1 Introduction

This document presents the parts of speech tagging and syntactic annotation guidelines of the Mbyá Treebank project (principal investigator: Guillaume Thomas, Department of Linguistics, University of Toronto). Mbyá is a Guaraní language of the Tupí-Guaraní family, with approximately 30,000 speakers in Argentina, Brazil and Paraguay (Ladeira, 2003).

Part of speech tagging and syntactic annotations in the treebank are meant to be compatible with the Universal Dependencies project (Nivre, 2016):

http://universaldependencies.org/

At the same time, our analysis of the morphology and syntax of Mbyá is rooted in Robert Dooley’s description of the language (Dooley, 2015). Dooley’s analysis of the syntax of Mbyá is not fully compatible with Universal Dependencies guidelines. This has lead us to depart from the letter of Dooley’s analysis in several respects. To a certain extent, this has also lead us to stretch Universal Dependencies guidelines. Although we try to indicate significant departures from both sources in the following document, the interested reader should consult the aforementioned references for more details.

2 Parts of Speech Tagging

In keeping with Universal Dependencies guidelines, the treebank makes use of two sets of Parts Of Speech (POS): language specific POS and so-called Universal Parts of Speech. Universal POS in use in the treebank are listed in table 1.

2.1 Parts of speech of open classes of lexical items

2.1.1 General Considerations

The four main POS of open classes of lexical items are nouns, verbs, adjectives and adverbs. In order to decide how to categorize a lexical item, one may use ontological criteria (does the item denote a thing, a property or an event?), morphological criteria (what inflectional and derivational affixes can the item be combined with) or syntactic criteria (what is its syntactic distribution)? This task is made more complex by the facts that these criteria do not always align with one another. In addition, one may wish to argue that certain lexical items belong to several POS, even when their interpretation is held constant. This type of cross-categorial flexibility is easily illustrated with adjectives and adverbs in English (e.g. a fast car vs. to drive fast).

We will discuss first the noun/verb distinction, then the categories of adjectives and adverbs.

2.1.2 Nouns and Verbs

The nature of the distinction between nouns and verbs in Tupí-Guaraní languages has been the subject of much debate, the main point of contention being the categorization of intransitive stative predicates. For a careful overview of different positions, see (Meira, 2006). See also (Dooley, 2015) for Mbyá, and
In these guidelines, we will only give a brief introduction to this debate as it applies to Mbyá Guaraní. As in other Tupí-Guaraní languages, transitive verbs cross-reference one of their arguments. Subjects are cross-referenced with a class A marker, while objects are cross-referenced with a class B marker:

(1) Ava a-exa.
    man A1.SG-see
    ‘I saw the man.’

(2) Ava xe-r-exa.
    man B1.SG-R-see
    ‘The man saw me.’

In the literature on Tupí-Guaraní languages, these markers have been analyzed as inflectional prefixes or pronominal proclitics. We treat them as prefixes. The choice of argument to be cross-referenced is governed by well-known principles that are not relevant to the present discussion. What is relevant though is that both classes of cross-reference markers illustrated in (1) and (2) are used to cross-reference subjects of intransitive predicates. Intransitive predicates whose subject is cross-referenced with a class A marker are known as active predicates, while those whose subject is cross-referenced with a class B marker are known as inactive. Dynamic predicates with agentive subjects tend to be lexicalized as active predicates, while stative predicates and dynamic predicates with non-agentive subjects tend to be realized as inactive predicates. There are however cases of active and inactive marking that do not fit this characterization, and that complicate the interpretation of the active/inactive distinction. See (Velázquez-Castillo, 20002) for an insightful discussion of this issue in Paraguayan Guaraní.

(3) A-nha.
    A1.SG-run
    ‘I ran.’
(4) Xe-kane’o.
   B1.SG-tired
   ‘I am tired.’

Crucially, class B markers are also used to cross-reference possessors on words that denote objects (rather than properties or events), and that would be translated into English as common nouns:

(5) Xe-irũ o-vaē.
    B1.SG-friend A3-arrive
    ‘My friend has arrived.’

