

Socio-Phonosemantic Digitization of Blaan Cultural Terminologies Through an Empirical Dataset and Ethical Data Governance Framework

Aisah B. Camar¹, Jeffrey T. Importante², Mary Jane U. Saligumba³, Mary Jane C. Tamba⁴, Ricaflor R. Avila⁵

¹ Department of General Teacher Training, North Eastern Mindanao State University–Cantilan Campus, Cantilan, Surigao del Sur 8317, Philippines.

Email: abcamar@nemsu.edu.ph

² College of Teacher Education, University of Antique–Tario Lim Memorial Campus, Tibiao, Antique 5707, Philippines.

Email: jeffrey.importante@antiquespride.edu.ph

³ College of Teacher Education, University of Antique, Sibalom, Antique 5713, Philippines.

Email: maryjane.saligumba@antiquespride.edu.ph

⁴ College of Arts and Sciences, University of Antique, Sibalom, Antique 5713, Philippines.

Email: maryjane.tamba@antiquespride.edu.ph

⁵ College of Education, Arts and Sciences, Capiz State University–Pontevedra Campus, Pontevedra, Capiz 5802, Philippines.

Email: rravila@capsu.edu.ph

Corresponding Author: Aisah B. Camar

Abstract: This study aims to document the socio-phonosemantic terminologies of the Blaan language to structure an empirical dataset foundational for culturally responsive artificial intelligence (AI) and digital information systems. Employing a descriptive-qualitative research design through ethnographic methodologies, the study executed rigorous data acquisition and validation. Informants were selected via purposive and convenience sampling from within the Blaan community to ensure cultural authenticity and structural integrity. Findings present a structured dataset revealing distinct linguistic features, including the presence of the schwa sound, frequent use of the phonemes /f/ and /v/, and multiple consonant clusters (e.g., /bl/, /fn/, /kw/). These phonetic nuances present specific computational challenges that require bespoke natural language processing (NLP) algorithmic training rather than rigid, pre-existing templates. Furthermore, the documentation of deep cultural significance, such as sensitive terminologies like *ki* and traditional *de jure* laws governed by tribal leaders, demonstrates the critical necessity of embedding Role-Based Access Control (RBAC) and decentralized administrative hierarchies into digital repositories. Ultimately, this research bridges localized indigenous knowledge with modern digital tools, establishing an ethical data governance framework. By aligning traditional customs with Indigenous Data Sovereignty, the study provides a vital blueprint for ensuring that technological integration protects, rather than exploits, marginalized ethnolinguistic identities.

Keywords: Artificial Intelligence; Indigenous Data Sovereignty; Natural Language Processing; Ethical Data Governance; Low-Resource Languages; Blaan language

1. Introduction

A critical gap exists in the rapid deployment of artificial intelligence (AI) and information and communication technology (ICT), wherein the digitization of low-resource indigenous languages frequently occurs without culturally responsive governance frameworks, risking data extraction devoid of community consent (Kwok, 2026; Shi &



DiFranzo, 2026). While the development of natural language processing (NLP) applications relies heavily on large, digitized datasets, underrepresented languages face significant computational barriers. These include linguistic fragmentation, a lack of standardized orthographies, and the profound difficulty of encoding dialectal variations and oral traditions into machine-readable formats (Ketere, 2025). Consequently, government and Big Tech initiatives often struggle with ethical data stewardship, leading to data colonialism that strips linguistic data of its cultural ownership and representation (Kwok, 2026). To ensure that multilingual language models and inclusive educational technologies do not reproduce existing global inequalities, there is an urgent need for an empirical, culturally grounded governance approach that integrates Indigenous rights, data boundaries, and accountable AI design directly into the information system's architecture (Shi & DiFranzo, 2026).

To build these equitable information systems, precision in data validation must begin with the recognition that language is not merely a medium of communication but the fundamental expression and repository of culture itself (Salazar, 1996). If culture represents the totality of thought, behavior, and knowledge defining a community's unique character, there is no culture without language (Salazar, 1996). In any language, specific words and expressions mirror the way of life, socialization, and adaptation of its speakers within their environment (Calinawagan, 2012). This perspective is drawn from deep-rooted traditions and beliefs that connect and strengthen unity, guiding a community toward a meaningful life (Rubrico, 2012). For ICT integration to be truly learner-centered and ethically sound, system development must document these sociolinguistic patterns (Labov, 1972). Ethnographic documentation, therefore, becomes a vital, empirical prerequisite for AI, grounding digital systems in the lived realities of ethnolinguistic groups (Batanes, 1999; Schmidt, 1993).

In the Philippines, particularly in Mindanao, various Lumad groups strive to safeguard their distinct languages and identities against the rapid pace of technological assimilation (Austro, 2011; Jocano, 1998). The Blaan, an ethnolinguistic group whose name derives from *bla* and *an* (meaning opposite people) (De Jong, 2009), is one such community actively striving against linguistic marginalization (Bautista, 2005; Valera, 2002). Primarily residing in South Cotabato, Sultan Kudarat, Sarangani, and Davao del Sur, the Blaan population is categorized into subgroups such as the To Lagad (highlanders) and To Baba (lowlanders), each contributing to the language's dialectal richness (Arcenas, 1993; Blount, 1977; Macatabon & Calibayan, 2016). The Blaan possess a profound cultural heritage encompassing unique phonological characteristics (Gonzales, 2003), traditional practices, and distinct societal laws governed by tribal leaders known as *Fulong* (Du, 2014). However, without a dedicated framework bridging their local knowledge with digital tools, the preservation of the Blaan language remains severely challenged.

Translating this rich cultural matrix into a digital environment requires more than a simple glossary; it demands an ethical data governance framework rooted in Indigenous Data Sovereignty (IDS). IDS refers to the inherent right of Indigenous peoples to control data from and about their communities, articulating collective rights to data access, privacy, and system stewardship (Kukutai & Walter, n.d.). Operationalizing IDS practices means minimizing harm and maximizing community benefits through intentional engagement with Indigenous researchers and local stakeholders (Calac & Mackey, 2025). Just as pioneering digital projects have successfully carved out space for indigenous representation by involving native language specialists throughout the entire AI development pipeline (Jones et al., 2023), the digitization of the Blaan language requires strict adherence to legal and ethical guidelines, including community consent and the right to withdraw data (Ketere, 2025).

Therefore, this study aims to systematically document the socio-phonosemantic terminologies of the Blaan language not merely as an anthropological exercise, but to structure an empirical dataset foundational for AI and digital information systems. By detailing Blaan phonetics, cultural terminologies, and societal rules, this research provides the essential blueprint for culturally responsive NLP development. Through this integration of indigenous knowledge and ICT, the study establishes a digital governance model that protects cultural privacy while advancing the structural integrity of modern technological applications.

2. Methods

This study employed a descriptive-qualitative research design, utilizing content analysis through an ethnographic and indigenous approach. Within the context of information systems and artificial intelligence (AI) development, this traditional ethnographic methodology serves as the rigorous data acquisition and validation phase required to construct an empirical, culturally responsive dataset. The research was conducted in Barangay Lampitak, Municipality of Tampakan, South Cotabato, a location chosen specifically because it is home to the majority of the Blaan people.

