A rethink of the relationship between salience and anaphora resolution

Eleni Miltsakaki

University of Pennsylvania
3700 Walnut St.
Philadelphia, PA 19104, U.S.A.
elenimi@seas.upenn.edu

Abstract
According to prior work on the interpretation of anaphoric elements, the form of linguistic expression used for reference to a discourse entity reflects the entity’s degree of salience. When both fully specified and underspecified forms of reference are available the most underspecified forms are used for reference to the most salient entity. In this paper, we present sentence completion data from Greek, a language whose range of anaphoric expressions includes null subjects and cliticized object pronouns (weak pronouns) as well as demonstrative expressions (strong pronouns). The results of our studies cast doubt on the currently assumed mapping between salience and form of linguistic expression. We found that discourse salience does not warrant the use of the most underspecified form (null subject) in Greek. Semantic focusing (Stevenson et al., 2000) may bring into focus an entity in an object (or other non-subject) position but the choice of antecedent for null or overt forms is sensitive to syntactic properties of the antecedent. We argue that discourse salience and choice of referring expression can be sensitive to different properties of discourse entities which may or may not converge depending on the language.

1. Introduction
Prior work on anaphora resolution has resulted in several theoretical accounts, most of which map the form of a referring expression to the degree of salience of the referent. Givón (1983) associates a gradient scale of topicality with choice of linguistic expression, with the most topical entity predicted to be referenced with a null pronoun when permitted by the language. In a similar vein, Ariel’s (Ariel, 1990) Accessibility Marked Scale associates null forms and pronouns with entities that fall on the high end of the accessibility scale. In Gundel et al. (1993), the proposed Givenness Hierarchy associates entities at the in focus end of the hierarchy with pronominal reference. Centering theory (Groß et al., 1995) associates backward-looking centers with the use of pronominal reference. Existing accounts on the relationship between discourse salience and choice of referring expression do differ in what they view as relevant factors that establish entities as more or less salient but they uniformly assume that more salient entities are referenced with the most reduced form, in most cases a null subject or a pronoun, depending on the language.

We frame our inquiry in the relationship between salience and reference form with respect to two main factors that have been associated with entity salience. A common focusing factor is the structural position of an entity in a clause, i.e., subjects rank higher in salience/accessibility than objects or other non-subject entities evoked in the discourse. For convenience, we will use the term structural focusing as an umbrella term to refer to the observed tendency of subject referents to be more salient and therefore more likely to be referenced with a pronoun than non-subject referents. Stevenson et al. (1994), on the other hand, put forward a semantic focusing account according to which verbs project their own focusing preferences. The basic claim of semantic focusing is that the verb by virtue of its semantic interpretation will focus entities according to their thematic role. Specifically, action verbs focus the entity associated with the thematic role patient, independently of its syntactic realization. In other words, the entity realizing the patient role will be in focus even if syntactically it is realized as the object of the verb.

Following the experimental design in Stevenson et al. (1994), we report the results of two studies on the interpretation of null subjects and strong pronouns in subject position in Greek. Our data are supportive of Stevenson’s et al’s hypothesis that an action verb focuses the entity associated with the patient role. However, our data also suggest that null subjects in Greek disprefer to pick antecedents in object position. In a free continuation task, Greek speakers showed a tendency to continue the given discourse with reference to the entity with the patient role which was in an object position but they were reluctant to use a null subject for reference to the patient when it was evoked in object position. Instead, they used a strong pronoun to refer to the patient-object. When a null subject was used in the continuation, the null subject picked the subject referent which in our data was the entity realizing the agent role.

The rest of the paper is organized as follows. Section 2 offers a brief overview of the range of pronominal forms available in Greek and a review of prior work on their referential properties. Section 3 describes the two sentence completion studies that we conducted and reports the results of each study. Section 4 discusses the results of the studies and their implications for our understanding on the relationship between salience and form of reference. Section 5 summarizes our conclusions.

