Mapping meaning in fantasy lexicons: a computational linguistic exploration of Enchanta - Encantadia language

ISSN 2657-9774; https://doi.org/10.36534/erlj.2025.02.04

Mikaela Louise P. de Guzman*, Jessa Mae A. Dela Cruz**
*Dr. Yanga's Colleges Inc., Philippines; mikee.deguzman@dyci.edu.ph
**Dr. Yanga's Colleges Inc., Philippines; jessa.delacruz@dyci.edu.ph

Abstract
This study investigates the constructed language Enchanta from the Filipino fantasy franchise Encantadia (2005, 2016, 2025), examining its lexicon through computational linguistic methods. Enchanta, devised by Suzette Doctolero, enhances narrative immersion by embedding a culturally symbolic language influenced by Philippine and Romance linguistic features. The lexicon comprises over eighty entries spanning kinship, spirituality, performative commands, and abstract concepts. Using Python-based tools such as NLTK for morphological tagging and phrase parsing, this study analyzes lexical frequency, semantic domains, and recurring syntactic templates (e.g., Verb–Modal–Object). High-frequency expressions such as Ssheda! and Avisala! highlight the language's performative and ritualistic functions. Findings indicate that Enchanta exhibits consistent morphosyntactic patterning and lexical productivity. These patterns can be interpreted through both generative and functional linguistic frameworks, though the present dataset is not designed to adjudicate between competing theoretical models. The study demonstrates that Enchanta operates as a structurally coherent conlang, offering insights into the intersection of fantasy linguistics, cultural semiotics, and computational analysis.

Keywords*: computational linguistics, fantasy lexicon, Encantadia lexicon, morphosyntax*

Introduction

Encantadia, a Filipino fantasy television franchise first aired in 2005 and reimagined in 2016, introduced viewers to a mythical realm divided into elemental kingdoms—air (Aera), water (Adamya), earth (Sapiro), and fire (Lireo), each ruled by powerful female warriors known as Sang'gres. Within this richly imagined universe, the language of Enchanta was born—not only to immerse audiences in its fantasy setting but also to deepen its cultural mythology and narrative coherence (Baldwin & Meyer, 2015).

Constructed languages, or conlangs, serve both artistic and communicative functions in media, often reflecting societal ideals and mythic frameworks through their lexical and syntactic structures (Okrent, 2009). Enchanta is one such language—developed specifically for the series with phonologically melodic forms, symbolic vocabulary, and thematic structuring that echo the lore and philosophy of Encantadia. Its lexicon reveals recurring motifs of kinship, spirituality, authority, and emotional expression. From greetings like Avisala to commands like Ssheda!, Enchanta embodies fantasy storytelling and functions as a tool for immersive world-building (Peterson, 2015).

This study applies computational linguistic methods to analyze the structure and semantic distribution of Enchanta's vocabulary. Using Python-based tools such as NLTK's tree module and WordCloud visualizers, the research generates frequency plots and thematic mappings to reveal lexical priorities, morphological tendencies, and symbolic logic embedded in the language's design (Bird, Klein, & Loper, 2009). Such methods bridge fantasy linguistics and empirical analysis, demonstrating how digital tools enrich our understanding of constructed languages in media.

Theoretical background

This study draws on multiple linguistic frameworks to interpret the structure and function of Enchanta.

Halliday and Hasan (1989, 3) conceptualize language as a *"social semiotic,"* emphasizing its role in encoding cultural meaning and communicative function. Language is shaped by context and reflects social relations and

ideology (Halliday & Hasan, 1989, 12-26). This perspective is particularly relevant to Enchanta, whose lexicon encodes ritual, authority, and kinship.

In contrast, the generative framework conceptualizes language as a rule-governed system. Syntax is described as "a set of rules for generating well-formed sentences" (Chomsky, 1965, 65), highlighting productivity and structural regularity. Enchanta's recurring phrase templates suggest similar rule-based organization (Chomsky, 1965, 64-65).

While these frameworks differ—one emphasizing function, the other form—both recognize systematic patterning in language. Enchanta's structure is therefore amenable to analysis under both perspectives.

Additionally, Jackendoff's (1997, 40-46) lexicalist hypothesis informs this study by proposing that grammatical information may be stored within the lexicon itself. This provides a useful interpretive lens for examining phrase-level constructions in Enchanta.

Methodology

The present study seeks to conduct a systematic analysis of the Enchanta language, as conceptualized within the televised fantasy series *Encantadia*. Specifically, it aims to quantify lexical frequency and identify the predominant semantic domains embedded in the Enchanta lexicon. Furthermore, the research investigates morphological and thematic structures that embody and reflect the socio-cultural constructs inherent to the Encantadia universe, as suggested by Crystal (2004). Lastly, the study highlights the potential of computational linguistics—particularly through parsing techniques and categorization frameworks—as a means of deepening scholarly understanding of invented languages in fictional media narratives.

