ローマ字・カタカナ・キリル文字による アイヌ語 Universal Dependencies の可能性

安岡 孝一 (京都大学)†
安岡 素子 (京都大学・京都外国語大学)

Universal Dependencies for Ainu Language in Latin Alphabet, Katakana, and Cyrillic

Koichi YASUOKA (Kyoto University)
Motoko YASUOKA (Kyoto University / Kyoto University of Foreign Studies)

要旨・既発表の有無

書写言語としてのアイヌ語は、ローマ字 (ラテンアルファベット)・カタカナ・キリル文字など、多彩な文字と記法によって記述されてきた。その一方、抱合語としてのアイヌ語は、日本語や欧米諸語とは全く異なる言語構造を持つことから、これらの言語向けの言語処理手法は、そのままではアイヌ語に適用できない。ならば Universal Dependencies は、どうだろう。言語横断的な文法構造記述として設計された Universal Dependencies は、書写言語としてのアイヌ語を、どの程度ちゃんと記述できるのだろう。『アイヌ神謡集』、『アイヌ語會話字典』、アイヌ語訳『五倫名義解』、『Аинско-русский словарь』を Universal Dependencies コーパスとして記述していく中で、われわれは、われわれの見積りが甘かったことを痛感すると同時に、それでも、アイヌ語 Universal Dependencies が、アイヌ語の言語処理に寄与することを確信した。本発表では、その一端について述べる。

本発表は、人文科学とコンピュータ研究会 (第 131 回) の発表「ローマ字・カタカナ・キリル文字併用アイヌ語 RoBERTa・DeBERTa モデルの開発」(c) 情報処理学会を拡張したものである。

1. アイヌ語 Universal Dependencies の概要

アイヌ語 Universal Dependencies は、Senuma and Aizawa (2017, 2018) がローマ字 (ラテンアルファベット) 版を開発[1]し、安岡 (2021a,b) がカタカナ[2]・キリル文字への拡張をおこなった。土台となった Universal Dependencies (UD) は、書写言語における品詞・形態素属性・依存構造 (係り受け関係) を、言語に関わらず記述する手法である [Marneffe et al. (2021)]。句構造を考慮せずに係り受け関係を記述することで、言語横断性を高めており、全ての文法構造を単語間のリンクで記述するのが特徴である。

依存構造解析それ自体は、Tesnière (1959) の構造的統語論に源を発し、Mel'čuk (1988) の有向グラフ記述によって、一応の完成を見た手法である。その最大の特長は、いわゆる動詞中

† yasuoka@kanji.zinbun.kyoto-u.ac.jp

(1) 瀬沼らによる公開は、知里 (1923)『アイヌ神謡集』の「ホテナオ」のみだった。

(2) われわれがカタカナ表記アイヌ語のデジタル化にかかわったのは、佐藤 (1996) が発端であり、それは JIS X 0213:2000 へのアイヌ語表記用カタカナ追加として結実した。

表 1　CoNLL-U の各フィールド

1. ID: 単語ごとに付与されたインデックスで、文ごとに 1 から始まる整数。縮約語に対しては、単語の範囲を示すのも可。
2. FORM: 語、または、句読記号。
3. LEMMA: 基底形、語幹。
4. UPOS: UD で規定された言語普遍な品詞タグ (表 2)。
5. XPOS: 言語固有の品詞タグ。
6. FEATS: UD で規定された言語普遍な形態素属性のリスト。言語固有の拡張も可。
7. HEAD: 当該の単語の係り受け元 ID。係り受け元が無い場合は 0 とする。
8. DEPREL: UD で規定された言語普遍な係り受けタグ (表 3)。HEAD が 0 の場合は root とする。言語固有の拡張も可。
9. DEPS: 複数の係り受け元を持つ場合、全ての HEAD:DEPREL ペア。
10. MISC: その他のアノテーション。

表 2　UD 品詞タグ (UPOS)

Open class words	**Closed class words**	**Other**
ADJ 形容詞	ADP 側置詞	PUNCT 句読点
ADV 副詞	AUX 助動詞	SYM 記号
INTJ 感嘆詞	CCONJ 並列接続詞	X その他
NOUN 名詞	DET 限定詞	
PROPN 固有名詞	NUM 数詞	
VERB 動詞	PART 接辞	
	PRON 代名詞	
	SCONJ 従属接続詞	

表 3　UD 係り受けタグ (DEPREL)

	Nominals	**Clauses**	**Modifier Words**	**Function Words**
Core arguments	nsubj 主語 obj 目的語 iobj 間接目的語	csubj 節主語 ccomp 節目的語 xcomp 節補語		
Non-core dependents	obl 斜格補語 vocative 呼称語 expl 形式語 dislocated 外置語	advcl 連用修飾節	advmod 連用修飾語 discourse 談話要素	aux 動詞補助成分 cop 繋辞 mark 標識
Nominal dependents	nmod 体言による連体修飾語 appos 同格 nummod 数量による修飾語	acl 連体修飾節	amod 用言による連体修飾語	det 決定語 clf 類別語 case 格表示

Coordination	**MWE**	**Loose**	**Special**	**Other**
conj 接続 cc 接続語	fixed 固着 flat 並列 compound 複合	list 細目 parataxis 隣接表現	orphan 親なし goeswith 泣き別れ reparandum 言い損じ	punct 句読点 root 親 dep 未定義

```
# text = kamuy tura okay=an
1     kamuy     kamuy     NOUN    名詞        _    3    obl     _    _
2     tura      tura      ADP     後置副詞    _    1    case    _    _
3     okay      okay      VERB    自動詞      _    0    root    _    SpaceAfter=No
4     =an       =an       PART    人称接辞    _    3    nsubj   _    _

# text = カムイ トゥラ オカヤン
1     カムイ    kamuy     NOUN    名詞        _    3    obl     _    _
2     トゥラ    tura      ADP     後置副詞    _    1    case    _    _
3-4   オカヤン  _         _       _           _    _    _       _    _
3     オカイ    okay      VERB    自動詞      _    0    root    _    _
4     アン      =an       PART    人称接辞    _    3    nsubj   _    _

# text = камуй тура окаян
1     камуй     kamuy     NOUN    名詞        _    3    obl     _    _
2     тура      tura      ADP     後置副詞    _    1    case    _    _
3-4   окаян     _         _       _           _    _    _       _    _
3     окай      okay      VERB    自動詞      _    0    root    _    _
4     ан        =an       PART    人称接辞    _    3    nsubj   _    _
```

