Analysing speech acts with the Corpus of Greek Texts
Implications for a theory of language

Dionysis Goutsos
Department of Linguistics
University of Athens
dgoutsos@phil.uoa.gr

Abstract

The paper aims at analyzing speech acts with the help of a Greek corpus and discussing the contribution of this analysis to a theory of language. The data comes from the Corpus of Greek Texts, a new reference corpus of Modern Greek. The analysis of an utterance found on a public sign suggests that illocutionary force may be indicated by the use of a lexicogrammatical pattern, achieving specific textual effects. The analysis of *forget* verb forms in Greek shows an uneven distribution, in which the marking of grammatical categories is related with preferred interpersonal and sequential functions. It is argued that the findings of corpus linguistic analysis have important implications for treating language as a form of historical praxis of a social and dialogical nature, in ways which concur with Voloshinov's ideas.

1. Introduction[1]

This paper has a double aim: first, to discuss certain linguistic expressions in Greek by using evidence from the Corpus of Greek Texts in order to show the relevance of corpus linguistics for the analysis of languages like Greek and, second, on the basis of this discussion, to reflect on the contribution of corpus linguistic analysis to a theory of language in the frame of current theoretical approaches. My starting point comes from the following two utterances in Greek, which are presented below with some indication of their linguistic context. The original examples are accompanied with phonetic transcription, a gloss and translation into English, as appropriate:[2]

(1)

| ΟΙ ΓΛΑΣΤΡΕΣ |
| ΔΕΝ ΕΙΝΑΙ ΤΑΣΑΚΙΑ |

οι	γλάστρες	δεν	είναι	τασάκια
i	γlástres	δen	íne	tasáca
the	pots	NEG	are	ashtrays

'plant pots are not ashtrays'

(2) 'Better facilities abroad'
    <K> but they also don't don't don't don't don't give any funding here for these regions to //have them better exploited
    <B> //yes
    <M> yes
    <K> that is what do they have more abroad, well they just a::re they've better exploited things ((they've got)) better roads
    <B> for sure
    <M> for sure
    <K> better facilities and so on
    <D> ντάξει όμως να σου πω κάτι <u>μην ξεχνάς</u> όμως ότι
    <D> dáksi ómos na su pó káti <u>min gseçnás</u> ómos óti
→   <D> //fine but, let me tell you something, <u>don't forget</u> though that
    <K> //it can be done here too
    <M> //THE BEAUTY THE BEAUTY IS THE BEAUTY that e::h is better ((inaudible)) that is I believe is at the same level as abroad
    <K> I believe that it is better here in us

The first utterance was found on a public sign placed in a plant pot located in the corridor of a major hospital in Athens. It is immediately clear, at least to the Greek speaker who is confronted with this sign, that there is a specific illocutionary force attached to the utterance, namely an injunction to passers-by not to put their cigarettes off in the plant pots. Speakers of Greek would also draw a further implication from this, i.e. that there indeed are some people who do follow the hideous practice of smoking in a hospital corridor.

The second utterance of interest to us here is the phrase *μην ξεχνάς ότι* 'don't forget that', uttered by speaker D in a longer turn in which he competes with two other speakers for the floor in a vivid discussion concerning a tourist place in the still 'undeveloped' Greek countryside. The three young male friends are arguing about how Greek tourist places compare to those 'abroad' and this is the point in the discussion where two of them (K and M) elaborate on a line of argument, which again would sound familiar to most Greek ears, namely that foreign tourist places are better because they have 'more facilities abroad', whereas Greek tourist places are more beautiful but less attractive because of the lack of facilities. At this point speaker D tries to chip in with several phrases, including *dáksi* 'OK, fine', *ómos* (twice) 'but, however', *na su pó káti* 'let me tell you something' and *μην ξεχνάς ότι* 'don't forget that'. This last one has the appearance of a reminder but seems to take on the force of a turn-taking device in the surroundings of similar phrases, which attempt to capture the attention of the listeners and thus grab the floor.

There are many things that could be said about these two fragments of language use, e.g. in a discourse analytic framework, in terms of conversation analysis, speech act theory, argumentation theory etc.[3] However, the question that centrally concerns me here is whether corpus linguistics has anything meaningful to say about these utterances. In my view, whereas discourse analytic approaches would suggest an interpretation of these texts (or text extracts) by bringing evidence from their expanded context, a corpus linguistic analysis has no other option but turning to the text and attempting a fresh reply to the question of the inveterate structuralist: is there something in the language of the utterances that indicates the speech act performed? Do we have any indications from the language 'itself' that this is the intended speech act in this context?

Discourse analytic and corpus linguistic approaches seem to work from opposite sides towards the same end of ascertaining how the interpretation of utterances comes about. In the case of utterances like (1) above, it must be noted, first of all, that, apart from the interpretation suggested, one could possibly construct alternative readings. For instance, the public sign could inform its readers about a fact of life or could be part of the instructions of a pot-maker to potential followers of the trade. The fact that these alternative possible meanings and their accompanying illocutionary forces are forcefully excluded in favour of the one suggested here obviously relates to aspects of the context. It is the immediate situational context that forces one interpretation rather than another, together with cultural knowledge about how hospitals operate in Greece etc. Pragmatic notions like Grice's maxims, the principle of relevance, procedures of inference drawing etc. may be relied upon in order to explain how such an interpretation is possible, as well as how other interpretations become less likely.

It is my contention in this paper that corpus linguistics can help us reverse this analytic perspective by focusing on the language of the text and the way in which it affects utterance interpretation in context, instead of drawing from the context in order to explicate the text. I will first attempt an interpretation of the two utterances presented above by precisely following this corpus linguistic approach. For this purpose, I will bring evidence from a corpus of Modern Greek, the Corpus of Greek Texts, the presentation of which will necessarily follow in the next section. I will then proceed to an analysis of the speech acts performed by the two utterances under investigation, by studying the patterns involved in their construction. Finally, I will draw

some conclusions in order to further discuss the implications of the approach followed for a theory of language (or 'the' theory of language, if one prefers).

2. The Corpus of Greek Texts

Although the need for a reference corpus of Modern Greek was brought up relatively early in the bibliography (Goutsos et al. 1994b), it was not before the beginning of the 2000s that such a corpus became available. This was the the ILSP Corpus, now developed as the Hellenic National Corpus (HNC),[4] which, however, is less than adequate for research purposes for several reasons. In particular, no spoken texts are included; more than 60% of the texts come from newspapers; the great majority of texts are not assigned to a specific text type in its classification scheme; most texts from books are extracts; and, no information is given on the structure of the corpus or the classification used.[5]

The Corpus of Greek Texts (CGT) has been developed as a new reference corpus for Modern Greek, especially designed as a resource for linguistic research and teaching applications (Goutsos 2003, forthoming, Goutsos and Pavlou forthcoming). CGT was initially developed as a common project of the University of Athens and the University of Cyprus and is now at its final phase of implementation at the University of Athens. Its aim has been to collect a substantial amount of data (30 million words) from a wide variety of text types, which are thought to be representative of basic genres and linguistic varieties in the language.

