

## When *pro*-Drop Languages Don't: On Overt Pronominal Subjects in Greek

Alexis Dimitriadis  
University of Pennsylvania  
*alexis@ling.upenn.edu*

### 1 Introduction

A speaker wishing to refer to a previously mentioned person or object must choose from the many types of referring and anaphoric expressions that any language provides; thus a speaker of English may have a choice between a full NP, a pronoun, and a demonstrative such as *this* or *that*. Speakers of languages that allow subject or object drop have the additional option of using a null pronoun.

The choice is generally said to be driven by the informational content of the candidate expressions, subject to some variant of the Gricean maxims of Quantity: the speaker must use an expression that is sufficiently informative for the hearer to determine its referent, and must not, for reasons of economy, use an expression that is more informative than required. This is said to explain why full NPs are used for non-salient entities, which are harder to identify, and reduced forms for the more salient entities.

But this process cannot in itself account for pronominal anaphora. For example, it cannot account for the interpretation of the following passage:

- (1) a. John<sub>i</sub> wanted to go fishing.  
b. He<sub>i</sub> called Bill<sub>j</sub> at 6am.  
c. He<sub>i</sub>/#he<sub>j</sub> spoke in a whisper.  
c'. Bill<sub>j</sub> spoke in a whisper.

Bill and John are at this point equally known to the speaker, and a singular masculine pronoun applies equally well to either of them. Yet the pronoun used in sentence (c) very strongly favors construal with the subject of the previous sentence, that is, with the antecedent *John*. Pronominal anaphora must take into account not just what is known about an entity, but also the manner in which the entity appeared in the earlier discourse.

Gundel et al. (1993) argue that the choice among referential forms is constrained by the Givenness Hierarchy of “cognitive statuses”:

in focus > activated > familiar > uniquely identifiable > referential > type identifiable

Their theory requires English pronouns to be in focus, full NPs to be (at least) uniquely identifiable, etc. By definition, membership to any status of the hierarchy

---

I would like to thank Ümit Turan and Ellen Prince for their invaluable help during the preparation of this paper.

entails membership to all lower statuses (but not vice versa), for example, a familiar entity is automatically uniquely identifiable, referential, and type identifiable. As a consequence, an expression that requires its antecedent to be referential is also usable for in-focus entities, etc.

Usage is said to be further constrained by the Gricean maxim of Quantity: using an expression that requires some particular cognitive status implies that an expression requiring a *higher* cognitive status would be inappropriate. Thus the form *that N*, which must refer to an entity that is familiar or higher, is not ordinarily used for an in-focus entity because its use implies an antecedent that is familiar, but not in-focus.<sup>1</sup>

The Cognitive Hierarchy is actually a mixed system: its lower half pertains to the informational status of an entity, while the top half pertains to the status of an entity in discourse. While it is clear that both factors are important to the determination of the appropriate anaphoric expression, it is less clear that the two ought to be conflated into one scale. Consider the following example:

- (2) a. I am going to get *a kitten*.  
 b. I will train *it* to fetch bottle caps and stuff.

Recall that membership in any status of the Givenness Hierarchy is supposed to entail membership in every lower status. Sentence (2a) must bring the entity *a kitten* in focus, since it can be referred to by the pronoun *it* in (2b), which requires an in-focus antecedent. But it is not clear that the entity in question, introduced by a non-specific indefinite, has the cognitive status “uniquely identifiable” or even “referential” in any useful sense.

The need to consider non-informational factors in pronominal use is particularly clear in the context of null-subject languages, which provide a three-way choice between full-NP, overt pronominal, and null subjects.

In Greek, a productive subject-drop language, a discourse-familiar subject can in principle be expressed by a null, by a full NP, or by the demonstrative *aftos* ‘this one’ or *ekinos* ‘that one’. The choice affects the preferred interpretation: In the following passage, the pro-drop subject of (c) strongly favors coreference with the subject of (b); while an overt pronominal subject as in (c’) must be construed with the object of (b).

- (3) a. O Giannis<sub>i</sub> ithele na pai gia psarema.  
 ‘John<sub>i</sub> wanted to go fishing.’  
 b.  $\phi_i$  pire telefono to Vasili<sub>j</sub> stis 6 to proi.  
 ‘He<sub>i</sub> called Bill<sub>j</sub> at 6am.’

---

<sup>1</sup>In fact, as the authors’ own data show, the expression *that N* is used with activated as well as familiar antecedents, even though English provides several expressions requiring activated antecedents, e.g., *this*, *that* and *that N*. I will return to this and other topics related to Gundel et al. (1993) in section 3.1.

- c.  $\phi_i/\#\phi_j$  milise psithirista.  
 ‘He<sub>i</sub>/#He<sub>j</sub> spoke in a whisper.’
- c'.  $\#Aftos_i/Aftos_j$  milise psithirista.  
 ‘#He<sub>i</sub>/He<sub>i</sub> spoke in a whisper.’

Note that in this case, Greek allows an overt pronoun in a context requiring a full NP in English. (Compare the judgements for sentences (1c) and (3c')).

Informational considerations cannot explain the choice between overt and null pronominal forms. First, the choice of pronominal affects interpretation even where an overt pronoun adds no information that can be used to distinguish between possible antecedents. For example, in the above passage the choice is between two third person singular masculine antecedents; yet an overt pronoun in (3c) selects one antecedent, and a null pronoun selects another.

