



HOKKAIDO UNIVERSITY

Title	Observations on “Northeastern” Hokkaido Ainu dialects : a statistical perspective
Author(s)	Ono, Yohei
Citation	北方言語研究, 9, 95-122
Issue Date	2019-03-15
DOI	https://doi.org/10.14943/88224
Doc URL	https://hdl.handle.net/2115/73722
Type	departmental bulletin paper
File Information	06_ono.pdf



Observations on “Northeastern” Hokkaido Ainu Dialects: A Statistical Perspective

Yohei ONO

(Graduate Student at the Open University of Japan)

Keywords: Ainu, Homogeneity Analysis, lexicostatistics, Neighbor-Net, ordinal scale.

1. Introduction

Hattori and Chiri (1960) recorded the lexicostatistical data of 19 Ainu dialects whose disappearance was imminent in the 1960's. Since then, statistical analyses (Asai 1974; Lee and Hasegawa 2013; Ono 2015) have been conducted based on their data. However, Ono (2019, to appear) indicated that previous statistical analyses results were obscured by the assumption of a nominal scale in Hattori and Chiri's (1960) data and that a statistical analysis assuming an ordinal scale succeeded in capturing the underlying information structure in their study, employing Biratori and Samani dialects as examples.¹

This paper focuses on the results of an analysis of five northeastern Hokkaido dialects that were omitted due to space restrictions in Ono (2019, to appear). Therefore, for details regarding the statistical methods utilized in this paper, the interested reader can refer to Ono (2015; 2019, to appear).

Since Asai (1974) classified Ainu dialects into three groups—Hokkaido Ainu dialects, typified by Nos. 1–13 in Figure 1, Northern Kurile dialects, typified by No. 22, and Sakhalin dialects, typified by Nos. 14–19—research on the Ainu language has made greater progress on southwestern Hokkaido Ainu dialects typified by Nos. 1–6, as compared to the other groups (Bugaeva 2012).

Although Asai (1974) classified the five northeastern Ainu dialects (i.e., Nos. 8–12 in Figure 1) that the author examines in this paper as belonging to the East Hokkaido group, previous statistical analyses have not succeeded in visualizing the underlying relationships among them based on Hattori and Chiri's (1960) data. This

¹ As the author will discuss in Section 2, Hattori and Chiri (1960) recorded the lexicostatistical data in the linguistic environments of Ainu in 1960's with various “symbols.” Therefore, statistical analyses based on their data require some quantification methods (i.e., to assign some “appropriate” numerical values to each symbols in terms of statistics). The different assumptions of scale type (i.e., ordinal scale and nominal scale in this paper) are related to the restriction of the quantification. The assumption of an ordinal scale constraints the quantification on the symbols on ordinal ranking as researchers impose but the assumption of a nominal scale does not. Ono (2019, to appear) dealt with this issue on the lexicostatistical data in Hattori and Chiri (1960) in detail.

paper applies statistical analysis assuming an ordinal scale—the same approach adopted in Ono (2019, to appear)—to the lexicostatistical data of each of the five northeastern Ainu dialects and attempts to capture the dialect relationships obscured by the assumption of the nominal scale by Hattori and Chiri (1960).

The primary results of this paper may be summarized as follows. First, our analyses demonstrate that Hattori and Chiri's (1960) lexicostatistical data contain significant information on an ordinal scale that classifies Asahikawa and Nayoro dialects as neither northeastern nor southwestern Hokkaido Ainu dialects; rather, they constitute an intermediate dialect group. Moreover, these two dialects display relationships with Sakhalin dialects. The results are consistent with the geography of the two dialects, which are located near the Taisetsu Mountain Range in central Hokkaido and situated between Hokkaido and Sakhalin.

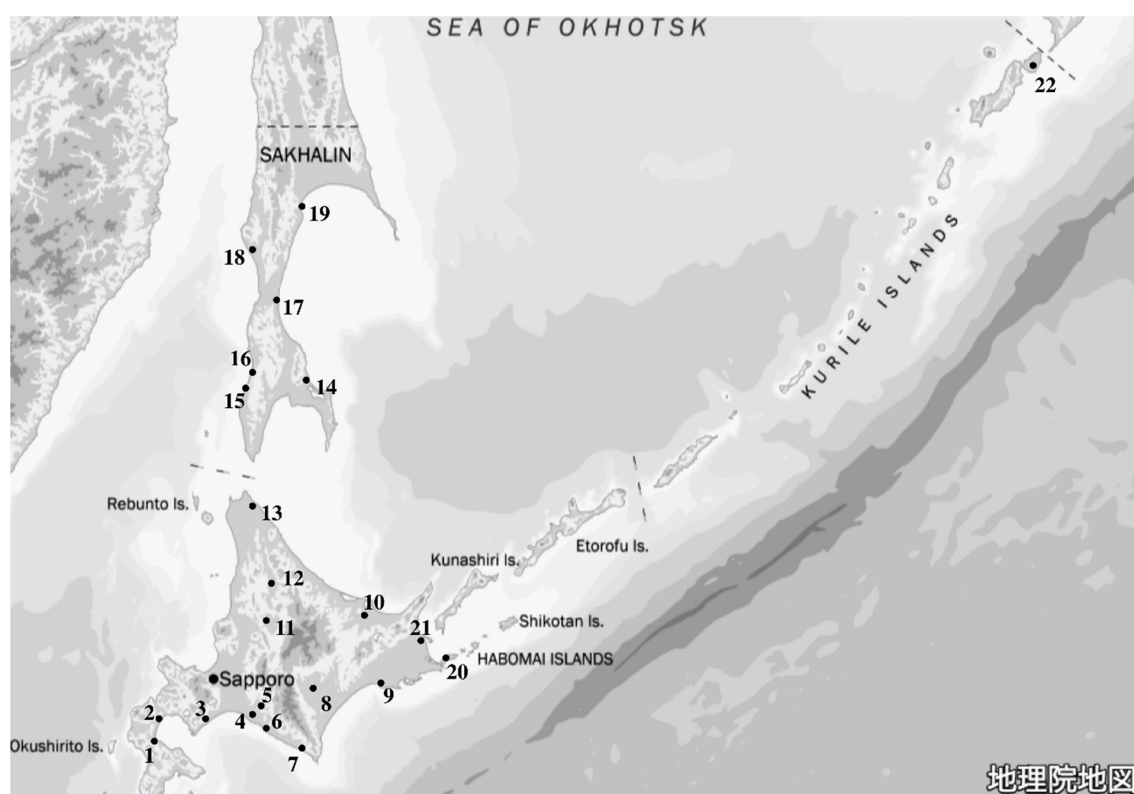


Figure 1. Map of a section of the region where the Ainu language is or was spoken (Geospatial Information Authority of Japan, 2018), edited by the author. 1: Yakumo, 2: Oshamambe, 3: Horobetsu, 4: Biratori, 5: Nukibetsu, 6: Niikappu, 7: Samani, 8: Obihiro, 9: Kushiro, 10: Bihoro, 11: Asahikawa, 12: Nayoro, 13: Soya, 14: Ochiho, 15: Tarantomari, 16: Maoka, 17: Shiraura, 18: Raichishka, 19: Nairo, 20: Nemuro, 21: Nobetsu, 22: Shumushu. Note: Hattori and Chiri (1960) present the investigation results of dialects 1 to 19 in this map.

Second, the statistical analyses illustrate one component corresponding to the distinction between the northern (i.e., Soya dialect) and southern parts of Hokkaido

(i.e., Samani dialect) from the viewpoints of northeastern Hokkaido Ainu dialects that is first illustrated with statistical methods.

Third, the author observes homogeneous structures among the Obihiro, Bihoro, and Kushiro dialects in all the statistical analyses of the four dialects (i.e., Asahikawa, Nayoro, Obihiro, and Bihoro dialects). However, the statistical analysis of the Kushiro dialect demonstrates a heterogeneous structure distinct from the Obihiro, Bihoro, and Kushiro dialects: the strong originality of the Kushiro dialect.

Fourth, further statistical investigation indicates that an Ainu dialect located to the east of Kushiro (e.g., Nos. 20, 21, and 22 in Figure 1) could provide significant information that previous statistical analyses assuming a nominal scale obscured. These statistical findings on Ainu dialects to the east of Kushiro coincide with the author’s joint research with Dr. Mika Fukazawa (Fukazawa and Ono 2015) on old documentation of the Ainu language and her geolinguistic and philological doctoral thesis on the *Kagake Monjo* (Fukazawa 2017) concerning Nemuro dialect, an Ainu dialect located to the east of Kushiro.

Therefore, given this marked improvement in statistical methodologies (i.e., statistical analysis with the ordinal assumption), this paper not only reconfirms current linguistic knowledge in Ainu dialectology but also partially reveals the underlying (or unknown) relationship among Ainu dialects that philological research (Fukazawa 2017) derived independently from the statistical approach.

Finally, the main results in this paper indicate that the five “northeastern” Ainu dialects (i.e., Nos. 8–12 in Figure 1) involve the underlying information structures with diversity that have been obscured by the assumption of the nominal scale in the previous statistical analyses.

The remainder of this paper is organized as follows. Section 2 discusses how the five symbols (i.e., “±”, “○”, “?”, “·”, and “()”) in Hattori and Chiri (1960) are the motivation for the statistical analysis with an ordinal scale used in this paper, as discussed in Ono (2019, to appear).

Section 3 demonstrates that statistical analyses with an ordinal scale more effectively delineate the relationships among 19 Ainu dialects from the perspective of Asahikawa and Nayoro dialects and the heterogeneous structure of Obihiro, Bihoro, and Kushiro dialects from the perspective of Kushiro dialect. The distinction between the northern and southern parts of Hokkaido is also illustrated.

Section 4 discusses this paper’s significance and focuses on the background of the heterogeneous structures of Obihiro, Bihoro, and Kushiro dialects from the perspective of Kushiro dialect.

In conclusion, this paper notes the future of complementary studies on Ainu linguistics and statistics.

2. Materials and Methods

Hattori and Chiri (1960) introduced seven symbols (“+”, “−”, “±”, “○”, “?”, “•”, and “()”) to record the lexicostatistical data in the linguistic environments of Ainu in the 1960’s. These symbols are explained as follows (Hattori and Chiri 1960: 307): “+” as *cognate residues*, “−” as *non-cognates*, “±” as *cognates and non-cognates*, “○” as *questionable etymology or choice*, “?” as *doubtful record*, “•” as *no answer given*, and “()” as *lacuna of the record*.² However, Ono (2019, to appear) confirmed that none of the four symbols “○”, “?”, “•”, and “()” represent “mere lack of information”; rather, they contain information among the word forms based on Hattori and Chiri’s (1960) substantive linguistic knowledge.

First, Table 1 demonstrates the original records of “stand” in 19 Ainu dialects and the cognacy judgments based on Yakumo dialect record (i.e., ’ás) according to their own linguistic knowledge (Hattori and Chiri 1960: 320). Although Hattori and Chiri (1960: 307) stated “○” as *questionable etymology or choice*, they also explained that “-as in /’etaras/ could be the same morpheme with /’ás/ in the dialects (e.g., Biratori) but its cognacy is not sufficient enough. Therefore, we determined the cognacy judgments as ‘○’”(1960: 336).

Second, Table 2 demonstrates the original records of “mother” in 19 Ainu dialects and the cognacy judgments based on Yakumo dialect record (i.e., hápo) according to their own linguistic knowledge (Hattori and Chiri 1960: 328). Although Hattori and Chiri (1960: 307) stated “?” as *doubtful record*, they also explained that “The original records in ‘mother’ can be translated as ‘Okaasan’ (familiar name of mother in Japanese). But the informant in Ochiho dialect reported ‘mother’ as ‘unu,’ whose word form we found in other dialects as ‘Hahaoya’ (formal name of mother in Japanese). Therefore, we determined the cognacy judgments among Ochiho dialect and the other dialects as neither ‘−’ nor ‘+’; rather as ‘?’”(1960: 337).

Third, Table 3 demonstrates the original records of “straight” in 19 Ainu dialects and the cognacy judgments based on Yakumo dialect records (i.e., ’íkne, ’o’upéka) according to their own linguistic knowledge (Hattori and Chiri 1960: 331). Although Hattori and Chiri (1960: 307) stated “•” as *no answer given*, they also explained that “The informant of Soya dialect answered ‘straight’ as /réwke somóki/ (i.e., not turn) and, he said, as neither /’o’upeka/-type nor /ku’anno/-type. If the investigator asked /’ittusne/-type, the informant of Soya dialect could remember. Therefore, we determined the cognacy judgments as ‘•’”(1960: 337).

