Abstract

It has been argued that the form of a referring expression can be affected by different pragmatic factors and inherent properties of the referent (Ariel 1990, Gundel et al. 1993). This paper investigates the effect of a number of pragmatic factors on the choice of subject form (overt/null) in Spoken Jordanian Arabic (JA). Goldvarb statistical analysis of the data shows that the choice of subject form in JA is significantly constrained by coreference relationships with the subject of the preceding sentence(s) in the discourse, the cognitive status of the subject in the immediate discourse, and the animacy property of the referent.

Keywords: null subjects, Arabic, pragmatic functions, spoken discourse.

ملخص

لقد أوردوت العديد من الدراسات أن شكل الاسم الدال قد يتأثر بعدد من العوامل البراجماتية، وذلك المتعلقة بخصائص الشيء نفسه الذي يدل إليه الاسم (Ariel 1990, Gundel et al. 1993). يهدف هذا البحث إلى دراسة عدد من هذه العوامل التي يمكن أن تؤثر على ظهور أو استكشاف فعل الجملة في اللغة العربية المحكية في الأردن. يشير التحليل الإحصائي البرمجي (Goldvarb) في هذه الدراسة إلى أن شكل الفاعل في الجملة يتأثر ودرجة دالية احصائية فيما إذا كانت دالة فعل الجملة مشتركة مع الفاعل في الجملة أو الفاعل الواردة سابقاً في السياق، أو في حالة التركيز الذهني على ما يدل عليه الاسم الفاعل، أو فيكون الفاعل يدل على شيء جماد أو عاقل.

الكلمات المفتاحية: فاعل مستتر، العربية، الوظائف البراجماتية، الخطاب.

* Assistant Professor, Yarmouk University.
Introduction:

Languages that allow the subject of finite clauses to be null have received an attention in the linguistic literature from different perspectives. Linguists working in the generative framework have shown that the pro drop option is controlled by idiosyncratic grammatical properties of the language. In particular, they find that the null subject option is dependent on rich agreement morphology between the subject and verb (Chomsky 1981, Rizzi 1986). Rizzi (1986) argues that null subjects are licensed by a head with strong phi-features. Arabic is a language known for rich agreement morphology between the subject and the verb. Therefore, I assume that the formal licensing of a null subject in Arabic goes along the lines suggested by Rizzi (1986) and the subsequent literature.

On the other hand, several studies have examined the subject in Arabic from a functional perspective (Fakhri, 1995; Khalil, 2000; Moutaouakil, 1989). These studies have focused on the information structure of the subject and its order in the sentence. However, most of these studies have investigated the subject in written discourse or brief constructed discourses (Eid, 1983). Only a few other studies have examined the information status of the subject in spoken discourse such as plays (Schulte-Nafeh, 2005; Parkinson, 1987) or naturally occurring conversations (Owens et. al., 2010). This paper aims to examine the overt/null realizations of the preverbal subject in Spoken Jordanian Arabic, a variety in which the variation in subject type (null or overt) has not been examined. In particular, the researcher provides a quantitative analysis of the effect of a number of pragmatic factors on the choice of an overt or null subject by native speakers of JA. Thus, this paper follows a variationist method in examining the pragmatic functions of subject form (null or overt) in JA. However, the study does not control for social factors as was done in examining the form of the subject in some heritage languages (Nagy 2015). Hence, a formal analysis of the derivation of the subject (null or overt) is beyond the scope of this paper. In what follows, I review the relevant literature (section 0), describe the data and methodology (section 0), present the results (section 0), and discuss the findings (section 0).
Literature Review:

Subject form and information structure:

Cognitive status and subject form:

The topic of discourse, defined as what the comment or the discourse is about (Brown and Yule, 1983; Reinhart, 1983), is the most salient entity in the discourse (Bosch, 1983). This salient cognitive status of the topic has been shown to be correlated with the form of referring expressions used in the discourse (Ariel 1988, 1990, 1994; Givon, 1983; Grosz and Sidner, 1986; Gundel et al., 1993; Prince, 1981; among others). According to Gundel et al. (1993), the form of a referring expression in natural discourse is determined by the cognitive status of the referent as used in the particular context of situation. They propose the Givenness Hierarchy in (1), which consists of a set of cognitive statuses that are involved in determining the form of the referent in natural language discourse (Gundel et al., 1993,P:275).

1. The Givenness Hierarchy:

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

\{
\{it\}  \{this\}  \{that N\}  \{the N\}  \{indefinite this N\}  \{a N\}
\}

Particularly relevant to this study is the “in focus” status. According to Gundel et al. (1993, P 279), entities “in focus” condition are in short-term memory and also at the current center of attention. They are “likely to be continued as topics of subsequent utterances.” Given the salient cognitive status of these elements, the form of the referring expression denoting them in the discourse will often be a null or an unstressed pronoun. For example, according to Gundel et al. (1993) in-focus antecedents in languages such as Spanish and Chinese are more likely to be correlated with null subjects or unstressed pronouns. Gelormini-Lezama and Almor (2011) also found a significant correlation between the salience of the antecedent and the form of the referring expression in Spanish.

