MDPI

Article

Definiteness Systems and Dialect Classification

Mike Turner

Department of World Languages & Cultures, The University of North Carolina Wilmington, Wilmington, NC 28403, USA; turnerml@uncw.edu

Abstract: In this article I explore how typological approaches can be used to construct novel classification schemes for Arabic dialects, taking the example of definiteness as a case study. Definiteness in Arabic has traditionally been envisioned as an essentially binary system, wherein definite substantives are marked with a reflex of the article *al-* and indefinite ones are not. Recent work has complicated this model, framing definiteness instead as a continuum along which speakers can locate referents using a broader range of morphological and syntactic strategies, including not only the article *al-*, but also reflexes of the demonstrative series and a diverse set of 'indefinite-specific' articles found throughout the spoken dialects. I argue that it is possible to describe these strategies with even more precision by modeling them within cross-linguistic frameworks for semantic typology, among them a model known as the 'Reference Hierarchy,' which I adopt here. This modeling process allows for classification of dialects not by the presence of shared forms, but rather by parallel typological configurations, even if the forms within them are disparate.

Keywords: definiteness; indefiniteness; specificity; referentiality; determination; article systems

1. Introduction

To date, most efforts at classifying Arabic dialects have been concerned with grouping dialects on the basis of shared forms. At times, these forms have been phonological, such as the reflexes of **q* that inform the well-known sedentary–bedouin division; at others, they have been morphological, such as the 1SG imperfective prefix *n-* that differentiates western from eastern varieties (see Palva 2006 on these, among others). In this paper I put forth an alternate proposal: that it may be beneficial to look past forms themselves, and add to our toolset the use of semantic typology as a metric for grouping and subgrouping dialects. In doing so, the possibility arises that formally dissimilar features in two or more varieties may actually have more in common than previously thought, at least to the extent that the features in question exhibit the same types of polysemy. This approach is not exclusive of existing classification schemes. Instead, it may be seen as a way to further test and refine previous characterizations, or otherwise break a tie when a classification decision is questionable.

Although the typological approach itself can theoretically be applied to any number of interrelated feature sets, I opt to focus here on the interplay between nominal morphosyntax and a set of semantic notions that I refer to with the umbrella term 'definiteness'. The choice to use the term holistically follows that of other works, including Lyons (1999), similarly titled *Definiteness*, and presumes Chafe's (1976, p. 39) definition of the same as "whether I think you already know and can identify the particular referent I have in mind". Nonetheless, to be clear, in speaking of 'definiteness systems', my focus is on a particular range of definite-indefinite meanings, including relevant subcategories, that can accompany common nouns in response to Chafe's question (whether or not the answer is affirmative). Definiteness is a useful feature set with which to test a typological classification approach for various reasons, among them that (1) it can be modeled with a reasonable degree of precision, (2) Arabic dialects are known to differ in the ways they express it, and (3) sufficient material exists such as to be able to model discrete dialects and compare them, at least on a preliminary basis.

Most discussion of definiteness in the Arabic dialectological literature is, as is often true of other features, primarily concerned with formal representations. These discussions can be subdivided into two primary types, the first being the shape and assimilation patterns of the so-called "definite article" *al-[1], and the second being the presence and shape of "indefinite articles" in dialects that exhibit them. In the case of the former, the article *al- typically receives little explicit semantic discussion, as it is usually presumed to indicate true definiteness. Indefinite articles have fared somewhat better, perhaps because they clearly depart from formal expectations imparted by the standard language, and differ within dialects themselves; Mion (2009) provides an excellent survey of these articles, and even provides a preliminary (form-focused) typology, though his paper stops short of placing them into a comparative semantic framework.

The organization of the present paper is as follows: I begin with a theoretical discussion of definiteness and models that can be used to envision it, especially as they apply to the Arabic case. Following that, and in keeping with the overall focus on meaning over form, I provide a tier-by-tier view of the primary semantic categories attested in the above models, providing evidence of variation in Arabic by drawing on material from the dialectological literature. The next section provides more complete models of a sample of discrete Arabic dialects, selected again to exhibit the extent of possible variation, and to allow for side-by-side comparison. Finally, I return to questions of dialect classification, including both how we can construct schemes from the present data and how these schemes might interact with classification proposals previously made.

Because linguistic examples are drawn from various sources, many of which exhibit different conventions, I have adapted them (with the exception of Nubi) into a single transcription system and provided my own interlinear glosses and free translations.[2] In addition, throughout this paper I follow Dryer's (2014, p. e234) in adopting an intentionally broad and more semantically oriented definition of the term 'article,' which is used interchangeably with 'marker' to refer to any morphosyntactic structure that adds referential meaning to a noun. As such, its use here should not be understood as a syntactic judgment of any particular form.

2. Modeling Definiteness

As a starting principle, definiteness (in the holistic sense) is presumed here to be a semantic property of nouns in all human languages, stemming from shared cognitive perceptions of the world, entities within it, and other humans' knowledge of them. This semantic view is distinct from the grammatical expression of definiteness, which may be realized differently (or not at all) on a language-by-language basis. Dryer (2005a, 2005b) respective overviews of definite and indefinite articles for the *World Atlas of Language Structures* (WALS) underscore this point, showing that common cross-linguistic definiteness systems include formal representation of (1) both definiteness and indefiniteness, (2) definiteness but not indefiniteness, (3) indefiniteness but not definiteness, and (4) neither definiteness nor indefiniteness. Despite the variability of possible arrangements, maps of the same data show that they are not distributed at random, but rather display areal characteristics, often bridging disparate language families that are geographically proximate, but then varying inside a single language family that is geographically distributed. As Arabic falls into the latter category, that it sees variability in the expression of definiteness is a reasonable initial assumption.

Although grammarians often speak of "definiteness and indefiniteness" in binary terms, scholars have nonetheless recognized that definiteness and its expression cannot adequately be envisioned on a bipartite basis. In the past half-century, various models have been offered as visualizations of the cognitive statuses that underlie nominal referentiality, a common component of which has been the subdivision of either the 'definite' or 'indefinite' categories—often both—into more precise subcategories. These models have also generally recognized the same ordering of categories, which form a sort of continuum along

which formal representations might be distributed. Here I briefly review some of these models and select one for the present task, then move more explicitly into the Arabic case.

2.1. The Wheel Model

Givón (1978, p. 298) proposes a wheel-shaped model that distinguishes six possible nominal statuses, which he identifies as (a) 'referential definite', (b) 'referential indefinite', (c) 'referential nondefinite', (d) 'nonreferential object', (e) 'generic predicate', and (f) 'generic subject', with the first and last categories bordering each other. Figure 1 shows this model as he envisioned it for standard English. The choice of a wheel is motivated by Givón's observation that, while languages often use a single morphosyntactic strategy (possibly including zero-marking) for two or more statuses at once, their distribution across categories is nearly always contiguous. One notes, for example, that the English 'indefinite article' *a* (or *an*) can indicate multiple underlying semantic statuses. Givón's terms are somewhat clumsy—it not immediately apparent how one would contrast 'indefinite' and 'nondefinite' without reviewing examples—but they do establish the basic principle of multiple semantic distinctions underlying a single form. He also rightly indicates that plural and singular forms do not have to follow the same patterning, and uniquely carves out space in his model for generic entities.[3]
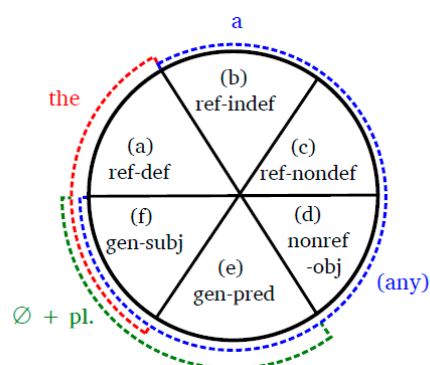

Figure 1. Givón's Wheel Model, for English (redrawn).

2.2. The Givenness Hierarchy

Gundel et al. (1993) approach the same issue more broadly, framing definiteness as a subcomponent of a larger set of meanings, including those indicated by personal and demonstrative pronouns, that they refer to as 'givenness'. They propose a 'Givenness Hierarchy' (Table 1) consisting of six cognitive statuses, wherein the more discursively 'known' or 'given' a referent it is, the further to the left of the hierarchy it will be. The three rightmost statuses in the Givenness Hierarchy might be seen as corresponding with the four statuses (a)–(d) of Givón's Wheel Model, showing a discrepancy in the choice of subdivision despite a general agreement that subdivisions should exist. One contribution of Gundel, Hedberg, and Zacharski is that they provide a formal representation of one of the 'indefinite' subcategories by giving informal English *this* as an indefinite article, a use that is further confirmed in Ionin (2006), who calls it a 'specific' marker. As it is useful to be able to provide semantically nuanced free translations, I make ample use of indefinite *this* in translations of Arabic examples in this paper.

