From "Who Am I With?" to "What Do I Do?": Modelling the Linguistic Profiles of Social and Professional Communities in Uzbek Digital Communication

Mirzaaxmedova Yulduzxon Anvarovna
Independent Researcher, Andijan State Institute of Foreign Languages, Uzbekistan

Abstract. This study models the linguistic profiles through which social and professional belonging are expressed in Uzbek digital communication and compares them as overlapping rather than discrete systems. The data are 10,000 annotated Uzbek YouTube comments tagged for value orientation, evaluative attitude and personal versus social positioning. Social values (59.68%) and identification (29.50%) form the semantic core of the corpus, and positive attitude (59.90%) dominates. The social profile is centred on collective belonging, values and identity ("Who am I with?"); the professional profile on activity, knowledge, competence and institutional belonging ("What do I do?"). The profiles meet in a shared zone of first-person plural possessives, address terms, evaluative units and value markers. Five intersection models are described, and the profile is reinterpreted as a communicative portrait built from marker combinations across lexical, grammatical and pragmatic levels.

Keywords: linguistic profile; social belonging; professional identity; Uzbek; YouTube comments; marker combination; communicative portrait; corpus sociolinguistics

1. Introduction

Digital platforms have become a main space in which speakers display where they belong. In short public YouTube comments, users position themselves as members of a nation, a family, a profession or an institution through lexical choice, grammatical form, evaluation and address. Because such texts accumulate at scale, the

communicative habits of language communities can be described on a corpus basis rather than from isolated examples (Baker, 2010; McEnery & Hardie, 2012).

This study treats the profile of a language community not as a list of words but as a system of interrelated markers recurring in the speech of a social or professional group, following the tradition that defines the speech community by shared norms of use (Gumperz, 1982; Hymes, 1974) and accounts in which identity is constructed in discourse (Bucholtz & Hall, 2005). For Uzbek, the social differentiation of language has been examined in descriptive sociolinguistics (Nurmonov, 2008; Mahmudov, 2013), and socially restricted vocabulary in digital communication has begun to attract attention (Raimova, 2026). However, no study has compared the profiles of social and professional communities on the basis of an annotated Uzbek digital corpus or shown how the two kinds of belonging intersect within a single comment.

Three questions are addressed:

1. Which semantic, evaluative and positioning markers form the core and periphery of the social and professional profiles?

2. Through which marker combinations do the two profiles intersect?

3. How can the profiles be generalised into a communicative portrait of Uzbek digital communication?

2. Literature Review

For Gumperz (1982), a speech community is defined not by a common lexicon or grammar but by shared norms and verbal behaviour formed in communicative activity. Hymes (1974) showed that situation, participants and purpose matter, since groups using the same language system may follow different norms. Labov (1972) demonstrated statistically that the regular use of socially correlated variants is a key

criterion for characterising a community, an approach later extended to lexical, pragmatic and discursive units.

A second line of research treats identity as a discursive accomplishment. Bucholtz and Hall (2005) see identity as emerging in interaction, where forms position speakers and construct social roles; Bucholtz (2009) traces how forms first index stance and then acquire social meaning. Eckert (2000) treats variation as social practice, and Chew (2013) views group and individual identity as shaped through language in a social and historical process. Shymko (2025), reviewing markers for sociolinguistic profiling, stresses that their reliability depends on context.

Corpus linguistics supplied the means for describing such profiles at scale: McEnery and Hardie (2012) and Baker (2006, 2010) developed frequency, collocational and discourse-analytic procedures suggesting that a group's profile is a system of interrelated signs rather than a sum of words. Research on computer-mediated communication added a further dimension. Herring (2004) argues that online communities develop their own norms, pragmatic strategies and graphostylistic resources (see also Herring & Androutsopoulos, 2015; Androutsopoulos & Juffermans, 2014). Pérez-Sabater (2012) shows that communicative habits and speech stereotypes become criteria for characterising virtual communities, and Seargeant and Tagg (2014) and Tagg (2015) show that social media users perform identity and community at several levels simultaneously.

In Uzbek linguistics, language and social experience (Mahmudov, 2013), sociolinguistic aspects of Uzbek (Nurmonov, 2008), pragmatics (Safarov, 2008) and professional self-identification (Kurbanova, 2020) have been addressed, yet studies that integrate linguistic markers into community profiles built from an Uzbek digital corpus are practically absent.

3. Materials and Methods

The empirical base is a corpus of 10,000 annotated Uzbek-language YouTube comments. Markers were not taken from a predefined list but described from usage regularities in authentic internet discourse, so the profiles are interpreted as dynamic portraits formed within a specific corpus and communicative environment rather than as fixed models. The analysis draws on three annotation layers:

- **Value orientation:** VAL_SOC (respect, solidarity, moral norms, social responsibility); VAL_ID (national, regional and collective belonging); VAL_FREE (individual choice and freedom of opinion); VAL_EDU (education); VAL_FAM (family relations, upbringing, generational continuity); VAL_TRD (customs, rituals, national tradition, heritage).