A similar claim is made by Ariel (1988, 1990). For her, the form of a referring expression is determined by the degree of accessibility of the
antecedent at the time of the utterance. “Accessibility” refers to the degree of availability of the antecedent in one’s working memory in the discourse model. Ariel (1990, 1994) suggests the Accessibility Marking Scale in (2) which marks the degree of accessibility of an antecedent of a referring expression. For instance, when an antecedent is highly “accessible” in the discourse, it will be encoded by the use of a null pronoun, while a less “accessible” referent will be denoted by a marked referring expression such as a definite NP.

2.  

zero < reflexives < cliticized pronouns < unstressed pronouns < stressed pronouns < stressed pronouns + gesture < proximal demonstrative (+ NP) < distal demonstrative (+ NP) < proximal demonstrative + NP + modifier < distal demonstrative + NP + modifier < first name or last name < definite description < full name.

The degree of accessibility of an antecedent can be affected by certain factors such as saliency\(^{(1)}\). Saliency refers to the importance of the antecedent in the discourse compared to other elements in the sentence. For example, according to Ariel’s Saliency Criterion shown in (3), elements that function as topics are more salient than non-topics:

3.  

**Saliency Criterion: (Ariel, 1990)**

a. Topics ≠ non-topics.

b. Grammatical subjects ≠ non-subject (e.g., objects, NP-internal possessives).

c. 1\(^{st}\) and 2\(^{nd}\) person 3\(^{rd}\) person.

d. Agent non-agent.

It is worth noting here that although both Ariel’s and Gundel’s models seem to agree on the idea that a cognitively salient antecedent in the discourse is encoded by the use of pronouns, the two models seem to classify pronouns differently. For example, while null pronouns signal a higher degree of accessibility of an antecedent than stressed pronouns on Ariel’s scale, Gundel et al. (1993) treat null and overt pronouns as marking equally high salience.

Gutman (2000) investigates the syntactic and discourse constraints on the use of null subjects in Hebrew. Following Ariel’s (1990) Accessibility model, she finds that pro-drop in Hebrew is constrained by Saliency and Unity. For example, given that Dafna in (4) is the topic of the discourse and thus is a salient entity, the use of a co-referential null subject form becomes licit.
4. \( \text{lo ta} \text{'r} \text{aminu ma kara le-Dafn} \text{ai} \): lifney \( \text{ju} \text{v} \text{u} \text{y} \text{i} \text{m} \) Dani xakar, neg. believe. 2fP what happened to-Dafna before week two. Dani investigated ota, ve-axarey xame\( \text{d} \text{akot} \) \( \text{ni} \text{m} \text{c} \text{e} \text{a} \text{f} \text{e} \text{m} \) her and-after five minutes found.f guilty.f ‘(You) wouldn’t believe what happened to Dafna: two weeks ago, Dani \text{in} \text{ves} \text{t} \text{i} \text{g} \text{a} \text{te}d her, and five minutes, (she) was found guilty!’ (Gutman, 2000, P:178).

In this paper analysis, I look at the variables that constrain the realization of null vs. overt subjects (pronouns and lexical NPs). Therefore, the researcher treats null subjects as signals of a highly accessible antecedent, but overt pronouns and lexical subjects are both treated as signals of less accessible antecedents. Testing the reliability of this methodology, the researcher made two separate runs in Goldvarb: In the first run, the researcher examined the factors favoring the realization of null subjects vs. overt subjects (collapsing overt pronouns with lexical NPs as one group), and, in the other run, the factors favoring the realization of null vs. overt pronominal subjects (i.e. NP subjects are excluded) are included. The researcher found that the results continue to hold. For example, in both runs, subject referents that function as the topic of discourse favored a null subject. The run examining the null vs. overt pronominal subject realization is also meant for examining features particularly relevant for pronouns such as person and point of view. This issue is discussed in section 0.

**Pragmatic anaphora:**

Working in a Gricean framework, Levinson (1987, 1991) proposes a set of pragmatic principles to account for intra-sentential anaphora. He suggested the following “general anaphora pattern”: “the more minimal the form, the stronger the preference for a coreferential reading”. (Levinson 1987, P:384). He summarizes this pattern in (5), where moving to the right favors a coreferential reading, while a reversion to the left implicates a disjoint reading. According to Levinson, the pronoun in the English example in (6) tends to pick its reference from the last relevant NP, while the switch to a full lexical NP (the man) signals a disjoint reference.