2. The Greek pronominal system
The pronominal system in Greek comprises weak and strong forms. Weak forms are null subjects and cliticized pronouns in object position. Strong forms are the demonstratives ‘aftos’ and ‘ekinos’ which are equivalent to ‘this’
and ‘that’ in English, only in Greek they are fully inflected for gender, number and case. Note that the definite article in Greek precedes both proper and common nouns and is also marked for gender, number and case. Null subjects pick the most salient antecedent in the prior discourse while the demonstratives ‘aftos’ and ‘ekinos’ pick the least salient antecedent, especially when exactly two antecedents are competing. In (1), the null subject must pick the subject ‘Yannis’ but the demonstrative must pick the object ‘Yorgo’. These observations have received empirical support in the works of Dimitriadis (1996) and Miltsakaki (2001).

(1) O Yannis-i proskalese ton
the(masc) John(nom) invited
Yorgo-j.
Yorgo(acc).
‘John invited George.’

a. null-i tu-j prosfere ena poto.
he(null) him(gen) offered a drink.
‘He offered him a drink.’

b. #null-j #tu-i prosfere ena poto.
he(null) him(gen) offered a drink.
‘He offered him-a drink.’

c. O Yorgos tu-i prosfere
the(masc) George(nom) him(gen) offered
ena poto.
a drink.
‘George offered him-a drink.’

d. Ekinos-j tu-i prosfere ena poto.
he(dem-masc) him(gen) offered a drink.
‘He offered him-a drink.’

Greek is also a free word language. To our knowledge, it is not known whether word order variation affects salience in Greek. Miltsakaki (2001) employs Rambow’s diagnostic (Rambow, 1993) to test if word order affects entity salience in Greek and finds that it does not. Applying Rambow’s diagnostic word order and grammatical function are contrasted in discourses (2) and (3). The discourse initial question in (2a) introduces two entities, i.e. prosfati diefthetisi and ikonomiki politiki in subject and object positions respectively. Both entities are ambiguous morphologically as they are both marked feminine. In (2b), the reply to the question contains a dropped subject and the predicate aneparkis is such that can take either prosfati diefthetisi or ikonomiki politiki as its subject. The test involves a native speaker’s interpretation of the dropped subject in the reply under two conditions. In the first condition, shown in discourse (2), the subject appears in the preverbal position. In the second condition, shown in discourse (3), the object has been fronted and the subject appears post-verbally. Independently of the surface positions of the subject and the object in the question, the null subject picks the subject referent, which is taken as an indication that word order does not affect entity salience in Greek. Note, however, that, to date, this observation has not been tested empirically.

(2) a. I prosfati diefthetisi-i tha veltiosi tin
the recent arrangement will improve the
ikonomiki politiki-j? economic policy?

‘Will the recent arrangement improve the economic policy?’

b. Ohi, (null-i) ine aneparkis.
No, (it) is inadequate.
‘No, it is inadequate.’

(3) a. Tin ikonomiki politiki-j tha ti-j veltiosi
the economic policy will CL-it improve
i prosfati diefthetisi-i?
the recent arrangement?

‘Will the recent arrangement improve the economic policy?’

b. Ohi, (null-i) ine aneparkis.
No, (it) is inadequate.
‘No, it is inadequate.’

3. Study 1: Interpretation of demonstratives

In this study, we investigate the interpretation of the strong pronoun ‘ekinos’ in Greek. In a sentence completion task, we introduce two competing antecedents, an agent in subject position and a patient in object position. To test the possible effect of word order the agent-subject appears both preverbally and postverbally.

One hundred native speakers of Greek participated in the study and their task was to complete written sentence fragments. The critical stimuli are SVO and OVS sentences with two masculine or feminine verb arguments. The masculine or feminine characters are evoked by role descriptions whose gender is clearly marked in the definite article that precedes the noun.1 To give an example, in the noun phrase ‘o giatros’ the form of the definite article is not known whether word order variation affects salience.

