Lexique
ISSN : 2804-7397
Éditeur : Université de Lille

36 | 2025
Varia

The diffusion of contemporary Arabic borrowings beyond the French *banlieues*

La diffusion des emprunts arabes contemporains au-delà des 'banlieues' françaises

Marylise Rilliard

http://www.peren-revues.fr/lexique/1992

DOI : 10.54563/lexique.1992

Référence électronique
Marylise Rilliard, « The diffusion of contemporary Arabic borrowings beyond the French *banlieues* », *Lexique* [En ligne], 36 | 2025, mis en ligne le 30 juillet 2025, consulté le 17 septembre 2025. URL : http://www.peren-revues.fr/lexique/1992

The diffusion of contemporary Arabic borrowings beyond the French *banlieues*

La diffusion des emprunts arabes contemporains au-delà des 'banlieues' françaises

Marylise Rilliard

PLAN

NOTES DE LA RÉDACTION

Received: October 2024 / Accepted: March 2025
Published online: July 2025

NOTES DE L'AUTEUR

I would like to thank my Arabic-speaking informants for their valuable help in identifying the Arabic forms and their characteristics, as well as the two anonymous reviewers and director of *Lexique* for their constructive feedback on this manuscript.

TEXTE

1. Introduction

1 Post-World War II immigration to France from North Africa has increased situations of multiculturalism and language contact in French urban areas, as Arabic-speaking immigrants are now the biggest ethnic minority in France (Blanc-Chaléard, 2001; Hargreaves, 1995). Major waves of immigration from the Maghreb to France started between WW1 and WW2, motivated by a need for foreign labor to support the economic revival of France (Blanc-Chaléard, 2001). This immigration intensified after WW2, but after the first oil crash of 1973, the beginning of the economic crisis, and the end of the "Trente Glorieuses", immigration from North Africa to France slowed down significantly and became mostly driven by family reunification policies (Bruneaud, 2005). Despite this slow down, in 2023, direct immigrants from the three Maghreb countries (Morocco, Algeria, Tunisia) were still the biggest immigrant group in France (Insee, 2024), and communities of North African origin—later generations included—constitute the biggest non-European minority in the country: they represent 29.7% of first- and 1.5-generation immigrants and 33% of second-generation immigrants, with a total of about 4.5 million individuals, or 6.6% of the French population, without counting third and fourth generations for which data is missing (Insee, 2021, 2022)[1].

2 These immigrant communities were historically often housed in *grands ensembles* ("projects"), usually built at the periphery of big cities and now known as *banlieues* or *cités*, where North Africans constitute the largest ethnic group (Blanc-Chaléard, 2001). This

context increased situations of language contact between French and (dialectal) Arabic in metropolitan France, and as a result, language mixing practices, including the use of contemporary Arabic borrowings—such as *heps/hebs* "prison", *shitan/sheitan* "devil", and *seum* "rage, lit. venom"—have been attested in the setting of these multiethnic working-class urban neighborhoods since the 1980s (e.g., Billiez, 1992 and Trimaille, 2003 in Grenoble, Goudaillier, 2002 and Gadet, 2017 in Paris, Melliani, 2000 in Rouen, etc.). Language mixing practices between French and dialectal Arabic have also long been attested in the post-colonial Maghreb countries (Bentahila & Davies, 1983; M'barek & Sankoff, 1988), with both languages serving as a donor language in the case of borrowing (e.g., Morsly, 1995; Mzoughi, 2015). In France, despite the situation of language contact being more recent, Arabic forms have now travelled beyond the limits of the spaces where they emerged into colloquial metropolitan French regardless of the origin and social class of the speakers, especially finding their way in the language of young people [2]. The use of Arabic borrowings by adolescents has become so salient that the phenomenon is discussed in national newspapers such as *Le Figaro* and *Le Monde* (e.g., Mouchotte, 2022; Santolaria, 2021; Vinçotte, 2024). However, the presence of these forms has seldom been studied in populations beyond the youth from the *banlieues* and other multiethnic urban neighborhoods (see Hebblethwaite, 2018 for an exception), raising questions as to who these new users are, where they live, and what forms they use.

3 Consequently, this paper sets out to investigate the diffusion of contemporary Arabic borrowings in metropolitan French through an online survey assessing speakers' familiarity—defined as knowledge and self-reported usage—of 38 forms from Arabic identified during a previous ethnographic study in a high school in Orange, France. There is a lot of scientific debate concerning what constitutes a borrowing, especially as opposed to one-word insertional code-switching (see for instance Poplack & Dion, 2012, and Poplack, Sankoff, & Miller, 1988 on how borrowing and code-switching involve different processes and are two different phenomena vs Stammers & Deuchars, 2012 and Myers-Scotton, 2002 on how they are points on the same integration continuum or two instantiations of the same process), these debates resting mostly on the argument of the phon-

ological, and, most importantly, the grammatical integration of the foreign form in the recipient or base language. It is therefore paramount to define what is considered a borrowing in this study. Arabic borrowings are understood in this work as Arabic one-word insertions and prefabricated multiword expressions used in verbal routines that function like a single semantic unit (Kesckes, 2014), and which do not require bilingual proficiency (Bullock & Toribio, 2009). The forms concerned are thus lexical borrowings, more specifically loanwords (Haspelmath, 2009; Winford, 2003). Loanwords are often considered to consist only of *established borrowings* (Haspelmath, 2009; Myers-Scotton, 2002), but here *nonce borrowings*, i.e., momentary other-language items that appear in the spontaneous speech of bilingual speakers (Poplack et al., 1988), are also included. All these forms are grammatically integrated into French or at least do not show evidence of non-integration. Moreover, this study focuses specifically on *contemporary* Arabic borrowings, i.e., forms mostly imported in the contemporary period and reflecting the settlement of Arabic speakers from North Africa in France after WW2 and media attention on Arab countries (Fasla, 2008; Hebblethwaite, 2018). This new wave of Arabic borrowings into French follows the first wave in the Middle Ages and Renaissance and the second wave in the Modern period of the 19th and 20th centuries, driven by the French colonization of North Africa (Fasla, 2008; Melander, 1932). Some borrowings from this latter period that are still highly relevant nowadays in the speech practices of the multiethnic, working-class, urban youth (e.g., *bled*) were also included. Importantly, these forms are salient to French speakers: most of them are still recognized as coming from Arabic—or at least from a foreign language. They are thus particularly relevant for this sociolinguistic study, since their socio-indexical potential seems to be what drives their diffusion (see Section 2).

4 While most previous studies of contemporary Arabic borrowings in France have offered a micro-perspective on the phenomenon, looking at how they are used in interaction to perform relational and identity work, this study offers a macro-perspective focusing on the sociodemographic factors that may explain familiarity with Arabic borrowings, inspired by survey studies in dialectology and the variationist tradition.

2. Arabic borrowings and social meaning

2.1. Investigating variation and social meaning

5 This study is anchored in a variationist approach to language use, relying on ideas from both the first and third waves of variationism. In the first wave, born out of William Labov's fourth floor experiment in New York City (1966), variation was seen as systematic and linguistic variables as indexing macrosociological categories such as class, gender, and ethnicity. The goal of first-wave sociolinguistic studies was to reveal the patterns of social variation of a given variable, and the use of a variable was understood as a reflection or consequence of belonging to a given group. Thus, there was very little room for speaker agency. However, subsequent waves, starting with the second, and later the third, highlighted the role of more local and situated factors in explaining variation (Eckert, 2012). While earlier approaches to linguistic variation focused on how variables index social categories at the macro and local levels, the third wave places meaning—instead of variables—as the point of departure (Eckert, 2012). In the third wave, variables are viewed as stylistic resources to construct social meaning and position oneself in the social space. Therefore, the speaker is an active social agent who uses linguistic variation as a semiotic resource to express their concerns, and not simply index social identities and categories. Because these concerns are changing, the meaning of variables is not fixed, and variables are characterized by their « indexical mutability » (Eckert, 2012, p. 94).

6 As mentioned above, the use of Arabic borrowings in metropolitan French is no longer the prerogative of North African (children of) immigrants living in the French *banlieues*, and has seemingly transcended class, age, and ethnic boundaries. A macrosociodemographic approach to variation taken alone hence seems to be shortsighted in explaining the diffusion of these lexical forms. However, the socio-indexical potential of these forms from a third-wave perspective seems to be better suited to explain their diffusion, as will be elabor-

ated further below. This study investigates the diffusion of contemporary Arabic borrowings in metropolitan French according to primarily macro-social characteristics of the survey respondents in a first-wave fashion, but understands these patterns of variation as only indirectly related to broad social categories « through [the] association [of linguistic variables] with qualities and stances that enter into the construction of categories » (Eckert, 2008, p. 455).

7 Although this study is situated in a variationist framework at the theoretical level, it does not rely on the traditional data collection methods of variationist studies (e.g., sociolinguistic interviews, participant observation), as these are not very well suited for a quantitative investigation of the phenomenon under study. Indeed, Arabic borrowings occur at low frequency in spontaneous speech (see Section 3)—as is often the case with lexical phenomena—so doing a quantitative, large-scale study based on production data is hard to implement at the methodological level. Sociolinguistic surveys allow researchers to easily collect large amounts of data, therefore obtaining a sample that is more representative of the whole population and can be submitted to statistical analysis, providing robust results that make it possible to draw conclusions about the population at large (Avanzi et al., 2016; Boberg, 2018; Dollinger, 2012). Though, sociolinguistic surveys present two major drawbacks: by inquiring directly about language, they say more about the social evaluation of the phenomenon under study than the phenomenon itself (see Labov's (1972) *observer's paradox*), and people's linguistic intuitions are often not good indicators of how they speak (Boberg, 2018; Dollinger, 2012). However, for several variables, production data and self-reporting data have been shown to be largely congruent (Dollinger, 2012), and some argue that surveys are most appropriate to investigate lexical variables, because lexical variation would be often less affected by language ideologies than grammatical or phonetic variation (Boberg, 2018). They have been successfully used to investigate lexical regionalisms in dialectology (e.g., Avanzi et al., 2016). Moreover, Arabic borrowings enjoying *covert prestige* (Trudgill, 1972), as we will see below, it was believed it could counterbalance the underreporting that can occur when investigating a socially marked variable or socially driven change in progress like Arabic borrowings in French (Dollinger, 2012; Glickman et al., 2018). Therefore, a ques-

tionnaire was deemed an appropriate tool to evaluate the diffusion of contemporary Arabic borrowings in metropolitan French.

2.2. Contemporary Arabic borrowings in metropolitan French

8 Despite being reported in virtually every corpus of French Contemporary Urban Vernaculars (henceforth CUV) (Rampton, 2011), contemporary Arabic borrowings have received little attention from a variationist perspective in metropolitan French, likely due to their scarcity in existing corpora. Some of the best descriptions of Arabic borrowings come from ethnographic works looking at interactional patterns and identity-building strategies among multiethnic youths (e.g., Billiez, 1985, 1992; Melliani, 2000; Trimaille, 2003; Tetreault, 2015) and rap studies (e.g., Hassa, 2010; Paine, 2012; Verbeke, 2017). To the researcher's knowledge, there are no studies that have quantitatively recorded borrowed Arabic occurrences in natural speech data.

9 Although such quantitative data is missing, eight weeks of ethnographic work in a high school in Orange resulted in the identification of only 38 different Arabic borrowing types, most of them being hapax. This speaks to the low frequency of occurrence of these forms. Studies on rap corpora confirm this empirical observation: they report very low rates of Arabic borrowings in rap music (~ 0.25-0.45%) (Rilliard & Adams, 2018; Verbeke, 2017)—in comparison, *verlan* is used almost three times as much in Paris-based rap (Verbeke, 2017). Arabic borrowings are likely even rarer in natural speech, since rap lyrics are « a hybrid of improvised speech and carefully written poetry » (Paine, 2012, p. 48) that serve as a display of a certain culture and language, as they are used by artists—primarily from the *banlieues*—to build identities and appeal to a specific audience (Hassa, 2010; Verbeke, 2017). Nevertheless, the results on speech data and rap music appear to be convergent as to who their greatest users are: their use or knowledge has been mainly associated with (young) male speakers of North African origin (e.g., Billiez, 1985; Hassa, 2010; Paine, 2012; Trimaille, 2003; Pooley & Mostefai-Hampshire, 2012; Verbeke, 2017). Rap studies also suggest rappers from Marseille are greater users of these forms, potentially due to demographic factors (Paine, 2012; Verbeke, 2017).