To ensure the authenticity and structural integrity of the digital dataset, informants were selected through convenience and purposive sampling. The criteria for these primary data stewards required that they be male or female members of the Blaan community, aged 50 and above, possessing good eyesight, hearing, and speech, alongside sufficient expertise in the Blaan language.

The data acquisition and validation protocol adhered to strict ethical governance guidelines, executed sequentially to ensure computational readiness. Initially, consent letters, data collection tools, and a guide questionnaire aligned with the study's objectives were prepared to standardize the personal interviews. Formal permissions were secured from the Municipal Mayor of Tampakan, the Barangay Captain of Lampitak, and the Director of the National Commission on Indigenous Peoples (NCIP) Region 12, ensuring community consent and alignment with Indigenous Data Sovereignty principles. Following informant consent, actual data collection was conducted via personal interviews over a one-month period, with each shared cultural term systematically listed in a tabular format. Furthermore, video recordings were utilized to capture verifiable multimedia data, ensuring the accurate preservation of pronunciation and conceptual meaning for future algorithmic training.

To maintain high precision in data validation, the gathered terminologies underwent a twenty-day validation process. Researchers consulted the informants to verify the correct pronunciation and orthography of each word, assisted by an interpreter fluent in both the Blaan and Filipino languages. The validated data was then compiled into a structured glossary featuring correct Blaan orthography, exact phonetic transcriptions, and standardized translations into both the Filipino and English languages, making the data highly accessible for bilingual database architecture.

Finally, the structuring of this dataset was informed by a multi-layered analytical framework. The cultural terminologies were subjected to sociolinguistic, phonological, semantic, and morphosyntactic analysis to uncover the foundational rules of the language. By integrating these perspectives, the study treated Blaan terminologies not only as linguistic artifacts but as discrete data entities, ensuring their complex social and cultural functions were accurately mapped and preserved for future integration into relational databases or natural language processing (NLP) models.

3. Results

To bridge the gap between indigenous cultural preservation and modern information systems, the following sociolinguistic findings are presented not merely as an anthropological glossary, but as a structured, empirical dataset designed for direct integration into natural language processing (NLP) models and relational databases. In this section, the collected cultural terms of the Blaan are presented, categorized according to body parts, types of vegetables, types of weather, numerals, colors, times, characteristics, terms related to marriage, superstitions or beliefs, types of food, terms related to settlement or agreements, and methods of healing. Also included are the phonetic transcriptions of each word, their translations in Filipino and English language, and the results of the analysis of the cultural terms of the Blaan. For clearer presentation of the data, a legend is used, with B indicating the Blaan language, F indicating the Filipino language, and E for English language.

CULTURAL TERMS OF THE BLAAN

TABLE 1. Parts of the Body

English	Filipino	Blaan	Phonetic Transcription
<i>hair</i>	buhok	<i>pék</i>	/wek/
<i>head</i>	ulo	<i>ulu</i>	/ʔu.luh/
<i>ear</i>	tainga	<i>klinge</i>	/kli.nge/
<i>forehead</i>	noo	<i>bunah</i>	/bunah/
<i>neck</i>	leeg	<i>liél</i>	/liʔel/
<i>eyelashes</i>	pilikmata	<i>klubeng</i>	/klu.bəng/
<i>teeth</i>	ngipin	<i>kifén</i>	/kifen/
<i>cheek</i>	pisngi	<i>fifi</i>	/fi.fi/
<i>lip</i>	labi	<i>salek</i>	/sa.lek/
<i>shoulder</i>	balikat	<i>blagél/balah</i>	/blagel/, /ba.lah/
<i>eyebrows</i>	kilay	<i>fulek</i>	/fu.lek/
<i>arm</i>	braso	<i>sigel/nalò</i>	/sigel/, /na.loʔ/
<i>elbow</i>	siko	<i>sigu</i>	/si.guh/
<i>waist</i>	beywang	<i>kwang</i>	/kwang/

<i>finger</i>	<i>daliri</i>	<i>namet</i>	/namet/
<i>stomach</i>	<i>tiyan</i>	<i>tién</i>	/ti ² en/
<i>chest</i>	<i>dibdib</i>	<i>kfu</i>	/kfu/
<i>underarm</i>	<i>kilikili</i>	<i>emak</i>	/ə.mak/
<i>breast</i>	<i>suso</i>	<i>toto</i>	/toto/
<i>navel or belly button</i>	<i>pusod</i>	<i>blutut</i>	/blu.tut/
<i>knee</i>	<i>tuhod</i>	<i>bukél</i>	/bukel/
<i>foot</i>	<i>paa</i>	<i>fued</i>	/fu ² ed/
<i>nail</i>	<i>kuko</i>	<i>tnugu</i>	/tnugu/
<i>leg</i>	<i>binti</i>	<i>usùbli</i>	/ ² usub.li ² /
<i>palm</i>	<i>palad</i>	<i>falél</i>	/falel/
<i>heel</i>	<i>takong</i>	<i>takéd</i>	/taked/
<i>hand</i>	<i>kamay</i>	<i>knamet</i>	/knamet/
<i>sole (of the foot)</i>	<i>talampakan</i>	<i>falél bli</i>	/falel.bli ² /
<i>skin</i>	<i>balat</i>	<i>kulit</i>	/kulit/
<i>but</i>	<i>puwet</i>	<i>kitut</i>	/kitut/
<i>back</i>	<i>likod</i>	<i>kagol</i>	/kagol/
<i>female genitalia</i>	<i>ari ng babae</i>	<i>kì</i>	/ki ² /
<i>male genitalia</i>	<i>ari ng lalake</i>	<i>batò</i>	/ba.to ² /

TABLE 2. Types of Vegetables

English	Filipino	Blaan	Phonetic Transcription
<i>eggplant</i>	<i>talong</i>	<i>sgatong</i>	/sgatong/
<i>onion</i>	<i>sibuyas</i>	<i>sibuyés</i>	/sibu.yes/
<i>squash</i>	<i>kalabasa</i>	<i>tambasal</i>	/tambasal/
<i>ginger</i>	<i>luya</i>	<i>malah tanà</i>	/ma.lah. tana ² /
<i>string beans</i>	<i>sitaw</i>	<i>blangol</i>	/blangol/
<i>chili</i>	<i>sili</i>	<i>malah</i>	/ma.lah/
<i>bitter melon or bitter gour</i>	<i>ampalaya</i>	<i>falyà</i>	/fal.ya ² /
<i>taro</i>	<i>gabi</i>	<i>kilong</i>	/ki.long/
<i>moringa</i>	<i>malunggay</i>	<i>kalmunggay</i>	/kal.munggay/
<i>tomato</i>	<i>kamatis</i>	<i>kalmati</i>	/kal.matih/
<i>round eggplant</i>	<i>bilog na talong</i>	<i>sgatong/bukuk</i>	/sgatong/, /bukuk/
<i>bottle gourd</i>	<i>upo</i>	<i>kabuk</i>	/labuk/
<i>malabar spinach</i>	<i>alugbati</i>	<i>lagbati</i>	/lag.ba.ti/

