

Giulia Guidotti

“L’Orientale” University of Naples

guidottigiulia@libero.it

On the possibility of applying the research methods of lexical availability studies to Middle Arabic and dialectal Arabic texts

Abstract

The aim of this article is to present and review the research methods employed in Laura Gago Gómez’s recent study, *Aproximación a la situación sociolingüística de Tánger-Arcila: variación léxica y grafemática* (2018). Moreover, this article intends to evaluate the possibility of applying the methods of Hispanic research on lexical availability to Middle Arabic texts, after considering what kind of Middle Arabic texts should be taken into account for this purpose.

Keywords

Arabic, Middle Arabic, Moroccan Arabic, lexical availability, lexical variation, graphematic variation.

Introduction

Studies on lexical availability have given interesting results since their birth in the 1950s. In an original manner, they do not focus on the most frequent words in speech and writing, but on the available lexicon, which is composed of words that are regarded as common even though they are not frequently used. This approach has also been applied to dialectology, mostly in the Hispanic context. However, in Arabic linguistics, lexical availability studies are still at an early stage. This is why Laura Gago Gómez’s work on the sociolinguistic situation of the area of Tangier-Asilah is very innovative. Gago Gómez employs the methods of lexical availability to see how social factors influence the available lexicon and the orthographic practices of Moroccan students. The elicited language

is *dārija*, that is Moroccan Arabic. This constitutes another original aspect of Gago Gómez's study. In fact, researchers generally either elicit Arabic dialects orally, or they investigate corpora of dialectal texts not specifically written for their investigation.

The purpose of the first part of this article is to present to a non-Spanish speaking audience the methods and results of Gago Gómez's recent study. This is then followed by suggestions for further research, in particular concerning the possibility of applying the methods of lexical availability studies to Middle Arabic texts. Firstly, some reflections are needed regarding what kind of Middle Arabic texts could be taken into account for this purpose. After this clarification, we will deal with the issue of applying Gago Gómez's methods to these written texts.

Although the present article does not go beyond theoretical considerations, it is hoped that these reflections will prove useful as a suggestion for further research in the fields of Arabic dialectology and sociolinguistics.

1. Laura Gago Gómez's, *Aproximación a la situación sociolingüística de Tánger-Arcila: variación léxica y grafemática* (2018)

Gago Gómez's book aims to study the current Moroccan sociolinguistic reality by focusing on the area of Tangier-Asilah. This area has not been properly considered by academic studies in recent years (Gago-Gómez 2018: 17), despite its interesting geolectal diversity (due to internal migration and rural depopulation) and strong influence of both French and Spanish on the local linguistic variety (caused by Tangier's turbulent history) (see Gago-Gómez 2018: 27–29).¹ In particular, her study analyses graphematic and lexical variation in the writing practices of young Moroccans. The informants are all secondary school students in their last year, given that Gago Gómez adopts the methods devised by the *Proyecto Panhispánico*, which considers pre-university students to represent the “pan-Hispanic educated standard” (Gago Gómez 2018: 39), highly representative of the pan-Hispanic population. Nevertheless, as Gago Gómez (2018: 40) is well aware, this is hardly true for the Moroccan context, since in 2009, only 29.1% of the population of the province of Tangier had a secondary degree. Because of this, according to Gago Gómez (2018: 40), “it is not possible to extend the results of this study to other social groups or to extrapolate them as a general characterization of the region”.² Both the urban

¹ On the Arabic dialect of Tangier, see mainly Marçais 1911, and Aguadé 2016 for its recent developments.

² Since one of the objectives of this article is to present the content of Gago Gómez's study to a non-Spanish speaking audience, I translate quotations from this book into English, along with other quotations, for consistency.

character of the sample, as well as the age and education of the informants, are expected to influence their lexical production, and thus the results of the study (Gago Gómez 2018: 41–42). The high schools selected for the study are situated in the centre and suburbs of Tangier, in the area between the centre and the suburbs, and in the town of Asilah (Gago Gómez 2018: 41). The position of the school is taken as a social variable along with sex, sociocultural level, first language, origin of the informant, and level of contact with the Spanish language (Gago Gómez 2018: 44–55). All of this information is provided by the students themselves through a questionnaire written in Standard Arabic, to be completed in the language of their choice before or after the test (Gago Gómez 2018: 44).

The methodological approach used by Laura Gago Gómez to consider the sociolinguistic dynamics of the province of Tangier-Asilah is lexical availability, which has developed particularly within Hispanic research. The study of available lexicon started in the 1950s from the issue of selecting the proper lexicon to be taught to learners of French as a second language. Choosing only frequent words was limiting, since many important and well-known words did not appear in frequency lists. As López Morales (2014: 2), pioneer of this discipline,³ explains, “[...] some words regarded as common, even usual were not actually frequent. This infrequency resulted from the fact that part of vocabulary, particularly nouns, was thematic; that is to say, their use was conditioned by the discourse theme. Only if the theme was favorable would certain words be realized in conversation”. Therefore, the available lexicon is essentially concrete, potential and “eminently psycholinguistic” (Gago Gómez 2018: 32–33). Informants are prompted to provide lists of available words through association tasks, starting from word cues called “centres of interest”. Following the norms of the *Proyecto Panhispanico*, Gago Gómez administers written tests⁴ in which students have to write the words that they associate with the given centres of interest with limited time. The order chosen by the informants must be taken into consideration for the analysis. In fact, this order provides information about the degree of availability of the words, since highly available words are more likely to appear first in lists of responses (López Morales 2014: 4).

