Applying Universal Dependency to the Arapaho Language

Irina Wagner[1], Andrew Cowell[1], Jena D. Hwang[2]

[1]University of Colorado Boulder, Department of Linguistics; [2]IHMC

{irina.wagner, james.cowell}@colorado.edu, jhwang@ihmc.us

Abstract

This paper discusses the use of Universal Dependency for annotations of a Native North American language Arapaho (Algonquian). While some relations of the universal dependency perfectly correspond with those in Arapaho, language specific annotations of verbal arguments elucidate problems of assuming certain syntactic categories across languages. By critiquing the influence of grammatical structures of major European and Asian languages in establishing the UD framework, this paper develops guidelines for annotating a poly-synthetic agglutinating language and sets a path to developing a more comprehensive cross-linguistic approach to syntactic annotations of language data.

1 Introduction

The recent initiatives to create a cross-linguistic scheme of annotation rely on Universal Dependency (UD) as a system of describing the syntactic connection between words (Nivre, 2015; de Marneffe et al., 2014). While research shows this annotation type is effective not only for monolingual parsers but also cross-linguistically across multiple platforms, the universality of this approach is based on the assumptions of similar syntactic structures of major, often European, languages (McDonald et al., 2013). Without doubt, those are also the languages that receive predominant attention in the computational sphere, the languages whose technological presence requires a thorough analysis and annotation. However, if the goal of natural language processing is truly to develop a universal cross-linguistic strategy for annotating and analyzing linguistic data, it is important to attend to lesser described languages that may present strikingly different syntactic structures and dependencies.

Applying the UD rules while annotating the data from the Arapaho (Algonquian) language, several specific features were observed to fall outside of the charted labels. Since the language does not have a fixed word order and allows discontinuous constituency, dependencies on the previous word were avoided and re-analyzed. The most problematic dependency distinction in this language is the variation in relations between a verb and its arguments. This paper examines the correlation of the dependency relations in the UD scheme and their practical application for the Arapaho data. Using the UD framework, we create guidelines for annotating this data. In considerations of space, this paper primarily focuses on the argument structures defined by the UD and their correspondences to the Arapaho syntactic patterns. An additional discussion of non-verbal roots and topicality problematizes some of the common assumptions in discounting pragmatic features while analyzing syntactic dependencies.

In the following pages, we first provide a short note on the Arapaho language and the procedures of annotations (2); discuss issues of mapping the labels for subject, objects, and noun modifiers of the UD onto the Arapaho dependencies (3); define the mechanism of analysis of non-verbal roots (4); and suggest further ways of developing these annotation guidelines (5).

2 Arapaho data and annotations

Arapaho is an Algonquian poly-synthetic agglutinating language spoken by less than 200 people in the Wind River Indian Reservation in Wyoming. Because the language is in critical condition, there have been attempts at documenting and preserving it. A large transcribed and annotated spoken corpus has been created and parts of it are now available in the Endangered Languages Archive[1]. A

---

[1]http://elar.soas.ac.uk/deposit/0194

total of around eighty thousand lines transcribed, translated, and grammatically analyzed is available for further processing. The current attempts at establishing the dependency scheme for this language initiate the new type of analysis of this data to allow machine processing.

2.1 Some features of the Arapaho language

The current paper largely relies on the previous description of the Arapaho grammar by Cowell and Moss (2008). There are several intriguing features of the grammar, but the ones most relevant to this study are its complex verbal morphology, split semantic and syntactic transitivity, and the system of obviation.

2.1.1 Verbal complexity

As is observed in many other poly-synthetic languages, Arapaho verbs are highly complex and mark multiple grammatical and semantic features. So, in example (1), a single verb demonstrates incorporation of not only the usual tense, aspect, mode, person, and number features, but also the manner of action and an incorporated object.

(1)   he'ih'ii-xoo-xook-bix-ohoe-koohuut-oo-no'
      "Their hands would go right through them and appear on the other side."

A single verb can be a full clause conveying a full thought. Verbal prefixes code grammatical as well as many semantic features, inhibiting the dependency analysis since this framework only considers the relations between individual words.

