



## Between Blad (Homeland) and Ghorba (Exile): An NLP-Assisted Semiotic Analysis of Cultural Identity Construction in Algerian Migrant Discourse

**Khansaa Mohammed Cherif**

Multilingualism Doctoral School, University of Pannonia, Hungary

email: [khansaa.mohammedcherif@phd.uni-pannon.hu](mailto:khansaa.mohammedcherif@phd.uni-pannon.hu)

ORCID: 0009-0001-4362-0301

DOI: <https://doi.org/10.36892/ijlls.v8i3.2652>

**APA Citation:** Mohammed Cherif, K. (2026). Between Blad (Homeland) and Ghorba (Exile): An NLP-Assisted Semiotic Analysis of Cultural Identity Construction in Algerian Migrant Discourse. *International Journal of Language and Literary Studies*, 8(3).405-429. <https://doi.org/10.36892/ijlls.v8i3.2652>

### Received:

05/04/2026

### Accepted:

25/05/2026

### Keywords:

Algerian Arabic; cultural identity; semiotic analysis; migration discourse; corpus linguistics

### Abstract

Migration constitutes both a socio-economic phenomenon and a discursive process through which cultural identities are continuously constructed and renegotiated. In Algerian migrant communities, the Darija terms *blad* (homeland) and *ghorba* (exile) function as central semiotic operators that organize experiences of belonging, displacement, memory, and affective rupture. Despite their prominence in everyday discourse, systematic analysis of how these terms structure cultural identity remains limited. This study investigates how *blad* and *ghorba* are discursively constructed in Algerian migrant narratives, identifies their semantic and collocational patterns, and evaluates the contribution of NLP-assisted corpus analysis. Adopting a convergent mixed-methods design, the research combines quantitative corpus analysis of 1,248 textual units collected from three Algerian diaspora Facebook groups in France with ten semi-structured qualitative interviews with Algerian migrants residing in France. Corpus processing was conducted using spaCy (v3.7) with custom normalization for Darija–French code-switched text, including frequency counts, collocation analysis via normalized pointwise mutual information (NPMI), and qualitative semiotic interpretation. Results show that *blad* is consistently associated with family, memory, warmth, and cultural authenticity, while *ghorba* is linked to loneliness, suffering, and spatial-cultural dislocation in France. Interview data reinforce this semiotic opposition while revealing reflexive hybrid negotiation and platform-specific identity performances. The study demonstrates both the value and the limitations of NLP-assisted methods for preserving cultural depth in migration linguistics and critical discourse analysis.

## 1. Introduction

Migration to France represents one of the most historically significant and enduring postcolonial migratory flows in Europe. Shaped by over a century of colonial entanglement, the Algerian War of Independence (1954 -1962), and successive waves of economic migration, political displacement, and family reunification, this movement continues to influence contemporary Franco-Algerian relations. Recent scholarship situates Algerian migration within ongoing postcolonial dynamics, highlighting persistent structures of inequality and identity navigation (Adamson et al., 2024). By the early twenty-first century, Algerians and their descendants remained one of the largest national-origin groups among France's postcolonial immigrant population. Yet despite their demographic weight, they continue to

occupy an ambivalent position within the French national imaginary, simultaneously viewed as a productive labor force and marginalized as cultural and religious others within a republican framework that prioritizes cultural assimilation (Ait El Caid et al., 2025).

This assimilationist pressure is not only external; it is internalized, negotiated, and resisted through everyday social and discursive practices. Abdelmalek Sayad's (2004) foundational concept of "double absence" the immigrant's simultaneous non-belonging in both host and origin societies remains highly relevant and continues to be revisited in contemporary studies (e.g., Abouyassine, 2023). This structural in-betweenness is profoundly linguistic and semiotic. As Bourdieu (1991) argued, language operates as a field of symbolic power. For Algerian migrants, Darija, the vernacular Arabic variety central to everyday life, serves as a key resource for claiming belonging, preserving cultural memory, and reconstructing identity amid French institutional pressures.

Within this linguistic repertoire, two terms carry exceptional semiotic density: *blad* (homeland) and *ghorba* (exile). These are not simple lexical items but culturally embedded semiotic operators through which migrants articulate orientations toward place, memory, and selfhood. *Blad* evokes familial roots, cultural authenticity, embodied memory, and emotional security. *Ghorba*, drawing on centuries of Arabic literary tradition, expresses existential estrangement and rupture. In contemporary migrant discourse, these terms retain powerful affective and ideological resonance.

The salience of *blad* and *ghorba* is particularly evident in digital spaces. Facebook has become a vital platform for diaspora community formation, affective expression, and transnational ties (Zrizi et al., 2023). Within Algerian Facebook groups, participants engage in fluid code-switching between Darija, French, and Modern Standard Arabic, a practice that mirrors the complex cultural positioning encoded in *blad* and *ghorba* (Ait El Caid et al., 2025). Although semiotic and discourse-analytical frameworks are well-established, systematic corpus-based studies of culturally specific lexical items in diaspora digital discourse remain limited. While earlier work demonstrated the value of combining corpus linguistics with critical discourse analysis (Bekkaoui, 2024), recent scholarship calls for greater integration of computational methods with semiotic interpretation in vernacular migrant communication.

This study addresses this gap. Drawing on a convergent mixed-methods design that combines NLP-based corpus analysis using spaCy (v3.7) with qualitative semiotic interpretation, it investigates how *blad* and *ghorba* construct and negotiate cultural identity across digital and interpersonal Algerian migrant discourse. The study is guided by three research questions:

RQ1: How are *blad* and *ghorba* discursively constructed in Algerian migrant narratives?

RQ2: What semantic and collocational patterns characterize their usage in the corpus?

RQ3: How can NLP-assisted corpus analysis support the identification of these patterns while preserving interpretive and cultural depth?

In pursuing these questions, the study aims to document the semiotic function of *blad* and *ghorba*, validate a replicable mixed-methods framework, and contribute to contemporary debates in migration linguistics. The remainder of this article is organized as follows: Section

2 reviews the theoretical and empirical literature; Section 3 details the methodology; Section 4 presents the findings; Section 5 discusses implications; and Section 6 offers conclusions and future directions.

## **2. Theoretical Framework and Literature Review**

### **2.1 Identity as Discursive Production: Hall, Bhabha, and Diasporic Subjectivity**

Any theorization of cultural identity in migrant discourse must begin with the foundational recognition that identity is not a stable, pre-given essence but a historically contingent, ongoing process of discursive production. Stuart Hall's (1990) seminal formulation remains theoretically indispensable: cultural identity, he argues, is not "an essence but a positioning," constituted within rather than outside representation, and therefore always incomplete, always in process, and always subject to the play of history, culture, and power. For diasporic subjects, Hall identifies two modalities of identity: one organized around shared cultural memory and continuity, and another organized around rupture, transformation, and the irreversible discontinuities produced by colonialism and migration. These modalities coexist in productive tension, generating the hybrid, ambivalent, and contested identities characteristic of diasporic experience. Although Hall's essay is now over three decades old, its theoretical architecture continues to demonstrate strong explanatory power, as evidenced by recent empirical studies in digital diaspora communication (Smets, 2018; Leurs & Ponzanesi, 2018; Bayramoglu, 2023).

Hall's framework requires extension, however, when applied to the analysis of vernacular lexical practice. Bhabha's (1994) concept of the "third space of enunciation" the hybrid, interstitial site in which cultural difference is negotiated, repeated, and displaced provides a complementary theoretical resource for understanding how *blad* and *ghorba* function as sites of ongoing reconstruction between ordinary and diasporic meanings. More recent postcolonial linguistic scholarship has developed this insight systematically. Abouyassine (2023) argues that translingual practice as a mode of creative and critical agency reframes the multilingual, code-switching discourse of Algerian Facebook communities as a sophisticated semiotic practice.

