



**HAL**  
open science

## A new local indicator of spatial autocorrelation identifies clusters of high rendaku frequency in Japanese place names

Thomas Pellard, Akiko Takemura, Hyun Kyung Hwang, Timothy Vance

### ► To cite this version:

Thomas Pellard, Akiko Takemura, Hyun Kyung Hwang, Timothy Vance. A new local indicator of spatial autocorrelation identifies clusters of high rendaku frequency in Japanese place names. *Journal of Linguistic Geography*, 2023, 11 (1), pp.1-7. 10.1017/jlg.2022.11 . hal-04009523

**HAL Id: hal-04009523**

**<https://hal.science/hal-04009523>**

Submitted on 1 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# A new local indicator of spatial autocorrelation identifies clusters of high rendaku frequency in Japanese place names

Thomas Pellard CRLAO, EHESS-Inalco-CNRS  0000-0001-6189-1677 thomas.pellard@cnrs.fr  
Akiko Takemura IFRAE, Inalco-Université Paris Cité-CNRS  0000-0003-0505-0448  
Hyun Kyung Hwang Tsukuba University  
Timothy J. Vance NINJAL  0000-0002-4915-6765

**Abstract** The methods of spatial statistics have been successfully applied to the study of linguistic variation, especially for detecting the existence of spatial patterns in the geographical distribution of linguistic features. However, the use of local indicators of spatial autocorrelation for detecting spatial clusters have been limited to continuous variables, and we propose to apply the new method of Anselin & Li (2019) for categorical variables to linguistic data. We illustrate this method with the case of Japanese *rendaku*, or sequential voicing, whose dialectal variation is still poorly documented. Focusing on regional differences in the frequency of *rendaku*, we examined the occurrence of *rendaku* for four lexemes in 4 921 place names from all Japan. A statistical analysis of local spatial association and an unsupervised density-based cluster analysis revealed the existence of two cluster areas of high *rendaku* frequency centered around Wakayama and Fukushima-Yamagata prefectures. This suggests that *rendaku* is more frequent in those dialects, and we recommend that further studies in the dialectal variation of *rendaku* start by looking at those areas.

**Keywords** spatial autocorrelation, local join count, cluster detection, *rendaku*

## 1 Introduction

The traditional methods of data analysis of dialectology and dialectometry (Wieling & Nerbonne 2015; Goebel 2018) have in recent years been supplemented by new quantitative approaches of spatial statistics.<sup>1</sup> These methods are particularly useful for determining the existence of spatial patterns in the geographical distribution of linguistic features and have been successfully applied to acoustic (Grieve 2013; Grieve et al. 2013), phono-syntactic (Grieve 2011), and lexical (Grieve et al. 2011) features. In particular, measures of local spatial autocorrelation have been able to identify clusters of locations exhibiting similar values for a given linguistic feature. However, the measures used hitherto, such as Getis-Ord  $G_i$  and  $G_i^*$  (Ord & Getis 1995) or local Moran's  $I$  (Anselin 1995), only apply to continuous variables, such as the relative frequency of variants measured at different locations, and not to categorical variables, where each observation is associated with one

of several values at a location, like the presence vs. absence of a feature. Until recently, only global indicators such as the global join count (see Section 2.2) were available for such discrete variables, but Anselin & Li (2019) recently proposed a local version of that indicator. In this article, we show how this local indicator can be applied to linguistic data in order to detect local spatial clusters with the example of Japanese *rendaku*.

*Rendaku*, or sequential voicing, designates the initial consonant alternation between different realisations of a lexeme in Japanese, i.e. cases where a lexeme is usually realized with an initial voiceless obstruent, but with an initial voiced obstruent when it appears in non-initial position within a compound (1).<sup>2</sup>

- (1) a. *umi* 'sea' + *kame* 'turtle' → *umigame* 'sea turtle'
- b. *neko* 'cat' + *se* 'back' → *nekoze* 'stoop'
- c. *ami* 'net' + *to* 'door' → *amido* 'window screen'

1. One of the first applications of spatial statistics to linguistic data is Lee & Kretzschmar (1993). See Grieve (2018) for an overview.

2. The consonant  $h$  alternates with  $b$  due to a sound change  $*p > h$  in the history of Japanese.

Rendaku is a complex phenomenon, largely irregular, that is well-studied from the historical, theoretical, and psycholinguistic points of view.<sup>3</sup> However, little research has been done on the extent of regional variation in the frequency of rendaku. Most of the few existing studies conclude that there is no clear difference across dialects (Tamaoka & Ikeda 2010; Asai 2014; Irwin & Vance 2015; Takemura et al. 2019), but these suffer from limitations in data coverage and/or methodology.