10 However, the use of contemporary Arabic borrowings goes beyond a claim to national or ethnic identification, as they are also used by speakers of other origins in what Rampton (1995) calls instances of *crossing*. They can be used to build a Muslim identity, or mark belonging to the (multiethnic) peer group, for instance (Billiez, 1985, 1992; Guerin, 2018; Hassa, 2010; Hebblethwaite, 2018; Melliani, 2000; Pooley, 2018; Pooley & Mostefai-Hampshire, 2012; Verbeke, 2017). Because of their association with *banlieues* or *cités*, contemporary Arabic borrowings also index affiliation with a socially stigmatized and underprivileged lower-class group, often with a migrant background (Hambye & Gadet, 2014), and a rebellious identity linked to life in the *banlieues* (Hassa, 2010; Melliani, 2000; Verbeke, 2017). This is reflected in the innovative lexicon of French CUV, which mostly pertains to sex, drugs, violence, inter-racial relations, negative emotions, etc., all themes that reflect the lives of young people in these disenfranchised spaces (Gadet, 2017; Hassa, 2010; Méla, 1997; Rilliard & Adams, 2018; Sloutsky & Black, 2008).

11 It is precisely this rejection of the 'legitimate culture' that Arabic borrowings and other features of French CUV index that makes them popular among young speakers of all social classes, adolescence being marked by the rejection of social and linguistic norms (Jamin, 2005; Méla, 1991; Trimaille, 2004). These forms are thus spreading through *covert prestige* (Trudgill, 1972). Rap music, a cultural artefact stemming out of the *banlieues*, has been one of the main factors driving the diffusion of features of French CUV beyond those spaces, diffusion now also enhanced by the widespread use of social media, especially among younger generations (Gasquet-Cyrus, 2013; Marchessou, 2018).

12 Yet not much—if anything—is known about the use of contemporary Arabic borrowings outside the *banlieues* and similar neighborhoods. Hebblethwaite (2018) investigated Parisians' degree of awareness of the Islamic lexical field using a sociolinguistic survey. Beyond the fact that Arabic speakers, unsurprisingly, were more familiar with these forms, he also found that the younger generation (below age 30) had a notably better knowledge of these terms than older speakers, supporting the idea that the use of Arabic borrowings may be an age-related phenomenon outside the *banlieues*, too. Still, the 'young' speakers in Hebblethwaite's (2018) study were on average older (up to

30 years old) than in most previous studies of contemporary Arabic borrowings, hinting at the spread of these forms beyond youth speech. Nevertheless, limitations of this study restrict our understanding of the phenomenon at a larger scale: it investigated awareness rather than usage of the forms and focused on the Islamic lexical field and speakers from Paris only. The questions concerning which forms have diffused beyond multiethnic working-class urban neighborhoods, who uses them and where thus remain open.

3. Methodology

13 An online survey was created to answer the above questions regarding the diffusion of contemporary Arabic borrowings in metropolitan French. It assessed respondents' familiarity with 38 Arabic forms identified during a previous phase of ethnographic fieldwork. This section begins by outlining the forms assessed in the survey, explaining how they were identified and describing their features. It then provides details about the online survey and its participants, followed by an explanation of the data analysis process.

3.1. Borrowed forms tested

14 The 38 forms assessed by the participants in the survey were mostly identified in spontaneous speech, but also self-reported, during participant observation, in a recorded video task and interview and in a self-reflexive essay. These data were collected over the course of an ethnographic fieldwork of eight weeks in a high school in the city of Orange in southern France during the spring of 2021.

15 These borrowed forms reflect overall and language specific trends in borrowing. The grammatical category and semantic field of the types are given in Tables 1 and 2 respectively. These categorizations are based on how the word is used in French, which can reflect semantic extensions or changes in grammatical category from its original use in Arabic. Since a lot of users of these forms have no or little knowledge of Arabic nowadays, it is believed that their meaning—and by extension their grammatical category—in French is more relevant to their diffusion than their original meaning in Arabic. Some types belong to two different semantic fields and were therefore classified

twice (e.g., *hamdoullah* categorized as Politeness as well as Religion and beliefs), hence the count of types above 38. Nouns are most represented, closely followed by interjections, expressions or idiomatic chunks, a lot of the latter being politeness or religious formula. The semantic fields identified overall closely reflect what has been found in previous work on contemporary Arabic borrowings in metropolitan French, whether in natural speech or rap music.

Grammatical category	Count of type	Percentage	Example	Meaning
Noun	16	42.1%	*hallouf*	pig
Interjection/Expression/Chunk	13	34.2%	*starfoullah*	may God forgive me
Adjective	5	13.1%	*khéné*	bad; that sucks
Verb	4	10.6%	*t'minik*	you are teasing/kidding
Total	**38**	**100%**		

Table 1. Type repartition by grammatical category

Semantic field	Count of type	Percentage	Example	Meaning
Negative emotion/affect	14	30.4%	*seum*	hatred, rage, resentment
Religions and beliefs	10	21.7%	*hamdoullah*	thanks to God
Vulgarity/insults	7	15.2%	*zeub*	dick
Politeness	6	13%	*bsartek*	congrats (lit. to your health)
Positive emotion/affect	4	8.7%	*hala*	great
Inappropriate behavior	3	6.5%	*khapta*	wasted, high
Country/culture of origin	2	4.3%	*zitoune*	olive
Total	**46**	**100%**		

Table 2. Type repartition by semantic field

16 The types were also categorized according to the register they belong to in Arabic (since in French they all belong to a sub-standard register), as well as according to their presence in a dictionary, whether traditional or urban, because these 'linguistic' factors were perceived as potentially affecting the diffusion of the forms. *Le Nouveau Petit Robert de la Langue Française* online was used as the traditional dictionary and *Le Dictionnaire de la Zone* online as the

urban dictionary[3]. The register to which a form belongs almost consistently aligns with its variety of origin, with colloquial and vulgar forms coming mainly from (North African) dialectal varieties of Arabic, and standard forms coming from Modern Standard Arabic. Accordingly, 18 forms were categorized as colloquial, 15 as standard and five as vulgar, while 18 did not appear in any dictionary, 12 in the urban dictionary and eight in the traditional dictionary. If a form appeared in a dictionary as a different part of speech or with a different meaning as the one it was used as or with when observed during fieldwork, it was still counted as appearing in this dictionary, because at least *a* form of the type is 'officially' attested.

3.2. Online Survey

17 As mentioned above, the diffusion of contemporary Arabic borrowings in metropolitan French was assessed by evaluating respondents' familiarity with the borrowed forms. The construct of Familiarity was operationalized as encompassing both knowledge and self-reported usage of the form, since it was believed knowledge of a form, i.e. having at least heard it even if not using it, already attests to some degree of familiarity with it. A Familiarity score served as a proxy to measure the diffusion of the forms. The survey was adapted from Secova, Gardner-Chloros and Atangana's (2018) questionnaire, which evaluates the popularity of various features of French CUV among typical users of such vernaculars by assessing their projected usage of these traits.

18 After explaining the goal and format of the survey, as well as obtaining written consent from the participants and letting them practice with an example word, the survey asked a few demographic questions about the participants' gender, age, geographical location, highest level of education, socioprofessional category, languages (L1s and L2s), and cultural identity. Then, respondents had to answer a series of questions about the 38 contemporary Arabic borrowings identified in the field. Each borrowing was first presented in writing, often with multiple spellings currently in use. It was also presented in an auditorily recorded carrier sentence (with transcription provided), since Arabic borrowings are mostly used in speech. The carrier sentences were as far as possible original, reconstructed or slightly

modified sentences recorded during the fieldwork phase to make sure the examples reflected an authentic way of using these lexical items. The most semantically neutral sentences were chosen to avoid participant bias based on elements that were not the variable under study. All the example sentences were recorded by the researcher to avoid speaker effect.

19 After being introduced to a borrowing, participants were asked to answer four multiple-choice prompts about their familiarity with this Arabic form. The first prompted them about their knowledge of the word ("Je connais ce mot (je l'ai déjà entendu ou utilisé): oui/non") and the next three about their usage of it: frequency of use, interlocutors with whom they use it, and modalities of use. Participants were asked to answer these last three prompts only if relevant for them. Appendix A provides a sample page of the survey for the target word *rhéné*. Appendix B lists the carrier sentences that were used on the survey for the 38 borrowed Arabic forms.

20 The multiple-choice answers were modified Likert scales adapted to each part of the prompt to provide the most representative Familiarity score. The initial yes/no question about knowledge of the word followed the same principle, where « no » scored 0 and « yes » scored 1. Words and expressions with higher scores reflect greater familiarity, and therefore greater diffusion. Lastly, there was a comment box at the end of each page for participants to leave any comments they wanted about the form. The last page of the survey also asked two optional open questions prompting respondents to share any other borrowed Arabic forms they may use or anything else they may want to add about their usage of Arabic borrowings ("Est-ce que vous utilisez d'autres mots empruntés à l'arabe quand vous parlez français ? Si oui, lesquels ? Est-ce qu'il y a quoi que ce soit que vous voulez ajouter à propos de votre usage des emprunts arabes en français ?").

21 To maximize completion of the survey and minimize the effects of focus fatigue, the survey was optimized to be as short as possible (about 15 minutes), and the borrowings were presented in random order to each participant. The survey was shared using snowballing methods, and, importantly, it was shared with the students at the Orange high school where previous fieldwork took place. Their

demographic profile is therefore well represented in the results, as we will see below.

3.3. Participants

22 All participants were aged 11 or older and living in metropolitan France at the time of the survey. They had also lived at least half of their lives in one of the 13 administrative regions of metropolitan France at that time and were native or near-native speakers of French, according to the selection criteria for participation in the survey. Moreover, participants who failed to answer all the questions they were expected to answer for more than 10% of the Arabic types were discarded [4]. Based on these criteria, 217 survey responses were collected and submitted to statistical analysis. Despite random sampling, the survey results are unbalanced, with some demographic groups more represented than others. Below is the repartition of participants according to the sociodemographic characteristics retained for analysis [5]. The socio-professional categories in Table 4 correspond to the categories from l'Institut National de la Statistique et des Études Économiques (https://www.insee.fr/fr/metadonnees/pcs2020/groupeSocioprofessionnel/2?champRecherche=true), to which an extra category was added to accommodate for unemployed participants or family members. Indeed, for students financially relying on someone, the socio-professional category of the main income provider was asked.

	Female		**Female total**	Male		**Male total**	No answer		**No answer total**	**Non-binary total**	**Total**
Ethnicity	Other	NA [6]		Other	NA		Other	NA		Other	
11-15 yo	6	4	**10**	4	1	**5**					**15**
16-19 yo	60	8	**68**	23	4	**27**	1	1	**2**	**1**	**98**
20-29 yo	7		**7**	6		**6**					**13**
30-39 yo	25	1	**26**	17	1	**18**					**44**
40-49 yo	17		**17**	4		**4**					**21**
50-59 yo	13		**13**	5		**5**					**18**
60+ yo	6		**6**	2		**2**					**8**
Total	**134**	**13**	**147**	**61**	**6**	**67**	**1**	**1**	**2**	**1**	**217**

Table 3. Participant repartition by age, gender and ethnicity

Socio-professional category	Count of participants

Farmer	5
Business owner	10
Executive/professional/company head	52
Employee	30
Laborer	19
Intermediate professional	64
Unemployed	37
Total	**217**

Table 4. Participant repartition by socio-professional category

Level of urbanization Region	Conurbation	Peri-urban area	Small-/medium-sized town	Rural area/ Village	**Total**
Auvergne-Rhône-Alpes	16	10	10	20	**56**
Bourgogne-Franche-Comté	1	1	1	2	**5**
Bretagne	1				**1**
Centre-Val de Loire	3		3		**6**
Grand Est	1		5	1	**7**
Hauts-de-France			1	1	**2**
Île-de-France	4	3		1	**8**
Nouvelle-Aquitaine		1	1		**2**
Occitanie	2	1	2	3	**8**
Pays de la Loire	1			2	**3**
Provence-Alpes-Côte d'Azur	8	6	56	49	**119**
Total	**37**	**22**	**79**	**79**	**217**

Table 5. Participant repartition by region and level of urbanization

3.4. Data analysis

23 Due to the quantity and uneven distribution of the online survey data collected, the following quantitative analysis is exploratory. It combines descriptive statistics and likelihood tests aimed at determining which sociodemographic factors help explain respondents' familiarity with contemporary Arabic borrowings, as well as which forms are most diffused. The results provided are based mainly on the total Familiarity score obtained from each respondent or for each

Arabic form. The software R 4. 4. 3 (R Core Team, 2025) was used to analyze the data.

24 Not all the information gathered in the Demographic Information part of the survey was used in this analysis. Independent variables with limited explanatory power and non-significant results (e.g., multilingualism), or which were confounded with other factors (e.g., level of education) were not retained.

4. Results

4.1. Diffusion by form

25 The results concerning the diffusion of the 38 Arabic forms or types will be presented first to provide some answers regarding which borrowed forms have diffused in metropolitan French, and what linguistic factors may be able to explain this. These results are based on the Familiarity score by form, i.e., the sum of the 217 respondents' Familiarity score for each form, as well as on the Knowledge score by form, comprising only the 217 answers to the first question about knowledge of the form.

