Exceptional Clitic Placement in Cypriot Greek: Results from an MET Study*

Panayiotis A. Pappas
Department of Linguistics, Simon Fraser University
panayiotis_pappas@sfu.ca

Abstract
Recent studies on the pattern of clitic placement in Cypriot Greek (Revithiadou 2006; Chatzikyriakidis 2010, 2012; Pappas 2010, 2011), have posited the existence of counterexamples to the rule that the pronoun is proclitic after a complementizer or other such function words. These counterexamples are associated with a specific set of lexical items: έντζε, ότι, επειδή, αφού, and γιατί. Equally unclear is the clitic pattern with preverbal elements such as focused DP subjects. I present here the results of an acceptability judgment study of 34 Cypriot speakers based on magnitude estimation tests (MET) in ten different syntactic environments and two different conditions (enclisis vs. proclisis), for a total number of data points N = 680. The results demonstrate that these exceptional patterns are integral parts of Cypriot Greek competence and highlight the role that lexical items can play in terms of creating sub-patterns of generalizations within larger schemes.

Keywords
clitics – Cypriot Greek – dialectology – experimental syntax – magnitude estimation

* This research was partially supported by SSHRC SRG 639510. I wish to thank Stavroula Tsipplakou for her help in tracking down sources on Cypriot Greek, Martin Wieling for his advice on the statistical modeling, an anonymous reviewer for insightful comments, as well as the audiences of GESUS 2011 and CWSL 2014 for their helpful feedback. I also thank the Department of English at the University of Cyprus for their support, as well as the participants for their time. The usual caveats apply.
1 Introduction

Although it has received a fair amount of attention in the literature over the past two decades, the pattern of object clitic placement in Cypriot Greek (henceforth CG) is still not well understood. It is true that there is general agreement about the basic pattern, but there are several details that have yet to be tidied up. These details are significant, because they present serious challenges to the (mostly) syntactic generalizations that have been advanced to account for the phenomenon. In the present study, I attempt to provide descriptive clarity to the pattern by examining the acceptability of proclitic and enclitic constructions in these specific environments. To accomplish this, I discuss the results of an acceptability judgment task experiment based on magnitude estimation testing (MET, cf. Johnson 2008). The results demonstrate clearly that these exceptions are integral parts of CG competence and not epiphenomena of performance, diglossia, or dialect contact. They also highlight the role that lexical items can play in terms of creating sub-patterns or constellations (cf. Joseph 1997) of generalizations within larger schemes. Finally, they confirm the validity of data found in sociolinguistic interviews, even when the patterns are quantitatively marginal, and re-enforce the notion that, in some cases, it is necessary to triangulate evidence from different methodologies (in this case, the historical record, sociolinguistic interviews, and experimental syntax) in order to fully understand a pattern of variation.

1.1 Background

The basic pattern of clitic placement in CG is first described in terms of a modern syntactic analysis by Agouraki (1997) who states that the language exhibits enclisis except in the context of “complementizers, negation, modality markers, wh-questions and syntactic xp-foci.” A very similar description of the pattern is also given by Terzi (1999). Agouraki (2001), however, notes that ἐντζε,1 a negative marker, is associated with enclisis and not proclisis as one would expect. Revithiadou (2006) presents an even more complex picture as she acknowledges the existence of variation with the complementizer ὅτι, and after certain wh-phrases. Agouraki (2009) revises the pattern slightly, stating that the set of XPs that can appear before the verb with proclisis is limited to a few items (existential and negative quantifiers, NPs, only-phrases, and proforms).

---

1 The spelling choice for this word conceals an etymological verdict as well, as is explained in the discussion. I choose not to write ἐντζαι, because I am not convinced that this word involves the coordinating conjunction.
and suggests that αφού is also exceptionally associated with enclisis. Pappas (2010, 2011) presents a much more complicated pattern, claiming that there are several function words (complementizers, negation, and modality markers) which are associated with enclisis instead of proclisis (έντζε, ότι, επειδή, αφού, and γιατί), and that the pattern of clisis with pre-verbal (stressed) elements exhibits a considerable amount of variation that cannot be accounted for by the distinction between focus and topic. Finally, Chatzikyriakidis (2010, 2012) lists focused DP subjects as pre-verbal elements that appear with proclisis (i.e., contra Agouraki 2009)2 and states that both proclisis and enclisis are available for the complementizer ότι and the causal conjunctions επειδή and γιατί. He also confirms that έντζε is surprisingly associated with enclisis (as is μήπως τζαι in his findings), and notes that the word ενώ appears with proclisis in its temporal use and enclisis in its contrastive use.

