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

Objectives

Although Universal Dependency scheme is both comprehensive and eloquent, it also assumes certain grammatical structures to be cross-linguistically universal. In our project :

- we **demonstrate the applicability** of these assumptions to annotations of the Arapaho language;
- and **propose language-specific guidelines** to account for particular features of Algonquian syntax.

Arapaho Language

Figure 1: Traditional Arapaho territory
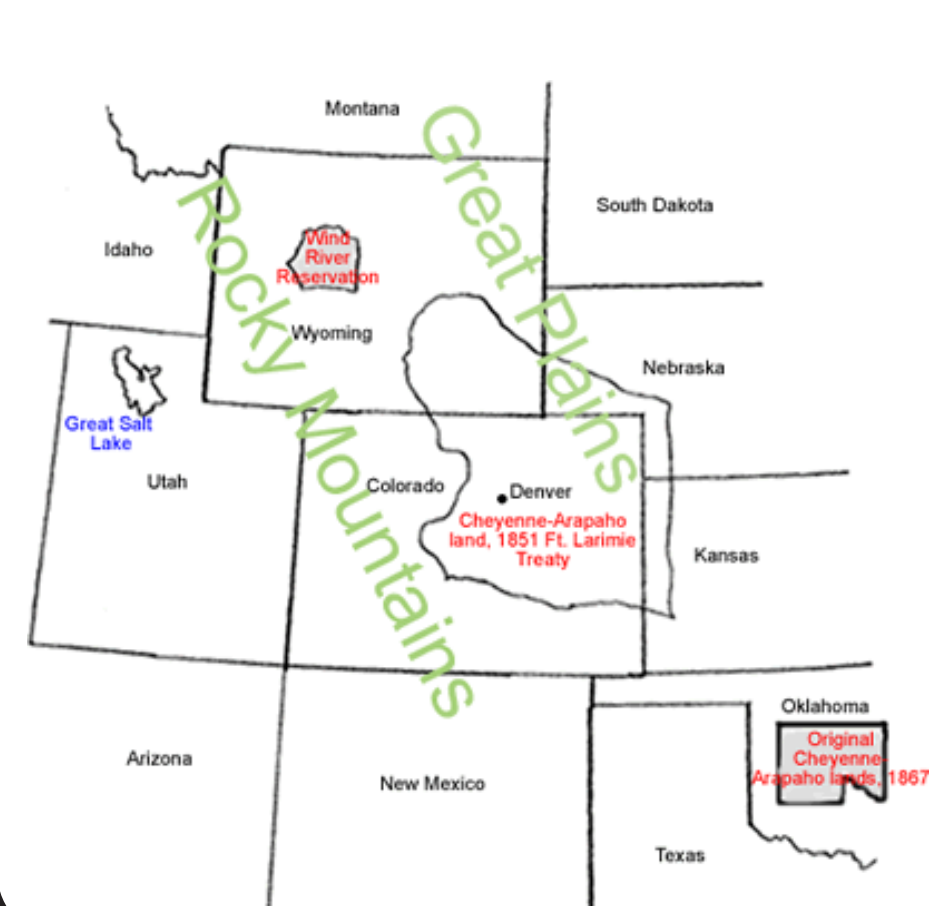
- Algonquian.
- Poly-synthetic agglutinating language.
- Critically endangered.
- Spoken in Wind River Indian Reservation in Wyoming, USA.
- Available transcribed and translated spoken corpus.

Problematic Features of Arapaho

Verbal Complexity

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

Transitivity

- Difference in syntactic and semantic transitivity is reflected in verbal morphology [2].
- All of the verbs may have two nominals they seem to agree with in meaning or none at all.

(1) nih-to'ow-oot.
PST-hit-3s/4
"S/he hit him/her"

(2) nih-'iikooko'uyei-3i' biino.
PST-pick things-3PL chokecherries
"They picked chokecherries"

Obviation

- When two or more third person animate nouns are present, one of them, the less pragmatically important one [5], is marked obviative, the verb agreeing with such noun is also marked.

(3) hee3eihok hiinoon 3eeyokooxuu.
say to s.o.-4/3s.subj his/her mother Tipi-pole Child
His mother said to Under-the-Tipi-Pole Child.

Annotations

- Manual annotations by graduate students in Linguistics.
- Kept in spreadsheet format.
- Checked by a fluent non-native speaker.
- Two phases of annotations:

Phase 1: **Initial annotations of traditional narratives.**

- About 2,000 lines.
- Patterns of Algonquian syntax used for the guidelines.

Phase 2: **Additional narratives and conversations.**

- Stanford Dependencies [4] used as the base, annotations converted.
- + 3616 lines of elicited personal and traditional oral narratives
- + 593 lines of conversational data.
- Universal Dependency [7] is adopted to account for conversational disfluencies.
- Previous annotations were converted according to the UD scheme.