For this reason, CGT has aimed at the following main characteristics:
- to be a well-defined collection of texts from a variety of genres that are central in Greek contexts of communication and important for the teaching of Greek as a first/second language
- to include a substantial percentage of spoken data, constituting the biggest existing collection of spoken Greek
- to contain a substantial percentage of data from Cyprus, offering for the first time a valuable resource for the study of geographical variation in Greek
- to become the basis for larger (e.g. monitor) corpora of the future
- to be available to researchers and learners through user-friendly applications.

According to these characteristics, CGT can be characterized as a general or reference corpus, a monolingual corpus, comprising two major geographical varieties (Standard Modern Greek and Cyprus Greek), a mixed corpus, including both spoken and written material, and a synchronic corpus, collecting data from two decades (1990 to 2010). It must be noted that texts are stored in their entirety, where possible, and no translated texts are included (again, to the extent that such an exclusion is possible).

As regards the size of CGT, the aim of the project is to collect 30 million words in total. Although this would seem a rather small size for current standards, it should be placed in the context of existing projects in Greek. The case of HNC indicates that the major priority in Greek is not to increase the size of the corpus but enhance the range of text types covered in it, avoiding at the same time a biased collection of genres. The amount of 30 million words is projected to cover the needs of linguistic research for the next decade, with the view of expanding CGT into a monitor corpus of Greek, in which new material could be constantly added and old data would be removed.[6]

Tables 1 and 2 below present a rough outline of the CGT structure, with regard to the medium of texts and the basic text types included. The figures given correspond to the number of words currently included (June 2009).

Medium	Number of words	Percentage
Book	6,190,045	22.73 %
Newspaper	8,054,039	29.58 %
Magazine	5,999,059	22 %
Electronic	1,598,291	5.87 %
Live	2,150,674	7.9 %
Radio	105,121	0.38 %
TV	675,485	2.5 %
Other	2,451,061	9 %
Total	27,223,775	100 %

Table 1: CGT structure according to medium

Mode	Text type	Number of words	Percentage
Spoken	News	291,382	1 %
	Interview	592,584	2 %
	Public speech	1,839,766	6.75 %
	Conversation	207,548	0.76 %
Written	Literature	2,455,080	9 %
	News	4,764,337	17.5 %
	Opinion articles	3,189,132	11.7 %
	Information item	100,570	0.36 %
	Academic	3,994,277	14.67 %
	Popularized	7,648,513	28 %
	Law and administration	1,472,700	5.4 %
	Private	186,210	0.68 %
	Procedural	145,770	0.53 %
	Miscellanea	335,906	1.65 %

Table 2: Classification of CGT texts according to text type

Apart from medium, mode and text-type, referred to above, CGT texts are also classified according to class (spontaneous-planned for spoken texts, information-non-information for written texts), geographical variety (standard Greek-Cyprus Greek), text sub-type (e.g. academic texts are further divided into arts, social and economic, and science, literature into poetry, novel, short story, biography, anecdote etc.). All texts are accompanied by metadata, employed to classify them for these categories. Classification also allows for a varied composition of sub-corpora, according to research needs and priorities.

CGT has been used in the successive phases of its development for linguistic studies on a variety of aspects of Greek grammar and lexis, including discourse markers (Georgakopoulou & Goutsos 1998), place adverbials (Goutsos 2007), shell nouns (Koutsoulelou & Mikros 2004-2005), male and female lexical pairs (Goutsos & Fragaki forthcoming) etc, as well as in pedagogical applications (Goutsos et al. 1994a). It is also currently being used for PhD research in more extended studies of Greek adjectives (Fragaki 2009), lexical clusters (Ferlas 2008) and academic vocabulary (Katsalirou forthcoming).

At present, CGT is freely available through its website (www.sek.edu.gr), although data collection and upload will be finalized in 2010. Future plans include the evaluation of CGT compilation practices and outcomes, which will feed back into CGT's structure. It is expected that the development of the CGT will radically change the picture we have of Greek, providing evidence for a more comprehensive, accurate and authoritative description of the language.

3. Lexicogrammatical patterns and the speech act of correction

Let us return now to the first utterance discussed above, after this necessary detour. If there are to be any linguistic indications of the speech act performed by the phrase *οι γλάστρες δεν είναι τασάκια* 'plant pots are not ashtrays', these must surely lie neither in the particular lexical items ('plant pots', 'ashtrays') nor in the structure *per se*. (Note that this is a typical S-V-complement structure of the type 'Socrates is good'). They must rather be identified in the complex pattern of interaction between definiteness, plurality and negation, as indicated by searching the CGT.

Specifically, since the phrase under discussion comes from the public sphere, a search was made in the spoken sub-corpus of CGT, which lies closer to this domain and includes approximately 2.5 million words from interviews, (Greek and Cyprus) Parliament speeches and TV and radio broadcasts (see Table 2 above). Patterns involving definiteness alone, marked in the subject part of the phrase ('<u>the</u> NP is NP'), plurality alone, marked in both the subject and the complement part of the phrase ('the <u>NP-pl.</u> is <u>NP-pl.</u>') or negation alone, marked in the copula part of the phrase by a special negation particle ('the NP is <u>not</u> NP'), do not seem to be associated with a particular speech act related to prohibition. Instead, according to evidence from CGT arrived at by various searches performed, it is the combination of these three grammatical categories in a particular lexical pattern that seems to be significant. In other words, the linguistic clues for the associated speech act must be related to the following pattern:

(3)	i γlástres	δen	íne	tasáca
	the NP-pl.	NEG	to be	NP-pl.
	subject	negation	copula	complement

We can formulate this pattern in Hunston & Francis' (1999) terms as 'the pl-N *not* V-link N', with the provisos a) that this is a pattern of Greek grammar (at least, until a similar claim is made for English or other languages) and b) that it is associated with the particular lexis (the negative marker *δεν* and the specific V-link *είναι*, which is the 3rd person plural of the present tense of the verb 'to be').

A search for this pattern in the spoken sub-corpus of CGT yielded 20 results, which can be seen in Appendix 1. What is immediately apparent from these examples is that the pattern under investigation forms part of broader textual patterning, that is it co-occurs with other patterns in a particular order in text. An analysis of the concordance lines suggests that these patterns can fall into three categories. First, most of the examples (namely 1-3, 5, 7, 11, 13-14, 16, 19-20 in the concordance of Appendix 1) use the pattern 'the pl-N *not* V-link N' to deny an assertion and then correct this with a closely related pattern 'V-link N' as in the following:

(4) οι αγώνες με κλασσικά αυτοκίνητα <u>ΔΕΝ είναι</u> αγώνες ταΧΥτητας είναι ΠΟΛΙτιστικές εκδηλώσεις[7]
vintage car races are NOT racing games, they are cultural events

(5) αυτά τα θέματα <u>δεν είναι</u> θέματα θρησκείας και Θεού, αλλά είναι θέματα απλής απλοποίησης
these issues are not issues of religion and God, but just issues of simplification

(6) Αυτά <u>δεν είναι</u> έργα για να επαίρεται κανείς, είναι βαρβαρότητα για να εντρέπεται
These are not deeds to be proud of, they are barbarity to be ashamed of

