DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

Equitable Metadata for Diverse Voices: Sustainable Computational Poetry Analysis with HathiTrust Extracted Features

Kahyun Choi[1], **You Peng**[1], and **Gyuri Kang**[2]

[1]School of Information Sciences, University of Illinois, Urbana-Champaign,, [2]Department of Information and Library Science, Indiana University, Bloomington

Abstract

The retirement of the HathiTrust Research Center (HTRC) infrastructure raises questions about continuing computational research on in-copyright collections in the HathiTrust Digital Library (HTDL). Since HTRC has actively supported inclusive research on underrepresented groups, the HTDL collections serve as a crucial test case for exploring post-HTRC workflows. To address this, we share an augmented dataset of American poetry by poets from historically underrepresented groups in the HTDL. Mapping this collection to HTRC Extracted Features (EF) v2.5 demonstrated that EF is highly reliable with high retrieval coverage, achieving a 100% match. Our computational linguistic analysis shows that EF effectively captures group-specific diversity, such as non-standard English, indigenous languages, and multilingual vocabularies. These findings indicate that adapting ML and NLP tools to properly handle such linguistic variation is essential to mitigate bias and marginalization. Although the lack of full-text limits structural analysis, EF remains a sustainable and highly useful resource for word-based research well beyond the HTRC's retirement.

1. Introduction

The HathiTrust Research Center (HTRC) has developed the HathiTrust Extracted Features (EF) dataset, a derived metadata representation of the HathiTrust Digital Library (HTDL). The EF dataset provides linguistic features, including token counts, part-of-speech tags, and page-level metadata, extracted from OCR-processed texts [1]. This derived metadata allows computational analysis of both public-domain and in-copyright materials under non-consumptive research conditions. In addition to EF, HTRC has also provided infrastructure, including the Data Capsule, a secure virtual environment that enables researchers to access and analyze full-text HTDL content. However, the HTRC infrastructure is scheduled for retirement in 2026 [2]. While EF access is expected to remain available, this transition raises important questions about whether EF alone can sustain computational research on HathiTrust collections in the future.

This change particularly affects research on historically underrepresented literary collections, which often involves in-copyright texts [3], [4]. Historically, literary analysis has focused on canonical texts, standard English, and some widely studied authors [5], [6], [7]. To address this imbalance, HTRC has supported inclusive initiatives such as SCWAReD [8]. Aligned with these efforts, our previous work introduced a curated collection of American poetry by poets from historically underrepresented groups in the HTDL [9], which we call the Poets from Historically Underrepresented Groups (PHUG) dataset. Providing a curated list of volume IDs was sufficient as long as the HTRC Data Capsule was available, allowing researchers to securely access and analyze the full-text

Short Paper

Available Aug 01, 2026

Correspondence to
Kahyun Choi
kahyun@illinois.edu

data. However, the retirement of HTRC will soon make this workflow no longer viable, raising questions about the continued computational analysis of culturally diverse, in-copyright texts.

To address this challenge, this paper introduces the Extracted Features (EF) version of the PHUG dataset. Using this derived metadata representation, we demonstrate how EF-based features can support computational literary analysis of HTDL materials without direct access to full text. Finally, through this case study, we explore how derived feature representations, such as EF, can serve as sustainable and equitable metadata for AI-driven analysis of culturally diverse literary collections in a post-HTRC era.

2. Datasets

2.1. HathiTrust Extracted Features

The HathiTrust Research Center (HTRC), established in 2011, supports large-scale computational analysis of works in the HathiTrust Digital Library (HTDL). The Center develops infrastructure, tools, and datasets that support massive text and data mining of the HTDL corpus for educational and non-profit research. Through these services, HTRC provides researchers with computational access to vast digital collections while adhering to copyright laws and non-consumptive research practices.

The HathiTrust Extracted Features (EF) dataset provides a set of structured features derived from OCR-processed pages. These features include volume-level metadata, page-level metadata, token counts, and part-of-speech (POS) tagged tokens generated using the Stanford NLP library [10], [11]. Files are distributed in JSON-LD format. The EF dataset contains more than 18 million volumes, 6.8 billion pages, and 3.2 trillion tokens, including both public-domain and in-copyright items.

