



Parsing Greek Language with Link Grammar

Chorozoglou G.¹, Zacharis N.¹, Mikros G.² and Papakitsos E.³

¹Department of Informatics and Computer Engineering, University of West Attica, Greece

²College of Humanities and Social Sciences, Hamad Bin Khalifa University, Qatar

³Department of Industrial Design and Production Engineering, University of West Attica, Greece

Corresponding Author: Papakitsos E.

ABSTRACT: Link Grammar is a computational approach-method for syntactic parsing developed in 1991 at Carnegie Mellon University, initially for the English language. It is a Context-Free Grammar according to Chomsky's classification. For a sentence to be syntactically parsed correctly by Link Grammar, links (connections) between adjacent words must have been declared and established, in terms of coding. Besides English, Link Grammar has been implemented in Arabic, German, Hebrew, Indonesian, Kazakh, Lithuanian, Persian, Russian, Thai, Turkish, and Vietnamese. In this paper, an initial attempt is presented to implement Link Grammar in the Greek language, by using the Link Grammar Parser. We aim to include most syntactic structures of Greek by providing numerous examples. By applying the same notation used in English grammar and adding some extra symbols, we can achieve syntactic parsing for a variety of sentences, as demonstrated herein.

KEYWORDS: Link Grammar, Syntactic Parsing, Greek Language, Parser

Received 14 Dec., 2024; Revised 26 Dec., 2024; Accepted 28 Dec., 2024 © The author(s) 2024.

Published with open access at www.questjournals.org

I. INTRODUCTION

The earliest writing systems were pictographic (pictograms, ideograms & hieroglyphics), representing objects. These were followed by logographic systems (symbols) representing words. Later, syllabic scripts, such as Linear B [1], corresponded to syllables, while alphabetic systems used symbols to represent phonemes or sounds. Alphabetic systems required fewer symbols (letters) to represent spoken language. Various letters form words, and words form sentences.

A foundational assumption in modern syntactic theory is that sentences in a language are not merely sequences of elements but are hierarchically structured component-based structures. Permissible structures are described by grammar, which is a set of phrase structure rules that determine how sentences are formed from phrase categories and how these categories, in turn, derive from lexical categories [2].

II. INTRODUCING LINK GRAMMAR

This article addresses syntactic parsing, identifying each component, and assigning syntactic roles (e.g., subject, object) using Link Grammar, developed by Daniel Sleator and Davy Temperley [3], in 1991 at Carnegie Mellon University for English language. Link Grammar is a Context-Free Grammar [4] according to Chomsky's classification. This is the first implementation of Link Grammar for Greek, with no prior literature on the topic. Besides English, Link Grammar has implementations in Russian, Thai, Arabic, German, Persian, Hebrew, Indonesian, Kazakh, Lithuanian, Turkish, and Vietnamese [5]-[7].

According to this grammar we aim to connect all words in a sentence with links, achieving a linkage. Then we say that the sentence belongs to grammar. Each link represents a line connecting two words. Each word has required links, described by an ordered sequence of letters (capital or lowercase) and special operators ('&', 'or', '@', parentheses, brackets) encoding different syntactic phenomena. The set of different links corresponding to a word is called its formula. All words with their formulas form the grammar's lexicon, which can continuously expand with new words.

The letters are followed by a '+' or '-', indicating a connection to the right or left. The order of the letters determines how far or close the connections are.

For a sentence to be syntactically parsed correctly by Link Grammar [8], the following must hold:

- Links between words must not intersect (*Planarity*).
- Links should suffice to connect all words in a sentence (*Connectivity*).
- The links in a word's type are ordered so that the closer the links are, the closer the connected words are (*Ordering*).
- No two links can join the same pair of words (*Exclusion*).

It should be noted that Greek, compared to English, has richer morphology (inflection) and relatively free word order [9]. This means that word types will be more complex [10]. To maintain some continuity with English Link Grammar, we will use similar English letters for links and new symbols to enhance types.

III. PARSING GREEK LANGUAGE WITH LINK GRAMMAR

Let's start by examining simple sentences to understand Link Grammar. In the following sentence the verb 'έφαγε' (ate) connects the subject 'γάτα' (cat) and the object 'ποντίκι' (mouse). Articles 'Η' (the, feminine) and 'το' (the, neuter) accompany the subject and object:

Η γάτα έφαγε το ποντίκι (1)

Using *Link Grammar*, we define the links **S** and **O** to specify the subject and object respectively. Articles belong to the category of *Determiners*, and we use the symbol **D**. The syntactic analysis is shown in Fig. 1.