5. Lexical NP > Pronoun > \( \text{Ø} \).
6. John came into the room. He sat down. The man coughed.

Levinson argues that anaphoric relations are attributed to the Gricean Maxims of Quantity and Manner. According to Levinson (1991, P:110-
Pragmatic Conditions on Subject ................................. Osam Omari

11), the maxim of Quantity, which he renames as the Informativeness Principle (I-Principle), encourages speakers to use “maximally informative and cohesive interpretations from minimal linguistic specifications”, while the Maxim of Manner (M-principle) “induces from the use of a prolix or marked expression an interpretation that is complementary to the one that would have been induced by the I-principle from the use of a semantically general expression”. Thus, the anaphoric expressions he in (6) will pick its specific referential features from antecedents in the context. However, the use of the lexical NP the man instead of a pronoun in (6) will implicate a non-coreference interpretation by the M-principle.

Building on Levinson’s (1987, 1991) neo-Gricean theory of anaphora, Huang (1991) argues that in addition to the effect of the I- and M-Principles proposed by Levinson, intersentential anaphora can be constrained by what he dubs ‘consistency constraints’. Relevant to this study is the antecedent salience constraint. Huang points out that the use of the overt pronoun ta in (7b) instead of a null form in (7a) would be interpreted as non-coreferential with the topic Xiaohua by Levinson’s M-principle, which predicts a disjoint reference when an overt pronoun is used in a position where a more minimal form (null pronoun) could have been used. However, Huang points out that Levinson’s M-principle does not give the right predictions in this context. Both the null pronoun (7a) and the overt form (7b) are interpreted as coreferential with the topic Xiaohua. Huang (1991, P:325-26) attributes the vanishing of Levinson’s M-principle predictions in this context to the saliency of the topic antecedent. In other words, the use of an overt pronoun instead of a null pronoun does not implicate a disjoint reference relation, which is due to the saliency of the referent Xiaohua as the topic.

7. a. Xiaohua, Xiaoming yì jin wù, Ø jiù.  
Xiaohua Xiaoming as soon as enter room EMP  
ba  men  guan shang le  
BA door close RV CRS  
‘Xiaohua1, as soon as Xiaoming2 enters the house, (he1) closes the door.’ (Huang, 1991, P:325).

b. Xiaohua, Xiaoming yì jin wù, ta jiù.  
Xiaohua Xiaoming as soon as enter room 3SG EMP  
ba  men  guan shang le  
BA door close RV CRS  
‘Xiaohua1, as soon as Xiaoming2 enters the house, he1 closes the door.’ (Huang, 1991, P:325).
Blackwell (1998, 2000) examines the pragmatic constraints on anaphora in Spanish conversational discourse within a neo-Gricean anaphora model (Huang, 1991, 1994; Levinson, 1987, 1991). She shows that the variation in the form of a referring expression in discourse is constrained by Levinson’s (1987, 1991) pragmatic model of anaphora plus the intervention of other factors such as the mutual knowledge of the speaker and addressee. She illustrates how Levinson’s model works in Spanish discourse with the following example (Blackwell, 1998, P:614):

8. Maria Jesus tenia gallina, Ø hizo un caldo, Ø lo puso en un termo.

Maria Jesus had.3s hen made.3s a broth it put.3s in the thermos

‘Maria Jesus had a hen, made a broth, and put it in a thermos.’

According to Blackwell, the coreferential relationship between the subject of the verb hizo ‘made’ and the subject of the preceding verb, tenia ‘had’ is encoded by the use of a minimal form (the null subject). Similarly, the coreferential relationship between the direct object of the verb puso ‘put’ and the object of the preceding verb hizo ‘made’ is encoded by the use of the pronoun lo.

Blackwell also shows that the use of minimal anaphoric expressions in Spanish can be constrained by other factors such as antecedent salience. Testing the applicability of Levinson’s theory of pragmatic anaphora in Spanish, she reports that 69% of the respondents chose the subject Juan as the antecedent of the pronoun él ‘him’ in (9), and only 20% of them chose Angel as a possible antecedent. She points out that this is surprising because, according to Levinson’s theory, the use of an overt pronoun will Q-implicate a non-coreferential relationship. However, speakers still favor to interpret the overt pronoun as coreferential with the subject antecedent Juan. Blackwell explains the “overriding” of Levinson’s Anaphora principle in this context as follows: “because Juan is foregrounded or in focus, as a result of its subject status, it may have been perceived as ‘what the utterance is about’, thus explaining why Juan was chosen most frequently as the antecedent” (Blackwell, 2000, P:407).

9. Cuando O llegaron al bar para darle la enhorabuena a Angel, when pro arrived-3pl to-the bar for give-inf.-him the congratulation to Angel Juan pidió una cerveza para él y otra para su novia Marta.

Juan ordered a beer for him and another for his girlfriend Marta.

‘When (they) arrived at the bar to congratulate Angel, Juan ordered a
beer for him and another one for his girlfriend Marta.” (Blackwell, 2000, P:405).

Owens et al. (2010) examine the function of the variation in the overt and null realization of the subject in a corpus of three Arabian Peninsula dialects. Employing a multivariate analysis methodology, they find that coreference and person/number are significant variables for the choice of subject type (overt or null). In particular, they find that null subjects are favored in contexts where the subject is coreferential with the subject of an adjacent preceding sentence. Overt subjects are also shown to be favored with 3rd person pronominal subjects.