TABLE 3. Types of Weather

English	Filipino	Blaan	Phonetic Transcription
<i>rainy</i>	<i>maulan</i>	<i>mulén</i>	/mu.lən/
<i>storm</i>	<i>bagyo</i>	<i>bagiyo</i>	/ba.gi.yoh/
<i>summer</i>	<i>tag-init</i>	<i>minit</i>	/minit/
<i>windy</i>	<i>mahangin</i>	<i>nosh</i>	/nosh/
<i>dusk or twilight</i>	<i>takipsilim</i>	<i>kmimel</i>	/kmi.mel/
<i>lightning</i>	<i>kidlat</i>	<i>smilà</i>	/smila ² /
<i>dawn</i>	<i>bukangliwayway</i>	<i>mtek faya</i>	/mtek.fayah/
<i>thunder</i>	<i>dagundong</i>	<i>lugung</i>	/lugung/
<i>morning</i>	<i>umaga</i>	<i>alnilaw nando</i>	/alnilaw.nan.doh/
<i>bright moon</i>	<i>maliwanag ang buwan</i>	<i>mnéng bulén</i>	/mneng bulen/
<i>morning</i>	<i>umaga</i>	<i>imabun</i>	/i.ma.bun/
<i>starry</i>	<i>mabituin</i>	<i>blatik</i>	/bla.tik/
<i>cloudy</i>	<i>maulap</i>	<i>labun</i>	/la.bun/

<i>northeast wind</i>	amihan	<i>dangan kamtu</i>	/da.ngan.kam.tu/
<i>rainbow</i>	bahaghari	<i>bluntu</i>	/blun.tuh/
<i>planting</i>	magtanim	<i>dangan mlah</i>	/da.ngan. mlah/
<i>drizzle</i>	ambon	<i>talmitik</i>	/tal.mi.tik/

TABLE 4. Numerals

English	Filipino	Blaan	Phonetic Transcription
<i>one</i>	isa	<i>satu</i>	/satuh/
<i>two</i>	dalawa	<i>lwe</i>	/lweh/
<i>three</i>	tatlo	<i>tlu</i>	/tluh/
<i>four</i>	apat	<i>fat</i>	/fat/
<i>five</i>	lima	<i>lima</i>	/limah/
<i>six</i>	anim	<i>nem</i>	/nəm/
<i>seven</i>	pito	<i>fitu</i>	/fituh/
<i>eight</i>	walo	<i>walu</i>	/waluh/
<i>nine</i>	siyam	<i>syam</i>	/syam/
<i>ten</i>	sampu	<i>stafalò</i>	/sta.fa.loʔ/
<i>twenty</i>	dalawampu	<i>lwefalò</i>	/lwe. fa.loʔ/
<i>one hundred</i>	isang daan	<i>mlatuh</i>	/mlatuh/
<i>one thousand</i>	isang libo	<i>mlibu</i>	/mlibuh/
<i>one million</i>	isang milyon	<i>sakati</i>	/saka.ti/

TABLE 5. Types of Colors

English	Filipino	Blaan	Phonetic Transcription
<i>red</i>	pula	<i>fulé</i>	/fule/
<i>orange</i>	kahel	<i>malung mlalel</i>	/malung.mla.lel/
<i>bughaw</i>	asul	<i>mlunu</i>	/mlunuh/
<i>brown</i>	kayumanggi	<i>bulaw</i>	/bu.law/
<i>yellow</i>	dilaw	<i>mlalel</i>	/mla.lel/
<i>milky white</i>	kulay gatas	<i>malung bukay</i>	/malung.bukay/
<i>black</i>	itim	<i>fitém</i>	/fitem/
<i>white</i>	puti	<i>bukay</i>	/bukay/
<i>green</i>	berde	<i>malung mlunu</i>	/malung.mlunuh/

TABLE 6. Times (based on *sunrise-du*)

English	Filipino	Blaan	Phonetic Transcription
<i>7:00 am</i>	alas 7 n.u	<i>fuksutdu</i>	/fuk.sut.duh/
<i>10:00 am</i>	10:00 n.u	<i>tnga mnagéd</i>	/tnga mnagəd/
<i>12:00 midnight</i>	12:00 hatinggabi	<i>nga buteng</i>	/nga bu.teng/
<i>1:00-4:00 pm</i>	1:00-4:00 n.h	<i>muyon faya</i>	/mu.yon fayah/
<i>12:00 noon</i>	12:00 tanghali	<i>lturu</i>	/lturuh/
<i>night</i>	gabi	<i>kifuh</i>	/ki.fuh/
<i>1:00 pm</i>	ala 1:00 n.h.	<i>amgung masol</i>	/amgung masol/
<i>morning</i>	umaga	<i>mtek fayah</i>	/mtek fayah/
<i>5:00 pm</i>	5:00 n.h.	<i>sdefdo</i>	/sdef.doh/
<i>6:00 pm</i>	6:00 n.g.	<i>kmimét</i>	/kmimet/

TABLE 7. Characteristics			
English	Filipino	Blaan	Phonetic Transcription
<i>kind / good-natured</i>	mabait	<i>fye balu</i>	/fyeɸ baluh/
<i>brave / courageous</i>	matapang	<i>bung labé</i>	/bung labéʔ/
<i>respectful / polite</i>	magalang	<i>mafé</i>	/mafɛɸ/
<i>strong</i>	malakas	<i>kamgis</i>	/kam.gis/
<i>helpful / supportive</i>	matulungin	<i>tmabeng</i>	/tma.bəng/
<i>weak / frail</i>	mahina	<i>mlungey nawa</i>	/mlu.ngey nawah/
<i>generous</i>	mapagbigay	<i>fulung mle</i>	/fulung mleɸ/
<i>agile / nimble</i>	maliksi	<i>mabaltik/malgaan</i>	/ma.bal.tik/, mal.gaan/
<i>clean / neat</i>	malinis	<i>mlaneb/mangtaf</i>	/mla.neb/, mang.taf/
<i>slow / sluggish</i>	mabagal	<i>mlinay/leet</i>	/mli.nay/, /leʔet/
<i>shy / timid</i>	mahiyaín	<i>too 'mya</i>	/to.o ʔmyah/
<i>arrogant / boastful</i>	mayabang	<i>fagsék</i>	/fagsek/
<i>rude / ill-mannered</i>	bastos	<i>lamafe</i>	/la.ma.fe/
<i>religious / devout</i>	relihiyoso	<i>bong kafloo</i>	/bong kaf.loo/
<i>dirty / messy</i>	madungis	<i>landé sik</i>	/landeʔ sik/
<i>obedient / compliant</i>	masunurin	<i>mimén</i>	/mimen/
<i>honest / truthful</i>	matapat	<i>faglut</i>	/fag.lut

