Towards a Hän morphological transducer

Maura O'Leary[1], Joseph Lukner[2], Finn Verdonk[2], Willem de Reuse[3], Jonathan Washington[2]

[1]Western Washington University
[2]Swarthmore College
[3]The Language Conservancy

Correspondence: Maura.O'Leary@wwu.edu, Jonathan.Washington@swarthmore.edu

First published `lexd` morphological transducer for a Dene language

Today:

- About the Hän language
- What a morphological transducer is
- Challenges in creating a morphological transducer for a Dene language
- Creating a guesser
- Evaluation
- Next steps & applications

The Hän language

The Hän language

- Dene / (Northern) Athabaskan
- Spoken around Eagle, Alaska, and Dawson City, Yukon Territory

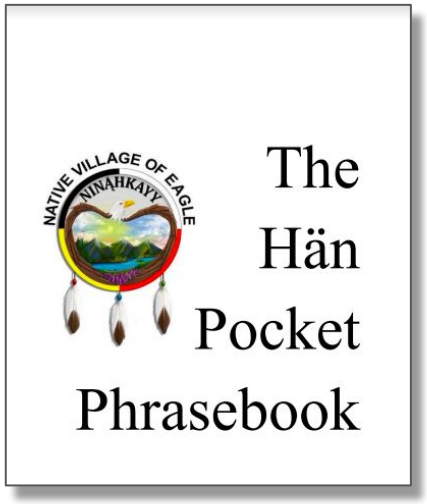

Indigenous Peoples and Languages of Alaska
This map shows the indigenous language regions of Alaska. Related languages of neighboring Canada and Russia are also shown. The language boundaries represent traditional territories at approximately 1900, though some shifts in language boundaries have occurred since that time. Boundaries are defined based on the similarity of sound systems and the ability of speakers from different regions to understand each other.
The colors of the individual languages reflect their classification into language families, each of which share a common ancestral language. Eighteen of the twenty indigenous Alaska languages on this map belong to either the Inuit-Yupik-Aleut or the Athabascan-Eyak-Tlingit families. Tsimshian is a member of the small Tsimshianic family, while Haida is an isolate, not demonstrably related to any other language in the world.
The language names appearing on this map are the English names generally accepted by most speakers today. Alternate names are listed in the legend. A selection of modern and historic indigenous villages are labeled with both indigenous and English names, as are major rivers, lakes, and islands. A comprehensive inventory of indigenous place names, including village sites and geographic features, would number in the tens of thousands.
As of 2010 few indigenous languages in Alaska are still spoken by children, but significant revitalization programs exist for some languages.
Although based largely, including language definitions, boundaries, and insets, on Michael Krauss's 1974 and 1982 Native Peoples and Languages of Alaska map, this map differs in several ways from that map. This map does not indicate indigenous populations or speaker numbers; language status in each village or size of village; and dialect boundaries. Finally, several language names have been revised to reflect current usage.
More information about Alaska languages and language maps can be found at www.uaf.edu/anla.
Table of Language Groups
Language Family
Language Name(s)
Aleut
Unangax̂, Unangan, Unangas, Unangam Tunuu, Aleut
Inuit–Yupik
Alutiiq, Sugpiaq
Central Yup'ik, Cup'ik, Cup'ig
Naukan Yupik
Siberian Yupik, St. Lawrence Island Yupik
Sirenik
Inupiaq, Iñupiaq
Tsimshianic
Tsimshian
Haida
Haida
Tlingit
Tlingit, Łingít
Eyak
Eyak
Athabascan
Ahtna
Dena'ina, Tanaina
Deg Xinag, Deg Hit'an
Holikachuk, Innoko
Koyukon, Denaakk'e
Upper Kuskokwim, Kolchan
(Lower) Tanana
Tanacross
Upper Tanana, Nabesna
Han
Gwich'in, Kutchin, Tukudh
Other Athabascan
Arctic Ocean
Chukchi Sea
Chukotka
Anadyr Gulf
Bering Sea
Norton Sound
Bristol Bay
Gulf of Alaska
Aleutian Islands
Inupiaq
Naukan Yupik
Sirenik
Siberian Yupik
Koyukon
Holikachuk
Deg Xinag
Upper Kuskokwim
Central Yup'ik
Dena'ina
Tanana
Tanacross
Ahtna
Upper Tanana
Gwich'in
Han
Eyak
Alutiiq / Sugpiaq
Unangax̂
Tlingit
Haida
Tsimshian
Hare
Northern Tutchone
Southern Tutchone
Kaska
Tagish
Tahltan
Russia
United States
United States
Canada
ANLC

