

Internet Applications for Endangered Languages: A Talking Dictionary of Ainu

Anna Bugaeva

1. Introduction: endangered languages and language documentation

There are an estimated 6,900 languages spoken in the world today and at least half of them are under threat of extinction. This is mainly because speakers of smaller languages are switching to other larger languages for economic, social or political reasons, or because they feel ashamed of their ancestral language. The language can thus be lost in one or two generations, often to the great regret of their descendants.

Over the past ten years a new field of study called “language documentation” has developed. Language documentation is concerned with the methods, tools, and theoretical bases for compiling a representative and lasting multipurpose record of languages. It has developed in response to the urgent need to make an enduring record of the world’s many endangered languages and to support speakers of these languages in their desire to maintain them. It is also fueled by developments in information and media technologies which make documentation and the preservation and dissemination of language materials possible in ways which could not previously be envisioned. The roles, needs and rights of language speakers and communities are also of central concern (<http://www.hrelp.org/>).

A number of the world’s famous universities have recently developed dedicated language documentation programmes and the supporting digital archives for preserving the data on endangered languages. A key role in these activities belongs to the Endangered Languages Documentation Programme (ELDP) of School of Oriental and African Studies (SOAS), University of London. ELDP was established in 2004 with a commitment of £20 million from Arcadia Trust to document as many endangered languages as possible

and to encourage the development of relevant skills across the world. Since then ELDP has offered up to £1 million in grants each year for the documentation of endangered languages across the world.

2. Internet applications for endangered languages: digital archiving vs. dissemination of digital materials via the internet

ELDP works in close cooperation with the Endangered Languages Archive (ELAR) which is a state of the art digital endangered language archive which catalogues, stores and makes accessible the documentations and descriptions of endangered languages resulting from the work of research grantees, students and others. ELAR at SOAS now has language documentation materials on over 70 endangered languages throughout the world and is expecting materials from at least 80 other ELDP-funded grantees; the present total data is about 8 TB in about 100,000 files including audio, video, transcriptions and other texts. ELAR aims to:

- provide a safe long-term repository of language materials
- enable people to see what documentation has been created for a language
- encourage international co-operation between researchers
- provide advice and collaboration (<http://www.hrelp.org/archive/>).

Although, typically, a digital archive such as ELAR will provide World Wide Web access to a catalogue of materials (<http://elar.soas.ac.uk/>) and, where appropriate, access to materials themselves, it should be emphasized that archiving is generally an entirely

native culture and language has changed to positive after the official adoption of “The Law for the Promotion of the Ainu Culture and for the Dissemination and Advocacy for the Traditions of the Ainu and the Ainu Culture” (1997) and the official recognition of the Ainu as the indigenous population of Hokkaido (2008).

Established in accordance with this law, was the Foundation for Research and Promotion of Ainu Culture which is, among other promotion activities, running fourteen Ainu language schools across Hokkaido in all regions with a high concentration of Ainu population and one school in Tokyo at the Ainu Cultural Center.

3.3. Previous description and research of Ainu

More than a century has passed since linguistic research of Ainu was begun, and this research has produced a few comprehensive dictionaries: Batchelor (1938), Chiri (1953, 1954 and 1962), Hattori (1964), Nakagawa (1995), Tamura (1996), and Kayano (1996), several Ainu grammars of Sakhalin (Murasaki 1979), Saru (Kindaichi 1931; Tamura 1988), Horobetsu (Chiri 1936), Shizunai (Refsing 1986) and Chitose (Bugaeva 2004; Satoo 2008) dialects and a few articles on some grammatical phenomena. Yet, none of those grammars is complete, as we are still at a rather early stage of

Ainu research.

Ainu was not a written language, but the Ainu folklore is extremely rich. Many Ainu texts were recorded by Japanese (K. Kindaichi, S. Tamura, K. Murasaki, H. Kirikae, H. Nakagawa, T. Satoo, O. Okuda and others), Ainu (Y. Chiri, M. Chiri, S. Kayano), English (J. Batchelor), Danish (K. Refsing), and Russian (N.A. Nevskij, A. Bugaeva, M.M. Dobrotvorskij, B. Pilsudski - the last two of Polish origin) scholars in Latin, Cyrillic and Japanese (katakana) alphabets. In fact, Ainu is a relatively well-documented language to be compared with other endangered languages, however, it should be emphasized that very rarely are Ainu texts also accompanied by the audio files (the exceptions are Tamura (1984... 2000), Kayano (1999), Bugaeva (2004) and some others). Therefore, generally, there is a very strong demand for Ainu texts of a ‘new generation’, i.e. fully glossed and annotated texts with both Latin and katakana transcriptions of Ainu, Japanese and English translations and attached audio and video files.

