

Kinship as Social Infrastructure: Address Terms and Cultural Relationality in Tetun Dili

Ujeius da Silva Soares^{1✉}, Adelcisa Alves Pereira¹

¹Department of Portuguese Language, Faculty of Education, Arts and Humanities,
Universidade Nacional Timor Lorosa'e, Dili, Timor-Leste

✉ ujeiusdasilvasoares.ok@gmail.com

doi: [10.65840/jllcd.v1i3.51](https://doi.org/10.65840/jllcd.v1i3.51)

Article Info

Submitted:

08/13/2024

Revised:

10/12/2024

Accepted:

11/25/2024

Available Online:

12/30/2024

Abstract. This article examines Tetun Dili address terms and kinship expressions as a cultural system of relational identity rather than as a simple vocabulary list. Based on a 37-entry documentary micro-corpus containing family terms, honorifics, professional titles, clan terms, ritual categories, and customary-governance expressions, the study combines data auditing, functional-domain coding, lexical-origin profiling, formality-gradient analysis, and cultural-pragmatic interpretation. The findings show that household and intimate kinship forms constitute the largest domain (29.7%), followed by clan, ritual, and ancestral authority (24.3%), honorific and professional address (21.6%), extended and affinal kinship (13.5%), community governance and customary law (8.1%), and symbolic social intimacy (2.7%). Indigenous Tetun forms dominate the corpus (56.8%), but Portuguese-derived and mixed forms reveal the continuing effects of colonial, religious, educational, and administrative contact. The study argues that Tetun Dili address practices function as social infrastructure: they organize intimacy, hierarchy, ritual authority, intergenerational belonging, and public respect in a multilingual postcolonial ecology. The article contributes a data-driven model for studying low-resource languages through relational lexicon, cultural pragmatics, and community-based linguistic documentation.

Keywords: Tetun Dili; address terms; kinship terminology; cultural pragmatics; language contact; Timor-Leste; relational identity; low-resource language



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0

[International License](https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Proverbs, address terms, kinship terms, honorifics, and ritual titles are often treated as small linguistic objects, but they carry unusually dense cultural work. In ordinary conversation, a single address form can do what a long explanation cannot: it can position the addressee as parent, elder, sibling, teacher, stranger, ritual authority, clan member, intimate friend, or social superior. For this reason, address terms should not be understood only as labels. They are relational instruments that transform grammar and vocabulary into social order. In a multilingual society such as Timor-Leste, the issue becomes even more important because Tetun Dili operates within a layered ecology of Indigenous linguistic inheritance, Portuguese colonial contact, Indonesian administrative memory, English globalization, Catholic institutional life, and local customary structures. The present article examines this ecology through a documentary micro-corpus of Tetun Dili address and kinship

expressions, asking how forms of address encode cultural relationality and how they reveal broader tensions between intimacy, respect, authority, and contact-induced change.

Recent work confirms that address and kinship terms remain productive entry points for studying identity, politeness, translation, social hierarchy, and intercultural pragmatics. Studies of Indonesian conversation show that address terms can manage participation and next-speaker selection, not merely name addressees (Hamdani, Barnes, and Blythe 2022; Blythe, Hamdani, and Barnes 2024). Translation studies show that address terms are often difficult to render because they encode interpersonal relations that do not travel cleanly across languages (Rozumko 2023). Comparative kinship studies similarly argue that kinship terminology indexes social reality, cultural categorization, and identity rather than only descent (Suryanarayan and Khalil 2021). In Southeast Asian and related contexts, current research on local address terms, language mixing, and politeness further shows that relational vocabulary is a site where locality, hierarchy, respect, age, gender, and institutional authority become linguistically visible (Gusanto, Budasi, and Dewi 2025; Wajdi et al. 2024; Zainal and Dewa 2022; Purba et al. 2024). These trends make Tetun Dili a compelling object for a contemporary language-and-culture article because its address system brings together kinship, ritual, colonial contact, national language policy, and low-resource documentation.

The topic is timely for three reasons. First, contemporary sociolinguistics has moved increasingly from static language description toward the study of how speakers accomplish identity, stance, hierarchy, solidarity, and belonging through interactional forms (Bucholtz and Hall 2005; Eckert 2012; Agha 2007; Silverstein 2003). Address terms are especially valuable in this shift because they are simultaneously lexical, pragmatic, and ideological. They can signal kinship, but they can also expand kinship beyond the household and into public life. They can index deference, but they can also soften distance, manage face, and make social inequality appear natural. Second, low-resource language research increasingly recognizes that documentation should not be limited to large corpora, machine translation, or orthographic standardization; it should also document everyday relational forms that make a language socially usable (de Jesus and Nunes 2024; Merx et al. 2025; Mager et al. 2023; Ranathunga et al. 2023). Third, studies of Timor-Leste remain strongly concerned with constitutional multilingualism, national identity, and language policy, yet micro-level address systems offer a more intimate view of how national and cultural belonging is enacted in daily life (Williams-van Klinken and Hajek 2006; Government of Timor-Leste 2002; INETL 2023).

Theoretically, the article builds on cultural pragmatics and linguistic anthropology. The basic assumption is that linguistic forms do not merely refer to social relations; they help produce, maintain, repair, and contest those relations (Duranti 1997; Hanks 1996; Keane 2016). A speaker who says *Aman*, *Mana*, *Mestri*, *Lian Nain*, or *Chefe de Suku* is not simply selecting a noun phrase. The speaker is activating a social frame that carries expectations of age, gender, profession, ritual legitimacy, leadership, familiarity, and accountability. This is why address terms are central to the classic problem of power and solidarity (Brown and Gilman 1960), to

politeness and face management (Brown and Levinson 1987; Kádár and Haugh 2013), and to the interactional organization of participation (Sacks, Schegloff, and Jefferson 1974; Schegloff 2007). In multilingual contexts, their importance is amplified because borrowed terms may carry institutional prestige, while Indigenous terms may carry genealogical, ritual, or affective force.

Tetun Dili is an especially revealing case because it is not merely a local language. It is one of the official languages of Timor-Leste, alongside Portuguese, and it functions as a major lingua franca in public life (Government of Timor-Leste 2002; Williams-van Klinken, Hajek, and Nordlinger 2002). Section 13 of the Constitution states that Tetum and Portuguese are official languages and that Tetum and other national languages shall be valued and developed by the state (Government of Timor-Leste 2002). At the same time, the 2022 Population and Housing Census highlights Timor-Leste's demographic and linguistic complexity in the post-independence period, showing why any analysis of a national language must remain sensitive to region, mobility, education, and multilingual repertoires (INETL 2023). Recent computational work on Tetun further underlines its low-resource status in digital language technology, including text retrieval, corpus construction, and machine translation (de Jesus 2024; de Jesus and Nunes 2024; Merx et al. 2025). These developments make the study of address terms doubly important: they matter for cultural interpretation and for future digital documentation.

