

**THE SYNTAX OF SPOKEN
ARABIC**

**A COMPARATIVE STUDY OF
MOROCCAN, EGYPTIAN,
SYRIAN, AND KUWAITI DIALECTS**

KRISTEN E. BRUSTAD

**GEORGETOWN UNIVERSITY PRESS
WASHINGTON, D.C.**

(2000)

1 THE DEFINITENESS CONTINUUM

1.0 Introduction

It is increasingly recognized that, in natural language, many grammatical features such as definiteness, number and animacy interact with each other, and behave more like continua than like strictly delimited categories (Comrie 1981, Givón 1979a, Croft 1990). In Arabic, nouns are said to be either definite or indefinite, but this dichotomy imperfectly represents the real world, in which entities can be more or less definite and specific. Natural language data from spoken Arabic in all four dialect areas contain entities that are neither wholly definite nor wholly indefinite, but rather lie somewhere in between, in an area that may be called 'indefinite-specific.' In this chapter, I will argue that speakers of Arabic exercise a degree of control in manipulating the rules of syntax to try to approximate these shades of meaning. This control allows speakers the flexibility they need to more closely describe the continuum of definiteness that the natural world presents.

1.1 Definite and Indefinite Markers

Definiteness in Arabic seems, at first glance, to be quite straightforward. Prescriptive and descriptive grammars alike describe the system of definiteness and indefiniteness as dichotomous: nouns are either definite or indefinite, and proper nouns are definite whether or not they are marked by the definite article. Nouns can be made definite with the addition of the definite article /i/ or /l-, or by the specification of a noun by the addition of another substantive to it to form a possessive construct (/iḍāfa/). In all of the dialects examined here, the definite article is /i/ or a phonetically determined variant. The following examples demonstrate:

	Indefinite Nouns	Definite Nouns
M	جَاهِم ضَيْف žāhum ḍīf came-he-to-them guest <i>A guest visited them</i>	ف الدار f əd-dār in the-house <i>in the house</i>

E	فستان يجنن fustān yigannin dress it-make-crazy <i>a stunning dress</i>	في البيت fi l-bēt in the-house <i>in the house</i>
S	تاخدي شقفة؟ tāxdi ša'fe? you-take piece <i>Will you take a piece?</i>	فات ع المطعم fāt 'a l-maṭ'am went-he into the-restaurant <i>He went into the restaurant.</i>
K	خوش مكان xōš makān good place <i>a good place</i>	ع البحر 'a l-baḥər on the-sea <i>on the beach</i>

In addition to the definite article, several of these dialects make use of other articles. Moroccan and Syrian employ the particle /ši/ *some*, and all four dialects permit limited use of the article /wāḥid/ (Moroccan /wāḥəd l-/ *one* (f /waḥda/).

In Moroccan, Harrell (1962) notes two "indefinite" articles: a "concretizing" article /wāḥəd l-/ *one*, and a "potential" article /ši/ *some* (1962:147, 189). Examples of these articles from my data include:

M1	كاين واحد الحاجة kāyn wāḥəd l-ḥāža there-is one thing <i>there is something</i>
M10	كيبقى كيقول شي كلمة قبيحة kaybqa kayqūl ši kəlma qbiḥa indic-he-keeps indic-he-says some word ugly <i>he keeps saying a some bad word</i>

Syrian Arabic shares with Moroccan the article /ši/, although Cowell classifies Syrian /ši/ as a partitive noun rather than an article (1964:467). Syrian speakers also use /wāḥid/ (f /waḥde/ (Grotzfeld 1965:76), primarily with human nouns. Examples of both /ši/ and /wāḥid/ from my data include:

S2 لازم نعمل له شي مقدّمة
 lāzim ni'mil-lu ši muqaddime
 must we-make for-him some introduction
We must give him some sort of preparation

S2 فيه واحد بدوي فات ع المطعم
 fī wāḥid badwi fāt 'a l-maṭ'am
 there-is one bedouin went-he into the-restaurant
There was a [certain] bedouin who went into the restaurant

Egyptian Arabic allows the use of /wāḥid/ with human nouns only (example from Woidich 1980b:32-3; translation mine):

E كان فيه واحد حطّاب
 kān fī wāḥid ḥiṭṭāb
 was-it there-is one wood-gatherer
There was a [certain] wood-gatherer

Holes notes the use of /wāḥid/ in Gulf Arabic preceding a noun to mean *a certain* (1990:114). My Kuwaiti data contain several examples in which /wāḥid/ modifies human nouns, including:

K3 راحت حگ واحد مطوّع
 rāḥat ḥagg wāḥid mṭawwa'c
 went-she to one religious-man
She went to a learned religious man

While the grammars of these dialects vary widely in their treatment of /ši/ and /wāḥid/, the recurrence of these articles across dialects invites comparative study of them. What is the motivation for marking nouns in these ways? What pragmatic role(s) do they play? Harrell's descriptive names, "concretizing" and "potential," allude to the fact that /wāḥid l-/ and /ši/ fulfill particular discourse functions, which may now be further specified and defined in light of developments in discourse theory in the years since Harrell wrote.

The traditional definite/indefinite dichotomy does not leave room to account for the function of these articles. Why do the dialects need "indefinite" articles such as /ši/ and /wāḥid/ if the unmodified noun is also indefinite? The contexts of the sentences cited above show that

the "indefinite" articles /ši/ and /wāḥid/ are in fact not wholly indefinite. Why are they used in some cases and not in others? Why does /wāḥid/ modify only human nouns? Answers to these questions may be found in typological and pragmatic approaches to syntax.

1.2 Definiteness, Indefiniteness, and Specification

Chafe gives a good description of 'definite' status: "I think you know and can identify the thing I have in mind" (1976:39). In other words, in order for a nominal phrase to assume definite status in discourse, it must meet one of several conditions: (a) it must have been *previously mentioned* in the discourse; or (b) it must be a member of a *universal set* of entities, such as the sun, that can be assumed to be known and identifiable without further specification, or (c) the speaker must have good reason to think that the entity is *retrievable* by the listener through knowledge shared by the interlocutors (Chafe 1976). Pragmatically, a definite noun usually represents *given* information, or information that has already been established in the discourse or can be assumed by the speaker to be present or active in the mind of his or her interlocutor. An indefinite noun, then, does not meet any of the above conditions, and represents an unknown, irretrievable entity.

