Revista Infoacceso
e-ISSN 2311-7605 (Online)
infoacceso.org/index.php/ri

Volume 13 · Issue 3 · 2026
Pages 2719–2725

Executing Identity in Digital Discourse: A Conceptual Structure for Sociolinguistic Research

Sarra BENMAROUF[1], Hayat BAGUI[1], Yasser YAHIAOUI[2]

[1] Department of Foreign Languages, Faculty of Literature and Languages, Training Engineering and Project Design (IFMP) Laboratory, University of Naama- Algeria

[2] Department of computer sciences, Faculty of exact sciences and applied sciences, Laboratory for the Development of Arabic Language Didactics in Algerian Universities, University of Naama- Algeria

Email: 1benmarouf.sarra@cuniv-naama.dz, 2bagui@cuniv-naama.dz, 3yahiaoui@cuniv-naama.dz

Received: 02/08/2026 | **Accepted:** 03/10/2026 | **Published:** 16/10/2026

***Abstract*—**This paper develops a sociolinguistic framework for analysing identity enactment in digital communication, arguing that digital identity is not a static trait transferred online but becomes recognizable through recurrent linguistic, multimodal, interactional, and platform-conditioned practices. It consolidates four analytical dimensions: linguistic styling, multimodal self-presentation, interactional positioning, and platformed visibility, connected to established perspectives on identity-in-interaction, performativity, audience design, indexicality, multimodality, and platform society. The contribution is conceptual and methodological; it specifies a bounded-corpus research design, a coding framework, procedures for distinguishing self-identification from audience attribution and analyst inference, and safeguards against overclaiming platform causation. Particular attention is given to multilingual digital discourse, where dialect choice, code-switching, orthography, emoji, and interactional stance jointly construct social positions. Identity meaning depends on context, audience, uptake, and platform conditions rather than form alone. The framework is intended as a reproducible basis for future empirical sociolinguistic research rather than a substitute for data.

***Keywords*—**digital identity; identity performance; sociolinguistics; computer-mediated communication; platform affordances; multilingualism

1. Introduction

Digital identity is often described as self-presentation, but this description is incomplete. In networked communication, people do not simply display identities they already possess. They select linguistic styles, images, symbols, interactional stances, and platform-specific practices through which identities become recognizable to audiences. At the same time, these performances occur within technical environments that regulate visibility, persistence, circulation, and audience composition.

The supplied paper proposes a four-dimensional model comprising linguistic styling, multimodal self-presentation, interactional positioning, and platformed visibility. This article develops that proposal into a methodological framework for sociolinguistic research. It is conceptual rather than empirical: it does not claim to have collected a dataset or measured algorithmic effects. Its purpose is to specify how future studies can operationalize digital identity as an enacted, relational, multimodal, and platform-conditioned phenomenon.

The motivation for this framework arises from a recurring gap in existing research: many studies of digital communication treat identity either as a purely linguistic phenomenon, analysed through code-switching or stylistic variation alone, or as a purely technical phenomenon, analysed through engagement metrics and platform architecture alone. Neither approach, taken in isolation, can explain how a single post, comment, or profile becomes socially meaningful as an identity performance. The framework developed here treats language as the

© 2026 The Author(s). Published by Revista Infoacceso.

central resource for identity work while insisting that language cannot be interpreted apart from the multimodal, interactional, and platform conditions in which it occurs.

2. Why a Sociolinguistic Model of Digital Identity?

Identity research has moved from essentialist models toward interactional, performative, and relational perspectives. Bucholtz and Hall (2005) argue that identity is produced through linguistic and semiotic practices and can involve macro-categories, local stances, and interactionally emergent positions. This is particularly relevant online because users can combine dialect, code-switching, spelling, punctuation, emoji, hashtags, images, and platform conventions within a single communicative act.

Digital contexts also complicate the concept of audience. Goffman's dramaturgical model remains useful as a metaphor for self-presentation, but online contexts differ from face-to-face encounters because posts can persist, be copied, become searchable, and reach unintended audiences. Bell's audience-design perspective is therefore relevant but must be extended to account for imagined, networked, and technically mediated audiences.

2.1 From essentialism to indexicality: a necessary theoretical shift

Earlier sociolinguistic work on identity often proceeded from a correlational logic: a given linguistic variable was assumed to correspond to a stable social category, such as gender, region, or class, independently of the communicative situation in which it occurred. This correlational logic is poorly suited to digital discourse, where the same lexical item, script, or emoji can carry sharply different social meanings depending on co-text, addressee, platform, and genre. The indexical-order framework proposed by Silverstein (2003) offers a more adequate alternative: social meaning is treated as emergent, layered, and context-dependent rather than fixed. A first-order indexical association, for example between a dialect feature and a regional origin, can be reinterpreted at a second order as an index of authenticity, humour, or resistance, depending on how it is deployed and received.