I will show below that in Greek, the function of the overt pronoun *aftos* is to signal *by its presence*, and not by any featural or referential information that it provides, that for its antecedent we should take not the most “prominent” potential antecedent from the previous sentence, but something somewhat less prominent. That is, that we should “skip” the most obvious candidate and move down to the next one.

In the following sections, I will define the above notion of “prominence” within the framework of Centering Theory; specifically, I will argue that the type of pronominal used determines where in the list of *forward looking centers* its antecedent will appear: *Pro* must be construed with the “compatible” antecedent that is highest in the list (where compatibility is determined by factors like featural agreement and agentivity). The proximal demonstrative *aftos* ‘this one’ is (almost) never used to refer to the “preferred center”; it is used instead when the antecedent occupies a lower position in the list of centers (thus *aftos* may or may not be the backward looking center). Finally, the distal pronominal *ekinós* ‘that one’ is often used with an antecedent that appeared several sentences earlier.

## 1.1 Centering theory

Centering theory models potential antecedents at any point in the discourse as a list of *forward looking centers* (Cf-list for short). The centers of each utterance are the available antecedents for the next utterance. These centers appear in the Cf list in order of prominence. Their ranking is in principle predictable from syntactic considerations alone: centers are ranked according to some variation of the following hierarchy. The centers in the Cf list are ranked according to some hierarchy like the following one:

SUBJECT > OBJECT2 > OBJECT > OTHER > DISCOURSE UNIT

Actually it is not entirely clear what factors determine this ranking. Turan (1994) discusses a number of factors, including point of view and the thematic role of a subject

or object, that may affect the ranking.

The most prominent center is the *preferred center* or Cp; it is assumed to be the most likely to be talked about in the next sentence.

Each utterance has a distinguished *backward looking center* (Cb). This supposed to be the center of the current utterance that the speaker is most concerned with. It is *defined* as the highest-ranked center of the previous utterance that is also realized in the current utterance.

An example: The Cf list and Cb are marked in the following passage. (The Cp is always the first element of the Cf list).

- (4) a. Keli<sub>i</sub> went to listen to a band<sub>j</sub>.  
       [ Cf = (Keli, band). No Cb ]  
       b. She<sub>i</sub> struck up a conversation with the band<sub>j</sub>'s stage manager<sub>k</sub>.  
       [ Cf = (Keli, conversation, stage manager, band). Cb = Keli ]  
       c. He<sub>k</sub> gave her<sub>i</sub> some free passes for their<sub>j</sub> next show.  
       [ Cf = (manager, Keli, passes, show, band). Cb = Keli ]

Centering theory characterizes the structure of discourse in terms of the behavior of the Cb. The Cb of an utterance may or may not be the same as its Cp; it also may or may not be the same as the Cb of the *previous* utterance. These two factors determine the type of the *centering transition* from the previous to the current utterance, according to the following table.

	Cb(U <sub>n</sub> ) = Cb(U <sub>n-1</sub> )	Cb(U <sub>n</sub> ) ≠ Cb(U <sub>n-1</sub> )
	or no Cb(U <sub>n-1</sub> )	
Cb(U <sub>n</sub> ) = Cp(U <sub>n</sub> )	Continue (CT)	Smooth Shift (SS)
Cb(U <sub>n</sub> ) ≠ Cp(U <sub>n</sub> )	Retain (RT)	Rough Shift (RS)

Thus sentence (4b) above corresponds to a Continue transition, since its Cb is also the Cp, and there is no previous Cb. Sentence (4c) corresponds to a Retain transition, since its Cb does not change from sentence (b), but Cb(4c) ≠ Cp(4c).

The type of transition is said to affect the coherence of the text. Continue transitions give the most coherent discourse, followed in order by Retain, Smooth Shift, and Rough Shift. Rough Shifts are quite rare in natural discourse. Centering theory predicts that in interpreting anaphoric expressions, hearers prefer to assign antecedents in a way that maximizes the coherence of the discourse.

The centering framework has relatively little to say about the acceptability of pronouns in discourse. The *Pronoun Rule* of Grosz et al. (1986) states that if an utterance U<sub>n</sub> contains any pronouns realizing a center of the previous utterance, then the Cb of U<sub>n</sub> must also be realized by a pronoun.

The Pronoun Rule has been extended and generalized since it was formulated as above. For example, Turan (1994) assumes the following as part of the “rules of centering theory”:

- (5) The appropriate use of a referential expression depends on the degree of salience which the antecedent in the Cf list is assigned by the speaker. Thus:
- a. a reduced expression (i.e., a null/unstressed pronoun depending on the availability of such forms in a language) is reserved for the most salient entity,
  - b. the most explicit expression, i.e., full NP is reserved for a less salient entity.

This description of centering is of necessity brief and incomplete. The standard reference on centering theory is Grosz et al. (1986). The reader is also referred to Walker et al. (1994) for a recent, particularly informative exposition.

## 2 A Corpus Study of Greek Pronouns

This section presents quantitative results establishing that Greek pronominal antecedents are selected on the basis of their position in the Cf list. I collected examples of the use of overt and null pronominal subjects from a corpus of four Greek texts drawn from the Greek part of the European Corpus Initiative (ECI) corpus. Most of the texts were novels written in Greek or translated to Greek. One was a set of lecture notes from a university-level electronics course.