4. ELDP-funded project “Documentation of the Saru dialect of Ainu” (2007-2009) IPF0128 (headed by A. Bugaeva)

In my ELDP project I proposed to create a digital cor-



Figure 2. Photograph of Mrs. Kimi Kimura (1900-1988).

pus with an easy web-access to Ainu data for linguists and for the members of Ainu community who over the last few years, have been experiencing a new sense of self-identity and are now in the process of reviving their culture and language (<http://www.hrelp.org/grants/projects/index.php?projid=124>).

Within the project, firstly, I was concerned with the creation (in cooperation with Prof. Nakagawa, Chiba University) of a digital web-accessible corpus which will consist of the previously unpublished audio recordings of Ainu folktales (genres *uweperker* 'prosaic folktale' and *kamuy yukar* 'divine epic') of the Saru Ainu dialect (South Hokkaido) which Prof. Nakagawa personally recorded in 1977-1988 with a very talented speaker and story-teller, Mrs. Kimi Kimura (1900-1988, born at Penakori Village, upper district of the Saru River) whose proficiency in Ainu considerably surpassed that of Japanese. The total recording time of the Ainu folktales (running text) to be deposited at the digital archive is about 7 hours (Nakagawa and Bugaeva, forthcoming at <http://lah.soas.ac.uk/projects/ainu/>). Importantly, the processing of these "old Ainu texts" would not have been possible without consultations with available native speakers of Ainu and it was probably our very last chance to produce an adequate reliable documentation of Ainu.

Secondly, the project was aimed at collecting all possible kinds of "new Ainu texts" with all available speakers. Although it was possible to collect some new folktales, the collection of conversational Ainu data appeared to be problematic due to the geographi-

cal location, older age and poor health condition of speakers, which made me propose the following substantial changes in original documentation project.

5. Documentation with revitalization in mind: "A talking dictionary of Ainu"

5.1. Origins, aims and primary assets of the project

In November 2007, I attended the 11th annual Ainu speech contest at the town of Shiranuka, Hokkaido. Here, I experienced a high level of presentations in Ainu and a community passion for learning the language, and at the same time recognized a very strong demand for Ainu conversational audio recordings because most previous Ainu documentation focused on recording folklore texts. For this reason, I decided to divert some funds from my ELDP project to support revitalization efforts through production of web-accessible materials of conversational Ainu. This was supported with joint efforts by the members of the Ainu community.

The aims were, firstly, to produce an easy-accessible and easy-usable multimedia with audio materials of conversational Ainu in order to meet the actual needs of the community as a primary audience and, secondly, to additionally provide the linguistic community with the underrepresented conversational Ainu data that is accountable for further linguistic analysis. It is obvious that there are basic differences in the structure and degree of complexity between colloquial genres of Ainu and folklore, mainly poetic, genres of