4.1.1. Most familiar forms

26 On average, any given Arabic form is known by about 131 respondents out of 217, with great disparities between the specific types (M = 131.36, SD = 65.46, range [7 – 216])[7]. This great variation between forms is not only reflected in the Knowledge score, but also in the Familiarity score for each type (M = 475.18, SD = 376.3584, range [16 – 1,523]). This means that, unsurprisingly, some forms are much more known and used than others. The eight most popular forms are known by at least 90% of the respondents, while the three least popular (*g3ar*, *houfik*, *habla*) are known by less than 10% of them. Table 6 below shows the 10 forms respondents are the most familiar with, i.e. the most diffused ones in the sample. They pertain mainly to the lexical fields of negative emotion/affect (N = 4) and politeness/greetings (N = 3, including two also pertaining to religion), partly reflecting the proportion of forms for each semantic field presented in the questionnaire. Thus, they do not tell us much about

a preference for specific semantic fields, since those were unbalanced [8].

Rank	Form	Grammatical category and meaning as used in French [9]
1	*kiffer*	(v) to like (a lot)
2	*seum*	(n) hatred, rage, resentment (lit. venom)
3	*niquer*	(v) to fuck; to fuck/win over someone; to deteriorate
4	*wesh*	(int) Hey!; What's up?; [discourse marker]
5	*bled*	(n) country/village of origin
6	*cheh*	(int) Sucks to suck!
7	*miskine*	(n) someone who elicits pity
8	*Inch'Allah*	(int) God willing
9	*hess*	(n) misery; shit
10	*salam (alaikum)*	(int) Greetings!; Hello!

Table 6. Top 10 most diffused Arabic forms

27 A Spearman's rank correlation smoothing out the amplitude differences between the scores of the two factors reveals a strong positive correlation between the Knowledge and Familiarity scores ($r(36) = .96$, $p < .001$), meaning that the forms that are the most known to respondents are also the most used. Whereas the two rankings are highly correlated, we can note some discrepancies between them. One worth noting is that two forms belonging to the semantic field of religion and beliefs (*wallah* and *hamdoullah*) are present in the top 10 for Knowledge but not present in the top 10 for Familiarity. Moreover, the forms *inch'Allah* and *salam (alaikoum)*, also religious terms, though present in the top 10 of both measures rank #3 and #6 respectively in the Knowledge ranking and only #8 and #10 in the Familiarity ranking. This points towards some qualitative differences between knowledge and usage of the forms based on semantic field.

4.1.2. The influence of linguistic factors

28 To assess whether there is a link between the register in Arabic or presence in a French language dictionary of a form and its diffusion, a two-way independent measures analysis of variance (ANOVA) was performed. Register did not turn out to be a significant factor, whereas presence of the form in a dictionary did ($F(2, 33) = 25.08$, p

= 2.38^{-7}). A post-hoc analysis using pairwise T-tests with pooled SD and Bonferroni adjustment was performed, and it revealed that familiarity is significantly higher with forms found in a traditional dictionary than it is with forms found in an urban dictionary ($p < .001$) or in no dictionary at all ($p < .001$), despite the fact that there are only eight forms in the "traditional" subgroup. There is no significant difference in familiarity with forms from the other two groups. The data is visualized in Figure 1. The black star indicates the mean score and the black line the median score, while the box displays the interquartile range. The whiskers mark the range of the remaining data. However, with few forms and uneven distributions of the data in the subgroups, these results must be considered as a suggestion rather than a conclusion on the potential role of these linguistic factors on the diffusion of the Arabic forms under study.

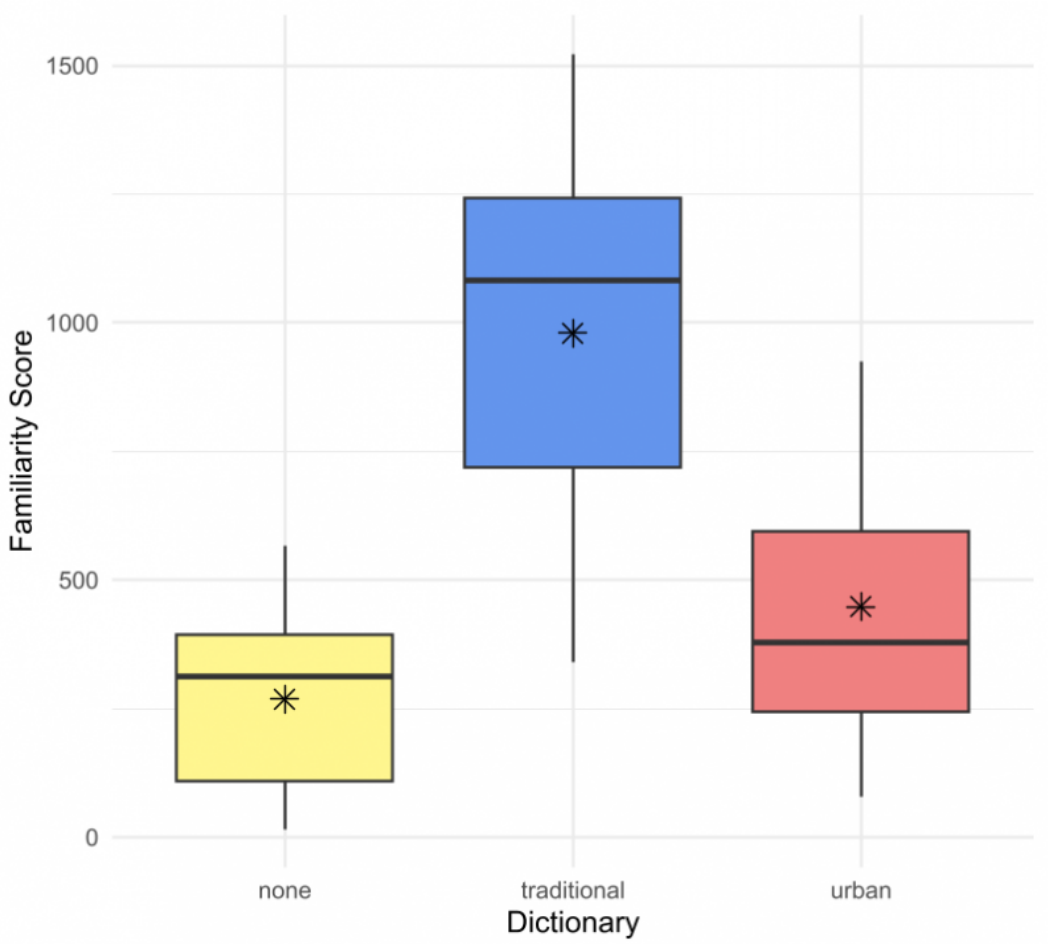

Figure 1. Familiarity scores by presence in type of dictionary

4.1.3. Most familiar forms in selected groups

29 Because the semantic fields of *verlan*—a lexical phenomenon similar to Arabic borrowings in its social trajectory—vary between young males and young females in multiethnic urban neighborhoods (Goudaillier, 2002; Melliani, 2000; Méla, 1997), the researcher investigated whether there were any qualitative differences between men and women in their familiarity with borrowed Arabic forms. A Spearman's correlation reveals a strong positive correlation between the two groups in their familiarity with the words tested ($r(36)= .97$, $p < .001$). Though, because it has been reported in a previous study that insults in Arabic are a sign of virility (Billiez, 1992), a further step was taken for a more fine-grained comparison: the Familiarity scores for males and females for the seven forms that belong to the semantic field of vulgarity and insults were compared[10]. A two-sample independent t-test reveals that females report being significantly less familiar than males with forms pertaining to this semantic field ($t(113.01) = -2.89$, $p = 0.0047$) with mean Familiarity scores of 12.03 and 16.16 respectively. Hence, there are at least some qualitative differences between these two groups in how they use contemporary Arabic borrowings.

30 A similar analysis was conducted for age, because the use and diffusion of contemporary Arabic borrowings appears to be an age-related phenomenon. The sample was split into two groups for the purpose of this test: the age groups 11-15 years old and 16-19 years old were classified as adolescents—age range closely matching the years of secondary education— and the other groups were classified as adults. This division was operated as such because the first group, adolescents, tends to move away from linguistic norms (Trimaille, 2004), while the second group, as part of the workforce, tends to use more standard language (Pooley & Mostefai-Hampshire, 2012). A Spearman's correlation between the Familiarity scores of both age groups on each of the 38 Arabic types was performed, again revealing a highly positive correlation ($r(36) = .85$, $p < .001$), and therefore no significant difference between the two age groups regarding which forms they report being most familiar with.

31 The results showing no major differences in the forms known and used by these different groups, no further investigation was carried

out concerning the forms themselves. The rest of the results presented here focus on the influence of sociodemographic factors on the overall familiarity with contemporary Arabic borrowings.

4.2. Diffusion by participant characteristics

32 These results are based on the total Familiarity scores obtained from participants, i.e., the sum Familiarity score of the 38 types for each respondent. Before delving into the sociodemographic characteristics of participants that significantly impact their familiarity with the forms, we can first note that the mean Familiarity score per respondent is 83.21 (out of a possible maximum of 456) with considerable variability between participants (M = 83.21, SD= 59.87, range [4 – 339]). The median score is 73, so overall, total Familiarity scores tend to be lower than 83.21 and the distribution of scores is slightly skewed right. The distribution is otherwise normal, allowing for the careful use of parametric statistical tests.

4.2.1. The influence of social factors

4.2.1.1. Diffusion by gender

33 Of the 217 respondents, 147 identified as female and 67 as male. Because only one respondent identified as non-binary and two preferred not to reveal their gender, only the sublevels Female and Male will be used for the gender analysis[11]. Figure 1 below illustrates the results obtained by gender. The dots mark the outliers. Here again, there is a lot of variation between participants, with a mean Familiarity Score of 78.10 for females (SD = 60.01) and 92.30 for males (SD = 58.75). There are also a few outliers, especially females, represented by the dots in Figure 2, with scores ranging from 4 to 339. These being true outliers, they were not taken out of the data. A first look at Figure 2 suggests males may be more familiar than females with the Arabic forms they were asked to evaluate. However, an independent samples t-test revealed no significant difference between the two groups (t (130.38) = -1.63, p = .11).

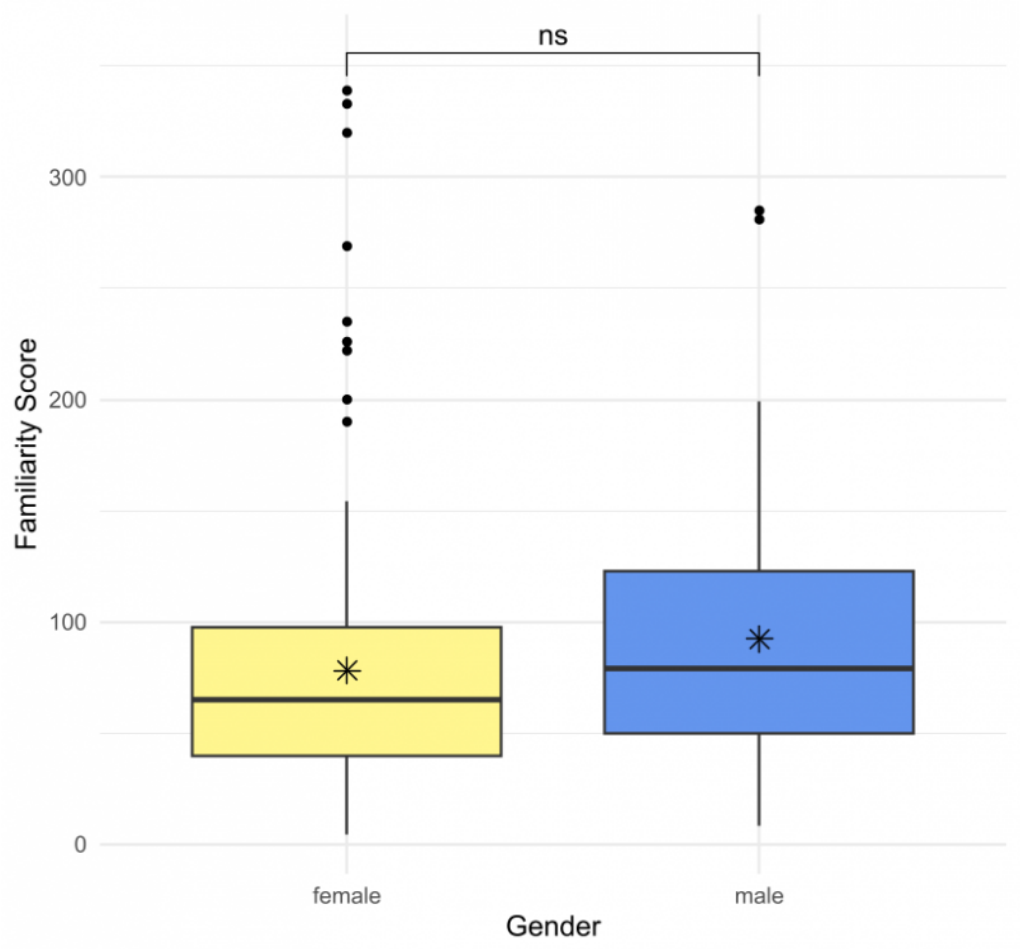

Figure 2. Familiarity scores by gender

4.2.1.2. Diffusion by age

34 Participants were categorized in age bins, corresponding roughly for the first two to the middle school and high school years respectively, and then to each decade of life up to approximately retirement age [12]. Figure 3a,b below visualize the results concerning the Familiarity score obtained by each age group. The large number of respondents of high school age is seemingly an artefact of the participation of the high schoolers from Orange in the survey.