図 1　アイヌ語 UD の CoNLL-U データ

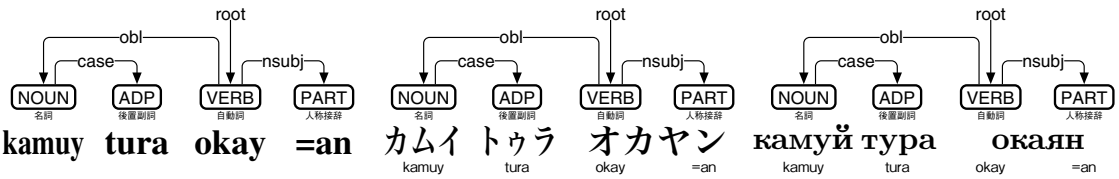


図 2　deplacy によるアイヌ語 UD の可視化

心主義によって言語横断的な記述が可能だという点にあり、Mel'čuk (1988) 依存文法をコンピュータ向けに洗練した UD においても、言語に関わらない記述、という特長が前面に押し出されている。UD における文法構造記述は、句構造を考慮せず、全てを単語間のリンクとして表現する。これにより、言語横断的な文法構造記述を可能としている。

UD 係り受けコーパスの交換用フォーマットとして、CoNLL-U と呼ばれるタブ区切りテキスト (文字コードは UTF-8) が規定されている。CoNLL-U の各行は各単語に対応しており、表 1 に示す 10 個のタブ区切りフィールドで構成される。ID・FORM・LEMMA は、単語そのものに関するフィールドである。UPOS・XPOS・FEATS は、単語の品詞と形態素属性に関するフィールドである。HEAD・DEPREL・DEPS は、単語の係り受けに関するフィールドである。

UD における係り受け関係は、単語間の有向グラフを HEAD と DEPREL で記述する。HEAD は、その単語に入る有向枝のリンク元 ID を示しており、DEPREL は、その有向枝における係り受けタグである。ただし、HEAD が 0 の場合、その枝に入るリンク元は存在しない。リンクの本数は単語の個数に等しく、各リンクのリンク先は、全て互いに異なっている。すなわち、各単語から出るリンクは複数の可能性があるが、各単語に入るリンクは 1 つだけである。なお、リンクはループしない。

UD の係り受けリンクは、Mel'čuk (1988) 依存文法の後裔にあたり、いわゆる動詞中心主義である。動詞をリンク元として、主語や目的語へとリンクする。修飾関係においては、被修飾語から修飾語へとリンクする。ただし、側置詞 (前置詞や後置詞) を体言の修飾語だとみなす [Nivre (2015)] 点が、Mel'čuk (1988) とは異なっている。また、コピュラ文においては動詞中心主義を採らず、補語をリンク元として、主語や繋辞へとリンクする。

UD は単語長を規定しておらず、各言語ごとに、自由に単語長を決めることができる。アイヌ語 UD では、田村 (1996)『アイヌ語沙流方言辞典』を、作業上の単語認定に用いている。なお、接尾辞・接頭辞については、人称接辞と動名詞接尾辞 (-i と-p) だけを語とみなし、それ以外は前後の語にくっ付けている。

アイヌ語 UD の例として、「kamuy tura okay=an」「カムイ トゥラ オカヤン」「камуй тура окаян」の CoNLL-U データを図 1 に示す。LEMMA と XPOS は『アイヌ語沙流方言辞典』に従っている(3)。また、これらの CoNLL-U を比較すべく，deplacy [安岡 (2020)] で可視化した (図 2)。UD 依存構造は全く同一だが、「オカヤン」や「окаян」は、文字の途中に単語境界がある点に注意されたい。

2. アイヌ語 Universal Dependencies コーパスの作成

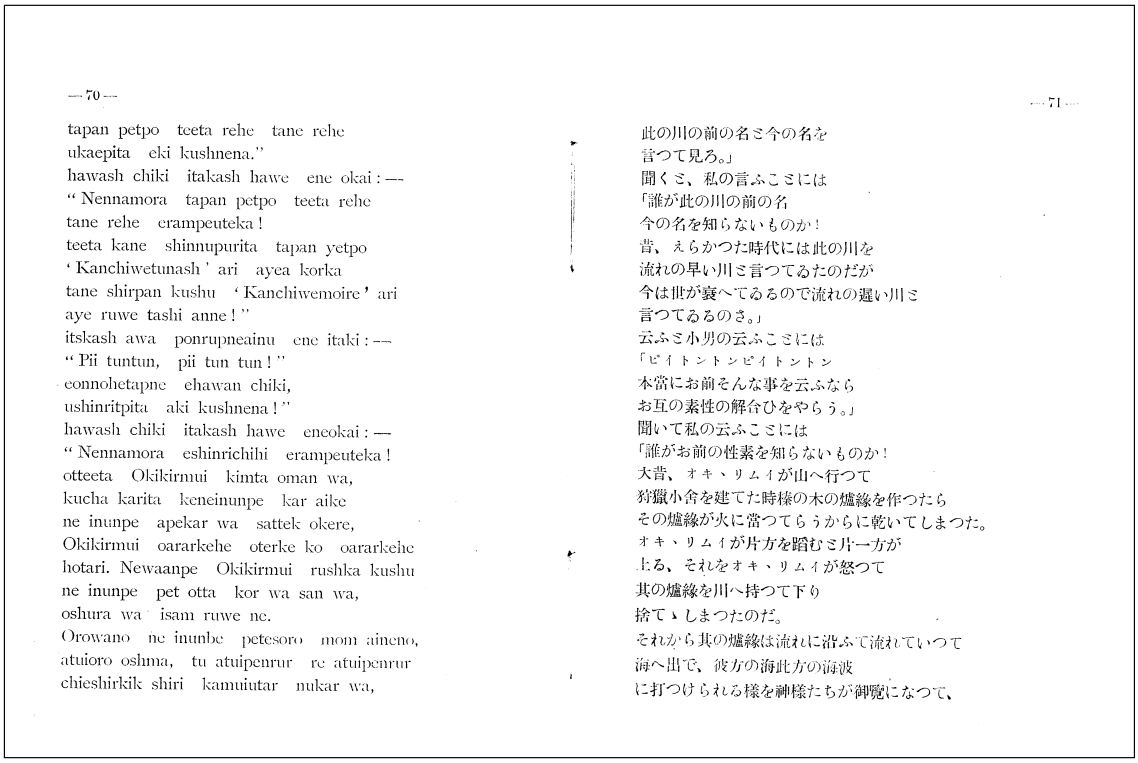
ローマ字・カタカナ・キリル文字で書かれたアイヌ語文書に対し、係り受け解析エンジン esupar のアイヌ語 DeBERTa モードで仮コーパスを作成し、その結果をアイヌ語 UD エディターで編集する、という手順で、アイヌ語 UD コーパスを作成した。以下、それぞれのアイヌ語 UD コーパスについて、概要を述べる。