It is practically impossible to assess the validity of each of these empirical claims, because there is very little information about how these datasets are constructed.3 For example, Revithiadou (2006: 83) mentions that the analysis is based on data from the cg Corpus, “with the help of native speakers of the varieties of Paphos and Nicosia” but does not provide any other details about how the corpus was constructed (how many speakers, of what age, how the data was elicited, etc.). Chatzikyriakidis (2012) states that the speakers in that study are between the ages of 20 and 35, some of whom have lived in London for more than a decade, and were asked to translate Standard Modern Greek (SMG) sentences into cg as well as provide grammaticality judgments for a set of prepared sentences in cg. However, the number of speakers is not given, and several aspects of the elicitation process remain unclear: Were the translations spoken, or written? Were the acceptability judgments given informally or by using a specific tool, such as a Likert scale? Are there any concerns that the viewing of the SMG stimuli may have affected the acceptability judgments of the Cypriot constructions? Agouraki (1997, 2001, 2009) does not provide any information about the construction of the dataset and neither does Terzi (1999). Finally, even though this type of detail is available in Pappas (2010, 2011), the number of tokens for the exceptional environments is too small for the results to be accepted without question. And though the source of data in syntactic studies

---

2 Although they do not address Cypriot directly, Condoravdi and Kiparsky (2001) also assert that in varieties of the same dialect group as Cypriot focused subjects appear with proclisis.

3 In highlighting these issues, it is not my intent to disparage the contribution of previous studies. I simply wish to emphasize the need for systematic methodology in dialectological research and for its detailed description in publications.
has been a general concern for the field (cf. Featherston 2007, and the responses to it), it is especially so when we are dealing with non-standard varieties that are stigmatized or proscribed in certain domains, as is the case with CG (cf. Papapavlou 2001; Pavlou and Papapavlou 2004; Tsiplakou et al. 2006; Rowe and Grohmann 2013). For in these situations it is very difficult to gauge the effect that the experimental setting has on participants and therefore replicability is compromised. At the very least then, a detailed description of the elicitation process is required.

At this point, one may ask whether the exceptions mentioned above are significant enough to merit such painstaking investigation. After all, according to Pappas (2010), these are constructions that occur quite rarely (roughly 3% of all clitic constructions). On the other hand, an accurate description of the entire pattern has methodological, theoretical, and applied implications. In terms of methodology, it is important to know whether these constructions that appear to be marginal in terms of usage evoke robust acceptability evaluations or not. It has long been a tenet in variationist linguistics that marginal patterns do not carry social meaning (cf. Tagliamonte 2006), but this does not necessarily mean that minor patterns found in data from sociolinguistic interviews should be ignored.

Theoretically, the pattern of clitic placement with focused DP subjects has important ramifications as to whether the phenomenon can be explained through syntax alone (as in Agouraki 1997, 2001, 2003; Terzi 1999), or if an explanation that takes into account both syntactic and prosodic considerations is required, as in Condoravdi and Kiparsky (2001) and Revithiadou (2006). The issue surrounding the lexical exceptions raises the question of whether these can be accounted for by a generalization that is consonant with the overall pattern, and distinguishes between phrase-initial positions and positions within the IP, or if each of them needs to be treated separately as is done in Chatzikyr- iakidis (2010, 2012). Both of these issues have repercussions for the history of Greek as well. The pattern of clitic placement in CG has many similarities with that of Medieval Greek, which also contains exceptions, several of which remain controversial (cf. Pappas 2004a; Condoravdi and Kiparsky 2004). Thus, determining which of the exceptions in CG are valid and which are not may also provide insight about the general development of clitic placement patterns in Greek. Finally, in the sociolinguistic environment of Cyprus, where there is significant tension between the non-local standard (SMG) and the native variety, the accurate understanding of a dialectal feature as iconic as clitic placement has significant implications in applied fields such as language pedagogy, speech pathology, and second language acquisition. In other words, the data matter.
For these reasons, I designed and conducted a study of acceptability judgment tasks for clitic placement in CG using a magnitude estimation test. In the following sections, I describe the set up of the study, its most significant results, and interpret them in light of previous findings.