² In the following sentences, unless italicized, the English translation of Japanese literature is by the author.

Table 1. An example of “○” in Hattori and Chiri (1960). The records are “stand” in Hattori and Chiri (1960: 320) and the cognacy judgments data are based on Yakumo dialect record (i.e., ’ás) according to their own linguistic knowledge.

“stand” in Hattori and Chiri (1960)		Cognacy Judgments based on Yakumo record in Hattori and Chiri (1960)	
Dialect	Word Form	Dialect	Cognacy Judgments
X1_Yakumo	’ás	X1_Yakumo	+
X2_Oshamambe	’ás	X2_Oshamambe	+
X3_Horobetsu	’ás	X3_Horobetsu	+
X4_Biratori	’ás	X4_Biratori	+
X5_Nukibetsu	’ás	X5_Nukibetsu	+
X6_Niikappu	’ás	X6_Niikappu	+
X7_Samani	’ás	X7_Samani	+
X8_Obihiro	’ás	X8_Obihiro	+
X9_Kushiro	’ás	X9_Kushiro	+
X10_Bihoro	’ás	X10_Bihoro	+
X11_Asahikawa	’ás	X11_Asahikawa	+
X12_Nayoro	’ás	X12_Nayoro	+
X13_Soya	’etáras	X13_Soya	○
X14_Ochiho	’etáras	X14_Ochiho	○
X15_Tarantomari	’etáras	X15_Tarantomari	○
X16_Maoka	’etáras	X16_Maoka	○
X17_Shirauro	’etáras	X17_Shirauro	○
X18_Raichishka	’etáras	X18_Raichishka	○
X19_Nairo	’etáras	X19_Nairo	○

Table 2. An example of “?” in Hattori and Chiri (1960). The records are “mother” in Hattori and Chiri (1960: 328) and the cognacy judgments data are based on Yakumo dialect record (i.e., hápo) according to their own linguistic knowledge.

“mother” in Hattori and Chiri (1960)		Cognacy Judgments based on Yakumo record in Hattori and Chiri (1960)	
Dialect	Word Form	Dialect	Cognacy Judgments
X1_Yakumo	hápo	X1_Yakumo	+
X2_Oshamambe	hápo	X2_Oshamambe	+
X3_Horobetsu	hápo	X3_Horobetsu	+
X4_Biratori	hápo	X4_Biratori	+
X5_Nukibetsu	hápo	X5_Nukibetsu	+
X6_Niikappu	hápo	X6_Niikappu	+
X7_Samani	hápo	X7_Samani	+
X8_Obihiro	hápo	X8_Obihiro	+
X9_Kushiro	hápo	X9_Kushiro	+
X10_Bihoro	hápo	X10_Bihoro	+
X11_Asahikawa	tóto	X11_Asahikawa	—
X12_Nayoro	tóto	X12_Nayoro	—
X13_Soya	hápo	X13_Soya	+
X14_Ochiho	unu	X14_Ochiho	?
X15_Tarantomari	nanna	X15_Tarantomari	—
X16_Maoka	nanna	X16_Maoka	—
X17_Shirauro	nanna	X17_Shirauro	—
X18_Raichishka	’onmo	X18_Raichishka	—
X19_Nairo	nanna	X19_Nairo	—

Table 3. An example of “・” in Hattori and Chiri (1960). The records are “straight” in Hattori and Chiri (1960: 331) and the cognacy judgments data are based on Yakumo dialect records (i.e., ’íkne, ’o’upéka) according to their own linguistic knowledge.

“Straight” in Hattori and Chiri (1960)		Cognacy Judgments based on Yakumo record in Hattori and Chiri (1960)	
Dialect	Word Form	Dialect	Cognacy Judgments
X1_Yakumo	’íkne, ’o’upéka	X1_Yakumo	+
X2_Oshamambe	’o’upéka	X2_Oshamambe	±
X3_Horobetsu	’o’úpeka	X3_Horobetsu	±
X4_Biratori	’owpeka	X4_Biratori	±
X5_Nukibetsu	’o’ópeka	X5_Nukibetsu	±
X6_Niikappu	’owpeka	X6_Niikappu	±
X7_Samani	’owpeka	X7_Samani	±
X8_Obihiro	’o’úpeka	X8_Obihiro	±
X9_Kushiro	’owpeka	X9_Kushiro	±
X10_Bihoro	’o’upeka	X10_Bihoro	±
X11_Asahikawa	’ittusne	X11_Asahikawa	±
X12_Nayoro	’o’úpeka	X12_Nayoro	±
X13_Soya	(réwke somóki)	X13_Soya	・
X14_Ochiho	ku’aNno, ’ukuruhne	X14_Ochiho	—
X15_Tarantomari	ku’anno	X15_Tarantomari	—
X16_Maoka	ku’anno, ’istusne	X16_Maoka	±
X17_Shiraaura	’ikuruhne	X17_Shiraaura	—
X18_Raichishka	e’iku’anno	X18_Raichishka	—
X19_Nairo	’o’ihtusno	X19_Nairo	±

Table 4. An example of “()” in Hattori and Chiri (1960). The records are “here” in Hattori and Chiri (1960: 324) and the cognacy judgments data are based on Yakumo dialect record (i.e., ’íkne, ’o’upéka) according to their own linguistic knowledge.

“here” in Hattori and Chiri (1960)		Cognacy Judgments based on Yakumo record in Hattori and Chiri (1960)	
Dialect	Word Form	Dialect	Cognacy Judgments
X1_Yakumo	téta	X1_Yakumo	+
X2_Oshamambe	téta	X2_Oshamambe	+
X3_Horobetsu	téta	X3_Horobetsu	+
X4_Biratori	téta	X4_Biratori	+
X5_Nukibetsu	téta	X5_Nukibetsu	+
X6_Niikappu	té’or	X6_Niikappu	()
X7_Samani	ta’anta	X7_Samani	—
X8_Obihiro	ta’anta	X8_Obihiro	—
X9_Kushiro	tanta	X9_Kushiro	—
X10_Bihoro	temanta	X10_Bihoro	○
X11_Asahikawa	téta	X11_Asahikawa	+
X12_Nayoro	téta, tánta	X12_Nayoro	±
X13_Soya	téta	X13_Soya	+
X14_Ochiho	teeta	X14_Ochiho	+
X15_Tarantomari	teeta	X15_Tarantomari	+
X16_Maoka	teeta	X16_Maoka	+
X17_Shiraaura	teeta	X17_Shiraaura	+
X18_Raichishka	teeta	X18_Raichishka	+
X19_Nairo	teyta	X19_Nairo	+

Finally, Table 4 demonstrates the original records of “here” in 19 Ainu dialects and the cognacy judgments based on Yakumo dialect record (i.e., téta) according to their own linguistic knowledge (Hattori and Chiri 1960: 324). Although Hattori and Chiri (1960: 307) stated “()” as *lacuna of the record*, they also explained that “All word forms except Niikappu dialect corresponded to ‘koko-ni’ in Japanese (i.e., koko

[here] + ni [particle]) but ‘té’or’ in Niikappu dialect corresponded to only ‘koko’ in Japanese. Although ‘te’ in ‘té’or’ and ‘te’ in ‘téta’ must be the same morpheme, Bihoro dialect distinguished ‘té’or’ as ‘koko’ in Japanese and ‘ta’ánta’ as ‘koko-ni’ in Japanese. Therefore, there is no guarantee that ‘té’or’ in Niikappu dialect means ‘koko-ni’ in Japanese. We determine the cognacy judgments as ‘()’ (1960: 336). Note that “koko” corresponds to the demonstrative noun in Japanese and “koko-ni” to the adverb.

We observed the four symbols represent some significant information on the linguistic environments around Ainu language in 1960’s rather than “mere lack of information.” Thus Ono (2019, to appear) summarized these four symbols as “ Δ ” (i.e., “not non-cognates” or “researcher does not necessarily decide the cognacy judgments between two word forms as ‘-’ [non-cognates]”) and proposed an ordinal ranking of the symbols as $+ > \pm > \Delta > -$ or $+ < \pm < \Delta < -$ in the statistical analysis.

In this Section, the author supplements the introduction of ordinal restrictions on Hattori and Chiri’s (1960) lexicostatistical data in two respects. First, it is natural that we understand these four symbols (i.e., “+”, “ $\pm$ ”, “ Δ ”, and “-”) as indicating some belief regarding Hattori and Chiri’s cognacy judgments. Second, unless the author imposes an ordinal restriction (i.e., $+ > \pm > \Delta > -$ or $+ < \pm < \Delta < -$) on these four symbols, they can be quantified as having certain values violating the order relation (e.g., “+” as 1.0, “ $\pm$ ” as 0.2, “ Δ ” as 0.8, and “-” as 0.6). Essentially, the values do not reflect the belief regarding Hattori and Chiri’s cognacy judgments on these symbols and, consequently, the symbols are calculated as a nominal scale.³

³ One of reviewers indicated a possibility that the linguistic environments, including the physical condition of an informant and the relationship between an informant and a researcher, could affect the serious bias on the data about the word forms in Hattori and Chiri (1960), and suggested that researchers should not summarize the four symbols (i.e., “○”, “?”, “·”, and “()”) as “ Δ ” in the ordinal restrictions (i.e., $+ > \pm > \Delta > -$ or $+ < \pm < \Delta < -$) but as “unpredictable answers on the symbols (i.e., “+”, “ $\pm$ ”, and “-”) in further investigation with the desirable conditions.” Moreover, the reviewer posed a question whether the ordinal restrictions on the symbols in this paper (i.e., $+ > \pm > \Delta > -$ or $+ < \pm < \Delta < -$) positioned the dialects (e.g., Asahikawa dialect and Soya dialect), whose data contain more “ Δ ” than other dialects, as an intermediate dialect as the results of the ordinal restrictions themselves. The author does not exclude the possibility in Hattori and Chiri’s (1960) data. However, from the statistical point of view on the former problem, the linguistic environments that the reviewer indicated expect to produce the four symbols *irrelevant* to the morpheme in word forms. Therefore, researchers need to analyze the patterns of the four symbols (i.e., “○”, “?”, “·”, and “()”) as to whether these symbols appear in the data of the dialect *randomly or not* to determine the effect of the linguistic environments to the data. Since, in case of the Asahikawa dialect, the symbol, “?” appears on “ta’ánpe”, “ta’ánkur”, “ta’ánutar” and “ta’ánta” whose word form corresponds to “that”, “he”, “they” and “there” respectively in Hattori and Chiri (1960: 314; 323; 324). As Hattori and Chiri (1960: 312) explained that “the word form in Asahikawa dialect (i.e., “ta’ánpe”) corresponds to /ta’ánpe/-type in other dialects that means ‘sore’ in Japanese: these two word forms might have the same meaning. Therefore, we determine the cognacy judgments among the Asahikawa dialect and the other dialects in Hokkaido as “?” (1960: 312). Furthermore, Hattori and Chiri (1960: 336) explained that “the word forms in the Asahikawa

Since Homogeneity Analysis (Gifi 1990) enables us to quantify these symbols with the assumption of both an ordinal and a nominal scale as implemented in R (2018) language through the “homals” package (de Leeuw and Mair 2009), Ono (2019, to appear) applies it to validate the assumption of an ordinal scale in Hattori and Chiri (1960). In the following analysis, the author also employs the “homals” package.

3. Results

Figures 2 and 3 present the results of the Neighbor-Net Analysis (Huson and Bryant 2006) calculated from the two Euclidean distance matrices that Homogeneity Analysis yielded from Hattori and Chiri’s (1960) lexicostatistical data on Asahikawa dialect with the assumption of an ordinal scale and a nominal scale, respectively.

