How similar are Dene languages?

Conor Snoek & Michaela Stang
University of Alberta

ABSTRACT. Dene is a family of languages spoken over a wide area of western North America, stretching west from northern Canada to the coast of Alaska and south to the Mexican border. Despite the distances between groups of speakers and substantial cultural diversity, Dene languages share grammatical similarities and widespread patterns of cognation. The grouping and classification of Dene languages, however, remains a long-standing problem. To achieve insight into the historical relationships among Dene languages, we propose the use of quantitative methods capable of discerning group structure in large bodies of data, such as clustering and multi-dimensional scaling. We apply these methods to a sample of 46 Dene languages and dialects. Since the fact of language relatedness is beyond question and the focus of interest is on degrees of relatedness, we use a sample of three semantic domains of human anatomy, mammals, and insects, instead of lists of purported universal meanings.

Keywords: Dene, dialectometry, classification, clustering.

1. INTRODUCTION. Dene or Athapaskan languages have long been understood to belong to a single, geographically dispersed, language family. The connection between Dene languages, and the languages Haida and Tlingit was formally proposed by Sapir (1915), who also suggested the name Na-Dene. There is considerable literature on Dene language chronology and reconstructions of Proto-Dene (see for example Dyen & Aberle 1974, Hoijer 1956, Kroeber 1959) and also on the extension into the larger Na-Dene and Dene-Yenisean families (Enrico 2004, Kari & Potter 2011). The attention devoted to these deep connections, while worthwhile and important, has meant that internal relationships within the Dene family have received less attention. The last concerted effort to produce an internal classification of the Dene languages was carried out by Hoijer (1963), who found that the sound correspondences exhibited by different groups did not render the kinds of distinction necessary for the construction of a phylogenetic tree. Hoijer was only able to distinguish a group of Dene languages in the southwestern part of the United States, the Apachean languages. These languages were geographically isolated, and at a great distance to the other Dene languages spoken in present-day Canada, Alaska, California and Oregon. Especially in Alaska and Canada, phonological differences appear to result from "areal diffusion of separate innovations from different points of origin" (Krauss & Golla 1981:68). This has led to a situation in which each Dene-speaking community is a "unique conglomerate" (ibid). As a consequence, the classification of Dene languages into sub-groups has been carried out predominantly along geographical divisions (Mithun 1999). Indeed, highest-level language groupings for the Dene family tend to be based on the three most salient geographic locations in which these languages are spoken: the Northern Dene (Alaska, Northwestern Canada), Pacific Coast Dene (Oregon, California), and Southern Dene (predominantly in Arizona and New Mexico, and to a lesser extent in the adjoining states).

While the last of these groups, generally referred to as Apachean, can be grouped according to the traditional historical linguistic criteria of shared sound changes and retentions (Hoijer 1938), it presents the lone exception in a family where patterns of sound correspondence cross-cut each other, in a manner extensive enough to obscure evidence of phylogenetic branching. Leading scholars, such as Michael Krauss and Victor Golla (1981) have come to the conclusion that the historical signal is so distorted that the possibility of identifying linguistic groupings above the level of individual Dene languages is near impossible.

These difficulties in classification have been postulated to result from sustained dialect admixture, especially among the northern Dene languages, spoken in Alaska and the Canadian Interior (Krauss 1973, 2005, Krauss & Golla 1981), which are subject to constant and dynamic interactions (Krauss 2005:118). Rejecting even the future possibility of the discovery of synapomorphic patterns capable of producing clearly defined phylogenetic branches, Krauss has suggested that the *Stammbaum* (phylogenetic tree) model cannot adequately capture the reality of Dene language history: "… it is fully to be expected that virtually the only meaningful internal relationships will be those which can be shown far better in terms of isogloss maps than by any other means" (1973:949). While we do not want to go quite that far, we accept Krauss' suggestion and follow the spirit of his idea in this comparative study of Dene lexical items. In fact, at an earlier point in his career, Krauss (1964:120) was convinced that it was possible to classify the Dene languages: "Further progress in classification will require careful study of the chronology of phonological changes in the sorting of inherited and diffused phenomena, collated with morphological, lexical, and ethnographic data." We wholeheartedly agree with Krauss' hypothesis here, and argue that with the modern methods of dialectometry and the advances in

the lexicographic description of Dene languages, the time is ripe for a fresh attempt to classify the Dene languages.

Our approach is to suspend the search for further and more telling sound correspondences, and instead address the question of language classification from the perspective of the phonological character of different languages based on aggregate measurements of individual phonological strings. We reinterpret Krauss' suggestion in light of new understandings of dialectal relationships and especially new research on quantitative dialectology known as DIALECTOMETRY (Goebl 2005). Krauss essentially advocates a dialectological approach in pointing out that language relationships may be identified through isoglosses, but fails to describe how language areas or groupings can emerge if isoglosses are not found to coincide. This problem, the identification of groupings in the case where isoglosses delimit non-commensurate areas, can be tackled through the use of aggregates of dialect features rather than individual features (Wieling & Nerbonne 2015). Even more precise assessments of language grouping can be obtained through numerical measures of phonological distance between dialects. We, therefore, follow through on Krauss' suggestion by implementing a dialectometric approach to the estimation of similarity among Dene languages. The dialectometric approach has several distinct advantages. Firstly, it produces information on the structure of internal groupings of Dene languages in a manner that goes beyond the predominantly regional arrangement of Dene languages that dominates current classificatory thinking in the field (Goddard 1997, Mithun 1999). Secondly, dialectometry relies on well-understood and easily accessible techniques, such as clustering, allowing for the results to be readily replicated.