However, not all indefinite nouns are created equal. Medieval Arab grammarians called the grammatical specification of nouns /taxṣīṣ/ *specifying*. Wright notes that /taxṣīṣ/ includes modification of an indefinite noun by adjectival phrases and annexation (1898ii:198). A Lebanese encyclopedia on Arabic grammar defines it thus:

التخصيص هو تقليل الاشتراك الحاصل في النكرات، ويكون بالوصف أو الإضافة، نحو: «إنه رجل علم»، بإضافة «رجل» إلى «علم» خففت من تنكيره، لأنه إذا قلنا «إنه رجل» كان شائعاً، أما إذا قلنا «إنه رجل علم» فإننا نكون قد أزلنا عنه بعض الشبوع.

Specification is the lessening of the commonality that occurs in indefinite entities, and [this] may occur through modification or the genitive (/iḍāfa/), such as: *He is a man of learning*, as the addition of 'man' to 'learning' has lessened the indefiniteness of [the former], because if we say, *He is a man*, [the statement] is general, but if we say, *He is a man of learning*, we will have eliminated from it some of that generality (al-Tūnǰi and al-'Asmar 1993:156).

The principle of partial specification of nouns is thus not new to the description of Arabic. Nor is it new to language typology. Croft calls it referentiality, and formalizes this definiteness hierarchy (1990:116):

Hierarchy of definiteness
 definite
 referential indefinite
 nonreferential indefinite

Croft's term *referential indefinite* describes al-Tūnjī and al-Asmar's phrase *a man of learning*, and his *nonreferential indefinite* describes their nonspecific noun *a man*.

Khan's work on Semitic provides a good model to use in the analysis of nominal syntax in spoken Arabic. Khan (1984) has adapted the work of Chafe (1976), Timberlake (1977), Hopper and Thompson (1980), and others to show that one or more of the features listed in Table 1-1 may operate to attract object marking and agreement pronouns in Semitic, such that the more individuated the noun, the greater the tendency of these markers to occur. He groups these qualities together under a rubric he calls individuation or salience, which he illustrates as follows (1984:470):

Table 1-1: Khan's Hierarchies of Individuation

<u>Individuated/Salient</u>		<u>Non-individuated/Non-salient</u>
1. Definite	>	Indefinite
2. Non-reflexive	>	Reflexive component
3. Specific	>	Generic
4. Concrete	>	Abstract
5. Qualified	>	Unqualified
6. Proper	>	Common
7. 1 st person > 2 nd > 3 rd > Human	>	Inanimate
8. Textually prominent	>	Incidental

The notion of individuation provides great explanatory power for the syntactic behavior of nouns in spoken Arabic. However, I will modify Khan's framework slightly to include those features that most influence the syntactic behavior of nouns in spoken Arabic. Reflexivity, while relevant to the individuation of nouns, is often expressed in Arabic through verb morphology. And since abstract nouns in Arabic

normally take the definite article, concreteness appears to have less central a role than specification and qualification in the syntactic marking of nouns. Parallel to Khan's textual prominence is physical prominence: nouns tend to be marked as more salient when they are present in the immediate environment (see further 4.2). Three other modifications are inspired by and adapted from the work of Cowell and Janda.

First, I will add to the list the feature quantification. Cowell contrasts agreement patterns of nouns denoting "collectivity or generality" with those denoting "heterogeneity or particularity" (1964:423). Belnap's study of number agreement in Cairene Arabic lends further support to this analysis (1991:68-72). Quantification involving numbers ten and lower also appears to have some relevance to the marking of new discourse topics (see 1.5).

Cowell also contrasts concepts of identification and classification, which play a role in the expression of possessive constructions. Identificatory annexion refers to the assigning of an entity to a specific possessor, while classificatory annexion assigns an entity to a set or group (1964:458). This notion of a proper identification contrasted to a generic one parallels Khan's hierarchy of specificity vs. genericness, and Khan's terms specific and generic are defined here to include Cowell's insight.

Finally, I will add to Khan's animacy hierarchy the concept of agency, defined here as the degree to which an individual or entity has the ability to act independently. Agency may be viewed as a kind of sociolinguistic parallel to textual prominence. The concept finds precedence in the work of linguists who have argued for the role of social status and power in language variation and change, among them Janda, who argues for the centrality of a feature she calls "virility" to salience in Slavic languages, which helps explain both synchronic patterns and diachronic changes in Slavic case and number agreement.¹ While virility represents an appealing concept in the age of feminist theory, the concept of agency applies more generally. Factors that contribute to agency include animacy and social status; thus for humans,

¹Laura A. Janda, UNC Chapel Hill, public lecture at Emory University, 10/16/1997, "Virility in Slavic: A Conspiracy of Factors Over Time and Space;" see also Janda 1999.

gender and age play a role as well. Children have less agency than adults, animals have less agency than humans, and inanimate objects are unlikely to be perceived as having agency at all. I will argue here and in Chapter 2 that agency provides a possible explanation for certain patterns of definite marking and agreement in some dialects.

The major objection to hierarchies of individuation or salience lies in their vagueness. The features that comprise such hierarchies need detailed investigation and definition involving a large corpus of naturally occurring data. Pending such a study, attention will focus here on those features which seem to have the greatest relevance to or explanatory power for the syntax of spoken Arabic. The list in Table 1-2, revised from Khan's (Table 1-1), includes syntactic, semantic, pragmatic, and sociolinguistic features, covering a range of levels that interact in the production of language. These features appear to play a role in pushing a noun toward higher or lower individuation, affecting the speaker's choice of syntactic marking of nouns in spoken Arabic when a choice of marking presents itself.

Table 1-2: Features Affecting Individuation

1. Agency: includes humanness/animacy, social status or power, perhaps gender and age groups as well.
2. Definiteness: syntactic marking or semantic status (e.g., proper nouns).
3. Specificity vs. genericness: the extent to which a speaker has a specific entity in mind.
4. Textual or physical prominence: the extent to which a noun plays a role important to the discourse, or is physically present and prominent.
5. Qualification: modification of a noun with adjectives and other descriptives.
6. Quantification vs. collectivity: the extent to which a noun is specifically quantified, especially with numbers from 2-10.