From each text, long contiguous tracts were analyzed, and from each such tract, all sentences meeting the selection criteria were collected. Overt subject pronominals are relatively rare, so for reasons of technical convenience, only instances of sentence initial overt pronominals were collected. Since null subjects are much more frequent, it was possible to accumulate a comparable number of tokens through searching much smaller tracts of text.

I collected three types of sentences: null subject sentences, sentences whose subject was the proform *aftos* ‘this one’, and sentences whose subject was the proform *ekinos* ‘that one’.<sup>2</sup>

Since the construction of interest was the free alternation between null subjects and overt pronouns, I collected only types of constructions which, considered in isolation, would have allowed a null subject in place of the overt pronoun. Thus I discarded all uses of *aftos* as part of a larger phrase, e.g., as the head of a relative clause, or when it is obviously a stressed pronoun.

For each sentence, I used my own intuitions to assign the reference of anaphoric expressions; then I computed the relevant centers and transition types according to the Centering Algorithm as described in Walker et al. (1994).

Discourse anaphora received special treatment: the antecedent of a discourse anaphor was assumed to be in the Cf list, but ranked lower than any overt center. Null and overt subjects were not distinguished for ranking purposes.

---

<sup>2</sup>For reasons that will be discussed in section 2.3, I only collected null subject sentences from passages in which there was no first or second person participant, hence all tabulated instances of *pro* and all its potential antecedents are in the third person.

Table 1: Centering transition types for *aftos*

CT:	11	SS:	46
RT:	19	RS:	2

Plus 4 w/o a Cb

(There was no prior Cb in 6 CT and 1 RT examples)

Table 2: Antecedent of *aftos*, by transition type

	CT	SS	RT	RS	(none)	Total
Cp( $U_{n-1}$ )	0	1	–	–	–	1
other in $U_{n-1}$	11	45	19	2	–	77
other	–	–	0	0	4	4

## 2.1 The pronominal *aftos* ‘this one’

Table 1 shows the distribution, by centering transition type, of sentences beginning with the pronominal subject *aftos*. While the majority of transitions in ordinary text tend to be of type Continue, with smaller numbers of type Retain and Smooth Shift and a few Rough Shifts, this sample consists primarily of Smooth Shifts, with fewer Retains and even fewer Continue transitions. That is, we find a strong, but by no means categorical, dispreference for the use of *aftos* in conjunction with a Continue transition.

Indeed, table 1 gives no evidence of a categorical restriction in operation: *Aftos* is not prohibited from referring to the Cb of the previous utterance: it does so in all the Retain cases. Similarly we see that in all the Smooth Shift cases, *aftos* is the Cb of the *current* utterance; and because *aftos* is the subject, it is always the Cp of the current utterance.

In fact *aftos* does conform to a categorical restriction on its use, namely, as can be seen in table 2, it is almost never construed with the Cp of the previous utterance. Due to the design of the study, only sentences of type Continue and Smooth Shift could have had the Cp of the previous utterance as the antecedent of *aftos*.<sup>3</sup> In table 2, a dash is used to indicate combinations that are *a priori* impossible; unattested but *a priori* possible combinations are indicated with the digit zero.

It can be seen that of the fifty-six utterances that could have had Cp( $U_{n-1}$ ) as the antecedent of *aftos*, only one actually did so. This is an effect much stronger than the tendency to avoid Continue transitions, observed in table 1; I will take it to be characteristic of the subject pronominal *aftos* (indeed, of all overt subject pronominals in Greek), and express it as follows:

- (6) **The Overt Pronoun Rule:** An overt pronominal subject in Greek should not be construed with the Cp of the previous utterance.

The paucity of Continue transitions in the distribution of *aftos* can be seen to be a consequence of the Overt Pronoun Rule. In particular, the Overt Pronoun Rule implies that *aftos* can only appear in a Continue transition if the preceding transition

<sup>3</sup>The reason is as follows: If *aftos* picked out Cp( $U_{n-1}$ ), it would be the Cb of utterance  $U_n$ . As a sentence-initial subject, *aftos* is also the Cp of utterance  $U_n$ ; we would then have Cb( $U_n$ ) = Cp( $U_n$ ), and the centering state must be Continue or Smooth Shift.

Table 3: Centering transition types for *ekinos*

CT: 10	SS: 12
RT: 7	RS: 1
Plus 10 w/o a Cb	

Table 4: Antecedent of *ekinos*, by transition type

	CT	SS	RT	RS	(none)	Total
Cp( $U_{n-1}$ )	2	(3)	–	–	–	5
other in $U_{n-1}$	8	9	1	0	–	18
other	–	–	6	1	10	17

was a Retain or Rough Shift;<sup>4</sup> since Retain and Rough Shift transitions are not very common, most Continue transitions follow another Continue or a Smooth Shift, and usage of *aftos* with Continue is necessarily rare. The following passage is an example of such a situation.