### **2.2 Barthesian Semiotics and the Mythologization of Migrant Space**

Roland Barthes's (1972) theorization of myth as "a type of speech" constitutes the central semiotic apparatus through which this study analyzes the ideological operations performed by *blad* and *ghorba*. For Barthes, myth operates as a second-order semiological system that appropriates an already constituted sign and transforms it into the vehicle for an ideological message, converting history into nature and presenting culturally produced meanings as self-evident and eternal. Applied to *blad* and *ghorba*, this framework illuminates how these terms function not simply as referential labels but as mythologized operators that naturalize particular orientations toward place, memory, and belonging.

### **2.3 Critical Discourse Analysis: Language, Power, and Digital Social Practice**

Norman Fairclough's (1992, 1995) foundational premise that discourse is simultaneously constitutive of and constituted by social structures establishes the critical analytical orientation

of the present study. His three-dimensional framework encompassing text, discursive practice, and social practice provides a systematic methodology for examining how *blad* and *ghorba* reflect and reproduce relations of belonging, exclusion, and cultural power within the Algerian diaspora.

#### **2.4 Algerian Migration, Postcolonial Identity, and the Politics of Belonging**

Algerian migration to France has historically unfolded within a complex nexus of postcolonial displacement, cultural attachment, and sociocultural negotiation. For many Algerian migrants, identity and belonging are not experienced as stable or territorially fixed conditions; rather, they are continuously reconstructed across intersecting linguistic, cultural, and symbolic spaces. This process of identity negotiation is particularly evident in diasporic discourse, where affective memories of the homeland coexist with experiences of marginalization, adaptation, and cultural hybridity. As Abouyassine (2023) argues, migrant identity formation is often articulated through emotionally charged narratives of home, mobility, exile, and historical rupture, reflecting the fragmented realities of transnational existence. Similarly, Bouyazdouzen (2025) emphasizes the centrality of diasporic memory in maintaining symbolic attachments to the homeland, particularly in contexts where migration transforms both individual and collective understandings of identity, belonging, and cultural continuity. Together, these perspectives demonstrate that Algerian migrant discourse functions not merely as a reflection of displacement, but as a dynamic site where meanings of *blad* (homeland) and *ghorba* (exile) are continuously negotiated and reimagined.

#### **2.5 Digital Diaspora Communication, Translanguaging, and Semiotic Identity**

The emergence of social media as a primary medium of diaspora communication has generated a substantial body of scholarship examining how digital platforms mediate multilingual identity practices. Leurs and Ponzanesi (2018) conceptualize “connected migrants” as subjects whose transnational identities are sustained and renegotiated through digital connectivity, providing a foundational framework for understanding the Algerian Facebook communities examined in this study. Subsequent research has significantly refined this perspective.

Code-switching and translanguaging practices in Algerian digital discourse have received specific scholarly attention in recent years. This development carries significant implications for understanding why migrants choose to deploy *blad* and *ghorba* in *Darija* rather than French or Modern Standard Arabic. Li Wei’s (2018, 2022) translanguaging framework, which understands multilingual practice as the deployment of a unitary, dynamic linguistic repertoire rather than the alternation between discrete language systems, provides a robust theoretical basis for analyzing these code-switching practices as coherent semiotic performances of diasporic identity rather than instances of linguistic deficit or interference.

#### **2.6 Corpus Linguistics, NLP, and Migration Discourse: Recent Methodological Developments**

The integration of corpus-based computational methods into migration discourse analysis has advanced considerably in recent years, driven by the increasing availability of tools capable of processing non-standard, code-switched, and morphologically complex texts. While the pioneering combination of corpus linguistics with critical discourse analysis established an

important methodological template, more recent scholarship has extended this model in significant ways. The application of NLP tools to Arabic and Arabic-inflected discourse represents a particularly dynamic methodological frontier. Habash (2022) provides a comprehensive overview of recent advances in Arabic NLP, including tools specifically designed for processing dialectal varieties such as Darija. This is methodologically significant given the orthographic variability, morphological complexity, and frequent code-switching with French that characterize Algerian digital discourse. Although the formalization of pointwise mutual information remains a foundational statistical measure for collocational strength, recent computational linguistics scholarship has refined and contextualized this approach within broader frameworks of distributional semantics and word embeddings (Lenci, 2018). Critically, as Bouyazdouzen (2025) argues, in a formulation that retains full methodological validity, corpus tools are highly effective at identifying linguistic patterns but cannot interpret them independently. This epistemological principle that computational pattern detection is necessary but not sufficient for discourse analysis governs the mixed-methods design of the present study and underscores the importance of integrating NLP-assisted analysis with sustained qualitative and semiotic interpretation.

### **2.7 The Conceptual Pair *Blad/Ghorba*: Toward a Cultural Semiotic Analysis**

Despite the evident centrality of *blad* and *ghorba* to Algerian migrant affective and discursive life, the systematic scholarly analysis of this lexical pair as semiotic operators remains strikingly limited. While Sayad's seminal work on *el ghorba* (2000) highlighted the mythification and collective misrecognition surrounding the experience of exile in Algerian migration narratives, contemporary research has tended to engage with these terms only incidentally rather than as primary objects of systematic semiotic investigation. Research on Arabic concepts of homeland and exile has largely focused on classical literary traditions and formal poetic contexts, while sociolinguistic and discourse-analytical scholarship on Algerian migration has rarely treated the *blad/ghorba* opposition as a structured cultural semiotic system in digital environments.

Studies on exile narratives within Algerian and broader Arab diasporic contexts highlight the enduring emotional and cultural significance of homeland, estrangement, and displacement in migrant storytelling, demonstrating how memories of *blad* continue to shape diasporic subjectivities across generations. Parallel research on digital multilingualism among North African communities, particularly the work of Ait El Caid et al. (2025), further reveals how vernacular expressions are strategically mobilized in online environments to construct fluid and hybrid identities that transcend national and linguistic boundaries. Moreover, cultural semantic and ethnolinguistic approaches offer valuable methodological tools for examining the culturally embedded meanings encoded within lexical choices, symbolic expressions, and discourse practices. Nevertheless, these frameworks have been applied only marginally to Arabic vernacular vocabulary and remain particularly underexplored in the context of Algerian diasporic digital discourse. Consequently, the semiotic and cultural meanings attached to key vernacular concepts such as *blad* and *ghorba* continue to receive limited scholarly attention, especially within NLP-assisted analyses of migrant identity construction in online communicative spaces.

The present study addresses this notable gap in the literature. By bringing together Hallian identity theory, Barthesian semiotics, Faircloughian critical discourse analysis, and NLP-based corpus methodology, it provides what appears to be the first systematic semiotic analysis of *blad* and *ghorba* as identity-organizing operators in Algerian digital migrant discourse. In doing so, it contributes to convergent scholarly conversations in migration linguistics, digital diaspora studies, and computational semiotics, while advancing a replicable methodological model that can be applied to the analysis of culturally embedded lexical pairs in other diasporic contexts.

### **3. Methodology**

#### **3.1 Research Design**

This study adopts a convergent mixed-methods design (Creswell & Creswell, 2018) that brings together corpus-based quantitative analysis and qualitative semiotic and discourse-analytical interpretation. The choice of this design is both epistemological and practical. Computational corpus tools offer systematic and replicable insights into the frequency, distribution, and collocational patterns of *blad* and *ghorba* across a large dataset. However, the ideological, affective, and culturally nuanced meanings carried by these terms cannot be fully captured through quantitative patterns alone. These deeper dimensions require sustained qualitative interpretation grounded in semiotic and critical discourse frameworks.