In the present study, we examine the occurrence of rendaku for four lexemes in 4 921 place names from all Japan and apply a statistical analysis of local spatial association followed by a cluster analysis in order to detect the presence of local spatial clusters of high rendaku frequency, i.e. areas where either place names exhibiting rendaku tend to co-occur together more frequently than would be expected if they were randomly distributed. Our results show the existence of two cluster areas of high rendaku frequency in different regions of Japan, which suggests that rendaku is more frequent in those dialects.

## 2 Materials and methods

### 2.1 Data

In the absence of any large coverage survey on rendaku in Japanese dialects, we chose to investigate the spatial variation of rendaku frequency in place names across Japan. We used the database assembled by Takemura et al. (2019), which derives from the comprehensive database of postal codes and the corresponding addresses provided by the Japan Post (2014). It contains all place names including any of the six lexemes listed in Table 1 in non-initial position, to the exclusion of place names from Okinawa prefecture and the Amami Islands of Kagoshima since these areas are traditionally Ryukyuan-speaking areas with peculiar place names (Shimoji & Pellard 2010). Each entry consists of the postal code of a location, the corresponding place name in Japanese script, its pronunciation in *kana*, the prefecture of the location, and an added annotation by Takemura et al. (2019) as whether the place name exhibits rendaku or not. The database contains a total of 7 725 different place names (Table 1).

Since the original database does not provide the geographical coordinates of the locations, the csv address matching service of the Center for Spatial Information Science of the University of Tokyo<sup>4</sup> was used to retrieve them. The geocoding process was successful for 7 470 entries, i.e. 96.7% of the 7 725 entries, and the 255 (3.3%) remaining ones were discarded. The lexemes chosen by Takemura et al. (2019) are commonly used in place names across Japan, and all of them undergo rendaku in at least

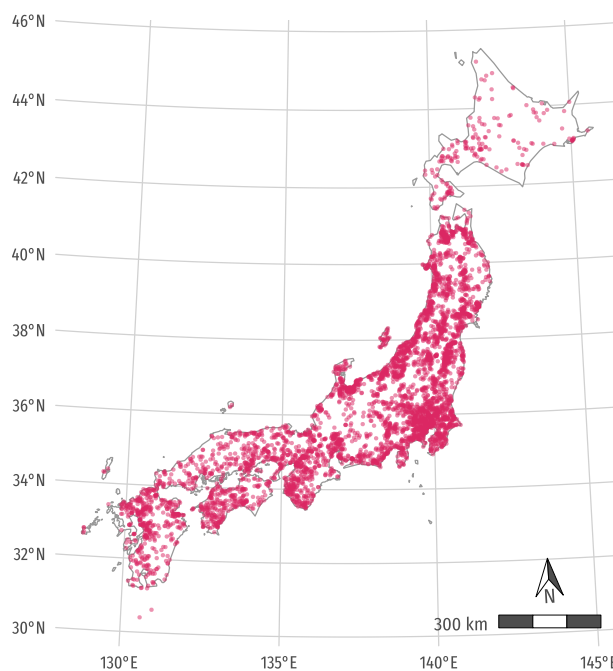


Figure 1: Map of the 4 921 locations investigated

some compounds. However, inspecting Table 1 reveals disparities in the rendaku frequency of the different lexemes ( $\chi^2(5, N = 7725) = 191.73, p < .001, V = .16$ ). Place names containing 'plain' (*hara ~ bara*) tend not to exhibit rendaku, while for other lexemes the opposite tendency is observed. On the other hand, 'valley' (*tani ~ dani*) exhibits a much higher rendaku frequency than other lexemes. This is problematic since the dependency between lexeme identity and rendaku frequency makes the data heterogeneous and the probability of rendaku not uniform. There is thus a risk that lexeme identity acts as a confounding variable, i.e. it could result in spurious clusters of either high or low rendaku frequency that would simply be spatial clusters of a certain lexeme. The two outliers 'plain' and 'valley' were thus removed from the geocoded data, resulting in an overall homogeneous data set (Table 2,  $\chi^2(3, N = 4921) = 0.17, p = .982, V = .01$ ). Figure 1 indicates the location of the place names in the final version of our database.

The spatial distribution of presence vs. absence of rendaku is difficult to interpret by simply mapping the data, partly due to the large number of data points, even if we examine separately place names with and without rendaku and use hexagonal bins (Figure 2), and even if we further map separately the different lexemes (Figure 3).<sup>5</sup> A statistical analysis is thus needed in order to assess the existence of clusters.