2.1 アイヌ神謡集

知里 (1923)『アイヌ神謡集』は、本文 124 ページに 13 編のアイヌ神謡を収録しており、見開き左ページ (偶数ページ) にローマ字で書かれたアイヌ語を、見開き右ページ (奇数ページ) に日本語訳を配置している (図 3)。各編の構成は以下のとおり。

1.「銀の滴降る降るまはりに」(11 ページ 230 行) × 2
2.「トワトワト」(7 ページ 136 行) × 2
3.「ハイクンテレケ ハイコシテムトリ」(6 ページ 121 行) × 2
4.「サンパヤ テレケ」(5 ページ 104 行) × 2
5.「ハリツクンナ」(4 ページ 83 行) × 2

(3) アイヌ語 UD の XPOS では、固有名詞を名詞から分離し、数詞を連体詞から分離した上で、複他動詞を他動詞に統合し、さらに記号を加えた [安岡 (2021b)]。



— 70 —

tapan petpo teeta rehe tane rehe
ukaepita eki kushnena."
hawash chiki itakash hawe ene okai : —
" Nennamora tapan petpo teeta rehe
tane rehe erampeuteka !
teeta kane shinnupurita tapan yetpo
' Kanchiwetunash ' ari ayea korka
tane shirpan kushu ' Kanchiwemoire ' ari
aye ruwe tashi anne ! "
itskash awa ponrupneainu ene itaki : —
" Pii tuntun, pii tun tun ! "
connohetapne ehawan chiki,
ushinritpita aki kushnena ! "
hawash chiki itakash hawe eneokai : —
" Nennamora eshinrichihi erampeuteka !
otteeta Okikirmui kimta oman wa,
kucha karita keneinunpe kar aike
ne inunpe apekar wa sattek okere,
Okikirmui oararkehe oterke ko oararkehe
hotari. Newaanpe Okikirmui rushka kushu
ne inunpe pet otta kor wa san wa,
oshura wa isam ruwe ne.
Orowano ne inunbe petesoro mom aineno,
atuioro oshma, tu atuipenrur re atuipenrur
chieshirkik shiri kamuiutar nukar wa,

— 71 —

此の川の前の名ミ今の名を
言つて見ろ。」
聞くミ、私の言ふこゝには
「誰が此の川の前の名
今の名を知らないものか！
昔、えらかつた時代には此の川を
流れの早い川ミ言つてゐたのだが
今は世が衰へてゐるので流れの遅い川ミ
言つてゐるのさ。」
云ふミ小男の云ふこゝには
「ピイトントンピイトントン
本當にお前そんな事を云ふなら
お互の素性の解合ひをやらう。」
聞いて私の云ふこゝには
「誰がお前の性素を知らないものか！
大昔、オキ、リムイが山へ行つて
狩獵小舍を建てた時様の木の爐縁を作つたら
その爐縁が火に當つてらうからに乾いてしまつた。
オキ、リムイが片方を蹈むミ片一方が
上る、それをオキ、リムイが怒つて
其の爐縁を川へ持つて下り
捨てゝしまつたのだ。
それから其の爐縁は流れに沿ふて流れていつて
海へ出て、彼方の海此方の海波
に打つけられる様を神様たちが御覽になつて、

図 3　『アイヌ神謡集』「ホテナオ」70〜71 ページ

```
# text = itskash awa ponrupneainu ene itaki : —
1	itsk	itak	VERB	自動詞	_	9	advcl	_	SpaceAfter=No
2	ash	=as	PART	人称接辞	_	1	nsubj	_	_
3	awa	awa	CCONJ	接続詞	_	1	mark	_	_
4	pon	pon	VERB	自動詞	_	6	amod	_	SpaceAfter=No
5	rupne	rupne	VERB	自動詞	_	6	amod	_	SpaceAfter=No
6	ainu	aynu	NOUN	名詞	_	8	nsubj	_	_
7	ene	ene	ADV	副詞	_	8	advmod	_	_
8	itak	itak	VERB	自動詞	_	9	acl	_	SpaceAfter=No
9	i	-i	PART	接尾辞	_	0	root	_	_
10	:	:	PUNCT	記号	_	9	punct	_	_
11	—	—	PUNCT	記号	_	9	punct	_	_
```

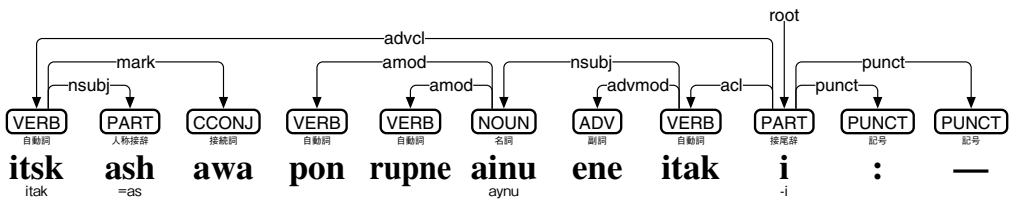


図 4　「itskash awa ponrupneainu ene itaki : —」のアイヌ語 UD

6.「ホテナオ」(3 ページ 66 行) × 2
7.「コンクワ」(6 ページ 125 行) × 2
8.「アトイカトマトマキ、クントテアシフム、フム!」(8 ページ 193 行) × 2
9.「トーロロハンロクハンロク!」(2 ページ 43 行) × 2
10.「クツニサクトンクトン」(2 ページ 30 行) × 2
11.「此の砂赤い赤い」(3 ページ 70 行) × 2
12.「カツパレウレウカツパ」(3 ページ 53 行) × 2
13.「トヌペカランラン」(2 ページ 43 行) × 2

『アイヌ神謡集』のローマ字表記は『アイヌ語沙流方言辞典』と異なっており、また、誤植も散見される[(4)]。われわれのアイヌ語 UD コーパスでは、『アイヌ神謡集』の表記をそのまま FORM に入れ、LEMMA と単語長を『アイヌ語沙流方言辞典』に合わせた。単語長の差は、MISC の `SpaceAfter=No` で吸収した。たとえば、図 3 左 70 ページ 10 行目「itskash awa ponrupneainu ene itaki : ―」に対しては、FORM は誤植も含めてそのままとし、LEMMA は「itak =as awa pon rupne aynu ene itak -i : ―」としている (図 4)。なお、『アイヌ語沙流方言辞典』に見当たらない単語については、基本的に片山 (2003) の単語認定に依っている。