This shift has direct implications for digital identity research. It implies that researchers cannot pre-assign identity meanings to particular languages, scripts, or platforms before examining how these resources are actually deployed in specific communicative events. It also implies that the same multilingual practice, such as switching from Modern Standard Arabic to a regional dialect within a single post, may construct different identity positions across different platforms, audiences, or historical moments. The framework developed in this paper is built on this indexical premise throughout.

3. The Four-Dimensional Model

3.1 Linguistic styling

Linguistic styling includes language choice, code-switching, dialect selection, orthographic variation, punctuation, lexical innovation, register, and metalinguistic commentary. These resources may index social meanings, but the meaning of a feature cannot be assumed independently of context. Silverstein's (2003) indexical-order perspective encourages researchers to examine how forms acquire and shift social meanings.

3.2 Multimodal self-presentation

Digital identity is rarely linguistic alone. Images, video, typography, emojis, hashtags, music, filters, gestures, and spatial arrangement contribute to how a user is perceived. A social-semiotic approach treats modes as distinct resources with different affordances rather than as decorative additions to language (Kress, 2015). The unit of analysis should therefore be multimodal where the research question concerns identity performance.

3.3 Interactional positioning

Identity is also constructed through relations with others. Alignment, disagreement, stance, evaluation, address terms, replies, mentions, and community participation can position a user as expert, novice, insider, outsider, ally, critic, supporter, or other social figure. This dimension is particularly important because self-presentation is not identical to identity attribution: other users can accept, reject, reinterpret, or contest a projected identity.

3.4 Platformed visibility

Platform architecture shapes what can be seen, by whom, for how long, and with what possibilities for circulation. Van Dijck, Poell, and de Waal (2018) conceptualize platforms as infrastructures that have become deeply embedded in social life. Duffy, Pruchniewska, and Scolere (2017) show how imagined platform affordances can shape self-branding practices. The model therefore treats visibility as a sociotechnical condition, not merely a numerical outcome.

4. Research Questions

The model can be operationalized through five research questions. RQ1: How are linguistic resources selected and styled in digital identity performances? RQ2: How do multimodal resources combine with linguistic choices? RQ3: How is identity positioned relationally through interaction? RQ4: How do platform features condition visibility and audience relations? RQ5: How do linguistic, multimodal, interactional, and platform dimensions interact rather than operate independently?

These questions deliberately avoid assuming that platforms determine identity. They treat platform features as conditions that shape possibilities and constraints. Similarly, they do not assume that a user's stated identity is the only valid identity outcome. Reception and uptake remain analytically important.

5. Methodological Design

A strong empirical implementation should use a bounded digital corpus. Sampling must specify platform, communicative genre, period, language(s), public/private status, and inclusion criteria. The corpus might contain public profiles, posts, comments, stories, short-form videos, or discussion threads, but researchers should not mix these units without justification.

Each item should be assigned metadata describing date, platform, genre, interactional position, and multimodal composition. For linguistic data, the researcher should preserve original orthography while creating a normalized analytical transcription only where necessary. This is particularly important for multilingual digital discourse because non-standard spelling can itself be an identity resource.

Ethical procedures should be defined before data collection. Public availability does not automatically mean that users expect their words to be reproduced in research. Researchers should minimize personally identifying details, consider paraphrasing searchable material, and follow institutional ethics requirements and platform terms where applicable.

5.1 Specifying corpus boundaries

Because the model is designed to travel across empirical contexts, it is necessary to specify in advance which boundary decisions a given study will make. Table 1 below proposes six boundary parameters that any implementation of the model should declare explicitly before data collection begins. These parameters function as a minimal transparency checklist rather than as a rigid protocol, since different research questions will justify different combinations of platform, genre, and language scope.

