



Expressing Specialized Knowledge in Arabic: A Sociolinguistic Study of fuṣḥa and Moroccan dārija

Author

Zakariae Azennoud

Sultan Moulay Slimane University, Morocco

Email: zakariae.azennoud@gmail.com

Corresponding author

El Ghazi Batla

Ibn Tofail University, Morocco

Email: elghazi.batla@uit.ac.ma

DOI: <https://doi.org/10.36892/ijlls.v8i5.2733>

APA Citation: Azennoud, Z., & El Ghazi, B. (2026). Expressing Specialized Knowledge in Arabic: A Sociolinguistic Study of fuṣḥa and Moroccan dārija. *International Journal of Language and Literary Studies*, 8(5), 130-141. <https://doi.org/10.36892/ijlls.v8i5.2733>

Received:

14/06/2026

Accepted:

24/07/2026

Keywords:

fuṣḥa; dārija;
corpus
analysis;
specialized
knowledge;
lexical and
syntactic
variation

Abstract

Human knowledge is a stratified system of ideas about subjects and objects. Some strata entail common, practical knowledge, embedded in daily practices of social interaction and related to basic necessities such as food, clothes, and traditions; others are specialized, involving, inter alia, conceptual reflections on scientific notions, visions of future destinies, and positions on intellectual controversies. Given that language and its colloquialisms constitute the primary medium through which human knowledge is conveyed, it remains unresolved, based on mere intuition and limited evidence, whether standard language or colloquial is more instrumental in articulating the higher strata of specialized knowledge. This study investigates this question through the lens of Arabic, focusing on Standard Arabic (fuṣḥa) and the Moroccan colloquial variety (dārija). The methodological framework comprises two complementary techniques; first, a structured interview is conducted with 15 graduate students of Arabic Studies, who are asked to respond to a set of specialized questions related to language, history, physics, religion, artificial intelligence, and media initially using dārija, then fuṣḥa. The data, drawn from the recorded and written responses, are collected for comparative corpus analysis with particular focus on lexical and syntactic dimensions. The findings of the study reveal that although participants are native speakers of dārija, they find it significantly challenging to express specialized knowledge using a unique colloquial variety, as their lexical content is heavily influenced by fuṣḥa structures. While fuṣḥa demands more cognitive effort and time, it yields output that is lexically specific and dense. Also, in dārija function words and short-length structures are dominant, which allow the speech to expand at ease, and linguistic creativity is notable. The present study seeks to enrich the sociolinguistic discussion on the role of Standard Arabic and Arabic colloquial varieties in their speakers' cognitive and communicative practices, particularly in regard to the expressibility of specialized knowledge.

1. Introduction

Popular attitudes contend, intuitively, that Standard Arabic (SA, henceforth) is more capable of expressing specialized knowledge than Arabic local dialects, though the data advocating this standpoint is not conclusively decidable. Specialized knowledge (SK)¹ is a genre-specific domain that requires high-quality information, such as medicine, science, and

tourism (Bondi, Cacchiani & Cavalieri, 2019). This knowledge is often conveyed in a technical or expert language and is conducive to particular situations, contextual factors, and communicative purposes (Schneider, 2025). Technical language is viewed as very different in style and structure from that of everyday literature (Charrow, Crandall & Charrow, 1982); lay people, therefore, are not exposed to it on a regular basis (LaBoucane-Benson & Hossack, 2007). In line with Charrow and colleagues (1982), Anna Joan Casademont (2014) maintains that the transmission of specialized knowledge calls for specific syntactic roles and semantic references within the phraseological frame. Comparing SA to an Arabic vernacular, it has not been empirically investigated where exactly the focal features tied to SK exist in linguistic expression. In his *al-Luḡa aš-šā'ira* (The Poetic Language), Al-Aqqad states that Arabic *Fuṣḥa*² is among the most expressive tongues, because through Arabic only, its words and styles, one can visualize the minutiae of an Arab's life and environment (Al-Aqqad, 2014 [1960]). Other pro-Arabic scholars attempted to derive arguments from various classical sources to support this view (Al-Msaddi, 2014; Anis, 1951; Saleh, 2012; Samarai, 2021, among them). Considering that SA was itself a multiplicity of dialects spoken by the indigenous people of Arabia in the pre-standardization times (Al-Rajihi, 1996; van Putten, 2022),³ to what extent the view above holds valid in comparison to the potentiality of modern Arabic dialects is not beyond dispute.

Arabic, in general, reflects a complex language system, in which differences among labels associated with it in various situations, like *classical*, *neo*, *standard*, *modern*, *native*, as well as the typological varieties it includes, such as genre, style and register, create a perplexity. Fortunately, the distinctions developed by a number of scholars managed successfully to reduce the ambiguity surrounding the actual boundary lines between standard and colloquial forms of Arabic on the one hand, and its genres, styles, and registers on the other hand.⁴ Regarding the textual typological criteria, eminent scholarship (e.g. Biber, 1988; Lee, 2001; Nunan, 2008; Sinclair & Ball, 1996), for example, states that genre is recognized as a culturally and conventionally established group of texts, and can be discerned by external, non-linguistic features. In the case of Arabic, there may exist a religious genre (e.g. Fridays' sermons), literary (e.g. novels), journalese (e.g. newspapers), etc.; all of them belong, now sociolinguistically, to a highly-valued H (high) variety of Arabic which is learned in schools as versus the low L (low) one used in ordinary conversations (Bassiouny, 2006; Mejdell, 2006). In respect to expressing SK, it is supposed that the H features which are invited in speech, a variety that may also be called Educated Spoken Arabic.

