Emotion verbs in Greek. From Lexicon-Grammar tables to multi-purpose syntactic and semantic lexica

Voula Giouli & Aggeliki Fotopoulou

Keywords: *emotion verbs, Lexicon-Grammar tables, syntactic structure, distributional properties, semantic classification*.

Abstract

We hereby present work aimed at giving an account of Greek verbs denoting emotion that is placed within a larger context, aimed towards defining and describing the semantic field of emotions by means of identifying, selecting, classifying and organizing a *core* lexicon of emotions in a conceptual Data Base. The ultimate goal is the exhaustive description of Modern Greek and the development of a wide-coverage lexical resource that will be appropriate for a range of Natural Language Processing Applications.

1. Introduction

This paper describes work in progress aimed at a formal description of the Greek verbs denoting emotion on the basis of formal syntactic and semantic criteria. A core lexicon of emotion verbs extracted from existing lexical resources and corpora has been manually updated and extended where appropriate in order to reflect contemporary lexis and usage. Additionally, linguistic and semantic criteria were employed in order to define the semantic class of emotion verbs formally. The present study focuses only on verbs directly referring to emotional states (e.g. *fear*, *love*) as opposed to those having an indirect reference that depends on context (e.g. words that indicate possible emotional causes (e.g. *monster*) or emotional responses (e.g., l*augh*, *cry*, *etc.)*). The so-identified verbs were then studied and classified on the basis of empirical evidence about their syntactic and semantic properties. The ultimate goal being to develop a sound lexical resource (LR) that not only depicts sub-categorization information and semantic properties of the verbs, but also comprises a conceptual system that organizes the semantic field of emotions (being, in this sense, an ontology of emotions), the encoding of emotion verbs in a conceptual Data Base (DB) conformant to the underlying theoretical framework was performed.

2. The Methodological Framework

The theoretical framework adopted within this study is that of Lexicon-Grammar (Gross 1975). Being a model of syntax limited to the elementary sentences of a natural language, the theory argues that the unit of meaning is not located at the level of the word, but at the level of elementary sentences of the form Subject – Verb – Object. This representation is meaningful not only for verbs, but also for nouns and adjectives. To this end, the elementary sentence is being transformed to its predicate-argument structure, and the main complements (subject, object) are separated from other complements (adjuncts) on the basis of formal criteria. Moreover, adverbial complements that are considered as crucial arguments in that they characterize certain verbs are also taken into account:

(1) John removed the cups *from the table*.

To cater for a more fine-grained classification, and the creation of homogenous word classes, this formal syntactic definition is further coupled with distributional properties associated with words, i.e., types of prepositions, features attached to nouns in subject and complement positions, etc. A set of transformation rules, construed as equivalence relations between sentences, further generate equivalent structures. All this information (syntactic structure, distributional properties and permitted transformational rules) is formally encoded in the so-called Lexicon-Grammar tables.

Being initially developed to serve as a means of linguistic description, this framework has, never-the-less, been proved to be applicable for the construction of robust computational lexica. And although it has been claimed (Mathieu, 2008) that the information is not *directly* exploitable for NLP applications due to the fact that certain pieces of information are not formally encoded or are implicit, a number of works (Hathout et al 1998, Danlos et al 2009) have successfully managed to reformat Lexicon-Grammar tables in efficient large-scale NLP lexica. It becomes evident, therefore, that the Lexicon-Grammar framework together with the requirement of substantial coverage leads to a *uniform and consistent description of elementary sentences* and the *formal* encoding of the morpho-syntactic and semantic properties of the predicates *across languages in a comparable manner*. In this respect, one of the main advantages of Lexicon-Grammar is that it not only allows comparisons between languages both at the lexical and the syntactic levels, but also facilitates the construction of cross-language resources. In this respect, segments of Lexicon-Grammars have already been built for many languages including Greek, and this work may be considered as an attempt to contribute to the description of Modern Greek also in the prospect of the automatic analysis of the language.

3. Defining a semantic class: a core lexicon of emotions

Verbs denoting emotion, also referred to in the literature as "verbs of psychological state" (Levin 1993) and more commonly known as psych-verbs, have been extensively treated in syntactic theory. Early studies (Lakoff 1970), (Postal 1971) have tried to give transformational accounts of the behavior of psych-verbs, whereas (Belletti et al 1988) attempts an analysis of certain properties exhibited by psych-verbs in Italian on the basis of their thematic information and syntactic configurations. Other theoretical studies have been engaged with the uniform and consistent description of the lexicon of emotions mainly for languages like English and French (Anscombre 1995), (Ruwet 1995), on the basis of their syntactic properties, lexical choices and semantic criteria. In Jackendoff (1990), 3 basic semantic classes are defined, which are positioned in the continuum defined as positive (e.g., joy, happiness, etc.), - neutral – negative (e.g., sorrow, fear, terror, etc.). On the other hand, Mathieu (2000) proposes the classification of French verbs denoting emotion on the basis of their syntactic and semantic properties. In a comparative study, (Mathieu et al, 2010), attempt to reveal properties shared among English and French emotion verbs on the grounds of syntax and semantics. At the same time, the idiosyncrasies of each language are also unveiled.