35 We can note in Figure 3a the great amplitude in the Familiarity scores for some of the groups, as well as overall decreasing Familiarity scores with increasing age, which is better illustrated in Figure 3b. This tendency is interrupted at the right extremity, as the oldest group (60+ yo) has a mean Familiarity score higher than the previous age group (50-59 yo), but this could be an artefact of the small

number of respondents in the 60+ year-old group. Alternatively, it could suggest that these two groups do not actually behave differently if the difference between their scores is not significant.

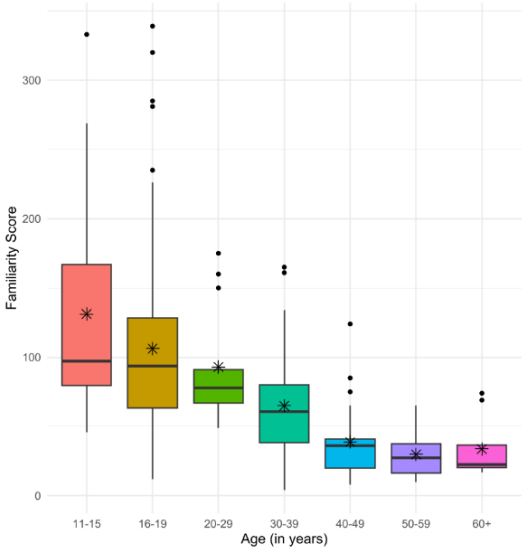

Figure 3a. Familiarity scores by age group

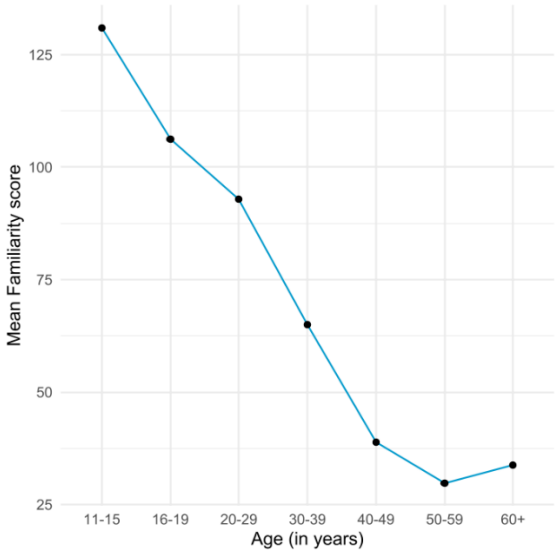

Figure 3b. Mean Familiarity scores by age group

36 A one-way independent measures analysis of variance (ANOVA) between age groups revealed age is a highly significant factor in determining Familiarity scores ($F(6, 210) = 13.40$, $p = 7.63^{-13}$). To reveal where the differences lie, a post-hoc analysis using pairwise T-tests with pooled SD and Bonferroni adjustment was performed. Despite the decreasing Familiarity scores with increasing age, the statistical analysis did not reveal any significant difference between any consecutive group, nor did it reveal any significant difference between any of the three youngest groups and any of the three oldest groups. However, as seen in Table 3, there is great variation between the sample sizes of the different age groups, with some small sample sizes, which can affect the robustness of a parametric test. As a matter of fact, none but one of the age groups with a sample size of less than 30 has a roughly normal distribution (50-59 yo), with two groups (20-29 yo and 60+yo) even presenting binomial distributions. Therefore, most t-tests performed in the pairwise comparison lack statistical power to be reliable, likely producing a Type I error.

Combining the two teenage groups on the one hand and the two oldest groups on the other to increase the sample size for these age groups did not yield any more significant differences between consecutive groups. In short, we cannot rule out that significant differences may exist between consecutive age groups, but the current data does not provide reliable evidence to support this hypothesis. However, despite small sample sizes and low statistical power, some age differences came back as significant. The four older age groups (from 30-39 yo to 60+ yo) all show a significant difference with each of the two youngest age groups (11-15 yo and 16-19 yo) with at least $p < .01$[13]. The fifties group also shows a significant difference at the $p < .05$ level with the twenties group[14], providing some supporting evidence for the tendency observed in Figure 3b.

37 Based on the previously cited literature and comments collected during fieldwork, an interaction was also suspected between age and gender. A two-way independent measures ANOVA with an interaction between gender and age was performed. While the interaction itself did not come back significant, age was here again significant ($F(6, 200) = 13.04$, $p = 1.96^{-12}$) and gender is approaching significance in this model ($p = .06$). While the interaction Age*Gender was not significant, the model was better when containing an interaction (based on the residual sums of squares) and better with both variables than with only gender or age. In other words, the variables gender and age explain Familiarity scores better together than they do on their own. The same interaction was tested after collapsing the age groups at the two ends of the age spectrum like above, but it was still not significant, and the sum of squares was higher for this model. Therefore, statistically, adolescent male speakers are not more familiar with Arabic borrowings than other speakers, but gender improves model fit.

4.2.1.3. Diffusion by social class

38 Social class is a complex concept encompassing a range of educational, social, and personal factors, and it cannot be reduced to just one measure (Fabricius, 2022). However, because highest level of education was confounded with age for younger speakers on the survey, the socio-professional category was used as a proxy for social class. This factor alone only approximately reflecting social class, the

results will first be presented for all seven socio-professional categories commonly used in France in census and official surveys, and then discussed in terms of social class. The category Farmer only counts five responses, which reflects the low prominence of this category in France rather than a sampling bias. Mean scores are split: for four of the socio-professional categories, they stand between 103 and 125 (Farmer, Business owner, Laborer, and Unemployed) and between 60 and 82 for the remaining three (Executive/professional/company head, Unskilled employee, Intermediate professional), suggesting potential significant differences between the groups. Figure 4 visualizes these results.

39 A one-way ANOVA between groups indeed revealed a significant difference at the .001 level ($F(6, 210) = 5.59$, $p = 2.12^{-5}$). The post-hoc analysis, also suffering from Type I error, only revealed four significant differences: between Executive/professional/company head and Business owner ($p < .05$), Laborer ($p < .001$), and Unemployed ($p < .05$), respectively, and between Intermediate professional and Laborer ($p < .01$). Executive/professional/company head being the highest socio-professional category, Business owner and Intermediate professional being in the middle range, and Laborer and Unemployed being in the lower range, there are thus significant differences between groups representative of the upper, middle, and lower classes (between adjacent as well as non-adjacent classes). Business owner is a category with a lot of variation in terms of social class, with some business owners economically and culturally closer to the working class than members of the middle class. Considering the small sample size for this group, it may be skewed towards that type of business owners, which could explain why this category scored similarly to the lower class rather than the middle class in this survey.

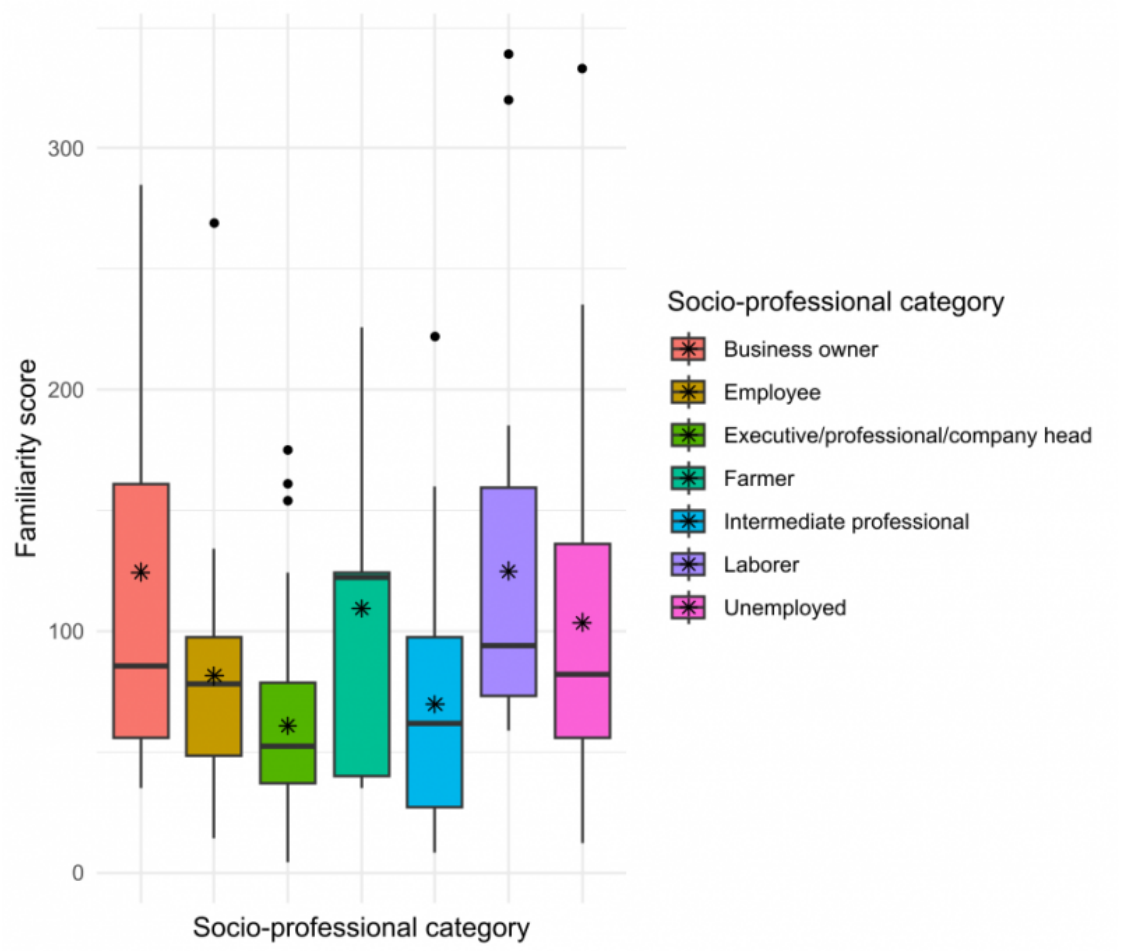

Figure 4. Familiarity scores by socio-professional category

40 To verify whether these differences are maintained at a larger scale, and to have a more robust statistical test, the socio-professional categories were combined into three levels roughly representative of the lower class, middle class, and upper class: Unemployed, Laborer, and Unskilled employee were grouped together as lower class, Intermediate professional and Business owner represent the middle class, and Executive/professional/company head is taken to represent the French upper class. The level Farmer was removed, as there is great variability in social class and revenues in that category (Insee, 2020), making it difficult to pin it to a specific social class. With only five respondents in that category, eliminating it does not have much impact on the results. Figure 5a shows the mean by social class. We can clearly see a decrease in familiarity with the 38 Arabic forms as social class increases. A one-way independent ANOVA between groups revealed that the difference in Familiarity scores between

groups is significant ($F(2, 209) = 8.29$, $p = 3.45^{-4}$), confirming that significant differences in familiarity with contemporary Arabic borrowings between social classes—based on socio-professional category only—do exist. A post-hoc pairwise t-test with Bonferroni adjustment was conducted. Figure 5b shows where the differences lie between the groups. All groups behave significantly differently but the middle and upper classes.

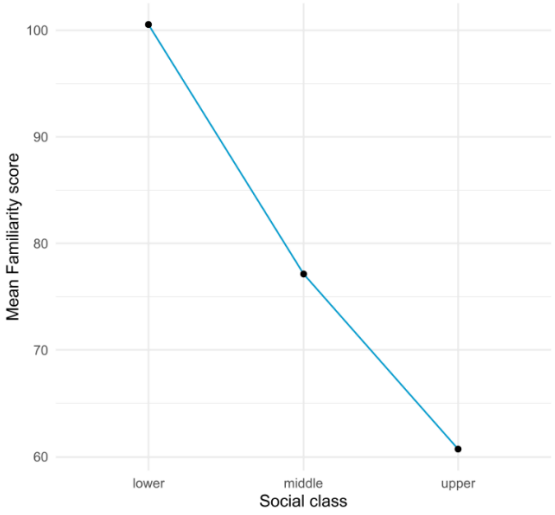

Figure 5a. Mean Familiarity scores by social class based on socio-professional category

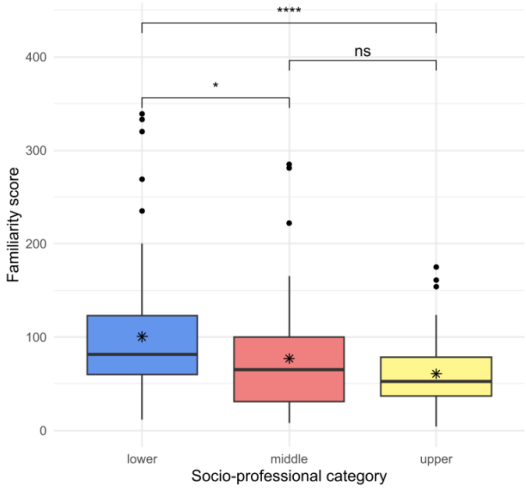

Figure 5b. Familiarity scores by social class based on socio-professional category

41 In addition, an interaction was suspected between age and social class, since Arabic borrowings first appeared in the speech of young speakers from multiethnic working-class neighborhoods. A two-way ANOVA between age and the newly defined variable social class with an interaction between age and social class was performed to verify this assumption. The interaction was not significant, but both factors were at the .001 level. While the interaction was not significant, based on the residual sums of squares, the model was better with the interaction and better than when both factors were tested individually, too. Thus, the variables age and social class together better predict Familiarity scores than each one on its own. The same interaction was tested after collapsing the age groups at the two ends of the age spectrum in this case too, but it was still not significant, and this

model increased the sum of squares. The assumption that there is an interaction between age and social class was therefore not supported.