To make the matter more complex, object denoting words can be used productively as predicates in Guaraní languages, in possessive and identificational or attributive predication:

(6) Xe-irũ.
    B1.SG-friend
    ‘I have a friend.’

(7) João xe-irũ.
    João B1.SG-friend
    ‘João is my friend.’

Likewise, words that denote states or properties are attested both as predicates or as arguments, although their use as arguments appears to be less productive than the predicative use of object words:

(8) Oo py o-ike h-aku gui.
    house in A3-enter EXPL-hot from
    ‘He entered the house because of the heat.’ (Dooley, 2016)

(9) Kuaray h-aku vaipa.
    sun B3-hot a.lot
    ‘The sun is very hot.’ (Dooley, 2016)

This states of affairs has lead some authors to categorize both object words and property words as nouns (Dietrich, 2017). An issue with this analysis is that all occurrences of inactive predicates would have to be treated as instances of nominal predication, which is counter-intuitive since property words are used as predicates more frequently than they are used as arguments, while object denoting words are used more frequently for reference than for predication.

How then should we deal with the categorial flexibility of object and property words? The strategy that we have adopted to answer this question is to encode the syntactic dimension of POS tagging (in this case, whether a word is used as an argument or as a predicate) at the level of universal POS, while its ontological and morphological dimensions are encoded at the level of language specific POS.

At the level of universal POS, words that are used as predicates are tagged as VERB, while words that are used as arguments are tagged as NOUN. At the level of language specific POS, non-predicative uses of object denoting words are tagged as nouns (n), while their predicative uses are tagged as nominal predicates (n:pred). Property denoting words that cross-reference their subject with a class B marker are tagged as inactive predicates (inact), and predicates that cross-reference their subject with a class A marker are tagged as verbs. Finally, language specific POS tags for verbs indicate their valency and
their inflectional class, see table 2. In addition, we use specific tags for two types of dependent verbs in complex predicates: \textit{v}s (Dooley’s (2015) secondary verb) and \textit{vpos} (Dooley’s (2015) postposed verb root).

\begin{tabular}{ll}
    \textit{vi:a} & intransitive active (dynamic) verb \\
    \textit{vi:d} & derived intransitive verb (always active) \\
    \textit{vd:i} & intransitive inactive indirect verb \\
    \textit{vd:a} & intransitive active indirect verb \\
    \textit{vt} & transitive verb \\
    \textit{vd} & transitive indirect verb \\
\end{tabular}

Table 2: Language specific verb tags

These conventions are illustrated in the following examples, repeated from (5), (7), (8) and (9), where morphological glosses are represented on the second line, language specific POS tags are represented on the fourth line, and the last line presents universal POS tags:

(10) Xe-irū o-vaē.
    B1.SG-friend A3-arrive
    n vi:a
    NOUN VERB
    ‘My friend has arrived.’

(11) João xe-irū.
    João B1.SG-friend
    nprop n:pred
    PROPN VERB
    ‘João is my friend.’

(12) Oo py o-ike h-aku gui.
    house in A3-enter EXPL-hot from
    n post vd:a inact post
    NOUN ADP VERB NOUN ADP
    ‘He entered the house because of the heat.’ (Dooley, 2016)

(13) Kuaray h-aku vaipa.
    sun B3-hot a.lot
    n inact intprt
    NOUN VERB PART
    ‘The sun is very hot.’ (Dooley, 2016)

2.1.3 Adjectives and Adverbs

While scholars generally agree on the existence of a Noun/Verb distinction in Guaraní languages, there is less agreement on the relevance of Adjectives and Adverbs to the grammar of these languages. A first reason to reject the existence of these categories in Mbyá is that many adjectival and adverbial modifiers are actually uninflfected property-denoting roots from the inactive inflection class. Secondly, although there are roots that can only function as modifiers and not as predicates, they typically function
as adjectives as well as adverbs. Furthermore, there are no morphological operations that map adjectives to adverbs, or vice versa, and no morphological operation that targets exclusively adjectives or adverbs (Dietrich, 2017). Consequently, it is reasonable to categorize these roots as modifiers, without invoking a categorial distinction between adjectives and adverbs.