2.1.2 Transitivity

The category of verbal transitivity is both syntactic and semantic (Cowell and Moss, 2008). To understand how many arguments are allowed in a verb's frame, one must examine both the morphological and the semantic structure of a verb. So, while semantically a verb *to'oo3ei* "to hit things" may appear transitive, grammatically it is intransitive, requiring only one argument, the subject, as in *too'oo3einoo* "I am hitting (unspecified) things." The transitivity of a verb is expressed in its inflection which must agree in person and number with its arguments. Truly transitive verbs carry inflections agreeing with both of its arguments:

(2)   Nih-to'ow-oo-t nuhu' **hinen-ino**
      PST-hit-3/4-3S this   man-OBV.PL
      "He hit these men"

Even though only one of the two arguments appears in the sentence, the verb *nihto'owoot* "s/he

hit him/her" is marked to agree both with the semantic agent and undergoer of the verb. This semantic distinction in the arguments is not observed in intransitive and semi-transitive verbs. Because such verbs demonstrate morphological agreement only with one nominal[2], other nominals are considered outside of the argument structure of a verb even if they specify the semantic patient or theme.

(3)   nih'ii-koo-ko'uyei-3i'   **biino**
      PST.IMPF-REDUP-pick things-3PL **chokecherries**
      "They were picking chokecherries."

So in the example (3), the noun *biino* "chokecherries" is not reflected in verbal morphology, but corresponds with its semantics by specifying the object of picking. Being outside of the argument structure of this verb, syntactically the noun is better understood as a verbal adjunct specifying the manner of action, while semantically it is still the patient. So the designation of the relationship between such arguments and verbs as *dobj* of the universal dependencies is wrong because it does not consider verbal morphology, whereas the label of *nmod* would not account for its semantic role.

2.1.3 Obviation

Unlike many languages, Arapaho does not rely on word order or case markers to disambiguate between overt nominals; rather it uses a system of obviation that incorporates a distinction based on animacy along with the combination of verbal morphosyntax and pragmatics to mark particular grammatical roles. This system clearly distinguishes between two third person referents by marking one of them (a less salient one in the discourse) as obviative and leaving the other referent unmarked (proximate). In Algonquian languages, the obviation is argued to be a pragmatic feature structuring discourse outside of a single clause (Goddard, 1984). Verbal morphology also shows agreement with these categories: the transitive verb inflection clearly marks which argument is acting on the other. So, instead of the usual three persons, Arapaho has four, with the fourth person being the obviative argument. In the example below, the obviative argument is the noun *hiinoon* "his mother" which corresponds with the verbal subjunctive inflection *-eihok* "4th person acting on 3rd singular."

---

[2]We use phrases "nominal" and "nominal expression" to refer to nouns, noun phrases and nominalized verbs that function as noun phrases.

(4) Hohou, hee3**eihok**   **hiinoon**
thank you say to s.o.-4/3S.SUBJ his/her mother
3eeyokooxuu.
Tipi-pole Child

"Thank you," his mother said to Under-the-Tipi-Pole Child.

As it is observed in this example, obviation does not correspond with the semantic or the syntactic role of an argument. Neither it depends on the transitivity of a verb. Rather, obviative status lines up with the semantic role of an obviative coded in verbal morphology. Based on this feature of transitivity and obviation, the current paper suggests employing the semantic labels in marking the syntactic relations.

2.2 Annotation procedures

We are not aware of previous attempts at dependency annotations with other Algonquian languages; however, dependency grammar has been one of the theoretical approaches in Algonquian syntax. The guidelines discussed below were created based on the annotations of a small set of Arapaho narratives. In the first phase of the project, the dependency relations were outlined based on annotations of a sample of several traditional narratives, totaling at about two thousand lines[3]. The annotators, one fluent non-native speaker and three graduate students in Linguistics well familiar with Arapaho language structure, were given a protocol established without the considerations of the UD framework but based purely on the Algonquian syntax patterns. Several problems using these syntax patterns clarified and specified the dependency relations, leading to the creation of a new set of labels.