The classification of Dene languages is an important part of the study of Dene (pre)history. The larger undertaking of tracing the history of Dene-speaking peoples is an

inherently interdisciplinary endeavour, also involving archaeology and anthropology, but one in which linguistics has a special place. Linguistic information is particularly important because the large geographic spread and non-contiguous nature of the Dene language communities results in a wide range of cultural diversity. Traditional means of clothing production, methods of subsistence, and housing vary tremendously, making it difficult to associate disparate Dene peoples on the basis of social and material culture. In fact, Dene peoples have shown a particular propensity to adopt the cultural practices of their geographic neighbors (Ives & Rice 2008). Despite cultural adaptations and geographic non-contiguousness, Dene speakers have retained a remarkable level of language similarity, sharing easily discernible grammatical structures and widespread patterns of cognation (Rice 2012). Thus, the social and material culture of Dene-speaking peoples are varied and do not indicate a common origin. In contrast, the linguistic data provide a strong indication of historical association, and the actual membership of individual languages in the family is beyond dispute.

The special place of linguistics in the study of Dene (pre)history emerges from these facts. This behooves us, as linguists, to present research results in a manner that speaks to the interdisciplinarity of these research questions. We therefore present our results in diagrams and maps that, we hope, allow for an intuitive understanding of the complex linguistic data that underlie them. *Gabmap* is a research tool which allows us to plot our data onto geographic maps annotated with information regarding the nature and strength of the relationship of the represented languages. We consider the results of these classifications as a temporary stepping stone to a fuller understanding of Dene cultural and linguistic history, which we believe will ultimately be attained through the conjunction of research in linguistics, archaeology, ethno-history, and anthropology. Therefore, we are advocating two kinds of transparency: firstly, a

transparency of data and methods to achieve a more open research process inclusive of all scholars of Dene, inside and outside of academia, and, secondly, a transparency across fields that aims to make itself understood to related disciplines.

In the following section we describe our methods in more detail. This is followed by a discussion of the sample, which requires special merit. We conclude with a discussion of the results.

2. METHODS. Each of the languages is represented by an ordered list of phoneme strings encoding individual referents. The similarity of languages to each other is estimated on the basis of a numerical measure of the similarity between phoneme strings encoding the same referent-concepts in different languages. The algorithms used for this are implemented in the software Gabmap (Nerbonne et al. 2011), freely available over the Internet.

Similarity of phoneme strings is operationalized by means of the simple LEVENSHTEIN DISTANCE. This is a measure of the distance between two words informally defined as the minimum number of single-character edits required to change one string into another phonologically (Sankoff & Kruskal 1983:18). The phoneme strings are aligned so that vowels will be compared with vowels and consonants with consonants. The distance between two strings is then established by comparing each character: if the characters are identical at an aligned location in each of the two strings, the distance will be measured as 0. If the two aligned characters are different, the distance will be measured as 1. Should only a diacritical mark indicating nasalization, tone, length or aspiration (e.g. /t/ vs. /tʰ/) distinguish the two characters, the distance will be measured as 0.5. The distance between two strings is the sum of the character distances. So as to compensate for differences in word length, the "distance of each

word pair is normalized by dividing it by the mean length of the word pair" (Nerbonne et. al 1999). In Table 1, two phoneme strings representing the concept 'thumb' are compared in the languages Ahtna and Kaska, and are found to have a Levenshtein distance of 3.5.

Compared referent-concept: 'thumb'							
Ahtna	l	a		kʰ	o	ts'	
Kaska (Frances Lake)	l	a:	s	ʧʰ	o	ʔ	
Levenshtein Distance	0	+ 0.5	+ 1	+ 1	+ 0	+ 1	= 3.5

TABLE 1. Measuring the simple Levenshtein distance between two strings

This measurement is taken between every possible pair of referent-concepts. The result is a distance matrix in which each cell lies at the intersection of two languages and contains the numeric distance value between them, as can be observed in the matrix excerpt in Table 2. This matrix serves as the input for the clustering algorithm.

	Deg Xinag	Koyukon	Dena'ina (Iliamna)
Deg Xinag	0	0.387154	0.431306
Koyukon	0.387154	0	0.372488
Dena'ina (Iliamna)	0.431306	0.372488	0

TABLE 2. Excerpt of distance matrix showing aggregate Levenshtein distances between

languages

From the matrix of distances, a dendrogram is constructed using hierarchical agglomerative clustering. This is a technique for finding groups in data which does not assume a predetermined number of clusters.[1] This kind of clustering has been shown to be highly sensitive to even small

variations in the data. To find stable results, the clustering procedure is repeated multiple times under varying conditions. Additional variations in the data (noise) are added artificially and the resulting dendrograms are recorded. After 100 repetitions, the likelihood of the repeated emergence of individual clusters can be assigned a percentage value indicating the stability of the cluster. This technique is known as FUZZY CLUSTERING (Nerbonne et al. 2008). Only clusters that appear in 60% or more of the dendrograms are deemed stable and considered further in the analysis below.

Finally, it is worth briefly mentioning how Gabmap handles missing values. This is important, because, as we describe below, Dene languages are often under-described and data availability varies considerable from language to language. Gabmap allows us to capitalize on the available data by drawing comparisons only between available strings. While missing data still lead to a less representative measurement of language similarity, aggregating the results from all available items allows us to achieve the highest possible precision, given the available information. The next section describes the dataset we used in more detail.

