Automatic Distinction of Arguments and Adjuncts for Modern Greek

1 Problem

Greek shows interesting asymmetries in the distribution of the PP/NP alternation.

Some examples for the PP/NP alternation in Greek:

Dative Alternation

1. *PP-Construction*: NP-nom V NP-acc ***se*** **(to) NP-acc**
2. *Genitive Construction*: NP-nom V NP-acc **NP-gen**

Benefactive Alternation

1. *PP-Construction*: NP-nom V NP-acc ***se*** **(to) /** ***gia*** **(for) NP-acc**
2. *Genitive Construction*: NP-nom V NP-acc **NP-gen**

Fortono (load) / *Alifo* (smear) Alternation

1. *PP-Construction*: NP-nom V NP-acc ***me*** **(with) NP-acc**
2. *Double Accusative Construction*: NP-nom V NP-acc **NP-acc**

Didasko (teach) Alternation

1. *PP-Construction*: NP-nom V NP-acc ***se*** **(to) NP-acc**
2. *Double Accusative Construction*: NP-nom V NP-acc **NP-acc**
3. *Genitive Construction*: NP-nom V NP-acc **NP-gen**

Drawing on this observation, I will try to combine existing computational methods (mixed symbolic/numerical models) with linguistic accounts of subcategorization properties in order to explore automatic ways to draw the distinction between arguments and adjuncts. The automatic distinction of arguments and adjuncts is a necessary step for the automatic acquisition of lexical knowledge, such as subcategorization frames and argument structure.

2 Motivating Examples

In this section, I describe with more details the dative and the benefactive alternation in Greek.

2.1 Dative Alternation

The term "dative alternation" is used in the literature for English and other languages to express the alternation regarding the categorial status of the indirect object, i.e. whether it is a PP or an NP. "Dative argument" refers to indirect objects (goals, beneficiaries, experiencers, possessors) regardless of case or categorial status (whether they are PPs or NPs).

The dative alternation is characterized in Greek by the prepositional frame and the double object frame (Tzartzanos 1989, Holton et al. 1997 among others) :

PP-Frame: NP-acc$_{<\text{theme}>}$ *se*[1]-PP$_{<\text{goal}>}$.

(1) O Petros edose ena milo s-ti Maria
The Peter-nom gave-3sg an apple-acc to-the Mary-acc
'Peter gave an apple to Mary'

Double Object Frame: NP-acc$_{<\text{theme}>}$ NP-gen$_{<\text{goal}>}$

(2) O Petros edose ena milo tis Marias
The Peter-nom gave-3sg an apple-acc the Mary-gen
'Peter gave Mary an apple'

Greek has lost the morphological distinction between genitive and dative case and has generalized the use of genitive. Verbs selecting for a single NP complement assign accusative case in Greek. Yet, there are certain verbs such as *milao* (talk), *aniko* (belong), *fenome* (seem) which assign genitive, but their complement can be an NP or a PP[2] :

(3) I Maria milise tu Petru
The Mary-nom talked-3sg the Peter-gen
'Mary talked to Peter'

(4) I Maria milise s-ton Petro
The Mary-nom talked-3sg to-the Peter-acc
'Mary talked to Peter'

[1] *Se* obligatorily incorporates an immediately following definite article and appears as a prefix *s-*. *Se* is also used as a locative (locational or directional) preposition.

[2] There are some verbs which take only genitive. With such verbs, the genitive cannot be replaced by a PP:

(i) Mu diafevgi to onoma tu
Me-cl,gen,fem/masc,sg escape-3sg the name-acc his-gen
'His name escapes me'

(ii) *Se emena diafevgi to onoma tu
To me-acc escape-3sg the name-acc his-gen

It is claimed in the literature (Markantonatou 1994, Bouba 1998, Anagnostopoulou 2001) that the alternation between a PP-dative and an NP-dative in Greek is similar to the "dative shift-alternation" in English. The genitive construction in Greek has two properties in common with the double object construction in English:

1. Sensitivity to animacy. The goal argument must be animate, i.e. it must be a recipient:

   (5) I Maria estile ena grama s-tin Italia
   The Mary-nom sent-3sg a letter-acc to-the Italy-acc
   'Mary sent a letter to Italy'