Table 1. The Givenness Hierarchy (Gundel et al. 1993).

in Focus	>	Activated	>	Familiar	>	Uniquely Identifiable	>	Referential	>	Type Identifiable
it		*that* *this* *this* N		*that* N		*the* N		indefinite *this* N		*a* N

2.3. The Reference Hierarchy

Drawing together advantages of both the Wheel Model and the Givenness Hierarchy, a more recent proposal by Dryer's (2014, p. e235) by the name of the 'Reference Hierarchy' (Table 2) combines the more limited scope and greater categorical distinctiveness of the former with the hierarchical implications of the latter. Dryer's model enjoys the unique advantage of having been constructed on the basis of a large corpus of real-world language data, which featured in his (Dryer 2005a, 2005b) work for the WALS database; as such, it is likely to be sufficient for the description of most languages (including Arabic). Like Givón before him, Dryer emphasizes the tendency of articles to be both polysemous and contiguous across a particular range of meanings; meanwhile, like Gundel, Hedberg, & Zacharski, Dryer relies on the notion of a hierarchical relationship whereby nouns that are more 'known' or 'given' are located further to the left. His choice of five categories is more akin to the Wheel Model, though he leaves out generics and splits 'referential definites' into 'anaphoric definites' and 'nonanaphoric definites'. Also like the Wheel Model, the Reference Hierarchy proposes three non-generic indefinite statuses, i.e., one more than the Givenness Hierarchy indicates. Finally, although Dryer's particular terminologies are lengthy, he does provide a set of 2 to 3-letter abbreviations (in heading of Table 2), which are particularly suitable for in-line reference and interlinear glosses.

Table 2. The Reference Hierarchy (Dryer's 2014), with abbreviations.

AD		ND		PSI		PNI		SNI
anaphoric definites	>	nonanaphoric definites	>	pragmatically specific indefinites	>	pragmatically nonspecific (but semantically specific) indefinites	>	semantically nonspecific indefinites

2.4. Applying the Reference Hierarchy

Because it captures the advantages of models before it, was specifically proposed as a response to cross-linguistic data, and allows for abbreviated reference to particular semantic statuses, I opt to use the Reference Hierarchy as the working model for the current paper, and hereby adopt the terms AD, ND, PSI, PNI, and SNI for their respective meanings. These abbreviations are henceforth used liberally in both glosses and prose. It is nonetheless worth pointing out that broad terminological consensus has yet to emerge within this field of inquiry, so I summarize each status as follows, for clarity:

1. Anaphoric definite (AD), which is a subset of both Givón's 'referential definite' and Gundel, Hedberg, & Zacharski's 'uniquely identifiable', refers to the status of a noun that the speaker presumes identifiable to the listener because the referent has already been explicitly introduced or implied in the present discourse. In English it is obligatorily marked with *the*, and optionally with the demonstrative adjectives *this* or *that*.

2. Nonanaphoric definite (ND), which is also a subset of Givón's 'referential definite' and Gundel, Hedberg, & Zacharski's 'uniquely identifiable', refers to the status of a noun that a speaker presumes identifiable to the listener because the referent is available through shared world knowledge. In English it is obligatorily marked with *the*.

3. Pragmatically specific indefinite (PSI), which corresponds with Givón's 'referential indefinite' and is a subset of Gundel, Hedberg, & Zacharski's 'referential', refers to the status of a noun that the speaker can uniquely identify but presumes the listener cannot. It has elsewhere been called 'specific', and in English is obligatorily marked with *a(n)*, or in more informal varieties with *this* (Ionin 2006).

4. Pragmatically nonspecific (but semantically specific) indefinite (PNI), which corresponds with Givón's 'referential nondefinite' and is a subset of Gundel, Hedberg, & Zacharski's 'referential', refers to the status of a noun that neither the speaker nor listener can uniquely identify, but which the speaker conceptualizes as being distinct

from others of its type. It has elsewhere been called 'existential', and in English is obligatorily marked with *(a)n*, but can also be marked with *some* (Israel 1999).

5. Semantically nonspecific indefinite (SNI), which corresponds with Givón's 'nonreferential object' and Gundel, Hedberg, & Zacharski's 'type identifiable', refers to the status of a noun that is fully unindividuated and is interchangeable with any other of its type. In English it is obligatorily and exclusively marked with *a(n)*.

Using the above definitions, it is possible to build a visual representation of a given language's definiteness system by representing the Reference Hierarchy as a series of blocks along which corresponding forms can be mapped. Figure 2 gives my interpretation of the system in spoken American English. The articles represented at top, *the* and *a(n)*, are obligatory; meanwhile, the forms at bottom represent auxiliary strategies. This strategy is maintained for other iterations of the model in this paper. The visual model has the added benefit of easing comparison between multiple systems, as is our purpose here, and explored further in Section 4.
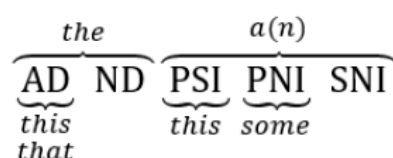

Figure 2. Forms represented along the Reference Hierarchy for American English.

2.5. Definiteness in Arabic

A handful of works to date have treated definiteness (or aspects of it) in Arabic specifically. Of these, Brustad (2000, pp. 18–43) is the most immediately relevant in both its focus on spoken Arabic and its comparative approach. She introduces the idea of a 'definiteness continuum' that includes not only meanings that are "wholly definite" or "wholly indefinite", but also exist within an intermediate range that she terms 'indefinite-specific'. Within the current framework, "wholly" definite and indefinite correspond with the statuses AD/ND and SNI, respectively; meanwhile, the indefinite-specific range that Brustad speaks of seems to cover both PSI and PNI. Looking at Moroccan, Egyptian, Syrian, and Kuwaiti dialects, Brustad identifies common patterns, among them the marking of true definites (AD/ND) with a reflex of *al-*, as well as the zero-marking of non-referential (SNI) nouns. Taken alone as a binary opposition, this initial observation corresponds with the way definiteness in Arabic is often framed.

At the same time, Brustad also establishes the presence of structures that add more nuance than the binary model allows, many of which vary by dialect. Within the indefinite-specific range, she documents use of reflexes of *wāḥid* 'one' for all four dialects, observing that it often marks a new topic that is subsequently adopted in the discourse. I qualify such referents as inherently PSI, in that new topics are necessarily known to the speaker—who can therefore expound upon them—but are presumed inaccessible to the listener. Nonetheless, as Brustad notes that *wāḥid* is often restricted to humans (e.g., *wāḥid badwi* 'a certain bedouin', p. 20), I am more inclined to read it in such cases as an indefinite pronoun modified by an adjective (i.e., *someone (who is a) bedouin*) rather than a truly inclusive article that can modify any common noun. The exception is in Moroccan, which I discuss more specifically in Section 3.3.

For Moroccan and Syrian varieties, Brustad locates an article *ši*, which she glosses as 'some (kind of)' and contends speakers use "to indicate that they have a particular type of entity in mind". Brustad also raises the possibility of interpreting dialectical *tanwīn* as a sort of indefinite-specific marker, citing Ingham (1994, pp. 47–50) comments on its semantic qualities in Najdi Arabic, and shows how both partitive structures and demonstrative adverbs can have the same semantic effect in Egyptian (Brustad 2000, pp. 30–31). Under the broad definition of 'article' used here—which, again, privileges semantic function over syntactic analysis—I consider such structures part of a given dialect's article system, and specifically include them in below models.

Elsewhere, Brustad complexifies uses of the article **al-*, typically seen to be a marker of true definiteness. Two principal qualifications arise from her data. The first of these is that while true definite (AD and ND) nouns are consistently represented with **al-*, anaphoric definites are often further marked with an unstressed demonstrative adjective (*hād-*, *ha-*, etc.) as a means of increasing their discursive prominence (112–139). I see this common strategy as akin to other auxiliary strategies for marking particular referential meanings, and thus class these as a type of AD marker. The second qualification involves the presence of **al-* in apparently indefinite contexts, which Brustad identifies as a common occurrence in Moroccan (e.g., *xəṣṣ-ni l-wəld* 'I need a son'; p. 36). I interpret this as evidence that the Moroccan reflex of **al-* is distributed over a wider range of referential statuses in general (see Section 4.9).

There are few other holistic studies of definiteness in spoken Arabic. Turner (2018) is comparative and concerned exclusively with spoken Arabic, and employs the same descriptive model as the current paper to explore variability in spoken Arabic; the reader is encouraged to refer to it for additional data presented within the Reference Hierarchy framework. Fassi Fehri (2012, pp. 205–31) provides a more traditional syntactic view of determination in Arabic and Semitic at large, and includes some spoken Arabic data. Remaining studies that have relevance for the study of definiteness in Arabic can be divided into two types. The first are those that focus on single varieties, such as Caubet (1983), Belyayeva (1997), and Fabri (2001) focused and theoretically nuanced descriptions of definiteness in Moroccan Arabic, Palestinian Arabic and Maltese, respectively. The second type of relevant studies are those that examine a single form through a multifunctional semantic lens, and include in turn accounts of its articular functions; among these, Wilmsen's (2014) expansive account of *ši* across Arabic varieties and Leitner and Procházka's (Forthcoming) examination of *fard* in the dialects of Iraq and Khuzestan stand out.

3. Points of Variation

Following from Brustad's observations that structures not traditionally recognized as articles can, on a semantic and pragmatic level, be used to indicate particular referential meanings, a set of metrics for locating these in situ is useful. Even for forms that have been recognized as articles—whether definite or indefinite—in previous literature, a semantics-first view allows us to more specifically delineate the range of meanings that they cover. The aim of this section is, accordingly, to walk through each of the semantic statuses along the Reference Hierarchy, describe how each can be located by discursive context, and identify some points of variation in regard to how each is expressed formally across spoken Arabic varieties.

