

Analysis of lexical ambiguity in Modern Greek using a computational lexicon

Panagiotis Gakis and Christos Panagiotakopoulos

Department of Primary Education, University of Patras, Greece

Kyriakos Sgarbas

Department of Electrical and Computer Engineering, University of Patras, Greece

Christos Tsalidis

“Neurolingo” Company

Abstract

Ambiguity is one of the most significant problems in Natural Language Processing. This difficulty may not be apparent to native speakers because of their natural ability at resolving it using contextual information and common sense knowledge. In contrast, current computer applications are still lacking the ability to disambiguate complex texts efficiently. The most common type of ambiguity is lexical ambiguity, and this is noticed even in highly inflectional languages such as Greek. In the present article, all the patterns of predictable lexical ambiguity in Modern Greek Language are registered, verified and quantified as occurred in the Neurolingo computational lexicon, after a study of morpho-syntactic characteristics that differentiate the ambiguous words.

Correspondence:

Panagiotis Gakis,
Department of Primary
Education, University of
Patras, University campus,
26504, Rio, Greece.

Email:

gakis@sch.gr

1 Introduction

Lexical ambiguity is a fundamental problem in Natural Language Processing (NLP); every practical text processing application has to deal with lexical ambiguity at some level. This article presents, in detail and systematically, all patterns of predictable lexical ambiguity in Modern Greek language as occurred in the Neurolingo¹ computational lexicon (Gakis *et al.*, 2012). Most of the frequent lexical ambiguity patterns of Modern Greek have been already specified in several related articles and with a large number of different approaches, e.g. stochastic tagging (Dermatas and Kokkinakis,

1995), two-level morphology (Sgarbas *et al.*, 1995), decision trees (Orphanos & Christodoulakis, 1999), part-of-speech (POS) tagging based on machine learning techniques (Petasis *et al.*, 1999, Papageorgiou *et al.*, 2000, Petasis *et al.*, 2001). Moreover, various other related methodologies (not specific to Modern Greek language) have been proposed, i.e: (1) the complement types (Bresnan, 1979), (2) the YAP parser (Church, 1980), (3) the grammatical relations (Kaplan and Bresnan, 1982), (4) the lexical disambiguation rules (Church, 1988), (5) the memory-based (POS) Tagger-Generator (Daelemans *et al.*, 1996), (6) the flexible POS tagger using an automatically

acquired language model (Màrquez and Padró, 1997).

However, this article presents for the first time all the patterns of predictable lexical ambiguity of Modern Greek organized in a systematic way and verified by a sufficiently large computational lexicon. Moreover, it addresses two additional questions: (1) how easily can the ambiguity be understood and (2) how efficiently a tagger can distinguish between multiple types of lemmas.

The presented results have been verified and quantified by the Neurolingo lexicon (Gakis et al., 2012), a general-purpose electronic lexicon for Modern Greek of the Machine Readable Dictionary type (i.e. not suitable for direct use by casual users, but a modular component for various NLP applications such as morphosyntactic taggers, parsers, semantic taggers, machine translation systems, etc.). It contains ~90,000 Modern Greek lemmas producing a total of 1,200,000 inflected word forms categorized by 67 POS and

morphological attributes, 79 domain and style attributes, 77 attribute sets, 19 accent rules, 191 suffix rules, and 306 grammar rules, as shown in Figs 1 and 2 and Table 1.

All lemmas in the Neurolingo lexicon were gathered by indexing four of the biggest Dictionaries of Modern Greek: (1) the Dictionary of Common Modern Greek Language by the Institute of Greek Studies of Aristotle University of Thessalonica (1998), (2) the Dictionary of Modern Greek by G. Babiniotis (1998), (3) the Greek Dictionary of the Modern Demotic Language by E. Kriaras (1995), and (4) the Major Greek Dictionary by Tegopoulos Fytrakis (1997). To ensure its completeness, the Neurolingo lexicon is being updated continuously and regularly by various sources (newspapers, WWW, etc.). It is currently at its fifth edition after 9 years of development since its first version in October 2003.

In this article, all patterns of predictable Lexical Ambiguity are presented as they occur in Modern

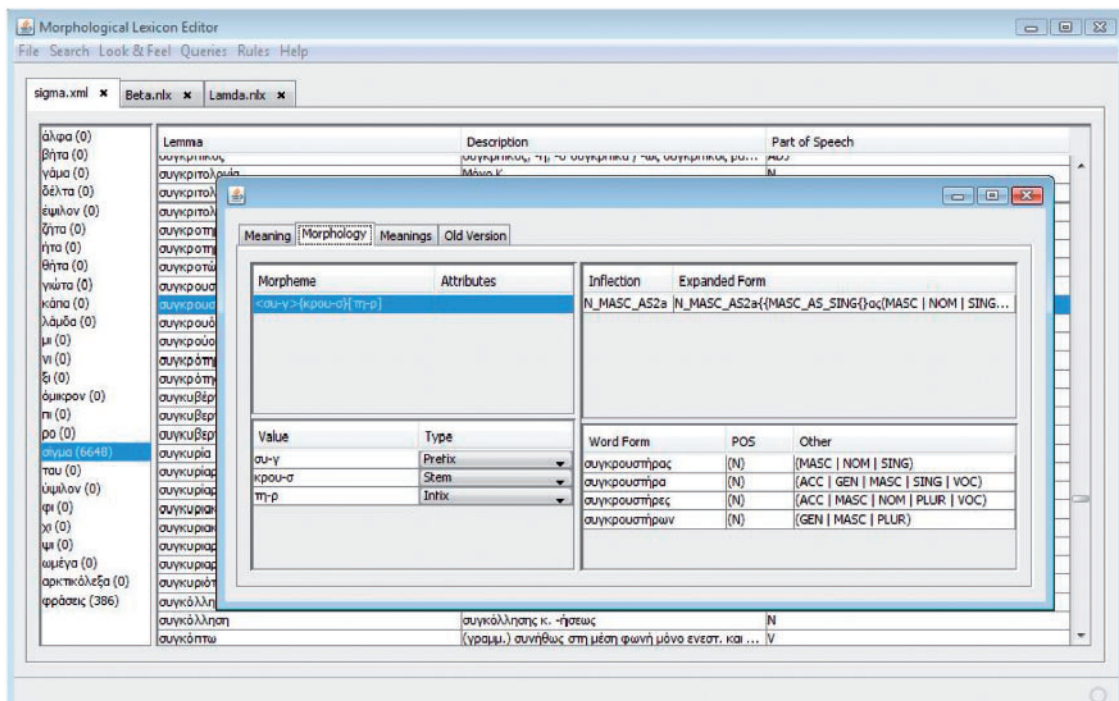


Fig. 1 Screen capture of the Neurolingo morphological lexicon editor

```

<lemma hw="συχρονικός" pos="ADJ">
  <ta>SPORTS</ta>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ού) " ma="ADJ+GEN+MASC+NO_COMP_SUP+SING"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ών) " ma="ADJ+GEN+MASC+NO_COMP_SUP+PLUR"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ούς) " ma="ACC+ADJ+MASC+NO_COMP_SUP+PLUR"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ός) " ma="ADJ+MASC+NOM+SING+NO_COMP_SUP"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ό) " ma="ACC+SING+ADJ+MASC+NO_COMP_SUP"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (έ) " ma="ADJ+MASC+NO_COMP_SUP+SING+VOC"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (οί) " ma="ADJ+MASC+NOM+NO_COMP_SUP+PLUR+VOC"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ή) " ma="ACC+ADJ+FEM+NOM+NO_COMP_SUP+SING+VOC"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ής) " ma="ADJ+FEM+GEN+NO_COMP_SUP+SING"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ές) " ma="ACC+ADJ+FEM+NOM+NO_COMP_SUP+PLUR+VOC"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ών) " ma="ADJ+FEM+GEN+NO_COMP_SUP+PLUR"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ό) " ma="ACC+ADJ+NEUT+NOM+NO_COMP_SUP+SING+VOC"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ού) " ma="ADJ+GEN+NEUT+NO_COMP_SUP+SING"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ά) " ma="ACC+ADJ+NEUT+NOM+NO_COMP_SUP+PLUR+VOC"/>
  <word hw="&lt;συχ-ε&gt; {χπο-ν} [ι-κ] (ών) " ma="ADJ+GEN+NEUT+NO_COMP_SUP+PLUR"/>
</lemma>

```

Fig. 2 Screen capture of morphological analysis results

Table 1 A sample of the attributes in Neurolingo lexicon

ATTRIBUTES		ATTRIBUTE_TAGS
Morphological		
Case	NOM, GEN, ACC, VOC	NOM_SING_MASC...
Number	SING, PLU	
Gender	MASC, FEM, NEUT	
Tense	PRES, IMPERF	PRES_IND...
Mood	IND, IMPERAT...	
Part of speech	N, ADJ, ADV, PREP...	
Degree	COMP, SUP, NO_COMP_SUP	
Person	A_P, B_P, C_P	B_P_SPOKEN...
Style attributes	LEARNED, RARE, ARCHAIC, DIALECTICAL...	
Stressing Attributes	NO_STRESS	
Terminological	ULT_PEN PHILOSOPHY, ATHLETICS...	