3. See Vance (2015) for an introduction to the topic, and Vance & Irwin (2016) for a thorough review of the issues and the relevant literature.  
4. <https://geocode.csis.u-tokyo.ac.jp/>

5. Additional geospatial data for mapping was obtained from the GADM database (v.4.1, Global Administrative Areas 2018).

Table 1: Rendaku rate by lexeme (original data of Takemura et al. 2019)

Form	Orthography	Gloss	Rendaku	No rendaku	Total
hara ~ bara	原	plain	33.9% (570)	66.1% (1112)	100.0% (1682)
kawa ~ gawa	川, 河	river	51.5% (937)	48.5% (883)	100.0% (1820)
ki ~ gi	木, 樹	tree	51.9% (482)	48.1% (447)	100.0% (929)
saki ~ zaki	崎, 埼, 寄, 先	cape	50.8% (441)	49.2% (427)	100.0% (868)
sawa ~ zawa	沢, 澤	marsh	50.3% (752)	49.7% (743)	100.0% (1495)
tani ~ dani	谷	valley	58.0% (540)	42.0% (391)	100.0% (931)
Total			48.2% (3722)	51.8% (4003)	100.0% (7725)

Table 2: Rendaku rate by lexeme (filtered and geocoded data)

Form	Orthography	Gloss	Rendaku	No rendaku	Total
kawa ~ gawa	川, 河	river	51.6% (923)	48.4% (865)	100.0% (1788)
ki ~ gi	木, 樹	tree	52.0% (473)	48.0% (436)	100.0% (909)
saki ~ zaki	崎, 埼, 寄, 先	cape	51.1% (435)	48.9% (417)	100.0% (852)
sawa ~ zawa	沢, 澤	marsh	51.7% (709)	48.3% (663)	100.0% (1372)
Total			51.6% (2540)	48.4% (2381)	100.0% (4921)

