

Corpus-based Analysis of Education Policy Discourse in Municipal Boards of Education in Kanagawa Prefecture

Mario DePavia

Fulbright Research Fellow

Waseda Graduate School of Applied Japanese Linguistics

mariodp@stanford.edu

Abstract

Corpus-based research is a well-established methodology informing language acquisition and language education research. Increasingly, corpora have been used meta-pedagogically to analyze education policy documents (Muldergig 2024) and discussions of language education policy in national assemblies, including Japan’s own Diet (Yamamoto 2011). However, discourse of municipal boards of education (MBoE)—the administrative assemblies closest to education policy implementation—remains an unexplored domain. Construction of a corpus of MBoE meeting minutes enables exploration of public education discussion and implementation across municipal borders.

In addition to the construction of a corpus of all MBoE minutes available on municipal websites across Kanagawa, this research demonstrates its utility by exploring several questions for which it can provide insight. Specifically, this research analyses policy discussions of Japanese Second Language Education (JSL). It maps the distribution of JSL discussions temporally and geographically. It further compares terminological variants that are commonly used conjunctionally to refer to JSL and JSL students (日本語指導, 日本語教育, 外国人児童, 海外にルーツを持つ子どもたち, etc.) and explores correlations between municipal population of foreign residents and frequency of these discussions.

Through exploration of policy dialogue, this research demonstrates how corpora can allow researchers to gauge how any given policy topic is prioritized in routine discussion of municipal boards or local governments.

1 Introduction

Article 21 of the Act on the Organization and Operation of Local Educational Administration authorizes MBoEs to manage eighteen enumerated categories of educational affairs followed by a nineteenth catch-all: “other educational affairs within the local public entity’s jurisdiction” 地方教育行政の組織及び運営に関する法律（昭和三十一年法律第百六十二号）（2026）. Each MBoE independently makes a set of decisions to answer the same set of administrative ambiguities. One can expect a diversity in MBoE policy given that each MBoE navigates unique local conditions such as geography, budget, and demographics. The language used in deliberations by which they arrive at decisions can thus be used to examine factors impacting local politics across Japan. However, the sheer scale and lack of standard measures, manually tracking how each MBoE changes these policies over time on a comprehensive scale is infeasible.

The field of quantitative linguistics provides a scalable answer to this challenge: by quantifying the frequency and spatial distribution of specific terminology within a corpus of official meeting minutes, one can model, evaluate and compare local educational priorities at scale. Recent work has established the viability of this paradigm for Japanese sub-national politics —Takamaru et al. (2026) construct a longitudinal corpus of all 47 prefectural assemblies which can be used to analyze discourse patterns across electoral terms. MBoEs operate at a lower level of localization, so analysis of their discourse could provide even higher-granularity insight to forces that drive and motivate local administrative priorities. While this is the case for any municipal assembly, the ability to glean how local interests impact the implementation of a supposedly equitable public education system further motivates the discourse to be sourced from boards of education.

The following paper outlines the construction of such a corpus from all MBoE meeting minutes that were accessible on March 23, 2026 in online archives from the 33 municipalities in Kanagawa prefecture. It details the distribution of documents and diversity in scope of documents available across municipalities as well as the total size and scale of the corpus. Finally, as a demonstration of the corpus’s utility, it presents findings from a case examination of discourse related to Japanese Second Language education (JSL) across Kanagawa.

2 Methodology

To minimize confounding variables caused by policies set by prefectural boards of education and in response to the limited amount of resources allocatable to this research, it was determined that the scope of the corpus should be restricted to a single prefecture. Given its wide range of population density, demographic diversity, and number of government-ordinance cities (政令指定都市), Kanagawa was selected in the expectation that these contrasts would support meaningful comparison of how differently-situated municipalities deliberate the same policy questions.