It must be noted that the second pattern ('V-link N') used to make the correction lacks the subject part, which is implied. (Greek, as a morphology-rich language, can have sentences without an explicitly expressed subject; see Joseph & Philippaki-Warburton 1989: 36-37). The opposition between denial and correction is further signalled by intonation as in example (4), by the use of a discourse marker (*αλλά* 'but') in (5) or by strong syntactic and phonological parallelism (*για να επαίρεται-για να εντρέπεται*: ja na epérete-ja na endrépete) as in (6). There

are thus strong cohesive links between the two parts of the argument (denial-correction), which are based on repetition and modification: lexical items are repeated and modified in the second part, so that a matching relation is established between denial and correction.[8]

Secondly, in only two instances (6 and 8 in the concordance of Appendix 1) the opposite pattern is observed, that is the negated statement follows the positive one as in:

> (7) Είναι θέματα καθαρά εσωτερικά. <u>Δεν είναι</u> θέματα πολιτικής σημασίας. Δεν είναι πολιτικές διαφορές
> They are clearly internal matters. They <u>are not</u> matters of political importance. They are not political differences

The effect of this order of lexicogrammatical patterns is to make an assertion and then clarify it through a counter-statement. What is noticeable here is again the forceful use of parallelism, involving repetition and modification.

Finally, a third group of examples (4, 9-10, 12, 15, 17-18 in the concordance of Appendix 1) only shows the denial part and leaves the correction to be implied as in:

> (8) Θεωρώ ότι είναι δίκαιο -<u>δεν είναι</u> επαίτες οι δικηγόροι, πρέπει να στηριχθούν με αξιοπρέπεια-
> I believe it is fair – lawyers <u>are not</u> beggars, they have to be supported with dignity –

> (9) πιστεύω ότι <u>δεν είναι</u> μαγικά, ταχυδακτυλουργικά κόλπα τα θαύματα. Θα επαναλάβω τι είναι.
> […] I believe miracles <u>are not</u> magic, juggling tricks. I will restate what they are.

This is a much more subtle expression of the denial-correction patterning, based on the principle of preference organization, as has been pointed out in conversation analysis (Pomerantz 1978, cf. Georgakopoulou & Goutsos 2004: 80, 117-8). Since correction is the preferred second pair part, as predicted by the first pair part, the utterance following the negated statement is understood as correcting the denial made in the previous statement, although explicit signalling for such a correction is not present.

To sum up, the corpus investigation of the pattern 'the pl-N *not* V-link N' in Greek has revealed that it is found along with other lexicogrammatical patterns, involved in larger textual patterning of three kinds:

a) DENIAL – CORRECTION: X is not Y, X is Z

        SUBJ are not COMP ⎤
    (but)   are   COMP ⎦— repetition & modification

b) STATEMENT – CLARIFICATION through counter-statement: X is Z, X is not Y

        SUBJ are   COMP ⎤
            are not COMP ⎦— repetition & modification

c) DENIAL: X is not Y – implicit CORRECTION as preferred second pair part

The denial-correction pattern is one of the more culturally popular rhetorical patterns, according to Hoey (2001), used to organize non-narrative discourse (Georgakopoulou & Goutsos 2004: 147). It can thus be claimed that the public sign referring to plant pots and ashtrays carries its illocutionary force precisely by its appeal to this popular rhetorical pattern. This is achieved through its association with the third kind of patterning indicated above. By using the specific lexicogrammatical pattern the producer of the sign indicates to the receiver that it falls into a

larger denial-correction rhetorical pattern: you should not do as other people have already done, but you should correct your behaviour and refrain from stubbing your cigarette in the plant pots. It is the commonly repeated use of this pattern in specific textual and contextual frames that is here drawn upon in order to arrive at the illocutionary force of prohibition (and to exclude that the utterance performs a different speech act of e.g. declaration or instruction).

In other words, this analysis suggests that the perlocutionary effect of an utterance may be related to the conventional repeated use of linguistic structures in actual texts for the achievement of specific communicative goals. As a result, our focus in the analysis of speech acts should be neither on individual words nor on abstract syntactic structures but must be oriented towards lexicogrammatical patterns, which function as parts of broader rhetorical patterns and thus conventionally encode common communicative goals. In addition, as has been shown here, lexicogrammatical patterns may be exploited in larger textual patterns for rhetorical effects.

4. Verb form preferences and functions: *to forget* or *not to forget*
Turning to the second example under discussion, it can be argued that in order to fully understand the role of the construction *μην ξεχνάς ότι* 'don't forget that' we need to refer to the distributional properties of the verb *ξεχνάω/ξεχνώ* [kseχnáo/kseχnó][9] 'to forget'. Let us note here that the construction *μην ξεχνάς ότι* is marked in Greek for negation (with the negator *μην* [min]), tense (non-past), aspect (incomplete) and person (2nd singular). Thus, it would be interesting to investigate how these and other relevant grammatical categories are distributed in terms of frequency in the overall instances of the verb forms.

To this effect, a search was made in a proportional sub-corpus of CGT, comprising 4.5 million words from all text categories included in the full CGT, i.e. spoken (news, interviews, speeches, lecture, conversation) and written (fiction, news, opinion articles, academic, non-fiction, law & administration, private etc.) texts. There are 513 types of the lemma *kseχnáo* in Greek;[10] their distribution with respect to modality (modal, non-modal types, use of particles θα, να, ας or other), tense (past, non-past), aspect (complete, incomplete, perfective) and negation (positive, negative) is shown in Appendix 2.

As expected, the distribution of the data is skewed across grammatical categories, although the direction of this skewedness cannot be arrived at on the basis of intuition alone. First of all, *kseχnáo* seems to occur more in negative (291) rather than positive forms, as can be seen in Table 3.

Positive forms	222	43,3%
Negative forms	291	56,7%
Total	513	100%

Table 3: Negation in the forms of *kseχnáo*

The association of the verb *forget* with negation has already been observed for English both in intuitive (Jørgensen 1990: 149) and corpus-based studies (Biber et al. 1999: 159, 174) and in this respect the use of the negator *μην* in the example under discussion is not at all untypical.