Table N°1: Corpus Boundary Parameters for Digital Identity Research

Parameter	Guiding question	Illustrative decision
Platform	Which platform(s) and why?	Single platform for internal comparability; multiple platforms only with a comparative design
Genre	Which communicative genre(s)?	Posts, comments, stories, or threads, not mixed without justification
Period	What time window?	A bounded period long enough to observe repetition, short enough to remain manageable
Language(s)	Which languages/varieties are included?	All languages present in the corpus, not pre-selected by assumed relevance
Public/private status	Is the material public or accessed by consent?	Public material handled per platform terms; private material requires consent
Inclusion criteria	What makes an item eligible?	Presence of identity-relevant linguistic, multimodal, or interactional features

Source: Author's own elaboration (2026)

5.2 From raw data to analytic record

Once boundaries are declared, each eligible item should be converted into an analytic record containing the original text or image, a preserved-orthography transcription, a normalized transcription where relevant, and metadata fields for date, platform, genre, and multimodal composition. Maintaining both the original and normalized forms side by side allows the researcher to return to the raw material at any later analytic stage without losing information that may itself be identity-relevant, such as expressive lengthening, mixed scripts, or non-standard punctuation.

6. Coding Framework

Table N°2: Coding Framework for Digital Identity Dimensions

Dimension	Primary codes	Evidence
Linguistic styling	language choice, dialect, code-switching, orthography, register	words, clauses, script, spelling, punctuation
Multimodal presentation	image, video, typography, emoji, hashtag, music	visual and audiovisual composition
Interactional positioning	stance, alignment, evaluation, address, uptake	posts, replies, mentions, comments
Platformed visibility	persistence, shareability, metrics, recommendation, audience structure	platform affordances and observable circulation
Identity outcome	self-claim, attributed identity, contested identity	explicit labels, interactional uptake, analyst interpretation

Source: Author's own elaboration (2026)

7. Illustrative Application: A Hypothetical Worked Example

To clarify how the model would function in practice, this section presents a constructed, hypothetical example. It is explicitly illustrative rather than empirical: no real user, post, or dataset is described, and no claims are made about actual Algerian digital discourse. Its purpose is solely to show how the six-stage analytical procedure introduced in Section 8 would be applied to a single communicative event.

Consider a hypothetical Instagram caption in which a young Algerian content creator writes a caption combining Darja, French, and English, accompanied by a filtered selfie, three emoji, and a trending hashtag, posted during a recognisable cultural event. Stage 1 would record the platform, genre, and date of this event as metadata. Stage 2 would identify the linguistic styling: the matrix language of the caption, the embedded French and English items, and any orthographic choices such as Latin-script transliteration of Darja. Stage 3 would analyse the multimodal composition: the filter style, the framing of the selfie, and the relationship between the image and the caption's verbal content.

Stage 4 would examine interactional positioning by reviewing how commenters responded: whether they replied in the same mixture of languages, whether they treated the post as humorous or aspirational, and whether any commenters contested the identity being projected. Stage 5 would document platform conditions such as whether the post appeared on a public or restricted account, whether it used a trending audio or hashtag associated with a particular imagined community, and what visibility affordances were available to the poster. Stage 6 would then integrate these observations: the analyst would ask whether the trilingual caption, the visual styling, and the audience uptake jointly constructed a recognisable identity position, for example a cosmopolitan, trend-aware, and locally grounded youth identity, rather than treating any single feature, such as the presence of English, as sufficient evidence of that identity on its own.

This hypothetical walkthrough illustrates two methodological commitments central to the model. First, no single dimension is treated as determinative: the linguistic mixture alone would not license a claim about identity without corroborating evidence from multimodal composition and interactional uptake. Second, the analyst's interpretation at Stage 6 is explicitly distinguished from the poster's self-identification and from commenters' attributed identities, following the three-source distinction introduced in Section 8.

8. Analytical Procedure

Stage 1 establishes the communicative event and its metadata. Stage 2 identifies linguistic styling without assigning identity meanings. Stage 3 analyses multimodal composition. Stage 4 examines interactional relations and audience uptake. Stage 5 documents platform conditions. Stage 6 integrates the dimensions to explain the identity position that becomes socially available.

The integration stage is essential. A language choice may index one identity only because it co-occurs with a particular image, stance, audience, and platform context. For example, a local dialect can signal intimacy in one setting but irony in another. English terminology can index expertise in one community and parody in another. The model therefore rejects one-to-one mappings between forms and identities.

Where identity attribution is coded, the researcher should distinguish three sources: self-identification, audience attribution, and analyst inference. These should never be collapsed into a single category. A user can self-identify as one thing while commenters position the user differently. Such divergence is itself valuable data for studying identity negotiation.

9. Identity, Performativity, and Repetition

Butler's performativity framework emphasizes repetition in the production of gendered identity categories. Applied more broadly to digital identity, the useful methodological insight is that identity becomes recognizable through repeated practices rather than a single isolated act. Researchers should therefore look for patterned recurrence: repeated linguistic styles, recurring visual motifs, stable interactional stances, or repeated community affiliations.