   (6) *I Maria estile ena grama tis Italias
   The Mary-nom sent-3sg a letter-acc the Italy-gen
   *'Mary sent Italy a letter'

2. Sensitivity to the semantic properties of the selecting verbal predicates. In particular, the central meaning is argued to involve transfer between a volitional agent and a willing recipient (Goldberg 1995). There are verbal classes in Greek which do not permit the genitive construction, like in English. Anagnostopoulou (2001) has examined more than 100 ditransitive verbs in Greek following Levin's (1993) classification of verbs in English. The investigation has shown that the prominent cases of verbs not allowing the dative alternation are verbs expressing "communication of propositions" (e.g. *parapempo* (refer), *paradehome* (admit))

   (7) Gia tin erotisi afti parepempsa ti Maria s-ton
   For the question-acc this-acc referred-1sg the Mary-acc to-the
   Petro
   Peter-acc
   'Regarding this question, I referred Mary to Peter'

   (8) *Gia tin erotisi afti parepempsa tu Petru ti
   For the question-acc this-acc referred-1sg the Peter-gen the
   Maria
   Mary-acc
   *'Regarding this question, I referred Mary Peter'

In Greek, unlike English, the indirect object in genitive cannot be nomitivized in the passive:

(9) Peter was given the book

(10) *O Petros dothike to biblio
The Peter-nom gave-3sg,passive the book-acc
'Peter was given the book'

In addition, Greek has clitic doubling of indirect (and direct object) NPs and in this respect it differs from English and other languages. In particular, when the indirect object is expressed as a genitive NP, it can be doubled by a pronominal clitic. The clitic and the NP match in features:

(11) I Maria tu edose tu
The Mary-nom,fem,sg him-cl,gen,masc,sg gave-3sg the
Petru to grama
Peter-gen,masc,sg the letter-acc,neutr,sg
'Mary gave Peter the letter'

Nevertheless, when the indirect object is realized as a PP, clitic doubling is not licit[3]:

(12) *I Maria tu edose s-ton
The Mary-nom,fem,sg him-cl,gen,masc,sg gave-3sg to-the
Petro to grama
Peter-acc,masc,sg the letter-acc,neutr,sg
'Mary gave the letter to Orestes'

Whenever the genitive construction is allowed, simple cliticization is possible too:

(13) I Maria tu edose ena
The Mary-nom,fem,sg him-cl,gen,masc,sg gave-1sg a
grama
letter-acc,neutr,sg
'Mary gave him a letter'

Whenever the genitive construction is not allowed, clitic doubling and cliticization are also infelicitous:

(14) *Gia tin erotisi afti (tis)
For the question-acc,fem,sg this-acc,fem,sg (her-cl,gen,fem,gen)
parepempsa (tis Marias) ton Petro
referred-1sg (the Mary-gen,fem,sg) the Peter-acc,fem,sg
*'Regarding this question, I referred her/Mary Peter'

Inanimate goals, which are generally not allowed to surface as genitive NPs, cannot be realized as clitic doubled genitives:

[3]In Greek, no preposition is possible in front of the doubled indirect object NP. The same can be observed with direct object clitic doubling.

(15) *O Petros (tis) estile (tis
The Peter-nom,masc,sg (her-cl,gen,fem,sg) sent-3sg (the
Italias) ena grama
Italy-gen,fem,sg) a letter-acc,neutr,sg
*'Peter sent her/Italy a letter'

Unlike genitive NPs, PPs are allowed in passive contexts:

(16) To grama dothike s-ton Petro
The letter-nom gave-3sg,passive to-the Peter-acc
'The letter was given to Peter'

(17) ?*To grama dothike tu Petru
The letter-nom gave-3sg,passive the Peter-gen
*'The letter was given Peter'

Clitic doubled and cliticized genitives are freely licensed in passives:

a. Clitic doubled genitive NP

(18) To grama tu dothike
The letter-nom,neutr,sg him-cl,gen,masc,sg gave-3sg,passive
tu Petrou htes
the Peter-gen,masc,sg yesterday
'The letter was given to Peter yesterday'

b. Simple clitic in genitive

(19) To grama tu dothike
The letter-nom,neutr,sg him-cl,gen,masc,sg gave-3sg,passive
htes
yesterday
'The letter was given to him yesterday'

Thus, in contexts in which full genitive NPs are legitimate (i.e. when the goal is animate and the verbal predicate indicates "change of possession"), cliticization and clitic doubling are optional. In contexts in which full genitive NPs are not allowed or they are ill-formed, cliticization or clitic doubling are obligatory.