Each sentence, containing exactly one verb and its two arguments, is followed by the first word of the second sentence which is either the masculine demonstrative form ‘ekinos’ when the first sentence introduces masculine characters or the feminine demonstrative form ‘ekini’ when the first sentence introduces feminine characters.2 Thus there are four conditions: a) masculine, strong pronoun continuation following an SVO sentence, b) masculine, strong pronoun continuation following an OVS sentence, c) feminine, strong pronoun continuation following an SVO sentence, and d) feminine strong pronoun continuation following and OVS sentence. A

---

1. We opted for role NPs instead of individual names in order to minimize referent ambiguity in the participants’ continuations.
2. Throughout the paper we use the terms ‘strong pronoun’ and ‘demonstrative’ interchangeably.
sample set of critical stimuli is given below. For ease of reading, in the translation, we represent the Greek strong pronoun with a capitalized English pronoun, a convention usually followed to represent a stressed pronoun in English.

(4) O travmatiofores kuvalise ton the(masc) stretcher-bearer carried the(masc) astheni. Ekinos... patient. He(dem)... ‘The stretcher-bearer carried the patient. HE...’

(5) Ton astheni kuvalise o the(masc) patient carried the(masc) travmatiofores. Ekinos... stretcher-bearer. He (dem)... ‘The patient, the stretcher-bearer carried. HE...’

(6) I daskala agaliase ti mathitria. the(fem) teacher hugged the(fem) student. Ekin... She(dem)... ‘The teacher hugged the student. SHE...’

(7) ti mathitria agaliase i daskala. the(fem) student hugged the(fem) teacher. Ekin... She(dem). ... ‘The student, the teacher hugged. SHE...’

In the continuations, we coded the referent of the strong pronoun in all conditions. Ambiguous references (5%) were excluded from the analysis. The results reveal a massive preference, almost 95%, for the strong pronoun to pick the patient-object antecedent across all conditions (see Figure 1). If, indeed, the demonstrative picks the less salient entity, then these results counter the prediction made by the semantic focusing account. However, we found the extremely high percentage of reference to the patient-object puzzling. From prior work on the interpretation of pronouns, we know that empirical investigations on the interpretation of pronouns rarely yield categorical results. This is because there always seem to be hidden factors who prevent categorical results when testing any single factor. Because of this overwhelming preference to interpret the subject of the second sentence as the patient of the preceding discourse, we suspected that the patient-object could indeed be fairly salient but, for some reason, referring to the patient-object with a strong pronoun was natural. To evaluate if our suspicion held any ground, we conducted a second, free continuation study, which indeed confirmed our suspicion and revealed patterns of reference far more interesting than what we originally set out to explore. We report the results of the free continuation study in the next section.

The analysis of the SVO and OVS condition does not reveal any significant effect of word order, although a small trend can be seen: in the VSO condition reference to the agent-subject was a little higher (8% of the OVS items referred to the agent-subject) than in the SVO condition (3% of the SVO items referred to the agent-subject). This small difference might be attributed to a processing error. The non-canonical order OVS out of context is unnatural and unexpected so it is possible that in a quick reading the post verbal subject was processed as an object. We will not pursue this further as word order variation did not have any effect and there were only few cases referencing the agent-subject.

4. Study 2: Free continuation

The aim of the second study is to investigate which entity would be picked for reference in the continuation and what type of referring expression would be selected. This study has a similar design to Study 1. Study 2 differs in that the continuation after the main clause is free, i.e., we didn’t give the beginning of the next sentence. Free continuations are known to be problematic because they tend to yield high numbers of irrelevant data (e.g., continuations with no reference to the critical entities). We took the risk because it was crucial to give the participants the choice to freely pick a referent and a referring expression. As in Study 1, half of the critical stimuli appeared in SVO order and half in OVS order.

As expected, a substantial part of the data was removed from the analysis. A total of 185 items (35%) of the continuations either contained an ambiguous reference or contained no reference to the critical entities. Examples of both types are shown in (8) and (9), respectively.

(8) tin kori filise i mama 0 the(fem-acc) daughter kissed the(fem-nom) mother in gia na figi order 0 to leave ‘The daughter, the mother kissed so that she could leave.’