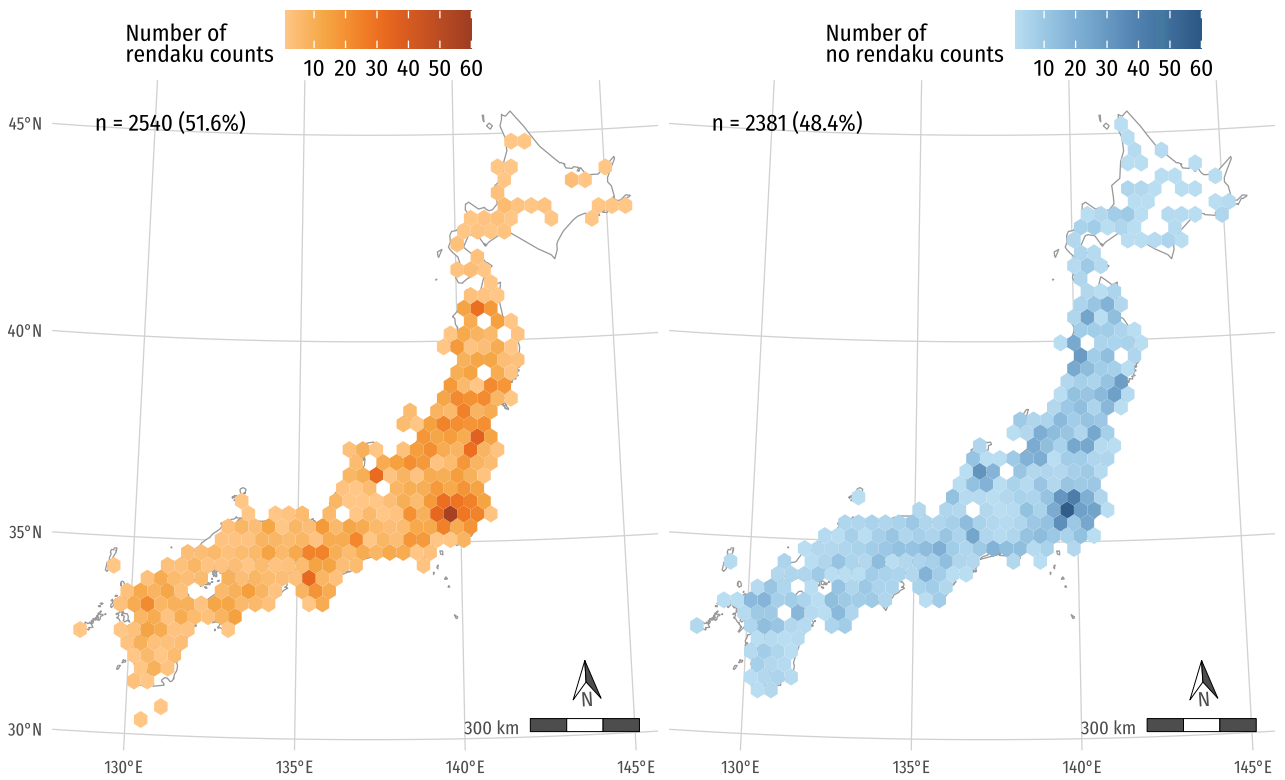


Figure 2: Presence vs. absence of rendaku in all place names



## 2.2 Statistical analysis

In order to investigate whether the spatial distribution of rendaku is random or follows a pattern, we need to measure the spatial autocorrelation for the variable of interest, i.e. the degree to which observations of presence vs. absence of rendaku agree between neighboring place names. First, we defined a location's neighbors as the set of all locations within a radius of 50 km of that location (Figure 4), which results in a circle with an area roughly equal to the mean area of Japanese prefectures without the Ryukyu Islands (7 982 km<sup>2</sup>).

For exploratory purposes, a global join count (Moran 1948) was computed, with three different measures:  $J_{BB}$ , the number of neighboring locations that exhibit rendaku ("black-black"),  $J_{WW}$ , the number of neighboring locations that do not exhibit rendaku ("white-white"), and  $J_{BW}$ , the number of neighboring locations that exhibit opposite values ("black-white").<sup>6</sup> The results (Table 3) indicate the existence of a statistically significant ( $z = 2.02$ ,  $p = 0.02$ , one-sided test) positive spatial autocorrelation for  $J_{BB}$ , i.e. place names exhibiting rendaku tend to co-occur more often than would be expected by chance alone.

However, though the global join count indicates the existence of rendaku clusters, it does not specify where these clusters are located. In order to locate these clusters, the local indicator of spatial association for binary marked points proposed by Anselin & Li (2019) was calculated for each location with a place name exhibiting rendaku as the number of its neighbors that also exhibit rendaku.<sup>7</sup> Then, a conditional permutation test with 999 repetitions was performed in order to obtain for each location the pseudo  $p$ -value of getting by chance an indicator at least as high as observed. Since many tests were performed, the false discovery rate of multiple hypothesis testing was controlled by applying the procedure of Benjamini & Hochberg (1995), in accordance with de Castro & Singer (2006). Locations with an adjusted  $p$ -value below the .05  $\alpha$ -level were considered to constitute cores of rendaku clusters, and the neighbors of those cores that exhibit rendaku were also considered to be part of the clusters since even though their indicator is not significant by itself, it contributes to make those of the cores significant.

Such cores and their neighbors exhibiting rendaku were then submitted to an unsupervised density-based cluster

6. Formally, with  $x_i$  and  $x_j$  the value of the variable  $x$  of presence ( $x = 1$ ) vs. absence ( $x = 0$ ) of rendaku at respectively locations  $i$  and  $j$  ( $i \neq j$ ),  $J_{BB} = \frac{1}{2} \sum_i \sum_j w_{ij} x_i x_j$ ,  $J_{WW} = \frac{1}{2} \sum_i \sum_j w_{ij} (1 - x_i)(1 - x_j)$ , and  $J_{BW} = \frac{1}{2} \sum_i \sum_j w_{ij} (x_i - x_j)^2$ . The value of  $w_{ij}$  is specified by a symmetrical binary spatial weights matrix where  $w_{ij} = 1$  if locations  $i$  and  $j$  are neighbors, and  $w_{ij} = 0$  otherwise.

7. Formally, the indicator is defined in way similar to the global join count as  $BB_i = x_i \sum_j w_{ij} x_j$ . The indicator  $BB_i$  boils down to 0 if there is no rendaku ( $x_i = 0$ ), and the contribution of other locations is zero when they are not a neighbor of  $i$  ( $w_{ij} = 0$ ) or when they do not themselves exhibit rendaku ( $x_j = 0$ ). Each neighbor exhibiting rendaku has a contribution of  $w_{i,j} x_j = 1$ .

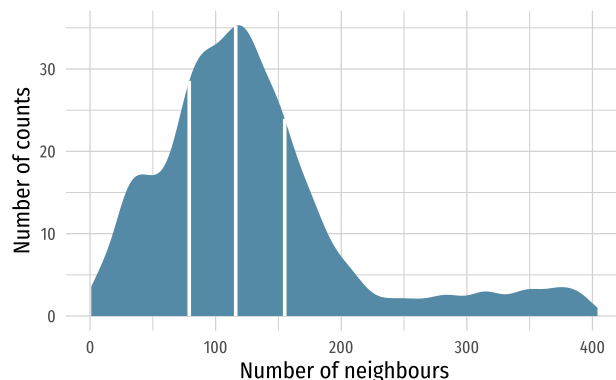


Figure 4: Density distribution of the number per location of neighbors within a 50 km distance; vertical white lines indicate the quartiles ( $Q_1 = 79$ ,  $Q_2 = 116$ ,  $Q_3 = 155$ ,  $\bar{x} = 128.8$ )

Table 3: Global join count statistics

Join	Observed	Expected	Variance
$J_{BB}$	87336 (27.6%)	84413.35 (26.6%)	2090813
$J_{WW}$	73287 (23.1%)	74173.9 (23.4%)	1839435
$J_{BW}$	156285 (49.3%)	158320.75 (50%)	83569

analysis with the DBSCAN algorithm (Ester et al. 1996) with parameter  $\epsilon = 50$  km and  $MinPts = 30$ .