- (7) a. Mia mera pernousē ap' to scholio i Elenitsa<sub>i</sub>, opos sichna-pikna to sinithize.  
 'One day Eleni<sub>i</sub> came by the school as she often did.'  
 b. O kenourios<sub>j</sub> tin<sub>i</sub> ide proti fora ke ksipastike.  
 'The new guy<sub>j</sub> saw her<sub>i</sub> for the first time and flipped.'  
 c. φ<sub>j</sub> Etrekse ksopiso tis<sub>i</sub> ke prospathise na tis<sub>i</sub> milisi: "Despinis..."  
 He<sub>j</sub> ran behind her<sub>i</sub> and tried to talk to her<sub>i</sub>: "Miss..."  
 [ Cb = he<sub>j</sub>, Cf = (he<sub>j</sub>, she<sub>i</sub>) ]  
 d. I Elenitsa<sub>i</sub> oute pou girise na ton<sub>j</sub> di.  
 'Eleni<sub>i</sub> didn't even turn to look at him<sub>j</sub>.'  
 [ Cb = he<sub>j</sub>, Cf = (she<sub>i</sub>, he<sub>j</sub>) : Retain ]  
 e. **Aftos**<sub>j</sub> epemene.  
 'He<sub>j</sub> insisted.'  
 [ Cb = he<sub>j</sub> : Continue ]

## 2.2 The pronominal *ekinos* 'that one'

The pronominal *ekinos* has a slightly more distal flavor compared to *aftos*. But excepting uses as the head of relative constructions and the like, the two pronominals are basically interchangeable in isolated sentences. Tables 3 and 4 show the distribution of uses by centering transition and by location of antecedent. Table 3 shows Continue transitions to be relatively rare, much as in the case of *aftos*. Table 4 shows that *ekinos*, like *aftos*, tends to obey the Overt Pronoun Rule.

The main difference between *aftos* and *ekinos* is in the bottom row of tables 2 and 4. While *aftos* overwhelmingly finds its antecedent in the preceding sentence, almost half (17 out of 40) of the uses of *ekinos* involve an antecedent that appeared

<sup>4</sup>Since *aftos* is automatically the Cp when it appears sentence-initially, and in a Continue transition the current Cp is also the Cb, the antecedent of *aftos* in a Continue transition must be the Cb of the previous sentence. But by the Overt Pronoun Rule, the antecedent of *aftos* may not be Cp of the previous sentence. It is only possible to satisfy both of these requirements when the previous sentence has different Cp and Cb, that is, when the preceding transition is a Retain or Rough Shift. Thus *aftos* can only appear with Continue transitions that follow a Retain or Rough Shift.

more than one sentence earlier. Thus *ekinos* appears to be specialized for long distance reference. In the following example, *ekini* does not have an antecedent in the previous sentence; in fact centering theory assigns **no** Cb to sentence (d).

- (8) a.  $\phi_i$  Eprepe na tis<sub>j</sub> eksigisi pos  $\phi_i$  eftase ke  $\phi_i$  kseskise tin kardia tis<sub>j</sub>.  
 ‘He<sub>i</sub> ought to explain to her<sub>j</sub> how he<sub>i</sub> got to the point where he<sub>i</sub> tore up her<sub>j</sub> card.’
- b. Ta logia den erchondan sto stoma tou<sub>i</sub>.  
 ‘The words were not coming to his<sub>i</sub> mouth.’
- c. Ta chilia tou<sub>i</sub> salevan choris na vgeni o paramikros ichos.  
 ‘His<sub>i</sub> lips moved without the smallest sound coming out.’
- d. **Ekini**<sub>j</sub> chtipise me nevro ti skoupa sto katofli tis.  
 ‘She<sub>j</sub> rapped with agitation the broom on her<sub>j</sub> doorstep.’

The number appearing in parentheses in table 4 (in the Smooth Shift column) indicates that three sentences coded as exceptions to the Overt Pronoun Rule contain a center that is a proper subset of the center explicitly appearing in the previous sentence. In the following example, *ekini<sub>i</sub>* refers to one of the characters collectively referred to by the null  $\phi_{i,j}$  (they) in the previous sentence:

- (9) a. Ke pia  $\phi_{i,j}$  den periorizondan mono sto stithos tis<sub>i</sub> ke stis roges tis<sub>i</sub>.  
 ‘And they<sub>i,j</sub> longer restricted themselves to her<sub>i</sub> breasts and nipples.’
- b. Molis ekline to porti tou katogiou piso tous<sub>i,j</sub>,  $\phi_{i,j}$  gimnonondan ki i dio olotela.  
 ‘As soon as the cellar door closed behind them<sub>i,j</sub>, they<sub>i,j</sub> would both strip completely.’  
 [ Cb = both ]
- c. **Ekini**<sub>i</sub> ksaplone s’ ena palio achirenio stroma, ki amesos tou fonaze:  
 ‘She<sub>i</sub> would lay down on an old straw mattress, and immediately say to him:’  
 [ Cb = She : Smooth Shift (subset) ]

In the terminology of Grosz et al. (1986), *ekini<sub>i</sub>* refers to an entity that is realized, but not directly realized, in the previous utterance. The relevance of this distinction to anaphoric accessibility does not appear to have been investigated within the centering literature. (But see Hirschberg (1985) for an analysis of subset relationships, Ward and Prince (1991) for their effects on topicalization).

If it can be demonstrated that an entity that is not realized directly is not as accessible as one directly realized, we can consider *ekinos* to be picking out, in the parenthesized cases also, an antecedent that is not at the very top of the Cf list; under this interpretation these sentences are not exceptions to the Overt Pronoun Rule. In fact, just such an accessibility effect is well known, and can be easily seen in English. In the following example (suggested by Ellen Prince), the entities *John<sub>i</sub>* and *Mary* cannot be the antecedents of an unstressed pronoun, even though no ambiguity would result since they are of different gender. (Accordingly, they can be referred to by a *stressed* pronoun, as in (10b')).

