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

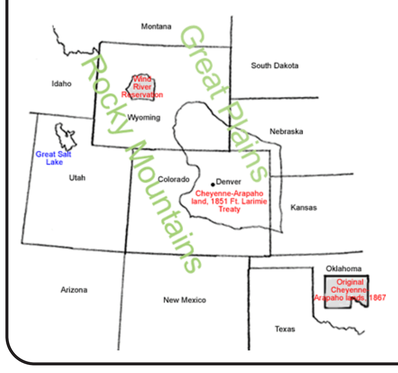
Objectives

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

- 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.
- 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 nominals are present, one of them, the less pragmatically important one, is marked obviative, the verb agreeing with such noun is also marked.

(3) hee3**eihok** **hiinoon** 3eeyokooxuu.
say to s.o.-4/3S.SUBJ his/her mother Tipi-pole Child
His mother said to Under-the-Tipi-Pole Child.

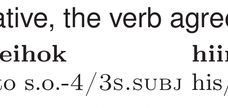
Annotations

- 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 (de Marneffe and Manning 2008) used as the base, annotations converted.
  - + 3616 lines of elicited personal and traditional oral narratives
  - + 593 lines of conversational data.
  - Universal Dependency 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:

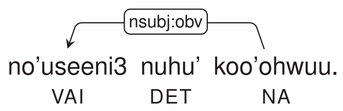
(4)


Bih'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*).

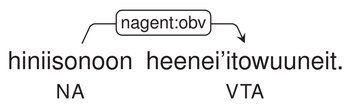
Subjects

- 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)


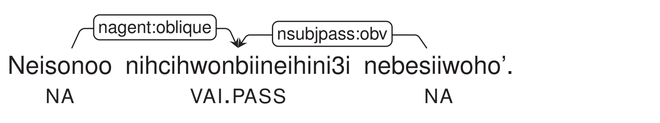
no'usee-ni3 nuhu' **koo'ohw-uu**.
arrive-4S this coyote-obv.
"This coyote came."

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

(6)


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

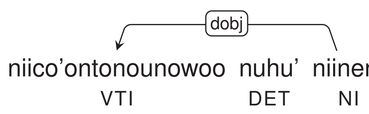
- When transitive verbs passivized, semantic agents may remain in the clause as obliques (*nagent:oblique*):

(7)


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"

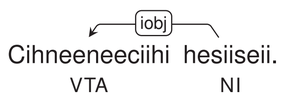
Objects

- 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)


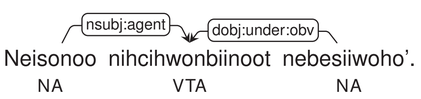
nii-co'on-tonoun-owoo nuhu' **niinen**.
IMPF-always-use-1S this piece of fat
"I always use this fat."

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

(9)


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

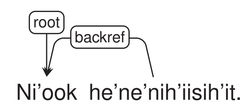
- Semantic label "undergoer" is used to further specify the direct objects of transitive animate verbs:

(10)


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

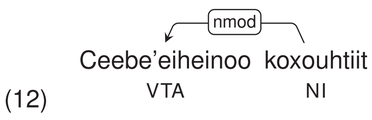
- *root* is a pragmatically independent word, not a particular part of speech.
- In predicative constructions, the *root* is the topic, while the predicate is labeled as *backreference*

(11)


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

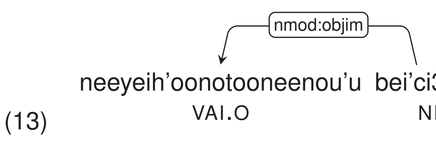
Noun Modifiers

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

(12)


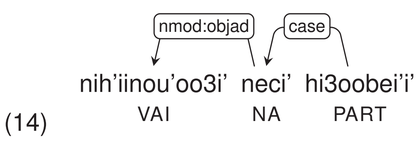
ceebe'eih-einoo **koxouhtiit**.
IC.beat-3S/1S handgame
"He beats me in handgame."

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

(13)


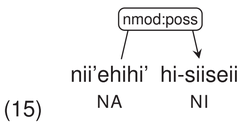
neeyeih-'oonotoonee-nou'u **bei'ci3ei'i**.
IC.try-REDUP-borrow things-12.ITER money
"Whenever we try to borrow money."

- 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'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'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

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.

Acknowledgements

The research is supported by the National Endowment for the Humanities grant, project number 1551671 "Arapaho Lexical Database and Dictionary." We thank the Northern Arapaho tribe for allowing us to conduct the work with their language.