2 Methodology

2.1 Design

The objective of the experiment is to test whether speakers of CG judge the acceptability of enclitic and proclitic constructions differently depending on the nature of the immediately preceding word. Based on the existing literature and especially on the Pappas (2010, 2011) data from sociolinguistic interviews which prove the existence of exceptional patterns in free-flowing conversations among native speakers, this study tested constructions involving the following environments:

- Function words: αφού, έντζε, επειδή, γιατί (causative conjunction and question word), and ότι (cf. examples 1–6 in Appendix A).
- Pre-verbal subjects: Even though the discussion in the literature classifies pre-verbal subjects as focused or not, this study, following Kiss (1998), tests three types of constructions: contrastive focus, information focus, and no focus on the pre-verbal subject (examples 7–9).
- Temporal αφού (example 10). In order to be able to determine whether the magnitude estimation test is capturing acceptability judgments based on CG competence rather than SMG, this construction was included because it is only available in the latter. In CG αφού cannot be used as a temporal conjunction, and the periphrasis μετά πού is used instead.

4 In order to keep the design from becoming too complex, we did not test other types of preverbal phrases such as prepositional phrases or fronted objects. These environments as well as other function words, such as ενώ, which are reported in the literature (as in Chatzikyriakidis 2010, 2012), but are not present in the sociolinguistic interviews will be examined in subsequent experiments.

5 The pronouns in the curly brackets indicate the two positions that were varied in the experiment.
Thus, there are ten (10) environments to be tested in two conditions, proclitic and enclitic, yielding 20 test tokens, as visualized in Table 1. These tokens were constructed with the help of four (4) native speakers of CG and graduate students in the Linguistics program of the Department of English at the University of Nicosia who were employed as ras for the project. Three different sentences were constructed for each environment; these were evaluated by the RAs and the ones considered best were selected for inclusion in the experiment. In addition to these, the RAs constructed 36 filler sentences, as well as one modulus sentence and four (4) training sentences (examples 11–14 in Appendix B) for the purposes of the met design. As can be seen in example (1), the modulus phrase was constructed to portray a common activity, and uses CG features in phonology, morphology, syntax, and lexicon, with the intention that this would be a clear indication to participants to judge the acceptability of sentences on the basis of CG grammar and not that of SMG.

(1) Να πάω στην φροντιαριά να πιάω
fut go.ISG.PRES to.the grocery comp get.ISG.PRES
κολοκάσι
jerusalem.artichoke

'I will go to the grocery store to buy some Jerusalem artichokes'

The training sentences were constructed in order to provide participants with a variety of acceptability judgments, ranging from SMG constructions (example 13 in terms of its syntax and vocabulary) to CG constructions found in the sociolinguistic interviews (examples 12 and 14). One sentence contained a clear violation of CG grammar, namely the interpolation of a PP between the future marker εννα and the verb (example 11). In addition to giving speakers a sense of the types of constructions they would hear in the experiment, these sentences also served as an indication of how well participants understood the purpose of the experiment. Participants who gave the SMG or ungrammatical construction a high score (above 35), or the actual CG constructions a low score (below 35) would be removed from the dataset. Once the entire set of 56
sentences was selected, each utterance, or brief exchange (for the purposes of context) was performed by the research assistants and recorded on a Marantz PDM 660 recorder. The entire set was then randomized and arranged in a powerpoint presentation so that there would be 10 seconds of silence between each utterance.

2.2 Participants
Participants were recruited among the family and friends of the research assistants and from the student population at the University of Cyprus. The former completed the experiment in their homes, while the latter did so in an office on campus. Both groups of participants were given the same set of instructions (adapted from Johnson 2008) which were rendered in Cypriot vernacular by the assistants to their relatives and friends and in English by the author to students from the University of Cyprus.

2.3 Procedure
The experiment proceeded as follows: First, the participants were asked to familiarize themselves with the task of magnitude estimation: they were presented with a set of three lines and their accompanying measurements in millimetres, but it was explained to them that these were relative numbers roughly showing the difference between the lines if we were to arbitrarily assign the number 50 to the first one. Then they were presented with a set of seven lines and asked to assign a number next to each one, keeping in mind the guideline of 50. At this point we discussed any questions or cleared up any issues that may have arisen. The next step was a training session in which the participants were asked to listen to a set of five sentences, and, keeping in mind that the modulus sentence had been deemed to be an everyday Cypriot vernacular utterance, that might be rated with the 50-line from the previous task, they were asked to use lines to evaluate these utterances according to how acceptable each one was in CG and in proportion to its acceptability. After this training session, participants were once again given the opportunity to ask questions about the process.