The methodological design is therefore grounded in the principle of methodological complementarity (Brezina, 2018). Rather than viewing quantitative and qualitative methods as competing approaches, this study treats them as mutually reinforcing. This design also responds to recent calls in migration linguistics for methodological frameworks that can effectively bridge computational precision with rich interpretive and cultural depth (De Fina, 2020).

#### **3.2 Data Sources and Corpus Construction**

##### **3.2.1 Digital Corpus**

The primary corpus for this study comprises 1,248 textual units collected between January 2023 and June 2024 from three active Algerian diaspora Facebook groups: *Algériens en France*, *Étudiants et cadres algériens en France*, and one additional expat community group whose exact name has been withheld for participant confidentiality. These groups were selected according to three key criteria: consistent activity levels (a minimum of fifty posts per week at the time of collection), a clear orientation toward Algerian diaspora experiences, and the regular presence of naturally occurring vernacular discourse involving *Darija*, French, and various mixed-code registers.

Facebook was chosen as the main source of data because it supports extended written interactions, enables sustained community-level identity negotiation over time, and has been widely recognized in the sociolinguistics literature as a significant platform for Arabic-speaking diasporas and code-switching practices (Androutsopoulos, 2013; Messaoudi, 2022). Additionally, data were collected using keyword-based purposive sampling, focusing on occurrences of *blad*, *ghorba*, and their morphological and orthographic variants (such as *bled*, *el blad*, *l-blad*, *el ghorba*, *ghourba*, and *la ghorba*). This sampling strategy was preferred over

random sampling because it allowed for a concentrated and analytically relevant dataset while remaining transparent and replicable. The resulting corpus includes not only original posts but also comment threads and replies, thereby capturing both individual expressions and the interactive processes through which meanings are negotiated and contested. Each textual unit was operationally defined as a discrete post or comment containing at least one target term, with a minimum of fifty surrounding tokens preserved to support detailed collocational and semiotic analysis.

### **3.2.2 Interview Data**

To complement and enrich the corpus analysis, ten semi-structured online interviews were conducted with Algerian migrants living in France. Each interview lasted between 45 and 70 minutes. The sample included five male and five female participants, ranging in age from 29 to 55 years, with a mean length of residency in France of 12.4 years. Participants came from diverse regional backgrounds in Algeria, held varied occupations, and represented different generational relationships to the migration experience. This provided a reasonable degree of sociodemographic variety for a qualitative study of this scale.

Recruitment was carried out through snowball sampling, beginning with personal contacts within the Facebook groups studied. While this approach facilitated access to individuals already engaged in the relevant discursive practices, it is important to acknowledge that purposive, network-based sampling does not produce a statistically representative sample. Moreover, interviews were conducted in the participants' preferred language, which in all cases involved natural code-switching between Darija and French, with occasional use of Modern Standard Arabic. This multilingual format was deliberately maintained rather than standardized, as the code-switching itself formed an important part of the data. It reflects the same hybrid identity practices observed in the digital corpus. The interviews focused on participants' personal narratives around *blad* and *ghorba*, exploring the affective meanings they associate with these terms, the contexts in which they use them, and how they understand their relationship to homeland and exile.

All interviews were recorded with explicit informed consent, transcribed verbatim using a modified version of Jefferson's (2004) transcription conventions, and preserved in their original multilingual form to maintain analytical fidelity.

## **3.3 Quantitative Corpus Analysis**

### **3.3.1 Pre-Processing and Normalization**

Corpus pre-processing was carried out using spaCy (version 3.7) alongside manually adjusted preprocessing procedures designed for Darija–French code-switched data. The corpus presented several linguistic challenges, particularly because Algerian Darija does not have a standardized written form and is often written using different spelling conventions across social media contexts. In many cases, participants alternated between Arabic script, Latin transliteration, and Franco-Arabic writing systems in which numbers are used to represent Arabic sounds, such as 3 for *ع* and 7 for *ح*. French lexical items were also frequently embedded within Darija syntactic structures, creating highly mixed linguistic patterns throughout the dataset.

To address these issues, the preprocessing stage included tokenization, text normalization, basic lemmatization, and stop-word filtering procedures informed by previous work in dialectal Arabic NLP (Habash, 2022; Ferrati, 2021). Because spelling variation was especially common in the corpus, orthographic variants of the target terms were manually standardized before analysis. For example, forms such as *bled*, *blad*, *l-blad*, and related spellings were grouped in order to ensure more consistent frequency counts and collocational analysis across the dataset.

### 3.3.2 Frequency and Collocation Analysis

Frequency counts were calculated for all normalized occurrences of *blad* and *ghorba*, including their orthographic variants identified during preprocessing. To allow comparison across different parts of the corpus, raw frequencies were normalized to occurrences per thousand tokens. This normalization helped reduce distortions caused by variation in post and comment length across the dataset.

Co-occurrence analysis was conducted using a symmetric  $\pm 5$ -token window around each target term. This window size was selected because it is commonly used in corpus linguistics research examining evaluative and emotionally charged language, while still providing enough contextual information to identify meaningful lexical associations (Brezina, 2018). Collocational strength was measured using normalized pointwise mutual information (NPMI) following Bouma (2009). NPMI was preferred over standard PMI because it reduces the tendency of PMI to overestimate rare word combinations, which is especially important in a multilingual and unevenly distributed corpus such as the present one. NPMI values range from  $-1$ , indicating complete avoidance of co-occurrence, to  $+1$ , indicating perfect co-occurrence. In this study, collocates with scores above 0.45 were treated as strongly associated with the target terms and were selected for further qualitative interpretation. The collocational findings were then examined alongside concordance lines and interview data to identify broader semantic and discursive patterns surrounding *blad* and *ghorba* within the corpus.

### 3.4 Qualitative Semiotic and Discourse Analysis

Qualitative analysis of the corpus data and interview transcripts combined semiotic analysis with critical discourse analysis to examine how *blad* and *ghorba* function within migrant identity discourse. The analysis was primarily informed by Barthes's concept of myth and Fairclough's three-dimensional model of discourse analysis. From a semiotic perspective, attention was given to the ways *blad* and *ghorba* moved beyond their literal meanings and became emotionally and culturally loaded expressions within migrant narratives. Rather than functioning only as references to homeland or exile, the terms frequently carried broader associations related to belonging, nostalgia, loss, identity, and social memory. The analysis therefore focused on how these meanings became normalized and emotionally reinforced through repeated patterns across both online discourse and interview narratives. Fairclough's framework was used to examine the data at three interconnected levels. The textual level focused on specific linguistic features such as lexical choice, pronoun use, modality, repetition, and sentence framing. The discursive practice level examined how meanings were produced and negotiated through Facebook interactions and personal storytelling. The social practice level considered the broader postcolonial and sociocultural context shaping the discourse,

including migration, marginalization, language ideology, and the maintenance of transnational cultural identity.

Interview data were analyzed using reflexive thematic analysis following Braun and Clarke (2019). Themes were developed inductively through repeated reading and coding of the interview transcripts rather than through predetermined categories. This approach allowed culturally specific expressions, metaphors, and code-switching practices to remain central to the analysis instead of being simplified into fixed analytical labels. After the initial themes were identified, they were compared with the collocational patterns emerging from the corpus analysis in order to identify points of convergence and divergence across the two datasets.

### **3.5 Triangulation and Analytical Validity**

To strengthen the reliability of the findings, the study combined corpus analysis with interview data through a process of methodological triangulation (Denzin, 2012). Patterns identified in the quantitative corpus analysis were compared with participants' interview narratives to examine whether similar meanings and associations appeared across both datasets. For example, when the corpus analysis showed strong associations between *blad* and themes such as family, memory, or belonging, the interview data were examined to determine how participants themselves described these experiences and emotional connections. The same process was applied to recurring patterns surrounding *ghorba*, particularly those related to loneliness, displacement, and emotional hardship.