In the second phase of the project, these new labels were further standardized based on the principles of the Stanford Dependencies (de Marneffe and Manning, 2008). Using this new set of relations, the annotations of the previous phase were converted and a total of 3616 lines of elicited personal and traditional oral narratives as well as 593 lines of conversational data were newly annotated. The disfluency of the conversational data indicated major issues with this annotation scheme which prompted us to turn to the UD-based system. The guidelines presented here have been used to re-mark the previous annotations of the data used in the second phase. No special software to perform
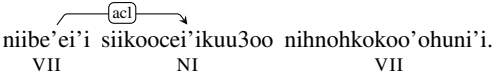
---

[3]What we call "lines" here refers to a ToolBox line which represent a single clause, or a complete thought.

the annotations has been used thus far, and all of the annotations are stored in a spreadsheet format.

Because the language is critically endangered, the resources available for this type of work are extremely limited. Importantly, it is not just that there are fewer recorded texts and conversations, but there are also fewer trained individuals able to perform any type of language annotation. So, during this particular project, most of the annotations were done by the first two authors of the paper with Andrew Cowell being the language expert due to his experience and acquired proficiency in the language. Over the course of six months, authors met regularly to discuss the annotations, solve the occurring issues, introduce and update labels. As a result, all of the current annotations are single annotated. The next part of the project includes more manual annotations using the guidelines proposed here as well as double annotations of at least a portion of the data to establish the inter annotator agreement. Having already annotated a few thousand lines of narratives, the focus of the following work will be on conversational data followed by machine learning.

3 Mapping the UD scheme

Out of the forty dependencies proposed by the UD, thirty Arapaho dependencies have one-to-one correspondence. Additional seventeen specifications and relations have been added to describe language-specific instances. The final scheme of Arapaho's nominal argument dependencies is presented in Table 1. Some of the dependencies were not used in the Arapaho scheme because such dependencies merely do not exist in this language. So, for example, the language does not have a grammatical category of an adjective; therefore, the dependency *amod* has not been used; instead descriptive verbs are analyzed as relative clauses, *acl*. Example (5) demonstrates the relative clause dependency where verb modifies the noun in the same manner that an adjective would.

(5)
⌒——[acl]——↘
niibe'ei'i siikoocei'ikuu3oo nihnohkokoo'ohuni'i.
VII   NI   VII

nii-be'ei-'i   siikoocei'ikuu3oo
IMPF-red-0PL rubber item
nih-nohk-okoo'ohuni-'i.
PST-INSTR-sealed with stiff object-0PL

"They would be sealed with the red rubber gasket."

In addition, there are no relative pronouns in the language, so the dependency relation *marker* is

also obsolete in the current scheme. Similarly, there is no category of a number or numeral; instead the number can be expressed by a verb or a particle, at which instance it is analyzed just like other particles with the dependency of *advmod* to the word that it modifies. In sum, omitted UD relations are the ones that are either expressed by some other dependency or non-existent in the Arapaho language.

Several UD dependencies perfectly line up with the Arapaho scheme. So such relations as noun modifiers, adverbial modifiers, adverbial clauses, determiners, appositives, relative clauses, case markers, and a few more have a direct correspondence. For example, an adverbial clause in Arapaho is very similar, if not the same, as adverbial clauses described for other languages in the UD. Arapaho adverbial clauses, as it is seen in the example below, lack a distinct word introducing it; instead, the head of an adverbial clause exhibits particular morphological markers indicating its dependency. So in the example (6) this distinction is made by the subjunctive mode indicating that the verb *bih'iyoohok* "when it is dark" is a dependent of the main verb of the sentence.

(6)

$$\overset{\frown}{\boxed{\text{advcl}}}$$

Bih'iyoohok ce'no'useeni'.
VII   VAI

Bih'iyoo-hok ce'-no'usee-ni'.
dark-SUBJUNCT back-arrive-1PL

"When it's dark, we'll come back."

In general, dependencies between function words and content words mirror the same dependencies in the UD framework, and most of these dependency labels are used.