Figure 2, which assumes an ordinal scale for Hattori and Chiri (1960), indicates less ambiguous structures (i.e., several shorter sides in Neighbor-Net correspond to the ambiguity for fit in Neighbor-Net Analysis) than Figure 3, which assumes a nominal scale.⁴ This result suggests that the statistical analysis with the assumption of an ordinal scale in Hattori and Chiri (1960) successfully captures the underlying information structure of the Asahikawa dialect. As shown in Figure 2, the Asahikawa dialect is more distant from the Net (i.e., line O) and has a relationship with Sakhalin dialects (i.e., line D) but is not classified as the northeastern or southwestern Hokkaido Ainu dialect groups.

dialect (i.e., “ta’áńkur” and “ta’áńutar”) might mean “sono-hito” and “sono-hito-tachi” in Japanese respectively. Therefore, we determine the cognacy judgments among the Asahikawa dialect and the other dialects in Hokkaido as “?” (1960: 336). The same analysis can apply to the case of “there” in the Asahikawa dialect (i.e., “ta’áńta”). We observed that the morpheme, “ta” appears in the four cases (i.e., “ta’áńpe”, “ta’áńkur”, “ta’áńutar” and “ta’áńta”) and causes the cognacy judgment, “?” *systematically*. Therefore, the author does not determine the symbol (i.e., “?”) as “unpredictable answers on the symbols (i.e., “+”, “±”, and “–”) in further investigation with the desirable conditions” which the reviewer suggested, in case of the Asahikawa dialect that the author will investigate in this paper. However, researchers need to pay attention to the problem that the reviewer suggested in case of the Soya dialect. The author will deal with this significant issue in another article. Furthermore, on the latter problem that the reviewer suggested, the effect of the different assumption on the scale type (i.e., nominal scale and ordinal scale in this paper) to the dialect relationships on the Ainu language is of concern from the linguistic point of view. Due to space restrictions, the author will also deal with these significant topics in another article.

⁴ In the following results of Neighbor-Net Analysis, a longer side from the Net with the bipartite by line O refers to the originality of the dialect; the bipartite of Neighbor-Net by line A: the relationships between the southwestern Ainu dialect group and the given dialect; the bipartite of Neighbor-Net by line B: the relationships between the northeastern Ainu dialect group and the given dialect; the bipartite of Neighbor-Net by line C: the relationships among Samani, Obihiro, Kushiro, and Bihoro dialect from the perspective of the four dialects; and the bipartite of Neighbor-Net by line D: the relationships between the Sakhalin Ainu dialect group and the given dialect, respectively.

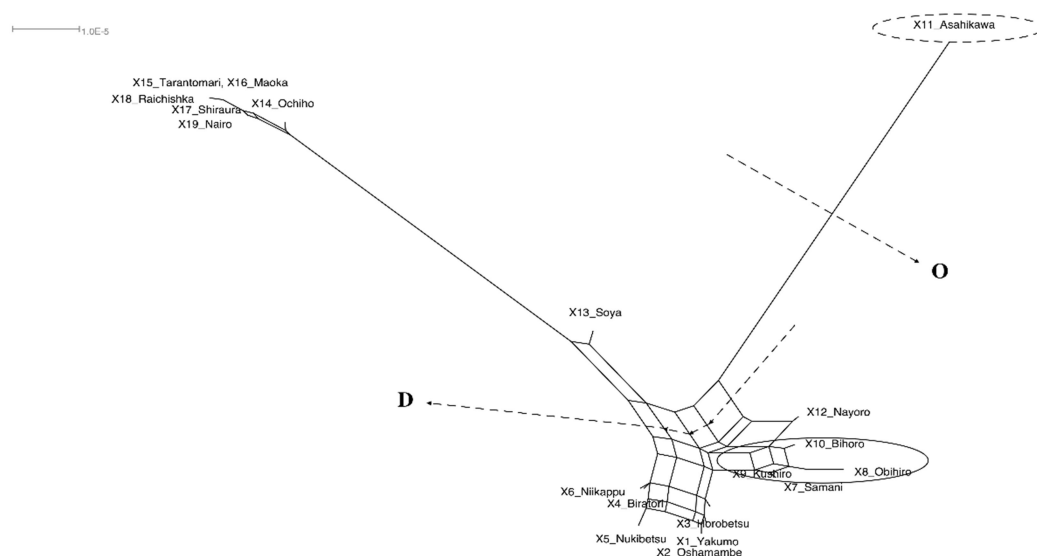


Figure 2. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Asahikawa dialect yielded by Homogeneity Analysis assuming an ordinal scale in Hattori and Chiri (1960).

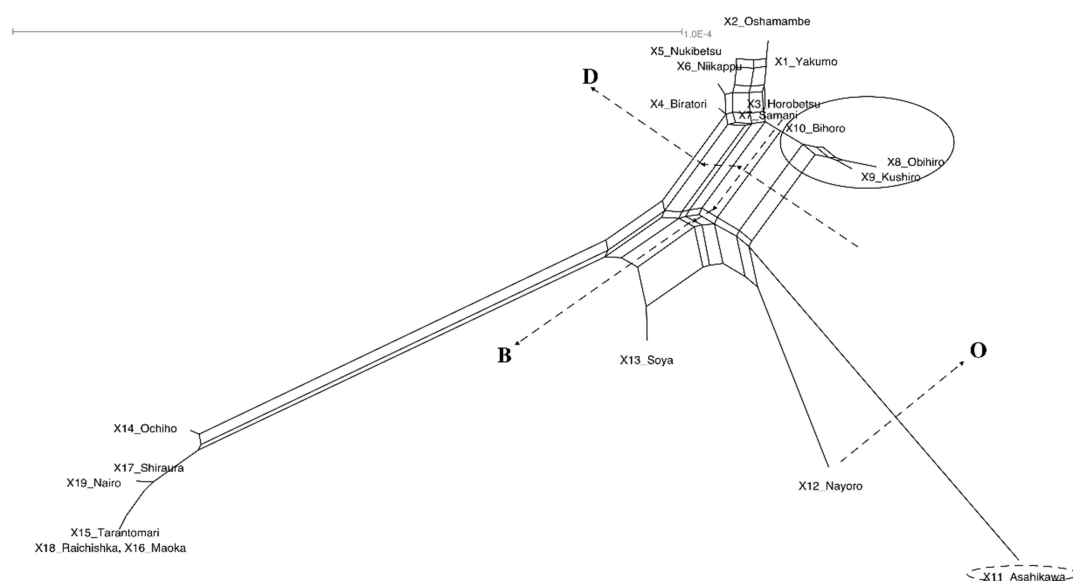


Figure 3. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Asahikawa dialect yielded by Homogeneity Analysis assuming a nominal scale in Hattori and Chiri (1960).

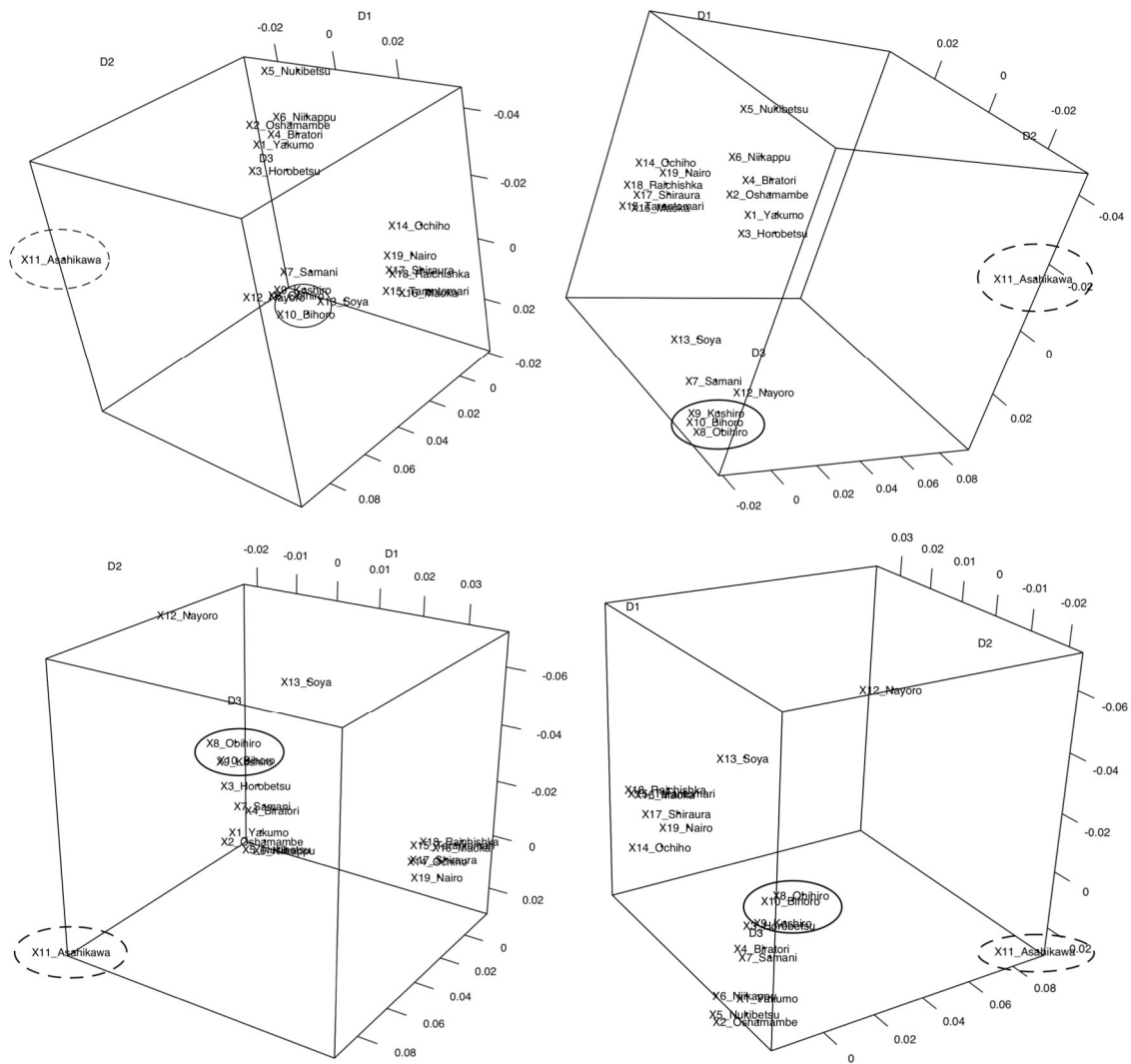


Figure 4. The result of three-dimensional plots for the Euclidean coordinate system for Asahikawa dialect yielded by Homogeneity Analysis in Hattori and Chiri (1960).
(Top) on ordinal scale (Bottom) on nominal scale

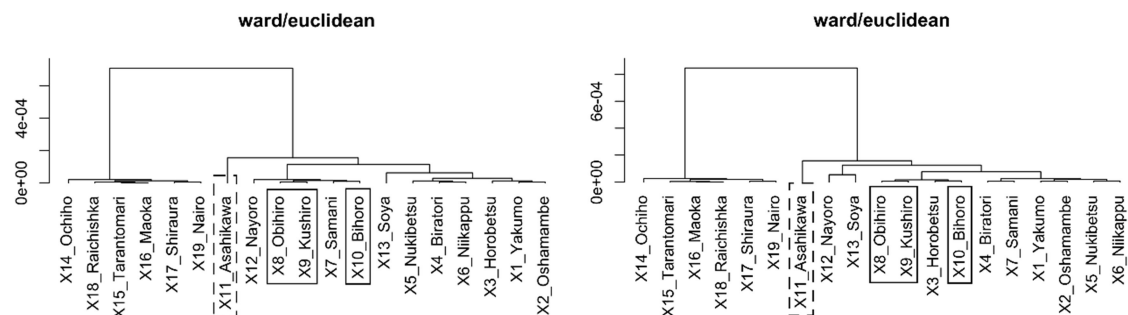


Figure 5. The results of cluster analysis of Euclidean distance for Asahikawa dialect.
(Left) on ordinal scale (Right) on nominal scale

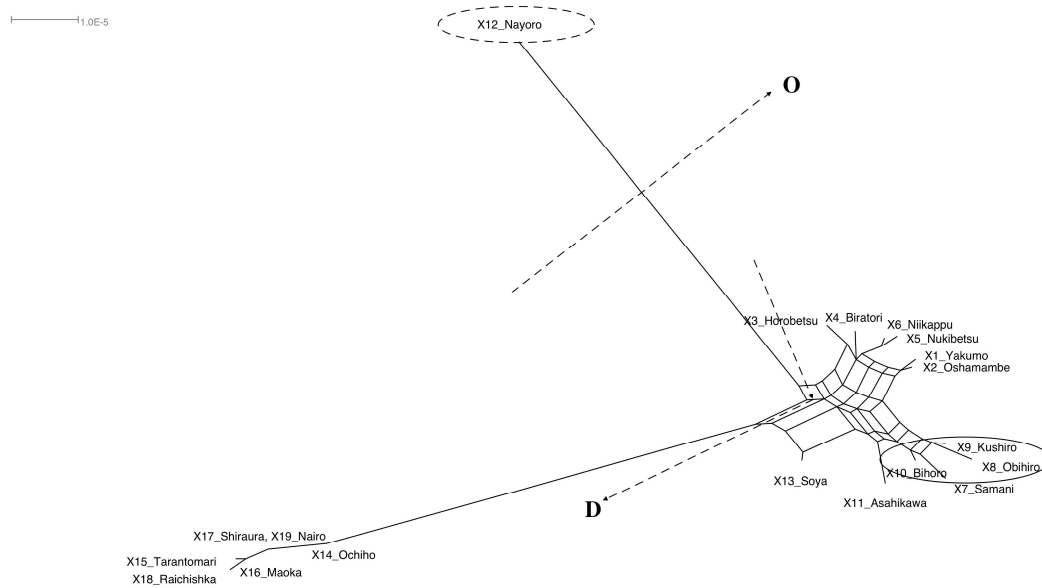


Figure 6. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Nayoro dialect yielded by Homogeneity Analysis assuming an ordinal scale in Hattori and Chiri (1960).