- **Evaluative attitude:** positive, neutral, negative.

- **Positioning:** personal (individual experience, subjective stance) and social (linking the opinion to a group or shared values).

Code-switching was also recorded. Automatic marker identification lies outside the scope of this article.

The analysis combined descriptive and statistical corpus methods with contextual and linguostylistic analysis. The unit of description was a set of semantically related markers. Frequency indicated how actively a marker is used, while diagnostic value was assessed through semantic content, contextual distribution, communicative function and combination with other markers. Profiles were compared by five criteria: lexical composition, stable speech patterns, communicative style, graphostylistic and multimodal resources, and communicative strategies. They were described at three levels: lexical (choice of socially or professionally oriented units), grammatical

(*men/biz* "I/we", possessive, plural and person-number forms) and pragmatic (displaying belonging, evaluating, expressing a collective stance, relying on professional experience).

4. Results

4.1. Quantitative layers

Table 1. Value orientations and evaluative attitude in the corpus (N = 10,000 comments)

Marker	Frequency	Share (%)
VAL_SOC (social values)	5968	59.68
VAL_ID (identification)	2950	29.50
VAL_FREE (freedom)	528	5.28
VAL_EDU (education)	216	2.16
VAL_FAM (family)	187	1.87
VAL_TRD (tradition)	137	1.37
Positive attitude	5990	59.90
Neutral attitude	3615	36.15
Negative attitude	378	3.78

Note. Attitude counts sum to 9,983, as in the source data.

Social values are the main semantic orientation, and the high share of identification shows that users present themselves as community representatives. Positive evaluation (approval, gratitude, respect, encouragement) is the most active layer, information exchange is substantial, and criticism is limited. Personal positioning occurs in 3,224 comments and social positioning in 2,943; the difference is small, but the former strengthens the "I" stance and the latter the "we" stance. In addition, 1,629 instances of code-switching mark multilingual communication.

4.2. The social profile

The core of the social profile is formed by collective values and belonging (VAL_SOC, VAL_ID). Its grammatical hallmark is the first-person plural: in *biz* "we",

xalqimiz “our people”, *yurtimiz* “our homeland” and *tilimiz* “our language”, possession and person-number merge the individual subject with a collective one. The characteristic model is **lexical identification + grammatical collectivity + value evaluation**. Evaluation forms the pragmatic layer: when an opinion about a group or national value combines with *ajoyib* “wonderful”, *barakalla* “well done” or *rahmat* “thank you”, it conveys not only *to whom* one belongs but *how* that belonging is evaluated.

The national-cultural layer appears in *ona tilimiz* “our mother tongue” and *milliy qadriyatlarimiz* “our national values”, where an object name carries collective possession. VAL_FREE markers show that individual stance and collective belonging can co-occur. In *Menimcha, bizning tilimizni asrash har birimizning vazifamiz* “In my opinion, preserving our language is the duty of each of us”, *menimcha* expresses individual stance, *bizning tilimiz* collective belonging and *vazifamiz* “our duty” shared responsibility: **personal stance → collective belonging → value evaluation**. The small family layer is semantically precise: in *ota-onani hurmat qilish* “respecting parents”, kinship fuses with respect and responsibility.

The social profile thus has a centre (value and belonging markers, supplemented by evaluation and positioning) and a periphery (family, tradition and freedom units). Low frequency does not mean low significance: peripheral markers can signal belonging very precisely.

4.3. The professional profile

The professional profile occupies a narrower semantic field, expressed not only by naming a profession but through experience, knowledge, institutional environment and field-specific evaluation. Six components were identified. *Terminological units* (*korpus* “corpus”, *annotatsiya* “annotation”, *sintaksis* “syntax”) signal an academic-

professional environment when they co-occur; the system of terms, not a single term, strengthens the profile. *Professionalisms* are marked by brevity and in-group intelligibility. *Institutional units* combine a professional setting with collective possession: *universitetimiz* "our university", *kafedramiz* "our department", *hamkasblarimiz* "our colleagues"; the first-person plural here points to a specific institutional group rather than a broad collective. *Activity verbs* (*o'qitmoq* "to teach", *tahlil qilmoq* "to analyse", *davolamoq* "to treat") express professional practice, as in *darsni tashkil etish* "organising a lesson".