Given these facts, we have decided to use the categories Adjectives and Adverbs only at the level of universal POS, where they reflect the syntactic function of modifier words: modifiers of nouns are tagged as \texttt{ADJ}, while modifiers of verbs, adjectives and adpositional phrases are tagged as \texttt{ADV}. The rest of this section describes how they are mapped to language specific parts of speech.

Adjectives are non-clausal modifiers of nouns. In Mbyá, we must distinguish two types of word forms used as adjectives. On the one hand, there exist inactive roots that can be used either predicatively, or as adnominal modifiers. When used predicatively, these roots are inflected with class B markers if at all, and they are tagged as verbs (Universal POS: \texttt{VERB}, language specific POS: \texttt{inact}). When used as adnominal modifiers, they can occur without inflection, in which case they are tagged as adjectives (Universal POS: \texttt{ADJ}, language specific POS: \texttt{inact}). The following examples illustrate:

(14) Avaxi o-nhotỹ r-xyx pora.
Corn A3-plant R-line good
\begin{verbatim}
NOUN  VERB  NOUN  ADJ
n   vt   n   inact
\end{verbatim}
‘He planted the corn in a good line.’ (Dooley 2015)

(15) Kova’e ára ma i-por’a vaipu.
DEM day BDY B3-good very
\begin{verbatim}
DET   NOUN   PART  VERB  PART
dem n   discprt   inact   intprt
\end{verbatim}
‘This day is very good.’ (Dooley 2015)

Adverbs are non-clausal modifiers of heads other than nouns. Just as in the case of adjectives, uninflected inactive roots can be used as adverbial modifiers, as illustrated in (16):

(16) Oro-va’u pora.
A2.PL.EXCL-happy a.lot
\begin{verbatim}
VERB  inact
\end{verbatim}
‘We were very happy.’ (Dooley 2015)

In addition to inactive roots, there is a small number of word forms in Mbyá that can be used as adnominal or adverbial modifiers, but that cannot be used predicatively. These word forms are tagged as modifiers (\texttt{mod}) at the level of language specific POS and as adjectives (\texttt{ADJ}) or adverbs (\texttt{ADV}) at the level of universal POS. An example is the word \textit{guaxu} (‘big’, ‘large’):

(17) O-vera va’e r-yku gui o-j-apo oja guaxu.
A3-shine REL  R-melted.solid from A3-B3-do pan big
\begin{verbatim}
VERB  SCONJ  NOUN  ADP  VERB  NOUN  ADJ
vi:a  rel  n  pos  vi:a  n  mod
\end{verbatim}
‘He cast the metal into a big pan.’ (Dooley 2015)
2.2 Other parts of speech

2.2.1 Particles

We distinguish Adverbs and Adjectives from Particles (PART). The set of expressions that we have characterized as particles is larger than Dooley’s (2015) proposed inventory. In deciding whether to categorize an expression as a particle, we follow the guidelines in (19):

(19) Particles categorization guidelines:

a. Expressions that denote objects, states, events or predicates thereof are not categorized as particles. Particles are typically function words, which express grammatical features of lexical words (e.g. aspect), discourse level information (e.g. discourse relations between propositional units), or illocutionary information (e.g. evidentiality).

b. Particles belong to closed classes of expressions.

c. Function words that are dependants of nouns but that cannot be categorized as determiners\(^1\) are categorized as particles.

d. Function words that are used flexibly as dependants of pronouns, nouns and verbs and that cannot be used predicatively are categorized as particles.