In reality, what seems confusing and controversial is the nature of structural consequences that might be obtained from language contact (Trudgill, 2009). Significantly, the contact-induced change within a linguistic ecology, such as the one of Arabic, is very frequently associated with the question of expressive adequacy. In the literature, some interconnections between SA and Arabic dialects are pinpointed as a product of diglossia, the concept as was applied by Charles Ferguson (1959). The similarities that may be found among modern SA and a colloquial are identified in numerous roots that are believed to be derived from the classical Arabic language (Gheibi & Heidari, 2023; Tachicart, Bouzoubaa & Jaafar, 2016). Abdolahad Gheibi and Shahla Heidari (2023) maintain that SA is the output of education, while Arabic dialects are acquired naturally through family upbringing. Granted that, SA is likely to express higher levels of knowledge with genuine complexity. Al-e Hussein (2022) conducted a study to highlight the power of *as-salīqa*, a speaker's natural, intuitive command of language. Indeed, Arab speakers, owing to fluency, may find it much easier to express notions of the world using their own vernaculars, but it has not been examined whether such native ability proves proportionate conceptual accuracy when it comes to specialized expression.

Woidich (2018) found that Moroccan Arabic (MA), or *dārija*, in some situations, uses a short number of words to connote a variety of meanings by employing such styles as metaphor and euphemism. According to another researcher, creativity, not only at the semantic level, but in addition to advanced discursive facets, is a driving mechanism of dialectal evolution

(Boucherit, 2017), in that the necessity to clarify and signify a symbol stimulates the speaker to generate a new form or associate a new meaning to an old one. Creativity also appears in the writing style of MA where distinct contractions are dominant (Aguadé, 2006). Another motive for MA's inherent tendency for intricate expression is its writing legacy. Miller (2017) states several examples where MA is deemed of rich literary heritage, including theatrical plays, Zajal, monumental anthologies of Malhun, Aita, publicity, and so forth. In these texts, MA appears to be impacted by the lexicon, semantics, and syntax of SA. In this case, however, the interaction with SA is amply intentional as it serves dramatic, emotional, or poetic aspects of the text. Nonetheless, some words may be mistakenly ascribed to SA in the absence of decisive evidence; others may exhibit clear lineage and borrowing trajectory. Proximal feminine singular pronoun *hādi*, which is used currently in MA, was spoken several hundreds of years ago in the Najdi area of the Arabian Peninsula (van Putten, 2026); the final *'Imālah* used in some Quranic readings (e.g. the verb *tas'ā*) is prevalent in several Shamali dialects in Morocco (e.g. where they use to say *'ays'ā* as equivalent to he begs). The diglossic situation and considerable linguistic diversity whose causes are traced back by some authors to the pre-Islamic times, specifically in North and Central Arabia (Corriente, 1975; Al-Jallad, 2009, 2018; among others), has been an overshadowed melting pot, where the so-called Middle Arabic (a form used in the Abbasid caliphate) was also used. With the same token, features of the current dialects may have been preserved as remnants of those times. Bearing that in mind, there is a reasonable likelihood that dialects can produce original structural forms across various domains of knowledge.

Although MA has been inspired by SA in some respects, it exposes complex morphology and syntax. The morpho-syntactic interface characteristic of MA may also account for the reason addressed in the literature that Maghrebi dialects still lag behind in respect to Natural Language Processing (Harrat, Meftouh & Smaïli, 2018). Furthermore, MA incorporates elements from foreign languages like French in a process of code-switching and borrowing, a result that is attributed to the colonial context in which the European languages acted, in the case of Morocco, as the superimposed medium (Versteegh, 2009). For example, MA speakers use to take the verb *déclarer* in an adapted structure as *diklarīt/diklara/ydiklari*. Aabi (2000) found that Moroccan speakers borrow words, and sometimes phrases, from French for lexical, syntactic, and pragmatic purposes and evidently feel in the meanwhile that an equivalent to a borrowing is not available in their mother tongue (Aabi, 2000). Relevant to the scope of the present study, the domains of communication where speakers of MA resort to SA or foreign languages have not been specified by researchers.

Based on the findings at which the previous literature arrived, we hypothesize that SA is more adequately disposed to express SK than the colloquial forms of Arabic, including MA. The hypothesis will be corroborated on the ground of empirical investigation of attested language produced in natural usage. A minor postulate may also state that MA, given its vast and firm legacy, produces distinctive structures and functions when talking about that knowledge, which is, in turn, not a falsifying proposition. To pursue this exploratory examination, it is crucial to address the following questions: How large is the common ground between SA and MA when expressing SK? Do speakers of MA use a variety that is distinct from SA when they talk about SK? Are the parts of its composition readily attributable to the SA system demarcated by grammarians? Do the answers to these questions have any implications for the potentiality of expressive adequacy in respect to SK? To answer these questions and test the hypothesis above, we conduct a comparative analysis of frequency, lexical density, specificity, and distribution (Brezina, 2018), along with investigating closely the lexico-syntactic structures of SA and MA as drawn from language-in-use data.