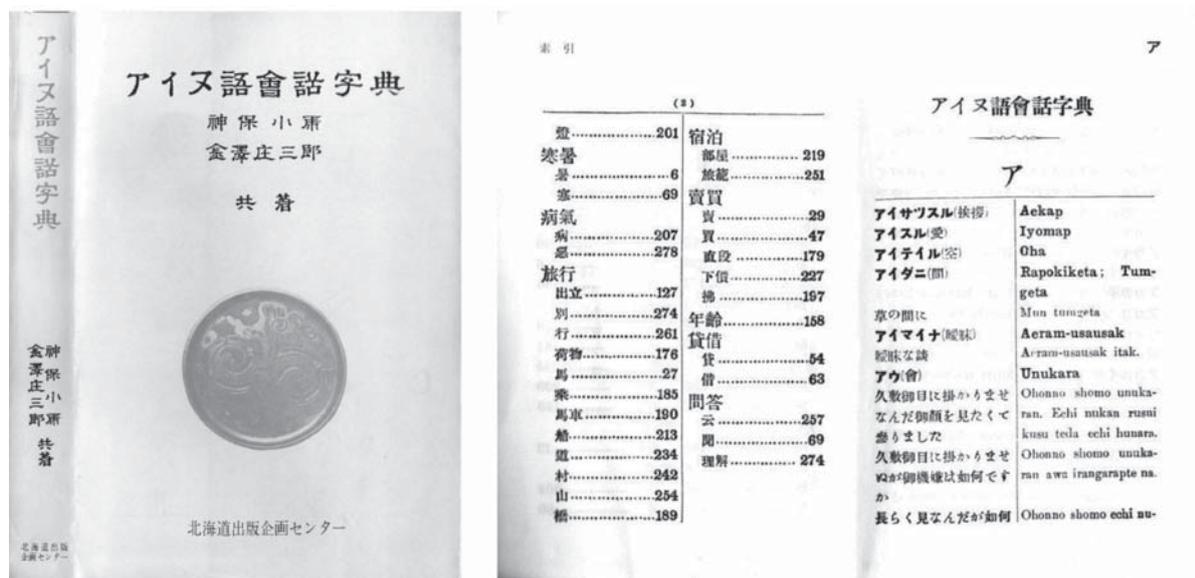


Figure 3. Book Cover of *Ainugo kaiwa jiten* [Ainu conversational dictionary].

Ainu.

As mentioned, collecting entirely new conversational Ainu data has been difficult, so I have selected an available written source as primary asset, namely Jinbo, K. and Kanazawa, S. (first edition: 1898) *Ainugo kaiwa jiten* [Ainu conversational dictionary]. Tokyo: Kinkoodoo Press (278 pp).

The dictionary was compiled by Shoozaburoo Kanazawa a postgraduate in linguistic studies at Tokyo University, with the supervision of Prof. Kotora Jinbo, a geologist actively promoting research on the Ainu language. Kanazawa visited Hokkaido about four times between 1895 and 1897, for a total of 150 days. The dictionary was compiled by selecting only commonly used words. Prof. Kotora Jinbo writes: “I have read this book extensively and added some notes and suggestions, allowing me to add my name to the cover of this dictionary.” (Jinbo and Kanazawa 2001: 1) The dictionary contains 3,847 entries, i.e. conversational phrases or words, and most of them presumably belong to the Saru dialect.

5.2. Shortcomings of the primary asset of the project

As the primary author Shoozaburoo Kanazawa notes, “there are many shortcomings. I was not able to spend much time in the editing process, and many word arrays are probably broken up. I eagerly expect corrections which may be made later.” (Jinbo and Kanazawa 2001: 1) Indeed, there are many transcription mistakes and misinterpretations, and the orthography of Ainu

(with Latin alphabet) and Japanese is far too outdated. This makes this precious source on conversational Ainu absolutely incomprehensible for the Ainu community members. The dictionary style follows the German Meyer’s Sprachfuher, and functions as both a dictionary (Japanese translations are listed in the *kana*-alphabetical order) and conversation phrase dictionary. When using as a conversation phrase dictionary, the user must find the topic of conversation from the index, however, it appears that half of the words are omitted from the Topical index.

5.3. Actual tasks and workflow

This project was started with the help of Mrs. Setsu Kurokawa (85), a good speaker of Saru dialect (Nukibetsu).

Together, the following tasks were completed:

- Typing original content (Ainu and Japanese translations) of the dictionary into a database.
- Correcting original Latin transcriptions and translations and adding them to the database.
- Preparing *katakana* transcriptions of Ainu for community members and adding them to the database.
- Making Ainu recording of the dictionary.
- Cutting the audio files cut into separate entries (3847 pieces).
- Making English translations.
- Transcribing the recordings of Mrs. Setsu Kurokawa as if they were completely new texts, since actual utterances deviated considerably from the



Figure 4. *Ainu fieldwork: Mrs Setsu Kurokawa (Ainu language consultant) and Anna Bugaeva (linguist). Hokkaido, Nukibetsu, 7 September 2008.*

