

Qwen3.5によるカタカナアイヌ語OCRの開発

安岡孝一(京都大学)・安岡素子(京都外国語大学)

はじめに

2026年2月16日に発表^{a)}されたQwen3.5は、最小0.8B(8億)パラメータの生成AIでありながら、マルチモーダルかつ多言語対応を謳っており、実際、日本語OCRとしても、そこそこの性能を発揮する。しかし、アイヌ語は読めない、多言語対応にアイヌ語が含まれておらず、アイヌ語の小書きカタカナ[1, 2]が読めないのだ(図1にGoogle Colaboratoryでのプログラムと結果を示す)。では、Qwen3.5にカタカナアイヌ語を追加学習するには、どのような手法が必要なのか。本稿では、カタカナアイヌ語OCRの開発という視点から、その手法を探る。

Qwen3.5-0.8Bへのトークン追加

Qwen3.5-0.8BのQwen2Tokenizerは、UTF-8のバイト列からトークンを組み上げる手法を採用しており、トークナイザ辞書にない文字に対しては、バイト列に対応するトークンを組み合わせる形で入出力をおこなう。たとえば、小書きカタカナ「プ」(UTF-8バイト列はe3 87 b7 e3 82 9a)は、トークン159(e3), トークン229(87), トークン115(b7), トークン3293(e3 82), トークン248(9a)を組み合わせて、[159, 229, 115, 3293, 248]という5つのトークン列で表現される。

しかし、複数のトークンで表現される文字は、入力

```
!pip install "transformers>=5.2.0" accelerate jinja2
img="https://qiita-image-store.s3.ap-northeast-1.amazonaws.com/0/428317/55d10bae-10b0-4a18-8e27-0907bc16361f.jpeg"
from transformers import pipeline
nlp=pipeline("image-text-to-text", "Qwen/Qwen3.5-0.8B", max_new_tokens=1024, device_map="auto")
d=nlp([{"role": "user", "content": [{"type": "image", "image": img}, {"type": "text", "text": "OCR Ainu sentences."}]}])
print(d[0]["generated_text"][1]["content"])
```

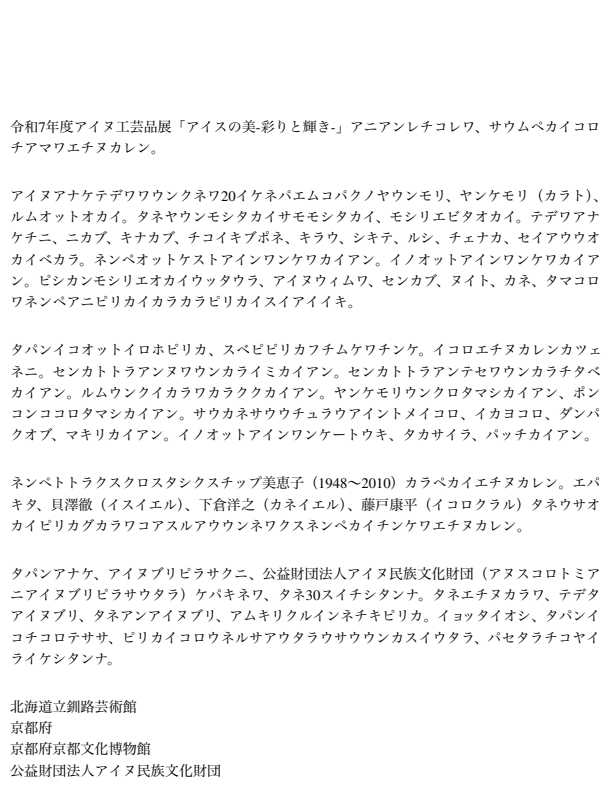
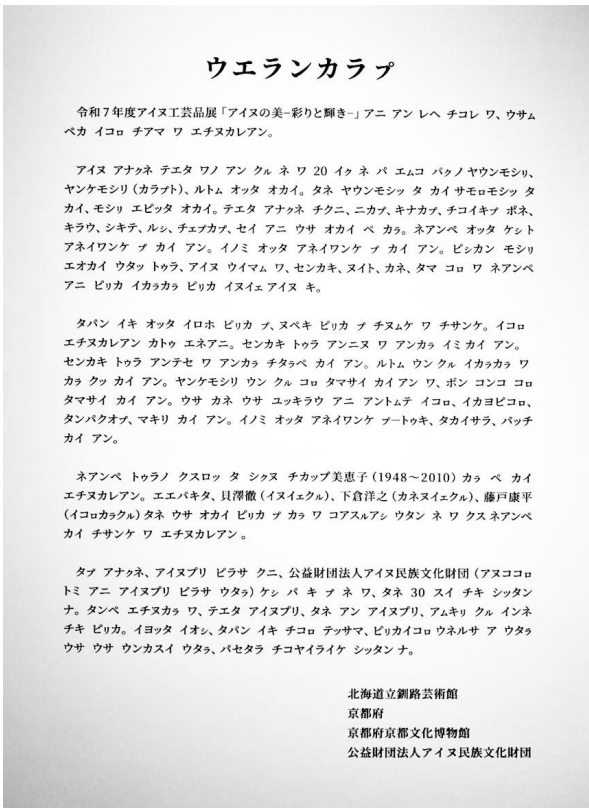