Because the goal of the section is simply to survey variation, it is more concerned with the fact that a strategy is attested at all than it is with that strategy's relative frequency of use. Nonetheless, as it is useful for comparative purposes (which follow in Section 4) to establish a baseline measure of how grammaticalized a given strategy is, I do also offer here initial readings of where each falls on a conceptual continuum that ranges between fully 'obligatory' and 'auxiliary'. While obligatory articles are easy to define—they are used by all speakers for all instances of the target meaning—and auxiliary articles can be understood as marked structures that are used by speakers for special emphasis, there is also an intermediate category of markers that are used so frequently with their corresponding meanings such as not to be highly marked, but are still not obligatory in all cases. I refer to these markers as 'conventionalized', a placeholder term used with an understanding that truly accurate frequency judgments will require more in-depth semantic study of individual varieties.

3.1. Anaphoric Definites

Anaphoric definites are easily located in extended discourse because they simply involve subsequent reference to an entity that has already been explicitly introduced. In the Hassaniya Arabic sentence given in (1), for example, the narrator introduces a certain

sba' 'lion' as a new referent; when it re-occurs in the text, the referent *sba'* is now necessarily AD, and is accordingly marked with **al-*:

raṣṣaf	*'lī-h*	*sba'*	...	*yraṣṣaf*	*'lī-h*	<u>*s-sba'*</u>	*w-*	*ygūl*	*-lu*	(1)
jump.PFV.3MSG	upon-3MSG	lion.PSI		jump.IPFV.3MSG	upon-3MSG	AD-lion	and-	say.IPFV.3MSG	-3MSG.DAT

'This lion jumped on him . . . <u>the lion</u> is jumping on him and saying to him . . . ' ([Heath 2003](), p. 116)

That the article **al-* is used here as the marker of anaphoric definiteness is not particularly surprising to anyone with knowledge of Arabic, formal or informal, and in most varieties it is indeed the sole obligatory marker of AD nouns. Nonetheless, the point of the example is to highlight contextual expectations. Importantly, when the same sort of discursive context is located elsewhere in the same variety, we find the variation of the type noted by Brustad elsewhere, namely in the auxiliary use of an unstressed demonstrative, as in (2):

w-	*žayna*	*l-*	*fullāni*	*-wāḥəd*	...	*wə-*	<u>*dāk*</u>	<u>*l-fullāni*</u>	*āba*	*yarḥal*	(2)
and-	come.PFV.1PL	to-	Fulbe	-PSI		and-	DEM.AD	AD-Fulbe	refuse.PFV.3MSG	leave.IPFV.3MSG

'We came to this Fulbe man . . . <u>the Fulbe man</u> refused to leave' ([Heath 2003](), p. 78)

While I am not aware of any survey beyond Brustad's, many Arabic varieties exhibit the demonstrative anaphoric reinforcement pattern in one way or another, and because demonstratives themselves vary widely in form, but mirror each other semantically, it is not particularly useful to list off all possible forms here (although see [Magidow 2016]() for a survey). On a typological level, it is also unsurprising that the demonstrative frequently plays this role, given demonstratives are a frequent source of definite articles in world languages ([De Mulder and Carlier 2011]()). Instead, what is more worthwhile to note in the Arabic case is the degree to which a variety has conventionalized the demonstrative as an AD marker, at which point it might be said be an article of its own. At least some Levantine dialects appear to meet this description, as is evident in the use of *hal-* (etymologically *hā* + *il-*) in (3), from Baskinta (Lebanon):

'in-na	*žār*	*ib-*	*haḍ-ḍay'a*	*b-iḥibb*	*an-nawm*	...	*nāyim*	<u>*haz-zalami*</u>	(3)
have-1PL	neighbor.PSI	in-	AD-village	IND-sleep.IPFV.3MSG	GEN-sleep		sleeping.PTCP	AD-man

'We have this neighbor in the village who loves to sleep . . . <u>the man</u> was sleeping' ([Abu-Haidar 1979](), p. 141)

Just how widespread this pattern is in the Levant warrants further study[4], but for the present purpose it is enough to point out that a dialect that could be shown to *obligatorily* mark AD nouns with a certain structure, but not ND ones, would be typologically distinct from most other varieties at present, and worthy of recognition of such. This phenomenon is also attested in the Nubi Arabic-based creole, wherein a postposed demonstrative reflex *'de* accompanies AD nouns. The major difference is that, in Nubi, the Arabic article **al-* has been lost entirely:

'bas	*'uo*	*'jowzu*	<u>*bi'niya*</u>	<u>*'de*</u>	(4)
well	3MSG	marry	girl	AD

'Well, he married <u>the girl</u> [previously mentioned]' ([Wellens 2003](), p. 67)

All of these strategies, of course, are overt, and they all incorporate either **al-* or a demonstrative (or a combination of both). The one major exception for AD nouns is the Central Asian cluster of dialects spoken in Uzbekistan (near Bokhara) and northern Afghanistan (near Balkh), which [Ingham]() ([2003]()) has suggested are branches of the same historical group (see also [Seeger 2013]()). These varieties neither have a reflex of **al-* nor have any obligatory compensatory strategy when nouns are AD, as in (5):

fad	*mara*	*kōnət*	. . .	*qōlət*	<u>*mara*</u>	. . .	(5)
PSI	woman	be.PFV.3FSG		say.pfv.3fsg	woman.AD		

'There was this woman . . . <u>the woman</u> said . . . ' ([Jastrow 2005](), p. 138)

That said, even these varieties use demonstratives anaphorically, as in *duk zaġīr* 'the child [previously mentioned]' ([Ingham 2003](), p. 33), so they do have at least an auxiliary means of overtly marking AD statuses. In this sense, the Central Asian group shares a typological feature with the larger dialect landscape, even if it is missing a 'core' Arabic feature in its lack of **al-*.

3.2. Nonanaphoric Definites

Nonanaphoric definites are uniquely identifiable to both the speaker and listener via world knowledge, and they can be distinguished from AD nouns in extended discourse in that they have not previously been introduced. Common nouns that are ND in most circumstances include 'the sun,' 'the world', 'the country', 'the king', and any other for which there is only likely to be one possible interpretation on the part of the listener, despite being new to the discourse; as such, they are relatively easy to locate. This semantic status shows the least variation from dialect to dialect, and is most often represented by **al-* to the exclusion of all other strategies (including demonstrative reinforcement). A typical example is in (6), from the Jazira area of Sudan, where 'the mayor' is unique and identifiable as the mayor of the implied town in the narrative despite only being mentioned for the first time:

rawwaḥ	*lə-*	<u>*l-ʿumda*</u>	*šakā*	*-ni*	*ʿalē*	. . .	(6)
go.PFV.3MSG	to-	ND-mayor	complain.about.PFV.3MSG	-1SG.OBJ	to.3MSG		

'He went to <u>the mayor</u> and complained about me to him' ([Hillelson 1935](), p. 48)

The primary exception to this pattern is, predictably, varieties that have lost the article **al-*; in such cases, the ND noun is unmarked. The Afghanistan Arabic utterances in (7), for example, provide first mention of 'the queen' with no further modification. Similar unmarked patterns can be identified in Nubi, as in *'hari ta 'shems* 'the heat of the sun' ([Wellens 2003](), p. 67). It is worth noting that these varieties, like others, do not see auxiliary use of demonstratives for ND nouns, even though they allow them for AD nouns.

<u>*malika*</u>	*li-*	*zōl*	*kasir*	*zīn*	*kōm*	*mi-ššūf*	(7)
queen.ND	OBJ-	Zal	very	wonderful	be.PFV.3MSG	IND-see.IPFV.3FSG	

'<u>The queen</u> thought that Zal was very wonderful' ([Ingham 2003](), p. 34)

3.3. Pragmatically Specific Indefinites

Pragmatically specific indefinites can be identified in extended discourse as referents that are mentioned for the first time, and not accessible to the listener via world knowledge, but for which the speaker can thereafter be seen to provide specific information. Strategies for marking PSI nouns are the most varied and innovative, particularly if we are to adopt a wide view of what an article is, and many have been under-recognized to date. Most of the "indefinite articles" of the dialectological literature are, in fact, PSI articles, whether exclusively or in a polysemic distribution with the PNI status.