For the real task, the participants got to listen to the modulus phrase once more, and then proceeded with each of the 56 utterances, and this time they were asked to draw a line and assign a value for each utterance before moving on to the next. They were also encouraged to make as fine a distinction according to how acceptable each utterance sounded. Both the PI and the RAS kept notes during the experiment to see if there were any constructions that gave the participants particular problems. There were very few issues, all of them of a technical nature (sound not loud enough, the program froze, etc.).
At the conclusion of the experiment, the participants were given an information sheet in which the precise objective of the task was discussed and then asked to sign the consent form. The experiment was administered to 38 participants between the ages of 19 and 47 (24 female, 14 male); all participants gave us permission to use the data.

3 Results

Once all the questionnaires were collected, each line was measured and the length (in millimetres) was entered in a spreadsheet in one column, while the number assigned by the participant was entered in a different column. Each score and sentence was also coded for the nature of the preceding word and the placement of the pronoun. Of the 38 participants, four had to be excluded from the dataset, either because their scores in the training section were contrary to common acceptability judgments (two) or because the values entered during the experiment proper did not follow the directions (also two): One participant only entered values above 1,000 for all tokens, while the other entered values either below 30 or above 200. This leaves 680 observations for each type of entry in the dataset, length of line, and numerical score.

The first question is whether the two types of entry are correlated. The results of a Pearson's product-moment correlation test ($r = .85, p<0.0001$) clearly show that there is indeed a very strong and significant correlation between the length of the lines that the participants drew and the scores that they assigned for each utterance (Figure 1). This internal consistency is a good indicator about the validity of the values that were entered. Since the score values are truly unbounded whereas the line values are constrained by the width of the page, the former are the ones used for the statistical modeling.6

Figure 2 shows the distribution of raw scores assigned to the proclitic and enclitic pronoun placement for each syntactic environment to allow the reader to form a visual impression of the data. The modulus score is represented by a dotted line to help with the interpretation of the difference between the means which are marked by thick bars. A first observation that, again, demonstrates the validity of the experimental method is that only in one of the ten constructions do the mean scores for enclisis vs. proclisis both fall

---

6 The models were also implemented for the line values, but since there were no significant differences in the results, I do not report them here.
below the line for the modulus score: this is the case for the temporal αφού constructions, which were included precisely in order to determine whether the participants are providing judgments on the basis of Cypriot grammar or that of the standard. In contrast to this, we see that in the other nine cases, the scores for enclisis vs. proclisis fall on either side of the modulus score to a smaller or lesser degree. In order to determine which of these differences are statistically significant, we need to build a statistical model.

The design of the experiment is one in which every participant provides a score for every token in the dataset, in other words, it is a repeated measures design, with both within subject and within item variability. The standard practice (Baayen 2008) is to analyze the data constructed by such experimental designs with linear mixed effects models which are able to take into consideration the contribution of several random effects, and their interaction (crossing) to the overall pattern of variation. As Barr et al. (2013) demonstrate, the statistical model that one constructs should be the maximal one necessitated by the
data. For most instances of psycholinguistic and experimental syntax datasets, this means that the model should include random intercepts for subjects (participants) and items (stimuli) as well as random slopes for subjects. Random intercepts allow the intercept term to vary across subjects, so that the model can generalize from the pool of subjects to the population from which they are drawn, while random slopes “allow subjects to vary with the treatment effect,” in order to account for the fact that participants vary in their responses to the stimuli.

These models are implemented by running the package *lme4* (Bates et al. 2013) in R (R Core Team 2013), using the formula $model = Y \sim X + (1 + X|Subject) + (1|Item)$, where $Y$ is the dependent variable, $X$ is the independent one, and the notations $(1 + X|Subject)$ and $(1|Item)$ introduce random intercepts and slopes per subject, and random intercepts per item, respectively.
The analysis that is necessary for the present research question is somewhat more complex. What we are asking is whether the score that participants have assigned to the proclitic construction in one environment (e.g., after the word ἐπειδὴ) is significantly different from the score that has been assigned to the enclitic construction in that same environment. In other words, we are interested in the effect of the interaction between clitic position and syntactic environment. The first consequence of this design feature is that the by-item variation is fully covered by this interaction and so does not represent a random effect. Thus the maximal model to be run is: \( \text{model} = \text{Score} \sim \text{Environment} \times \text{Clitic} + (1 + \text{Environment} \times \text{Clitic}|\text{Subject}) \). The second consequence is that, since the \texttt{lme4} package uses one of the factors as the baseline reference to which all other factors are compared, the significance of all nine interactions cannot be tested with a single run of the model (Martin Wieling p.c., Alexander 2010). Instead, it is necessary to perform a separate analysis with each of the nine different environments as the baseline reference. Finally, given that the values entered for \textit{score} are skewed (as are the ones for \textit{line}), the results were logarithmically transformed so that parametric tests, such as mixed effect modeling, could be applied.