Prior research on Dili Tetun address patterns shows that speakers may use a range of strategies even for the same addressee and that polite forms expand as Tetun moves into domains previously dominated by other languages (Williams-van Klinken and Hajek 2006). This article extends that insight by treating the address lexicon not only as a pragmatic system but also as a cultural map. The corpus includes terms for nuclear family, extended family, affinal relations, honorifics, professional titles, clan houses, ancestors, ritual leaders, customary law, and playful intimacy. The presence of forms such as *Inan-Aman*, *Uma Lisan*, *Bei'ala*, *Lian Nain*, *Tara Bandu*, *Senhor*, *Bapak/Ibu*, and *Lafaek* shows that Tetun Dili relational vocabulary cannot be reduced to either traditional kinship or modern honorifics. It is a layered social infrastructure in which the household, the clan, the school, the church, the state, and the peer group intersect.

The article therefore asks three research questions. First, what functional domains structure the available corpus of Tetun Dili address and kinship terms? Second, how do lexical origin and formality level reveal language contact and cultural authority in the corpus? Third, what model of relational identity emerges when the distributional evidence is read together with cultural-pragmatic interpretation? These questions are designed to avoid two common weaknesses in small-corpus cultural analysis. One weakness is purely anecdotal interpretation, in which a few striking examples are used to support broad claims. The other is shallow quantification, in which counts are reported without explaining how they relate to culture, power, and communicative practice. The present study combines both approaches: it reports the data structure transparently and then interprets why that structure matters.

The proposed contribution is conceptual and methodological. Conceptually, the article advances the idea of kinship as social infrastructure. This means that kinship and address terms are not only linguistic reflections of pre-existing relations; they are tools through which relations are made legible, emotionally charged, and normatively ordered. Methodologically, the article demonstrates how a small documentary spreadsheet can be converted into a publishable cultural-linguistic analysis when it is audited, normalized, visualized, and interpreted with theoretical care. For journals concerned with language, literature, and cultural dynamics, this approach is useful because it connects local data with larger Scopus-facing debates on cultural pragmatics, multilingual identity, low-resource language documentation, digital futures, and the politics of relational speech (O'Connor and Joffe 2020; Krippendorff 2018; Rozumko 2023).

2. Method

The study used a documentary micro-corpus design, combining qualitative content analysis, cultural-pragmatic interpretation, and descriptive visual analytics (Braun and Clarke 2006; Elo and Kyngäs 2008; Hsieh and Shannon 2005; Krippendorff 2018; Saldaña 2021). The primary data came from a spreadsheet entitled *Sistem Sapaan dan Istilah Kekerabatan dalam Bahasa Tetun Dili*. The workbook contains two sheets: a main data table and a source-reference sheet. The main table includes seven fields: family or social category, Tetun term, English translation, definition or context of use, formality level, lexical origin, and source index. The corpus contains 37 lexical records. Each record was treated as an analytic unit. Because the data were not collected through live ethnography, the article does not claim to reconstruct natural conversation. Instead, it analyzes the relational taxonomy embedded in the available documentary data and reads that taxonomy through sociolinguistic, cultural-pragmatic, and language-contact theory.

The audit procedure followed three steps. First, all records were inspected for lexical interpretability. No row was removed from the corpus because each contained an identifiable Tetun Dili term or culturally relevant expression. However, one record contained an origin field marked as not found in the source. This item was retained for relational analysis because the term and function were interpretable, but it was marked as origin-unverified in the lexical-origin profile. Second, the raw family and social categories were normalized into six functional domains: household and intimate kinship; extended and affinal kinship; honorific and professional address; clan, ritual, and ancestral authority; community governance and customary law; and social intimacy and symbolic affiliation. Third, the lexical-origin field was normalized into Indigenous Tetun, Portuguese-derived, mixed Tetun-Portuguese, Indigenous/contact variety, mixed Portuguese-Indonesian, Indonesian-derived, and unverified/source gap.

The coding scheme was developed through a combination of deductive and inductive logic. Deductively, the analysis drew on existing work on address systems, kinship terms, politeness, language ideology, and interactional pragmatics (Brown and Gilman 1960; Brown and Levinson 1987; Agha 2007; Suryanarayan and Khalil 2021; Hamdani, Barnes, and Blythe 2022). Inductively, the categories were adjusted

to the actual corpus. For example, Uma Lisan and Bei'ala could not be adequately placed under ordinary extended family because the definitions frame them through clan identity, ancestral communication, and sacred houses. Likewise, Tara Bandu is not merely a lexical title; it is a customary law practice regulating social order and resource management. The coding therefore had to preserve the distinction between household kinship, clan belonging, ritual authority, and community governance.

The study used descriptive statistics rather than inferential statistics. Frequencies and percentages were calculated for functional domains, lexical origins, formality bands, and source distribution. Cross-tabulations were then used to examine how formality and origin intersected with functional domains. The visual analysis included three figures: a ranked bar chart of functional domains, a bar chart of lexical origins, and a heatmap of formality across domains. The purpose of these visualizations was not decorative. Each figure was used to clarify an analytical decision: domain distribution shows where relational density is concentrated; origin profile shows how contact history is lexicalized; formality heatmap shows how authority is unevenly distributed across kinship, honorific, ritual, and governance categories.

Because the corpus is small, interpretive caution is necessary. The study does not claim that 37 records represent the entire Tetun Dili address system, the entire Tetun language, or all Timorese cultural identities. It also does not treat the source-indexed spreadsheet as a fully verified ethnographic archive. Rather, it treats the workbook as a curated documentary dataset that can be strengthened through transparent auditing and careful interpretation. This stance follows qualitative content-analysis principles: categories must be explicit, evidence must be traceable, and conclusions must not exceed the data (Hsieh and Shannon 2005; Schreier 2012; Saldaña 2021; Nowell et al. 2017). It also follows recent discussions of qualitative reliability, which emphasize transparency and reflexivity rather than mechanical certainty (O'Connor and Joffe 2020).

The interpretive stage focused on cultural function. Each term was read according to four questions: What relation does the term name or activate? What degree of formality does it imply? What kind of authority or intimacy does it support? What does its lexical origin reveal about contact, history, or institutional domain? These questions allowed the analysis to connect micro-level lexical data to macro-level cultural dynamics. Thus, Aman and Inan-Aman were read as household kinship; Banin and Fetosaa-umane as affinal and inter-family relation; Mana, Mama, Mestri, Mestre, and Senhor as honorific or professional address; Uma Lisan, Lian Nain, Lulik, and Bei'ala as ritual and ancestral authority; Tara Bandu and Chefe de Suku as customary and administrative governance; and Lafaek as playful symbolic intimacy. The result is a relational model that treats address terms as social infrastructure rather than as isolated vocabulary.

