Subject-verb order in spoken Arabic: Morpholexical and event-based factors

JONATHAN OWENS
University of Maryland and Bayreuth University

ROBIN DODSWORTH
North Carolina State University

TRENT ROCKWOOD
University of Maryland

ABSTRACT

This article explores the relationship between the global functions of variable subject-verb order and morpholexical class of subjects in the spoken Arabic of the Arabian peninsula. Using corpus-based methods, it is shown that lexical class—pronoun, pronominal, noun—definiteness, and the discourse-defined lexical specificity of a noun all correlate significantly with subject-verb or verb-subject word order. The global function of the two orders is explored using an array of measures to show that verb-subject order prototypically presents events, while subject-verb signals available referentiality. Using the quantitatively based study of Anthony Naro and Sebastiao Votre ([1999]. Discourse motivations for linguistic regularities: Verb/subject order in spoken Brazilian Portuguese. Probus 11:75–100.) on Brazilian Portuguese as a point of comparison, a typological framework is developed for understanding languages with variable subject-verb order.

This article examines subject-verb (SV) word order in three varieties of Arabic spoken on the Arabian peninsula, showing that SV order is broadly sensitive to two sets of factors. The first encompasses formal and categorical properties of the nominals in subject position, properties such as lexical class (pronoun, noun, or pronominal), definiteness, and length of subject phrases. The second pertains to the discourse-pragmatic status of the subject. In addition to detailing how the

Research for this article was supported by the University of Maryland Center for Advanced Study of Language. An earlier version of this article was presented at the 35 NWAV (New Ways of Analyzing Variation) conference, Columbus, Ohio, in November 2006. We are indebted to three reviewers for their critical comments and to Nathanael Lynn for his assistance in coding the Al-Rawi Emirati data.
interaction of these two factors influences SV word order in Peninsular Arabic, we note cross-linguistic parameters of variation in word order. From this, the specifics of word order variation in spoken Arabic will emerge, and at the same time, suggestive cross-linguistic parameters of variation will be proposed.

PERSPECTIVES ON PRAGMATICALLY BASED WORD ORDER

Information status, discourse function: givenness/newness, continuity

Discourse-based studies of SV word order typically recognize two key variables—information status and morpholexical class—with the former generally treated as the integrating factor that morpholexical class instantiates. Information status has been characterized in a number of ways. A seemingly universal factor conditioning variable SV order is the given/new distinction (the Prague school, Halliday & Hassan, 1976; Prince, 1981). Several versions of this distinction have been employed to show that subjects conveying new referents occur postverbally more often than subjects with previously mentioned or implied referents. For example, Ferrari (1990) found presentative/existential clauses in Brazilian Portuguese—defined as clauses that use either ter ‘have’ or ser ‘be’ and convey new information—to be almost categorically verb-subject (VS). Further, in Ferrari’s data, 84% of nonpresentative/existential VS clauses have subjects that have not been previously mentioned in the discourse, and 95% of SV clauses involve subjects that have been mentioned previously or evoked by a semantic frame.

The given/new distinction is closely tied to the topic/focus dichotomy insofar as the focus is typically considered new information relative to the topic. For example, would-be topics that are, in Prince’s (1981) terms, brand-new and brand-new anchored are often introduced with the Spanish presentative hay and the French il y a so as to appear postverbally and in a separate clause from the focus/comment (Ocampo, 1993; see (14), (15)). Under this strategy, new “topics” are placed in focus position.

Givón (1983) generalized the basic dichotomy of given/new into broad measures of topic continuity in discourse, one of which, referential distance (see Table 1), can be adduced here:

Referential distance: gap between X and previous mention, limit 10 or 20 sentences.

One of Givón’s major contributions to word-order variation was to attribute to the alternative SV/VS orders global informational-status meanings. Thus, longer referential distance is interpreted as indicating a lower degree of topic continuity. In these terms, in Klamath, Nez Perce, and Biblical Hebrew, topic discontinuity is signaled more by SV than VS, as shown in Table 1. The opposite logic works for Polish. In Polish, VS is associated with topic discontinuity, and VS is the typical order for introducing new entities in texts. These studies minimally indicate that the link between topic continuity and word order is realized.
differently in different languages and therefore must initially be defined anew for each.5

**Morpholexical class**

In addition to communicative function, explanations of pragmatically determined word order typically recognize the grammatical properties of subjects and verbs, particularly morpholexical class. From a nonvariationist perspective, Kiss (1998) argued that in Hungarian “identificational” focus, exemplified in (1a), a preverb constituent shows absolute grammatical constraints on the type of modifier that can occur in it. Universal quantifiers, also phrases, and even phrases cannot occur in the identificational focus, as in (1b) (Kiss, 1998:248, 252).

(1) a. tegnap este Marinak mutattam be Pétert
    Last night Mary.DAT introduced.I Perf PERF Peter.ACC
    ‘It was to Mary that I introduced Peter last night’.

b. Mari még egy kalapot is nézett ki maganak
    Mary even a hat.ACC also picked out herself.DAT
    ‘*It was even a hat that Mary picked out for herself’.

Kiss explained the restriction semantically: identificational focus picks out a unique individual from a set. The even quantifier implies a referent beyond that restricted by even, so that the uniqueness restriction is violated (1998:252).

Looking at SV variation from within a Givónian quantitative framework, a number of studies indicate that morpholexical class membership alone is often a predictor of word order. In Nez Perce (Rude, 1992:203), a pronoun subject is SV to a statistically far greater degree than it is VS (Table 1).6 In Klamath, referential indefinites, which include quantifier + N subjects, are always SV (Meyer, 1992:173, 178). In O’odham (Papago), a large class of quantifiers and indefinites (including all, many, someone, one, how many) are either categorically SV (quantifiers) or overwhelmingly so (indefinites) (Payne, 1992:148, 156). Payne’s careful analysis of indefinites (termed h-words in O’odham, as they begin with /h/) is instructive. Payne classified them into referential and nonreferential tokens, but in the end concluded, “the majority [85/91] of h-phrases occur in preverbal position, regardless of whether they express referential or non-referential information.” What emerges in these

<table>
<thead>
<tr>
<th>Language</th>
<th>SV</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klamath (Meyer, 1992)</td>
<td>7.8</td>
<td>3.9</td>
</tr>
<tr>
<td>Nez Perce (Rude, 1992)</td>
<td>8.6</td>
<td>5.9</td>
</tr>
<tr>
<td>Polish (Jacennik &amp; Dryer, 1992)4</td>
<td>2.8</td>
<td>6.6</td>
</tr>
<tr>
<td>Biblical Hebrew (Fox, 1983)</td>
<td>9.7</td>
<td>4.8</td>
</tr>
</tbody>
</table>
languages, which generally are considered to have a pragmatically based word order, is relatively fixed word order subsets based on morpholexical class. These studies suggest that even within languages with pragmatically based word order, there is strong prima facie evidence for a link between word order and morpholexical class, even if individual manifestations differ across languages.7

Naro and Votre’s (1999) study of Brazilian Portuguese links morpholexical class with information status. Subjects in their corpus are classified within the general framework set out by Prince (1981) and are found to fall into two categories—evoked and available—on the basis of the relative frequencies of SV and VS order:

SV: Evoked (previously mentioned): common nouns, pronouns, demonstratives

Against the statistically dominant SV pattern in Brazilian Portuguese, subjects classified as available may occur in VS rather than SV order. Naro and Votre’s (1999) findings brought out the nonrandom intersection between information status and morpholexical class in determining word order.

**SV ORDER IN SPOKEN ARABIC**

We examine variation in SV order in spoken Arabian Peninsular Arabic from a combined discourse informational and morpholexical perspective. Our overall organizational strategy is to proceed from unambiguously measurable factors to factors explicable largely in qualitative terms, given the data available, both our own and in data sets from spoken Arabic generally. This progression mirrors a movement from discretely definable lexical units toward a larger discourse context. Finally, we extrapolate, on a preliminary basis, general elements determining SV order in pragmatically controlled languages on the basis of a detailed comparison between our results and those of Naro and Votre (1999).