The attempt to model the data with the formula in (1) did not succeed, however, because the model did not converge, a common difficulty with large models (Martin Wieling, p.c.). The most extensive model that did converge was calculated using the formula in (2), where \((1|\text{Subject})\) introduces a random intercept for subjects, and \((0 + \text{Clitic}|\text{Subject})\) introduces a random slope for subjects but only taking into account the variability according to the position of the clitic.

\[
\begin{align*}
(1) \quad \text{model} & = \text{Scorelog} \sim \text{Environment} \times \text{Clitic} + (1 + \text{Environment} \times \text{Clitic}|\text{Subject}) \\
(2) \quad \text{model} & = \text{Scorelog} \sim \text{Environment} \times \text{Clitic} + (1|\text{Subject}) + (0 + \text{Clitic}|\text{Subject})
\end{align*}
\]

Table 2 provides a summary of the results. I present the mean score value for each clitic position in each environment, its log-transformation equivalent, and the \(t\)-value for the comparison between the log-values of the two means (pro-clitic vs. enclitic) in that syntactic environment. The significance of these values, which is noted in the last column, has been calculated using the \texttt{pvals.fnc} function (Baayen 2008). I have also taken into account a Bonferroni adjustment (cf. Gries 2013) since we are comparing results from ten separate runs of a model and not a single one. Thus, even though the \(p\)-value that is normally used as a

\[\text{7 } \text{The baseline factor is chosen alphabetically.}\]
threshold for significance is 0.05, here we divide it by 10 (the number of runs), so that $p = 0.005$.

I begin with the acceptability of temporal αφού, because of its unique status in the experiment. This is a construction that is not available in CG, but very common in SMG, and so its purpose is to indicate whether the participants accessed their CG grammars in performing the judgment task, or if they are influenced by the grammar of the standard. The results are encouraging, since we see that not only is the difference between the enclitic and proclitic position not significant, but also that this is the only case in which neither of the two scores approaches the modulus value of 50. Thus it appears that participants did indeed recognize the exceptional status of this stimulus and judged it accordingly. On the other hand, we do see some influence of the standard, since
the proclitic construction is given the higher score (42 vs. 32). This should be kept in mind as we interpret the rest of the results.\footnote{One reviewer notes that the influence may not be from the smg construction, but from other temporal conjunctions in cg, such as έμαν. Although this may be probable, it is not as likely, since, as we have seen, αφού does not have a temporal interpretation in cg.}

The results of the linear mixed effect modeling show that participants clearly prefer the enclitic placement of the pronoun in the following environments: after αφού, after έντζε and after a pre-verbal subject that does not carry focus intonation. On the other hand, they clearly prefer the proclitic placement of the pronoun after γιατί when it is used as a wh-word, and after a pre-verbal subject that carries information-focus intonation. When the preceding element is γιατί, ότι, or a subject that carries contrastive focus, then there is no preference (the $p$-values are well above 0.005). In the case of επειδή, however, we see a marginal result, as the enclitic construction is preferred but at the $p = 0.008$ level.

4 Discussion

Table 3 summarizes the results of four different types of studies with respect to these exceptional environments. Most of the conclusions listed in the last column are uncontroversial, but others need to be explained. While the function words αφού and έντζε are unattested in the records of Medieval Greek (Pappas 2004b), the enclitic pattern in the corpus of interviews (Pappas 2010) and in the current study is clear enough to allow us to speculate that this is the canonical pattern of cg. For ότι, we see a preference for enclisis in the Medieval Cypriot documents and in the interviews, but no preference in the elicitation studies or in the current study. This leaves us with two possible scenarios: Either the enclitic pattern of the interviews is the real one, and the met results are skewed due to the influence of the standard in a formal setting, or the no preference pattern is the real one and the interview results are skewed due to the small number of observations. At this point, unfortunately, it is not possible to decide between these two options.

