



SAINT LOUIS UNIVERSITY
SCHOOL OF ACCOUNTANCY, MANAGEMENT,
COMPUTING, AND INFORMATION STUDIES
Computer Science and Computer Applications Department



Prelim Activity:
WRITE-UPS FOR TUWALI ELECTRONIC LEXICON

9363 | CSE 30

1:30 - 3:00 TF

Group 3:

Balang, Patrick Andrew
Binwag, Dana Wynnette
Bromeo, Emmanuel Paolo
Bumacod, Mark John
Calsiman, Carl Allan
Cendaña, Juliana April
Munar, Dulthe Carlo Jr.

Submitted to:
Sir Dalos Miguel

INTRODUCTION

Tuwali is one of the most spoken languages in Ifugao provinces. It originated in the Southern region part of Ifugao, particularly in the municipality of Kiangnan (Hohulin & Burquest, 2011). Kiangnan is widely known as a cultural and historical center of the Ifugao people since it is the oldest town in the province of Ifugao. Tuwali is part of the Austronesian Language Family, Malayo-Polynesian and Northern Luzon Group of Languages (Hohulin & Burquest, 2011; Taleon, 2020). There are approximately 25,000 speakers of the native Tuwali language as of the year 2000, and it is still increasing (Hohulin & Burquest, 2011). The language is actively being used in daily communication and is also used in social events, religious gatherings, and businesses. The strong influence of more widely spoken languages like Ilocano, Tagalog, and English has led many speakers to become bilingual. It broadened communication, but it became a challenge in passing the native Tuwali language to the next generation (Hohulin & Burquest, 2011).

Tuwali language is more than just a way of speaking; it carries the local knowledge, tradition, and cultural identity of the natives of Ifugao (Taleon, 2020). The language shows the qualities, ways of thinking, and cultural practices that are difficult to express in other languages. If Tuwali is not passed or maintained, it may be lost, so valuable cultural and linguistic knowledge may disappear (Taleon, 2020). By using modern technology, efforts to record, protect, and use the language are important to keep it alive.

One way of keeping the language alive is to build a complete Tuwali lexicon (Hohulin & Burquest, 2011). The lexicon will include words, their pronunciation, root words, affixes, English translations, Tagalog translations, and meanings. It can help support digital tools, especially machine translation systems. With a lexicon, a system can be developed to translate between Tuwali, Tagalog, and English. It will help preserve the language and make it more accessible to students, researchers, teachers, and to the public.

METHODOLOGY

Data Collection

The researchers gathered Tuvuli words from different sources, namely songs, poems, stories, video, chant, articles and dictionaries. To get a wide and balanced set of data, the study used both oral and written materials. Oral sources include songs, chants, and folk stories that are still circulating in the community. These oral sources helped preserve older words and traditional expressions. Written sources such as poems, articles, and dictionaries were also used to collect standard forms of the words. Videos were also used to know how Tuvuli is spoken. This was important because pronunciation sometimes changes compared to the written form. By collecting words from many different sources, the study made sure the data reflected how Tuvuli is used in everyday life and it also ensures that the sources are reliable so that it can be used for teaching and research.

Table 1. Summary of the Raw Resources Distribution

Type of Sources	Number of Source
Song	2
Poem	6
Stories	3
Video	2
Chant	2
Others	3

Data Processing and Cleaning

The researchers preprocessed and cleaned the data in order to make the list of words more consistent and accurate. Words that had different spellings were compared and standardized to one accepted form, based on common usage and existing references. Duplicate words were removed so that the same word would not appear twice in the table. Each word was also labeled to identify if it is a noun, verb, or another part of speech. This step was important because a clean and consistent

dataset is needed for building a good lexicon. It also makes the data ready for future use in computer systems like translation tools.

Lexicon Organization

The cleaned dataset was organized into a structured lexicon. The lexicon was not just a list of words, but a table with several columns that provide complete information about each entry. These included:

Table 2. Summary of the created lexicon

Column	Description
Word	The Tuwali words collected
Root Word	The basic form of the word.
Affixes	Any added parts such as prefixes, infixes, or suffixes.
English Translation	The meaning of the word in English.
Tagalog Translation	The meaning of the word in Tagalog.
Meaning	A short explanation of the word.
Part of Speech	The grammatical role of the word (e.g., noun, verb).

This structure makes the lexicon more useful for many purposes. It can help learners understand the language, teachers prepare materials, and researchers study how the language works. At the same time, the clear and organized format makes it easier to use the lexicon in technology. For example, separating root words and affixes can help a translation system recognize how new words are formed. The study created a Tuwali lexicon that supports both language preservation and digital use. The method ensures that the lexicon is complete, easy to understand, and ready for future projects such as building a bidirectional translation system between Tuwali and English.