Lexicon as data: cultural and linguistic semiotics

The Enchanta lexicon comprises over eighty unique words and phrases spanning thematic domains such as:

- **Kinship** (*Ada*, *Ado*)
- **Spiritual Authority** (*Emre*, *Sanctre*)
- **Performative Command** (*Ssheda!*, *Ena-i!*)
- **Abstract Ideals** (*Amarteya*, *Corra*)

These elements demonstrate a ritualistic phonology and consistent morphological roots akin to ceremonial or liturgical language structures (Halliday & Hasan, 1989). Some features parallel natural languages, including:

- **Definite Markers** (*ivi*) reminiscent of Tagalog case particles (Schachter & Otanes, 1972)
- **Gendered/Societal Identifiers** (*Mo-re*, *Sang'gre*, *Encantado*)
- **Temporal Anchoring Lexemes** (*Yanarteya*, *Adoyaneva*)—serving as narrative timestamps

By treating this lexicon as a linguistic dataset, the study utilizes categorization based on semantic fields. Tools from Natural Language Toolkit (NLTK) enable morphological tagging and tree structures to assess lexical depth and distribution, while visualization libraries facilitate interpretation through symbolic mapping

Corpus and data sources

The dataset was derived from a curated corpus of *Encantadia* dialogue across three versions of the series:

- 2005 original (Episodes 1–20)
- 2016 reboot (Episodes 1–10)
- 2025 continuation (Episodes 1–5)

A total of 312 Enchanta utterances were manually identified and transcribed from subtitle files and verified against audiovisual dialogue.

The lexicon (84 entries) was compiled from:

1. Manual transcription of dialogue

2. Official GMA materials
3. Encantadia Wiki (cross-verified)

Translation Basis

Translations were derived from:

- **OS** – on-screen subtitles
- **OM** – official materials
- **CW** – community wiki (verified)
- **CI** – contextual inference

Context-based translations are interpretive and acknowledged as a limitation.

Phrase structure and lexical composition in Enchanta

Each phrase was segmented and aligned with the lexicon to identify its components and syntactic roles, revealing patterned constructions that reflect internal logic akin to natural languages (Crystal, 2004). This consistency validates Enchanta's designed linguistic cohesion as more than narrative artifice—it becomes a model for linguistic structure in media-based conlangs (Okrent, 2009).

Phrase	**Structure breakdown**	**Syntactic roles**
Ssheda musni verom	Ssheda (Stop), Musni (Give), Verom (Object?)	Command → Modal → Object
Ilantre ivi e correi?	Ilantre (Where), Ivi (Definite), E Correi (Love)	Interrogative → Marker → Abstract Concept
Ivi sanctre	Ivi (Definite), Sanctre (Death)	Subject → Predicate
E correi diu	E Correi (Love), Diu (You)	Verb Phrase → Recipient
Avisala meiste	Avisala (Greetings), Meiste (Formal particle)	Ritual Greeting → Modifier

These constructions exhibit root-syntax templates such as Verb + Modal + Object or Interrogative + Marker + Abstract, suggesting a consistent phrase formation system. Such regularity aligns with generative grammar principles that foreground syntax as rule-governed and productive (Chomsky, 1965). Moreover, the presence of performative verbs and pragmatic markers in Enchanta supports Halliday's functionalist view that language is shaped by its social purposes (Halliday & Hasan, 1989).

Implications for lexicon expansion

Phrase-level analysis reveals:

- The productive recombination of root terms into complex expressions;
- Opportunities for syntactic construction mapping in a conlang dictionary;
- Evidence of grammatical consistency and pragmatic encoding.

These findings support Jackendoff's lexicalist hypothesis, which posits that much grammatical information

resides within the lexicon, not merely syntax (Jackendoff, 1997). By treating phrases like *Ssheda musni verom* and *E correi diu* as compositional units, the Enchanta lexicon demonstrates capacity for lexically driven grammar development.

Phrase frequency and distribution

Using a curated corpus of Encantadia dialogue, Python-based frequency analysis identified *Ssheda!* (Stop!) and *Avisala!* (Greetings) as dominant. This mirrors how illocutionary acts recur in ritual and expressive registers (Austin, 1962). Compound expressions like *E correi diu* and *Ivi sanctre* appeared frequently in scenes requiring emotional or dramatic emphasis—reinforcing linguistic functions through narrative pragmatics.

