cmp` relation. We mark the relation between the second nominal and the root noun with `obl` relation. We thank one of the reviewers for pointing



I went to the library yesterday.

Figure 4: Serial verb construction

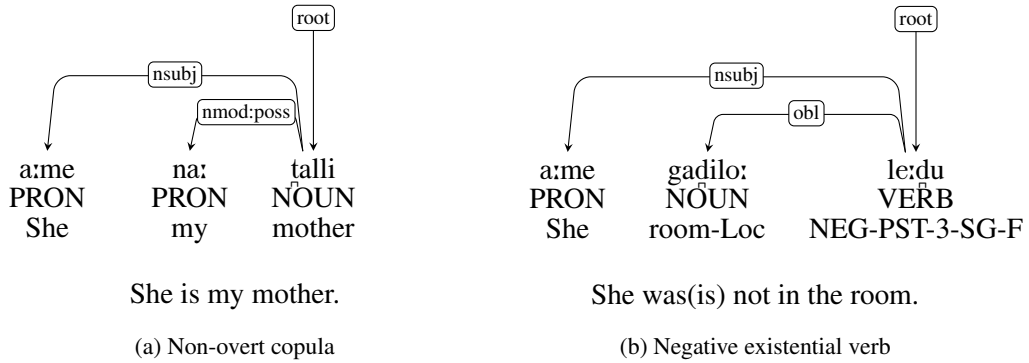


Figure 5: Non-verbal predication in Telugu.

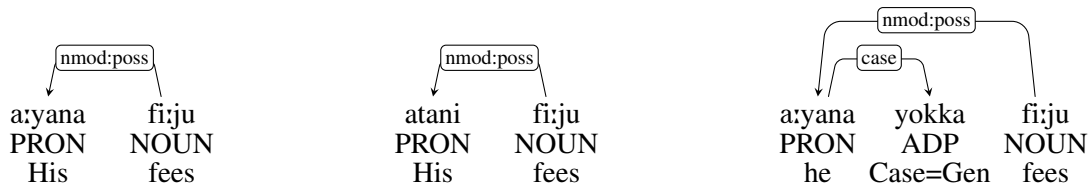
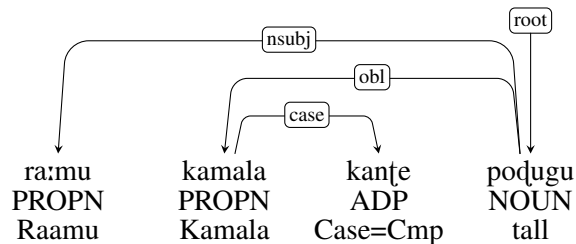


Figure 6: Genitive formation strategies



Ramu is taller than Kamala.

Figure 7: Comparative construction without an overt copula

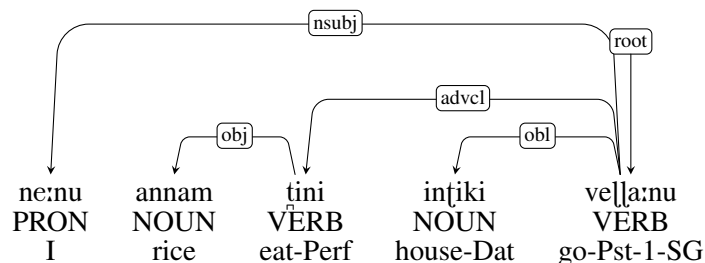
**Dative subjects:** Typically, NPs that occur at the sentence-initial position are in nominative case (unmarked). However, stative verbs such as “to know” and intransitive verbs such as “to want” do not show any agreement with any of the NPs in the sentence. In such cases, the NP in initial position is marked with dative case (Sridhar, 1979; Nizar, 2010). We mark the syntactic relation between the verb and the dative NP with `nsubj:nc`. Although dative NP occurs in sentence initial position, the free word order allows the dative NP to be moved to a non-final position in the sentence. A dative NP (annotated as

this mistake.



nsubj:nc) can also occur as the experiencer NP in non-verbal sentences (cf. figure 3b).<sup>9</sup>

**Adverbial clauses:** Telugu forms adverbial clauses through converbs. The final verb in the sentence is a finite verb which is treated as the root of the sentence. The subject of the embedded clause can be co-referential (cf. Figure 8) when the non-finite verb is marked for perfective or progressive aspect. In such a case, we annotate the subject to be the dependent of the main verb. Subjects of the main and subordinate clauses cannot be co-referential when the non-finite verb is marked for conditional or concessive moods.