2. Method

The method of the present study is carried out by two research instruments, an interview and corpus analysis. The interview was performed with 15 students who graduated from the

major of Arabic Studies at various Moroccan universities. It includes mainly seven open-ended questions specialized in language, history, physics, religion, artificial intelligence, and media, and answered in two phases (see Appendix). In the first phase, participants were asked to respond purely in MA. For practical reasons, as MA is the mother tongue of the participants, the interview was executed in an oral immediate way.⁵ In the second phase, after a week, they were asked to fill in a structured form whose questions are SA translation of the previous MA interview. This procedure is believed to enable a valid linguistic comparison. In this vein, it is worth noting that the method did not tend to measure the participants' possession of deep disciplinary knowledge in the abovementioned fields, but only to study their ability to verbalize the SK in the Arabic varieties.

MA recorded responses were transcribed into a uniform, ready-made input for corpus analysis, and the written SA answers were refined from typos. The two documents, which are now two subcorpora, were run in #LancsBox X 5.5.1., a natural language processing tool developed at Lancaster University (Brezina & Platt, 2025). Corpus analysis centered on the lexical component of the language, by running a wordlist of 1-grams and 2-grams, concordance, and collocations. Inferences were further complemented by in-depth scrutiny of contextualized syntactic patterns in the subcorpora of Arabic in order to arrive at well-founded corpus-informed conclusions.

3. Results

The first output of lexical analysis is lexical density, which is the proportion of unique lexical words (types) to the total number of words (tokens). In the analysis, the top occurring 2,000 words were taken from each subcorpus for assessment, because texts of varying lengths would give inaccurate results, whereas the relative frequency was normalized on the basis of 10,000 words as it is a representative value of the raw count of words in both subcorpora.

Lexical density is measured following this conventional calculation:⁶

$$\text{relative frequency} = \frac{\text{absolute frequency}}{\text{number of tokens in corpus}} \times \text{basis for normalization}$$

Lexical density for the SA subcorpus is thus calculated as follows:

$$\text{Types} \div \text{Extracted tokens} \times 10,000 = \text{Relative frequency of types per 10,000 words}$$

For better visualization, the relative frequency can be transformed into a percentage, or in what is technically called type-token ratio (TTR), as presented in the table below.

	Raw count	Extracted tokens	Types	Relative frequency of types per 10,000 words	TTR
SA	2,025	2,000	1,186	5,930	59.3%
MA	3,174	2,000	1,019	5,095	50.9%

Table 1. Lexical distribution of tokens and types across the SA and MA subcorpora.

Table 1 shows that TTR in the SA subcorpus is higher than MA, that is, the lexical use in the former is denser than the latter. The density of lexis also indicates that the words used are diverse and genre-specific. Still, 50.9% (that is, 5,095 typical word forms per 10,000) representing MA also displays lexical uniqueness.

Lexical uniqueness is attributed to many internal sources underlying the overall lexicosemantic composition of texts, such as the prominence of some keywords, dominance of particular collocations and fixed expressions, foregrounding of particular themes, and the use of specific writing styles. To explore the internal lexical distribution in the subcorpora, the analysis proceeds to the second result, the wordlist. The wordlist, presented in Table 2, provides an insight into the top 50 occurring words in the subcorpora.

Rank	Word (SA)	Romanized	Gloss	Fre.	Word (MA)	Romanized	Gloss	Fre.
1	من	<i>min</i>	from	61	في	<i>fī</i>	in, at, ..	76
2	في	<i>fī</i>	in, at, ..	56	لي	<i>lī</i>	for me	67
3	على	<i>'alā</i>	on, at, ..	33	من	<i>min</i>	from	62
4	عن	<i>'an</i>	about, on, ..	23	ما	<i>Mā</i>	that, not	49
5	إلى	<i>'ilā</i>	to, till, ..	21	على	<i>'lā</i>	on, at, ..	46
6	الفيزياء	<i>al-fīzyā'</i>	physics	18	واحد	<i>wāḥd</i>	one of	43