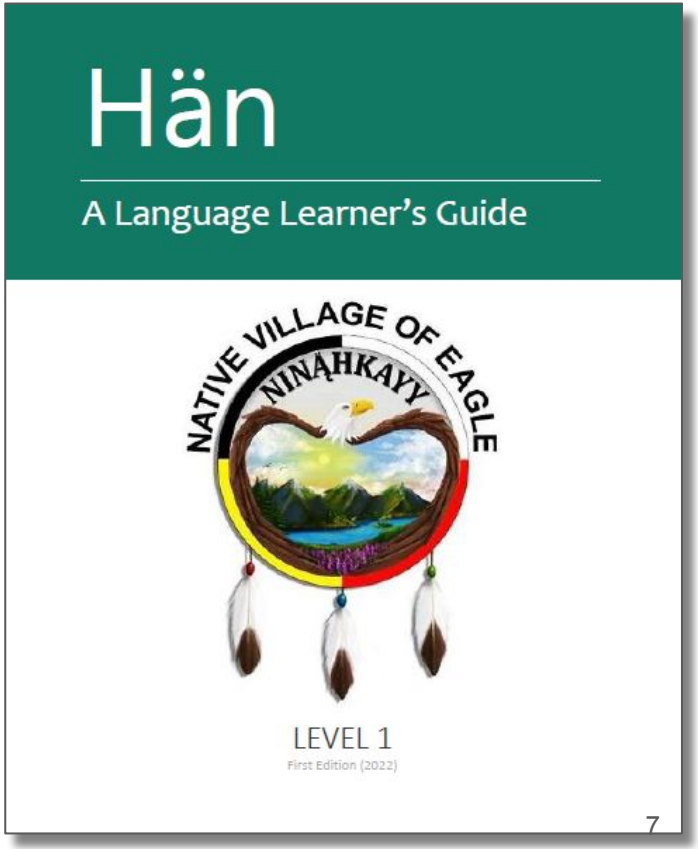
Critically endangered

- Only 5 remaining speakers, all over 70 years old
- No longer used on a daily basis
- However, children and grandchildren of the remaining speakers now want to learn and revitalize the language

Previously, in Hän revitalization

- Flashcards, phrasebook, textbook





Previously, in Hän revitalization

Language lessons

Where our data comes from

~2300 verb paradigms
elicited by Willem de Reuse
between 2006-2012

Where our data comes from

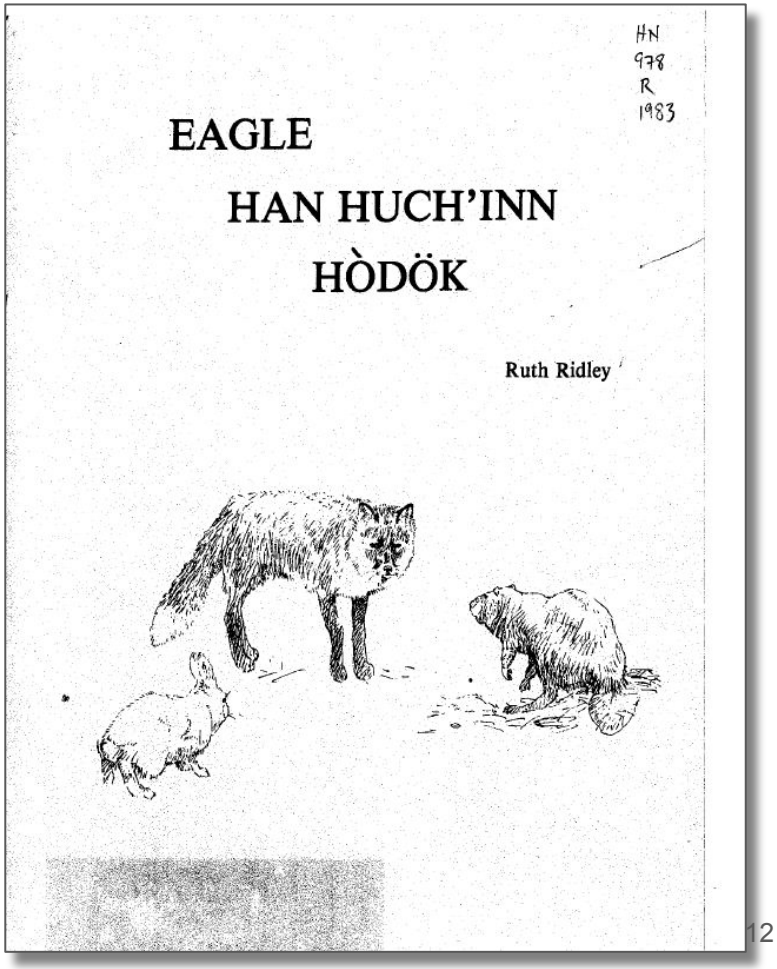
SEE, look at O, 3i O, A		
Imperfective mode, O, Eagle dialect, <see also Customary mode> [LP: falling stem tone]		
nök-'įį, LP, RR, TM, nök-'įh, SM, CS, CSi	tr'ënoh'įį, LP, RR, tr'ënowh'įį, EBU, tr'ënoh'įh, CSi	Notes: '1 saw 3, 1 am seeing 3, 1 am looking at', RR, TM: sentex EBU: '1p are seeing 3', sentex
nąh'įį or näh'įį, RR, näh'įį, CS, EBU, TM, näh'įh, CSi	näh'įį, LP, RR, EBU, hinäh'įh, CSi	Notes: rck pl. CSi, sentex
Onoh'įį, LP, TM, Onäh'įh, CSi, yënoh'įį, RR, BU, yinoh'įį, EBU, 3-3o, (yë)no'į', CS	Ohënoh'įį, LP, RR, hinowh'įį, EBU, Ohënäh'įh, CSi	Notes: CSi: Imperfective and Perfective confused in 3!; EBU: '3p are seeing 3', RR form gotten twice, sentex