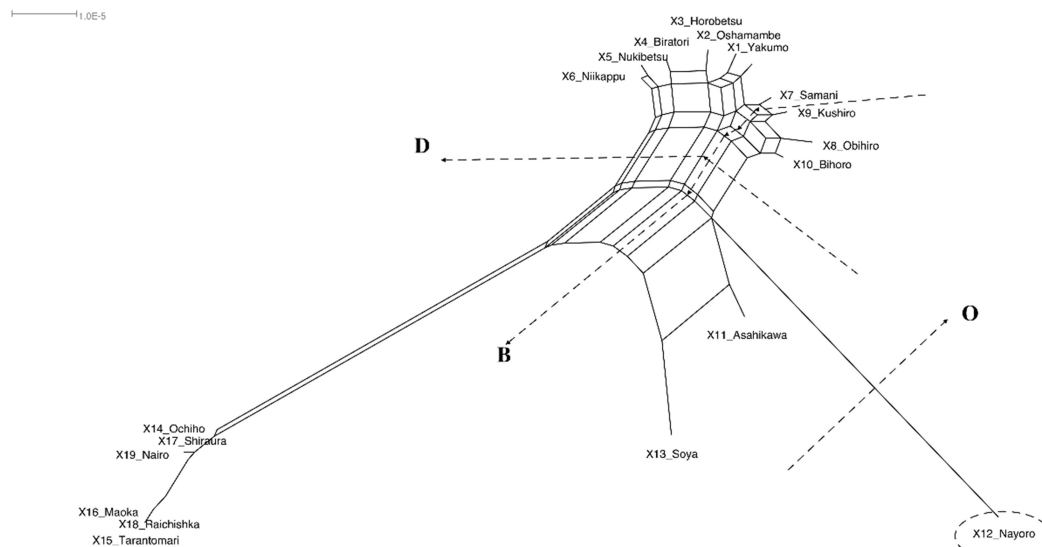


Figure 7. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Nayoro dialect yielded by Homogeneity Analysis assuming a nominal scale in Hattori and Chiri (1960).

This suggests reconsideration of a classification of Asahikawa dialect based on previous statistical research showing that it belongs to the northeastern Ainu group. Instead, the statistical analysis indicates that the Asahikawa dialect is an intermediate dialect between northeastern and southwestern Hokkaido Ainu dialect groups. Moreover, this suggestion is consistent with the geographical location of Asahikawa dialect in Figure 1.

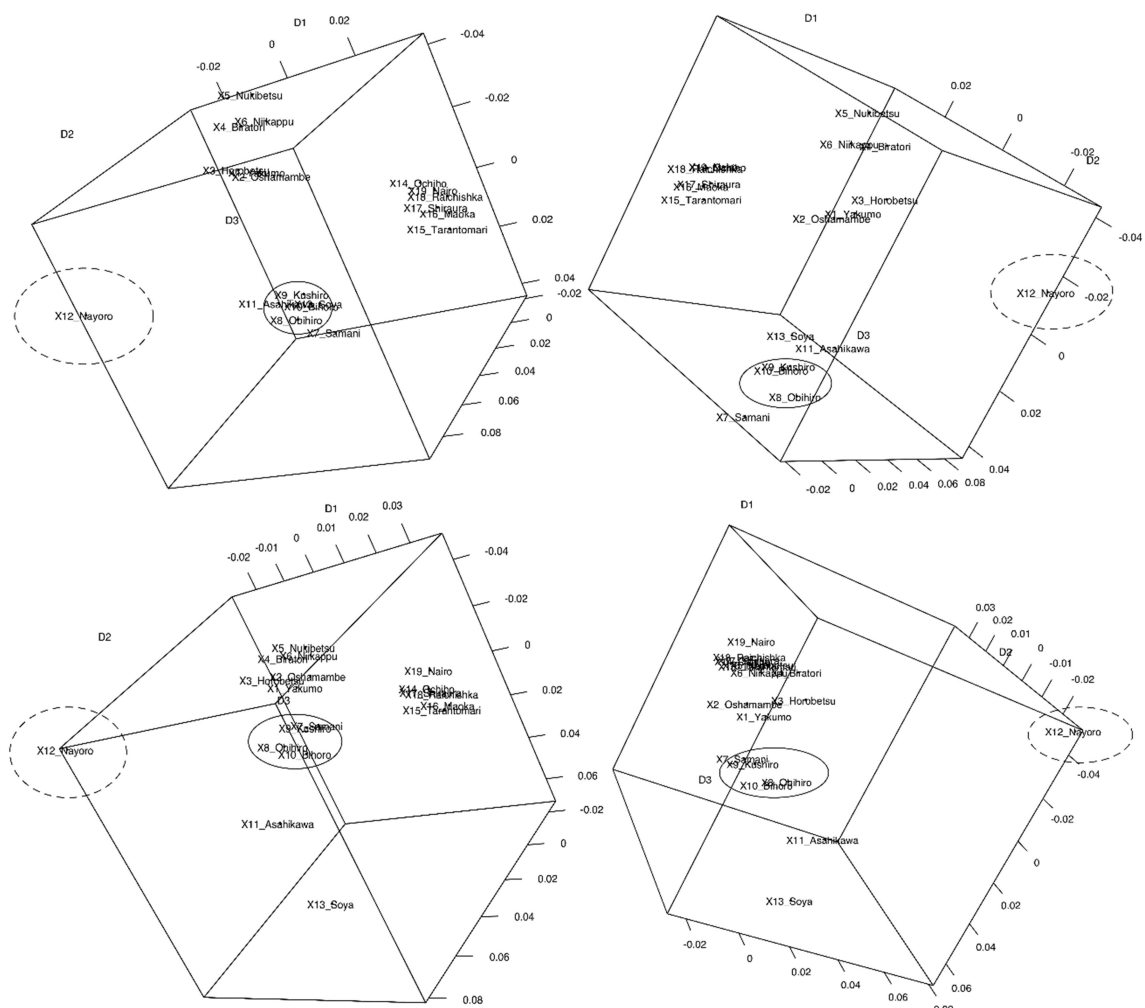


Figure 8. The result of three-dimensional plots for the Euclidean coordinate system for Nayoro dialect yielded by Homogeneity Analysis in Hattori and Chiri (1960).
(Top) on ordinal scale (Bottom) on nominal scale

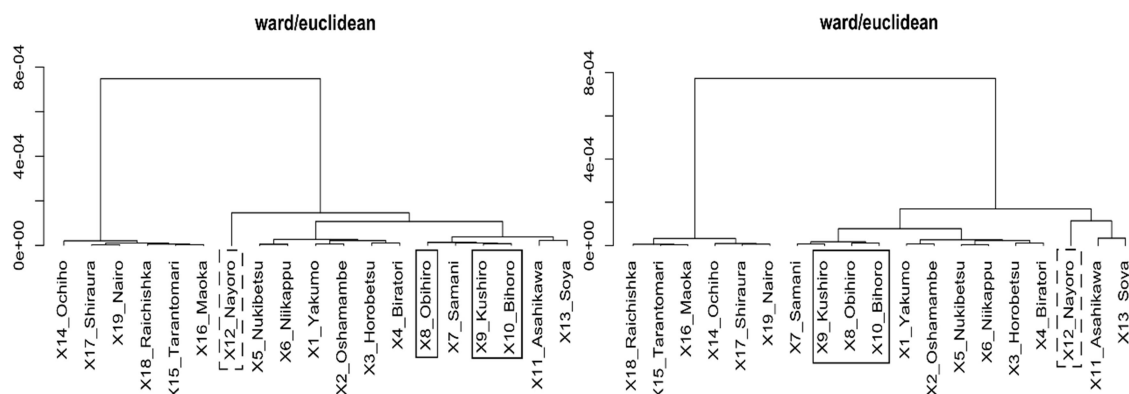


Figure 9. The results of cluster analyses of Euclidean distances for Nayoro dialect.
(Left) on ordinal scale (Right) on nominal scale

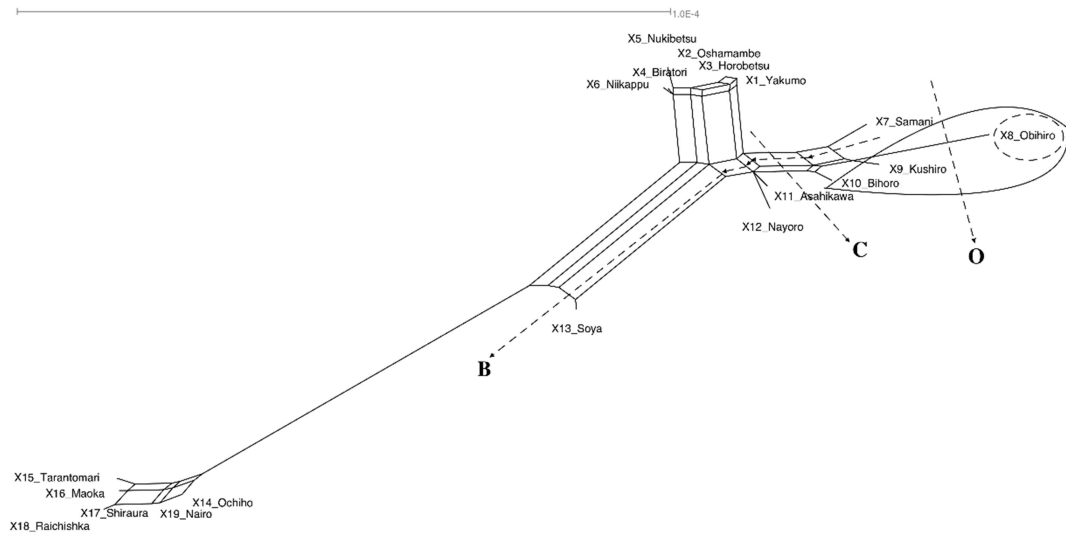


Figure 10. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Obihiro dialect yielded by Homogeneity Analysis assuming an ordinal scale in Hattori and Chiri (1960).

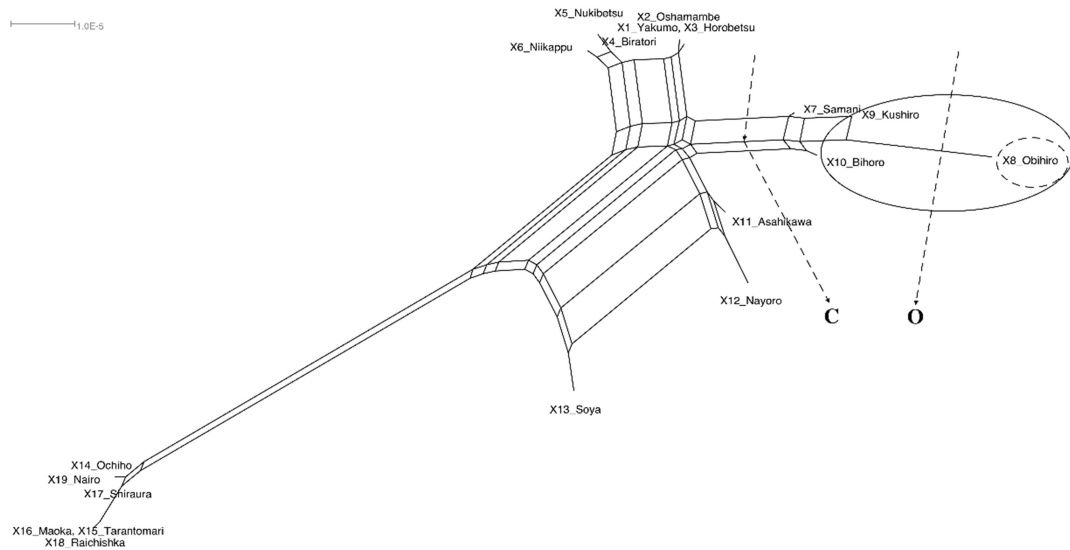


Figure 11. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Obihiro dialect yielded by Homogeneity Analysis assuming a nominal scale in Hattori and Chiri (1960).