7	أو	'aw	or	16	هو	huwa	it, he, ..	41
8	التي	allatī	which, that, ..	15	اللغة	l-luḡa	language	28
9	أن	'an	that, to, ..	15	ديال	dyāl	of	26
10	الكون	al-kawn	universe	14	هي	hiya	it, she, ..	23
11	اللغة	al-luḡa	language	14	يعني	ya 'nī	means	23
12	لا	lā	not	13	التاريخ	at-tārīḥ	history	22
13	التاريخ	at-tārīḥ	history	12	الإنسان	al-'inṣān	human	22
14	فهم	fahm	understanding	12	لينا	līnā	for us	21
15	ينفع	yanfa'	benefit	11	مثلا	maḡalan	for example	21
16	الإنسان	al-'inṣān	human	11	أن	'an	that, to, ..	20
17	هي	hiya	she, it, ..	10	أنه	'annahu	that he/it	19
18	هذا	hādā	this	10	ولا	walā	and not	19
19	ما	mā	that, not, ..	9	هاذ	hād	this, these,	18
20	هو	huwa	he, it, ..	8	ممکن	mumkin	perhaps	18
21	إذا	'idā	if	8	الإعلام	al-'i'lām	media	18
22	الاصطناعي	al-'iṣṭinā'ī	artificial	8	ولكن	walākin	but	17
23	دراسة	dirāsa	study	7	لا	Lā	not	17
24	كل	kull	all, every, ..	7	الفيزيك	l-fīzīk	physics	16
25	بين	bayn	between	7	العالم	l-'ālam	world	15
26	لكن	lākin	But	7	الأشياء	al-'aṣyā'	things	15
27	إن	'in	if	7	الاصطناعي	al-'iṣṭinā'ī	artificial	15
28	أكثر	'akṭar	more	7	الفيزياء	al-fīzyā'	physics	14
29	يمكن	yumkin	perhaps	7	الذكاء	al-dakā'	intelligence	14
30	الذكاء	al-dakā'	intelligence	7	يكون	yakūn	be, become	14
31	العام	al-'āmm	general	7	فيه	Fīh	in it	14
32	مجموعة	maḡmū'a	a group of	6	يمكن	ymkn	perhaps	14
33	بها	bihā	in it	6	لغة	luḡa	language	13
34	ذلك	dālika	that	6	الدين	ad-dīn	religion	13
35	القرآن	al-qur'ān	the Qur'an	6	علاقة	'alāqa	relation	13
36	الرأي	al-ra'y	opinion	6	عندو	'andū	it/he owns	13
37	الناس	al-nās	people	6	باش	bāš	for	13
38	خلال	ḥilāl	during	6	الظواهر	l-ẓawāhir	phenomena	12
39	التجارب	al-taḡārib	experiments	5	ديالو	dyālū	of it/his	12
40	علم	'ilm	science	5	مجموعة	maḡmū'a	a group of	12
41	هناك	hunāka	there	5	الماضي	al-mādī	past	11
42	بعض	ba'ḍ	some	5	بيها	bihā	by/in it	11
43	و	wa	and	5	ز عما	za'mā	as if	10
44	العلمي	al-'ilmī	scientific	5	بالنسبة	bin-nisba	for	10
45	لذلك	liḍālika	for this	5	كيفاش	kīfāš	how	10
46	عليه	'alayhi	on it	5	خلال	ḥilāl	during	10
47	أداة	'adā	medium	5	إلى	'ilā	to, till, ..	10
48	استعماله	isti'māluh	its use	5	العام	al-'āmm	general	10
49	كذلك	kaḍālika	likewise	5	أو	'aw	or	10
50	الإعلام	al-'i'lām	media	5	تعلم	ta'allum	learning	9

Table 2. 50 top-ranking individual words (1-grams) in the SA and MA subcorpora.

Table 2 demonstrates that function words are dominant in MA with respect to general frequency. This finding aligns with the previous results in that fewer content words may determine low magnitude of lexical density in language use. In other words, the more content words a text uses, the denser it becomes. Another thing to notice in the result above is that around 39 out of 50 items tabulated in the MA column also exist in SA.

Most of the eleven remaining unique words share base morphology with SA; for example, the MA لينا is comparable to the SA لينا, بها, and كيفاش. In a phonetic practice, such words as those used in MA are transcribed in a different way (e.g. *līna* in MA versus *lanā* in SA), and though this linguistic aspect does not belong firmly to the scope of the present study, merging lexis and syntax, it highlights several relations among the varieties of Arabic and directs the analyst to historical accounts of language variation and evolution.

The frequency of 2-grams (a sequence of two consecutive words) holds ampler implication for interpretation in the latest regard, that is the similarities and differences between SA and MA with regard to SK expression. The results, as are laid out in Table 3, illustrate more notably the interactional space between SA and MA, relative to particular domains of SK, performance significantly indicative of interdependent tacit competence.

Rank	2-gram (SA)		Frequency	2-gram (MA)		Frequency
1	الاصطناعي	الذكاء	7	الاصطناعي	الذكاء	12
2	العام	الرأي	6	من	مجموعة	11
3	من	مجموعة	5	خلال	من	8

4	خلال	من	5	العام	الرأي	8
5	بها	يعبر	4	أجنبية	لغة	7
6	أغراضهم	عن	4	واحد	هو	6
7	هي	اللغة	4	هي	لي	6
8	من	والاستفادة	4	نظري	في	6
9	ينفع	الاصطناعي	4	هو	ما	6
10	طريق	عن	4	لغة	تعلم	5
11	يعبر	أصوات	3	الأشياء	من	5
12	كل	بها	3	واحد	هي	5
13	قوم	كل	3	واحد	في	5
14	عن	قوم	3	إيجابي	هو	5
15	على	التعرف	3	هو	التاريخ	4
16	هو	ما	3	العالم	تفسر	4
17	من	يمكننا	3	لينا	يمكن	4
18	الكريم	القرآن	3	الأمور	من	4
19	والطاقة	المادة	3	علاقة	كأين	4
20	ذلك	من	3	لي	الأشياء	4

Table 3. 20 top-ranking co-occurring words in the SA and MA subcorpora.⁷

Table 3 also grounds standard clusters of words used by speakers in response to specialized questions. Overall, the results reported above require qualitative discussion and close reading of substantial lexical and syntactic context where they have been framed, a task to be undertaken in the section that follows.

4. Discussion

The lexical analysis results described in the previous section have a number of implications for discussion. To start with the diversity of lexical items, it has been found that SA texts are denser than MA, with around 5,930 types per 10,000 tokens (Table 1). We can ascribe this trait to several reasons, among them education. This goes in line with Gheibi and Heidari (2023) who claimed that SA expressivity is constructed by formal instruction, hence Educated Spoken Arabic, which different from dialects whose adequacy lies in expressing daily and immediate needs. Lexical density indicates that words are not repeated in the language, and hereby specificity is characteristic of dense texts (Baker, 2004; Biber & Egbert, 2019; Jarvis, 2013). Therefore, the result obtained about density is an argument for the fact that SA lexically varies according to a specific genre of knowledge.