For all but one municipality, the meeting minutes were uploaded in PDF format. For these municipalities, the architectures of MBoE meeting minutes archive web pages were mapped in advance to allow for the collection of all documents on the same date (March 23, 2026). This also prevented other incidental documents, such as meeting presentation materials uploaded to BoE archives, from being erroneously included in the corpus. The text of these files was then extracted using the Python package PyMuPDF (Artifex Software, Inc. 2026). In instances where this package produced CID errors (e.g. cid:1234), an empty string, or an unreasonably low character count ($n < 50$), the native extraction was re-run using the package pdfplumber (Singer-Vine and The pdfplumber contributors 2025). If the same problems persisted, documents were then uploaded to Google Drive where Optical Character Recognition (OCR) was employed Google’s server-side OCR engine. As a final sanity check, the length and Japanese character content of the outputted text from OCR extraction was compared with the native PyMuPDF extraction. The favorable extraction was included in the final corpus.

Atsugi was the only municipality to directly upload meeting minutes as text on webpages. These were extracted by converting HTML to text using the BeautifulSoup Python library. These meetings were then parsed using regular expressions matching patterns occurring at the beginning and end of Atsugi’s MBoE meetings (e.g. 開会時刻, 閉会時刻) to prevent leakage of HTML code into the corpus.

Originally, dates of each meeting were extracted from the archival pages from which they originated using regular expressions. However, it was discovered that these dates did not always accurately reflect the dates listed on the documents. Thus, the dates were extracted directly from the document texts using regular expressions. As a sanity check against spuriously-matched dates mentioned within the meeting minutes, the meeting frequency for each municipality was calculated according to the extracted dates, and anomalies in frequency were manually investigated. The text files were then given HTML-style heading tags demarcating the meeting date. All meetings from a single municipality were then concatenated using Python and the resulting file was given a heading tag indicating the municipality from which they originated. Finally, each municipal file was concatenated to form the final corpus. This corpus was then tokenized using MeCab (Kudo 2005) with IPADIC through the text mining software KH Coder for further analysis. This included the forced extraction of JSL terminology to prevent premature segmentation. Supplementary data analysis and graphical generation was performed using Python.

Originally, the JSL terminology examined was based on variations encountered in the literature, including: ”with foreign ties/roots” (外国につなが(り)・る) / 外国にルーツ), ”with ties/roots overseas” (海外につなが(り)・る) / 海外にルーツ), and ”Japanese language instruction/education” (日本語指導 / 日本語教育). Following a preliminary examination of the data, two frequently observed terms—”foreign pupils” (外国人児童) and ”Japanese language support” (日本語支援)—were added to the scope. One false positive was excluded (「オンラインで海外につながる」).

3 Corpus Specifications

In total, 5,333 meetings were collected. The average archive documented 11.5 years of meetings, with the longest spanning 21.2 years and the shortest spanning 0.9 years.¹ On average, municipalities held 13.5 meetings per year, with the most frequent holding an average of 23.0 meetings annually and the sparsest holding an average of 8.3 meetings per year (see Table 1 for municipal breakdown). The average length of a document was 16302 characters and did not follow a normal distribution (see Figure 2). MeCab segmentation yielded 52.8M tokens (19.4M after KH Coder’s part-of-speech filtering to content words) across 70,574 types.

¹This short timespan was due to a municipal policy in Ayase that limits the public online archive to only the past year’s meeting minutes.

Table 1: Overview of Kanagawa Prefecture Municipal Board of Education Meeting Minutes Corpus

Municipality	Cnt	Start	End	Yrs	Cnt/Yr
Yokohama	386	2009	2026	16.8	23.0
Kawasaki	288	2011	2025	14.6	19.8
Sagamihara	289	2009	2026	16.7	17.3
Yokosuka	143	2015	2026	11.0	13.0
Hiratsuka	72	2020	2026	5.8	12.5
Kamakura	251	2008	2025	17.4	14.4
Fujisawa	134	2015	2026	10.7	12.5
Odawara	326	2004	2025	21.2	15.4
Chigasaki	24	2024	2026	1.8	13.0
Zushi	258	2005	2025	20.6	12.5
Miura	141	2015	2026	11.0	12.8
Hadano	237	2008	2025	17.5	13.5
Atsugi	203	2010	2025	15.9	12.7
Yamato	264	2006	2025	20.0	13.2
Isehara	120	2016	2026	9.7	12.3
Ebina	194	2013	2025	12.5	15.5
Zama	68	2021	2026	5.0	13.6