Where our data comes from

Additional data elicited between 2016 and 2022 by Maura O'Leary and Blake Lehman

Where our data comes from

Short stories written by Ruth Ridley
(the youngest remaining speaker)



Transducer

What is a morphological transducer?

- A finite-state model of a language's morphology
- Performs:
  - **analysis**: valid forms of a language receive one or more morphological analyses
  - **generation**: and a valid form is output when an analysis is input

noh’ɨɨ<v><tv><perf><s_1pl><o_3pl>

generation ↓ ↑ *analysis*

hutr’ënäh’ì’

- Preferable to ML approaches (Butt 2020), no large dataset available anyway

5 Challenges in creating a transducer for a Dene language

#1 Prefixational morphology

- Traditional approach: `lexc`
  - Continuation lexicons
  - Ordering of morphemes and tags the same by default
  - Really hard to do non-suffixational morphology

- Hän: a prefixational language
  - To get prefixes, previous approaches have used: flag diacritics, intricate continuation lexicons, simplified “zones” of morphology (Harrigan et al., 2017; Arppe et al., 2017; Holden et al., 2022)
  - More complex code (less maintainable), increased transducer size, slow compilation, slow runtime speeds

Kyrgyz:

канат-тар-ыбыз-дан

wing-PL-POSS.1PL-ABL

‘From our wings’

канат<n><pl><px1pl><abl>

Hän:

hu-tr’ë-n-oh-’įį

3plO-1plS-theme-1plS-see.IMPF

‘We see them.’

noh’įį<v><tv><impf><s_1pl><o_3pl>

rather than:

<o_3pl><s_1pl>noh’įį<v><tv><impf>

#1 Prefixational morphology

- Our approach: `lexd` formalism, designed for non-suffixational morphology
  - Uses patterns (see below), rather than continuation lexicons

Verb Template:

Disjunct prefix	Pl. Subj.	Object	Deictic Subject	Reflexive	Directive	Future/ Inceptive	Gender/ Qualifier	Theme	Conjugation Marker, Subject, Classifier	**Stem**

Can be easily rendered as a pattern in `lexd`:

```
(VerbStem-Tv(1) subject(1) object?(1) subject(2) object?(2) :VerbStem-Tv(2)
aspect(1) VerbStem-Tv(3) VerbStem-Tv(4) subject(3) [ :{NOV} ] VerbStem-Tv(5) [
<v><tv>: ] VerbStem-Tv(2): aspect(2) subject(4)
object?(3))[^[3Ssub,non3Ssub],^[impf,perf,incp,fut,opt],^[sg,pl],^[1,d,0cl,
ɬ],^[0cm,dh,gh,n]]
```

#2 Distributed morphological features

- Subject morphology is spread over three spots in the verb structure with a four-column lexicon format

	Plural Subject (3pl)	Deictic Subject (1pl)	Subject marking	Subject tags
tr'ënoh'įį 'we see'		tr'ë>	oh>	<s_1pl>
nihënoh'įį 'they are looking at us'	hë>		oh>	<s_3pl>

- Stem-specific morphology encoded in a verb lexicon of five parts

	Disjunct Prefix	Directive Prefix	Gender/Qualifier	Theme Prefix	Stem
udohkät 'ask'		u>		d>	kät
nä'aww 'eat'	nä>				'àww
jënohtlòt 'boil'			jë>	n>	tlòt

#3 Verb and subject features: Stem alternations

- Unpredictable verb stem alternations depending on aspect marker and (sometimes) plurality

Verb Gloss	Imperfective	Perfective	Future
’see’	’įį	’į̀’	’į̀ww

Verb Gloss	Singular (Imperfective)	Plural (Imperfective)
’go’	haa	jèww

- Alternations not predictable / not able to be treated as phonology

#3 Verb and subject features: Subject conjugations

- Subject markers also take different forms based on classifier, conjugation marker, aspect, number, and person

Classifier -l-

Mode: / Subject:	**Imperfective**	**0- Perfective**	**dh- Perfective**	**gh- Perfective**	**Progressive and Future**	**Optative**
1sg	ök-	ök-	dhök-	ök-	ök-	ok-
2sg	ą̈h-	ą̈h-	dhą̈h-	ą̈h-	ą̈h-	ǫh-
3sg, 1pl, 3pl	oh-	äh-	dhoh-, oh-, eh-	ą̈h-	äh-	uh-
2pl	äh-	äh-	dhäh-	äh-	äh-	äh-