¥ref 2196					
¥or-ft-j	どこから来たか				Japanese translation as in <i>Ainugo kaiwa jiten</i> (original text used)
¥or-tx	Nak wa ek?				Ainu transcription in the Roman alphabet as in <i>Ainugo kaiwa jiten</i> (original text used)
¥tx	hunak wa e= ek?				Ainu text by Setsu Kurokawa (Latin transcription)
¥mb	hunak wa e= ek				Morphological boundaries
¥ge	where from 2SG.S= come.SG				Morpheme-to-morpheme interpretation
¥ps	n.interr conj ppx vi				Categorization of parts of speech (English)
¥ft-e	Where did you come from?				English translation
¥kana	フナク ワ エ=エク?				Ainu language (<i>katakana</i> transcription)
¥wj	どこ ~から 君は 来る				Word-to-word interpretation (Japanese)
¥ps-j	【疑問】【格助】【人接】【自】				Categorization of parts of speech (Japanese)
¥ft-j	君はどこから来たの?				Modern Japanese translation
¥tp-j	問答				Topic categorization (Japanese)
¥sbtp-j	どこ				Sub-topic categorization (Japanese)
¥tp-e	Question and Answer				Topic categorization (English)
¥sbtp-e	Where				Sub-topic categorization (English)

Figure 5. An example of data in Toolbox.

original dictionary: Mrs. Setsu Kurokawa often corrected mistakes in the original; see Figure 5, cf. ¥tx and ¥or-tx lines, note a grammatically correct obligatory use of the person-number verbal cross-referencing prefix *e=* <2SG.S> in ¥tx and a use of the different interrogative particle *hunak* ‘where’ which are absent in ¥or-tx.

- Transferring all the data from Excel into Toolbox⁽¹⁾, interlinearization and annotation of Mrs. Setsu Kurokawa’s recorded texts in Toolbox.
- Undertaking an additional fieldtrip for consulting on the meanings of some unclear entries. Investigating the history of the dictionary compilation.
- Assigning all entries to the preexisting Topical categories/subcategories and creating new Topical categories/subcategories.

5.4. Evaluation and evolution of the project

Key issues for developing multimedia included:

- the importance of input from community at all stages of the project and possible interim and final evaluation by the community members;
- the importance of hiring a programmer experienced or willing to develop multimedia for endangered languages: inviting David Nathan (SOAS, University of London), a media programmer with a background in linguistics, to cooperate with the project;
- “do not attempt to make a significant multimedia product without a graphic designer.” (Nathan 2004: 158)

In the beginning, a tentative multimedia product “Ainu morning talk” which draws upon a piece of my Toolbox data and the respective audio recordings by Mrs. Setsu Kurokawa was prepared and presented to a group of Ainu community members studying Ainu.

Strong positive reactions and valuable feedback, as well as kind words of encouragement, were received. Ainu community members suggested that I should also provide for Ainu phrases a word-to-word interpretation in Japanese, categorization of parts of speech and modern Japanese translations.

5.5. Outcome

As an outcome of the ELDP project a web-accessible Ainu-Japanese-English conversational dictionary (*A Talking Dictionary of Ainu: A New Version of Kanazawa’s Ainu Conversational dictionary* [音声付きアイヌ語辞典—新編 金澤版アイヌ語会話辞典]) has been created in collaboration with Japanese co-editor Shiho Endo (Chiba University) and programmer David Nathan (ELAR director, SOAS, University of London) and with an art input of the Ainu community, i.e. web-design by Tamami Kaizawa and photography by Koichi Kaizawa; released in June 2010.

The Web dictionary has two possible views: Community View with *katakana* transcriptions of Ainu, a word-to-word interpretation in Japanese, categorization of parts of speech in Japanese and modern Japanese translations for the Ainu community as a major target audience and Linguist View with the information about morphemic boundaries, glosses,



Figure 6. A screenshot of the tentative multimedia product “Ainu morning talk” made by Anna Bugaeva and David Nathan at the Documentary Linguistics Workshop (Tokyo University of Foreign Studies, 9-13 February 2009) and its evaluation by a group of Ainu community members who attend advanced Ainu classes held by Prof. Nakagawa (Ainu Cultural Center in Tokyo, 28 February 2009).