TABLE 8. Terms Related to Marriage			
English	Filipino	Blaan	Phonetic Transcription
<i>dowry</i>	dote / ari-arian	<i>sunggud</i>	/sung.gud/
<i>dance</i>	para sa kasal		
<i>courtship visit /</i>	sayaw	<i>maral</i>	/maral/
<i>formal visit of the</i>	pamanhikan	<i>falsut</i>	/fal.sut/
<i>suitor to the bride's</i>			
<i>family</i>			
<i>coconut toddy /</i>	tubaw	<i>utob</i>	/ʔu.tob/
<i>fermented coconut</i>			
<i>sap</i>			
<i>Kris / wavy dagger</i>	kris	<i>fais</i>	/faʔis/
<i>clothes / clothing</i>	damit	<i>albong</i>	/al.bong/
<i>guitar</i>	gitara	<i>faglong blaán</i>	/fag.long blaʔan/
<i>dance</i>	sayaw	<i>sayéw</i>	/sayew/
<i>string (e.g., musical</i>	kord	<i>sansikit</i>	/san.si.kit/
<i>instrument string)</i>			
<i>comb / hair ornament</i>	suklay/palamuti sa buhok	<i>swat</i>	/swat/
<i>sword</i>	espada	<i>sundeng</i>	/sun.dəng/
<i>Kris without</i>	kris na walang buhok ng	<i>kabasi</i>	/ka.ba.si/
<i>horsehair (specific</i>	kabayo		
<i>type of kris)</i>			

TABLE 9. Superstitions/Beliefs			
English	Filipino	Blaan	Phonetic Transcription
<i>bountiful harvest /</i>	masaganang ani	<i>sabak</i>	/sa.bak/
<i>abundant harvest</i>			
<i>ritual / ceremony</i>	ritwal	<i>malém</i>	/ma.ləm/
<i>warning / caution</i>	babala	<i>almugen</i>	/al.mugən/

<i>call / song of the Alimukon bird</i>	(huni ng ibon Alimukon)		
<i>conceived / pregnancy begins</i>	naglihi	<i>kalmien</i>	/kal.mi'en/
<i>evil eye / curse / bewitchment</i>	usog/buyag	<i>taban</i>	/ta.ban/
<i>lord / master / deity</i>	panginoon	<i>dwata</i>	/dwa.tah/
<i>deity / nature spirit / fairy</i>	diwata	<i>meleb</i>	/me.ləb/
<i>conception / early pregnancy</i>	paglihi	<i>fandita</i>	/fan.di.tah/
<i>curse / spell to prevent earthquakes</i>	sumpa para masugpo ang lindol	<i>asnumfa</i>	/as.num.fah/

TABLE 10. Types of Food

English	Filipino	Blaan	Phonetic Transcription
<i>banana heart cooked in coconut milk</i>	ginataang ubod ng saging	<i>gnatà bulig</i>	/gna.ta' bulig/
<i>cooked in broth / soup-based dish</i>	sinabawan	<i>snabéw</i>	/snabəw/
<i>coconut heart / coconut pith</i>	ubod ng niyog	<i>ubol</i>	/ʔu.bol/
<i>broth / soup</i>	sabaw	<i>sabéw</i>	/sabəw/
<i>grilled native chicken / roast chicken</i>	inihaw na native na manok	<i>dnaléng anuk</i>	/dnaleng a.nuk/
<i>fried / deep-fried</i>	prito	<i>sansal</i>	/san.sal/
<i>cooking rice in bamboo / bamboo-cooked rice</i>	pagsaing sa kawayan	<i>lmutut</i>	/lmu.tut/

TABLE 11. Terms for Kasfalah/Agreement of Settlement

English	Filipino	Blaan	Phonetic Transcription
<i>mediator / arbitrator</i>	tagapamitan	<i>samfalah</i>	/sam.falah/
<i>conflict / dispute</i>	hidwaan	<i>dmumut</i>	/dmu.mut/
<i>council of elders</i>	lupon ng matatanda	<i>dadfulong</i>	/dad.fulong/
<i>paid fine / penalty given</i>	multa na binigay	<i>fné</i>	/fnə'/
<i>adultery / wife-stealing / marital dispute</i>	pag-agaw sa asawa	<i>mled</i>	/mləd/
<i>killing / homicide / murder</i>	patayan	<i>stafi</i>	/sta.fih/
<i>double return of dowry</i>	dobleng pagbalik ng dote	<i>fulè sungud</i>	/fu.le' sung.gud/
<i>land dispute / land grabbing</i>	agawan sa lupa	<i>kasfalah</i>	/kas.falah/
<i>family quarrel / domestic dispute</i>	awayan sa pamilya	<i>santulen</i>	/san.tulen/
<i>resentment / grudge</i>	sama ng loob	<i>kasdumot</i>	/kas.dumot/

TABLE 12. Methods of Healing			
English	Filipino	Blaan	Phonetic Transcription
<i>toothache</i>	sakit sa ngipin	<i>do a navo</i>	/do a na.voh/
<i>headache</i>	sakit sa ulo	<i>ika at afol</i>	/i.ka at afol/
<i>stomachache / abdominal pain</i>	sakit sa tiyan	<i>fdu sawa</i>	/fdu sawah/
<i>praying to the spirits / ritual prayer to the spirits</i>	pagdasal sa mga mele	<i>fun kayo (puno)</i>	/fun ka.yoh/
		<i>fun e'el (sapa)</i>	/fun ʔe'el/
		<i>fun bulul (gubat)</i>	/fun bu.lul/
		<i>fun nos (hangin)</i>	/fun nos/
<i>witchcraft / sorcery / hex</i>	kulam	<i>kambong</i>	/kam.bong/

The Blaan language like other languages has its own distinct sounds that identify their unique characteristics. Due to the frequent use of these letters as well as the pronunciation of words, their identity as Blaan is recognized.

Some of the letters and sounds, as well as words that are considered unique to the Blaan language, are the following:

SCHWA SOUND /ə/

In phonetics, especially in phonology, the schwa is a central vowel sound located at the center of the vowel chart (mid-central vowel), represented by the symbol ə in the IPA. In Filipino orthography, it is symbolized by the diacritical mark E (ë) and can be found in languages like Ilokano, Mëranaw, Kankana-ey, and others.

The Blaan language also exhibits this, with some of their cultural terms containing the schwa sound, as seen in the following examples:

B: <i>klubëng</i> /klu.bəng/	B: <i>ëmak</i> /ə.mak/	B: <i>mulën</i> /mu.lən/
F: <i>pilikmata</i>	F: <i>kilikili</i>	F: <i>maulan</i>
E: <i>eyelashes</i>	E: <i>underarm</i>	E: <i>rainy</i>
B: <i>nëm</i> /nəm/	B: <i>tmabëng</i> /tma.bəng/	B: <i>almugën</i> /al.mugən/
F: <i>anim</i>	F: <i>matulungin</i>	F: <i>babala</i>
E: <i>six</i>	E: <i>helpful</i>	E: <i>warning</i>
B: <i>melëb</i> /me.ləb/	B: <i>snabëw</i> /snabəw/	B: <i>sabëw</i> /sabəw/
F: <i>diwata</i>	F: <i>sinabaw</i>	F: <i>sabaw</i>
E: <i>faïry</i>	E: <i>cooked in broth</i>	E: <i>broth</i>
B: <i>f në</i> /fənəʔ/	B: <i>mlëd</i> /mləd/	
F: <i>multa</i>	F: <i>pag-agaw sa asawa</i>	
E: <i>fine</i>	E: <i>adultery</i>	
B: <i>mnëng bulën</i> /mnəng.bulən/		
F: <i>maliwanag ang buwan</i>		
E: <i>bright moon</i>		

It can be observed in the listed words that the schwa sound is a significant vowel sound in the Blaan language, and it can be found in three positions: initial, medial, and final.