#3 Verb and subject features: Matching

- We use **filter tags**, a `lexd` feature, to match verb entries to subject markers
  - Many permutations as a result - 173 entries in the subject lexicon

- Each morpheme in the subject lexicon has filters for **verb classifier, conjugation marker, aspect, person** (to a minimal extent)**,** and **number**.

```
[ł,impf,0cm,non3Ssub,sg]:ök>
```

- Each morpheme in verb lexicon has filters for **verb classifier, aspect, conjugation marker, and number**

```
nähaa:haa[0cl,impf,0cm,sg]
```

#4 Spelling variations

- Spelling is not perfectly consistent across sources:
  - Orthography developed in 1977 by linguists John Ritter and Michael Krauss with speakers Louise Paul and Ruth Ridley
  - Not all speakers are comfortable writing in the orthography
  - Linguists' elicitation notes sometimes reflect phonemes and sometimes reflect allophones
- Use the tag "Dir/LR", but keep the same lemma:
  - automatically removed from generator transducer
  - retained in analyser transducer

```
choo:choo     # "big"

choo:choh     # "big"  Dir/LR
```

```
analysis:    ^choo/choo<adj>$
             ^choh/choo<adj>$

generation:  ^choo<adj>/choo$
```

#4 Encoding variations

- Different ways sources encode characters:
  - The character ‘ą̀’ (a + ˛ + ¨ + ̀) could be encoded as:
    - ‘a’ with a series of diacritics after (in several possible orders),
    - precomposed ‘ä’ or ‘ą’ with additional diacritics added (in several possible orders)

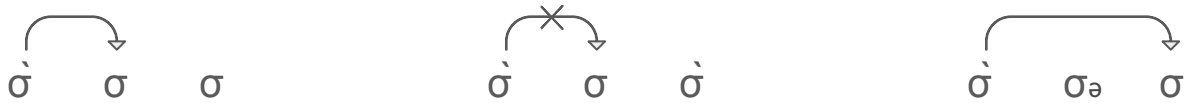
- Solution: **"spellrelax" rules** (compose-intersected with analyser)
  - consistent with NFKD and available Hän keyboard (from Yukon Native Language Centre)

```
.o. [ ?* [ i ˛ (->) į ] ?* ]          # i + ˛ treated the same as composed į

.o. [ ?* [ ˛ ¨  (->) ¨ ˛  ] ?* ]      # ˛ + ¨ treated the same as ¨ + ˛
```

#5 Tone spreading

- The pattern:
  - Underlying low tones spread to the next syllable unless spreading would create a sequence of 3 low tones
  - Spreading skips over schwas
  - Tone spreading crosses word boundaries
  - Diacritics are generally used on all pronounced low tones, not just underlying ones



#5 Tone spreading

- The challenge:
  FSTs operate at the level of the word; not easily possible to condition twol rules across a word (token) boundary

- The solution: **“spellrelax”** rules
  - Accepting a low tone diacritic (or not) on the first non-schwa syllable for any word

```
.o. [ [b|d|l|h|z|r|j|g|’|t|s|c|k|ł|w|m|n|y]* [ a (->) a ` ] ?* ]
                        #  low tone “à” allowed in first syllable in place of high tone (unmarked) “a”

.o. [ [b|d|l|h|z|r|j|g|’|t|s|c|k|ł|w|m|n|y]* [ i (->) i ` ] ?* ]
                        #  low tone “ì” allowed in first syllable in place of high tone (unmarked) “i”
```

An additional part of the implementation process: The guesser

What is a guesser

- **A version of a transducer that accepts any hypothetical verb**
- Leverages morphological patterns of transducers with regular expressions

```
[ł,0cm]: [ł,0cm]<GUESSER_ł_0cm_nthm>: [ł,0cm]:n> /([a-z’¥ˆ\]])+/[ł,0cm]
[l,0cm]: [l,0cm]<GUESSER_l_0cm_nthm>: [l,0cm]:n> /([a-z’¥ˆ\]])+/[l,0cm]
```

- A form of a verb not in the transducer:
  shënähtthee ‘you all are barking at me’

- The returned set of analyses includes the correct analysis:

```
<GUESSER_0cl_0cm>nähtthee<v><tv><impf><s_3sg><o_1sg>
/<GUESSER_0cl_0cm_nthm>tthee<v><tv><impf><s_2pl><o_1sg>
/<GUESSER_d_0cm_nthm>tthee<v><tv><impf><s_2pl><o_1sg>
/<GUESSER_ł_0cm_nthm>tthee<v><tv><impf><s_2pl><o_1sg>
```