Municipality	Cnt	Start	End	Yrs	Cnt/Yr
M. Ashigara	220	2010	2026	15.7	14.0
Ayase	12	2024	2025	0.9	12.9
Hayama	75	2020	2025	5.9	12.7
Samukawa	66	2020	2025	5.3	12.4
Oiso	293	2005	2026	20.7	14.1
Ninomiya	200	2010	2025	15.6	12.8
Oi	67	2021	2026	4.8	14.0
Matsuda	89	2016	2026	9.3	9.5
Yamakita	63	2021	2026	5.0	12.6
Kaisei	89	2018	2025	7.2	12.4
Hakone	57	2019	2026	6.8	8.3
Manazuru	133	2015	2026	10.7	12.4
Yugawara	176	2013	2026	12.8	13.8
Aikawa	236	2007	2025	18.9	12.5
Kiyokawa	95	2018	2026	7.8	12.1
Nakai	64	2021	2026	4.8	13.4
Total/Avg	5,333	—	—	11.5	13.5

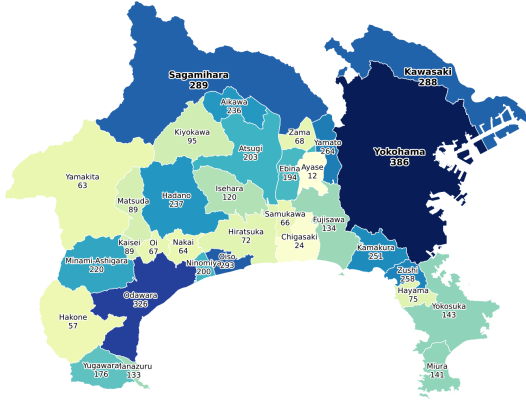
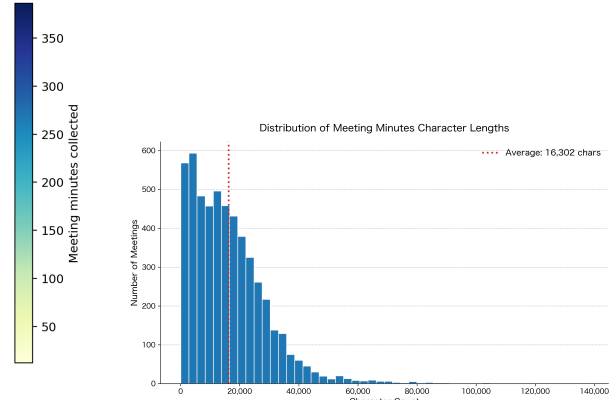
Corpus coverage: meeting minutes per municipality
Kanagawa Prefecture (33 municipalities, 5,333 meetings)

Figure 1: Geographic distribution of corpus coverage: meeting minutes collected per municipality across Kanagawa Prefecture (33 municipalities, 5,333 meetings).

Figure 2: Distribution of meeting minutes character lengths across the corpus ($n = 5,333$). The dotted line marks the corpus average of 16,302 characters.

4 Privacy considerations and corpus release design

Release and distribution of the corpus conducted in this research is subject to consideration of privacy and redistribution rights of MBoE documents released to the public. As public entities, MBoEs have already performed a considerable amount of redaction of sensitive information present in these documents. The corpus is frequented with omission denoted with notes such as "Topics withheld" (「非公開」, $n=9085$). Further, the corpus was composed entirely of publicly available documents and thus does not pose an inherent risk of divulging new information. However, MBoEs may wish to retroactively censor a meeting record retroactively deemed too sensitive to make available to the public. Immortalizing and proliferating a version of MBoE minute archives as they existed in March 2026 undermines MBoEs to perform this retroactive omission. While the proper release model to address these concerns is considered, the corpus remains private.