**Morpholexical identity in Arabic**

The Arabic corpus employed here consists of four recordings from different areas of the Arabian peninsula, totaling about 30,000 words (Table 2). In broad dialectological terms, two geographic areas are represented in the corpus. Kuwait and the Emirates belong to Gulf Arabic dialects, which are distinguished from the Jedda dialect. For instance, Jedda always has *k, whereas Gulf Arabic changes *k to č in the context of front vowels (e.g., čeef ‘how’ vs. Jedda keef).

In addition, a set of published texts (Al-Rawi, 1990) recorded in 1987 from Emirati Arabic, is used, comprising some 4,250 words. These texts consist of 3 folk tales; 6 accounts of traditional activities, such as how to make a sheep-hair tent; and 1 story about an earlier recollection of life in Dubai. All were
tape-recorded from women, all over 40 years of age. The 10 texts are spoken by 8 individuals, 2 narrating 2 texts each. We term these texts the Al-Rawi Emirati corpus. To the extent possible, we use these texts to complement the finding from our own corpus, using the same quantitative methods.

Grammatical subjects in spoken Arabic are invariably marked by SV agreement in terms of person, number, and, where appropriate, gender. All verbal subjects were coded for definiteness. Definite nouns, those marked by the definite article (2) or occurring in a definite genitive (idaafa) construction (3), are distinguished from indefinite nouns ((4) and (5)) and proper nouns (6). The subject noun is italicized in the examples.

(2) in-naaga ɪl-lɪ tə-rkab miš ɪl-bišiɪr yə-rkab
DEF-camel which 3F-ridden not DEF-male camel 3MSG-is ridden
‘The female camel is the one that is ridden (in a race). The male camel is not ridden’.

(E)

(3) fa-trat ɪl-saaj-ha ṭ ẓawwil
period recovery-her 3F-last long
‘The period of her recovery will last a long time’. (K)

(4) ti-ji-i-na ṭanāawb-aat
3F-comes-1PL duty-PL
‘Our special turn of duty comes’ (lit. turns of duty come to us). (E)

(5) yi-ḥi-i-oon-a, ba’i-deni i-sīir, ɪtámar, wa
3-put-PL-it then 3MSG-becomes, “tamar” and yi-tlaʕi minn-a dibs
3MSG-comes out from-3MSG molasses
‘When it [the ripe dates] become very black and dry, they put it in it [the container] and then it becomes ‘tamar’ and molasses excretes from it’. (E)

(6) šeex zayd kaan mɔnazzil qaraar
sheikh Zayd 3MSG-was AP-set down decision
‘Sheik Zayd had stipulated a decree’. (E)

In addition, nouns are distinguished from pronominals and personal pronouns. Pronominals include quantifiers, partitives, numerals, demonstrative pronouns, (e.g., ḥad ‘someone’, ma ḥad ‘no one’, kaθiir min ‘many of’ (7), waahid ‘someone’, ayy X ‘anyone’, baściD X’some of’ (8), kull ‘all’ (9)), and demonstratives when used alone as subjects. Personal pronouns are illustrated in (10).
Many girls marry directly from secondary school.

And some of the games differ from others.

We all stay together.

When he raised the car on two tires, right, he turns.

Constructions that are categorically SV, including headless relatives and wh-word subjects, are excluded from the statistical analysis, as are subjects of such nonverbal predicates as active participles or existential constructions (see (24) and (25)). Because of structural restrictions on the occurrence of overt subjects in them, relative clauses are also excluded, though subjects in conditional and time adverbial clauses, where the choice of subject occurrence is free, are included. In addition, we deal with overt subjects only. Arabic is a prodrop language, and in the current corpus, null subject constructions outnumber overt subject constructions by a 2-to-1 margin. It remains to be worked out how null subjects interact with and upon the SV/VS parameter treated here.

Statistical overview

A multivariate regression analysis using Goldvarb X (Sankoff, Tagliamonte, & Smith, 2005) was carried out to identify factors significantly correlated with SV/VS word order. Given the lack of quantitative treatments of this topic in Arabic, we began with a relatively large number of exploratory independent variables.

Morpholexical class: noun, pronoun, pronominal. Pronouns, typically having “given” referents, were hypothesized to favor SV order, as in Brazilian Portuguese. Among nouns, definite nouns were expected to favor SV order, indefinites VS.

Transitivity: copula, intransitive, transitive, transitive without object. Intransitives and copula constructions, mirroring presentatives will favor VS, and transitives will favor SV because the typical new element will align on the opposite side of the subject; in its verb-object (V-O) order, the object parallels the S of VS in its informational status, tending, like O, to be new.10

Number of words in clause: up to four (the mean and median for clauses containing overt subjects) versus more than four words. The SV clauses were expected to contain more words than VS clauses, as VS is hypothesized to coincide with introduction of new referents and SV with elaborating descriptions of discourse-old referents.

Complement of gaal ‘say’: yes/no. Complements of gaal were expected to favor SV because one function of the subject of the complement clause is to demarcate the narrated proposition in the complement clause from the narrator subject of the verb say itself.
Dialect: Emirates, Kuwait, Jedda. We had no a priori reason to predict any particular dialect-based pattern, though given the general tendency of modern Arabic dialects to have fixed SV word order (Egypt, northern Mesopotamia), a tendency toward SV among some of those tested here can be expected.

Relationship between subject referent and previous subject referent: same/different. Different subject referents were hypothesized to favor VS, new subjects being introduced in VS order; same subject referents were expected to favor SV, typically having a contrastive function.

Occurs in a conditional or adverbial clause: yes/no. Typically containing discourse-old information, adverbial clauses were hypothesized to favor SV order.

Immediately follows a change of speaker: yes/no. In our dialogistic texts (vs. the narrative texts), a new speaker will continue the same topic. This will favor SV order, as the conversation continues previously invoked referents, and may invoke the “discourse contrast.”

Table 3 shows morpholexical class, transitivity, gaaL complement, and clause length to be significant factors. The input value over 0.8 indicates a strong overall tendency toward SV. Morpholexical class is a highly contrastive factor group, with nouns favoring VS and pronouns and pronominals favoring SV. We will not further explore the transitivity or gaaL variables here.

Turning to a more detailed look at the morpholexical class categories, the high factor weights associated with pronouns and pronominals indicate that they overwhelmingly favor SV order. Nouns, on the other hand, show considerable variation between SV and VS. Table 4 breaks the distribution of nouns into the four classes noted: proper, indefinite, definite, and possessed (idaafa). Here it emerges that it is not nouns in general, but rather definite nouns—proper, definite, and definite by possession—that favor SV. Indefinites, on the other hand, favor VS.

**Discourse status: General vs. specific nouns**

Figure 1 extrapolates the results in Tables 3 and 4, moving from categories most favoring SV to those favoring VS. In this scale, up to “definite noun” there is a dominant clustering in favor of SV. Both definite and indefinite nouns, on the other hand, occur in large numbers in both SV and VS word order. We therefore turn to a closer investigation of the SV/VS variation among these nouns.

We begin by examining a property of the nouns themselves. In Naro and Votre’s study, subject nouns were classified as “available,” because these nouns are readily accessible to the hearer (1999:88). It appears that Naro and Votre treated “availability” as an inherent lexical property; for instance, a noun of unique

<table>
<thead>
<tr>
<th>Category</th>
<th>SV</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>pronoun</td>
<td>&gt;</td>
<td>pronomin &gt;</td>
</tr>
<tr>
<td>pronomin</td>
<td>&gt;</td>
<td>proper noun &gt;</td>
</tr>
<tr>
<td>proper noun</td>
<td>&gt;</td>
<td>definite noun &gt;</td>
</tr>
<tr>
<td>definite noun</td>
<td>&gt;</td>
<td>indefinite noun</td>
</tr>
</tbody>
</table>

**FIGURE 1.** General trends: word order and subject type.
reference such as Saturday is considered available. In this section, we expand upon the availability parameter, looking at it in terms of discourse context. We use the descriptive labels general and specific, as the term availability suggests an insight into the cognitive disposition of speakers that we do not have access to. In our treatment, general and specific pertain to discourse-specific entities. One and the same word can be general in one context and specific in another. Judgments need to be made on an individual basis. In (11), il-bint is part of a discourse about marriage problems in general. Il-bint represents a generic class.