Furthermore, an interesting correlation between negation and tense can be found in the Greek data from the proportional CGT, as can be seen in Figure 1 below:
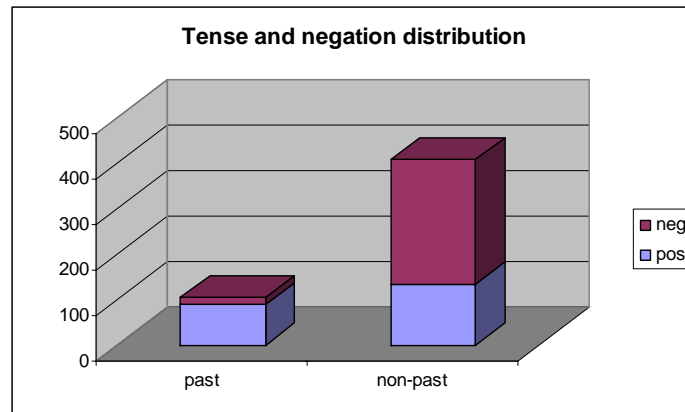
Figure 1: Tense and negation distribution in the forms of *kseχnáo*

As shown in Figure 1, most forms of *kseχnáo* are found in non-past (408) rather than past tenses (105).[11] However, when the verb is found in past tenses it usually occurs in a positive form (88 or 84%), while the reverse is true for the non-past forms of the verb, where uses with negative markers are more frequent (274 or 67%). It seems then that we have a double preference of the verb for negation and non-past tense and in this sense our example is again typical of general use.[12]

Verb forms in Greek are also marked for modality, according to Klairis & Babiniotis' grammar. Klairis & Babiniotis (2005: 436 ff., 474 ff.) distinguish between:

a) modal verb forms, i.e. those that are marked for epistemic or deontic modality and are combined with free grammatical morphemes used as modality markers (the traditionally called 'particles' *θα* [θa], *να* [na], *ας* [as] and imperative types, as shown in the Table of Appendix 2). A further case, not mentioned by Klairis & Babiniotis, should be added here, that of subordinate (e.g. conditional or concessive) clauses with modal verb forms. (These are listed under 'other' in Appendix 2; cf. Holton et al. 1997: 207 ff.). Modal verb forms take the negative marker *μην* [min].

b) non-modal verb forms, which are not marked for modality and are not combined with modality markers. Non-modal verb forms take the negator *δεν* [δen].

This analysis of Greek verb forms is economical, since it subsumes distinctions of mood, which, at any rate, is not clearly morphologically distinguished in Greek,[13] under modality: the indicative mood is identified with non-modal forms, whereas subjunctive and imperative moods are associated with non-modal forms. Figure 2 shows the interaction of the three grammatical categories discussed so far in the distribution of *kseχnáo* verb forms:
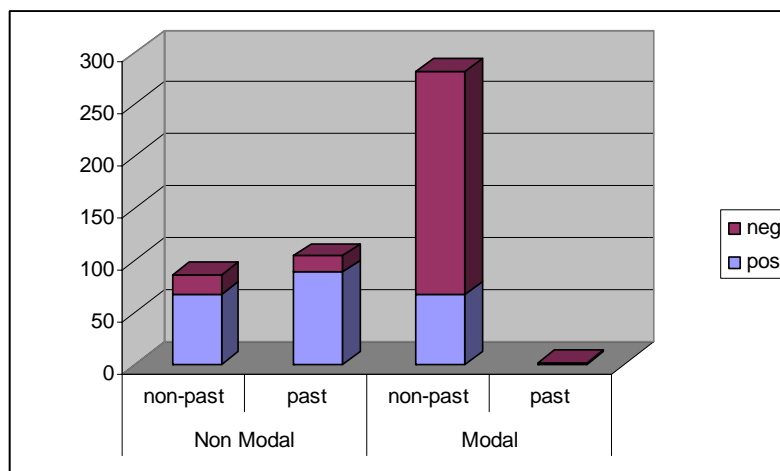

Figure 2: Modality, tense and negation distribution in the forms of *kseχnáo*

As can be seen in Figure 2, non-modal verb forms of *kseχnáo* are much fewer (190 or 37%) than modal forms (323 or 63%). The most frequent modal forms occur almost always (322) in non-past tenses and of these most (255 or 79%) occur in negative sentences. It seems thus that the preponderance of non-modal forms occurs with positive forms, whereas the opposite is true for modal forms of the verb.

In other words, as shown in Figure 3 below, negative non-past modal forms take up almost half (255 or 49.7%) of all instances of the verb forms for *kseχnáo*. Of the rest, more than one third is occupied by positive past non-modal forms and two quarters are equally taken by positive non-past non-modal and non-past modal forms.
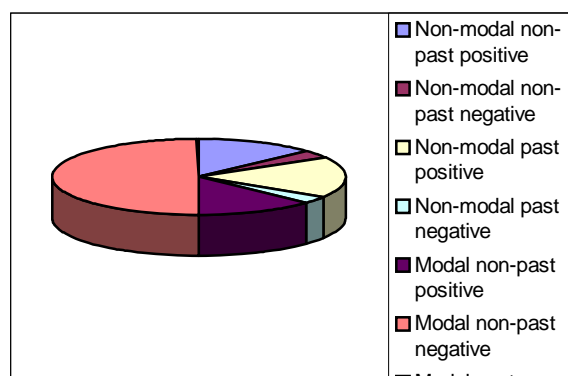

Figure 3: Distribution of modality, tense and negation in the forms of *kseχnáo*

Our investigation so far has shown that the verb *kseχnáo* in Greek tends to occur in negative non-past modal forms and this is precisely the case with the utterance in question. There are, however, two further grammatical categories that need to be explored, namely grammatical person and number, which are always marked in Greek verb forms through inflectional morphemes at the end of the word. Figure 4 below shows the distribution of the verb forms with respect to person and number.
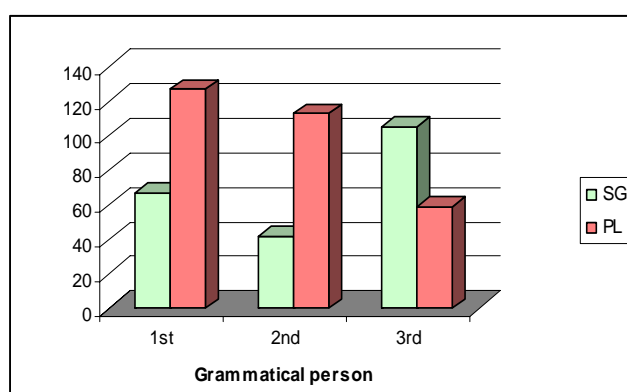

Figure 4: Distribution of grammatical person and number in the forms of *kseχnáo*

As can be seen, the verb is mostly found in the 1st and 2nd persons plural and then the most frequent forms are those of the 3rd person singular with the rest to follow.[14] It is also interesting to look at the interaction of modality and negation with person and number, as shown in Figure 5:
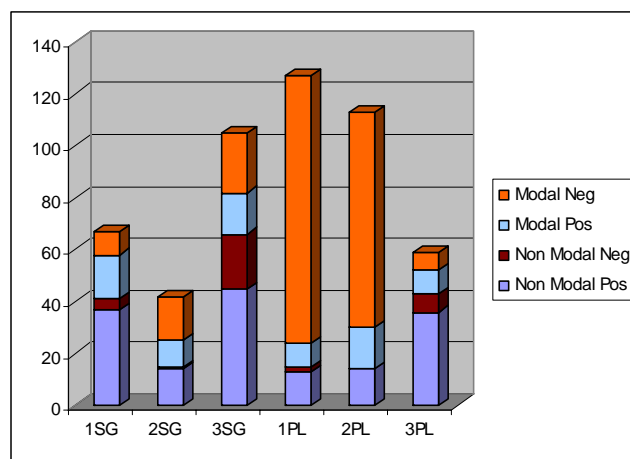
Figure 5: Modality and negation in each grammatical person in the forms of *kseχnáo*

What comes out as particularly prominent in Figure 5 is that the great majority of the verb's occurrences are found in the 1st and 2nd person plural in modal and negative sentences (186 in both persons or 36%). These are followed by 1st, 3rd singular and 3rd plural non-modal positive forms (118 or 23%) and the rest. In this sense, the example under discussion is untypical, since it is found in the 2nd person singular, although within this category most occurrences (17 or 40%) are indeed modal and negative, as happens with *μην ξεχνάς ότι* 'don't forget that'.