3. DATA. This study uses data collected through the database building project the Pan-Athapaskan Comparative Lexicon (PACL). PACL is an interdisciplinary database at the University of Alberta, supervised by Dr. Sally Rice of the Department of Linguistics and Dr. John Ives of the Department of Anthropology. The goals of PACL are: (a) to build a better cross-linguistic lexical set to compare contemporary and historical relations among Dene communities, (b) to make an accessible and hands-on living archive for Dene speakers to assist them in revitalisation work, and (c) to create a framework to help identify converging lines of evidence to answer questions about ancient Dene relations and migrations (Rice et al. 2013).

Currently, PACL entries are collected from dictionaries, unpublished field notes, and other text sources, all of which are publicly available. The accessibility of the data is a point that bears emphasising in the context of Dene historical linguistics, a field in which discussion and research has been marred by general lack of transparency and unavailability of data. This lack can now be amended thanks to the greater ease of access to digitized lexical resources on Dene, especially those resources made available through the Alaskan Native Language Center (2015), as well as more recently published print dictionaries. We hope that this trend continues and want to emphasize that the making public of data and methods has to be the way forward in Dene linguistics, and that the wider involvement of researchers outside and inside of the academia can only be a benefit to the field as a whole.

Each entry in PACL is annotated with as much information from its source as possible and additional labels and notes are made to help with future recall and analyses. To create the larger PACL database, entries are added according to semantically defined domains in order to manage and mitigate the overwhelming task of lexical collection. As we continue with PACL, we will keep incorporating more data to help continuously build a better and more complete picture of Dene language relatedness.

For this study, we used a sample of the larger database, focusing on three semantically defined lexical sets. These included terms referring to parts of the human body, mammals, and insects. These three semantic domains were chosen because they contain terms most frequently available for Dene languages in PACL. This resulted in 5,588 unique phoneme strings, found across 346 referents. Every item has been transcribed from its source's original orthography into International Phonetic Alphabet notation, ensuring proper alignment and distance measurement in Gabmap (discussed in Section 2). An illustration of the database is given in Table 3.

Language	Referents		
	LEG	BEAVER	RABBIT
Ahtna	ts'ən	cʰɑʔ	qɑχ
Gwich'in (Tl)	tθ'àn?	tsʰè:ʔ	kʷeh
Tanacross	ɣolʔ	tsʰàʔ	ga
Hare	w'én	sáʔ	kah
Dene Sųłiné	tθ'ẽn	tsʰá	kah
Witsuwit'en	qʰecən	tsʰɑ	qəχ
Hupa	ts'ɪŋʔ	tʃʷaʔ	na:q'ɪtʰahk'ɪɬɪxun
Tolowa	ts'e:n	chi:ntʰeł	gɑ:mɑʔʂ
Navajo	tʃǎ:t	tʃʰa:ʔ	kah
Jicarilla Apache	tʃátí	tʃʰə́:	kə̃h
Plains Apache	dʒà:h	t'shá	kʰáltsuɾe
Tsuut'ina	wus	tsʰò	nìtł'ótòɣà

TABLE 3. Sample of languages and referents from PACL

Each referent-concept varies in number of total items, meaning that many of the concept lists are incomplete. This is due to two main factors: under-documentation of languages and range of ecosystems. Many Dene languages are moribund or extinct and suffer from under-documentation, and many of the concepts we search for are simply not recorded and are missing from PACL. The other factor which leaves gaps in PACL is the range of ecosystems in which Dene communities are in. For example, the types of mammals and insects found in the Arctic Circle will vary from those in the American Southwest or the Pacific Coast; scorpions and tarantulas are not found in Northern Dene communities and caribou are not found in Apachean or Pacific Coast Dene communities, and thus we do not expect to have entries for the respective languages. However, many of the referent-concepts are shared among all the languages, especially those of body parts, which are semantically more stable (Snoek 2013).
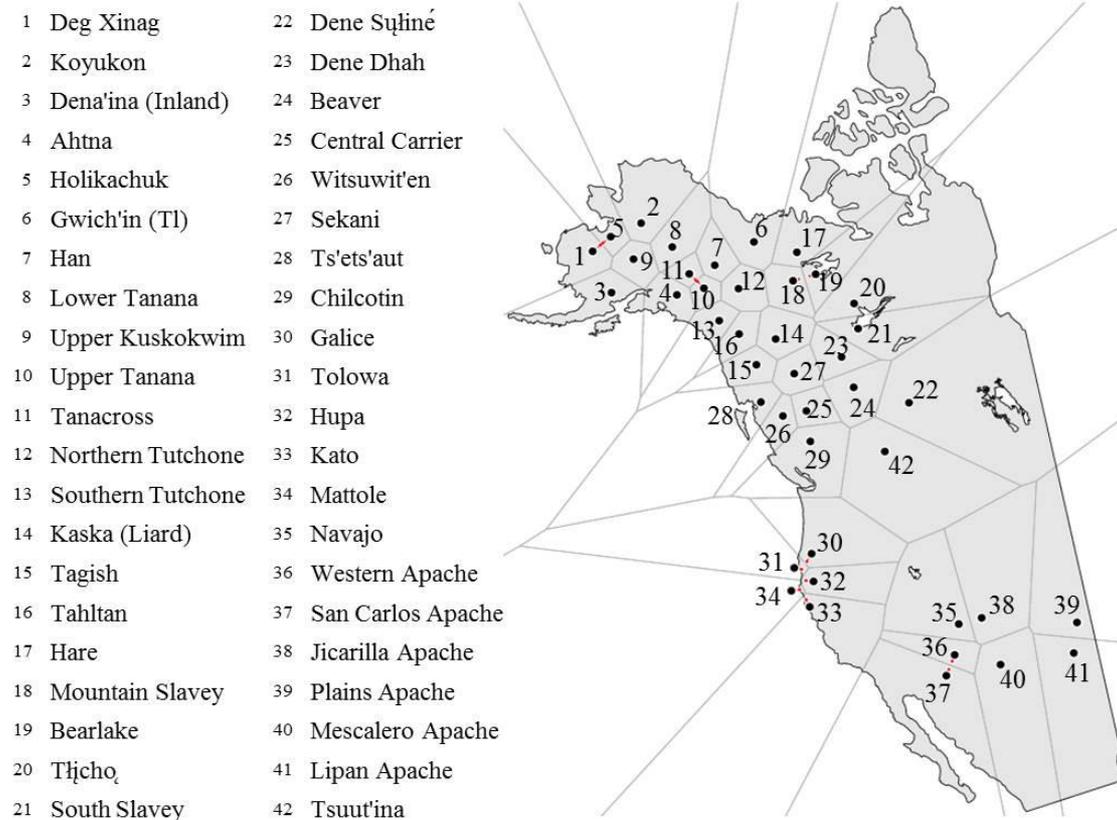
1	Deg Xinag	22	Dene Sųłiné
2	Koyukon	23	Dene Dhah
3	Dena'ina (Inland)	24	Beaver
4	Ahtna	25	Central Carrier
5	Holikachuk	26	Witsuwit'en
6	Gwich'in (Tl)	27	Sekani
7	Han	28	Ts'ets'aut
8	Lower Tanana	29	Chilcotin
9	Upper Kuskokwim	30	Galice
10	Upper Tanana	31	Tolowa
11	Tanacross	32	Hupa
12	Northern Tutchone	33	Kato
13	Southern Tutchone	34	Mattole
14	Kaska (Liard)	35	Navajo
15	Tagish	36	Western Apache
16	Tahltan	37	San Carlos Apache
17	Hare	38	Jicarilla Apache
18	Mountain Slavey	39	Plains Apache
19	Bearlake	40	Mescalero Apache
20	Tłįchǫ	41	Lipan Apache
21	South Slavey	42	Tsuut'ina