(11) mumkin il-bint t-ruuh t-guul hagg umm-ha
possibly DEF-girl 3FS-goes 3FS-tells POSS mother-her
‘It’s possible that the girl [any girl in that situation] will go and tell her mother’. (K)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Weight</th>
<th>% SV</th>
<th>N (SV/VS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitive without object</td>
<td>0.65</td>
<td>90</td>
<td>62/7</td>
</tr>
<tr>
<td>Transitive</td>
<td>0.65</td>
<td>88</td>
<td>312/41</td>
</tr>
<tr>
<td>Intransitive</td>
<td>0.45</td>
<td>74</td>
<td>269/96</td>
</tr>
<tr>
<td>Copula</td>
<td>0.28</td>
<td>54</td>
<td>105/89</td>
</tr>
<tr>
<td>goal Complement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.91</td>
<td>98</td>
<td>47/1</td>
</tr>
<tr>
<td>No</td>
<td>0.47</td>
<td>75</td>
<td>701/232</td>
</tr>
<tr>
<td>Number of Words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater than 4</td>
<td>0.62</td>
<td>83</td>
<td>317/63</td>
</tr>
<tr>
<td>Up to 4</td>
<td>0.43</td>
<td>72</td>
<td>431/170</td>
</tr>
<tr>
<td>Morpholexical Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronominal</td>
<td>0.69</td>
<td>90</td>
<td>70/8</td>
</tr>
<tr>
<td>Pronoun</td>
<td>0.67</td>
<td>90</td>
<td>409/44</td>
</tr>
<tr>
<td>Noun</td>
<td>0.3</td>
<td>60</td>
<td>269/181</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>981</td>
</tr>
</tbody>
</table>

Input = 0.83; log likelihood = -433.309; chi-square/cell = 1.0600.

<table>
<thead>
<tr>
<th>% SV</th>
<th>% VS</th>
<th>N: SV/VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proper</td>
<td>81</td>
<td>19</td>
</tr>
<tr>
<td>Definite</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>Possessed</td>
<td>68</td>
<td>32</td>
</tr>
<tr>
<td>Indefinite</td>
<td>36</td>
<td>64</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>41</td>
</tr>
</tbody>
</table>
On the other hand, the formally same noun in (12) describes a particular girl in a party. Here it is considered specific.

(12) wi tu-dxul il-bint yaʕni ti-msik al-kaasa
and DEF-girl DM 3F-grabs DEF-glass
‘and the girl [in the story] enters, and ya know grabs the glass’ (J)

Judging whether a given subject is about a general class or a specific event or action entails looking at the larger context of the story, as well as individual parts of the clause. Two points can be noted here. First, the microstructure of the clause will often give clues as to the status of the noun. For example, the subject in (13) was considered specific, as the word ihni ‘here’ localizes the referent.

(13) yiʕjib-ni ʕañir il-burțqaal ihni
3MS-pleases-1SG juice DEF-orange here
‘I like orange juice here’ (lit. orange juice here pleases me). (K)

Second, the nature of the larger discourse context is an important criterion for classifying individual subjects. As a larger frame of reference, we will use the idea of episode, or section of discourse relating to a particular topic; we will assume this term rather than attempt to motivate it operationally. Consider the following episode from the Jedda recording, cited at length here as it will take part in the discussion that follows as well. The episode is first about parties to find brides in Saudi Arabia, and it then moves on to a description of the wedding ceremony, which is excerpted in (14).

(14) yibda ilʕariis w ahyaanan i-jiib-u
appears DEF-bridegroom and sometimes 3-bring-PL
l-mumlik, i-guüm il-mumlik y-imsak, aaa, yadd
DEF-officiant 3M-rise DEF-officiant 3M-take aaa hand
il-hurma w yi-gra l-quɾʔaan, w baʕdeen i-șiir
DEF-woman and 3M-read DEF-Koran and then 3M-happen
il-faráh w ii-ji abuu-ha w yii-j-u
DEF-party and 3M-come father-her and 3-come-PL
axwaana-ha
brothers-her
‘the bridegroom appears and sometimes they bring the religious officiant and then the officiant stands up and takes the hand of the woman [in the hand of the bridegroom] and he reads the Quran and then the party takes place and her father comes and her brothers come...’ (J)

In our classification, actors in episodes describing fixed frames—culturally familiar, formulaic events—represent specific, not general entities. In (14), for instance, the individual subjects are considered specific actors because they fulfill an essential role in the wedding event. Abuuha ‘her father’ would not be felicitously substituted for by another term. In the same passage, it would be
grammatically possible to substitute the indefinite *mumlik* for the definite *il-mumlik*, though this would change the meaning, suggesting that any officiant could do the job or that there was more than one officiant at the event and an arbitrary one performed the ceremony. The definite article conveys the idea of the officiant who has been chosen for that particular event and indicates that in such an event, one officiant typically presides. There is a genericness to the event, but it resides in the overall frame of “wedding ceremony.” The individual participants are specific to that event. Where an event frame has been established, individual participants will, in general, be marked by a definite article, and this marking establishes their “given” status in discourse about that event.

This episode-dependent definiteness marking is further in evidence in the Al-Rawi Emirati corpus. The corpus is presented in such a way that each story is a self-contained episode: a folktale or a set of instructions about how to make a tent, for instance. In this corpus, the definite nouns predominate, 71 definites as compared with only 6 indefinites. Most of the stories are about specific individuals participating in events or descriptions of specific processes. Text 3 (Al-Rawi, 1990:128–130), for instance, is a description of wedding ceremony, event by event. The sentence:

\[(15) \text{ʕaad y-aw in-naas xaðoo-ha} \]
\[DM \text{ came-3PL DEF-people take-3PL-3F} \]
\[\text{‘Then the people came again and took her’}. \quad (E)\]

contains the word *naas* ‘people’, which in our corpus tends to be used in a general sense.14 Here, however, it refers to the people of the village, who have a role to play in the unfolding wedding. The adverb ʕaad ‘again’ references their previous appearance in the event. In general, the subjects in the Al-Rawi texts are specific individuals, and we therefore forego presenting a statistical count. Interestingly, however, in the one text where lexical noun SV order equals VS order, Text 1, each SV token represents a general statement, a part of a description about how life once was, as in (16).

\[(16) \text{tawwa in-naas širb-aw min il-beelar} \]
\[formerly DEF-people drank-3PL from DEF-reservoir \]
\[\text{‘Formerly people drank from the water reservoir’}. \quad (E)\]

Even allowing for an episode-specific interpretation of the general/specific contrast, it still is striking that nouns most frequently classified as general are those with an inherent general lexical meaning, such as *naas* ‘people’, and inversely, those classified as specific have a specific reference (Lyons, 1995). The dichotomy thus clearly has semantic and text-pragmatic bases. Representative examples of our semantic-pragmatic classification of nouns are as follows.

\[(17) \text{Characteristic nouns classified as general or specific} \]
\[\text{general} \]
It should be noted that discourse specificity is independent of the definite/indefinite distinction. In (18) and (19), the indefinite nouns are specific, as they imply essential individuals in the story. In (18), the bicycle is a main participant in the entire episode, and in (19), the suggestion is that one particular man, not just anybody, represents the respective tribes in the court.

(18) kaan ʕində-na seekal
3MS-was at-1PL bicycle
‘we had this bicycle’ (E)
(19) iruuħ rayyaal min ha l-jibiila
3MS-goes man from this DEF-tribe
‘a man from this tribe goes [to the court]’ (E)

By contrast, in the following, the indefinite noun bint ‘girl’ is considered general.

(20) iða bint ti-ʕrif it-γanni ti-msik l-makrafoon
if girl 3F-knows 3F-sing 3F-grabs DEF-microphone
wi t-γanni
and 3F-sings
‘If a girl [any girl at the ceremony] knows how to sing, she takes the microphone and sings’. (J)

Possessed nouns tend overwhelmingly to be specific, for example, in (14), abuuha ‘her father (of girl in story)’, so statistics will be given only for the definite and indefinite nouns, excluding possessed nouns.