Mapping to the UD scheme

- Out of 40 UD dependencies, **30 have one-to-one correspondence.** For example, the adverbial clause dependency:

(4)
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."

- Added **17 Arapaho-specific** dependencies.
- Some dependencies were eliminated because they do not exist (e.g., *xcomp, neg, amod, nummod, aux, auxpass, cop, expl, mark, mwe*).

Subjects

- Subjects are not distinguished syntactically by word order.
- Obviation by itself also does not specify syntactic dependency but disambiguates between nominals.
- Only intransitive animate (VAI), intransitive inanimate (VII) and transitive inanimate verbs (VTI) exhibit agreement with syntactic subjects.

(5)
no'useeni3 nuhu' koo'ohwuu.
VAI   DET   NA

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

- Because transitive animate verbs (VTA) are marked for obviation and the direction of action, the agent distinction (*nagent(:obv)*) is used instead:

(6)
hiniisonoon heenei'itowuuneit.
NA   VTA

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

- When transitive verbs passivized, semantic agents may remain in the clause as obliques (*nagent:oblique*):

(7)
Neisonoo nihcihwonbiineihini3i nebesiiwoho'.
NA   VAI.PASS   NA

ne-isonoo nih-cih-won-biin-eihi-ni3i ne-besiiwoho'
1s-father PST-to here-ALLAT-give-PASS-4PL 1s-grandfathers.obv
"My grandfathers were given (sth) by my father"

Objects

- Almost any semantic role can be expressed by an "object."
- Any "classic" object can be demoted and not marked verbally.
- Overt nominal expressions of an object are not always necessary.
- The only true direct object is the inanimate object of a VTI:

(8)
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."

- Ditransitive constructions verbally mark only animate nominals, so the overt inanimate nominal is actually an indirect object:

(9)
Cihneeneeciihi hesiiseii.
VTA   NI

Cih-nee-neeciih-i he-siiseii
EMPH.IMPER-REDUP-lend-1s.IMPER 2s-eyes
"Lend me your eyes."

- Semantic label "undergoer" is used to further specify the direct objects of transitive animate verbs:

(10)
Neisonoo nihcihwonbiinoot nebesiiwoho'.
NA   VTA   NA

ne-isonoo nih-cih-won-biin-oot ne-besiiwoho'
1s-father PST-to here-ALLAT-give-3s/4 1s-grandfathers.obv
"My father came to give [me] to my grandfather."

Non-Verbal Roots

- *root* is a pragmatically independent word, not a particular part of speech.
- In predicative constructions, the *root* is the topic, while the predicate is labeled as *backreference*

(11)
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."

Noun Modifiers

- *nmod* relation is used to disambiguate between objects of transitive verbs and objects of intransitive verbs.

(12)
Ceebe'eiheinoo koxouhtiit
VTA   NI

ceebe'eih-einoo **koxouhtiit**.
IC.beat-3s/1s   handgame
"He beats me in handgame."

- Specification *:objim* (implied object) is used with objects of semi-transitive verbs:

(13)
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."

- Implied and incorporated objects can be modified by an adverbial participle, like prepositions, and are labeled as *:objad* – object of an adverbial:

(14)
nih'iinou'oo3i' neci hi3oobei'i'
VIAU   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"

- *nmod:instr* (instrumentals) are introduced by instrumental verbal prefixes and corresponding particles: *hi'-, nohk-,* and *nii3-*.
- *nmod:poss* marks the possessor:

(15)
nii'ehihi' hi-siiseii
NA   NI

nii'eihihi' hi-siiseii
little bird' 3s-eyes
"Little bird's eyes"

Future Research

- More manual annotations of conversational data.
- Double annotations to ensure inter-annotator agreement.
- Developing POS and dependency correspondences.
- Subject this scheme to machine learning.

Conclusions

1. Dependencies of nominal arguments should not rely purely on syntax but also include semantics and pragmatics.
2. Under-specification of semantic relationship for Arapaho leads to misrepresentation of some dependency relationships.
3. The guidelines developed for the Arapaho language demonstrate the use of both the universal patterns of syntax and the language-specific ones.

References

[1] *Universal Dependency Annotation for Multilingual Parsing*, Sofia, Bulgaria, 2013.

[2] A. Cowell and A. Moss. *The Arapaho Language.* Westview Press, Boulder, 2008.

[3] M.-C. de Marneffe, T. Dozat, N. Silveira, K. Haverinen, F. Ginter, J. Nivre, and C. D. Manning. Universal Stanford Dependencies: A cross-linguistic typology. *In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14),* pages 4585–4592, 2014.

[4] M.-C. de Marneffe and C. D. Manning. Stanford typed dependencies manual. Revised: April 2015:1–18, 2008.

[5] I. Goddard. The obviative in Fox narrative discourse. pages 273–286, 1984.

[6] J. Nivre. 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, 2015.

[7] UniversalDependencies.org. Universal dependency relations (single document), 2014.

Acknowledgments