The other analyses support this finding. Asahikawa dialect is located relatively farther from the other dialect groups in the three-dimensional plot at the top of Figure 4 than those at the bottom.

The result of cluster analysis on the left in Figure 5 (i.e., on the ordinal scale) demonstrates that Asahikawa dialect is an intermediate dialect between the northeastern and southwestern Hokkaido Ainu dialect groups in contrast to the right in Figure 5 (i.e., on the nominal scale).

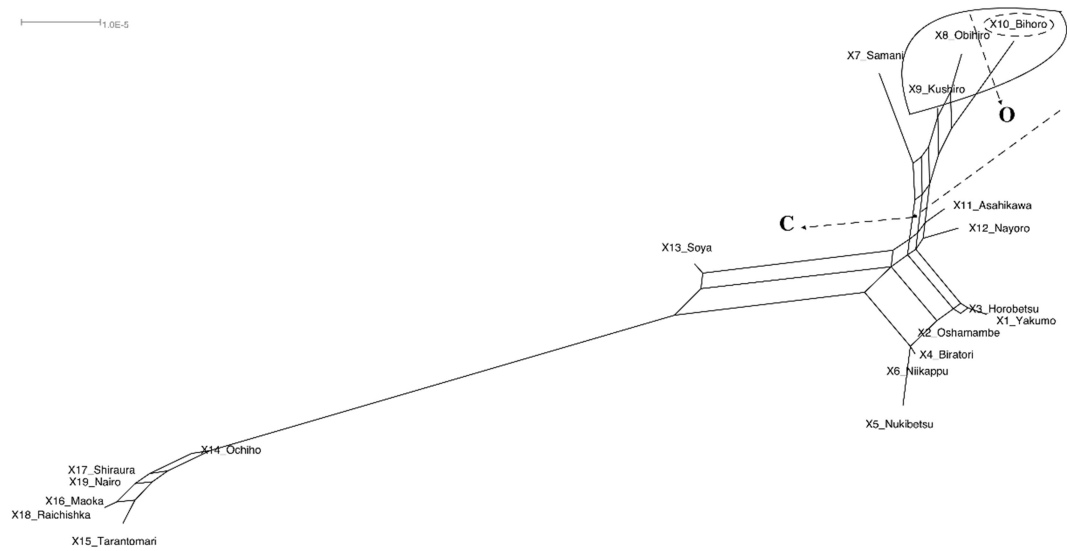


Figure 14. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Bihoro dialect yielded by Homogeneity Analysis assuming an ordinal scale in Hattori and Chiri (1960).

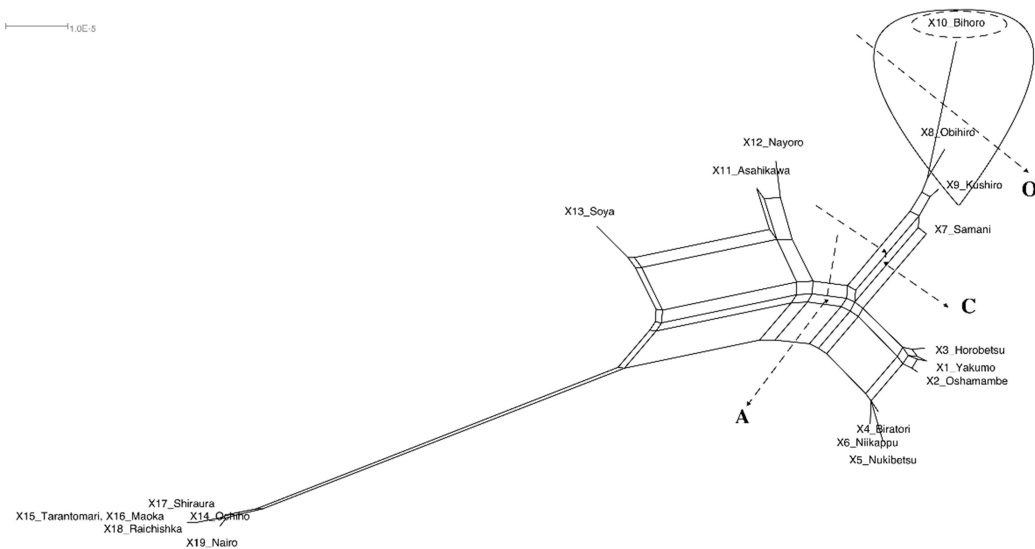


Figure 15. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Bihoro dialect yielded by Homogeneity Analysis assuming a nominal scale in Hattori and Chiri (1960).

The same analysis can be applied to the case of Nayoro dialect. Figure 6, which assumes an ordinal scale in Hattori and Chiri (1960), displays less ambiguous structures than Figure 7, which assumes a nominal scale. This result also suggests that the statistical analysis with the assumption of an ordinal scale in Hattori and Chiri (1960) succeeded in capturing the underlying information structure of Nayoro dialect.

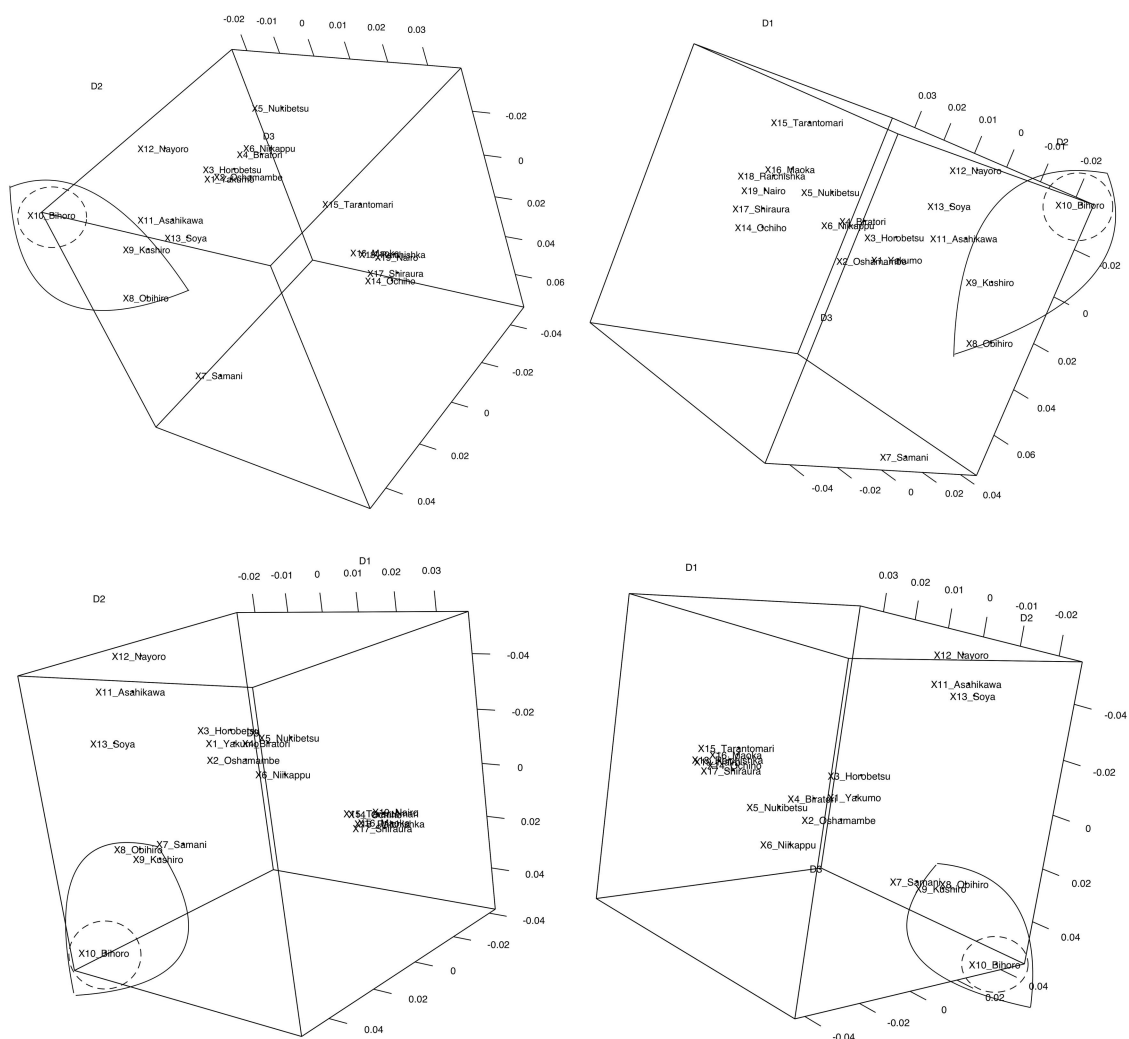


Figure 16. The result of three-dimensional plots for the Euclidean coordinate system for Bihoro dialect yielded by Homogeneity Analysis in Hattori and Chiri (1960). (Top) on ordinal scale (Bottom) on nominal scale

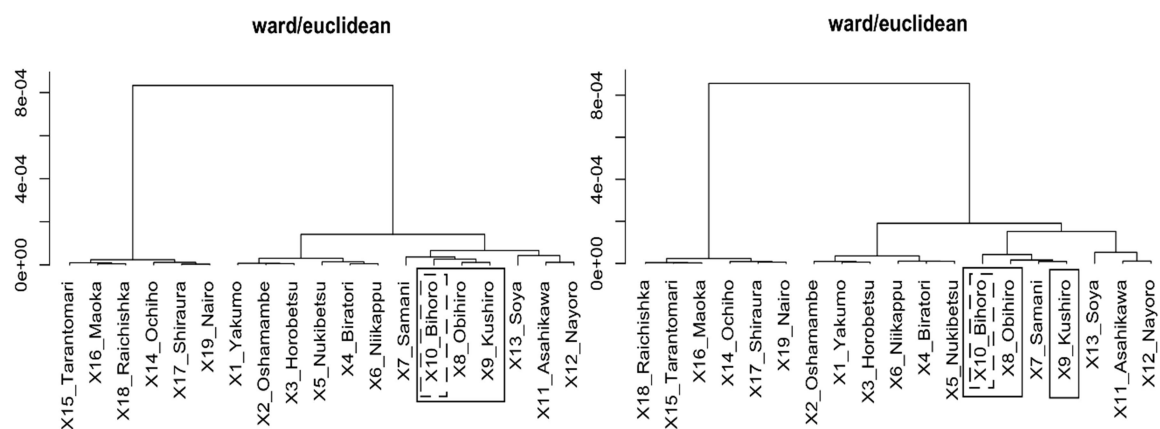


Figure 17. The results of cluster analyses of Euclidean distances for Bihoro dialect. (Left) on ordinal scale (Right) on nominal scale

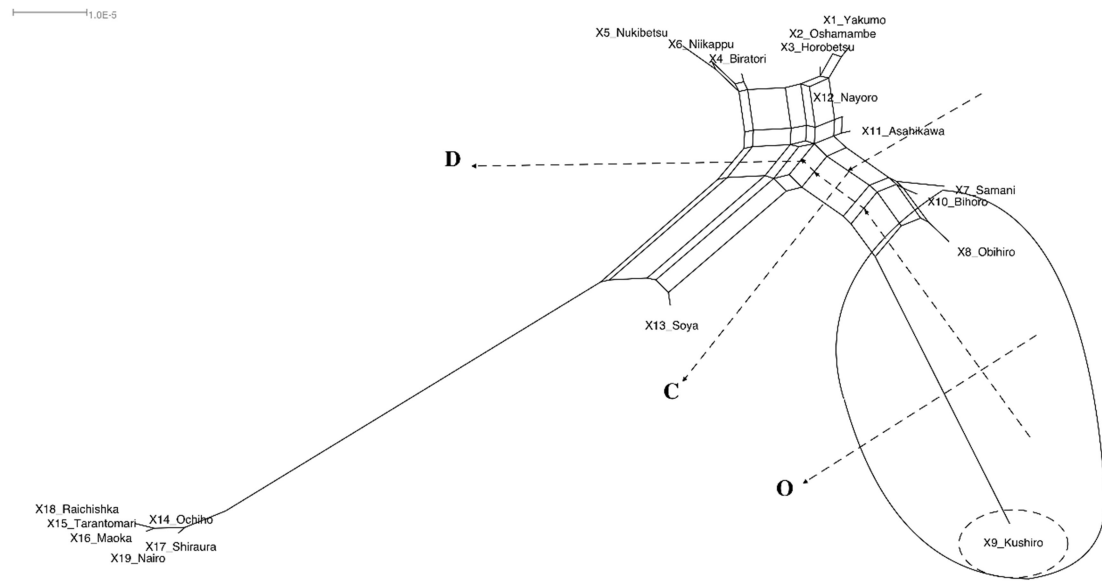


Figure 18. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Kushiro dialect yielded by Homogeneity Analysis assuming an ordinal scale in Hattori and Chiri (1960).