5 Case Examination of Japanese Second Language

JSL is one such policy domain in which a large degree of freedom is given to municipal MBoEs, making it a possible topic for a study examined using such a corpus. Diversity in nomenclature used in discourse of JSL highlights the lack of standardization in policy: MEXT frequently employs the terms "Japanese language instruction" (日本語指導) and "children and students requiring Japanese language instruction" (日本語指

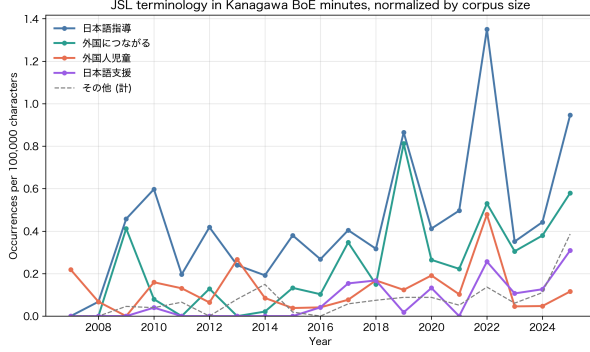


Figure 3: Annual frequency of JSL terminology in Kanagawa MBoE minutes, normalized per 100,000 characters of corpus text.

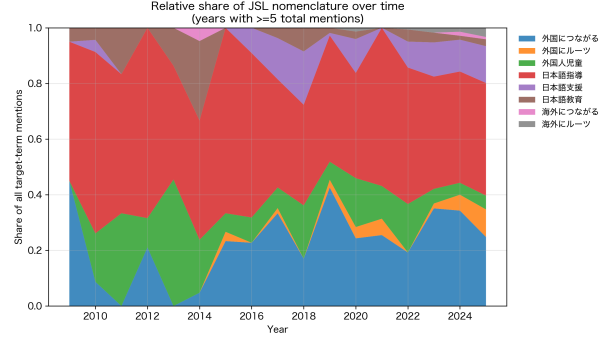


Figure 4: Relative distribution of JSL terminology over time. To minimize statistical noise, years with fewer than five total mentions across all variants were excluded.

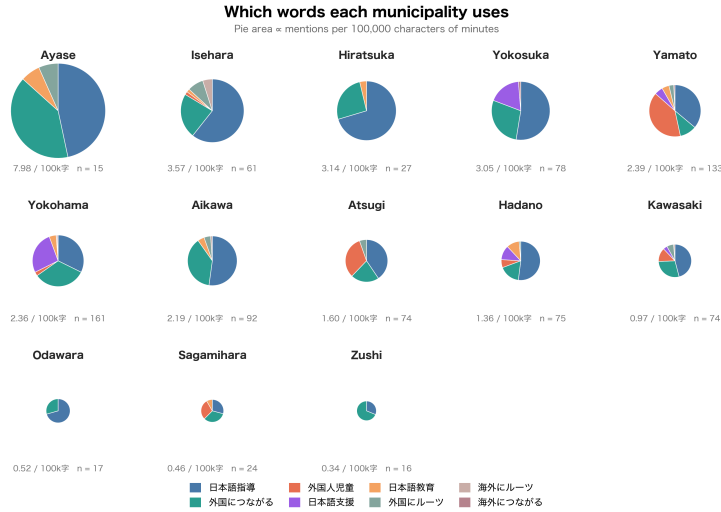


Figure 5: Terminological preferences by municipality. Pie area is proportional to mentions per 100,000 characters of minutes; municipalities are ordered by normalized mention rate.