PHONEME /f/

Many words in the Blaan language contain the phoneme /f/, which is rarely heard in other languages in the Philippines. Like in Tagalog, the phoneme /f/ is either absent or not used, so the /f/ in Blaan can be considered a

peculiarity that is part of their identity. The following are examples of Blaan cultural terms that feature the phoneme /f/.

B: <i>kifen</i> /kifen/	B: <i>fifi</i> /fi.fi/	B: <i>fulek</i> /fu.lek/
F: <i>ngipin</i>	F: <i>pisngi</i>	F: <i>kilay</i>
E: <i>teeth</i>	E: <i>cheek</i>	E: <i>eyebrows</i>
B: <i>kfu</i> /kfu/	B: <i>fued</i> /fu ² ed/	B: <i>falel</i> /falel/
F: <i>dibdib</i>	F: <i>paa</i>	F: <i>palad</i>
E: <i>chest</i>	E: <i>foot</i>	E: <i>palm</i>
B: <i>falel bli</i> /falel.bli ² /	B: <i>falya</i> /fal.ya ² /	B: <i>mtek faya</i> /mtek.fayah/
F: <i>talampakan</i>	F: <i>ampalaya</i>	F: <i>bukangliwayway</i>
E: <i>sole</i>	E: <i>bitter gourd</i>	E: <i>dawn</i>
B: <i>fat</i> /fat/	B: <i>fitu</i> /fituh/	B: <i>stafalo</i> /sta.fa.lo ² /
F: <i>apat</i>	F: <i>pito</i>	F: <i>sampu</i>
E: <i>four</i>	E: <i>seven</i>	E: <i>ten</i>
B: <i>lwefalo</i> /lwe. fa.lo ² /	B: <i>fule</i> /fule/	B: <i>fitem</i> /fitem/
F: <i>dalawampu</i>	F: <i>pula</i>	F: <i>itim</i>
E: <i>twenty</i>	E: <i>red</i>	E: <i>black</i>
B: <i>futsutdu</i> /fuk.sut.duh/	B: <i>kifuh</i> /ki.fuh/	B: <i>mtek fayah</i> /mtek.fayah/
F: <i>ala 7 n.u.</i>	F: <i>gabi</i>	F: <i>umaga</i>
E: <i>7:00 am</i>	E: <i>night</i>	E: <i>morning</i>
B: <i>sdefdo</i> /sdef.doh/	B: <i>fye balu</i> /fyeh.baluh/	B: <i>mafe</i> /mafeh/
F: <i>5 n.h.</i>	F: <i>mabait</i>	F: <i>magalang</i>
E: <i>5:00 pm</i>	E: <i>kind / good-natured</i>	E: <i>respectful</i>
B: <i>fulung mle</i> /fulung.mleh/	B: <i>mangtaf</i> /mang.taf/	B: <i>fagsek</i> /fagsek/
F: <i>mapagbigay</i>	F: <i>malinis</i>	F: <i>mayabang</i>
E: <i>generous</i>	E: <i>clean/ neat</i>	E: <i>arrogant / boastful</i>
B: <i>lamafe</i> /la.ma.fe/	B: <i>bong kaflo</i> /bong.kaf.loo/	B: <i>faglut</i> /fag.lut/
F: <i>bastos</i>	F: <i>relihiyoso</i>	F: <i>matapat</i>
E: <i>rude</i>	E: <i>religious</i>	E: <i>honest</i>
B: <i>falsut</i> /fal.sut/	B: <i>fais</i> /fa ² is/	B: <i>faglong blaan</i> /fag.long bla ² an/
F: <i>pamanhikan</i>	F: <i>Kris</i>	F: <i>gitara</i>
E: <i>courtship visit</i>	E: <i>Kris/ sword</i>	E: <i>guitar</i>
B: <i>fandita</i> /fan.di.tah/	B: <i>asnumfa</i> /as.num.fah/	B: <i>samfala</i> /sam.falah/
F: <i>paglihi</i>	F: <i>pagsugpo</i>	F: <i>tagapamagitan</i>
E: <i>conception</i>	E: <i>suppression</i>	E: <i>mediator</i>

The presented Blaan cultural terms containing the phoneme /f/ can be found in three positions: initial, medial, and final. However, the phoneme /f/ is most commonly found in the initial position.

PHONEME /v/

The Blaan cultural term used in healing that contains the phoneme /v/ is found in the phrase *do a navo /do a na.voh/*, which means toothache. This term refers specifically to the healing process for the teeth. It can be said that every language has its own unique set of sounds or phonemes, that are distinct to that language. The presence of /v/ in Blaan, typically found only in foreign languages, illustrates this uniqueness and underscores the distinctiveness of Filipino as the national language.

CLUSTER

Based on the collected cultural terms of the Blaan, there are also words that contain consonant clusters, such as:

/fn/	/tm/
B: fne /fnəʔ/ F: multa na ibinigay E: <i>fine</i>	B: tmabeng /tma.bəŋg/ F: matulungin E: <i>helpful</i>
/fd/	/tl/
B: fdu sawa /fdu.sawah F: sakit sa tiyan E: <i>stomachache</i>	B: tlu /tluh/ F: tatlo E: <i>three</i>
/lm/	/kw/
B: lmutut /lmu.tut/ F: pagsaing sa kawayan E: <i>bamboo-cooked rice</i>	B: kwang /kwang/ F: balakang E: <i>hips</i>
/ml/	/kf/
B: mlinay /mli.nay/ F: mabagal E: <i>slow</i>	B: kfu /kfu/ F: dibdib E: <i>chest</i>
/kl/	
B: klinge /kli.nge/ F: tainga E: <i>ear</i>	B: klubeng /klu.bəŋg/ F: pilikmata E: <i>eyelashes</i>
/bl/	
B: blagel /blagel/ F: balik E: <i>shoulder</i>	B: blutut /blu.tut/ F: pusod E: <i>belly button</i>
B: blatik /bla.tik/ F: bituin E: <i>star</i>	B: bluntu /blun.tuh/ F: bahaghari E: <i>rainbow</i>
B: blangol /blangol/ F: sitaw E: <i>string beans</i>	