Taken as a group, it is clear that these features affect the syntactic behavior of nouns, but in a way that is not (yet) possible to describe formally. It is more convenient to formalize them using a continuum, which has the additional advantage of privileging speaker control in determining the marking of a given noun. In Figure 1, the features on the right tend to reflect (or, from the speaker's point of view, attract) higher individuation or salience, the ones on the left, less:

Figure 1: The Individuation Continuum

Unindividuated:	Partly individuated	Individuated:
- agency/animacy		+ agency/animacy
- definite		+ definite
- specific		+ specific
- prominence		+ prominence
- qualified		+ qualified
+ collectivity		+ quantified 2-10

The importance of this continuum lies less in predicting the marking on the noun itself than in explaining the syntactic behavior of the noun in the larger sentential context, as I will show. In other words, the higher the individuation of a noun in the mind of the speaker, the more that noun will attract certain kinds of syntactic marking.

1.3 Definiteness and Individuation

The correspondence between individuation and definiteness in Arabic is not direct, because syntactic rules also affect definite marking. For example, abstract and generic nouns in Arabic are marked with the definite article. Thus an unspecified noun referring to any member of the class may be marked with /-(i)l-/, while a specified human noun may not be marked definite. In the following introduction to a joke, the unnamed but somewhat specified bedouin is marked indefinite with the indefinite article /wāḥid/, while the generic restaurant and waiter are both marked definite.²

S2 فيه واحد بدوي فات ع المطعم. قال له للكارسون، انطيني بوظة
 fī wāḥid badwi, fāt 'a l-maṭ'am. qāl-lu la-l-garsōn, inṭini būza
 there-is one-ms bedouin-ms, entered-he into the restaurant. said-he
 to-him to-the-waiter, give-me ice-cream
There's this bedouin who went into a restaurant. He said to the waiter, give me ice cream.

Grammar views definiteness as a dichotomy: nouns are either definite or not. But the real world that a speaker knows and desires to represent is far from black and white: some nouns may be somewhat defined or specified but not entirely, either from the point of view of

²The joke is cited in full, and its nominal markings further discussed, in 1.5.

the speaker or in the assumptions the speaker makes about the listener's knowledge. Definiteness is a grammatical category in which the speaker has a degree of control, and speakers of Arabic need to represent a range of undefined, partially defined, and fully defined entities. Definite and indefinite represent the black and white ends of a definiteness continuum, while the grey area of partially defined or specified entities falls in between. I will call this range indefinite-specific, after Wald (1983).³ It includes nouns that are syntactically indefinite, but carry a degree of specific reference that may be reflected in a number of possible syntactic constructions. In other words, an indefinite noun that carries a degree of individuation or specificity may attract a degree of definite or specifying syntactic marking. Or, as described from a different perspective, a speaker may be influenced by these features to mark a noun in a certain way. The rest of the chapter will explore the use of definite and indefinite markers in the dialects, and show how spoken Arabic uses various combinations of syntactic markings to indicate a range of indefinite-specific status.

1.4 Indefinite-Specific Marking

The semantic range from highly individuated (e.g. definite, specific, or animate) nouns to non-individuated (e.g. indefinite, non-specific, or inanimate) nouns includes varying degrees of definiteness and specificity. The more specific the reference of the noun, the greater the probability that the noun will be syntactically marked with some kind of article or specifying construction. This section will examine two kinds of indefinite-specific marking, the article /ʃi/ in Moroccan and Syrian, and the indefinite suffix /-in/, found in some Gulf regions.

1.4.1 Indefinite-Specific Article /ʃi/

Both Syrian and Moroccan speakers often identify a noun in the indefinite-specific range with the referential indefinite article /ʃi/ *some (kind of)*. In the following sentence, /ʃi/ lends a degree of specificity to the noun /muqaddime/ *introduction*.

³Wald uses the term 'indefinite specific' to explain the use of *this* in spoken English in sentences such as *I saw this guy ...*

- S2 لازم نعمل له شي مقدمة لحتى ما ينصدم
lāzīm na'mil-lu ʃi muqaddime la-ḥatta mā yinšidim
necessary we-make-for-him some introduction so-that neg he-be-shocked
We must arrange some kind of preparation for him so that he won't be shocked

Analogous Moroccan examples include the following /ʃi kəlma qbīḥa/ *some nasty word*, /ʃi masā'il qbīḥa/ *some nasty things*, and /ʃi nās ḏifān/ *some guests*:

- M10 كيبقى يقول شي كلمة قبيحة، كيقول شي مسائل قبيحة
kaybqa yqūl ʃi kəlma qbīḥa, kayqūl ʃi masā'il qbīḥa
indic-he-keeps he-say some word ugly, indic-he-says some things ugly
He keeps saying some nasty word, he says some nasty things
- M2 وانا عندي شي ناس ضيفان
w āna 'ndi ʃi nās ḏifān
and I at-me some people guests
While I had some guests

Here, as in the Syrian example, the article /ʃi/ indicates the partial specificity of the nouns it modifies. It is worth noting that all of the nouns in the Moroccan examples are modified, reflecting the relevance of qualification to the individuation of a noun. The Moroccan data thus provide evidence of correspondence among qualification, specificity, and syntactic marking on nouns. All these examples demonstrate that speakers use /ʃi/ to indicate that they have a particular type of entity in mind.

1.4.2 Nunation as Indefinite-Specific Marking

In formal Arabic, nunation or /tanwīn/ refers to the endings /-un/, /-an/, and /-in/ that function as indefinite case markers on nouns and adjectives. Nunation thus represents part of the case-marking system of formal Arabic. However, spoken Arabic has no case-marking system. In a number of dialects, nunation in the form of /-an/ (/tanwīn fathā/) occurs in certain fixed adverbial expressions, such as /'abadan/ *ever, at*

all, and /dāyman/ *always*. Adverbial /-an/ is of higher productivity in the educated registers of spoken Arabic.