(9) I mitera filise tin kori sto the(fem-nom) kissed the(fem-acc) daughter on the metopo forehead ‘The mother kissed the daughter on the forehead’

In the legitimate continuations, we coded the referent of the subject in the continuation and the type of referring expression. In addition, taking into consideration prior work on the potential role of clause type in the choice of referring expression, we coded the type of clause in the continuation as main or subordinate. In subordinate clause continuations, we, also, coded the connective. The coding of the type of clause and type of connective is motivated by prior work of Miltsakaki (2002b) and Stevenson et al. (2000). Miltsakaki (2002b) claims that the traditional notion of sentence (i.e., the main clause and dependent subordinate clauses) defines a syntactic locality in which semantic focusing affects the interpretation of null subjects and demonstrative forms. In later work (Miltsakaki, 2002a), it was shown experimentally that, at least for English, a subject pronoun in subordinate clauses may pick either the main clause subject or the main clause object as its antecedent but intersententially a strong tendency is observed for a subject pronoun to pick the subject referent of the preceding clause. In Stevenson et al. (2000). Stevenson and her collaborators investigate the interactions between the semantic focusing properties of the verbs and their interaction with the semantics of connectives. They found that semantic focusing on the patient is reinforced when the discourse continues to establish
Figure 1: Object/patient as the antecedent of the ambiguous strong pronoun.

The results of the second study, Figure (2), show that in 52% of the data, the subject in the continuation referred to the patient-object and in the 48% of the data the subject in the continuation referred to the agent-subject. Given that the participants were free to continue the discourse in any way they found natural, the slight preference for reference to the object-patient supports the semantic focusing hypothesis by Stevenson and her collaborators, indicating that conceptually the patient was indeed salient despite being evoked in object position. If subjects were conceptually more salient than objects by virtue of their syntactic function, we would expect a clear bias for reference to the preceding subject. On the contrary, the continuations reveal that in more than half of the cases the object-patient was picked in the continuation in subject position.

Interestingly, looking closer at the distribution of type of referring expression we observe that, in main clause continuations, the most common referring expression for reference to the object-patient is the strong pronoun or noun phrase, (10), whereas in subordinate clause continuations the object-patient is referenced with a null subject, (12). The preferred referring expression for reference to the agent-subject is with a null subject both in main and subordinate clause continuations. Overall, the most common continuation (44% of all continuations) is with null reference to object-patient in a subordinate clause, (11).

The results of the coding of the connective in the subordinate clause continuation reveal a strong tendency for causal continuations, which according to Stevenson et al. (2000) enhance the salience of the object-patient. 47% of all subordinate clause continuations established a causal relation with the connectives ‘because’, ‘since’ and ‘when’. Another 18% of the subordinate clause continuations established a purpose relation with the connectives ‘so that’ and ‘in order to’ clause and another 11% were relative clause continuations. The remaining cases had fewer than 3 instances of a different connective such as ‘while’, ‘for as long as’, ‘before’ as well as a few free adjuncts. For the majority of the main clause continuation the coordinate conjunction ‘and’ was used to connect to the previous clause (89%). Other connectives included, ‘however’, ‘but’ and ‘as a result’. As in Study 1, word order variation showed no effects and will not be discussed any further.

5. Discussion

Taking the results of the two studies together, we observe the following. First, the strong pronoun in Study 1 is not used for a causal relation but weakened when the discourse continues with a narrative (e.g., Then...).

The semantic focusing hypothesis predicts that at the end of the first sentence the patient-object is in focus. This hypothesis, combined with Milsakaki’s hypothesis that semantic focusing tends to apply within the boundaries of a sentence, yields the prediction that, in subordinate clause continuations, there will be more references to the patient-object with a null subject. In contrast, in main clause continuations, Milsakaki’s hypothesis predicts that the preceding main clause subject is more salient and therefore reference to the patient-object requires a stronger form.