/tn/

B: **tnugu** /*tnuguh*/
F: kuko
E: *nail*

/fy/

B: **fye** balu /*fye baluh*/
F: mabait
E: *good/ kind*

/kn/

B: **k'namet** /*kənamet*/
F: kamay
E: *hand*

/lt/

B: **lturu** /*lturuh*/
F: 12 ng tanghali
E: *12:00 noon*

/sg/

B: **sgatong** /*sgatong*/
F: talong
E: *eggplant*

/sd/

B: **sdefdo** /*sdef.doh*/
F: 5 ng hapon
E: *5:00 pm*

/sh/

B: **nosh** /*nosh*/
F: mahangin
E: *windy*

/sm/

B: **smila** /*smila*²/
F: kidlat
E: *lightning*

/km/

B: **k'mimel** /*kəmi.mel*/
F: takipsilim
E: *dusk / twilight*

B: **k'mimet** /*kəmimet*/
F: 6 ng gabi
E: *6:00 pm*

/mt/

B: **m'tek** faya /*mətek.fayah*/
F: bukangliwayway
E: *dawn*

/mn/

B: **m'nəng** bulən /*mənəng bulen*/
F: maliwanag ang buwan
E: *bright moon*

/lw/

B: **lwe** /*lweh*/
F: dalawa
E: *two*

/my/

B: too **mya** /*to.o ʔmyah*/
F: mahiyain
E: *shy*

/sw/

B: **swat** /*swat*/
F: suklay
E: *comb*

The consonant clusters found in the cultural terms of the Blaan are /bl/, /fn/, /fd/, /fy/, /lm/, /lt/, /lw/, /tm/, /tl/, /kw/, /kf/, /kl/, /km/, /kn/, /my/, /ml/, /mt/, /mn/, /sd/, /sh/, /sm/, /sw/, and /sg/, which are found only in the initial position.

In linguistics, a cluster refers to two or more adjacent consonants within a syllable of a word. This is similar to consonant clusters in Tagalog (derived from the words "kambal" or "two" and "katinig" or "consonant"). However, in

Tagalog, only two adjacent consonants within a syllable are considered a cluster. There are no more than three adjacent consonants in a syllable in Tagalog, unlike in Filipino, where having three to four consonants within a syllable is possible.

DIPHTHONG

A diphthong is any vowel that is followed by a semi-vowel /y/ or /w/ within a syllable. The diphthongs in Filipino are /iw/, /iy/, /ey/, /oy/, and /uy/.

There are also diphthongs found in the cultural terms of Blaan, such as:

/ay/

B: kalamunggay /*kal.munggay*/

B: malung bukay /*malung.bukay*/

F: malunggay

F: kulay gatas

E: *moringa*

E: *milky white*

B: m'linay /*məli.nay*/

F: mabagal

E: *slow*

/ey/

B: m'lungey nawa /*məlu.ngey nawah*/

F: mahina

E: *weak*

/aw/

B: almilaw nando /*almilaw.nan.doh*/

B: bulaw /*bu.law*/

F: umaga

F: kayumanggi

E: *morning*

E: *brown*

/ew/

B: sayew /*sayew*/

B: s'nabew /*sə nabəw*/

F: sayaw

F: sinabawan

E: *dance*

E: *cooked in broth*

B: sabew /*sabəw*/

F: sabaw

E: *soup*

Based on the presented cultural terms of the Blaan, the diphthongs that can be found are /ay/, /ey/, /aw/, and /ew/. When these words are syllabified, they become “ka/la/mung/gay” /*kal.munggay*/, “ma/lung/bu/kay” /*malung.bukay*/, “m’li/nay” /*məli.nay*/, “m’lu/ngey/na/wa” /*məlu.ngey nawah*/, “al/mi/law/nan/do” /*almilaw.nan.doh*/, “bu/law” /*bu.law*/, “sa/yew” /*sayew*/, “s’na/bew” /*sə nabəw*/, “sa/bew” /*sabəw*/ . These specific diphthongs in Blaan are only found in the final position.

DIPHTHONG GLIDE

A diphthong glides in a monosyllabic word, starting with the pronunciation of one vowel, moving toward the position of another vowel, and is often written with two letters representing the beginning and the end of the glide.

Some of the cultural terms of the Blaan also exhibit diphthongs, as follows:

/ie/

B: tien /*ti'en*/

B: liel /*li'el*/

B: kalmien /*kal.mi'en*/

F: tiyan	F: leeg	F: naglihi
E: <i>stomach</i>	E: <i>neck</i>	E: <i>conceived</i>

/ee/

B: fun eel /*fun ʔeʔel*/

F: sapa

E: *creek / stream*

/oo/

B: too mya /*to.o ʔmyah*/

F: mahiyain

E: *shy*

The diphthongs found in the cultural terms of the Blaan include /ie/ and /oo/. The diphthong /ie/ may be influenced by Spanish suggesting that the Spanish influence extended to the southern part of the island of Mindanao.

Meanwhile, the diphthongs /oo/ and /ee/ are adjacent and not separated by a hyphen. According to the Blaan Orthography, when spelling words that have identical vowels within the syllable, like /oo/, no hyphen is used as it would only cause confusion.

If the words were spelled out, they would become “*tien*” /*tiʔen*/, “*too/ mya*” /*to.o ʔmyah*/, “*kal/mien*” /*kal.miʔen*/, “*fun/ eel*” /*fun ʔeʔel*/, “*liel*” /*liʔel*/ . Notice that the diphthongs /ie/, /ee/, and /oo/ are within the syllable.

SILENT SOUND

The Blaan language also has words that contain silent sounds or sounds that are not pronounced. An example of a cultural term that exhibits this characteristic is “*kfu*” /*fuh*/, which means chest. It can be observed that the letter /k/ appears when written, but when the word is pronounced, it becomes /*fuh*/ without the /k/ sound being pronounced. Similarly, in the word “*lmumut*” /*mu.mut*/, which means cooking rice in bamboo, it can be observed that the letter /l/ is present, but when the word is pronounced, it becomes /*mu.mut*/ only.

The silent sounds or unpronounced sounds in the mentioned cultural terms of Blaan occur only in the initial position of the word.

MONOSYLLABIC WORDS

Monosyllabic words are words that consist of only one syllable but are considered meaningful. In the cultural terms of the Blaan language, such monosyllabic words can also be observed, such as:

B: liel / <i>liʔel</i> /	B: tien / <i>tiʔen</i> /	B: kfu / <i>kfu</i> /
F: leeg	F: tiyan	F: dibdib
E: <i>neck</i>	E: <i>stomach</i>	E: <i>chest</i>
B: ki / <i>kiʔ</i> /	B: nosh / <i>nosh</i> /	B: du / <i>duh</i> /
F: ari ng babae	F: mahangin	F: araw
E: <i>female genitalia</i>	E: <i>windy</i>	E: <i>morning</i>
B: fat / <i>fat</i> /	B: nem / <i>nəm</i> /	B: syam / <i>syam</i> /
F: apat	F: anim	F: siyam
E: <i>four</i>	E: <i>six</i>	E: <i>nine</i>
B: tʼlu / <i>tluh</i> /	B: swat / <i>swat</i> /	
F: tatlo	F: suklay	
E: <i>three</i>	E: <i>comb</i>	

It is rare for a language to have monosyllabic words or single-syllable words that carry a complete meaning or serve as root words. In the cultural terms of the Blaan language, some examples that exhibit this linguistic feature are *liel* /li²el/, *tien* /ti²en/, *kfu* /kfu²h/, *ki* /ki²/, *nosh* /nosh/, *du* /duh/, *fat* /fat/, *nem* /nəm/, *syam* /syam/, *tlu* /tluh/, and *swat* /swat/, which proves that the Blaan language is truly unique.