4.2.1.4. Diffusion by ethnicity

42 Because the use of contemporary Arabic borrowings has often—but not only—been associated in previous studies to the expression of a North African identity, the analysis by ethnicity will compare participants who could be reasonably identified as North African with participants of all other ethnicities. As recording information about someone's race or ethnicity on a survey is prohibited in France, alternative, imperfect ways to gather this information through the online questionnaire were used. To simplify this difficult endeavor, the focus was put only on North African speakers versus other ethnicities/races. Hence, respondents who reported having some working knowledge of Arabic and using it at least sometimes were counted as North African, since Arabic is barely taught in educational contexts in France (Barontini, 2016), and therefore its speakers are most likely to be native or heritage speakers of the language. Considering France's immigration history, such speakers can be considered with a certain level of confidence to be Maghrebi. Additionally, respondents who identified as Tunisian, Algerian, Moroccan, Arab or Muslim in response to the question pertaining to cultural identification on the survey were considered to be North African, too, bringing respondents in that sub-level to 20. Despite this low number, the distribution for this group is normal, and an independent t-test between North African respondents and respondents with other racial/ethnic backgrounds (N = 197) returned a highly significant difference between the two groups ($t(20.03) = 6.15$, $p = 5.20^{-6}$). This difference is illustrated in Figure 6.

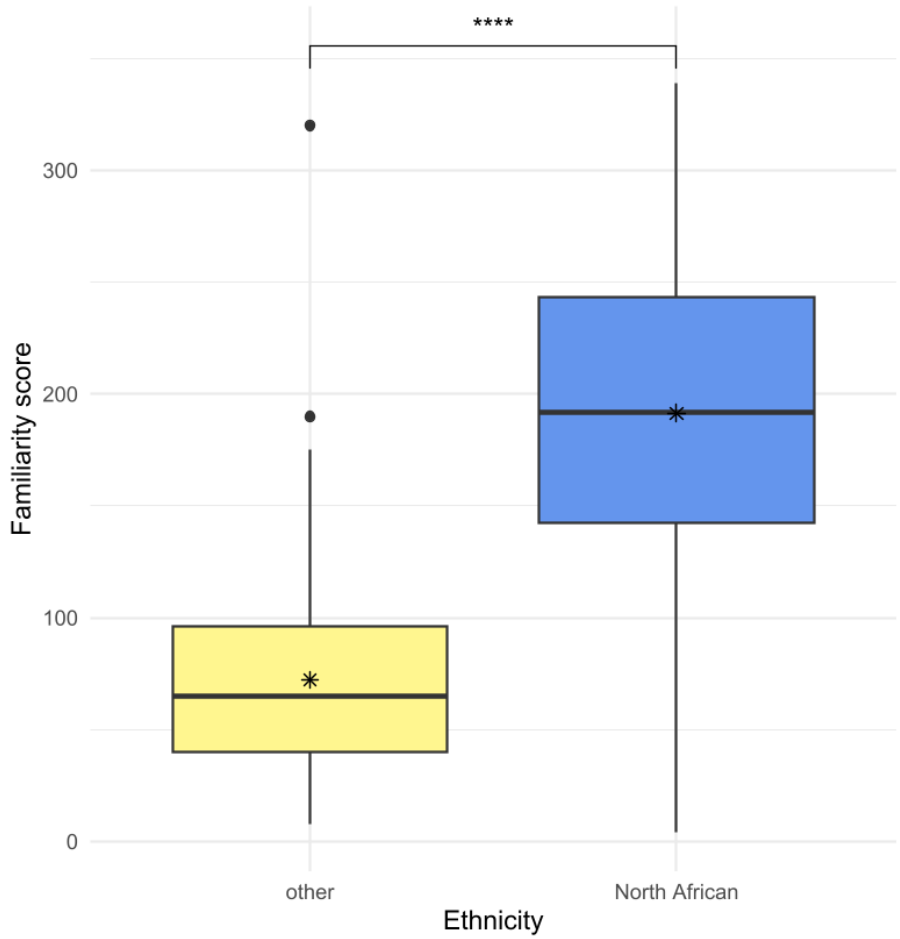

Figure 6. Familiarity scores by ethnicity

43 Unsurprisingly, the respondent with the highest Familiarity score (339) is an Arabic speaker, who also identifies with Arab culture. More surprisingly, the respondent with the lowest Familiarity score (4) also identifies Arabic as one of her languages but does not identify with Arab culture. Additionally, her comments suggest she is linguistically conservative, as she mentions that most of the Arabic words presented in the survey come from "street slang" and not the "noble" Arabic language (her words). This linguistic conservatism may explain her low Familiarity score, as she would likely not use or even maybe know these slang terms [15].

4.2.2. The influence of geographical factors

44 The next two sections investigate the role of the geographical place of residence on familiarity with the 38 Arabic forms tested, and

therefore the diffusion of contemporary Arabic borrowings to different spaces.

4.2.2.1. Diffusion by level of urbanization

45 Respondents were asked to report on the level of urbanization of the area in which they had lived for most of their lives, ranging from a conurbation to a rural area or village. Great variation within the sublevels of this independent variable is again noticeable in Figure 7, but there appears not to be as much variation between the four groups. Three groups—the two most urbanized and the least urbanized—seem to behave similarly, with means within a close range (between 84 and 90) and standard deviations within the same range, too (51 to 66). Only one group appears to behave differently, Small-/medium-sized town, with a mean about 15 points below the highest mean (M = 74.53). This is still within a close range compared to the variation observed between means for the sublevels of other independent variables. A one-way ANOVA between groups confirmed that there is no significant difference between the four levels of urbanization. However, these results need to be considered with caution. While rural areas seem to behave similarly to the most urbanized areas, there is a confound with age, specifically in the sublevel Rural area/village. Indeed, 45 out of 79 respondents in this sublevel—more than half—are between 11 and 19 years old, and we saw above that the two youngest age groups are significantly more familiar with contemporary Arabic borrowings than any other group. Therefore, the mean score for Rural area/village reflects the respondents' age more than it does the level of urbanization of their environment, and a more balanced sample likely would have yielded a lower mean.

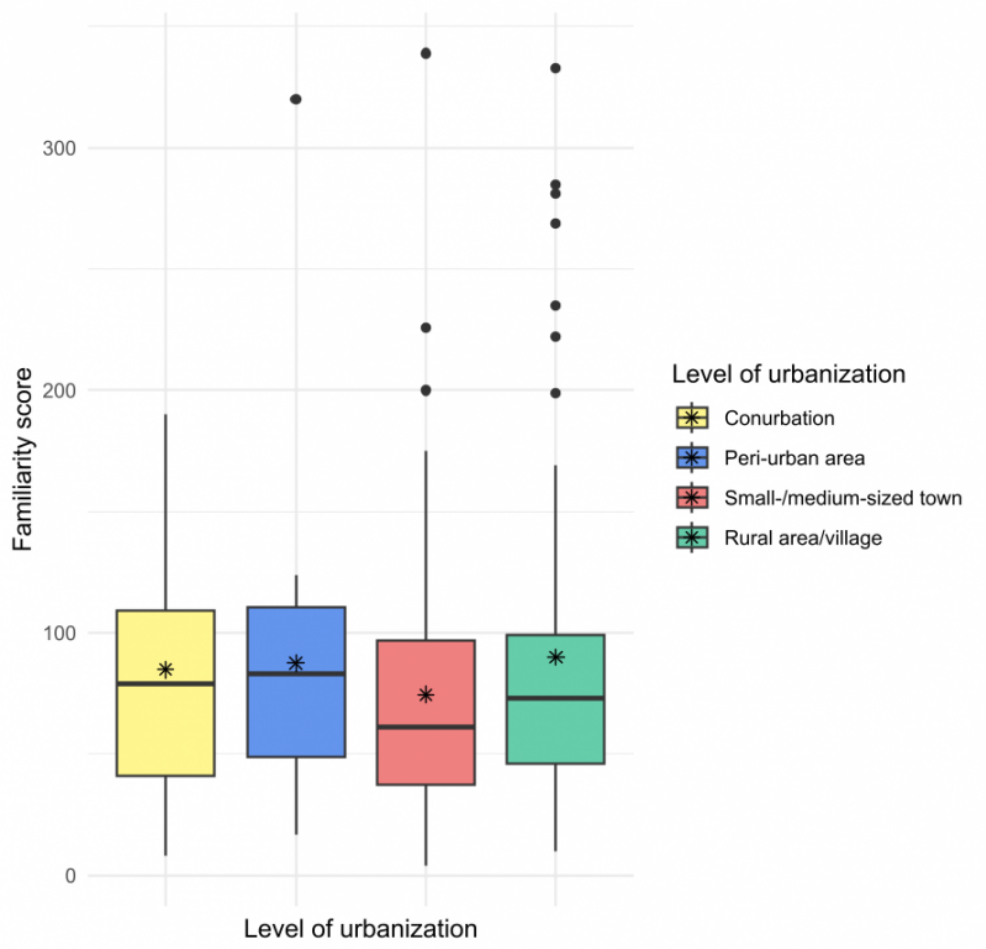

Figure 7. Familiarity scores by level of urbanization

4.2.2.2. Diffusion by region

46 The diffusion of the 38 Arabic forms to the 13 metropolitan regions of France was also investigated. Two regions did not receive any responses, and all others but two received less than ten responses. This is due to a sampling bias, with the two regions where the researcher conducted fieldwork and has lived receiving more responses because of the researcher's networks in these regions.

47 The map[16] in Figure 8 illustrates Familiarity scores and standard deviations for regions that have at least five respondents. We can note that Île-de-France, which includes Paris, has the highest mean Familiarity score. However, with only eight respondents from this region, this mean is only indicating a potential tendency that would need to be confirmed with more data[17]. The small sample sizes for most sublevels preventing a robust statistical test, the following

analysis will focus on the regions Auvergne Rhône-Alpes and Provence-Alpes-Côte d'Azur (PACA), which, although unbalanced (N = 56 for Auvergne Rhône-Alpes, N =119 for PACA), both count over 50 respondents and have roughly similar distributions (skewed positively).

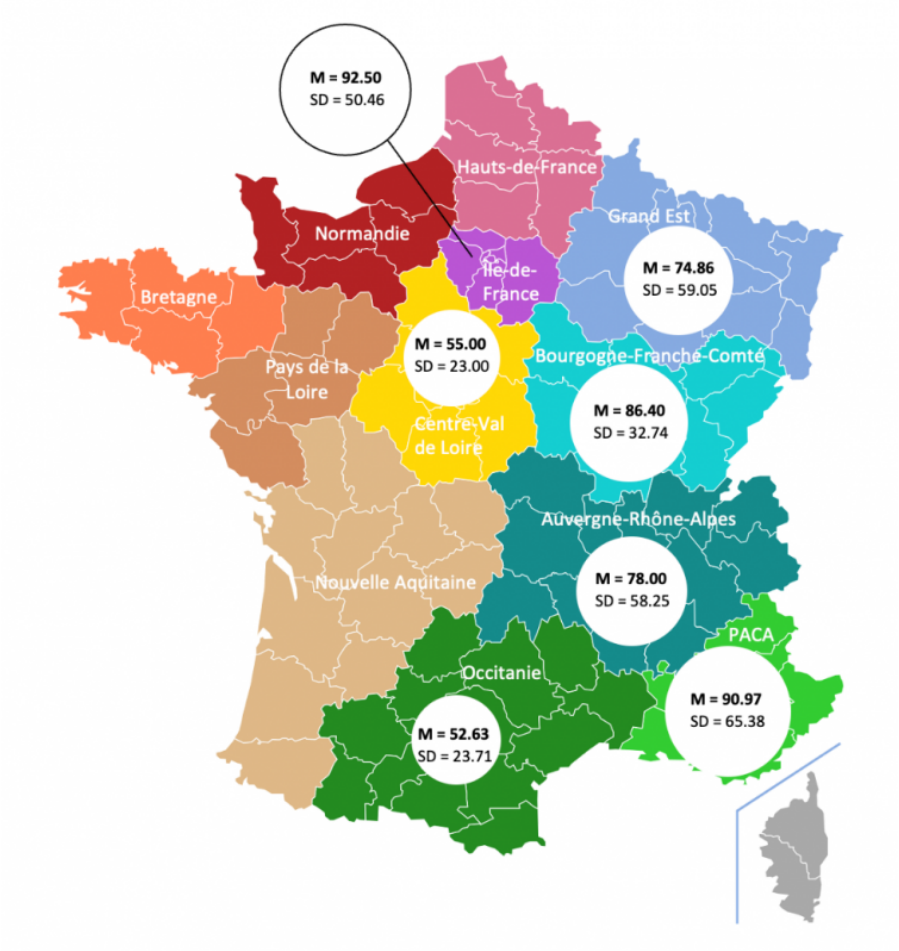

Figure 8. Mean Familiarity scores and standard deviations by region

48 Figure 9 displays Familiarity scores for Auvergne-Rhône-Alpes and Provence-Alpes-Côte d'Azur. The distribution of their scores appears rather similar, with almost identical medians and a slightly higher mean and larger spread for PACA. An independent t-test between the two regions reveals there is no significant difference between them.