EF enables broad-scale computational analyses, such as lexical, stylistic, and genre-based studies, across the HTDL collections without violating copyright restrictions. Previous studies have used EF-based features for tasks such as word similarity [12], topic modeling [13], and literary genre identification [14]. Despite its usability, the linguistic features in EF are derived from OCR-processed text, which may contain OCR errors. These errors can influence subsequent tokenization and POS tagging. In addition, general-purpose POS taggers, such as the Stanford NLP parser, are typically trained on modern prose and may perform less reliably on non-standard, domain-specific texts, including poetry. These inherent limitations should be considered when interpreting token-level results.

In October 2024, HathiTrust Executive Director Mike Furlough announced that HathiTrust will cease funding and data access for the HTRC at the end of 2026. Since the announcement, HTRC has been preparing a transition and sunsetting process. Unlike the discontinued access to Data Capsule, access to HTDL, Bookworm, and derived datasets, including EF, will remain available. As a result, EF is expected to remain an important resource for computational analysis of the HathiTrust collections after the retirement of the HTRC infrastructure.

2.2. Poets from Historically Underrepresented Groups (PHUG) Dataset

In this study, we build on the PHUG dataset introduced by Kang and Choi [9], a curated collection of American poetry by poets from historically underrepresented groups in the HTDL. The dataset comprises 4,723 poems from 120 collections by African American (AA), Asian American (APA-AA), Pacific Islander (APA-PA), Latin American (LXA), and Native American (NA) poets, with a detailed breakdown provided in Appendix Table A1. Most importantly, it provides poem-level boundary annotations, including the start and end page numbers of each poem. This structural metadata enables large-scale computational poetry analysis through retrieving poem records in HTRC EF.

To construct the derived metadata representation, we cross-referenced the PHUG dataset with EF. Initially, we used EF v2.0, yielding 92% coverage, with 8 volumes and 375 poems missing. By upgrading to EF v2.5, retrieval coverage reached 100%; all 4,723 poems across 120 volumes were matched to their corresponding EF data. This shows the high reliability and robustness of the EF dataset.

We also observed meaningful qualitative improvements in token extraction in v2.5, resulting from HTRC's recently updated OCR pipeline. When comparing the exclusive words extracted from both versions, we found that EF v2.5 more accurately recovers orthographically complex forms, especially words with apostrophes and diacritics. For instance, words that appeared in EF v2.0 in simplified or degraded forms such as *Wen*, *shiyazhf*, *Hooulu*, *tia*, *que*, and *Dine* are correctly captured in EF v2.5 as *W'en*, *Shiy'azh'i*, *Ho'oulu*, *tía*, *Qué*, and *Diné*. These examples demonstrate how improved OCR processing preserves critical orthographic markers that the earlier algorithms often missed. Such enhanced token recognition allows a more robust linguistic analysis across diverse and non-standard language forms. The matching extracted features and analysis code are publicly available via Zenodo and GitHub, respectively.[1]

3. Linguistic Diversity Analysis

3.1. Method

To identify unique linguistic characteristics and culturally specific words for each poet group, we conducted an exclusive word analysis using token frequency from EF. We chose this approach because general POS taggers perform poorly on poetry, and we preserved original capitalization to identify culturally significant proper nouns and capture deliberate stylistic emphasis. We recognized that a strict zero-tolerance criterion (a threshold of 0) was too restrictive and would eliminate highly distinctive terms that appeared only rarely in other groups.