THE CULTURAL MEANINGS OF THE TERMS IN THE BLAAN LANGUAGE

Almugen

The Almugen bird is believed to signal danger through its call. If they hear its call, they will no longer proceed with their journey. Also, they will not immediately continue their journey if, upon leaving, they suddenly remember something they forgot at home and have to return. After going back, they will wait for a few minutes before setting out again to ward off bad luck or to avoid misfortune.

nga'

Their dog, called 'ye'm', accompanies them during hunting because it serves as their guide in the forest, and they believe it can see evil spirits or 'meleh'. Its barking is believed to drive these spirits away.

All young animals or offspring are referred to as 'nga'. For example: piglet = 'nga sde', foal (young horse) = 'nga' kura'.

ki

They are very sensitive when it comes to mentioning the private parts of men and women. They consider it vulgar or insulting if these are mentioned casually.

The word 'kt', which means female genitalia is forbidden to be spoken in society as it is deemed indecent. Even women themselves are not allowed to mention it, even in ordinary conversations or discussions.

Numbers

They only have specific terms for the numbers 1–10, 100, 1,000, and 1 million. For numbers like eleven, their way of counting is by combining 10 + 1. They simply join the terms together. For example, 11 = *stafalo satu*. 'Stafalo' means 10 and 'satu' means 1.

For 20, they blend the term for 2 and attach it to the term for ten, with some reduction. They combine the two words. For example, 20 = *lwefalo*, where 'lwe' means 2 and 'falo' comes from 'stafalo' (10), with 'sta' omitted, leaving 'falo' to be attached to 'lwe'.

For 21, it becomes *lwefalo satu*. This pattern continues up to 29. For 30, it becomes *tlufalo*, where 'tlu' means 3 and is joined with 'falo'.

20 = <i>lwefalo</i>	50 = <i>limafalo</i>	80 = <i>walufalo</i>
30 = <i>tlufalo</i>	60 = <i>nemfalo</i>	90 = <i>syamfalo</i>
40 = <i>fatfalo</i>	70 = <i>fitufalo</i>	100 = <i>mlatuh</i>

They write numbers using straight lines enclosed in a box. For example, 23 is written as IIII (four straight lines like that) and three straight lines.

CONDUCTING THE COURTSHIP (FALSUT) AND MARRIAGE (SONG)

They no longer practice courtship. When a young man falls for a maiden, he will inform his parents that he has someone in mind and wishes to go to her house for the courtship. If his parents agree, the young man will go to the maiden's house and ask her parents for permission to marry their daughter. If the parents agree, they will then state the *sunggud* (dowry) that the young man needs to give.

The young man will then go back to his parents and tell them what *sunggud* is being requested. If his parents can provide the *sunggud* in full, the wedding will take place on the same day as the *sunggud* presentation. In other words, the wedding date depends on the *sunggud* that the young man's family can provide.

The *Bung Fulong* will perform the wedding ceremony. As part of the ceremony, the man will place one of his feet on the woman's head, symbolizing that the woman should not be prideful towards the man.

The sign that they are officially married is not a kiss, but the touching of their foreheads, which is the sign that they are now husband and wife.

It is also possible for both families to agree that their children will marry each other (*alnabu=buya*), assuming they are a boy and a girl. If they are both of the same gender, the agreement is directly canceled (*palso*).

If the young man and woman run away together (*astangal*), the *sunggud* still stands, and if they return or come back, the young man's family will provide it. If they fail to give the *sunggud*, the young man will be killed.

The child of a *Fulong* cannot marry anyone who is not also a child of a *Fulong*.

BELIEF IN *SANDUGO* OR ADOPTION OF THOSE WHO WISH TO BECOME PART OF THE ETHNOLINGUISTIC GROUP (*SADYANDI*)

All the *Fulong* (tribal leaders or chiefs) from each town or barangay will be called to serve as the primary witnesses for the *sandugo* that will take place. The *Fulong* of the barangay will present the person who will be adopted by their ethnolinguistic group.

The ritual involves immersing a metal or iron object in water, which will be held by the *Fulong* and the one to be adopted. This symbolizes that their brotherhood is strengthened or solidified because of the ritual. Only then will the *fulong* say, 'He is now our brother in the ethnolinguistic group; let us accept him and treat him as a sibling.'

BELIEF IN *KASFALAH* OR AGREEMENT (*AREGLO*)

Bung Fulong will conduct or mediate the agreement and it will take place at his own house. The *Fulong*s from each barangay will be called as primary witnesses.

In the agreement, the one who called for it will state the fine or compensation they are demanding from the person being complained about. This depends on the severity of the case or the wrongdoing committed by the one being complained about.

If the family of the person being complained about is unable to provide the fine, *Fulong* will help find a way for it to be provided. The *kasfalah* will only be completed once the fine (*fne*) is given. Even if it takes days or weeks, the parts of the *kasfalah* will remain in place. *Fulong* will provide food for everyone.

If the case involves a wife running off with another man, the man she ran off with will give double the fine (*fule sunggud*). Similarly, if the wife leaves or rejects her husband (*bana*), the fine will also be returned doubled to the husband. If it is the man who refuses, no fine will be given. This means it will be considered (*kwits*).

If the man the wife ran off with does not give or is unable to provide double the fine, he will be killed.

THE *DE JURE* LAW BEHIND THE KRIS (*FAIS*)

De jure refers to rules or policies that are considered legal or mandatory for everyone to follow; these are grounded in law and often passed down within the family, such as to the head or *Fulong*.

The *de jure* refers to rules or policies that are considered legal or should be followed by all, in accordance with the law passed down to family members such as the leader (*Fulong*). In the Blaan society, one of these rules involves the role of the leader (*Fulong*), who alone has the right to use a kris (*fais*) with a horse's tail. No one is allowed to attach a horse's tail to their kris (*fais*) except the leader (*Fulong*) of the ethnolinguistic group. Attaching a horse's tail to the kris (*fais*) of non-leaders or ordinary citizens is a violation of the group's laws.

Another law that must be followed in Blaan culture is the marriage of a *Fulong*'s child. Only the child of a *Fulong* can marry the child of another *Fulong*.

If there is an occurrence of *astangal* (elopement) between a maiden and a young man, the dowry (*sunggud*) that the woman's parents requested will remain in place. If the elopers return home, the family of the man must provide the dowry, and if they fail to do so, the man will be killed.

4. Discussion

The novelty of this study lies in its proposition that empirical ethnographic documentation must serve as the structural architecture for information systems involving marginalized languages. While current literature highlights the profound computational barriers, algorithmic biases, and risks of data colonialism inherent in training AI on

underrepresented groups (Ketere, 2025; Kwok, 2026), this research provides a tangible dataset to overcome these hurdles. By explicitly linking Blaان sociolinguistic phenomena with digital data governance, the study contributes a culturally grounded blueprint essential for accountable AI design (Shi & DiFranzo, 2026).