図1: 『アイヌの美—彩りと輝き—』ウエランカラブ(左)をQwen3.5-0.8BでOCRした結果(右)

^{a)}<https://qwen.ai/blog?id=qwen3.5>

はできて、めったに出力されない。図1右で「プ」などのいわゆるアイヌ語小書きカタカナが出力されていないのは、これが原因である(表1)。

表1: Qwen3.5-0.8B のトークン (小書きカタカナ)

ア	149368	へ	[159, 229, 116]
イ	54198	ホ	[159, 229, 117]
ウ	158936	ム	[159, 229, 118]
エ	74607	ヤ	65203
オ	149185	ユ	68850
カ	247021	ヨ	74469
ク	[159, 229, 108]	ラ	[159, 229, 119]
ケ	158810	リ	[159, 229, 120]
シ	[159, 229, 109]	ル	[159, 229, 121]
ス	[159, 229, 110]	レ	[159, 229, 122]
ツ	24404	ロ	[159, 229, 123]
ト	[159, 229, 111]	ワ	[2687, 106]
ヌ	[159, 158830]		
ハ	[159, 229, 113]		
ヒ	[159, 229, 114]		
フ	[159, 229, 115]		
プ	[159, 229, 115, 3293, 248]		

Qwen3.5-0.8B は、248320 種類のトークンを、それぞれ 1024 次元の「単語ベクトル」で表している。トークン 0~248319 のうち、0~248043 が UTF-8 バイト列によるトークン、248044~248076 が制御用の特殊トークン、248077~248319 が未使用である。ここにアイヌ語小書きカタカナを追加するにあたって、われわれは、特殊トークンを 248063~248095 に移動し、該当する「単語ベクトル」も同時に移動した。その上で、半濁点 (e3 82 9a) をトークン 248044 に追加し、アイヌ語小書きカタカナ UTF-8 のアタマ 2 バイト (e3 87) をトークン 248045 に追加し、「ク」「シ」「ス」「ト」「ヌ」「ハ」「ヒ」「フ」「へ」「ホ」「ム」「ラ」「リ」「ル」「レ」「ロ」をトークン 248046~248061 に追加し、さらに「プ」をトークン 248062 に追加した。Qwen2Tokenizer のトークナイザ辞書に「プ」(e3 87 b7 e3 82 9a) を追加するには、「フ」(e3 87 b7) と半濁点 (e3 82 9a) を組み上げなければならず、しかも特殊トークンをまたいではいけないため、こういう手順でのトークン追加が必要だった。なお、各トークンの「単語ベクトル」の初期値は、たとえば「プ」であれば、「プ」の「単語ベクトル」から「ヤ」を引いて「ヤ」を加えた値とした。

トークン追加をおこなうと、モデルの動作が不安定となる(一度「プ」を出力すると、そのまま無限に「プ」を出力しつづけてしまう)ため、さらに追加学習をおこなった。追加学習に用いたのは、国立アイヌ民族博物館の「ごあいさつ」(図3)を段落で区切った画像 65 枚と、大小カタカナの比較画像 17 枚である。なお、各画像の長辺が 1024 ドット以下になるよう、一部の画像は縮小した上で、全ての画像に読み情報を手作業でテキスト付与した。tr1 0.29.1 での追加学習 (3 epochs) に要した時間は、NVIDIA GeForce RTX 3080 Laptop (16GB) 1GPU で 19 分程度だった。