As far as Greek is concerned, Greek verbs (Antoniou, 1984), (Valetopoulos 2005) nouns (Gavrilidou 2002) and adjectives (Valetopoulos 2005), denoting emotion have been treated within the Lexicon-Grammar framework. The study hereby proposed, however, will be based on previous work by our research group at the Institute for Language and Speech Processing, that gives an account of Greek nouns denoting emotion (Pantazara et al. 2008) (Fotopoulou et al., 2009) from a Lexicon-Grammar perspective. This previous work consists in the classification of nouns denoting emotion based on their syntactical and semantic properties. A grammar was elaborated the focus being on the combinatorics of nouns denoting emotion (85

items) with specific verbs expressing diverse modalities (aspect, intensity, control, manifestation or verbal expression). The long-term goal is the construction of an integrated LR that encompasses nouns, verbs, adjectives and adverbs.

4. Greek Verbs denoting emotion: selection and classification

After defining and consolidating the *core lexicon of emotion verbs*, their classification was performed on the basis of their syntactic, distributional and semantic properties as attested in corpus data. For this reason, a subset of the Hellenic National Corpus (HNC), a large POS-tagged and lemmatized reference corpus of the Greek language was employed. To this end, all sentences that contain the verbs under investigation were automatically extracted on the basis of lemma information. Further manual validation of the so-extracted data resulted in the removal of sentences in which the meaning of the verbs at hand was not an emotional one. This corpus was further enriched with a collection of documents originating from various sources over the internet, such as blogs, forums, life-style websites, and from other sources of opinionated texts. The latter data were considered appropriate for this study, since they depict everyday language usage and they reflect some emotional involvement either of the writer or of the discourse participants. The so-collected corpus that currently amounts to c. 300K words also bears annotations at various levels of linguistic analysis: POS-tagging and lemmatization, chunking and dependency parsing. Annotations have been added semi-automatically by means of a Natural Processing pipeline that was developed in-house for the Greek language. Extensive searches in the corpus were performed to verify the usage of the selected verbs.

4.1. *Verb selection and initial encoding*

Initially, Greek verbs denoting emotion were extracted from *Antilexicon* (Vostantzoglou 1962), a concept-based Lexicon for Modern Greek that is in accordance with the lexicographic tradition hallmarked by Roget's thesaurus. More precisely, 1168 verbs and multi-word expressions were identified as lexicalizations of 57 concepts systematized under the semantic class of emotion. This material underwent extensive corrections and modifications catering thus for the exclusion of *lemmas* or *usages* that are *obsolete*. Similarly, multi-word entries were also left out.

At the next phase, a closer investigation over the extracted data, however, revealed a number of entries that are not clear instances of emotion verbs, i.e., verbs directly referring to emotional states of the type *αγαπώ* (=*love*), *φοβάμαι* (=fear), etc. A number of verbs were identified as having only an indirect reference that depends on the context (e.g. words that indicate possible emotional causes as for example *βασίζομαι* (=rely on) or emotional responses (e.g. words that indicate possible emotional responses as *κλαίω* (=cry), *γελάω* (=lough), *δακρύζω* (=shed tears), etc.). To limit, therefore, the scope of the study – at least at this stage – to clear instances of emotion verbs only, a formal distinction between *direct affective words* and *indirect affective words* was in order. To this end, a set of *appropriate lexical semantic tests* were employed as a formal device guiding the final selection of a *core vocabulary of emotions* that covers the grammatical category of verbs, and the delineation of the semantic class of emotions, leaving all other cases of semantically and conceptually related verbs for future treatment. As a result, only verbs denoting *emotion*, or *emotional state*, as for example *αγαπώ* (=love), *αγανακτώ* (=resent), *ντροπιάζω* (=embarrass), *μισώ* (=hate), etc., were assigned a label *EMOTION* and were selected for further treatment. Building on similar lexicographic efforts adopted by the NLP community (Strapparava et al. 2004), the

remaining entries were assigned a label denoting the relation that links them with a specific emotion. The following labels have been used: *EMOTION-ELICITINGSITUATION*, *EMOTIONALRESPONSE*, *BEHAVIOUR*, *ATTITUDE AND SENSATION*. These entries were left aside for future treatment.