### **3.6 Ethical Considerations**

This study followed established ethical guidelines for research involving online discourse and human participants. Ethical procedures were informed by the recommendations of the British Association for Applied Linguistics (BAAL, 2021) and the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679). Ethical approval was obtained from the relevant institutional review process before data collection began.

The Facebook data used in the study were collected only from publicly accessible groups and pages that could be viewed without private membership approval. In line with current ethical recommendations for internet-based research (Franzke et al., 2020), all identifying information was removed before analysis. Usernames, profile information, and other potentially identifying details were anonymized, and some quotations were slightly paraphrased where necessary to reduce the possibility of traceability through online search engines. All interview participants provided informed consent before participating in the study and were informed that participation was voluntary and that they could withdraw at any stage without consequence. Participants were also allowed to review their interview transcripts and request corrections or clarifications before analysis was finalized. Audio recordings and transcripts were stored securely on password-protected and encrypted devices, and the data will be deleted following the completion and publication of the research in accordance with data protection and minimization principles. No financial compensation was provided to participants.

## **4. Results**

### **4.1 Overview**

The findings are presented in two complementary sections: quantitative corpus findings and qualitative interview findings. The corpus analysis identifies the main linguistic and collocational patterns associated with *blad* and *ghorba* across the Facebook dataset, while the interview analysis provides deeper insight into how participants emotionally and culturally interpret these terms in relation to their lived experiences. Bringing these two forms of analysis together allows for a broader understanding of how *blad* and *ghorba* function in Algerian migrant discourse. The quantitative findings highlight recurring lexical and semantic patterns, whereas the qualitative data help explain the meanings, memories, and identity positions attached to those patterns. Across both datasets, a consistent contrast emerges between the two terms. *Blad* is generally associated with belonging, family, familiarity, and emotional attachment, while *ghorba* is more frequently connected to displacement, loneliness, distance, and emotional struggle. At the same time, participants often moved between these positions in fluid and sometimes contradictory ways, suggesting that the relationship between *blad* and *ghorba* is more complex than a simple binary opposition. Nevertheless, the findings indicate that the tension between these two terms plays a central role in how cultural identity is discussed and negotiated across both online and interpersonal migrant discourse.

## 4.2 Quantitative Corpus Findings

### 4.2.1 Corpus Overview and Frequency Distribution

After preprocessing and normalization, the corpus consisted of 1,248 textual units with a total of approximately 187,000 tokens. Across the dataset, *blad* and its orthographic variants (*bladi*, *bled*, *l-blad*, *el blad*) occurred 682 times, while *ghorba* and its related forms (*lghorba*, *el ghorba*, *ghourba*, *la ghorba*) appeared 566 times. The relatively close frequency distribution between the two terms suggests that both occupy an important position in discussions of migration, belonging, and identity within the corpus.

When normalized per thousand tokens, *blad* appeared at a rate of 3.65 occurrences and *ghorba* at 3.03 occurrences. Although the difference is not substantial, *blad* appeared slightly more frequently overall, particularly in posts centered on memory, family, nostalgia, and return. *Ghorba*, by contrast, appeared more often in discussions of loneliness, emotional difficulty, distance, and adaptation to life abroad. Moreover, code-switching between Darija and French was common throughout the corpus and appeared in 41% of *blad* instances and 37% of *ghorba* instances. This finding reflects the multilingual and translanguaging character of Algerian diaspora discourse on social media, where speakers frequently move between Darija and French within the same sentence or expression. Examples included phrases such as *fi blad* (“in the homeland”), *dans la ghorba* (“in exile”), and hybrid constructions such as *Bladna hiya l’authenticité* (“Our homeland is authenticity”). These mixed linguistic structures illustrate how participants use multiple linguistic resources simultaneously when discussing identity, belonging, and migration experiences.

### 4.2.2 Collocational Analysis: NPMI Results

Table 1 presents the strongest collocates associated with *blad* and *ghorba* based on NPMI scores above 0.45 and a minimum co-occurrence frequency of five instances within the  $\pm 5$ -token window described in Section 3.3.2.

**Table 1.** Strongest Collocates Associated with Blad and Ghorba

Term	Collocate	Co-occurrences	NPMI	Semantic Association
<b>Blad</b>	ahl (family/relatives)	214	0.68	Family and belonging
<b>Blad</b>	hnin (nostalgia/longing)	167	0.61	Memory and emotional attachment
<b>Blad</b>	al-Jaza'ir (Algeria)	112	0.52	National identity and origin
<b>Blad</b>	rju3 (return)	98	0.54	Desire for return
<b>Blad</b>	dyar (homes/homeland)	76	0.49	Familiarity and rootedness
<b>Ghorba</b>	waḥda (loneliness)	189	0.72	Isolation and emotional rupture
<b>Ghorba</b>	Fransa (France)	142	0.63	Migration and displacement
<b>Ghorba</b>	m3anah (suffering)	121	0.57	Hardship and emotional struggle
<b>Ghorba</b>	ghorba (intensified exile)	89	0.52	Reinforced sense of exile
<b>Ghorba</b>	flus (money)	67	0.48	Economic motivation

The collocational patterns surrounding *blad* were strongly connected to themes of family, nostalgia, homeland, and return. The strongest association was observed with *ahl* (“family” or “relatives”, NPMI = 0.68), suggesting that participants often discussed *blad* through relationships, kinship, and emotional attachment rather than through purely national or political references. Other frequent collocates, such as *hnin* (“longing”, NPMI = 0.61) and *rju3* (“return”, NPMI = 0.54), further reinforced the idea of *blad* as an emotionally valued and memory-centered space.

By contrast, the collocational profile of *ghorba* was more strongly associated with loneliness, hardship, and displacement. The strongest collocate, *waḥda* (“loneliness”, NPMI = 0.72), indicates that *ghorba* was commonly framed as an emotional condition rather than simply a geographical situation. The frequent association with *Fransa* (“France”, NPMI = 0.63) also shows that the term was often linked directly to migration experiences in the French context rather than to exile in a general sense. An additional pattern emerged through the collocate *flus* (“money”, NPMI = 0.48), which appeared alongside *ghorba* but not prominently with *blad*. This indicates that discussions of exile were shaped not only by emotional concerns but were also connected to economic realities and the financial motivations underlying migration. In many cases, participants frequently framed *ghorba* as simultaneously emotionally difficult and economically necessary, reflecting the ambivalent nature of migration experiences within the corpus.

#### 4.2.3 Semantic Category Distribution

All concordance lines containing *blad* and *ghorba* were examined and grouped into semantic categories that emerged directly from the corpus during the coding process. Through repeated

reading and comparison of contexts, four dominant categories were identified. Table 2 summarizes the distribution of these categories across the two target terms.

**Table 2.** Semantic Category Distribution Across Blad and Ghorba Instances

Semantic Category	Blad (%)	Ghorba (%)	Illustrative Example (translated)
Family and memory	43.8	9.2	“Blad is where my roots live, where my childhood still breathes.”
Emotional warmth and authenticity	29.4	11.7	“In blad the air feels different, real, not artificial.”
Alienation and suffering	7.2	48.5	“Ghorba is eating my soul day after day.”
Hybridity and negotiation	19.6	30.6	“We live between blad and ghorba, building something of our own.”

The category distribution reinforces the patterns already observed in the collocational analysis. Most instances of blad were connected to family, memory, emotional attachment, and feelings of authenticity or familiarity. References to blad frequently positioned the homeland as a space associated with childhood memories, emotional security, and social belonging. Negative associations appeared only occasionally in discussions of blad. In contrast, ghorba was much more strongly associated with alienation, emotional exhaustion, loneliness, and personal struggle. Many participants described ghorba as an ongoing emotional condition linked to distance from family, cultural displacement, and the pressures of migrant life.

