Automatic Content Generation Based on Predefined Grammatical Rules

Radu Bucea Manea Tonis[1]

Abstract: While traditional methods rely on deciphering ancient texts and comparative linguistics, computational techniques have increasingly played a crucial role in language reconstruction. One powerful approach uses Definite Clause Grammars (DCGs) in Prolog, a Turing-complete logic programming language designed for symbolic reasoning. DCGs extend Prolog's capabilities by offering a declarative way to define grammatical rules. Unlike conventional context-free grammars, DCGs allow for additional logic and flexibility, making them well-suited for modeling natural languages. In the case of reconstructing lost languages, researchers can define grammar rules based on fragments of known linguistic evidence, such as surviving texts, inscriptions, or spoken dialects closely related to the extinct language. The significance of this study lies mostly in its potential application for detecting deceptive content, such as "fake news."

Keywords: Prolog; Definite Clause Grammar (DCG); Linear A.

1. Introduction

Automatic content generation based on predefined grammatical rules involves using structured language models that rely on formal syntactic patterns to produce text. Rather than learning from vast datasets, these systems work by combining sentence templates, vocabulary lists, and grammar rules to generate linguistically accurate and

[1] Danubius International University, School of Behavioral and Applied Sciences - Informatics, Address: Bd. Galaţi, Nr. 3, 800654, Romania, Corresponding author: radumanea@univ-danubius.com.

predictable outputs. Common techniques include syntax trees, rule-based parsing, and lexical selection based on part-of-speech tags.

While these systems excel in applications where clarity and consistency are paramount—such as language education tools, structured reports, or basic chatbot responses—they often lack creativity and contextual flexibility. However, with evolving approaches that integrate rule-based logic into broader AI frameworks, we're seeing new possibilities for personalized, multilingual, and hybrid content generation. It's a space where linguistic precision meets evolving innovation.

Linear A is the undeciphered script used by the Minoan civilization on Crete, predating Linear B. Unlike its successor, which has been linked to early Greek, Linear A remains a linguistic enigma, with no known direct descendants and limited corpus material to study. But here's the twist: even in its cryptic nature, the idea of predefined grammatical rules can still be a powerful lens through which scholars attempt to decode it.

Rule-based content generation can be assisted by applying patterns and structural consistencies across known inscriptions. Researchers analyze symbol frequency, sequence regularity, and placement on tablets to hypothesize grammatical constructs—essentially creating "synthetic grammars" for simulation. These patterns, processed through computational linguistic models, may identify repetition, syntax scaffolding, or even clues to semantic categories.

2. Literature Review on Old Europe Linguistic Research

Csaba Barnabas Horvath examines the linguistic landscape of Western Europe before the spread of Indo-European languages. His research synthesizes insights from archaeology, historical linguistics, and genetics to reconstruct the migrations and interactions of prehistoric language families. (Horvath, 2019) argues that Vasconic (related to Basque) and Tyrsenian (associated with Etruscan) languages existed in Western Europe before the expansion of Indo-European languages. He also explores the correlation between Y-DNA lineages and linguistic shifts, suggesting that genetic evidence supports the presence of distinct pre-Indo-European populations. The paper integrates findings from material culture to trace the movement of language families across Europe from the Copper Age to the Roman conquest.

In (Verhasselt, 2009, pp. 211–239), the linguistic elements in Greek that do not adhere to established phonetic laws are examined. It reviews key theories regarding

the origins of these elements, such as the Pelasgian Theory, which reconstructs Pre-Greek as an Indo-European language with distinct phonetic rules; the Anatolian-Luwian Hypothesis, which posits that foreign elements in Greek vocabulary originate from an Anatolian substratum; the Aegean Theory, currently the most widely accepted, suggesting that Pre-Greek is neither Indo-European nor Semitic; and the Kartvelian Theory, a more recent perspective that identifies Pre-Greek as predominantly related to Kartvelian (South Caucasian) languages. Gertjan Verhasselt emphasizes how Greek civilization incorporated elements from a pre-Greek culture, including place names, cultural concepts, and even deities. Additionally, it addresses the ongoing debate surrounding the origins of Poseidon and the broader significance of linguistic borrowings.

