Collaborative Multilingual Discourse Analysis in Science Classrooms: A Systemic Functional Approach

Anonymous ACL submission

Abstract

This paper examines how a Systemic Functional Linguistics (SFL) framework can inform the design of computational linguistic features to analyze dialogic interactions in multilingual classrooms. Classroom discourse with multilingual learners (MLs) often involves translanguaging, defined as the fluid use of linguistic resources across and beyond named languages (e.g., English, Spanish). However, conventional natural language processing (NLP) methods (i.e., approaches based on structural features such as part-of-speech tags, syntactic dependency relations, or discourse markers) typically represent linguistic form without explicit alignment to the functional meanings captured by SFL metafunctions: ideational, interpersonal, and textual. To address this gap, we map these metafunctions onto computational linguistic features derived from part-of-speech patterns (POS), Universal Dependencies (UD) relations, and dialogue act (DA) categories. Through multiple analyses, we show that POS and UD features align most strongly with ideational meaning, DA features align most clearly with interpersonal meaning, and textual meaning remains weakly represented across feature types. These findings clarify which computational linguistic features most effectively capture functional aspects of language use and provide a foundation for developing more inclusive and scalable NLP approaches to analyzing translanguaging in classroom interaction.

1 Introduction

Classroom interaction involving multilingual learners (MLs) is frequently characterized by *translanguaging*, in which speakers draw on their full linguistic repertoires to make meaning (García and Lin, 2017). Translanguaging differs from code-switching. Whereas code-switching presupposes alternation between distinct language systems, translanguaging conceptualizes MLs' mental grammars as structured, unitary repertoires of features from which speakers select according to communicative intent (Otheguy et al., 2015). From this perspective, named languages are social constructs, not separate "systems" in the mind. Instead, what appear as mixed-language utterances reflect how speakers draw from a single, integrated linguistic repertoire.

Speaker	Utterance	Meta-functions
Yesenia	¿Qué pasó? ¿Qué vieron? *[What happened? What did you see?]*	Ideational, Interper-sonal
Trevor	Matando a las … *[Killing the …]* … um … urchins	Ideational, Interper-sonal
Yesenia	¿Están matando a los erizos porque qué pasó con los erizos? *[They are killing the urchins because what happened with the urchins?]*	Ideational, Textual
Trevor	Se comen la *kelpo [They eat the kelp (la alga marina)]*	Ideational

Table 1: Excerpt of multilingual classroom discourse annotated with SFL metafunction labels.

These practices are especially visible in small-group disciplinary discussions, where MLs collaboratively construct explanations, negotiate epistemic stance, and organize ideas during joint problem solving (Mercer, 2004; Chi, 2009; Sun et al., 2019).

Recent work in computational discourse analysis has sought scalable methods for examining classroom interaction (Rosé et al., 2019; Demszky et al., 2025). However, many NLP systems are trained on standardized monolingual corpora and rely on structural representations that assume stable language boundaries (Joshi et al., 2020; Blodgett et al., 2020). As a result, these systems may inadequately capture translanguaging in classroom discourse.

To address this gap, we draw on Systemic Func-

tional Linguistics (SFL), a theory that links linguistic form to social meaning (Halliday, 1993, 1994). SFL conceptualizes language as a semiotic system organized around three metafunctions: ideational meaning (construing experience and logical relations), interpersonal meaning (enacting social roles and stance), and textual meaning (organizing information flow and cohesion). Because these metafunctions specify how linguistic choices realize how, why, and for what purpose(s) language is used to make meaning, they provide a principled framework for connecting structural features to functional interpretation (Table 1).

While SFL has supported extensive qualitative analyses of classroom discourse and academic genres (Halliday and Hasan, 1985; Martin and Rose, 2007; Schleppegrell, 2004; Khote and Tian, 2019; Troyan et al., 2022), its constructs have rarely been operationalized for computational analysis of multilingual interaction. Prior NLP approaches typically extract part-of-speech (POS) patterns, syntactic dependencies, or discourse markers without explicit alignment to metafunctional meaning. This disconnect makes it difficult to interpret how automatically extracted linguistic features relate to the functional meanings emphasized in SFL analyses of translanguaged classroom discourse. In this study, linguistic features refer to automatically extractable structural and pragmatic indicators derived from NLP pipelines, including POS patterns, dependency relations, and dialogue act signals.

To this end, the paper makes three contributions: (i) we introduce a corpus of small-group science discussions involving multilingual learners annotated by bilingual experts for SFL metafunctions; (ii) we examine how POS patterns, Universal Dependencies (UD) relations, and dialogue act (DA) indicators align with these metafunctions; and (iii) we examine how the linguistic composition of the corpus distributes across metafunction-aligned linguistic features to provide a quantitative characterization of translanguaging within the dataset.