The most complicated dependency relations tend to be between the content words, and especially the relations between the verb and its arguments. From the UD scheme, only one of such relations matches the Arapaho scheme with some modifications: *nsubj* and *csubj* correspond to subjects of intransitive verbs and transitive inanimate verbs. Similarly, subjects of passive verbs also correspond to the *nsubjpass* and *csubjpass* dependencies. Additional provisions are made in Arapaho scheme to account for the obviation status. In the following section, we discuss all of the provisions and additions made to the argument dependencies.

3.1 Subjects

While there is some correspondence between the UD's *nsubj* and subjects in Arapaho, it is, nonetheless, problematic to analyze subjects based purely on syntax since there is no syntactic features that would index the particular verbal arguments. Because nominals can take any position in the sentence and because they are not marked by a case corresponding with its syntactic role, the only certain way of finding a subject is in the person and number verbal agreement. The proximate and obviative distinction also does not clarify the syntactic role of the nominal, so with transitive verbs, the proximate form can be either agent or undergoer, and thus roughly correspond to either subject or object in English. In other words, the distinction of subject is not really important in the Arapaho language, especially with transitive verbs, and a relationship that is based on obviation would mark the dependencies more clearly. In response to this, the current dependency scheme adopted the UD dependency of *nsubj* and *csubj* with the additional marker *:obv* to index the obviative arguments of intransitive verbs expressed in the verbal morphology. The proximate counterparts are not marked. In the example (7), the obviative noun agreeing with the verb is such subject.

(7)

$$\overset{\frown}{\boxed{\text{nsubj:obv}}}$$

no'useeni3 nuhu' koo'ohwuu.
VAI   DET   NA

no'usee-ni3 nuhu' **koo'ohw-uu**.
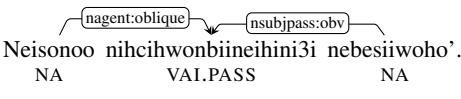arrive-4S this coyote-obv.
"This coyote came."

Similarly, the *nsubj* and *csubj* dependency is also used for animate arguments of transitive inanimate verbs (VTI) and inanimate arguments of intransitive inanimate verbs (VII). However, transitive verbs exhibit a double marker on indicating both the proximate and obviative participants, as well as the direction of action (agential relationship) between the two. The proximate participant can be either agent or patient, as can the obviative participant. So, an additional label employing the semantic distinction, *nagent* (nominal agent) is introduced.

(8)

$$\overset{\frown}{\boxed{\text{nagent:obv}}}$$

hiniisonoon heenei'itowuuneit.
NA   VTA

hi-niisonoon heen-ei'itowuun-eit.
3S-father.obv REDUP-tell s.o.-4/3S

His father tells him.

174

UD	Arapaho Dependencies	Notes
nsubj	nsubj(:obv)	Nominal subjects of VII, VAI, and VTI verbs.
csubj	csubj(:obv)	Clausal subjects of VII, VAI, and VTI verbs.
nsubjpass	nusubjpass(:obv)	✓
csubjpass	csubjpass(:obv)	✓
	agent	Proximate agent of VTA expressed by a noun.
✗	nagent:obv	Obviative agent of VTA expressed by a clause.
	nagent:oblique(:obv)	Oblique agents of passive verbs
	cagent	Proximate agent of VTA expressed by a clause.
✗	cagent:obv	Obviative agent of VTA expressed by a clause.
	cagent:oblique(:obv)	Oblique agents of passive verbs
	dobj	Inanimate nominals as objects of VTI
dobj	dobj:under	Animate proximate nominals, undergoers of VTA
	dobj:under:obv	Animate obviative undergoers of VTA
iobj	iobj	Secondary objects of VTA not expressed in the verb
ccomp	ccomp	Additional specification of dependency (e.g., dobj, dobj:under, iobj, nmod) is required
xcomp	✗	
nmod	nmod	Adjuncts of verbs
	nmod:impobj	Implied objects of VAI.O, VAI.T, and incorporated verbs
	nmod:objad	Objects of adverbial particles and some verbal prefixes
	nmod:instr	Objects of instrumental particles and instrumental verbal prefixes

Table 1: Mapping of the UD argument labels and Arapaho nominal argument labels.