- (10) a. John<sub>i</sub> and Mary<sub>j</sub> are twins.  
 b. #She<sub>j</sub> likes linguistics, but he<sub>i</sub> does not.  
 b'. **She**<sub>j</sub> likes linguistics, but **he**<sub>i</sub> does not.  
 b''. They<sub>i,j</sub> love linguistics.

Sentence (10a) shows that *John and Mary*, as an aggregate, can be referred to by an unstressed pronoun, just like simple subjects in that position. Moreover, *John* and *Mary* above are less accessible than object NPs; this can be seen by the fact that the preferred interpretation of the subject *she* in (11b) is *Carmen<sub>k</sub>*, not *Mary<sub>j</sub>*.

- (11) a. John<sub>i</sub> and Mary<sub>j</sub> went to see Carmen<sub>k</sub>.  
 b. She<sub>k/#j</sub> was wearing jeans.

The question here is how to represent, within centering theory, the accessibility pattern of complex and set antecedents. Ideally the behavior sketched above should be a consequence of the way subset entities are ranked in the Cf list. While a detailed treatment is beyond the scope of this paper, note that the behavior of the above examples can be captured by ranking subset entities in the Cf list below all overtly realized entities.

Assuming then that subset entities appear lower in the Cf list than the collective expression that describes them, we can conclude that the three parenthesized sentences of table 4 are not exceptions to the Overt Pronoun Rule.

### 2.3 The null subject *pro*

Null subjects form the great majority of pronominal subjects in Greek, and not surprisingly, their use is harder to characterize than that of the overt pronouns. Since overt pronominal subjects are as we saw incompatible with a preceding Cp, ideally we would expect the null subject *pro* to always select the Cp of the previous sentence as its antecedent. But although, as I will show, *pro* does by and large tend to take the previous Cp as its antecedent, exceptions occur with frequency much greater than seen with the overt pronominals.

The most numerous class of such exceptions involves first and second person participants. (Greek verbs agree in number and person with their subject). In passage (12), sentence (b) has a null subject that is construed with the character Artemis, who was neither the Cp nor the Cb of the previous utterance. The next sentence contains a null subject referring to the narrator, who is the Cb of (b) but not its Cp. Both sentences are exceptions to the prediction that *pro* should be construed with the Cp of the preceding sentence.

- (12) a. Etsi to<sub>k</sub> akousa ki ego<sub>i</sub> ap' ta chilia tis Artemis<sub>j</sub> ekino to magiko apogevma.  
 'That's how I<sub>i</sub> heard about it<sub>k</sub> from Artemis'<sub>j</sub> lips that magical evening'  
 [ Cb = it<sub>k</sub>, Cf = (me, it<sub>k</sub>, lips, Artemis, evening) ]  
 b.  $\phi$ <sub>j</sub> M'<sub>i</sub> iche pari ap' to cheri, ke  $\phi$ <sub>j</sub> me<sub>i</sub> travikse ksopiso tis; n' anevoume sto

lofo pano ap' ti mikri mas poli.

'(She<sub>j</sub>) had taken me<sub>i</sub> by the hand, and (she<sub>j</sub>) pulled me<sub>i</sub> after her<sub>j</sub> to climb the hill above our little town.'

[ Cb = me, (Cf = she, me, hand, hill, town) ]

- c.  $\phi_i$  Alafiasa ap' ta prota vimata, eno i kardia mou chtipouse isame pou me ksekoufane.

(I<sub>i</sub>) was agitated from the first steps, while my heart beat as if it would drive me deaf.

[ Cb = I : Continue ]

Fortunately, this behavior is not peculiar to Greek, and has been addressed in work on other languages. Turan (1994) notes (after Di Eugenio (1990)) that *pro* can have a non-subject as its antecedent as long as the subject of the previous utterance does not match the person or number features of *pro*. In the following example (adapted from Turan (1994)), either the overt *aftos* or the null *pro* can be used in sentence (13b) to refer to the object of the previous sentence; but when the higher-ranked subject does not contrast in number, as in (14), the overt pronoun is required.

- (13) a. O Achmet ke o Murat kalesan ton Ali<sub>i</sub> gia fagito.

'Ahmet and Murat invited Ali to dinner.'

- b. (Alla)  $\phi_i$ /aftos<sub>i</sub> den borouse na pai giati  $\phi_i$  iche doulia.

(But) (he<sub>i</sub>) could not go because (he<sub>i</sub>) was working.

- (14) a. O Achmet kalese ton Ali<sub>i</sub> gia fagito.

'Ahmet invited Ali to dinner.'

- b. (Alla) # $\phi_i$ /aftos<sub>i</sub> den borouse na pai giati  $\phi_i$  iche doulia.

(But) # $\phi_i$ /he<sub>i</sub> could not go because (he<sub>i</sub>) was working.

Turan (1994) adopts a conclusion that claims, essentially, that as long as the antecedent can be uniquely identified by its  $\varphi$ -features, pronominal anaphora is allowed regardless of the position of the antecedent on the Cf list.

