

# **A Toolkit for Analysis of Ainu Language**

**Michal Ptaszynski**<sup>1</sup> Fumito Masui<sup>1</sup> and  
Yoshio Momouchi<sup>2</sup>

1) Kitami Institute of Technology, Department of Computer Science

2) Professor Emeritus at Hokkai-Gakuen University

# A Toolkit for Analysis of Ainu Language

- Ainu language is a language of Ainu people, mostly living in northern Japan.
- Population of Ainu = about 23 thousand people.
- Number of native speakers = less than
- hundred (Hohmann, 2008).
- Ainu language is critically
- **endangered** (Moseley, 2010).
  
- Purpose of this research:
  - ↓ Create language analysis toolkit including POS tagger, translation support tool and shallow parser.
  - ↓ Help in linguistic and language anthropology research and support translators of Ainu texts.
- **Contribute to the process of reviving Ainu language.**



# A Toolkit for Analysis of Ainu Language

**DICTIONARY VIEW**

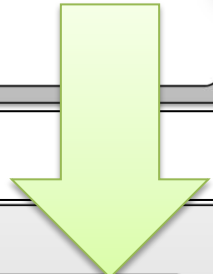
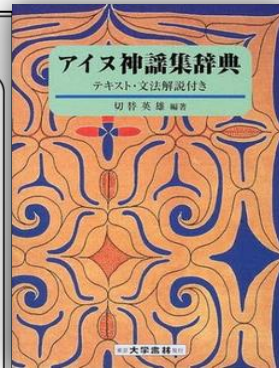
acikara [間投詞] (人をののしるとき用いる。原注 1-(3)、11-(1) 参照)  
interjection used in cursing or abusing somebody. Ref: 1-(3), 11-(1)

> acikara ta 汚い! 1-32. 11-46.  
filthy!

aehomatup [名詞] 「急変」(突然起こった事故) 2-46.  
noun sudden change

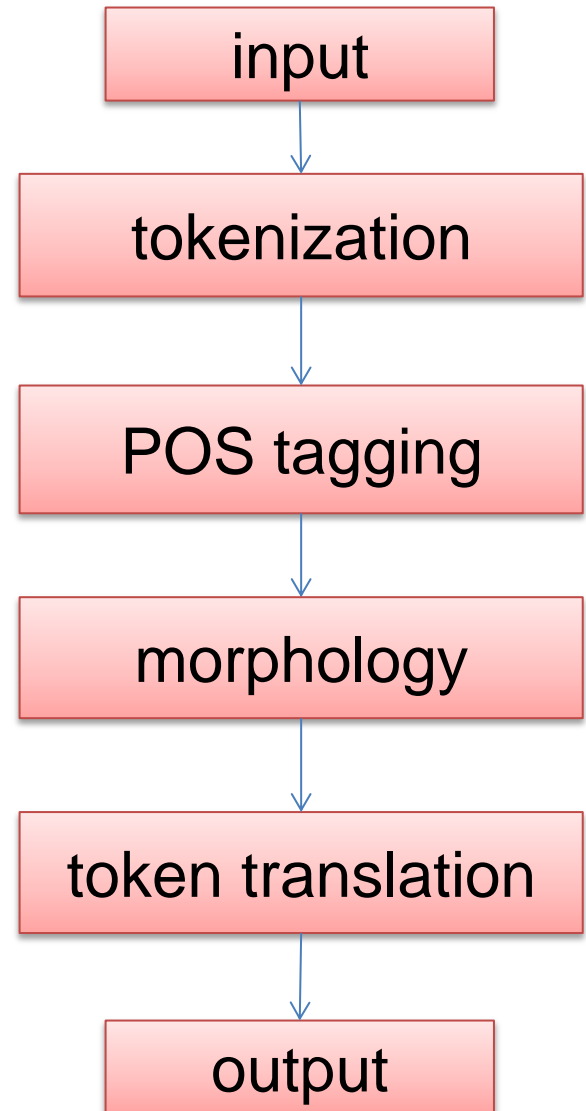
aekirusi [名詞] 「目次」  
noun contents

aep [名詞] 「食べ物」  
noun food



**DATABASE VIEW**

```
<word>acikara</word>
  <pos>間投詞</pos>
  <tr>人をののしるとき用いる</tr>
  <ex>acikara ta 汚い! </ex>
<word>aehomatup</word>
  <pos>名詞</pos>
  <tr>急変(突然起こった事故)</tr>
<word>aekirusi</word>
  <pos>名詞</pos>
  <tr>目次</tr>
<word>aep</word>
  <pos>名詞</pos>
  <tr>食べ物</tr>
```