3. Results and Discussion

3.1. Corpus structure and functional-domain distribution

The documentary corpus contains 37 analyzable terms. The distribution is small but culturally dense. Household and intimate kinship forms constitute the

largest domain, with 11 records or 29.7% of the corpus. Clan, ritual, and ancestral authority follows with 9 records or 24.3%. Honorific and professional address accounts for 8 records or 21.6%. Extended and affinal kinship contains 5 records or 13.5%. Community governance and customary law contains 3 records or 8.1%, while social intimacy and symbolic affiliation contains 1 record or 2.7%. The distribution immediately shows that the corpus is not simply a list of biological family terms. Instead, it moves from the household to the clan, from intimate address to ritual authority, and from ordinary kinship to customary governance.

Table 1 presents the basic audit and inclusion structure of the dataset. The purpose of this table is to clarify that the article's empirical base is a small but complete documentary corpus. The table also distinguishes between lexical interpretability and origin verification, because one term has an unverified origin field but remains interpretable as a relational item.

Table 1. Corpus audit and inclusion protocol

Stage		Operational decision	Records
Initial corpus	documentary	All lexical rows in the main sheet before normalization	37
Retained lexical records		Rows with interpretable Tetun term and contextual definition	37
Origin-unverified records		Rows retained for relational interpretation but marked as unverified in origin field	1
Normalized domains	functional	Analytical categories used to group raw family/social labels	6
Normalized origin classes	lexical-	Origin classes used to map Indigenous, borrowed, mixed, and unverified forms	7
Normalized bands	formality	Pragmatic bands used to interpret respect, intimacy, and ritual authority	5

The audit confirms that the dataset is best understood as a documentary micro-corpus rather than as a large sociolinguistic sample. This distinction is methodologically important. A large corpus would allow broader statistical inference; a small curated corpus allows more careful cultural interpretation. The strength of the dataset lies not in size but in its ability to show how relational categories cluster across household, clan, honorific, professional, and customary domains. This is consistent with current cultural-linguistic work that treats small but well-audited datasets as legitimate when claims are proportionate to the evidence (Krippendorff 2018; Schreier 2012; O'Connor and Joffe 2020; Nowell et al. 2017). It also aligns with language-documentation approaches that view lexical documentation as a record of social practice, not merely as a list of isolated items (Himmelman 1998; Woodbury 2011; Austin 2016).

Table 2 reports the coding frame used to normalize the raw categories. The raw spreadsheet contains categories such as *Keluarga Inti*, *Keluarga Besar*, *Gelar Kehormatan*, *Sapaan Kehormatan*, *Sistem Kekerabatan*, *Uma Lisan*, *Leluhur*, *Kepemimpinan Tradisional*, and *Tara Bandu*. These labels were analytically useful but too granular for corpus-level interpretation. The normalized coding frame therefore consolidates them into six functional domains while retaining the cultural

difference between household kinship, extended kinship, honorific address, ritual authority, customary governance, and symbolic intimacy.

Table 2. Functional-domain coding frame for Tetun Dili address and kinship terms

Functional domain	Operational definition	Records	Share (%)
Household and intimate kinship	Parents, spouse, siblings, step relation, and affective intra-household address	11	29.7
Clan, ritual, and ancestral authority	Clan house, ancestor, sacred authority, ritual speaker, lineage names, and symbolic ancestral belonging	9	24.3
Honorific and professional address	Respectful address, professional titles, gendered honorifics, and formal public address	8	21.6
Extended and affinal kinship	Extended family, uncles, aunts, affinal relations, intermarried families, and kinship networks	5	13.5
Community governance and customary law	Village leadership, customary regulation, collective action, and public agreement	3	8.1
Social intimacy and symbolic affiliation	Playful or symbolic address used to mark close social affiliation	1	2.7

The coding frame matters because it prevents two distortions. First, it prevents the analysis from collapsing ritual and ancestral expressions into ordinary kinship. Uma Lisan and Bei'ala are related to family, but their cultural function exceeds household reference because they connect persons to clan house, lineage memory, and ancestral presence. Second, it prevents modern professional titles such as *Profesora*, *Mestre*, and *Mestri* from being treated as separate from honorific practice. In interactional terms, these professional titles also manage respect, institutional identity, and role-based hierarchy. The frame therefore allows the article to map relational speech across private, public, ritual, and institutional scales.

To make the domain structure visible, Figure 1 translates the normalized counts into a ranked distribution. Whereas Table 2 explains how the categories were defined, Figure 1 shows their relative weight in the corpus. The figure is necessary because the three largest domains—household kinship, ritual authority, and honorific address—together account for 28 records, or 75.7% of the corpus. This concentration suggests that Tetun Dili relational vocabulary in the dataset is organized around three major social sites: the family, the sacred or ancestral order, and the public etiquette of respect.

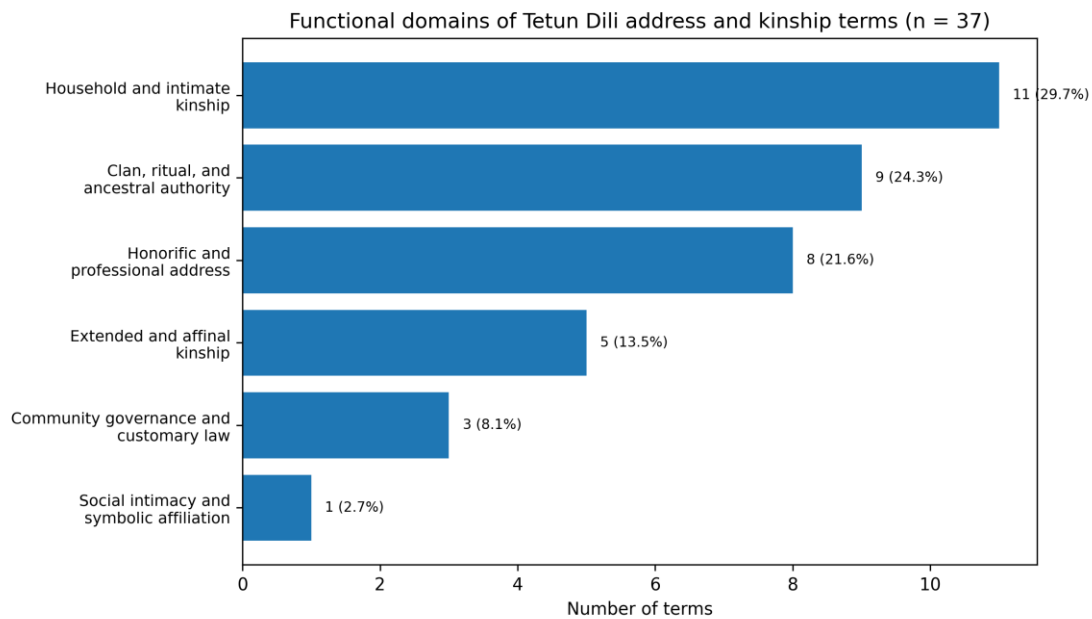


Figure 1. Functional domains of Tetun Dili address and kinship terms.