Another type of nunation, usually realized as /-in/ and occurring on indefinite nouns in a number of bedouin dialects, especially in their poetic register, are considered to be vestiges of the case-marking system of formal Arabic. Yet this suffix provides no case information. Holes mentions "the vestige of" a suffix /-in/, found in "the speech of some less educated Gulf speakers, and in dialect poetry," which marks indefinite nouns only when occurring in "Noun + Attributive Adjective" phrases (1990:115). In his study of Najdi Arabic, Ingham includes the indefinite marker /-in/ as part of noun phrase structure (1994:47):⁴

indefinite	bēt	'house' or 'a house'
indefinite (marked)	bēt-in	'a (particular) house'
definite	al-bēt	'the house'
possessed	bēt-i/-ik	'my/your house etc.'

This schema suggests that /-in/ functions as an indefinite-specific marker, and Ingham's description of the usage patterns of this ending further confirms this analysis (1994: 49; IND = indefinite):

(i) where a nominal follows another nominal as in *bēt-in kibīr* (house-IND large) 'a large house;' (ii) where a modifying prepositional phrase follows a nominal as in *wāhd-in min ar-rabu'* (one-IND from the group) 'one of the group' (i.e. one of my friends) ... (iii) where a noun is followed by a modifying clause as in *kalmit-in gāl-ō-hā-li* (word-IND said-they-it-to-me) 'a word which they said to me.'

Nunation in Najdi Arabic occurs on indefinite nouns modified by an adjective or relative clause, in other words, on *specified* indefinite nouns. The fact that this ending occurs on indefinite nouns that are modified in some way is significant, for it indicates that such a noun is not purely indefinite, but has a degree of specificity. Urban Kuwaiti seems to have lost the ending except in highly formalized contexts such as poetry; most published material on this phenomenon does not distinguish between poetic use of /-in/, which belongs to a special artistic register, and examples taken from naturally-occurring speech. Al-Ma'tūq's study of the tribal dialect of the 'Ajmān in Kuwait reports

⁴It occurs on plural nouns as well: /mgaddm-in-in/ *submitted-p-indef* (Ingham 1994:167).

similar use of /-in/, but her examples too are taken largely from proverbs and poetry, rather than extemporaneous speech (1986:190-91).

This function of nunation appears to be quite old. Evidence for the use of nunation with /-an/ as an indefinite-specific marker may be found in early Spanish Arabic texts. Corriente reports that this kind of nunation has "the function of linking constituents," these constituents being an indefinite noun with a following adjective or relative clause (1977:121-2). Corriente's examples clearly show contexts in which an indefinite noun is qualified or partly specified, resulting in an indefinite-specific noun phrase marked with /-an/. These phrases include (121-2):

kalban abyad <i>a white dog</i>	'ala qalban kāfir <i>with an unfaithful heart</i>
bi-xāṭiran yattaqad miṭl al-nār <i>with a mind as bright as fire</i>	

Shumaker's (1981) study of the indefinite suffix /-an/ in Galland's fourteenth or fifteenth century manuscript of *Alf Layla wa Layla* (*The Thousand and One Nights*, ed. Mahdi 1984) establishes patterns of /-an/ as a syntactic marker of textually prominent entities that appear to correspond to indefinite-specific patterns. Her conclusions suggest that /-an/ functions in the text as a type of indefinite-specific article. Examples from the text support this analysis; in the following passages, indefinite nouns that are specified with adjectival or relative phrases end in /-an/. From the "Story of the Porter and the Three Girls" (Arabic text from Mahdi 1984i:126; transcription and translation mine):

... اد وقتت عليه امرأة ملتفة في ايزار موصلية مشعر بحرير
[id waqafat 'alayhi imra'at-an multaffa fi 'izār muṣīlī muṣa'ar
bi-ḥarīr ...
when stood-she over-him woman-an wrapped in shawl Mosuli
fringed with-silk]
*When suddenly there stood before him a woman wrapped in a
Mosuli shawl fringed with silk*

The suffix /-an/ on /imra'at-an/ *woman* is marked in the Arabic text, and the noun is partly specified by the following phrase. Another part of the story contains an instance of nunation marking an indefinite noun modified by a relative clause (Mahdi 1984i:138):

فقال جعفر يا امير المؤمنين، هولاي ناساً قد دخل فيهم السكر ولا يعلمون من نحن

[fa-qāl ja'far yā 'amīr al-mūminīn, hawlāy nās-an qad daxala fihim as-sukr wa lā ya'lamūn man naḥnu

so-said-he Ja'far, O Caliph, those are people-an perf entered-it in-them the-drunkenness and neg they-know who we]

Ja'far said, Caliph, those are people in whom drunkenness has set, and they do not know who we are

It may be argued that these examples belong to an artistic register and do not necessarily reflect everyday speech. At the very least, though, these texts provide evidence for the indefinite-specific function of nunciation as part of some register of Arabic in the medieval period.

Evidence thus exists supporting the use of /tanwīn/ as an indefinite-specific marker in a number of geographic locations from an early period, and surviving until today in parts of the Peninsula, perhaps also in Levantine /ḥadan/ *someone, anyone*, which may be a reflex of this indefinite-specific nunciation.

Egyptian dialects do not appear to have any articles that specify or individuate an indefinite noun. In order to express the notion of specificity when referring to an indefinite noun, Egyptian speakers commonly employ the adverb /kida/ *thus, so* as a modifier:

E2 ... شفت حاجة كدا

šuft ḥāga kida

saw-I thing like-this

I saw something ... or I saw this thing ...

In addition, Egyptians use several nouns that function in similar fashion, that is, to lend certain kinds of specificity to an indefinite noun, and thus can be identified as specialized indefinite-specific markers, since they have more specialized meanings than does /ši/. The meanings of these articles involve plurality or intensification: /iṣṣi/ *some* or *a bunch of*, in a series, *something* in certain idiomatic phrases, and /ḥittit/ *what a ...!* (Badawi and Hinds 1986:25,190). Clearly, /iṣṣi/ is related to /ši/. Badawi and Hinds give three contexts for this article; the first is of interest here (1986:25):

E ... عنده مما جميعه اشفي جبنة واشفي زيتون واشفي سردين ...
'andu mimma gamī'u 'iṣṣi gibna w-'iṣṣi zatūn wi-'iṣṣi sardīn ...
[at-him of-what all-it 'iṣṣi cheese and-'iṣṣi olives and-'iṣṣi sardines]
he has something of everything-cheese, olives, sardines ...

