

Multi-word expressions in Modern Greek: present and future



Anna Anastassiadis (University of Thessaloniki),
Angeliki Fotopoulou (ILSP/"Athena" RIC)&
Tita Kyriacopoulou (University of Paris-Est,
Descartes)

Berlin, ICLG12

Multi-word expressions in Modern Greek

Terminology

- EN. idioms, frozen/fixed & semi-fixed/semi-frozen expressions, (set) phrases, formulae, collocations, cranberry collocations, phraseologisms, phraseological units, etc.

- GR Fixed and semi-fixed expressions, stereotyped expressions, idioms, phrases, expressions, phraseologisms, cooccurrences, collocations etc.

->The definition of phrase/expression is rather confusing. The denomination varies according to the criteria followed in order to define and classify the phrases.

Multi-word expressions in Modern Greek

General

- From phraseology to NLP applications

Dictionaries

Dictionary of Modern Greek (1998): **ΦΡ.** -> wide range of fixedness which comprises lexical collocations but also proverbs.

Dictionary of Common Greek (M. Triandafyllidis Foundation 1998): distinction between **phrase** and **expression** according to their relation with the literal sense.

Dictionary of Modern Greek (2014): **ΣΥΜΠΛ.**-> for multi-word compounds and **ΦΡ.** -> for idioms.

Multi-word expressions in Modern Greek

- Anastassiadis-Symeonidis 1986
 - Nominal expressions
- Fotopoulou, 1993
 - Verbal expressions

Multi-word expressions in Modern Greek

Select papers

- ❑ **Linguistic research:** Μότσιου 1987, Moustaki 1995, Fotopoulou 1993, Γαβριηλίδου, Ζ. & Νάκας, Α. 2002, Sfetsiou, 2003, Αναστασιάδη-Συμεωνίδη & Ευθυμίου 2006, Θώμου 2006, Pantazara et al. 2008, Γεωργακόπουλος et al. 2009, Fotopoulou et al. 2009, Φούφη 2012, Fotopoulou et al. 2014
- ❑ **Interdisciplinary** (mostly psycholinguistic): Μινη 2009, Mini et al. 2011, Anastassiadis-Syméonidis & Voga 2011, Diakogiorgi & Fotopoulou 2012
- ❑ **Computational Linguistics:** Gavriilidou 1997, Valetopoulos 2007, Σφέτσιου 2007, Φωτοπούλου et al. 2008, Linardaki et al. 2010, Παπαγεωργίου 2011, Μορφοπούλου 2011, Τζιάφα 2012, Ιωαννίδου 2013, Σαμαρίδη 2014, Φούφη 2014, Fotopoulou et al. 2014
- ❑ **Teaching:** Αναστασιάδη-Συμεωνίδη & Ευθυμίου 2006, Δημοπούλου 2010, Χέλμη 2011, Θώμου 2014
- ❑ **Monolingual lexicography:** Ιορδανίδου 2001 and **bilingual:** Gavriilidou 1997, Moustaki et al. 2008.

Multi-word expressions in Modern Greek

Questions

- ❑ 1) Can multi-word expressions be distinguished from free structures?
- ❑ 2) Is fixedness an absolute or graded concept? Is it entirely or partially expressed?
- ❑ 3) In which grammatical categories do multi-word expressions belong?
- ❑ 4) Does a group of nouns or verbs comprise other sub-categories?
- ❑ 5) Is automatic extraction of multi-word expressions possible?
- ❑ 6) How are multi-word expressions organized in a mental lexicon and how are they processed?

Multi-word expressions in Modern Greek

Definitions

- A **multi-word expression** is a sequence of words which constitutes a semantic unit in the general language, e.g. *μαύρη αγορά/black market*, *κρέμα γάλακτος/cream* or in the scientific and technical vocabulary, e.g. *βαρύ ύδωρ/heavy water*, *αγωγή του πολίτη/citizenship education*, *κάθομαι σ'αναμμένα κάρβουνα/pins and needles*.
- Compound phonological structure
- Compound lexical and morphological structure
- They are formed by more than one words but they constitute a single unit at the semantic level.

Multi-word expressions in Modern Greek

- There are multi-word structures in all grammatical categories:
 - **nominal**
 - Compounds, collocations: παιδική χαρά/*park*, δυνατός καφές/*strong coffee*
 - **verbal**
 - Fixed: κάθονται σ' αναμμένα κάρβουνα/*pins and needles*
 - Collocations, structures with support verbs: τρώω χαστούκια
 - **adjectival**: καθωσπρέπει
 - **adverbial**: επί παντός επιστητού/*on just about everything*
- **Proverbs, sayings, clichés**

Multi-word expressions in Modern Greek

- Nominal multi-word expressions:
 - Αναστασιάδη-Συμεωνίδη 1986, Kyriakopoulou 2011, Φούφη 2014.
- Proper nouns, the so-called named entities: initialisms and acronyms.
- Verbal multi-word expressions
 - Fixed expressions, (Fotopoulou 1993, Moustaki 1995, Θώμου 2006 Mini 2009, Dimopoulou 2010)
 - Expressions with support verbs: Fotopoulou 1985, 1992, Τσολάκη 1997, Gavriilidou 2004, Πανταζάρα 2005, Sfetsou, ...)
 - Proverbs (Τσακνάκη 2005, Χιώτη 2010), sayings etc.
 - Non-inflected multi-word expressions (see also Αναστασιάδη-Συμεωνίδη & Ευθυμίου 2006).