All analyses were performed using the R Statistical Software (v4.2.1, R Core Team 2022) and the spdep (v1.2.5, Bivand et al. 2013) and dbscan (v1.1.10, Hahsler et al. 2019) packages. Geospatial data was processed using the R package sf (v1.0.8, Pebesma 2018).

## 3 Results

The statistical analysis of local spatial association indicates the existence of 167 cores of clusters, which represent 3.4% of all locations and 6.6% of locations exhibiting rendaku. These are surrounded by a total of 376 neighbors, which themselves represent 7.6% of all locations and 14.8% of locations exhibiting rendaku. Conversely, 78.6% of place names exhibiting rendaku do not belong to a cluster. Figure 5 shows the location of the cores of clusters.

These cores and neighbors form two different geographical cluster areas, as revealed by both the visual inspection of the data (Figure 6) and the application of the clustering algorithm. Table 4 and Table 5 summarize the number of cores and neighbors by cluster area and respectively prefecture and lexeme.



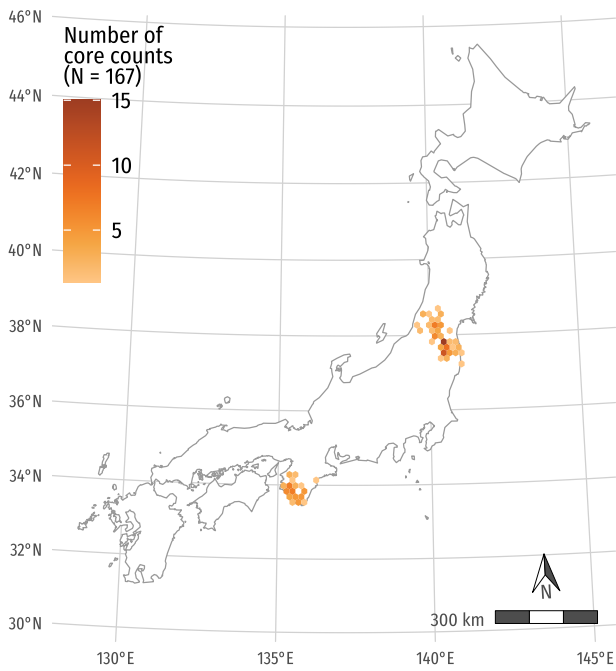


Figure 5: Cores of rendaku clusters

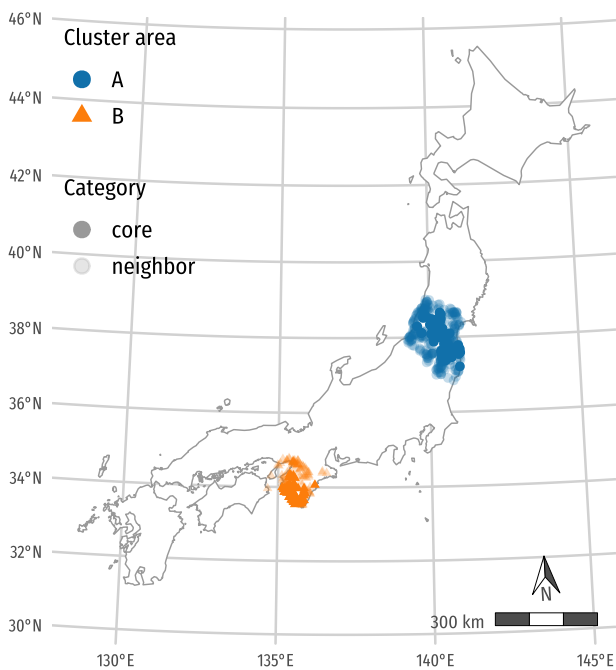


Figure 6: Cores (full colours) of rendaku clusters and their neighbors (dimmed colours)

Table 4: Number of cores and neighbors by area and prefecture

Area	Prefecture	Cores	Neighbors	Total
A	Fukushima	60	89	149
A	Yamagata	41	99	140
A	Miyagi	4	46	50
A	Niigata	1	35	36
A	Ibaraki	0	2	2
A	Tochigi	0	1	1
A	Total	106	272	378
B	Wakayama	55	31	86
B	Nara	3	23	26
B	Mie	2	10	12
B	Ōsaka	1	29	30
B	Hyōgo	0	10	10
B	Tokushima	0	1	1
B	Total	61	104	165
Total		167	376	543