導が必要な児童生徒) (Ministry of Education, Culture, Sports, Science and Technology 2026), while other actors such as public interest groups and academics use phrases including “Children with foreign ties/外国につながる子供たち” (YOKE 2026), “Children with foreign roots/外国にルーツを持つ子どもたち” (Mori 2018), or “Children with roots overseas/海外にルーツを持つ子供たち” to refer to these students and “academic support/学習支援” or “education/教育” to refer to their pedagogical support. Further, this variation in nomenclature lends itself to a quantitative linguistic study as the differences in terminology raise questions over what factors impact favored verbiage: how does the usage of each term change over time? Do municipalities demonstrate high bias toward particular nomenclature, suggesting localized policy cultures? Does verbiage correlate with context sentiment? Such terminological competition also invites a prototype-theoretic reading, in which one variant comes to serve as the cognitive reference point for the category (Rosch 1973, 1978).

5.1 JSL Terminology

Phrases examined in the study occurred a total of 916 times across the corpus of 85,929,343 characters. In order to account for variability in scope of data over time, statistics used to compare the usage of these terms were normalized by frequency per n characters. In aggregate across all municipalities, “Japanese language instruction/日本語指導” occurred most frequently (see Figure 3; the relative share of each variant over time is shown in Figure 4). No significant trend was observed among ordinance cities (see Figure 5).

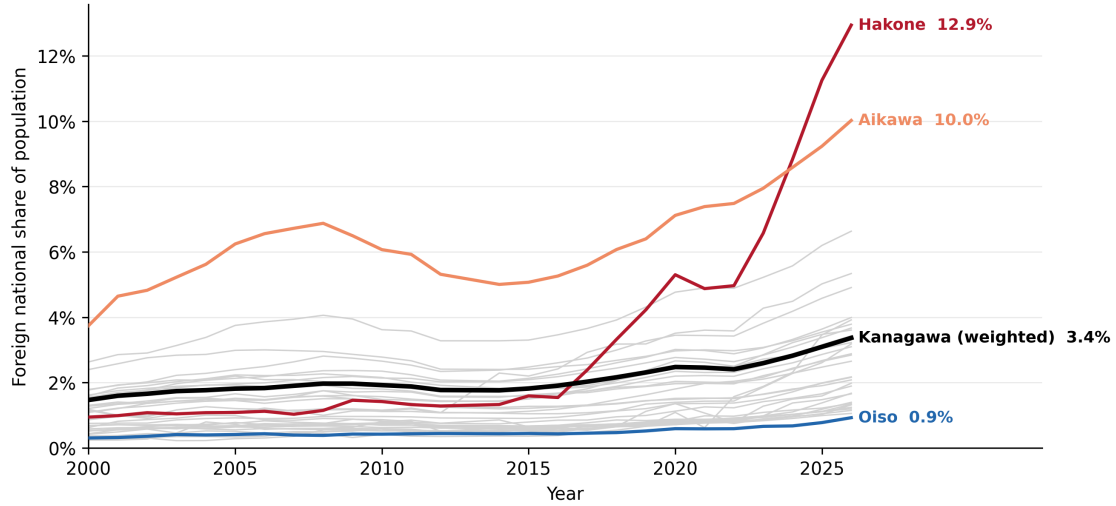


Figure 6: Foreign national population share across Kanagawa’s 33 municipalities, 2000–2026 (grey lines), with the population-weighted prefectural trend (black) and municipalities discussed in the text labeled.

5.2 Foreign National Population

By using demographic data released by the Kanagawa Prefectural Government, one can track changes in the ratio of foreign national residents to total population in each municipality in Kanagawa as early as 2001 (Kanagawa Prefectural Government 2026a,b). Across the prefecture, the population-weighted foreign share rose from 1.47% in 2000 to 3.37% in 2026, more than doubling over the duration for which data were available. This proportional growth occurred unevenly across municipalities (Figure 6): Hakone reached the highest share (12.94%), while Oiso remained lowest (0.93%).

5.3 Discourse vs. Foreign Resident Percentage

To test whether attention to JSL students tracks this demographic variation, mention rates were correlated with foreign-population share across municipalities. Municipalities with larger average foreign shares devoted significantly more deliberative attention to JSL students ($r = .54$, $p < .01$, 95% CI [.24, .74]; Spearman $\rho = .47$).