A common source for PSI articles is, as is common in world languages ([Heine 1997](), pp. 66–83), a numeral **wāḥid* 'one' or **fard* 'one, an individual'. The former of these is best associated with Moroccan and western Algerian varieties, where a reflex of **wāḥid* is typically obligatory for new, pragmatically salient referents of which the speaker has unique knowledge. Unique to this structure, however, is that **wāḥid* accretes with **al-*, yielding a sort of double-marked structure. [Caubet]() ([1983](), p. 83) gives the Moroccan article as a fused *wāḥad-əl*, which is a plausible reading in most cases, but I venture that the article *l-* itself might also be considered a PSI marker, especially as it can be syntactically detached from *wāḥad* but still coincide with a clear PSI meaning, as in (8) from Anjra (Morocco):

hāda	*wāḥd*	*əl-ʿāyəl*	*u-*	*l-ʿāyla*	*ma-*	*yžəbru*	*-ši*	*fāyn*	*ysəknu*	(8)
EXIST	PSI	PSI-boy	and-	PSI-girl	NEG-	find.IPFV.3PL	-NEG	REL.LOC	live.IPFV.3PL	

'There was <u>this boy</u> and <u>this girl</u> who couldn't find anywhere to live' ([Vicente 2000](#), p. 221)

The articular use of **wāhid* to mark PSI referents is also attested in eastern varieties of Hassaniya, as spoken in Mali, though here it is suffixed rather than prefixed, and is not obligatory. It has not been explicitly recognized as such, but is regularly apparent in contexts such as (9), recorded in Gao, where further specification of the noun *blad* 'place' makes it clear that the speaker has unique knowledge of it. A similar structure is documented in Nubi, e.g., *mas'kin ʿwai* 'a certain poor man' ([Wellens 2003](#), p. 64).

dxalna	*blad*	*wāḥəd*	*yəngāl*	*-lu*	*hari-bomo*	*fī-h*	*s-sbuʿa*	*yāsrīn*	(9)
enter.PFV.1PL	country	PSI	call.PASS.IPFV.3MSG	-3MSG.DAT	Hari-Bomo	in-3MSG	SNI-lions	many.PL	

'We entered <u>this place</u> that's called Hari-Bomo; there are a lot of lions there' ([Heath 2003](#), p. 110)

The article **fard*, of similar semantic provenance, is widely recognized in the dialectogical literature, where it is most often associated with Mesopotamian varieties. [Blanc](#) ([1964](#), 118) locates this article in Baghdad, and describes phonological variants of it associated with particular sectarian groups, but gives limited semantic information, saying "its presence contrasts fairly clearly with that of the article /l/ or other determination marks, but the degree to which it contrasts with absence of any mark is yet to be determined." Recent work by [Leitner and Procházka's](#) ([Forthcoming](#)) significantly expands on the functions of **fard*, showing that it is a polyfunctional lexeme with multiple senses, one of which is to mark a noun that is "new for the hearer and important for the subsequent discourse." This quintessentially PSI sense for **fard* is attested throughout Iraq and Khuzestan, as in (10), from Basra, where the speaker starts a story by introducing a particular *ṭālib* 'student':

fadd	*yōm*	*fadd*	*ṭālib*	*rāḥ*	*li-*	*l-madrasa*	*mitʾaxxir*	(10)
PSI	day	PSI	student	go.PFV.3MSG	to-	ND-school	late	

'One day <u>this student</u> went to school late' ([Denz and Edzard 1966](#), p. 78)

[Mion](#) ([2009](#)) locates reflexes of **fard* in other Arabic varieties, too, including those of Mardin and Tunis, but in most of these cases the reflex is less apparently referential and simply implies 'one, the same' (though potential for future reanalysis remains). Nonetheless, it is attested with a clear PSI meaning in Central Asian varieties, as in *fad mara* 'a [certain] woman' in (5), above.

These are the only structures regularly called 'articles' in the literature, to my knowledge, that meet the semantic parameters of PSI, but under the broad definition we can easily expand the field of extant PSI articles. The first sort of novel article is derived from the demonstrative adverb, but has the same pragmatic effect of indicating a referent that is identifiable to the speaker, but not the listener. Brustad offers this interpretation of *kida* in Cairene (*šuft ḥāga kida* 'I saw this thing … '), a view that is supported by numerous examples in [Woidich](#) ([2006](#), p. 236). The same function can also be located elsewhere in Egypt, as in (11), from Bani Swayf:

bi-niʾmil	*-laha*	*ḥuwəyza*	*sġayyaṛa*	*kida*	*ʿala*	*ʾaddi-ha*	(11)
IND-do.IPFV.1PL	-3FSG.DAT	box	small.F	PSI	upon	size-3FSG.POSS	

'We make <u>this little box</u> for it, in its size' ([Behnstedt and Woidich 1988](#), p. 16)

Furthermore, there is evidence for a parallel strategy in some Yemeni varieties, which use the demonstrative adverb *hākaḏāha* (and similar; see [Watson and ʿAmri 1993](#), pp. 418–19) to the same effect; in (12), for example, the speaker introduces a *bugʿah* 'place' and immediately provides more information, a hallmark of a PSI noun:

ḥna	*fiʿil-na*	*l-ʿiris*	...	*fi*	<u>*bugʿah*</u>	*hākaḏāhā*	*nisammī*	*-ha*	*mafraj*	(12)
1PL	do.PFV.1PL	AD-wedding		in	place	PSI	call.IPFV.1PL	3SG.OBJ	mafraj	

'We had the wedding … in <u>this place</u> that we call a *mafraj*' ([Watson and ʿAmri 2000](), p. 242)

Another structure that qualifies as a PSI marker on the basis of its semantic associations is the so-called 'dialectical *tanwīn*' (DT) of the dialectological literature. Even though the origins of this marker remain an object of debate, its functions are relatively similar across varieties. Stokes ([2020](), p. 637) summarizes DT as "the morpheme, typically realized as *in* or *an*, that is suffixed to a morphologically indefinite noun, primarily when followed by some type of adnominal adjective or clause". The fact that DT is restricted to indefinites is alone sufficient to establish that it has some relationship with the semantics of refentiality; in addition, that it typically proceeds an adnominal element—which, on a pragmatic level, individuate the noun as distinct from others of its type—calls for a PSI or PNI reading of the resulting phrase. As such, it is not surprising that it can be located with nouns that clearly meet the parameters of a PSI referent, as in (12) from the Jezira (Sudan):

qibēl	*ǧa*	*fōq-i*	<u>*gana-yan*</u>	<u>*šukri*</u>	*rākib*	*-lu*	*baʿīr-an*	*ḫūri*	(13)
earlier	come.PFV.3MSG	up-1SG	boy-PSI	Shukri	riding.PTCP	-3MSG.DAT	camel-PSI	yellow	

'Earlier this *Shukri* <u>boy</u> came up to me riding <u>this yellow camel</u>' ([Hillelson 1935](), p. 60)

While DT can accordingly be read as a sort of PSI article, in most cases it is still syntactically conditioned, in that it depends on the presence of an adnominal attribute (regardless of the speaker's ability to uniquely identify the referent). There is nonetheless evidence that some varieties have moved toward fully semanticizing DT, as in Najdi, for which Ingham ([1994](), p. 50) gives examples such as *ligēt bēt-in* 'I've found a [certain] house'. It is also possible to locate varieties in which a reflex of DT (which only occurs in this sense) accretes with another PSI article such as **wāḥid*, as can be seen in (14), from Tillo (Anatolia):

yeḥkaw	*baḥs*	*āv*	*ət-tattūn*	<u>*ḥakkoyət*</u>	*-ən-wəḥde*	(14)
tell.IPFV.3PL	about	DEM	AD-tobacco	story	-PSI-PSI	

'They tell <u>this story</u> about that tobacco' ([Lahdo 2009](), p. 229)

Finally, it is worth pointing out that in many varieties, underlying PSI referents are simply unmarked. Such nouns have the same underlying semantic properties, but are not overtly marked as such, either because a marker is unavailable or the speaker chooses not to use it. A typical example is in (15), from Baskinta (Lebanon).

ʿin-na	<u>*žār*</u>	*ib-*	*haḍ-ḍayʿa*	*b-iḥibb*	*an-nawm*	...	(15)
have-1PL	neighbor.PSI	in-	AD-village	IND-sleep.IPFV.3MSG	GEN-sleep		

'We have <u>this neighbor</u> in the village who loves to sleep' ([Abu-Haidar 1979](), p. 141)

3.4. Pragmatically Nonspecific Indefinites

Pragmatically nonspecific indefinites are neither uniquely identifiable to the speaker nor the listener, but are conceived of by the speaker as being distinct from others of their type in the world at large. Though a speaker of a variety that marks overtly PNI nouns can signal them as such in any desired context, from an observer's perspective this semantic status is most easily located where the speaker speculates about the potential nature of a unique referent not yet located; as such, it is often the object of verbs such as 'find', 'obtain', and 'make'. The most easily identifiable PNI article is *ši*, conventionalized in Levantine (16) and Moroccan (17), and which carries this sense exclusively when used as an article:[5]

kān	*yaʿmil*	*-lu*		<u>*ši*</u>	*mašḥra*	*w-*	*išīl*	*išwayyit*	*faḥm*	(16)
be.PFV.3MSG	make.IPFV.3MSG	-3MSG.DAT		pni	kiln	and-	take.out.IPFV.3MSG	bit	charcoal

'He would make himself <u>some sort of kiln</u> and produce a bit of charcoal' (Abu-Haidar 1979, p. 145)

ma-	*gāl*	*-līya*	*-š*	*smīyt-u,*	*gāl*	*-li*	*kāyn*	<u>*filəm məzyān*</u>	(17)
NEG-	say.PFV.3MSG	-1SG.DAT	-NEG	name-3MSG.POSS	say.PFV.3MSG	-1SG.DAT	EXIST PNI	film good

'He didn't tell me its name; he told me there's <u>some good film</u> [playing]' (Caubet 1993, p. 338)

The article **fard*, described above as a conventionalized marker of PSI statuses, is also attested with a PNI meaning, making the form itself polysemous, as in (18), from Baghdad. The *bayt* 'house' in question here is semantically specific, but the speaker has not located it yet. Reflexes of **fard* are used comparably in Central Asian varieties, as in *fad- ōrd* 'some place' (Ingham 2003, p. 34).

də-ndawwir	*ʿala*	<u>*fəd*</u>	<u>*bayt*</u>	*l-*	*il-ʾiğār*	*āni*	*w-*	*zawiğt-i*	(18)
ASP-search.1PL.IPFV	for	PNI	house	for	GEN-rent	1SG	and-	wife-1SG.POSS