2.2 アイヌ語會話字典

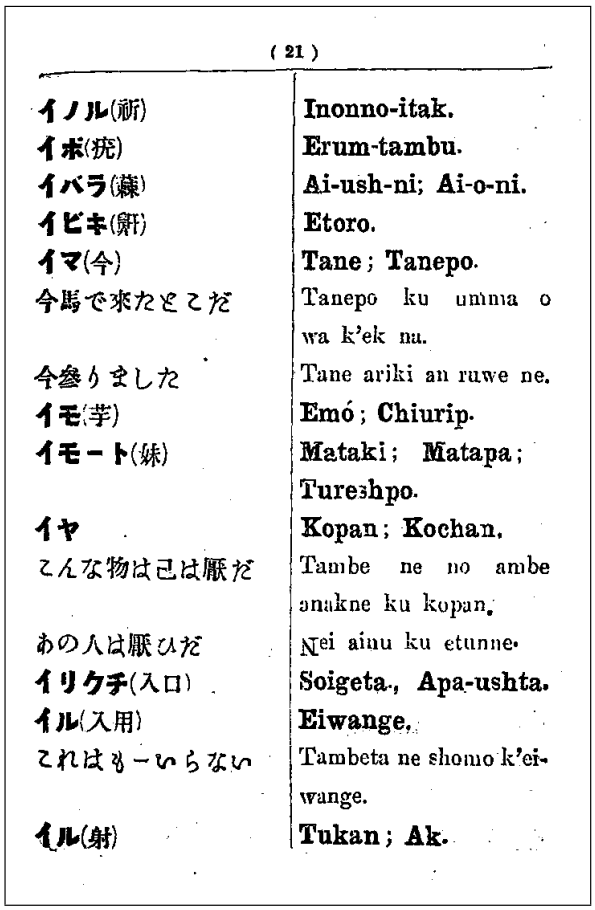
神保・金澤 (1898)『アイヌ語會話字典』は、本文 2 段組 278 ページの段組左側に日本語を、右側にローマ字のアイヌ語訳を配置している (図 5)。Bugaeva (2011) は『アイヌ語會話字典』を拡張する形で、トピック別アイヌ語会話辞典 (全 3847 見出し) を公開している。

『アイヌ會話字典』のローマ字表記は『アイヌ語沙流方言辞典』と異なっており、特に単語長の認定が全く違う。われわれのアイヌ語 UD コーパスでは、『アイヌ會話字典』の表記をそのまま FORM に入れ、LEMMA と単語長を『アイヌ語沙流方言辞典』に合わせた。単語長の差は、FORM 中の空白や、MISC の `SpaceAfter=No` で吸収した。たとえば、図 5 右側 18〜19 行目「Tambeta ne shomo k'eiwange.」に対しては、FORM は「ta ne」に空白を含みつつ、LEMMA は「tan pe tane somo k= eywanke .」としている (図 6)。

2.3 Аинско-русский словарь

Добротворский (1875) アイヌ語・ロシア語辞典の補遺第 12 章 (図 7) には、キリル文字で書かれた樺太アイヌ語の対話文が収録されている [寺田・安田 (2019)]。この対話文については、阪口 (2021) によるローマナイゼーションと日本語訳、および詳細な解説があり、これを参照しつつアイヌ語 UD コーパスの作成をおこなった。図 7 左ページ本文 3 行目「Танъ котàнъ охтà утáса—анъ кусý áреги анъ.」に対するアイヌ語 UD を、図 8 に示す。なお、阪口 (2021) は「охтà」のローマナイゼーションを「ohtà」としているが、われわれは『アイヌ語沙流方言辞典』に合わせて「or ta」とした。樺太アイヌ語を沙流アイヌ語に合わせるかどうかについては、もちろん議論の余地があると考えられる。

(4) 図 3 の左 70 ページには、6 行目「yetpo」→「petpo」、10 行目「itskash」→「itakash」、12 行目「eonnohetapne」→「sonnohetapne」の誤植がある [佐藤 (2004)]。右 71 ページには、15 行目「性素」→「素性」、18 行目「らうから」→「からから」の誤植がある。



(21)

イノル(祈)	Inonno-itak.
イボ(疣)	Erum-tambu.
イバラ(蘇)	Ai-ush-ni; Ai-o-ni.
イビキ(鼾)	Etoro.
イマ(今)	Tane; Tanepo.
今馬で來たところだ	Tanepo ku umma o wa k'ek na.
今參りました	Tane ariki an ruwe ne.
イモ(芋)	Emó; Chiurip.
イモート(妹)	Mataki; Matapa; Tureshpo.
イヤ	Kopan; Kochan.
こんな物は己は厭だ	Tambe ne no ambe anakne ku kopan.
あの人は厭ひだ	Nei ainu ku etunne.
イリクチ(入口)	Soigeta., Apa-ushta.
イル(入用)	Eiwange.
これはもーいらない	Tambeta ne shomo k'eiwange.
イル(射)	Tukan; Ak.

図 5　『アイヌ語會話字典』21 ページ

```
# text = Tambeta ne shomo k'eiwange.
```

1	Tam	tan	DET	連体詞	_	2	det	_	SpaceAfter=No
2	be	pe	NOUN	形式名詞	_	6	obj	_	SpaceAfter=No
3	ta ne	tane	ADV	副詞	_	6	advmod	_	_
4	shomo	somo	ADV	副詞	_	6	advmod	_	_
5	k'	k=	PART	人称接辞	_	6	nsubj	_	SpaceAfter=No
6	eiwange	eywanke	VERB	他動詞	_	0	root	_	SpaceAfter=No
7	.	.	PUNCT	記号	_	6	punct	_	_