At the same time, the findings also revealed a significant number of instances that did not fit neatly into a simple opposition between homeland and exile. The category labeled “hybridity and negotiation” included posts and comments in which participants described living between cultures, adapting to migration, or creating identities that combined elements of both Algeria and France. These examples suggest that while the contrast between blad and ghorba remains central in the discourse, many speakers also negotiate more fluid and hybrid understandings of belonging and identity. A chi-square test comparing the category distributions across the two terms showed statistically significant differences between blad and ghorba ( $\chi^2(3) = 412.7$ ,  $p < .001$ , Cramér’s  $V = 0.58$ ), indicating a strong relationship between the target terms and the semantic categories in which they appeared.

#### 4.2.4 Temporal and Modal Patterns

The analysis of tense, verbal aspect, and temporal markers revealed clear differences in how blad and ghorba were positioned across time within the corpus. Table 3 summarizes the main temporal and modal orientations associated with each term.

**Table 3.** Temporal and Modal Orientation of Blad and Ghorba Instances

Orientation Category	Blad (%)	Ghorba (%)	Illustrative Examples
Past-oriented (memory and nostalgia)	51.2	18.4	<i>kān fi blad</i> (“I was in the homeland”)
Future-oriented (return and aspiration)	22.7	14.9	<i>nrja3 lblad</i> (“I will return to the homeland”)
Present-oriented (ongoing experience)	26.1	66.7	<i>fi ghorba</i> (“in exile”)
Conditional and modal expressions	12.4	34.8	<i>lawla ghorba</i> (“if not for exile”)

Note: Some instances were coded in more than one category, so percentages may exceed 100%.

The temporal patterns associated with *blad* were largely oriented toward the past. More than half of the instances referred to memories, childhood experiences, family history, or earlier stages of life in Algeria. These references often used past tense verbs and nostalgic expressions, suggesting that *blad* was frequently constructed as a remembered and emotionally valued space rather than as an immediate lived reality. Future-oriented forms also appeared regularly, especially in statements expressing hope, imagined return, or long-term aspirations connected to the homeland. On the other hand, *ghorba* was much more strongly associated with the present. Most occurrences described exile as an ongoing emotional and social condition tied to everyday migrant life. Present-tense constructions were especially common in discussions of loneliness, work pressures, adaptation, and emotional fatigue.

Conditional and modal expressions also appeared more frequently with *ghorba*. Participants often used hypothetical or counterfactual statements to reflect on alternative life possibilities, personal sacrifices, or imagined futures outside migration. Expressions such as *lawla ghorba* (“if not for exile”) suggested a recurring tension between economic necessity and emotional belonging. Overall, these temporal patterns indicate that *blad* was commonly linked to memory, nostalgia, and imagined return, whereas *ghorba* was more closely connected to present experience, uncertainty, and the mediation of daily life abroad.

#### 4.2.5 Code-Switching and Multimodal Associations

Table 4 presents the main linguistic and multimodal elements that appeared alongside *blad* and *ghorba* in the corpus, including French lexical items, emoji usage, religious expressions, and insertions from Standard Arabic.

**Table 4.** Co-occurring Linguistic and Semiotic Elements in *Blad* and *Ghorba* Discourse

Co-occurring Element	Blad Frequency	Ghorba Frequency	Interpretive Pattern
French spatial terms (en France, ici, là-bas)	89	204	More strongly associated with <i>Ghorba</i> and the present location in France
Positive affect emoji (❤️, 🇩🇿, 😊)	156	41	More frequently in <i>blad</i> discourse
Melancholic emoji (😭, 📢, 📢, 📢)	34	98	More frequent in <i>ghorba</i> discourse
Religious expressions (Allah, mektoub, inshallah)	67	42	Present in both contexts
Standard Arabic insertions	52	38	Slightly more common with <i>blad</i>

The results show that the distinction between *blad* and *ghorba* was expressed not only through vocabulary, but also through emoji use, language choice, and other multimodal features commonly found in online communication. On one hand, positive emoji especially hearts and Algerian flag symbols appeared much more frequently in posts discussing *blad*. These visual markers were often used to express pride, affection, nostalgia, or emotional closeness to homeland and family. In comparison, melancholic emojis such as crying faces and broken hearts appeared more frequently alongside *ghorba*, reinforcing themes of loneliness, emotional exhaustion, and separation.

French spatial expressions such as *en France*, *ici*, and *là-bas* were strongly associated with *ghorba*. This suggests that discussions of exile were often grounded in the immediate reality of life in France rather than framed as an abstract idea of migration. Participants frequently linked *ghorba* to their present surroundings, daily routines, and experiences of adaptation abroad. Additionally, religious expressions such as *Allah*, *mektoub*, and *inshallah* appeared in both *blad* and *ghorba* contexts, although they were slightly more frequent in discussions of homeland. In many cases, these expressions seemed to function as emotional and spiritual resources through which participants framed separation, patience, hope, and belonging. The findings additionally revealed occasional shifts into Standard Arabic, particularly in emotionally charged or reflective posts. These shifts often appeared in moments where participants adopted a more formal, poetic, or culturally symbolic style of expression.

### 4.3 Qualitative Interview Findings

The interview data, consisting of approximately 9.5 hours of transcribed conversations, generated four main themes through reflexive thematic analysis. These themes largely supported the patterns identified in the corpus analysis while also providing more detailed insight into the emotional, cultural, and personal meanings participants attached to *blad* and *ghorba*.

#### 4.3.1 Theme 1: *Blad* as a Space of Authenticity, Memory, and Emotional Belonging

Across all interviews, participants described *blad* as more than a geographical place. It was often associated with familiarity, emotional comfort, childhood memory, family relationships, and a sense of being fully understood within one's social environment. Participants frequently spoke about their homeland through sensory memories, daily routines, and emotionally meaningful experiences rather than through political or national language alone.

One participant explained:

*“Blad is not a country on the map. It is the smell of my mother’s chorba in winter, the sound of the adhan in the morning, and people who know you without needing explanations. Even if things change there, in my mind it still feels pure.”*

This type of narrative appeared repeatedly throughout the interviews. Homeland was often remembered through sounds, smells, food, family interactions, and emotional familiarity. These descriptions closely reflected the corpus findings in which *blad* frequently co-occurred with terms related to nostalgia, family, and memory. Participants also tended to describe *blad* through an idealized perspective shaped by distance and memory. Several acknowledged that Algeria itself had changed over time, yet the emotional image of homeland remained stable in their narratives. In this sense, *blad* functioned less as a current physical reality and more as a symbolic and emotional reference point connected to identity, belonging, and continuity.

#### **4.3.2 Theme 2: Ghorba as Emotional Rupture and Everyday Struggle**

Across the interviews, *ghorba* was consistently described as an emotionally difficult and deeply personal experience rather than simply a condition of living abroad. Participants often connected *ghorba* to loneliness, separation from family, emotional exhaustion, and the feeling of living between two worlds without fully belonging to either.

One participant described the experience in the following way:

*“Ghorba is when you are alive, but something inside you disappears slowly. You work, you make money, you continue your life, but every Eid you miss, and every family moment that happens without you... leave a mark.”*

This kind of narrative closely reflected the patterns identified in the corpus analysis, particularly the strong associations between *ghorba*, loneliness, suffering, and emotional strain. Many participants described exile as an ongoing condition that accumulates over time through repeated absence from family events, social gatherings, and important life moments. The interviews also showed that discussions of *ghorba* often involved a tension between economic opportunity and emotional sacrifice. Several participants acknowledged that migration allowed them to work, support relatives, or improve their financial situation, while at the same time describing feelings of isolation and emotional loss. In this sense, *ghorba* was frequently presented as both necessary and painful.