The following example further demonstrates the mismatch between subject and agent in Arapaho. Here, the verb is in passive voice, and the "subject" of the verb is "my grandfathers." However, this "subject' is obviative, and it is the oblique agent ("my father") which is proximate.

(9)

nagent:oblique ⟶ nsubjpass:obv

Neisonoo nihcihwonbiineihini3i nebesiiwoho'.
NA   VAI.PASS   NA

ne-isonoo nih-cih-won-biin-**eihi-ni3i**
1S-father PST-to here-ALLAT-give-PASS-4PL
ne-besiiwoho'
1S-grandfathers.obv

"My grandfathers were given (sth) by my father"

Since the verb is passivized and thus intransitive, only one argument is reflected in its morphology, the obviative subject *nebesiiwoho'* "my grandfathers." The label of *nagent* is kept with an additional marker *:oblique* to indicate that the argument *neisonoo* "my father" is not expressed in verbal morphology. Importantly, such oblique agents are different from noun modifiers, which are discussed further in the paper, because they specify the actor of the verb rather than its manner.

The subject relationship is not clearly defined in the Arapaho language. Instead, it is possible to talk about nominal expressions that are indexed by verbal morphology either as sole arguments of

(syntactically) intransitive verbs or agential arguments (proximate or obviative) of the transitive animate verbs. We propose to account for this distinction as well as the distinction in obviation, which is clearly marked in nominal and verbal morphology.

3.2 Objects

The prototypical objects of transitive verbs do not easily fit the *dobj* relation in Arapaho. This is primarily because Arapaho verbs commonly undergo complex secondary derivation to produce verb stems which allow one to promote an animate argument to a core argument, marked inflectionally on the verb disregarding its semantic role. Thus, benefactives, recipients, goals, and even themes are typically the "object" marked inflectionally on the verb. Conversely, other arguments that would be classic "direct" objects in English are demoted, and not marked inflectionally on the verb. On the other hand, because the promoted animate argument is marked inflectionally, it can also easily be dropped from overt mention in the sentence, while unmarked items are much more likely to be mentioned explicitly.

Thus, when the manual for universal dependencies notes that *dobj* is the most patient-like argument of a verb, this is in direct tension with the tendencies of Arapaho transitive verb dependencies.

Additionally, when it notes that "if there is just one object, it should be labeled *dobj*, regardless of the morphological case or semantic role that it bears" (UniversalDependencies.org, 2014), this raises additional problems, since the actual 'object' marked on the Arapaho verb is highly likely not to appear in the sentence. The only exception and full correspondence to the UD's definition of the direct object is the inanimate object of an inanimate transitive verb (VTI):

(10)

niico'ontonounowoo nuhu' niinen.
VTI　　　　　　 DET　 NI

nii-co'on-tonoun-owoo nuhu' **niinen**.
IMPF-always-use-1S　 this　 piece of fat

"I always use this fat."

Because the verb is transitive inanimate, it requires two arguments, only one of which (the animate agent) is marked inflectionally. The second argument can only be expressed by an inanimate noun and can either precede or follow the verb. So the overt nominal in the example above represents a prototypical direct object for transitive inanimate verbs. Meanwhile, transitive animate verbs can have up to three arguments (e.g., ditransitive verbs), with the two animate arguments being expressed inflectionally on the verb. So technically, ditransitive constructions may have only one overt nominal not corresponding to either of the person markers in verbal inflection. According to the UD definition cited above, such a nominal should be considered a direct object. In the following example, the true "object" of the Arapaho verb is "you," (since it is in imperative form) while "your eyes" is not marked on the verb, and is thus from the perspective of Arapaho grammar an oblique form.

(11)

Cihneeneeciihi hesiiseii.
VTA　　　　　 NI

Cih-nee-neeciih-i　　　　　　 **he-siiseii**
EMPH.IMPER-REDUP-lend-1S.IMPER 2S-eyes

"Lend me your eyes."

There is no direct agreement between the secondary object *hesiiseii* "your eyes" and the verb. Ideally, this should be represented by *iobj* relation which emphasizes the indirect syntactic relation between the verb and the nominal.