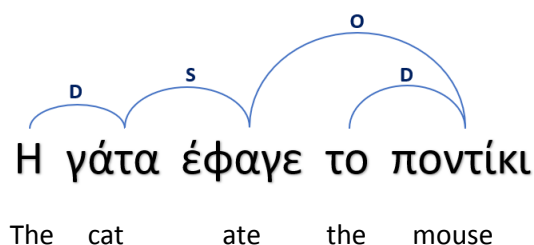


Figure 1: Parsing of a simple S-V-O sentence (1)

It should be noted that the lines connecting the various words in sentences are not directional as in Dependency Grammars, but simple lines. Since Greek has a nearly free word order, the above sentence can be written differently:

Το ποντίκι έφαγε η γάτα (2)

This does not mean that the *mouse* ate the *cat*, but the object 'ποντίκι' (mouse, accusative case) precedes the subject 'γάτα' (cat, nominative case). If we denote **S** for the subject, **V** for the verb and **O** for the object, we have the following ways of writing the verb together with languages [11] that use it (Table 1):

Table 1: Examples of languages and word-ordering

i	V – S – O	Philippines - Filipino
ii	S – O – V	Japan - Japanese
iii	S – V – O	France - French
iv	O – S – V	Papua New Guinea - Tobati
v	O – V – S	Peru - Urarina
vi	V – O – S	Brazil - Hixkaryana

In Greek we usually use (iii) case. But there are also cases where we use (i) & (vi) [*Έφαγε η γάτα το ποντίκι* - *Έφαγε το ποντίκι η γάτα*] or (v) [*Το ποντίκι έφαγε η γάτα*] in the spoken word for emphasis. The syntactic analysis for the second sentence (2) is shown in Fig. 2:

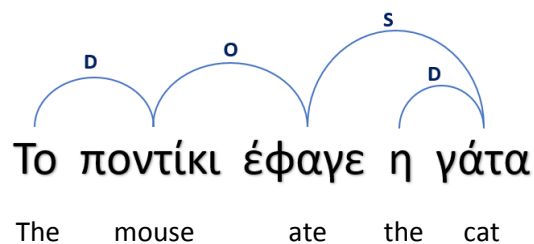


Figure 2: Parsing of a simple O-V-S sentence (2)

Enriching the first sentence we can add two adjectives for object and subject. So, the clause becomes:

Η μαύρη γάτα έφαγε το άσπρο ποντίκι (3)

With the following syntactic analysis (Fig. 3):

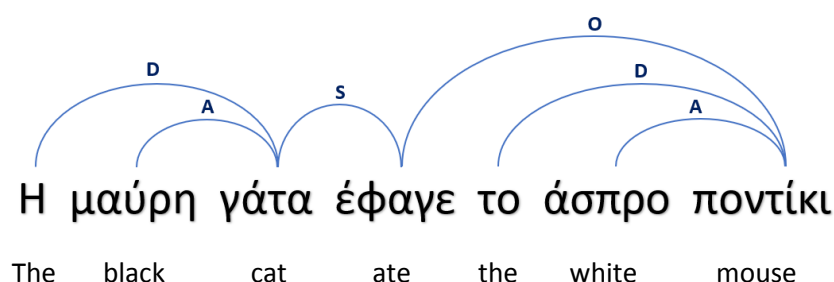


Figure 3: Parsing of a common S-V-O sentence (3)

Link **A** refers to adjectives. Syntactic analysis according to *Link Grammar* does not produce syntactic trees but lines connecting words. We will try to make the dictionary, i.e., the formulas for all words (Table 2):

Table 2: Words of sentence (3) and their formulae

Word	Formula
η	D+
μαύρη	A+
γάτα	{A-} & D- & (S+ or S-)
έφαγε	(S- & { O+ }) or (O- & S+)
το	D+
άσπρο	A+
ποντίκι	{A-} & D- & (S+ or S-)

The parentheses define priority, brackets indicate optional elements, and ‘**or**’ specifies alternatives. The dictionary can be enriched with new words and new links. The order of the links relates to the proximity of terms connected to the word. The further to the left the links are, the closer the connected terms are. The ‘+’ signifies a connection to the right, while the ‘-’ signifies a connection to the left.

So far, we have not considered the case of nouns and the conjugation of verbs. We know that the subject requires the nominative case, while the object requires the accusative one. This means we need agreement in case and gender for the syntactic analysis to be correct. To achieve agreement in number, case, and gender, we use lowercase letters following uppercase ones, which we call subscripts. Thus, we have:

- Singular Number: **s**,
- Plural Number: **p**
- Nominative Case: **n**,
- Genitive Case: **g**,
- Accusative Case: **a**
- Masculine Gender: **m**,
- Feminine Gender: **f**,
- Neuter Gender: **n**

The link, e.g., **Dsam**, refers to a connection of a singular article in the accusative case and masculine gender with the corresponding noun. In Table 3, we can see some examples of such connections.

Table 3: Examples of links with substrings

τον Dsam+ (the)	μαθητή Dsam- (student)	τον μαθητή (the student)
τα Dpnn+ (the)	βιβλία Dpnn- (books)	τα βιβλία (the books)
τις Dpaf+ (the)	γάτες Dpaf- (cats)	τις γάτες (the cats)

The connection ‘*τον μαθητής*’ is not possible because the word ‘*τον*’ has the link **Dsam+**, and the word ‘*μαθητής*’ has the link **Dsnm-**, meaning they do not match in case (**a** vs. **n**). In Table 4 we have all the substrings for link **D**.