Greek. Many grammatically similar word forms may correspond to different lemmas, even to lemmas with different POS. Lexical ambiguity in Modern Greek results from the inflectional complexity of the language consisting of 10 parts of speech, five of which are declinable and five indeclinable. The declinable ones include a huge selection of morphological word forms. Evidence of this is the fact that >300 inflectional word forms can be derived from a verb—both active and passive voice word

forms included. An adjective can produce ~100 morphological word forms, including the comparative and superlative types (Gakis *et al.*, 1999). It is worth noticing that some archaic word forms, which are still in use in colloquial Modern Greek, are included in the lemma. This fact increases the ambiguity and the complexity in the processing of a Greek text if such a huge lexicon is consulted.

Tables 2 and 3 display the number of lemmas, word forms, and related statistics for the lexicon.

Table 2 Content statistics of the lexicon

Part of speech	Number of word forms
Nouns	60,511
Adjectives	22,844
Verbs	9,245
Participles	865
Adverbs	7,830
Other parts of speech	420

Table 3 Statistics concerning lexical ambiguity

Part of speech	Number of word forms
Number of unique inflected word forms	8,73,701
Ambiguous word forms (from different Lemmas)	39,119
Ambiguous word forms (from the same Lemma)	4,758
Total number (for all ambiguous words)	9,17,578

2 Ambiguity

The term ‘ambiguity’ may have many interpretations, i.e. lexical ambiguity, semantic ambiguity, syntactic ambiguity, relative ambiguity, and so forth. (Kanakis, 2007; Lyons, 1995, 1968). Lexical ambiguity occurs when a word has more than one lexical entries or when it is used with different meanings in transposition. Van Eijck and Jaspars (1996) determines the lexical ambiguity as the lack of information about word meaning. John Lyons (Lyons, 1977) recognizes two types of lexical ambiguity: (1) the homonymy (e.g. the noun [όρος, ο = term] and the noun [όρος, το = a mountain] are homonyms) and (2) the polysemy (e.g. the verb [δίνω = I give] may have a lot of meanings: [exist || turn away || strike || make angry]).

In addition to the interpretations that are attributed to the term ‘lexical ambiguity’, concerning the computational processing, lexical ambiguity is closely related to linguistic data representation (Boguraev and Pustejovsky, 1990). To be more precise, ‘lexical ambiguity’ means that ambiguous words are two or more lexical types with a common graphemic form (i.e. spelling), but they belong to different lemmas and/or differ in one or more morpho-syntactic attributes, mainly with regard to the part of speech.

According to this definition, the type [απαντήσεις] is lexically ambiguous, as it can be either a verb (< [απαντώ = to answer]) or a noun (< [απάντηση = answer]). Also, the type [ματιών] is lexically ambiguous because it may have been derived from different lemmas (< [μάτι = eye], [ματιά = glance]). Moreover, the type [κόρη = daughter] presents lexical ambiguity, owing to its ambiguous declension (nominative or accusative or vocative) in the same lemma [κόρη = daughter] and not because of its polysemy (daughter || pupil of an eye). The same type does not have lexical ambiguity if, for example, it presents itself in the noun phrase [η κόρη = the daughter]: the presence of the article disambiguates the case, whereas (however) the word still retains its semantic ambiguity.

Sometimes the same word corresponds to different entities, and this presents an additional source of ambiguity. For example, the word [κανάλι = channel] is a morphological entity with various cases (nominative, accusative, or vocative), and it also has more than one meaning at the semantic level (television station || deep ditch || communication medium || difficulty, etc.).

Lexical ambiguity also occurs during the conversion of spoken utterances to written text and vice versa (Crystal, 2000). For example, the graphemic form λόγια (< [lo-gi-a = scholar], [lo-gia = words]) represents two different words, discriminated only by pronunciation, which is not represented in the written text.

Syntactic ambiguity arises when a phrase can be parsed in more than one way. For example, the sentence: *He has hurt the woman with the flowers* could mean that: (1) He has hurt [the woman with the flowers], or (2) He has hurt [the woman] [with the flowers]. Such phrases differentiate the meaning of the sentence because different grammatical structures can be assigned to the same word sequence.

Ambiguity can also affect a whole sentence in the wide frame of communication (pragmatics level). The reason is that the sentence—without the linguistic frame—is still abstract. Thus, even simple statements like [Θα είμαι εκεί = “I will be there”] can function differently (i.e. as a notification, as a promise, or as a threat, etc.), according to the

linguistic and unlinguistic frame or context (such as the linguistic action in which it is registered).

3 Ambiguity and Computational Systems

Native speakers usually resolve ambiguity easily, but in computational linguistics, ambiguity resolution is a crucial problem (Filippaki-Warburton, 1992). Many NLP applications require extraction and processing of the meanings of texts, in addition to processing their surface forms (Orphanos and Tsalidis, 1999). Such a task needs a significant amount of information about the world and the domain of discourse (Grishman, 1986). The knowledge-based approach to NLP concerns methods of acquisition and representation of such domain knowledge and its application to disambiguate input utterances (Mahesh and Nirenburg, 1997). Lexical and syntactic ambiguity resolution prevents the need for special parsing, reconciles a number of apparently conflicting results concerning the roles of lexical and contextual information in sentence processing, explains differences among ambiguities in terms of ease of resolution and provides a more unified account of language comprehension than what was previously available (MacDonald *et al.*, 1994).

It is common for syntactic structures to be represented by phrase structure rules (Chomsky, 1965). However, many cases occur in which ambiguity is introduced in the part of speech—as a lemma is the major feature (Pollard and Sag, 1987)—such as an individual morphological attribute that a lexical type may have. When the part of speech is ambiguous, the parser is forced to examine many more syntactic rules and, eventually, to produce all the phrasal structures that these rules dictate, expecting that one analysis will eventually prevail (Allen, 1987). A successful recognition of neighbouring structures could be doubtful, if these contain more ambiguous components. Modern Greek is a language with a lot of particularities, an element that makes even more difficult and complicated its processing by NLP systems. For example the word [απαντήσεις] may be: (1) the head of a noun phrase, (2) the head of a verb phrase [απαντήσεις<You answer]. Moreover, if it is a

noun, then an additional morphological ambiguity is involved, as the word can be in nominative, accusative, or vocative case of plural (Orphanos, 2000). Extending the same example, in the phrase: [Οι απαντήσεις του έχουν καλά στοιχεία = His answers have good elements], a lexical parser accompanied by a morphosyntactic tagger produces the structural analysis shown in Fig. 3.

Figure 3 uses xml formalism to represent information. Each token is specified inside tags <t>...</t>.

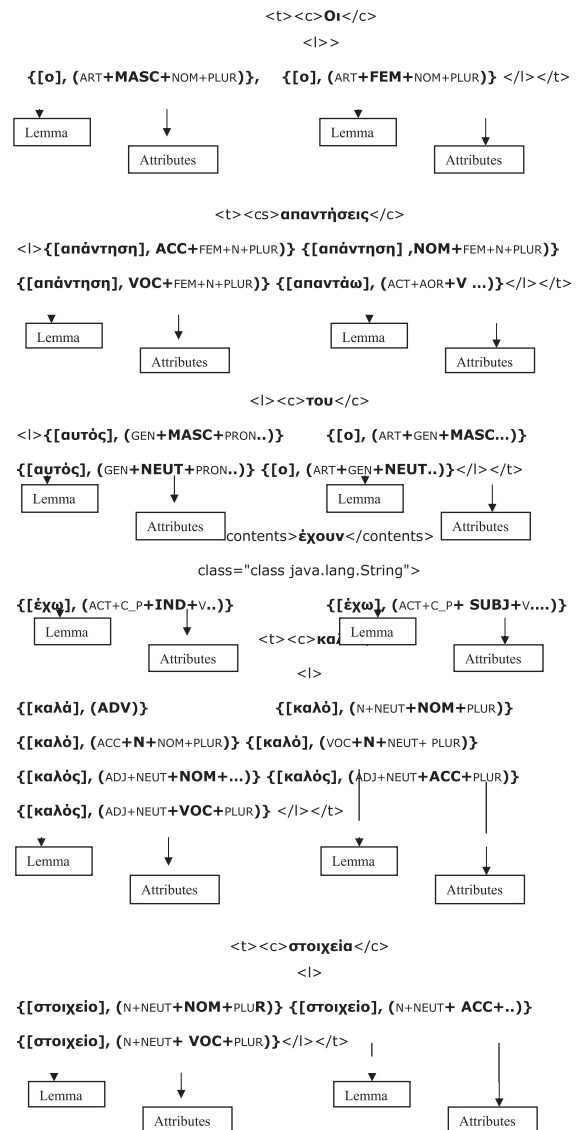


Fig. 3 Structural analysis performed by the lexical parser

The token structure is comprised by the token's text (content) using the tag. The lemma information defined inside tags <l>...</l> is a list of pairs of the form {[lemma headword] (plus (+) separated morphological attributes)}.