Finally, for reasons of completeness, it is necessary to also investigate the distribution of *kseχnáo* verb forms with respect to aspect and tense. The frequency of these is shown in Table 4 below:

	past	*non-past*	
Incomplete	8 (2.8%)	277 (97.2%)	285
Complete	83 (42%)	116 (58%)	199
Perfective	14 (48%)	15 (52%)	29
Total	105	408	513

Table 4: Tense and aspect in the forms of *kseχnáo*

As can be seen in Table 4, incomplete aspect forms tend to almost exclusively appear in non-past tenses, whereas complete and especially perfective aspect forms do not seem to show any preference for a particular tense. This is again interesting for the purposes of discussion here, since it suggests, among else, that past tense forms tend to predominantly occur in the complete aspect i.e. in types mainly associated with foreground rather than background events in a narrative sequence (Georgakopoulou & Goutsos 2004: 108). This would seem to imply that past tense forms are reserved for narrative actions, in opposition to non-past forms, which seem to be predominantly related with dynamic actions seen as processes and hence are found in incomplete aspect forms.

Let us summarize our findings so far: the verb forms of the verb *kseχnáo* show a marked preference for negation, modality, non-past tense and 1st and 2nd person plural (cf. Tao 2001: 137). To put it simply, forgetting in Greek is mostly what we and you should not or cannot do and to a much less extent what other people did or I did.

Examples (10) to (12) below must thus be regarded as typical of the use of the verb in Greek:

(10) Άλλωστε, μην ξεχνάμε ότι το 1ο δεκαήμερο του ζωδίου σας φιλοξενεί τον Κρόνο…
At any rate, we should not forget that the 1st ten days of your star sign are host to Saturn …

(11) Πολλαπλασίασε μάλιστα τις εξαγωγές της, που έφθασαν τα 250 εκατ. δρχ. <u>Ας μην ξεχνάμε</u> όμως ούτε στιγμή ότι καθημερινά απαξιώνεται και αν δεν γίνει γρήγορα κάτι…
It also multiplied its exports, which reached 250 m. dr. Let us not forget, though, not even for a moment, that everyday it loses its value and, if something is not done soon …

(12) Επίσης, χάρη στη σύνθεσή της, λειτουργεί ως φυσικό διεγερτικό. Βέβαια, <u>μην ξεχνάτε</u> ότι μία σοκολάτα 100 γραμμ. δίνει περίπου 525 θερμίδες …
In addition, thanks to its ingredients, it works as a natural stimulant. Of course, don't (you-PL) forget that a 100 gr. chocolate bar gives some 525 calories …

As these examples suggest, the verb is predominantly used in expressions that act as reminders ('we should not forget' equals 'we should remember', 'we should be reminded'). However, the speech act of reminding also has an important textual role: *forget*-expressions in Greek indicate transition to a new topic or a new aspect of the argumentation. It is significant that these expressions co-occur with discourse markers like *άλλωστε* 'at any rate', *όμως* 'although' or *βέβαια* 'of course', which also have a predominant sequential role in indicating transition. As a result, the illocutionary effect of the related speech act is not to forbid forgetfulness, as a literal understanding of the expression would suggest, but to bring another aspect of the situation to the reader's attention. It must also be noted that the use of the 1st person plural in examples like (10) and (11) above has a significant interpersonal dimension, since it is a form of the so-called inclusive plural, which attempts to implicate the reader in the writer's argumentation.[15]

This use of *forget*-expressions in Greek is what is being exploited in the following example from a TV script, involving the 2nd person singular in a negated, non-modal phrase:

(13) Καίτη: Έχεις δίκιο, το έχω σκεφτεί κι εγώ. <u>Μην ξεχνάς</u> όμως ότι πριν παντρευτούμε δούλευα ως γραμματέας. Θα κάνω λοιπόν …
Katy: You're right, I've thought about it too. Don't (you-SG) forget though that before we got married I worked as a secretary. I will do then…

The fictional character of Katy agrees with her interlocutor, only to express then her disagreement with the other person's assessment of the situation. The sequential role of the expression thus takes up an additional interpersonal role of expressing opposition. This analysis allows us to understand how our original example (repeated here as 14) takes on the force of a turn-taking device, as suggested at the beginning of this paper:

(14) <D> ντάξει όμως να σου πω κάτι <u>μην ξεχνάς</u> όμως ότι
fine but, let me tell you something, <u>don't forget</u> though that

The use of discourse markers like *ντάξει, όμως* etc. was found above to typically accompany *forget*-expressions in Greek. Their overuse in (14) contributes to the oppositional force of the speech act performed and thus we would expect speaker D to now express forceful disagreement with his interlocutors, a prediction which is borne out in the text that follows.

Apart from this dominant use, the verb forms of *ksexnáo* are also found in positive past non-modal forms, as in examples (15) and (16) and to much less extent in positive non-past modal and non-modal forms, as in examples (17) and (18).

(15) Είναι λέει πολύ απασχολημένος. <u>Ξέχασε</u> και τα ελληνικά του.
He says he's very busy. He's also <u>forgotten</u> (PAST-COMP) his Greek

(16) Α, <u>ξέχασα</u> να σου πω. Ο Αλέξανδρος είναι νέος γιατρός στην πόλη μας
Oh, I <u>forgot</u> to tell you. Alexandros is a new doctor in our town

(17) Παίρνω ένα ταξί να γυρίσω σπίτι και <u>ξεχνάω</u> μέσα το πορτοφόλι με όλες τις πιστωτικές μου κάρτες.

I get into a taxi to come home and <u>forget</u> (PRES-INCOMP) in it my wallet with all my credit cards.

(18) έχει μπει σε μια καινούργια τροχιά, προσπαθώντας να <u>ξεχάσει</u> το άσχημο παρελθόν της, χωρίς όμως να αφήνει ατιμώρητους τους υπευθύνους
… it has entered a new phase, trying to <u>forget</u> her awful past, but without leaving unpunished those responsible

In examples like (15) to (18) above, the verb forms of *forget* are related to narrative functions i.e. the marking of foregrounded events in a story, although secondary uses like the marking of an asides as in (16) ('I forgot to mention') are also found. It only makes sense that these functions would be related to 1st and 3rd person singular, as found above, since these are the persons mainly associated with story-telling.