The results of the two studies, Figure (2), show that in 52% of the data, the subject in the continuation referred to the patient-object and in the 48% of the data the subject in the continuation referred to the agent-subject. Given that the participants were free to continue the discourse in any way they found natural, the slight preference for reference to the object-patient supports the semantic focusing hypothesis by Stevenson and her collaborators, indicating that conceptually the patient was indeed salient despite being evoked in object position. If subjects were conceptually more salient than objects by virtue of their syntactic function, we would expect a clear bias for reference to the preceding subject. On the contrary, the continuations reveal that in more than half of the cases the object-patient was picked in the continuation in subject position.

Interestingly, looking closer at the distribution of type of referring expression we observe that, in main clause continuations, the most common referring expression for reference to the object-patient is the strong pronoun or noun phrase, (10), whereas in subordinate clause continuations the object-patient is referenced with a null subject, (12). The preferred referring expression for reference to the agent-subject is with a null subject both in main and subordinate clause continuations. Overall, the most common continuation (44% of all continuations) is with null reference to object-patient in a subordinate clause, (11).

(10) O yios esproxe ton patera kai the(masc-nom) pushed the(masc-acc) father and the o pateras tu fonaxe father at-him(gen) yelled

‘The son pushed the father and the father yelled at him’

(11) I mitera filise tin kori epidi 0 perase tis the mother kissed the daughter because 0 passed the exetasis exams

‘The mother kissed the daughter because she passed the exams.’

(12) O amintikos klotsise ton epithetiko kai sti the defender kicked the attacker and in-the sinehia tu pire ti mpala.

continuation of-him took the ball

‘The defender kicked the forward and he, then, took the ball from him.’

The results of the coding of the connective in the subordinate clause continuation reveal a strong tendency for causal continuations, which according to Stevenson et al. (2000) enhance the salience of the object-patient. 47% of all subordinate clause continuations established a causal relation with the connectives ‘because’, ‘since’ and ‘when’. Another 18% of the subordinate clause continuations established a purpose relation with the connectives ‘so that’ and ‘in order to’ clause and another 11% were relative clause continuations. The remaining cases had fewer than 3 instances of a different connective such as ‘while’, ‘for as long as’, ‘before’ as well as a few free adjuncts. For the majority of the main clause continuation the coordinate conjunction ‘and’ was used to connect to the previous clause (89%). Other connectives included, ‘however’, ‘but’ and ‘as a result’. As in Study 1, word order variation showed no effects and will not be discussed any further.
reference to the less salient entity. If the patient-object was the less salient entity then we would expect to see limited references to it in the free continuation study. Contrary to this expectation, in Study 2 there was 52% reference to the patient-object, all in subject position. Secondly, the distinction between main and subordinate clauses significantly affects the choice of linguistic expression. Null subjects in main clause continuations preferred the subject antecedent while strong pronouns in main clause continuations preferred the object antecedent. This pattern was not attested in subordinate clause continuations.

So, what do we make of these results? Our findings clearly show that a mapping between choice of linguistic expression and entity salience is much more complicated than current theories suggest. If subjects are more salient than objects then the use of a null subject for reference to the object in the subordinate clause continuations is puzzling. If patients are more salient than agents then the use of a strong pronoun in the main clause continuations in both studies is also puzzling. We would like to put forward two possible explanations for our findings.

First, we would like to hypothesize that entity accessibility within the boundaries of a syntactic sentence is primarily determined by the semantic relations that are established by the predicates of the main and subordinate clauses. Note that (adverbial) subordinate clauses are always interpreted with respect to the main clause of the sentence.3 It is plausible, therefore, to propose that the semantics of the relations between the predicates of the main and subordinate clauses determine the interpretation of the anaphoric expressions. If this hypothesis is correct, we have an explanation about the low percentage of the use of a strong pronoun in a subordinate clause continuation for reference to either the agent-subject of patient object of the main clause (less than 1%). This line of argument is also consistent with Kehler’s (Kehler, 2002) theory of pronominal interpretation which, roughly, suggests that the interpretation of pronouns should fall out naturally from the semantic interpretation of the coherence relation established between clausal propositions. In the main clause continuations, on the other hand, our findings suggest that null subjects prefer the subject referent and strong pronouns prefer the object referent. Note that in the main clause continuations a new syntactic unit is started whose relation to the prior discourse cannot be predicted ahead of time. Put in Centering terms, we would like to suggest that when a new syntactic unit is processed, the null subject picks the subject referent and by doing so it signals topic continuation, assuming here that in our decontextualized data the subject referent is also the topical entity of the first sentence. The use of a strong pronoun of reference to the object of the preceding sentence is used to signal that a topic shift from the subject referent to the object referent of the preceding sentence. The semantic focusing account put forward by Stevenson and her collaborators may still be at play here in making the patient-object a more likely topic for the continuing discourse but it does not establish the patient-object as the new topic.