To address this, we empirically tested a cross-group overlap threshold, increasing it incrementally. We observed no significant changes beyond a threshold of 2; a threshold of 3 yielded few new words, and those added tended to have relatively low frequencies. Therefore, we set the final threshold at 2. This slightly relaxed approach helped retain

[1]

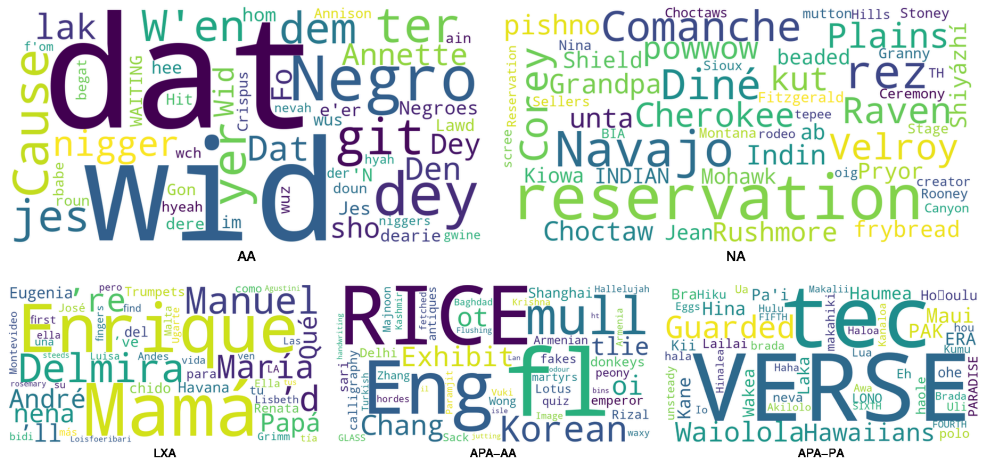

Figure 1. *Word clouds of the top 50 exclusive words for each poet group (threshold of 2).*

important, high-frequency words that are distinctive in a particular group, such as *Dat* and *dere* in the AA group, without compromising overall cross-group distinctiveness. We visualized the top 50 exclusive words per group with a threshold of 2 in Figure 1. Frequency counts and ranked word lists are provided in Table B1 in the Appendix.

3.2. *Results and Discussion*

In the African American (AA) group, the most frequent exclusive words represent African American English (AAE) variants and African American Vernacular English (AAVE). For example, 'dat', which indicates a common pronunciation of 'that' in AAE [15], appears 154 times and 'wid' referring to 'with' in AAVE has 124 counts, along with 99 counts of 'dey' meaning 'they' in AAVE. 'fo (Fo)' meaning 'for', reflects a common characteristic of AAE known as r-dropping, in which the final r/ is not pronounced [15].

Asian American (APA-AA) group has the fewest exclusive word counts, indicating that the group is less homogeneous than the other groups. There are more words referring to various cultures and locations than other groups. 'RICE (rice)' appears 28 times across the corpus, which holds cultural, agricultural, and culinary significance in Asian countries [16], [17]. 'Korean' (16 counts), 'Shanghai' (13), 'Armenian' (8), 'Kashmir' (7), 'Baghdad' (7), 'Turkish' (7), 'Delhi' (6), 'Nepal' (6) and 'Armenia' (6) demonstrate the diversity of ethnicity and nationality within the same group. Some popular surnames for specific ethnic groups, such as 'Chang' (16), 'Wong' (9), and 'Zhang', (8) are also included in the APA-AA exclusive word list. Names are important indicators for inferential ethnic classification, which overcomes the limitation of an aggregate Asian American category [18].

Pacific Islander (APA-PA) group's exclusive words display the most linguistic distinctiveness. The words 'Waiolola' and 'Waiololi', both of which count 43 times, are respectively a female and male figure from *The Kumulipo*, the creation chant in Hawaiian religion, written by Liliuokalani, Queen of Hawaii. The former translates as 'broad stream' and the latter as 'narrow stream' [19]. However, direct translations do not deliver the significance of the mythic figures throughout the narrative poem. These examples indicate

that preserving linguistic originality in multilingual poetry is important because direct translations can hinder multiple interpretations of original words and their musicality in the poem [19].