Some gender-related differences also appeared across the interviews. Female participants tended to focus more on emotional absence, family relationships, missed ceremonies, and disrupted social ties, while male participants more frequently emphasized work, financial

pressure, and professional responsibility. Although the interview sample was too limited to draw broader conclusions, these patterns suggest that experiences of migration and exile may be shaped differently across social and gendered contexts.

### 4.3.3 Theme 3: Hybrid Identity and Negotiated Belonging

Most participants described their identities in ways that moved beyond a simple opposition between *blad* and *ghorba*. Rather than feeling fully attached to one place and disconnected from the other, many spoke about living between cultures, languages, and social worlds. These narratives reflected an ongoing process of reconstruction shaped by migration, adaptation, family history, and changing relationships with both Algeria and France.

One participant explained this experience as follows:

*“At work I feel French sometimes, at home I feel Algerian, but honestly, I do not feel completely one or the other. France gave me opportunities and a future, but there is still a distance inside me. And when I return to Algeria, it is not always the same Algeria I remember.”*

This sense of living “between” identities appeared repeatedly across the interviews and corresponded closely to the hybridity category identified in the corpus analysis. Participants often acknowledged emotional attachment to Algeria while also recognizing the role that France had played in shaping their education, careers, language practices, and everyday lives. Several interviewees also reflected critically on their own nostalgic image of *blad*. Some explained that returning to Algeria after many years abroad could feel unfamiliar or emotionally complicated because both the country and the individual had changed over time. In these cases, *blad* was no longer imagined as a fixed or untouched space, but as something shaped by memory, distance, and personal interpretation. These findings suggest that participants were not simply reproducing a rigid opposition between homeland and exile. Instead, many actively negotiated complex forms of belonging that combined attachment, ambivalence, adaptation, and emotional uncertainty.

### 4.3.4 Theme 4: Digital Versus Interpersonal Discourse – Platform-Specific Identity Performance

A recurring pattern across the interviews was the distinction participants made between what they shared publicly on Facebook and what they expressed in private conversations with family or close friends. Several participants explained that social media was often used to maintain collective connection, nostalgia, and community belonging, while more personal feelings related to loneliness or emotional suffering were usually kept private.

One participant described this difference in the following way:

*“On Facebook, we usually share the good side of life, the memories, food, music, and old photos. It helps people feel connected. But the difficult side of ghorba, the sadness and pressure, I do not really post that online. Those feelings stay for close people.”*

This distinction suggests that online discourse may encourage more collective and emotionally positive representations of the homeland, especially through shared memories, cultural symbols, humor, and nostalgic content. In contrast, the more painful or emotionally vulnerable aspects of migration were often described as private experiences discussed mainly within trusted interpersonal relationships. These findings also help explain some of the patterns observed in the corpus analysis. Public Facebook discourse appeared to emphasize themes associated with *blad*, such as belonging, warmth, nostalgia, and cultural pride, while the more emotionally complex dimensions of *ghorba* were less visible in public online interactions. As a result, relying only on social media data could produce a partial picture of migrant experience by overrepresenting communal nostalgia and underrepresenting emotional struggle, ambivalence, and isolation.

#### **4.4 Triangulation: Convergence and Complementarity**

The comparison between the corpus findings and the interview data revealed both strong convergence and important points of complementarity between the two forms of analysis. Across both datasets, similar semantic and emotional patterns consistently emerged. In the corpus analysis, *blad* was strongly associated with family, memory, nostalgia, and belonging, while *ghorba* was more frequently linked to loneliness, suffering, and emotional separation. These same patterns appeared repeatedly in the interview narratives, where participants described homeland through memories, family relationships, and emotional attachment, and described exile through isolation, sacrifice, and everyday emotional pressure.

The temporal patterns identified in the corpus were also reflected in the interviews. Discussions of *blad* were often framed through memories of the past, childhood experiences, and imagined return, whereas discussions of *ghorba* focused more on present-day realities and ongoing emotional struggles. This consistency across the two datasets strengthened the interpretation that the opposition between *blad* and *ghorba* plays a meaningful role in how participants understand and describe migration, belonging, and identity. At the same time, the interview data added dimensions that could not be fully captured through corpus analysis alone. Participants reflected on the emotional complexity of migration, the changing meaning of homeland over time, gendered experiences of exile, and the difference between public online expression and private emotional experience. These themes provided important context for interpreting the computational findings and helped explain why certain patterns appeared more strongly in Facebook discourse than in personal narratives.

### **5. Discussion**

#### **5.1 The Blad/Ghorba Binary as a Semiotic Structure**

Across both the corpus and interview data, *blad* and *ghorba* emerged as emotionally saturated categories through which Algerian migrants interpreted mobility, attachment, and displacement. *Blad* was consistently associated with family, nostalgia, familiarity, and cultural continuity, whereas *ghorba* was more frequently linked to loneliness, rupture, uncertainty, and emotional strain. These patterns support research suggesting that migrant discourse is often organized through affective spatial imaginaries in which language, memory, and belonging become closely interconnected (Leurs & Ponzanesi, 2018; Canagarajah, 2021). In this context,

references to the homeland functioned less as neutral geographical descriptions than as symbolic reconstructions shaped by emotional proximity and selective remembrance.

The interviews reinforced this pattern by showing how participants repeatedly described Algeria through sensory and relational memories connected to food, sound, family life, and childhood experience. The homeland therefore appeared not simply as a physical location, but as an affective space sustained through memory and emotional continuity. This aligns with Georgiou's (2019) analysis of how diaspora communities construct affective attachments to homeland through digital communication and transnational belonging, as well as with Fuller's (2024) examination of how migration discourse is shaped by emotional and cultural dimensions that extend beyond geographical reference. The collocational association between *blad* and kinship-related vocabulary further reflects the extent to which homeland discourse was grounded in emotional continuity rather than material presence. By contrast, *ghorba* was often articulated as an ongoing condition shaped by adaptation, distance, and social fragmentation. The predominance of present-oriented constructions suggests that exile was experienced less as a completed transition than as a continuing process of negotiation and adjustment, a pattern similarly identified in research describing migration as a durational and embodied condition (Busch, 2021).

At the same time, the findings complicate any rigid separation between homeland and exile. Participants frequently moved between attachment to Algeria and adaptation to life in France, producing forms of identification that were fluid, reflexive, and context-dependent. These shifting positions resonate with translanguaging and diaspora research emphasizing the instability of fixed national identities within multilingual and transnational settings (Li Wei, 2018; Zhu & Li, 2022). Several participants openly acknowledged that their understanding of *blad* was shaped as much by nostalgia as by lived experience, while others described feeling emotionally distant from both France and Algeria at different moments in their lives. Rather than functioning as a fixed binary, the *blad/ghorba* opposition operated as a flexible semiotic framework through which participants interpreted movement, memory, and self-positioning across changing social realities.

## **5.2 Postcolonial Dimensions: Structural Inequality and Mythologized Belonging**

The meanings attached to *blad* and *ghorba* were also shaped by broader postcolonial conditions linked to migration, language, and social inequality. The strong collocational association between *ghorba* and references to France indicates that exile was not constructed as an abstract emotional condition, but as one embedded within the historical and social realities of Algerian migration to France. In this sense, the discourse surrounding *ghorba* reflected not only personal displacement, but also unequal relations of linguistic legitimacy, cultural recognition, and institutional belonging patterns similarly identified in research on postcolonial migration and diaspora identity in France (Dovchin, 2020; Canagarajah, 2021).