For analysis purposes, the whole content of the lexicon (Fig. 2) (including the morphosyntactic attributes and the lemmas they belong to) was incorporated in a database. The ambiguous words were found by recalling all word forms with the same spelling but different morphosyntactic attribute. The patterns of lexical ambiguity and their classification into categories were based on the morphosyntactic features that differentiate these ambiguous words.

4 Lexical Ambiguity through a Computational Lexicon

The role of a computational lexicon is to attribute all (if possible) the word forms of a language (Radford, 1988). The model adopted for the Neurolingo lexicon has been specifically designed for Modern Greek. The contents of the lexicon are organized into morphological lemmas. The Neurolingo lexicon includes—as far as possible—the formal word forms of all lemmas and at the same time the attribution of the whole morphosyntactic characterization. The basic morphosyntactic attribute of a word-form is its POS tag. The POS tag determines what other grammatical attributes characterize a word-form: gender, number and case for nouns, adjectives, articles, pronouns and present perfect participles; voice, tense, mood, number, and person for verbs. The first word-form of a morphological lemma, the headword, plays the role of lemma representative; referring to the headword is the same as referring to the lemma. As the Neurolingo morphological lexicon is monolingual, grammatical, and morphosyntactic, annotations are assigned only to Greek words.

When a word-form is looked-up in the morphological lexicon, there are three possible outcomes: (1) the word-form is matched to one morphological lemma, (2) the word-form is matched to two or more morphological lemmas, and (3) the word-form is not found. Instead of trying to perform morphosyntactic

analysis to assign unambiguous morphosyntactic annotations to word-forms, we use relaxed pattern forms as explained later in the text.

The project plan for the generation and recognition processes combined several planning algorithms. The lexicography editor analyzes only isolated words. On the basic level, the whole outset of morphological description of Modern Greek suffixes has been standardized by their classification in distinct word forms according to separate parts of speech. We only focus on the morphological analysis and not on the morphological production. The system is based on pre-selected lists of characteristics and rules with definite number of values. The architecture of the lexicon is built on a five-level model.

The 'ambiguity report' is one of the most important reports produced by the electronic morphological analysis system. It expands each lemma and produces the full list of word forms. The word forms are classified in five classes depending on the existence of another word with the same textual representation (i.e. they are lexicographically similar). Each word form can belong to zero or more of these classes. The five classes are as follows:

- (1) Words with different POS, e.g. [έντεκα, το = eleven {noun, neuter}] || [έντεκα = eleven {adjective}].
- (2) Words with different hyphenation pattern, e.g. [ήπιες] < [ή-πιες = you drank] || [ή-πι-ες < {ήπιος = mild}].
- (3) Words with different morphemic structure, e.g. <Πα-ρα->{σκευ-}(ή) < [παρασκευή = preparation], {Πα-ρα-σκευ-}(ή) < [Παρασκευή = Friday].
- (4) Words in different lemmas. These words can be further sub-classified in classes depending on the event that they have in the same part of speech. This means that we want to know whether there are equal word forms with the same part of speech belonging to different lemmas, e.g. [απαντήσεις] < [απάντηση = answer {noun}] || [απαντώ = to answer {verb}].
- (5) Double words, i.e. words that are not related at all. e.g. [όρος] < [όρος = term {male}] || [όρος = mountain {neuter}].

Table 4 Ambiguity in corpora

POS ambiguity schemes	Examples words	% Occurrence in the corpus
Pronoun-article	το, τον, τη, την, τις	25.38% (5,20,611 tokens)
Pronoun-preposition	με, σε	3.78% (77,627 tokens)
Adjective-adverb	λίγο	3.92% (80,585 tokens)
Preposition-particle-conj	για	2.16% (44,295 tokens)
Verb-noun	ερωτήσεις, γέννα	1.71% (35,066 tokens)
Adjective-adverb-noun	μέσα, άδεια	1.22% (25,040 tokens)
Adjective-noun	επίπεδο	6.30% (1,29,263 tokens)
Adverb-conjunction	καθώς	0.63% (12,963 tokens)
pronoun-adverb	μόνο	0.20% (4,305 tokens)
Verb-adverb	δω	0.08% (1,800 tokens)

After the morphological declination, we have counted the words that belong to more than one lemmas and classify them in patterns according to the POS or other morphosyntactic attributes of their lemma. The etymology of ambiguous words has also been considered so that the classification is created with the criterion of predictable or symptomatic ambiguity.

To calculate the frequencies, a random corpus of 2,050,614 tokens has been set up, consisting of sentences from student writings, literature texts, newspaper articles, and articles from technical, financial, and sports magazines. Tokens were attributed to the corpus, and morphosyntactic tags were assigned to each word token automatically by the lexicon, with all the attributes available to each known word. An example of a sentence tagged by the lexicon is given in Table 4.

5 Lexical Ambiguity—Patterns

Analysis of the computational lexicon's data shows that the morphological complexity of Modern Greek creates a wide variety of lexical ambiguity—either predictable or symptomatic (i.e. non predictable).

5.1 Predictable ambiguity

This pattern of ambiguity includes words of identical spelling and etymological form and is separated in two subcategories:

- The first one includes the ambiguity noticed in the same lemma. For example, the type [κρίνω

= judge] is used for the present tense (indicative or subjunctive mode) for continuous or concise future (indicative mode) as well as for the perfective aspect tense (subjunctive mode). This pattern includes the ambiguity that is noticed in (a) singular genitive case of male in the genitive case of neuter. For example, the word type [καλού = of good (singular)] is either genitive of neuter or male, (b) plural genitive case of male, female, and neuter of (1) adjectives, e.g. [καλών = of good (plural)], (2) pronouns, e.g. [εκείνων = of those], and (3) participles, e.g. [δυστυχισμένων = of unhappy]. This pattern of ambiguity is aggravating for computational systems, it is one of the natural characteristics of Greek morphology and it is not annotated. The ambiguity between adjective and adverb is also predictable.

- The second one includes the ambiguity that can be noticed in words belonging to different lemmas with the same or different part of speech. This category is described in detail in Section 6.

5.2 Symptomatic ambiguity.

This category includes words of identical spelling but different etymology. For example, the word [βάλτε] may be singular vocative of the noun [βάλτος = swamp] or imperative for perfective aspect tense of the verb [βάζω = to put]. Although the second interpretation is much more frequent than the first, this research did not take into consideration word frequencies within texts, as it was based on a lexicon, not a corpus. Thus, all the