The universal parts of speech tag for particles is ‘PART’. In the language specific inventory of parts of speech, particles have been sorted in subcategories, which reflect their semantic properties:

(20) Types of particles:

a. Aspect particles: \(\text{aaspr}t\): \(já\), \(ma\) (completive), \(\text{tema}\) (continuative), \(ta\) (prospective), …

b. Discourse particles: \(\text{discprt}\): \(\text{h'}avь\) (‘then’), \(ma\) (spacing particle), \(\text{minha}\) (similative), \(ri\), (‘indicate allusion to something activated in the context’, Dooley 2015) \(\text{rimа}\) (“corrects a presupposition”, Dooley 2015), …

c. Focus particles: \(\text{focprt}\): \(\text{anho}\) (‘only’), \(\text{avei}\) (additive), \(\text{ju}\) (repetitive), \(\text{jevy}\) (repetitive), \(\text{ne}́\и\) (“not even”), \(\text{teri}\) (still), \(\text{rive}\) (‘only’), …

d. Illocutionary particles: \(\text{illocprt}\): \(\text{je}\) (reportative), \(\text{ke}\) (imperative), \(\text{ko}\) (epistemic), \(\text{na}\) (imperative), \(\text{nda}'u\) (confirmative), \(\text{nda}'u\) (interrogative), \(\text{rа}'е\) (mirative), \(\text{rа}'а\у\) (epistemic), \(\text{pa}\) (interrogative), \(\text{po}\) (epistemic), \(\text{rа}ко\) (epistemic), \(\text{tu}\) (mirative), \(\text{ty}\) (mirative), …

e. Intensifiers: \(\text{intprt}\): \(\text{rei}\), \(\text{ae}\), \(\text{katuve}\), \(\text{merами}\), \(\text{ete}\), \(\text{nho}\), \(\text{katu}\), \(\text{voи}\), …

f. Modal particles: \(\text{modprt}\): \(\text{mate}\) (deontic), \(\text{ra}'а\га\) (counterfactual), \(\text{teй}\) (frustrative), …

g. Quantificational particles: \(\text{quaptprt}\): \(\text{kuеy}\) (associative/group plural), \(\text{rа}\) (free choice), \(\text{меме}\) (internal pluractional, universal quantifier), …

h. Question particles: \(\text{qprt}\): \(\text{nda}'u\), \(\text{pa}\), \(\text{ty}’ь\), …

i. Tense particles: \(\text{temprt}\): \(\text{a}ра\'а\е\) (distant past, indirect evidence), \(\text{kаrамboаe}\) (distant past, direct evidence), \(\text{kuevе}\) (intermediate past), \(\text{kури}\) (current past or future), \(\text{rа}'а\) (future), \(\text{раmо}\) (near past), \(\text{voи}\) (‘early’), …

---

\(^1\)Because they do not have the semantic function of quantifiers or (in)definiteness markers.
A few comments are in order on at least some of these subcategories. Discourse particles are composed of two types of particles: (i) discourse connectives that express relations between sentences or clauses, but do not have the distribution of conjunctions and (ii) spacing particles. The latter is a category of expressions identified in Dooley’s (2015) grammar, whose function is to indicate boundaries between constituents in the information structure of the clause (Dooley 2015: §24.4.3.2). The expressions that we have categorized as focus particles are similar to focus sensitive particles in better studied languages, both in their interpretation and in their syntactic flexibility, which allows them to modify predicates as well as arguments and modifiers. Illocutionary particles include a group of expressions that one may tentatively analyze as illocutionary modifiers, such as evidentials, miratives, and imperative markers. Quantificational particles include functional dependants of nouns and other categories that express plurality or quantificational notions, but their distribution differs from that of Mbyá determiners (they are postposed to the noun, along with other particles, and unlike adnominal quantifiers and demonstrative determiners).

Finally, note that many verbal particles may occur in construction with arguments of the verb, in which case they are interpreted as spacing particles (Dooley 2015: §24.4.3.2). This phenomenon does not affect the glossing and tagging of the particles, although it is reflected in the dependency annotation. See section 3 of this document for more details.