Table 5 shows the significant factors in a Goldvarb analysis incorporating the same exploratory variables as used in the previous discussion for the entire corpus, except that (i) a variable encoding general/specific is added, and (ii) morpholexical class is removed, because only nouns are evaluated. The total nouns are fewer than those in Table 3, as we excluded those for which the general/specific status was unclear. The general/specific variable is significant: general nouns favor SV, specific ones VS. Further, definite nouns significantly favor SV, indefinite VS. These results lend support to the view that lexical semantic and discourse-pragmatic factors both influence word order in spoken Arabic. The significance of transitivity and number of words mirrors the results for the entire corpus (Table 3). The nonsignificance of the variable marking
complements of *gaal* ‘say’ is probably due to the low token count: there are only nine overt noun subjects in this position (eight occurring in SV constructions).

Similar regression analyses were applied to the Al-Rawi corpus, with some differences required by the smaller corpus size and text format. The variables “*gaal* complement” and “occurrence in adverbial clause” were excluded, as were copular constructions and transitive constructions without objects in the transitivity factor group. Dialect and change of speaker information are either not provided or not relevant in the Al-Rawi corpus. The remaining factors are morpholexical class, number of words, general/specific, and same/different subject referent. The results for significant factors are shown in Table 6.

Although the overall set of independent variables only slightly biases the sample toward SV, two factors stand out: shorter clauses favor VS, and pronouns and pronominals strongly favor SV. Thus, two of the three factors significant in our full corpus (Table 3) are also significant in this corpus.\(^{15}\) In fact, these two variables display a more dichotomous breakdown of SV/VS categories: short clauses and those with noun subjects are strongly VS whereas longer clauses and those with pronouns and pronominals are SV. Because there are only six tokens of indefinite noun subjects, not a great deal can be read into its lack of significance.

The analysis of nouns only, taking away subject type and adding general/specific, gives results parallel to those in our own corpus (Table 5), with general showing 69% (9 of 13) and specific 21% (10 of 47) SV order. However, as noted, with only six indefinite subject nouns in the entire corpus, not a great deal can be read into this.

All in all then, allowing for inherently unresolvable issues arising from small sample size and different presentational formats, nothing in the Al-Rawi corpus contradicts our own results. Any significant variable influencing SV/VS in the

| Table 5. Significant factors contributing to SV order in Goldvarb analysis, nouns only |
|-----------------------------------------|---------|---------|---------|
| Factor                                | Weight  | % SV    | N (SV/VS) |
| General/Specific                       |         |         |          |
| General      | 0.64     | 69      | 74/34    |
| Specific     | 0.43     | 49      | 101/107  |
| Definiteness                       |         |         |          |
| Definite     | 0.6      | 63      | 144/83   |
| Indefinite   | 0.27     | 35      | 31/58    |
| Transitivity                          |         |         |          |
| Transitive without object             | 0.88     | 92      | 12/1     |
| Transitive     | 0.69     | 76      | 64/20    |
| Intransitive           | 0.5      | 55      | 72/59    |
| Copula        | 0.26     | 31      | 27/61    |
| Number of Words in Clause             |         |         |          |
| Greater than 4 | 0.61     | 64      | 75/42    |
| Up to 4        | 0.44     | 50      | 100/99   |
| Total                       |         |         | 316      |

Input = 0.572; log likelihood = −177.536; chi-square/cell = 0.6825.
Al-Rawi corpus is also significant in ours, and the direction of significance relative to SV/VS tends in the same direction in both cases. Where the Al-Rawi corpus does not have parallel significant categories, the lack of adequate tokens in her data renders interpretation opaque. Given the independent status of the Al-Rawi corpus, separated from ours by some 20 years, the results of the two Goldvarb analyses become highly corroborative.

**DISCOURSE FUNCTIONAL INTERPRETATIONS**

We can recapitulate the results of our quantitative analysis as follows. Pronouns and pronominals as well as personal nouns are overwhelmingly SV; general nouns strongly tend to be SV and indefinite nouns VS. A statistically characterized category that does not display a clear tendency is the definite, specific noun category. To broaden our understanding of this category, we move away from a strictly statistical orientation toward a discourse characterization. Here, observations from the previous section will be worked into functional generalizations, and new discourse functions will be identified and integrated into the description.

**SV word order**

We can isolate three general factors impinging on SV order, beginning with the most categorical ones. A first element favoring SV order is the lack of specificity of reference of the subject. This pertains in particular to our class of pronominals, which is heavily weighted toward SV. In the case of quantifiers, as illustrated in (8), no individual in particular is identified because the semantics of the quantifier specify an undefined individual in a group. Further, the regression analysis in Table 5 showed that among nouns as well, nouns with the semantic-pragmatic property of “general” favor SV order.

Headless relative pronouns in as far as they define general classes of objects and individuals (e.g., “those who …”) could be accommodated in this second category
as well as Q-word subjects, which do not have a specific referent. Both of these are invariably SV.

A second factor, situational givenness, often denoted by a pronoun, is relevant to SV order. The use of first and second person pronouns is usually situationally rather than referentially motivated. The first person pronoun announces attitudinal or turn-specific nuances such as degree of commitment toward a point of view, claiming the floor for speaking, or assertion of authority (see text in Appendix). A second person often has a generic referent or indexes the speaker’s direct addressee. Again, in these functions, the subjects do not represent participants in events but rather are derived from the status of the speaker and addressee as participants in the speech situation, or in the case of generics, are open placeholders. Third person pronouns are inherently referential via their anaphoric function.

Finally, SV order signals contrasts. On the one hand are explicit contrasts between nouns belonging to the same set (Myhill & Xing, 1993:37). In (2), for instance, there is a contrast between the female and male camel. In (21), the speaker is contrasting the status of a woman in Saudi Arabia as opposed to a man implicitly and explicitly her status in the era of the Prophet as opposed to today. In all instances, the repeated noun subject *al-hirma* occurs in SV order. After referring critically to the *šeexs* (i.e., men) of the *mutatawwi‘ī*, the nickname of an organization charged with maintaining public morals, she goes on to contrast the women’s situation at the time of the Prophet.

(21) a. min ayyaam ar-rasuul ʕalee-h is-samaat wi s-samaam
from days DEF-prophet on-3MSG DEF-prayers and DEF-peace
kaan-at *al-ma‘īta* t-biiʕ
was-3F DEF-woman 3F-sell
‘in the days of the Prophet, peace be upon Him, a woman would sell …’ (J)

Three turns later the constraints on women are introduced in SV order:

(21) b. *al-hirma* ma t-suug, *al-hirma* ma ti-rkab
DEF-woman NEG 3F-drive, DEF-woman NEG 3F-ride
sayyaara illa ma‘a-ha muhram, *al-hirma* ma
car unless with-3F male escort DEF-woman NEG
t-saafir illa ma‘a-ha riijaal yi-waddii-ha wi
3F-travel unless with-3F man 3MSG-bring-3F and
i-jiib-ha,
3MSG-take-3F
‘A woman doesn’t drive, a woman doesn’t ride a car unless she has a male escort with her, a woman doesn’t travel unless she has a man to bring and take her, she can’t sit in a restaurant on her own, she can’t drink a cup of tea anywhere unless there is a male escort with her’. (J)

There is thus a double contrast between the woman (*al-hirma*) and the preceding subjects:
Men of mutaqawwif vs. the woman (al-hirma) of Saudi Arabia today
Women of the Prophet’s day vs. the woman (al-hirma) of Saudi Arabia today

A second type of contrast is in play when SV order picks up “semiactive” participants (Chafe, 1987, 1994) and brings them back into a conversation in SV order. This can broadly be termed discourse contrast, the subject simultaneously reintroducing a semiactive participant and signaling a contrast with the subject of the immediately preceding predicate. The following story is about a woman who picked up the speaker as a hitchhiker. A woman driver is introduced in the story.

(22) a. wəg-f-at əs-sayyaara, in-zeen, ba’d-deen əna jii-t, a-a harma
stopped-3FSG DEF-car okay then I came-Iaaa woman
‘The car stopped, okay, then I came over [to the car], [it was] a woman [driver]’! (E)

The initial novelty resides in a woman picking up a male hitchhiker. There were other riders in the car, and the story then turns to what transpired inside the car among the passengers. When it returns to the subject of the woman, describing her driving habits some 318 words later, the subject noun is reintroduced in SV order.

(22)b. …ʕāt-t-ee-hum gam ʕalič-oon, il-harma yam ṭa-jaawiz
gave-1SG-3PL gum 3-chew-PL DEF-women when 3F-passes
‘… I gave them gum to chew on. The lady, when she would pass (other cars) …’ (E)

In all in our corpus, there are 33 noun subject tokens that we have classified as marking contrast, all being SV, 22 specific as in (22), 11 general as in (21).