Results and discussion

The quantitative findings reveal that Enchanta, while a constructed and fictional language, operates with linguistic coherence comparable to natural languages. High-frequency phrases such as Ssheda! and Avisala! confirm the lexicon's performative emphasis, especially in commands and ceremonial expressions. The recurrence of key root terms across compound utterances suggests the presence of an internal grammar—one that reflects both morphological productivity and thematic consistency.

Phrase parsing illuminated syntactic patterns: Verb–Modal–Object sequences in Ssheda musni verom, and Interrogative–Marker–Emotion structures in Ilantre ivi e correi?. These configurations mirror grammatical forms common in Austronesian and Romance languages, indicating the influence of Filipino and Spanish linguistic features on the language's design. Such patterns reinforce Enchanta's cultural hybridity while validating its application for linguistic modeling and symbolic analysis.

Table 1: Phrase frequency comparison table

Enchanta Phrase Frequency Comparison Table			
Series Phrase	**Translation / Function**	**Lexicon Match**	**Notes**
Ssheda!	Stop / Halt / Desist	Present	High-frequency command; appears in multiple episodes
Avisala!	Greetings / Goodbye	Present	Used in both formal and casual farewells
Ena-i!	Hurry! / Move quickly	Present	Urgent directive; often in battle or escape scenes

Furthermore, the overlap between the curated lexicon and actual usage in series dialogue shows how root terms serve as functional building blocks. Missing idioms like E correi diu ("I love you") point to an opportunity for lexicon expansion—where affective and narrative nuance can be better captured through compositional analysis.

Overall, this study highlights Enchanta as a symbolically rich and grammatically patterned lexicon, providing fertile ground for linguistic inquiry, especially at the intersection of computation, creativity, and language pedagogy.

NLTK output

A sample parse using NLTK was conducted for the phrase *Ssheda musni verom* using a custom CFG:

S → CMD MODAL OBJ
CMD → Ssheda
MODAL → musni
OBJ → verom

Resulting structure:

(S (CMD Ssheda) (MODAL musni) (OBJ verom))

Frequency distribution:

Token	Frequency
Ssheda	28
Avisala	21
Ena-i	15
Emre	12
Ivi	11

This confirms recurring lexical and syntactic patterns.

Results and discussion

The corpus consisted of 312 utterances, with:

- 87 (27.9%) performative commands
- 43 (13.8%) ritual/greeting expressions

The most frequent terms were:

- *Ssheda!* – 28 (9.0%)
- *Avisala!* – 21 (6.7%)

These findings indicate that Enchanta prioritizes performative and ritual speech, consistent with its narrative function.

Phrase analysis revealed structured patterns:

- Verb–Modal–Object (*Ssheda musni verom*)
- Interrogative–Marker–Abstract (*Ilantre ivi e correi?*)

Such regularity aligns with generative grammar's rule-based syntax (Chomsky, 1965, p. 65), while also supporting Halliday's view of language as socially functional (Halliday & Hasan, 1989, pp. 12-26).

The recombination of root lexemes suggests lexical productivity. These findings are consistent with Jackendoff's lexicalist hypothesis, though they do not decisively distinguish it from alternative syntactic explanations.

Conclusion

This research explored the phrase-level architecture of the Enchanta lexicon through the lens of computational linguistics, demonstrating how a fictional language can reflect deliberate thematic structure and syntactic design. Using Python to quantify phrase frequency, parse grammatical patterns, and compare dialogue expressions with lexicon entries, the study confirms that Enchanta is not only narratively expressive but linguistically coherent.

This study demonstrates that Enchanta exhibits systematic lexical and syntactic patterning despite its fictional origin. Computational analysis reveals consistent phrase structures, semantic clustering, and high-

frequency performative expressions.

These patterns can be analyzed using both generative and functional linguistic frameworks. However, the study does not provide independent empirical support for one theoretical model over another. Instead, it applies these frameworks as interpretive tools to illuminate the language's internal logic.

Future research may expand the corpus, generate full parse trees, and explore comparative analysis with Philippine languages to further evaluate competing linguistic accounts.

These findings underscore the potential of constructed languages to contribute to formal linguistic research and education—particularly when grounded in cultural mythology and applied through digital tools. Future research may build on this foundation by generating full parse trees, tracking discourse features across larger corpora, or integrating Enchanta into comparative studies with indigenous Philippine languages.

The convergence of fantasy lexicon, cultural symbolism, and computational methodology in this study offers a dynamic model for interdisciplinary engagement within language studies.

References

Austin, J. L. (1962). *How to do things with words* (J. O. Urmson & M. Sbisà, Eds.). Oxford University Press.

Baldwin, B., Meyer, T. (2015). Fantasy language and culture in fictional television. *Journal of Media Semiotics, 8*(2), 101-118.