Figure 1 shows that household kinship is the most frequent domain, but it is not overwhelmingly dominant. This is the central result of the first subsection. If the corpus were only a family vocabulary, household and extended kinship would dominate the distribution. Instead, ritual and honorific forms are nearly as visible as household terms. The presence of *Lian Nain*, *Uma Lisan*, *Bei'ala*, *Lulik*, *Tara Bandu*, *Chefe de Suku*, and *Fulidai-dai* indicates that relational identity in Tetun Dili is strongly mediated by institutions of ritual, ancestry, leadership, and communal obligation. Kinship is therefore not only domestic; it extends outward into cultural governance.

The smaller domains are also analytically important. Extended and affinal kinship contains terms such as *Família boot*, *Tiu*, *Tia*, *Banin*, and *Fetosaa-umane*. These terms show that kinship expands through marriage, alliance, age, and extended household networks. Community governance and customary law, although smaller, contains high-value expressions because it links address and identity to collective decision-making and resource regulation. The single playful-intimacy record, *Lafaek*, is especially interesting because it shows that relational terms can also produce intimacy through joking or symbolic affiliation. Thus, a small category can still carry theoretical significance when it reveals a distinct pragmatic function.

3.2. Lexical origin and the contact architecture of address

The lexical-origin profile reveals a layered contact system. Indigenous Tetun forms account for 21 records, or 56.8% of the corpus. Portuguese-derived forms account for 8 records, or 21.6%. Mixed Tetun-Portuguese forms account for 3 records, or 8.1%. Indigenous/contact-variety forms account for 2 records, or 5.4%. Indonesian-derived and mixed Portuguese-Indonesian forms each account for 1 record, or 2.7%, and one record contains an unverified origin field. This distribution

provides one of the strongest pieces of evidence in the article: Tetun Dili address and kinship terms are anchored in Indigenous relational categories but remain visibly shaped by Portuguese and Indonesian contact.

Table 3 presents the lexical-origin profile. The table shows that Indigenous Tetun is the largest source, especially in household kinship and ritual authority. This is expected because terms for parents, siblings, clan houses, ancestors, sacredness, and customary relations tend to be deeply rooted in local culture. However, the presence of Portuguese-derived forms is not incidental. Terms such as *Família*, *Profesora*, *Mestre*, *Mestri*, *Senhor*, *Madrasta*, *Tiu*, and *Tia* point to colonial, religious, educational, and administrative contact. They also show that borrowing is not just a lexical process; it is a cultural-pragmatic process through which institutional roles and social hierarchies enter the address system.

Table 3. Lexical-origin profile of the Tetun Dili relational corpus

Lexical-origin class	Records	Share (%)
Indigenous Tetun	21	56.8
Portuguese-derived	8	21.6
Mixed Tetun-Portuguese	3	8.1
Indigenous/contact variety	2	5.4
Mixed Portuguese-Indonesian	1	2.7
Indonesian-derived	1	2.7
Unverified/source gap	1	2.7

The table should not be read as a simple contrast between authentic Indigenous terms and foreign borrowed terms. Such a contrast would be reductive. Contact languages often become locally meaningful through repeated use, adaptation, and pragmatic specialization (Thomason 2001; Matras 2009; Bakhtin 1981). In Tetun Dili, Portuguese-derived terms may signal modern education, formal respect, church-mediated history, or urban social prestige. Indonesian-derived *Bapak/Ibu* similarly reflects modern regional contact and administrative familiarity. Mixed forms such as *Família boot* demonstrate that contact elements can be embedded into Tetun syntactic and semantic patterns. The result is not lexical impurity but layered social meaning.

To clarify this layering visually, Figure 2 presents the lexical-origin profile as a bar chart. The visual transition is useful because it shows both dominance and diversity. Indigenous Tetun remains clearly dominant, but the contact-derived layer is too large to ignore. Portuguese-derived and mixed Portuguese-Tetun forms together account for 11 records, or 29.7% of the corpus. If Indonesian and mixed Portuguese-Indonesian forms are added, contact-visible items account for 13 records, or 35.1%. This means that over one-third of the corpus shows explicit traces of historical or regional contact.

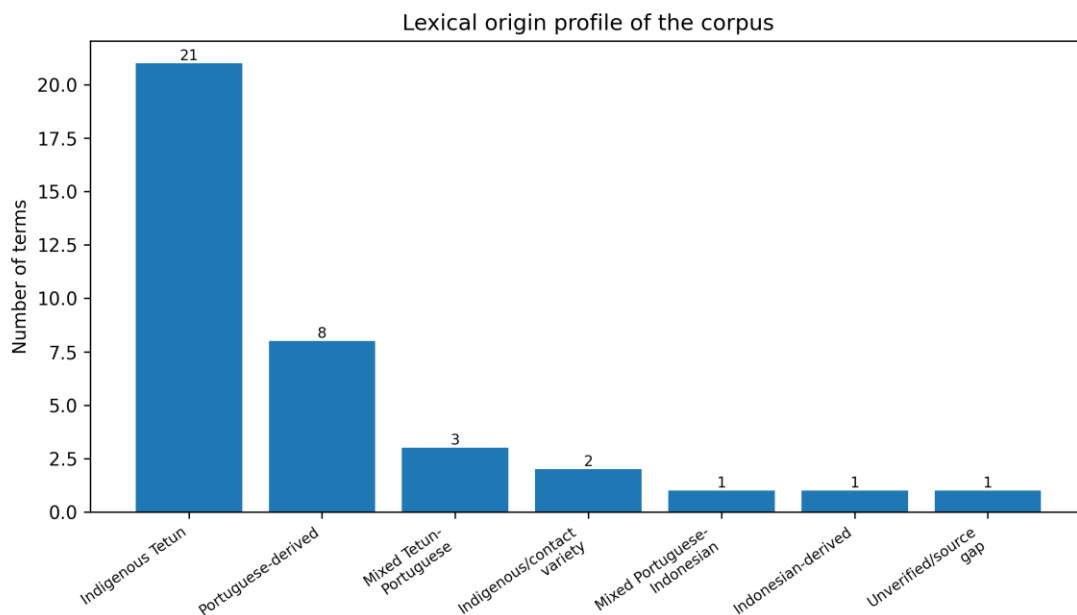


Figure 2. Lexical origin profile of the corpus.