The methods of lexical availability have been applied to several disciplines, such as dialectology,⁵ after their invention in the field of foreign language teaching. Yet, before Gago Gómez’s book, only once they had been used with a variety of Arabic, in a study by Muhammad Hasan Amara (1999) concerning Hebrew and English borrowings in Palestinian Arabic.⁶

³ See López Morales 1995–6 and 2014 on the history of lexical availability studies.

⁴ On the preference for written or oral tests, see Borrego Nieto-Fernández Juncal 2002 and Arnal Purroy 2008: 18.

⁵ See Borrego Nieto-Fernández Juncal 2002 and Arnal Purroy 2008 for example.

⁶ See below for a discussion of this study.

The first part of Gago Gómez's linguistic study concerns graphematic variation within the corpus of available lexicon. In discussing the relatively recent consideration of orthographic variability as a sociolinguistic variable, she follows Rutkowska and Rössler's article (2012) on orthographic variables. These can be correlated with extra-linguistic variables, since "the written level is subject to social factors in the same manner as other language levels" (Gago Gómez 2018: 61). Thus, Gago Gómez's analysis aims to examine the influence of social variables on the orthographic choices of the students participating in the study. In particular, quoting Aguadé (2006: 255), "when writing in dialect Moroccans have two opposite possibilities: either to preserve as much as possible the orthography of Classical Arabic or to innovate trying to represent the phonemes of the spoken language: the result is generally a fluctuation between both tendencies". In the first case, Moroccans follow the etymological principle, namely, they try to make their orthography of *dārija* words closer to the orthography of Standard Arabic, while, in the second case, the phonetic-phonological principle is applied (Gago Gómez 2018: 61). Actually, other options are possible. In fact, some students write their answers in both Arabic and Latin script, or alternate between the two scripts (Gago Gómez 2018: 78),⁷ but this interesting issue is not discussed by Gago Gómez, who focuses solely on Arabic script. It is also worth mentioning that the informants are not always consistent with their orthographic choices, because of the limited time available for completing the test. This represents one of the peculiarities of the corpus of this study, together with the particular lexicon that is elicited and together with the fact that the informants, despite their mid-high level of education, are neither writers nor journalists (Gago Gómez 2018: 66).⁸

Gago Gómez divides the resulting graphematic alternations into two variables: the *tafṣīḥ* variable and the dialectalising variable. Regarding the first one, she takes into consideration "conservatory graphemes": namely, *hamza*, *tā*⁹ and *alif al-wiqāya*.

The variables that I have called "conservatory" imply certain values that are connected with the Standard, since they represent phonemes that are absent from the variety of the students. Moreover, this consideration is supported by the fact that the writing of these graphemes requires an additional effort, not only because of their absence from the idiolect of the informant, but also because,

⁷ On the use of Latin script for writing Moroccan *dārija*, see Benítez Fernández 2003, Moscoso García 2009, Caubet 2012 and 2017. On biscriptality in general, see Bunčić-Lippert-Rabus 2016.

⁸ However, also other kinds of texts can show similar inconsistencies. See for example Hoogland 2013b: 70–71 on the *Code de la Route* and on the text of a theatre play.

⁹ Actually, fricative [d] and [t] can be found as allophones of /d/ and /t/ in some dialects of Northern Morocco (as the one of Chefchaouen) because of Berber influence. See Vicente 1999: 320–323 and Aguadé 2003: 67–70. Also Gago Gómez (2018: 82–83) mentions this peculiarity.

from a graphematic point of view, they constitute additions (though minimal): [additions] of graphetic strokes (Gago Gómez 2018: 70–71).¹⁰

With respect to the dialectalising variable, Gago Gómez mainly considers the alternation between the graphemes <ṭ> and <ḍ>, because of the typically Northern and Tangerine trait of the desonorization of /ḍ/ in /ṭ/, which is a strongly stigmatized feature (Gago Gómez 2018: 84–85). Therefore, shifts like substituting <ṭ> for <ḍ> point to a dialectalising trend, while shifts such as the hyper-corrected substitution of <ḍ> for <ḍ> point to a tendency towards levelling (Gago Gómez 2018: 88–91). The cases of substitution of <ṭ> for <ḏ>, of <ḍ> for <ḍ> and of <ḍ> for <ḏ> are excluded from the analysis, since some of them may be due to graphetic mistakes (only one diacritic or line is missing) and not because of a conscious choice (Gago Gómez 2018: 86).

After calculating the median of the *tafṣīḥ* variable and of the dialectalising variable in relation to social factors, Gago Gómez (2018: 96 *et passim*) uses statistical tests to see if the differences between informants who belong to different social categories are relevant.¹¹ The social factors that cause the differences with the greatest statistical significance for both the *tafṣīḥ* and the dialectalising variables are the origins of the informant and the position of the school. In fact, for both variables, a significant difference is found between students from Tangier and its surroundings, and students from Southern regions (Gago Gómez 2018: 100–105; 112–114; 116). Regarding the location of the school, which has a certain correlation with the origin of the informants because of the demographic dynamics of Tangier (Gago Gómez 2018: 54), informants studying in the city centre (where Tangerines are the majority) tend to use fewer traits of Standard Arabic, compared to informants studying in the semi-centre and in the suburbs (Gago Gómez 2018: 116). The native language does not give results with statistical significance because of the not sufficiently diversified sample (Gago Gómez 2018: 100; 112), while sex seems to constitute a non-relevant factor for the students' linguistic behaviour (Gago Gómez 2018: 96–97; 108). Concerning the socio-cultural level, the middle-way group appears to be the one that differs from the other two since it uses fewer dialectalizing as well as fewer normative traits, and it seems to be leading a linguistic shift towards the elision of the stigmatized desonorization of /ḍ/ (Gago Gómez 2018: 115; 219; 222).