評価と考察

追加学習後のモデル Qwen3.5-0.8B-AinuOCR で、図1左の画像を OCR してみたところ、図2の結果が得られた。小書きの「プ」も出力されており、かなり良くなっている気がする。比較のため、原テキスト(「プ」を1文字とみなし、連続する空白と改行を1文字にまとめた結果、総文字数 1124 字)を手作業で準備し、レーベンシュタイン距離を見てみたところ、図1右が 412 (単純誤り率 36.7%) だったのに対し、図2は 144 (単純誤り率 12.8%) と、かなりの改善が見られた。

結果に気を良くしたわれわれは、トークン追加と追加学習を Qwen3.5-2B に適用し、さらに Qwen3.5-4B にも適用してみた。ただし、RTX 3080 LapTop の

ウエランカラ
令和7年度アイヌ工芸品展「アイヌ之美・彩りと輝き」-アニアレヘチコレワ、ウサムベカイコロ チアマワエヌカクレアン。
アイヌアナケネテエタワノアンクルネワ 20 イクネバエムコバノヤウンモシリ、ヤンケモシリ(カララト)、ルムトオツカイ。タネヤウンモシリタタ
カイ、モシリエビタオカイ。テエタアナケネチクニ、ニカブ、キナカブ、チオイキブボネ、キラウ、シケチ、ルシ、チエバカブ、セイアノウサオカイベカラ。ネアンベオタクシ
アネイワンケブカイアン。イノモオッタアネイワンケブカイアン。ピシカンモシリ
エオカイウタツトウラ、アイヌウイマムワ、センカヌ、ヌイト、カネ、タマコロワネアンベ
アニビリカイカラカラビリカイヌイアイヌキ。
タバンイキオッタイロホビリカワ、ヌベキビリカワチヌクワチサンケ。イコロ
エチヌカクレアンカトウエネアニ。センカヌトウラアンヌワアンカワイミカイアン。
センカヌトウラアンテセワアンカワチタラベカイアン。ルムトウクルイカラカラワ
カラクツカイアン。ヤンケモシリウクルコロタマサイカイアンワ、ボンコンコロ
タマサイカイアン。ウサカネウサヌチラウアニアントメイコロ、イカヨビコロ、
ダンバクオプ、マカリカイアン。イノモオッタアネイワンケパートウキ、タカサイラ、バッチ
カイアン。
ネアンベトウラノクスツタシクズチカブ美恵子 (1948~2010) カラベカイ
エチヌカクレアン。エバキタ、貝澤徹 (イスイエクル)、下倉砂之 (カネヌイエクル)、ウシクラン平
(イコロカクル) タヌウサオカイビリカブカラワコアスルアシウタンネワクスネアンベ
カイチサンケワエヌカクレアン。
タバアナケネ、アイヌブリピラサクニ、公益財団法人アイヌ民族文化財団 (アヌコロ
トミアニアヌブリピラサウタラ) ケシバキネフワ、タネ 30 スイチキシタン
ナ。タバエチチカラワ、テエタアイヌブリ、タネアンアイヌブリ、アムキリクルインネ
チキビリカ。イヨツタイオン、タバンイキチコロチサザマ、ビリカコロウヌサアウタラ
ウサウサウカスイウタラ、バセタラチコヤイライケシタンナ。
北海道立銅路芸術館
京都府
京都府京都文化博物館
公益財団法人アイヌ民族文化財団