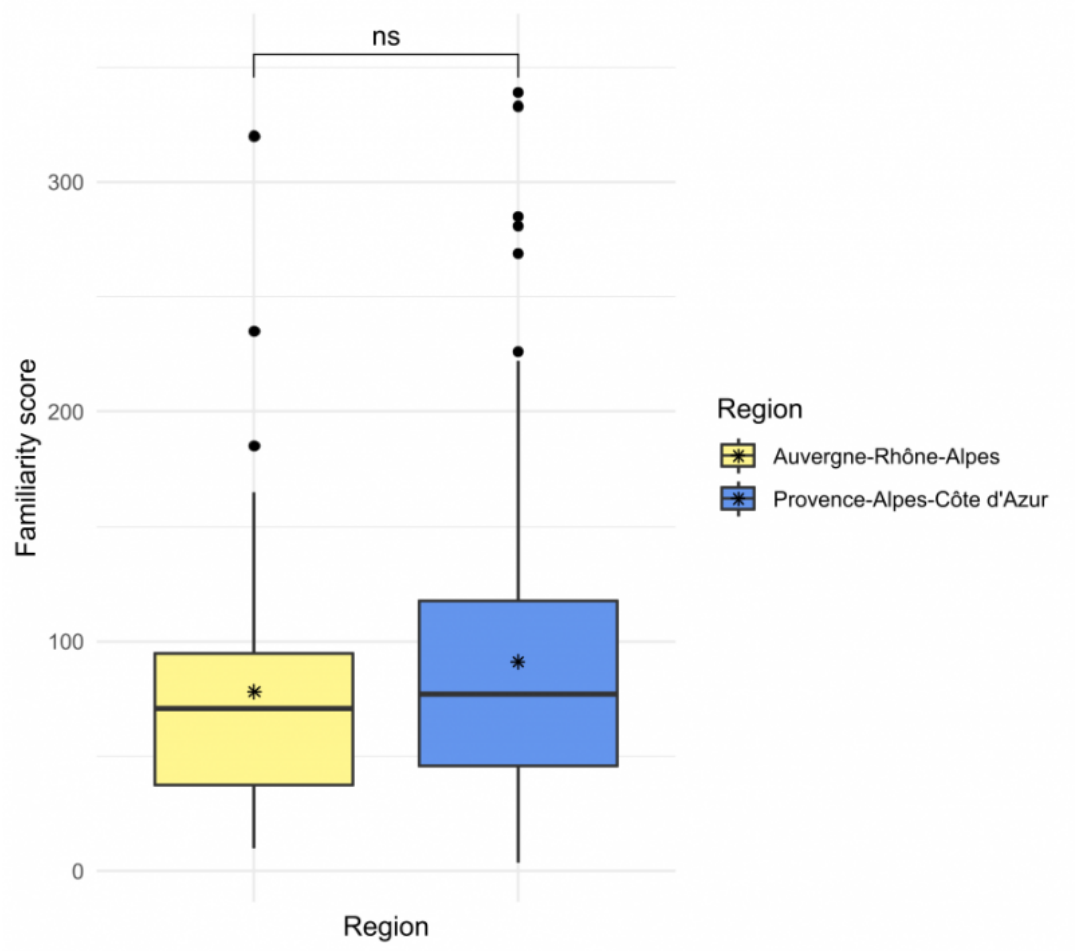

Figure 9. Familiarity scores by region for Auvergne-Rhône-Alpes and PACA

5. Discussion

49 Familiarity scores, and to a lesser extent Knowledge scores were used in this survey as a measure of diffusion of contemporary Arabic borrowings in metropolitan French. From the above results, we can first note that there are great discrepancies in familiarity with Arabic borrowings depending on the form and speaker. Overall, though, more than half of the respondents knew any given borrowed form. However, while nominal knowledge and familiarity were highly correlated, knowledge of a form does not necessarily mean usage. For instance, the analysis by types revealed that respondents may know Arabic words or formulas relating to Islam but not use them. Indeed, if speakers do not identify with Islam or Islamic culture, it is fair to believe they are not very likely to use these forms that are often routine formulas, even if they know them. This aligns with comments

from participants both in the survey and during fieldwork who mentioned that they did not find it appropriate to use Islamic Arabic terms if one is not religiously or culturally Muslim.

50 With an average Familiarity score by respondent just above 83 (out of a maximum possible score of 456), it appears borrowed Arabic forms are not used with high frequency, but some of the 38 forms tested have diffused more widely among the French population. Indeed, borrowings that can be found in the traditional dictionary *Le Petit Robert*, such as the top 4 forms *kiffer*, *niquer*, *seum*, and *wesh*, are significantly more diffused than other Arabic borrowings. Their presence in a traditional dictionary most certainly attests to their diffusion, rather than factors in it. The register the form belongs to in Arabic did not emerge as a significant factor in the diffusion of the borrowings. Since most participants in the survey do not know Arabic (N = 200), it is likely that characteristics of the forms as used in French are more relevant to their diffusion. The least familiar items in the list reflect quite idiosyncratic uses. To summarize, most of the 38 Arabic types evaluated in the survey have diffused geographically throughout France's regions and a diversity of urbanized landscapes, as well as within different demographic groups, despite most respondents' knowledge and reported usage of these forms being limited, although this varies according to different factors.

51 Concerning the influence of socio-demographic factors on the diffusion of these forms, we can first note that the results of the online survey confirm the main assumption from previous research: the use (and knowledge) of Arabic borrowings is an age-related phenomenon, with Familiarity scores decreasing with increasing age, and younger speakers being statistically more familiar with Arabic forms than older speakers, supporting Hebblethwaite's (2018) findings. Arabic borrowings being non-standard lexical items, these results align with the fact that adolescents overall use more non-standard language than adults (Trimaille, 2004). Yet, the results also show that the age effect is not categorical. Arabic borrowings are not only a feature of youth speech; Adults know and use these forms as well, admittedly to a lesser degree, and have similar form preferences. This is reminiscent of Rampton's (2011) findings in England, who observed that adult speakers of British English retained features of CUV they acquired in adolescence but overall used them to a lesser degree as adults. It is

thus possible that adults who used contemporary Arabic borrowings in French when they were younger carried some of these forms over into adulthood, and their use may still index some form of youthfulness. The two older age groups and a large part of the 40-49-year-old group would not have been exposed to Arabic borrowings in their teenage years, at least not to a sustained degree, since CUV French really became a largely recognized phenomenon in the mid 1990s (Trimaille, 2004). This could possibly explain their lower Familiarity scores. It is also important to consider that the use of particular Arabic forms at any given time may be fleeting. Thus, adults may be less familiar with forms that are in use among the youth today, as these forms may not have been in circulation when these adult speakers were younger, to the benefit of other forms. It is also possible that Arabic borrowings are spreading to adult groups through the young generation, which appears to be supported by some survey responses. A few respondents mentioned knowing—and sometimes using—some of the forms thanks to their kids or students. Lastly, we can note that although college-aged respondents (20-29 yo) were less familiar with the Arabic forms than high school-aged respondents, there were no significant differences between these speakers, which could suggest that the twenties mark a linguistic transition into adulthood, as previously suggested (Pooley & Mostefai-Hampshire, 2012). However, with only 13 respondents in the 20-29-year-old group, the statistical analysis lacks power and we cannot draw any definitive conclusions.

52 As for the results concerning the diffusion of Arabic borrowings according to the gender of the speaker, the results demonstrated that gender (male/female) is only significant in explaining differences in Familiarity scores if combined with age. In other words, the contemporary Arabic borrowings under study are more diffused among men only if we also consider their age. This points towards young males being more familiar with the borrowings, even if there was no significant interaction between any age group and gender. The claim that Arabic usage or knowledge is more strongly related to males and marks virility (Billiez, 1992; Pooley, 2018; Pooley & Mostefai-Hampshire, 2012) is thus only partially supported. Moreover, it is possible that women also use Arabic borrowings to index a certain type of masculinity or non-normative femininity (Rilliard, 2023),

which this quantitative analysis cannot show. If the gender groups were more balanced though, the difference between males and females may be significant overall. Lastly, the claim that specifically insults in Arabic index masculinity, which was based on qualitative, ethnographic work (Billiez, 1992), is supported by this quantitative analysis, since males are significantly more familiar with the eight forms pertaining to the semantic field of vulgarity/insults than females. This also supports the more broadly observed link between masculinity and vulgarity (Bourdieu, 1977). However, a greater number of forms in this semantic field would help to strengthen this claim.

53 Additionally, contemporary Arabic borrowings have diffused to all socio-professional categories, but when grouping the socio-professional categories into three groups roughly reflecting social class, we note a significant difference between social classes, with higher Familiarity scores as social class decreases. Thus, although Arabic borrowings have spread beyond the multiethnic lower working classes where they first emerged, it appears they remain more prominent in the language practices of these social groups. As noted by Labov (1966), the working class is less amenable to the pressures of standard language, which allows for the diffusion of non-standard forms like Arabic borrowings[18]. Nevertheless, just like *verlan* has spread to other social classes thanks to the covert prestige of CUV French among the youth, and thanks to its power to mark solidarity with *banlieues* populations and awareness of trends and social issues for social elites (Méla, 1991, 1997; Sloutsky & Black, 2008), contemporary Arabic borrowings have diffused to all social classes as well.

54 Regarding the influence of ethnicity, even though very few respondents could be (imperfectly) identified as North African, these respondents were significantly more familiar with the 38 Arabic types tested than the rest of the participants[19]. This finding supports both Hebblethwaite's (2018) and Pooley and Mostefai-Hampshire's (2012) results suggesting that Arabic speakers and speakers who have an immigration background have greater knowledge of (borrowed) Arabic forms. It also supports the claim made by several researchers that contemporary Arabic borrowings are used in French by speakers to index a North African identity (e.g., Billiez, 1985, 1992; Hassa, 2010; Melliani, 2000; Tetreault, 2015) or a more specific national identity

(Moroccan, Algerian, or Tunisian—Pooley & Mostefai-Hampshire, 2012), although these identities are not the only ones that have been linked to the use of Arabic forms, as seen in Section 2.2. Those other identities that contemporary Arabic borrowings have the potential to index are likely more relevant in attempting to explain the diffusion of these forms beyond the *banlieues*. Collecting more responses from French speakers with a Maghrebi background would help strengthen this claim and allow us to make some important comparisons between subgroups. First, males and females within that group could be compared, since other banlieue French features have been strongly associated with young males of North African descent specifically (e.g., Fagyal, 2010; Jamin, 2005). Additionally, we could compare the practices of Arabic speakers who identify with Arab culture(s) and those who don't, since the respondent with the lowest Familiarity score suggests that knowledge of (some) Arabic may not be enough to guarantee familiarity with and mostly usage of Arabic borrowings. Indeed, the internalization of the assimilationist discourse in France can lead to a devaluation of the heritage language, leading to its limited usage and a subsequent breakdown in intergenerational transmission (Bruneaud, 2005; Hargreaves, 1995; Pooley & Mostefai-Hampshire, 2012). However, other linguistic ideologies related to the status of the different varieties of Arabic can also be at play, as suggested above.

55 Concerning the geographical diffusion of Arabic borrowings, the results pertaining to level of urbanization are of particular interest. Arabic borrowings originated in French CUVs—so as an urban phenomenon—which prompted the researcher to assess whether they have now diffused not only beyond urban working-class neighborhoods such as *banlieues*, but also beyond densely populated urban areas. The results of the survey confirm that they have: they are familiar to respondents at all levels of urbanization, and there is no significant difference between groups, suggesting Arabic borrowings are no longer a specifically urban phenomenon. Nevertheless, as mentioned above, most respondents from rural areas and villages are under 20 years old, artificially driving the mean score for this level of urbanization up. A more balanced sample may have revealed a bigger difference between more urbanized and less urbanized areas, like the lower mean score for the level Small-/medium-sized town seems to

suggest. Furthermore, it is believed that these younger respondents for Rural area/village are primarily high schoolers from Orange (see end of Section 5), so there may additionally be a locality effect with the forms. In other words, this result remains to be verified before definitive claims can be made about the diffusion of contemporary Arabic borrowings to less densely populated areas.