I ate rice and went home.

Figure 8: Adverbial clause

**Relative clauses:** There is no relative pronoun in Telugu and relative clauses are formed through verbal adjectives. There are no expletive nominals in Telugu and cleft constructions are formed through pronominalized verbal adjectives. We analyze cleft sentences as relative clauses (cf. figure 9).

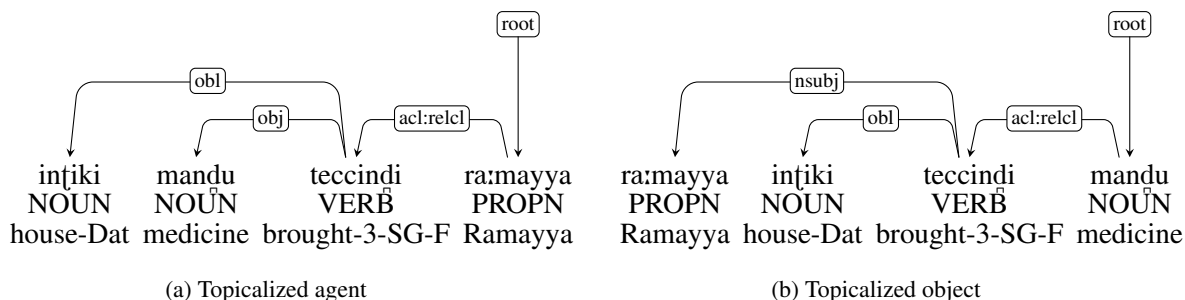


Figure 9: Cleft constructions derived from a simple sentence: *ra:mayya inṭiki mandu tecca:ḍu*. (Ram-mayya brought medicine home.)

**Nominalized clauses:** Non-finite verbs are nominalized by adding *-atam*. A nominalized verb can then be the head of a subordinate clause which can be the subject or object of the main verb (cf. figure 2). We annotate a nominalized verb clause as *csubj* (functions as subject) and *ccomp* (functions as object), respectively.

## 8 Tagging and Parsing Experiments

As a demonstration of the usefulness of our treebank in real world settings, we evaluated POS tagging and parsing models trained using UDPipe (Straka et al., 2016). UDPipe is a free, open-source, and language agnostic pipeline for training and evaluating NLP models for lemmatization, POS tagging and dependency parsing.

We split our treebank into 80-10-10 for training, development, and testing; and trained POS tagging and parsing models. Both training and evaluation was performed with UDPipe-1.2 on a Linux machine.

<sup>9</sup>We follow the Persian UD annotation guidelines (Seraji et al., 2016) in this case and name the dependency relation as *nsubj:nc*.

We report POS tagging accuracy, Labeled Attachment Score (LAS), and Unlabeled Attachment Score (UAS) on test set (after parameter tuning on development set) in Table 2. We also trained and evaluated a second parsing model on gold POS tags and found that the LAS and UAS scores are better than the joint model for predicting POS tags and dependency relations. We expect the POS tagging results to be high since nouns and verbs make the bulk of the part-of-speech tags in Telugu.

Input features	POS Acc.	LAS	UAS
Tagging + Parsing	90.43%	74.76%	87.79%
Parsing (Gold POS tags)	–	78.50%	89.74%

Table 2: Preliminary tagging and parsing results with UDPipe.

Previous work on Telugu dependency parsing – trained and evaluated with Paninian dependency labels – report the highest LAS of 70.15% (Husain et al., 2010) and best UAS of 90.5% (Kanneganti et al., 2016). While our LAS results are higher than both the previous results, the UAS results are slightly lower; however, a direct comparison is not possible due to the unavailability of the training data for these results and also due to different annotation schemes.