# A Toolkit for Analysis of Ainu Language

- Output

**Sentence:**

Ci nukar wa ci eramesinne pet esoro hosippa as .

**Translation:**

When I saw this I was relieved and came back with the river current.

# A Toolkit for Analysis of Ainu Language

- Output
  - tokenization
  - POS tagging

## Sentence:

Ci nukar wa ci eramesinne pet esoro hosippa as .

## Translation:

When I saw this I was relieved and came back with the river current.

## BASELINE VERTICAL OUTPUT

ci	人称接語, 私(たち)が
nukar	二項動詞, ~が~を見る
wa	接続助詞, (…し)て
ci	人称接語, 私(たち)が
eramesinne	二項動詞, ~が~を見て安心する
pet	名詞, 川
esoro	後置詞的副詞, ~ (川)に沿って下流へ
hosippa	一項動詞の複数形, ~が帰る
as	人称接語, 私(たち)が
.	記号, .

# A Toolkit for Analysis of Ainu Language

- Output
  - tokenization
  - POS tagging
  - morphology

**Sentence:**  
Ci nukar wa ci eramesinne pet esoro hosippa as .

**Translation:**  
When I saw this I was relieved and came back with the river current.

**BASELINE VERTICAL OUTPUT**

ci	人称接語, 私 (たち) が
nukar	二項動詞, ~が~を見る
wa	接続助詞, (…し) て
ci	人称接語, 私 (たち) が
eramesinne	二項動詞, ~が~を見て安心する
pet	名詞, 川
esoro	後置詞的副詞, ~ (川) に沿って下流へ
hosippa	一項動詞の複数形, ~が帰る
as	人称接語, 私 (たち) が
.	記号, .

**TAMURA STANDARD WITH MORPHOLOGY ANALYSIS**

ci	人称接辞, 私 (たち) が
nukar	単他動詞, ~が~を見る
wa	接続助詞, (…し) て
ci	人称接辞, 私 (たち) が
eramesinne	単他動詞, ~が~を見て安心する
e	補充接頭辞, ~で (道具, 手段, 場所, 原因, 理由)
rame	名詞の変異形, 心 ……を導く)
ram	名詞, 心
sinne	自動詞; 助詞的普通の副詞, ~が落ち着いて
pet	名詞, 川 ……いる; まるで~のように
esoro	助詞的普通の副詞, ~ (川) に沿って下流へ
hosippa	自動詞複数形, ~が帰る
as	人称接辞, 私 (たち) が
.	記号, .

# A Toolkit for Analysis of Ainu Language

- Output
  - tokenization
  - POS tagging
  - morphology
  - POS standard change rules (5 standards)

**Sentence:**  
Ci nukar wa ci eramesinne pet esoro hosippa as .

**Translation:**  
When I saw this I was relieved and came back with the river current.

**BASELINE VERTICAL OUTPUT**

ci	人称接語, 私(たち)が
nukar	二項動詞, ~が~を見る
wa	接続助詞, (...し)て
ci	人称接語, 私(たち)が
eramesinne	二項動詞, ~が~を見て安心する
pet	名詞, 川
esoro	後置詞的副詞, ~(川)に沿って下流へ
hosippa	一項動詞の複数形, ~が帰る
as	人称接語, 私(たち)が
.	記号, .