Tools Used

Several digital tools and platforms were utilized to build and process the Tuwali lexicon efficiently:

1. Google Sheets – Served as the primary platform for compiling and organizing the lexicon. Features such as filtering, sorting, and formula-based operations were used to manage entries and facilitate data exploration.
2. ChatGPT Assistance – Provided support in generating definitions and meanings for Tuwali words that had already been translated into English, helping to enrich the lexicon with accurate semantic information.
3. Google Translate (embedded in Sheets) – Used to translate English words into Tagalog, enabling a bidirectional translation component in the lexicon.
4. Google Colab with Python – Employed for more advanced text processing tasks, including:
 - Extracting root words and affixes from Tuwali entries.
 - Formatting text and preparing data for analysis.
 - Web scraping relevant lexical resources to supplement the lexicon.
 - Conducting data exploration and analysis, such as calculating the number of words under each part of speech, identifying pronunciation patterns, and other linguistic insights.

These tools collectively enabled the systematic compilation, processing, and analysis of the Tuwali lexicon, ensuring both accuracy and usability for future applications, including bidirectional machine translation.

Validation

To ensure the accuracy and reliability of the Tuwali lexicon, validation was conducted through consultations with knowledgeable native speakers and language experts. Assistance was provided by an assistant professor from Ifugao State University, Nayun, Lamut, Ifugao, who helped to confirm the correctness of translations, part-of-speech annotations, and meanings. Additionally, a college student from Saint Louis University with permanent residency in Baguinge, Kiangnan, Ifugao, provided a native-speaker perspective and assisted in cross-checking pronunciation, usage patterns, and translations. Collectively, these consultations not only helped verify word meanings and usage but also supported the translation of Tuwali words into English and

Filipino, ensuring that the lexicon accurately reflects the language and is suitable for applications.

RESULTS AND DISCUSSIONS

Overview of the Lexicon

The compiled Tawali lexicon contains a total of 5755 entries. Each entry includes the Tawali word, root word, affixes, English and Tagalog translations, meaning, and part of speech. This structure allows for easy reference and further computational processing for applications such as bidirectional machine translation.

Table 3. Summary of the Lexicon Distribution by Part of Speech

Part of Speech	Number of Words	Percentage (%)
Noun	781	13.57%
Pronoun	156	2.71%
Verb	4,251	73.87%
Adverb	150	2.61%
Adjective	306	5.34%
Conjunction	53	0.92%
Preposition	34	0.59%
Interjection	20	0.35%
Determiner	1	0.02%

This overview provides a quantitative snapshot of the lexicon, showing how words are distributed across different parts of speech. It also sets the foundation for further analysis, such as extracting root words, affixes, and pronunciation patterns.

Sample Entries

Table 4. Sample Entries from the Tuwali Lexicon

Tuwali	Part of Speech	English Translation	Usage (Tuwali Sentence)	Usage (English translation)
gagala	Adverb	Quickly	Gagala on immey algo.	The days go by so quickly.
appa	Interjection	Oops!	Appa ya ugge daka tinibo.	Oops, I didn't see you.
bahu	Adjective	New	Maphod nan bahun bulwatim.	Your new clothes are good.
hin	Conjunction	If	Hin ume ka, alam tun tudok ku.	If you go, get my letter.
maka	Verb	Come	Maka tu.	Come here.
para	Preposition	For	Hituwe ke ya para ke Juan.	This is for Juan.
pun-agahan	Noun	Hospital	Pun-agahan nan banbana.	The hospital is full.

A selection of sample entries is presented in Table 4 above to show the structure and the content of the Tuwali lexicon. These entries were chosen to highlight the different parts of speech that include verbs, nouns, adverbs, interjections, and etc, as well as the morphological phenomenon. The table includes the Tuwali word, its part of speech, English translation, and sample sentences.

Table 5: Root affix relation

Root Affix Relation		
Root Word	Affixes	Count
tibo	-ni-	7
tigo	-ni-	6

bunga	-um-	3
tubu	-mu-	3

Table 6: Top affixes

Top Affixes	
Affixes	Count
-in-	34
-im-	24
-um-	18
-ni-	15
-al-	9

There are several patterns in the morphology, phonology, and the lexical usage, and in terms the most occurring are affixes like -in-, -im-, -um-, -ni-, which could indicate that these are central in verb derivation or some other grammatical processes. Root-affix pairings such as tibo with -ni- and bunga with -um- suggest certain root words favor specific affixes, which could highlight a consistent derivational rule in the dialect.