Richard Janke examines the impact of Hittite and Luwian linguistic elements on Minoan Linear A inscriptions. His study posits that Linear A, often regarded as undeciphered, exhibits structural and phonetic similarities with Anatolian languages, particularly Hittite and Luwian. This (Janke, 2022) supports the hypothesis that Minoan Linear A was not an isolated script but rather part of a broader linguistic network within the eastern Mediterranean. Key findings include the concept of linguistic addstratum, where Linear A, Hittite, and Luwian coexisted as synchronous influences, each impacting the others without direct derivation. Additionally, the study highlights the religious and ritual contexts of Linear A inscriptions, many of which appear in libation invocations, suggesting that religious terminology was likely borrowed from Anatolian sources. Finally, it speculates that Minoan scribes may have employed various writing conventions, potentially influenced by Hittite cuneiform practices.

Naomi Carless Unwin delves into the historical and cultural connections between Caria and Crete, highlighting their shared mythologies, religious practices, and maritime exchanges. In (Unwin, 2017, pp. 1-31), ancient traditions that often link Caria with Crete are investigated, particularly through common cults and deities. Additionally, material evidence indicates a sustained interaction between Carian and Cretan communities, driven by trade and migration that shaped the cultural exchanges between these two regions.

Margalit Finkelberg examines in (Finkelberg, 1990) the linguistic features of Linear A inscriptions found on ritual vessels. The study suggests that Linear A shares phonetic similarities with Mycenaean Linear B, indicating a closer relationship between the two scripts than previously assumed. Also, many inscriptions contain names and terms that appear in both Linear A and Linear B texts, reinforcing the

idea of continuity between Minoan and Mycenaean writing systems. While Linear A remains undeciphered, the study argues that applying phonetic values from Linear B has led to significant progress in identifying words and meanings.

3. Methodology

I used Prolog - a logic programming language designed for NLP - to analyze words from the Minoan language (Lin. A) and to determine whether OM (Old Minoan) sentences are well-formed or not. This can be achieved by constructing a specially designed file called grammar. In the chosen file, one can define the syntactic structure of an utterance (for example, SVO or VSO) and analyze morphologically the parts of speech (e.g., nouns, verbs, adjectives, etc.). This way, I could write a recursive parser to define the OM grammar, formalized in nine clauses, according to (Bucea-Manea-Ţoniş et. al., 2020):

```
s --> vp, np.
vp --> v, np.
vp --> v.
np --> nex.
np --> det, nex.
ne --> n.
ne --> n, suf.
nex --> ne.
nex --> ne, obj.
```

The first clause outlines the expression s used for identifying sentences with a VSO order. The second clause discusses the verb particle vp, which includes both the verb itself v and a predicative noun np. The subordinate clause consists of a prefix serving as a determiner det or a suffix suf along with a nominal expression nex. A noun and a qualifier/adjective obj in either the dative or genitive case form a nominal expression. After (Younger, 1998), the term SU-KI-RI-TE-I-JA, found on a pot and potentially labeled Sybritan, may trace its origins to the place name SU-KI-RI-TA, but I strongly suggest a possible IE etymology that links it to SACER (lat. for sacred, holy, or consecrated) and Su-kīrti (सुकीर्ति) meaning good fame or noble reputation.

I considered determinants like i- and j- (i + consonant, j + vowel), which prefix a noun with the meaning of from/to after (Duhoux, 1997, pp. 29-30; 289-94), and those like -te and -ti as being from/to, as demonstrated by (Valério, 2007, pp. 3-14). Changes from U-NA-RU-KA-NA-SI to U-NA-RU-KA-NA-TI seem to reflect a shift to the plural, according to (Younger, 2024).

The objects are in the genitive or dative case, as indicated by (Janke et. al., 2018). Therefore, we need only to initialize our lexical cases in Prolog, as may be seen in Table 1:

Table 1. Minoan vs. Koine Greek grammars

Minoan (proposed) grammar	Greek (Koine) grammar
% Sentence structure s --> simple_s. s --> simple_s, s. % Simple sentence structure simple_s --> vp, np. % Noun phrase structure np --> n, obj. np --> det, n, obj. % Includes article (determiner) % Verb phrase structure vp --> v, obj. vp --> v. % Lexicon % Nouns n --> [o-su-qa-re]. % Proper noun (subject) n --> [i-pi-na-ma]. % Proper noun n --> [ja-sa-sa-ra-me].% Proper noun n --> [a-re-pi-re-na]. % Proper noun % Verbs v --> [a-ta-i-301-de-ka]. % Verb v --> [u-na-ka-na-si]. % Verb % Objects (cases) obj --> [su-ki-ri-te-i-ja]. % Genitive obj --> [ku-pa3-ri-ja]. % Genitive obj --> [pa-sa-ri-ja]. % Genitive obj --> [a-si-da-to-i]. % Dative obj --> [sa-ke-ri]. % Dative obj --> [da-ma-te]. % Dative obj --> [si-ru-te]. % Dative % Determiners (Articles)	% noun phrase np --> art, n. np --> n. % verb phrase vp --> v, np. vp --> v. % lexicon art --> [ὁ]. % “the” (masculine singular) art --> [ἡ]. % “the” (feminine singular) art --> [τό]. % “the” (neutral singular) n --> [ἀνήρ]. % “man” n --> [γυνή]. % “woman” n --> [παιδίον]. % “child” v --> [βλέπει]. % “sees” v --> [ἀγαπᾷ]. % “loves” v --> [ἔχει]. % “has”

```
det --> [i_].  % Possible article (context-
dependent)
det --> [j_].  % Possible article

% Suffixes (morphology)
suf --> ['-te']. % Suffix variant (quoted for
Prolog safety)
suf --> ['-ti']. % Suffix variant        %
sentence structure
s --> np, vp.
```

The grammar is consulted in Prolog using the command:

?- consult('C:/Users/radub/Desktop/Minoan/Minoan_Greek_gram.pl').

To test if the VSO command applies to a phrase according to grammar, type:

| ?- phrase(s, [u-na-ka-na-si, o-su-qa-re]).

no

,or:

| ?- phrase(s, [u-na-ka-na-si, ja-sa-sa-ra-me, pa-sa-ri-ja]).

true ?

yes

To test a more complex sentence, such as a libation, we will try the following:

| ?- phrase(s, [a-ta-i-301-de-ka, o-su-qa-re, su-ki-ri-te-i-ja]).

true ?

yes

After loading the grammar into memory, it is possible to run all combinations of terms in grammatically correct sentences with the following command:

| ?- findall(Sentence, phrase(s, Sentence), Sentences).

Sentences = [[a-ta-i-301-de-ka, o-su-qa-re, su-ki-ri-te-i-ja],

[u-na-ka-na-si, ja-sa-sa-ra-me, si-ru-te],

Explanation:

- findall/3 collects all possible sentences based on our grammar.
- A sentence represents an individual sentence structure.

- Sentences stores all valid sentences that Prolog generates.
- phrase(s, Sentence) ensures Prolog follows the DCG rules to produce well-formed Minoan Greek structures.

4. Conclusion

Based on the literature review and our research, we can conclude that Classical Greek seems to be part of the Anatolian group of languages, with a substratum of a pre-IE Aegean language (Thyrenian?) and an IE addstratum from the beginning of the Iron Age (circa 900 BC, e.g., Dorians), supported by the following factors:

- Geographical context: The birthplace of Greek mythology, such as the tale of Zeus, is situated closest to Anatolia. This area served as a crossroads for tin trade routes since the Bronze Age, connecting the East to the West (from Cyprus to the eastern coast of Spain) and the North to the South (from the British Isles and the Baltic to Egypt).
- Linguistic evidence: The terminations and declensions, such as -os(gr.)/-atta(hit.)/-assa(luv.), indicate a close relationship between the Greek language and a Lydian/Carian substratum. Furthermore, the symbols used in Carian inscriptions do not correspond phonetically with those in Linear B, except for four distinct instances. This suggests that the Carians migrated to Crete (Caphtor) before the arrival of the Greeks on the island.
- Historical context: Herodotus connects the Cretans to the Carian people through the mythological narrative of Zeus and Europa, a Phoenician queen, from whom two significant figures descended: Minos, inheriting Crete, and Sarpedon, inheriting Caria. The eruption of the Thera volcano around 1600 B.C. is believed to have prompted a mass migration of the Cretan population towards the Anatolian coasts, particularly the islands of Kos and Kios. This migration transformed the once-flourishing Minoan commercial fleet into a band of marauding Carians and Achaean seafarers, part of the well-known Sea Peoples. Those who migrated to Canaan may have laid the foundation for a new maritime empire, also known as the Phoenician thalassocracy.