FIGURE 1. Languages sampled with relative location according to Gabmap

The languages covered in this study total 42, representing all the Dene languages for which we have access to sources. Figure 1 is a list of the languages and their relative geographic location. Where there is significant documentation on dialectal variation for one language, we have indicated the dialect in parentheses which represents the language used for this study. Please note in Figure 1 that the boundaries on this map are artificially placed by Gabmap and do not represent real geographic ranges for these languages, while also excluding the locations of many other indigenous languages found in this area.

4. RESULTS. As indicated above, Gabmap was used to calculate language similarity on the basis of the aggregated distances between phoneme strings. Gabmap provides visualizations of these results in the form of cluster dendrograms and maps. The complete dendrogram resulting from the fuzzy clustering is given at the end of this section in Figure 5. In order to make transparent our results, we have chosen to represent the resulting clusters on geographic maps, which are color coded to aid visual identification of the groups and associated regions. Many of the clusters reach stability values that exceed the pre-determined threshold. In Figures 2-4, the numbers above a branch indicate the percentage of times that groupings re-emerged through the fuzzy clustering algorithm. In this section, each of the clusters is described in detail.

As shown in Figure 2, the language cluster formed by Deg Xinag and Holikachuk, Lower Tanana and Upper Kuskokwim, Ahtna, Koyukon, and Dena'ina (Inland) is very stable, with each branch re-emerging 100% of the time. This group of Western Alaskan languages agrees with common geographic groupings; however, Gwich'in and Han, Tanacross and Upper Tanana, and Northern Tutchone are generally associated with the Alaskan grouping (Mithun 1999:344). Through this analysis, they are clustered separately and with relative stability (Figure 2). Northern and Southern Tutchone are generally grouped as two closely related dialects, but according to phonemic differences, they do not fall under a cluster, let alone a similar branch. Southern Tutchone instead forms a close relationship to Tagish. It must be noted however that the latter language is very under-documented, and consequently this grouping must be considered tentative despite the high reliability value. This placement of Tagish contrasts with Rice's classification (in Mithun 1999), where Tagish is placed within the Northern Cordillera Languages along with Tahltan and Kaska. Instead, these last two languages cluster with Sekani; Kaska and Sekani represent a particularly stable cluster, reappearing 100% of the time.
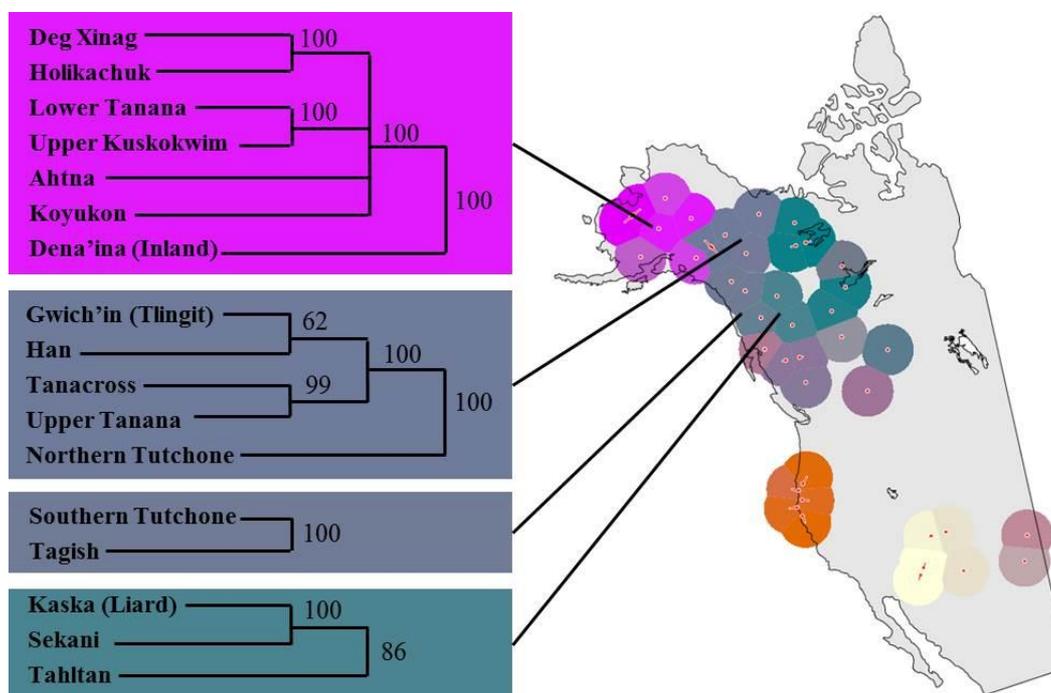
FIGURE 2. Dene languages of Alaska and northern Canada