Table 7: Meaning Shifts

Meaning Shifts	
Root Word Affixes	Meaning
ahu	small dogs
aaga	many medicine
aga	medicine/cure

There are also meaningful semantic shifts likely due to affixation, for example, the root *aga* transforming into *aaga*, which means many medicines, demonstrating how affixes can convert roots into nouns or modify the meaning.

Table 8: Translation Consistencies

Translation Consistencies		
Tuwali Words	English Words	Tagalog Words
a-ahhu	1	1
aaga	1	1
aamod da	1	1
aam'od	1	1

Translation consistency check reveals that most words have reliable English and Tagalog translations, with only minor polysemy for words like *aapo*, and *abaka*. The part of speech analysis shows that there is a strong verb dominance followed by noun, adjectives, adverb, pronoun, conjunction, preposition, and interjection.

Table 9: Highest frequency

High Frequency Root		High Frequency Words	
nomnom	24	baba	5
tigo	20	bola	5
ala	19	tag-e	4
bangad	18	dama	4

The correlation between affixes and POS demonstrates that certain affixes are strongly associated with verbs; an example of this is *-al-*, and while others may appear in noun formation.

Insights and Observations

The construction of the Tualì lexicon revealed several important patterns about the language. One of the most notable findings was the dominance of verbs, which made up 73.87% of the total entries. This high proportion reflects how Tualì is strongly action-oriented, emphasizing processes and activities that mirror the daily lives of its speakers, from agricultural work to storytelling traditions.

Morphological analysis showed that approximately 18% of the words in the lexicon contained affixes such as -um-, -in-, and -ni-, underscoring the agglutinative nature of the language. These affixes not only adjust grammatical markers like tense or aspect but also reshape the meanings of root words entirely, pointing to the richness of Tualì morphology.

Another key observation was the variation between spoken and written forms. The collection process revealed phonological variations in roughly 9% of entries when comparing data from oral sources such as chants, songs, and folk stories with standardized spellings found in dictionaries. While these pronunciation insights informed our understanding of authentic usage, they were not included as a separate column in the final lexicon. This nevertheless highlights the continuing importance of oral tradition in shaping authentic language use and preserving regional distinctions.

In addition to structural features, cultural vocabulary formed a significant portion of the lexicon. Approximately 21% of entries were tied to ecological and community life, including words describing agricultural tools, native plants, bundles of firewood, and farming practices. This illustrates how Tualì is deeply rooted in the environment and culture of the Ifugao people.

The expressive aspect of the language was also evident in the collection. A small but meaningful set of entries consisted of interjections and exclamations, such as appa ("oops"), which made up 0.35% of the lexicon. While numerically modest, these words are essential for capturing the emotional tone and spontaneity of spoken Tualì.

Finally, the lexicon revealed translation challenges. Approximately 11% of the entries did not have direct equivalents in either English or Tagalog and required longer

descriptive definitions to properly convey meaning, particularly for culturally specific concepts.

Taken together, these insights highlight that the Tuvuli lexicon is not simply a list of words, but a living record of linguistic structure and cultural identity. Its verb-centered system, extensive use of affixes, oral variations, and culturally specific vocabulary demonstrate the richness of the language. At the same time, the need for descriptive translations shows the uniqueness of Tuvuli expressions and the difficulty of capturing them in other tongues. The process of constructing the lexicon thus emphasized both the structural complexity of the language and its role in preserving the heritage and values of the Ifugao people.

CONCLUSIONS

The development of the Tuvuli lexicon has demonstrated the depth, richness, and cultural significance of the language. By systematically compiling entries from dictionaries, oral traditions, songs, poems, chants, stories, videos, and articles, the study created a structured digital resource that documents vocabulary, root words, affixes, meanings, and parts of speech. The analysis revealed that verbs dominate the language, making up 73.87% of entries, which highlights its action-oriented nature, while the frequent use of affixes confirms its agglutinative character.

The research process also acknowledged phonological variations between oral and written forms, which emerged during data collection and validation. This underscores the continuing importance of oral tradition in maintaining linguistic authenticity, even though pronunciation data was not included as a separate column in the final lexicon. At the same time, the strong presence of cultural and ecological terms reflects the deep connection of Tuvuli with the lifeways of the Ifugao people.

The study also identified practical challenges, particularly in translating culturally specific concepts directly into English or Tagalog, which often required descriptive definitions to preserve meaning. Despite these challenges, the lexicon serves as both a linguistic tool and a cultural repository. It provides a foundational resource for future work on indigenous language preservation and opens pathways for practical applications in education, communication, and local governance.

RECOMMENDATIONS

Based on the outcomes of this study, several recommendations are proposed to further enhance the Tawali lexicon and support its ongoing development and application. First, it is recommended that the lexicon be continuously expanded through the inclusion of additional entries drawn from both oral traditions and contemporary usage. Engaging actively with the community, especially native speakers, elders, and cultural practitioners, will help ensure the resource remains authentic, comprehensive, and reflective of the language as it is used today.