2 Related Work

2.1 Systemic Functional Linguistics

Systemic Functional Linguistics (SFL), developed by Michael Halliday and colleagues, models language as a semiotic system in which meaning is shaped by social activity and realized through patterned linguistic choices (Halliday, 1993, 1994; Ballard, 1980; Halliday and Hasan, 1985; Martin and Rose, 2007). SFL situates language use within cultural and situational contexts and specifies how ideational, interpersonal, and textual meanings are instantiated through linguistic resources and registers.

A central claim of SFL is that metafunctional meaning is realized through lexicogrammar. Harman's (2017) model of language in context foregrounds lexicogrammatical resources, such as verbs, prepositional phrases, and adverbials, as links between linguistic form and functional interpretation (see Figure 4 in Appendix A). From this perspective, metafunctions are enacted through participant and process types, mood and modality systems, and cohesive and thematic organization in addition to registers (e.g., field, tenor, mode).

Systemic Functional Linguistics has been widely used to examine disciplinary knowledge construction and scaffolding practices in classroom discourse (Schleppegrell, 2004; Hammond and Gibbons, 2005; Jones and Lock, 2011; Humphrey and Macnaught, 2015; Khote and Tian, 2019). However, SFL constructs have rarely been operationalized for computational analysis of multilingual interaction, leaving open questions about whether the lexicogrammatical realizations described in SFL can be approximated through interpretable linguistic features extracted from classroom discourse.

2.2 Translanguaging and Multilingual NLP

In sociolinguistics, code-switching refers to the alternation between named languages within or across turns (García and Wei, 2015), while translanguaging conceptualizes mental grammars as unitary repertoires from which speakers select features without regard for externally defined language boundaries (Otheguy et al., 2015). In this view, named languages are sociocultural objects that should not be treated as linguistically bounded systems.

Recent advances in multilingual NLP have expanded models' capacity to handle code-switched input. However, many NLP approaches continue to treat languages as separable entities evaluated against monolingual or standardized benchmarks (Joshi et al., 2020; Blodgett et al., 2020). Although multilingual models increasingly accommodate code-switched data, they often struggle with mid-clause shifts and discourse-pragmatic functions such as stance and alignment (Zhang et al., 2023).

From a translanguaging perspective, this diffi-

culty reflects modeling assumptions about language boundaries. Computational pipelines that segment, parse, or classify input according to named languages may underrepresent the repertoires through which multilingual learners construct meaning (García and Lin, 2017). While SFL-informed qualitative studies have documented these practices in multilingual classrooms (Khote and Tian, 2019; Troyan et al., 2022), few studies have examined how metafunctional constructs align with automated linguistic features in translanguaged educational discourse. Therefore, the present study examines whether interpretable linguistic features derived from multilingual classroom discourse align with expert judgments of metafunctional meaning.

3 Data

3.1 Dataset

The dataset consists of 204 utterances drawn from ten small-group science classroom transcripts collected in Fall 2024 and Spring 2025 across three Midwestern K–12 school districts. We generated transcripts from longer classroom sessions that were part of an IRB-approved study on translanguaging in science classrooms. For the present analysis, we selected focal segments capturing small-group, collaborative sense-making. 123 students participated in the research study. The classrooms varied in grade level, science content, and linguistic demographics. The corpus contains utterances in English, Spanish, or both within the same turn, which reflects translanguaging practices in classroom interaction.

To characterize the linguistic composition of the corpus, we performed automatic language identification at the utterance level. Short non-lexical responses (e.g., *mhmm*, *oh*, *aha*) were categorized as *Other*. This analysis indicated that 56.86% of utterances were primarily English and 12.75% primarily Spanish, while 30.39% consisted of short discourse markers or non-lexical responses. Three utterances (1.47%) contained lexical material from both languages within the same turn.

Token-Level Language Distribution To characterize lexical language use more precisely, we also performed token-level language identification using NLTK language detection. Across the corpus, English accounted for 454 tokens (84.9%) and Spanish for 81 tokens (15.1%). Spanish lexical material appeared both in fully Spanish utterances and in occasional multilingual turns.

3.2 Human Annotation

Two bilingual annotators fluent in English and Spanish independently annotated all utterances using a finalized annotation protocol. Both annotators had prior knowledge of SFL and experience analyzing multilingual classroom discourse. The annotation unit was a single speaker turn defined as one sentence or short utterance. Annotators identified all metafunctions expressed in each utterance and used multiple labels when meanings overlapped:

- **Ideational**: construing experience, describing processes or phenomena, evaluating feasibility, or providing factual information.
- **Interpersonal**: expressing stance, attitude, uncertainty, persuasion, alignment, or disagreement.
- **Textual**: sequencing ideas, referencing prior talk, or using connectors and cohesive devices to organize discourse.

3.3 SFL Label Distribution

We examined the distribution of metafunction labels based on adjudicated annotations. Multifunctional utterances were common. Combined ideational–interpersonal labels represented the largest category (20.98%), followed by singular ideational meaning (18.54%) and full three-way combinations (17.56%). Purely textual or purely interpersonal turns were comparatively rare.