In Latin American (LXA) group, ethnic personal words frequently appear: 'Enrique' (80 counts), 'Delmira' (73), 'Manuel' (64), María (48), 'André' (41), 'Eugenia' (35), 'Renata' (24), 'Lisbeth' (16), 'Luisa' (14), 'Ugarte' (14), and 'Agustini' (12) [20]. This shows a balanced mix of male and female names, but there are more words referring to women: 'Mamá' (98), 'nena' (40), 'Ella' (17), 'ella' (14), and 'tía' (13), respectively translating as 'mom', 'little girl', 'she/her', and 'aunt' [21]. Location words, such as 'Havana' (30), 'Montevideo' (16), 'Andes' (16), 'Minas' (10), and 'Juarez' (10), demonstrate that LXA poets have different national backgrounds, including Cuba, Uruguay, Brazil, and Mexico. Interestingly, unlike APA-AA group, in which there are words referring to ethnicity and country names, LXA group has more words of city names.

Native American (NA) group has a lot of words related to NA history and culture. The most frequent exclusive word 'reservation' with 87 counts and the third most frequent word 'rez' with 45 counts both refer to a designated land area managed by a Native American tribe. In addition, popular tribe names appear in the word list: 'Navajo' (51 counts), 'Comanche' (35), 'Diné' (35), 'Cherokee' (23), 'Choctaw' (20), 'Mohawk' (17), 'Kiowa' (15), 'Sioux' (13), 'Choctaws' (13), 'Stoney' (12), and 'Navajos' (11). Some words, such as 'Directions' (22) [22], 'powwow' (21; a Native American ritual), 'fry-bread' (18; traditional Native American food), INDIAN (Indian) (17), 'Ceremony' (14), and 'tepee' (11; a traditional Native American tent) represent Native American culture and history [23], [24], [25], [26], [27]. Many counts of community-related words, such as 'Grandpa' (21), 'pishno' (21; meaning 'we', 'us', 'ours') [28], and 'Granny' (16), as well as nature- and landscape-related words, including 'Raven' (27), 'Plains' (24), 'Rushmore' (20), and 'shiprock' (11), indicate NA group's iconic interests in family and the natural environment [29], [30], [31].

4. Conclusion

This study explores the feasibility of sustainable computational research on underrepresented literary collections using only derived metadata following the retirement of the HTRC infrastructure. We share an augmented PHUG dataset mapped to HTRC Extracted Features (EF) v2.5. Achieving 100% coverage across all 120 volumes and 4,723 poems demonstrates the high reliability of the EF dataset for this purpose. Our analysis revealed unique linguistic diversity across poet groups, indicating that ML/NLP tools can more effectively process these collections when designed to accommodate diverse linguistic forms and culturally specific expressions. Although the lack of full text limits structural analysis, EF shows promising potential for word-based research. Before the HTRC retirement, we plan to expand the currently small APA-PA corpus to better represent the group.

Acknowledgments

This work is supported by the Institute of Museum and Library Services (Grant No. RE-252382-OLS-22) under the Laura Bush 21st Century Librarian Program.

Appendix A. PHUG Dataset Statistics

Table A1. *Number of volumes and poems by poet group in the PHUG dataset.*

Poet Group	Volumes	Poems
African American (AA)	34	1,494
Asian American (APA-AA)	22	761
Pacific Islander (APA-PA)	3	134
Latin American (LXA)	26	948
Native American (NA)	35	1,386
Total	**120**	**4,723**

Appendix B. Exclusive Word Lists

Table B1. *Top 50 exclusive words for each poet group (threshold of 2). Values in parentheses indicate the word's total occurrences across the other four groups combined.*