Summarizing these observations, SV order has three prominent macrofunctions. In the first and second, the identity of the individuals is either irrelevant (pronominal quantifiers, general nouns) or given by speech or text situation (pronouns). The pronouns convey attitudes and degrees of commitment toward what the interlocutors are saying, and the quantifiers might be thought of as subject placeholders signaling that an event is occurring, though the identity of the one(s) carrying it out are important only as representatives of a general class. In the third, it is used to pick out, highlight, and contrast individuals, A versus B, whose identity is known but whose referential status needs to be specified. A common denominator linking these three cases can be termed, adopting a term from Naro and Votre (1999), “available referentiality.” Availability is situationally (first and second person pronouns), quantificationally (pronominals), textually (third person pronouns, contrastive nouns), or semantic-pragmatically (general nouns) sanctioned.

VS order
Turning to VS order, it is useful to distinguish indefinite and definite nouns. As seen in Table 4, indefinites occur more often in VS than in SV order. One
prominent function of VS order is to introduce new referents into the discourse. Beginning with these, in the majority of cases in the Arabic corpus, the introduction of the presented noun takes place in VS order under two discourse conditions. First, the clause may introduce a key element in a story. In (4), for instance, the subject $mənaawəbaat$ ‘rotational duty’, refers to a specific activity, which in fact becomes the central theme of the unfolding story. This can be seen in the dialogue subsequent to the excerpt in (4), given in (23), where aspects of rotational duty, including its synonym $twaari$ are added.

(23) A: ti-jii-na $mənaawəbaat$ ana a-noowəb bi yoom
   3F-comes-1PL duty-PL I I-take turn on-day
AR: $twaari$
A: hee
AR: le t-$twaari$ ya$ə$nī
A: ‘Our special turn of duty comes. I take my turn on a day …’
AR: ‘emergency duty’
A: ‘yeah’
AR: ‘for emergency duty that is’. (E)

Second, the clause may be part of a listlike sequence of actions or states. In (5), the speaker describes the process of making $dibs$ ‘molasses’, beginning with ripe and ending with dried dates. The results in the process, $tamar$ ‘dates’ and $dibs$ ‘molasses’, are introduced in their turn in the appropriate place of production. The story is not about $tamar$, nor $dibs$, but about the entire process, and the relevant elements will be introduced in VS order.

The VS order coincides in terms of predicate-S sequence with the prototypical presentative construction, in which a nonverbal predicate, either existential $fi$ (24) or the predicate $ʕind$ ‘have’ (25) (lit. ‘at-X’) occurs before the subject.

(24) $fi$ $had$ i-$təbbəl$
   exist one 3M-drums
‘There is a drummer’ (lit. someone who drums). (E)

(25) $ʕind-a$ $ʃəssa$ u $ʕind-a$ $tifag$ u $ʕind-a$
   at-3MSG stick and at-3MSG pistol and at-3MSG
$ʃəfəf$
   sword
‘and he has a stick and a pistol and a sword’. (E)

This construction is a prototypical means of introducing individual referents. It typically is an indefinite subject (214 of 244 presentative subjects in our corpus), an unmodified noun (25), except, as in (24), that it may be qualified by a relative clause. The indefinite noun + relative clause consist not infrequently of a general operator (nouns such as $had$ ‘someone’, $naas$ ‘people’, $ʃey$ ‘thing’). This
ostensibly contradicts the tendency of VS (in this case, predicate-S) not to have a quantifier or general noun as subject. However, in this case, the lexical content is specified in the following descriptive relative clause. In (24), it is not ‘any old person who drums’, but rather ‘a drummer in this particular occasion’ (describing the *razfa*, a type of dance). For comparative purposes, regarding indefinite VS order, numerous studies, going back at least to Hetzron (1973) have shown that many languages have this order for discourse-new presentatives. It is not only indefinite nouns, however, that occur in large numbers in VS order; definite nouns do as well. To account for this further distribution, it is relevant to first examine the distribution of word order across three morpholexical classes in the Al-Rawi Emirati corpus in Table 7. Compared with the parallel data for our corpus (percent SV in Table 3), all of the three morpholexical classes in the Al-Rawi corpus show lower SV rates. This is true for nouns in particular. Further, among 71 definite subject nouns in the Al-Rawi corpus, 30% are SV and 70% are VS, as compared with 65% SV and 35% VS among definite nouns in our corpus.

The stronger VS presence in the Al-Rawi corpus can be understood, we suggest, in terms of a particular type of episode, *event-based frames*. Myhill (1992:265) argued that “there is a universal correlation between temporal sequencing and verb-subject word order.” Temporal sequencing is implied in certain frames or schema (Chafe, 1987:29; Lambrecht, 1994:99; Minsky, 1975; Quakenbush, 1992; Sperber & Wilson, 1995:88). In the Al-Rawi corpus, 9 of the 10 episodes are about sequentially ordered events or processes, either folktales or procedural instructions about how to perform various traditional activities. In each of these 9 stories, VS order among lexical nouns dominates. The final story has equal SV and VS (see (16) for a characteristic SV example from 1 of these episodes). The per-episode statistics for definite and indefinite nouns are given in Table 8.

In our own corpus, event-based frames constitute a small minority of all episodes, though when they occur the effect is striking. The excerpt given in (14) from a description of the speaker’s recollection of weddings in Saudi Arabia in the 1980s is characteristic. In this episode, it is not one individual action that is implied but rather a sequence of actions in a larger episode. Table 9 breaks down the SV/VS sequence in this episode by morpholexical category.

The VS dominance in this event frame is best seen in contrast with the SV/VS distribution in another (non-event-based) episode from the same recording, shown in Table 10. This episode is a very argumentative segment in which one of the interlocutors criticizes Saudi Arabia, often quite vociferously from a number of perspectives including its outmoded educational system and its terminology for the Arabian or Persian Gulf. This episode, like many in our corpus, contains a large number of speaker turns. (See the Appendix for an excerpt from this episode). In our own corpus, the episode-specific distribution in Table 10 is the norm, though intraspeaker variability can be high.
Definite, specific nouns

We have concluded that: (a) situational givenness, realized as first and second person pronouns, favors SV; (b) lack of specificity, realized by pronominals and general nouns, favors SV; (c) contrastive subjects are realized as definite nouns favoring SV; and (d) the introduction of new referents, particularly in event-based frames, is often affected by indefinite nouns in VS word order.

### Table 7. Word order and morpholexical class in the Al-Rawi Emirati corpus

<table>
<thead>
<tr>
<th></th>
<th>% SV</th>
<th>% VS</th>
<th>N: SV/VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronouns</td>
<td>79</td>
<td>21</td>
<td>27/7</td>
</tr>
<tr>
<td>Pronominals</td>
<td>79</td>
<td>21</td>
<td>15/4</td>
</tr>
<tr>
<td>Nouns</td>
<td>30</td>
<td>70</td>
<td>29/69</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>151</td>
</tr>
</tbody>
</table>

### Table 8. SV/VS frequency, definite and indefinite nouns, in 10 episodes, Al-Rawi corpus

<table>
<thead>
<tr>
<th>Episode</th>
<th>Frequency SV/VS</th>
<th>% VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3/3</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>4/8</td>
<td>67</td>
</tr>
<tr>
<td>3</td>
<td>3/8</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>2/7</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>1/2</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>0/1</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>2/3</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>9/14</td>
<td>61</td>
</tr>
<tr>
<td>9</td>
<td>0/1</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>0/6</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>24/53</td>
<td>69</td>
</tr>
</tbody>
</table>

### Table 9. Number of SV and VS clauses in two Jedda event-based frames

<table>
<thead>
<tr>
<th>Episode</th>
<th>Subject</th>
<th>SV</th>
<th>VS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parties to find brides, grooms</td>
<td>Pronoun</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Pronominal</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Definite N</td>
<td>3</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Indefinite N</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Possessed N</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>13</td>
<td>15</td>
<td>28</td>
</tr>
</tbody>
</table>

Definite, specific nouns

We have concluded that: (a) situational givenness, realized as first and second person pronouns, favors SV; (b) lack of specificity, realized by pronominals and general nouns, favors SV; (c) contrastive subjects are realized as definite nouns favoring SV; and (d) the introduction of new referents, particularly in event-based frames, is often affected by indefinite nouns in VS word order.
Still, definite, specific nouns, unless they are contrastive, continue to escape a unitary profile. Specific nouns alone in our corpus divide almost evenly between SV and VS clauses (101 vs. 107). Incorporating further discourse-specific parameters, however, goes some way toward a better characterization of this class. Here we note three factors.