The situation is different for the two causal conjunctions επειδή and γιατί, which require us to consider several factors. The two words have different histories of development (Babiniotis 1998). The former is a continuation of an Ancient Greek formation that was used as a temporal conjunction at first, and then developed a causal meaning, whereas the latter is a Medieval Greek
exceptional clitic placement in cypriot greek

Table 3 Comparison of clitic placement in Medieval Greek (Pappas 2004b), sociolinguistic interviews (Pappas 2011), elicitation studies (Agouraki 1997, 2001, 2009; Revithiadou 2006; Chatzikyriakidis 2010), and the present study.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Medieval Greek</th>
<th>Interviews</th>
<th>Elicitation</th>
<th>MET</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>αφού</td>
<td>unattested</td>
<td>enclisis</td>
<td>enclisis</td>
<td>enclisis</td>
<td>enclisis</td>
</tr>
<tr>
<td>έντζε</td>
<td>unattested</td>
<td>enclisis</td>
<td>enclisis</td>
<td>enclisis</td>
<td>enclisis</td>
</tr>
<tr>
<td>οτι</td>
<td>enclisis</td>
<td>enclisis</td>
<td>no preference</td>
<td>no preference</td>
<td>unclear</td>
</tr>
<tr>
<td>επειδή</td>
<td>enclisis</td>
<td>enclisis</td>
<td>no preference</td>
<td>enclisis (?)</td>
<td>enclisis</td>
</tr>
<tr>
<td>γιατί</td>
<td>unattested</td>
<td>enclisis</td>
<td>no preference</td>
<td>unclear</td>
<td>unclear</td>
</tr>
<tr>
<td>γιατί.wh</td>
<td>unattested</td>
<td>proclisis</td>
<td>proclisis</td>
<td>proclisis</td>
<td>proclisis</td>
</tr>
<tr>
<td>subj.contrastive</td>
<td>unclear</td>
<td>unclear</td>
<td>unclear</td>
<td>no preference</td>
<td>unclear</td>
</tr>
<tr>
<td>subj.information</td>
<td>unclear</td>
<td>unclear</td>
<td>unclear</td>
<td>proclisis</td>
<td>proclisis</td>
</tr>
<tr>
<td>subj.neutral</td>
<td>enclisis</td>
<td>enclisis</td>
<td>enclisis</td>
<td>enclisis</td>
<td>enclisis</td>
</tr>
</tbody>
</table>

formation from the words για and οτι. This etymological information, together with the absence of γιατί from Medieval Cypriot texts strongly suggest that this word must be a recent borrowing from SMG into Cypriot, especially since it is also used as a wh-word. Despite the fact that in SMG both are used with proclitic placement, we see that in Cypriot the causal γιατί has also acquired the enclitic pattern that characterizes επειδή, while wh-γιατί is associated with proclisis, as is the case with its Cypriot counterpart (ίντα που, cf. Pappas 2010, 2011). At the same time, it seems that these borrowings have also had an effect on the usage of επειδή, whose enclitic pattern is preferred but only marginally over the proclitic one. For these reasons, the pattern for επειδή is listed as “enclisis,” while for γιατί it is listed as “unclear.”

In the matter of fronted subject DPs, the pattern that is revealed in the MET study is also complex. Participants clearly prefer proclitics when the phrase has information focus, and enclitics when the phrase is neutral (focus on the verb), while they do not show a preference when the phrase carries contrastive focus. It is difficult to assess what the other datasets contribute to answering these questions, as the number of tokens is limited and it is not easy to classify utterances according to these criteria. According to Agouraki (2009) these constructions should all have received low evaluations. Chatzikyriakidis (2010) found that proclisis is allowed with pre-verbal subjects that carry contrastive focus, but, interestingly, he does not state that enclisis is prohibited. Pappas (2011) only examines whether the pre-verbal phrase is stressed or not, and finds that even when it is stressed, enclitic placement occurs 39% of the time. The MET
study results provide an explanation for this latter pattern, if we assume that the stressed elements in that study contain both instances of information and contrastive focus, and that in the latter case both enclisis and proclisis are equally acceptable. Nonetheless, it is clear that the results of this study raise more questions about the role of pre-verbal DPs in the pattern of clitic placement than they provide answers. In fact, the more general issue of the left periphery and information structure in Cypriot requires much deeper investigation.

The conclusions provided in the last column of Table 3 do not only provide a clearer description of clitic placement in CG, but also have important methodological, theoretical, and applied implications. The results of all the studies combined prove, beyond doubt, that the “troublesome” exceptions are real. The sociolinguistic interviews show that they are produced as part of natural conversations, and the MET results show that they influence the acceptability of an utterance at a highly significant level. Even more, for some of them, the historical record reveals that their exceptional status has remained more or less stable for well over 500 years. It is impossible, therefore, to dismiss these patterns as accidents of performance or the result of dialect mixing. They clearly constitute aspects of competence in Cypriot grammar. Accepting this has implications that go beyond the purposes of descriptive linguistics. For example, the Cyprus Acquisition Team (cf. Grohmann 2011) has an ambitious program of understanding how the pattern of clitic placement in Cypriot is acquired, given its clinical and pedagogical repercussions. An accurate description, then, of the pattern used by adults provides the necessary baseline for such applied considerations.