図2: 図1左を Qwen3.5-0.8B-AinuOCR した結果

ウエランカラブ
令和7年度アイヌ工芸品展「アイヌの美・彩りと輝き」アニアンレヘチコレワ、ウサム
ベカイコロチアマワエチヌカレアン。
アイヌアナクネテエタワノアンクルネワ20イクネバエムコバツノヤウンモシリ、
ヤンケモシリ(カラプト)、ルトムオッタオカイ。タネヤウンモシツタカイサモロモシツタ
カイ、モシリエビツタオカイ。テエタアナクネチクニ、ニカヂ、キナカヂ、チコイキバボネ、
キラウ、シキテ、ルシ、チエツカヂ、セイアニウサオカイベカラ。ネアンベオッタケスト
アネイワンケンカイアン。イノミオッタアネイワンケンカイアン。ピシカンモシリ
エオカイウタツウラ、アイスウイマムワ、センカキ、ヌイト、カネ、タマコロワネアンベ
アニビリカイカラカラビリカイヌイエアイヌキ。
タパンイクオッタイロホビリカチ、ヌベキビリカチヌメケワチサンケ。イコロ
エチヌカレアンカトウエネアニ。センカキトウラアンニワアンカライミカイアン。
センカキトウラアンデセワアンカラチタラベカイアン。ルトムウンクルイカラカラワ
カラククカイアン。ヤンケモシリウンクルコロタマサイカイアンワ、ボンコンコロ
タマサイカイアン。ウサカネウサユキウラウニアントムテイコロ、イカヨビコロ、
タンバクオバ、マキリカイアン。イノミオッタアネイワンケートウキ、タカイサラ、パッチ
カイアン。
ネアンベトウラノクスロツタシツヌチカッブ美恵子(1948~2010)カラベカイ
エチヌカレアン。エエバキタ、貝澤徹(イヌイエクル)、下倉洋之(カネヌイエクル)、藤戸康平
(イコロカラクル)タネウサオカイビリカチカラワコアスルアシウタンネワクスネアンベ
カイチサンケワエチヌカレアン。
タナアナクネ、アイヌプリピラサクニ、公益財団法人アイヌ民族文化財団(アヌコロ
トミアアイヌプリピラサウタラ)ケシバキブネワ、タネ30スイチキシツタン
ナ。タンベエチヌカラワ、テエタアイヌプリ、タネアンアイヌプリ、アムキルクインネ
チキビリカ。イヨッタイオシ、タパンイキチコロテッサマ、ビリカイコロウネルサアウタ
ウサウサウンカスイウタラ、バセタラチコヤイライケシツタンナ。
北海道立銅路芸術館
京都府
京都府京都文化博物館
公益財団法人アイヌ民族文化財団

ウエランカラブ
令和7年度アイヌ工芸品展「アイヌの美・彩りと輝き」アニアンレヘチコレワ、ウサム
ベカイコロチアマワエチヌカレアン。
アイヌアナクネテエタワノアンクルネワ20イクネバエムコバツノヤウンモシリ、
ヤンケモシリ(カラプト)、ルトムオッタオカイ。タネヤウンモシツタカイサモロモシツタ
カイ、モシリエビツタオカイ。テエタアナクネチクニ、ニカヂ、キナカヂ、チコイキバボネ、
キラウ、シキテ、ルシ、チエツカヂ、セイアニウサオカイベカラ。ネアンベオッタケスト
アネイワンケンカイアン。イノミオッタアネイワンケンカイアン。ピシカンモシリ
エオカイウタツウラ、アイスウイマムワ、センカキ、ヌイト、カネ、タマコロワネアンベ
アニビリカイカラカラビリカイヌイエアイヌキ。
タパンイクオッタイロホビリカチ、ヌベキビリカチヌメケワチサンケ。イコロ
エチヌカレアンカトウエネアニ。センカキトウラアンニワアンカライミカイアン。
センカキトウラアンデセワアンカラチタラベカイアン。ルトムウンクルイカラカラワ
カラククカイアン。ヤンケモシリウンクルコロタマサイカイアンワ、ボンコンコロ
タマサイカイアン。ウサカネウサユキウラウニアントムテイコロ、イカヨビコロ、
タンバクオバ、マキリカイアン。イノミオッタアネイワンケートウキ、タカイサラ、パッチ
カイアン。
ネアンベトウラノクスロツタシツヌチカッブ美恵子(1948~2010)カラベカイ
エチヌカレアン。エエバキタ、貝澤徹(イヌイエクル)、下倉洋之(カネヌイエクル)、藤戸康平
(イコロカラクル)タネウサオカイビリカチカラワコアスルアシウタンネワクスネアンベ
カイチサンケワエチヌカレアン。
タナアナクネ、アイヌプリピラサクニ、公益財団法人アイヌ民族文化財団(アヌコロ
トミアアイヌプリピラサウタラ)ケシバキブネワ、タネ30スイチキシツタン
ナ。タンベエチヌカラワ、テエタアイヌプリ、タネアンアイヌプリ、アムキルクインネ
チキビリカ。イヨッタイオシ、タパンイキチコロテッサマ、ビリカイコロウネルサアウタ
ウサウサウンカスイウタラ、バセタラチコヤイライケシツタンナ。
北海道立銅路芸術館
京都府
京都府京都文化博物館
公益財団法人アイヌ民族文化財団