2.2.2 Pronouns
Personal pronouns are tagged as pro in the language specific tagset and PRON in the universal tagset. This category also includes indefinite, interrogative, negative and total (universal) pronouns:

(21) Xee nd-a-exa-i mba’e-ve rei.
    1.SG NEG-A1.SG-see-NEG thing-NEG INT
    pro vt pro intprt PRON VERB PRON PART

‘I didn’t see anything at all.’ (Dooley 2015)

2.2.3 Numerals
Numerals used as adnominal modifiers are tagged num in the language specific tagset, and NUM in the universal tagset:

(22) Yma je o-iko mokoi ava-kue.
    old HSY A3-live two man-PL
    inact illocprt vi:a num n
    ADV PART VERB NUM NOUN

‘A long time ago there lived two men.’ (GUN001R001I001)

2.2.4 Quantifiers and Determiners
Quantifiers and demonstratives were tagged as quant and dem at the language specific POS level. These expressions are multi-functional, and can be tagged as adjectives, adverbs, determiners or pronouns at the universal POS level:

(23) Upe opy ke e-ike eme.
    DEM inside.a.house IMP 2.SG.IMP-enter NEG.IMP
    dem n illocprt vi:a negprt
    DET NOUN PART VERB ADV

‘Do not enter this room.’ (GUN001R007I001)

---

2.2.5 Adpositions

Adpositions in Mbyá are postpositional. They are tagged as post in the language specific tagset, and ADP in the universal parts of speech tagset:

(25) Tetã gui a-ju.
    city from A1.SG-come
    n post inact
    NOUN ADP VERB

‘I came from the city.’ (Dooley 2015)

2.2.6 Nominalizers, relativizers and adverbializers

Clausal nominalizers, relativizers and adverbializers (words that introduce adverbial clauses) are all analyzed as subordinating conjunctions (SCONJ) in the universal POS tagset.

Mbyá makes ample use of category changing operations of nominalization in the construction of clausal subordination. The clitic a, in particular, is frequently used to nominalize clausal complement of verbs or postpositions:

(26) A-a ta xe-ru o-ĩ a py.
    VERB PART NOUN VERB SCONJ ADP
    vi:a asprt n vi:a nmlzer post

‘I am going where my father is.’ (Dooley 2015)

Note that we have decided to represent clausal nominalizers (language specific tag: nmlzer) as clitics. This allows us to represent nominalization as a syntactic dependency, which in turns makes it possible to represent the syntactic scope of nominalization more accurately, while respecting the Universal Dependency definition of syntactic dependencies as relations between words.3

Relative clauses are formed with the clitic va’e. Its language specific tag is ‘rel’. Va’e may also be analyzed as a nominalizer, insofar as it licenses the use of nominal temporal suffixes. The following example illustrates the use of va’e in the formation of headless relative clauses:

(27) Aa ma i-kuai re-raa va’e-rā.
    here BDY B3-be.PL A2.SG-take.away REL-FUT
    dem discrpt inact vt rel
    ADV PART VERB VERB SCONJ

‘Here are the things that you will take away.’ (Dooley 2015)

Finally, the grammar of Mbyá also includes clausal subordinators that do not nominalize their complement. These subordinators are most frequently used to introduce adverbial clauses, although some of

3To illustrate, the nominalizer a may be phonologically dependent on an adverbial modifier of the verb that it nominalizes. In such examples, the fact that a nominalizes the whole clause that is headed by the verb would not be captured syntactically in the treebank if a were transcribed as a suffix attached to the adverb.
them can be used to introduce complement clauses. Two frequent non-nominalizing subordinators are
the switch-reference markers vy (same subject) and rā (different subject):

(28) A-nh-aka’õ uru a-juka vy.
A1-OBJ-behead chicken A1-kill SS
vt n vt subordconn
VERB NOUN VERB SCONJ
‘I beheaded the chicken, killing it.’ (Dooley 2015)

These subordinators have been tagged as subordinating connectives (subordconn) in the language spe-
cific tagset.