Multi-word expressions in Modern Greek

Processing axes

- ❑ Vocabulary-Lexicology
 - ❑ Morphology (mostly nouns, inflection)
 - ❑ Syntax (+ dictionary)
- > Interdisciplinary: syntax – psycholinguistics
- > Natural Language Processing

Multi-word expressions in Modern Greek

Theoretical approaches 1

O. Jespersen in *Philosophy of Grammar* (1924) was the first to make the distinction between combinational freedom and fixedness

According Lexicon-Grammar for the classification of French idioms, by M. Gross (1982), were:

- 1) the semantic criterion, according to which the meaning of an idiom is not derived from the meaning of its parts, and
- 2) the lexical-structural criterion, according to which one or more elements of the clause are lexically invariable in relation to the verb.

(1a) τα φόρτωσα στον κόκορα

[literally 'I loaded them on the rooster'] meaning 'I did not act at all, as I was feeling lazy

(1b)*τα φόρτωσα στους κόκορες **fixed**

[τα φόρτωσα στο κάρο] 'I loaded them on the car'

Non compositionality – non substitutability

Multi-word expressions in Modern Greek

Theoretical approaches 2

(2) σκοτείνιασε (το πρόσωπό + η όψη + το χαμόγελο + το βλέμμα + τα μάτια) του

(2α) *σκοτείνιασε το κεφάλι του

(2β) *σκοτείνιασε το μάτι του

[his (face + face + smile + gaze + eyes) darkened]

Compositionality - non substitutability

Transparency - degree of fixedness

Gradation in fixedness extends from non-analysable to partially analysable (G. Gross, 1996, Sag et al., 2001).

Multi-word expressions in Modern Greek

Theoretical approaches 3

Nunberg et al. (1994) in HPSG (Head-driven Phrase Structure Grammar) study the expressions, in priority, from a semantic aspect, they localize many dimensions in a “prototypical idiom”, such as **conventionality**, syntactic restriction, figuration (metaphor, metonymy, hyperbole etc. – proverbiality, etc.).

Conventionality, which is then divided into other parameters, is necessary.

Multi-word expressions in Modern Greek

Theoretical approaches 4

Construction Grammar (Fillmore et al., 1988): a phrase is idiomatic if the speakers attribute to this phrase a certain meaning. On the other hand, if a person knows only the vocabulary and the grammar of a language, cannot be aware of this phrase or its meaning or if this phrase constitutes a conventional, acceptable phrase.

-> it seems that even if the theoretical basis is different, there are common characteristics.

Multi-word expressions in Modern Greek

Theoretical approaches - discussion

- ❑ A phrase could be considered as fixed according to a theoretical framework, but according to another it could be less fixed (semi-fixed) or collocation or none of the two, it could be a simple use of the verb.
- ❑ According to Mel'cuk (1995), *spill the beans* is an idiom, which means that it is a **non-analysable** sequence, whose meaning is not deducted from the meaning of its constituents.
- ❑ According to Nunberg et al. (1994), *spill the beans* is an idiomatically combining expression, which means that it has an important degree of analysability.

Multi-word expressions in Modern Greek

- ❑ Fixedness - Degree of fixedness
- ❑ Non compositionality vs compositionality.
- ❑ Transparency vs opacity

Multi-word expressions in Modern Greek

Fixedness

Fixedness is a graded concept which characterizes all the multi-word expressions.

G. Gross, 1996: a sequence of simple words is a multi-word expression when at least one of its syntactic, distributional or semantic properties is not deducted from the properties of its constituents:

χωρικά ύδατα/territorial waters

* *χωρικό ύδωρ/territorial water*

Multi-word expressions in Modern Greek

Transparency vs opacity¹

Types:

- Transparent multi-word expression: *πήρε τον κατήφορο/go downhill,*
- Opaque: *την πάτησα/have it bad,*
- With compositional meaning: *η ζωή του κρέμεται από μια κλωστή/his life is hanging in the balance*
- Non-compositional meaning: *βράσε ρύζι!*

Transparent multi-word expression: *πήρε τον κατήφορο [από τότε που πέθανε η μάνα του]/he went downhill [since his mother died]*

Multi-word expressions in Modern Greek

Transparency vs opacity²

In an opaque multi-word expressions we can localize the point of opacity (Anastassiadis-Syméonidis & Voga 2011):

- α) presence of an unknown word, e.g. *στα κουτουρού/blindly*,
- β) presence of the weak form of personal pronouns e.g. *τα πήρα στο κρανίο/I went mad*,
- γ) necessity of use of the extra-linguistic environment, e.g. *είναι του δρόμου*.