This Pearson value was taken using averages from each municipality over time to address serial correlation: the percentage of foreign residents in city A in 2015 cannot be considered independent from its high percentage in 2014. Similarly, a municipality’s vocabulary and topics of discussion are likely influenced by discussion in the recent past. To address this serial correlation, both foreign resident percentage and discussion rates were averaged across each municipality ($n=33$). JSL key-term mention rate correlates with percentage of foreign residents in a city at a Pearson $r = .54$, 95% CI [.24, .75] and a Spearman $\rho = .47$, indicating the association is not driven by a single outlying municipality.

To test if this association was specific to JSL terminology, the same procedure was applied to three placebo term sets irrelevant to foreign-resident policy: lunch items (牛乳, ミルク, お茶), stationery (ペン, 鉛筆), and transport (自転車, 電車). At the term-set level, none of these placebos showed a meaningful association ($r = .06, -.09, -.06$; all 95% CIs spanning zero) when compared to the JSL set. However, correlation between demographic trends and discourse focus does not necessarily represent a shift in prioritization; future researchers are called to examine how this correlation compares to other shifts in discourse frequency correlated with demographic changes.

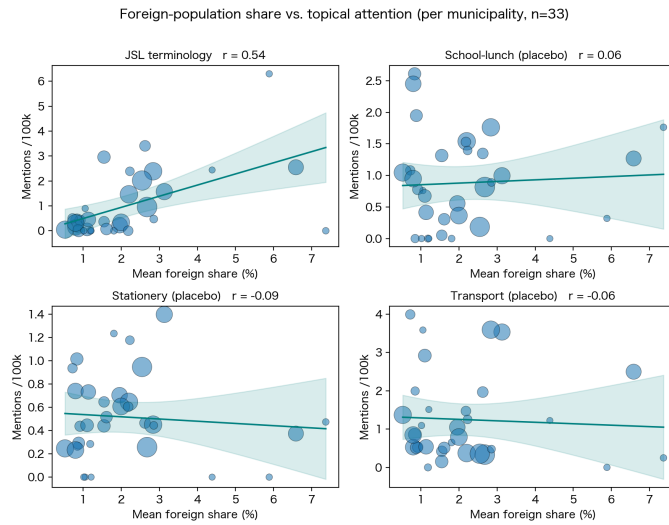


Figure 7: Foreign-population share and topical attention across 33 Kanagawa municipalities, 2004–2026. Each point is a municipality (sized by corpus volume); JSL terminology (top left) is compared against three placebo term sets. Shaded bands show 95% CIs of the fitted regression.

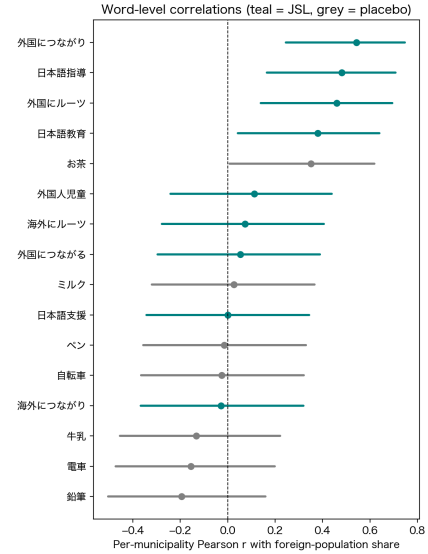


Figure 8: Word-level pooled city-year correlations with foreign-population share (Pearson r , 95% CI). JSL terms in green, placebo terms in grey.

6 Conclusion

In conclusion, this research constructed a novel tool for research on education discourse occurring at the local level. The utility of the corpus it outlines is demonstrated by a case examination of one aspect of public education (JSL) administered by Municipal Boards of Education (MBoE). The construction of this resource allows for the examination of a new frontier of questions concerning the implementation of public education, semantics of local policy discourse, and how the prioritization of different policy topics change over time. Further research can be performed to examine these questions as well as to expand the scope of the corpus to encompass municipalities in multiple prefectures to further explore how regional factors influence municipal discourse.