Constraining guesser outputs based on null morphemes

- Lots of over-guessing, especially 3sg impf of 0-classifier verb
  - 3sg impf 0-classifier prefix = ∅-
  - So the entirety of any input could be analyzed as the stem (with the conjugation “∅- + input”)
- Some heuristics to the rescue:
  - Vowels and some consonant clusters don’t seem to appear at the beginning of roots
- twol rules that restrict guesser possibilities
  - using the /<= operator to exclude any path matching the pattern from the compiled transducer

```
"restrict guessed forms with vowel-initial stems"
Vowel:Vowel /<= %{NOV%}: _ ;

"no hC- or nC- initial stems guessed by guesser"
C1:C1 /<= %{NOV%}: _ Cons:Cons ;
     where C1 in ( h n ) ;
```

- Result: fewer incorrect guesses

Lexicon and evaluation[1]

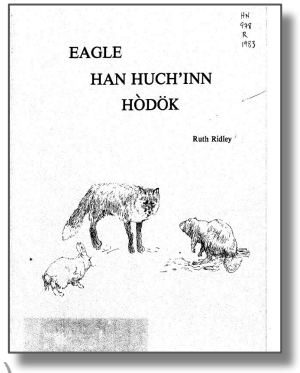
[1] All reports of code and performance are based on the code at revision `b334130`, dated 2025-01-17.

Lexicon

- current smallish lexicon:
  - for implementation of morphology
  - covers other common words in our test corpus
- "unique" excludes:
  - spelling variants
  - context-dependent stems
- lemmas:
  - currently 3sg impf
  - will transition to 1sg impf based on recent speaker judgements

part of speech	unique	total
nouns	167	183
verbs	15	64
adjectives	18	20
prepositions	15	17
adverbs	6	8
conjunctions	3	4
modal words, determiners, pronouns, numerals, anthroponyms, etc.	22	23
total	246	319

Corpora



Several texts used to evaluate transducer:

- **Short stories** written by native speaker Ruth Ridley (Ridley, 1983, 2018)
  - ~3.3k tokens
  - manually transcribed with some OCR augmentation
- **Elicited sentences** accompanying verb paradigms (de Reuse 2015b)
  - ~11.5k tokens (4.5k sentences, on average very short)
  - extracted by script
  - filtered to exclude English, author comments, organisational codes, non-sentence Hän material

Coverage

- naïve coverage:
  raw percentage of tokens analysed by the transducer, regardless of accuracy

corpus	tokens	ambiguity	coverage
stories	3275	1.08	60.40%
elicited sents	11479	1.10	21.87%

- stories corpus has much higher coverage; reasons:
  - uninflected POSs (common nouns, prepositions, etc.) less common in sents corpus, easily included in transducer
  - elicited sentences include full range of verbs vs handful of common, domain-specific verbs
  - Elicited sentences have dialect variation which is not yet integrated into transducer
- sentences corpus main data source for transducer lexicon; coverage good sign
- to improve coverage: more verbs, more spellrelax, more phonology

Size & Speed

- Compared to other Dene transducers, lexd approach appears to be:
  - much **faster**
  - much **smaller**
  - much **easier to maintain**
- Reason: less unwieldy approaches (no flag diacritics)

Generator content	
19824	states
23105	arcs
4286	analysis-form pairs

Compiled transducer size	
generator	375kB
analyser	879kB
guesser	6986kB

speed	
analysis (3.3k tokens)	125ms
compilation	(652MB)
- 1 thread	30s
- 4 threads	14s
non-cyclical expansion of generator	280ms

Next Steps & Applications

Next steps for transducer

Moving forward:

- Expand the transducer lexicon
- Complete our account of Hän morphology
- Use spellrelax rules to account for phonological alternations
- Account for systematic spelling and vocabulary differences found between the the Eagle (Alaska) and Moosehide (Yukon) dialects of Hän, so that pedagogical resources we produce will be equally accessible to both communities