Furthermore, objects of a transitive animate verb (VTA) and transitive inanimate verb (VTI) are different from the point of view of the grammar[4] and their respective part of speech designation[5]. Hence, a further specification of the *dobj* is necessary for transitive verbs. To stay consistent with the labels proposed for the *nagent* and *cagent* relations, the additional labels employed are *:under* and *:obv*.

(12)

Neisonoo nihcihwonbiinoot nebesiiwoho'.
NA　　　 VTA　　　　　　 NA

ne-isonoo nih-cih-won-biin-oot
1S-father PST-to here-ALLAT-give-3S/4
ne-besiiwoho'
1S-grandfathers.obv

"My father came to give [me] to my grandfather."

In the example (12), the object clearly marked on the verb is the fourth person, or the obviative. Specifying this dependency relation disambiguates between the nominals and enables the correct translation of the sentence.

So, in the current scheme the distinction between different types of objects is further clarified. The *iobj* is reserved only for the secondary objects of the ditransitive verbs which show no verbal agreement. Meanwhile, the *dobj* is used to mark the dependency relation between the transitive inanimate verb and its object, which is also not specified in the verbal morphology. Label *dobj:under* with the additional specification of obviation indicates the dependency relation between transitive animate verbs and the undergoers specified in the verbal morphology.
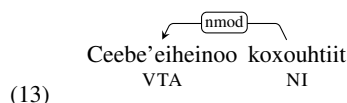
3.3 Noun Modifiers

The dependency relation of noun modifier corresponds rather well to the noun modifiers in Arapaho. It is primarily used for the disambiguation between direct or indirect objects of transitive verbs and the implied, incorporated objects, or otherwise, adjuncts.

Having argued that some overt nominals of transitive animate nouns play a role of a secondary, or indirect object, we now also argue that such label in the same context can be inappropriate as well. Using the UD rules for distinguishing the dependency in the example below would lead to analyzing the nominal *koxouhtiit* "handgame" as a direct object of the main verb. But as one can see from

---

[4] VTI objects are not reflected in verbal morphology.

[5] Only the animate nominal expressions (NA) can be the objects of VTA.

the translation, it would also lead to a wrong analysis. Similarly, the indirect object analysis would also be incorrect. Indeed, annotating this noun as an oblique or an adjunct, *nmod*, is the only way of ensuring the correct analysis and translation.

(13)
```
        ┌──nmod──┐
Ceebe'eiheinoo koxouhtiit
     VTA          NI
```
ceebe'eih-einoo **koxouhtiit**.
IC.beat-3S/1S   handgame

"He beats me in handgame."

When adjuncts are used with semi-transitive verbs, the *nmod* relation is further suffixed with *:objim* to note that the noun modifier further specifies the under-specified objects of semi-transitive verbs. Essentially, while these nominals are analyzed and marked as noun modifiers, for a successful translation they need to be marked as direct objects, which we have argued against in the previous section. In order to avoid the incorrect translation as well as incorrect analysis, the label *nmod:objim* is used. In the following example, the noun *bei'ci3ei'i* "money" semantically is the object of the semi-transitive verb. However, as we argue, marking it as direct or indirect object would violate the principles of Arapaho syntax.
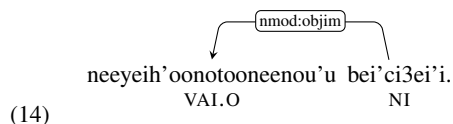
(14)
```
           ┌────nmod:objim────┐
neeyeih'oonotooneenou'u bei'ci3ei'i.
        VAI.O                  NI
```
neeyeih-'oonotoonee-nou'u   **bei'ci3ei'i**.
IC.try-REDUP-borrow things-12.ITER money

"Whenever we try to borrow money."