Figure 2 confirms that Tetun Dili relational vocabulary is best understood as an ecology rather than as a sealed system. Indigenous terms provide the core of kinship, ritual authority, and ancestral belonging, while borrowed and mixed forms mark professional, institutional, and modern honorific domains. This pattern aligns with Williams-van Klinken and Hajek's observation that Dili Tetun address practices display considerable variation and adaptation as the language expands across domains (Williams-van Klinken and Hajek 2006). It also resonates with recent low-resource language research: language technology and documentation must preserve culturally central vocabulary rather than flattening it into general translation equivalents (de Jesus and Nunes 2024; Merx et al. 2025; Mager et al. 2023).

The contact architecture is particularly visible in professional titles. *Profesora*, *Mestre*, and *Mestri* do not merely translate 'teacher.' They attach educational authority to a historically layered lexicon. In a postcolonial state where Tetun and Portuguese are official languages and where Indonesian and English have also shaped public life, titles can become indexes of institutional history (Government of Timor-Leste 2002; INETL 2023). The corpus therefore illustrates a broader sociolinguistic principle: borrowed address forms often cluster in domains where power, schooling, administration, and modern public status are salient (Agha 2007; Irvine and Gal 2000; Silverstein 2003).

3.3. Formality, ritual density, and relational authority

The formality distribution strengthens the argument that Tetun Dili address terms operate as relational infrastructure. Ritual/cultural authority is the largest formality band, with 13 records or 35.1%. Formal/respectful forms account for 11 records, or 29.7%. General/standard and familiar/affective forms each contain 6 records, or 16.2%. Playful/intimate usage appears in 1 record, or 2.7%. The

dominance of ritual/cultural and formal/respectful bands shows that the dataset is heavily oriented toward social regulation rather than casual naming alone.

Table 4 summarizes the formality bands and their pragmatic meanings. The table shows that formality in this corpus is not a single scale from casual to formal. It is a cultural structure with several dimensions. Formal/respectful forms manage interpersonal respect. Ritual/cultural authority forms manage sacred, ancestral, or customary legitimacy. General/standard forms manage neutral reference. Familiar/affective forms manage closeness. Playful/intimate forms create solidarity through teasing or symbolic association. The formality system is therefore multidimensional: it differentiates not only degrees of politeness but also types of social relation.

Table 4. Formality bands and pragmatic orientation

Formality band	Records	Share (%)
Ritual/cultural authority	13	35.1
Formal/respectful	11	29.7
General/standard	6	16.2
Familiar/affective	6	16.2
Playful/intimate	1	2.7

The strongest evidence comes from the cross-domain pattern. Honorific and professional address is entirely formal/respectful in the coding. Clan, ritual, and ancestral authority is entirely ritual/cultural. Household kinship is mixed: it includes familiar/affective, general/standard, formal/respectful, and one ritual/cultural item. This pattern shows that household kinship is flexible, while ritual and professional domains are more specialized. In pragmatic terms, Aman or Inan may be used in everyday family contexts, but Lian Nain or Lulik carries a stronger cultural force. A teacher title may be polite and institutional, but an ancestral or sacred title invokes a deeper normative order.

Figure 3 presents this formality-gradient pattern as a heatmap. The transition from Table 4 to Figure 3 is necessary because the table identifies band totals, while the figure shows how those bands are distributed across functional domains. The heatmap makes visible that ritual authority is concentrated in clan, ancestral, and customary domains, while formal respect is concentrated in honorific and professional address. Household terms are the most internally diverse, confirming that the family is both an affective and normative site.

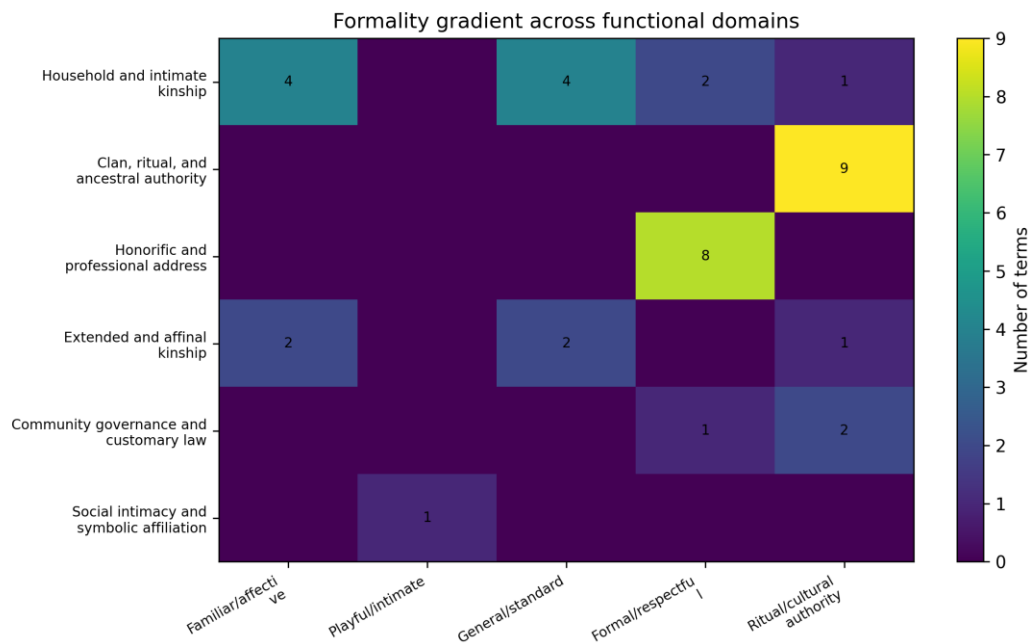


Figure 3. Formality gradient across functional domains.

Figure 3 demonstrates that Tetun Dili address practices organize relational authority through domain-specific formality. This is a crucial finding because it prevents the analysis from treating all respect forms as equivalent. *Mestri*, *Senhor*, and *Bapak/Ibu* point toward formal or institutional respect. *Lian Nain*, *Uma Lisan*, *Bei'ala*, and *Lulik* point toward sacred and ancestral legitimacy. *Aman*, *Ama/Inan*, *Maun*, *Alin*, and *Fen* point toward household roles that can be intimate, standard, or socially respectful depending on context. The dataset therefore supports a layered model: respect is not only interpersonal; it is institutional, genealogical, ritual, and affective.

This finding also has implications for translation and digital language resources. If an automatic system translates *Lian Nain* simply as 'traditional speaker' or *Lulik* simply as 'sacred,' it may miss the relational work carried by the term. Low-resource MT and information retrieval systems increasingly need domain-sensitive lexical resources, especially for cultural terms that do not map neatly onto high-resource-language equivalents (de Jesus 2024; Merx et al. 2025; Ranathunga et al. 2023). The corpus examined here offers precisely the kind of micro-lexical data that can enrich glossaries, educational materials, and culturally grounded NLP resources.