	aa_poets		apa-aa_poets		apa-pa_poets		lxa_poets		na_poets	
Rank	Word	Count (Elsewhere)	Word	Count (Elsewhere)	Word	Count (Elsewhere)	Word	Count (Elsewhere)	Word	Count (Elsewhere)
1	dat	154 (2)	fl	29 (0)	VERSE	55 (0)	Mamá	98 (0)	reservation	87 (1)
2	wid	124 (0)	RICE	28 (1)	tec	45 (0)	Enrique	80 (0)	Navajo	51 (1)
3	Negro	110 (0)	Eng	23 (0)	Guarded	43 (0)	Delmira	73 (0)	rez	45 (0)
4	dey	99 (0)	mull	19 (0)	Waiolola	43 (0)	Manuel	64 (1)	Comanche	35 (2)
5	git	67 (0)	Korean	16 (2)	Waiololi	43 (0)	'd	55 (2)	Diné	35 (0)
6	Cause	66 (2)	Chang	16 (0)	Hawaiians	34 (0)	're	49 (0)	Corey	30 (0)
7	ter	63 (1)	Exhibit	16 (1)	Maui	22 (0)	María	48 (1)	Raven	27 (1)
8	yer	62 (0)	oi	15 (0)	PAK	20 (0)	André	41 (1)	Velroy	26 (0)
9	jes	61 (0)	ot	15 (0)	Hina	18 (1)	nena	40 (0)	Plains	24 (1)
10	W'en	59 (0)	tlie	13 (1)	Kane	17 (0)	Papá	39 (0)	Cherokee	23 (2)
11	dem	59 (0)	Shanghai	13 (0)	ERA	16 (0)	Qué	36 (0)	kut	22 (0)
12	nigger	56 (0)	donkeys	12 (1)	Haumea	16 (0)	'll	35 (2)	powwow	21 (0)
13	Dat	54 (1)	sari	12 (0)	Wakea	14 (0)	Eugenia	35 (0)	Grandpa	21 (0)
14	lak	54 (0)	Delhi	11 (0)	Laka	14 (0)	've	30 (1)	Indin	21 (0)
15	Den	44 (0)	emperor	11 (2)	Pa'i	13 (0)	Havana	30 (0)	unta	21 (0)
16	Annette	42 (0)	Rizal	11 (0)	Kii	11 (0)	chido	28 (0)	pishno	21 (0)
17	Fo	38 (1)	Lotus	10 (1)	Bra	11 (0)	del	27 (2)	Rushmore	20 (2)
18	Dey	37 (0)	calligraphy	10 (2)	Hooulu	10 (0)	Trumpets	25 (0)	Choctaw	20 (1)
19	sho	36 (0)	fakes	10 (0)	ohe	10 (0)	para	24 (0)	frybread	18 (0)
20	Wid	36 (0)	Wong	9 (1)	haole	9 (1)	Renata	24 (0)	Pryor	18 (1)
21	Negroes	35 (1)	quiz	9 (2)	Lailai	9 (0)	tu	22 (2)	Shiyázhí	18 (0)
22	hom	35 (0)	peony	9 (0)	polo	9 (2)	como	19 (0)	Shield	18 (1)
23	Jes	34 (0)	antiques	9 (1)	Eh	9 (2)	José	19 (0)	INDIAN	17 (2)
24	'N	33 (1)	Armenian	9 (1)	neva	9 (0)	first	18 (0)	ab	17 (0)
25	e'er	32 (2)	martyrs	8 (2)	LONO	9 (0)	una	17 (1)	Mohawk	17 (1)
26	im	31 (1)	Majnoon	8 (0)	PARADISE	9 (1)	Ella	17 (2)	Kiowa	16 (0)
27	hee	31 (1)	Sack	8 (0)	Hiku	8 (0)	Grimm	17 (2)	beaded	16 (2)
28	dearie	29 (0)	Zhang	8 (0)	makahiki	8 (0)	Andes	16 (1)	Jean	16 (2)
29	WAITING	29 (2)	Hallelujah	7 (1)	unsteady	7 (2)	Montevideo	16 (0)	Fitzgerald	16 (1)
30	wus	28 (0)	Vuki	7 (0)	Uli	7 (0)	Lisbeth	16 (0)	Granny	16 (0)
31	Lawd	28 (0)	Baghdad	7 (1)	Lua	7 (0)	más	15 (0)	Stoney	14 (0)
32	dere	27 (2)	Krishna	7 (0)	Ua	7 (0)	vida	15 (1)	Montana	14 (2)
33	Gon	27 (2)	hordes	7 (1)	hou	7 (1)	LA	15 (2)	mutton	14 (0)
34	wuz	26 (0)	Kashmir	7 (0)	hala	7 (1)	Malta	15 (0)	Ceremony	14 (1)
35	Hit	25 (1)	GLASS	7 (1)	brada	7 (0)	fingers	15 (0)	Sellers	14 (0)
36	roun	23 (0)	Armenia	7 (1)	Kumu	7 (0)	find	15 (0)	Stage	14 (1)
37	Crispus	23 (0)	Image	7 (1)	Hulu	6 (0)	bidi	15 (0)	Nina	14 (2)
38	babe	22 (2)	Turkish	7 (2)	Awa	6 (0)	Las	14 (2)	Rooney	13 (0)
39	doun	22 (0)	waxy	7 (2)	Akilolo	6 (0)	Ugarte	14 (0)	Choctaws	13 (0)
40	ain	22 (1)	Paramjit	7 (0)	Haha	6 (0)	ella	14 (0)	TH	13 (0)
41	wch	22 (1)	fetched	7 (2)	Eggs	6 (0)	Luisa	14 (0)	Hills	12 (2)
42	Annison	21 (0)	Flushing	7 (0)	Io	6 (2)	su	13 (0)	BIA	12 (0)
43	hyeah	21 (0)	Lan	6 (0)	Kanaloa	6 (0)	ven	13 (0)	oig	12 (0)
44	begat	21 (2)	handwriting	6 (2)	Hinalea	6 (0)	pero	13 (1)	Sioux	12 (0)
45	der	20 (1)	il	6 (0)	Haloa	6 (0)	tía	13 (0)	creator	12 (1)
46	niggers	20 (0)	jutting	6 (2)	Brada	6 (0)	Loisfoeribari	13 (0)	scree	12 (0)
47	nevah	20 (0)	odour	6 (1)	Makalii	5 (0)	tus	12 (0)	tepee	11 (0)
48	hyah	20 (0)	ht	6 (0)	SIXTH	5 (0)	Agustini	12 (0)	Canyon	11 (2)
49	f'om	20 (0)	isle	6 (2)	FIFTH	5 (0)	rosemary	12 (1)	rodeo	11 (2)
50	gwine	20 (0)	bins	6 (1)	FOURTH	5 (2)	steeds	12 (2)	Reservation	11 (0)