The first consists of semantically specific cases, 10 tokens in our corpus, where temporal subjects occur in VS clauses:

\[(26) \text{bid-at} \text{ it-kuun} \text{ is-saa}'a} \text{sitta} \text{ wa} \text{ naßs} \\
\text{started-3F 3F-be DEF-time six and half} \\
\text{‘The time got to be half past six’ (E)}\]

A second factor derives from the presentational status of VS in predicate-subject order, which when linked to an active verb calls attention to the event itself. In stylistic terms, a dramatic highlighting results. For example, in the following description (27) of how a juicer works, the subject of the first clause, \(l-\text{igšuur}\) ‘the skins’, is introduced as an isolated topic, then repeated in an SV clause (available referent), and the culminating action, in italics, is presented with a VS clause.

\[(27) \text{l-igšuur} \text{ kill-ha} \text{ il-jøzy} \text{ il-xalfi} \text{ min} \text{ al-}\text{Yaşṣaara} \text{ maalt} \\
\text{DEF-skin.PL all-F DEF-piece DEF-back from DEF-juicer POSS} \\
\text{il-jizar} \text{ haaði} \text{l-igšuur} \text{ kill-ha} \text{ kill-ha} \text{ it-ruuh hinaak wa} \text{ð} \\
\text{DEF-carrots this.F DEF-skin.PL all-F all-F 3F-goes there and} \\
\text{yi-nzil-l-ič} \text{ il-juus} \text{ şaafii} \\
\text{3MSG-comes down-to-2FSG DEF-juice pure} \\
\text{‘All the skins, the back part of the carrot juicer, these skins all of them all of them go to that side and the juice comes out for you pure’. (K)}\]

In a second example (28), the speaker is describing events leading up to a girl’s engagement. The speaker emphasizes the element of chance: \(miin idugg il-baab Yaşsan yaxtub-ik\) ‘who is going to knock on the door to ask for your hand’. Then, as if to emphasize the unpredictability of the event, the speaker uses the
following VS clause to introduce the decisive event (the mother of the groom coming to ask about marriage on behalf of her son).

(28) baʾdeen ti-żii-ki l-umm 
    then 3F-come-you.F def-mother
‘then the mother (of the would-be-groom) comes to you’. (J)

Third, as seen in Table 5 and as noted for the Al-Rawi corpus, there is a correlation between general subjects and longer clauses and specific subjects and shorter clauses. Elaboration and explanation are more likely to occur in SV than in VS format, although the presentation of an event is effected parsimoniously. Example (27) illustrates this division of labor. The situation, the description of the machine that makes carrot juice, is given in SV order, but the concise result is presented in a VS clause containing four words.

The distribution of morpholexical classes relative to discourse-functional categories can be recapitulated in the following representation (Figure 2). The SV order can broadly be said to signal available referentiality, whereas VS has the presentational function of slotting in referents one by one as they appear in narratives, descriptions, or at certain dramatic points in events. In the figure, the relative strength of association between word order and morpholexical class is represented by shading: the darker the line, the stronger the association.

The available/presentational paradigm represented in Figure 2 echoes the basic given/new dichotomy (Prince, 1981) that has been influential in studies of discourse status of nominals, including that of Naro and Votre (1999), which we will examine more closely. The Arabic data provide evidence that the discourse-functional dimension, though critical, does not operate independently from the morpholexical dimension. That discourse status of subject as available/presentational can tease apart morpholexical categories is exemplified in (29), taken from the same episode from which (14) was extracted. In this example, we see a strongly SV-disposed lexical category, pronominals, being overridden into VS order by the “right” discourse context: the pronominal subject aktar min occurs after the verb:

(29) yimkin y-ii-j-u aktar min arbaʾa aw xamsa min l-ummahaat
    perhaps 3-come-PL more than four or five from DEF-mothers
‘Perhaps more than four or five of the mothers might come’. (J)

At the end of our analysis, there remains a residue of definite specific noun subjects (about 70 in our corpus) that occur in SV clauses and that we have not associated with a discrete discourse or semantic-pragmatic function. They are not contrastive, and they have specific reference (e.g., (22)). We believe that understanding these
requires a more detailed contrastive study of genre. The VS clauses are closely associated with discourse functions, including the introduction of new referents and the presentation of unfolding events and procedures. Beyond the specific functions already summarized, we suspect SV clauses are, by contrast, associated with a dialogistic mode of conversation: taken as a whole, our own texts are distinguished from Al-Rawi’s in containing a great deal more dialogue (see Appendix). This intuition requires greater research.

MORPHOLEXICAL CLASS, DISCOURSE, AND WORD ORDER: A COMPARATIVE SUMMARY

Earlier, we discussed that morpholexical class is a broad predictor of SV/VS word order in Peninsular Arabic, echoing a pattern that has been documented in a number of languages. Finer-grained comparisons with other languages based on corpus data are difficult, however, either because such studies are lacking or because a different categorical vocabulary is used. For this reason, the study of Naro and Votre (1999) will be invoked for a detailed comparison, as we have orientated our own study according to theirs, precisely to bring out typological similarities and differences with another language with pragmatically controlled word order.

Three points of overlap and contrast emerge between our corpus and that of Naro and Votre (1999). First, in its lexical realization, our class “available referentiality” corresponds in part to their “evoked,” comprising pronouns and previously mentioned nouns (cf. our “contrast” category). These correspond both in terms of lexical class and in their SV word order. Second, there is a lexical and
semantic overlap between our “pronominals” and their “indefinites,” as well as between our “general nouns,” and their “irrelevant reference.” In their word order implications, however, these two classes are treated in opposite ways in the two languages: VS in Brazilian Portuguese, SV in Peninsular Arabic. In (30), the Portuguese ‘all’ (which we include among pronominals) is postposed in VS order, for example. This can be contrasted to our example (9).

(30) Sumiram esses homem todo que botaram
‘All those men who installed it have disappeared’ (lit. have disappeared all those men)
(Naro & Votre, 1999:96)

Third, in both varieties, VS is a favored order for the introduction of brand new referents.

Naro and Votre (1999) had one more dimension in their treatment of word order: the global information status of the subject in the flow of discourse. They described this in terms of high and low communicative tension, identifying degree of topicality with communicative tension. For them, in Brazilian Portuguese, SV order indicates a high degree of topicality and VS a lower degree. Importantly, in our comparative account, VS order in Brazilian Portuguese appears whenever a low topicality subject is used; for instance, in (31), the hero of a Kung Fu film is suddenly confronted by a group of men whose identity for the continuing narration is unimportant:

(31) Ele estava perdido, assim. apareceu uns homens
‘He was lost, like that. Some men appeared’ (lit. appeared some men) (Naro & Votre, 1999: 80)

Though they showed correlations between word order and morpholexical class, Naro and Votre (1999) ultimately explained SV and VS order in terms of high topicality versus low topicality.

Lower topicality corresponds to local events. In a larger episode, the topic is established; the topical noun occurs in SV order; and nontopical, marginal, background subjects enter and leave the scene sporadically (see (30) and (31)). They illustrated this with a flow diagram (1999:84), which we reproduce schematically in Figure 3. As represented in Figure 3, an individual episode will be dominantly SV, punctuated by occasional lapses into VS.

A general discourse-based meaning of SV/VS word order provides a generalizing link to the individual category-based observations based on morpholexical class. The overall pattern Naro and Votre established is quite different from what we describe for Peninsular Arabic. Here the overall episode will trend toward SV or VS, correlating in part with the choice of lexical noun subject. The VS order further flags high points in individual events ((27) and (28)).