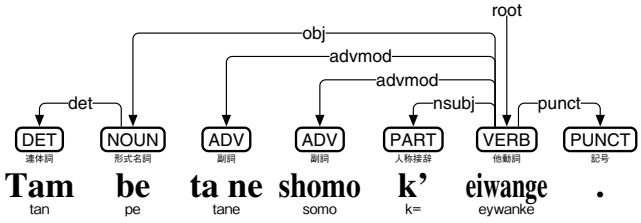
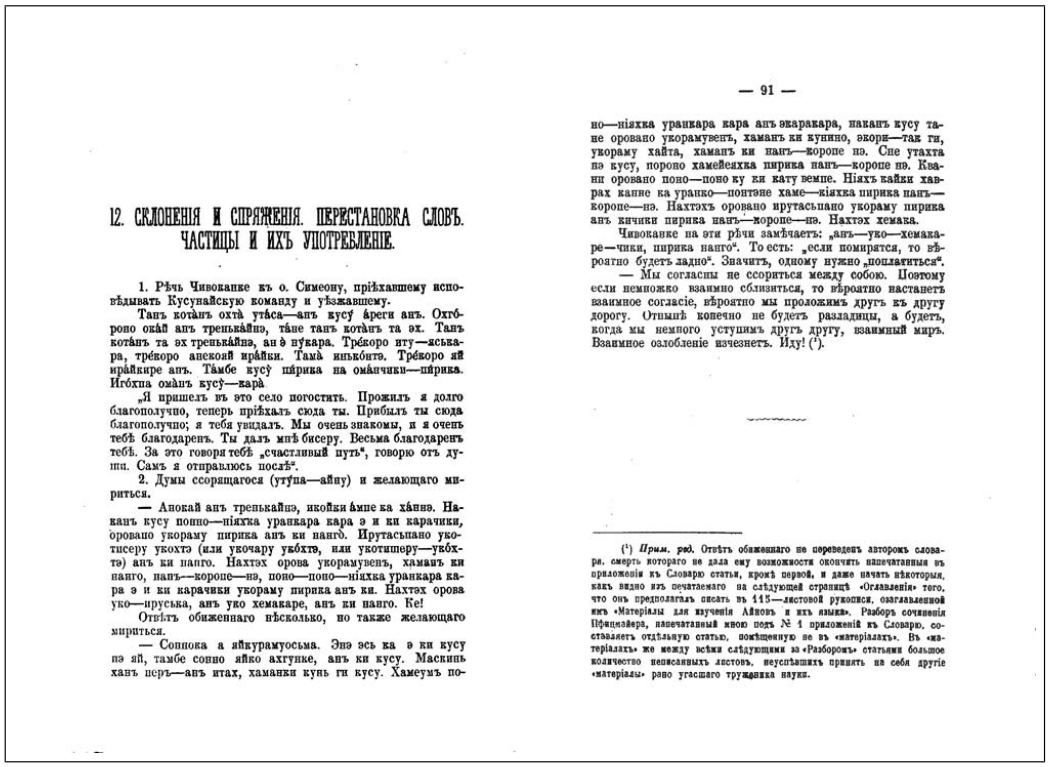


図 6　「Tambeta ne shomo k'eiwange.」のアイヌ語 UD



12. СКЛОНЕНІЯ И СПРЯЖЕНІЯ. ПЕРЕСТАНОВКА СЛОВЪ. ЧАСТИЦЫ И ИХЪ УПОТРЕБЛЕНІЕ.

1. Рѣчь Чивоканке къ о. Симеону, пріѣхавшему исповѣдывать Кусунайскую команду и уѣзжавшему.

Танъ котàнъ охтà утáса—анъ кусу́ áреги анъ. Охтóрано окáй анъ тренькáйпэ, тáне танъ котàнъ та эх. Танъ котàнъ та эх тренькáйпэ, ан э пу́кара. Трéкоро иту—аскара, трéкоро апескай ирáйки. Тамà инькóнтэ. Трéкоро яй ирáйкире апъ. Тáмбе кусу́ пи́рика на омáнчики—пи́рика. Игóхпа омàнъ кусу́—кара

„Я пришелъ въ это село погостить. Прожилъ я долго благополучно, теперь пріѣхалъ сюда ты. Прибылъ ты сюда благополучно; я тебя увидалъ. Мы очень знакомы, и я очень тебѣ благодаренъ. Ты далъ мнѣ бисеру. Весьма благодаренъ тебѣ. За это говоря тебѣ „счастливый путь", говорю отъ души. Самъ я отправлюсь посла́ть.

2. Думы ссорящагося (утýпа—айну) и желающаго мириться.

— Анокай анъ тренькайпэ, икойки áмпе ка хáнпэ. Наканъ кусу понно—пíяхка уранкара кара э и ки карачики, оровано укораму пирика анъ ки пангò. Ирутасыпано укотисеру укохтэ (или укочару укóхтэ, или укотишеру—укóхтэ) анъ ки пангo. Нахтэх орова укорамувенъ, хамàнъ ки пангo, напъ—коропе—пэ, поно—поно—пíяхка уранкара кара э и ки карачики укораму пирика анъ ки. Нахтэх орова уко—ируська, анъ уко хемакаре, анъ ки пангo. Ке!

Отвѣтъ обиженнаго нѣсколько, но также желающаго мириться.

— Сопнока а айкурамуосьма. Энэ эсь ка э ки кусу пэ яй, тамбе сонно айко ахгунке, анъ ки кусу. Маскинь ханъ перъ—анъ итах, хаманки кунь ги кусу. Хамеумъ по-

но—пíяхка уранкара кара анъ экаракара, наканъ кусу тане оровано укорамувенъ, хамàнъ ки кунино, экори—так ги, укораму хайта, хамàнъ ки нанъ—коропе пэ. Сне утахта пэ кусу, пороно хамéйехха пирика нанъ—коропе пэ. Квани оровано поно—поно ку ки кату вемпе. Ниихъ кайки хаврах канпе ка уранко—понтэне хаме—пíяхка пирика нанъ—коропе—пэ. Нахтохъ оровано ирутасыпано укораму пирика анъ кичики пирика нанъ—коропе—пэ. Нахтэх хемака.

Чивоканке на эти рѣчи замѣчаетъ: „анъ—уко—хемакаре—чики, пирика пангò". То есть: „если помирятся, то вѣроятно будетъ ладно". Значитъ, одному нужно „поплакаться".

— Мы согласны не ссориться между собою. Поэтому если немножко взаимно сблизиться, то вѣроятно настанетъ взаимное согласіе, вѣроятно мы проложимъ другъ къ другу дорогу. Отнынѣ конечно не будетъ разладицы, а будетъ, когда мы немного уступимъ другъ другу, взаимный миръ. Взаимное озлобленіе изчезнетъ. Йду! (1).

(1) *Прим. ред.* Отвѣтъ обиженнаго не переведенъ авторомъ словаря, смерть котораго не дала ему возможности окончить напечатанныя въ приложеніи къ Словарю статьи, кромѣ первой, и даже начать нѣкоторыя, какъ видно изъ печатаемаго на слѣдующей страницѣ «Оглавленія» того, что онъ предполагалъ писать въ 115—листовой рукописи, озаглавленной имъ «Матеріалы для изученія Айновъ и ихъ языка». Разборъ сочиненія Пфицмайера, напечатанный имъ подъ № 1 приложеній къ Словарю, составляетъ отдѣльную статью, помѣщенную не въ «матеріалахъ». Въ «матеріалахъ» же между всѣми слѣдующими за «Разборомъ» статьями большое количество написанныхъ листовъ, неуспѣвшихъ принять на себя другіе «матеріалы» равно угасшаго труженика науки.