Given the self-evident fact that pronouns cannot be construed with antecedents that have incompatible  $\varphi$ -features,<sup>5</sup> and the findings of Di Eugenio (1990) and Turan (1994), I will assume the following working hypothesis:

(15)  **$\varphi$ -invisibility hypothesis:**

In selecting an antecedent, pronominals ignore potential antecedents with incompatible  $\varphi$ -features.<sup>6</sup>

---

<sup>5</sup>An obvious qualification is necessary: a pronoun may be allowed, or required, to carry gender or number features that do not match those of the expression that introduced its antecedent.

- (i) a. I am going to dinner with my family<sub>i</sub>  
b. They<sub>i</sub> are waiting for me outside.

This kind of quirk is unproblematic for centering theory, which stresses the distinction between entities (centers) and the expressions that introduce them.

<sup>6</sup>It can be assumed for the time being that the relevant  $\varphi$ -features are those carried by the pronominal, if overt, and by the subject agreement on the verb. (But see below).

Table 5: Centering transition types for third-person *pro*

CT: 87	SS: 8
RT: 7	RS: 1
Plus 8 w/o a Cb	

Table 6: Antecedent of third-person *pro*, by transition type

	CT	SS	RT	RS	(none)	Total
Cp( $U_{n-1}$ )	79	6	4	0	-	89
other in $U_{n-1}$	7	2	1	1	0	11
other	1	0	2	0	8	11

In view of this assumption, the hypothesis that *pro* should pick out the Cp (that is, the highest element on the Cf list) of the previous utterance must be modified as follows:

(16) ***Pro* anaphora proposal:**

*Pro* is construed with the highest antecedent in the Cf list that has compatible features with it.

Under this proposal, the selection of antecedents for *pro* is based on position in the Cf list, as proposed above for the overt pronominals. Centers with incompatible  $\varphi$ -features are simply skipped over by the selection process.

At this point it is necessary to examine the effects of the  $\varphi$ -invisibility hypothesis on what has already been claimed about the selection of antecedents to overt pronominals. Should the Overt Pronoun Rule be modified to say that overt pronouns never select the most highly ranked *compatible* antecedent? It turns out that such a modification would be empirically incorrect. Overt subjects can select the second highest element of the Cf list even when the top element has incompatible  $\varphi$ -features. Consider again passage (7); the masculine subject *aftos* in sentence (e) is construed with the antecedent *he<sub>j</sub>*, occupying second place in the Cf list of sentence (d), even though *he<sub>j</sub>* is the highest-ranked masculine center. (The antecedent *Elenitsa<sub>i</sub>* would have required the feminine form *afti*). Similarly, the singular subject *aftos* in sentence (13b) is acceptable, although its antecedent is the highest-ranked singular center in the previous sentence. Thus the Overt Pronoun Rule must continue to rule out the Cp as an antecedent for overt pronouns regardless of  $\varphi$ -features.

Having thus refined our expectations of the behavior of null subjects, we can put them to the test with some quantitative data. In order to eliminate the distracting effects of person feature mismatch, I collected samples of null subject sentences only from passages that did not have a first or second person participant. The results, shown in tables 5 and 6, are still not as clear-cut as those for the overt pronouns. Although there is a strong tendency for third person *pro* to be construed with the Cp of the previous utterance, there is still a significant number of exceptions, few of which can be explained by a mismatch of agreement features.

In a number of the exceptional cases, however, another type of incompatibility seems to be involved; when *pro* is the subject of a verb that selects for a sentient or agentive subject, it appears to skip over high-ranked centers that are non-agentive. Two such examples are provided here. In passage (17), sentence (b) includes an inan-

imate *pro* subject referring to the cabin of (a), but no animate centers; the agentive null subject of sentence (c) skips over the entire Cf list of sentence (b) and takes its antecedent from sentence (a), the closest source of an agentive antecedent.

- (17) a. I kambina<sub>i</sub> tou<sub>j</sub> vriskotan sti gefira V, sti deksia plevra tou pliou.  
His<sub>j</sub> cabin<sub>i</sub> was on bridge 5, on the right side of the ship.
- b. Proorismeni gia dio atoma  $\phi_i$  itan poli evrichori ke iche idiotiko banio.  
Intended for two people, (it<sub>i</sub>) was very spacious and had a private bathroom.  
[ Cf = (cabin, bathroom) ]
- c.  $\phi_j$  Dieschise enan steno diadromo, estripse s' enan allo ke meta anikse tin porta aristera.  
(He<sub>j</sub>) crossed a narrow hallway, turned into another and then opened the door to the left.

In the following passage, sentence (18b) contains *two* null subjects; one of them ( $\phi_i$ ) is construed with the inanimate Cp of the preceding sentence; the other ( $\phi_j$ ), which requires an animate antecedent, skips down the Cf list and selects the animate antecedent *Max*. It might be added that this passage is not at all difficult to process.

- (18) a. Ostoso to telefteo onoma<sub>i</sub> itan afto pou travikse tin prosochi tou Max<sub>j</sub>.  
However, the last name<sub>i</sub> was the one that attracted Max<sub>j</sub>'s attention.  
[ Cf = (name, attention, Max) ]
- b.  $\phi_j$  Skeftotan pos  $\phi_i$  teriaze apolita sti gineka pou to efere.  
(He<sub>j</sub>) thought that (it<sub>i</sub>) fit perfectly the woman that carried it.