References

- YOKE. 外国につながる子ども・若者支援 [support for children and youth with foreign connections]. <https://www.yokeweb.com/kodomo/>, 2026. Accessed: 2026-07-06.
- Artifex Software, Inc. Pymupdf, 2026. URL <https://github.com/pymupdf/pymupdf>. Python bindings for the MuPDF library.
- Kanagawa Prefectural Government. 県内外国人統計 (外国人登録者統計) [statistics on foreign residents in kanagawa prefecture]. <https://www.pref.kanagawa.jp/docs/k2w/cnt/f4695/index.html>, 2026a. Updated 23 June 2026. Accessed 7 July 2026.
- Kanagawa Prefectural Government. 住民基本台帳人口 [resident registry population statistics]. <https://www.pref.kanagawa.jp/docs/v2x/cnt/f100643/index.html>, 2026b. Accessed 7 July 2026.
- Takumitsu Kudo. Mecab : Yet another part-of-speech and morphological analyzer. 2005. URL <https://api.semanticscholar.org/CorpusID:61584143>.
- Ministry of Education, Culture, Sports, Science and Technology. 「日本語指導が必要な児童生徒の受入状況等に関する調査 (令和 7 年度)」の結果について [regarding the results of the “Survey on the Acceptance of Students Requiring Japanese Language Instruction (FY2025)”]. 総合教育政策局国際教育課 (Comprehensive Education Policy Bureau, International Education Division), May 2026. INSERT_URL_HERE.
- Yujiro Mori. The development of social relationships of children with roots in foreign countries : through the educational practice based on social constructivism. 同志社政策科学研究, 19(2):169–184, 03 2018. ISSN 18808336. doi: 10.14988/pa.2017.0000017003. URL <https://cir.nii.ac.jp/crid/1390853649845627264>.

- Jane Mulderrig. Critical policy discourse analysis in higher education. In Kathy Hytten, editor, *Oxford Research Encyclopedia of Education*. Oxford Academic, New York, NY, January 2024. doi: 10.1093/acrefore/9780190264093.013.1702. URL <https://doi.org/10.1093/acrefore/9780190264093.013.1702>. Accessed: 2026-07-06.
- Eleanor H. Rosch. Natural Categories. *Cognitive Psychology*, 4(3):328–350, 1973.
- Eleanor H. Rosch. *Cognition and Categorization*, chapter Principles of Categorization, pages 27–48. Lawrence Erlbaum, NJ, Hillsdale, 1978.
- Jeremy Singer-Vine and The pdfplumber contributors. pdfplumber. <https://github.com/jsvine/pdfplumber>, 2025. Python library for extracting text, tables, and metadata from PDFs. Version 0.11.10.
- Keiichi Takamaru, Hokuto Ootake, Yuzu Uchida, and Yasutomo Kimura. Construction of japanese prefectural assembly minutes datasets across three electoral terms: Comparative analysis of 2011, 2015, and 2019 four-year periods. In Stelios Piperidis, Núria Bel, Henk van den Heuvel, Nancy Ide, Simon Krek, and Antonio Toral, editors, *Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)*, pages 6216–6223, Palma, Mallorca, Spain, May 2026. European Language Resources Association (ELRA). doi: 10.63317/5hsgnayssvei.
- Saeri Yamamoto. 国会における日本語教育関係議論のアクターと論点. *日本語教育*, 149:1–15, 2011. doi: 10.20721/nihongokyoiku.149.0_1. URL https://www.jstage.jst.go.jp/article/nihongokyoiku/149/0/149_1/_article/-char/ja.
- 地方教育行政の組織及び運営に関する法律（昭和三十一年法律第百六十二号）. 地方教育行政の組織及び運営に関する法律. e-Gov 法令検索, 令和 8 年 4 月 1 日施行版, 2026. URL <https://laws.e-gov.go.jp/law/331AC0000000162>. 2026 年 4 月 1 日確認.