MA also encode relatively dense lexis, in that 50% of the tokens found in the MA subcorpus are unique words, but it does not demonstrate marked keyness; a large number of SA genre-specific words barely have distinctive equivalents in MA, such as *الاصطناعي* (*al-ʾiṣṭināʾī*), *التجارب* (*al-taḡārib*), and *الإعلام* (*al-ʾi lām*). It is the same reason why MA speakers would also resort to foreign language phrases as *intelligence artificielle*.

It is certain that although both varieties share these words, in form and meaning, with some originality leaning towards SA. Ismaili (2025) believes that this shared ground between SA and MA plausibly justifies the need to take into account the repository of the child's dialect and engage it in teaching SA and fostering its proficiency.

Lexical originality can be attributed to the sources from which speakers of Arabic varieties derive SK. Knowledge in SA is often built, in addition to formal education, from books, mainstream media, technical websites, and so on, while MA knowledge is taken from daily life conversation, informal oral journalese, social media, podcasts, etc. It is undoubtable that such trajectories as these have significant effect on the expression output of speakers since the former sources are highly prestigious and valued, and well-organized. Interdependency, especially on the part of MA sources, also is manifest, in that numerous sequences of SA-driven speech writing are traceable in the MA productions. Now that MA sources of knowledge have the charm of easy and casual language, speakers try to recycle the input they draw from them in their own way, or influenced by foreign-language mannerisms.

In this case, results show higher morphological creativity in the MA subcorpus. For example, MA speakers drop some long vowels and merge words in smooth cohesion, in such instances as in *f + waḥd* and *li + hiya* (see Table 3).

It was found that the MA texts are longer than the SA. Although the same questions were asked of participants, speakers felt the need to lengthen their response by a total of around 1,000 words. In this context, two findings came out in a cause-and-effect relationship. To elaborate, the heavy use of function words (Table 2) extends the speech, as some of these words, like relative and possessive pronouns, tend to add extra information about the antecedent subject or idea (Salim & Ahmad, 2015). Consider the example below:

Example (1)

اللغة نقدر ونعرفوها نقول أنها أحد الأدوات التي تمكننا من التواصل في المجتمع ديالنا، سواء بالكلام ويمكن لنا حتى هاذ اللغة تكون بالإشارة، لأننا يمكن لنا نعبرو بها في الحديث ديالنا اليومي، لكنشيرو بيدينا ببعض الأشياء، ممكن حتى الحاجب إلى غير ذلك، بمعنى كل هاذ الأشياء هاذي كدخل في اللغة، وهاذي شي كامل الغاية ديالو هو التواصل.

MA, being the mother tongue of the participants, brings a psychological effect on the production. The fact that speakers use MA naturally (Al-e Hussein, 2022) underpins a simplified terrain where they employ language freely and flexibly. The freedom of language use manifests very clearly in the phonetic character of MA data.

As shown above (Table 2 and Table 3), a plenty of words are extremely similar to the SA orthography. It is, nonetheless, remarkable how MA speakers diacritize words that resemble SA forms. For instance, the word *perhaps* is diacritized in SA as يُمكنُ, yet in MA three short vowels are dropped and replaced by *sukūn* (no-vowel symbol), pronounced thus as يُمَكُنْ. The fact supports some researchers' attitudes about the propensity of MA to be interpreted in a diversity of ways in accordance with a particular context of speech (Fannan, 2005; Woidich, 2018).

At the level of syntax, SA is often found as either verb-subject-object (VSO) language, with verbs located at the beginning of the clause, or subject-verb-object (SVO). In the latter formation, MA follows a comparable structure in that it uses a nominal sentence that starts with a noun rather than a pro-dropped verb. Compare examples (2) for MA and (3) for SA:

Example (2)

كاين علاقة بين الفيزيك والدين مثلا كنلقاو الفيزيك كيهتم بالمادة، حنا كنعرفو بعض الأشراف الدينية مثلا يقدر يفيدنا باش ندرسوهم جيولوجياً.. العلاقة لي كتجمع لنا الفيزيك بالدين هي أنه الفيزياء كتفكك الظواهر والدين كيجمع المعنى.

Example (3)

كلاهما ينتمي إلى مجال معرفي مختلف؛ فالدين يقوم على الإيمان والتجربة الروحية، ويعالج قضايا الغاية والمعنى، بينما الفيزياء تعتمد على المنهج العلمي القائم على الملاحظة والتجريب والنسبية.

Notably, not only do the examples above show that SA and MA follow the SVO formation, but the majority of sentences in both subcorpora are structured in this way. This reveals that this syntactic choice may be determined by the domains of SK, requiring the foregrounding of nouns.

Another point of similarity exists in the postposition of adjectives. Both MA and SA systematically place adjectives directly after nouns as illustrated in the examples below, respectively:

Example (4)

ديك التقاليد منتشرة.. الماضي ديالك وعلى الجدود ديالك.

Example (5)

لا يتم إلا عبر مجموعة من العمليات العقلية المجردة، والتمارين المرهقة.