In the current study, the researcher examines the effect of pragmatic factors such as coreference and antecedent salience in the choice of the form of the subject (null and overt) in JA. The researcher shows that the use of a null subject in JA is constrained by Levinson’s “general anaphora principle” and antecedent salience (Huang, 1991).

**Switch-reference and disambiguity:**

Eid (1983) argues that the overt/null realization of the subject pronoun in Egyptian Arabic has certain communicative functions. She proposes that an overt subject pronoun can serve two functions: disambiguating a referent and signaling a subject switch. As shown in (10a), the verb in the relative clause ʃ atam ‘he insulted’ is inflected for 3rd person masculine singular, and both al-walad ‘the boy’, the object of the matrix clause, and ʕ ali, the subject of the matrix clause, are possible antecedents for the null subject in the relative clause. According to Eid, when the subject of the relative clause is null, it will be interpreted as coreferential with the closest antecedent al-walad, as seen in (10a). However, when the subject is overt, it serves as cue to the hearer that the antecedent of the subject is the remote NP ʕ ali, the subject of the matrix clause, as illustrated in (10b) (Eid, 1983, P:289).

10. a. ʕ ali kallim il-walad illi Ø ʃ atam-u imbaarih.
   ‘Ali talked to the boy that insulted.3ms-him yesterday’

   b. ʕ ali kallim il-walad illi huwwa ʃ atam-u imbaarih.
   ‘Ali talked to the boy that he (the boy) insulted him (Ali) yesterday.’

In conjoined structures such as (11), Eid claims that unlike in (10), neither of the two clauses involves reference ambiguity. The null realization of the subject in (11a) indicates that the referent of the subject of the conjoined clause is coreferential with the matrix subject,
whereas the overt subject pronoun in (111) signals a subject-switch where the referent of the subject of the conjoined clause changes to be Samiir, the object subject of the matrix clause, as seen in (111).

11. a.ʕ ali dˁ arab Samiir wi Øʃ atam-u
   Ali hit.3ms Samiir and insulted.3ms-him
   ‘Ali hit Samiir and he (Ali) insulted him.’

   1.ʕ ali dˁ arab Samiir wi huwwaʃ atam-u
   Ali hit.3ms Samiir and insulted.3ms-him
   ‘Ali hit Samiir and he (Samiir) insulted him.’

Eid’s analysis of the function of overt pronouns in Egyptian Arabic is based on constructed discourses, and it is hard to obtain similar examples in naturally spoken data. For example, Parkinson (1987) finds only few examples (14/188) in his corpus of the type in (10b), where the subject of the relative clause is overt.

Parkinson (1987) investigates the constraints on the presence/absence of subject pronouns in a corpus of two Egyptian Arabic plays. Testing Eid’s (1983) previous claims about the effects of disambiguity and subject switch, he notes that the disambiguity function does not seem to be testable in naturally occurring data due to a rarity of such ambiguous instances. Out of 14 occurrences of relative clauses with an overt subject, only two examples were found to be subject to the ambiguous possibilities discussed by Eid. However, he finds that subject switch has a highly significant effect on the null/overt realization of the subject.

Parkinson also notes that first person pronouns favor an overt subject more than second or third person pronouns. While the current study confirms Parkinson’s finding, the researcher also provides an answer for this tendency for this type of subject to be overt. This issue will be discussed in section 0.

Schulte-Nafeh (2005) investigates the use of overt/null pronominal subjects in Egyptian Arabic. She argues that the overt/null realization of the subject in Arabic is not controlled by the person feature marked on the predicate; rather it is controlled by the context of discourse. This is based on the fact that null subjects are licit in zero-copula sentences, where the predicate in these clauses is not marked for person. For example, a null subject is possible in the conjoined clause in (12) although the participle predicate sakta “being quiet. fs” is not marked for person.

12.ʔana katma fiiʔalb-i uw Ø sakta.
I repressing.fs in heart-my and being quiet.fs

‘I am repressing my feelings and keeping quiet.’ (Schulte-Nafeh, 2005, P:198).

The disambiguity function will not be addressed in this study due to methodological considerations; the contexts discussed by Eid (1983) would be rare in natural discourse. However, this study does investigate how subject switch is encoded in natural discourse. As discussed in section 0, subject switch contexts favor the use of an overt subject (pronoun and NP).