The second part of the linguistic study considers lexical variability. Gago Gómez's aim is "to examine, on the one hand, the distribution of foreign words in the different centres of interest examining its correlation with the sociolinguistic and historical context of the community under consideration" and, on the other

¹⁰ The use of these graphemes in borrowings is taken into consideration as well (Gago Gómez 2018: 72–73).

¹¹ Numerical data and statistical tests are described in detail throughout the book.

hand, “to determine if there are differential uses, depending on the usual social factors” (Gago Gómez 2018: 220). The words that are taken into consideration are those borrowed from languages that have some function in the Moroccan sociolinguistic context (French, English, Berber, and Spanish), and only those whose integration into *dārija* is “uncertain or partial” (Gago Gómez 2018: 122–123). Thus, this excludes ancient borrowings from Romance or Latin as well as proper names and words that belong also to Modern Standard Arabic (Gago Gómez 2018: 123–124). In fact, Gago Gómez (2018: 121–122) intends to take into account only the foreign words that are *perceived* as such by the informants. They are divided into categories according to their likely etymology.¹²

In general, the results show a high presence of foreign words (18.26% of the corpus), above all in those centres of interest that are more susceptible to borrowing (such as “Parts and furniture of the house”, “Means of transport” and “Clothes”) (Gago Gómez 2018: 133; 220), and above all from Spanish, French, and English (in that order) – not surprisingly, considering the geolectal variety under study (Gago Gómez 2018: 135–136). Concerning the distribution of foreign words among different centres of interest, the highest percentages of Spanish words are found in “Food and drinks” and “Parts and furniture of the house”, while French is confirmed as being the key to the professional and academic world (i.e. “Professions and jobs” and “School: furniture and materials”) and English appears to be progressively entering the linguistic inventory of young Moroccans (i.e. “Games and entertainment”) (Gago Gómez 2018: 140–141; 147–150; 214). Regarding the correlation between the production of foreign words and social factors, the origin of the informant confirms to be an influential one. In fact, students from Southern regions use several French words and fewer Spanish words, compared to students from the area of Tangier (Gago Gómez 2018: 187–200). Moreover, paralleled with the results for graphematic variation, it is the students living in the centre of the city who use more Spanish words (Gago Gómez 2018: 200–212). To this point, Gago Gómez (2018: 208) underlines:

This symmetrical attitude in the use of these linguistic variables – normative traits and foreign words – is consistent in the context of a situation of social bilingualism: students take a stance in favour of one variety or the other (Classical Arabic or foreign languages, especially French), depending on the identity that they want to project. Likewise, this symmetry ought to be correlated with linguistic purism, in the sense that the informants who show a purist attitude will more likely avoid foreign words (or words that are perceived as foreign) and opt for normative forms.

¹² See Gago Gómez 2018: 125–128 on the difficulty of recognizing the etymology of many foreign words. For this reason, she inserts words that may be Spanish or French into a specific category called “Words from Romance languages”, while syntagms composed of words with different etymologies are considered among the “Hybrids”.

As opposed to the results for graphematic variation, sex is a significant factor for lexical variation. In fact, the use of Spanish and French words appears to be correlated to gender roles (women use more of these words in the centres of interest “Clothes”, “Parts and furniture of the house” and “The kitchen: furniture and tools”, while men stand out in “The land” and “Means of transport”) (Gago Gómez 2018: 152–161). Furthermore, women use more French words in general, compared to men (Gago Gómez 2018: 161; 215–216). Also in this case, the first language of the informants does not influence their use of foreign words (Gago Gómez 2018: 179–187), and not even their degree of contact with Spanish affects their employment of Spanish or other foreign words (Gago Gómez 2018: 171–179). As for the socio-cultural level, the only statistically significant difference lies in the use of French words. In fact, informants from the high and medium groups use more French words than those of the low group, probably because of the values of prestige and social advancement associated with this language (Gago Gómez 2018: 161–171). On the whole, the use of French words proves to be the linguistic variable that is influenced by the largest number of social factors: sex, socio-cultural level, origin of the informant, and the position of the school (Gago Gómez 2018: 215–216).

Gago Gómez’s book provides a new perspective on the writing practices of young Moroccans, the study of which is especially interesting because of the possible ongoing process of the emancipation of *dārīja* from Standard Arabic (Gago Gómez 2018: 218). Specifically, the originality of this study lies in the fact that the results of the study on orthographic and lexical variation are correlated with social factors, which indeed influence orthographic choices and the use of foreign words. In other words, looking at the matter from another point of view, the current sociolinguistic situation of the area of Tangier-Asilah is investigated by eliciting *dārīja* in writing, not orally, and thus this allows to focus on orthographic choices as well. Moreover, the elicited lexicon constitutes a peculiar branch of lexicon, which had not been properly considered yet in Arabic studies. A further enlargement and, above all, diversification of the sample will allow to gain information regarding matters that would need more data to be properly studied, such as the use of Berber borrowings (Gago Gómez 2018: 216) and the correlation between first language and graphematic variation (Gago Gómez 2018: 115–116).