The languages spoken in the Mackenzie drainage, sometimes identified as the Slavey languages, form a stable cluster within our model, shown in Figure 3. Though the whole cluster is very stable, the relatedness of languages within exhibits a greater degree of variation, only returning with 62% likelihood. Rice argues that these languages, here Bearlake, Mountain Slavey, Hare, Dene Dháh and South Slavey, are best treated as a dialect complex, since there is a 'chain of mutual intelligibility' (1989:9). She also does not consider Tłįchǫ (Dogrib) or Dene Sųłiné to be part of this dialect complex (in Mithun 1999), which is in line with our findings. Figure 5 shows that neither of these two languages clusters closely with any other language in the Mackenzie region, though they do show regional association being part of the larger branch in which the other Mackenzie languages are found.
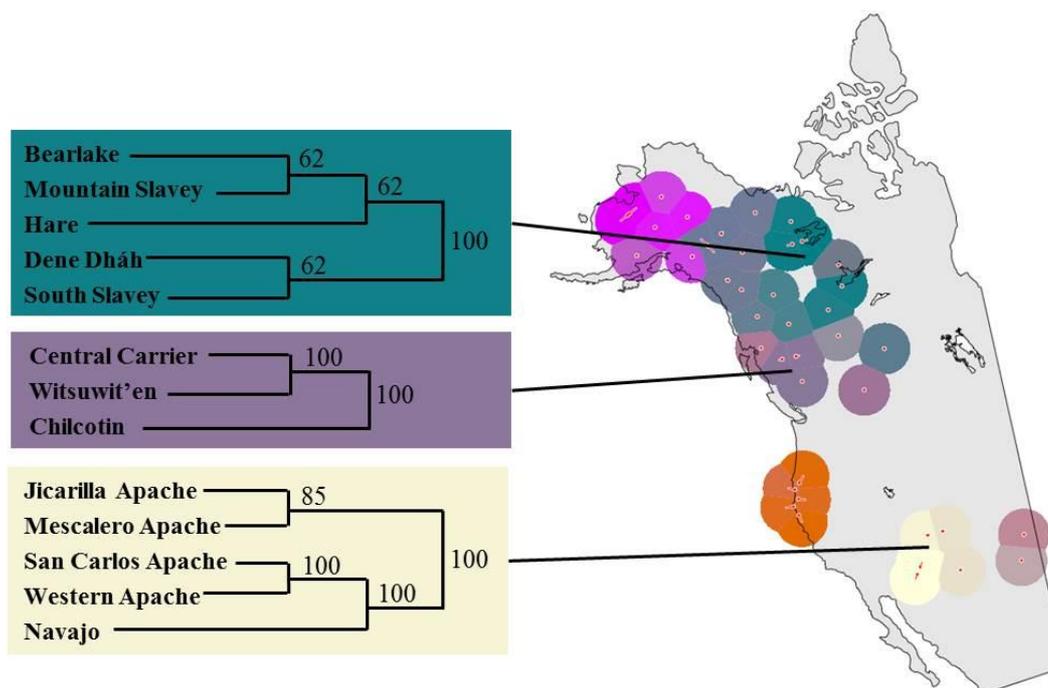
FIGURE 3. Dene Languages of Canadian interior, British Columbia, and the American Southwest

Moving into central British Columbia, Central Carrier, Witsuwit'en, and Chilcotin form a stable cluster (Figure 3), and this follows in line with geographic clustering as well. Nicola, an extinct language that was non-contiguous with the other B.C. Dene languages, is often grouped with these languages, and we are unfortunately unable to test this due to lack of data.