References

[1] HathiTrust Research Center, "Extracted Features v.2.5." [Online]. Available: https://go.illinois.edu/EF25

[2] HathiTrust Research Center, "2026 HTRC Transition Guide." [Online]. Available: https://htrc.atlassian.net/wiki/spaces/COM/pages/1324810243/2026+HTRC+Transition+Guide

[3] K. Ibacache, "Building an underrepresented collection," in *Building Library Collections and Curriculum Centered on Indigenous Knowledge*, Routledge, 2024. doi: 10.4324/9781032660561-2.

[4] K. P. Alexander, D. Terry, J. Kirby, R. J. Wittmann, and A. Neatrour, "The invisible default: Examining representation in digital collections," *Information Technology and Libraries*, vol. 44, no. 3, 2025, doi: 10.5860/ital.v44i3.17306.

[5] G. M. Gugelberger, "Decolonizing the canon: Considerations of third world literature," *New Literary History*, vol. 22, no. 3, pp. 505–524, 1991, [Online]. Available: https://www.jstor.org/stable/469201

[6] J. Marx, "Western literary canon," in *The Cambridge Companion to Postcolonial Literary Studies*, 2004, p. 83.

[7] L. Zhang, "Canon and world literature," *Journal of World Literature*, vol. 1, no. 1, pp. 119–127, 2016, doi: 10.1163/24056480-00101012.

[8] I. Magni *et al.*, "The SCWAReD projects: Scholar-Curated Worksets for Analysis, Reuse & Dissemination." 2025.

[9] G. Kang and K. Choi, "A dataset of American poetry by poets from historically underrepresented groups in the HathiTrust Digital Library," *Journal of Open Humanities Data*, vol. 12, no. 1, 2026.

[10] J. A. Walsh *et al.*, "*The library is open!*": *Open data and an open API for the HathiTrust Digital Library*. in CHR. 2023, pp. 703–714.

[11] J. A. Walsh *et al.*, "The HathiTrust Research Center Extracted Features Dataset (2.5)," HathiTrust Research Center, 2025. doi: 10.13012/PXP0-F135.