図4: 図1左を Qwen3.5-2B で OCR した結果(左)と Qwen3.5-2B-AinuOCR した結果(右)

ウエランカラブ
令和7年度アイヌ工芸品展「アイヌの美・彩りと輝き」アニアンレヘチコレワ、ウサム
ベカイコロチアマワエチヌカレアン。
アイヌアナクネテエタワノアンクルネワ20イクネバエムコバツノヤウンモシリ、
ヤンケモシリ(カラプト)、ルトムオッタオカイ。タネヤウンモシツタカイサモロモシツタ
カイ、モシリエビツタオカイ。テエタアナクネチクニ、ニカヂ、キナカヂ、チコイキバボネ、
キラウ、シキテ、ルシ、チエツカヂ、セイアニウサオカイベカラ。ネアンベオッタケスト
アネイワンケンカイアン。イノミオッタアネイワンケンカイアン。ピシカンモシリ
エオカイウタツウラ、アイスウイマムワ、センカキ、ヌイト、カネ、タマコロワネアンベ
アニビリカイカラカラビリカイヌイエアイヌキ。
タパンイクオッタイロホビリカチ、ヌベキビリカチヌメケワチサンケ。イコロ
エチヌカレアンカトウエネアニ。センカキトウラアンニワアンカライミカイアン。
センカキトウラアンデセワアンカラチタラベカイアン。ルトムウンクルイカラカラワ
カラククカイアン。ヤンケモシリウンクルコロタマサイカイアンワ、ボンコンコロ
タマサイカイアン。ウサカネウサユキウラウニアントムテイコロ、イカヨビコロ、
タンバクオバ、マキリカイアン。イノミオッタアネイワンケートウキ、タカイサラ、パッチ
カイアン。
ネアンベトウラノクスロツタシツヌチカッブ美恵子(1948~2010)カラベカイ
エチヌカレアン。エエバキタ、貝澤徹(イヌイエクル)、下倉洋之(カネヌイエクル)、藤戸康平
(イコロカラクル)タネウサオカイビリカチカラワコアスルアシウタンネワクスネアンベ
カイチサンケワエチヌカレアン。
タナアナクネ、アイヌプリピラサクニ、公益財団法人アイヌ民族文化財団(アヌコロ
トミアアイヌプリピラサウタラ)ケシバキブネワ、タネ30スイチキシツタン
ナ。タンベエチヌカラワ、テエタアイヌプリ、タネアンアイヌプリ、アムキルクインネ
チキビリカ。イヨッタイオシ、タパンイキチコロテッサマ、ビリカイコロウネルサアウタ
ウサウサウンカスイウタラ、バセタラチコヤイライケシツタンナ。
北海道立銅路芸術館
京都府
京都府京都文化博物館
公益財団法人アイヌ民族文化財団

ウエランカラブ
令和7年度アイヌ工芸品展「アイヌの美・彩りと輝き」アニアンレヘチコレワ、ウサム
ベカイコロチアマワエチヌカレアン。
アイヌアナクネテエタワノアンクルネワ20イクネバエムコバツノヤウンモシリ、
ヤンケモシリ(カラプト)、ルトムオッタオカイ。タネヤウンモシツタカイサモロモシツタ
カイ、モシリエビツタオカイ。テエタアナクネチクニ、ニカヂ、キナカヂ、チコイキバボネ、
キラウ、シキテ、ルシ、チエツカヂ、セイアニウサオカイベカラ。ネアンベオッタケスト
アネイワンケンカイアン。イノミオッタアネイワンケンカイアン。ピシカンモシリ
エオカイウタツウラ、アイスウイマムワ、センカキ、ヌイト、カネ、タマコロワネアンベ
アニビリカイカラカラビリカイヌイエアイヌキ。
タパンイクオッタイロホビリカチ、ヌベキビリカチヌメケワチサンケ。イコロ
エチヌ
カレアンカトウエネアニ。センカキトウラアンニワアンカライミカイアン。センカキトウ
ラアンデセワアンカラチタラベカイアン。ルトムウンクルイカラカラワカラククカイ
アン。ヤンケモシリウンクルコロタマサイカイアンワ、ボンコンコロタマサイカイ
アン。ウサカネウサユキウラウニアントムテイコロ、イカヨビコロ、タンバクオバ、マキリ
カイアン。イノミオッタアネイワンケートウキ、タカイサラ、パッチカイアン。
ネアンベトウラノクスロツタシツヌチカッブ美恵子(1948~2010)カラベカイエチヌカレ
ン。エエバキタ、貝澤徹(イヌイエクル)、下倉洋之(カネヌイエクル)、藤戸康平(イコロ
カラクル)タネウサオカイビリカチカラワコアスルアシウタンネワクスネアンベカイ
チサンケワエチヌカレアン。
タナアナクネ、アイヌプリピラサクニ、公益財団法人アイヌ民族文化財団(アヌコロトミア
アイヌプリピラサウタラ)ケシバキブネワ、タネ30スイチキシツタンナ。タネ
ベエチ
ヌカラワ、テエタアイヌプリ、タネアンアイヌプリ、アムキルクインネチキビリカ。イ
ヨッタ
イオシ、タパンイキチコロテッサマ、ビリカイコロウネルサアウタラウサウサ
ウ
ンカスイウタラ、バセタラチコヤイライケシツタンナ。
北海道立銅路芸術館 京都府 京都府京都文化博物館 公益財団法人アイヌ民族文化財団