On the other hand, the multi-word expression *πνίγομαι σε μια κουταλιά νερό* is transparent. Also, the transparency/opacity can play a different role in the comprehension and production of multi-word expressions

Multi-word expressions in Modern Greek

NLP applications

Transparency vs opacity 3

In a pilot research (Mini, 2009 and Mini and al. 2011), the processing of expressions by schoolchildren demonstrated that the comprehension is more easily achieved when the expression is rather transparent but the most important aspect is the context and the daily use of the expression.

Multi-word expressions in Modern Greek

Compositionality vs non compositionality

- ❑ The meaning of a multi-word expression is never totally compositional for the auditor, which means that it is not deduced by the meaning of its constituents nor by the combination of its constituents according to the syntactic rules (Anastassiadis-Symeonidis & Voga 2011).
- ❑ The meaning of the multi-word expressions which is non-compositional is given directly without previous processing because multi-word expressions are considered to have meaningless constituents, according to olden psycholinguistic theories.

Multi-word expressions in Modern Greek

Un-fixedness 1

- ❑ The speaker's freedom is inversely proportional to the degree of fixedness of multi-word expressions. Un-fixedness may either take the form of **puns**, i.e. conscious, deliberate, ephemeral, or **folk etymology**, i.e. unconscious, not deliberate deviation.
- ❑ In the case of **puns**, un-fixedness is based on either polysemy / homonymy, e.g. ελληνική γλώσσα: a) Greek language, b) sole (kind of fish), or
- ❑ lexical substitution, e.g. Σαν βγεις στον πηγαιμό για το Παγκράτι 'As you set out for Pangrati'. The lexical substitution presupposes the existence of a very well known word sequence coming from the literature. Palimpsest according to Galisson (1995: 105).
- ❑ The subjacent lexicalized text of the poem by Cavafy: *As you set out for Ithaca*.

Multi-word expressions in Modern Greek

Un-fixedness 2

- In the case of **folk etymology**, a mechanism of cognitive origin, the speaker normalizes a formula or an unknown and opaque sequence making it transparent.
- Opacity may cause folk etymology, e.g. *βρώμα και δυσωδία* 'filth and stench',
- *βρώμα* 'food' in the 4th century AD, today 'filth/dirtiness' (Anastassiadis-Syméonidis 2003).

Multi-word expressions in Modern Greek

Nominal expressions: definition

1. Multi-word compound common nouns

The sequence of at least two simple words with a separator (space, hyphen, apostrophe).

κινητό τηλέφωνο/mobile phone,

παιδί-θαύμα/child prodigy and

αντ' αυτού/instead of him are compound nouns (Kyriakopoulou, 2005).

Multi-word expressions in Modern Greek

Nominal expressions : structures

2. Multi-word compound common nouns

The most frequent types of multi-word compounds in Modern Greek are nominal (Anastassiadis-Symeonidis 1986: 134, 147):

- a) **Adj + Noun**: *ψυχρός πόλεμος/cold war*,
- b) **Noun + (Definite Article (gen)) + Noun (gen)**: *γλυκό του κουταλιού, φακοί επαφής/lens*
- c) **Noun + Noun**: *λέξη-κλειδί/keyword*.

(Fotopoulou et al. 2008) have added 2 more types:

- (d) **Noun + [Prep + N]**: *φόνος εκ προμελέτης/ premeditated murder*,
- (e) **[Prep + Noun] + Noun**: *διά βίου μάθηση/long-life learning*.

Multi-word expressions in Modern Greek

Nominal expressions : degree of opacity

The most simple type of expressions is the entirely fixed expressions, e.g. *αποδιοπομπαίος τράγος/scapegoat*.

In the majority of multi-word compound nouns, only a part of the structure is fixed and the rest is a free structure.

For instance, in sequences

*υπαίθρια/τοπική/κεντρική/λαϊκή/ελεύθερη/μαύρη
αγορά/open/local/central/open/free/black market,*

there is **a grade of opacity** but at the same time there is lexical freedom.

Multi-word expressions in Modern Greek

Nominal expressions - morphology – inflection 1

The inflection of a multi-word compound depends on:

- The inflection of its simple units.
- The syntactic rules of Modern Greek. The second noun either remains uninflected or it is at the same case as the first noun, e.g. *οι λέξεις-κλειδί*, *οι λέξεις κλειδιά/keyword*.

Multi-word expressions in Modern Greek

Nominal expressions - morphology – inflection 2

- Noun+ (definite article - genitive) + Noun (genitive), the information about the inflection is related only to the first noun.
- Adjective + Noun, morphological information is given for each one of the simple constituents.

Anastassiadis-Symeonidis 1896, Fotopoulou et al., 2008, Foufi 2012, Kyriakopoulou 2011, Gavriilidou Z. 1997.