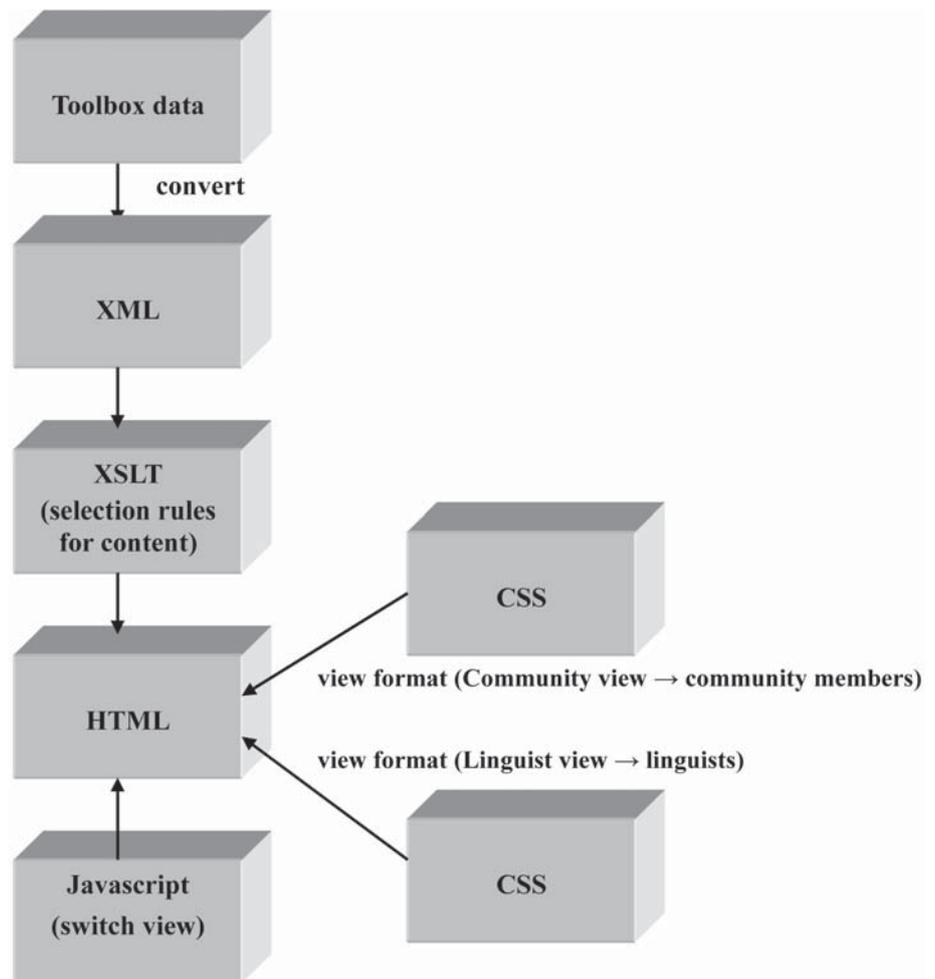


Figure 7. Architecture of A Talking Dictionary of Ainu: A New Version of Kanazawa's Ainu Conversational dictionary 『音声付きアイヌ語辞典—新編 金澤版アイヌ語会話辞典』

parts of speech and English translations for linguists, see Figure 8, cf. Figure 5. Almost all entries are accompanied by audio files recorded with the speaker of Ainu Mrs. Setsu Kurokawa.

A *Talking Dictionary of Ainu* has been deposited to the archive (ELAR) for a safe long-term preservation (<http://elar.soas.ac.uk/>) and disseminated via the internet for enabling easy access to the language materials by Ainu community and linguists (<http://lah.soas.ac.uk/projects/ainu/>).

In addition to the internet dissemination, we are also planning to publish with Hokkaido Kikaku Sentai

a paper edition of the dictionary with a CD-ROM to facilitate its use by those community members who live in remote areas and do not have internet access.

The Ainu language materials will provide an important and irreplaceable resource in the future for Ainu community members who want to find out more about their heritage, and for linguists and other researchers who want to understand more about the diversity of human knowledge and experience.

6. Concluding remarks

- The idea of “a talking dictionary” is not new (e.g.

A talking dictionary of Ainu: a new version of Kanazawa's Ainu conversational dictionary
with recordings of Mrs Setsu Kurokawa
金澤版 アイヌ語会話辞典・音声 黒川セツさん

Under development July 2010, please visit again soon
Last updated 15 July 2010 DJN

Introduction 序文 Dictionary - 辞典

Editors: Anna Bugaeva and Shiho Endo (編集) ブガエワ・アンナ、遠藤志保
Speaker: Setsu Kurokawa (発音) 黒川セツ
Multimedia developer: David Nathan (マルチメディア開発) ネイザン・デイヴィッド

Community view

私は百姓で御坐りますあちらから来ました
Kuani anakne toi-ta guru ku ne ruwe ne; toani wa ek. 2195

kuani anakne toyta kur ku= ne ruwe ne. 2195
クアニ アナクネ トイタクルク=ネルウエネ。
わたしは 耕作する人 わたしは ~である ことである
【代名】【助動】【自】【名】【入接】【他】【形名】【他】
「私は百姓です。」 'I am a farmer/peasant.'