**TAMURA STANDARD WITH MORPHOLOGY ANALYSIS**

ci	人称接辞, 私(たち)が
nukar	単他動詞, ~が~を見る
wa	接続助詞, (...し)て
ci	人称接辞, 私(たち)が
eramesinne	単他動詞, ~が~を見て安心する
e	補充接頭辞, ~で(道具, 手段, 場所, 原因, 理由)
rame	名詞の変異形, 心 ...を導く)
ram	名詞, 心
sinne	自動詞; 助詞的普通の副詞, ~が落ち着いて
pet	名詞, 川 ...いる; まるで~のように
esoro	助詞的普通の副詞, ~(川)に沿って下流へ
hosippa	自動詞複数形, ~が帰る
as	人称接辞, 私(たち)が
.	記号, .



# A Toolkit for Analysis of Ainu Language

- Output
  - tokenization
  - POS tagging
  - morphology
  - POS standard change rules (5 standards)
  - token translation

**Sentence:**  
Ci nukar wa ci eramesinne pet esoro hosippa as .

**Translation:**  
When I saw this I was relieved and came back with the river current.

---


**BASELINE VERTICAL OUTPUT**

ci	人称接語, 私(たち)が
nukar	二項動詞, ~が~を見る
wa	接続助詞, (...し)て
ci	人称接語, 私(たち)が
eramesinne	二項動詞, ~が~を見て安心する
pet	名詞, 川
esoro	後置詞的副詞, ~ (川)に沿って下流へ
hosippa	一項動詞の複数形, ~が帰る
as	人称接語, 私(たち)が
.	記号, .

---

**TAMURA STANDARD WITH MORPHOLOGY ANALYSIS**

ci	人称接辞, 私(たち)が
nukar	単他動詞, ~が~を見る
wa	接続助詞, (...し)て
ci	人称接辞, 私(たち)が
eramesinne	単他動詞, ~が~を見て安心する
e	補充接頭辞, ~で (道具, 手段, 場所, 原因, 理由
rame	名詞の変異形, 心 ...を導く)
ram	名詞, 心
sinne	自動詞; 助詞的普通の副詞, ~が落ち着いて
pet	名詞, 川 ...いる; まるで~のように
esoro	助詞的普通の副詞, ~ (川)に沿って下流へ
hosippa	自動詞複数形, ~が帰る
as	人称接辞, 私(たち)が
.	記号, .





# A Toolkit for Analysis of Ainu Language

- Output

- tokenization
- POS tagging
- morphology
- POS standard

**Sentence:**  
Ci nukar wa ci eramesinne pet esoro hosippa as .  
**Translation:**  
When I saw this I was relieved and came back with the river current.

**BASELINE VERTICAL OUTPUT**

ci	人称接語, 私(たち)が
nukar	二項動詞, ~が~を見る
wa	接続助詞, (...し)て
ci	人称接語, 私(たち)が
eramesinne	二項動詞, ~が~を見て安心する
pet	名詞, 川
esoro	後置詞的副詞, ~ (川)に沿って下流へ

chan WITH ABBREVIATED ENGLISH POS TAG NAMES AND TRANSLATION SUPPORT

(5 st) ci nukar wa ci eramesinne pet esoro hosippa as .  
paffix vt pj paffix vt n adv(postp) vi(pl) paffix sym  
私が見る て 私が 見て安心する 川 に沿って下流へ 帰る 私が .

- token



- horizontally/  
vertically

ci	人称接辞, 私(たち)が
eramesinne	単他動詞, ~が~を見て安心する
e	補充接頭辞, ~で(道具, 手段, 場所, 原因, 理由)
rame	名詞の変異形, 心 ...を導く)
ram	名詞, 心
sinne	自動詞; 助詞的普通の副詞, ~が落ち着いて
pet	名詞, 川 ...いる; まるで~のように
esoro	助詞的普通の副詞, ~ (川)に沿って下流へ
hosippa	自動詞複数形, ~が帰る
as	人称接辞, 私(たち)が
.	記号, .

# **A Toolkit for Analysis of Ainu Language**

**Michal Ptaszynski**<sup>1</sup> Fumito Masui<sup>1</sup> and  
Yoshio Momouchi<sup>2</sup>

1) Kitami Institute of Technology, Department of Computer Science

2) Professor Emeritus at Hokkai-Gakuen University