2. On the possibility of applying these methods to Middle Arabic texts

2.1. What kind of Middle Arabic should be taken into account?

Some of the linguistic shifts described by Laura Gago Gómez in her book bring to mind similar features found in Middle Arabic texts. For instance, the hyper-corrected orthography of *hamza*, the use of *alif al-wiqāya* after proper

names ending with the sound *-u* or *-o*, and the confusion between interdental and alveolars are typical Middle Arabic features. At the same time, Gago Gómez's reflections are closely related to the observations of authors who are concerned with the orthographic conventions for writing *dārija* in Arabic script. In their turn, some of these texts in Moroccan Arabic show a mixture of Standard and dialectal features that remind us of Middle Arabic texts.

Before addressing the issue of the possibility to apply the methods described by Gago Gómez to Middle Arabic texts, we must ask ourselves: *what kind of Middle Arabic writings should be taken into account for this purpose?* This question is strictly connected to another wider issue: *may the above-mentioned writings in dārija be considered Middle Arabic texts?*

The definition of Middle Arabic, as given by Joshua Blau (1981b: 187) – who was the first scholar to investigate systematically Middle Arabic texts and to overcome the previous terminological ambiguity about this language variety – is “the mixed language of mediaeval texts, containing Standard Arabic, Neo-Arabic, and [...] pseudo-correct features”. However, as it is now universally acknowledged, Middle Arabic is not restricted to medieval texts, as these features are found also in contemporary texts. According to Ignacio Ferrando (2001: 147), “Middle Arabic is not circumscribed to a specific historical phase; instead, it existed, it exists, and presumably, it will exist throughout the entire life of the Arabic language”. Therefore, the current definition of Middle Arabic is rather broad:

Middle Arabic encompasses all the attested written layers of the language which can be defined as entirely belonging neither to Classical Arabic nor to colloquial Arabic, and as an intermediate, multiform variety, product of the interference of the two polar varieties on the continuum they bound, a variety that, for this very reason, has its own distinctive characteristics. Since the mixing is achieved to variable extents, one actually has to deal with a whole set of mixed varieties (Lentin 2008: 216).

In recent years, several studies have focused on written dialectal literature. However, only a few scholars, such as Kees Versteegh, have explicitly addressed the issue of the possibility of considering these writings as Middle Arabic texts. Versteegh mentions some official Dutch brochures addressed to the Moroccan minority, which were meant to be written in Moroccan Arabic, but are actually influenced by Standard Arabic phraseology and sentence structure:

It is certainly not current usage to call a text such as the one just quoted “Middle Arabic”. Yet there is an unmistakable similarity between these contemporary examples and the [medieval] texts [...]. The common denominator in all instances of mixed language and at all levels of written production is the centripetal force of the standard language. Whether authors deliberately use colloquial

features or simply fail to attain the level of grammatically correct speech, they always remain within the framework of the standard language (Versteegh 2014: 168).

For this reason, Versteegh rejects this kind of texts as evidence of “genuine” dialect literature. Instead, they represent “attempts to use vernacular elements while remaining within the sphere of influence of the standard language” (Versteegh 2014: 168).¹³

Other scholars deal with this issue less explicitly. For instance, Ángeles Vicente (2003: 174) affirms that “the importance and status of Classical Arabic have decisively influenced ALL THE WRITTEN MANIFESTATIONS of the distinct varieties of Arabic language, in such a manner that we can always find some traits that are due to the passage from spoken language to written language, namely, the use of classical forms alongside oral forms. This is why written sources do not show the ‘purest’ dialectal registers [...]” (emphasis added). Analogously, concerning contemporary writings, Høigilt and Mejdell (2017: 11) speak of “recent writing practices that revive the practice of writing in dialect, OR AT LEAST A VARIETY THAT CANNOT EASILY BE DEFINED AS *FUṢḤĀ*” (emphasis added). Moving onto the Moroccan context, Jan Hoogland (2013b: 61) offers his own definition of written *dārija* as “any written Arabic text (and for the purpose of this study, written in Arabic script) that is not fully according to MSA rules in terms of orthography, phonetics, morphology, syntax or lexicon, and that contains one or more dialect features belonging to Moroccan Arabic [...]”. Similarly, Ignacio Ferrando (2012: 355) and Leila Abu-Shams (2012: 295), in reference to the written *dārija* in novels and proverbs, respectively, mention authors who use a “compromising” *dārija*, thus avoiding terms perceived as too local and preferring dialectal words that are extant in Standard Arabic as well. All these descriptions of dialectal writings of mixed nature remind us of the mixed character of Middle Arabic, above all in its broader definition. Interestingly, Catherine Miller (2012: 431–432) states that the transcriptions of oral interviews in Moroccan newspapers remind her of the “phenomenon of oral alternations that we can hear nowadays on the radio”, namely, Mixed Arabic. Middle Arabic and Mixed Arabic (a written variety and an oral variety, respectively)¹⁴ are strictly connected to each other since they are “different manifestations of one and the same sociolinguistic phenomenon” (Den Heijer 2012: 3),¹⁵ namely, diglossia.

Nevertheless, an essential distinction must be drawn between texts that mix features of Standard and dialectal Arabic (regardless of whether their

¹³ Another interesting example offered by Versteegh (2014: 169) is the *Wikipedia Masry*, whose Egyptian Arabic is influenced by the Standard norms.

¹⁴ See for instance Den Heijer 2012: 8.