Table 5: Number of cores and neighbors by area and lexeme

Area	Lexeme	Cores	Neighbors	Total
A	sawa ~ zawa	59	125	184
A	saki ~ zaki	19	29	48
A	kawa ~ gawa	14	92	106
A	ki ~ gi	14	26	40
A	Total	106	272	378
B	kawa ~ gawa	49	62	111
B	ki ~ gi	9	23	32
B	saki ~ zaki	3	16	19
B	sawa ~ zawa	0	3	3
B	Total	61	104	165
Total		167	376	543

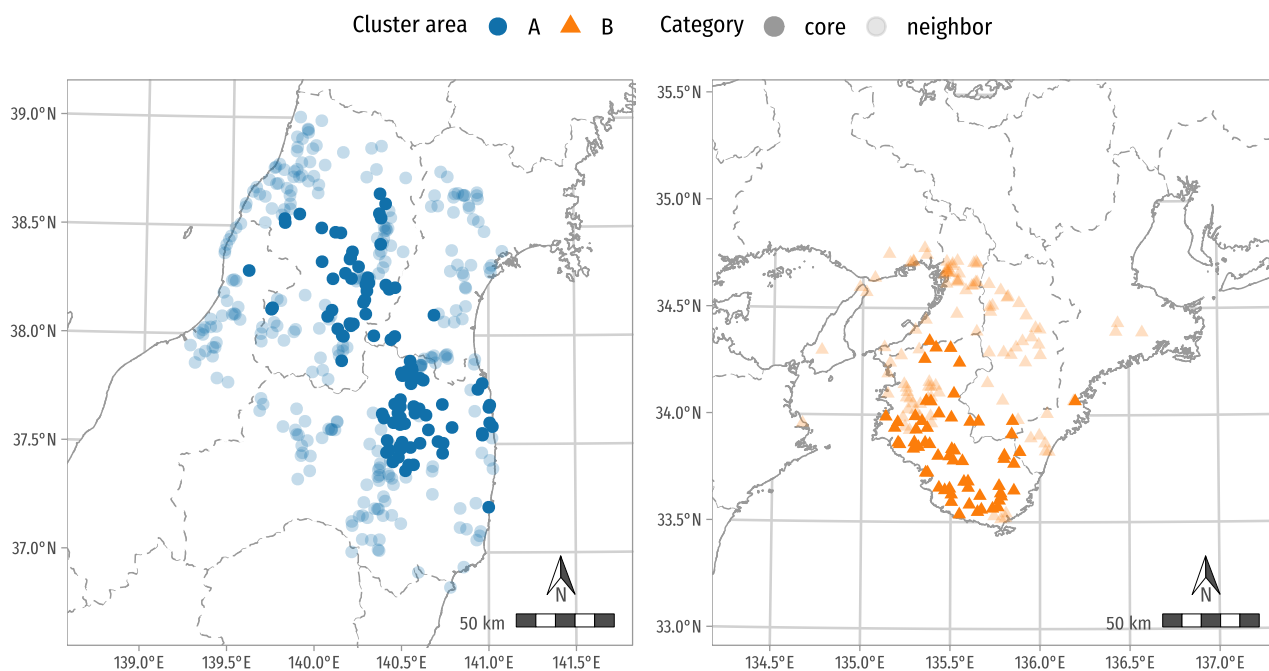


Figure 7: Cores (full colours) of individual rendaku cluster areas and their neighbors (dimmed colours)

## 4 Discussion

We observed that place names exhibiting rendaku are not distributed completely at random across Japan but tend to co-occur in some areas, and we detected two different geographical rendaku cluster areas. This contrasts with the findings of [Takemura et al. \(2019\)](#) that there is no clear geographical pattern. The major problem with that preliminary study is that it ignored the problems of spatial autocorrelation and calculated the mean rendaku rate by prefecture, though there is little justification for believing that dialectal zones and their differences in rendaku rate strictly follow the boundaries of modern administrative units. This probably had the unfortunate consequence of masking the existence of rendaku clusters distributed over several prefectures, as revealed in our more fine-grained study. It also suffers from a lack of homogeneity in the rendaku rate of the different lexemes studied.

[Tamaoka & Ikeda \(2010\)](#) conclude that dialect has no influence on rendaku frequency. Their study is based on an experiment with 405 speakers from six different localities (Kagoshima, Ōita, Fukuoka, Yamaguchi, Hiroshima, and Shizuoka). However, their conclusion is expected from our perspective since none of these prefectures belongs to a rendaku cluster area.