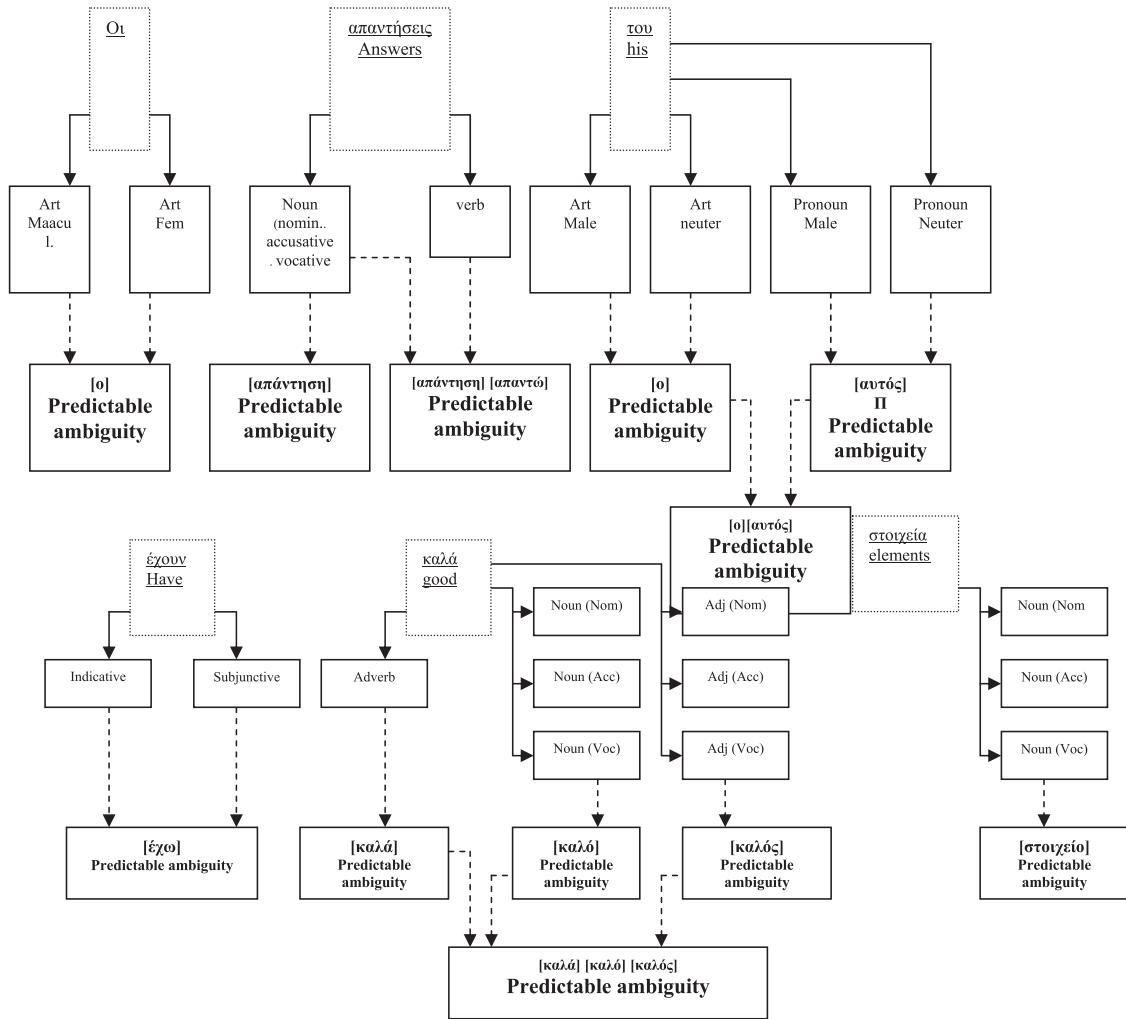


Fig. 4 Patterns of lexical ambiguity

grammatically plausible types contained in the lexicon have been considered and attributed.

The following example concerns different example types. The suffix (-ια) (with morphological attributes: nominative, accusative, vocative, neuter gender) is in other lemmas either hyphenated (-ια) or non-hyphenated (ια). For example, the word [ήπια] is either a verb <[πίνω = to drink] (first person, singular, perfective aspect, indicative), or an adverb <[ήπια = mildly], and an adjective (nominative, accusative, vocative, plural, neuter || nominative, accusative, vocative, singular, female also) [ήπιος = mild].

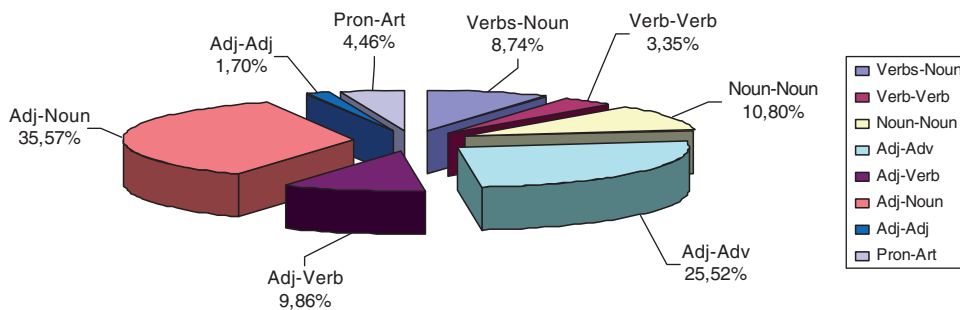
In Fig. 4, we present the patterns of lexical ambiguity that occurred in the phrase: [Οι απαντήσεις του έχουν καλά στοιχεία = His answers have good elements]

6 Patterns of Predictable Lexical Ambiguity

This section describes in detail the categories of predictable lexical ambiguity as shown in Table 5 and Fig. 5.

Table 5 The lexical ambiguity results

Lexical ambiguity		
Noun – verb	Verb – verb	Noun – noun
8.73% (8,084 word forms)	3.35% (3,103 word forms)	10.79% (9,992 word forms)
Adjective – adverb	Adjective – verb	Adjective – noun
25.5% (23,659 word forms)	9.85% (9,127 word forms)	35.55% (32,926 word forms)
Adjective – adjective	Pronoun – article	
1.7% (1,576 word forms)	4.46%	

**Fig. 5** Quantitative results of lexical ambiguity

6.1 Noun–noun

This pattern includes nouns with the same theme but different suffix. They were derived from 9,992 word forms that correspond to the 10.79% of the total lexical ambiguity as shown in Table 5. The 14 subcategories of the ‘noun-noun’ category sum up to 85.54%. The remaining 14.46% corresponds to symptomatic ambiguity and was not included in the patterns. The 14 subcategories are as follows:

- 6.1.1 The parisyllabic nouns, of the female or male gender, that are accented on the antepenult, with suffix **-ος**, form the suffixes: **-ου** (genitive of singular), **-ο** (accusative of singular), and **-ων** (genitive of plural) in the same way as the parisyllabic nouns, neuter gender, that are accented on the antepenult, with suffix **-ο**. The only difference for these word forms is the gender, e.g. {αγριόσκυλο, αγριόσκυλο, αγριόσκυλων < [αγριόσκυλο = fierce dog], [αγριόσκυλος = wild dog]}, {εξόδο, εξόδο, εξόδων < [έξοδο = cost],

[έξοδος = exit]}. Quantitative analysis in the lexicon revealed that this ambiguity pattern takes up the greater section of the pattern ‘noun–noun’ (27.70%).

- 6.1.2 The parisyllabic nouns, of female or male gender, that are accented on the last syllable or on the antepenult, with suffix **-ός/-ος**, the parisyllabic nouns, of female gender, that are accented on the last syllable or on the antepenult, with suffix **-ή** or **-ά**, and the parisyllabic nouns, of male gender, that are accented on the antepenult, with suffix **-ας** all form a common genitive of plural with suffix **-ών** or **-ων**: {αγωγών < [αγωγός = conduit], [αγωγή = education]}, {γραμματικών < [γραμματικός = secretary], [γραμματική = grammar]}, {αρωγών < [αρωγός = helper], [αρωγή = assistance]}, {αυλών < [αυλός = pipe], [αυλή = playground]}, {επωδών < [επωδός = refrain], [επωδή = incantation]}, {νομών < [νομός = prefecture], [νομή = pasture]}, {πομπών < [πομπός