Table 4: Substrings of link D

Dsnm	Dsnf	Dsnn	Dpnm	Dpnf	Dpnn
Dsgm	Dsgf	Dsgn	Dpgm	Dpgf	Dpgn
Dsam	Dsaf	Dsan	Dpam	Dpaf	Dpan

We can understand that connections are more numerous compared to the English language. It should be noted that in Greek language, some articles are shared across certain cases and genders. For example, in the neuter gender the nominative case is the same as the accusative case in both singular and plural forms. In Table 4, all such cases are listed because, although they may appear to be the same words, they represent different cases and correspond to subjects and objects. A similar logic applies to adjective links, e.g., **Asnf+** for the adjective ‘άσπρη’ (*white*), which is singular number, nominative case and feminine gender.

Similarly, for subject link **S**, we account for singular or plural numbers and first, second, or third person, as in Table 5.

Table 3: Substrings of link S

Ssa	Spa
Ssb	Spb
Ssc	Spc

3.1 Simple Examples in Greek Language

Based on all the above, we will begin syntactic analysis with simple examples and attempt to analyze as many grammatical phenomena as possible. We will use the Link Grammar Parser described on the website of Link Grammar Parser [12].

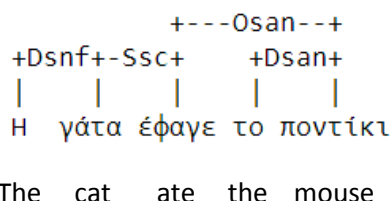


Figure 4: Example of parsing a simple S-V-O sentence with Link Grammar

In Fig. 4, the subject (‘Η γάτα’ – The cat) appears in the nominative case, and the object (‘το ποντίκι’ - the mouse) in accusative case. In the **Ssc** link, **s** denotes the singular number, and **c** refers to the third person. Including adjectives results in a more complex analysis (Fig. 5).

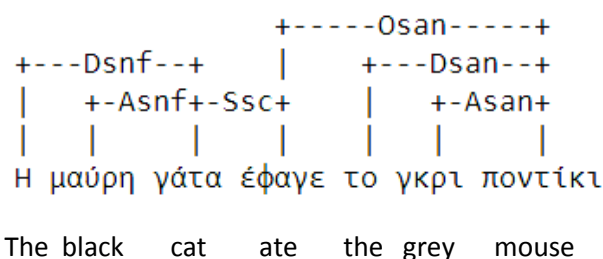


Figure 5: Example of parsing a common S-V-O sentence with Link Grammar

The English language has a relatively fixed word order, whereas Greek features more flexible syntax, inflections, and greater morphological complexity. The following three examples [Fig. 6 (a), (b), (c)] are characteristic.

+---Ssc---+
+Dsan+-Osan-+ +Dsnf+
| | | | |
Το ποντίκι έφαγε η γάτα

The mouse ate the cat

(a)

+-----Osan-----+
+---Ssc---+ |
| +Dsnf+ +Dsan+
| | | | |
Έφαγε η γάτα το ποντίκι

Ate the cat the mouse

(b)

+-----Ssc-----+
+---Osan---+ |
| +Dsan+ +Dsnf+
| | | | |
Έφαγε το ποντίκι η γάτα

Ate the mouse the cat

(c)

Figure 6: Examples of parsing a simple sentence of different word-order with Link Grammar

In (a), we have the case **O – V – S**, in (b), the case **V – S – O**, and in (c), the case **V – O – S**. These are commonly used in spoken language and are rarely encountered in written language. Indicatively, we can observe the formula of the nouns ‘γάτα’, ‘ποντίκι’ and the verb ‘Έφαγε’:

Γάτα (cat): ($\{ @Asnf- \} \& Dsnf- \& Ssc+ \}$) or ($\{ @Asnf- \} \& Dsnf- \& \{ Ssc- \}$) or ($\{ @Asaf- \} \& Dsaf- \& \{ Osaf- \}$ or $Osaf+ \}$);

Ποντίκι (mouse): ($\{ @Asnn- \} \& Dsnn- \& Ssc+ \}$) or ($\{ @Asan- \} \& Dsan- \& Osan- \}$) or ($\{ @Asan- \} \& Dsan- \& Osan+ \}$);

Έφαγε (ate): ($\{ Ssc- \} \& \{ O*a*+ \}$) or ($\{ Ssc+ \} \& \{ O*a*- \}$) or ($\{ O*a*+ \} \& \{ Ssc+ \}$) or ($Ssc+ \& \{ O*a*+ \}$);

We observe that it covers all the cases shown in Fig. 5, with the subjects and objects in various positions, adjectives, as well as the case of the nouns ‘γάτα’ (cat) and ‘ποντίκι’ (mouse), which share the same form in the nominative and accusative. By enriching the lexicon and analyzing the structure of the Greek language, the forms will become more complex.