2.2 Benefactive Alternation

The benefactive alternation resembles the dative alternation and is often subsumed under it. It is characterized in Greek by the prepositional frame and the genitive frame:

PP-Frame: NP-acc$_{<\text{theme}>}$ *se/gia*-PP$_{<\text{beneficiary}>}$

Genitive Frame: NP-acc$_{<\text{theme}>}$ NP-gen$_{<\text{beneficiary}>}$

Similarities to goals: Beneficiaries and goals surface in Greek as NPs in genitive or as PPs. Clitic doubled and cliticized genitives are freely allowed.

Difference from goals: According to Anagnostopoulou (2001), when the beneficiary is realized as a PP, it can be introduced by the preposition *gia* (for) or by the preposition *se* (to)[4]. Unlike beneficiaries, PP$_{<\text{goal}>}$ can be introduced only by the preposition *se*.

a. Genitive Construction

Genitive NP

(20) O Petros magirepse tis Marias mia supa
The Peter-nom cooked-3sg the Mary-gen a soup-acc
'Peter cooked Mary a soup'

Clitic doubled NP

(21) O Petros tis magirepse tis
The Peter-nom,masc,sg her-cl,gen,fem,sg cooked-3sg the
Marias mia supa
Mary-gen,fem,sg a soup-acc,fem,sg
'Peter cooked Mary a soup'

Simple clitic

(22) O Petros tis magirepse mia
The Peter-nom,masc,sg her-cl,gen,fem,sg cooked-3sg a
supa
soup-acc,fem,sg
'Peter cooked her a soup'

b. PP-Construction

se-PP

(23) ?O Petros magirepse mia supa s-ti Maria
The Peter-nom cooked-3sg a soup-acc to-the Mary-acc
'Peter cooked a soup for Mary'

gia-PP

[4]For some native speakers of Greek, the examples with the *se*-PP are not really natural.

(24) O Petros magirepse mia soupa gia ti Maria
The Peter-nom cooked-3sg a soup-acc for the Mary-acc
'Peter cooked a soup for Mary'

The benefactive alternation (genitive frame, *se*-PP frame and *gia*-PP frame) is found in Greek mostly with verbs of creation, such as *ftiahno* (make), *magirevo* (cook), and verbs of obtaining, such us *agorazo* (buy), *kalo* (call). There are verbal predicates, however, which allow only the *gia*-PP frame:

(25) I Maria danistike ena vivlio gia ton Petro
The Mary-nom borrowed-3sg a book-acc for the Peter-acc
'Mary borrowed a book for Peter'

(26) *I Maria danistike ena vivlio tu Petru
The Mary-nom borrowed-3sg a book-acc the Peter-gen
*'Mary borrowed Peter a book'

(27) *I Maria danistike ena vivlio s-ton Petro
The Mary-nom borrowed-3sg a book-acc to-the Peter-acc
*'Mary borrowed a book to Peter'

According to Anagnostopoulou (2001), the preposition *gia* can add a benefactive argument to all kinds of different predicates, while *se*-PPs and genitive constructions have a restricted distribution.

Moreover, Anagnostopoulou (2001) claims that *gia*-PPs are licit in passive constructions, while genitives are ungrammatical and *se*-PPs are ill-formed.