3.4. Kinship beyond the household: clan, ancestry, and customary governance

The most conceptually important result is that kinship extends beyond biological relation. Terms such as *Uma Lisan*, *Uma Lulik*, *Bei'ala*, *Ratu*, *Mape*, *Mane Walu*, and *Uma Metan* show that kinship is embedded in house, clan, ancestor, symbol, and name. In many societies, kinship is not simply a map of descent. It is a cultural technology for organizing memory, land, obligation, inheritance, ritual, and moral accountability (Basso 1996; Assmann 2011; Connerton 1989; Vansina 1985). The Tetun Dili corpus fits this broader anthropological pattern. The house is not only

a physical building; it is a site where identity is stored and transmitted. Ancestors are not only past relatives; they are active figures in the moral imagination of descendants.

Table 5 provides representative terms and cultural functions. It is not intended to replace the full corpus. Rather, it selects terms that illustrate the movement from household kinship to cultural authority. The table includes Inan-Aman as parental collectivity, Banin as affinal extension, Lian Nain as ritual leadership, Uma Lisan as clan identity, Lulik as sacred order, Tara Bandu as customary regulation, and Lafaek as playful intimacy. Together, these items show that the address and kinship system is not only referential but performative: it activates relations that can be affectionate, hierarchical, sacred, administrative, or humorous.

Table 5. Representative Tetun Dili relational terms and cultural functions

Term	English gloss	Functional domain	Cultural-pragmatic function
Inan-Aman	Parents	Household and intimate kinship	Names the parental pair as a collective moral and domestic unit.
Banin	Parents-in-law	Extended and affinal kinship	Extends kinship through marriage and affinal obligation.
Mana	Older sister / respectful sister	Honorific and professional address	Turns sisterhood into a polite public address form.
Mestre / Mestri	Teacher	Honorific and professional address	Indexes schooling, knowledge authority, and institutional respect.
Uma Lisan	Clan / lineage group	Clan, ritual, and ancestral authority	Links kinship to house, lineage, memory, and collective identity.
Lian Nain	Traditional speaker / ritual leader	Clan, ritual, and ancestral authority	Locates communicative authority in ritual legitimacy and ancestral mediation.
Lulik	Sacred / holy	Clan, ritual, and ancestral authority	Marks sacred objects or places protected by customary norms.
Tara Bandu	Customary law / public agreement	Community governance and customary law	Connects relational language with peace, norms, and resource governance.
Lafaek	Crocodile	Social intimacy and symbolic affiliation	Functions playfully among close friends, showing symbolic intimacy rather than literal reference.

The table reveals a critical cultural pattern. Some terms define persons by their household role. Others define them by their institutional position. Others define them by their relation to ancestors, sacred houses, customary law, or symbolic animals. This is why the article uses the phrase kinship as social infrastructure. Infrastructure usually refers to roads, bridges, energy, or digital systems. Here, the term is used analytically to describe the relational pathways through which a society organizes recognition, respect, authority, and belonging. Address terms are linguistic infrastructure because they make social movement possible: one knows how to approach an elder, a teacher, a ritual leader, a stranger, a friend, or a chief.

Tara Bandu is especially important because it connects language, culture, law, and ecology. In the corpus, Tara Bandu is defined as customary law or public

agreement regulating social norms, peace, and natural-resource management. This aligns with wider discussions of customary practice in Timor-Leste, where Tara Bandu is often understood as a mechanism for communal regulation and ecological restraint (Miyazawa 2013; AP-PLAT 2021). Its presence in an address-and-kinship dataset shows that relational terms do not stop at interpersonal etiquette. They can enter environmental governance and collective accountability. This makes the article relevant to contemporary Scopus trends in language ecology, Indigenous knowledge, sustainability, and cultural governance.

The playful term *Lafaek* adds another dimension. Literally glossed as crocodile and used among close friends in a playful way, it suggests that symbolic animal terms can mark closeness rather than mere insult or classification. This is consistent with pragmatic theories of meaning: the force of a term depends on relationship, situation, tone, and shared background, not only dictionary semantics (Kecskes 2014; Goffman 1967; Hanks 1996). A form that might be aggressive in one context can be affectionate in another. Therefore, the corpus shows that Tetun Dili relational language cannot be analyzed only by semantic category. It must be analyzed as interactional practice.

3.5. Toward a model of Tetun Dili relational identity

The results support a five-layer model of Tetun Dili relational identity. The first layer is household intimacy, represented by terms such as *Aman*, *Ama/Inan*, *Maun*, *Alin*, *Fen*, and *Inan-Aman*. This layer organizes everyday family roles and early moral socialization. The second layer is extended and affinal relation, represented by *Tiu*, *Tia*, *Banin*, *Família boot*, and *Fetosaa-umane*. This layer shows that family is not confined to the nuclear unit but expands through marriage, alliance, seniority, and social obligation. The third layer is public respect, represented by *Mana*, *Mama*, *Senhor*, *Bapak/Ibu*, *Mestre*, *Mestri*, and *Profesora*. This layer connects address to gender, age, professional identity, and institutional hierarchy.

The fourth layer is ritual and ancestral authority, represented by *Uma Lisan*, *Lian Nain*, *Bei'ala*, *Lulik*, *Ratu*, and related clan names. This layer carries the deepest cultural density because it links persons to memory, lineage, house, sacredness, and ancestral presence. The fifth layer is communal governance, represented by *Chefe de Suku*, *Fulidai-dai*, and *Tara Bandu*. This layer moves from person-to-person address to collective norm-making. It reveals how relational language participates in the organization of community, leadership, peace, and resource use. Across these five layers, Tetun Dili address terms create a social map that links the intimate, the formal, the sacred, and the political.

The model also helps explain why address terms are central to cultural literacy. A learner who memorizes Tetun words without understanding relational function will know translations but not social meanings. For example, translating *Mana* as sister, *Senhor* as sir, *Lian Nain* as ritual leader, and *Uma Lisan* as clan house is useful but insufficient. Each term must be learned as part of a social scene: who may use it, to whom, in what setting, with what expectations, and with what risks if misused. This point has direct implications for Tetun language teaching,

intercultural communication, translation, and AI-supported lexical resources. Cultural pragmatics must be part of language documentation, not an optional addition (Duranti 1997; Kádár and Haugh 2013; Merx et al. 2025).

The model also complicates postcolonial language-contact analysis. Portuguese-derived forms do not simply indicate colonial residue, and Indigenous forms do not simply indicate untouched tradition. The corpus shows a more dynamic ecology. Some Portuguese-derived terms have become ordinary tools of respect and institutional identity. Some Indigenous terms are highly formal or sacred. Some mixed forms show localized adaptation. Indonesian-derived *Bapak/Ibu* appears as a modern honorific influenced by regional contact. Such layering is typical of multilingual societies in which language choice is bound to history, education, religion, state formation, and mobility (Blommaert 2010; Appadurai 1996; Irvine and Gal 2000). The value of the corpus is that it makes this layering visible at the level of individual address terms.