Educational content (VAL_EDU, 216 comments) is related to but distinct from professional content: *talaba* "student" acquires professional meaning in *bo'lajak mutaxassislarni tayyorlash* "training future specialists". *Professional evaluation* (*malakali* "qualified", *tajribali* "experienced", *mas'uliyatli* "responsible") assesses persons by professional criteria, as in *malakali mutaxassis* "a qualified specialist". *Professional address forms* (*ustoz* "teacher, mentor", *hamkasb* "colleague", *mutaxassis* "specialist") name status while performing pragmatic functions: *ustoz* can express respect, *hamkasb* professional equality.

4.4. Comparative description

Table 2. Main linguistic features of the social and professional profiles

Linguistic feature	Social profile	Professional profile
Semantic centre	Community, value, identification	Profession, activity, knowledge, qualification
Main positioning	National, regional, family, social	Professional, institutional, educational
Leading vocabulary	Social and value-related units	Terms, professionalisms, profession names
Grammatical marker	*biz, bizning, xalqimiz, yurtimiz*	*universitetimiz, kafedramiz, hamkasblarimiz*
Object of evaluation	Person, community, value, social phenomenon	Specialist, activity, knowledge, professional result
Pragmatic orientation	Solidarity, support, belonging	Professional experience, competence, activity
Type of identification	"We" = broad social community	"We" = specific professional group
Main topics	Society, nation, language, culture, family, tradition	Education, profession, work, science, specialisation
Evaluation criteria	Moral and social	Professional and activity-based
Point of intersection	Collective positioning	Institutional-professional positioning

Positioning and evaluation are shared, but differently oriented: *U juda yaxshi inson* "He/she is a very good person" evaluates social qualities, whereas in *U juda yaxshi mutaxassis* "He/she is a very good specialist" the same *yaxshi* "good" targets competence. The social profile answers "Who am I with?", the professional profile "What do I do and which professional group do I belong to?". Stance differs too: social-profile comments rely on shared norms (*tilimizni asrashimiz kerak* "we must preserve our language"), while professional-profile comments speak from experience (*mening tajribamda* "in my experience", *mutaxassis sifatida* "as a specialist").

4.5. Intersection points and marker combinations

The two kinds of belonging rarely appear in isolation. In *Universitetimiz talabalari xalqaro tanlovda yaxshi natija ko'rsatdi* "Students of our university showed good results in an international competition", *universitetimiz* expresses institutional belonging, *talabalari* the educational-professional layer and *yaxshi natija* "good result" positive evaluation. In *fidoyi ustoz* "a devoted teacher" or *halol mutaxassis* "an honest specialist", the noun denotes status and the attribute a social-normative evaluation. In *Rahmat, ustoz!* "Thank you, teacher!", a professional address expresses respect and gratitude, and in *ona tilimizni o'qitish* "teaching our mother tongue" national belonging combines with professional activity. *Men o'qituvchi sifatida…* "I, as a teacher…" presents an individual professional subject, while *Biz, o'qituvchilar, bu masalaga boshqacha qaraymiz* "We, teachers, view this issue differently" combines a collective pronoun, a profession name and a group stance. Table 3 generalises these combinations.

Table 3. Linguistic intersection models of the social and professional profiles

Intersection model	Main linguistic means	Meaning expressed
Professional + social	Profession name + value markers	Social evaluation of professional activity
Professional + identificational	*biz*, *-imiz* + professional unit	Belonging to a professional group
Professional + personal	*men*, *mening* + professional unit	Professional identification based on personal experience
Professional + evaluative	Profession name + positive/negative evaluation	Attitude towards a professional or an activity
Professional + national	Professional unit + national-cultural marker	Linking professional activity to national-cultural content

Positioning, evaluation and value markers thus bridge the profiles, and lexical boundaries are permeable: *yoshlar* "young people" is a social group, but *yosh mutaxassislar* "young specialists" specifies it professionally.

4.6. The communicative portrait

The results generalise into a communicative portrait comprising the subject's stance ("I", "we" and "I, as a specialist"), value orientation, evaluation, lexical choice and grammatical and pragmatic layers. In *Universitetimiz ustozlari yoshlarni milliy qadriyatlar ruhida tarbiyalamoqda* "The teachers of our university are educating young people in the spirit of national values", *universitetimiz* expresses institutional and collective positioning, *ustozlari* professional belonging, *yoshlarni* a social group, *milliy qadriyatlar* value content, and *tarbiyalamoqda* professional activity. The portrait grows through combination: *ustoz* indicates professional belonging; *ustozimiz* "our teacher" adds collective positioning; *fidoyi ustozimiz* "our devoted teacher" adds positive evaluation.

5. Discussion

Regarding the first question, the weight of VAL_SOC and VAL_ID and the dominance of positive evaluation show that Uzbek YouTube comments are oriented primarily towards collective values and community belonging. This specifies for Uzbek the norms that online communities develop (Herring, 2004; Herring & Androutsopoulos, 2015). The lower visibility of professional markers reflects the platform rather than their unimportance: YouTube is a mass communication environment where users display social and cultural belonging more actively than professional status, illustrating that norms depend on setting (Hymes, 1974).