3Complement clauses are different in this respect because they form the argument of the matrix verb. The predicates of complement clauses are not interpreted with respect to the higher verb. The higher verb simply provides attribution information, i.e., it attributes the content of the complement clause to certain individuals. In ‘Mary said that John left’, the event of John’s leaving is not related to the event of Mary’s saying. The higher verb attributes the content of the complement clause to Mary. Also ‘Mary said’ is not ‘the main clause’. In isolation the clause is uninterpretable because it is missing one argument.
An alternative route of explanation is to consider the relation between parallel structures and choice of referring expression. As suggested by an anonymous reviewer, it is possible that the strong pronoun is used to counteract the parallelism effect, making the use of the Greek strong pronoun comparable with the use of a stressed pronoun in English in (13).

(13) John-i insulted Bill-j and then HE-j hit HIM-i.

This line of explanation is consistent with the data in the main clause continuations. In subordinate clause continuations a strong pronoun (or a stressed pronoun in English) is not required to counteract parallelism effect but it can be argued that subordinate clauses do not trigger parallelism effects because the clauses themselves are not syntactically parallel. To fully explore the relationship between parallelism effects, choice of referring expression and salience, further studies are required. Here we have only investigated the referential properties of null subjects and strong pronouns in subject position in Greek. Once we shift our attention to referential properties of weak clitic pronouns (used in object position in Greek), parallelism effects quickly lead to further complications. Kousta (2003) has shown experimentally that, in parallel structures, a Greek clitic pronoun (preceeding the verb) can effortlessly pick an object antecedent as can be seen in the conconexaion in (14). Note, though, that in (14) the introduction of a new subject referent in the second sentence already signals that the preceding subject is no longer in focus.

(14) O Yanis-i thavmaze ton the(masc-nom) admired the(masc-acc) Vasili.
Vasili-i. I Anna ton-i...
The(fem-nom) Anna him...
‘John-i admired Bill-j. Anna ... him-j.’

6. Summary
In this paper we investigated the referential properties of null subjects and strong pronouns (demonstratives) in Greek. Our findings challenge the commonly held assumption that choice of linguistic expression maps to degree of entity salience. Our data show that strong pronouns are used for reference to object antecedents even when these are predicted to be in focus. Conversely, null subjects showed a strong tendency to pick subject antecedents. We concluded that Greek pronouns are sensitive to the syntactic position of the antecedents.

The conclusions drawn here are also supported by studies in other languages which have a similar repertoire of referring expressions. Similar mismatches between degree of salience and direct mapping to anaphoric forms have been discussed in (Kaiser, 2003) for Finnish, in (Kaiser and Trueswell, 2004) for Dutch and more recently in (Alonso-Ovalle et al., 2002) for Spanish and (Mayol, 2006) for Catalan.

We argued that semantic focusing is not sufficient to warrant reference with a null subject. We suggested that reference to a subject referent with a null subject signals topic continuation. In cases in which an object referent is conceptually brought to focus, it must subsequently be referenced with a strong pronoun and this move signals a topic shift.

Acknowledgments
We would like thank Martha Lampropoulou for the invaluable help in collecting the sentence completion data. We are grateful to three anonymous reviewers for their constructive discussion of the present work.

7. References