In addition, some of these implied or incorporated objects with overt nominal expressions can be modified by an adverbial particle similar to a preposition in English.
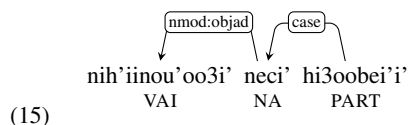
(15)
```
     ┌──nmod:objad──┐ ┌─case─┐
nih'iinou'oo3i' neci' hi3oobei'i'
     VAI         NA      PART
```
nih-'iinou'oo-3i'   **nec-i'**   **hi3oobei'-i'**
PST-float around-3PL **water-LOC** **under sth-LOC**

"They were floating around under the water"

In the example above, particle *hi3oobei'i'* "under" is a dependent of the adjunct *neci'* "water-LOC." This relation is reflected in the locative case marker on the noun showing a direct dependency with the particle. The Arapaho dependency scheme additionally distinguishes the instrumental case since there are special case markers defined by an adverbial or an adverbial prefix. So where the prefixes *hi'-, nohk-,* and *nii3-* are present or where the corresponding adverbial particles appear, the nominal adjunct is considered to be instrumental (*nmod:instr*). So in example (5), the relation between the head of the relative clause *si-ikoocei'iikuu3oo* "rubber item" and the main verb is *nmod:instr*.
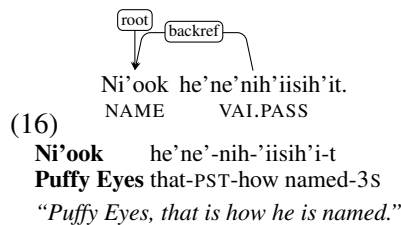
Finally, an additional dependency *poss*, possessor modifier, is being used for possessive constructions with an overt possessor. Similar to Finnish (Tsarfaty, 2013; Haverinen et al., 2014), in Arapaho, it is possible to distinguish between the subject and the object of a possession. However, unlike in other languages, no special genitive construction exists to mark this type of relation. Instead, the possessor and possessed appear side-by-side. The possessed in such constructions has a third (or fourth) person possessive marker. So in a phrase *nii'ehihi' hi-siiseii* "little bird his-eyes" the possessor is "little bird" since the possessive prefix *hi-* "his" directly references this third person. The dependency relation marked here is possessor nominal modifying another nominal.

The examples above demonstrate that not all of the arguments that may semantically appear similar to the dependencies established in UD are the same in Arapaho. While under-specification of the semantic relationships can be beneficial in establishing some commonalities cross-linguistically, it can also result in misrepresentation of some of the relations and lowered efficiency in machine learning (Lipenkova and Souček, 2014). The major underspecification for the Arapaho language is the omission of proximate-obviative distinction: while we realize that it could potentially be problematic in cross-linguistic applicability, omitting this distinction disregards one of the main features of Algonquian syntax, and renders automatic translation of English transitive verbs into Arapaho effectively impossible.

4 Non-verbal roots

Adopting the relation of a root as the independent word in a clause or sentence allows us to avoid issues arising from securing the root node with verbs. So, like in the UD scheme, our annotations do not attach the node of a root to a particular part of speech even though they are usually

represented by verbs. The main reason for doing this is avoiding the potential analysis of what is not there (Nivre, 2015; Hajicova et al., 2015; Osborne and Liang, 2015). In our annotations, the *root* often represents a pragmatically independent word, as for example in predicative type constructions (Cowell and Moss, 2008). Such constructions are used to topicalize one of the verbal arguments or the manner of action (i.e., verbal particles) similar to existential constructions in other languages. However, instead of marking the predicate as a root of the sentence as it is done in the Russian TreeBank (de Marneffe et al., 2014), the topicalized nominal or the particle is the root in Arapaho. The relation between the root and the predicate is *backreference*:

(16)

Ni'ook he'ne'nih'iisih'it.
NAME   VAI.PASS

Ni'ook   he'ne'-nih-'iisih'i-t
Puffy Eyes that-PST-how named-3S

"Puffy Eyes, that is how he is named."

In example (16), the argument of the verb *he'ne'nih'iisih'it* "that is how he is named" is not realized overtly, and the verbal prefix *ne'-* "that is" references back to the topical argument, making the verb actually a dependent of it. Were we to analyze distinct morphological elements, this prefix would act as a copula between the two. Overall, the reasoning for treating such topicalized elements (which sometimes may take other than the clause-initial position) comes from the combination of the pragmatics and morphology: nearly all of the verbal clauses with prefixes *ne'-* "that" and *nee'ees-* "that is how" are backreference dependents of such roots.