In looking to define the contrast between our data and Naro and Votre’s, one might attempt to look for differences at the general, discourse level. If, following Naro and Votre, topicality is taken as a general heuristic, it might be claimed, for
instance, that the two languages simply represent topicality in different ways. For instance, in Brazilian Portuguese topicality resides in SV, but in Peninsular Arabic, it resides in VS. However, it is hard to see what common denominator exists to postulate such a transformation of topicality from SV in one case to VS in the other. Recall, for instance, that pronouns and evoked nouns are strongly SV in Brazilian Portuguese, and these are, for Naro and Votre, topical. These same (in the case of pronouns) and analogous (in the case of evoked nouns) classes in our data also tend to be SV. In fact, in Peninsular Arabic, pronouns and lexical SV nouns are in some sense topical in the sense of Naro and Votre (see, for example, the function of contrast that is widespread in SV nouns in Peninsular Arabic). Differences in or redefinitions of topicality alone cannot account for the contrasts and similarities between our two cases.

What emerges here is an argument for a language- or variety-specific characterization of variable SV/VS order. Clearly, in our data, and in that of Naro and Votre, VS and SV can only be understood relative to each other. This alone relativizes certain global claims about the meaning of variable word order, for instance Myhill’s (1992) claim that VS order implies sequentiality. In many cases, including Peninsular Arabic, it does, but in Brazilian Portuguese, VS order provides a secondary counterpoint to topicality.

Nonetheless, broad themes emerge from this study. One is the role of morpholexical class in predicting word order. Pronouns appear to cluster in one order, as do pronominals. Lexical nouns are more variable, a point well recognized in studies on word order typology. The role of lexical and discourse specificity of the subject noun is less well explored in word order studies, though the broad parameter of availability is applicable both to our data and that of Naro and Votre. More general nouns tend to pattern with pronominals in both sets.

A second general theme is the role of the overall discourse function of SV/VS word order. This is a parameter that Givón tried to capture with his characterization of “topic continuity.” Though accepting that discourse function
is crucial, at this point, we distance ourselves from an assumption that there exists a broad enough arsenal of comparative studies to know all of the universal parameters that are in play. Comparing Brazilian Portuguese, as interpreted by Naro and Votre (1999), and our own data, two different factors are at work: topicality in the former case and presentation of referent in an event in the latter.

We suggest unifying these two parameters componentially, starting with the observation that in both Brazilian Portuguese and Peninsular Arabic, topicality in SV order shares the property of available referentiality. We propose this characteristic as the link in SV order between Brazilian Portuguese and Peninsular Arabic. The VS order has three aspects. First, Brazilian Portuguese and peninsular Arabic share a universal tendency to present new referents in VS order. Brazilian Portuguese alone uses VS for backgrounded topics and Peninsular Arabic alone uses VS for the presentation of events. This latter function shades into the presentational one, a division represented graphically by the broken line in the two cells between them in Figure 4. In these latter two functions, VS order is maximally differentiated: backgrounding in the case of Brazilian Portuguese and presentation of key referents in events in the case of Peninsular Arabic.

Coupling this with the morpholexical realization of these nouns, the typology in Figure 4 describes the two cases.

### FIGURE 4. Discourse status and morpholexical realization.

<table>
<thead>
<tr>
<th>discourse status</th>
<th>morpholexical realization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Brazilian Portuguese</td>
</tr>
<tr>
<td>SV: available referentiality:</td>
<td>pronouns, evoked nouns</td>
</tr>
<tr>
<td>VS: backgrounded topics:</td>
<td>indefinites (overlapping with our pronouns, general and generically available nouns)</td>
</tr>
<tr>
<td>VS: presentation of new referents:</td>
<td>indefinite nouns</td>
</tr>
<tr>
<td>VS: presentation of events:</td>
<td>lexically/discourse specific nouns (definite or indefinite)</td>
</tr>
</tbody>
</table>

NOTES

1. Cf. Pashova (2003) for a corpus-based treatment of Modern Standard Arabic in which the parameters of sequentiality (VS) and simultaneity/habituality are argued to play a role in literary Arabic as well.
2. However, Herring (1990) warned against a simple given/new reading of the topic/focus distinction for two reasons. First, there are subtypes of topics and foci that differ in their information status; for
example, shifted (as opposed to continuous) topics relay new or reactivated information. Second, Herring pointed out that the basic word order of a language in part predicts the relative order of topics and foci, and so the information-structure properties of a given SV order cannot be assumed to match across languages. Table 1 summarizes the results of four studies that have related SV/VS word order to referential distance. As expected, a given word order carries different information-based functions across languages.

3. Others include persistence, same subject/different subject from preceding subject, and in what order an entirely new entity is introduced (presentative).

4. Jacenik and Dryer (1992) distinguished two VS sequences—VS and VSX. The first is VS only, and the latter is VS + complement(s). There are considerable differences between VS and VSX; though for our expository purposes, only one measure will be exemplified.

The Biblical Hebrew study distinguished human and nonhuman subjects, and definite and indefinite nouns, without giving a common measure for them. We have chosen definite human nouns for the statistic, as they are the most frequent.

5. A reviewer suggests that languages with dominant SVO order may use VS for less topical subjects, and, symmetrically, VSO dominant languages may use SV for lower topicality subjects. Although this observation holds for some languages (Polish, Portuguese, Spanish among SVO languages, perhaps Biblical Hebrew for VS), its status as a generalization appears problematic to us. An obvious counterexample is found in English, a SVO dominant language. Analyzing inverted VS sentences of the type “Sitting and talking with an elderly man was your little brother” (1994:245), Birner found that inversion is constrained as “discourse older X – Verb – Subject = discourse newer” (1994:245). Here VS order marks discourse newness, not lower topicality.

The VSO languages often have SVO alternatives. Siewirska (1998:493) in a sample of 171 languages found that among her 24 dominant VSO languages (examining lexical noun only), 68% had an SVO alternative (though OVS at 58% is nearly as common an alternative). Classical Arabic is one such language, prototypically VSO with SVO order as well. In Classical Arabic, however, SV signals discourse relevant topicality, not backgrounded topicality. Herring (1990:167) in a survey of nine VSO languages found that an SV order was used to indicate focus or shifted topicality. More generally, Tallerman (1998:636), after an extensive survey of SV word order in Celtic and other languages that have been ascribed VSO status (at some level), concluded, “It seems more likely that there are no ‘true’ VSO languages, but that certain properties tend to recur amongst languages with VSO word order.” Given this state of research consensus, a generalization proposing a mirror-image treatment of topicality of the subject constituent between SVO and VSO languages appears to us premature.

6. The bare figures are Pro-V 14, V-Pro 3, versus Noun-V 80, V-Noun 100, giving a significance value of p < .004.

7. In a number of quantitative studies on word order variation, although morpholexical class is routinely included among the variables (e.g., Fox, 1983, on Biblical Hebrew; Jaggar, 1983, on Hausa), the delicacy of classificatory analysis is too coarse to allow the relative contribution of competing variables to be compared statistically.

8. The data use a very adequate phonemic transcription. Some editing apparently did occur, in contrast to our sample, though no corrections are undertaken. Al-Rawi (1990) noted grammatical incongruities, such as non-reference-induced shifts from second person plural to second person singular. This indicates that the basic integrity of the texts was maintained unedited.

9. Abbreviations are as follows: J = Jedda, E = Emirates, K = Kuwait, BP = Brazilian Portuguese, DEF = definite, DM = discourse marker, F = feminine, IMP = imperative, M = masculine, PL = plural, POSS = possessor particle, SG = singular, SV = subject-verb order, VS = verb-subject order, 1 = first person, 2 = second person, 3 = third person.

10. Thus, indefinite objects outnumber definite ones in our corpus 347 to 258.

11. A Goldvarb analysis substituting person (1, 2, 3) for subject type also yielded significant categories parallel to those in Table 3. First and second persons (being pronouns) aligned at SV, whereas third person, containing dominantly overt nouns, overproportionally has VS. Morpholexical class and person interact inherently and hence cannot be included in the same analysis.

12. The complement of gaal is an interesting, though specialized context that applies to only a relatively small number of tokens. Transitivity requires extensive treatment of its own, though we will touch on it.