2.2.7 Conjunctions

Coordinating conjunctions are tagged as coordconn in the language specific tagset and CCONJ in the
universal tagset. Note that the coordinating conjunction ha’e is a homonym of the third person pronoun
ha’e:

(29) Nha-ma’ê monde r-e ha’e nhuã r-e.
A1.INCL-look trap R-ABL and trap R-ABL
vt n post coordconn n post
VERB NOUN ADP CCONJ NOUN ADP
‘We looked at the mundéus and at the traps.’ (GUN001R004I001)

Mbyá has a rich inventory of sentence initial discourse connectives built by combining the word ha’e
with postpositions and adverbializers. Following Dooley (2015), we analyze ha’e in these expressions
as a pronoun that refers to a situation described by a previous sentence or discourse unit (cf. this in the
English discourse connective contrary to this):

(30) Ha’e vy je mboapy xi’y o-juka.
3 SS HSY three coati A3-kill
pro subordconn illocpt num n vt
PRON SCONJ PART NUM NOUN VERB
‘And then, he killed three coaties. (GUN001R002I001)

3 Dependency annotation used in the treebank

3.1 General Principles

Syntactic annotation in the Mbyá Treebank is done in dependency grammar, in the style of Universal
Dependencies (version 2.4). Our analysis of the syntactic structure of Mbyá sentences is based on
Dooley’s (2015) description of the language, to the extent that it can be recast in Universal Dependencies.

In dependency grammar, the syntactic structure of a sentence is represented as a set of grammatical
relations between words, which are called dependencies. Dependencies are binary and asymmetric: one
member of the relation is a governor (also called head), and the other member is a dependant. Every
dependency relation has a type that indicates the nature of the grammatical relation between the head
and its dependant, as illustrated in the following example:
This example also illustrates the principle that all words in a sentence are dependants of another word, save for the head of the sentence, which is the dependant of an abstract root node.

In Universal Dependencies, dependency relations are assumed to hold primarily between content words. Function words are dependants of content words. To illustrate, in the sentence ‘The dogs are sleeping,’ the noun ‘dogs’ is the subject of the verb ‘sleeping,’ while the determiner is a dependant of the noun and the auxiliary is a dependant of the verb.

Guidelines for syntactic annotations in Universal Dependencies are available at the following URL:

http://universaldependencies.org/guidelines.html

This document focuses on issues specific to the annotation of Mbyá texts.

3.1.1 Core arguments and obliques

Nominal arguments are related to their verbs through the relations nsubj (nominal subject) and obj (object). Nominal dependants of verbs that function as oblique arguments or adjuncts are related to the verb by the relation ob1 (oblique):

Clausal subjects and clausal complements are related to the verb by the relation csubj and ccomp, respectively. The following example illustrate the latter:

We consider that an argument is clausal when it is headed by a verb, or by a predicative noun. We make an exception to this rule for headless relative clauses, which we treat as non-clausal arguments.
This allows us to mark the difference between individual denoting clausal arguments and proposition denoting clausal arguments through the dependency relations that introduce them:

(34) ‘It ate everything it found.’ (GUN001R012I001)

3.1.2 Modifiers

Applying Universal Dependency guidelines to the annotation of Mbyá modifiers is not trivial, due to the frequent use of particles whose syntactic behaviour does not fall squarely within the realm of adjectival or adverbial modification. We will first discuss less problematic cases of modification, and then move on to particles.

Words that modify a word of category NOUN but are not particles are introduced by relations amod. Words that modify a word of category VERB, ADJ or ADV but are not particles are introduced by advmod, see example (34). This relation is also used to introduce non-particles that modify adpositional phrases.