¹⁵ On the relationship between Middle and Mixed Arabic and on some of their common features, see Mejdell 2012.

language is called “written dialect” or “Middle Arabic” by their authors or by scholars) and those that show examples of code-switching, thus alternating words or entire sentences in Standard and dialectal Arabic. For instance, some journalists from the Moroccan newspaper, *Nichane* used to put *dārija* expressions in quotation marks (Miller 2012: 433 and 2017: 105), thus demarcating them clearly. Analogously, in Egyptian *Fuṣḥāmmiyya*, described by Gabriel Rosenbaum (2000: 77), “alternations [...] occur at the word, intra-sentential, and sentence levels, but not word-internally”. Besides, texts written in *Fuṣḥāmmiyya* do not contain hyper-corrections, which, according to Rosenbaum (2000: 77), prevents us from calling their language “Middle Arabic”. As a matter of fact, pseudo-corrections are mentioned in virtually all definitions of Middle Arabic provided by scholars,¹⁶ and could therefore be regarded as diagnostic features of this variety. Nevertheless, Versteegh, speaking of dialectal literature that actually adopts “a literary form of the dialect” (Versteegh 2014: 166),¹⁷ labels the language of writings which do not exhibit pseudo-corrections as “Middle Arabic”:

[...] literary writers have an intimate knowledge of Classical Arabic, and their use of colloquial language is always intentional. Thus, in their writings there are no examples of pseudo-correction due to lack of grammatical education. THIS KIND OF MIDDLE ARABIC is therefore much more akin to those Middle Arabic texts in which colloquial elements are used for the purpose of *couleur locale* (Versteegh 2014: 166; emphasis added).

Another issue is the use of Latin script for writing Arabic varieties. Nowadays, this practice is extremely common for writing Moroccan *dārija* on mobile phones and computers.¹⁸ One may wonder if dialectal texts that show similarities with Middle Arabic texts, as the ones mentioned above, may be labelled as “Middle Arabic writings” even though they use Latin script. In this respect, a significant precedent exists, which is Judaeo-Arabic, the first Middle Arabic variety to be studied (Blau 1981a). It is written in Hebrew script, yet this peculiarity does not prevent us from labelling it as “Middle Arabic”.

If we look at Middle Arabic as a continuum whose poles are dialectal Arabic and Standard Arabic, both Middle Arabic texts *stricto sensu* and dialectal writings will be positioned at some point on this imaginary line. However, differences exist between them regarding the direction of the imaginary arrow that would represent the author’s intention – namely, if his/her linguistic ideal is Standard Arabic or if his/her purpose is to write in dialect. In both cases, s/he does not achieve his/her goal, since s/he writes neither in “pure” Standard Arabic, nor in “pure”

¹⁶ See for instance Blau’s definition above.

¹⁷ See also Vicente’s opinion above.

¹⁸ Caubet (2012) interestingly comments on the different use of Arabic and Latin scripts depending on the support. See also the articles quoted in note 7.

dialectal Arabic (see Hoogland 2013a: 176). In the second case, as Versteegh underlines, pseudo-corrections are not found. Nevertheless, the incorporation of these writings among Middle Arabic texts appears reasonable, considering the (perhaps inevitable) influence of Standard Arabic, which seems to appear as soon as orality turns into writing, whether through a pen or through a keyboard.

Regardless how they are defined, dialectal writings are worth considering for our purposes along with Middle Arabic texts *stricto sensu*. In fact, it must be kept in mind that the language elicited in Gago Gómez's study is *dārija*, not Standard Arabic. Therefore, the informants' answers, in which the imaginary arrow is supposed to point to *dārija*, are related more closely to dialectal writings than to Middle Arabic ones.

2.2. Lexical variation

As Gago Gómez (2018: 218) indicates, her analysis of the use of foreign words concerns social bilingualism. However, upon closer inspection, it is more adequate to speak of transglossia, since Moroccan Arabic and foreign languages coexist in parallel with the diglottic continuum Moroccan Arabic-Standard Arabic.¹⁹

As mentioned previously, only Amara (1999) used the analysis of lexical availability with an Arabic variety before Gago Gómez's study. His results on the use of Hebrew borrowings in Palestinian Arabic, correlated with social factors including age, education, occupation, and contact with Jews, are highly interesting, and show some similarities with Gago Gómez's results. For example, the prestige of Hebrew, which is associated with work, modernity, and technology (Amara 1999: 93; 98), somehow resembles the role of French according to the Moroccan investigation. Analogously, the use of Hebrew words is correlated to gender roles (Amara 1999: 92–93) as in the Moroccan context. However, Amara's sample is much more diversified than Gago Gómez's one. Therefore, he can also consider demographic variables such as age, education, and occupation. His collection of data is based on structured oral interviews in which subjects are requested to answer open-ended questions about work, school, leisure time, the village, and the importance of electricity (Amara 1999: 94). This kind of method is criticized by López Morales (2014: 4), who, as mentioned before, emphasizes the importance of the order of appearance of available lexical units. Besides, by definition, available lexicon may not appear in spontaneous discourses, despite its availability, since its use is conditioned by the discourse theme (López Morales 2014: 2–3). Conversely, Gago Gómez admits that association tasks are somehow artificial, and that the names of the centres of interest considerably influence the resulting available lexicon. Therefore, "the low proportion of certain items

¹⁹ See, for instance, the use of "conservatory graphemes" in foreign words (Gago Gómez 2018: 72–73).