図 7　『Аинско-русский словарь』補遺第 12 章

```
# text = Танъ котàнъ охтà утáса—анъ кусу́ áреги анъ.
```

1	Танъ	tan	DET	連体詞	_	2	det	_	_
2	котàнъ	kotan	NOUN	名詞	_	3	nmod	_	_
3	ох	or	NOUN	位置名詞	_	5	obl	_	SpaceAfter=No
4	тà	ta	ADP	格助詞	_	3	case	_	_
5	утáса	u-tasa	VERB	自動詞	_	9	advcl	_	SpaceAfter=No
6	—	-	PUNCT	記号	_	5	punct	_	SpaceAfter=No
7	анъ	=an	PART	人称接辞	_	5	nsubj	_	_
8	кусу́	kusu	SCONJ	接続助詞	_	5	mark	_	_
9	áреги	ar-iki	VERB	自動詞	_	0	root	_	_
10	анъ	=an	PART	人称接辞	_	9	nsubj	_	SpaceAfter=No
11	.	.	PUNCT	記号	_	9	punct	_	_

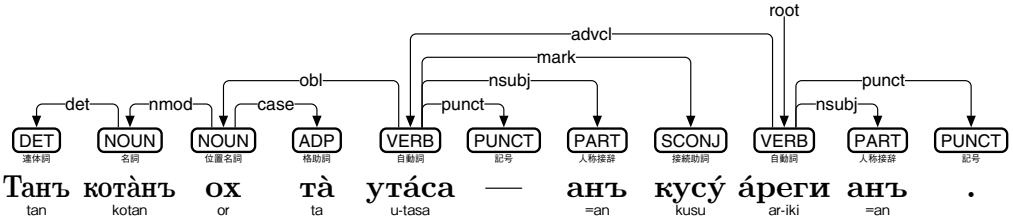


図 8　「Танъ котàнъ охтà утáса—анъ кусу́ áреги анъ.」のアイヌ語 UD

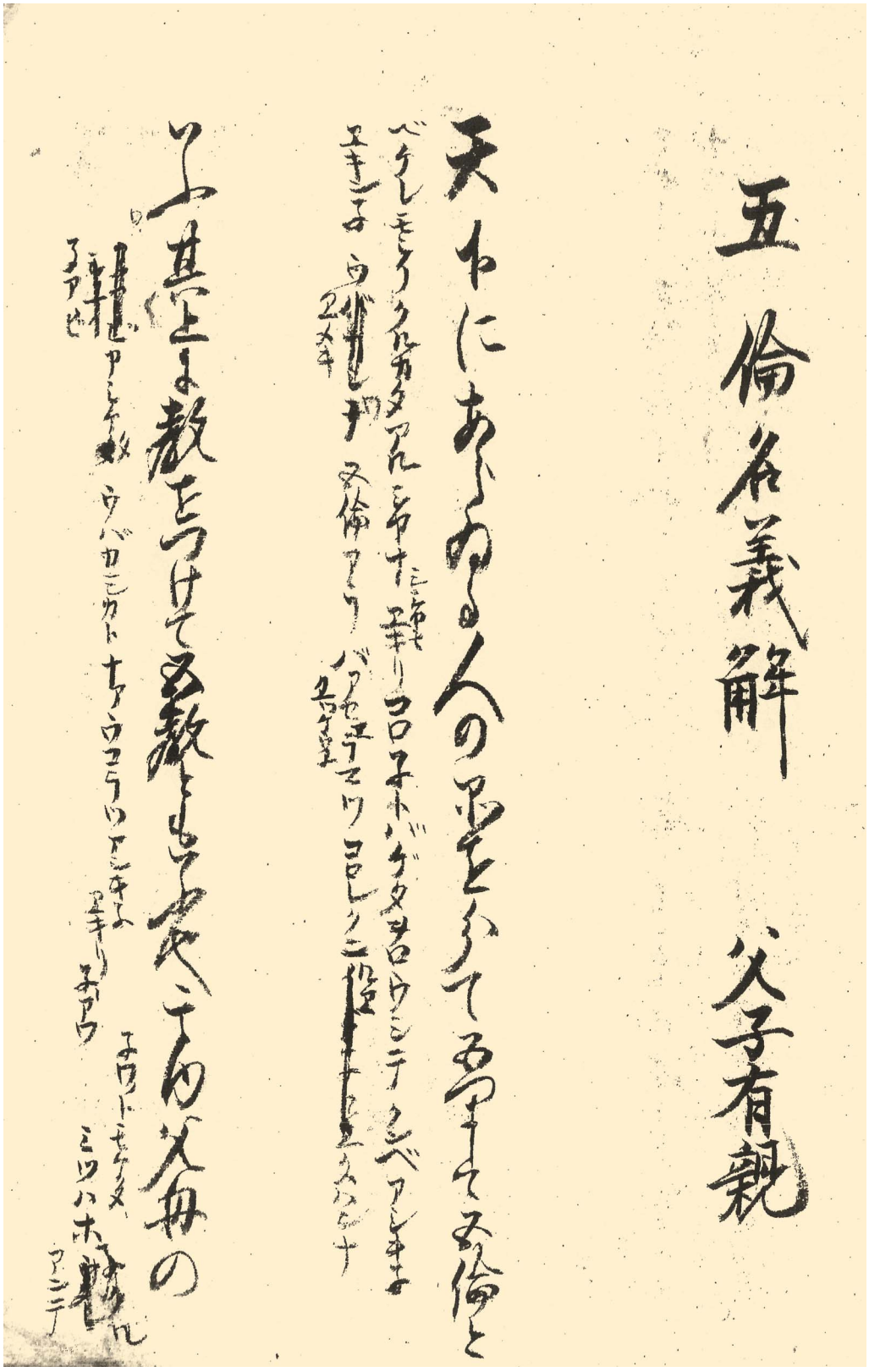


図９　アイヌ語訳『五倫名義解』冒頭部

2.4　アイヌ語訳『五倫名義解』

加賀家文書館 (別海町) 所蔵のアイヌ語訳『五倫名義解』(整理番号 K3-21) は、室・空谷 (1855, 1858)『五倫名義解』に加賀伝蔵がアイヌ語訳を施したもので、文久〜慶応年間に書かれたものである [深澤 (2014a)]。以下に示す 5 章と刊記で構成される。

1.「父子有親」6 ページ　　12 文× 2
2.「君臣有義」6 ページ　　14 文× 2
3.「夫婦有別」6 ページ　　13 文× 2
4.「長幼有序」7 ページ　　13 文× 2
5.「朋友有信」8 ページ半　13 文× 2
6. 刊記　　5 ページ半　11 文× 2