In such cases, /iṣṣi/ intensifies the following noun, and in doing so specifies it; however, /iṣṣi/ is not regularly used to specify indefinite nouns, but only in contexts where quantity or variety is stressed.

Other intensifiers are more commonly used in Egypt, among them /ḥittit/, literally a *piece of*. While this particle is mainly understood as an intensifier expressing astonishment or admiration, part of its function is specification:

E3 لقينا فار قد كدا هو، حتة فار قد كدا هو

la'ēna fār 'add kida-hō, ḥittit fār kida-hō

found-we mouse size like-this, piece-of mouse like this

We found a mouse this big, what a mouse this big!

1.5 Definiteness and First Mention: New Topic

On the continuum of definiteness, near the indefinite-specific range, there appears to be an area reserved for the first mention in discourse of 'textually prominent' entities (one of Khan's features, see Table 1-1). The first mention of a nominal entity in discourse may be indefinite, indefinite-specific, or, in some cases (and in Moroccan in particular), definite in marking. Not all first-mention nouns are of equal importance to the discourse; some nouns play a more prominent role than others. Discourse analysts distinguish between figure and ground, or central and marginal entities (and events) in discourse respectively (Hopper and Thompson 1980). An entity that plays an important role in a text is likely to be marked in a way that reflects that status, whereas an entity that is part of the background is less likely to be so marked.

In the following joke, the specified and textually prominent figure *bedouin* is marked with an 'indefinite' article /wāḥid/, while the inanimate, unspecified, even generic nouns /l-maṭ'am/ *restaurant* and /l-garsōn/ *waiter* are marked with the definite article, and inanimate /būza/ *ice cream* has no article in its first occurrence:

- S2 فيه واحد بدوي فات ع المطعم. قال له للغارسون، انطينني بوظة. جاب له صحن بوظة صار ياكله بالخبز. بلش الغارسون يتضحك عليه. قال له البدوي علويش تتضحك؟ على طبيخك المسكع؟
 fī wāḥid badwi fāt ‘a l-maṭ‘am. qāl-lu la-l-garsōn, inṭīni būza. žāb-lu šaḥn būza šār yāklū bi-l-xəbəz. ballaš ig-garsōn yitdaḥḥak ‘alēh. qāl-lu l-badawi ‘alwēš titdaḥḥak? ‘alā ṭabīxak l-msagga? !
 there-is one-ms bedouin-ms entered-he into the restaurant. said-he to-him to-the-waiter, give-me ice-cream. brought-he to-him plate of-ice-cream began-he he-eat-it with-the-bread. he-started the-waiter he-laughs at-him. said-he-to-him the-bedouin on-what you-laugh? at cooking-your the-cold?!

There's this bedouin who went into a restaurant. He said to the waiter, give me ice cream. He brought him a plate of ice cream, he began to eat it with bread. The waiter started laughing at him. He said, what are you laughing at? Your ice-cold cooking?!

The speaker telling this joke introduces its subject, /badwi/ (*a*) *bedouin*, with the indefinite /wāḥid/ *one*, for two reasons: first, because the bedouin's existence is previously unknown, and second, because he plays a key role in the joke and therefore specific referentiality must be established. On the other hand, /il-maṭ‘am/ *the restaurant* and /il-garsōn/ *the waiter* are marked definite at their first mention, the former due to its generic status and the latter because any restaurant may be presumed to have a waiter working in it. Were the speaker to have said /fāt ‘a maṭ‘am/ *he went into a restaurant*, it might imply that the restaurant had some importance to the story, in which case further specification would be expected. In contrast, *ice cream*, which also has some textual prominence, is less individuated because it is inanimate, and so its first mention is indefinite but not marked with an indefinite-specific article.

These 'textually prominent' entities fall in the indefinite-specific range; but represent a special case within it. Such entities may be called "new topics," making the article /wāḥid/ a "new topic" article.⁵

Evidence from Syrian suggests that /wāḥid/ modifies only human nouns. Egyptian data exhibit a similar pattern, with the usage of /wāḥid/

⁵I have adapted the term 'new topic' from Wald (1983).

also syntactically restricted to human nouns:

- E2 كان فيه واحدة ست ...
 kān fī waḥda sitt
 was-it there-is one woman
There was althis woman ...

In this Kuwaiti example (repeated from 1.1.1 above), /wāḥid/ also modifies a human noun, /mṭawwa‘/ *learned religious man*.

- K3 راحت حگ واحد مطوع
 rāḥat ḥagg wāḥid mṭawwa‘
 went-she to one religious-man
She went to a learned religious man

Evidence thus suggests that "new topic" status as marked syntactically by the article /wāḥid/ is restricted in Egyptian, Syrian, and Kuwaiti dialects to textually prominent, highly individuated, and specifically human, nouns.

In contrast to the restricted use of /wāḥid/ in the other dialects, the Moroccan article /wāḥəd l-/ is not restricted to human or even animate nouns, but extends to inanimates as well:

- M10 كايين واحد النوع آخر د الحوت ...
 kāyn wāḥəd n-nū‘ āxur d l-ḥūt
 there-is-ms one the-kind other gen the-fish
There's this other kind of fish ...

In Moroccan, /waḥəd/ is obligatorily followed by the definite article /l-/: /wāḥəd l-/, and it is used with a much broader range of entities, not only human but also inanimate entities if they are textually prominent. The following Moroccan passage introduces two characters in similar fashion, and a third differently (all marked in boldface):

- M11 حاجيت لك، هادا واحد الراجل ما عندوش الولاد، عنده غير
 المرا ومرته عمرها ما ولدت. ناضت ولدت واحد البنت
 ḥāžit-lk, hāda wāḥəd ər-rāžl mā ‘əndūš l-wlād, ‘əndu ġir l-mra
 w martu ‘ammrha mā wəldat. nāḏət wəldat wāḥəd l-bənt
 told-story-I-to-you, this **one man** neg-at-him the-children, at-him
 only **the-wife** and wife-his life-her neg bore-children-she. arose-
 she bore-she **one girl**

I tell you a story, this is a man who doesn't have children. He has only the wife, and his wife never bore children. Then she had a daughter.