'I'm looking for <u>some house or the other</u> for me and my wife to rent' (McCarthy and Raffouli 1965, p. 17)

Exhibiting similar polysemy, if we are to read it as a type of article, is dialectical *tanwīn*, which can also indicate a PNI meaning. This is evident in (19), from the Jezira (Sudan), where the speaker has no particular *arnab* 'rabbit' in mind, but implies God might:

allāh	*yəddī*	*-na*	*-lēna*	<u>*arnab*</u>	<u>*-an*</u>	*nitʿašša*	*b-a*	(19)
God	bring.IPFV.3MSG	-1PL.OBJ	1PL.DAT	rabbit	-PNI	have.dinner.IPFV.1PL	with-3FSG

'May God bring us <u>some rabbit</u> that we can have for dinner' (Hillelson 1935, p. 46)

Beyond these articles, I am not aware of any other regularly occuring PNI markers, and most varieties simply leave PNI nouns unmarked, as in (20) from Sanaa. This is not to rule out that partitive-like structures, in particular, might sometimes bridge into this meaning; Sanaani itself does, for example, occasionally use a form *zārat* with plurals or as part of the SNI indefinite pronoun *zārat wāḥid* (Watson and ʿAmri 2000, p. 114).

hānāk	*tilgā*	<u>*maktab*</u>	*illī*	*hum*	*yixarrijū*	*-k*	*bi-*	*s-siyāḥa*	(20)
there	find.IPFV.2SG	office	REL	3PL	take.out.ipfv.3pl	-2msg.obj	for	GEN-tourism

'There you'll find <u>some office or the other</u> that can take you out for tourism' (Watson and ʿAmri 2000, p. 26)

3.5. Semantically Nonspecific Indefinites

Semantically nonspecific indefinites are, by definition, interchangeable with any other entity of their type, and cannot be discursively prominent. As such, they are nearly always the object of a verb or preposition and not typically modified. Across Arabic varieties, SNI nouns are most commonly unmarked. The word *ḥbal* 'rope' in the Hassaniya example in (21) is typical:

gaṛṛanna	*lə-ḥmīr*	*...*	*kull*	*tlāta*	*fə-*	<u>*ḥbal*</u>	(21)
bind.PFV.1PL	AD-donkeys		every	three	to-	rope

'We bound the donkeys . . . each three with <u>a rope</u>. (Heath 2003, p. 110)

As a general rule, articles that fulfill the PSI or PNI function are not used to indicate SNI entities, though pragmatic considerations may occasionally let PNI markers bridge into this meaning.[6] In the case of *tanwīn*, which is both semantically and syntactically conditioned, the fact that SNI nouns are unmodified means there is no syntactic impetus for it to appear with them, and I am not aware of any examples that show it being used alone with any sense other than the PSI one noted in Section 3.3.

The primary exception to the general tendency of Arabic varieties to leave SNI nouns unmarked is, perhaps unexpectedly, in varieties that instead mark them with **al-*, at least in some circumstances. Moroccan is most notable for this, as in (22) and (23), where both the *tūr* 'bull' and *səlhām* 'cloak' are non-referential, being mentioned only once in passing:

dbəḥ		**ṯ-ṯūr**,	*'ṛəḍ*	*'la*	*n-nās*	(22)
slaughter.PFV.3MSG		SNI-bull	invite.PFV.3MSG	PREP	ND-people	

'He slaughtered <u>a bull</u>, invited people over … ' (Brustad 2000, p. 37)

lyūm	*huwa*	*lābəs*	**s-səlhām**	(23)
today	3MSG	wearing.PTCP	SNI-cloak	

'Today he is wearing <u>a cloak</u>' (Harrell 1966, p. 190)

That this pattern is attested and permissible is sufficient to call the view of **al-* as a *universal* definite article in Arabic into question.[7] That said, within Moroccan it is possible to find SNI nouns both with **al-* and with no marking at all. I have elsewhere argued that the marked pattern is more common with type-focused uses of SNI nouns and that the unmarked one is mostly reserved for delineating a specific quantity (Turner 2018, pp. 184–88). It is probably not prudent to call **al-* obligatory in this sense, but it is frequent.

4. Systems in Comparison

Taking the above data into account, it seems fair to say that there are a wide variety of strategies for expressing discrete definiteness values in Arabic dialects. This observation alone has implications for descriptive practice, as being aware of extant diversity within a linguistic group is always helpful in delineating which grammatical categories one should check for in fieldwork and comment on in publications. The greater promise of explicitly collecting such data, that said, is that it opens the door for new comparative approaches. In this section, I provide provisional sketches of the overall arrangement of definiteness systems in a sample of ten Arabic varieties, in addition to the Nubi Arabic-based creole, allowing for side-by-side comparison, before moving into the final discussion of how we might use such characterizations for classification. The rough order of sketches here is from more simplex systems to more complex ones, as I estimate them to be.[8]

4.1. Libyan

Libyan Arabic dialects, including those spoken in the eastern Benghazi area (Elfitoury 1976; Owens 1984) and Tripoli further west (Grand'Henry 2000; Yoda 2005), show a very strict binary division between definite (AD and ND) nouns, marked with *(i)l-*, and indefinite (PSI, PNI, and SNI) nouns, which are invariably unmarked. A review of texts in Grand'Henry (2000) confirms this impression, and I am not able to locate any regular auxiliary strategies. Figure 3 gives the distribution of forms in Libyan.

Figure 3. Reference Hierarchy for Libyan.

4.2. Egyptian

Egyptian varieties show the same basic pattern of obligatorily marked definite (AD and ND) nouns, and Brustad (2000, p. 140) specifically notes "the absence of an anaphoric demonstrative article in Egyptian." Brustad's data are from Cairo, but texts from Behnstedt and Woidich (1988) show the same patterns elsewhere in Lower Egypt. Although it does not have any obligatory means for indicating indefinite meanings, speakers of Egyptian do have the auxiliary marker *kida* for PSI referents (see Section 3.3). Figure 4 gives the distribution of forms in Egyptian, with the obligatory *il-* represented at top and the auxiliary *kida* at bottom.

Figure 4. Reference Hierarchy for Egyptian.

4.3. Kuwaiti

Kuwaiti Arabic (Figure 5) also shows the formal distinction between true definites marked with *il-* and unmarked indefinites, but also allows for regular auxiliary marking of AD nouns with an unstressed anaphoric demonstrative *ha-* (Brustad 2000, pp. 120–21), which accretes with the definite article. Brustad does not identify any Kuwaiti structures that would express meanings in her 'indefinite-specific' range (i.e., PSI and PNI), and I am likewise unable to locate any in her texts.
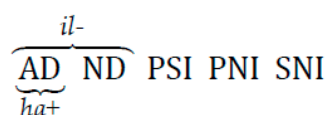
Figure 5. Reference Hierarchy for Kuwaiti.

4.4. Hassaniya

Hassaniya Arabic varieties are found across a wide expanse of western Africa; Cohen (1963) provides a description of the Hassaniya of southwestern Mauritania, Heath (2003) a collection of texts from further east in Mali, and Aguadé (1998) a brief overview of speech in southern Morocco. The latter shows features more similar to Moroccan (below), so I do not consider them here. More western varieties (Figure 6), including those in Mauritania and Gao, show a relatively simplex distribution of forms that looks much like Kuwaiti, i.e., an obligatory definite marker *il-* and auxiliary marking of AD referents with a demonstrative *ḏāk* or *ḏīk* (inflected for gender). Malian varieties around Gao (Figure 7), however, exhibit additional complexity in that they have a relatively frequent PSI marker *wāḥīd* (see Section 3.3). Heath (p. 8) asserts that "the grammar of Malian Hassaniya differs little from that of Mauritanian dialects," but the current framework does raise the question of whether grammatical marking of PSI nouns might be a useful metric for internal classification of Hassaniya.
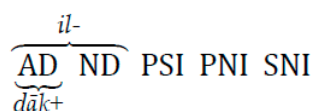
Figure 6. Reference Hierarchy for western Hassaniya varieties.
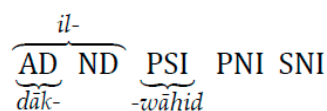
Figure 7. Reference Hierarchy for eastern Hassaniya varieties.

4.5. Sanaani

There is so much linguistic diversity in Yemen that I am hesitant to make broad pronouncements about "Yemeni," and thus base my judgements here only on Watson and 'Amri (2000) texts from Sanaa. In them, Sanaani (Figure 8) can be seen to obligatorily mark AD and ND statuses together with *il-*, like other varieties above, and also allows for auxiliary marking of AD with a preposed demonstrative *ḏayyik* (etc.).[9] In addition, Sanaani has an auxiliary strategy, described in Section 3.3, wherein PSI referents can be further differentiated with what is elsewhere a demonstrative adverb *hākaḏā(yā)*. This marker is similar in function to the Egyptian PSI marker *kida*.
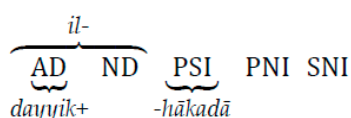
Figure 8. Reference Hierarchy for Sana'ani.