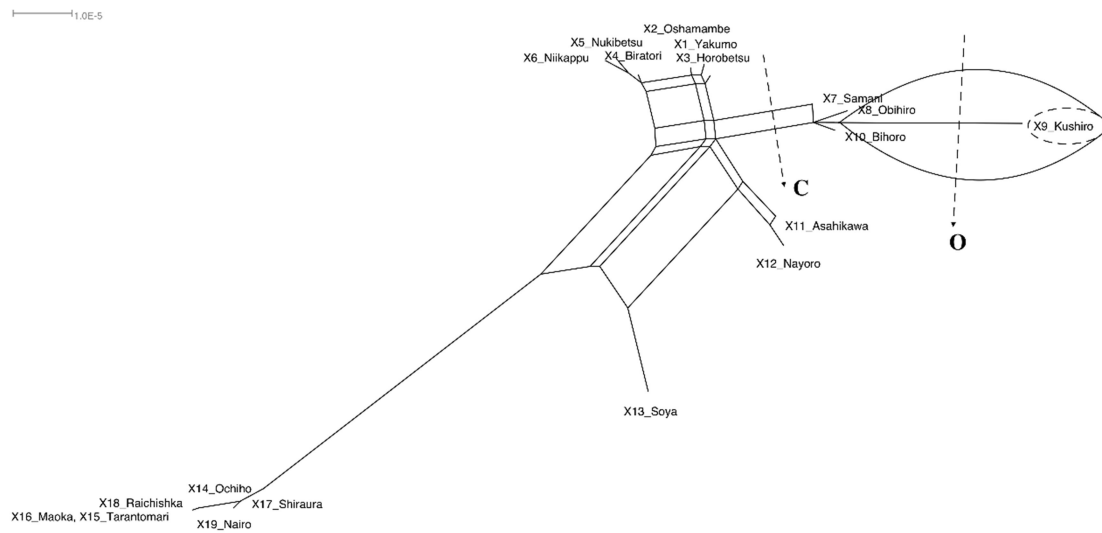


Figure 19. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Kushiro dialect yielded by Homogeneity Analysis assuming a nominal scale in Hattori and Chiri (1960).

Since Nayoro dialect in Figure 6 is more distant from the Net (i.e., line O) and bears a relationship with Sakhalin dialects (i.e., line D) but not with the northeastern or southwestern Hokkaido Ainu dialect groups, it suggests that the classification of Nayoro dialect should be reconsidered, similar to Asahikawa dialect.

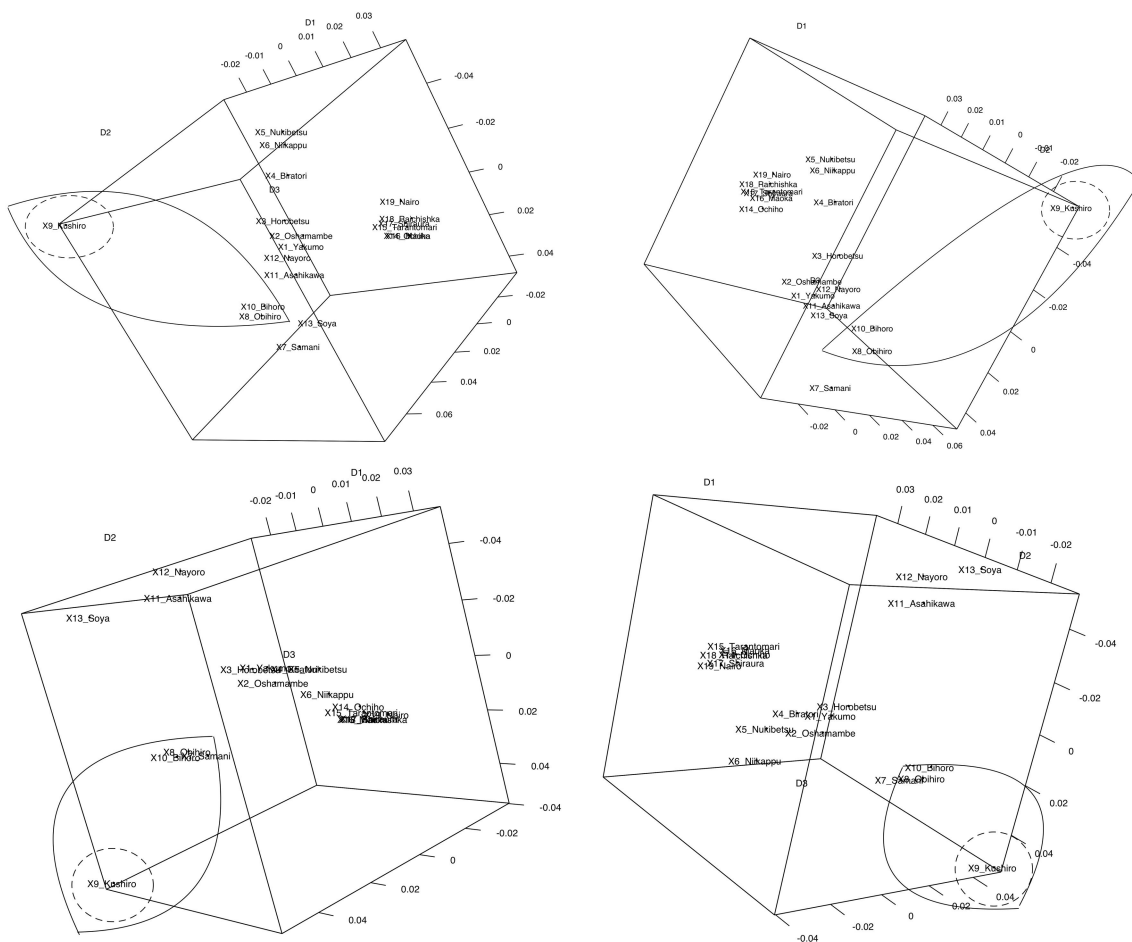


Figure 20. The result of three-dimensional plots for the Euclidean coordinate system for Kushiro dialect yielded by Homogeneity Analysis in Hattori and Chiri (1960). (Top) on ordinal scale (Bottom) on nominal scale

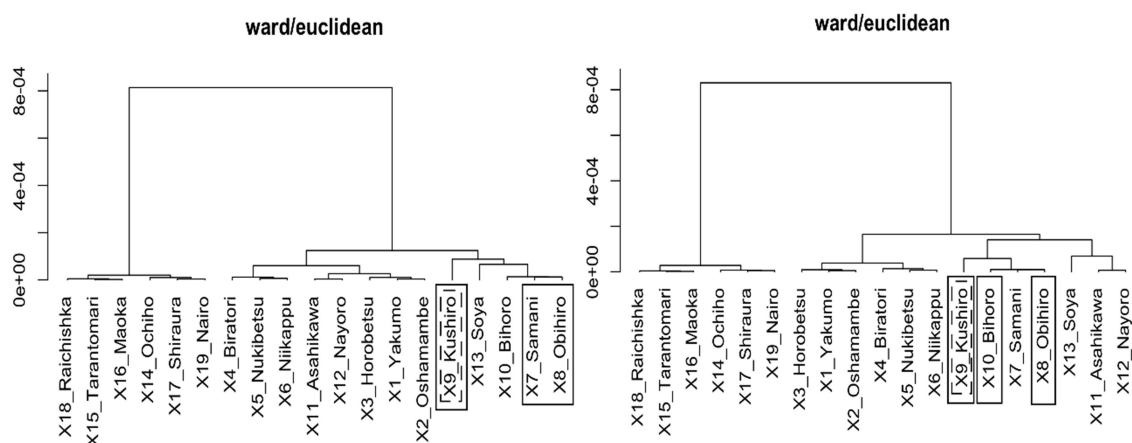


Figure 21. The results of cluster analyses of Euclidean distances for Kushiro dialect. (Left) on ordinal scale (Right) on nominal scale

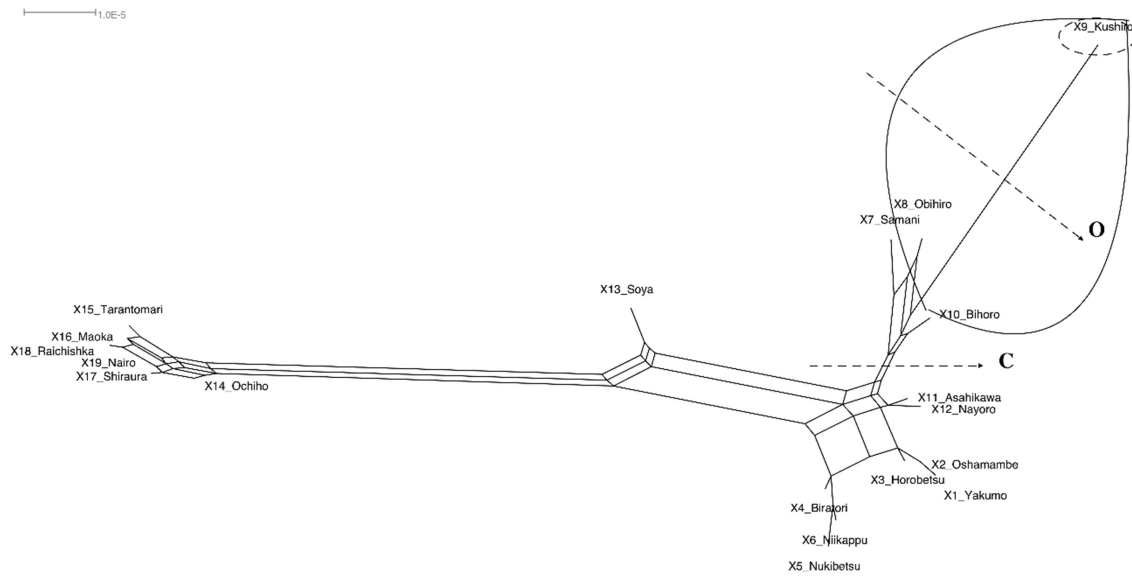


Figure 22. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Kushi dialect yielded by Homogeneity Analysis assuming an ordinal scale in Hattori and Chiri (1960) with the dimension value of 4.