(28) I supa magireftike gia ti Maria
The soup-nom cooked-3sg,passive for the Mary-acc
'The soup was cooked for Mary'

(29) ?*I supa magireftike s-ti Maria
The soup-nom cooked-3sg,passive to-the Mary-acc
*'The soup was cooked to Mary'

(30) *I supa magireftike tis Marias
The soup-nom cooked-3sg,passive the Mary-gen
*'The soup was cooked Mary'

In contrast to (18) and (19), Anagnostopoulou (2001) observes that theme passivization in the presence of a genitive beneficiary does not become grammatical under clitic doubling and cliticization:

(31) *O kafes tis ftiahtike tis
The coffee-nom,masc,sg her-cl,gen,fem,sg made-3sg,passive the
Marias (apo ton Petro)
Mary-gen,fem,sg (by the Peter-acc,masc,sg)
*'The coffee was made Mary (by Peter)'

(32) *O kafes tis ftiahtike (apo ton
The coffee-nom,masc,sg her-cl,gen,fem,sg made-3sg,passive (by the
Petro)
Peter-acc.mac,sg)
*'The coffee was made her (by Peter)'

2.3 Questions

Is the dative and the benefactive alternation one and the same phenomenon?

a. Do genitive goals and genitive beneficiaries realize the same syntactic function (argument or adjunct)?

b. Do *se*-PP goals and *se*-PP beneficiaries realize the same syntactic function (argument or adjunct)?

2.4 Some Previous Accounts

Holton et al. (1997): According to Holton et al., the indirect object in genitive case represents an animate, often human entity which is indirectly affected by the action expressed by the verb. In the examples below, the direct object is the recipient and the beneficiary respectively[5]:

(33) O Yanis edose ena oreo brahioli tis Meris
The John-nom gave-3sg a beautiful-acc bracelet-acc the Mary-gen
'John gave Mary a beautiful bracelet'

(34) I Eleni magirepse tu Niku makaronia
The Helen-nom cooked-3sg the Nick-gen spaghetti
'Helen cooked spaghetti for Nick'

Holton et al., however, point out that from the above examples it is clear that the genitive case does not clearly mark the precise semantic relation of the indirect object to its verb. The precise semantic function of the indirect object is derived by considering the semantics of the verb and of the indirect

[5]Holton et al. observe that sources, like recipients and beneficiaries, surface as genitive NPs too:

(i) Hthes piran tis Elenis ena ekatomirio drahmes
Yesterday took-3pl the Helen-gen a million-acc drachmas-acc
'Yesterday they took a million drachmas from Helen'

object as well as the total linguistic and pragmatic context. When the goals and the beneficiaries are expressed by PPs, they merely consider them to be indirect objects.

Kleris & Babiniotis (1999): In their descriptive grammar for Modern Greek, Kleris and Babiniotis consider goals to be indirect objects of the verbal predicate, contrary to beneficiaries which are considered adjuncts.

Anagnostopoulou (2001): Anagnostopoulou states that constructions where the beneficiary is introduced by the preposition *se* behave, in a number of respects[6] , like double object constructions. However, the fact that neither clitic doubling nor simple cliticization of the genitive beneficiary are grammatical in a passive context (contrary to dative alternation) leads her to the conclusion that the syntax of benefactive constructions is more complicated than sometimes assumed.

Levin (1993): According to Levin, in English the benefactive alternation differs from the dative alternation in involving the benefactive preposition *for* rather than the goal preposition *to* in the prepositional variant. The benefactive alternation is found basically with verbs of creation or verbs of obtaining. Levin in her comments refers to studies of this alternation which show that the first object in a double object construction with a benefactive alternation verb is *less "object-like"* in some respects than the first object in a double object construction with a dative alternation verb.

---

[6]

a. The *se*-construction is only permitted with verbs of creation on a par with the genitive construction. Both types are disallowed in many other environments which still permit the *gia*-construction. While the *se*-PPs and the genitive construction are limited to verbs of creation, *gia*-PPs show no such restriction.

b. Passives do not license *se*-benefactive PPs similarly to genitives and unlike *gia*-PPs. In this respect, they behave as double object datives and they differ from *gia*-PPs.

c. The third property of *se*-benefactive PPs concerns their distribution in nominalizations based on roots denoting creation. Nominalizations related to the form with *se*-PP or the genitive construction are, unlike nominalizations related to the form with *gia*-PP, not felicitous:

(i) To magirema tu fagitu gia ti Maria
The cooking-nom/acc the food-gen for the Mary-acc
'The cooking of the food for Mary'

(ii) *To magirema tu fagitu tis Marias
The cooking-nom/acc the food-gen the Mary-gen

(iii) *To magirema tu fagitu sti Maria
The cooking-nom/acc the food-gen to the Mary-acc

This is the third piece of evidence for Anagnostopoulou that the *se*-PP is a double object dative rather than a true PP in benefactive constructions.