To summarize the discussion of the second example, the achievement of participants' communicative goals is related to the predominant preferences among verb forms. As was found, the verb *forget* in Greek shows a clear tendency to occur in negative, non-past, modal expressions, especially in 1st and 2nd person plural forms. These expressions serve particular sequential and interpersonal functions, which are implicated in textual patterns of argumentation. When the verb is found in less frequent forms (e.g. past, non-modal, 1st and 3rd singular), it is associated, instead, with a narrative function.[16] The analysis suggests that the various forms of the verb are related to patterns and through them to specific discursive acts. Most of these forms prefer certain interpersonal functions (e.g. the inclusive first or the second person plural or the narrative first and third person singular) and are associated with preferred sequential outcomes (e.g. turn-taking or narrative foregrounding). *Forget*-forms thus do not just impart informational content but are used for addressing the listener in acts of reminding, warning, disagreeing, narrating etc. and in this sense are characterized by addressivity.

5. Conclusions and implications for a theory of language
Bringing it all together, I would like to suggest that the conclusions that can be drawn from corpus linguistic analyses like the above have an important contribution to make to language theory. The Greek examples discussed above suggest that speech acts are indicated in discourse through the use of commonly repeated lexicogrammatical patterns, associated with particular functions. These patterns constitute complex units of meaning, in the sense of Sinclair (1996), which combine:
- a specific form, involving grammatical categories like negation, definiteness, modality, grammatical person and number etc.
- frequency preferences
- particular semantic orientation (negative/positive, modal/non-modal etc.)
- associated interpersonal and sequential functions
- related rhetorical effects in an overall textual patterning.
It is this combination of characteristics which defines the speech act potential for each unit of meaning.

The contribution of the analysis of Greek examples is twofold. First of all, it suggests that these complexes of meaning are linguistically and culturally defined: although parallels can be found in other languages e.g. in patterns of negation or the behaviour of mental verbs like *forget*, it seems that languages and cultures part ways when it comes to dominant ways of expression and associated perlocutionary effects. Secondly, the analysis highlights the importance of individual forms in the process of signalling discourse functions. As we found out, the information encoded through the marking of grammatical categories in individual verb forms) is crucial in establishing their predominant functions in discourse. In this sense, even the individual word may carry traces of discursive acts through statistical preferences for particular

uses and meanings. Thus, whereas languages with reduced inflection like English may rely more on larger patterns, inflected languages like Greek may condense a large amount of information in the word morphology. It is obvious that much more corpus research is needed in order to confirm this hypothesis.

More generally, the analysis of Greek examples has pointed out that, as Adolphs puts it, "even minor variations in form can be linked to a particular variable in the function of an utterance" (2008: 2). Adolphs concludes that much more attention needs to be paid to individual lexical items and phrases, which recur in specific discourse patterns, whereas "it may be not necessary to resort to complex inference processes that rely on wide contextual knowledge" (2008: 14). Her view is supported by the line of argumentation developed in this paper. As was argued, we do have many indications from language form that a particular function or speech act is intended in a certain context. Knowledge about preferred frequencies, patterns and functions is thus indispensable in the micro-analysis and interpretation of texts.

By accepting the latter suggestion we support that corpus linguistic analysis has a significant contribution to make to a theory of language in the frame of existing linguistic approaches to the form-function relation in language. To somewhat simplify the picture, two major positions have been taken in respect to this central issue of linguistics. On the one hand, formal linguistic approaches, along with traditional semantic and pragmatic theories, attempt to discover fixed correspondences between forms and functions and, in this sense, are more concerned with discovering the system underlying language use. By attempting to relate specific forms to specific meanings in a pre-existing, fixed way, these approaches fit well with the code model of communication, which has come under serious attack in recent years (e.g. by integrational linguistics, among others: see Harris 1998).

On the other hand, discourse analytic views, including ethnographic, anthropological, sociolinguistic and other approaches, have emphasized the multifunctionality of forms and the crucial dependency of meaning on local context. It is this local, ever-renewed, ad hoc nature of interpretation that is foregrounded here with the implication, at least in strongest versions, that it is irrelevant whether a system exists or not. In this point of view what really matters are the nuances that meaning takes in each individual context as it is constructed anew by participants in a communicative event.

The history of linguistic thought can provide a significant parallel to this modern dilemma. As early as 1929, Voloshinov sums up his contemporary approaches to language into an opposition between Saussure's "abstract objectivism" and Humboldt's "individualistic subjectivism", as he labels the two conflicting views. In the former language is seen as "a stable, immutable system of normatively identical linguistic forms which the individual finds ready-made" (Voloshinov 1973: 57), whereas, according to the latter, language is realized in individual speech acts, as an "ever-flowing stream of speech in which nothing remains fixed and identical to itself" (1973: 52). There could hardly be a better description of the antithesis in current linguistic thought between formalism and functionalism or competence-based and performance-based approaches to language.

What are the lessons that we have learned from corpus linguistics with respect to this dilemma? In other words, how is corpus linguistic analysis placed in the map of current approaches to language? As one may have guessed, my suggestion is that corpus linguistics is closer to a Voloshinovian view of language, which gives emphasis on the recurring links between form and function. According to Voloshinov, linguistic items are imbued with dialogical and social meaning, by being the product of repeated uses produced in specific contexts in time. In his famous quote, "each word […] is a little arena for the clash and criss-crossing of differently oriented social accents" (Voloshinov 1973: 41). In this sense, no linguistic item is fixed in its role, as seems to be implied by formal approaches, but carries traces of previous discursive acts with particular orientation. It thus holds a meaning potential

that can be creatively exploited in individual speech events and can become the locus for negotiation and opposition. However, contrary to functionalist assumptions, individual lexical items are also not neutral with respect to pre-existing meanings, nor are they pliable to any kind of manipulation in discourse. They rather come to each discourse participant with pre-set preferences of occurrence and function, as defined by their history in language.

As was argued in this paper, corpus linguistic analysis concurs with this view by suggesting that language relies on lexicogrammatical patterns or form-function complexes that contain traces of their possible use in discourse. These traces may be indications of their preferred illocutionary force, their semantic orientation, their predominant uses in terms of frequency, their potential for rhetorical exploitation in text etc. All these aspects constitute cues for the speaker and clues for the listener, assisting them in the process of meaning-making. As was particularly seen in the Greek case, traces may also appear in individual words rather than patterns and thus particular word forms may be equipped with specific meaning potential.

Voloshinov also points out two important aspects of language, which have now come to gain increased attention through corpus analysis, namely addressivity and evaluation. As he emphasizes, the word is "a two-sided act"; it constitutes "the product of the reciprocal relationship between speaker and listener, addresser and addressee" (1973: 85, 86).[17] Some of the most obvious ways in which addressivity may be manifest in language use have been explored in our analysis above, including the perlocutionary effects associated with lexical patterns and the predominant patterning of individual words such as *kseχnáo* with grammatical person and through this with discursive acts and textual effects. With respect to evaluation, Voloshinov also succinctly states that "there is no such thing as word without evaluative accent" (1973: 103) and that "every utterance is above all an *evaluative orientation*" (1973: 105). Aspects of evaluation have been brought up above in the analysis of the semantic orientation of patterns and words and the rhetorical effects produced by them. It is significant that evaluation has become a central theme in analysis through corpora in languages like English (Bednarek 2006) or Greek (Fragaki 2009).