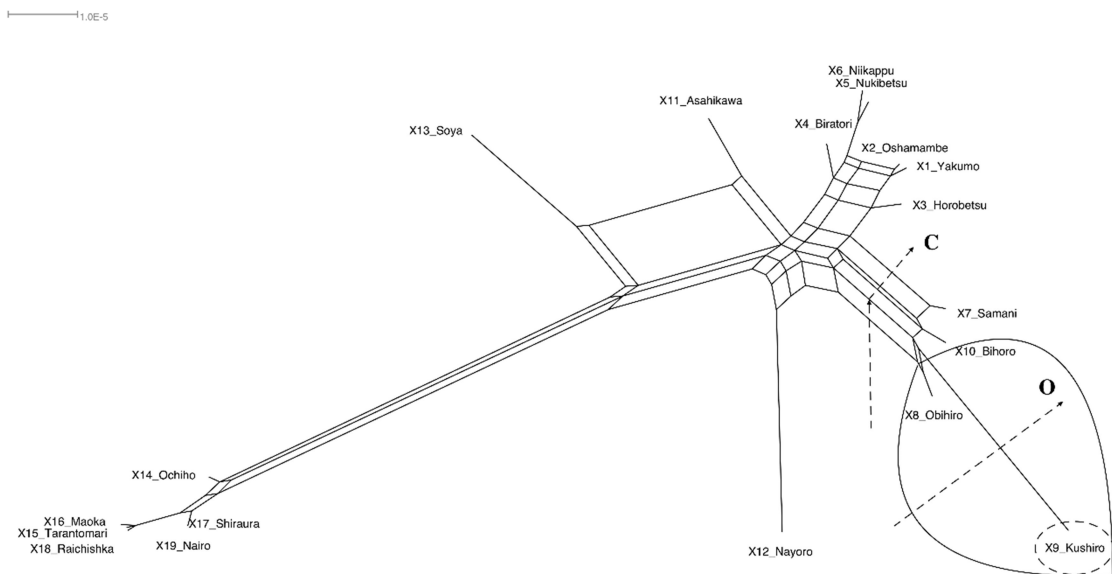


Figure 23. The result of Neighbor-Net Analysis applied to the Euclidean distance matrix for Kushi dialect yielded by Homogeneity Analysis assuming a nominal scale in Hattori and Chiri (1960) with the dimension value of 4.

Instead, the statistical analysis indicates that Nayoro dialect is an intermediate dialect between the northeastern and southwestern Hokkaido Ainu dialect groups. This suggestion is also consistent with the geographical location of Nayoro dialect in Figure 1.

The other analyses also support this finding. Nayoro dialect is located relatively

farther from the other dialect groups in the three-dimensional plot at the top of Figure 8 than those at the bottom. Moreover, the result of cluster analysis on the left in Figure 9 (i.e., on the ordinal scale) demonstrates that Nayoro dialect is an intermediate dialect between the northeastern and southwestern Hokkaido Ainu dialect groups in contrast to the right in Figure 9 (i.e., on the nominal scale).

Then, the author will focus on other results of Obihiro and Bihoro dialect. Note that all the results of the Neighbor-Net Analysis with the assumption of an ordinal scale in Asahikawa, Nayoro, Obihiro, and Bihoro dialects indicate less ambiguous structures than with the assumption of a nominal scale. Therefore, these results indicate that statistical analyses with the assumption of an ordinal scale have successfully visualized the relationships among the 19 Ainu dialects, at least from the perspectives of these four Ainu dialects.

Next, let us focus on the relationships among Obihiro, Bihoro, and Kushiro dialects in detail. These three dialects are surrounded by circles in each Neighbor-Net Analysis and 3-dimensional plot and by rectangles in each cluster analysis. Notably, they indicate certain homogeneous structures; the three dialects are located closely in both the Neighbor-Net Analysis and three-dimensional plots in the analyses of the lexicostatistical data for Asahikawa, Nayoro, Obihiro, and Bihoro dialects (see Figures 2–17).

However, the author observes different structures in the lexicostatistical data results for Kushiro dialect in Figures 18–21. First, Kushiro dialect is more distant from the Net (i.e., line O) in Figure 18 (i.e., on the ordinal scale) and located farther from Obihiro and Bihoro in the three-dimensional plot at the top of Figure 20 (i.e., on the ordinal scale) than those obtained for the other four dialects.

Second, Figure 18, which assumes an ordinal scale in Hattori and Chiri (1960), displays more ambiguous structures than Figure 19, which assumes a nominal scale. This result is inconsistent with our current observation that assuming an ordinal scale in Hattori and Chiri (1960) successfully captures the underlying information structure of the other four northeastern Hokkaido Ainu dialects.

Since the current statistical analysis with Homogeneity Analysis was performed with three dimensions following Ono (2015), the author attempted to analyze the data on the Kushiro dialect along four dimensions to investigate this phenomenon. Figures 22 and 23 are the results of Neighbor-Net Analysis calculated by the two Euclidean distance matrices that Homogeneity Analysis yielded from the same data in Hattori and Chiri (1960) with four dimensions assuming an ordinal scale and a nominal scale, respectively.

It is surprising that Figure 22, which assumes an ordinal scale in Hattori and Chiri (1960), displays *less* ambiguous structures than Figure 23, which assumes a nominal scale in Hattori and Chiri (1960), since Figure 18, which assumes an ordinal scale with

a dimensional value of 3, displays *more* ambiguous structures than Figure 19, which assumes a nominal scale. Essentially, the expanding number of dimensions reduces ambiguity in the Neighbor-Net Analysis results in the case of an ordinal scale and increases it in the case of a nominal scale.

Unless the fourth dimension, which the author added, contains significant information regarding the classification of Ainu dialects, the expanding number of dimensions normally increases the ambiguity of Neighbor-Net Analysis (i.e., several shorter sides in the Neighbor-Net increase). Therefore, the author reanalyzed Hattori and Chiri’s (1960) lexicostatistical data with the assumption of an ordinal scale and a dimensional value of 4. Due to space restrictions, the author will discuss the analysis in another paper but the primary results regarding northeastern Hokkaido Ainu dialects are summarized as follows:

Four main components in the analyses of northeastern Hokkaido Ainu dialects have been observed: (1) the distinction between Sakhalin dialects and other dialects; (2) the distinction between the northeastern and southwestern Hokkaido Ainu dialect groups; (3) the distinction between the northern and southern parts of Hokkaido; and (4) the originality of each given dialect whose features tend to increase further to the east each dialect is located in Hokkaido.⁵⁶ Note that these four components affect each other in practice. Since the strong originality of the Kushiro dialect ranks as the third component in Figure 18, it fails to capture the component corresponding to the distinction between the northern and southern parts of Hokkaido in the fourth dimension. Therefore, the ambiguity of the Neighbor-Net Analysis has been reduced in Figure 22.

Figures 24–26 present an example of Homogeneity Analysis with the assumption of ordinal scale and a dimensional value of 4 for Obihiro dialect. The first dimension clearly corresponds to (1) the distinction between Sakhalin dialects and other dialects, the second dimension to (2) the distinction between the northeastern and southwestern Hokkaido Ainu dialect groups, the third dimension to (3) the distinction between the northern (i.e., Soya dialect) and southern parts of Hokkaido (i.e., Samani dialect), and the fourth dimension to (4) the originality of each given dialect.

⁵ The statistical analysis assuming an ordinal scale in this paper succeeded in visualizing the last two components that present statistical analyses (Asai 1974; Lee and Hasegawa 2013; Ono 2015) assuming a nominal scale in Hattori and Chiri (1960) have obscured. However, statistical analyses assuming an ordinal scale from the perspectives of southwestern Hokkaido Ainu dialect groups in Hattori and Chiri (1960) do not clearly demonstrate these two components. Due to space limitations, the author will deal with this interesting issue in another article.

⁶ The author notes each eigenvalue of the fourth component, whose value can be considered as the strength of the component in the statistical analyses of the Obihiro, Bihoro, and Kushiro dialects, as follows: 0.000312, 0.000318, 0.000404 that corresponds to the case for Obihiro, Bihoro, and Kushiro dialect respectively.

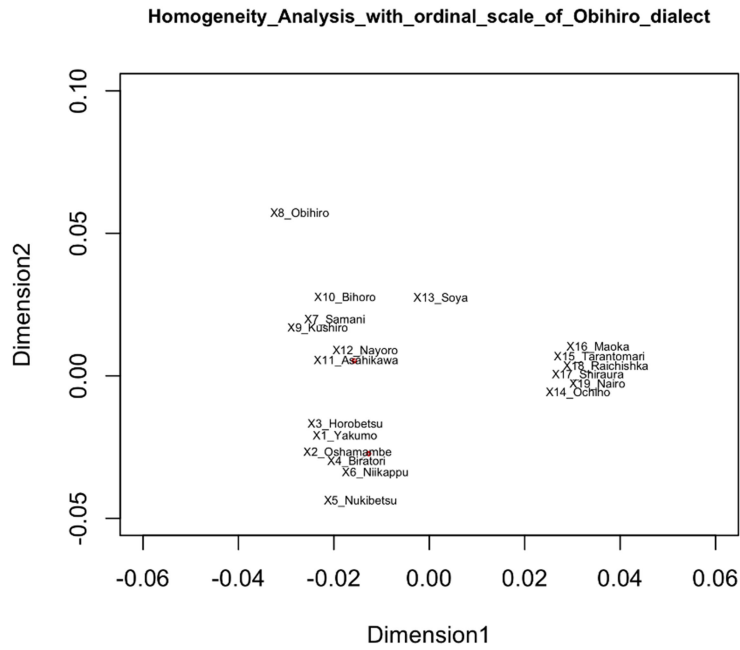


Figure 24. The result of Homogeneity Analysis for Obihiro dialect on an ordinal scale with the dimension value of 4. The horizontal axis is the first dimension calculated by Homogeneity Analysis and the vertical axis is the second dimension. Each dimension is normalized by its corresponding eigenvalue.

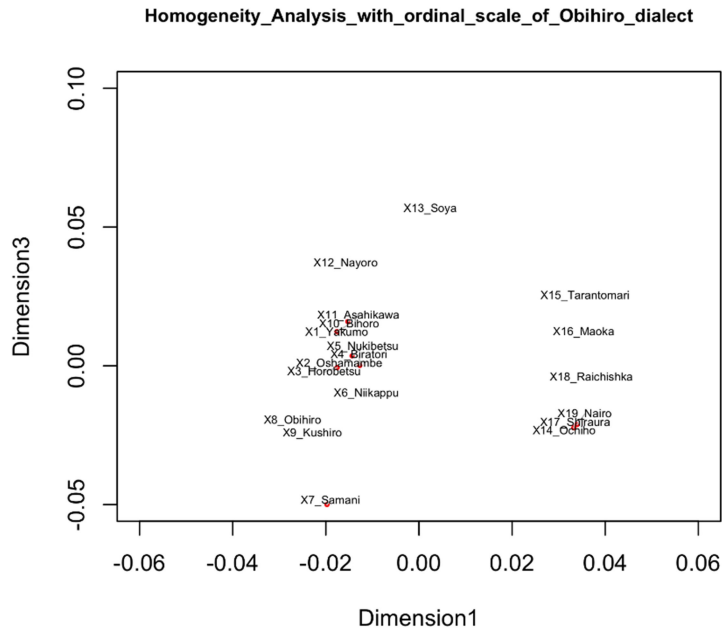


Figure 25. The result of Homogeneity Analysis for Obihiro dialect on an ordinal scale with the dimension value of 4. The horizontal axis is the first dimension calculated by Homogeneity Analysis and the vertical axis is the third dimension. Each dimension is normalized by its corresponding eigenvalue.

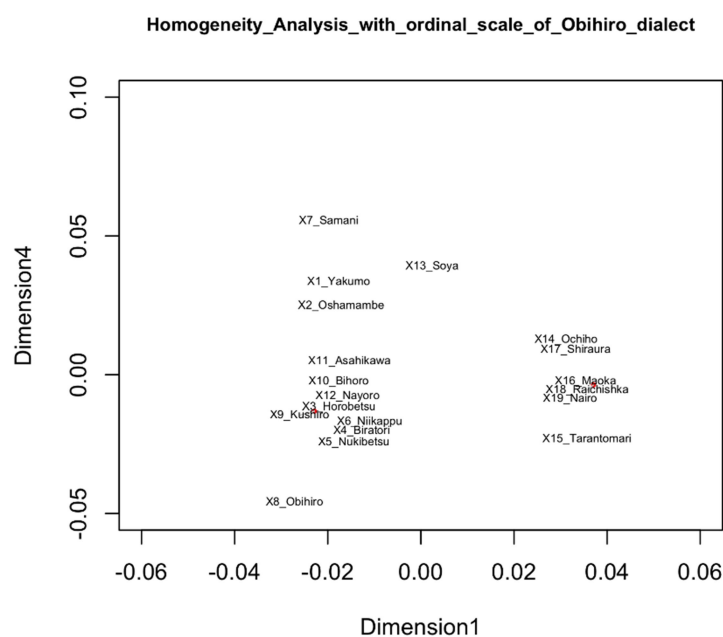


Figure 26. The result of Homogeneity Analysis for Obihiro dialect on an ordinal scale with the dimension value of 4. The horizontal axis is the first dimension calculated by Homogeneity Analysis and the vertical axis is the fourth dimension. Each dimension is normalized by its corresponding eigenvalue.