This story opens with the introduction of a main character identified by the article /wāḥəd l-/ *a*. The next characters introduced are /l-wlād/ *children*, marked definite because of their high animacy, except that their existence is negated, and so they do not constitute a topic. The wife /l-mra/ is not marked as a new topic, however, presumably because her existence may be inferred, as most men are married, and she thus constitutes a retrievable entity that does not need to be singled out for introduction. The next topic introduced is /l-bənt/ *girl*, marked with /wāḥəd l-/ as a new topic and important to the story. The function of /wāḥəd l-/ as a new topic marker is substantiated by the fact that the man and the girl are the only two characters who reappear in the story.

Two entities are introduced in the next passage. The first, /wəld l-malik/ [*a*] *king's son* is human, and also has an important role to play in the story, and so the speaker takes care to establish textual prominence for him by marking the phrase with /wāḥəd/. The second identity, /šəržəm/ *window* is neither animate nor topical, and thus is not marked by the new-topic article.

M11 ناض الـأ واحد ولد الملك - وراه عندها شرحم ف حال هكا. طلت
م الشرحم
nāḍ a-lalla wāḥəd wəld l-malik -- w-rah 'əndha šəržəm f-ḥāl
hāk^wa. ʔəllat m š-šəržəm
arose-he O lady one son of-the-king -- and see-now at-her window
like like-this. leaned-out-she from the-window
*So, my lady, this prince up and -- She had a window like that.
She leaned out of the window*

In the third passage, the repetition of /mra x^wra/ *another wife* with the addition of the 'new-topic' article /wāḥəd l-/ may be due to the necessity of marking this second wife as a significant character in the story. After omitting the article in her first mention of the woman, the speaker may have subconsciously felt that the character needed more of an introduction:

M11 ناض اتجوج مرا خرى اتجوج واحد المرا خرى
nāḍ tžəwwež mra x^wra tžəwwež wāḥəd l-mra x^wra
arose-he married-he woman other married-he one the-woman other
He up and married another woman, married this other woman

Further evidence that /wāḥid/ functions as a new topic marker lies in the use of /ši/ without /wāḥəd/ in both Moroccan and Syrian. While partially individuated, textually prominent persons may be introduced with /ši wāḥ(i)d/ *someone*, speakers of both dialects can use /ši/ to mean *someone* when the reference is non-individuated, non-specific, not textually prominent, and of low social status. An example from Caubet's Moroccan texts (1993ii:7):

M شي كيكمي السبسي، شي كيددل جوانات، شي ...
ši kayəkmi əs-səbsi, ši kaybəddəl žwānāt, ši ...
some indic-he-smokes the-hashish-pipe, some indic-he-exchanges
joints, some ...
*Someone is smoking the hashish pipe, someone is exchanging
joints, someone ...*

From my Syrian data:

S2 ... إلا يعني عالم اسوأ منها بكتير كثير يعني شي إلن ماضي كتير
مشرشح منشان يقبلوا ياخذوا يعني ها البنات
... 'illa ya'ni 'ālam 'aswa' min[h]a bə-ktīr ktīr ya'ni še 'ilon māḍi
ktīr mšaršəḥ minšān yiqbalu yāxdu ya'ni ha l-binət
... except that-is world worse than-her by-a-lot a-lot that-is thing
to-them past alot sordid in-order-to they-accept they-take that-is
this the-girl
... except for people who are much worse than her, that is, people
who have a very sordid past, to accept to marry the girl

The marking of an indefinite noun with the new-topic article /wāḥid (l-)/ appears to be motivated by the high degree of textual prominence played by that noun. The derivation of the article itself, /wāḥid/ *one*, further supports the notion that quantification, individuation, and textual prominence are all related to each other as well as to the syntactic marking of nouns.

1.6 Definite Marking in Moroccan⁶

Moroccan speech in particular is characterized by the occurrence of unexpectedly “definite” nouns that fulfill none of the conditions specified by Chafe (see 1.2). First-mention nouns are normally indefinite in Arabic, as in most languages that mark definiteness. But in the following passage (repeated from above), even though *children* is semantically indefinite, since the man in question has none, and his wife is introduced for the first time, both /l-wlād/ *the children* and /l-mra/ *the wife* both have definite marking in this passage:

- M11 حاجيت لك، هادا واحد الراجل ماعندوش الولاد، عنده غير المرا
 ḥāžit-lk, hāda wāḥəd ər-rāžəl ma‘andūš l-wlād, ‘andu ġi[r] l-mra
 told-I-to-you this one the-man neg at-him the-children, at-him
 only the-wife
I’ll tell you a story, this is a man who has no children. He has only a wife

The next passage contains the first mention of /bīt/ *room* in the story; therefore, it would be expected to be indefinite. On the other hand, the noun /bīt/ has highly individuated reference in this case, which ‘attracts’ the definite article here.

- M11 باها ملك سلطان وناض دار لها البيت ديالها بوحدها
 bb^wāha malik səlṭān w nāḍ dār-lha l-bīt dyālha b^wuḥdha
 father-hers king sultan and arose-he made-he for-her the-suite
 gen-hers by-herself
Her father [was] a king, a sultan, and he up and made her a suite of her own

Similarly, /l-wəld/ *the son* in the following refers to a nonexistent child. The importance and (future) agency of the desired son attract definite marking to a noun that should otherwise be indefinite:

- M11 ما عندهاش الولد. ناض گال لها انا خصني الولد
 mā ‘ndhāš l-wəld. nāḍ gāl-lha ‘ana xəššni l-wəld
 neg-at-her the-son arose-he said-he to-her I needed-for-me the-son
She didn’t have [a] son. He up and told her, I need [a] son

⁶See Caubet (1983) for a more detailed, formalized approach to definiteness in Moroccan Arabic.