The significant parallels between Voloshinov's position and the findings of corpus linguistic analysis, such as those suggested by research in the Corpus of Greek Texts, have important implications for the role of corpus linguistics and its contributions to a theory of language. The theoretical conclusions of the claims made by analyses through corpora point to a fundamentally Voloshinovian or Bakhtinian and –dare I say– Marxist view, which gives emphasis on language as a form of historical praxis, a systematic accident produced by repeatable and recurring forms of utterances, produced in particular contexts and bearing the marks of their social and dialogical origin.

References

Adolphs, S. (2008). *Corpus and Context. Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam/Philadelphia: John Benjamins.

Bakhtin, N. M. (1986). *Speech Genres and Other Late Essays*. Austin: University of Texas Press.

Bednarek, M. (2006) *Evaluation in Media Discourse. Analysis of a Newspaper Corpus*. London: Continuum.

Biber, D. (1993). "Representativeness in corpus design". *Literary and Linguistic Computing*, 8, 1-15.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education.

Ferlas, E. (2008). "Lexical clusters in the Corpus of Greek Texts". Paper given at the Formulaic Language Research Network (FLaRN) conference, 19 June 2008, University of Nottingham.

Fragaki, G. (2009). The evaluative role of the adjective and its use as a marker of ideology. [In Greek]. PhD dissertation, University of Athens.

Georgakopoulou, A. & D. Goutsos (1998). "Conjunctions versus discourse markers in Greek: The interaction of frequency, positions and functions in context". *Linguistics* 36, 887-917.

Georgakopoulou, A. and D. Goutsos (2004). *Discourse Analysis: An Introduction*. Edinburgh: Edinburgh University Press. (2nd edition).

Goutsos, D. (2001). "Sequential and interpersonal aspects of English and Greek answering machine messages". *Pragmatics*, 11, 357-77.

Goutsos, D. (2003). "Corpus of Greek Texts: Design and implementation" [In Greek]. *Proceedings of the 6th International Conference of Greek Linguistics, University of Crete, 18-21 September 2003*. Available at: http://www.philology.uoc.gr/conferences/6thICGL/gr.htm (accessed: 1 July 2009).

Goutsos, D. (2007). "Basic place adverbs in electronic corpora: Preliminary remarks" [In Greek]. In Department of Linguistics, University of Athens (ed.) *Studies Dedicated to Dimitra Theophanoulou-Kontou* [In Greek]. Athens: Kardamitsa, 36-46.

Goutsos (forthcoming). "Corpus of Greek Texts: A reference corpus for Modern Greek". *Corpora*.

Goutsos, D. & G. Fragaki (forthcoming). "Lexical choices of gender identity in Greek genres: The view from corpora". In A. Georgakopoulou and V. Lytra (eds.) Language & Identities in Greek. Special issue of *Pragmatics*.

Goutsos, D., Hatzidaki, R. and P. King (1994a). "A corpus-based approach to Modern Greek language research and teaching". In I. Philippaki-Warburton, K. Nicolaidis and M. Sifianou (eds.) *Themes in Greek Linguistics: Papers from the First International Conference on Greek Linguistics. Reading, September 1993*. Amsterdam/Philadelphia: John Benjamins, 507-13. Reprinted in Teubert, W. & R. Krishnamurthy (eds.) *Corpus Linguistics. Critical Concepts in Linguistics*. Volume 6. London & New York: Routledge, 150-56.

Goutsos, D., King, P. and R. Hatzidaki (1994b). "Towards a Corpus of Spoken Modern Greek". *Literary and Linguistic Computing*, 9, 215-23.

Goutsos, D. & P. Pavlou (forthcoming). "CGT: Building a reference corpus of Greek". *Proceedings of the 30th International Conference on Functional Linguistics, University of Cyprus, 18-21 October 2006*.

Harris, R. (1998). *Introduction to Integrational Linguistics*. Oxford: Pergamon.

Hatzigeorgiu, N., Gavrilidou, M., Piperidis, S. Carayannis, G., Papakostopoulou, A., Spiliotopoulou, A., Vacalopoulou, A., Labropoulou, P., Mantzari, E., Papageorgiou, H. & I. Demiros (2000). "Design and implementation of the online ISLP corpus". *Proceedings of the LREC 2000 Conference*, 1737-42.

Holton, D., Mackridge, P. & I. Philippaki-Warburton (1997). *Greek. A Comprehensive Grammar of the Modern Language*. London: Routledge.

Hoey, M. (2001). *Textual Interaction: An Introduction to Written Discourse Analysis*. London: Routledge.

Hunston, S. & G. Francis (1999). *Pattern Grammar*. Amsterdam/Philadelphia: John Benjamins.

Jørgensen, E. (1990). "Remember and forget with gerund and infinitive as objects". *English Studies*, 71 (2), 147-51.

Joseph, B. D. & I. Philippaki-Warburton (1989). *Modern Greek*. London: Routledge.

Katsalirou, A. (forthcoming). General academic vocabulary in the teaching of Greek as a foreign language. PhD dissertation, Aristotle University of Thessaloniki.

Klairis, C. & Babiniotis, G. (2005). *Grammar of Modern Greek. Structural-Functional and Communicative* [In Greek]. Athens: Ellinika Grammata.

Koutsoulelou, S. and G. Mikros (2004-2005). "The word γεγονός as a shell-noun. Use and function in Greek electronic corpora" [In Greek]. *Glossologia* 16, 65-95.

Pomerantz, A. (1978). "Compliment responses: Notes on the co-operation of multiple constraints". In J. Schenkein (ed) *Studies in the Organization of Conversational Interaction*. New York: Academic Press, 79-112.

Renouf, A. (2007). "Corpus development 25 years on: from super-corpus to cyber-corpus". In R. Facchinetti (ed) *Corpus Linguistics 25 Years on*. Amsterdam: Rodopi, 27-49.

Sinclair, J. M. (1996) "The search for units of meaning". *Textus* IX (1), 75-106.

Tao, H. (2001). "Discovering the usual with corpora: The case of remember". In R. Simpson and J. Swales (eds.), *Corpus Linguistics in North America: Selections from the 1999 Symposium.* Ann Arbor: University of Michigan Press, 116-44.

Tao, H. (2003). "A usage-based approach to argument structure. 'Remember' and 'Forget' in spoken English". *International Journal of Corpus Linguistics*, 8 (1), 75-95.

Triandafyllidis, M. (1941). *Modern Greek Grammar* [In Greek]. Athens: OESV.

Voloshinov, N. (1973 [1929]). *Marxism and the Philosophy of Language*. Cambridge, MA: Harvard University Press.