Conclusion: These accounts fail to provide clear-cut diagnotics, and, sometimes, definitions for the functional syntactic labels they use such as direct object, indirect object, ditransitive verb and the like. Furthermore (and perhaps as a consequense of the previous problem), they fail to draw a clear-cut line, if one exists, between the arguments of a verbal predicate and its adjuncts.

3 Task Definition - Future Work

I will investigate empirically the syntactic behavior of verbal predicates that show the PP/NP alternation in Greek with the help of data-intensive methods. My work will be based on data from large unrestricted corpora and on the application of statistical methods for NLP-tasks.

Step 1: The context-free grammar describing Greek sentence structure is being developed in the YAP formalism of Schmid (1999) and it will be compiled into a grammar in the formalism of Carroll & Rooth (1998). The grammar will be trained on a large Greek newspaper corpus with the probabilistic context-free parser by Carroll & Rooth (1998). The parser will utilize the Greek context-free grammar and a lexicalized probability model to produce parse forests[7] for sentences. Each sub-tree will be annotated with information about the lexical head and the probability. After the parsing of the corpus, the maximum probability parses will be extracted from the parse forests. The maximum probability parses will be used to determine the verb and its frame. For example, the frame token of the English sentence "Nobody excelled him in that judgement" would be defined as (Schulte im Walde 1998):

act*excelled subj*nobody obj*him pp*in*judgement

Step2: Based on the verb frame extraction, I will try to acquire verbs that show the PP/NP alternation. Observing all the possible contexts where the alternation occurs in the corpus (e.g. active context with PP, NP, clitic doubled NP, clitic; passive context etc.) promises to be interesting. The frame extraction could supply evidence concerning the linguistic diagnostics for argumenthood. The linguistic diagnostics that are used for Greek (especially for the cases of PP/NP alternation) to determine whether a complement is an argument or an adjunct are not accurate, and they give rise to relative, and not absolute, acceptability judgements. Therefore, the core difficulty concerning the discrimination between arguments and adjuncts is to define precisely enough the notion of argument and adjunct so that it can be used automatically.

Merlo & Leybold (2001) propose a methodology that retains both the ability to combine partial diagnostics, typical of automatic induction meth-

[7] A parse forest is a compact representation of all possible parse trees.

ods, and the linguistic insight of the diagnostic tests for the distinction between argument and adjunct for the case of PPs attached to the verb. Drawing on the linguistic diagnostics proposed in the literature (Schuetze 1995) for English, they choose four tests that:

a. capture the semantic difference between arguments and adjuncts[8]

b. can be easily expressed by simple statistical concepts (the diagnostics are estimated using corpus counts).

The diagnostic tests are:

Head dependence: Arguments depend on their lexical heads, while modifiers do not. Therefore, arguments can occur only with a specific verbal head by which they are lexically selected. On the other hand, modifiers can appear with a far greater range of different heads than arguments. Merlo & Leybold (2001) capture this insight by counting the number of different verbs that co-occur with a given PP in a corpus. A low number indicates argument status, while a high number indicates modifier status.

Optionality: In most cases, PP-arguments are obligatory elements whose absence leads to ungrammaticality, while modifiers are optional. Thus, the predictive power of a verb about its complements should be greater for arguments than for modifiers.

Iterativity and Ordering: Arguments cannot be iterated and must be adjacent to the lexical head. Therefore, in a sequence of several PPs only the first one can be an argument while the others must be modifiers.