It is not within the scope of this paper to explain this pattern structurally. However, one sub-generalization that is rather obvious is that three of the five lexical exceptions are causal conjunctions (αφού, επειδή, γιατί) and ὅτι has strong connections to causality: It functioned as a causal conjunction as well as a complementizer in Ancient Greek (Jannaris 1968), and, more recently, it was involved in the formation of causal γιατί as seen above, as well as of its more archaizing counterpart διότι. A possible explanation can be given for the use of επειδή, which, according to standard grammars at least (Holton et al. 1997), should appear at the beginning of utterances with the main clause following, whereas γιατί clauses appear after the main clause. It may be then, that constructions with επειδή were originally (in Medieval Greek) a type of parataxis and not hypotaxis (in the sense of Culicover and Jackendoff 1997) which, in turn, would mean that επειδή would follow the pattern of coordinating conjunctions, i.e., enclisis. A similar position is stated by Jannaris (1968) and Mackridge (1993) about ὅτι in Medieval Greek, namely that it functions as a link between clauses and not as a complementizer. Chatzikyriakidis (2010, 2012)
notes that a similar pattern existed with the Medieval Spanish causal conjunction *ca*, for which it is has been argued, most recently by Bouzouita (2008), that it could introduce both subordinate and coordinated sentences. He argues that a similar explanation could hold for the Medieval Greek and Cypriot usages of ὅτι and επειδή, although the two cases receive slightly different treatment in his system. It is not clear, however, whether the exact nature of these structures can be uncovered through a more thorough examination of the Medieval texts.

Whether or not this is a correct hypothesis for Medieval Greek, it must be noted here that in CG as in colloquial SMG, the επειδή clause is commonly placed after the main clause, as in example (2). This means that even if these conjunctions were originally associated with a different syntactic structure, there is no evidence that this is still the case. Instead, the organizing principle behind these exceptions appears to be their semantic relationship, by which they form a *constellation*, in the sense of Joseph (1997: 158–159) and defined as “generalizations that are not wide-ranging ones but rather are localized or fragmented.” To extend the metaphor, the strength of the gravitational pull of επειδή is, undoubtedly, significant, if we judge from the pattern of causal γιατί. For we see that Cypriot speakers have incorporated γιατί as an alternative to επειδή, while retaining the enclitic pattern of the dialectal prototype as well as the proclitic pattern of the standard.9

(2) Δεν μπορεί να τους συλλάβει, επειδή

neq be.able.3SG.PRES comp them arrest.3SG.PRES because

αξιοποιούν τη σύγχρονη τεχνολογία

use.3PL.PRES the modern technology

‘[The government] is not able to arrest them because they use modern technology’ (Goutsos 2010: wopc16–1297)

For ἐντζε, much depends on its etymology, which is far from settled. This marker, which does not appear in written form until the late 19th century (Liapis 2007), has a mysterious provenance. Agouraki (2001) claims that this is a two-word cluster comprising the negative marker εν and the coordinating conjunction τʃαι, but the former belongs to a separate CP whose IP has been elided. Chatzikyriakidis (2010, 2012) treats it as a compound, which, despite

---

9 An anonymous reviewer correctly notes that other accounts are possible, such as one in which the neutral pattern seen with causal γιατί is explained as the balancing out of the tension between the pattern associated with its form (i.e., proclisis with wh-γιατί) and the pattern associated with its function (i.e., the enclisis with επειδή).
deriving its meaning and function from the first part, retains the enclitic pattern of the second. On the other hand, Chatziioannou (2010) claims that the word is a direct descendant from Medieval Greek ουκ, while Liapis (2007), expanding on an idea of Menardos (1969), argues that έντζε is the product of compounding between Cypriot εν and the Pontic Greek negative marker κε when Pontic Greeks were encouraged to move to Cyprus by land grants in the 15th century AD. None of these proposals is unassailable, but it is interesting that all of them recognize the enclitic placement that follows έντζε as an exceptional pattern that either merits exceptional treatment (as in Agouraki 2001) or as a distinctive characteristic that can be used to locate the word’s origin (as in Liapis 2007). In other words, this is a case in which the two aspects of this lexical item are so intertwined that the one cannot be explained without the other. In this respect, the situation with έντζε is more reminiscent of collocation patterns (as in Ke and Yao 2008) than a straightforward syntactic mechanism.