各ページには、日本語が 3 行ずつ書かれており、その横にカタカナでアイヌ語訳が記されているが、各章の表題はアイヌ語訳されていない (図 9)。アイヌ語訳に小書きカタカナは使われておらず、拗音も促音も小書きにしない上、末子音が母音を伴って書かれている。深澤 (2014b) が指摘するとおり、母音の混同 (イとエ、ウとヲ) も散見される。しかも書き直しが多く、非常に読みにくい。図 9 のアイヌ語訳に対するアイヌ語 UD を、図 10 に示す。ただし、「アルシヤナ」に「earsayne」を当てていいのか、「ウバカシ」は「uwepakasnu」なのか、など多くの疑問点が残っており [安岡・安岡 (2023)]、現在も引き続き作業中である。

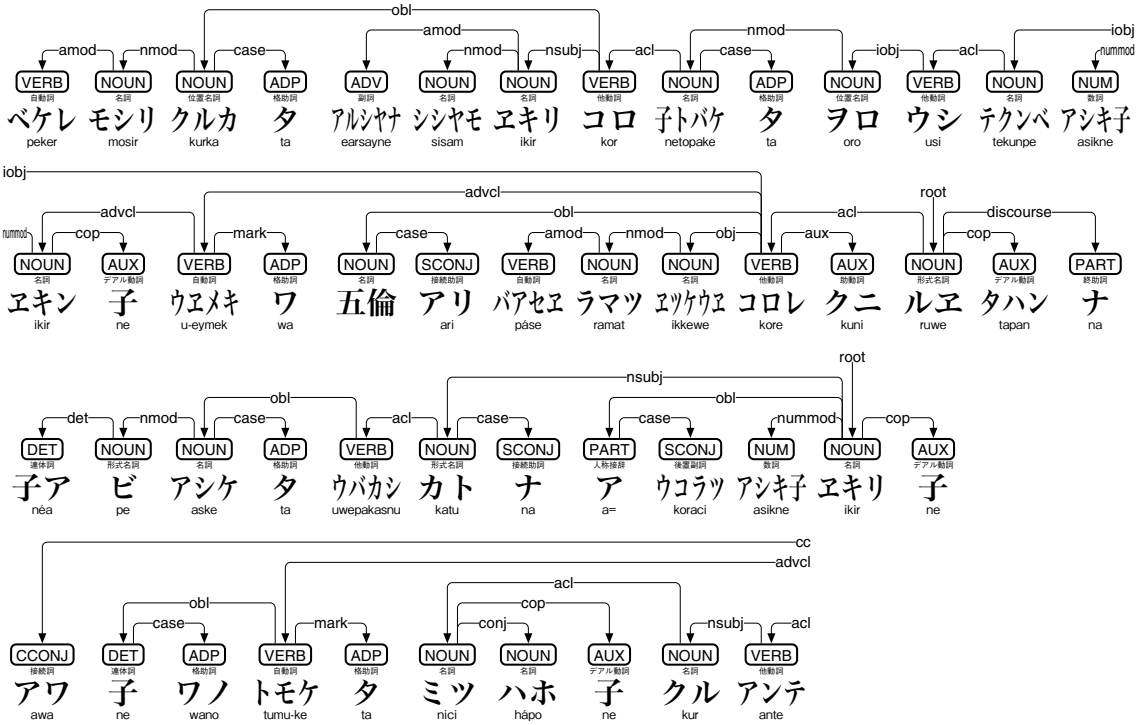


図 10　アイヌ語訳『五倫名義解』冒頭部のアイヌ語 UD

2.5　国立アイヌ民族博物館ガイドブック

ウポポイ (民族共生象徴空間) はアイヌ語を第一言語としており [小林 (2023), 深澤 (2023)]、その方針は、国立アイヌ民族博物館 (2020a,b, 2021) にも踏襲されている。

図 11　『国立アイヌ民族博物館ガイドブック』21〜22 ページ

国立アイヌ民族博物館 (2020a)『国立アイヌ民族博物館ガイドブック』は、見開き左ページ (奇数ページ) にアイヌ語で解説を書き、見開き右ページ (偶数ページ) にその日本語訳と英語訳を載せる、という方針で編集されている (図 11)。アイヌ語はカタカナで書かれており、表題に限ってローマ字が添えられている。図 11 の「カムイ トゥラ オカヤン」「kamuy tura okay=an」に対するアイヌ語 UD を、図 1・2 に示す。ただ、各解説には執筆者が記されており、アイヌ語 UD コーパスを作成した場合、著作権処理をどうおこなうべきか悩ましい。

3.　アイヌ語 Universal Dependencies の可能性

われわれが作成したアイヌ語 UD コーパスは、係り受け解析エンジン esupar の訓練に用いている。esupar のアイヌ語モジュールを訓練して解析精度を上げることで、さらなるアイヌ語 UD コーパスの作成が楽におこなえる。いわば循環システムだと考えてよい。このようなシステムがうまくいっているのは、UD の言語横断性に加え、単語長と LEMMA を田村 (1996)『アイヌ語沙流方言辞典』に押し込んだ点が、功を奏したと言える。

ただ、樺太アイヌ語や釧路アイヌ語など、多種多様なアイヌ語を、全て沙流アイヌ語に押し込んでいいものだろうか。この点は、われわれにとっても非常に悩ましい。多種多様なアイヌ語をそのまま言語処理しようとすると、それぞれの分量が少なくなってしまうため、解析精度が下がってしまう。多種多様なアイヌ語を保持したままで解析精度を維持するには、FORM に原文を入れた上で、LEMMA を『アイヌ語沙流方言辞典』に接地する、という両天秤な手法しか、うまくいくやり方を見つけきれていない。

アイヌ語 UD は万能ではない。実際、いくつかの文を記述する際に「綻び」が出てきているのも、また事実である。たとえば「shichorpok chikushte shienka chikushte」[5]は、佐藤 (2004) の指摘どおり「shi」を分離する方が適切 (図 12) なのだが、これはアイヌ語 UD としては、かなり特異な事例である。このような特異な事例を踏まえつつ、より適用範囲の広いアイヌ語コーパスを作成していくには、どうすべきか。われわれの今後の研究に期待されたい。