- = transmitter], [πομπή = procession]], {(εθνο);φρουρών < [(εθνο)φρουρός = guard], [(εθνο)φρουρά = garrison]], {καπνών; < [καπνός = smoke], [κάπνα = smog]], {εμπόρων < [έμπορος = trader], [έμπορας = dealer]]. This ambiguity pattern takes up the 21.88% of the category ‘noun-noun’.
- 6.1.3 The parisyllabic nouns, female gender, that are accented on the penult, with suffix **-α** a common genitive of plural with the parisyllabic nouns, female gender, that are accented on the last syllable, with suffix **-ιά** form: {δουλειών < [δουλεία = slavery], [δουλεία = job]], {ευλογιών < [ευλογία = blessing], [ευλογιά = variola]], {διακονιών < [διακονία = service], [διακονιά = begging]]. This pattern of lexical ambiguity encompasses word types that differ in stress. It occupies the 14.89% of the ‘noun-noun’ category.
- 6.1.4 The parisyllabic nouns, of male gender, that are accented on the penult, with suffix **-ας** present lexical ambiguity with the parisyllabic nouns, of female gender, that are accented on the penult, with suffix **-α**. They have common declension differentiating in the morphological attributes of gender and case of singular number. Their common attribute is the number (singular and plural). Thus, we have the word types: {κεφάλας, κεφάλα, κεφάλες, κεφαλών} that may belong to either of the lemmas [κεφάλας = someone with big head] and [κεφάλα = big head]. In the first case, their attributes are (nominative, singular, male [κεφάλας] < [κεφάλας]), and in the second case, they are (genitive, singular, female [κεφάλας] < [κεφάλα]). The same pattern follow the nouns [καράφας = bald-head] and [καράφα = baldness], [ηδονοβλεψίας = peeper] [ηδονοβλεψία = voyeurism]. These nouns occupy the 10.48% of the ‘noun-noun’ lexical ambiguity pattern.
- 6.1.5 The parisyllabic nouns, neuter gender, that are accented on the penult, with suffix **-ι** and **-ιο** form a common nominative, accusative, and vocative in plural {ακρωτήρια < [ακρωτήρι = cape], [ακρωτήριο = cape]]. This pattern of lexical ambiguity occupies the 4.88% of the ‘noun-noun’ category, but it does not seem to create problems in semantic processing, as it concerns nouns of neutral gender, with the same meaning. During syntactic analysis, however, ambiguities of this pattern are still problematic, as they produce different parsing trees.
- 6.1.6 The parisyllabic nouns, neuter gender with suffix **-ο** [πιατέλο = dish] form common suffixes with the parisyllabic nouns, female gender, that are accented on the last syllable or on the penult, with suffix **-α** [πιατέλα = large flat dish]. They only differ in their morphological attributes: (1) **-α** (nominative, accusative, vocative of singular [<πιατέλα]) or (nominative, accusative, vocative of plural [<πιατέλο]) and (2) **-ων** (genitive of plural, female gender [<πιατέλα], or (genitive of plural, neuter gender [<πιατέλο]), {πλευρών < [πλευρά = side], [πλευρό = rib]]. This pattern occupies the 2.19% of the ‘noun-noun’ category.
- 6.1.7 The parisyllabic nouns, of female gender, that are accented on the last syllable, with suffix **-ιά** form a common plural genitive with the parisyllabic nouns, neuter gender, that are accented on the penult or at the last syllable, with suffix **-ί/ι**. For example the word type [φασολιών] results from the female [φασολιά = beans] and the neuter [φασόλι = bean]. The following nouns belong to the same pattern: {σουβλιών < [σουβλιά = nudge], [σουβλί = stabber]], {ποδιών < [ποδιά = apron], [πόδι = foot]], {ματιών < [ματιά = look], [μάτι = eye]], etc. This ambiguity pattern occupies the 1.80% of the ‘noun-noun’ category.

- 6.1.8 The parisyllabic nouns, female gender, that are accented on the last syllable, with suffix **-ή** form a common genitive of plural with the parisyllabic nouns, female gender, that are accented on the penult, with suffix **-η**: {**αποχών** < [απόχη = [net] [αποχή = abstention]}. This pattern takes up the 0.57% of the ‘noun-noun’ category.
- 6.1.9 The parisyllabic nouns, male gender, that are accented on the penult, with suffix **-ης** produce identical plural with the parisyllabic nouns, female gender, that are accented on the penult with suffix **-η**: {**γρανίτες**, **γρανιτών** < [γρανίτης = granite], [γρανίτα = water-ice]}. This pattern occupies the 0.41% of the ‘noun-noun’ category.
- 6.1.10 The parisyllabic nouns, neuter gender, that are accented on the penult, with suffix **-ος** present lexical ambiguity with the parisyllabic nouns, female gender, that are accented on the penult, with suffix **-η**, as they form the suffixes: (1) **-η** (nominative, accusative, vocative of singular if the lemma is female or nominative, accusative, vocative of plural if the lemma is neuter) and (2) **-ων** (common genitive of plural) in the same way: {**πλάτη**, **πλατών** < [πλάτος = width], [πλάτη = back]}. This pattern of lexical ambiguity occupies the 0.40% of the ‘noun-noun’ category.
- 6.1.11 The parisyllabic nouns, male gender, that are accented on the last syllable, with suffix **-ός** form both the genitive of singular and plural in the same way as the parisyllabic nouns, neuter gender, that are accented on the antepenult, with suffix **-ι**: {**ανιψιού**, **ανιψιών** < [ανιψιός = nephew], [ανίψι = niece]}. This pattern occupies the 0.29% of the ‘noun-noun’ category.
- 6.1.12 The parisyllabic nouns, neuter gender, that are accented on the penult, with suffix **-ι** form a common genitive of singular and plural with the parisyllabic nouns, neuter gender, that are accented on the last syllable, with suffix **-ί**: {**τζαμιού**, **τζαμιών** < [τζάμι = pane], [τζαμί = mosque]}. This pattern takes up the 0.05% of the ‘noun-noun’ category.
- 6.1.13 The non-parisyllabic nouns, male gender, that are accented on the last syllable, with suffix **-άς** form common plural with the parisyllabic nouns, female gender, that are accented on the penult, with suffix **-άδα**: {**αμαξάδες**, **αμαξάδων** < [αμαξάς = coachman], [αμαξάδα = coach ride]}. The word form [αμαξάδες] presents additional lexical ambiguity while it is nominative and accusative of plural. This pattern of lexical ambiguity is rare and poorly represented in the lexicon; thus, no percentage has been calculated.
- 6.1.14 The parisyllabic nouns, male gender, that are accented on the penult, with suffix **-ης** form a common genitive of plural with the parisyllabic nouns, neuter gender, that are accented on the last syllable with suffix **-ό**: {**ποτών** < [πότης = drinker], [ποτό = drink]}. This pattern of lexical ambiguity is poorly represented, and no percentage has been calculated.

6.2 Noun-verb

- 6.2.1 The nominative, accusative, and vocative cases of plural of the parisyllabic nouns, female gender, that are accented on the antepenult with suffix **-η**, resemble to the second person of future tense and to the subjunctive of perfective aspect tense of verbs with suffix **-ω**: {**αμφισβητήσεις** < [αμφισβήτηση = dispute], [αμφισβητώ = to doubt]}, {**πράξεις** < [πράξη = fact], [πράττω = to do]}, {**ανακαλύψεις** < [ανακάλυψη = discovery], [ανακαλύπτω = to discover]}, {**αναζητήσεις** < [αναζήτηση = search], [αναζητώ = to look for]}. This pattern dominates almost totally the ‘noun-verb’ category occupying a percentage of 97.06%.

6.2.2 The nominative, accusative and vocative cases of singular of the parasyllabic nouns, female gender, that are accented on the penult or the last syllable with suffix **-α** resemble to the second person of the imperative present of verbs with suffix **-άω**: {κοπάνα < [κοπάνα = wash-tub], [κοπανώ = strike]}, {κλώσα < [κλωσώ = brood], [κλώσα = sitter]}, {γέννα < [γέννα = childbirth], [γεννώ = to give birth to]}, {γλίστρα < [γλίστρα = slide], [γλιστρώ = to slide]}. Also, in this category, the parasyllabic nouns, female gender, that are accented on the last syllable, with suffix **-α**, form the nominative and genitive of singular with suffixes **-ά** and **-άς**, respectively. These types resemble to the second and third person of singular of verbs with suffix **-άω**, in active voice: {φορά, φοράς < [φορά = direction], [φορώ = to wear]}. Also, for the parasyllabic nouns, male gender that are accented on the last syllable, with suffix **-ας** and verbs with suffix **-αω**, their common suffixes **-ά**, **-άς** correspond to the nominative and genitive of singular of the noun and the first and second person of present indicative tense of active voice of the verb: {πολυλογάς, πολυλογά < [πολυλογάς = babbler], [πολυλογώ = to babble]}. Finally, verbs with suffix **-αω** have common morphological types with parasyllabic nouns, female gender, that are accented on the penult, with suffix **-α**. The common suffixes are **-α**, **-ες**, with morphological attributes: (1) nominative singular or plural (for nouns), or (2) first and second person of imperfect singular in active voice (for verbs): {βρομούσα, βρομούσες < [βρομούσα = slut], [βρομώ = to be nifty]}, {πατούσα, πατούσες < [πατούσα = sole], [πατώ = to tread]}. This pattern of lexical ambiguity occupies the 2.51% of the 'verb-noun' category.

6.2.3 The parasyllabic nouns, male or female gender, that are accented on the penult, with suffix **-ος** form the vocative of singular the same way as verbs of the first or second conjugation and suffix **-ε**, form the second person of present or imperative of perfective aspect tense: {μέθυσε < [μέθυσος = drunkard], [μεθάω = to get drunk]}, {έλεγχε < [έλεγχος = control], [ελέγχω = to control]}, {παράγραφε < [παράγραφος = paragraph], [παραγράφω = to erase]}. This pattern of lexical ambiguity occupies the 0.28% of the 'noun-verb' category.

6.3 Noun—participle

The word form [ορίζοντας = horizon] is singular nominative of the noun [ορίζοντας = horizon] and present participle (indeclinable) of the verb [ορίζω = to fix]. The ambiguity emerges between the participle of passive voice (nominative, accusative and vocative of neuter) and the adverb. The common suffix is **-α**. This pattern of lexical ambiguity takes up the 0.1% of the lexical ambiguity in total.