The article also contributes to low-resource language scholarship. Recent work has made major progress in Tetun corpora, information retrieval, and machine translation (de Jesus and Nunes 2024; de Jesus 2024; Merx et al. 2025). However, computational resources often prioritize sentence-level news data, translation pairs, or web text. Such resources may underrepresent culturally dense relational terms unless they are deliberately documented. Address and kinship lexicons should therefore be treated as critical data for low-resource language development. They can support culturally sensitive translation, glossary-building, heritage education, and community-controlled language technologies. Without them, digital systems may reproduce surface translation while missing social meaning.

Finally, the findings contribute to broader debates on Southeast Asian language and culture. Similar patterns appear in Indonesian, Balinese, Bugis, Javanese, and other regional address systems, where kinship terms often extend into non-kin relationships and where address choices encode respect, age, intimacy, and hierarchy (Errington 1988; Hamdani, Barnes, and Blythe 2022; Gusanto, Budasi, and Dewi 2025). What makes Tetun Dili distinctive in this dataset is the visibility of clan houses, ancestral authority, and customary law alongside modern professional and contact-derived honorifics. The system is therefore neither purely traditional nor simply modern. It is a relational ecology shaped by cultural continuity, colonial contact, national language policy, and everyday interaction.

4. Conclusion

This article has shown that Tetun Dili address and kinship terms function as social infrastructure. Based on a 37-entry documentary micro-corpus, the analysis found that household and intimate kinship forms account for 29.7% of the data, clan, ritual, and ancestral authority for 24.3%, honorific and professional address for 21.6%, extended and affinal kinship for 13.5%, community governance and customary law for 8.1%, and symbolic social intimacy for 2.7%. Indigenous Tetun forms remain the lexical core, but Portuguese-derived, Indonesian-derived, and mixed forms reveal a contact ecology shaped by colonial history, schooling, public administration, and regional multilingualism. The main argument is that Tetun Dili

relational vocabulary cannot be reduced to translation equivalents. It organizes intimacy, respect, ritual legitimacy, professional status, clan belonging, and customary governance.

The study contributes a five-layer model of relational identity: household intimacy, extended and affinal relation, public respect, ritual and ancestral authority, and communal governance. This model is useful for language documentation, cultural pragmatics, intercultural education, and low-resource language technology. Future research should expand the corpus through fieldwork, record actual interactional use, verify regional variation with native speakers, and compare Tetun Dili with other Timorese and Austronesian address systems. Digital projects on Tetun should also include culturally dense address and kinship terms so that machine translation, information retrieval, and educational resources preserve not only lexical meaning but relational meaning. In this sense, the future of Tetun documentation is not only computational; it is cultural, pragmatic, and deeply social.

Acknowledgement

The author expresses sincere appreciation to the academic community concerned with Timorese languages, Southeast Asian cultural pragmatics, and low-resource language documentation. The author also acknowledges the value of the uploaded documentary dataset as a starting point for a more systematic analysis of Tetun Dili address terms and kinship expressions.

Author Contributions

Ujeius da Silva Soares: conceptualization, methodology, data auditing, coding design, formal analysis, visualization.

Adelcisa Alves Pereira: writing-original draft, writing-review and editing, and final manuscript approval.

Funding

This research received no external funding.

Data Availability Statement

The data analyzed in this article were derived from the author-provided workbook on Tetun Dili address terms and kinship expressions. The cleaned coding structure and figures can be supplied as supplementary material upon request.

Declaration on the Use of Artificial Intelligence

Artificial intelligence support was used in a limited capacity for language refinement, structural organization, and document preparation. The AI tool was not used to fabricate data, alter the corpus, or make independent scholarly decisions. All analytical decisions, interpretation, and final manuscript approval remain the responsibility of the author.

References

- Agha, Asif. 2007. *Language and Social Relations*. Cambridge: Cambridge University Press.
- AP-PLAT. 2021. Integrating Tara Bandu Law into Timor-Leste's National Adaptation Planning. Asia-Pacific Climate Change Adaptation Information Platform.
- Appadurai, Arjun. 1996. *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis: University of Minnesota Press.
- Assmann, Jan. 2011. *Cultural Memory and Early Civilization: Writing, Remembrance, and Political Imagination*. Cambridge: Cambridge University Press.
- Austin, Peter K. 2016. Language Documentation 20 Years On. *Language Documentation & Conservation* 10: 147-170.
- Bakhtin, Mikhail M. 1981. *The Dialogic Imagination: Four Essays*. Edited by Michael Holquist. Austin: University of Texas Press.
- Basso, Keith H. 1996. *Wisdom Sits in Places: Landscape and Language among the Western Apache*. Albuquerque: University of New Mexico Press.
- Blommaert, Jan. 2010. *The Sociolinguistics of Globalization*. Cambridge: Cambridge University Press.
- Blythe, Joe, Fakry Hamdani, and Scott Barnes. 2024. Tactile Engagement of Prospective Next Speakers in Indonesian Multiparty Conversations. *Language in Society* 53 (4): 673-704. <https://doi.org/10.1017/S0047404523000441>.
- Braun, Virginia, and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3 (2): 77-101. <https://doi.org/10.1191/1478088706qp063oa>.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Brown, Roger, and Albert Gilman. 1960. The Pronouns of Power and Solidarity. In *Style in Language*, edited by Thomas A. Sebeok, 253-276. Cambridge, MA: MIT Press.
- Bucholtz, Mary, and Kira Hall. 2005. Identity and Interaction: A Sociocultural Linguistic Approach. *Discourse Studies* 7 (4-5): 585-614. <https://doi.org/10.1177/1461445605054407>.
- Connerton, Paul. 1989. *How Societies Remember*. Cambridge: Cambridge University Press.
- de Jesus, Gabriel. 2024. Text Information Retrieval in Tetun: A Preliminary Study. arXiv. <https://doi.org/10.48550/arXiv.2406.07331>.
- de Jesus, Gabriel, and Sérgio Sobral Nunes. 2024. Data Collection Pipeline for Low-Resource Languages: A Case Study on Constructing a Tetun Text Corpus. In *Proceedings of the 2024 Joint International Conference on Computational*

- Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 4368-4380. Torino, Italia: ELRA and ICCL. <https://doi.org/10.63317/3bqdt48jvtjm>.
- Duranti, Alessandro. 1997. *Linguistic Anthropology*. Cambridge: Cambridge University Press.
- Eckert, Penelope. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology* 41: 87-100. <https://doi.org/10.1146/annurev-anthro-092611-145828>.
- Elo, Satu, and Helvi Kyngäs. 2008. The Qualitative Content Analysis Process. *Journal of Advanced Nursing* 62 (1): 107-115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>.
- Errington, J. Joseph. 1988. *Structure and Style in Javanese: A Semiotic View of Linguistic Etiquette*. Philadelphia: University of Pennsylvania Press.
- Goffman, Erving. 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*. Chicago: Aldine.
- Government of Timor-Leste. 2002. *Constitution of the Democratic Republic of Timor-Leste*. Dili: Government of Timor-Leste.
- Gusanto, Athanasia, I Gede Budasi, and Kadek Sintya Dewi. 2025. The Terms of Address Used by Local People 'Krama Desa Adat' in Lemukih Village. *Journal of English Language and Education* 10 (1): 518-526.
- Hamdani, Fakry, Scott Barnes, and Joe Blythe. 2022. Questions with Address Terms in Indonesian Conversation: Managing Next-Speaker Selection and Action Formation. *Journal of Pragmatics* 200: 194-210. <https://doi.org/10.1016/j.pragma.2022.08.010>.
- Hanks, William F. 1996. *Language and Communicative Practices*. Boulder, CO: Westview Press.
- Himmelmann, Nikolaus P. 1998. Documentary and Descriptive Linguistics. *Linguistics* 36 (1): 161-195. <https://doi.org/10.1515/ling.1998.36.1.161>.
- Hsieh, Hsiu-Fang, and Sarah E. Shannon. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15 (9): 1277-1288. <https://doi.org/10.1177/1049732305276687>.
- INETL. 2023. *Timor-Leste Population and Housing Census 2022 Main Report*. Dili: Timor-Leste National Institute of Statistics.
- Irvine, Judith T., and Susan Gal. 2000. Language Ideology and Linguistic Differentiation. In *Regimes of Language: Ideologies, Politics, and Identities*, edited by Paul V. Kroskrity, 35-83. Santa Fe: School of American Research Press.
- Kádár, Dániel Z., and Michael Haugh. 2013. *Understanding Politeness*. Cambridge: Cambridge University Press.

- Keane, Webb. 2016. *Ethical Life: Its Natural and Social Histories*. Princeton: Princeton University Press.
- Kecskes, Istvan. 2014. *Intercultural Pragmatics*. Oxford: Oxford University Press.
- Krippendorff, Klaus. 2018. *Content Analysis: An Introduction to Its Methodology*. 4th ed. Thousand Oaks, CA: Sage.
- Mager, Manuel, Elisabeth Mager, Katharina Kann, and Ngoc Thang Vu. 2023. Ethical Considerations for Machine Translation of Indigenous Languages: Giving a Voice to the Speakers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 4871-4897. Toronto: Association for Computational Linguistics.
- Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.
- Merx, Raphaël, Adérito José Guterres Correia, Hanna Suominen, Trevor Cohn, and Ekaterina Vylomova. 2025a. Low-Resource Machine Translation: What For? Who For? An Observational Study on a Dedicated Tetun Language Translation Service. In *Proceedings of the Workshop on Low-Resource Machine Translation*. Association for Computational Linguistics.
- Merx, Raphaël, Hanna Suominen, Lois Hong, Nick Thieberger, Trevor Cohn, and Ekaterina Vylomova. 2025b. TULUN: Transparent and Adaptable Low-Resource Machine Translation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics.
- Miyazawa, Nobuyuki. 2013. Customary Law and Community-Based Natural Resource Management in Post-Conflict Timor-Leste. In *Land and Post-Conflict Peacebuilding*, edited by Jon Unruh and Rhodri Williams, 497-514. London: Earthscan.
- Nowell, Lorelli S., Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods* 16 (1): 1-13. <https://doi.org/10.1177/1609406917733847>.
- O'Connor, Cliodhna, and Helene Joffe. 2020. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods* 19: 1-13. <https://doi.org/10.1177/1609406919899220>.
- Purba, Ridwin, Corry Corry, Herman Herman, Magdalena Ngongo, Hisarma Saragih, Tutiariani Nasution, and Rohdearni Wati Sipayung. 2024. Simalungun Addressing Terms Based on the Kinship System of the Tolu Sahundulan Lima Saodoran. *Research Journal in Advanced Humanities* 5 (3): 37-49. <https://doi.org/10.58256/fm64kj51>.
- Ranathunga, Surangika, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. 2023. Neural Machine Translation for Low-Resource Languages: A Survey. *ACM Computing Surveys* 55 (11): 1-37.

- Rozumko, Agata. 2023. Interpersonal Relationships in Translation: Address Terms in the English and Polish Translations of Henning Mankell's *The Dogs of Riga*. *Journal of Pragmatics* 206: 19-28. <https://doi.org/10.1016/j.pragma.2023.01.013>.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50 (4): 696-735. <https://doi.org/10.2307/412243>.
- Saldaña, Johnny. 2021. *The Coding Manual for Qualitative Researchers*. 4th ed. London: Sage.
- Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Schreier, Margrit. 2012. *Qualitative Content Analysis in Practice*. London: Sage.
- Silverstein, Michael. 2003. Indexical Order and the Dialectics of Sociolinguistic Life. *Language & Communication* 23 (3-4): 193-229. [https://doi.org/10.1016/S0271-5309\(03\)00013-2](https://doi.org/10.1016/S0271-5309(03)00013-2).
- Suryanarayan, Neelakshi, and Amr Khalil. 2021. Kinship Terms as Indicators of Identity and Social Reality: A Case Study of Syrian Arabic and Hindi. *Russian Journal of Linguistics* 25 (1): 125-146. <https://doi.org/10.22363/2687-0088-2021-25-1-125-146>.
- Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.
- Vansina, Jan. 1985. *Oral Tradition as History*. Madison: University of Wisconsin Press.
- Wajdi, Majid, Aprianoto Aprianoto, Ratna Dewi Harahap, Mulyono Mulyono, and Wisman Hadi. 2024. How Indonesian Terms of Address 'Bapak or Ibu' and 'Kamu' Are Used in Indonesian Educational Settings. *Journal of Education, Social & Communication Studies* 1 (1): 1-9.
- Williams-van Klinken, Catharina, and John Hajek. 2006. Patterns of Address in Dili Tetum, East Timor. *Australian Review of Applied Linguistics* 29 (2): 21.1-21.18. <https://doi.org/10.2104/ara10621>.
- Williams-van Klinken, Catharina, John Hajek, and Rachel Nordlinger. 2002. *Tetun Dili: A Grammar of an East Timorese Language*. Canberra: Pacific Linguistics.
- Woodbury, Anthony C. 2011. Language Documentation. In *The Cambridge Handbook of Endangered Languages*, edited by Peter K. Austin and Julia Sallabank, 159-186. Cambridge: Cambridge University Press.
- Zainal, Nurfitriana, and Kumara Tungga Dewa. 2022. The Sociolinguistics Analysis of Address Terms in English and Enrekang-Durinese. *Scientia: Social Sciences & Humanities* 1 (1): 124-132.