**Subject form and predicate type:**

Several studies have shown that the overt/null realization of subject pronouns is correlated with the type of predicate (Enríquez, 1984, as cited in Stewart, 2003, P:199; Posio, 2011). Posio (2011) provides a clause-level functional account for the variation in the frequency of subject pronouns with different types of verbs in spoken Peninsular Spanish. He finds that first person pronominal subjects of “cognitive verbs” such as creo ‘I think’ tend to be overt. Following Silva-Corvalán (1997, 2001), Posio explains this tendency as a result of the “focus of attention” on the speaker to express his/her personal point of view towards the subject matter. The “focus of attention” in Posio’s proposal refers to the relative importance of the participant at the clause level. For instance, according to Posio, subject pronouns are more likely to be overt with cognitive verbs such as creo ‘I think’ because the subject of these verbs is the element under “focus of attention”, expressing the speakers’ personal point of view towards the subject matter. However, the subject in a highly transitive clause such as *I killed the cat* is more likely to be absent since the “focus of attention” will be drawn on the object of the sentence.
**Data and methodology:**

**Data:**

The data for this study are drawn from a corpus of interviews with seven native speakers of JA. All the interviews were conducted by the researcher in a public Jordanian University. The length of the interviews was between 25-35 minutes. Recordings were made using an analogue recorder and were then digitized.

All of the informants were between (18-22) years of age and were enrolled in an undergraduate program. The researcher has taken some steps to decrease the level of formality of the interview. For example, the topics raised in the interview were informal, such as informants’ daily routine, experience with school, lifestyle, socializing, and future plans. In addition, the speakers were selected with the help of their professors who introduced the researcher as a friend and a colleague who had worked in the university. As well, almost all of the interviews (6/7) were conducted with one or two of the interviewee’s friends present. This helped to create a comfortable atmosphere, as noted by the female speakers, in particular.

Following a normal sociolinguistic methodology (Tagliamonte, 2006), statistical analysis in this study was carried out using the Windows application Goldvarb X (Sankoff et al., 2005). A total of 1154 verbal sentences were extracted. They were transcribed and coded in a Microsoft Excel sheet by the author. A time-stamp was also added to track back the utterances for a close analysis. The data were then copied into a token file created by Goldvarb X.

**Coding and methodological considerations:**

This study examines the factors that may constrain the overt/null realization of the subject in verbal sentences. Thus, zero-copula sentences such as in (13) are not coded in the study. (3)

13. (ʔana) taʕbaan.(4)
   I tired.ms
   ‘I am tired.’

One of the reasons behind excluding this type of sentences is that when the subject of these sentences is null, it is difficult to determine whether the predicate is a complement of a true sentence with a null subject or simply a phrase.

Each verbal sentence was coded for the following pragmatic factors:
Pragmatic Conditions on Subject

1. Subject type: full NP, null pronoun, overt pronoun, demonstrative pronoun.

2. Coreference: whether or not the referent of the subject of the target sentence is the same as the referent of the subject of the previous sentence. For this factor, a subject of a preceding zero-copula sentence was considered.

3. Discourse topic: whether the referent of the subject is active in the conversation (old topic) or not (new topic).

4. Person/number: 1st person (singular and plural), 2nd person (singular and plural), 3rd person (singular and plural).

5. Epistemicity of the verb: whether or not the verb is used to state the speaker’s point of view. Only verbs with a 1st person singular subject were counted.

6. Animacy of the subject: animate or inanimate.

7. Transitivity of the verb: transitive with object, intransitive/transitive without object, copula (kaan ‘was’/ṣaar ‘became’).

8. Clause type: matrix, coordinate, embedded, conditional, adverbial, interrogative.

9. Tense/aspect of the verb: perfect, imperfect, imperfect with a future auxiliary.

The following contexts were not coded/included in the analysis:

1. Imperative clauses: In addition to being rare in the data, these sentences are nearly categorical with a null subject.

2. Quotations: subjects in direct quotations were not coded. For example, the subject of the verb bahib ‘like’ in the following example was not coded because it was a direct quotation of what a third party said:
   a. “bahib ṭaxaalif il-qawaanin.”
      like.1s break.subjn the rules ‘(I) like to break the rules.’

3. Relative clauses: the subject of these clauses is obligatorily null, as shown in the following example:
   a. il-walad, illi *huu/Ø, iʃ tara sayyaarah. the-boy that he/ Ø bought.3ms car
4. The subject of a complement of a control predicate. For example, the subject of the verb *yinjaḥ* ‘succeed’ in (a) below is not coded because it is obligatorily null and coreferential with the subject of the matrix verb. As well, the subject of the verb *ʔiid-u* ‘redo’ in (b) is not coded because the subject is obligatorily overt.

a. *il-walad, haawal *huu/Ø yinjaḥ.*
   the-boy tried.3ms he/Ø succeed.3ms.subjn
   ‘The boy tried to succeed.’

   the-instructor let.3ms the-students redo-3mp the-test
   ‘The instructor let the students redo the test.’