Consequently, the so-produced verb list was further enriched with new entries that were extracted from existing standard lexicographic resources for Greek via synonyms identification. Additionally, specialized electronic resources available for the English language were also consulted, namely FrameNet (Fillmore 2001) and WordNet (Fellbaum 1998), and Greek data were supplemented by obtaining translations of the English entries. More precisely, verbs occurring in frames involving an argument of the type *Experiencer* (that is, a participant characterized as aware of or experiencing something) were considered. Similarly, specific WordNet synsets were also consulted. Levin's (1993) seminal work on verb classification was also employed. The afore-mentioned process resulted in a list of 339 verbs denoting emotion that were finally selected for further study.

4.2. *Syntactic Classification of Greek emotion verbs*

The syntactic classification of the verbs was performed on the basis of the following axes: (i) syntactic information (subcategorisation); (ii) selectional restrictions imposed over their Subject and Object complements; and (iii) transformation rules. Elementary sentences of the form subject-verb-essential complements were considered as units for syntactic as well as for semantic composition of emotion verbs. *A*long these lines, emotion verbs in Greek were identified to appear in both transitive and intransitive constructions being represented as (a) N_0 *V*, (b) N_0 *V* N_1 , (c) N_0 *V* Prep N_1. According to the Lexicon-Grammar notation N_0 denotes a Noun in Subject position of a given verb *V*, whereas, N_1 denotes its Object. On the basis of the above-mentioned criteria, four major classes of Greek emotion verbs have been identified which we recall here briefly the definitional constructs:

1st class: N0 (hum) V (E + Prep N1)*: αγανακτώ* (=resent). This category encompasses verbs like *αγαλλιάζω* (=rejoice), *δυσανασχετώ* (=resent), *ντρέπομαι* (=be ashamed), *φρίττω* (=shudder) with N_0 in Subject position being obligatorily [+human]. Examples:

(2) a. *Οι* πολίτες/**Τα βιβλία αγανακτούν*.
The citizens /*The books *resent*.

b. *Οι* πολίτες *αγανακτούν με* τα μέτρα λιτότητας.
The citizens *resent (with)* the austerity measures.
The citizens *resent* the austerity measures.

(3) *Η* τοπική κοινωνία *έφριξε* με το στυγερό έγκλημα.
The local community *shuddered* with the heinous crime

2nd class: N0 (hum) V (E+N1): αγαπώ (=love). This class encompasses verbs with a N_0 in Subject position that is obligatorily [+human], whereas N_1 is non-restrictive, in the sense that it can be either [+human] or [-human]. Examples:

(4) Ο Γιάννης *αγαπάει* τη Μαρία/τα βιβλία/τη μουσική.[1]
John *loves* Mary / books / music.

(5) *Η μουσική/*Το βιβλίο *αγαπάει* τον Γιάννη.
*The music /*The book *loves* John.

One basic defining property shared among verbs falling in this class lies in the fact that depending on semantic context, their argument structure may be on occasions reduced by one argument, omitting thus the non-restrictive Object. For example:

(6) a. Γιάννης *αγαπάει* τη Μαρία. (transitive)
John *loves* Mary.

b. Δεν γεννήθηκα για να μισώ αλλά για ν' αγαπώ. (intransitive)
I was not born to hate (I was born) to love.

A second property of this type of psych-verbs is that some verbs in this class possess a passive morphology. Depending on context, these verbs allow for passive transformations:

(7) a. Αντίπαλοι και οπαδοί, όλοι αγάπησαν τον Χούλιο. [+active]
Opponents and supporters, all loved Julio.

b. Ο Χούλιο αγαπήθηκε από αντιπάλους και οπαδούς. [-active]
Julio was loved by opponents and supporters.

3rd class: N0 V N1 (hum) ⇔N1 (hum) V(E+PrepN1) : ανησυχώ(=worry). The third class comprises verbs that establish a homogeneous group with specific morpho-syntactic properties, as they have only [+active] morphology (*νευριάζω* (=pish), *ανησυχώ* (=worry), *ηρεμώ* (=calm), etc.) while lacking the [-active] one. These verbs occur in two syntactically distinct, yet semantically equivalent structures, allowing, thus, for inchoative alternation. In the transitive construction, they present an argument structure with N_0 in Subject position and N_1 [+human] in Object position. In the intransitive construction, N_0 in Subject position is [+human]. Examples:

(8) a. Ο Γιάννης/ο θόρυβος *νευρίασε* τη Μαρία /*το σπίτι.[2]
John / The noise *make* Maria / * the house *angry*.

b. Η Μαρία/*το σπίτι *νευρίασε* με τον Γιάννη/με το θόρυβο.
Maria / * home *was angry* at John / the noise.