Problems with Greek using the methodology of Merlo & Leybold (2001):

a. Unlike English, Greek has a relatively free word order. Arguments must not necessarily be adjacent to the selecting verbal head (this is not the case with nominal heads). Therefore, the measure of ordering is not reliable for Greek.

b. The corpora that Merlo & Leybold (2001) use correspond to the subsets of verb attachments extracted from two PP-attachment corpora. One

[8]According to Grimshaw (1981), a complement is an argument, if its interpretation depends exclusively on the head which is associated. The interpretation of adjuncts, on the other side, remains relatively constant when associating with different heads. Pollard & Sag (1987) expressed the notion of the semantic difference between arguments and adjuncts similarly. They stated that arguments are limited to co-occurence with a semantically restricted class of heads, while adjuncts can co-occur with a relatively broad range of heads.

corpus contains data encoding information for attachment of single PPs in the form of four head words (verb, object noun, preposition and noun inside the preposition) for each instance of PP attachments found in the corpus. They also use an auxiliary corpus of sequences of three PPs, where each data item consists of the two PPs following the head noun and verb and of the third preposition. For the case of the prepositional constructions (e.g. NP-nom V NP-acc *se* NP-acc) in Greek, I won't use restricted corpora like Merlo & Leybold (2001) did, since, as mentioned in (a.), the word order in Greek is not fixed.

In contrast to Merlo & Leybold (2001), I would like to use additional linguistic diagnostics that require more complex syntactic manipulation of the sentence, such as passivization, cliticization, nominalization, topicalization etc. (Markantonatou 1994, Bouba 1998, Anagnostopoulou 2001). It is challenging to try to capture this kind of diagnostics with statistical concepts. Moreover, based on the corpus frequencies, I will attempt to determine if alternating verbs have frame preferences.

I hope that the combination of frame induction for the alternating verbs and the corpus frequencies will supply valuable information in order to define the notion of argument and adjunct precisely enough so that it can be used automatically.

References

Anagnostopoulou, Elena (2001): The Syntax of Ditransitives. Evidence from Clitics. Ms., University of Crete; to appear in *Mouton de Gruyter*.

Bouba, Maria (1998): Diptota Rimata sta N. Elinika. In: *Proceedings of the 12th Annual Meeting of the English Department of the University of Thessaloniki*. Thessaloniki.

Carroll, Glenn & Mats Rooth (1998): Valence Induction with a Head-Lexicalized PCFG. In: *Proceedings of the 3rd Conference of Empirical Methods in Natural Language Processing*. Granada.

Goldberg, Adele (1995): *A Construction Grammar Approach to Argument Structure*. The University of Chicago Press, Chicago/London.

Grimshaw, Jane (1981): *Argument Structure*. MIT Press, Cambridge MA / London.

Holton, David, Peter Mackridge & Irene Philippaki-Warburton (1997): *Greek: A Comprehensive Grammar of the Modern Language*. Routledge, London/New York.

Kleris, Christos & Georgios Babiniotis (1999): *Gramatiki tis Neas Elinikis: Domoliturgiki-Epikinoniaki*. Elinika Gramata, Athens.

Levin, Beth (1993): *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press, Chicago/London.

Markantonatou, Stella (1994): Diptota Rimata: Mia Lexiko-Simasiologiki Prosegisi. In: *Proceedings of the 15th Annual Meeting of the Department of Linguistics of the University of Thessaloniki.* Thessaloniki, pp. 360–371.

Merlo, Paola & Matthias Leybold (2001): Automatic Distinction of Arguments and Modifiers: the Case of Prepositional Phrases. Ms., University of Geneva and American Management Systems; to appear in the Proceedings of CoNLL-2001.

Pollard, Carl & Ivan Sag (1987): *An Information-based Syntax and Semantics.* CSLI Lecture Notes, Stanford CA.

Schmid, Helmut (1999): YAP: Parsing and Disambiguation with Feature-Based Grammars. PhD thesis, University of Stuttgart, Stuttgart, Germany.

Schuetze, Carson (1995): 'PP Attachment and Argumenthood', *MIT Working Papers in Linguistics* **26**, 95–151.

Schulte im Walde, Sabine (1998): Inducing Lexicons with the EM Algorithm. Arbeitspapiere des Instituts fuer Maschinelle Sprachverarbeitung, Universitaet Stuttgart.

Tzartzanos, Achilleas (1989): *Neoeliniki Sintaxis (tis Kinis Dimotikis).* Kiriakidhes, Thessaloniki.