**Results:**

Goldvarb binominal analysis identified a number of factor groups that are significant for the realization of overt vs. null subjects. These findings are summarized in Table (1), where a value greater than .50 indicates that this variable favors an overt subject, and a value under .50 indicates that this variable has a disfavoring effect. All of the factors selected as significant will be discussed in section 0 below.
Table (1):  
**Factors contributing to the favoring of an overt subject**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Factor weight</th>
<th>% N (null and overt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same as previous subject</td>
<td>0.19</td>
<td>5.6</td>
</tr>
<tr>
<td>Different from previous subject</td>
<td>0.66</td>
<td>49.7</td>
</tr>
<tr>
<td>Discourse topic:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referent is not active/new topic</td>
<td>0.78</td>
<td>56.4</td>
</tr>
<tr>
<td>Referent is active/old topic</td>
<td>0.13</td>
<td>7.3</td>
</tr>
<tr>
<td>Point of view:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker expressing his/her point of view</td>
<td>0.81</td>
<td>55.9</td>
</tr>
<tr>
<td>Not expressing a point of view</td>
<td>0.46</td>
<td>18.7</td>
</tr>
<tr>
<td>Animacy:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animate</td>
<td>0.43</td>
<td>28.9</td>
</tr>
<tr>
<td>Inanimate</td>
<td>0.85</td>
<td>77.2</td>
</tr>
<tr>
<td>Transitivity:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitive with object</td>
<td>0.41</td>
<td>26.6</td>
</tr>
<tr>
<td>Intransitive/transitive without object</td>
<td>0.53</td>
<td>35.5</td>
</tr>
<tr>
<td><em>Kaaum</em> ‘was’, <em>sā'ār</em> ‘became’</td>
<td>0.56</td>
<td>68.8</td>
</tr>
<tr>
<td>Clause type:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverbial</td>
<td>0.27</td>
<td>22.5</td>
</tr>
<tr>
<td>Matrix</td>
<td>0.48</td>
<td>36.9</td>
</tr>
<tr>
<td>Coordinate</td>
<td>0.75</td>
<td>26.6</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.51</td>
<td>38.5</td>
</tr>
<tr>
<td>Embedded</td>
<td>0.57</td>
<td>42.0</td>
</tr>
<tr>
<td>Interrogative</td>
<td>0.33</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Total N: 1154  
Log likelihood = -393.083  Significance = 0.262
Discussion:

Coreference and discourse topic

According to Levinson’s general anaphora principle (5), repeated below in (14), the use of a null form signals a coreferential relationship with an adjacent antecedent, while the use of a pronoun or a lexical NP form implicates a non-coreferential relationship.

14. Lexical NP > Pronoun > Ø

This prediction is supported by the results of this study. As shown in Table (1), overt subjects are favored in contexts where the referent of the target verb is different from the referent of the previous verb. This tendency is illustrated by the following excerpt:

15. il-waadidi bi=staq fil muwazzi ʕ. Øi kaan yishtaq fil b-iss=s ʕuudiyyih.

the-father work.3ms distributor. was.3ms working. 3ms in-the-Saudi. Arabia
Ø, rij iʕ. Ø, fatah dukkaanih. returned.3ms. opened.3ms store.
‘My father works as distributor, (He) was in Saudi Arabia. (He) returned (from Saudi Arabia). (He) opened a store.’
The speaker in (15) introduces the new referent il-waadid ‘the father’ to the discourse in a full lexical form. However, the subsequent reference to il-waadid is encoded by a null subject form which picks its reference from subject NP.

Similarly, the use of a null subject with the verbs kammal ‘continued’ tˁ iliʕ ‘left’, and if tara ‘bought’ in (16) below implicates a coreference relationship with the subject NP antecedent Hassan. The switch to the full NP subject form xaal-i ‘my uncle’ signals a disjoint reference relation with the previously introduced subject (Hassan) and introduces a new referent into the discourse.

16. (…) fi= waaqad. hassan, ism-uh. Ø, maa kammal b-iʃ-j et j. Ø, tˁ iliʕ gabul
there one. Hassan name-his. neg continue in-the-army. left before maa yitqaʕ ad, uw Ø, if tara taxi. if taraa-l-uh, ʕiyyah xaal-i
that retire.3ms, and bought.3ms taxi. bought.3ms-to-him it uncle-my iθ-θaani. uw Ø, biʃ tayʕ il ʕalat -h, uw lissah Ø, biʃ tayʕ il ʕalat -h
the-other. And work.3ms on-it and still work.3ms on-it.
‘(...) There is one. His name is Hassan. (He) didn’t continue in the army. (He) quit before retirement, and (he) bought a taxi. My other uncle bought it for him. (He) drives it (the Taxi), and (he) still drives it.’

However, although Levinson’s anaphora principle could successfully predict the form of the subject for the verbs kammal ‘continued’ tˁ iliʕ ‘left’, and if tara ‘bought’ in the above example kammal ‘continued’, other factors come into play to affect the form of the referring expression in the discourse. The null subject of the subsequent verb bif taʕil ‘he works’, which is repeated two times in the example above, does not pick its antecedent from the adjacent subject NP xaal-i ‘my uncle’; rather it picks its reference from the non-adjacent subject NP Hassan. This can be explained by the cognitive status of Hassan in the current discourse segment. Given that Hassan is “in-focus”, as the topic of the discourse segment, it is encoded by the use of a null subject form.