Multi-word expressions in Modern Greek

Nominal expressions - morphology – inflection 3

□ Multiflex platform (Savary)

Inflection codes for this type of multi-word compounds (AN)

□ *σχολική(σχολικός.A1:fs)* ,

- *A1* is the inflection code which corresponds to the inflection vectors of the adjective *σχολικός* and *fs* is morphological information according to the DELA formalism.

For the whole multi-word expression the entry in the electronic morphological dictionary DelaGR is:

σχολική(σχολικός.A1:fs) τάντα(τάντα.N22:fs),NC_AN

DELA (Dictionnaire Electronique du LADL) Laboratoire d'Automatique Documentaire et Linguistique (LADL).

Multi-word expressions in Modern Greek

Nominal expressions - lexical particularities

- Lexical particularities AN
 - **Δεύτερη Παρουσία* 'Second Coming'
 - *Δευτέρα Παρουσία*

 - *δεύτερη δέσμη* 'second orientation'
 - **δευτέρα δέσμη*

αγγλικό(αγγλικός.A1:Ans)

χιούμορ(χιούμορ.N305:Nns),NC_sing+Abst

'English humor'

Multi-word expressions in Modern Greek

Nominal expressions- lexical variations

- Lexical variations AN
 - *εθελουσία έξοδος* 'voluntary exit' (learned form)
 - *εθελούσια έξοδος* (neutral form)
- Variations with synonyms AN
 - *δίσεκτο έτος* 'leap year' (learned form)
 - *δίσεκτος χρόνος* (neutral form)

But

δούρειος ίππος 'Trojan horse' (learned form)
**δούρειο άλογο* (neutral form)

- Spelling variations AN
 - *ασφαλιστική εταιρεία* 'insurance company'
 - *ασφαλιστική εταιρία*

Multi-word expressions in Modern Greek

Nominal expressions -named entities

□ Named Entities

■ Ch. Symeonidis 1992

- i) names of persons or saints: *Μέγας Αλέξανδρος, Μέγας Κωνσταντίνος, Άγιος Δημήτριος,*
- ii) place names (names of countries, of cities, Symeonidis 2010, names of regions etc.): *Ανατολική Ευρώπη, Εύξεινος Πόντος,*
- iii) names of mountains or hydronyms: *λίμνη Βόλβη,*
- iv) organizations: *Αγροτική Τράπεζα, Βρετανικό Μουσείο, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης,*
- v) authorial texts: *Πράσινη Βίβλος, Εθνικός Ύμνος,*
- vi) events: *Γαλλική Επανάσταση, Ολυμπιακοί Αγώνες.*

Multi-word expressions in Modern Greek

Nominal expressions - truncated forms 1

■ Acronyms

□ *Αθλητική Ένωση Κωνσταντινουπόλεως* → *ΑΕΚ* → *ΑΕΚτζής/αεκτζής* 'οπαδός της ΑΕΚ'

□ **Initialisms** are uninflected parts of the speech, the gender and the number of the main noun determine the gender and the number of the initialism

ο Γενικός Γραμματέας → *ο Γ.Γ.* 'General Secretary'

Multi-word expressions in Modern Greek

Nominal expressions-truncated forms 2

■ Acronyms

- *Ελληνικά Ταχυδρομεία* → *ΕΛ.ΤΑ.*
 - *Ελληνική Αστυνομία* → *ΕΛ.ΑΣ.*
 - *Δημοτικό Περιφερειακό Θέατρο* → *ΔΗ.ΠΕ.ΘΕ.*
 - *βιβλίο περιπτέρου* → *βίπερ*
 - *προγνωστικά ποδοσφαίρου* → *ΠΡΟΠΟ*
- The electronic dictionary (DelaGR) contains 193 initialisms and 15 acronyms.

Multi-word expressions in Modern Greek

Verbal expressions – fixed expressions 1

- ❑ Biber et al. (2003), multi-word expressions are blocks of words or else lexical combinations which are often repeated and frequently used.
- ❑ G. Gross (1996) proposes criteria (features and restrictions) for the French language.
- ❑ Anasstassiadis - Symeonidis & Efthimiou (2006): semantic opacity and non-compositionality are not absolute but graded phenomena.