We can very easily form tenses that are constructed periphrastically, such as the future tense, the present perfect, the past perfect, and the future perfect. We begin with the sentences:

- Έχω φάει την σαλάτα (4)
- Είχα φάει την σαλάτα (5)
- Η Μαρία θα φάει την σαλάτα (6)
- Θα έχω φάει την σαλάτα μέχρι αύριο (7)

which have the analyses shown in Fig. 7 and 8.

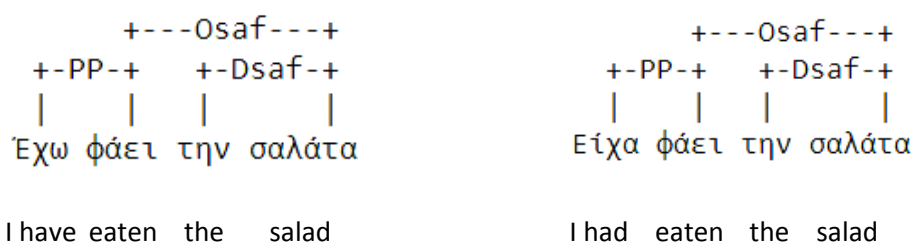


Figure 7: Examples of parsing a simple sentence of different tenses with Link Grammar

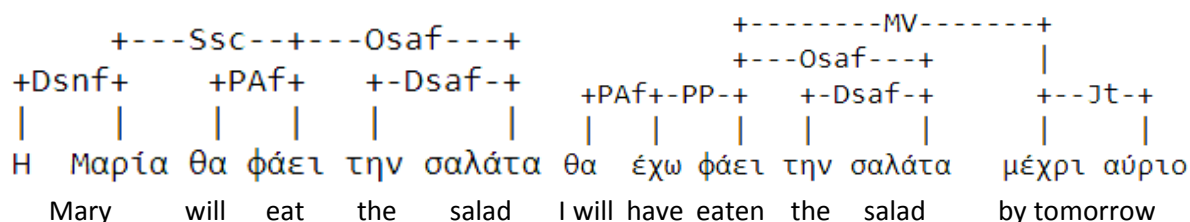


Figure 8: Examples of parsing a common sentence of different tenses with Link Grammar [sentence (6) on the left; sentence (7) on the right]

We observe the new links **PP**, **PAf**, **MV** and **Jt**. The **PP** link connects the past participle with the auxiliary verb 'έχω' - have ('είχα' – had) for the formation of the present perfect (and past perfect), the **PAf** link connects the particle 'θα' (will) to the verb for the formation of the future tense, the **MV** link pertains to modifiers, and the **Jt** link forms prepositional sentences:

θα (will): *PAf+*;

Έχω (have) είχα (had): *{PAf-} & {Ssa-} & {PP+ or O*a*+}*;

φάει (eaten): *{PAf-} & {Ssc-} & {PP-} & {O*a*+} & {MV+}*;

μέχρι (by): *MV- & Jt+*;

Αύριο (tomorrow): *J-*;

It is noteworthy that sentences (4), (5) and (7) are elliptical, meaning the subject is omitted. In the formulas we see that **Ssa** is in brackets, indicating that the subject may be absent.

The links **P**, **M**, **N**, **KT** and **DP** refer to the predicate, the modifiers of nouns and adjectives (as opposed to **MV** for verbs), negations, possessive pronouns, and demonstrative pronouns, respectively. Their usage is illustrated in the following examples (8)-(10), with their analyses shown in Fig. 9 and 10.

Ο Κώστας δεν είναι ο άντρας μου (8)

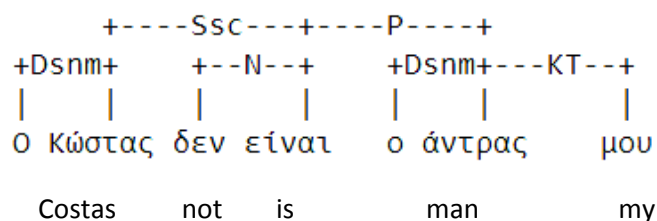


Figure 9: Example of parsing a sentence of negation (8) with Link Grammar

Τα βιβλία της Φυσικής του Κώστα (9)

Αυτά τα ωραία κόκκινα βιβλία (10)

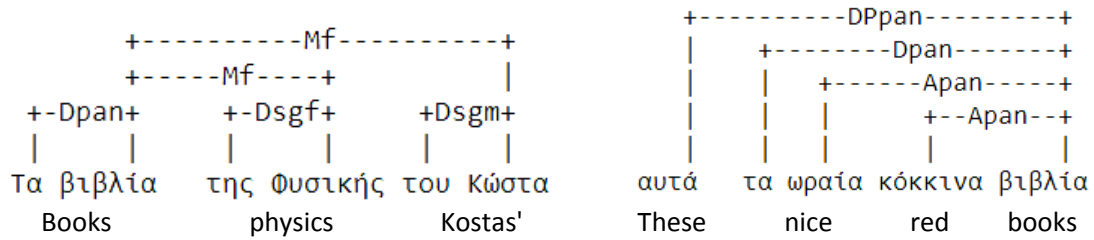


Figure 10: Examples of partial parsing of phrases [(9) on the left; (10) on the right] with Link Grammar