This section has discussed the effect of coreference and saliency of the referent in the choice of the referring expression. The following section discusses the influence of certain inherent factors of the referent such as animacy and person and the semantic category of the verb on the choice of referring expressions.

**Animacy, Person/number, and predicate type:**

As shown in section (5.1) above, the choice of the referring expression (null or overt) is affected by whether the referent has been in focus in the immediate discourse. However, other studies have shown that the choice of the form of referring expression may be affected by non-contextual factors such as the animacy of the referent (Prat-Sala and Branigan, 2000, Fukumura and van Gompel 2011, Vogels et al. 2014). Fukumura and van Gompel (2011: p, 1478) report that pronouns are used to refer to animate referents more frequently than noun phrases. In addition, based on data from Hebrew, Artstein (1999) suggested the animacy hierarchy in (17). He generalizes that “elements higher on the person/animacy scale tend to be realized as null subjects, while those lower on the scale tend to be realized overtly” (Artstein 1999,P:4). The results of this study provide further support to these two finding. As shown in Table (1), when the referent of the subject is inanimate, the overt form of the subject is strongly favored. This finding is also consistent with other studies have reported about other Arabic varieties. Examining data from three varieties of Arabic spoken in the Arabian Peninsula, Owen et al. (2010) find that when the referent of the subject is inanimate, overt subjects are the favored form of referring expressions.
17. 1st/2nd person > Proper Noun 3rd > Human 3rd > Animate 3rd > Inanimate 3rd

To test for the effect of person and number on the form of the referring expression, the author made a separate run in which NP subjects and demonstrative pronouns were excluded from the person group because including them would shift the weight value to favoring the 3rd person since all the NP subjects in the data are 3rd person. The 1st person pronouns were divided into two groups (singular and plural) because the point of view factor group mainly matters for 1st person singular pronouns. Both singular and plural 3rd person pronouns were collapsed into one group. Number did not matter for the 2nd person pronouns since only singular forms were found in the data. The finding for the person factor group is presented in Table (2) below.

Table 2: The contribution of person/number factors to the favoring of an overt Subject

<table>
<thead>
<tr>
<th>Factors</th>
<th>Factor weight</th>
<th>(%) N (null and overt) Person/number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st sg</td>
<td>0.898</td>
<td>22.8</td>
</tr>
<tr>
<td>1st pl</td>
<td>0.429</td>
<td>26.1</td>
</tr>
<tr>
<td>2nd</td>
<td>0.073</td>
<td>4.1</td>
</tr>
<tr>
<td>3rd</td>
<td>0.234</td>
<td>5.1</td>
</tr>
</tbody>
</table>

As shown in Table (2), 1st person subject contexts favor an overt pronoun, while 2nd and 3rd subject contexts tend to favor a null subject. This finding is consistent with similar observations in other studies of Arabic dialects (Parkinson, 1987). Parkinson (1987,P:356) notes that “speakers are using pronouns more often when they themselves are the subject of their sentences, than when either their addressee is the subject or when some other person or thing is the subject.” In addition, following a qualitative methodology, Shulte-Nafeh (2005,P:193) observes that one of the obligatory positions for an overt 1st person subject pronoun in Egyptian Arabic is with the verb ʔazʕun ‘I think’. This is otherwise surprising because 1st and 2nd person antecedents, unlike 3rd person subjects, are salient entities, being participants in the conversation, and thus they are more likely to be referred to by a null form (Ariel, 1990).
However, the findings for person in the above table do not conform to Artstein’s hierarchy in (17). This hierarchy treats 1st person and 2nd person pronouns as equal in terms of their overt/null occurrence in the discourse. Accordingly, a null form of referring expression is expected to be favored when the referent is a 1st or 2nd person. However, as shown in Table (2), 1st person and 2nd person singular referents in JA favor an overt form rather than a null form. Thus, although Artstein’s hierarchy has been heavily based on Hebrew, a language closely related to Arabic, its predictions at least for the parallel treatment of 1st and 2nd pronouns are not supported by this study. It appears here that Hebrew and Arabic in this context employ different pragmatic interpretation of the world. In JA, the speaker appears to treat himself/herself differently from the hearer.6)

To further understand the effect of the person factor on the form of a referring expression, the author coded the predicate type. Since the “first person singular subject is the prototypical site for expression of speaker point of view” (Scheibman, 2002, P:63), the researcher coded the predicate type for first person singular subjects (null and overt). Predicates which the speaker used to express his/her opinion, such as bahki ‘I say’ baʔayyid ‘I agree’, and baʃuuuf ‘I see’, were coded as epistemic,7) while predicates that do not state the speakers’ point of view were coded as non-epistemic. As shown in Table (1), epistemic predicates tend to favor an overt 1st subject pronoun, while non-epistemic tend to disfavor it. This is consistent with the findings reported about other languages such as Spanish (Posio, 2011). “Lexical items in the cognitive group (e.g. think, know) with 1st and 2nd subjects have subjective function in discourse, and these usages do not extend to a subject that is not a speech act participant” (Scheibman, 2002,P:89). This particular function of these verbs seems to constrain the type of subject used in discourse. Thus, since the speaker is under the “focus of attention” expressing his/her own stance, these verbs tend to favor an overt subject (Posio, 2011).
Transitivity:

As shown in Table (1), the verbs *kaan* ‘was’ and *sˁaar* ‘became’ favor an overt subject. These verbs function as copulas with complements such as a noun, an adjective, or a prepositional phrase, as illustrated in the following examples:

18. *bikuun il-j aw raaʔiʕ* . be.fut.3ms the-weather great ‘The weather will be great.’