Methodologically, frequency and diagnostic value must not be equated. VAL_FAM and VAL_TRD markers are rare yet highly specific, and professional terms, though less frequent than social markers, reflect professional communicative norms. This supports reading frequency together with distribution and function (Baker,

2006; McEnery & Hardie, 2012) and echoes Shymko's (2025) point that marker reliability is context-dependent.

Regarding the second question, the two profiles are intersecting fields with distinct centres and a shared periphery. Uzbek morphology is decisive: the possessive *-imiz* on institutional and professional nouns (*kafedramiz*, *ustozlarimiz*) answers at once "to whom do I belong?" and "to which professional environment?". Grammatical means thus act as implicit markers of belonging, since speakers rarely state membership explicitly. This supports the view of identity as positioned in discourse (Bucholtz & Hall, 2005) and of forms indexing stance before acquiring social meaning (Bucholtz, 2009). Evaluation, moreover, specifies belonging semantically: the same *yaxshi* indexes moral-social qualities in one context and competence in another.

Concerning the third question, the portrait's distinctiveness lies in the relations between markers, and its multilayered character reflects the fact that digital users perform several identities at once (Seargeant & Tagg, 2014; Tagg, 2015). In line with Eckert (2000), the profile is better seen as a social practice than a static inventory.

The study is limited to one platform, so the dominance of social markers may be platform-specific; quantitative results cover value, evaluation, positioning and code-switching only, and inter-annotator agreement is not reported.

6. Conclusion

Based on 10,000 annotated Uzbek YouTube comments, the social profile is centred on collective belonging, values and identity, with social values (5,968 comments) and identification (2,950) at its core, supported by predominantly positive evaluation (59.90%) and the shift from "I" to "we". The professional profile is centred on activity, knowledge, competence and institutional belonging, built from terms, professionalisms, institutional units, activity verbs, professional evaluation and address

forms. Five intersection models show that their boundaries are contextual and gradient rather than categorical. A linguistic profile is therefore best interpreted as a configuration rather than a sum of markers, and the communicative portrait of Uzbek digital communication emerges from the functional combination of units across levels.

References

Androutsopoulos, J., & Juffermans, K. (2014). Digital language practices in superdiversity: Introduction. *Discourse, Context & Media, 4–5*, 1–6.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh University Press.

Bucholtz, M. (2009). From stance to style: Gender, interaction and indexicality in Mexican immigrant youth slang. In A. Jaffe (Ed.), *Stance: Sociolinguistic perspectives* (pp. 146–170). Oxford University Press.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies, 7*(4–5), 585–614.

Chew, P. G.-L. (2013). Introduction: A sociolinguistic history of early identities in Singapore. In *A sociolinguistic history of early identities in Singapore* (pp. 1–18). Palgrave Macmillan.

Eckert, P. (2000). *Linguistic variation as social practice: The linguistic construction of identity in Belten High*. Blackwell.

Gumperz, J. (1982). *Discourse strategies*. Cambridge University Press.

Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for virtual communities in the service of learning* (pp. 338–376). Cambridge University Press.

Herring, S. C., & Androutsopoulos, J. (2015). Computer-mediated discourse 2.0. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 127–151). Wiley-Blackwell.

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. University of Pennsylvania Press.

Kurbanova, Z. (2020). Kasbiy autoidentifikatsiya shakllanishining bosqichlari [Stages in the formation of professional self-identification]. *Oriental Art and Culture*, (II), 175–179.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Mahmudov, N. (2013). *Til va nutq* [Language and speech]. O'zbekiston milliy ensiklopediyasi.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Nurmonov, A. (2008). *O'zbek tilining sotsiolingvistik jihatlari* [Sociolinguistic aspects of the Uzbek language].

Pérez-Sabater, C. (2012). The linguistics of social networking: A study of writing conventions on Facebook. *Linguistik Online, 56*(6), 81–93.

Raimova, N. O. (2026). Ijtimoiy chegaralangan leksikaning jamiyatdagi roli va muhimligi [The role and importance of socially restricted vocabulary in society]. *ALFRAGANUS xalqaro ilmiy jurnali*, (2(11)), 65–70.

Safarov, Sh. (2008). *Pragmalingvistika* [Pragmalinguistics]. Fan.

Seargeant, P., & Tagg, C. (Eds.). (2014). *The language of social media: Identity and community on the Internet*. Palgrave Macmillan.

Shymko, V. (2025). Identifying key idiolect markers in sociolinguistic profiling: A scoping review and analytical framework for real-world applications. *SAGE Open, 15*(2), 1–32.

Tagg, C. (2015). *Exploring digital communication: Language in action*. Routledge.