βιβλία (books): ((({@Apnn-} & Dpnn- & {KT+} & Spc+) or ({@Apan-} & Dpan- & {DPpan-} & Opan-)) or ({@Apan-} & (Dpan-) & {DPpan-} & {@M+}) or ({@Apnn-} & (Dpnn-) & {@M+}) or ({@Apan-} & Dpan- & Opan+) or {Dpan- or Dpnn-} or (({Dpnn-} & {Dpnn-} & Apnn- & Mf-)) or ({Dpnn-} & {DPpnn-} & Apnn- & Mf+));

της (of): Dsgf+;

Φυσικής (Physics): {@Asgf-} & Dsgf- & {KT+} & {Mf-};

αυτά (these or those): DPpnn+ or DPpan+;

ωραία (nice): Asnf+ or (Apnn+ or Apan+);

κόκκινα (red): Apnn+ or Apan+;

The operator '@' indicates that the link is repeated as many times as necessary. In sentence 9, the word 'βιβλία' (books) has two modifiers (Mf), and in sentence 10, before the word 'βιβλία' (books), there are two consecutive adjectives (A) ('ωραία' - nice and 'κόκκινα' - red).

Concluding this section, we can examine a sentence (11) that is ambiguous, with its two analyses shown in Fig. 11.

Ο Γιάννης είδε την Μαρία με το τηλεσκόπιο (11)

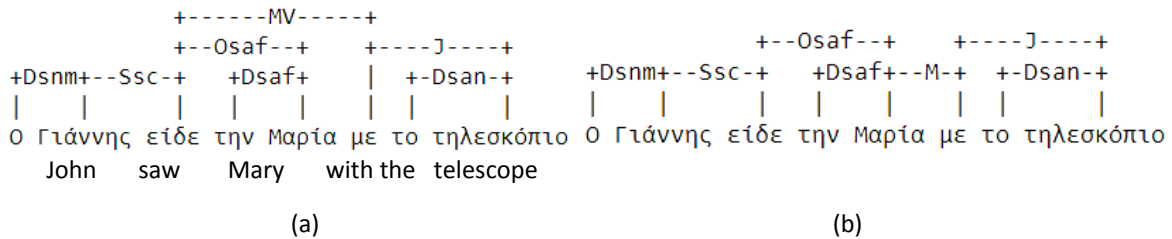


Figure 11: Examples of different parsing of the same sentence (11) with Link Grammar

In (a), the phrase 'με το τηλεσκόπιο' (with the telescope) is associated with the verb 'είδε' (saw), while in (b), it is associated with 'Μαρία' (Mary). This occurs because the preposition 'με' (with) has a link to the left to qualify verbs (MV-) and another to qualify 'Μαρία' (M-), matching the corresponding links in the verb (MV+) and the noun (M+).

3.2 Subordinate Clauses and Coordination

Subordinate clauses are sentences that depend on another sentence (the main clause) and cannot stand independently. For their analysis according to Link Grammar, we will use links such as **TH**, **C**, **TO**, **I** and **QI**. The clause (12):

Η δασκάλα δεν πιστεύει ότι ο μαθητής ξέρει την απάντηση (12)

has the analysis shown in Fig. 12.

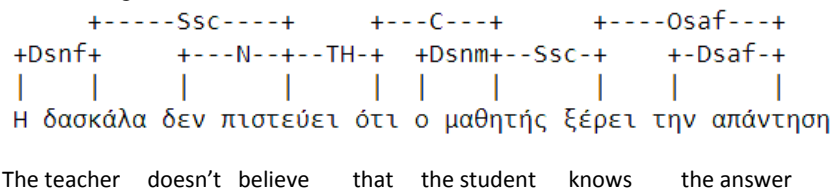


Figure 12: Example of parsing a composite sentence (12) with Link Grammar

The verb 'πιστεύει' (believe) has the link **TH+**, while the conjunction 'ότι' (that) has a link **TH-** to the verb and a **C+** link to the subject of the subordinate clause, 'μαθητής' (student).

Similarly, sentences 13 and 14 have their analyses in Fig. 13 (a) and (b) respectively. A new link appears, **PT**, which ensures that in the second subordinate clause, the subject (neuter gender and singular number) matches in gender and number with the predicate, which is an adjective.