There is no previous mention of /tūr/ *bull* in the passage from which the next example is taken, nor does it appear to have specific or generic reference, and yet it is marked definite. In this case, either animacy has influenced the ‘indefiniteness’ of the noun and pushed it towards the definite end of the continuum, or the speaker assumes that a slaughtered bull is an expected and retrievable part of celebrating a son’s birth:

- M11 دبع التور، عرض على الناس
 dbəḥ t-tūr, ‘raḍ ‘la n-nās
 slaughtered-he the-bull, invited-he the-people
He slaughtered a bull, invited people

Finally, even though the reference of /l-ḥūt/ *fish* in the following sentence is semantically indefinite, generic, and nonspecific, it carries definite marking. What attracts definite marking here may be the textual importance of *fish* to the passage on the depressed fishing economy of Larache from which this sentence is taken:

- M10 ماشي تشوف بعينك تقول راه كاين الحوت
 maši tšūf b-‘īnk tqūl rāh kāyn l-ḥūt
 fut you-see with-eye-your you-say see-here there-is the-fish
You’ll see with your own eyes and say there are fish

The examples just cited may be contrasted to the following passage, in which an indefinite noun does occur. Here true indefinite nouns are not marked with the so-called “indefinite” articles: /blād/ *a place* is without marking, due to its indefinite, unspecific, and inanimate status in the story; in other words, it is not individuated.

- M11 والنهار اللي ولدت امك هربوا عليها خلّوها بالكروش وخاوا بلاد
 وعمروا بلاد
 w-n-nhār lli wəldat mm^wk hərbu ‘līha xəllāwha b-l-kərš w xwāw
 blād w ‘amməru blād
 and-day that bore-she mother-your deserted-they on-her left-they-
 her with-the-belly and emptied-they place and settled-they place
The day that your mother gave birth they deserted her. They left her pregnant and moved to another town.

Of the four dialects, Moroccan speech is clearly the most influenced by the specificity and animacy factors, both of which may trigger the use of definite marking. In the following, all animate nouns are marked with either the definite article or a possessive pronoun, even though not all are semantically definite and specific:

M11 عندك سبعة د خواتاتك، وتزوج باك على أمك على الولد، باش تولد الولد

‘əndk səb’a d xwātātk w tʒuwʒ bbʷāk ‘la mmʷk ‘la l-wəld, bāš təwləd l-wəld

at-you seven gen sisters-your and married-he father-your on mother-your for the-son in-order-to she-bear the-son

You have seven sisters, and your father took a second wife for [a] son, for her to bear [a] son

Kuwaiti speech contains instances of this phenomenon as well. In these examples, /waladha/ *her son* has specific reference, at least in the mind of the speaker, as does /l-’ahal/ *the people* in the second.

K3 عندها ولدها ضابط
‘indha waladha ḍābiṭ
at-her son-her officer
She has [a] son, an officer

K4 هاذا شغلکم انتو يا اهل
hāda šuġulkum ‘intu ya l-’ahal
this business-your-p you-p O-the-people
This is your doing, people

The difference between specific and non-specific reference is demonstrated by the following passage. The woman in this passage replies twice to the slave’s question, saying at first /wlədt bənt/ *I had a girl*, and the second time, /wlədt l-wəld/, literally *I had the son*.

M11 مشى داك العبد عندها، قال لها قال لك سيدي شنو ولدت؟ كات له ولدت بنت. قال لها گولي لي شنو ولدت، راه إيلا ولدت البننت غ ندبحك وندبجها. تا شافته زايد لها بالموس كات له هدا، ولدت الولد
mša dāk l-‘əbd ‘əndha, gāl-lha gāl-lik sīdi šnu wlədti? gā[l]t-lu wlədt bənt. gāl-lha gūli-li šnu wlədti rāh ‘ila wlədti l-bənt ga-ndəbhk w ndbəḥha. ta šāft-u zāyd-lha b-l-mūs, gā[l]t-lu hda, wlədt l-wəld

went-he that-one the-slave at-her said-he to-her said-he-to-you master-my what bore-you? said-she to-him bore-I girl. said-he-to-her tell-me what bore-you see-here if bore-she the girl will-I-slay-you and I-slay-her. until saw-she-him coming at-her with-the-knife, said-she-to-him calm-down, bore-I the son

The slave went to her, said to her, ‘My master says, what did you bear?’ She told him, ‘I had a girl.’ He told her, ‘Tell me what you had--if you had a girl, I will slay you and slay her.’ Until she saw him coming at her with the knife. She told him, ‘Calm down, I had a son.’

The discrepancy in marking between the indefinite *girl* and definite *son* may be attributable to the social importance of the male child, giving him a higher degree of individuation. In any case, it is clear that these features operate as semantic continua, leaving a fair degree of control to the speaker.

Another feature that seems to attract definite marking to Moroccan nouns is that of inalienable possession. Diem (1986) investigates the concepts of alienability and inalienability in Semitic, and suggests that this broad distinction is useful in determining the grammaticality of sentences containing pseudo-verbs /‘and/ and /li-/ , both meaning *to have*, in spoken Arabic. These concepts are indeed important, and in fact they are more broadly applicable than Diem suggests, especially to Moroccan. Diem cites two contrasting examples from a Fez informant, the one on the right judged to be grammatical, and the one on the left ungrammatical (1986:278):

M	عنده بآ مشهور * *‘ndu bba mšhūr * Er hat einen berühmten Vater [He has a famous father]	M	عنده دار كبيرة ‘ndu dār kbīra Er hat ein grosses Haus [He has a big house]
---	--	---	---

The ungrammaticality of the second example, Diem claims, is due to the inalienable nature of *father*. A more precise explanation is that this inalienable nature interferes not with the construction *to have*, but with the reference of the noun itself: /bba/ *father* is interpreted by my informants as *my father*:

- M1 عَيْطَت عَلِي بَا
 ‘əyyətt ‘la bba
 called-I on father
I called my father (elicited)

When pushed to make some sense out of Diem's "ungrammatical" example, my Moroccan informants came up with the interpretation *He thinks my father is famous*. Moreover, they found the following sentence acceptable:

- M1 عنده بَاه مشهور
 ‘ndu bbāh mšhūr
 at-him father-his famous
He has a famous father or His father is famous

The concept of inalienability applies to /bba/ *father* as a noun, regardless of its syntactic position, in that a father cannot be indefinite or "unassigned." The word /bba/ may not be interpreted as *a father* but only as *my father*. An unusual "double" genitive construction cited by Harming shows a similar pattern (1980:132):

- M باباها ديمَا
 bābāhā de-yimma
 father-her gen Mother
der Vater meiner Mutter [my mother's father]

Here, too, the double marking of the possession seems to be motivated by the need to mark the noun /bābā/ *father* as definite.