Multi-word expressions in Modern Greek

Verbal expressions – fixed expressions 2

- Stereotypes/fixed verbal expressions can be characterized by:
 - Lack, absence of constituents: of the definite articles : *παίρνω πόδι* (kick out).
 - Learned forms (Iordanidou 2001): *η ισχύς εν τη ενώσει*
 - Features of oral speech: *τα χασα, για όνομα του Θεού!, δεν έχω μούτρα να+verb*
 - Structural patterns: with diminutive: *την προσέχει τη ζωούλα του*, simile: *βρίζει σα χαμάλης*, repetition: *κόσμος και κοσμάκης, είδα κι απόειδα*
- Transformations are not allowed:
 - Passivization: *πληρώνω τη νύφη* – **η νύφη πληρώθηκε από μένα*, [pay the bill / pick up the bill]
 - Pronominalization: *πλήρωσα τα σπασμένα* - **τα πλήρωσα*, [pay the bill / pick up the bill]

Multi-word expressions in Modern Greek

Verbal expressions - fixed expressions 3

- ❑ Cliticization: *πλήρωσα τα σπασμένα* - **τα σπασμένα, τα πλήρωσα,*
- ❑ Raising: *πλήρωσα τα σπασμένα* - **είναι τα σπασμένα που πλήρωσα,*
- ❑ Relativization: *πλήρωσα τα σπασμένα* - **τα σπασμένα που πλήρωσα,*
- ❑ Question: *πλήρωσα τα σπασμένα* - **τι πλήρωσα;*
- ❑ None of the constituents of a stereotyped expression can be actualized separately: *πλήρωσα τα σπασμένα* - **πλήρωσα αυτά τα σπασμένα/τα σπασμένα μου.*
η Μαρία έγινε Τούρκος 'θύμωσε' – **η Μαρία έγινε Τουρκάλα.*
- ❑ As for modalities: *καρφί δε μου καίγεται, δεν έχω μούτρα*
va+verb

Multi-word expressions in Modern Greek

Verbal expressions –fixed expressions 4

A. Fotopoulou (PhD 1993. Further discussion:1997, 1998, 2001...), Mini (PhD, 2009, 2012), Morfopoulou (MA Diss. 2011)

- > 6000 verbal expressions with distributional and transformational properties
- > 20 Classes of Fixed expressions

Multi-word expressions in Modern Greek

Verbal expressions –fixed expressions 5

- ❑ Fotopoulou's study (1993) on modern Greek idiomatic expressions the three features mentioned by M. Gross (1982) can be seen: 1) the possible combination of fixed and free constituents within an idiomatic expression;
- ❑ 2) the possibility for some lexical freedom associated to a restricted alternation of one or more constituents of the idiomatic expression and
- ❑ 3) the presence of a continuum between idiomatic and non-idiomatic expressions.
- ❑ Based on Greek fixed expressions, the **classification** proposed aims at clarifying the syntactic and semantic properties of idioms emphasising the range and the variety of these properties: (1) Idiom fixedness can be limited to only certain constituents of the sentence

Multi-word expressions in Modern Greek

Verbal expressions – fixed expressions 6

- (1) Fixedness of this expression
 - NO *δίνει τόπο στην οργή*
 - NO **gives** way to anger
 - (NO swallows one's pride/anger)

relies on the relation of the verb *δίνω* /give and its arguments.
The first component (= subject) is not fixed;

- (2) There are small groups of fixed constructions with similar meanings, allowing some degree of constituent variation within the idiomatic expression.
- (3) It appears that fixed sentences are set upon a continuum, starting from free structured combinations and ending with fixed expressions specified as prototypical, i.e. semantically opaque and structurally fixed. For example:

Multi-word expressions in Modern Greek

Verbal expressions – fixed expressions 7

Structure	Example (Greek)	Word for word translation	Translation
V C0	Το ποτήρι ξεχείλισε	The glass overflowed	‘My patience has been exhausted.’
VC0 Ngen	Δεν ιδρώνει τ’ αυτί του N.	It does not sweat the ear of N	Do not care about what others tell him
N1 V C0 =: Cl(acc) VC0	Τον φοβήθηκε το μάτι μου	Him feared the eye mine	I was scared stiff by him or his actions

Multi-word expressions in Modern Greek

Verbal expressions – fixed expressions 8

- The classification of sentences aimed -among others- at exploiting these data for the development of language processing software. This aim justifies the collection of many such data, which in practice means that expressions that do not really seem to be so fixed can also be included in the classification:
froncer les sourcils 'frown' (Lamiroy 2003).

Multi-word expressions in Modern Greek

Verbal expressions – fixed expressions 9

Other relatives studies (Fotopoulou, 1989, 1990, 1992, 1993, 1997)

- The **types in genitive** of stereotyped phrases are systematically analyzed (free or stereotyped):
του** πέταξε σπόντες* and *χαίρει **άκρας υγείας.
- In some categories of stereotyped phrases in the form of stereotyped argument + free argument in genitive, we study the word order.
- The relationship between fixed expressions and Light/Support verb Constructions is examined

Multi-word expressions in Modern Greek

Verbal expressions – fixed expressions 10

Chioti 2010. This study focuses on the influence of foreign languages (mainly the ones of neighbouring countries) on fixed expressions and especially on the influence of Turkish on Greek. Furthermore, a classification based on semantic and lexical features, such as diminutives, proper nouns, and categories such as plants, animals and abstract concepts is attempted.

Dimopoulou 2010. Semantic analysis of stereotypes and idioms, within the semantic field of nature, and the practical use of the relative conclusions in teaching Modern Greek as a second / foreign language.