## 9 Conclusion

In this paper, we presented the first publicly available treebank for Telugu annotated in Universal Dependencies framework. We annotated POS tags and dependency relations from scratch for 1328 sentences. We trained and evaluated two parser models using UDPipe on the training split of the treebank and found that the parser performs within the range reported in the previous experiments.

As a part of future work, we intend to add morphological annotations to the treebank. It would also be interesting to compare different parsers on this treebank data. We are currently working towards expanding the treebank to include at least 100,000 tokens from Telugu Wikipedia. We plan to achieve this in a semi-automated fashion by running a trained parser model on Wikipedia sentences and then, manually checking and correcting for errors. We are also in the process of augmenting the treebank with fine-grained POS tags designed for Indian languages (Choudhary and Jha, 2011). The average sentence length in our treebank corpus is rather small ( $\sim 5$  tokens per sentence) whereas Wikipedia sentences are typically much longer. We intend to analyze how accurate can an automatic parser trained on grammar book examples would be when faced with longer sentences, with possibly complicated syntactic structures.

## Acknowledgments

The first author was supported by ERC Advanced Grant 324246 EVOLAEMP and a NRC LightHouse grant which is gratefully acknowledged. We thank Çağrı Çöltekin, Lilja Øvrelid, and Viswanath Naidu for all the discussions and input which proved useful in the development of the treebank. Finally, we thank the anonymous reviewers for the comments that helped improve the paper.

## References

- Elena Badmaeva and Francis M Tyers. 2017. A Dependency Treebank for Buryat. In *15th International Workshop on Treebanks and Linguistic Theories (TLT15)*. pages 1–12.
- Rafiya Begum, Samar Husain, Arun Dhvaj, Dipti Misra Sharma, Lakshmi Bai, and Rajeev Sangal. 2008. Dependency Annotation Scheme for Indian Languages. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*. pages 721–726.
- Yevgeni Berzak, Jessica Kenney, Carolyn Spadine, Jing Xian Wang, Lucia Lam, Keiko Sophie Mori, Sebastian Garza, and Boris Katz. 2016. [Universal Dependencies for Learner English](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, pages 737–746. <http://www.aclweb.org/anthology/P16-1070>.