The results discussed here also have important methodological implications. Tagliamonte (2006) correctly advises that we should not investigate patterns that are near-categorical (e.g., with an 85–15 % distribution), because they typically do not convey social meaning. This does not mean, however, that marginal patterns should be ignored, because as this study has shown they may contain information that is important for understanding other aspects of grammar. In addition, the results from the met study provide us with some insight about the validity of such marginal patterns from sociolinguistic interviews: For items that are unique in the dialectal system (such as έντζε and αφού), the marginal pattern is clearly confirmed, but for items that are shared between the dialect and the standard (such as στι and γιατί) there is no clear outcome. In either case, however, they can form a solid foundation for further research. This truly highlights the importance of using participant observation methods for gathering dialectal data for higher level phenomena (i.e., morphosyntax and above), and especially in circumstances where there are several metalinguistic factors at play, as is the case in CG. Finally, this study shows that we can triangulate evidence from the historical record, free-flowing conversations and more deeply probing elicitation or experimental techniques in order to compensate for the disadvantages of each approach. This way we can effectively capture patterns that are rare, but nonetheless significant for our understanding of dialectal variation.
References


Jannaris, Antonius N. 1968. *An historical Greek grammar: Chiefly of the Attic dialect as written and spoken from classical antiquity down to the present time, founded upon the ancient texts, inscriptions, papyri and present popular Greek*. Hildesheim: G. Olms.


Appendix A. Test Sentences

(1) Αφού {τον} αγαπώ {τον} έννα τον πάρω
since him love.1sg.pres him fut him take.1sg.pres
‘Since I love him, I will marry him’

(2) Ἐντζέ {μου} δια {μου} ο κύρης μου το αυτοκίνητο
NEG to.me give.3sg.pres to.me the father my the car
‘My father doesn’t give me the car’

(3) Επειδή {με} επόφκαλέ {με}, έν είμαι με το
because me infuriate.3sg.pst me neg be.1sg.pres with the
Γιώργο
George
‘I am no longer with George because he infuriates me’

(4) Εν είμαι με το Γιώργο γιατί {με} επόφκαλε
neg be.1sg.pres with the George because me infuriate.3sg.pst
{με}
me
‘I am no longer with George because he infuriates me’

(5) Γιατί {τον} εκάλεσες {τον}, αφού ξέρεις ότι
why him invite.2sg.pst him since know.2sg.pres that
εμαλώσαμε;
quarrel.1pl.pst
‘Why did you invite him, since you know that we quarrelled?’

(6) Ωραία τσάντα! Εμαθα ότι {την} αγόρασες {την} που
nice bag learn.1sg.pst that her buy.2sg.pst her from
την Αγγλία
the England
‘Nice bag! I heard that you bought it in England’

(7) Ο Χάρης {με} έδει {με} στο κλαμπ, ό ο Γιάννης
the Haris me see.3sg.pst me at.the club no the Yanis
‘It was Haris that saw me in the club, not Yanis’
(8) Ωραίο ρολόι! Αρέσκει σου; Ο Πάμπος {το} nice watch be.pleasing.3SG.PRES to you the Pambos it
έπικασε {το} buy.3SG.PST it
'Α: Nice watch! B: You like it? It's Pambos that bought it'

(9) Ο Χάρης {με} είδε {με} στο κλαμπ αλλά ε μου the Haris me see.3SG.PST me at the club but neg to me
εμίλησε speak.3SG.PST
'Haris saw me at the club but he did not speak to me'

(10) Αφού {το} αγόρασα {το} έμαθα ότι τον άλλο μήνα έννα after it buy.1SG.PST it learn.1SG.PST that the next month fut have.3SG.PRES sales
έχει εκπτώσεις
'After I bought it, I found out that there will be a sale next month'

Appendix B. Training Sentences

(11) *Εννα που την πυρά φυρτώ fut from the fire faint.1SG.PRES
'It’s so hot, I am going to faint'

(12) Εψές τηγάνησα αγρέλια με τ’ αφκά yesterday fry.1SG.PST asparagus with the eggs
'Yesterday I made an asparagus omelet'

(13) Θα πάω στο Λιντλ να πάρω σφουγκαρίστρα fut go.1SG.PRES to the Lidl comp buy.1SG.PRES mop
'I am going to Lidl (market) to buy a mop' (With smg morphology and vocabulary)

(14) 'Εφερέ μου η αδερφή μου μια τσάντα που την Αγγλία, bring.3SG.PST me the sister my one handbag from the England
πελαμός crazy
'My sister brought me an awesome bag from England'