6.4 Adjective—noun

This pattern encompasses the common morphological types of nouns and adjectives with the same etymology. To be exact,

- 6.4.1 The noun-adjectives and the adjective (male or female gender) have frequent presentation in the lexicon and occupy the 61.71% of the 'adjective-noun' category: {ηθική < [ηθικός, -η, -ο = moral (adj.)], [ηθική = ethics (noun)]}.
- 6.4.2 The adjectives with suffixes **-ός**, **-ή**, **-ό** and the parasyllabic nouns, male gender, that are accented on the last syllable, with suffix **-ής** produce types with lexical ambiguity in their common suffixes: **-ής**, **-ή**, **-ές**, **-ών**: {κατασκευαστής, -ή, -ές, -ών < [κατασκευαστός = produced], [κατασκευαστής = maker]}, {αγοραστής, -ή, -ές, -ών < [αγοραστός = bought], [αγοραστής =

buyer}}, {καπνιστής, -ή, -ές, -ών < [καπνιστός = smoked], [καπνιστής = smoker]}. The adjectives with suffixes -ός, -ή, -ό and the parisyllabic nouns, male gender, that are accented on the last syllable, with -άς, have the suffix -ά in common which may be: (1) singular genitive of noun or (2) singular nominative, accusative or vocative, neuter gender: {ψαρά < [ψαρός = grizzled], [ψαράς = fisherman]}. This pattern includes all the noun-adjectives (for all the genders, for both singular and plural) and the adjectives (that are accented on the last syllable or the antepenult or the penult) with suffixes -ος, -η, -ο / -ος, -α, -ο / -ύς, -εια, -υ: {θρησκευτικά, -ών < [θρησκευτικός = religious], [θρησκευτικά = the religious education]}, {ακουστική, -ής, -ές, -ών < [ακουστικός = auditory], [ακουστική = acoustics]}, {άδεια, -ας, -ες, -ων < [άδειος = empty], [άδεια = leave]}, {ταχεία, -ας, -ες, -ων < [ταχύς = quick], [ταχεία = express train]}. This pattern of lexical ambiguity occupies the 35.28% of the 'adjective-noun' category.

- 6.4.3 The adjectives with suffixes -ός, -ή, -ό and the parisyllabic nouns, male gender, that are accented on the penult, with suffix -ης, have the genitive of plural in common: {γνωστών < [γνωστός = known], [γνώστης = knowledgeable]}, {γλυπτών < [γλυπτός = sculptural], [γλύπτης = sculptor]}, {δεκτών < [δεκτός = acceptable], [δέκτης = receiver]}, {αποδεκτών < [αποδεκτός = accepted], [αποδέκτης = receiver]}. The same pattern encompasses the adjectives with suffixes -ος, -η, -ο and the parisyllabic nouns, male gender, that are accented on the antepenult, with suffix -ας which form the genitive in the same way: {υπερηλίκων < [υπερήλικος = very old], [υπερήλικας = very old]}, {τριπόδων < [τρίποδος = three-legged],

[τρίποδας = tripod]}. This pattern of lexical ambiguity occupies the 2.14% of the 'adjective-noun' category.

- 6.4.4 The adjectives with suffixes -ής, -ιά, -ί and the parisyllabic nouns, neuter gender, that are accented on the penult, with suffix -ι, form their genitive cases the same way, both in singular and plural: {χρυσαφίου, -ών < [χρυσαφής = golden], [χρυσάφι = gold]}, {ασημιού, -ών < [ασημής = silver], [ασήμι = silver]}, {κερασιού, -ών < [κερασής = cerise], [κεράσι = cherry]}. This pattern of lexical ambiguity occupies the 0.15% of the 'adjective-noun' category.
- 6.4.5 The adjectives with suffixes -ων, -ον and the parisyllabic nouns, male gender, that are accented on the antepenult, with suffix -ας form common types with suffixes -ες, -ων which can be either (1) the plural of the male adjective or (2) the plural of the male noun: {ειδήμονες, -όνων < [ειδήμων = expert (adj)], [ειδήμονας = expert (noun)]}. This pattern of lexical ambiguity occupies the 0.10% of the 'adjective-noun' category.

6.5 Adjective—adjective

- 6.5.1 The adjectives that are accented on the last syllable, with suffixes -ής, -ές and adjectives that are accented on the last syllable, with suffixes -ός, -ή, -ό form the following ambiguous suffixes:
- (1) -ή with morphological attributes:
 - (a) singular accusative of male,
 - (b) plural nominative of neuter,
 - (c) singular nominative of female;
 - (2) -ής with morphological attributes:
 - (a) singular nominative of male or female,
 - (b) singular genitive of female;
 - (3) -ούς with morphological attributes:
 - (a) singular genitive of male or female,
 - (b) plural accusative of male;
 - (4) -ές with morphological attributes:
 - (a) plural nominative, accusative,

or vocative of female, (b) singular nominative, accusative, or vocative of neuter;

- (5) **-ών**, which denotes plural genitive in all genders of adjectives with suffixes **-ής**, **-ές** or adjectives with suffixes **-ός**, **-ή**, **-ό**. For example, {**ακριβή**, **-ής**, **-ούς**, **-ές**, **-ών** < [ακριβής = accurate], [ακριβός = expensive]}, {**ψευδή**, **-ής**, **-ούς**, **-ές**, **-ών** < [ψευδή = false], [ψευδός = lying]}.

This pattern includes a small number of lemmas. However, these lemmas are frequently used and occupy the 63.29% of the wider lexical category (adjective—adjective).

- 6.5.2 The adjectives that are accented on the last syllable, with suffixes **-ιός**, **-ιά**, **-ιό** form the suffixes of female and the plural of male and neuter in the same way as the adjectives that are accented on the last syllable, with suffixes **-ής**, **-ιά**, **-ιό**: {**δεξιά**, **-ιάς**, **-ές**, **-ιών**, **-ιοί**, **-ιούς**, **-ιά** < [δεξιός = right], [δεξής = efficient]: [δεξής] may also have the sense of [right]. The frequency of this pattern is high and occupies the 36.50% of the wider lexical category (adjective—adjective).
- 6.5.3 The adjectives, which are accented on the penult, with suffixes **-ης**, **-α**, **-ικο** form the singular and plural of neuter the same way as the adjectives, which are accented on the antepenult, with suffixes **-ικος**, **-η**, **-ο**: {**βρομιάρικο**, **-ου**, **-α**, **-ων** < [βρομιάρης = rascal], [βρομιάρικος = unclean]}, {**λυσσάρικο**, **-ου**, **-α**, **-ων** < [λυσσάρης = rapid], [λυσσάρικος = mad]}, {**ξεχασιάρικο**, **-ου**, **-α**, **-ων** < [ξεχασιάρης = forgetful], [ξεχασιάρικος = absent-minded]}. This pattern occupies the 0.19% of the wider lexical category (adjective—adjective)

6.6 Adjective—adverb

The neuter of adjectives form their plural nominative, accusative, and vocative cases for all three

degrees, the same way as the equivalent adverb with suffix **-α**: {**απλά**, **-ούστερα** **-ούστερα**, **-ούστατα** < [απλός = simple], [απλά = simply]}, {**βαθιά**, **-ύτερα**, **-ύτατα** < [βαθός = deep], [βαθιά = deeply]}. Predictable lexical ambiguity was also found in the word forms: {**λίγο**, **λιγότερο**, **ελάχιστα** < [λίγος = short], [λίγο = shortly]}, {**πολύ**, **περισσότερο** < [πολύς = much], [πολύ = very]}, {**ευθύς** < [ευθύς = direct], [ευθύς = directly]}. Finally, the non-Greek words that have been incorporated in oral and written speech but are not embodied to Greek morphology belong to this category pattern (adjective-adverb). For example, [σtάνταρ = standard], [σπέσιαλ = special].

This pattern occupies the 33.48% of the lexical ambiguity in total.