Standard NLP models often impose rigid computational templates that fail to capture indigenous linguistic fragmentation and orthographic variations, severely limiting the efficacy of educational and communicative technologies (Ketere, 2025). The Blaان dataset reveals complex linguistic features that challenge standard western-centric text processing. These include the prominent schwa sound (/ə/), unique consonant clusters (e.g., /bl/, /fn/, /sg/, /kw/), and specific silent letters found in terms like *kfu* (/fuh/) and *lmumut* (/mu.mut/). For ICT integration to be effective, machine learning algorithms must be customized to recognize these exact phonetic transcriptions and multiple consonant configurations occurring at the initial stages of word processing. Documenting these precise features establishes the empirical baseline required to train bias-reducing algorithms, ensuring the digital representation of the Blaان language is structurally sound and linguistically authentic (Gonzales, 2003).

A major contribution of this study to information governance is the translation of cultural taboos into digital privacy protocols. The findings reveal that certain Blaان terminologies, such as *ki* (female genitalia), are considered highly indecent and are strictly forbidden from casual conversation, even among women in the community. In a digital repository, educational app, or AI application, exposing such terms indiscriminately constitutes an ethical breach and inflicts cultural harm (Calac & Mackey, 2025). Therefore, the system architecture must implement Role-Based Access Control (RBAC) mechanisms. By establishing strict data boundaries informed directly by ethnographic research (Shi & DiFranzo, 2026), database administrators can ensure that sensitive data is preserved securely in the backend for historical archiving but restricted from public-facing NLP translations. This approach directly embeds the community's sociocultural protocols into the system's cybersecurity framework.

Finally, this study aligns traditional community governance with Indigenous Data Sovereignty (IDS), which asserts the inherent right of indigenous populations to govern and steward their own data (Kukutai & Walter, n.d.). In Blaان society, *de jure* laws mandate that only the *Fulong* (tribal leader) holds the authority to mediate high-level agreements (*kasfalah*) and wield specific symbols of power, such as the *fais* (kris) adorned with a horse's tail. Translating this societal structure into an ICT framework requires a decentralized data governance model where digital administrative rights mirror the tribal hierarchy. The *dadfulong* (council of elders) must be integrated as the primary stewards of the database. By granting them the authority to determine what information is open-source and what remains protected, system developers ensure that the digital space is led by native language specialists (Jones et al., 2023) and that technological deployment empowers the community rather than facilitating unauthorized data extraction (Kwok, 2026).

5. Conclusion

The digitization of indigenous knowledge systems requires more than linguistic translation; it demands an ethical, culturally grounded data governance architecture. Integrating empirical ethnographic data directly into computational frameworks ensures that artificial intelligence and digital repositories protect against data colonialism while maintaining the structural integrity of the community's heritage. By embedding sociolinguistic phenomena and traditional societal laws as foundational dataset parameters, information systems can achieve both technical precision and ethical compliance.

To operationalize this framework, software developers and IT administrators must collaborate directly with local indigenous councils, granting them sovereign administrative control over their digitized cultural assets. Furthermore, developers are advised to engineer customized natural language processing algorithms capable of processing complex phonetic structures without defaulting to rigid, pre-existing computational templates. System architectures must concurrently implement strict Role-Based Access Control (RBAC) mechanisms to secure sensitive cultural terminologies from unauthorized public extraction.

Future research should focus on the practical deployment of these culturally responsive datasets through the development of bilingual digital applications or inclusive, learner-centered educational technologies. Additionally, scholars must investigate how integrating these ethical data governance models into broader pedagogical frameworks impacts indigenous cultural preservation and long-term digital sustainability.

References

1. Arcenas, M. (1993). *The Vanishing Nomads: The Blaans' Past, Present and Future*. Philippine Exchange Assistance Center Foundation.

2. Austro, L. (2011). Language Preservation in the Philippines: Issues and Perspectives. *Philippine Journal of Linguistics*, 19(3), 215-232.
3. Batanes, L. (1999). *Culture and Identity: The Social Context of Language in the Philippines*. University of the Philippines Press.
4. Bautista, M. (2005). *Linguistic Practices and Cultural Identity: Case Studies from the Blaan and Other Philippine Ethnic Groups*. University of Santo Tomas Publishing House.
5. Blount, B. (1977). *Culture and Language: A Study of the Blaan*. University of the Philippines Press.
6. Calac, A. J., & Mackey, T. (2025). A systematic review of responsible stewardship of research and health data from Indigenous communities. *npj Digital Medicine*.
7. Calinawagan, E. (2012). Diak maawatan: Unang sulap sa iluko varayti ng Filipino. *Salindaw: Varayti at baryasyon ng Filipino. Sentro ng Wikang Filipino: UP-Diliman*.
8. De Jong, R. (2009). The Bilaan Tribe of Southern Mindanao. *Things Asian*. <http://thingsasian.com/story/bilaan-tribe-southern-mindanao>
9. Du, N. (2014). Pagdalumat Sa Mga Sinaunang Di- Materyal Na Kultura Ng Mga B'laan Sa Brgy. Pisan, Kabacan, Cotabato Mula Sa Kanilang Kwentong Bayan. *Asia Pacific Journal of Multidisciplinary Research*, 2(6).
10. Gonzales, A. (2003). The Blaan Language: An Overview of Phonological Features. *Journal of Philippine Linguistics*, 10(2), 121-140.
11. Jocano, L. (1998). *Filipino Indigenous Ethnic Communities: Patterns, Variations, and Typologies*. Punlad Research House, Inc.
12. Jones, P.-L., Mahelona, K., Duncan, S., & Leoni, G. (2023). Kia tangata whenua: Artificial intelligence that grows from the land and people. *Ethical Space: The International Journal of Communication Ethics*.
13. Ketere, C. (2025). Decolonizing artificial intelligence: Reducing algorithmic biases in African educational and linguistic contexts. *Journal of Contemporary Issues in Open Distance and E-Learning*.
14. Kukutai, T., & Walter, M. (n.d.). Indigenous data sovereignty. In *The Data Journalism Handbook*.
15. Kwok, J. Y. (2026). Data colonialism and indigenous languages in AI: A critical review of existing initiatives and their struggles with data sovereignty. *AI & SOCIETY*.
16. Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.
17. Macatabon, R., & Calibayan, M. (2016). Varayti at Varyasyon ng Wikang Blaan sa Bacong, Tulunan, Hilagang Cotabato at Lampitak, Tampakan, Timog Cotabato, Philippines. *Asia Pacific Journal of Multidisciplinary Research*, 4(3). <http://www.apjmr.com/>
18. Rubrico, J. (2012). *Indigenization of Filipino: The case of the Davao City variety*. University of Malaya.
19. Salazar, Z. (1996). Ukol sa Wika at Kulturang Pilipino. In P. C. Constantino & M. M. Atienza (Eds.), *Mga Piling Diskurso sa Wika at Lipunan* (pp. 19-45).
20. Schmidt, M. (1993). *Etnolinguistics: The Study of Language and Culture in the Philippines*. National Bookstore.
21. Shi, H., & DiFranzo, D. (2026). Culturally-grounded governance for multilingual language models: Rights, data boundaries, and accountable AI design. *arXiv*.
22. Valera, T. (2002). Language and Ethnic Identity: Insights from the Blaan People. *Asian Ethnology Journal*, 15(1), 22-38.