Bird, S., Klein, E., Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Crystal, D. (2004). *The Language Revolution*. Polity Press.

Halliday, M. A. K., Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford University Press.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. MIT Press.

Okrent, A. (2009). *In the Land of Invented Languages: Adventures in Linguistic Creativity, Madness, and Genius*. Spiegel & Grau.

Peterson, D. J. (2015). *The Art of Language Invention: From Horse-Lords to Dark Elves, the Words Behind World-Building*. Penguin Books.

Schachter, P., Otanes, F. T. (1972). *Tagalog Reference Grammar*. University of California Press.

Appendices

All computational analyses were conducted using Python's Natural Language Toolkit (NLTK), with scripts developed for frequency distribution and context-free grammar parsing. A sample implementation is provided in Appendix for reproducibility.

Phrase	Translation	Frequency (n)	%	Lexicon match	Notes
Ssheda!	Stop	28	9.0%	Present	Most frequent command
Avisala!	Greeting	21	6.7%	Present	Ritual expression
Ena-i!	Hurry	15	4.8%	Present	Urgent directive
Emre!	God	12	3.8%	Present	Spiritual invocation
Ivi	Marker	11	3.5%	Present	Functional syntax

Enchanta lexicon (Encantadia, 2005; 2016, 2025)			
Theme	**Enchanta term**	**Translation / meaning**	**Notable linguistic feature**
Kinship & identity	Ada	Mother	Simple root; common in familial contexts
	Ado	Father	Phonological parallel to *Ada*
	I-lo	Grandfather	Affix *I-* possibly age-related marker
	I-la	Grandmother	Gendered variation of *I-lo*
	Mo-re	A person (any gender)	Inclusive reference
	Sang'gre	Royalty	Societal role-specific lexeme
Commands & expression	Ssheda!	Stop / Halt / Desist	Exclamatory with sharp phonology
	Ena-i!	Fast! (imperative)	Hyphenated form indicates emphatic action
	Ashtadi!	Bad behavior!	Expressive reprimand
Greetings & ritual	Avisala	Greetings / Goodbye	Ritualistic dual-use form
	Emre	God / Bathala	Root for divine authority
Temporal concepts	Yanarteya	Past	*-teya* suffix suggests temporal framing
	Adoyaneva	Future	Prefix *Ado-* (root?) with future affix *-yaneva*
Places & mythology	Encantadia	Land of new beings	Root *En-* repeated in language terms

	Enchanta	The language of Encantadia	Meta-lexeme; language as cultural emblem
	Lireo	Kingdom of fairies (Diwata)	Proper noun with ethereal phonetics
Emotional/ abstract	Corra	Heart	Symbolic core of emotional vocabulary
	Amarteya	Freedom	Possibly from Spanish *amar* (to love) + *teya*
	Detrumvia	Liar	Polysyllabic moral judgment lexeme
Nature & objects	Ganto	Horse	Lexical borrowing or stable root term
	Kantao	Bracelet	Noun used in physical description or symbolism
Functional syntax	ivi	Definite marker (like 'si' in Tagalog)	Syntactic particle for nominal emphasis

Enchanta phrase frequency comparison table			
Series phrase	**Translation / function**	**Lexicon match**	**Notes**
Ssheda!	Stop / Halt / Desist	Present	High-frequency command; appears in multiple episodes
Avisala!	Greetings / Goodbye	Present	Used in both formal and casual farewells
Ena-i!	Hurry! / Move quickly	Present	Urgent directive; often in battle or escape scenes

Emre!	God / Bathala	Present	Invoked in prayers and blessings
Musti maste Lireo!	Peace to Lireo	Partial (*Musti*)	Compound phrase; modal + place name
Ivi sanctre	He is dead	Partial (*ivi*, *Sanctre*)	Combines definite marker and death term
Estasectu!	Get ready	Not listed	Possible future addition to command lexicon
Ebi meshne	You will regret this	Not listed	Threat phrase; expressive and dramatic
Avisala meiste	Formal goodbye	Partial (*Avisala*)	Extended farewell variant
E correi diu	I love you	Not listed	Emotional declaration; potential for abstract category
Ssheda musni verom	I won't give you what you want	Partial (*Ssheda*, *Musti*)	Complex command with modal negation
Ilantre ivi e correi?	Where is love?	Partial (*Ilantre*, *ivi*)	Interrogative + definite marker + abstract noun
Paneya	Bread	Present	Everyday object; appears in domestic scenes
Detrumvia	Liar	Present	Used in confrontations and accusations
Asshenti	Listen	Present	Often used in tense or emotional dialogue
Yanarteya	Past	Present	Temporal reference; used in flashback scenes
Adoyaneva	Future	Present	Temporal reference; used in prophecy or planning