Similarly, my Moroccan informants have difficulty producing an indefinite form of the word /xa-/ *brother*. In the next passage, the speaker marks /xāy/ *my brother* and /‘ammi/ *my uncle* for possession, resulting in definite noun phrases, even though the sentence clearly indicates that she has neither a brother nor an uncle:

- M9 ما عندي خاي ما عندي عمي ما عندي تا شي واحد ماش ينوب علي
 mā ‘ndi xāy mā ‘ndi ‘ammi mā ‘ndi ta ši waḥəd māš ynūb ‘liyya
 neg at-me brother-my neg at-me uncle-my neg at-me even any
 one will act-on-behalf of-me
I don't have a brother, I don't have an uncle, I don't have anyone who would act on my behalf

Moroccan speakers thus consistently avoid using terms for male relatives in the indefinite. Diem's concept of inalienability, applied to the individuation continuum, explains this pattern. In turn, it is worth noting that these persons have a high degree of agency, which also may operate to attract specific, definite marking. The individuation hierarchy explains why certain kinds of nouns cannot remain "unassigned" or unspecified. The higher the individuation of a given noun or noun phrase, the more likely it is to receive definite syntactic marking. In the case of Moroccan, certain nouns with high animacy indicating close familial relations seem to take definite marking even in contexts low in definiteness.

The continuum of individuation also explains another feature particular to Moroccan speech. In most varieties of Arabic, both members of a definite noun-adjective phrase must agree in definiteness, such that both will carry the definite article, as the following Egyptian and Syrian examples show:

- | | | | |
|----|-------------------------|----|------------------------|
| E1 | التايير الموف | S2 | المجتمع السوري |
| | it-tayēr il-mōv | | il-mužtama' s-sūri |
| | the-outfit the-mauve | | the-society the-Syrian |
| | <i>the mauve outfit</i> | | <i>Syrian society</i> |

However, my Moroccan data include examples of asymmetrically definite constructions such as the following:

- M10 كيتباعوا ف الحانوت عصري ...
 ... kaytbā'u f l-ḥānūt 'ašri
 ... indic-they-are-sold in-the-store modern
 ... *they are sold in a modern store*

The phrase /l-ḥānūt 'ašri/ *at the modern store* consists of a definite noun (/l-ḥānūt/) modified by an indefinite adjective (/‘ašri/), a construction not permitted under the syntactic rules of Arabic. If it were an isolated occurrence, it could be dismissed as a performance error; however, several such examples occur in my data. Moreover, Harrell notes the occurrence of asymmetrically definite noun-adjective phrases in Moroccan, and notes that the adjective tends to remain unmarked unless the reference is quite specific (1962:166). In fact,

specificity may not be the only factor involved. The following example contains a noun-adjective phrase in which the noun is specified, but the adjective is indefinite:

M1 عندها ذوقها خاص
 ‘əndha dūqha xāṣṣ
 at-her taste-her special

She has her own taste or She has a special taste of her own

The asymmetric definiteness of the phrase /dūqha xāṣṣ/ *her own special taste* may be explained by the low animacy and high abstractness of the noun /dūq/ *taste*, rendering the noun relatively low in individuation, and attracting the indefinite form of the adjective. A Moroccan informant confirms that the adjective is marked definite only in highly individuated contexts, such as the following:

M3 عندها الدار المخبيرة في الشارع
 ‘əndha d-dār l-mxayyera f š-šārə‘
 at-her the-house the-choice in the-street

She has the best house on the street (elicited)

These Moroccan examples show that indefinite marking can interact with definite marking when the noun in question is not highly individuated. The reverse can also happen: at times definite marking can interact with indefinite to give higher specificity or individuation to an otherwise indefinite noun. Chapter 3 will show that this latter phenomenon occurs as well in relative clause structures in all four dialects.

1.7 Summary

The data cited in this chapter show clearly the inadequacy of traditional categories of definite and indefinite, and suggest that definite and indefinite marking represent two ends of a continuum of definiteness, which includes an indefinite-specific range that may be expressed syntactically by one of several strategies.

The articles /ši/ and /wāḥid/ do not mark true indefinite nouns, but rather nouns that lie somewhere between definite and indefinite, in an indefinite-specific range. This range is represented in the syntax of

all four dialects, in Moroccan and Syrian by the article /ši/, in all four dialects by the article /wāḥid (l-)/, which functions as a “new-topic” article, and in Syrian, Egyptian and Kuwaiti by the interaction of the semantics with other definite and indefinite markers, such as the use of definite relative pronouns with morphologically indefinite nouns.⁷

The geographical distribution of indefinite articles in Moroccan and Syrian, both of which share /ši/, the related Egyptian article /iṣi/, and the apparent indefinite-specific function of the suffix /-in/ in the Arabian Peninsula—even though it is disappearing in urban dialects—are facts that invite further diachronic research. The fact that Moroccan and Syrian dialects share the article /ši/ and Tunisian and Iraqi share a similar article /fard/ suggests that these articles have fulfilled this function for a very long time. Likewise, evidence that nunation plays a role as specifying marker in Andalusian texts, a medieval *Thousand and One Nights* manuscript, and the bedouin dialects of the Gulf, suggests that this function has long been part of spoken Arabic. The absence in Egyptian and urban Kuwaiti of an indefinite-specific article suggests that these dialects may have undergone a linguistic levelling process which reduced the number of syntactic markers of definiteness and specificity.

Overall, Moroccan and Syrian dialects show greater richness and variation in nominal syntactic marking, leading to speculation of a degree of cross-dialect hierarchy in which Moroccan shows highest definite and indefinite-specific marking, while Egyptian shows the least. The fact that Cairene Arabic falls on the lower end of the spectrum of dialects in the range of specifying articles is paralleled also by its relative paucity of demonstrative forms, as Chapter 4 will show.

⁷Other dialects appear to have indefinite-specific articles as well: Tunisian and Iraqi appear to share a specifying article /fard/ *one* (see Marçais 1977 for a description of Tunisian and Erwin 1963 for Iraqi).