is not necessarily due to their poor functionality in the linguistic community under consideration, but could be due to the configuration of the test itself [...]” (Gago Gómez 2018: 120). Besides, taking into account an oral corpus, Amara (1999: 96–97) is able to examine also the use of Hebrew words according to different speech styles, which is not possible working with association tasks. With written Middle Arabic texts, a similar analysis could be performed by considering the proximity or the distance of the language of the text from Standard Arabic. In this perspective, different approaches may be adopted. First, considering Middle Arabic texts *stricto sensu*, one could focus on the proportion of dialectal words, following the model of lexical availability studies applied to Spanish dialectology.²⁰ Instead, taking into account texts written in dialectal Arabic (or that are supposed to be written in this variety), the proportion of classicisms may be investigated. Moreover, one could take into account the presence of foreign words – in particular for contexts such as the Moroccan one, where transglossia is especially evident. In order to obtain reliable results, the lexical availability method may be applied to texts, or text sections, concerning only one domain, along the lines of Amara’s interviews. However, a major difficulty arises. For the purposes of obtaining sociolinguistic information such as that found in Amara’s and Gago Gómez’s results, sociolinguistic data is needed. Regarding dialectal texts, this kind of data is often available,²¹ particularly in the case of professional writers,²² but the same does not hold true for Middle Arabic texts, especially older ones. Moreover, the different texts included in the corpus should be comparable, and therefore concern the same topic – which is to say, the same centre of interest. It is only if these conditions are fulfilled that the methods of lexical availability studies could be applied to Middle Arabic and dialectal Arabic texts.

2.3. Graphematic variation

Given the written nature of the collected data, Gago Gómez is able to investigate graphematic variation as well as lexical variation. As she indicates (2018: 218), the analysis of graphematic variation concerns the phenomenon of diglossia, since, from the graphematic point of view, the written answers of the informants are influenced by Standard Arabic even though an Arabic dialect is elicited.

Concerning diglossia, in recent years, a process of destandardization appears to be taking place in the Arabic-speaking world, where dialects are gradually

²⁰ See Borrego Nieto-Fernández Juncal 2002 and Arnal Purroy 2008 as examples.

²¹ For example, Caubet (2012) and Miller (2012) study individual differences in writing practices and mention the names of the individuals (such as journalists or rappers).

²² See for example Ferrando 2012.

entering written domains previously restricted to *fuṣṣḥā*.²³ This is particularly true for Morocco, where *dārija* has been increasingly used in newspapers²⁴, in keyboard-to-screen communication, and in social networks.²⁵ However, as Alexander Elinson (2017) underlines, this is not an entirely new phenomenon, since the practice of writing dialectal Arabic in Morocco is rooted in the *zajal* tradition.

Nonetheless, when prompted to write in *dārija*, Gago Gómez's informants are influenced by the orthographic conventions of Standard Arabic, as it happens in dialectal writings. In fact, as previously mentioned, students use the "conservatory graphemes" *hamza*, *tā'* and *alif al-wiqāya* in their answers, and these graphemes are also sometimes found in writings in *dārija*.²⁶ The "centripetal force of the standard language" (Versteegh 2014: 168) is exerted even on the orthographic conventions of dialectal texts written in Latin script: in fact, also in this kind of writings some occurrences of *hamza* (written as <2>) are found (see Moscoso García 2009: 218), and sometimes gemination is not graphically indicated, probably because of the influence of Arabic script (see Moscoso García 2009: 219 and Caubet 2012: 401).

From a graphematic perspective, the lists of words provided by the informants closely resemble Middle Arabic texts because of the presence of pseudo-corrections. In fact, these lists include Standard Arabic words in which the choice of the "seat" of *hamza* does not comply with Standard Arabic rules, as well as words in which an *hamza* is added though it is not present in the Standard Arabic word (Gago Gómez 2018: 79). Moreover, the letter *tā'* is also found in words that do not contain any dental fricative, in neither Moroccan nor in Standard Arabic (Gago Gómez 2018: 84). Similarly, the graphemes indicating voiced alveolar stops (<d> and <ḍ>) are sometimes substituted by the ones

²³ See Høigilt-Mejdell 2017. A recent article by Kindt and Kebede (2017) interestingly comments on the acceptance of Egyptian and Moroccan Arabic as written varieties, analysing the results of two surveys carried out in Cairo and Rabat in 2013 and 2015.

²⁴ See Benítez Fernández 2012 and Miller 2012 and 2017 on the language of the newspapers *TelQuel* and *Nichane* and on their orthographic conventions.

²⁵ See especially Benítez Fernández 2003, Moscoso García 2009 and Caubet 2017. The latter article is particularly worth mentioning since Caubet speaks of "a novel passage to literacy" (2017: 137), having witnessed "a qualitative change in the texts published on the internet: writing of long elaborate prose texts in *dārija*, passing from basic communication to literacy proper" (2017: 122). In this process of "non-institutional language planning" (Høigilt-Mejdell 2017: 13), it is interesting to note the coherent development of orthographic conventions. So, for instance, the ones that are referred to as rare or unusual by Moscoso García (2009), in reference to Internet forums and early *MSN* chats, are not mentioned in more recent studies regarding blogs and social networks.

²⁶ For example, the use of *hamza* in written *dārija* is mentioned by Hoogland (2013b: 63) and Caubet (2017: 125), the use of *tā'* by Aguadé (2006: 257–258) and Caubet (2017: 135) and the *alif al-wiqāya* by Aguadé (2006: 265), Hoogland (2013b: 64), and Miller (2017: 105).

indicating voiced dental fricatives (<ḍ> and <ẓ>) (Gago Gómez 2018: 90), which are non-existent in the informants' dialect.²⁷

Nevertheless, there are evident differences among Gago Gómez's corpus and Middle Arabic texts, not only in the scope of the act of writing, but also in its modalities. The informants are prompted to provide a large number of words in limited time and, as Gago Gómez (2018: 66) underlines, "the quickness [...] explains, at least partially, the informants' graphematic inconsistencies, which tend to be fewer, though not absent, in texts analysed in previous studies".²⁸ Instead, Middle Arabic texts are meant to be read by an audience and usually have literary content. Therefore, we can assume that more attention is devoted to their formal accuracy, compared to the words listed by the students. Nonetheless, also in Middle Arabic texts one finds frequent occurrences of omitted diacritics (as in Gago Gómez 2018: 86) and other deviations from Standard Arabic orthography apparently due to negligence or oversight.