6.7 Adjective—verb

- 6.7.1 The adjectives, that are accented on the last syllable, with suffixes **-ής**, **-ές** and the second singular person of present of verbs with suffix **-ώ** (second conjugation, second class) form the suffix **-είς**, with morphological attributes: (1) second singular person of present (for the verb) or (2) plural nominative, accusative or vocative, female gender (for the adjective): {**αδρανείς** < [αδρανής = inert], [αδρανώ = to be inert]}, {**δυστυχείς** < [δυστυχής = unhappy], [δυστυχώ = to be unhappy]}. This pattern dominates the category (adjective—verb) occupies the 92.13% of its total.
- 6.7.2 The singular vocative, male gender, of the adjectives with suffixes **-ος**, **-η**, **-ο** and the second imperative person, active voice (first or second conjugation), or/and some types of past tenses of verbs have common morphological types: {**έκθετε** < [έκθετος = exposed], [έκθέτω = display]}, {**ξέμπλεκε** < [ξέμπλεκος = non-confused man], [ξέμπλεκω = unravel]}, {**πρόσθετε** < [πρόσθετος = additional], [προσθέτω = put in]}, {**μέθυσε** < [μεθάω = get drunk], [μέθυσος = drunkard]}. This pattern occupies the 7.85% of

the lexical ambiguity category ‘adjective – verb’.

6.8 Adjective – participle

The adjective [επιτυχής = successful] forms common suffixes with the participle, active voice (ancient-declination, but still in use in Modern Greek [επιτυχών]: {επιτυχών < [επιτυχής = successful], [επιτυχών = effective]}). This pattern of lexical ambiguity is rare and poorly represented in the lexicon; thus, no percentage has been calculated.

6.9 Participle – adverb

The neuter of participles, passive voice (plural nominative, accusative, vocative for the three degrees) resembles the adverb of all participles with suffix –α: {ανεπτυγμένα < [ανεπτυγμένος = someone that has been developed], [ανεπτυγμένα = develop]}}, {βεβιασμένα < [βεβιασμένος = someone that has been in hurry], [βεβιασμένα = quickly]}}, {επανειλημμένα < [επανειλημμένος = repeated], [επανειλημμένα = repeatedly]}. This pattern of lexical ambiguity is rare and poorly represented in the lexicon; thus, no percentage has been calculated.

6.10 Participle – verb

- 6.10.1 The participle [διαγραφείς = someone that has been deleted] (singular nominative, male gender, ancient declination but still in use in Modern Greek, passive voice, perfective aspect) is identical to the verb type {διαγραφείς < [διαγράφω = delete]} (second singular person, future tense, indicative mood or second person, perfective aspect, subjunctive mood).
- 6.10.2 The word [λήξαν] can be a participle of neutral gender (an archaic type but still in usage today), or a verb (informal style, third person, perfective aspect tense, indicative mood): {λήξαν < [λήξας = someone that that has ended]}.

This pattern of lexical ambiguity is rare and poorly represented in the lexicon; thus, no percentage has been calculated.

6.11 Verb – verb

Some verbs form the declination of the perfective aspect tense (types without augment) of the active voice in the same way. For example, the verbs [καταλαμβάνω = take] and [καταλαβαίνω = understand], [εμπλέκω = confuse] and [μπλέκω = mix up].

The verbs [διώκω = run after] και [διώχνω = send away] are declined in an identical way in both the future and perfective aspect tense of the active and passive voice.

This pattern occupies the 1.62% of the lexical ambiguity in total.

6.12 Adverb – verb

The word [εντάξει] is either an adverb (which means ‘OK’), or the third person singular, (of the indicative mood, future tense or of the subjunctive mood, perfective aspect tense) of the verb [εντάσσω = enlist]. This pattern of lexical ambiguity is rare and poorly represented in the lexicon; thus, no percentage has been calculated.

6.13 Pronouns and other Parts of speech

- 6.13.1 The pronouns μου, σου, του, μας, σας and τους are word forms of personal pronouns and word forms of possessive pronouns as well.
- 6.13.2 The word forms [κάμποσο = enough], [μόνο = only] and [τόσο = so] are pronouns and adverbs.
- 6.13.3 The word form [όσο = as long as] in addition to being a relative pronoun, is a relative adverb and a temporal conjunction.
- 6.13.4 The indefinite pronoun [ένας] may be a cardinal numeral or an indefinite article.
- 6.13.5 The word form μιας is an indefinite pronoun, a numeral, or a temporal conjunction.

- 6.13.6 The singular of the genitive, of male or neuter gender, of the definite article [αυτός] is the same as the informal-style adverb [αυτού = there].

The frequency of this lexical ambiguity category (pronoun – other Parts of speech) takes up the 3.90% of the total.

7 Description of Symptomatic Lexical Ambiguity

In Table 6, we present patterns of symptomatic lexical ambiguity. The most frequent patterns of lexical ambiguity are the following:

- 7.1 The lemma [πισίνα = swimming pool] (noun, female gender) and the lemma [πισινός = backside] (adjective, male gender) form the plural genitive in the same way. The adjective (male gender) [άσφαλτος = infallible] and the noun (female gender) [άσφαλτος = coal - tar] have a common declination.
- 7.2 The nouns [όρος = mountain] (neuter gender) and [όρος = term] (male gender) have the following word forms in common: (1) [όρος] (with the morphological attributes: singular nominative, accusative, vocative, neuter gender, or singular nominative, male gender), (2)

[όρους] (with the morphological attributes: plural accusative, male gender or singular genitive, neuter gender).

- 7.3 Symptomatic lexical ambiguity (verb – noun) is found in the following word forms: {πελεκάνε < [πελεκάνος = pelican], [πελεκώ = hew]}, {βάλανε < [βάλανος = acorn], [βάζω = put]}, {πάταγε < [πάταγος = din], [πατώ = step]}, {βάλτε < [βάλτος = swamp], [βάζω = put]}, {έρανε < [έρανος = collection], [ραίνω = sprinkle]}, {άσε < [άσος = ace], [αφήνω = leave]}.
- 7.4 The verbs [δρω = act] and [εδράζω = strengthen] form common word forms in all suffixes of the singular and in the third person of perfective aspect tense: {έδρασα, –εσ, –ε, –αν < [δρω = act], [εδράζω = strengthen]}. The word [ξέρανε] can be a second person, singular, imperative mood, perfective aspect or third person, singular, indicative mood, perfective aspect of the verb [ξεραίνω = dry]. It can also be a third person, plural, imperfect tense of the verb [ξέρω = to know]. The word form ((ξανα)δέστε is a second person, plural, imperative mood, perfective aspect tense of the verb [(ξανα)βλέπω = see (again)] and a second person, plural, imperative mood, perfective aspect tense of the verb [(ξανα)δένω = bind (again)]. The word form [πέστε] is a second

Table 6 Statistics concerning symptomatic ambiguity

Symptomatic ambiguity							
Part of speech 1	Part of speech 2	Word forms	Example	Part of speech 1	Part of speech 2	Word forms	Example
Interjection	Participle	3	άι	Pronoun	Article	50	τα
	Adjective	12	καλέ		Noun	57	σου
	Adverb	3	ίσα		Adjective	39	ίδια
	Verb	3	ορίστε		Verb	8	εμείς
	Noun	104	γιούχα		Adverb	18	κάμποσο
Preposition	Pronoun	16	με	Conjunction	Conjunction	3	
		29	συν		Noun	11	μόλο
		12	υπό		Verb	4	όντας
					Adverb	33	πριν
Art	Noun	5	η	Pronoun	Participle	9	να
					Pronoun	6	όσον

person, imperative mood, present tense, active voice of the verb [λέω = say], or it is a second person, imperative mood, perfective aspect, active voice of the verb [πέφτω = fall]. The lemmas [κτίζω = build] and [εκτίω = serve]; [βοσκάω = pasture] and [βόσκω = graze]; [ανεβάζω = take up] and [ανεβαίνω = walk up]; [γυρνάω = return] and [γυρίζω = turn]; [κατεβαίνω = climb down] and [κατεβάζω = bring]; [νιώθω = feel] and [νιώνω = become young]; [πίνω = drink] and [πιέζω = press] form common word forms in unique and unpredictable patterns.

- 7.5 The word [ανάθεμα] is either an exclamation, or the nominative, accusative, or vocative cases of the neuter, singular of the noun [ανάθεμα = anathema].
- 7.6 The word [ήπια] can either be: (1) a verb with morphological attributes: first person, singular, indicative mood, perfective aspect [πίνω = drink], (2) an adjective with morphological attributes: nominative, accusative, vocative, singular, female gender [ήπιος = mild], or (3) an adjective with morphological attributes nominative, accusative, vocative, plural, neuter gender [ήπιος = mild]. Furthermore, it can also be an adverb. The word forms [ήπιες] and [ήπια] can be either the female of an adjective or a verb (indicative mood, perfective aspect).
- 7.7 The informal-style word [δω] that derives from the adverb [εδώ = here] is also a verb [βλέπω = see] (either first person, singular, subjunctive mood, perfective aspect tense, or first person, singular, indicative mood, future tense).
- 7.8 The word type [δικών] can be a possessive pronoun or the genitive clause, plural of the noun [δίκη = trial].
- 7.9 The declination of the singular, female gender of the interrogative pronoun [πόσος = how] is the same as the one of the noun [πόση = drinking], singular, female gender.