Of considerable interest are the clustering and closest relations of the Southern Dene, or Apachean. As mentioned in the introduction, Apachean languages are the rare exception among the Dene languages, in that they can be grouped according to the traditional historical linguistic criteria of the comparative method (Hoijer 1938), and thus the internal branching of these languages is quite well-established on independent grounds. Shown in Figure 3, our analysis also finds a two-way division among the Southern Dene languages, with Navajo, San Carlos, and Western Apache forming one branch, just as Hoijer had found. The other branch however,

groups Mescalero and Jicarilla Apache, which Hoijer had assigned to different groups. In our estimates, Lipan Apache bears no greater similarity to any of the other Apachean languages, contrary to Hoijer's conclusions, which had placed Lipan closer to Jicarilla. Above these internal sub-groupings, Jicarilla and Mescalero Apache, San Carlos and Western Apache, and Navajo cluster closely and stably as a single branch, which in turn falls under the larger division which includes all Northern Dene groups, except for the Western Alaskan cluster. Lipan Apache does cluster with the aforementioned Southern Dene languages, but falls outside of any close clusters, just like Tłįchǫ, Dene Sųłiné, and Beaver (Figure 4). These 'outlier' languages fall under a larger grouping, which also includes the languages spoken in Alaska and the Canadian interior (see Figure 5). While further analysis is needed to produce more precise branching patterns indicative of Northern and Southern relatedness, our results provide good, if tentative, evidence for the association of the Dene languages spoken in the Canadian interior with the Apachean branch.
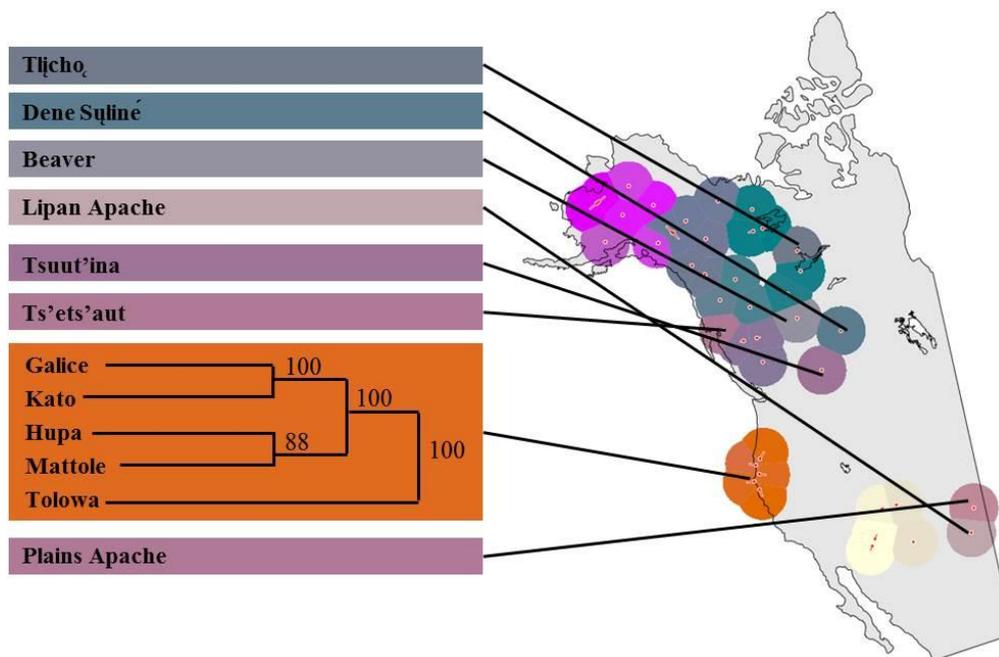

FIGURE 4. Pacific Coast languages and isolates

According to the dendrogram produced through Gabmap, Plains Apache (Kiowa Apache) lies furthest away from all Dene languages, in that sense being as closely related to all languages as other Apachean languages. Geographic groupings often put Plains Apache automatically in with the Apachean cluster, splitting Lipan and Plains Apache into an Eastern branch (Mithun 1999). Bittle argued based on morphological and phonological comparison with Apachean languages that "[Plains] Apache has long been separated from the remaining Apachean languages, and has undergone a separate development" (1956:180), which our phonological comparison concurs with.