(9) a. Η Μαρία/άδικη κριτική *θύμωσε* τον Μάριο /*την ταινία.
Maria / the unfair criticism *made* Mario / * the movie *angry*.

b. Ο Μάριος/*η ταινία θύμωσε με τη Μαρία.
Marios / * the film *was angry* at Maria.

4th* class*: N0 V N1(hum) ⇔ : N1=hum Vmp (E + Prep N0): αγχώνω (=stress). This class comprises verbs with N_0 in Subject position and N_1 that is obligatorily [+human] in Object position. Verbs in this class take part in either the passive or in the middle construction. Examples:

(10) a. Η Μαρία/οι εξετάσεις *αγχώνουν* τον Γιάννη/*το παιχνίδι.
Maria / The exams *stress* John / * the game.

b. Ο Γιάννης *αγχώνεται* (με τις εξετάσεις)
John *is stress.*

As a final remark, classes (1) and (2) project the Experiencer of the emotion as their structural subject and the Theme or Stimulus as their structural object. Similarly, classes (3) and (4) realize the Theme/Stimulus as the subject and the Experiencer as their object, their distinguishing property being their participation in unaccusative and middle constructions, the latter being linked to the implicit presence of an Agent (middle) and the absence of an Agent (unaccusative).

4.3. *Semantic classification of emotion verbs*

To render the intended resource useful for NLP applications, the classification of verbs in terms of their lexical semantic properties was in order. We aimed at defining a classification scheme that will be applicable for the effective cross-language description of the semantic field of emotions. The approach taken started from a rather coarse classification that gradually takes more fine-grained attributes into account. Following previous works (Mathieu et al. 2010), our scheme initially caters for the *polarity* of the emotion denoted. Polarity subsumes one of the following values: (a) *positive, i.e.* verbs which express a pleasant feeling; (b) *negative,* i.e., verbs which express an unpleasant feeling; and neutral, i.e., verbs expressing a feeling the polarity of which is context-dependent (e.g., surprise).

Moreover, the initial 57 concepts initially identified from the Greek concept-based dictionary, were mapped onto a more concise or coarse schema that subsumes concepts that can be classified under a more generic one, and proper assignments to entries are applied thereof. This classification is centred on a set of 8 basic emotions (Plutchik 1991): *anger*, *fear*, *sadness*, *disgust*, *surprise*, *anticipation*, *acceptance*, and *joy*. Verbs were further assigned feature *strength* with possible values: *low*, *medium,* and *high*, catering, thus, for the distinction between near synonyms such as *φοβάμαι (= I am scared), πανικοβάλλομαι (=panic), etc.*

5. Encoding in a Lexical Resource

The syntactic classification of Greek verbs denoting emotion has been encoded in Lexicon-Grammar tables that describe formally from a linguistic point of view their argument structure, distributional properties and possible transformations. Each class is represented by a table that includes all lexical items of the class. The properties that are involved in the definition of tables are of three types: syntactic, distributional and semantic. Verbs with more than one usages or meanings, have been treated as separate lexical items, and syntactic and semantic properties are assigned to each meaning thereof (as for example, the verb *αγαπώ* (=love) and *αγαπώ* (=like). A set of features that are appropriate for the description of a syntactic category concerning grammatical (i.e., past participle form) or syntactic information (i.e., passive transformation, inchoative alternation, etc.) and lexical choice (i.e., preposition a given verb is subcategorized for) is applied to all entries under consideration, and their linguistic validity was checked on the basis of corpus evidence. At the intersection of a row corresponding to a lexical item and a column corresponding to a feature, the cell is set to '+' if it is valid or '-' if is not.

The lexical database that is being developed also provides information that is commonplace to lexicographic projects, i.e., grammatical category, and spelling variants. As it has already been mentioned, during the verb selection process, verbs that pertain to the core lexicon of emotions were assigned a label *EMOTION*, and were kept in the database for further study. To ensure compatibility and interoperability with standardized tools the final resource will be conformant to widely-accepted standards (Lexical Markup Framework ISO (TC37/SC4)) and will be represented in XML format.

6. Conclusions

We have hereby presented ongoing work aimed at the formal description and classification of the Greek verbs denoting emotion. The analysis of the empirical data extracted from corpora aims at the development of LR will be applicable for a number of NLP applications ranging from Word sense Disambiguation to text understanding, Natural Language Generation, sentiment analysis, etc. Current and future work concentrates on extending the set of syntactic and semantic properties of the verbs at hand, and to integrate other grammatical categories (i.e., nouns, adjectives) and multi-word expressions.