However, performativity should not be interpreted as unlimited freedom. Repetition occurs within normative structures. Some identity performances are more readily recognized or valued than others. This is where language ideology and platform power intersect with identity performance. A user may repeatedly perform an identity while receiving limited visibility or negative evaluation because the identity conflicts with dominant norms.

10. Audience, Context, and Contestation

Digital audiences are heterogeneous. A single post can be read by friends, colleagues, strangers, imagined communities, and algorithmically selected viewers. This complicates the classic notion of audience design. Researchers should therefore document both intended audience cues and observable audience responses when available.

Contestation is especially important. Identity is not complete when a post is published. Replies, quote-posts, comments, corrections, ridicule, support, and silence can modify its meaning. The framework treats uptake as an interactional outcome rather than assuming that self-presentation is automatically accepted.

11. Platform Affordances and Visibility

Platform affordances should not be treated as purely technical features. Their social meaning depends on how users understand and exploit them. The literature on platform-specific self-branding demonstrates that users develop strategies in relation to perceived affordances (Duffy et al., 2017). Van Dijck et al. (2018) further emphasize that platforms are part of wider social and economic structures.

Algorithmic visibility is particularly difficult to infer. A researcher should not claim that an algorithm promoted a particular identity unless the research design provides evidence about recommendation or ranking processes. Observable metrics such as likes, shares, or views can indicate engagement but do not by themselves reveal algorithmic causation. This distinction is crucial for methodological integrity.

12. Reliability and Validity

Reliability in digital identity research requires explicit coding rules. A subset of the corpus should be independently coded by another researcher for categorical dimensions such as presence of code-switching, type of stance, or explicit identity claim. Agreement can be quantified where appropriate. For interpretive categories, researchers should provide definitions, examples, and an audit trail rather than relying only on numerical agreement.

Validity depends on triangulation. Linguistic analysis can be combined with interviews, participant explanations, comment analysis, platform documentation, or longitudinal observation. The strongest claims are those supported by more than one evidential route. If only posts are available, conclusions should remain at the level of discourse and representation rather than audience impact or psychological identity.

12.1 Developing a coding manual iteratively

A coding manual for this model should not be finalized before any data is examined, nor should it remain permanently open to revision once coding has begun. A defensible procedure is to draft an initial manual from the four theoretical dimensions, apply it to a small pilot sample, document every case where two coders disagree or where a single coder hesitates, revise category definitions in response to these cases, and then freeze the manual before coding the remaining corpus. Any subsequent ambiguous case should be logged rather than resolved by silently stretching a category's definition, since undocumented category drift is one of the most common sources of unreliable qualitative coding.

12.2 An illustrative audit-trail entry

An audit trail entry should record, at minimum, the item identifier, the coding decision, the category applied, and a brief justification. For example, an entry might read: 'Item 014, classified as intra-sentential code-switching rather than borrowing, because the embedded English phrase retains English morphology and is not lexicalized elsewhere in the corpus; classification revisited after pilot disagreement on Item 009.' Entries of this kind allow a second researcher, or the original researcher at a later date, to reconstruct why a given classification was made without needing to rely on memory.

12.3 Distinguishing discourse-level from audience-level claims

A recurring validity threat in this field is the tendency to slide from a discourse-level observation, such as the presence of a particular linguistic or multimodal feature, to an audience-level claim, such as an assertion about what viewers believed, felt, or did. The model requires that these two levels of claim be kept explicitly separate in the analytical write-up. A study that has only textual data can support claims of the form 'the post makes available a particular identity position'; it cannot, without additional evidence, support claims of the form 'viewers adopted that identity' or 'the post persuaded its audience.' Maintaining this distinction is treated here as a validity safeguard rather than an optional stylistic preference.

13. Proposed Conceptual Relations

The model predicts reciprocal relations among its four dimensions. Linguistic styling can contribute to identity positioning; multimodal resources can reinforce or contradict linguistic cues; interaction can confirm or contest projected identities; and platformed visibility can amplify or constrain which performances reach audiences. These relations are probabilistic and contextual rather than deterministic.