図5: 図1左を Qwen3.5-4B で OCR した結果(左)と Qwen3.5-4B-AinuOCR した結果(右)

表 2: 各モデルで図 1 左を OCR した際の評価

	Qwen3.5-xB	Qwen3.5-xB-AinuOCR
0.8B	412 (36.7%)	144 (12.8%)
2B	122 (10.9%)	88 (7.8%)
4B	126 (11.2%)	88 (7.8%)
9B	217 (19.3%)	89 (7.9%)

※レーベンシュタイン距離 (単純誤り率)

おわりに

小書きカタカナを含むアイヌ語 OCR モデルとして、Qwen3.5-0.8B-AinuOCR, Qwen3.5-2B-AinuOCR, Qwen3.5-4B-AinuOCR を開発した。これらのモデルは <https://huggingface.co/KoichiYasuoka> で公開している⁹⁾ので、ぜひ使ってみてほしい。なお本稿は、京都大学未踏科学研究ユニット「フィールドワークとデータサイエンスを融合した超学際研究ユニット」の成果である。

参考文献

- [1] 佐藤知己: 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 (1996 年 11 月 25 日).
- [2] Addition of forty eight characters, JTC1/SC2/WG2 N2092 (September 13, 1999).
- [3] アカント ウン コタン —阿寒湖畔のアイヌ文化—, 白老: 国立アイヌ民族博物館 (2023 年 3 月 14 日).

ホシキ イタク

イッソロレ。
 ウボボイ オッタ エチシネエ クム バセ オンカムイ テラ オリバク イタク アンキ。
 第4回テーマ展示 アナタネ 「地域からみたアイヌ文化展 アカント ウン コタン —阿寒湖畔のアイヌ文化—」
 アニ アン レヘ チコレ ワ イコロ ウサム ベカ チアマ ワ エチヌカレアシ。
 タン テンブ オッタ アカント ウン コタン コタン ウン クル ビリカノ ハリキキ カト エチヌカレアシ
 ワ アカント ウン コタン タ アン アイヌプリ カイ ウキヨコマ カイ イカラ カイ ウボボ カイ リムセ
 カイ ウサオカイベ エチヌカレアシ ナ。
 アカン アナク 「シボロシリ シボロコタン」 ネ ワ 山本多助エカシ エネ ハウキ。 「カムイカラモシリ カムイカラ
 ビリカト アンワ シノツ ラムコロ ウワツチウタラ ウシノツ ウシネワ ヌボツイ クシリボツケイ ユ セセケイ
 ボロンノ アン」。 アカント ウン コタン オッタ イヌエ カイ チケメカラ カイ ウボボ カイ リムセ カイ
 アンヌカ ヲ ネ。 オロ タ コタン ウン クル アイヌプリ テイタプリ エオリバカ カネ アシップリ
 タネアンプリ アニ カラ ベ カイ オカイ。 タン テンブ オッタ コタン ウン クル ハリキキ カト エチヌカ
 エヤシカイ。
 シリーズ展示「地域からみたアイヌ文化展」オッタ カンナ カンナ アイヌモシリ タ ウサ アイヌ アン ワ ウサ
 コタン アン ワ ウサ アイヌイタク アン ワ ウサ アイヌプリ アン シリ エチヌカラ ナンコロ ナ。
 テエタ ワノ タネ タ バクノ アイヌプリ アン カト アンエイコシラムユイバ ルユイ。

(アイヌ語: 網路方言)
2023/4 3チュウ 14ト
アヌココ アイヌ イコロマケル サバネクル エナナンカラ

[4] アイヌ・モシリ, 才 3 号 (1958 年 3 月), p.7.