Ο Γιάννης νομίζει ότι ο Πέτρος τον βλέπει (13)
 Ο Κώστας ρώτησε αν οι φοιτητές νομίζουν ότι το μάθημα είναι δύσκολο (14)

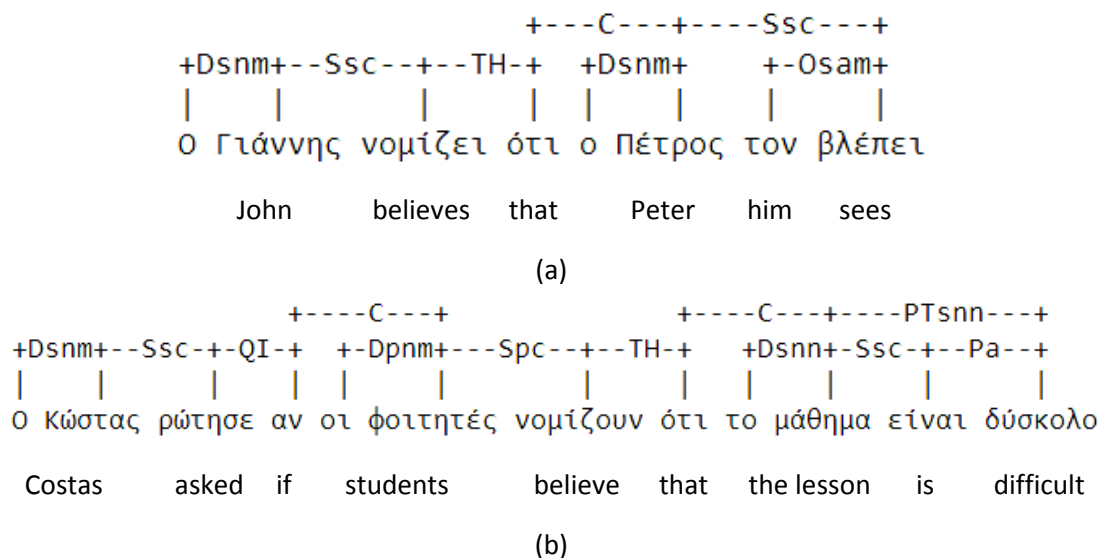


Figure 13: Examples of parsing composite sentences with Link Grammar

This mechanism prevents analyzing sentences like the following:

*Το μάθημα είναι δύσκολη (15)

The word 'μάθημα' (lesson) is neuter in gender, the adjective 'δύσκολη' (difficult) is feminine in gender, and sentence 15 does not belong to the grammar. In Fig. 13 (a), the word 'τον' (him) is a clitic form of the personal pronoun 'αυτός' (he) masculine in gender, third person, singular, and serves as the object in the subordinate clause.

In a similar way we can analyze sentences 16 and 17 as shown in Fig. 14 (a) and (b) respectively. The links **TO** and **I** are used to connect the particle 'να' (to) with the infinitive and verb:

Είπε ο Γιάννης ότι η Μαρία διάβασε το βιβλίο (16)
 Ο Γιάννης κανόνισε να φύγει η Μαρία (17)

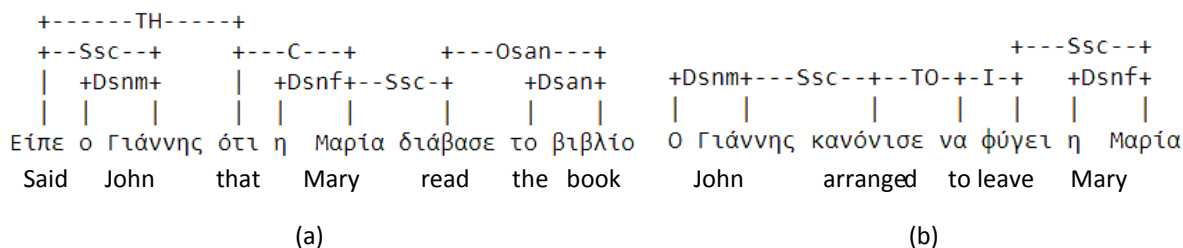


Figure 14: Examples of parsing common sentences with Link Grammar

κανόνισε (arranged): {Ssc-} & {TO+} & {TH+};

να (to): TO- & I+;

φύγει (leave): I- & {Ss*+} & {PTs-};

Coordinating conjunctions like 'και' (and) and 'ή' (or) are used to connect verbs and/or nouns and adjectives. These roles are fulfilled by the links **VJl** and **VJr** for verbs, and **SJl** and **SJr** for nouns and adjectives (**I** for left).

r for right). Fig. 15 (a) and (b) shows the results of sentences 18, 19 respectively, and Fig. 16 shows the results of sentence 20:

Οι ψεύτικες αξίες και επικίνδυνες ιδέες (18)
 Πήρε αυτά τα βιβλία και έφυγε (19)