The Facebook groups examined in this study also functioned as spaces of cultural continuity in which participants collectively maintained forms of Algerian identity through *Darija*, code-switching, humor, music, food references, and nostalgic imagery. These practices extended beyond casual online interaction and became mechanisms for sustaining shared cultural memory across distance, consistent with research on digital diaspora communities and

transnational identity performance (Leurs & Ponzanesi, 2018; Georgiou, 2019). Within this context, the continued use of Darija alongside French carried an additional symbolic dimension. Multilingual practices frequently operated as assertions of linguistic legitimacy within environments where vernacular migrant speech is often socially marginalized. As Dovchin (2020) and Zhu and Li (2022) argue, multilingual speakers regularly negotiate tensions between dominant monolingual expectations and everyday translanguaging practices. The use of Darija in these digital spaces can therefore be understood not only as communication, but also as an affirmation of cultural continuity and vernacular identity.

The corpus findings additionally revealed a persistent economic dimension within discussions of ghorba. Terms related to money and work frequently co-occurred with exile discourse, indicating that migration was often framed simultaneously as emotional sacrifice and material necessity. Participants repeatedly described life abroad as financially stabilizing yet emotionally exhausting, reflecting tensions between economic mobility and affective displacement that have been widely documented in studies of postcolonial and transnational migration (Busch, 2021). Religious expressions such as Allah, mektoub, and inshallah further complicated this discourse by functioning as symbolic resources through which uncertainty, endurance, and separation were interpreted. Their recurrence across both blad and ghorba contexts suggests that spiritual language remained closely connected to emotional coping and cultural continuity throughout the migration experience.

### **5.3 Methodological Contribution: NLP-Assisted Analysis and Its Interpretive Limits**

One of the central methodological contributions of this study lies in its integration of NLP-assisted corpus analysis with qualitative discourse interpretation. NLP-based preprocessing and collocational analysis made it possible to examine large quantities of Darija–French code-switched data despite substantial orthographic variation and transliteration inconsistency, challenges that remain significant within dialectal Arabic NLP research (Ferrati, 2021; Habash, 2022). Corpus analysis identified recurring semantic and temporal patterns surrounding blad and ghorba, particularly in relation to kinship, nostalgia, loneliness, and economic pressure, providing an empirical foundation for subsequent qualitative interpretation.

At the same time, the findings demonstrate the limits of computational analysis when separated from contextual interpretation. Statistical associations alone could not account for the emotional and ideological meanings attached to exile, belonging, or memory. While the corpus revealed a strong relationship between ghorba and loneliness, it did not explain how participants experienced displacement in everyday life or how these meanings were shaped by broader postcolonial conditions. These dimensions became visible primarily through the interviews, which revealed forms of ambivalence, vulnerability, and emotional negotiation that were less apparent in public Facebook discourse.

Platform context also shaped the discourse itself. Facebook interactions tended to foreground nostalgia, collective humor, and affiliative identity performance, whereas more emotionally difficult experiences were often confined to private conversations and interpersonal relationships. Without the qualitative component, the analysis would likely have produced a simplified and excessively nostalgic account of migrant identity discourse. The findings therefore support approaches in discourse studies and migration linguistics that position

computational methods as interpretive support rather than analytical replacement, as Fuller (2024) demonstrates in her review of corpus and critical discourse approaches to migration, where computational pattern detection gains analytical value only when combined with contextual and cultural interpretation. NLP-assisted analysis can identify large-scale linguistic patterns across multilingual datasets, but understanding the social and emotional significance of those patterns still requires contextual interpretation grounded in lived experience and social theory.

#### **5.4 Gendered Dimensions of Blad and Ghorba Discourse**

The interviews also revealed gender differences in how participants discussed blad and ghorba. Female participants more frequently framed exile through emotional absence, caregiving responsibilities, family separation, and disrupted social intimacy, while male participants tended to foreground work, financial pressure, and professional obligation. Although these distinctions emerged from a relatively limited sample, they nevertheless suggest that migration and belonging are mediated through gendered social expectations and forms of affective labor.

#### **5.5 Theoretical Contribution: Toward a Cultural Semiotics of Diaspora Lexis**

Beyond its empirical findings, the study proposes a broader framework for understanding how culturally specific lexical items function within diaspora discourse. The findings suggest that blad and ghorba operate not merely as descriptive vocabulary, but as semiotic structures through which migrants interpret displacement, continuity, attachment, and memory. Their meanings extended beyond direct reference to homeland and exile and instead functioned as culturally embedded categories shaped by emotional experience, historical positioning, and collective narrative.

This perspective brings together insights from diaspora studies, semiotics, and critical discourse analysis to examine how migrant subjectivities are constructed through language and symbolic meaning rather than through fixed national categories alone. The findings reinforce Hall's (1990) argument that diasporic identity is continuously produced through representation, memory, and narration, while also supporting research emphasizing the role of emotionally charged discourse in sustaining transnational forms of belonging across digital and multilingual environments (Georgiou, 2019; Canagarajah, 2021). Within this framework, blad and ghorba functioned as interpretive resources through which participants organized experiences of mobility, separation, and cultural continuity across everyday discourse.

#### **5.6 Limitations and Directions for Future Research**

Several limitations should nevertheless be acknowledged. The corpus was restricted to three Facebook groups within the Algerian diaspora community in France, meaning that the findings reflect a specific digital and sociocultural context rather than migrant discourse more generally. Identity performance and emotional expression may differ substantially across other platforms, including WhatsApp, TikTok, YouTube, or spoken everyday interaction, where communicative norms and audience expectations are structured differently. Research on digital migration discourse has repeatedly shown that platform affordances influence both the linguistic and affective dimensions of online identity construction (Leurs & Ponzanesi, 2018; Georgiou, 2019).

The interview sample was also relatively small and exploratory in nature, limiting broader claims regarding gender, generation, class, regional background, or migration history. The gendered patterns identified here should therefore be interpreted cautiously and examined further through larger and more demographically diverse participant groups. In addition, the preprocessing of Darija–French code-switched discourse required normalization decisions concerning transliteration and orthographic variation, reflecting broader methodological challenges within dialectal Arabic NLP and computational sociolinguistics, particularly given the absence of a fully standardized written form for Darija (Habash, 2022; Ferrati, 2021).

## **6. Conclusion**

This study has shown that *blad* and *ghorba* function as more than lexical references to homeland and exile within Algerian migrant discourse. Across both the corpus data and the interview narratives, the two terms emerged as culturally and emotionally saturated frameworks through which participants articulated memory, distance, attachment, and everyday experiences of migration. While *blad* was consistently associated with family, familiarity, and retrospective forms of belonging, *ghorba* was more frequently connected to social isolation, adaptation, and the affective pressures of migrant life in France. The findings nevertheless complicate any rigid opposition between the two terms. Participants often described forms of attachment shaped simultaneously by mobility, multilingualism, and long-term cultural negotiation, reflecting the unstable and processual character of diasporic identity formation.

The study also highlights the interpretive value of combining NLP-assisted corpus analysis with qualitative discourse analysis in the study of multilingual migration discourse. Computational methods made it possible to identify recurring collocational and semantic patterns across a linguistically variable dataset, but the broader ideological and emotional meanings attached to those patterns became visible only through qualitative interpretation. In this sense, the findings support corpus-assisted approaches that integrate computational scalability with contextual and culturally grounded analysis. More broadly, the study suggests that culturally embedded lexical items can operate as semiotic structures through which migrants interpret displacement, continuity, and belonging across transnational contexts. Rather than functioning as peripheral expressions within everyday discourse, *blad* and *ghorba* emerged as central symbolic resources in the ongoing construction of postcolonial migrant subjectivity.