Table N°3: Proposed Conceptual Relations Among Model Dimensions

Analytical relation	Question	Example of evidence
Language → identity	What social meaning is indexed?	Dialect choice plus stance and contextual uptake
Image ↔ language	Do modes reinforce or contradict?	Caption and visual persona
Interaction → identity	How is identity accepted or contested?	Replies, evaluations, alignment
Platform → visibility	What conditions exposure?	Persistence, recommendation, shareability
Visibility → identity	Does increased exposure alter performance?	Longitudinal change in style or self-presentation

Source: Author's own elaboration (2026)

14. Implications for Sociolinguistic Research

The model expands sociolinguistic analysis beyond the text itself while retaining language as a central resource. It is particularly suitable for multilingual digital environments where code-switching, orthographic creativity, emoji, hashtags, and visual styles operate together. It also allows researchers to compare platforms without assuming that identity has the same structure everywhere.

For Algerian digital discourse, the framework can support studies of Darja, Standard Arabic, French, English, and mixed practices, but it requires empirical evidence for each proposed social meaning. Regional, generational, and community differences should be investigated rather than generalized from a small set of posts.

14.1 Implications for comparative and longitudinal research

Because the model separates linguistic, multimodal, interactional, and platform dimensions analytically, it also supports comparative designs that would otherwise conflate these levels. A researcher could, for example, compare how the same four dimensions operate across two platforms with different affordances, or track a single set of accounts longitudinally to observe whether linguistic styling or interactional positioning shifts as platform features change. In both cases, the model's insistence on declaring corpus boundaries in advance, as specified in Section 5, becomes essential: comparative and longitudinal claims are only interpretable if the researcher has been explicit about what was held constant and what was allowed to vary between comparison points.

14.2 Implications for language policy and educational research

Beyond descriptive sociolinguistics, the model has implications for research on language attitudes and language policy in multilingual settings such as Algeria. If identity meanings are shown to be indexically constructed rather than fixed to particular languages, this has consequences for how language policy debates frame the presence of French, English, or regional varieties in public and digital communication. A policy argument that assumes a given language intrinsically carries prestige, modernity, or authenticity is vulnerable to exactly the kind of essentialist assumption that the indexical approach developed in Section 2 sets out to avoid. Empirical applications of the model could therefore contribute evidence to language-policy debates without requiring the researcher to adopt a normative position on which language ought to dominate a given domain.

15. Limitations

The model is intentionally broad and therefore requires narrowing in empirical studies. A project that attempts to code every linguistic, visual, interactional, and technical variable may become unmanageable. Researchers should select dimensions that correspond directly to their research questions.

A second limitation is access to platform data. Researchers may not be able to observe algorithmic ranking, private interactions, or the full audience of a post. The model therefore distinguishes observable platform conditions from inferred algorithmic effects.

A third limitation concerns the hypothetical illustration presented in Section 7. While that example is useful for clarifying the analytical procedure, it cannot substitute for empirical validation against a real corpus. The model's categories, though grounded in established sociolinguistic and multimodal theory, have not yet been tested for inter-coder reliability on an actual dataset, and this remains a necessary next step before the framework can be considered fully validated.

16. Conclusion

Digital identity is best analysed as an enacted and relational process shaped by language, multimodality, interaction, and platformed visibility. The four-dimensional model developed here provides a coherent sociolinguistic structure for studying that process. Its principal methodological contribution is to prevent reductionism: identity is not reduced to language, language is not reduced to identity, and visibility is not equated with algorithmic causation.

The framework can support empirical research across social media, messaging environments, discussion forums, and video-sharing platforms. Its strongest use will be in studies that combine close discourse analysis with transparent sampling, explicit coding, ethical procedures, and triangulated evidence. In that form, digital identity becomes not merely a topic of description but a testable sociolinguistic research problem.

17. References

Bell, A. (1984). Language style as audience design. Language in Society, 13(2), 145–204. https://doi.org/10.1017/S004740450001037X

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. Discourse Studies, 7(4–5), 585–614. https://doi.org/10.1177/1461445605054407

Butler, J. (1990). Gender trouble: Feminism and the subversion of identity. Routledge.

Duffy, B. E., Pruchniewska, U., & Scolere, L. (2017). Platform-specific self-branding: Imagined affordances of the social media ecology. Proceedings of the 8th International Conference on Social Media & Society. https://doi.org/10.1145/3097286.3097291

Goffman, E. (1959). The presentation of self in everyday life. Doubleday.

Kress, G. (2015). Semiotic work: Applied linguistics and a social semiotic account of multimodality. AILA Review, 28, 63–82. https://doi.org/10.1075/aila.28.03kre

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. Language & Communication, 23(3–4), 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2

van Dijck, J., Poell, T., & de Waal, M. (2018). The platform society: Public values in an online world. Oxford University Press. https://doi.org/10.1093/oso/9780190889760.001.0001