It is possible that adjectives come before nouns in MA (e.g. mntāšra dik taqalīd), yet the postpositionality of adjectives may be a feature of SK expression, so is the case in SA.

In contrast, most implicit accusative adverbials used in MA are resemble SA. Consider the following MA example:

Example (6)

لأنه يمكن لنا مثلا ندويو على الفيزياء نسبياً في تفسير بعض الظواهر.

Example (6) includes two instances, *maṭalan* (مثلاً) and *nisbiyyan* (نسبياً), that are also used in SA. Extensive borrowing of formal structures amplifies an influential consequence of

diglossic situation, and speakers of MA are not always able to justify their heterogeneous grammatical choices. In this vein, SK context is one of the circumstantial determiners where the manifestations of diglossia rise to the surface.

Along with the above-discussed syntactic features shared by both SA and MA, there is the collocational structures that come in SK expression as a unity of form and meaning. Some of the collocations have been explored in Table 3, and they are underlaid by what may called sequential circulation of language use. That is to say, there are some structures that occur in specialized domains of knowledge as fixed expressions and the role of speakers is confined to reproducing these expressions verbatim. In the case of Arabic, this often results in transliteration and word-for-word usage mostly symmetrical on the lexical spectrum.

5. Conclusion

The present study used an empirical framework to explore and compare SA and MA potentiality in expressing SK. In this enterprise, SA demonstrates remarkable uniqueness at the level of lexis, while MA usage is processed by recurrence to several phenomena, including borrowing, code-switching, and innovative styles. SA also reveals originality of word choice relevant to genre-specific use, which justifies a dense and diverse quality of its texts, whereas MA relies on function words and elaborated language, which in turn encourages the length of its texts. On the other side, the linguistic variation leans towards a group of structural similarities, firmly connected with collocational sequences and syntactic conventions such as the position of subjects and adjectives in the sentence, whose etymology are hardly attributed to either variety. Overall, the findings of the present study have pertinent implications for such domains of critical discussion as language etymology, diglossia, second language learning theories, and communicative competence. However, the limitations still persist regarding the size of the corpus; its sample, which is produced by only 15 participants, is not highly representative of the Arabic Studies graduate community. The bias may also be located in the asymmetry of data collection, in that some differences in expressing SK can be attributed to language modality, spoken versus writing expression. Overall, the study generates falsifiable demonstrations and several questions remain open for further research; one example would be addressing to what extent the originality of language components (e.g., lexis, semantics, and syntax) contributes to its potentiality in expressing higher strata of human knowledge. To investigate this and other relevant questions, researchers are recommended to triangulate their research approaches and expand the scope of data collection.

References

- Aabi, M. (2000). *The syntax of Moroccan Arabic/French and Moroccan Arabic/Standard Arabic code switching* [Unpublished doctoral dissertation]. University of Sheffield.
- Aguadé, J. (2006). Writing dialect in Morocco. *Estudios de Dialectología Norteafricana y Andalusí*, 10, 253–274. <http://hdl.handle.net/10498/7405>
- Al-Aqqad, M. A. (2014 [1960]). *Al-luya aš-šā'ira* [The poetic language]. Hindawi.
- Al-e Hussein, A. M. (2022). Mental dictionary and standard forms between colloquial Arabic dialects and classical Arabic. *Buḥūth fī al-Luġha al-'Arabiyya*, 14(26), 153–170. <https://doi.org/10.22108/rall.2021.121380.1261>
- Al-Jallad, A. (2009). The polygenesis of the Neo-Arabic dialects. *Journal of Semitic Studies*, 54(2), 515–536. <https://doi.org/10.1093/jss/fgp011>
- Al-Jallad, A. (2018). The earliest stages of Arabic and its linguistic classification. In R. Bassiouney & E. Benmamoun (Eds.), *The Routledge handbook of Arabic linguistics* (pp. 315–330). Routledge.
- Al-Msaddi, A. (2014). *Al-huwiyya al-'arabiyya wa al-'amn al-luyawī: Dirāsah wa-tawṭīq* [The Arab identity and linguistic security: A study and documentation]. Arab Center for Research and Policy Studies.
- Al-Rajihī, A. (1996). *Allāhajāt al-'arabiyyah fī al-qirā'āt al-qur'āniyyah* [Arabic dialects in the Quranic readings]. Dār al-Ma'rifah al-Jāmi'iyyah.