Appendices

Appendix 1: Concordance for the pattern 'the pl-N *not* V-link N' in the spoken sub-corpus of CGT:

1. οι αγώνες με κλασσικά αυτοκίνητα **ΔΕΝ είναι** αγώνες ταΧΥτητας είναι ΠΟΛΙτιστικές εκδηλώσεις (.) μια και τα
2. Το πραξικόπημα και η εισβολή **δεν είναι** γεγονότα που κάποτε συνέβησαν και σήμερα αποτελούν μακρινό
3. αυτοί που πιστοποιούν τις υπογραφές **δεν είναι** ειδικοί υπάλληλοι επί τούτω, αλλά υπάλληλοι σε άλλα επαγγέλματα.
4. Θεωρώ ότι είναι δίκαιο –**δεν είναι** επαίτες οι δικηγόροι, πρέπει να στηριχθούν με αξιοπρέπεια–
5. το διαχειριστικό συμβούλιο; Αυτά **δεν είναι** έργα για να επαίρεται κανείς, είναι βαρβαρότητα για να εντρέπεται
6. Αυτά είναι θέματα του Κινήματος, **δεν είναι** θέματα της κυβέρνησης. Μπορώ να αρκεστώ, όμως, σε ένα σχόλιο:
7. επειδή καμιά φορά αυτά τα θέματα **δεν είναι** θέματα θρησκείας και Θεού, αλλά είναι θέματα απλής απλοποίησης
8. Είναι θέματα καθαρά εσωτερικά. **Δεν είναι** θέματα πολιτικής σημασίας. Δεν είναι πολιτικές διαφορές ή
9. πουλιούνται και σε πάγκους // Όχι **δεν είναι** θεωρίες, στο γυμνάσιο τα μαθαίνουμε. Έχετε βρεθεί ποτέ στην Ερμού
10. ΠΡΟΕΔΡΕΥΩΝ (Παναγιώτης Σγουρίδης): **Δεν είναι** καρπούζια, κύριε Γιαννόπουλε. </p> <p> Ο κ. Βαθειάς έχει το λόγο.
11. [δεν πρ- περίπου τα [όχι παιδιά **δεν είναι** κολόνες αυτές αυτ- αυτές [ναι αυτές εδώ δίπλα ((δείχνει)) είναι
12. θεμέλιο πάνω στο οποίο πιστεύω ότι **δεν είναι** μαγικά, ταχυδακτυλουργικά κόλπα τα θαύματα. Θα επαναλάβω τι είναι.
13. πλέον δισεκατομμύρια δραχμές. Αυτά **δεν είναι** μακέτες, είναι μια πραγματικότητα. Δίνουν λύσεις.
14. τα μηνύματα που στέλλει η κυβέρνηση **δεν είναι** μηνύματα αναγνώρισης των θυσιών των εγκλωβισμένων και των παιδιών
15. υπεύθυνη δουλειά.</p> <p> Όλα αυτά **δεν είναι** μόνο θέματα του παρελθόντος. Καραδοκούν καθημερινά σε κάθε μας
16. και τέτοιων επιστολών, και – **δεν είναι** μόνο θύματα πλάνης, αλλά μπορεί να πάθουν και κακό? – εφόσον,
17. κάτοικοι της ελεύθερης Αμμοχώστου **δεν είναι** πολίτες δεύτερης κατηγορίας. Και πρέπει να γίνει ξεκάθαρο:
18. στο σύνολό τους, μα οι πρυτάνεις **δεν είναι** πολιτικοί για να γράφουν και να εφαρμόζουν πολιτική. Εμείς έχουμε
19. Εκεί οι περισσότερες μεταβιβάσεις **δεν είναι** σπίτια, αλλά είναι αγροτεμάχια, που είναι πολύ κάτω από πέντε
20. στοιχεία αυτά, κύριοι συνάδελφοι, **δεν είναι** στοιχεία που εφεύρε το ΑΚΕΛ. Είναι στοιχεία που έχουμε πάρει

Appendix 2: Distribution of verb forms for the verb *ξεχνάω* 'to forget' in the proportional CGT

Modality	- modal				+ modal											
Tense	-past		+past		-past								+past		-past	
					θα		να		ας		other		other		Imperative	
Negation Aspect	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg	pos	neg
Incomplete	56	15	6	2	0	0	10	48	0	45	0	29	0	0	4	70
Complete			68	14	4	16	33	38	3	0	5	9	0	1	8	0
Perfective	11	4	14	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	67	19	88	16	4	16	43	86	3	45	5	38	0	1	12	70

Notes

---

[1] Research for this paper and participation in the Corpus Linguistics Conference 2009 were supported by the University of Athens research programme 'Kapodistrias', code no 70/4/7607.

[2] The transliteration of Greek is broadly phonetic and follows the conventions of Goutsos (2001). For the transcription symbols used, see Georgakopoulou & Goutsos (2004: vii).

[3] Georgakopoulou & Goutsos (2004), among many others, offers a concise introduction to most of these approaches.

[4] Documentation of the ILSP corpus includes Hatzigeorgiu et al. (2000). The related webpage address can be found at: http://hnc.ilsp.gr.

[5] A more complete discussion of these shortcomings and more details on Greek corpora can be found in Goutsos (forthcoming). Information on the composition of the HNC can be gleaned from the website: http://hnc.ilsp.gr/subcorpus.asp# (accessed: 1 July 2009).

[6] In this sense, CGT can be argued to be ahead of the 20 million *Birmingham Corpus* (1980-1986), but still far behind the dynamic corpora of the 1990s, in Renouf's (2007) terms.

[7] Capitalization in the example indicates syllables of prominence, according to the transcription conventions followed for CGT spoken texts.

[8] Repetition with variation has been found to be common in oral narrative discourse (Georgakopoulou & Goutsos 2004: 120). In Hoey's (2001: 31, 84) terms, the relation between the two parts is one of matching contrast.

[9] For reasons of co-articulation the verb *kseχnáo* appears as *min gseχnás* in the example under discussion.

[10] This frequency is remarkably more frequent (114 per million) than the respective figure given for English (more than 50 per million) in Biber et al. (1999: 370, cf. 663, 701).

[11] The distinction between past and non-past verb forms is well-established in modern Greek grammars; here, as in the rest of the analysis and the Table of Appendix 2, we follow the distinctions of Klairis & Babiniotis (2005: 443).

[12] For English Jørgensen finds that there is "clearly very little use in the language *for positive forget* with retrospective function" (1990: 151), while Tao (2003: 84), instead, observes that *forget* clauses deal much more with past events than clauses with *remember*.

[13] Klairis & Babiniotis' (2005) analysis goes against the grain of distinguishing subjunctive forms in Greek both in traditional (Triandafyllidis 1941) and contemporary grammars (Joseph & Philippaki-Warburton 1989: 179 ff. and Holton et al. 1997: 205 ff.). Note also that, θα is considered to be a subjunctive marker in Holton et al. (1997), but not in Joseph & Philippaki-Warburton (1989).

[14] Cf. Tao (2003), who finds that the subject of *forget* in English is more exclusively associated with the first person, usually the current speaker.

[15] Tao (2003: 91) finds that *forget in English* is more closely associated with negative semantic connotations, since it can be more face-threatening than *remember*, especially in the cases of second person subjects (e.g., 'you mustn't forget') in interpersonal communication.

[16] In this sense, the 1st person singular, which is customarily used as the lemma form in Greek, is in this particular case most unrepresentative of the actual use of the lemma.

[17] Surely, this view is related to Bakhtin's emphasis on "the traces of addressivity and the influence of the anticipated response, dialogical echoes from others' preceding utterances, faint traces of changes of speech subjects that have furrowed the utterance from within" (1986: 99).