Also

▣ *Moustaki, 1993, 1995, Helmi, 2011, Samaridi, 2014.*

Multi-word expressions in Modern Greek

Verbal expressions – collocations-light/support verb constructions

- The intermediate space between free and fixed expressions consists of verbal structures of the (**V** + **Noun**) type that present semantic transparency but still have lexical and syntactic restrictions. These structures are called *semi-fixed expressions, collocations* or *lexical collocations*.

Δίνω ένα βιβλίο 'give a book'

-> **free construction**

Δίνω κουράγιο

-> **collocation -> V^{sup} construction**

Δίνω τόπο στην οργή

-> **fixed construction**

Multi-word expressions in Modern Greek

Verbal expressions - collocations 1

Lexical collocations

Sinclair & Carter (1991), collocation is the co-occurrence of two or more words in short distance

González-Rey (2002) defines lexical collocations as the combinations of words which “prefer” to coexist in discourse having anaphoric-declarative function while keeping the same meaning as simple units.

-> semi-free sequences and simultaneously semi-fixed with compositional meaning.

Multi-word expressions in Modern Greek

Verbal expressions - collocations 2

- Gurrutxaga & Alegria (2011):
 - i) semi-compositional combinations where the noun keeps its literal meaning while the verb is a support verb or it has a special meaning in this combination and
 - ii) compositional combinations with lexical restrictions where there is no possibility of replacing the verbs with its synonyms.

Multi-word expressions in Modern Greek

Verbal expressions - collocations 3

Thomou (2006) Lexical collocations in Modern Greek as a foreign language, mostly studies the structures in the form of verb + nominal complement

- “The lexical collocations are lexical combinations of a certain type which during the lexical production are situated between on one hand the **free phrases** and on the other hand the **stereotyped phrases/fixed**”(2006: 30).
- Adjective in the position of the attribute:
ο καφές είναι δυνατός/coffee is strong.

Multi-word expressions in Modern Greek

Verbal expressions - collocations 4

Thomou (2002: 253-255): lexical collocations vs fixed expressions.

- Criteria:
- a) expansion between two elements:
γερά, ατσάλινα νεύρα/strong, steel nerves,
- b) conjunction with other adjectives:
δυνατός και μυρωδάτος καφές/strong and aromatic coffee,
- c) gradation from quantitative adverbs:
πολύ δυνατός καφές/very strong coffee,
- d) adjective in the position of the attribute:
ο καφές είναι δυνατός/coffee is strong.

- => *δυνατός καφές/strong coffee* **lexical collocation**

Multi-word expressions in Modern Greek

Verbal expressions – light/support verb constructions 1

A series of studies investigate the syntactic and lexical organisation of light/support verb constructions. Although these structures are a separate domain in literature (M. Gross 1981, Grimshaw & Mester 1988, Radford 1997), they belong to the space between fixed and free structures, considering their degree of fixedness and their syntactic inflexibility.

It has been argued that the semantic core is the noun and that the verb bears a grammatical characterisation and supports the noun. Depending on which axis of analysis one focuses, they are called *semi-fixed expressions*, *collocations* or *Vsup/light verb* constructions. There is also a gradation in these structures.

Multi-word expressions in Modern Greek

Verbal expressions – light/support verb constructions 2

For the Greek language it has been found that, for example, the predicative nouns of emotion (Pantazara et al. 2008, Fotopoulou et al. 2009, Valetopoulos 2007) present constraints that would place them under the category of fixed expressions, at least for NLP applications, where the notion of fixedness is broader

- *Κοκκίνισα* από (ντροπή +* ανία + * φόβο)
'to be ashamed of'
- *Πλέω σε πελάγη* (ευτυχίας + *αισιοδοξίας + *χαράς + *έρωτα)
'I'm sailing on seas of (happiness + optimism + joy + love)'
- **Πλέω* (στην ευτυχία + στην αισιοδοξία + στη χαρά + στον έρωτα)
'I'm sailing on (happiness + optimism + joy + love)'

Multi-word expressions in Modern Greek

Verbal expressions – light/support verb constructions 3

Others studies

- *Δίνω* 'give' first approach (Tsolaki, 1997)
- First approach of some structures of the verbs ***έχω/to have, κάνω/to do, είμαι/to be*** and **their variants**, which incur to the initial structure some modifications related to the aspect of the sentence: *έχω κουράγιο-παίρνω κουράγιο-χάνω το κουράγιο* (Fotopoulou, 1985).
- With support verb *έχω/have* and its equivalents (Pantazara (2005)).
- Light verbs (support verbs) + sentiment predicative nouns -> 300 nouns with distributional and transformational properties (Fotopoulou et al. 2009, Pantazara et al. 2008).