- Al-Sharkawi, M. (2010). *The ecology of Arabic: A study of Arabicization*. Brill.
- Anis, I. (1951). *Min 'asrār al-luya [Secrets of language]*. Maktabat al-Anglū al-Miṣriyya.
- Baker, P. (2004). Querying keywords: Questions in difference, frequency, and sense in keyword analysis. *Journal of English Linguistics*, 32(4), 346–359. <https://doi.org/10.1177/0075424204269894>
- Bassiouny, R. (2006). *Functions of code switching in Egypt: Evidence from monologues*. Brill.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D., & Egbert, J. (2019). Incorporating text dispersion into keyness analyses. *Corpora*, 14(1), 77–104. <https://doi.org/10.3366/cor.2019.0162>
- Bondi, M., Cacchiani, S., & Cavalieri, S. (Eds.). (2019). *Communicating specialized knowledge: Old genres and new media*. Cambridge Scholars Publishing.
- Boucherit, A. (2017). L'expressivité comme moteur d'évolution d'un système verbal aspectuel: Le cas de dialectes arabes. *La Linguistique*, 53(2), 279–296.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Brezina, V., & Platt, W. (2025). *#LancsBox X [Software]*. Lancaster University. <http://lancsbox.lancs.ac.uk>
- Casademont, A. J. (2014). On the elements activating the transmission of specialized knowledge in verbs. *Terminology*, 20(1), 92–116. <https://doi.org/10.1075/term.20.1.05joa>
- Charrow, V., Crandall, J. A., & Charrow, R. (1982). Characteristics and functions of legal language. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage: Studies of language in restricted semantic domains* (pp. 175–190). Walter de Gruyter.
- Corriente, F. (1975). Marginalia on Arabic diglossia and evidence thereof in the Kitab al-Agani. *Journal of Semitic Studies*, 20(1), 38–61. <https://doi.org/10.1093/jss/XX.1.38>
- Fannan, A. (2005). *Min qaḍāyā al-lisāniyyāt al-'arabiyyah wa al-lahziyya al-mayribiyya: Lahzat ad-dār al-baydā' namūdajān [On Arabic linguistics and Moroccan dialectology: Casablanca dialect as a case study]*. Info-Print.
- Ferguson, C. A. (1959). Diglossia. *Word*, 15(2), 325–340. <https://doi.org/10.1080/00437956.1959.11659702>
- Gheibi, A., & Heidari, S. (2023). The phenomenon of diglossia between eloquent language and vernacular language in the novel of “Dilshād, sirat al-ju' wa as-shaba” by Bushra Khalfan based on Charles Ferguson's theory. *Buḥuth fī al-Lugha al-'Arabiyya*, 15(29), 103–120. <https://doi.org/10.22108/rall.2023.137974.1465>
- Harrat, S., Meftouh, K., & Smaïli, K. (2018). Maghrebi Arabic dialect processing: An overview. *Journal of the International Science and General Applications*, 1(1), 1–8.
- Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.
- Hussein, T. (1933). *Fī al-'adabī al-jāhilī [On the pre-Islamic literature]*. Dār al-Fārūq.
- Ismaili, Y. (2025). Al-inziyāḥ al-luyawiy li-l-muta'allim bayna al-faṣīḥah wa-d-dāriḥa fī al-mamlaka al-mayribiyya wa-'āliyyāt tawḍīfih ta'līmiyyan [Moroccan learners' linguistic deviation between Standard and colloquial Arabic and strategies of its educational use]. *Icesco*, 2(1), 285–312.
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(1), 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- LaBoucane-Benson, P., & Hossack, F. (2007). Understanding technical language: A literature review. *A Department of Native Counselling Services of Alberta*.
- Lee, D. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology*, 5(3), 37–72. <https://doi.org/10.64152/10125/44565>

- Mejdell, G. (2006). The use of colloquial in modern Egyptian literature—a survey. In L. Edzard & J. Retsö (Eds.), *Current issues in the analysis of Semitic grammar and lexicon II* (pp. 195–213). Harrassowitz Verlag.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory, and practice*. Cambridge University Press.
- Miller, C. (2017). Contemporary Dārīja writings in Morocco: Ideology and practices. In J. Høigilt & G. Mejdell (Eds.), *The politics of written language in the Arab world: Writing change* (pp. 90–115). Brill.
- Moscatti, S., Spitaler, A., Ullendroff, E., & von Soden, W. (1964). *An introduction to the comparative grammar of Semitic languages*. Harrassowitz.
- Nunan, D. (2008). Exploring genre and register in contemporary English. *English Today*, 24(2), 56–61. <https://doi.org/10.1017/S0266078408000217>
- Owens, J. (2006). *A linguistic history of Arabic*. Oxford University Press.
- Saleh, A. A. (2012). *Al-samā' al-luyawī al-'ilmī 'inda al-'arab wa-mafhūm al-faṣāḥa [Scientific linguistic listening among the Arabs and the concept of eloquence]*. Dar Mofem.
- Salim, A. A., & Ahmad, M. A. (2015). Semantic structure of relative clauses in English and Arabic: A contrastive study in Sudanese context. *European Academic Research*, 3(8), 9307–9345.
- Samarai, F. (2021). *Al-zumla al-'arabiyya wa al-ma'nā [Form and meaning of the Arabic sentence]*. Dar Ibn Katheer.
- Schmitt, N. (2010). *An introduction to applied linguistics*. Hodder Education.
- Schneider, F. (2025). Exploring the notion of specialized language from diverse perspectives. *Journal of Language Science and Practice*, 1(1), 14–19.
- Sinclair, J., & Ball, J. (1996). *Preliminary recommendations on text typology*. EAGLES (Expert Advisory Group on Language Engineering Standards). <http://www.ilc.cnr.it/EAGLES96/texttyp/texttyp.html>
- Tachicart, R., Bouzoubaa, K., & Jaafar, H. (2016). Lexical differences and similarities between Moroccan dialect and Arabic. In *2016 4th IEEE International Colloquium on Information Science and Technology* (pp. 331–337). IEEE.
- Trudgill, P. (2009). Contact, isolation, and complexity in Arabic. In E. Al-Wer & R. de Jong (Eds.), *Arabic dialectology: In honour of Clive Holes on the occasion of his sixtieth birthday* (pp. 173–185). Brill.
- Van Putten, M. (2022). *Quranic Arabic: From its Hijazi origins to its classical reading traditions*. Brill.
- Van Putten, M. (2026). When did Faṣīḥ become Qabīḥ? Rehabilitating Classical phonological and morphological features. *Journal of Semitic Studies*, 71(1), 201–242. <https://doi.org/10.1093/jss/fgaf033>
- Versteegh, K. (1997). *The Arabic language*. Columbia University Press.
- Versteegh, K. (2006). *Encyclopedia of Arabic language and linguistics*. Brill.
- Versteegh, K. (2009). Loan verbs in Arabic and the DO-construction. In E. Al-Wer & R. de Jong (Eds.), *Arabic dialectology: In honour of Clive Holes on the occasion of his sixtieth birthday* (pp. 187–200). Brill.
- Woidich, M. (2018). Lexical creativity in Arabic dialects: The case of “behind, buttocks.” *Wiener Zeitschrift für die Kunde des Morgenlandes*, 108, 241–264. <https://www.jstor.org/stable/26808019>