Two languages are isolates within the larger cluster uniting all Canadian and Alaskan Dene languages: Ts'ets'aut and Tsuut'ina. Ts'ets'aut, an extinct language whose geographically closest Dene language community was the Tahltan, is most distant from all other Dene languages. Rice placed Ts'ets'aut within the Cordillera with other Northwestern Canada languages, but did not propose any close relations outside of proximity (Mithun 1999), which our results agree with. Tsuut'ina is closer to the remaining languages of this grouping than Ts'ets'aut, but remains otherwise isolated. Many theories have positioned Beaver as the closest linguistic and genetic relation to Tsuut'ina (Dempsey 2001, Goddard 1915, Hoijer 1963, Osgood 1936), though mutual intelligibility between the two is doubted (Cook 1984:2). Though Beaver and Tsuut'ina do not cluster closely together through this study, both appear as outliers, making it difficult to either support or deny hypotheses of their relationship to one another.
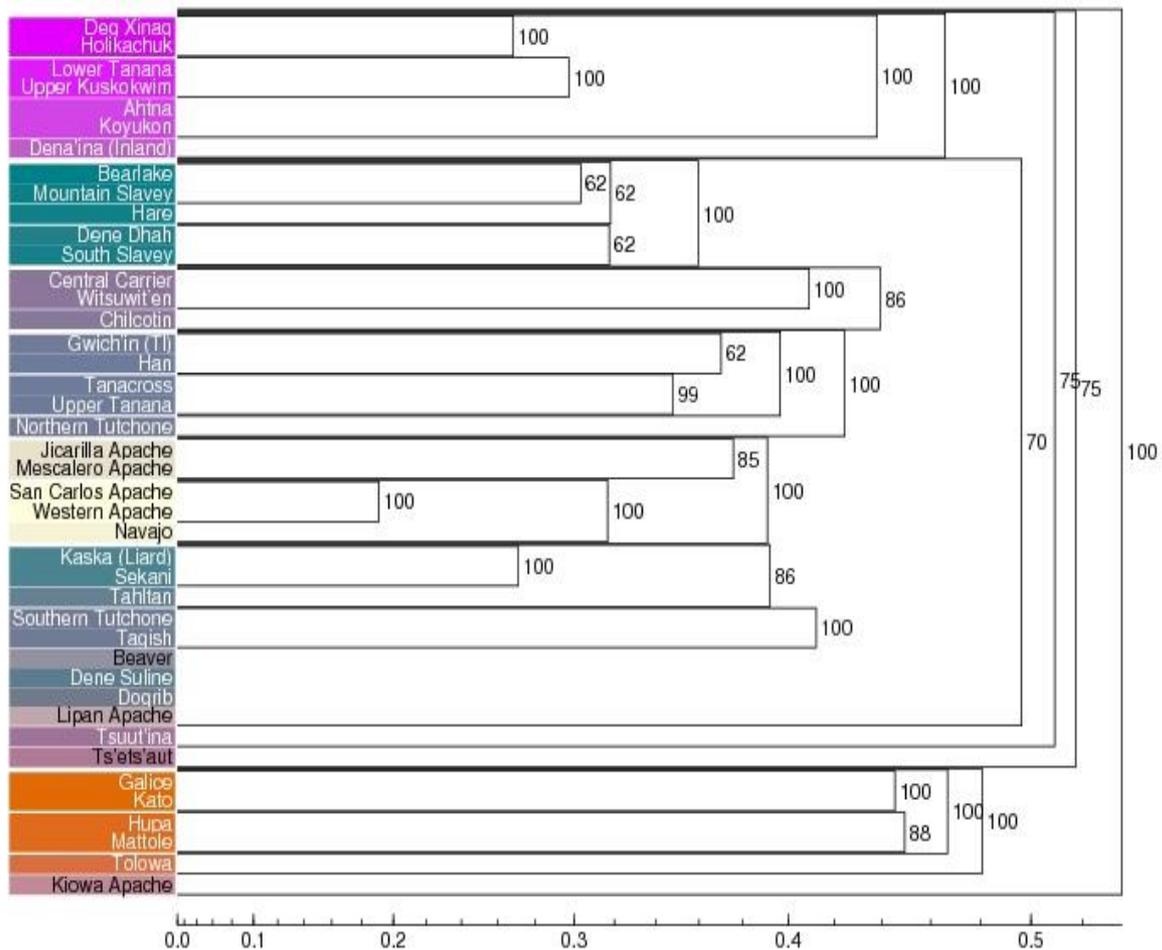
FIGURE 5. Fuzzy cluster dendrogram (the language Tłįchǫ appears as Dogrib here)

Finally, the Pacific Coast languages, Galice and Kato, Hupa and Mattole, and Tolowa, form a
stable cluster. This Pacific Coast cluster is isolated, representing a group of languages that is not
associated with any other branch of the Dene language family.

5. DISCUSSION AND CONCLUSION. The results from the application of dialectometric
measurements and methods to the question of sub-grouping among the Dene languages are
promising and show that detailed information on language relationships beyond mere regional

association can be obtained through careful sampling and analysis. The branchings and groupings our method uncovered do not represent a radical departure from the current classifications of the Dene Language family, rather, they are distinct in their details. At the present state of knowledge on the Dene language family, it is precisely these details that are missing. However, we do not consider our task completed. Instead, even more careful and detailed analysis is needed. For example, our methods assumed that the terms drawn from different semantic domains can be treated as a uniform source of data, or rather, that our methodology is powerful enough to draw general conclusions from varied data. While this might be true, we intend our future work to test this assumption by comparing clustering results across semantic domains. We are also conscious of the fact that we have included no semantic, syntactic or morphological data in our study, and understand that a fuller picture of the historical relationships among Dene languages will have to include these.

Overall, we hope to have shown the benefit of using published data and replicable methods. By opening the field of Dene historical linguistics to a wider scholarly audience both inside and outside of academic linguistics, we hope to invigorate interest in the field and inspire the kind of detailed etymological analysis which will ultimately be needed for a fuller understanding of Dene linguistic and migratory history.

REFERENCES

ALASKAN NATIVE LANGUAGE CENTER (2015). https://www.uaf.edu/anla/.

BITTLE, WILLIAM E. 1956. *The Position of Kiowa-Apache in the Apachean Group*. Los Angeles: PhD Dissertation, University of California.

COOK, EUNG-DO. 1984. *A Sarcee Grammar*. Vancouver: University of British Columbia Press.

DEMPSEY, HUGH. 2001. Sarcee. *Handbook of North American Indians: Plains*, ed. by Raymond J. Demaillie, Vol. 13, 629-637. Washington: Smithsonian Institution.

DYEN, ISIDORE AND DAVID FRIEND ABERLE. 1974. *Lexical reconstruction: The case of the Proto-Athapaskan kinship system*. Cambridge: Cambridge University Press.

ENRICO, JOHN. 2004. Toward Proto-Na-Dene. *Anthropological Linguistics 46*(3). 229–302. Indiana: American Indian Studies Research Institute.

GODDARD, IVES. 1997. The classification of the native languages of North America. *Handbook of North American Indians: Languages*, ed. by Ives Goddard, and William Sturtevant, Vol. 17, 290-324. Washington: Smithsonian Institution.

GODDARD, PLINY EARLE. 1915. Sarcee texts. *University of California Publications in American Archaeology and Ethnology* 11(3). 189-277. Berkeley: University of California Press.