4.6. Levantine

Levantine varieties again show the pattern of marking AD and ND nouns with *il-*, and allow for additional delineation of AD nouns with an unstressed demonstrative *ha-*, but differ from varieties above in that they have a conventionalized article *ši* that denotes PNI referents (see Section 3.4). As discussed in Section 3.1, varieties of the Levant also make particularly productive use of anaphoric *ha-*, some perhaps to the extent that the resulting fused marker *hal-* should be considered its own, exclusive marker of AD statuses. Figure 9 gives a more conservative interpretation of the distribution of forms in Levantine, and Figure 10 offers the secondary analysis.
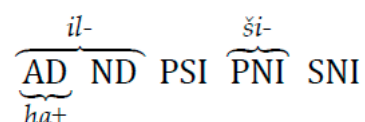
Figure 9. Reference Hierarchy for Levantine varieties.
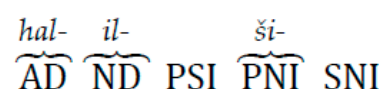
Figure 10. Possible reading for some Levantine varieties.

4.7. Iraqi

Arabic varieties in Iraq (Figure 11) have been described as having an indefinite *fard (Blanc 1964, p. 118), and Leitner and Procházka's (Forthcoming) focused semantic analysis supports the notion that this polyfunctional lexeme acts as a conventionalized PSI/PNI article in most Iraqi dialects (see Sections 3.3 and 3.4). Texts in Iraqi varieties also regular show the use of demonstrative *ha-* as an auxiliary AD marker alongside the oblitary definite marker *il-*, as is common elsewhere.
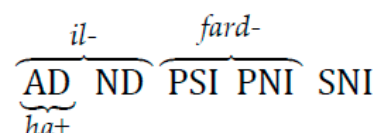
Figure 11. Reference Hierarchy for Iraqi varieties.

4.8. Najdi

The expression of definiteness in Najdi Arabic (Figure 12), as described in (Ingham 1994), somewhat parallels the formal distribution given for Iraqi above. For AD and ND nouns, *il-* is the obligatory article, with auxiliary marking of AD nouns possible with *ha-*. As a dialect that has so-called dialectical *tanwīn*, PSI and PNI nouns that are adnominally modified with adjectives, relative clauses, or prepositional phrases obligatorily have the marker *-in*. There is also evidence, described in Section 3.3, that at least some Najdi speakers can use DT on a purely semantic basis, i.e., without the noun being followed by any sort of modifier.
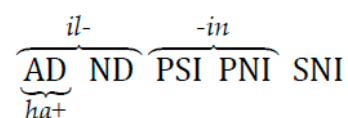
Figure 12. Reference Hierarchy for Najdi varieties.

4.9. Moroccan

Moroccan varieties (Figure 13) represent a relatively complex case, the main complications of which are that (1) the article *l-* is not restricted to definite (AD and ND) nouns

and (2) both PSI and PNI meanings are uniquely distinguished with overt, highly conventionalized articles. While the reflex of **al-* in all the above varieties is restricted and can thus truly be considered a definite article, in Moroccan it is conventionally extended to PSI referents (see Section 3.3) and is frequently used with SNI nouns as well (see Section 3.5). For PSI nouns, *l-* accretes with an article *wāḥad*, which is similar in function to the optional article found in eastern Hassaniya (Section 4.4); meanwhile, for PNI nouns, an article *ši*—identical in form and meaning to that attested in the Levant (Section 4.6)—is used. Moroccan also allows for auxiliary indication of AD nouns with the proximal and distal anaphoric demonstratives *hād-* and *dāk-*, the former of which is uninflected.
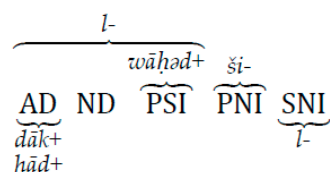
Figure 13. Reference Hierarchy for Moroccan varieties.

4.10. Central Asian

Central Asian varieties combine known strategies from elsewhere in Arabic with the unique feature of not having a reflex of **al-*; among others, this latter feature has probably played a role in these varieties being characterized as "metatypized" (Ratcliffe 2005), particularly given other nearby languages also lack true definite articles. There is evidence that Central Asian Arabic varieties can, like many others, use unstressed demonstratives for anaphoric (AD) reference (see 3.1). In addition, these dialects also show a reflex of **fard* that has the same PSI/PNI semantic scope of **fard* in Iraqi varieties (4.7). Central Asian also shows its own reflex of dialectical *tanwīn*, which sees the same syntactic conditioning as elsewhere (i.e., before adnominals), but has a wider semantic range because it can also occur with true definites.[10] It is not attested with SNI nouns, but considering these are unlikely to be adnominally modified in the first place (see Section 3.5), it would not be unreasonable to say that DT in Central Asian has fully lost its referential dimensions, and can be envisioned purely as a syntactic linker, hence the question mark in Figure 14.
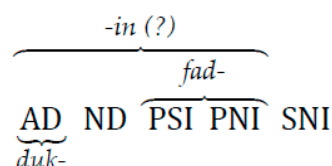

Figure 14. Reference Hierarchy for Central Asian varieties.

4.11. Nubi

Finally, while Wellens (2003), among others, has classified Nubi (Figure 15) as an Arabic-lexifier creole rather than a "true" Arabic variety, it is worthwhile to consider points of overlap with the above varieties in its expression of definiteness. Like Central Asian, Nubi has lost the article **al-*, differentiating it from the greater body of Arabic; nonetheless, also like Central Asian, the markers it does use have commonality with strategies attested in Arabic at large. The "definite article" *'de* that Wellens identifies is, in my reading, primarily an AD article, and shares semantic scope with the many other demonstrative forms that mark anaphoric definiteness in Arabic dialects. In addition, the apparently polysemic PSI/PNI article *'wai* has clear parallels with the postposted use of *wāḥid* in Hassaniya (Section 4.4).
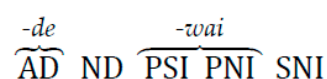

Figure 15. Reference Hierarchy for Nubi.

5. Definiteness and Classification

In theory, if the definiteness systems of Arabic dialects can be modeled, they should be relatively easy to classify. In practice, various complications arise that mean any attempt at classification will necessarily be subject to caveats and in need of ongoing refinement. As indicated more than once above, some of the systems themselves need more focused study to confirm how fully applicable the provisional models I have provided are to the dialect group as a whole. Scholars of Levantine Arabic, for example, face an open question as to just how close the unstressed anaphoric demonstrative complex *hal-* has come to acting as an obligatory article; similarly, scholars of Moroccan and Iraqi dialects may be able to further quantify uses of their respective indefinite articles in the same way by looking at them through a primarily semantic lens.

A related question is the concept of 'obligatory' vs. 'auxiliary', which I have attempted to frame here as a sort of continuum, the intermediate range of which might be described as 'conventionalized'. For the purpose of grouping and classification, it seems that obligatory articles—those that are required when a speaker wants to denote a particular referential meaning—should take priority, as they represent a sort of linguistic consensus on the part of the speaker community that is not present for other markers. Nonetheless, is not always immediately clear what 'obligatory' means. It seems unwise to treat it as an absolute notion that only a single contrasting token would disqualify, especially when diglossic practices allow speakers to switch between registers (and their respective definiteness systems) at will. Instead, it seems more reasonable to look at the preponderance of the evidence: what forms *most often* arise in everyday conversation between native speakers of the variety in question? I suggest that these highly conventionalized strategies should also be prioritized for the purposes of classification.

This is not to say, either, that less frequent auxiliary strategies have no value, else I would not have included them here. To the contrary, it does appear worthwhile to point out that a majority of Arabic varieties optionally use unstressed demonstratives for anaphoric definite meanings, and that both varieties that do not (such as Egyptian) and varieties that oblige them (such as some in the Levant) are the outliers. It does seem relevant to note that not just one, but at least two, Arabic varieties (Egyptian and Sanaani) show the same typological pattern of co-opting a demonstrative adverb as a marker of specific indefinites, even if these are not required or even all that frequently used, statistically speaking, to express that meaning. Most importantly, although these are synchronic patterns, all fully crystallized innovations were presumably in flux at one time, so for the historical record alone it is worth noting that such strategies exist.

With these qualifications in mind, then, we can approach the question of classification more directly. I propose that there are two primary methodologies for grouping dialects when looking at a set of interrelated semantic features, as is the case with definiteness. The first is a 'single-tier' approach, meaning we simply limit our view to a particular type of meaning within the Reference Hierarchy, survey the forms that are attested for it, and order them into groups. This approach is not particularly distinctive from the survey I provided in Section 3, and can be useful as a starting point for hypotheses, especially because it is suitable for identifying outliers. The Central Asian group, for example, clearly stands out in that it does *not* obligatorily mark definite (AD/ND) nouns (see Sections 3.1 and 3.2), and Moroccan clearly stands out in that it *can* mark full indefinite (SNI) nouns (see Section 3.5). Nonetheless, while this approach might be initially useful for looking beyond forms and toward semantic function—e.g., for noting that *ši* and **fard* have at least partial semantic overlap—it is not particularly useful for comparing systems as whole.