Notes

[1] However, the verb *αγαπώ (=love)* with [-human] in object position is treated as a separate entry in dictionaries.
[2] Verbs in this category also enter the *genitive restructuring property (Guillet el al. 1981), Ο Γιάννης θύμωσε τη Μαρία [$_{prep}$ με [$_{n}$ τη συμπεριφορά του]] = Η συμπεριφορά του Γιάννη θύμωσε τη Μαρία (John made Maria angry with his behavior = John's behavior made Maria angry).*

References

Anscombre, J.-C. 1996. 'Noms de Sentiment, Noms d'Attitude et Noms d'Abstraits.' In N. Flaux, M. Glatigny and D. Samain (eds.), *Les noms abstraits, Actes du Colloque de Dunkerque (1992)*. Lille: Presses universitaires du Septentrion.

Antoniou, J. 1984. *Syntaxe et métaphore des verbes psychologiques en grec*. Doctorat de 3e cycle, Université Paris 7.

Danlos, L. and B. Sagot 2009. 'Constructions pronominales dans Dicovalence et le lexique-grammaire: Integration dans le Lefff.' In M. De Gioia (ed.) *Actes du 27e Colloque international sur le lexique et la grammaire*. Special issue of *Lingvisticæ Investigationes* 32.2.

Fellbaum, C. (ed.) 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fillmore, C., C. Wooters, and K. Baker 2001. 'Building a Large Lexical Databank Which Provides Deep Semantics.' *In Proceedings of the Pacific Asian Conference on Language, Information and Computation*. Hong Kong.

Fotopoulou, A., M. Mini, M. Pantazara and A. Moustaki 2009. 'La combinatoire lexicale des noms de sentiments en grec moderne.' In I. Novacova and A. Tutin (eds), *Le lexique des émotions*. Grenoble: ELLUG.

Gavrilidou, Z. 2002. 'Détermination des noms de sentiment en grec moderne.' *Langages* 145: 83–96.

Guillet, A. and C. Leclere 1981. 'La restructuration du sujet.' *Langages* 65.

Gross, M. 1975. *Méthodes en syntaxe*. Paris: Hermann.

Hathout, N. and F. Namer 1998. 'Automatic Construction and Validation of French Large Lexical Resources: Reuse of Verb Theoretical Linguistic Descriptions.' In *Proceedings of the 1st LREC Conference, Granada, Spain*.

Jackendoff, R. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.

Levin B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: The University of Chicago Press.

Lakoff, G. 1970. 'Global Rules.' *Language* 46.

Mathieu, Y. Y. 2000. *Les verbs de sentiments. De l'analyse linguistique au tratement automatique*. Paris: CNRS Editions.

Mathieu, Y. Y. and C. Fellbaum 2010. 'Verbs of Emotion in French and English.' *Emotion* 70.

Matthieu, C. and E. Tolone 2008. 'A generic tool to generate a lexicon for NLP from Lexicon-Grammar tables.' In *Actes du 27e Colloque international sur le lexique et la grammaire, L'Aquila, 10-13 Septembre 2008*.

Pang, B., L. Lee and S. Vaithyanathan 2002. 'Thumbs up? Sentiment Classification using Machine Learning Techniques.' In *Proceedings of EMNLP 2002*.

Pantazara, M., A. Fotopoulou, A. Moustaki and M. Mini 2008. 'La description des noms de sentiments du grec moderne.' *Lingvisticae Investigationes* 31.2.

Plutchik, R. 1991. *The Emotions*. New York, NY: University Press of America.

Postal, P. M. 1971. *Cross-Over Phenomena*. New York: Holt, Rinehart & Winston.

Ruwet N. 1995. 'Les verbes de sentiment forment-ils une classe distincte dans la grammaire?' In H. B. Shyldkrot and L. Kupferman (eds), *Tendances récentes en linguistique générale et française*. Amsterdam-Philadelphia: Benjamins.

Strapparava, C., and A. Valitutti 2004. 'WordNet-Affect: an affective extension of WordNet.' In *Proceedings of the 4th LREC Conference, Lisbon, May 2004*.

Turney, P. 2002. 'Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.' In *Proceedings of 2006 International Conference on Intelligent User Interfaces (IUI06)*.

Valetopoulos, F. 2005. 'Ce que vous pensez des autres: La grammaire locale de la jalousie et de l'admiration.' *Lidil* 32: 67–82.

Vostantzoglou, Th. 1962. *Antilexicon i Onomastikon tis Neas Ellinikis Glossis*. Athens.