(35) ‘And then the angry man bared its teeth at the man.’ (GUN001R002I00)

Relative clauses are related to their head by the dependency acl:

(36) ‘Then we saw a man with only one hand.’ (GUN001R023I001)

Adverbial clauses are related to their head by the dependency advc1:
Because particles can be dependants of nouns as well as verbs or modifiers, they do not fall squarely within the categories of adjectival or adverbial modifiers. Because of this, we decided to introduce all particles by a single dependency relation dep:mod, which we interpret as a modifier relation that is not specified for the category of its head. The following examples illustrate:

(38) ‘He stayed in the forest for a long time.’ (GUN001R003I001)
3.2 Complex Predicates

Complex predicates are composed of a main predicate followed by one or several dependent predicates. The dependent predicate may be a postposed root (Dooley 2015: §12.1.1) or a dependent serial verb (Dooley 2015: §12.3.2). The former are tagged as vpos in the language specific tagset. The latter are marked by the suffix -vy or one of its allomorphs (-py, -my, -ngy, -ny). Both types of dependent verbs are related to their head by the relation compound:svc, see example (41).

3.3 Adpositions and Subordinators

Following Universal Dependencies guidelines, adpositions are introduced by the relation case, while clausal subordinators (subordinating conjunctions and relativizers) are related to their heads by the dependency mark, see example (40).

Note that the relativizer va’e often modifies nouns, in what appear to be relative clauses with nominal heads. We treat these constructions as nominal dependants, and do not use the acl dependency, although the relativizer is still related to its head by the dependency mark:

(42) ‘That is him!’ said the woman. (GUN001R003I001)

The relativizer va’e is frequently combined with the nominal temporal prefixes -kue and -rã. The resulting expressions, va’ekue and va’erã, can be used as temporal modifiers of independent predicates. We analyze these constructions as forms of insubordination, and we relate va’ekue and va’erã to their head with the relation mark:
‘I will lend my horse,’ he said.’ (GUN001R016I001)

3.4 Conjunctions and Sentence Connectives

Most sentences in narratives start with a sentence initial discourse connective. These connectives are composed of the pronoun ha’e, which is generally followed by an adposition or a subordinator. The form and interpretation of these connectives is described in Dooley (2015), §21.5 and §23.1. The head of these complex connectives is the pronoun ha’e. They are related to the main predicate of the clause by the relation obl:sentcon:

“But the older brother said: ‘You stay here.” (GUN001R001I001)

The reader might wonder why the switch reference marker rā in this example was tagged as a subordinating conjunction, rather than as a post-position, since its head is a pronoun rather than a clause. The use of subordinating conjunctions to introduce pronouns appears to be restricted to sentence initial connectives, and is likely due to the fact that pronouns in this constructions refer to propositions or discourse units, rather than to individuals. Because of these idiosyncracies, we consider that sentence initial connectives do not provide evidence for a reanalysis of subordinating conjunctions as post-positions.

3.5 Reported speech

In Universal Dependencies, the complement of a verb of saying in reported speech is related to the verb by a parataxis dependency. When the reported clause follows the verb of saying, the latter is treated as the head. When the verb of saying follows the reported clause, the main predicate of that clause is treated as the head. In our corpus, we treat postposed verbs of saying as heads, when there is no concurrent verb of saying, since the OV order is attested in Mbyá and the reported sentence can be analyzed as the complemented of the clause final verb of saying, see example (43). In addition, in order to distinguish complement relations in reported speech from other forms of parataxis, we use the relation parataxis:rep to relate a verb of saying to its complement. However, postposed verb of saying that repeat a verb that precedes the speech report are treated as dependants of the head of the reported clause, in keeping with Universal Dependencies guidelines:
(45) ‘Then his grandfather said: ‘Go far away,’ he said.’ (GUN001R007I001)

3.6 Nominal Predicates

In Mbyá, nouns can be used as identification/classification predicates, possessive predicates and meteorological event predicates without copula, see e.g. hogue (‘leaf’/‘have leaves’) and pytă (‘dark’/‘get dark’). Nominal predicates have been tagged as VERB in the universal tagset, and their modifiers are treated as adverbs rather than adjectives:

(46) ‘One day, he went there and the cabbages had beautiful leaves.’ (GUN001R003I001)

(47) ‘Very late at night, he heard a scratching noise on the window.’ (GUN001R003I001)

References