To better support language learners, educators, and linguistic researchers, future iterations of the lexicon should consider integrating pronunciation guidance, such as phonetic transcriptions or linked audio recordings, as well as notes on regional and contextual variations in word usage. This addition would capture the spoken dimension of Tawali and provide a more complete linguistic profile.

Second, the lexicon should be fully digitized and made widely accessible through online platforms or mobile applications. Open access to schools, cultural organizations, and local government units can promote Tawali literacy, aid in the translation of local documents, and serve as a foundational resource for educational programs. Digitization will also enable the lexicon to be utilized in computational applications, including machine translation systems and speech recognition tools.

Finally, it is encouraged that future research builds upon this work through comparative studies with other indigenous Philippine languages. Such investigations can highlight shared morphological traits and unique linguistic features, thereby enriching the academic understanding of regional languages and contributing to broader language preservation and revitalization efforts across the Philippines.

ACKNOWLEDGEMENT

The researchers would like to extend their deepest gratitude to all those who contributed to the success of this study. Sincere appreciation is given to their families and friends for their unwavering encouragement, understanding, and support throughout the process. They also acknowledge their peers and classmates for their cooperation, insightful feedback, and motivation that inspired them to improve their work.

Most importantly, the researchers wish to express their heartfelt gratitude to Mrs. Julie M. Guay of Ifugao State University, whose expertise, guidance, and constructive suggestions greatly enhanced the quality and credibility of this study. The careful evaluation, valuable recommendations, and willingness to share time and knowledge were instrumental in refining the research and ensuring its reliability. This contribution is deeply acknowledged and appreciated.

RAW RESOURCES

- Songs
 - ITETEM
<https://lifesunfathomabledepths.blogspot.com/2011/07/itetem-ifugao-tuwali-song.html>
 - HAON YA TAGUD IFUGAO
<https://www.scribd.com/document/234240257/Ifugao-Songs>
- Poems
 - Ahiboto'
<https://ifugaoword.blogspot.com/2013/09/ahiboto-attempt-on-children-poetry-in.html>
 - Mayon
<https://ifugaoword.blogspot.com/2015/11/a-poem-on-mayon-volcano-in-ifugao.html>
 - Tayo (Ifugao Native Dance)
<https://ifugaoword.blogspot.com/2015/11/>
 - Ambet
<https://ifugaoword.blogspot.com/2016/>
 - Uhhun di Duntug / Ibabaw ng Bundok
<https://ifugaoword.blogspot.com/2014/>
 - Limericks an Nalpu hi Muyung (Limericks from the forest)
<https://ifugaoword.blogspot.com/2014/03/limericks-nalpu-hi-muyung-limericks.html>
- Stories
 - Hay Nalpuwan di Galiwgiwon (The Origin of the Galiwgiwon)
<https://ifugaoword.blogspot.com/2020/06/hay-nalpuwan-di-galiwgiwon-legend-of.html>
 - "Nan Batawel Dowel" : An A'apo (Short Story for Children)
<https://ifugaoword.blogspot.com/2014/05/>
 - Pamhod Ina
<https://bloomlibrary.org/player/25VWSIGaf5?lang=ifk>
- Video
 - Tuwali cover of Dia Masa Lalumu by Vionita
<https://www.youtube.com/watch?v=7MOKJMIxd7M>
 - Ed-eddoy

<https://web.facebook.com/watch/?v=2377110112383549>

- Chant
 - Hudhud
<https://filameducation.com/wp-content/uploads/Hudhdud.pdf>
 - Gopah!
<https://ifugaoword.blogspot.com/2020/06/the-gopah.html>
- Others
 - Webonary dictionary
<https://www.webonary.org/tuwali-ifugao/>
 - SIL Dictionary
<https://sil-philippines-languages.org/prog/ifk/dict/index-english/main.htm>
 - BIBLE
<https://www.bible.com/bible/2123/GEN.1.IFKWB>

REFERENCES

- Hohulin, E. L., & Burquest, D. A. (2011). The cross-referencing system of Tuvāli Ifugao: Form and function. SIL International.
https://www.sil.org/system/files/reapdata/15/75/56/157556649957600365164698508152071376873/Hohulin_Burquest_ifk_Cross_Referencing.pdf
- Taleon, K. A. (2020). A phonological sketch of Tuvāli Ifugao. ResearchGate.
https://www.researchgate.net/profile/Kristine-Anne-Taleon/publication/344170392_A_Phonological_Sketch_of_Tuwali_Ifugao/links/5f58568692851c250b9fd47f/A-P_honological-Sketch_of_Tuwali-Ifugao.pdf