Given the behavior of examples like the above, I will conclude that agentivity, sentience or something like that is among the features that are subject to the  $\varphi$ -invisibility rule (15); that is, that non-animate (or whatever) centers are ignored if a *pro* looking for an antecedent should be animate (or whatever). Assuming (necessarily without evidence) that this factor is also taken into account in the computation of antecedents to *overt* pronominal subjects<sup>7</sup>, we can formulate the following rule, which should properly be considered a clarification of the  $\varphi$ -invisibility hypothesis (15).

- (19) **Agentivity rule:** An antecedent for a pronominal subject must match it in terms of agentivity as well as number and person.

A number of the remaining exceptions can be blamed on what Turan (1994) describes as "discourse point of view." Turan claims that in contexts involving changes in point of view, signalled by verbs of perception like *see* or verbs that represent inner thought, like *remember*, nonsubjects can appear to be ranked higher than subjects. For example, in the following passage the pronoun *he<sub>j</sub>* in sentence (c) refers to

<sup>7</sup>While there is no question that the antecedent of an anaphoric expression should have the appropriate properties with respect to agentivity, animacy etc., it does not follow that the process of antecedent selection is itself sensitive to these factors.

The Overt Pronoun Rule as stated only makes negative predictions; in the absence of a positive theory of antecedent selection for overt pronominals, the set of  $\varphi$ -features that are relevant to the selection process cannot be determined.

Charles, despite the tendency of English pronouns in such environments to be construed with the Cb (or Cp?) of the previous sentence (cf. the discussion of passage (1)).

- (20) a. “As you wish”, said Dr. Wiley<sub>i</sub>.  
 b. He<sub>i</sub> remembered Charles<sub>j</sub> as a rash but thorough student.  
 c. Apparently he<sub>j</sub> hadn't changed.  
 d. Although Dr. Wiley<sub>i</sub> knew that...  
 (Robin Cook, 1982. *Fever*, p. 81. Cited in Turan (1994)).

Turan (1994) derives from this the following rule:

(21) **(Discourse) Point of View [POV] rule:**

In a discourse segment with a subjective point of view, entities are ranked through the represented mind of a subject of consciousness.

It must be stressed that while all the processes discussed so far concern the selection of antecedents from a fixed Cf list, the POV rule affects the construction of the Cf list itself. Unfortunately the POV rule, as given, is rather vague. It appears to advocate, for the class of “subjective” discourse segments, abandoning the syntactically-based ranking of the Cf list in favor of some subjective way of ranking centers. But the claim to a well-defined way of ranking centers is one of the strengths of centering theory. How exactly ought the subjective ranking be done? It is generally granted that discourse participants (including the owner of the point of view) can be referred to as if they are present in the Cf list. But in other cases, including passage (20), the Cf list is interpreted as if the owner of the point of view is *absent* from the Cf list.<sup>8</sup> Thus at this point in its development, the Point of View explanation must remain a “garbage can” category that can accommodate cases for which no more constrained account can be found.

Where does all this leave the hypothesis that *pro* (tries to) pick out the top of the Cf list as its antecedent? The applicability to the null subject data of the various refinements proposed above is indicated in table 7. A good fraction of the exceptions can be explained by extending the rule of  $\varphi$ -invisibility to agentivity and the like; two more cases can be dismissed by taking appositive relative clauses to constitute separate utterances; but eleven cases remain that must be consigned to the POV proposal, which is broad enough that almost any exception can be accommodated by it.

In summary, the behavior of *pro* is more complicated than that of the overt pronominals. Much of the explanation offered in this section is speculative, and needs to be backed by future quantitative work. But it appears plausible that the hypothesis can be maintained that *pro*, like the overt pronominals, selects its antecedent on the basis of its position in the Cf list.

---

<sup>8</sup>One might also wonder how the owner of the point of view could ever fail to rank themselves as the most important entity in their own consciousness!

Table 7: Non-Cp antecedents of *pro*, by excuse

	antecedent type	
	in $U_{n-1}$	other
Phi-features	3	0
Agentivity (rel clause)	3	3
Point of View	2	0
	3	8

### 3 Conclusions

I have argued that the subject pronouns of Greek select their antecedent on the basis of its position on the Cf list. In particular, they have differing preferences for the part of the Cf list from which they draw their antecedent:

Null subjects appear to take as their antecedent the highest-ranked center that has compatible grammatical and agentivity features.

*Aftos* is normally incompatible with the Cp of the previous sentence, and should be construed with lower-ranked centers.

*Ekinos* is also incompatible with the previous Cp; it frequently occurs as a long-distance anaphor with no antecedent in the previous sentence.

It should be noted that these conditions are not mutually exclusive: it is not rare for more than one pronominal to be able to access the intended antecedent; for example, when *pro* is not construed with the Cp because of incompatible  $\varphi$ -features, both *pro* and the overt pronoun *aftos* are possible. (Passage (13) is such an example).

A number of apparent exceptions to these generalizations were argued to be due to complicating factors. In section 2.2 I suggested that centers that are (realized, but) not directly realized in the Cf list of the preceding sentence should be considered to be ranked lower than the aggregate center that includes them; this would account for most of the apparent exceptions to the distribution of *ekinos*. In section 2.3 I reviewed the role of  $\varphi$ -features and point of view to null pronoun anaphora, as discussed by Turan (1995), and suggested that agentivity, or a related attribute, may constitute a relevant  $\varphi$ -feature.