Despite its potential, reconstructing lost languages using Definite Clause Grammars (DCGs) faces certain limitations. The biggest challenge is the insufficient data available, as many ancient languages have left behind scant evidence. Furthermore,

linguistic evolution is a complex phenomenon that necessitates extensive interdisciplinary collaboration between computational linguists and historians. Nevertheless, with advancements in artificial intelligence and natural language processing, DCGs in Prolog could act as a foundational tool for the reconstruction of extinct languages. By integrating formal grammar models with historical context, researchers may ultimately unveil the mysteries of forgotten tongues, breathing new life into lost languages.

The approach to analyzing undeciphered scripts like Linear A—particularly through predefined grammatical rules—provides a compelling framework for identifying deceptive content, such as fake news. In both contexts, researchers work with a limited semantic understanding, relying significantly on patterns, syntactic structures, and symbol frequency. Just as scholars studying Linear A infer grammar and potential meanings from recurring forms and the placement of symbols, analysts examining online content utilize grammatical regularities, stylistic cues, and anomalies in sentence construction to detect manipulation. Rule-based systems can identify the overuse of emotionally charged verbs, suspicious modal constructions, or syntactic irregularities that frequently characterize misleading information.

By generating synthetic grammars for both ancient inscriptions and contemporary digital text, we engage in a shared methodology grounded in structured pattern recognition. This analytical approach not only seeks to uncover falsehoods but also achieves a greater level of explainability. Unlike opaque AI models, rule-based systems provide transparent insights into the reasons a particular piece of text may raise concerns. The symmetry between deciphering Linear A and identifying misinformation is not merely metaphorical; it is methodological, deeply rooted in linguistic rigor, structural precision, and a commitment to extracting truth from ambiguity.

References

Bucea-Manea-Țoniș, Ro., Lianu, C., Bucea-Manea-Țoniș, R., Gudei, C. & Lianu, Co. (2020). *Innovation and digitalization from a managerial perspective.* Economic Publishing House.

Duhoux, Y. (1997). *LA > B da-ma-te = Déméter? Sur la langue du linéaire A.* (pp. 29-30: 289-94) Minos.

Finkelberg, M. (1990). *Minoan Inscriptions on Libation Vessels. Department of Classics*. Tel-Aviv University. https://www.academia.edu/5650986/Minoan_Inscriptions_on_Libation_Vessels.

Horvath, C. B. (2019). Redefining Pre-Indo-European Language Families of Bronze Age Western Europe: A Study Based on the Synthesis of Scientific Evidence from Archaeology, Historical Linguistics and Genetics. *European Scientific Journal*, *15*(26). Doi:10.19044/esj.2019.v15n26p1

Janke, R. (2022). The influence of Hittite and digraphia on Minoan Linear A proto-Greek libation invocations. *Preliminary Findings Book*. Western University. https://www.researchgate.net/publication/359722498

Janke, R., Solcà, A., Bengtson, J., Binnberg, J., Gingras, J-P., Canada, R. & Koslenko, R. (2018). High *Correlation Linear A-Linear B vocabulary, grammar and orthography*., In Linear A. Board of Editors/Conseil des rédacteurs. https://www.researchgate.net/publication/347268190_High_Correlation_Linear_A-Linear_B_vocabulary_grammar_and_orthography_in_Linear_A_Board_of_EditorsConseil_des_redacteurs/citations.

Unwin N.C. (2017). Approaching the Topic of Carian–Cretan Interaction. Caria and Crete in Antiquity. *Cultural Interaction between Anatolia and the Aegean*. pp. 1-31. https://www.cambridge.org/core/product/0180814C922F1B9BA44DC7EB9175620D

Valério, M. (2007). Diktaian Master. *A Minoan Predecessor of Diktaian Zeus în Linear A*. *Kadmos,* 3-14.

Verhasselt, G. (2009). The Pre-Greek Linguistic Substratum. An Overview of Current Research. *Les Études Classiques*. *77*(3), 211–239.

Younger, J. (1998). Review of Aegean Prehistory VII: Neopalatial, Final Palatial, and Postpalatial Crete. *American Journal of Archaeology*. Doi: https://doi.org/10.2307/506138.

Younger, J. (2024). Linear A Texts & Inscriptions in phonetic transcription Commentary. *Comments, corrections, questions*. https://www.academia.edu/117949876/Linear_A_Texts_and_Inscriptions_in_phonetic_transcription.