These considerations do not impede the application to Middle Arabic texts of the methods employed by Gago Gómez to study graphematic variation in the students' lists of words. In comparison to the study of lexical variation, these methods for studying graphematic variation could be applied also to texts of different content and subject. However, one should consider the subject of the texts as a variable itself. For example, it could be speculated that pseudo-corrections may be greater in number in Middle Arabic texts concerning religion or politics than in texts of lighter content. Moreover, as mentioned previously with reference to lexical variation, sociolinguistic data is needed to study how social factors influence the use of "conservatory graphemes" and pseudo-corrections in Middle Arabic and dialectal texts. Such data is sometimes available for dialectal writings,²⁹ but not for Middle Arabic ones.

3. Conclusions

Gago Gómez's study on lexical and graphematic variation in a corpus of available lexicon is definitely worth consideration, since it deals with the sociolinguistic situation of the area of Tangier-Asilah by using research methods that had been previously applied to the Arabic-speaking world only in one study by Amara (1999). Furthermore, thanks to the written nature of her investigation, Gago Gómez also analyses the orthographic conventions adopted by the informants

²⁷ However, see note 9.

²⁸ Hoogland (2013b) provides some examples of inconsistencies within texts of a single author as regards the orthography of *dārīja*.

²⁹ See for instance Benítez Fernández 2003, in which some peculiar orthographic choices found in text messages in Latin script are correlated with the sociolinguistic background of the writers.

for writing their Arabic dialect, which is of particular interest for the Moroccan context, where *dārija* appears to be more and more accepted as a written variety.

In this article, we made some initial reflections on the possibility of applying these methods to dialectal Arabic and Middle Arabic texts (in the broader definition of the latter, that may also encompass the former). On the whole, both analyses appear to be applicable to these kinds of texts, but a number of limitations need to be considered. With respect to lexical variation, its application could lead to interesting results, such as those achieved by Amara regarding Hebrew borrowings in Palestinian Arabic. Nevertheless, the study of lexical variation in a corpus of texts that have not been produced for this specific purpose may be problematic, considering López Morales' remarks about the necessity of taking into consideration the ordering of the provided available words. Besides, one should take into account texts concerning the same subject, in order to compare the lexicon related to a certain domain. This limitation does not apply to graphematic variation. Therefore, the latter could be more readily studied within a corpus of Middle Arabic or dialectal Arabic texts. However, for both lexical and graphematic variation, sociolinguistic data about the authors would be needed in order to conduct a study such as Gago Gómez's one, which achieves results on the influence of social factors such as origin and socio-cultural level on both orthographic practices and presence of foreign words in the available lexicon. Sociolinguistic data is usually easily obtainable with respect to contemporary dialectal writings, but the same does not hold true for older Middle Arabic texts. The availability of such data could potentially lead to the desirable union of different fields of research. In fact, it would be possible to apply sociolinguistic analyses also to texts that have so far been analysed only from a strictly linguistic point of view (taking into account their orthography, phonetics, morphosyntax and lexicon), as is the case with Middle Arabic texts.

References

- Abu-Shams, Leila. 2012. Proverbios marroquíes: Su difusión por escrito y en internet. In Meouak, Sánchez & Vicente, 287–304.
- Aguadé, Jordi. 2003. Estudio descriptivo y comparativo de los fonemas del árabe dialectal marroquí. *Estudios de dialectología norteafricana y andalusí* 7. 59–109.
- Aguadé, Jordi. 2006. Writing dialect in Morocco. *Estudios de dialectología norteafricana y andalusí* 10. 253–74.
- Aguadé, Jordi. 2016. The Arabic dialect of Tangier across a century. *Arabic varieties: Far and wide: Proceedings of the 11th international conference of AIDA: Bucharest, 2015*, ed. by George Grigore and Gabriel Bițună, 21–27. Bucharest: Editura Universității din București.
- Amara, Muhammad Hasan. 1999. Hebrew and English borrowings in Palestinian Arabic in Israel: A sociolinguistic study in lexical integration and diffusion. *Language and society in the Middle East and North Africa: Studies in variation and identity*, ed. by Yasir Suleiman, 81–103. London–New York: RoutledgeCurzon.