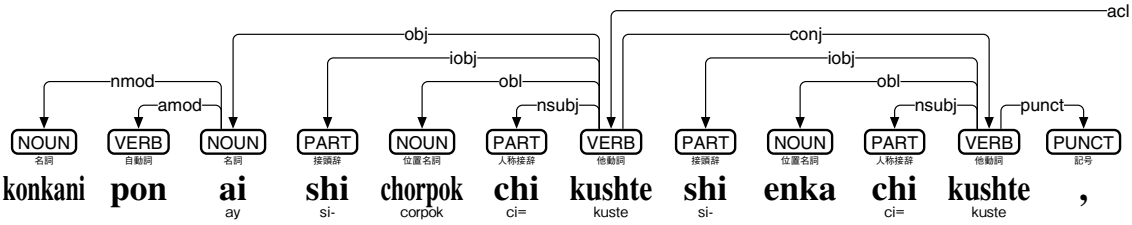


図 12　「konkani ponai shichorpok chikushte shienka chikushte,」のアイヌ語 UD 改良案

謝　辞

本発表に用いた係り受け解析エンジン esupar は、学際大規模情報基盤共同利用・共同研究拠点公募型共同研究『単語間に区切りのない書写言語における係り受け解析エンジンの開発』の成果である。また、アイヌ語 UD エディターとコーパス管理システムの開発、およびそれらを用いたコーパス作成作業は、文部科学省『AI 等の活用を推進する研究データエコシステム構築事業』の支援を受けている。

文　献

Hajime Senuma, and Akiko Aizawa (2017). “Toward Universal Dependencies for Ainu.” *NoDaLiDa 2017 Workshop on Universal Dependencies*, pp. 133–139.

Hajime Senuma, and Akiko Aizawa (2018). “Universal Dependencies for Ainu.” *LREC 2018: Eleventh International Conference on Language Resources and Evaluation*, pp. 2354–2358.

知里幸恵 (1923).『アイヌ神謡集』　郷土研究社, 東京.

安岡孝一 (2021a).「アイヌ語 Universal Dependencies 再考」　東洋学へのコンピュータ利用, 第 34 回研究セミナー, pp. 25–53.

[5] 知里 (1923)『アイヌ神謡集』「銀の滴降る降るまはりに」4 ページ 2 行目。

安岡孝一 (2021b).「Universal Dependencies によるアイヌ語テキストコーパス」 情報処理学会研究報告, 2021-CH-127:5, pp. 1–8.
佐藤知己 (1996).「アイヌ語を記述するのに必要な文字セットについて」 JIS 符号化文字集合調査研究委員会第 2 分科会 (WG2) 資料, JCS-2-8-02.
Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman (2021). "Universal Dependencies." *Computational Linguistics*, 47:2, pp. 255–308.
Lucien Tesnière (1959). *Éléments de Syntaxe Structurale*. Paris: C. Klincksieck.
Igor A. Mel'čuk (1988). *Dependency Syntax: Theory and Practice*. New York: State University of New York Press.
Joakim Nivre (2015). "Towards a Universal Grammar for Natural Language Processing." *CICLing 2015: 16th International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 3–16.
田村すず子 (1996).『アイヌ語沙流方言辞典』 草風館, 東京.
安岡孝一 (2020).「Universal Dependencies にもとづく多言語係り受け可視化ツール deplacy」人文科学とコンピュータシンポジウム「じんもんこん 2020」論文集, pp. 95–100.
佐藤知己 (2004).「知里幸恵『アイヌ神謡集』の難読個所と特異な言語事例をめぐって」 北海道立アイヌ民族文化研究センター研究紀要, 10, pp. 1–32.
片山龍峯 (2003).『「アイヌ神謡集」を読みとく』 片山言語文化研究所, 武蔵野.
神保小虎・金澤庄三郎 (1898).『アイヌ語會話字典』 金港堂書籍, 東京.
Anna Bugaeva (2011). "Internet Applications for Endangered Languages: A Talking Dictionary of Ainu." 早稲田大学高等研究所紀要, 3, pp. 73–81.
М. М. Добротворский (1875). *Аинско-русский словарь*. Казань: Университецкая типография.
寺田吉孝・安田節彦 (2019).「M. M. ドブロトゥヴォールスキーのアイヌ語・ロシア語辞典 (26)」 北海学園大学学園論集, 178, pp. 121–149.
阪口諒 (2021).「『アイヌ語ロシア語辞典』中のアイヌ語樺太方言テキスト」 千葉大学大学院人文社会科学研究科研究プロジェクト報告書, 第 358 集, pp. 43–55.
室鳩巣・空谷茂潤 (1855).『五倫名義解』 此君園, 江戸.
室鳩巣・空谷茂潤 (1858).『五倫名義解』 宗谷御用所, 宗谷.
深澤美香 (2014a).「加賀家文書のアイヌ語資料と加賀伝蔵」 千葉大学大学院人文社会科学研究科研究プロジェクト報告書, 第 274 集, pp. 21–48.
深澤美香 (2014b).「加賀家文書における表記の特徴と傾向」 千葉大学大学院人文社会科学研究科研究プロジェクト報告書, 第 274 集, pp. 49–72.
安岡孝一・安岡素子 (2023).「アイヌ語訳『五倫名義解』Universal Dependencies への挑戦」東洋学へのコンピュータ利用, 第 36 回研究セミナー, pp. 3–37.
小林美紀 (2023).「アイヌ語を第一言語に」 国立アイヌ民族博物館 (編)『ウアイヌコロ コタン アカㇻ ウポポイのことばと歴史』 国書刊行会, 東京 pp. 97–111.
深澤美香 (2023).「国立アイヌ民族博物館のアイヌ語による展示解説文と「私たち」」 国立ア

イヌ民族博物館 (編)『ウアイヌコロ コタン アカㇻ ウポポイのことばと歴史』 国書刊行会, 東京 pp. 112–153.
国立アイヌ民族博物館 (2020a).『国立アイヌ民族博物館ガイドブック』 国立アイヌ民族博物館, 白老.
国立アイヌ民族博物館 (2020b).『アヌココロ アイヌ イコロマケンル an=ukokor aynu ikor oma kenru』, 国立アイヌ民族博物館パンフレット (日本語), 白老.
国立アイヌ民族博物館 (2021).『ゴールデンカムイ トゥラノ アㇷ゚カシアン』, 国立アイヌ民族博物館第 2 回特別展示, 白老.

関連 URL

Universal Dependencies for Ainu	https://github.com/KoichiYasuoka/UD-Ainu
係り受け解析エンジン esupar	https://github.com/KoichiYasuoka/esupar
アイヌ語 UD エディター	https://koichiyasuoka.github.io/UD-Ainu/editor/
トピック別アイヌ語会話辞典	https://ainu.ninjal.ac.jp/topic/
国立国会図書館デジタルコレクション	
『アイヌ神謡集』	https://dl.ndl.go.jp/pid/1909336
『アイヌ語會話字典』	https://dl.ndl.go.jp/pid/993685