Instead, I offer that a preferable approach is to look at the distribution of forms holistically, in what might be called a 'multi-tier' approach. It is still necessary, of course, that we prioritize some features over others as a means of subgrouping, but as a general principle I hold that each primary subgroup should be selected to describe as many varieties as possible while whittling away the outliers. One possible schema, based on the

comparative systems given in Section 4 (minus Nubi), and taking into account the above points about obligatory and conventionalized forms, as is follows:

1. Dialects with strict formal distinction between true definites and indefinites . . .
    a. No highly conventionalized marking of indefinites . . .
        i.   No attested auxiliary strategies: **Libyan, Kuwaiti**
        ii. Attested auxiliary strategies: **Egyptian, Hassaniya, Sanaani**
    b. Highly conventionalized marking of some indefinites . . .
        i.   Marking syntactically determined: *tanwīn* **dialects; Najdi**
        ii. Marking semantically or pragmatically determined . . .
            1. Single marker for specific (PSI) and existential (SNI) indefinites: **Iraqi**
            2. Marker for existential (SNI) indefinites only: **Levantine**
2. Dialects with lax formal distinction between true definites and indefinites . . .
    a. Marked definites: **Moroccan**
    b. Unmarked definites: **Central Asian**

There are admittedly other ways in which this same set of metrics could be ordered, and the varieties in question consequently be grouped, but this one has a few advantages. The first is that the present classification does give some credence to traditionalist views of Arabic as having a normative system where **al-* is a "definite article," while leaving room for exceptions and, at the same time, expanding the profile of what a "normative" dialect is by showing that a majority of these do have at least some means of marking indefinite referents, a pattern that stretches from the Atlantic to the Gulf. A second advantage is that the classification serves to group together varieties that might not necessarily share features, but which do share basic semantic patterns, in turn opening the door for diachronic questions, especially when these varieties are geographically distant from each other. I do not mean to imply by this a hereunto undiscovered genetic relationship between Moroccan and Central Asian varieties, but I do mean to point out that both groups have seen the strict categorical distinction between definites and indefinites unravel, and they are both at the far ends of the Arabic-speaking world.

Interpreted this way, the definiteness data align most closely with a 'core-periphery' classification model, in that a strict formal distinction between definites and indefinites is maintained across a large, contiguous cultural area and frays only at its edges. Within the core area, there is frequent variation in the particular means of marking referential indefiniteness, and somewhat of a northern–southern split as one moves from unmarked or optional marking strategies of Egypt, Yemen, and the Gulf to the more conventionalized strategies of the Levant and Mesopotamia, but the strict and exclusive association of **al-* with definiteness goes unchallenged. Meanwhile, on the geographic fringes of this core, dialects break away typologically by either (1) extending **al-* to indefinite meanings or (2) detaching it from definite meanings.[11] The concept of peripheral dialects has been explored in volumes such as Owens (2000) and Anghelescu and Grigore (2007), and even though such varieties are just as often defined by what they are not than what they have in common, the addition of definiteness as a metric does at least support the idea of the 'core' against which they are defined as a viable linguistic entity.

Other classification proposals do not align as well with a scheme based on definiteness systems. The oft-proposed east–west division of dialects (see Palva 2006) is not easily evident here, especially given that the minimal expression of indefiniteness in Hassaniya varieties fall into the same general pattern as dialects much further east, including those of Egypt, Yemen, and Kuwait. The bedouin–sedentary division (again see Palva) is tenable only on the basis of the *tanwīn* feature, which is largely limited to bedouin-type varieties and is unique among indefinite markers in that it is conditioned by syntactic factors in addition to semantic ones. Nonetheless, in a purely typological sense, the presence of a conventionalized indefinite marker actually places DT-expressive bedouin varieties such as

Najdi closer to the indefinite-marking sedentary dialects of the Levant and Mesopotamia than it does to other bedouin varieties that lack it, such as western Hassaniya or Kuwaiti. Finally, one may consider whether, within the sedentary dialects, an urban-rural division is relevant; this too seems unlikely, given the systems found in a given geographic region *do* tend to be contiguous across urban and rural areas. The Levantine PNI article *ši*, for example, is used by speakers both in Beirut and small mountain villages in the same way that the Moroccan PSI article *wāḥǝd* is found both in the old cities and rural countryside.

In summary, the system-level configuration of definiteness marking does ultimately seem to be an areal pattern, and even minor differences between systems might consequently be useful for further subdividing clusters of geographically adjacent dialects. This possibility has already been raised for eastern vs. western Hassaniya (Section 4.4), as well as Levantine (Section 4.6) varieties. I also offer the observation that somewhere between central Algeria and Tunisia, dialects see an abrupt shift from complex, Moroccan-like systems (Section 4.9) to simplex, Libyan-like systems (Section 4.1). Precisely where these lines may lie—and why—is a question for future studies to address. Many of the systems in question seem to be the product of innovation, whether via semantic extension or leveling, and whether prompted by contact or otherwise. As it seems reasonable to expect that groups that innovate together, along the same timeline and to the exclusion of nearby groups, are indeed more likely to share history and social ties, further studies on definiteness and referentiality in spoken Arabic will be of value to the larger project of dialect classification.

6. Conclusions

In this paper I have outlined the process of building a novel classification scheme for Arabic dialects, using semantic typology as a metric for grouping rather than relying on the presence of forms alone. Taking definiteness as a case study, I discussed a selection of possible models, and adopted Dryer's (2014) 'Reference Hierarchy' as the most suitable of these for the task of envisioning definiteness systems in Arabic. I thereafter showed that, for expression of each semantic status along the Reference Hierarchy, the dialectological literature attests multiple strategies across the Arabic-speaking world. This variability can be made more useful for classification by modeling the semantic distribution of forms for discrete dialects holistically and then placing those models side by side, in turn allowing us to look past the forms themselves and instead class the dialects by shared typological characteristics. Key metrics that emerge are whether varieties maintain a strict formal delineation between true definites and indefinites, whether they overtly distinguish referential indefinites, and whether the latter is subject to syntactic conditions beyond the semantic ones. This particular classification approach does not align well with some traditional proposals, such as a east–west or bedouin–sedentary split, but it does lend some credence to the idea of a 'core' dialect area that contrasts with a 'periphery'.

Notes

[1] For the sake of simplicity, I use *al-* to refer to this article and all its various phonological realizations in the dialects. The same is true of other markers that have a common etymological source, such as *wāḥid* and *fard*. The precise shapes of their reflexes are not particularly relevant to a semantic analysis, and are already well documented.

[2] Readers should refer to the original sources, cited alongside the examples, for further context. As I draw some conclusions independent of those of the original authors, errors of interpretation are my own.

[3] In the present article I do not treat generics or plurals, although there is strong evidence for variation among them as well, with likely diachronic implications; see Turner (2018, pp. 232–35).

[4] Rosenhouse (1984, p. 82) notes the same pattern in Bedouin varieties of northern Israel, stating that "often this attachment is so strong that it seems to lose the demonstrative function and serve only for definition of the noun". To this I would add the caveat that it highlights *anaphoric* definition specifically.

[5] The lexeme *ši* itself is polyfunctional, as explored in detail in Wilmsen's (2014). Wilmsen (51–53) calls this particular use 'partitive *ši*' and notes its "indefinite determiner function as marking a quality somewhere between indefinite and definite," as is descriptive of PNI in the current framework.

[6] For example, *ʾəndək ši stīlu?* 'do you have some sort of pen?' is often used in the sense of 'do you have a pen [I can borrow]?' in Moroccan speech. Even though from the speaker's perspective there need be nothing particular about the 'pen' in question, allowing that there might be is a polite deferral to the listener. Leitner and Procházka's (Forthcoming) call this discursive strategy "mitigation" and locate it as a use of *fard* in Iraq and Khuzistan.

[7] Although using *al-* with singular SNI nouns is not possible in most varieties, a much greater number allow it with unquantified SNI plurals; see, for example, *s-sbūʿa* 'lions' in example (9). In this light, Moroccan might be seen as having simply leveled a more widespread plural paradigm to singulars.

[8] For this and the following models, I use a hyphen [-] to indicate the syntactic position of the marker in relation to the noun, and a plus sign [+] to indicate both the marker's syntactic position *and* that it accretes with other markers in the same semantic range. Like in Figure 2 (for English), forms given at top are either obligatory or highly conventionalized, whereas forms at bottom represent more marked auxiliary strategies.

[9] Demonstratives in Sanaʿani are highly variable (see Watson and ʿAmri 2000, p. 20); they appear to be used interchangeably in this sense, and are inflected for gender.

[10] For example, *duk parvardigōr-in ki lā-yi fi raḥim umm-i ḥāvī-ni* 'the protector who protected me in my mother's womb' (Ingham 2003, p. 36).

[11] Similar "fraying" of the definiteness system occurs in Arabic varieties of southern Iran (Matras and Shabibi 2007) and southern Turkey (Akkuş 2016), where unmarked definite head nouns are attested, albeit under different syntactic constraints. In Maltese, the strict association of *al-* with definites has been lost for adjectival attributes; see Fabri (2001).

References

Abu-Haidar, Farida. 1979. *A Study of the Spoken Arabic of Baskinta*. Leiden and London: E. J. Brill.

Aguadé, Jordi. 1998. Relatos en hassaniyya recogidos en Mhamid (valle del Dra, sur de Marruecos). *Estudios de dialectología norteafricana y andalusí* 3: 203–15.

Akkuş, Faruk. 2016. The Arabic Dialect of Mutki-Sason Areas. In *Arabic Varieties: Far and Wide: Proceedings of the 11th International Conference of AIDA, Bucharest, 2015*. Edited by Gheorghe Grigore and Gabriel Bițună. Bucharest: Editura Universității din București, pp. 29–40.