- Arnal Purroy, María Luisa, 2008. Los dialectismos en el léxico disponible de los jóvenes aragoneses. *Estudios de disponibilidad léxica en los jóvenes aragoneses*, ed. by María Luisa Arnal Purroy, 17–49. Zaragoza: Institución Fernando el Católico.
- Benítez Fernández, Montserrat. 2003. Transcripción al árabe marroquí de mensajes de teléfono móvil. *Estudios de dialectología norteafricana y andalusí* 7. 153–63.
- Benítez Fernández, Montserrat. 2012. TelQuel: Una fuente contemporánea para el estudio del árabe marroquí. In Meouak, Sánchez & Vicente, 403–17.
- Blau, Joshua. 1981a. *The emergence and linguistic background of Judaeo-Arabic*. 2nd edition. Jerusalem: Ben-Zvi Institute for the Study of Jewish Communities in the East.
- Blau, Joshua. 1981b. The state of research in the field of the linguistic study of Middle Arabic. *Arabica* 28/2–3, *Numéro Spécial Double: Études de Linguistique Arabe*. 187–203.
- Borrego Nieto, Julio, and Fernández Juncal, Carmen. 2002. Léxico disponible: aplicaciones a los estudios dialectales. *Actas del IV congreso de Lingüística General (Cádiz del 3 al 6 de abril 2000)*, ed. by María Dolores Muñoz Núñez, II, 297–306. Cádiz: Universidad de Cádiz.
- Bunčić, Daniel; Lippert, Sandra L.; and Rabus, Achim (eds.) 2016. *Biscriptality: A sociolinguistic typology*. Heidelberg: Universitätsverlag Winter.
- Caubet, Dominique. 2012. Apparition massive de la *darja* à l'écrit à partir de 2008–2009: Sur le papier ou sur la toile : Quelle graphie? Quelles régularités? In Meouak, Sánchez & Vicente, 377–402.
- Caubet, Dominique. 2017. Morocco: An informal passage to literacy in *dārja* (Moroccan Arabic). In Høigilt & Mejdell, 116–41.
- Den Heijer, Johannes. 2012. Introduction: Middle and Mixed Arabic, a new trend in Arabic studies. In Zack & Schippers, 1–25.
- Elinson, Alexander. 2017. Writing oral and literary culture: The case of the contemporary Moroccan *zajal*. In Høigilt & Mejdell, 190–211.
- Ferrando, Ignacio. 2012. Apuntes sobre el uso del dialecto en la narrativa marroquí moderna. In Meouak, Sánchez & Vicente, 349–58.
- Gago Gómez, Laura. 2018. *Aproximación a la situación sociolingüística de Tánger-Arcila: Variación léxica y grafemática*. Zaragoza: Prensas de la universidad de Zaragoza.
- Høigilt, Jacob, and Mejdell, Gunvor (eds.) 2017. *The politics of written language in the Arab world: Writing change*. Leiden–Boston: Brill.
- Hoogland, Jan. 2013a. L'arabe marocaine, langue écrite. *Évolution des pratiques et représentations langagières dans le Maroc du XXI^e siècle*, ed. by Montserrat Benítez Fernández, Catherine Miller, Jan Jaap de Ruiter, Youssef Tamer, 175–87. Paris: l'Harmattan.
- Hoogland, Jan. 2013b. Towards a standardized orthography of Moroccan Arabic based on best practices and common ground among a selection of authors. *Árabe marroquí: De la oralidad a la enseñanza*, ed. by Paula Santillán Grimm, Luis Miguel Pérez Cañada and Francisco García Moscoso, 59–76. Cuenca: Ediciones de la universidad de Castilla-La Mancha.
- Kindt, Kristian Takvam, and Kebede, Tewodros Aragie. 2017. A language for the people? Quantitative indicators of written *dārja* and *'ammiyya* in Cairo and Rabat. In Høigilt & Mejdell, 18–40.
- Lentin, Jérôme. 2008. Middle Arabic. *Encyclopedia of Arabic Language and Linguistics*, ed. by Kees Versteegh, III, 215–24. Leiden–Boston: Brill.
- López Morales, Humberto. 1995–6. Los estudios de disponibilidad léxica: Pasado y presente. *Boletín de Filología de la Universidad de Chile* 35, *Homenaje a Rodolfo Oroz*. 245–259.
- López Morales, Humberto. 2014. Lexical availability studies. *Lexical availability in English and Spanish as a second language*, ed. by Rosa María Jiménez Catalán, 1–11. New York: Springer.
- Marçais, William. 1911. *Textes arabes de Tanger*. Paris: Ernest Leroux.

- Mejdell, Gunvor. 2012. Playing the same game? Notes on comparing spoken contemporary Mixed Arabic and (pre)Modern written Middle Arabic. In Zack & Schippers, 235–245.
- Meouak, Mohamed; Sánchez, Pablo; and Vicente, Ángeles (eds.) 2012. *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*. Zaragoza: Universidad de Zaragoza.
- Miller, Catherine. 2012. Observations concernant la présence de l'arabe marocaine dans la presse marocaine arabophone des années 2009–2010. In Meouak, Sánchez & Vicente, 419–440.
- Miller, Catherine. 2017. Contemporary *dārija* writings in Morocco: Ideology and practices. In Høigilt & Mejdell, 90–115.
- Moscoso García, Francisco. 2009. Comunidad lingüística marroquí en los foros y chats. Expresión escrita, ¿norma o anarquía? *Al-Andalus Maghreb: Estudios árabes e islámicos* 16. 209–26.
- Rosenbaum, Gabriel. 2000. “*Fuṣḥāmmiyya*”: Alternating style in Egyptian prose. *Zeitschrift für Arabische Linguistik* 38. 68–87.
- Rutkowska, Hanna, and Rössler, Paul. 2012. Orthographic Variables. *The handbook of historical sociolinguistics*, ed. by Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre, 213–36. Malden–Oxford: Wiley-Blackwell.
- Versteegh, Kees. 2014. *The Arabic Language*. 2nd edition. Edinburgh: Edinburgh University Press.
- Vicente, Ángeles. 1999. Los fonemas interdental en los dialectos árabes magrebíes. *Al-Andalus Maghreb: Estudios árabes e islámicos* 10. 317–33.
- Vicente, Ángeles. 2003. Fuentes para el estudio de los dialectos árabes. *Estudios de dialectología norteafricana y andalusí* 7. 173–195.
- Zack, Liesbeth, and Schippers, Arie (eds.) 2012. *Middle Arabic and Mixed Arabic: Diachrony and synchrony*. Leiden–Boston: Brill.