5 Conclusion

In this paper we demonstrate the use of Universal Dependency scheme with a language typologically different from the ones often included in machine-learning technologies. In using the UD framework, several unmentioned issues stemming from the reliance on the word order were noticed. For example, in the current annotation scheme, we reanalyzed the relationship of *parataixis* to account for verbs of citations, so that dependency would be traced from such a verb to the root of the whole clause. Similarly, the discourse marker dependencies were modified to include and analyze interjections. Unfortunately, it is outside of the scope of the paper to discuss these issues, but we hope that expanding this project to annotating conversational data and applying the annotated data to machine learning methods will further reveal some additional insights on analysis of discontinuous constituency.

In critiquing the UD, we, nonetheless, want to stress the eloquence of such an approach. Unlike the phrase structure annotations, UD allows us to account for the inconsistent phrase structures and dislocated tokens so often encountered in the Arapaho language. At the same time, however, we argue that to adequately account for the many linguistic nuances in annotations of such a morphologically and syntactically complex language like Arapaho, it is often necessary to include the semantic and pragmatic levels of analysis.

Acknowledgments

References

Andrew Cowell and Alonzo Moss. 2008. *The Arapaho Language*. Westview Press, Boulder.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. Stanford typed dependencies manual. Revised: April 2015:1–22.

Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal Stanford Dependencies: A cross-linguistic typology. *In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4585–4592.

Ives Goddard. 1984. The obviative in Fox narrative discourse. In William Cowan, editor, *Papers of the Fifteenth Algonquian Conference*, pages 273–286. Carleton University Ottawa.

Eva Hajicova, Marie Mikulova, and Jarmila Panevova. 2015. Reconstructions of Deletions in a Dependency-based Description of Czech: Selected Issues. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 131–140, Uppsala, Sweden.

Katri Haverinen, Jenna Nyblom, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Anna Missilä, Stina Ojala, Tapio Salakoski, and Filip Ginter. 2014.

Building the essential resources for Finnish: the Turku Dependency Treebank. *Language Resources and Evaluation*, pages 1–39.

Janna Lipenkova and Milan Souček. 2014. Converting Russian Dependency Treebank to Stanford Typed Dependencies Representation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 143–147, Gothenburg, Sweden. Association for Computational Linguistics.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency Annotation for Multilingual Parsing Ryan. In *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 92–97, Sofia, Bulgaria.

Joakim Nivre. 2015. Towards a universal grammar for natural language processing. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9041:3–16.

Timothy Osborne and Junying Liang. 2015. A Survey of Ellipsis in Chinese. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 271–280, Uppsala, Sweden.

Reut Tsarfaty. 2013. A Unified Morpho-Syntactic Scheme of Stanford Dependencies. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 578–584.

UniversalDependencies.org. 2014. Universal dependency relations (single document). `http://universaldependencies.org/u/dep/all.html`.

Appendix: Abbreviations

0PL	inanimate plural
12	first person plural inclusive
3s/4 or 1PL/2PL	the first number indicates the person and number acting on the following person and number (he to him; we to you.pl)
ALLAT	allative
DET	determiner
DETACH	detached adverbial prefix
DIM	diminutive
EMPH	emphatic
FUT	future tense
IC	phonological initial change
IMPER	imperative
IMPF	imperfect
ITER	iterative
LOC	locative
NA	animate noun
NAME	name
NARRPAST	narrative past tense
NI	inanimate noun
PART	particle
PASS	passive voice
PL	plural
PST	past tense
REDUP	reduplication
REL	relative prefix
S	singular
SELFBEN	self-benefactive
VAI	animate intransitive verb
VAI.O	animate intransitive verb with an implied object
VAI.PASS	animate intranstive passive verb
VAI.T	animate intransitive verb with a specific implied object
VII	inanimate intransitive verb
VTA	transitive verb with animate object
VTI	transitive verb with inanimate object