Applications in Revitalization

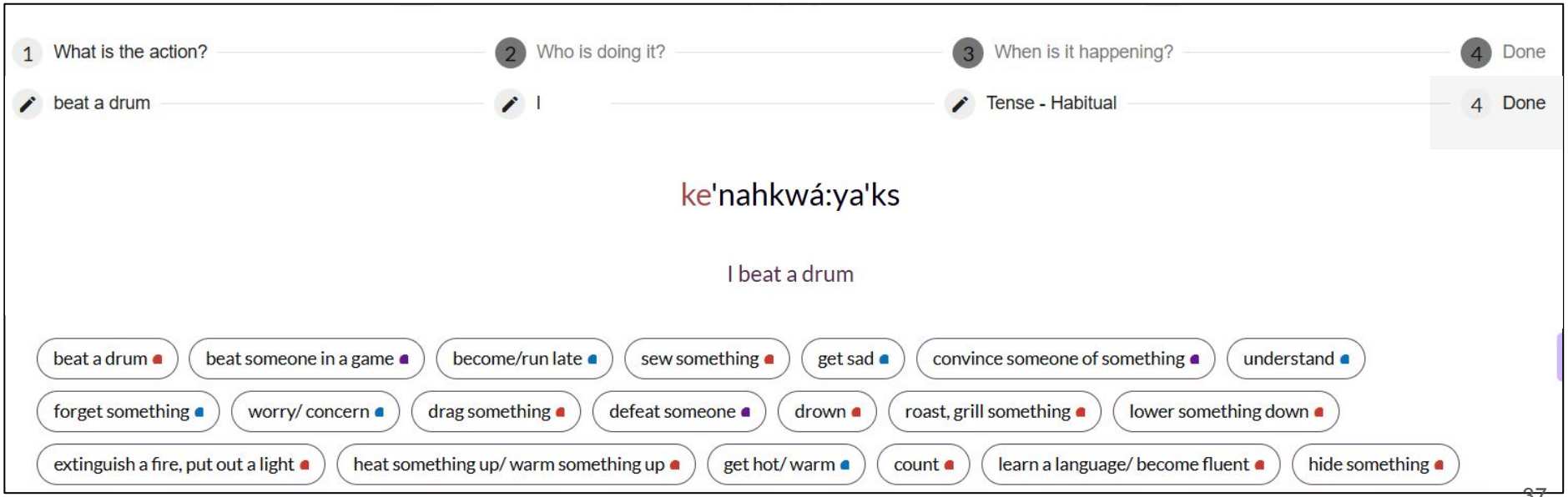
Use the transducer to build dynamic tools that can be used by language learners:

- verb-form generator (Example: Kanien'kéha (Mohawk))

Applications in Revitalization

Use the transducer to build dynamic tools that can be used by language learners:

- verb-form generator (Example: Kanien'kéha (Mohawk))



Applications in Revitalization

Use the transducer to build dynamic tools that can be used by language learners:

- verb-form generator
- paradigm generator (Example: Spanish)

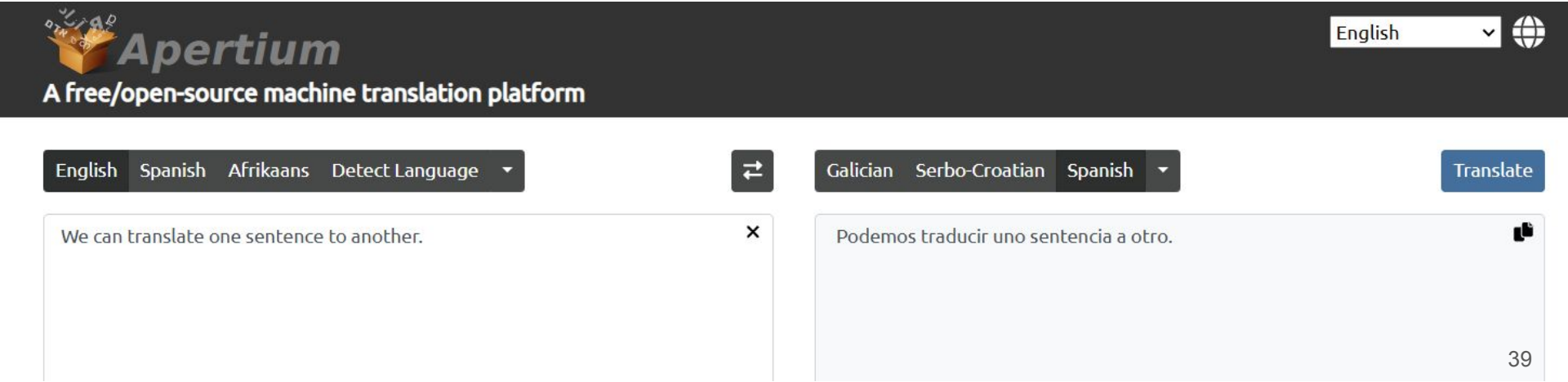
Present

	Singular	Plural
First (1st)	voy	vamos
Second (2nd)	vas	vais
Third (3rd)	va	van

Applications in Revitalization

Use the transducer to build dynamic tools that can be used by language learners:

- verb-form generator
- paradigm generator
- sentence-level translator (Example: Many languages)



Conclusion

- First morphological transducer for a Dene language written in `lexd`, which shows advantages over previous approaches to Dene morphology using `lexc`:
  - cleaner code
  - small transducer
  - fast compilation and runtime speeds
- Our hope is that an efficient transducer will allow us to create helpful and easy-to-use language resources to aid the revitalization of the Hän language.