どこから来たか
Nak wa ek? 2196

hunak wa e= ek? 2196
フナクワ エ=エク?
どこ ~から 君は 来る
【疑問】【格助】【入接】【自】
「君はどこから来たの?」 'Where did you come from?'

.....から登りました
.....wano k'ek. 2197

Pipausi kotan or wano k= ek. 2197
ピバウシ コタン オロワノ ケク(ク=エク)。
ピバウシ 村 ところ ~から わたしは 来る
【固名】【名】【位名】【格助】【入接】【自】
「ピバウシから来ました。」 'I came from the village Pipausi.'

Linguist view

私は百姓で御坐りますあちらから来ました 2195
Kuani anakne toi-ta guru ku ne ruwe ne; toani wa ek. 2195

kuani anakne toyta kur ku= ne ruwe ne. 2195
ku-an-hi anakne toy-ta kur ku= ne ruwe ne.
1SG.S-exist.SG-thing/person TOP-COP earth-dig person 1SG.A= COP INFR.EV COP
prn adv.prt vi n ppx vt nmiz vt
'I am a farmer/peasant.' 「私は百姓です。」

どこから来たか 2196
Nak wa ek? 2196

hunak wa e= ek? 2196
hunak wa e= ek?
where from 2SG.S= come.SG
n.interr conj ppx vi
'Where did you come from?' 「君はどこから来たの?」

.....から登りました 2197
.....wano k'ek. 2197

Pipausi kotan or wano k= ek. 2197
Pipausi kotan or wa-no ku= ek.
Pipausi village place from-ADV 1SG.S= come.SG
prp.n n nl pp ppx vi
'I came from the village Pipausi.' 「ピバウシから来ました。」

Figure 8. Screenshots of *A Talking Dictionary of Ainu: A New Version of Kanazawa's Ainu Conversational dictionary* 『音声付きアイヌ語辞典—新編 金澤版アイヌ語会話辞典』 (<http://lah.soas.ac.uk/projects/ainu/>)

Hercus and Nathan 2002) since such dictionaries have already been developed for some endangered languages of Australia⁽²⁾, Pacific⁽³⁾, North America⁽⁴⁾ and Asia⁽⁵⁾. However, it is definitely a first attempt of this kind for Ainu.

- The project emerged as a direct response to a strong demand for Ainu conversational materials in the Ainu community which is now in the process of revitalizing its language and culture.
- “Multimedia projects are typically more time consuming and expensive than other activities” (Nathan 2004: 156). However, it was chosen here because multimedia provides the most effective way to mobilise language materials via the internet. In my paper, I have shown how a large-scale (3,847 entries) multimedia product may possibly be created in two years.
- Not only communality’s initiation but also its support and ongoing participation are important for the success of a community-oriented project.
- Working together with an Ainu language consultant Mrs. Setsu Kurokawa, I was able to revise and give new life to old but very precious materials which were otherwise incomprehensible to a broad audience because of outdated transcriptions and orthography in Japanese translations and because of numerous misinterpretations.
- The fact that Mrs. Setsu Kurokawa’s productive skills in Ainu considerably surpassed my expectations proves that it is never too late to give your fieldwork another chance in the case of moribund languages!

NOTE

- (1) Toolbox is a data management and analysis computer program (software) for field linguists. It is especially useful for maintaining lexical data, and for parsing and interlinearizing text, but it can be used to manage virtually any kind of data.
- (2) Yuwaalaraay at <http://lah.soas.ac.uk/projects/gw/>
- (3) Khinina-ang Bontok (Philippines): <http://htq.minpaku.ac.jp/databases/bontok/>
- (4) [http://www.nativeweb.org/resources/languages_linguistics/native_american_languages/and_especially_Lenape_\(Delaware,_USA\)_http://www.talk-lenape.org/](http://www.nativeweb.org/resources/languages_linguistics/native_american_languages/and_especially_Lenape_(Delaware,_USA)_http://www.talk-lenape.org/)
- (5) Yami (Taiwan) at <http://yamibow.cs.pu.edu.tw/>

References

- Batchelor, John (1938) 1995. *An Ainu-English-Japanese dictionary*. Tokyo: Iwanami Shoten.
- Bugaeva, Anna (2004) *Grammar and folklore texts of the Chitose*