- Riyaz Ahmad Bhat, Rajesh Bhatt, Annahita Farudi, Prescott Klassen, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, Ashwini Vaidya, Sri Ramagurumurthy Vishnu, and Fei Xia. 2017. The Hindi/Urdu Treebank Project. In *Handbook of Linguistic Annotation*, Springer, pages 659–697.
- Miriam Butt. 2010. The Light Verb Jungle: Still Hacking Away. *Complex predicates in cross-linguistic perspective* pages 48–78.
- Narayan Choudhary and Girish Nath Jha. 2011. Creating Multilingual Parallel Corpora in Indian languages. In *Language and Technology Conference*. Springer, pages 527–537.
- Çağrı Cöltekin. 2015. A Grammar-book Treebank of Turkish. In *Proceedings of the 14th workshop on Treebanks and Linguistic Theories (TLT 14)*. pages 35–49.
- Jan Hajič, Eva Hajičová, Marie Mikulová, and Jiří Mírovský. 2017. Prague Dependency Treebank. In *Handbook of Linguistic Annotation*, Springer, pages 555–594.
- Samar Husain, Prashanth Mannem, Bharat Ambati, and Phani Gadde. 2010. The ICON-2010 tools contest on Indian language dependency parsing. *Proceedings of ICON-2010 Tools Contest on Indian Language Dependency Parsing, ICON 10*:1–8.
- Silpa Kanneganti, Himani Chaudhry, and Dipti Misra Sharma. 2016. Comparative Error Analysis of Parser Outputs on Telugu Dependency Treebank. *Proceedings of CICLING 2016* .
- Bhadriraju Krishnamurti. 2003. *The Dravidian Languages*. Cambridge University Press.
- Bhadriraju Krishnamurti and John Peter Lucius Gwynn. 1985. *A Grammar of Modern Telugu*. Oxford University Press, USA.
- John Lee, Herman Leung, and Keying Li. 2017. [Towards Universal Dependencies for Learner Chinese](#). In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*. Association for Computational Linguistics, Gothenburg, Sweden, pages 67–71. <http://www.aclweb.org/anthology/W17-0408>.
- Aibek Makazhanov, Aitolkyn Sultangazina, Olzhas Makhambetov, and Zhandos Yessenbayev. ????. Syntactic Annotation of Kazakh: Following the Universal Dependencies Guidelines- A Report.
- Kadri Muischnek, Kaili Müürisep, and Tiina Puolakainen. 2016. Estonian Dependency Treebank: From Constraint Grammar tagset to Universal Dependencies. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*. pages 1558–1565.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D Manning, Ryan T McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. pages 1659–1666.
- Milla Nizar. 2010. Dative Subject Constructions in South-Dravidian Languages. *Unpublished master's thesis, University of California, Berkeley, Berkeley, CA* .
- Martha Palmer, Rajesh Bhatt, Bhuvana Narasimhan, Owen Rambow, Dipti Misra Sharma, and Fei Xia. 2009. Hindi Syntax: Annotating Dependency, Lexical Predicate-Argument Structure, and Phrase Structure. In *The 7th International Conference on Natural Language Processing*.
- Avinesh PVS and Gali Karthik. 2007. Part-of-speech Tagging and Chunking Using Conditional Random Fields and Transformation Based Learning. *Proceedings of the Workshop on Shallow Parsing for South Asian Languages* .
- Sampo Pyysalo, Jenna Kanerva, Anna Missilä, Veronika Laippala, and Filip Ginter. 2015. Universal dependencies for Finnish. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*. Linköping University Electronic Press, 109, pages 163–172.
- Loganathan Ramasamy and Zdeněk Žabokrtský. 2012. [Prague Dependency Style Treebank for Tamil](#). In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. İstanbul, Turkey, pages 1888–1894. <http://www.lrec-conf.org/proceedings/lrec2012/summaries/456.html>.
- Mojgan Seraji, Filip Ginter, and Joakim Nivre. 2016. Universal dependencies for persian. In *LREC*.
- Shikaripur N Sridhar. 1979. Dative Subjects and the Notion of Subject. *Lingua* 49(2-3):99–125.

- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. BRAT: A Web-based Tool for NLP-assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 102–107.
- Milan Straka, Jan Hajic, and Jana Straková. 2016. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. pages 4290–4297.
- Umut Sulubacak, Memduh Gokirmak, Francis Tyers, Çağrı Çöltekin, Joakim Nivre, and Gülşen Eryiğit. 2016. Universal Dependencies for Turkish. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. pages 3444–3454.
- Juhi Tandon, Himani Chaudhry, Riyaz Ahmad Bhat, and Dipti Sharma. 2016. [Conversion from Paninian Karakas to Universal Dependencies for Hindi Dependency Treebank](#). In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*. Association for Computational Linguistics, Berlin, Germany, pages 141–150. <http://anthology.aclweb.org/W16-1716>.
- Ashwini Vaidya, Sumeet Agarwal, and Martha Palmer. 2016. Linguistic Features for Hindi Light Verb Construction Identification. In *COLING*. pages 1320–1329.
- Viveka Velupillai. 2012. *An introduction to linguistic typology*. John Benjamins Publishing.
- Chaitanya Vempaty, Viswanatha Naidu, Samar Husain, Ravi Kiran, Lakshmi Bai, Dipti Sharma, and Rajeev Sangal. 2010. Issues in Analyzing Telugu Sentences towards Building a Telugu Treebank. *International Conference on Intelligent Text Processing and Computational Linguistics (CICLING)* pages 50–59.
- Veronika Vincze, Katalin Simkó, Zsolt Szántó, and Richárd Farkas. 2017. Universal Dependencies and Morphology for Hungarian-and on the Price of Universality. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. volume 1, pages 356–365.