7.10 The personal or possessive pronoun [σου] is also a noun (indeclinable noun, foreign word).

7.11 The personal pronouns [με] and [σε] may also be prepositions.

7.12 The word form [παρά] can be a preposition, a comparative conjunction, or a noun [παράς = money] (genitive, accusative, vocative, singular).

8 Conclusion

A detailed study of lexical ambiguity for Modern Greek has been presented, involving both theoretical description and experimental verification and quantification using an electronic lexicon.

The presented patterns and classification describe the deep lexical ambiguity of Modern Greek. These patterns are an important source for word disambiguation (for the correct interpretation of ambiguous words, of clauses in complex sentences, of the constitution of word forms, and of the syntactic role of constituents) (Koskenniemi, 1983).

Lexical ambiguity results are shown in detail and according to their patterns in Table 5 and Fig. 5. The statistics (i.e. the quantified frequencies) refer only to predictable forms of lexical ambiguity, as symptomatic ambiguity involves unrelated word forms that cannot be counted accurately, as explained in Section 5. Each ambiguity class complicates ulterior linguistic processing (Orphanos and Christodoulakis, 1999). If a word displays POS ambiguity (e.g. verb – noun, adjective – verb, pronoun – clitic etc.), the resulting analysis is more complicated. POS is the most important morphosyntactic feature because it defines the extra morphosyntactic trait (subject, object, pre-phrase, etc.). For example, the word [κλειδώσεις] can be either a noun-phrase (subject or object) or a verb-phrase.

Finally, concerning further research, one of the immediate applications of the analysis presented in this article is the construction of a grammar checker for Modern Greek that will be able to correctly attribute wider structures with word forms that belong to more than one lemma, once they are disambiguated.

Acknowledgements

The authors wish to thank the anonymous reviewers of the previous version of this paper for their valuable comments, which are copied below, along with the according revisions in the current version for their convenience.

References

- Allen, J.** (1987). *Natural Language Understanding*. California: Benjamin Cummings.
- Babinotis, G.** (1998). *The Dictionary for the Modern Greek* [in Greek]. Athens: Lexicology Center.
- Bresnan, J.** (1979). *Theory of Complementation in English Syntax*. New York: Garland.
- Boguraev, B. and Pustejovsky, J.** (1990). *The role of knowledge representation in Lexicon Design, Proceedings of the 13th conference on computation linguistics*. Finland August, 1990, pp. 36–41.
- Chomsky, N.** (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Church, K. W.** (1980). On parsing strategies and closure. In: *Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics and Parasession on Topics in Interactive Discourse*. Philadelphia, PA, June 19–22, 1980, pp. 107–111.
- Church, K.** (1988). *A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text, Proceeding of the Second Conference on Applied Natural Language Processing*. Austin, TX, pp. 136–43.
- Crystal, D.** (2000). *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell. Trans. Xydopoulos, G. (2000). Athens: Patakis.
- Daelemans, W., Zavrel, J., Berck, P., and Gillis, S.** (1996). MBT: A memory-based part of speech Tagger-Generator. In Ejerhed, E. and Dagan, I. (eds). *Proceedings of 4th Workshop on Very Large Corpora*. ACL SIGDAT, 1996, Copenhagen, Denmark, pp. 14–27.
- Dermatas, E. and Kokkinakis, G.** (1995). Automatic stochastic tagging of natural language texts. *Computational Linguistics*, 21: 137–163.
- Filippaki-Warburton, E.** (1992). *Introduction to the Theoretical Linguistics*. Athens: Nefeli.
- Fitrakis** (1997). *The Major Greek Dictionary* [in Greek]. Athens: Patakis.
- Fromkin, V. and Rodman, R.** (1988). *An Introduction to Language*. 4th ed. New York: Holt, Rinehart & Winston, Inc.
- Gakis, P., Orphanos, G., Michalakou, A., and Iordanidou, A.** (1999). *Morphosyntactic Ambiguity in Modern Greek: The Case of Adjective-Noun-Adverb, Proceedings of the 20th Annual Meeting of the Department of Linguistics* [in Greek]. Thessaloniki, Greece.
- Gakis, P., Panagiotakopoulos, Ch., Sgarbas, K., and Tsalidis, C.** (2012). Design and implementation of an electronic lexicon for Modern Greek. *Literary and Linguistic Computing*, 27(2): 155–69.
- Grishman, R.** (1986). *Computational Linguistics: An Introduction*. Cambridge: University Press.
- Kanakis, K.** (2007). *Introduction to the Pragmatism. Gnostic kai Social Aspects of the Linguistic Usage*. Athens: Eikostou Protou.
- Kaplan, R. and Bresnan, J.** (1982). Lexical-functional grammar: a formal system for grammatical representation. In Bresnan, J. (ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MA: The MIT Press.
- Koskenniemi, K.** (1983). *Two-Level Morphology: A General Computational Model for Word-form Recognition and Production*. Helsinki: University of Helsinki, Department of General Linguistics. Publication 11.
- Kriaras, E.** (1995). *Greek Dictionary of the Modern Demotic Language* [in Greek]. Thessaloniki: Center for Greek Language.
- Lyons, J.** (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Lyons, J.** (1977). *Semantics*, vol. 1 and 2. Cambridge: Cambridge University Press.
- Lyons, J.** (1995). *Introduction to Linguistics* [In Greek, trans. M. Arapopoulou, Ed. G. Karanassios]. Athens: Patakis.
- Márquez, L. and Padró, L.** (1997). *A Flexible POS Tagger Using an Automatically Acquired Language Model, Proceedings of EACL/ACL '97*. Madrid, Spain.
- MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S.** (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4): 676–703.
- Mahesh, K. and Nirenburg, S.** (1997). Knowledge-Based Systems for Natural Language Processing. In: *The Computer Science and Engineering Handbook*. Maine: CRC Press (published in cooperation with ACM).

- Orphanos, G.** (2000). *Computational Morphosyntactic Analysis of Greek*. PhD thesis, University of Patras.
- Orphanos, G. and Christodoulakis, D.** (1999). *POS Disambiguation and Unknown Word Guessing with Decision Trees*, *Proceedings of the Annual Meeting of The European Chapter of The Association of Computational (EACL 1999)*. Pennsylvania, USA, June 1999, pp.134–141.
- Orphanos, G. and Tsalidis, C.** (1999). *Combining hand-crafted and corpus-acquired Lexical Knowledge into a Morphosyntactic Tagger*, *Proceedings of the 2nd CLUK Research Colloquium*. Essex, UK.
- Papageorgiou, G., Prokopidis, X., Giouli, B., and Piperidis, S.** (2000). A Unified POS Tagging Architecture and its Application to Greek. In: *Proceedings of the 2nd Language Resources and Evaluation Conference*. Athens, May 2000, pp. 1455–62.
- Petasis, G., Paliouras, G., Karkaletsis, V., Soyropoulos, D., and Androutopoulos, I.** (1999). *Resolving Part-Of-Speech Ambiguity in the Greek language Using Learning Techniques*, *Proceedings of the ECCAI Advanced Course on Artificial Intelligence (ACAI '99)*. Chania, Greece, July 1999.
- Petasis, G., Vichot, F., Wolinski, F., Paliouras, G., Karkaletsis, V., and Soyropoulos, D.** (2001). *Using Machine Learning to Maintain Rule-based Named - Entity Recognition and Classification Systems*, *Proceedings of the 39th Conference of Association for Computational Linguistics (ACL-EACL 2001)*. Toulouse, France, July 2001, pp. 418–25.
- Pollard, C. and Sag, A.** (1987). *Information-Based Syntax and Semantics*, vol. 1. Stanford University, California: CSLI Publications.
- Radford, A.** (1988). *Transformational Grammar: A First Course (Cambridge Textbooks in Linguistics)*. Cambridge: Cambridge University Press.
- Sgarbas, K., Fakotakis, N., and Kokkinakis, G.** (1995). A PC-KIMMO-based morphological description of Modern Greek. *Literary and Linguistic Computing*, 10(3): 189–201.
- Triantafyllidis, M.** (1998). *Dictionary of Common Modern Greek Language* [in Greek]. Thessaloniki: Institute of Greek studies of Aristotle University of Thessalonica, Manolis Triantafyllidis Foundation.
- Van Eijck, J. and Jaspars, J.** (1996). *Ambiguity and Reasoning*. Technical Report CS-R9616. Amsterdam: Dutch National Research Institute for Mathematics and Computer Science.

Note

- 1 For more info about Neurolingo Company and Neurolongo Lexicon, see <http://www.neurolingo.gr>.