Multi-word expressions in Modern Greek

Verbal expressions – light/support verb constructions⁴

- Verbes supports et intensité en grec moderne (Gavrilidou 2004)
- 1000 predicative nouns combined with the light verb *κάνω*/to do' with distributional and transformational properties (Sfetsiou 2007, T. Kyriacopoulou, V. Sfetsiou 2003, 2009)

Multi-word expressions in Modern Greek

Verbal expressions –

Psycholinguistics and Linguistics research 1

- ❑ The relationship between the expressions and their meaning is not abstract, provided that native speakers give a meaning depending on their intuition.
- ❑ Flores d'Arcais (1993) proposed the **point of uniqueness**: the time when the auditor recognizes that it's a multi-word expression.
- ❑ In the expression *έδεσε το γάιδαρό του*, the point of uniqueness that most people chose was the word *γάιδαρος* (Anastassiadis-Symeonidis & Voga 2011).

Multi-word expressions in Modern Greek

Verbal expressions – Psycholinguistics and Linguistics research 2

- The comprehension of a multi-word unit does not require a previous processing of its literal meaning and then its idiomatic meaning. That is the reason why the processing is carried out faster than in its literal meaning (Anastassiadis-Symeonidis & Voga 2011: 29).
 - The frequency of use and the productivity of its constituents contribute to the recognition of multi-word units. Mostly the extent of the word family where its main constituents belong.
- => Psycholinguistic experiments are carried out between almost equally frequent units.

Multi-word expressions in Modern Greek

Verbal expressions – Psycholinguistics and Linguistics research 3

- ❑ The study presented by Mini 2009 and Mini, Diakogiorgi and Fotopoulou 2011 and Diakogiorgi & Fotopoulou 2012 herein contains two, intricately linked but distinct studies: a linguistic study and a psycholinguistic study.
- ❑ The linguistic study aims at investigating the extent to which idiomatic phrases are fixed, through a thorough linguistic analysis of 470 phrases with fixed subject in Greek.

Multi-word expressions in Modern Greek

Verbal expressions - Psycholinguistics and Linguistics research 4

- A linguistic model of fixedness is developed which, being the product of the aforementioned analysis, will be presented immediately afterwards.

-> *spectrum of fixedness*

- **typical phrases:** *τα φόρτωσα στον κόκορα* [literally 'I loaded them on the rooster'] 'I did not act at all, as I was feeling lazy'
- **quasi phrases:** *ραγίζει η καρδιά μου* [literally 'my heart cracks'] 'I am in deep grief' or 'it broke my heart'
- **conventionalised phrases:** *τον τρώει το σαράκι της ζήλειας* [literally 'The woodworm of the jealousy eats him']

Multi-word expressions in Modern Greek

Verbal expressions- Psycholinguistics and Linguistics research 5

The aim of the psycholinguistic study, on the other hand, is precisely to assess the empirical adequacy of this model in Greek elementary school students, aged between 7 years and 6 months (7.6) and 9 years and 5 months (9.5). A series of experiments was conducted.

Multi-word expressions –Identification and Extraction for NLP applications

Several Approaches : Statistical, Linguistic and Hybrid (Statistical-Linguistic Models)

Statistical:

- ❑ - Log-likelihood score (Dunning 1993)
- ❑ - Mutual Information (Pearce 2002)
- ❑ - Latent semantic analysis (Katz and Giesbrecht 2006)
- ❑ - Permutation entropy (Zhang et al. 2006)
- ❑ - Word alignment (Medeiros et al. 2010)
- ❑ - Similarity measures (Gurrutxaga and Alegria 2012)

Linguistic knowledge:

- ❑ - Regular expressions (Leech et al. 1994)
- ❑ - Local Grammars (Breidt et al. 1996)
- ❑ - Semantic Tagging (Piao et al. 2003)
- ❑ - Syntactic rules (Conrado et al. 2011)

Hybrid:

- ❑ - POS taggers and shallow parsers to first identify MWEs, then feed the predictions into a statistical classifier (Baldwin 2005)
- ❑ - Perform first MWEs probabilistic identification, then use local grammars that includes MWE identification via specialized annotation schemes (Constant et al. 2013).

Multi-word expressions in Modern Greek NLP applications- recognition and extraction¹

Automatic recognition and extraction of multi-word nominal expressions from corpora” (Fotopoulou, Giannopoulos, Zourari, Mini 2008).

- A first approach to the development of an algorithm that would automatically detect multi-word expression (MWE). The algorithm that was developed for the present project is based in a combination of automatic MWE extraction methods (Sag et al. 2002). Combination of two methods:
- Word-based, knowledge-driven extraction: lexical sequences of a predetermined type are extracted (i.e., nominal compounds)
- Statistical extraction based on words: extraction of statistically idiosyncratic lexical sequences.

Multi-word expressions in Modern Greek

NLP applications – recognition & extraction²

Using the World Wide Web to identify multiword expressions (in Greek and Spanish) (Post doc Linardaki 2009).

Towards the Construction of Language Resources for Greek Multi-word Expressions: Extraction and Evaluation (Linardaki, Ramisch, Villavicencio and Fotopoulou 2010).