19. *sˁaar-at il-maadilih sahlih* . became-3fs. The-course easy ‘The course has become easy.’

Following Posio (2011:P:786), the researcher account for the tendency to have overt subject with this type of verb as a result of the “focus of attention”. Since “focus of attention” in these clauses is directed to the subject rather than to the complement, the subject tends to be overt.

Clause type:

Previous studies have shown that the overt/null realization of the subject can be constrained by the type of clause (Prince, 1999; Owens *et. al.*, 2010). For example, Owens *et. al.* (2010:35) finds that complement clauses in Peninsular Arabic favor an overt subject. However, although coordinate clauses in JA appear to favor an overt subject, as shown in Table (1), clause type does not otherwise seem to constitute an independent factor. A cross tabulation for the clause type and the reference factor groups shows that the favoring of an overt subject in coordinate clause contexts is an epiphenomenon of the reference factor group. In other words, the clause type factor is not an independent factor constraining the form of the subject. The following examples illustrate this pattern:

20. a. *ʔabuu-i mitqaʕ-id min il-j eʕf uw Ø maa biʕ1-iʕ* . father-my retired from the-army and neg work.3ms-neg ‘My father is retired from the army, and (he) does not work.’

b. *ʔana badrus hon uw ʔuxt-i btudrus fii j aamʕah ʔaanyih* . I study.1s here and sister-my study.3fs in university another ‘I study here, and my sister studies in another university.’
As shown in (20a), the speaker tends to use a null subject for the coordinate clause because the subject is coreferential with the matrix clause subject. However, the speaker uses a lexical form in the coordinate clause in (200a) since the subject has a new referent. Thus, it appears that the coreference factor has a stronger quantitative effect on the overt/null realization of the subject than the type of clause.

**Summary and Conclusions:**

In this study, I have examined the pragmatic factors that constrain the variation in overt and null subjects in JA. I have shown that this variation is constrained by pragmatic anaphora, antecedent salience, predicate type, and animacy. With regard to pragmatic anaphora, null subjects are favored in contexts where they co-refer with a preceding subject or when their antecedent has a salient position as the topic of the discourse segment.

Predicate type was also shown to be a factor constraining subject type in JA. Null subjects are disfavored with copula-type predicates such as kaan ‘was’ and sˁaar ‘became’. This is explained as a result of the “focus of attention” at the sentence level being drawn to the subject rather than the complement, which thus favors an overt subject type. Predicates used to express speakers’ opinions have shown an effect on subject type. Speakers favor using an overt form of the 1st person subject pronoun when they express their opinion.

I have also shown that the overt/null realization of first person pronouns for JA does not conform to Artstein’s (1999) hierarchy. The hierarchy predicts 1st and 2nd person pronouns to behave similarly in terms of their overt/null occurrence. However, this study shows that 1st person subject contexts favor an overt pronoun, while 2nd and 3rd subject contexts tend to favor a null subject form. I believe that further research is needed to confirm whether Hebrew and Arabic view the pragmatic world differently with respect to their use of null/overt pronouns.
References:


Pragmatic Conditions on Subject


Pragmatic Conditions on Subject


Margins:

(1) Only saliency is relevant here. Other factors suggested by Ariel, such as Unity, are not relevant to the current discussion.
Golvarb assesses the effect of factor groups in the data for a particular dependant variable. Individual factors within factor groups selected as statistically significant are given values from 0 to 1 (Tagliamonte 2006).

Existential clauses such as those in (a) below also belong to this class, and thus they are not coded. Owens et al. (2010: 28) report that these clauses were included in their analysis of null/overt subjects in Peninsular Arabic. However, the subject of these clauses is obligatorily overt. Moreover, existential sentences in spoken Arabic belong to the zero-copula clause type, which were not coded in the same study.

a. fii sayyaarah ʕind il-beit.
   there car next.to the-house.
   ‘There is a car next to the house.’

Adjectives and participles functioning as complements of copular verbs in JA are marked for number and gender, but are not for Person (14). However, Prepositional Phrases and noun complements do not carry any form of agreement with the subject.

In the person factor group, NP subjects and demonstrative pronouns were not coded as 3rd person; rather they were given a different code to allow for possible exclusion from the analysis.

Thanks to an anonymous reader for paying my attention to this point.

These predicates were most recurring in the data.