[12] D. Mimno, "Word Similarity Tool, Extracted Features in the Wild," HathiTrust Research Center Wiki. [Online]. Available: https://htrc.atlassian.net/wiki/spaces/COM/pages/43289591/Extracted+Features+in+the+Wild

[13] P. Organisciak, "Within-Book Topic Modeling, Extracted Features in the Wild," HathiTrust Research Center Wiki. [Online]. Available: https://htrc.atlassian.net/wiki/spaces/COM/pages/43289591/Extracted+Features+in+the+Wild

[14] HathiTrust Research Center, "Extracted Features 2.0 Use Cases and Examples, Use Case 1: Using EF to identify poetry and prose volumes," HathiTrust Research Center Wiki. [Online]. Available: https://htrc.atlassian.net/wiki/spaces/COM/pages/43288220/Extracted+Features+2.0+Use+Cases+and+Examples

[15] J. Graham, N. L. Day-Vines, and K. Zaccor, "Dis, Dat, and Dem: Addressing linguistic awareness for counselors of African American English speakers," *Journal of Multicultural Counseling and Development*, vol. 50, no. 2, pp. 73–81, 2022, doi: 10.1002/jmcd.12241.

[16] D. Q. Fuller, "Pathways to Asian civilizations: Tracing the origins and spread of rice and rice cultures," *Rice*, vol. 4, no. 3–4, pp. 78–92, 2011, doi: 10.1007/s12284-011-9078-7.

[17] K. H. Ko, "The influence of rice agriculture on East Asian culture and language," *European Journal of East Asian Studies*, vol. 15, no. 1, pp. 86–107, 2016, doi: 10.1163/15700615-01501001.

[18] D. S. Lauderdale and B. Kestenbaum, "Asian American ethnic identification by surname," *Population Research and Policy Review*, vol. 19, no. 1, pp. 283–300, 2000, doi: 10.1023/A:1026582308352.

[19] B. N. McDougall, "Mo'okū'auhau versus colonial entitlement in English translations of the Kumulipo," *American Quarterly*, vol. 67, no. 3, pp. 749–779, 2015, doi: 10.1353/aq.2015.0054.

[20] M. Aceto, "Ethnic personal names and multiple identities in Anglophone Caribbean speech communities in Latin America," *Language in Society*, vol. 31, no. 4, pp. 577–608, 2002.

[21] SpanishDict, "SpanishDict." [Online]. Available: https://www.spanishdict.com/

[22] Akta Lakota Museum & Cultural Center, "Native American Four Directions." [Online]. Available: https://aktalakota.stjo.org/lakota-culture/native-american-four-directions/

[23] A. Axtmann, "Performative power in Native America: Powwow dancing," *Dance Research Journal*, vol. 33, no. 1, pp. 7–22, 2001, doi: 10.2307/1478853.

[24] J. A. Cross, "Native American landscapes in the Plains and Northwest Coast," in *Ethnic Landscapes of America*, Springer, 2017, pp. 47–64.

[25] D. Mihesuah, "Indigenous health initiatives, frybread, and the marketing of nontraditional "traditional" American Indian foods," *Native American and Indigenous Studies*, vol. 3, no. 2, pp. 45–69, 2016.

[26] L. Roy, "Four directions: An indigenous educational model," *Wicazo Sa Review*, vol. 13, no. 2, pp. 59–69, 1998, [Online]. Available: https://www.jstor.org/stable/1409146

[27] S. Willis, "The four directions," *International Journal of Art & Design Education*, vol. 24, no. 1, pp. 31–42, 2005, doi: 10.1111/j.1476-8070.2005.00421.x.

[28] Wiktionary contributors, "pishno." [Online]. Available: https://en.wiktionary.org/wiki/pishno

[29] A. L. Booth, "We are the land: Native American views of nature," in *Nature Across Cultures: Views of Nature and the Environment in Non-Western Cultures*, Springer, 2003, pp. 329–349.

[30] C. Byington, *Grammar of the Choctaw Language*. McCalla & Stavely, printers, 1870.

[31] H. N. Weaver and B. J. White, "The Native American family circle: Roots of resiliency," in *Cross-Cultural Practice with Couples and Families*, Routledge, 2019, pp. 67–79.