56 Another aspect of geographical diffusion is diffusion by region. The current data is limited and only allows us to talk about two regions with confidence—Auvergne-Rhône-Alpes and Provence-Alpes-Côte d'Azur. While the 38 Arabic forms tested in the survey at first glance seemed to be more diffused in PACA, it turned out there is no significant difference between the two regions, suggesting these borrowings have diffused equally in both regions. We can suggest two reasons why respondents in PACA may be slightly more familiar with these borrowings. First, the 38 types that respondents were asked to evaluate in the online survey were recorded in PACA during fieldwork. There could be an effect of locality, as some of these forms may be more present in the linguistic landscape of this region compared to Auvergne-Rhône-Alpes, which could explain why the forms are more familiar to respondents in PACA. Moreover, the average survey respondent in PACA is younger than in Auvergne-Rhône-Alpes, which also skews the results towards higher Familiarity scores for PACA. A more balanced sample may have reduced the difference between the two regions. Yet, despite the 'advantages' conferred to PACA, Familiarity scores in the two regions are not dissimilar enough for this difference to be statistically significant. Verbeke (2017) found a higher proportion of Arabic borrowings in Marseille-based rap compared to Paris- and Brussels-based rap, which he attributed at least in part to Marseille's history as a port of entry for immigration and the consequently large number of people of North African origin in Marseille. While carefully crafted song lyrics are only comparable to self-reported speech practices to some extent, the current data suggests that PACA, where Marseille is located, may not behave so differently from Auvergne-Rhône-Alpes, despite its history of immigration. Furthermore, although the results for Île-de-France are only indicative because of the very low number of respondents (N = 8) and should be considered with extreme caution, the Paris area displays a mean Familiarity score almost identical to PACA (M = 92.50 vs. M = 90.

97, respectively), suggesting there may not be a difference regarding the use of Arabic borrowings between these two regions either, as opposed to what was observed by Verbeke (2017) and anecdotally suggested by others (Goudailler, 1997; Paine, 2012). However, Marseille, as the largest urban center in PACA, may behave differently from the rest of the PACA region. It is believed that most of the responses collected for PACA are primarily from students, but also faculty/staff, at the high school in Orange where fieldwork was conducted, which also corresponds to lower levels of urbanization. Indeed, the researcher received a large wave of responses from the region right after sharing the survey with the school. With more responses from Marseille, the results for PACA might look different from those for other regions, here especially Auvergne-Rhône-Alpes.

6. Conclusion

57 This study aimed to evaluate the diffusion of contemporary Arabic borrowings in metropolitan France using an online survey assessing respondents' familiarity—knowledge and self-reported usage—with 38 Arabic forms identified during a previous fieldwork phase. Specifically, it was interested in knowing which forms have made their way into colloquial metropolitan French beyond the urban spaces of the *banlieues* and similar neighbourhoods where they first emerged, who uses them, and where. Although the survey data collected only provides partial answers to these questions, some valuable insight was gained.

58 Unsurprisingly, Arabic forms found in a traditional dictionary appear to be more diffused than other types. Moreover, while the use of contemporary Arabic borrowings is still largely an age-, class-, and ethnicity-related phenomenon in colloquial contemporary French, the data show clear diffusion of these forms to a wider variety of speakers in terms of their social and geographical characteristics. Importantly, familiarity with these forms may not be as marked by gender differences outside multiethnic working-class urban neighbourhoods as it reportedly is inside them—although the gender differences in usage of CUV features is questioned by some researchers even in these spaces (Trimaille & Billiez, 2007)—and the urban nature of these forms seems challenged by the data. Therefore,

after crossing linguistic boundaries, contemporary Arabic borrowings appear to have crossed social and geographical borders in contemporary France.

59 However, the results of this study must be considered with caution considering the uneven sampling and subsequent limited data, which impacted the statistical analysis, ultimately limiting the conclusions that can be drawn. This lack of data especially affected the analysis by region. Additional data from all regions would allow researchers not only to compare the frequency of use of contemporary Arabic forms per region, but also to determine whether there are qualitative differences between regions, with some forms perhaps being more localized. Therefore, more data should be collected in the future for a more robust statistical analysis, with the view of performing regression modeling, which allows for a more complex and connected picture of the influence of each independent variable, as well as for the evaluation of random effects such as form and subject. The loss of data and granularity ensuing from the necessary operations on the data to be able to conduct such a statistical analysis with the current dataset affecting the 'quality' of the picture, an exploratory approach was preferred here to preserve the fine-grained details.

60 On the methodological level, the way some social variables were operationalized could also be rethought to increase their validity. Specifically, the factor social class, a complex social phenomenon encompassing many aspects of a person's life, should ideally be determined from a composite score, not from a single measure. Additionally, identifying respondents of North African descent proved difficult based on the information collected in the socio-demographic part of the questionnaire, almost certainly resulting in some being left out. Better ways of identifying respondents' ethnic origins that remain within the French legal framework should be designed, such as asking respondents about their family's immigration history for instance. The question in the survey pertaining to knowledge of a given form could also be formulated differently to ask instead about whether respondents understand the form (its meaning in French), which could potentially be a better measure of knowledge. One must also not forget that self-reported usage is only representative of real

usage to some extent, so the results of this study must be interpreted cautiously regarding usage.

61 Lastly, the survey data was not fully exploited. Participants' responses pertaining to their detailed and specific usage of the borrowed Arabic forms (frequency, interlocutors, and modalities) could be exploited in further studies to get a more detailed insight into how these forms are used on a scale larger than the one currently offered by ethnographic studies, which have so far been the ones investigating the 'how'. Furthermore, responses to the open-ended questions—only just mentioned here—should be further explored in a qualitative analysis to determine which other Arabic forms are in use nowadays, as well as to better understand the different forces that drive the use (or not) of contemporary Arabic borrowings in metropolitan French, including language attitudes and ideologies.

BIBLIOGRAPHIE

Avanzi, M., Barbet, C., Glikman, J., & Peuvergne, J. (2016). Présentation d'une enquête pour l'étude des régionalismes du français. *SHS Web of Conferences*, 27. https://doi.org/10.1051/shsconf/20162703001

Barontini, A. (2016). Arabe maghrébin et berbère : les absents omniprésents. In C. Hélot & J. Erfurt (Eds.), *L'éducation bilingue en France : politiques linguistiques, modèles et pratiques* (pp. 458-472). Lambert-Lucas.

Bentahila, A., & Davies, E. E. (1983). The syntax of Arabic-French code-switching. *Lingua*, 59(4), 301-330.

Billiez, J. (1985). La langue comme marqueur d'identité. *Revue européenne des migrations internationales*, 1(2), 95-104. https://doi.org/10.3406/remi.1985.982

Billiez, J. (1992). Le parler véhiculaire interethnique de groupes d'adolescents en milieu urbain. In *Actes du colloque international des langues et des villes* (pp. 117-126). Didier Erudition.

Blanc-Chaléard, M.-C. (2001). *Histoire de l'immigration*. La Découverte.

Boberg, C. (2018). Surveys: The use of written questionnaires in sociolinguistics. In C. Mallinson, B. Childs & G. Van Herk (Eds.), *Data collection in sociolinguistics: Methods and applications* (pp. 280-301). Routledge.

Bourdieu, P. (1977). L'économie des échanges linguistiques. *Langue française*, 34(1), 17-34. https://doi.org/10.3406/lfr.1977.4815

Bruneaud, J. F. (2005). *Chroniques de l'ethnicité quotidienne chez les Maghrébins français*. L'Harmattan.

Bullock, B. E., & Toribio, A. J. (2009). Themes in the study of code-switching. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp 1-17). Cambridge University Press.

Dollinger, S. (2012). The written questionnaire as a sociolinguistic data gathering tool: Testing its validity. *Journal of English Linguistics*, 40(1), 74-110. http://dx.doi.org/10.1177/0075424211414808

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453-476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41, 87-100. https://doi.org/10.1146/annurev-anthro-092611-145828

Fabricius, A. H. (2022). Social class, social capital, social practice and language in British sociolinguistics: Unravelling historical and ethnographic complexities. In Cieri, C., Yaeger-Dror, M., & Drager, K. (Eds.), *Dimensions of linguistic variation*. (pp. 260-283) Oxford University Press. https://doi.org/10.1093/oso/9780197533499.003.0014

Fagyal, Z. (2010). *Accents de banlieue : aspects prosodiques du français populaire en contact avec les langues de l'immigration*. L'Harmattan.

Fasla, D. (2008). Contribución al estudio de los arabismos en francés magrebí: Identidad versus modernidad. *Sociolinguistic Studies*, 2(1), 61–95.

Gadet, F. (Ed.) (2017). *Les parlers jeunes dans l'Île-de-France multiculturelle*. Ophrys.

Gasquet-Cyrus, M. (2013). Perspectives dynamiques sur la ségrégation sociolinguistique en milieu urbain : le cas de Marseille. *Glottopol*, *21*, 9-21.

Glickman, J., Benzitoun, C., Goldman, J.-P., Scherrer, Y., Avanzi, M., & Boula de Mareuil, P. (2018). Donnez votre Français à la Science ! Internet et la documentation de la diversité linguistique : présentation de la plateforme et premiers résultats. *SHS Web of Conferences*, 46. https://doi.org/10.1051/shsconf/20184602003

Goudaillier, J.-P. (1997). *Comment tu tchatches ! Dictionnaire du français contemporain des cités*. Maisonneuve et Larose.

Goudaillier, J.-P. (2002). De l'argot traditionnel au français contemporain des cités. *La linguistique*, 38(1), 5-24.

Guerin, E. (2018). Les « emprunts urbains contemporains » : une approche sociolinguistique d'un phénomène lexical. *SHS Web of Conferences*, 46, 1-14. https://doi.org/10.1051/shsconf/20184605003

Hambye, P., & Gadet, F. (2014). Contact and ethnicity in "youth language" description: In search of specificity. In R. Nicolai (Ed.), *Questioning language contact: Limits of contact, contact at its limits* (pp. 183-216). Brill.

Hargreaves, A. (1995). *Immigration, 'race' and ethnicity in contemporary France*. Routledge.

Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In M. Haspelmath & U. Tadmor (2009),

Loanwords in the World's Languages : A Comparative Handbook (pp. 35-54). De Gruyter.

Hassa, S. (2010). Kiff my zikmu: Symbolic dimensions of Arabic, English and verlan in French rap texts. In M. Terkourafi (Ed.), *Languages of global hip hop* (pp. 44-66). Continuum International.

Hebblethwaite, B. (2018). Rap and the Islamic lexical field in Parisian French: A study of Arabic religious language contact with vernacular French. In F. Gadet (Ed.), *Les métropoles francophones européennes en temps de globalisation* (pp. 167-184). Classiques Garnier.

Insee. (2020, April 28). Emploi et revenus des indépendants. Édition 2020. https://www.insee.fr/fr/statistiques/4470766?sommaire=4470890. Consulted on September 15, 2024.

Insee. (2021, November 25). *France, portrait social. Éditions 2021.* https://www.insee.fr/fr/statistiques/5432149?sommaire=5435421. Consulted on September 15, 2024.

Insee. (2022, January 18). *Bilan démographique 2021.* https://www.insee.fr/fr/statistiques/6037741?sommaire=6036447. Consulted on September 15, 2024.

Insee. (2024, August 29). *L'essentiel sur… les immigrés et les étrangers.* https://www.insee.fr/fr/statistiques/3633212. Consulted on September 15, 2024.

Jamin, M. (2005). *Sociolinguistic variation in the Paris suburbs*. Doctoral Dissertation, University of Kent at Canterbury.

Kesckes, I. (2014). *Intercultural pragmatics*. Oxford University Press.

Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, 1, 97-120. https://www.jstor.org/stable/4166672

Marchessou, A. (2018). Strasbourg, another setting for sociolinguistic variation in contemporary French. *Journal of French Language Studies*, *28*(2), 265-289. https://doi.org/10.1017/S095926951800008X

M'barek, M. N., & Sankoff, D. (1988). Le discours mixte arabe/français: emprunts ou alternances de langue? *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *33*(2), 143-154. https://doi.org/10.1017/S0008413100012810

Méla, V. (1991). Le verlan ou le langage du miroir. *Langages*, 101, 73-94. https://doi.org/10.3406/lgge.1991.1802

Méla, V. (1997). VERLAN 2000. *Langue française*, 114, 16-34. https://doi.org/10.3406/lfr.1997.5381

Melander, J. (1932). Les mots d'emprunt orientaux en français. *Studia neophilologica* 5(2), 89-102. https://doi.org/10.1080/00393273208586769

Melliani, F. (2000). *La langue du quartier : appropriation de l'espace et identités urbaines chez des jeunes issus de l'immigration maghrébine en banlieue rouennaise*. L'Harmattan.