Acknowledgements

Thanks first and foremost go to the speakers of Hän who have shared their language with us over the past 20 years, including many who are no longer with us. Speakers involved in the data used in this project are, alphabetically: **Angie Joseph-Rear**, **Adeline (Juneby) Potts**, **Archie Roberts**, **Bertha Ulvi**, **Charlie Silas**, **Charlie Stevens**, **Danny David**, **Doris Roberts**, **Edith Josie**, **Edward Roberts**, **Ethel Beck**, **Eliza Malcolm**, **Harry David, Jr.**, **Isaac Juneby**, **Joseph Susie Joseph**, **Louise Paul**, **Matthew Malcolm**, **Percy Henry**, **Richard Nukon**, **Richard Silas**, **Ruth Ridley**, **Sarah Malcolm**, **Stanley Roberts**, **Susie Paul**, **Timothy Malcolm**, and **Willie Juneby**.

We also thank linguists whose work with Hän has helped us at various stages, including (alphabetically) **Blake Lehman**, **David Shinen**, **Geoffrey O'Grady**, **Gordon Marsh**, **John Ritter**, **Jordan Lachler**, and **Michael Krauss**.

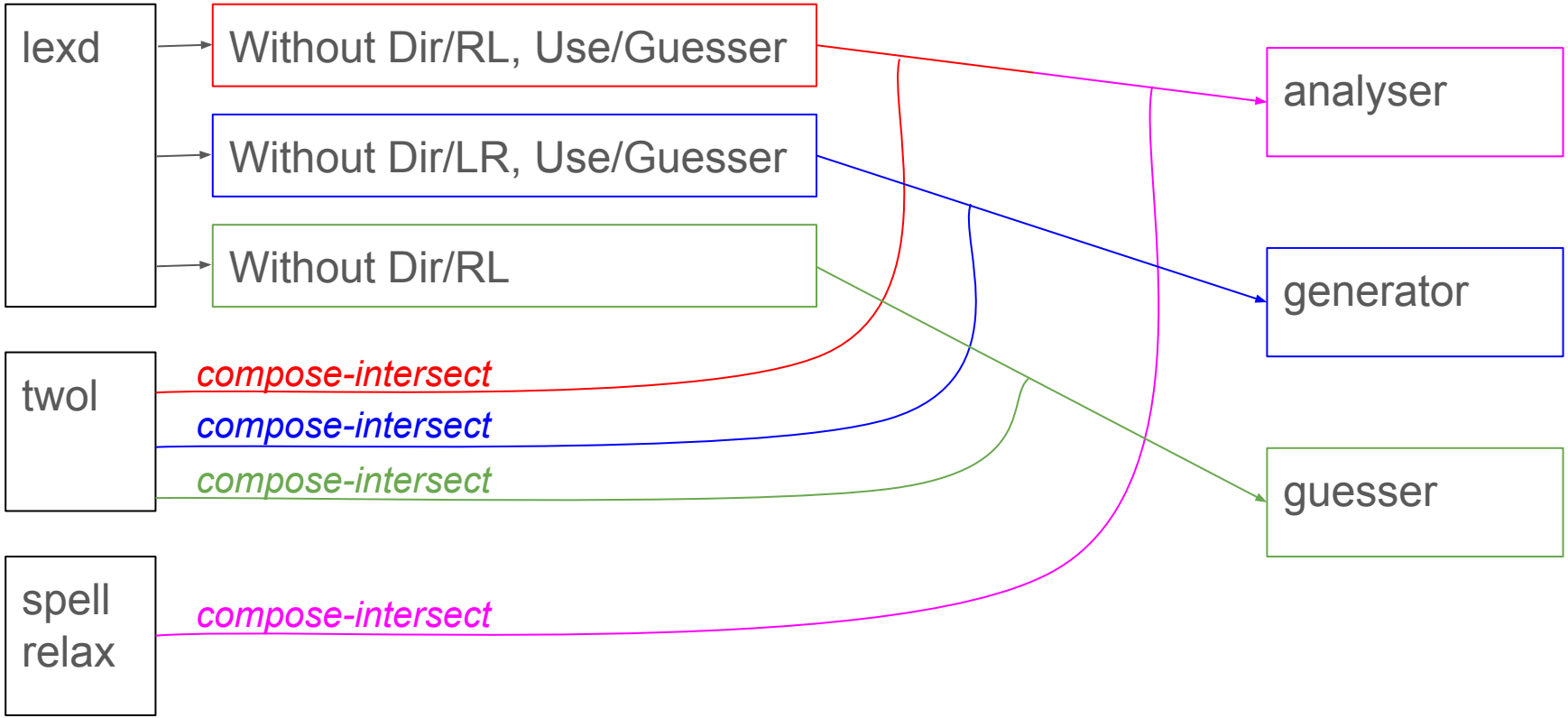
Also, community members **Georgette McLeod** and Eagle Chief **Karma Ulvi**, who have been heavily involved in the revitalization of Hän.

We also recognize work done by research assistants **Ryan Baldwin** and **T Sallie** to expand the lexicon of the transducer.