13. As a point of orientation, we can cite Linell (1998:187), “relatively bounded sequences within the more comprehensive speech event or encounter as a whole.” In previous work, episode boundaries have been found to correlate with morphosyntactic variables including Chinese definite versus indefinite NPs (Liu & Zhang, 2004), pronominal versus nominal reference to central characters in Japanese (Sieg, 2004), and NPs versus zero anaphors in Japanese (Obana, 2003).
We have undertaken an extensive classification of 12 texts in our corpus according to episode to show that SV and VS order can be sensitive to episode, though closely integrating this into our discussion of discourse determinants of SV/VS word order would take us too far outside the immediate issues of this article. Our reference to episode in our texts refers to this classification, however.

14. We counted the occurrences of naas/in-naas as subject of a verb in the first 125 pages of Holes’s (2005) collection of Bahraini Arabic texts. All 13 tokens are in SV order. There is 1 VS token, but in a quantifier phrase, ista’lmalo ... kill in-naas ‘all the people use’ (2005:100).

15. Four factors are significant in Table 3, but one, the complement of gaal was left out of the Al-Rawi analysis for the reasons stated in note 12. The nonsignificance of transitivity is interesting, though the lack of adequate tokens in parallel categories makes interpretation difficult.

16. Ocampo (1993:351–362), describing French and Spanish presentatives, noted that a biclausal strategy is employed in both languages, in which the new referent is introduced alone in one clause (e.g., in an avoir ‘have’, qui ‘who/that’, haber ‘have’, estar ‘be’ or similar clause), followed by the information about it in the relative clause. This same strategy holds in the analogous construction for our data.

17. Sasse (1987) embedded presentational in the larger category of thetic expressions, a type of predication whose function is to present a state or event. As will be seen, the Peninsular Arabic VS construction, although having a presentational component, has an overall discourse function wider than those identified as typically thetic by Sasse (1987:566).

18. Note that these figures are more complete than those in Table 6. This is because in the Varbul analysis data gaps in a number of the variables occurred, excluding the token from analysis.

19. Holes (2009) notes that for sequenced actions in Bahraini Arabic, equally a Peninsular Arabic dialect, VS is the norm, at least, as Holes qualifies it, among uneducated speakers.

20. A reader suggested that previous mention may play a significant role in determining the occurrence of definite nouns. Responding to this suggestion from a Givónian perspective, in both of our data sets where VS is dominant, previous mention does not play a role in favoring a stronger SV sequence in a subsequent mention. In our Jedda episode (Table 9), there are 6 SV nouns, 3 previously mentioned, 3 new. There are 10 VS nouns. One is new, 6 are previously mentioned, as object, introduced in an existential clause, or as object of a preposition, and 3 are “anchored” by a suffix pronoun cross-referential with a noun (e.g., the –ha of abuu-ha her father in (14)). Anchored nouns are half new, the noun itself, and half old, the linking suffix pronoun. Here, of the 12 previously mentioned nouns, 3 are SV and 3 are VS, with 3 falling in the midway category of anchored.

In the Al-Rawi Emirati texts, a count of the clear cases of all definite and indefinite nouns relative to the new/previous mentioned parameter gives the following breakdown:

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>Previously mentioned</td>
<td>17</td>
<td>26</td>
</tr>
</tbody>
</table>

In simple numerical terms, previously mentioned VS nouns actually outnumber new VS ones. Peninsular Arabic contrasts starkly with Brazilian Portuguese in this respect. In Brazilian Portuguese, Ferrari (1990) noted that 95% of VS nouns are in first-mention situations.

21. The independence of narrated events from the speech situation as a factor influencing word order is also in evidence in pronoun usage. In all of our events, there are 13 tokens of VS order, with the subject the 1SG pronoun, ana, gaal ifaddal, okal-t ana ‘He told me, please (eat the dates) and I ate (ate I)’. Two of these tokens come in the fixed phrase, ana wiiya- ‘I and X’. This subject phrase is categorically VS, though there are so few tokens that we have left them in the counts. The other 11 all are used in a context where the subject “I” is a participant in an event, not “I” as speaker in the immediate speech situation.

22. Verb form, perfect or imperfect, has sometimes been adduced as a factor affecting word order as well. Dahlgren (1998:229–231) presented an array of statistics from published texts that appear to show that VS word order in Arabic strongly correlates with perfect verb form. Also, Benmamoun, using elicited data, reported that Moroccans favor VS perfect and SV imperfect (2000:62). In our texts, however, imperfects can be consistently VS as well. In the Jedda text (14), the decisive feature is the event frame, in this case expressed mainly in the imperfect tense. Holes (2009) on Bahrain Arabic similarly reports that VS imperfects are common.

23. In fact, Naro and Votre did not treat pronouns at all, except to note that they are invariably SV. They did note that pronouns are well suited to their SV = topicality informational category, because they are given.
Note that a statistical count of which pattern is more frequent does not necessarily help elucidate the issue. In Naro and Votre’s Brazilian Portuguese (excluding pronouns), SV outnumbers VS by 656 to 391. In our data, SV (excluding pronouns) outnumber VS by a somewhat larger 491 to 196. On this basis, it might be argued that SV is more important in Peninsular Arabic than in Brazilian Portuguese because it is proportionally more frequent. We argue, however, that the key point is the functionality of the order, not simple frequency.

Note that this designation covers all morpholexical realizations of SV word order in both varieties. An alternative to this is to define SV in Peninsular Arabic as representing topicality, pace Naro and Votre, runs into the problem that general and generic nouns although characteristically SV order, represent a lower degree of topicality.

REFERENCES
Hetzron, Robert. (1973). The presentative movement or why the ideal word order is SVOP. In C. N. Li (ed.), Word order and word order change. Austin: University of Texas Press. 346–388.


APPENDIX

**Jedda text: dominant SV**

This excerpt illustrates the frequent turn changes and the development of dialogue through frequent interlocutor interaction. As far as SV order goes, the speakers underline their differing opinions with frequent use of *ana*, the first person singular pronoun. Those marking a verbal subject are in italics.

H2-B:  
ana, ana šaayfat-ik šweyya yašni naazla ūala l-ʔarab [laughter]  
H2-A:  
[laughter] la la la? w-alla ana muš naazla, *ana* ma ansa inn-i ʕarabiyya  
H2-B:  
naʕam  
H2-A:  
bass *ana* ma hibb l-ʔašyaa? Ili kaanu yhuṭṭuu-ha f-raas-na, w  
nimsik-ha ūala asaas inn-a msallamaat, w ašyaa? hagiigiyya w  
hiyya ma fii šeyy, ma dri  
H2-B:  
bi-l-faks.  
H2-A:  
kull waahad u raʔy-u(h)  
H2-B:  
yašni *ana* ašuuf il-manaahij illi kaanat ūind-ana  
H2-A:  
mm-hmm.  
H2-B:  
w illi ana darast-aha fi l-madrasa, wa fi l-jaamiya, bi-l-ʔašks yašni,  
θaqaafa, aa...aa...aaθa... xaθiyya, kweyyisa, hatta inn-i yani hina  
lamman jiit, aa... darrast θalaba  
H2-A:  
eywa.  
H2-B:  
fi maʕaahid al-luθa hina  
H2-A:  
eeh.  
H2-B:  
istʔarabt inn-u ma ūind-ahum al-xalθiyya θ-θaqaθiyya Ili ūind-i  
an-a  
H2-A:  
illi hiyya masalan?  
H2-B:  
‘I, I see you’re a bit down on the Arabs [laughter]’  
H2-A:  
‘[laughter] No, no, no, really. I’m not down on them. I can’t forget  
that I’m an Arab’.  
H2-B:  
‘okay’  
H2-A:  
‘It’s just, I don’t like what they used to put in our heads, and we’d  
take them because they were simply delivered up to us, as if they  
were truths whereas they were nothing, I don’t know’  
H2-B:  
‘To the contrary’  
H2-A:  
‘Each has a right to their opinion’  
H2-B:  
‘I mean, I see that the curriculum that we had’  
H2-A:  
‘uh huh’  
H2-B:  
‘and that I studied at school and at the university, to the contrary,  
I mean, culture, aa a I had a good (cultural) background. To the  
extent that when I came here [to the U.S.] and taught students’  
H2-A:  
‘Ya’  
H2-B:  
‘At the language institutes here’  
H2-A:  
‘Ya’  
H2-B:  
‘I was surprised that they [the students] didn’t have the cultural  
background that I did’.

H2-A:  
‘For instance’?