#### 3.1 Centering and the Givenness Hierarchy

My proposal of a scale of potential antecedents, from which pronominals select their antecedent—subject to various constraints—on the basis of its position, is not unlike the Givenness Hierarchy of Gundel et al. (1993), already described in the Introduction. I cannot hope to match the breadth or the size of their study; but in this section I will discuss what I perceive as some inadequacies of their model, and argue that my analysis provides a better explanation of the behavior of pronominal subjects.

Table 8: Distribution of Japanese forms according to highest status. Reproduced from Gundel et al. (1993).

	IN FOCUS	ACTIVATED	FAMILIAR	UNIQUE	REFERENTIAL	TYPE	TOTALS
$\phi$	87	1					88
kare	4						4
kore	1	1					2
sore		1					1
are							
kono N	1	7	1				9
sono N	18	15	1				34
ano N		1	1				2
N	14	32	17	71	45	44	223
TOTALS	125	58	20	71	45	44	363

The notion of cognitive status has the drawback that, as stated by Gundel et al., cognitive status “is not uniquely determined by syntactic structure.” The Cf list of centering theory, on the other hand, is at least in principle syntactically determined. Thus an account of pronominal use that is based on the ranking of centers is more concrete as well as utilizing independently-motivated structures.

There are other problems with the Givenness Hierarchy; the appropriateness of a single, implicationaly ordered scale including both informational and structural statuses was already questioned in the Introduction.

Another problem is that the Givenness Hierarchy only puts *lower* bounds to the cognitive status of the antecedents with which a given expression may construed. Since upper bounds are automatically imposed by the operation of the maxim of Quantity, they should be apply in a uniform way to all anaphoric expressions. But the authors' own data fail to conform to this prediction: for example, as table 8 shows, they report that in Japanese the form *sono N* is used equally often for in-focus and activated antecedents, while the form *kono N* is used almost exclusively with activated antecedents. Thus the two forms require the same minimum status of their antecedent; the Maxim of Quantity, which ought to apply to both of them alike, cannot account for the difference in the high end of their range.

A similar problem was already identified in footnote 1: the data of Gundel et al. (1993) show that the expression *that N* is used with activated as well as familiar antecedents, even though English provides several expressions requiring activated antecedents, including *this*, *that* and *that N*. Why should the maxim of Quantity fail to trigger the inference that the antecedent of *that N* is not activated?

Similar comments apply to definite NPs (in any of the languages in their sample), whose distribution is much broader than that of the other expressions. It can be seen in table 8 that NPs in Japanese are used in significant numbers for entities at all cognitive statuses, and that all other expression types are essentially restricted to one or two cognitive statuses. This pattern is repeated in every language that does

not have both definite and indefinite determiners; in English and Spanish, indefinite NPs can be used with the two lowest statuses “referential” and “type identifiable,” and definite NPs can be used with all other statuses.

The above issues demonstrate that the maxim of Quantity cannot by itself determine the upper end of the range of use of anaphoric expressions. The claim that such upper bounds are (uniformly) imposed by the maxim of Quantity is also incompatible with the Overt Pronoun Rule, which crucially invokes an explicit, lexically specified upper bound in claiming that Greek overt pronominals are specifically constrained to pick out antecedents that are not the most prominent in the Cf list.

A final reason to favor a centering-based approach over the Givenness Hierarchy is the tendency of *aftos* to take its antecedent from the immediately preceding sentence. There is nothing in the Givenness Hierarchy that will explicitly enforce such behavior (although some of the higher statuses may well be impossible to attain for anything but the most recent antecedents). As mentioned in the previous section, the importance of the immediately preceding utterance to centering theory should be taken as support for the relevance of centering theory to pronominal anaphora.

## References

- Di Eugenio, Barbara (1990) “Centering Theory and the Italian Pronominal System,” in *Proceedings of the 13th International Conference on Computational Linguistics (COLING '90)*, Helsinki, Finland, 270–275.
- Grice, H. P. (1968) “Logic and Conversation,” in *Studies in the Way of Words*, Harvard University Press. Reprinted everywhere.
- Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein (1986) “Towards a Computational Theory of Discourse Interpretation,” ms., Harvard/U. Penn.
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski (1993) “Cognitive Status and the Form of Referring Expressions in Discourse,” *Language* 69:2, 274–307.
- Hirschberg, Julia (1985) *A Theory of Scalar Implicature*, Doctoral dissertation, University of Pennsylvania.
- Prince, Ellen F. (1994) “Subject-Prodrop in Yiddish,” ms., University of Pennsylvania.
- Turan, Ümit Deniz (1994) “Ranking Forward-Looking Centers in Turkish: Universal and Language-Specific Properties,” to appear in E. Prince, A. Joshi and M. Walker (eds.), “Centering in Discourse”.
- Turan, Ümit Deniz (1995) *Null vs. Overt Subjects in Turkish Discourse: A Centering Analysis*, Doctoral dissertation, University of Pennsylvania.
- Walker, Marilyn, Masayo Iida, and Sharon Cote (1994) “Japanese Discourse and the Process of Centering,” *Computational Linguistics* 21, 193–232.
- Ward, Gregory L., and Ellen F. Prince (1991) “On the Topicalization of Indefinite NPs,” *Journal of Pragmatics* 16:2, 167–177.