Finally, thanks to the **ComputEL-8 organizers** and **four anonymous reviewers** for their feedback.

code available at:
https://github.com/SwatLangTech/apertium-haa

Appendix: architecture



References

Antti Arppe, Christopher Cox, Mans Hulden, Jordan Lachler, Sjur N. Moshagen, Miikka Silfverberg, and Trond Trosterud. 2017. Computational modeling of verbs in Dene languages: The case of Tsuut'ina. *Working Papers in Athabascan Linguistics.*

Cody Scott Christopherson. 2023. A finite-state morphological analyzer for Q'eqchi' using Helsinki Finite-State Technology (HFST) and the Giellatekno infrastructure.Master's thesis, Brigham Young University.

Willem de Reuse. 2015a. A guide to the Hän verb paradigms. Unpublished ms.

Willem de Reuse. 2015b. Hän Athabascan verb paradigms. Unpublished ms. Fairbanks: Alaska Native Language Archives, University of Alaska.

Willem de Reuse and Kathy Joy Las. 2014. Han verb prefix paradigms: digitized and rechecked version of field notes from Jeff Leer (1974, 1977, 1980) and Leer and Ridley (1982). Unpublished ms.

Theodore B. Fernald, Nabil Kashyap, and Jeremy Fahringer. 2016. Navajo verb generator. Development version.

Mikel L. Forcada, Mireia Ginestí-Rosell, Jacob Nordfalk, Jim O'Regan, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, and Francis M. Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation*, 25(2):127–144.

Atticus G. Harrigan, Katherine Schmirler, Antti Arppe, Lene Antonsen, Trond Trosterud, and Arok Wolvengrey. 2017. Learning from the computational modelling of Plains Cree verbs. *Morphology,* 27:565–598.

Joshua Holden, Christopher Cox, and Antti Arppe. 2022. An expanded finite-state transducer for Tsuut'ina verbs. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5143–5152, Marselha, França. European Language Resources Association.

Sardana Ivanova, Anisia Katinskaia, and Roman Yangarber. 2019. Tools for supporting language learning for Sakha. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 155–163, Turku, Finland. Linköping University Electronic Press.

Anisia Katinskaia, Javad Nouri, and Roman Yangarber. 2018. Revita: a Language-learning Platform at the Intersection of ITS and CALL. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018),* Miyazaki, Japan. European Language Resources Association (ELRA).

Anna Kazantseva, Owennatekha Brian Maracle, Ronkwe'tiyóhstha Josiah Maracle, and Aidan Pine. 2018. Kawennón:nis: the wordmaker for Kanyen'kéha. In *Proceedings of the Workshop on Computational Modeling of Polysynthetic Languages,* pages 53–64, Santa Fe, New Mexico, USA. Association for Computational Linguistics

Tanmai Khanna, Jonathan N. Washington, Francis M. Tyers, Sevilay Bayatlı, Daniel G. Swanson, Tommi A. Pirinen, Irene Tang, and Hèctor Alòs i Font. 2021. Recent advances in Apertium, a free/open-source rule-based machine translation platform for low-resource languages. *Machine translation,* 35:475–502.

András Kornai. 2013. Digital language death. PLoS ONE, 8.

Blake Lehman. 2018. Tone-prominence interaction in Hän. Master's thesis, University of California, Los Angeles.

Blake Lehman and Maura O'Leary. 2019. Unexpected Athabaskan pronouns. *UCLA Working Papers: Schuhschrift: Papers in Honor of Russell Schuh*, pages 122–137.

Krister Linden, Miikka Silfverberg, Erik Axelson, Sam Hardwick, and Tommi Pirinen. 2011. *HFST—Framework for Compiling and Applying Morphologies*, volume 100 of *Communications in Computer and Information Science,* pages 67–85. Springer.

Jonathan Manker. 2014. The syntax of sluicing in Hän. In *Proceedings of the 2012 Athabascan Languages Conference.*, Fairbanks, AK: Alaska Native Language Center.

Miriam Butt. 2020. *Building Resources: Language Comparison and Analysis.* Keynote Speech: SIGTYP 2020.

Keren Rice. 2000. *Morpheme Order and Semantic Scope: Word Formation in the Athapaskan Verb*. Cambridge University Press, Cambridge.

Ruth Ridley. 1983. *Eagle Han Huch'inn Hòdök (Stories in Eagle Han Huch'inn)*. Alaska Native Language Center, University of Fairbanks, Fairbanks, Alaska.

Ruth Ridley. 2018. Hän children's stories. Unpublished manuscript.

Conor Snoek, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterud. 2014. Modeling the noun morphology of Plains Cree. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages,* pages 34–42, Baltimore, Maryland, USA. Association for Computational Linguistics.

Daniel Swanson and Nick Howell. 2021. Lexd: A finite-state lexicon compiler for non-suffixational morphologies. In Mika Hämäläinen, Niko Partanen, and Khalid Alnajjar, editors, *Multilingual Facilitation*, pages 133–146.

Jonathan Washington, Felipe Lopez, and Brook Lillehaugen. 2021. Towards a morphological transducer and orthography converter for Western Tlacolula Valley Zapotec. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 185–193, Online. Association for Computational Linguistics.