Appendix: Interview questions

The interview questions are listed in SA, followed by MA and English translations, respectively.

1. ما تعريفك للغة؟
كيفاش يمكن لك تعرّف اللغة؟

(What is your definition of language?)

2. هل يضيف تعلم اللغات الأجنبية شيئاً إلى مخزون الإنسان المعرفي؟
واش تعلم شي لغة أجنبية جديدة كايضيف شي حاجة للمخزون المعرفي ديال الإنسان؟

(Does learning foreign languages add anything to the learner's knowledge?)

إذا كان الجواب "نعم"، فما الإضافة؟
إلا كان الجواب "أه"، شنو بالضبط؟

(If yes, what benefits does it have?)

3. التاريخ هو تجربة خاضها الإنسان وانقضت، فما الفائدة من دراسته؟
التاريخ هو شي تجربة فاتت وماتت، شنو لفائدة من الدراسة ديالو؟

(Since history concerns events that have already passed, what is the benefit of studying it?)

4. كيف تستطيع الفيزياء تفسير ظواهر الكون والوجود؟
كيفاش physique تقدر تفسر العالم اللي كانتواجدو فيه؟

(How can physics account for the phenomena of the universe and existence?)

5. هل تعتقد أن للمعرفة الفيزيائية علاقة بفهم الدين؟
واش كايينة شي علاقة بين physique وفهم الدين؟

(Do you think that there is a relationship between physical knowledge and the understanding of religion?)

إذا كان الجواب "نعم"، فأين تتجلى هذه العلاقة؟
بحالاش؟

(If yes, where does this relationship lie?)

6. هل تظن (ين) أن الذكاء الاصطناعي ينفع المستعمل أم يضره؟ وكيف يحدث ذلك؟
الذكاء الاصطناعي واش كايمنع اللي كايستعملوا أولاً كايضرو؟

(Do you view artificial intelligence as a useful or harmful tool?)

كيفاش؟

(Elaborate?)

7. كيف ينور الإعلام الرأي العام؟
كيفاش يمكن للإعلام ينور الرأي العام؟

(How does the media enlighten public opinion?)

Notes

1. From now on, we use the notion specialized knowledge to refer to such language matter described as technical, expert, or educated.
2. We use SA in the present paper to refer to the form of Arabic that is accepted as an elevated, grammaticized, and official language. Although it is considered a modern variety, speakers of SA may use some classical structures.
3. Abduh Al-Rajihi (1996) and Marijn van Putten (2022) focus on the Arabic variation in the Quranic reading traditions. In this way, they track down the dialectal origins of abundant Arabic language usage.
4. One remarkable feature of SA discussed by scholars (Al-Sharkawi, 2010; Holes, 2004; Hussein, 1933; Owens, 2006; Moscati et al., 1964; Versteegh, 1997; Versteegh, 2006; among others) is the case system, the final short vowel endings indicating mood and case which determine the syntactic roles in the Arabic sentence.
5. Another motivation to execute the MA in the oral mode arises from several presumptions; typing in MA may cause a challenge to respondents, given the fact that MA has not an agreed-upon orthography among lay users and also would be a barrier to practical computer manipulation of the textual data. Moreover, having been provided papers, respondents would have referred extensively to SA, being it the medium of instruction they used at university. Procedurally, we bore in mind the asymmetrical issue of SA writing vs. MA speech and its implications for analysis. For this reason, few measures were taken, such as handing MA speakers a blueprint of the topics and questions to be asked, allowing them sufficient time to respond. On the other hand, speakers with an Arabic Studies major were chosen on the basis of their formal ability to express knowledge in the language with which they were academically trained.

6. For more details about lexical and statistical characterization of words in corpus analysis, see Brezina (2018), McEnery and Hardie (2011), Schmitt (2010), and Biber (1988).

7. Due to space limitations, we refrain from Romanizing the 2-grams as well as the Arabic excerpts presented in Section 4.

Authors' Bio

Zakariae Azenoud is a Lecturer in English at the Polydisciplinary Faculty of Khouribga, Sultan Moulay Slimane University, Morocco. His research interests include corpus linguistics, discourse analysis, semiotics, sociolinguistics, and foreign language acquisition.

El Ghazi Batla (corresponding author) is a Doctoral student specializing in Arabic applied linguistics, affiliated with the laboratory of Didactics, Languages, Media and Dramatology at Ibn Tofail University in Kenitra, Morocco.