Similarly, [Irwin & Vance \(2015\)](#) record the absence of presence of rendaku in 32 examples of 31 lexemes in the speech of 67 speakers of dialects from four different regions (Yamagata, Ehime, Miyazaki, and Hyōgo) but do not find a significant difference in average rendaku rate across locations. From our perspective, no such difference

is indeed expected between Ehime, Miyazaki, and Hyōgo: Ehime and Miyazaki do not belong to a rendaku cluster area, and Hyōgo only contains a few neighbors. On the other hand, Yamagata contains several cores of rendaku clusters, and our results are thus at variance with theirs.

[Asai \(2014\)](#) finds rendaku to be slightly more frequent in Northeastern Japanese dialect recordings (58%) than in Standard Japanese (50%).<sup>8</sup> His sample of Northeastern Japanese consists in dialects from Aomori, Akita, Iwate, Miyagi, Yamagata, and Fukushima, which substantially overlaps with our rendaku cluster area A. Such results are thus compatible with ours, and at variance with those of [Irwin & Vance \(2015\)](#).

The existence of a large rendaku cluster area in North-east Japan is interesting since it is known that in dialects of that region, the voiceless vs. voiced distinction is realized differently than in Standard Japanese, i.e. in intervocalic position voiceless stops are realized as voiced and voiced ones as prenasalized ([Miyashita et al. 2016](#)). It is however unlikely that this cluster area is the result of place names without rendaku but intervocalic voicing erroneously recorded as exhibiting rendaku. Though the consonant *s* does not undergo intervocalic voicing, half of the lexemes examined have an initial *s-*, and as much as 13.3% of the place names in cluster area A involve a lexeme with an initial *s-*.

8. [Asai \(2014\)](#) did not perform a significance test, and he does not present all the relevant figures.



## 5 Conclusions

This article shows the benefits of fine-grained analyses of spatial statistics for the study of linguistic variation and how a new measure of local spatial autocorrelation can be applied to linguistic categorical variables for detecting and localizing clusters. We were able to show that the frequency of *rendaku* varies across Japanese regions, and we detected the existence of two cluster areas of *rendaku* that could not be detected by simply looking at the data or by averaging frequency by prefecture, and that had not been identified by previous studies.

Our study used place names to investigate the spatial variation of *rendaku* frequency. This raises the problem of the reliability of our data as a representative sample of dialects. Place names are often ancient and can predate the formation of the modern dialects, and the origin of their current pronunciation is unknown. Moreover, there is no assurance that the recorded place names readings that we used are always the same as the pronunciations in actual use locally.

Nevertheless, our statistical procedure detected a spatial pattern that cannot be reasonably attributed to chance only, and which needs to be explained in any case. Moreover, we think of place names a proxy for a pandialectal study covering all Mainland Japan, which is unfeasible by the means of traditional dialectology. We recommend that the spatial cluster areas of *rendaku* that we detected be subjected to conventional dialectological surveys in order to try to replicate our results with experimental data. In other words, we suggest that further dialectal studies of *rendaku* start by looking at the areas of Fukushima-Yamagata and Wakayama.

## Data availability statement

The data and code used for this study are openly available in an Open Science Framework repository at <http://doi.org/10.17605/OSF.IO/MA8Q7>.