[5] 中川裕: アイヌ人によるアイヌ語表記への取り組み, 表記の習慣のない言語の表記 (東京外国語大学アジア・アフリカ言語研究所, 2006 年 3 月), pp.1-44.

付録 アイヌ語のカタカナに関して

本稿では、小書きカタカナを追加するやり方で、カタカナアイヌ語 OCR の開発をおこなったが、追加するのは小書きカタカナだけでいいのか。たとえば『アカント ウン コタン』[3] のホシキ イタク (図 3 右上) 末尾には、「エナナンカラ」という文字列が見える。館長の佐々木史郎(エトウナンカラ)が、この「ごあいさつ」では山本多助のアイヌ文字 [4, 5] を使用しており、「トウの合字」以外にも「シュの合字」「チャの合字」「イエの合字」が見える¹⁰⁾。

Qwen3.5 のトークナイザには、「トウ」(e3 83 88 e3 82 a5)を表すトークン 220263 が準備されており、これを使うという戦略もあるが、そうすると「トウ」という文字列と「ㇿ」の見分けがつかなくなる。その戦略は、「プ」と「ヅ」を見分けるべく頑張る、という本稿の主旨に反する。かと言って「ㇿ」をトークナイザ辞書に追加しようにも、この字が UCS に含まれていない(つまり UTF-8 が決まっていない)ので、現状どうしようもない。

そんなわけなので、このような合字を含む段落は、とりあえず、追加学習から取り除いておいた。結果として、われわれのアイヌ語 OCR は、これらの合字が読めない(図 6)。今後の課題としておきたい。

ホシキ イタク

イッソロレ。
 ウボボイ オッタ エチシネエ クム バセ オンカムイ ヤラ オリバク イタク アンキ。
 第4回テーマ展示 アナタネ 「地域からみたアイヌ文化展 アカント ウン コタン —阿寒湖畔のアイヌ文化—」
 アニ アン レヘ チコレ ワ イコロ ウサム ベカ チアマ ワ エチヌカレアシ。
 タン ヤンブ オッタ アカント ウン コタン コタン ウン クル ビリカノ ハリキキ カン エチヌカレアシ
 ワン アカント ウン コタン タ アン アイヌプリ カイ ウキヨコマ カイ イカラ カイ ウボボ カイ リムセ
 カイ ウサオカイベ エチヌカレアシ ナ。
 アカン アナク 「シボロシリ シボロコタン」 ネ ワ 山本多助エカシ エネ ハウキ。 「カムイカラモシリ カムイカラ
 ビリカト アンワ シノツ ラムコロ ウワツチウタラ ウシノツ ウシネワ ヌボツイ クシリボツケイ ユ セセケイ
 ボロンノ アン」。 アカント ウン コタン オッタ アイヌエ カイ チケメカラ カイ ウボボ カイ リムセ カイ
 アンヌカ ヲ ネ。 オロ タ コタン ウン クル アイヌプリ テイタプリ エオリバカ カネ アシップリ
 タネアンプリ アニ カラ ベ カイ オカイ。 タン ヤンブ オッタ コタン ウン クル ハリキキ カン エチヌカ
 エヤシカイ。
 シリーズ展示「地域からみたアイヌ文化展」オッタ カンナ カンナ アイヌモシリ タ ウサ アイヌ アン ワ ウサ
 コタン アン ワ ウサ アイヌイタク アン ワ ウサ アイヌプリ アン シリ エチヌカラ ナンコロ ナ。
 テエタ ワノ タネ タ バクノ アイヌプリ アン カン アンエイコシラムユイバ ルユイ。

(アイヌ語: 網路方言)
2023/4 3チュウ 14ト
アヌココ アイヌ イコロマケル サバネクル エナナンカラ

図 6: 『アカント ウン コタン』ホシキ イタク (左) を Qwen3.5-2B-AinuOCR した結果 (右)

⁹⁾各モデルのリポジトリに、開発に用いた maker.py も含めておいた。

¹⁰⁾<https://qiita.com/KoichiYasuoka/items/1be160bc15d73edf6593>