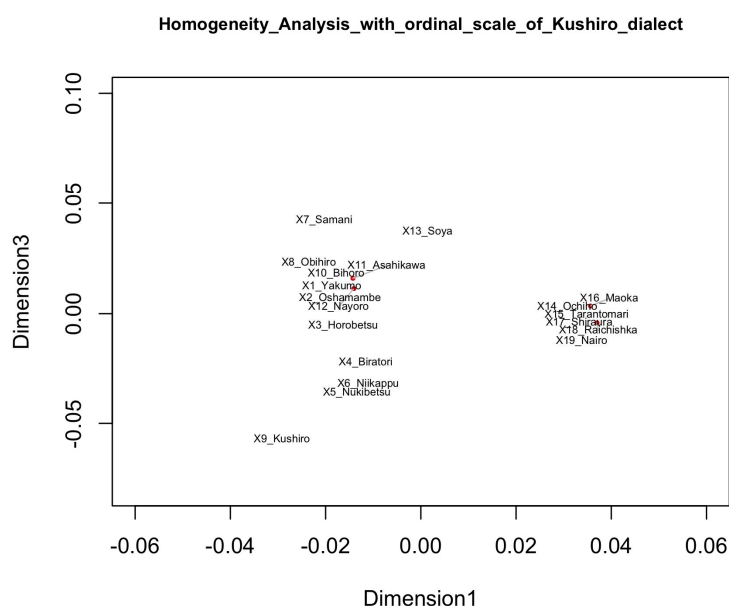


Figure 27. The result of Homogeneity Analysis for Kushiro dialect on an ordinal scale with the dimension value of 4. The horizontal axis is the first dimension calculated by Homogeneity Analysis and the vertical axis is the third dimension. Each dimension is normalized by its corresponding eigenvalue.

Furthermore, Figure 27 demonstrates an example of Homogeneity Analysis with the assumption of ordinal scale and a dimensional value of 4 for Kushiro dialect. The horizontal axis is the first dimension corresponding to (1) the distinction between Sakhalin dialects and other dialects and the vertical axis is the third dimension corresponding to (4) the originality of Kushiro dialect (in this case, the originality of Kushiro dialect is stronger than that of Obihiro in terms of each eigenvalue).⁷

Section 4 discusses the significance of this paper and states a conclusion.

4. Discussion and Conclusions

This Section discusses what the main results of this paper suggest for current research on the Ainu language from the perspectives of statistics and linguistics.

From the statistical perspective, the results indicate that the statistical analysis with the assumption of an ordinal scale in Ono (2019, to appear) successfully captures the underlying information structures of the Asahikawa and Nayoro dialects and reveals some hidden (or previously unknown) relationships or components in Hattori and Chiri's (1960) lexicostatistical data on which the improved statistical methodologies provide new insight.⁸

Notably, the component corresponding to the distinction between the northern and southern parts of Hokkaido is first illustrated with the statistical analysis with the assumption of the ordinal scale in Hattori and Chiri (1960).

Moreover, the result that there is one component, considered to be the originality of each given dialect and whose features tend to increase further to the east the dialect is located in Hokkaido, coincides with the author's previous joint research with Dr. Mika Fukazawa (Fukazawa and Ono 2015) based on old documentation of the Ainu language and her geolinguistic and philological doctoral thesis on the *Kagake Monjo*

⁷ "wordcloud" package (Ian 2018) was utilized to avoid overlaps of the symbols in Figures 24-27.

⁸ One of reviewers suggested a need for further statistical investigations with the ordinal assumption based on the data in Asai (1974). Since Asai (1974: 66) has made some modifications on word forms in the Asahikawa, Obihiro, and Kushiro dialects whose informants have brought invaluable contributions in the research on the Ainu language, statistical analyses based on the data type in Hattori and Chiri (1960) with these modifications will be promising in future research. However, there still remain several obstacles on the modifications. Since Asai (1974) has not demonstrated each cognacy judgment among the word forms but focused on whether there is one common word form on the word among Ainu dialects or not, as described in the definition on "relation index" in Asai (1974: 61-62), consequently, Asai (1974)'s each cognacy judgment among the word forms was not demonstrated in the case that there are two or more word forms on the word in either or both dialects. Therefore, the modifications of Hattori and Chiri's (1960) data based on Asai (1974) require researchers recovering each cognacy judgment among the word forms in Asai (1974) from his "data matrix" (Asai 1974: 92; Table 1). The author will deal with this interesting and novel issue in another article.

(Fukazawa 2017) concerning Nemuro dialect, an Ainu dialect located to the east of Kushiro.

Further improvements in statistical methodologies on the data type in Hattori and Chiri (1960) will be the issue concerning the components among the Ainu dialects. From the linguistic perspective, the main results may be summarized under four points. First, the four components identified in this paper can potentially contribute to determining the origin of the Ainu language. The last two components that have been not clearly observed in the statistical analyses from the viewpoints of each dialect in the other groups will be of concern in future research (i.e., the distinction between the northern and southern parts of Hokkaido; and the originality of each given dialect whose features tend to increase further to the east each dialect is located in Hokkaido).

Second, the component corresponding to the distinction between the northern and southern parts of Hokkaido will be the matter of concern in Ainu linguistics. While Kindaichi (1955: 728) has already noted the differences among the northern and southern parts of Hokkaido whose classification was different in some points from the statistical findings in this paper, Asai (1974: 48) commented Kindaichi’s classification as “*Kindaichi’s vague classification was based on a few words and a few phonological phenomena; thus we cannot agree with his descriptions on several points.*”

However, as demonstrated in Section 3, previous statistical analyses results, including Asai (1974), were obscured by the assumption of a nominal scale in Hattori and Chiri’s (1960) data and that the statistical analysis assuming an ordinal scale succeeded in capturing the component: the distinction between the northern and southern parts of Hokkaido.

Third, the strong originality of the Kushiro dialect, as demonstrated in Figures 18-23, will also be of concern in Ainu linguistics. The author notes that Kindaichi (1955: 728) has also commented on the originality of the Kushiro dialect as “Kushiro may be related to the North dialect but it may be counted as a small independent one, East dialect.” As a result, Kindaichi’s insightful views have corresponded to the statistical findings in this paper.

Finally, the fourth component in the analysis suggests that Ainu dialects located to the east of Kushiro in Hokkaido constitute an unknown problem. Geolinguistic and philological research on the *Kagake Monjo* (Fukazawa 2017) has already contended with these issues independently of this paper.

Furthermore, the results denote the significance of North Kurile dialect (No. 22 in Figure 1). As Fukazawa (2017) indicates, Ainu dialects located to the east of Kushiro are situated in connection with North Kurile (and Kamchatka), Hokkaido, and Sakhalin, where the Ainu language is or was spoken.

Therefore, it will be promising to reanalyze Hattori and Chiri (1960), including the systematic review process of the lexicostatistical data related to Ainu dialects

located to the east of Kushiro and North Kurile and reexamine the cognacy judgments based on current Ainu linguistics. The progress of the statistical methodologies should also be productive, as indicated by the primary results of this paper.

The main results in this paper suggest that the five “northeastern” Ainu dialects (i.e., Nos. 8–12 in Figure 1) contain the underlying dialect relationships with diversity that the statistical analyses with the assumption of the ordinal scale succeeded in capturing but the previous statistical analyses with the assumption of the nominal scale have obscured.

In conclusion, the author hopes that complementary studies of Ainu linguistics and statistics will prove invaluable to the discipline in the future.

Acknowledgements

The author is very grateful to the editors and two highly conscientious reviewers; all errors are of course my own.

References

- Asai, Tōru (1974) Classification of dialects: Cluster analysis of Ainu dialects. *Bulletin of the Institute for the Study of North Eurasian Culture*, 8, 45–136.
- Bugaeva, Anna (2012) Southern Hokkaido Ainu. In Nicolas, Tranter (ed.) *The languages of Japan and Korea*, 461–509. London: Routledge.
- de Leeuw, Jan and Mair Patrick (2009) Gifi method for optimal scaling in R: The package homals. *Journal of Statistical Software*, 31(4), 1–20.
- Fukazawa, Mika (2017) *Kagake monjo ni'okeru Ainugo no bunkengaku-teki kenkyū* [A philological study of the Ainu language in the Kaga family's archives], Doctoral Thesis. Chiba: Chiba University. (in Japanese)
- Fukazawa, Mika and Ono Yohei (2015) Ainugo shohōgen ni mirareru sanshuno seibun [Three Components in the Ainu Dialects] Research presentation at the Symposium, *Typological and Historical/Comparative Research on the Languages of Japanese Archipelago and its Environs*, conducted at the meeting of National Institute for Japanese Language and Linguistics, Tachikawa, Tokyo, 14 June 2015.
- Gifi, Albert (1990) *Nonlinear multivariate analysis*. New York: John Wiley and Sons.
- Geospatial Information Authority of Japan (2018) Ministry of Land, Infrastructure, Transport and Tourism. <https://maps.gsi.go.jp> [accessed on September 2018].
- Hattori, Shirō and Chiri Mashiho (1960) Ainugo shohōgen no kisogoi tōkeigakuteki kenkyū [A lexicostatistical study on Ainu dialects]. *Kikan minzokugaku kenkyū* [The Japanese Journal of Ethnology], 24(4), 307–342. (in Japanese)
- Huson, Daniel and Bryant David (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(1), 254–267. Oxford: Oxford University Press.

- Ian, Fellows (2018) wordcloud: Word Clouds. R package version 2.6.
<https://CRAN.R-project.org/package=wordcloud>
- Kindaichi, Kyōsuke (1955) Ainu-go [the Ainu language] In Ichikawa, Sanki and Hattori Shirō (eds.) *Sekai gengo gaisetsu [An introduction to the languages of the world]*, 725-749. Tokyo: Kenkyūsha. (in Japanese)
- Lee, Sean and Hasegawa Toshikazu (2013) Evolution of the Ainu language in space and time. *PloS One*, 8(4), e62243.
- Ono, Yohei (2015) Statistical reanalysis of the classification of Ainu dialects: On the data of Hattori and Chiri (1960), *Journal of the Center for Northern Humanities*, 8, 25–41. (in Japanese)
- Ono, Yohei (2019, to appear) The ordinal scale on lexicostatistical data in Ainu dialects: Towards a new interdisciplinary research among the humanities and statistics. *Journal of the Center for Northern Humanities*, 12.
- R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
<https://www.R-project.org/>.

Summary

A previous paper (Ono 2019, to appear) demonstrated that the assumption of an ordinal scale revealed relationships among 19 Ainu dialects in Hattori and Chiri’s (1960) lexicostatistical data that were indicated philologically numerous times but without sufficient statistical clarity. This paper focuses on northeastern Hokkaido Ainu dialects that have not been as adequately researched as southwestern Hokkaido Ainu dialects (Bugaeva 2012). The author applies a statistical analysis based on an ordinal scale to five northeastern Hokkaido Ainu dialects (i.e., Asahikawa, Nayoro, Obihiro, Bihoro, and Kushiro dialects), comparing the results to those obtained employing a nominal scale.

This paper’s primary results may be summarized in four points. First, previous statistical analyses (Asai 1974; Lee and Hasegawa 2013; Ono 2015) classified the Asahikawa and Nayoro dialects as northeastern Hokkaido Ainu dialects. However, the statistical analysis in this paper positions both dialects as intermediate dialects between these two groups, rather than as northeastern Hokkaido Ainu or southwestern Hokkaido Ainu. Furthermore, the author observed that both these dialects are interlinked with Sakhalin dialects. Second, the component corresponding to the distinction between the northern and southern parts of Hokkaido is first illustrated statistically. Third, the author observed a homogeneous structure among Obihiro, Bihoro, and Kushiro dialects from the statistical analysis of four northeastern

Hokkaido Ainu dialects. However, the analysis of the Kushiro dialect demonstrated a heterogeneous structure among these three dialects, identifying some similarity with the Soya and Sakhalin dialects. Fourth, these results statistically suggest that Ainu dialects located to the east of Kushiro in Hokkaido present a previously unknown problem. This paper's suggestion coincides with the geolinguistic and philological research on the *Kagake Monjo* (Fukazawa 2017) concerning Nemuro dialect, an Ainu dialect located to the east of Kushiro.

(linguistics.dialectometry@gmail.com)