Anghelescu, Nadia, and George Grigore, eds. 2007. *Peripheral Arabic Varieties*. Romano-Arabica, VI–VII. Bucharest: Center for Arab Studies.

Behnstedt, Peter, and Manfred Woidich. 1988. *Die ägyptisch-arabischen Dialekte: Texte. Niltaldialekte, Oasendialekte*. Wiesbaden: Reichert, vol. 3.

Belyayeva, Dina. 1997. Definiteness Realization and Function in Palestinian Arabic. In *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics, Salt Lake City, 1996*. Edited by Mushira Eid and Robert R. Ratcliffe. Amsterdam: John Benjamins Publishing, vol. X, pp. 47–67.

Blanc, Haim. 1964. *Communal Dialects in Baghdad*. Cambridge: Harvard University Press.

Brustad, Kristen. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, DC: Georgetown University Press.

Caubet, Dominique. 1983. *La détermination en arabe marocain*. Paris: Université Paris 7, Dép. de recherches linguistiques, Laboratoire de linguistique formelle.

Caubet, Dominique. 1993. *L'arabe Marocain: Syntaxe et Catégories Grammaticales, Textes*. Paris/Louvain: Peeters.

Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In *Subject and Topic*. Edited by Charles N. Li. New York: Academic Press, pp. 159–82.

Cohen, David. 1963. *Le Dialecte Arabe Ḥassānīya de Mauritanie, Parler de La Gebla*. Paris: C. Klincksieck.

De Mulder, Walter, and Anne Carlier. 2011. The Grammaticalization of Definite Articles. In *The Oxford Handbook of Grammaticalization*. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press.

Denz, Adolf, and Dietz Otto Edzard. 1966. Iraq-Arabische Texte Nach Tonbandaufnahmen Aus al-Hilla, al-ᶜAfač Und al-Basra. *Zeitschrift Der Deutschen Morgenländischen Gesellschaft* 116: 60–96.

Dryer, Matthew. 2005a. Definite Articles. In *The World Atlas of Language Structures*. Edited by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie. Oxford: OUP Oxford, pp. 154–57.

Dryer, Matthew. 2005b. Indefinite Articles. In *The World Atlas of Language Structures*. Edited by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie. Oxford: OUP Oxford, pp. 158–61.

Dryer, Matthew. 2014. Competing Methods for Uncovering Linguistic Diversity: The Case of Definite and Indefinite Articles (Commentary on Davis, Gillon, and Matthewson). *Language* 90: e232–49. [CrossRef]

Elfitoury, Abubaker Abdalla. 1976. A Descriptive Grammar of Libyan Arabic. Unpublished Ph.D. thesis, Georgetown University, Washington, DC, USA.

Fabri, Ray. 2001. Definiteness Marking and the Structure of the NP in Maltese. *Verbum* 2: 153–72.

Fassi Fehri, Abdelkader. 2012. *Key Features and Parameters in Arabic Grammar*. Amsterdam: J. Benjamins Pub. Co.

Givón, Talmy. 1978. Definiteness and Referentiality. In *Universals of Human Language: Vol. 4, Syntax*. Edited by Joseph Harold Greenberg, Charles Albert Ferguson and Edith A. Moravcsik. Palo Alto: Stanford University Press, pp. 291–330.

Grand'Henry, Jacques. 2000. Deux Textes Arabes de Benghazi (Libye). *Oriente Moderno* 19: 47–57. [CrossRef]

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69: 274–307. [CrossRef]

Harrell, Richard S. 1966. *A Dictionary of Moroccan Arabic: Arabic-English*. The Richard Slade Harrell Arabic Series: No.9; Washington: Georgetown University Press.

Heath, Jeffrey. 2003. *Hassaniya Arabic (Mali): Poetic and Ethnographic Texts*. Leipzig: Otto Harrassowitz Verlag.

Heine, Bernd. 1997. *Cognitive Foundations of Grammar*, 1st ed. New York: Oxford University Press.

Hillelson, Sigmar. 1935. *Sudan Arabic Texts*. Cambridge: Cambridge University Press.

Ingham, Bruce. 1994. *Najdi Arabic: Central Arabian*. Amsterdam: John Benjamins Publishing Company.

Ingham, Bruce. 2003. Language Survival in Isolation: The Arabic Dialect of Afghanistan. In *AIDA Proceedings: Fifth International Conference of the Association Internationale de Dialectologie Arabe*. Cádiz: Universidad de Cádiz Publicationes.

Ionin, Tania. 2006. This Is Definitely Specific: Specificity and Definiteness in Article Systems. *Natural Language Semantics* 14: 175–234. [CrossRef]

Israel, Michael. 1999. Some and the Pragmatics of Indefinite Construal. *Proceedings of Berkeley Linguistics Society* 25: 121–32. [CrossRef]

Jastrow, Otto. 2005. Uzbekistan Arabic: A Language Created by Semitic-Iranian-Turkic Linguistic Convergence. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*. Edited by Éva Ágnes Csató, Bo Isaksson and Carina Jahani. Hove: Psychology Press, pp. 133–39.

Lahdo, Ablahad. 2009. *The Arabic Dialect of Tillo in the Region of Siirt (South-Eastern Turkey)*. Acta Universitatis Upsaliensis Studia Semitica Upsaliensia 26. Uppsala: Uppsala University, Department of African and Asian Languages.

Leitner, Bettina, and Stephan Procházka. Forthcoming. The polyfunctional lexeme /fard/ in the Arabic dialects of Iraq and Khuzestan: More than an indefinite article. *Brill's Journal of Afroasiatic Languages and Linguistics* 13.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Magidow, Alexander. 2016. Diachronic Dialect Classification with Demonstratives. *Al-Arabiyya* 49: 91–115.

Matras, Yaron, and Maryam Shabibi. 2007. Grammatical Borrowing in Khuzistani Arabic. In *Grammatical Borrowing in Cross Linguistic Perspective*. Berlin: Mouton de Gruyter.

McCarthy, Richard. J., and Faraj Raffouli. 1965. *Spoken Arabic of Baghdad, Part Two (A): Anthology of Texts*. Beirut: Librairie Orientale.

Mion, Giuliano. 2009. L'indétermination nominale dans les dialectes arabes. Une vue d'ensemble. In *Miscellanea Arabica 2009*. Edited by Angelo Arioli. Rome: Edizioni Nuova Cultura, pp. 215–31.

Owens, Jonathan. 1984. *A Short Reference Grammar of Eastern Libyan Arabic*. Wiesbaden: Otto Harrasowitz.

Owens, Jonathan. 2000. *Arabic as a Minority Language*. Berlin: Walter de Gruyter.

Palva, Heikki. 2006. Dialects: Classification. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden and Boston: Brill, vol. 1, pp. 604–13.

Ratcliffe, Robert. 2005. Bukhara Arabic: A Metatypized Dialect of Arabic in Central Asia. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*. Edited by Éva Ágnes Csató, Bo Isaksson and Carina Jahani. Hove: Psychology Press, pp. 141–51.

Rosenhouse, Judith. 1984. *Bedouin Arabic Dialects: General Problems and a Close Analysis of North Israel Bedouin Dialects*. Weisbaden: Otto Harrasowitz.

Seeger, Ulrich. 2013. Zum Verhältnis Der Zentralasiatischen Arabischen Dialekte Mit Einem Bisher Unveröffentlichten Text Aus Südchorasan. In *Nicht Nur Mit Engelszungen: Beiträge Zur Semitischen Dialektologie-Festschrift Für Werner Arnold Zum 60. Geburtstag*, 1st ed. Edited by Renaud Kuty, Ulrich Seeger and Shabo Talay. Wiesbaden: Otto Harrassowitz, pp. 313–22.

Stokes, Phillip W. 2020. A Fresh Analysis of the Origin and Diachronic Development of 'Dialectal Tanwīn' in Arabic. *Journal of the American Oriental Society* 140: 637–64. [CrossRef]

Turner, Michael. 2018. Definiteness in the Arabic Dialects. Ph.D. dissertation, The University of Texas at Austin, Texas, USA.

Vicente, Ángeles. 2000. *El Dialecto Árabe de Anjra (Norte de Marruecos)*. Zaragoza: Universidad de Zaragoza.

Watson, Janet C. E., and ʿAbd al-Salām ʿAmri. 1993. *A Syntax of Ṣanʿānī Arabic*. Leipzig: Otto Harrassowitz Verlag.

Watson, Janet C. E., and ʿAbd al-Salām ʿAmri. 2000. *Waṣf Ṣanʿā: Texts in Ṣanʿānī Arabic*. Wiesbaden: Harrassowitz.

Wellens, Inneke Hilda Werner. 2003. An Arabic Creole in Africa: The Nubi Language of Uganda. Ph.D. dissertation, Radboud University, Nijmegen, The Netherlands.

Wilmsen, David. 2014. *Arabic Indefinites, Interrogatives, and Negators: A Linguistic History of Western Dialects*. Oxford: Oxford University

Woidich, Manfred. 2006. *Das Kairenisch-Arabische. Eine Grammatik*. Leipzig: Otto Harrassowitz Verlag.

Yoda, Sumikazu. 2005. *The Arabic Dialect of the Jews in Tripoli (Libya): Grammar, Text and Glossary*. Leipzig: Otto Harrassowitz Verlag, vol. 35.