The possibility of automatic recognition of multi-word expressions. Various statistical correlation measures are used (Mutual Information, χ^2 etc.) in order to localize possible compounds. Verification and evaluation is achieved via research on the web.

Multi-word expressions in Modern Greek

NLP applications – recognition & extraction³

- Extraction of a list of candidate MWEs (mainly light verb construction and multi-word prepositions) from a corpus of sentences analyzed as dependency trees;
- Manual classification of candidate MWEs;
- Mapping of different types of MWEs in the train and test sets of a Greek dependency treebank and reporting on effect on parsing accuracy (Prokopidis et al. 2015).

Multi-word expressions in Modern Greek

NLP applications – named entities 1

Tool for the Recognition of Named Entities which is based on grammatical information and the use of rules.

The development of this tool was based on empirical data (Annotated Textual Corpus). Based on these data we have extracted Greek Named Entities.

Multi-word expressions in Modern Greek

NLP applications – named entities 2

Consequently, extension of the annotating schema, use of quantitative methods for the recognition of named entities -> new tool for the recognition of named entities which will be based on rules but it will also optimize statistical information concerning the co-occurrence of words in certain context. During our research, we developed a more detailed annotation tool for named entities and also a new corpus.

Publications

- Γιούλη et al. 2001. "οικΟΝΟΜία": Ένα Σώμα Σχολιασμένων Οικονομικών Κειμένων.
- Boutsis et al. 2000. A system for Recognition of Named Entities.
- Demiros et al. 2000. Named Entity Recognition in Greek Texts.
- Giouli et al. 2000. Multi-domain Multi-lingual Named Entity Recognition: Revisiting & Grounding the resources issue.

Multi-word expressions in Modern Greek NLP applications – conceptual lexicon

Encoding MWEs in a conceptual lexicon of Modern Greek
(Fotopoulou, Markantonatou and Giouli, 2014)

Work in progress aimed at the development of a conceptual lexicon of Modern Greek (MG) and the encoding of MWEs in it. Morphosyntactic and semantic properties of these expressions were specified formally and encoded in the lexicon. The resulting resource will be applicable for a number of NLP applications.

Multi-word expressions in Modern Greek

NLP applications – electronic morphological dictionary

- ❑ Electronic morphological dictionary for Modern Greek (DelaGr), Translation and Natural Language Processing Laboratory, Translation Department, School of French, Aristotle University of Thessaloniki, following the DELA formalism (Courtois 1990, Courtois & Silberztein 1990).
- ❑ It consists of: the dictionary of simple lexical units DELAS (one word) and the dictionary of compound lexical units (multi-word) DELAC, which are accompanied by grammatical, semantic and morphological information.

Multi-word expressions in Modern Greek

NLP applications – DELAC 1

It is divided to: **dictionary of canonical forms** of simple and compound words (DELAS and DELAC respectively) and the dictionary of inflected forms of simple (DELAF) and compound (DELACF) words. Part of the dictionary (approximately 30%) has been incorporated in the Greek version of the corpus processor Unitex which is accessible on the website <http://www-igm.univ-mlv.fr/~unitex/>.

- **Delac**
- *κινηματογραφικός(κινηματογραφικός.A1:ms)
αστέρας(αστέρας.N11:ms),NC_AN+Hum*
- *εξοχικό(εξοχικός.A1:Ans)
σπίτι(σπίτι.N35:Nns),NC_AN+Conc+[Lieu]*

Multi-word expressions in Modern Greek

NLP applications – DELAC 2

- **Delacf**
- *κινηματογραφικός αστέρας, κινηματογραφικός αστέρας. N+Hum:Nms*
- *κινηματογραφικοί αστέρες, κινηματογραφικός αστέρας. N+Hum:Nmp*
- *κινηματογραφικού αστέρα, κινηματογραφικός αστέρας. N+Hum:Gms*
- *κινηματογραφικού αστέρος, κινηματογραφικός αστέρας. N+Hum:Gms*
- *κινηματογραφικών αστέρων, κινηματογραφικός αστέρας. N+Hum:Gmp*
- -----

Multi-word expressions in Modern Greek

NLP applications – DELAC 3

Ambiguity

- In the inflection of multi-word compound nouns a linguistic form can often correspond to more than one cases in singular and plural.
- Syncretism (Spyropoulos & Kakarikos 2008) mostly remarked at the feminine and neutral genders.
- *αγροτική επιχείρηση, αγροτική επιχείρηση. N:Nfs*
- *αγροτική επιχείρηση, αγροτική επιχείρηση. N:Afs*
- *αγροτική επιχείρηση, αγροτική επιχείρηση. N:Vfs*

Multi-word expressions in Modern Greek

Perspectives –Tasks

- Unification of terminology
- Degree of fixedness
- Selection of some common features for NLP applications
- Applications on big data with grammatical and statistical methods
- Creation of new electronic dictionaries: every multi-word expression will be accompanied by features coming from varied theoretical approaches.