- dialect of Ainu (Idiolect of Ito Oda)*. + 1 audio CD. ELPR Publication Series A-045. Suita: Osaka Gakuin University.
- Bugaeva, Anna and Shiho Endo (eds.), speaker: Setsu Kurokawa, multimedia developer: David Nathan (2010) *A Talking Dictionary of Ainu: A New Version of Kanazawa’s Ainu Conversational dictionary*. Japanese title: 『音声付きアイヌ語辞典—新編 金澤版アイヌ語会話辞典』 [Onsei-tsuki ainugo jiten: shinhen Kanazawa-han ainugo kaiwa jiten], SOAS, University of London <http://lah.soas.ac.uk/projects/ainu/>
- Chiri, Mashiho (1936) 1974. *Ainu gohoo gaisetu* (An outline of Ainu grammar). *Chiri Mashiho chosakushuu*, v. 4, 3-197. Tokyo: Heibonsha.
- Chiri, Mashiho (1953, 1954 and 1962) 1975 and 1976. *Bunrui ainugo jiten* [A classificational dictionary of the Ainu language], vol. 1-3, reprinted by Tokyo: Heibonsha.
- Jinbo, Katora and Kanazawa, Shoozaburoo (2001) *Ainu go kaiwa jiten* [Ainu conversational dictionary]. Sapporo: Hokkaido kikaku sentaa. First edition: (1898), Tokyo: Kinkoodoo Press.
- Hattori, Shiroo (ed.) (1964) *An Ainu dialect dictionary*. Tokyo: Iwanami Shoten.
- Hercus, Luise A. and Nathan, David (2002) Paakantjy. Multimedia CD-ROM. Canberra: ATSIC
- Kayano, Shigeru (1996) *Kayano Shigeru no ainugo jiten* [An Ainu dictionary by Kayano Shigeru]. Tokyo: Sanseido.
- Kayano, Shigeru (1998) *Kayano no ainu shinwa shuusei* [A collection of Ainu myths by Kayano]. 1-10 vol, Tokyo: Heibonsha.
- Kindaichi, Kyoosuke (1931) 1993. *Ainu yuukara gohoo tekiyoo* [An outline grammar of the Ainu epic poetry]. *Ainu jojishi yuukara no kenkyuu* 2, 1-233, Tokyo: Tooyoo Bunko. Reprinted in 1993. *Ainugogaku koogi* 2 (Lectures on Ainu studies 2). *Kindaichi Kyoosuke zenshuu. Ainugo I*, v. 5, 145-366. Tokyo: Sanseido.
- Murasaki, Kyooko (1979) *Karafuto ainugo. Bunpoo-hen* [Sakhalin Ainu. Grammar volume]. Tokyo: Kokushokan-kookai.
- Nakagawa, Hiroshi 1995. *Ainugo chitose hoogen jiten* (A dictionary of the Chitose dialect of Ainu). Tokyo: Soofuukan.
- Nakagawa, Hiroshi and Anna Bugaeva (forthcoming) *A web-accessible corpus of folktales of the Saru dialect of Ainu by Mrs. Kimi Kimura (1900-1988)*, London University, SOAS, ELDP at <http://lah.soas.ac.uk/projects/ainu/>
- Nathan, David (2004) Planning multimedia documentation. In: Peter K. Austin (ed.) *Language documentation and description*, vol. 2, London: SOAS, ELDP, 154-168.
- Refsing, Kirsten (1986) *The Ainu language. The morphology and syntax of the Shizunai dialect*. Aarhus: Aarhus University Press.
- Satoo, Tomomi. 2008. *Ainugo bunpoo no kiso* [The basics of Ainu grammar]. Tokyo: Daigakushorin.
- Tamura, Suzuko. (1984...2000). *Ainugo shiryoo 1-12*. [Ainu audio materials 1-12]. Tokyo: Waseda daigaku gogakuyo-oiku kenkyuujo.
- Tamura, Suzuko. (1988). *Ainugo* [The Ainu language]. *Gengogaku daijiten* [Encyclopedia of linguistics]. Kooji Takashi, Rookoo Koono and Eiichi Chino (eds.). Tokyo: Sanseido.
- Tamura, Suzuko. 1996. *Ainugo Saru hoogen jiten* [A dictionary of the Saru dialect of Ainu]. Tokyo: Soofuukan.