GOEBL, HANS. 2005. Dialektometrie. *Quantitative Linguistik/Quantitative Linguistics*, ed. by Reinhard Köhler, Gabriel Altmann, and Rajmund G. Piotrowski, 498-531. Berlin & New York: Walter de Gruyter.

HOIJER, HARRY. 1938. The Southern Athapaskan languages. *American Anthropologist* 40(1). 75-87.

HOIJER, HARRY. 1956. The chronology of the Athapaskan languages. *International Journal of American Linguistics* 22(4). 219-232.

HOIJER, HARRY. 1963. The Athapaskan languages. *Studies in the Athapaskan Languages, University of California Publication in Linguistics* 29. 1-29. Berkeley: University of California Press.

IVES, JOHN W. AND SALLY RICE. 2008. Correspondences in archaeological, genetic and linguistic evidence for Apachean history. Unpublished manuscript.

JAIN, ANIL K. AND RICHARD D. DUBES. 1988. *Algorithms for Clustering Data*. Englewood (NJ): Prentice Hall.

JAIN, ANIL K., M. NARASIMHA MURTY AND PATRICK J. FLYNN. 1999. Data clustering: A Review. *ACM Computing Surveys* 3(31), 264-323.

KARI, JAMES AND BEN A. POTTER. (eds.). 2011. *The Dene-Yeniseian connection*. Fairbanks : University of Alaska Department of Anthropology / Alaska Native Language Center.

KRAUSS, MICHAEL E. 1964. Proto-Athapaskan-Eyak and the problem of Na-Dene: the phonology. *International Journal of American Linguistics* 30(2), 118-131. Chicago: The University of Chicago Press.

KRAUSS, MICHAEL E. 1973. Na-Dene. *Linguistics in North America*, ed. by T.A. Sebeok, 903–978. The Hague: Mouton.

KRAUSS, MICHAEL E. 2005. Athabaskan tone. *Athabaskan Prosody*, ed. by Sharon Hargus and Keren Rice, 51-136. Amsterdam: John Benjamins.

KRAUSS, MICHAEL E., AND GOLLA, VICTOR. 1981. Northern Athapaskan languages. In J. Helm (Ed.), *Handbook of North American Indians: Subarctic*, ed. by June Helm, Vol. 6, 67-85. Washington: Smithsonian Institution.

KROEBER, ALFRED L. 1925. *Handbook of the Indians of California*. Washington, D.C.: Bureau of American Ethnology Bulletin 78.

KROEBER, ALFRED L. 1959. Reflections and tests on Athabascan glottochronology. *Ethnographic Interpretations: 7-11*, 235-310. University of California Publications in American Archaeology and Ethnology, 47(3). Berkeley: University of California Press.

MITHUN, MARIANNE. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.

NERBONNE, JOHN, WILBERT HEERINGA AND PETER KLEIWEG. 1999. Edit Distance and Dialect Proximity. *Time Warps, String Edits, and Macromolecules - The Theory and Practice of Sequence Comparison*, ed. by David Sankoff and Joseph B. Kruskal. V-XV. Stanford (CA): CSLI Publications.

NERBONNE, JOHN, RINKE COLEN, CHARLOTTE GOOSKENS, PETER KLEIWEG, AND THERESE LEINONEN. 2011. Gabmap — A Web Application for Dialectology. *Dialectologia Special Issue II*.

NERBONNE, JOHN, RINKE COLEN, CHARLOTTE GOOSKENS, PETER KLEIWEG, AND THERESE LEINONEN. 2011. Gabmap. https://www.gabmap.nl/.

NERBONNE, JOHN; PETER KLEIWEG, FRANZ MANNI, WILBERT HEERINGA, CHRISTINE PREISACH, LARS SCHMIDT-THIEME, HANS BURKHARDT, AND REINHOLD DECKER. 2008. Projecting dialect distances to geography: Bootstrap clustering vs. noisy clustering. *Proceedings of the 31st Annual Meeting of the German Classification Society*. Berlin: Springer.

OSGOOD, CORNELIUS. 1936. *The Distribution of the Northern Athapaskan Indians*. Yale University Publication in Anthropology 7. London: Humphrey Milford, Oxford University Press.

RICE, KEREN. 1989. *Grammar of Slavey*. Berlin: Mouton de Gruyter.

RICE, KEREN. 2012. Linguistic Evidence regarding the Apachean migration. *From the Land of Ever Winter to the American Southwest: Athapaskan Migration, Mobility, and Ethnogenesis*, ed. by Deni J. Seymour, 249-270. Salt Lake City: University of Utah Press.

RICE, SALLY, JOHN IVES, AND CONOR SNOEK. 2013. Pan-Athapaskan comparative lexicon. Online: http://www.linguistics.ualberta.ca/Research/Projects/PanAthapaskanComparativeLexico.aspx/

SANKOFF, DAVID AND CLYDE KRUSKAL. 1983. *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. Boston: Addison-Wesley.

SAPIR, EDWARD. 1915. The Na-Dene languages, a preliminary report. *American Anthropologist*, *17*(3), 534–558. Washington D.C.: American Anthropological Association.

SNOEK, CONOR. 2013. Using semantically restricted word-lists to investigate relationships among Athapaskan languages. *Approaches to measuring linguistic differences*, ed. by Lars Borin and Anju Saxena, 231-248. Trends in Linguistics Studies and Monographs 265. Germany: de Gruyter Mouton.

WIELING, MARTIJN AND JOHN NERBONNE. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1. 243-264.

[1]There is not enough space to explain the method in detail here. Interested readers are referred to Jain and Dubes (1998) and Jin et al. (1999).