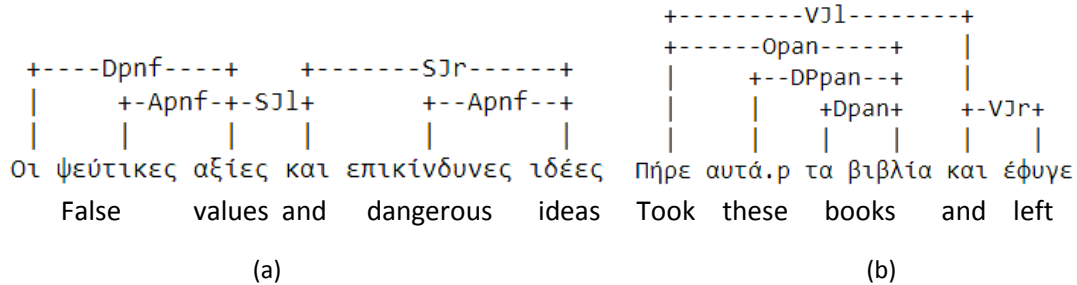


Figure 15: Examples of partial parsing of different phrases with Link Grammar

Τα βιβλία της Φυσικής ή τα τετράδια της Χημείας της Φρόσως (20)

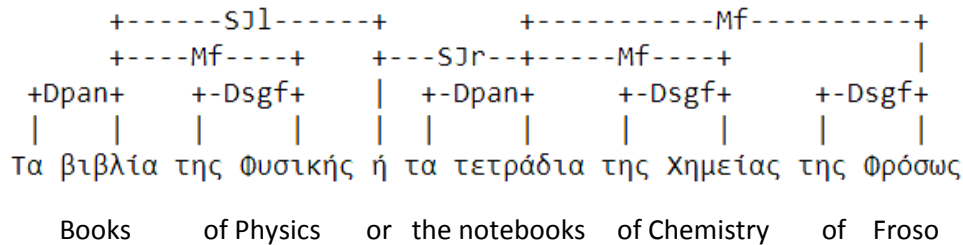


Figure 16: Example of partial parsing of a long phrase with Link Grammar

3.3 Relative and Interrogative Sentences

A challenging aspect of syntactic analysis involves relative clauses. The links we will use are **R**, **RS** & **B**. The following sentences (21-22) are examples of relative clauses:

Η γάτα που έφαγε το ποντίκι ήταν μαύρη (21)
 Η γάτα που με γρατζούνισε ήταν μαύρη (22)

where the results are shown in Fig. 17 (a) and (b) respectively:

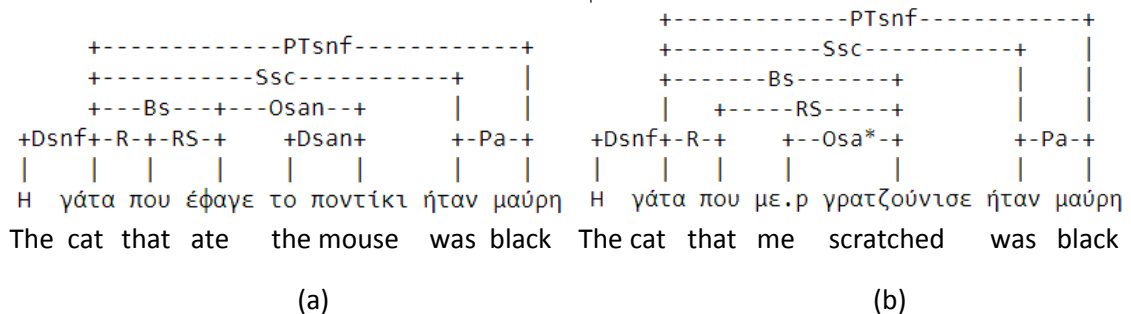


Figure 17: Examples of parsing sentences 21 (a) and 22 (b) with Link Grammar

as well as sentences 23-24:

Η γάτα η οποία έφαγε το ποντίκι ήταν μαύρη (23)
 Το καρπούζι το οποίο έφαγα ήταν ωραίο (24)

where the results are shown in Fig. 18 (a) and (b) respectively:

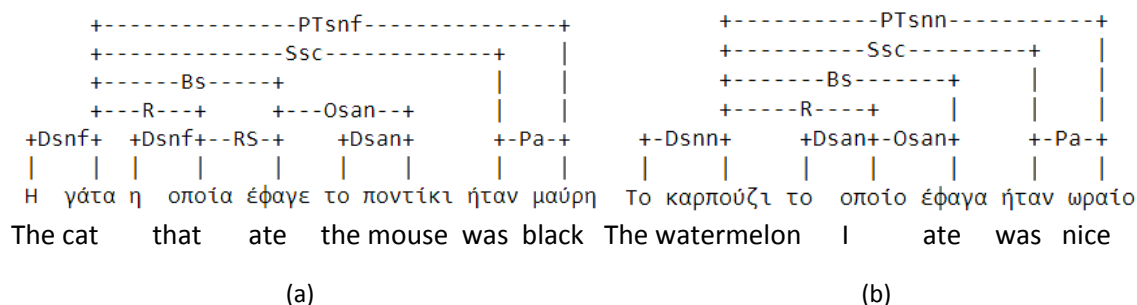


Figure 18: Examples of parsing sentences 23 (a) and 24 (b) with Link Grammar

Interrogative sentences refer to questions of total or partial ignorance. Using interrogative pronouns, we can form them. Below, in Fig. 19, we see some examples:

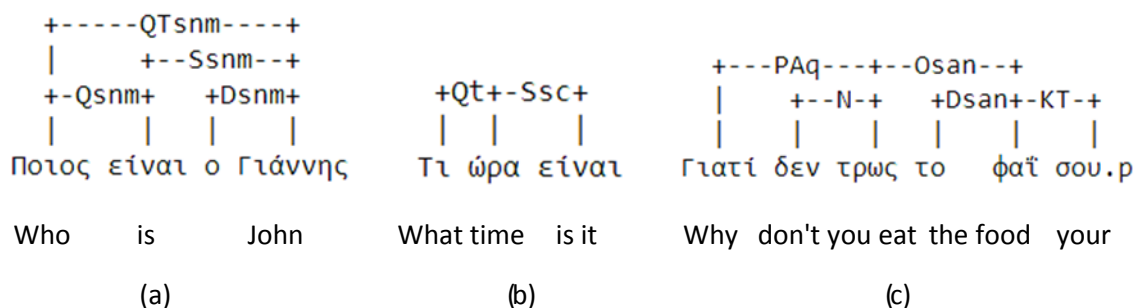


Figure 19: Examples of parsing short sentences with Link Grammar

We observe that in (a), we have the **Q** link, which connects to the verb, and the **QT** link, which ensures that the type of the interrogative pronoun matches the subject. Similarly, in (b). In (c), the interrogative particle ‘γιατί’ (why, wh-word) connects to the verb.

IV. FURTHER WORK AND CONCLUSION

There was an effort, presented herein, to implement an initial approach to the Greek language using Link Grammar, which is based on links between words. Using the structure of the language as a foundation, we created a lexicon governed by specific rules to enable syntactic analysis of sentences in Greek. Compared to the English language, which has a relatively fixed word order, Greek features more flexible syntax, inflections, and greater morphological complexity. Additionally, agreement between nouns, adjectives, and verbs must be maintained. The recognition and connection of syntactic modifiers, such as adjectives and adverbs, must also be carefully considered.

The complexity and diversity of a morphologically rich/complex language, like Greek, make its full modeling challenging, but the way the lexicon can be constructed allows us the flexibility to expand it and include more word types. The elements **LEFT-WALL** and **RIGHT-WALL**, which are special symbols used to mark the boundaries of a sentence (beginning and end) during analysis, were not yet utilized.

The ultimate goal is for the lexicon of Greek to be developed to a point where it can take its place alongside other languages on the LG Parser's website.

REFERENCES

- [1]. Theodorides, K., Revisiting the tradition of Palamedes as inventor of the alphabet. *Academia Letters*, 2022. Article 5305.
- [2]. Jurafsky, D. and Martin, J., *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd Edition). New Jersey: Prentice-Hall, 2009.
- [3]. Sleator, D. and Temperley, D., Parsing English with a Link Grammar. In *Proceedings of the Third International Workshop on Parsing Technologies*, 1993, p. 277-292, Tilburg, Netherlands and Durbuy, Belgium.
- [4]. Sleator, D. and Temperley, D., Parsing English with a Link Grammar. Carnegie Mellon University, Computer Science technical report CMU-CS-91-196, October 1991.
- [5]. Casbeer, W., Dehdari, J. and Lonsdale, D., A Link Grammar parser for Arabic. In *Perspectives on Arabic Linguistics: Papers from the annual symposium on Arabic linguistics*, Volume XX, March 2006. Mustafa A. Mughazy (Ed.), Kalamazoo, Michigan.

- [6]. Dehdari, J. and Lonsdale, D., A link grammar parser for Persian. In the First International Conference on Aspects of Iranian Linguistics, 2005. Leipzig, Germany.
- [7]. Istek, O. and Cicekli, I., A Link Grammar for an Agglutinative Language. In Proceedings of Recent Advances in Natural Language Processing (RANLP), 2007, p. 285-290. Borovets, Bulgaria.
- [8]. Chorozoglou, G.V., et al., Parsing Reviews with Link Grammar - Use of R for Sentiment Analysis. In eRA-2021 14th International Scientific Conference 'Industry 4.0', 2021. University of West Attica, Athens, Greece.
- [9]. Chorozoglou, G., et al., Review of Parsing in Modern Greek - A New Approach. International Journal of Computational Linguistics, 2021. **12**(1): p. 1-8.
- [10]. Triantopoulou, T., Application of the Affix Grammars over a Finite Lattice (AGFL) formalism for the description of the Modern Greek noun phrase' (in Greek). Ph.D. Dissertation, University of Patras - Department of Electrical & Computer Engineering, Greece, 1997.
- [11]. World Atlas of Language Structures, n.d., <https://wals.info/>
- [12]. Link Grammar Parser, n.d., <https://github.com/opencog/link-grammar/>