## References

- Anselin, Luc. 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27(2). 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- Anselin, Luc & Li, Xun. 2019. Operational local join count statistics for cluster detection. *Journal of Geographical Systems* 21(2). 189–210. <https://doi.org/10.1007/s10109-019-00299-x>.
- Asai, Atsushi. 2014. *Rendaku seiki no keikō to teichakuka*. *NINJAL Research Papers* 7. 27–44. <https://doi.org/10.15084/00000523>.
- Benjamini, Yoav & Hochberg, Yosef. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1). 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bivand, Roger S. & Pebesma, Edzer & Gómez-Rubio, Virgilio. 2013. *Applied spatial data analysis with R*. 2nd edn. New York: Springer. <https://doi.org/10.1007/978-1-4614-7618-4>.
- de Castro, Marcia Caldas & Singer, Burton H. 2006. Controlling the false discovery rate: A new application to account for multiple and dependent tests in local statistics of spatial association. *Geographical Analysis* 38(2). 180–208. <https://doi.org/10.1111/j.0016-7363.2006.00682.x>.
- Ester, Martin & Kriegel, Hans-Peter & Sander, Jörg & Xu, Xiaowei. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In Simoudis, Evangelos & Han, Jiawei & Fayyad, Usama M. (eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231. Menlo Park: AAAI Press. <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>.
- Global Administrative Areas. 2018. *GADM database of global administrative Areas*. Version 3.6. <http://www.gadm.org>.
- Goebel, Hans. 2018. Dialectometry. In Boberg, Charles & Nerbonne, John & Watt, Dominic (eds.), *The handbook of dialectology*, 123–142. Hoboken: Wiley. <https://doi.org/10.1002/9781118827628.ch7>.
- Grieve, Jack. 2011. A regional analysis of contraction rate in written Standard American English. *International Journal of Corpus Linguistics* 16(4). 514–546. <https://doi.org/10.1075/ijcl.16.4.04gri>.
- Grieve, Jack. 2013. A statistical comparison of regional phonetic and lexical variation in American English. *Literary and Linguistic Computing* 28(1). 82–107. <https://doi.org/10.1093/lc/fqs051>.
- Grieve, Jack. 2018. Spatial statistics for dialectology. In Boberg, Charles & Nerbonne, John & Watt, Dominic (eds.), *The handbook of dialectology*, 415–433. Hoboken: Wiley. <https://doi.org/10.1002/9781118827628.ch24>.
- Grieve, Jack & Speelman, Dirk & Geeraerts, Dirk. 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change* 23(2). 193–221. <https://doi.org/10.1017/S095439451100007X>.
- Grieve, Jack & Speelman, Dirk & Geeraerts, Dirk. 2013. A multivariate spatial analysis of vowel formants in American English. *Journal of Linguistic Geography* 1(1). 31–51. <https://doi.org/10.1017/jlg.2013.3>.
- Hahsler, Michael & Piekenbrock, Matthew & Doran, Derek. 2019. *dbSCAN: Fast density-based clustering with R*. *Journal of Statistical Software* 91(1). 1–30. <https://doi.org/10.18637/jss.v091.i01>.
- Irwin, Mark & Vance, Timothy J. 2015. *Rendaku across Japanese dialects*. *Phonological Studies* 18. 19–26. <https://doi.org/10.1017/S0014180100000000>.

- [// www . researchgate . net / publication / 332141508 \\_ Rendaku\\_Across\\_Japanese\\_Dialects](http://www.researchgate.net/publication/332141508_Rendaku_Across_Japanese_Dialects).
- Japan Post. 2014. *Yūbin bangō dēta*. <https://www.post.japanpost.jp/zipcode/download.html>.
- Lee, Jay & Kretzschmar, William A. 1993. Spatial analysis of linguistic data with GIS functions. *International Journal of Geographical Information Systems* 7(6). 541–560. <https://doi.org/10.1080/02693799308901981>.
- Miyashita, Mizuki & Irwin, Mark & Wilson, Ian & Vance, Timothy J. 2016. Rendaku in Tōhoku Japanese: The Kahoku-chō survey. In Vance, Timothy J. & Irwin, Mark (eds.), *Sequential voicing in Japanese: Papers from the NINJAL Rendaku Project*, 173–194. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.176.10miy>.
- Moran, Patrick A. P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B (Methodological)* 10(2). 243–251. <http://www.jstor.org/stable/2983777>.
- Ord, J. K. & Getis, Arthur. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27(4). 286–306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>.
- Pebesma, Edzer. 2018. Simple features for R: Standardized support for spatial vector data. *The R Journal* 10(1). 439. <https://doi.org/10.32614/rj-2018-009>.
- R Core Team. 2022. *R: A language and environment for statistical computing*. Computer software. <https://www.R-project.org>.
- Shimoji, Michinori & Pellard, Thomas (eds.). 2010. *An introduction to Ryukyuan languages*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa.
- Takemura, Akiko & Pellard, Thomas & Hwang, Hyun Kyung & Vance, Timothy J. 2019. Rendaku in place names across Japanese dialects. *Reports of the Keio Institute of Cultural and Linguistic Studies* 50. 79–89. hal: hal-01991521.
- Tamaoka, Katsuo & Ikeda, Fumiko. 2010. Whiskey or bhiskey? Influence of first element and dialect region on sequential voicing of *shoochuu*. *Gengo Kenkyū* 37. 65–79. [https://doi.org/10.11435/gengo.137.0\\_65](https://doi.org/10.11435/gengo.137.0_65).
- Vance, Timothy J. 2015. Rendaku. In Kubozono, Haruo (ed.), *Handbook of Japanese phonetics and phonology*, 397–441. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9781614511984.397>.
- Vance, Timothy J. & Irwin, Mark (eds.). 2016. *Sequential voicing in Japanese: Papers from the NINJAL Rendaku Project*. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.176>.
- Wieling, Martijn & Nerbonne, John. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1(1). 243–264. <https://doi.org/10.1146/annurev-linguist-030514-124930>.