## References

- Abouyassine, B. (2023). Theorising diaspora: Nationalist discourse in shaping diasporic subjects. *International Journal of Language and Literary Studies*, 5(4), 269–297. <https://doi.org/10.36892/ijlls.v5i4.1493>
- Adamson, F. B., Chung, E. A., & Hollifield, J. F. (2024). Rethinking the migration state: Historicising, decolonising, and disaggregating. *Journal of Ethnic and Migration Studies*, 50(3), 559–577. <https://doi.org/10.1080/1369183X.2023.2269769>
- Ait El Caid, H., Oukhouya, H., & El Himdi, K. (2025). Through the lens of immigrants: Maghrebi immigrants' assessment of integration policy in France and Spain. *Journal of Immigrant & Refugee Studies*, 1–17. <https://doi.org/10.1080/15562948.2025.2589423>
- Androutopoulos, J. (2013). Code-switching in computer-mediated communication. In Herring, S. C., Stein, D., & Virtanen, T. (Eds.), *Pragmatics of computer-mediated communication* (pp. 667–694). De Gruyter Mouton.
- Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*.
- Barthes, R. (1972). *Mythologies* (A. Lavers, Trans.). Hill and Wang.
- Bayramoglu, Y. (2023). Border countervisuality: Smartphone videos of border crossing and migration. *Media, Culture & Society*, 45(3), 595–611. <https://doi.org/10.1177/01634437221136013>
- Bekkaoui, T. (2024). Crossing borders, shifting identities: The transnational journey of Wafa Faith Hallam from Morocco to America. *International Journal of Language and Literary Studies*, 6(4), 216–229. <https://doi.org/10.36892/ijlls.v6i4.1864>
- Bhabha, H. K. (1994). *The location of culture*. Routledge.
- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the Biennial GSCL Conference 2009* (pp. 31–40). Gesellschaft für Sprachtechnologie und Computerlinguistik.
- Bourdieu, P. (1991). *Language and symbolic power*. Harvard University Press.
- Bouyazdouzen, Y. (2025). Voicing the Mediterranean: Home, identity, and diaspora in Abdulhamid Bajouki's *Uyūn al-Manfā Moro Jaime*. *International Journal of Language and Literary Studies*, 7(2), 379–403. <https://doi.org/10.36892/ijlls.v7i2.2057>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. <https://doi.org/10.1017/9781316410899>

- British Association for Applied Linguistics. (2021). Recommendations on good practice in applied linguistics (4th ed.). BAAL. <https://www.baal.org.uk/wp-content/uploads/2021/08/BAAL-Good-Practice-Guidelines-2021.pdf>
- Brookes, G., & McEnery, T. (2020). *Corpus linguistics and 21st-century media discourse*. Routledge. <https://doi.org/10.4324/9780429437021>
- Busch, B. (2021). The language portrait in multilingualism research: Theoretical and methodological considerations. De Gruyter. <https://doi.org/10.1515/9781501511295>
- Canagarajah, S. (2021). Translanguaging as spatial repertoires: Expanding the paradigm beyond structuralist orientations. *Applied Linguistics*, 42(4), 607–629. <https://doi.org/10.1093/applin/amab006>
- Caubet, D. (2004). *Les mots du bled: Création contemporaine en arabe maghrébin*. L'Harmattan.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
- De Fina, A. (2020). Narrative and identities. In A. De Fina & A. Georgakopoulou (Eds.), *The Cambridge handbook of discourse studies* (pp. 349–370). Cambridge University Press. <https://doi.org/10.1017/9781108348195.017>
- Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 6(2), 80–88. <https://doi.org/10.1177/1558689812437186>
- Dovchin, S. (2020). The psychological damages of linguistic racism and international students in Australia. *International Journal of Bilingual Education and Bilingualism*, 23(7), 804–818. <https://doi.org/10.1080/13670050.2020.1759504>
- Fairclough, N. (1992). *Discourse and social change*. Polity Press.
- Fairclough, N. (1995). *Media discourse*. Edward Arnold.
- Ferrati, A. (2021). Global media and cultural identity: Opportunities and challenges for Morocco in the digital era. *International Journal of Language and Literary Studies*, 3(3), 109–120. <https://doi.org/10.36892/ijlls.v2i3.673>
- Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., & Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>
- Fuller, J. M. (2024). Media discourses of migration: A focus on Europe. *Language and Linguistics Compass*, 18(3), e12526. <https://doi.org/10.1111/lnc3.12526>
- Georgiou, M. (2019). City of refuge or digital order? Refugee recognition and the digital governmentality of migration in the city. *Television & New Media*, 20(6), 600–616. <https://doi.org/10.1177/1527476419857683>

- Habash, N. (2022). Introduction to Arabic natural language processing (2nd ed.). Springer. <https://doi.org/10.1007/978-3-031-02193-2>
- Hall, S. (1990). Cultural identity and diaspora. In J. Rutherford (Ed.), *Identity: Community, culture, difference* (pp. 222–237). Lawrence & Wishart.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). John Benjamins. <https://doi.org/10.1075/pbns.125.02jef>
- Kianpour, M., Triandafyllidou, A., Allen, T., Mazrouei, S., & Shams, M. (2025). Social media narratives, diasporic identity and collective memory: A critical synthesis of the literature. *Ethnicities*, 26(1), 36–62. <https://doi.org/10.1177/14687968251386072>
- Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, 4, 151–171. <https://doi.org/10.1146/annurev-linguistics-030514-125254>
- Leurs, K., & Ponzanesi, S. (2018). Connected migrants: Encapsulation and cosmopolitanization. *Popular Communication*, 16(1), 4–20. <https://doi.org/10.1080/15405702.2017.1412777>
- Li Wei. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, 39(1), 9–30. <https://doi.org/10.1093/applin/amx039>
- Mautner, G. (2009). Corpora and critical discourse analysis. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 32–46). Continuum.
- Mendoza-Collazos, J., & Zlatev, J. (2022). A cognitive-semiotic approach to agency: Assessing ideas from cognitive science and neuroscience. *Biosemiotics*, 15(1), 141–170. <https://doi.org/10.1007/s12304-022-09473-z>
- Messaoudi, L. (2022). Sociolinguistics in the Maghreb: Diversity, contact, and change. In E. Benmamoun & R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics* (2nd ed., pp. 378–394). Routledge.
- Pavlenko, A. (2006). Bilingual minds: Emotional experience, expression, and representation. *Multilingual Matters*. <https://doi.org/10.21832/9781853598746>
- Sayad, A. (2000). *La double absence: Des illusions de l'émigré aux souffrances de l'immigré*. Seuil.
- Sayad, A. (2004). *The suffering of the immigrant*. Polity Press.
- Silverstein, P. A. (2004). *Algeria in France: Transpolitics, race, and nation*. Indiana University Press.
- Smets, K. (2018). The way Syrian refugees in Turkey use media: Understanding “connected refugees” through a non-media-centric and local approach. *Communications*, 43(1), 113–123. <https://doi.org/10.1515/commun-2017-0041>

- Wierzbicka, A. (1997). Understanding cultures through their key words: English, Russian, Polish, German, and Japanese. Oxford University Press.
- Zappavigna, M. (2018). Searchable talk: The linguistic functions of hashtags. Bloomsbury Academic.
- Zhu, H., & Li Wei. (2022). Translanguaging in performance or performance in translanguaging. *Journal of Multilingual and Multicultural Development*, 47(1), 1–14. <https://doi.org/10.1080/01434632.2022.2066109>
- Zrizi, H., Ait Idir, L., & Aouaki, S. (Eds.). (2023). Diasporic voices in contemporary Maghrebi literature: (Re)membering, forgetting, and identity transfer. Multidisciplinary Research Laboratory in Humanities and Social Sciences, Faculty of Arts and Humanities, Hassan II University Casablanca.

### **Author's Biography:**

***Khansaa Mohammed Cherif*** is a doctoral researcher in multilingualism and psycholinguistics at the University of Pannonia, Hungary. Her research interests include multilingual studies, language identity, bilingual cognition, sociolinguistics, psycholinguistics, and the role of multilingualism across cognitive, cultural, and interdisciplinary contexts.