Morsly, D. (1995). El-Watan, El-Moudjahid, Algérie-Actualités, El-Djeich, Liberté, Le Matin... La Presse

algérienne de langue française et l'emprunt à l'arabe. *Plurilinguismes*, 9(1), 35-53. https://doi.org/10.3406/pluri.1995.986

Mouchotte, L. (2022, May 6). «Wesh», «khalass», «moula»… Pourquoi les mots arabes plaisent-ils tant aux jeunes? *Le Figaro*. https://www.lefigaro.fr/langue-francaise/expressions-francaises/wesh-khalass-moula-pourquoi-les-mots-arabes-plaisent-ils-tant-aux-jeunes-20220506. Consulted on September 15, 2024.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press.

Mzoughi, I. (2015). Intégration des emprunts lexicaux au français en arabe dialectal tunisien. Doctoral Dissertation, Université de Cergy-Pontoise.

Paine, S. (2012). The quadrilingual vocabulary of French rap. *Multilingualism in Popular Arts*, 3(1), 48-69.

Pooley, T. (2018). Le français et l'acquisition transethnique spontannée (*crossing*) de l'arabe parmi des adolescents à Lille. In F. Gadet (Ed.), *Les métropoles francophones européennes en temps de globalisation* (pp. 167-184). Classiques Garnier.

Pooley, T., & Mostefai-Hampshire, Z. (2012). Code-crossing and multilingualism among adolescents in Lille. *French Language Studies*, 22, 371-394. https://doi.org/10.1017/S0959269512000208

Poplack, S., & Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, 24(3), 279-315. http://dx.doi.org/10.1017/S0954394512000018X

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, 47-104. https://doi.org/10.1515/ling.1988.26.1.47

Rampton, B. (1995). Language crossing and the problematisation of ethnicity and socialisation. *Pragmatics*, 5(4), 485-513. http://dx.doi.org/10.1075/prag.5.4.04ram

Rampton, B. (2011). From "Multi-ethnic adolescent heteroglossia" to "Contemporary urban vernaculars." *Language & Communication*, 31(4), 276–294. https://doi.org/10.1016/j.langcom.2011.01.001

Rilliard, M. (2023). Contemporary Arabic borrowings and identity formation in French: Interactional, variationist, and pedagogical perspectives. Doctoral Dissertation, The University of Texas at Austin.

Rilliard, M., & Adams, T. (2018, September 12-13). "Du kiff-kiff au final": Arabic borrowings in French rap [Conference]. *French in Multilingual Urban Centers*, Aston University.

Santolaria, N. (2021, March 7). Parentologie : « Cheh », l'argot que les enfants ramènent de la cour de récré. *Le Monde*. https://www.lemonde.fr/m-perso/article/2021/03/07/parentologie-cheh-ou-le-nouvel-argot-de-la-cour-de-recre_6072227_4497916.html. Consulted on October 1, 2024.

Secova, M., Gardner-Chloros, P., & Atangana, F. (2018). 'Il parle normal, il parle comme nous': Self-reported usage and attitudes in a banlieue. *Journal of*

French Language Studies, 28, 235-263. doi:10.1017/S0959269518000078

Sloutsky, L., & Black, C. (2008). Le verlan, phénomène langagier et social : récapitulatif. *The French Review*, 82(2), 308-324. https://www.jstor.org/stable/25481548

Stammers, J. R. & Deuchar, M. (2012). Testing the nonce borrowing hypothesis: Counter-evidence from English-origin verbs in Welsh. *Bilingualism: Language and Cognition*, 15(3), 630-643. https://doi.org/10.1017/S1366728911000381

Tetreault, C. (2015). *Transcultural teens: Performing youth identities in French cities*. Wiley Blackwell.

R Core Team (2025). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/

Trimaille, C. (2003). *Approche sociolinguistique de la socialisation langagière d'adolescents*. Doctoral Dissertation, Université Stendhal-Grenoble 3.

Trimaille, C. (2004). Études de parlers de jeunes urbains en France. *Cahiers de sociolinguistique*, 9(1), 99-132. https://doi.org/10.3917/csl.0401.0099

Trimaille, C., & Billiez, J. (2007). Pratiques langagières de jeunes urbains : peut-on parler de « parler » ? In C. Molinari & E. Galazzi (Eds.), *Les français en émergence* (pp. 95-109). Peter Lang.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language & Society*, 1, 179-195. https://doi.org/10.1017/S0047404500000488

Verbeke, M. (2017). Represent your origins: An analysis of the diatopic determinants of non-standard language use in French rap. *International Journal of Francophone Studies*, 20(3-4), 209-236. https://doi.org/10.1386/ijfs.20.3-4.209_1

Vinçotte, A. (2024, February 6). «Wesh», «miskine»... Pourquoi ces mots plaisent-ils tant aux jeunes bobos ? *Le Figaro*. https://www.lefigaro.fr/langue-francaise/actu-des-mots/wesh-miskine-pourquoi-ces-mots-plaisent-ils-tant-aux-jeunes-bobos-20240206. Consulted on October 1, 2024.

Winford, D. (2003). *An introduction to contact linguistics* (Vol. 33). Blackwell.

ANNEXE

Appendix A: Sample survey page

Mot : rhéné/khéné

Audio : « C'est rhéné les vacances en Corse. »

Je connais ce mot (je l'ai déjà entendu ou utilisé):

☐ oui ☐ non

SI NON, passez au prochain mot sur la page suivante.

SI OUI, j'utilise ce mot :

(Choisissez l'une des options)

☐ Jamais (passez à la page suivante)

☐ Rarement

☐ De temps en temps

☐ Souvent

☐ Très souvent

(Choisissez l'une des options)

☐ Uniquement entre ami.e.s

☐ Uniquement en famille et entre ami.e.s

☐ Avec tout le monde, mais dans les contextes informels uniquement

☐ Avec tout le monde, peu importe le contexte

(Choisissez l'une des options)

☐ Uniquement à l'oral

☐ Uniquement à l'écrit (ex. SMS, réseaux sociaux, etc.)

☐ À l'oral et à l'écrit

Commentaires (optionnel) :

Appendix B: Carrier sentences for Arabic forms in online survey

Dis pas ça, miskine !

Mais wesh elle est plus grande que moi !

Il m'a niqué ma scolarité.

Wallah, ça m'énerve !

On a gagné, il faut pas avoir le seum, hein !

Elle est partie au bled.

Non, mais si vous kiffez, c'est bien, hein.

C'est elle le sheitan !

Cheh ! Je t'ai bien mis l'œil !

C'est la hess !

Yallah ! Il faut qu'on y aille, là !

Inch'Allah, j'espère que je te vois pas là-bas.

Nous c'est le zbeul, hein, nous il se passe rien.

Oh le zgeg !

C'est rhéné les vacances en Corse !

Oh je dois avoir une tête de zeub.

Vas-y, tu t'minik là !

Hamdoullah, on n'a pas cours !

Vas-y, bsartek !

Bismillah !

Starfoullah !

Je suis complètement rapta.

Habla, il est sérieux ?

T'as un gros g3ar.

Allah habibi !

Ah c'est hala ça, elles tuent tes Converses !

Il lui a fait une hagra.

C'est sa meuf, sa halal.

Ouais, t'es un hallouf.

Oh kah, toi aussi tu t'énerves quand je parle de ta mère.

Hashek, dis pas ça !

Allez là, bleh !

Salam alikoum !

Il lui a mis une zitoune.

On aurait dit un blédard.

Zeubi, je marche pas là !

Il lui a lâché un 3ain !

Elle faisait flipper elle, c'était sheitan.

Oh t'es hallouf toi !

Houfik !

NOTES

1 Calculation based on data from 2020, the last year with comparable data for the different generations of immigrants.

2 But see for instance Trimaille (2004) on the difficulty of defining 'youths'.

3 *Le Petit Robert* is less conservative than most of its traditional counterparts, which oriented the choice of this dictionary, while *Le Dictionnaire de la Zone*, now also existing in a paper version, is a reference for urban slang in France.

4 Some questions were optional for those who responded « no » to the Knowledge question. However, some participants responded « yes » to this question and then did not appropriately answer the following questions regarding their usage of the form. These respondents were discarded if they failed to fully answer when expected for more than four borrowing types out of 38. Few participants have 100% response rate, so respondents with 10% or fewer missing answers were kept for analysis, or a statistical analysis would not have been possible.

5 A comprehensive table of respondents' characteristics would be hardly readable considering the number of independent variables and subgroups under study, so it is split in several, more readable tables.

6 NA stands for North African

7 Based on the Knowledge score by form.

8 It tells us, however, about the preferred semantic fields among the Orange high schoolers.

9 The meanings of the 38 forms were determined using the researcher's previous knowledge, traditional dictionaries (e.g., *Le Petit Robert de la langue française*), online crowdsourced (urban) dictionaries (e.g., *Le Dictionnaire de la Zone*, *Wiktionnaire*), and the help of Arabic-speaking informants.

10 Five of these forms are vulgar in Arabic too.

11 This applies to all tests that involve gender as a variable.

12 In 2022, when the data was collected, the legal retirement age in France was 62.

13 $p < .001$ for all but one comparison.

14 Significance levels not depicted on the graph for readability.

15 This respondent identified as a speaker of dialectal Arabic and was able to identify the forms as coming mainly from 'street language'. Hence, it is likely she knows more forms than she acknowledged in the survey, and her score may reflect her language ideologies more than her actual familiarity with the forms, even considering dialectal variation, which may have made some of these forms unknown to her.

16 Background map from Wikimedia Commons (https://upload.wikimedia.org/wikipedia/commons/8/81/France_proposal_regions_%282014%29_map.svg).

17 The skewed-right but otherwise normal distribution—which mirrors the overall distribution of scores and the distribution of most independent variables' sublevels—is a good sign that this tendency may remain true with more data.

18 He also points out that the working-class men he observed in NYC are especially resistant to the linguistic standard because they perceive their ways of speaking as virile, echoing the above discussion about virility and masculinity.

19 They were very likely more respondents of North African origin in the sample, but they could not be identified with the limitations imposed by the survey format.

RÉSUMÉS

English

Contemporary Arabic borrowings in metropolitan French first emerged in the speech practices of speakers from multiethnic working-class urban neighborhoods (Billiez, 1992; Gadet, 2017), often referred to as *banlieues* or *cités*. While in these neighborhoods their use has been mainly associated with young male speakers of North African origin (e.g., Pooley & Mostefai-Hampshire, 2012), the increasing diffusion of Arabic borrowings beyond the *banlieues* raises questions such as: which forms are used, by whom, and where? Employing an online survey to address these questions, this study investigates the diffusion of contemporary Arabic borrowings in metropolitan French by assessing speakers' familiarity with 38 Arabic forms. Based on 217 survey responses, the results show that there is great variation in the diffusion of individual forms. Moreover, while the use of Arabic borrowings is still largely an age-, class-, and ethnicity-related phenomenon in contemporary colloquial French, the data show clear diffusion of these forms to a wider variety of speakers in terms of their social and geographical characteristics, and challenges previous findings and assumptions regarding the gendered and urban nature of the phenomenon. Therefore, after crossing linguistic boundaries, contemporary Arabic borrowings also appear to have crossed social and geographical borders in contemporary France.

Français

Les emprunts arabes contemporains dans le français métropolitain ont d'abord émergé dans les pratiques langagières de locuteurices issu.e.s de quartiers urbains populaires et multiethniques (Billiez, 1992 ; Gadet, 2017) souvent désignés sous le nom de *banlieues* ou *cités*. Alors que dans ces quartiers, leur utilisation a été principalement associée à de jeunes locuteurs masculins d'origine maghrébine (Pooley & Mostefai-Hampshire, 2012), la diffusion croissante des emprunts à l'arabe au-delà des banlieues soulève des questions telles que : quelles formes sont utilisées, par qui, et où ? À l'aide d'un questionnaire en ligne, cette étude tente de répondre à ces questions en examinant la diffusion des emprunts arabes contemporains dans le français métropolitain en évaluant la familiarité de locuteurices avec 38 formes issues de l'arabe. Sur la base des 217 réponses à l'enquête, les résultats montrent qu'il existe une grande variation dans la diffusion des formes individuelles. De plus, alors que l'utilisation d'emprunts arabes est encore largement un phénomène lié à l'âge, à la classe sociale et à l'ethnicité dans le français familier contemporain, les données montrent une diffusion claire de ces formes à une plus grande variété de locuteurices en termes de leurs caractéristiques sociales et géographiques et remettent en question les conclusions et présupposés antérieurs concernant la nature urbaine et genrée du phénomène. Par conséquent, après avoir franchi les frontières

linguistiques, les emprunts arabes contemporains semblent avoir franchi les frontières sociales et géographiques dans la France contemporaine.

INDEX

Mots-clés
arabismes, emprunt lexical, contact de langues, variation sociolinguistique

Keywords
Arabic loanwords, lexical borrowing, language contact, sociolinguistic variation

AUTEUR

Marylise Rilliard
Institut für Romanistik - Universität Wien
marylise.rilliard@univie.ac.at