Because utterances frequently expressed overlapping ideational, interactional, and textual meanings, metafunctions were not treated as mutually exclusive categories in subsequent analyses. A full visualization of the adjudicated metafunction label distribution appears in Appendix C.

Rank-Order Labeling and Adjudication When multiple metafunctions were present, annotators assigned the primary communicative function first. For example, the student utterance *No creo que se va a funcionar [I don't think that will work]... probably actually... yeah, it probably would* was labeled primarily as interpersonal (expressing uncertainty), secondarily as textual (resequencing stance), and tertiarily as ideational (evaluating feasibility).

After independent annotation, disagreements were resolved through adjudication by the primary author. Adjudication involved reviewing both annotators' labels and written justification notes. In

ambiguous cases, decisions were guided by the dominant communicative purpose of the utterance. For example, when an utterance both introduced new content and expressed stance (e.g., a speculative explanation), adjudication prioritized the function most central to advancing the turn sequence. This approach prioritized communicative intent, dialogic engagement (Lane and Kent, 2018), and maintaining consistency across annotators.

Reliability Inter-rater reliability was assessed using Cohen's kappa for each metafunction. Agreement was moderate for ideational meaning ($\kappa = 0.551$), high for interpersonal meaning ($\kappa = 0.693$), and moderate for textual meaning ($\kappa = 0.444$) (Landis and Koch, 1977). Overall micro-averaged agreement was $\kappa = 0.595$, with exact set-level agreement of 0.620 (Table 2). Following adjudication, a single gold-standard label set was used for subsequent analyses.

4 Methods

This study evaluates whether computationally extractable linguistic features align with theoretically defined SFL metafunctions in translanguaged classroom discourse. Our methodological design follows principles of construct validity, in which observable indicators are evaluated for systematic association with theoretically defined categories (Cronbach and Meehl, 1955; Messick, 1995).

4.1 Pre-processing and Feature Extraction

Utterances were processed using the Stanza NLP pipeline (Qi et al., 2020). For English and Spanish utterances, Stanza produced tokenization, lemmatization, part-of-speech (POS) tags, and Universal Dependencies (UD) relations (de Marneffe et al., 2021).

Token-level annotations were aggregated to the utterance level by computing counts and normalized proportions of POS categories and UD relations. Dialogue act indicators were generated automatically using a large language model (LLM) and incorporated as additional pragmatic features.

4.2 Linguistic Features

We constructed interpretable linguistic indicators spanning POS categories, UD relations, and dialogue act (DA) features. POS and UD features capture grammatical structure, while DA indicators capture conversational actions such as questioning, agreement, or disagreement.

Guided by SFL theory (Harman, 2017), we selected features likely to approximate ideational participant–process structures, interpersonal interactional moves, and textual cohesion devices. The complete feature inventory appears in Appendix B.

POS and UD Structure POS tagging and UD parsing provide structural representations widely used in NLP (Honnibal and Curran, 2007; Qi et al., 2020). POS categories encode lexical classes (e.g., nouns, verbs, adjectives), while UD relations specify grammatical roles (e.g., nominal subjects, objects, obliques), making participant–process configurations observable.

Guided by SFL theory, particularly lexicogrammatical realizations of participants, processes, and clause linkage (Harman, 2017), we selected POS and UD features that plausibly approximate these constructs and can be reliably extracted from English–Spanish discourse.

To approximate textual organization, we incorporated UD relations associated with coordination, subordination, discourse markers, and parataxis, which capture clause linkage and local sequencing. In total, we extracted 22 POS-based features and 15 UD-based features for analysis.

Dialogue Acts Dialogue Act (DA) features provide a complementary pragmatic layer beyond structural representations as they allow the characterization of communicative moves such as statements, questions, agreements, disagreements, and acknowledgments (Jurafsky and Martin, 2025).

To operationalize DA categories for our data, we perform automated utterance-level classification using LLMs. Although LLMs generally underperform task-specific fine-tuned models on fine-grained DA classification tasks (Qamar et al., 2025), they may provide a compute-efficient alternative for identifying frequently occurring dialogue-act labels (Martinenghi et al., 2024; Finch et al., 2023). DA labels were generated automatically using ChatGPT-5.1 following the annotation protocol described by Finch et al. (2023).

The resulting labels were used as pragmatic linguistic features representing conversational actions. We selected 20 DA-based features corresponding to questioning, stance, agreement, disagreement, repair, and alignment categories.

DA labels were treated as automatically generated pragmatic features and analyzed alongside POS and UD features in the feature–metafunction

correlation analysis.

4.3 Evaluation

Each utterance was represented by POS-based features, UD relations, and DA indicators alongside binary presence markers for ideational, interpersonal, and textual metafunctions. Spearman rank correlations were computed to quantify associations between individual features and each metafunction.

We selected Spearman correlation because it captures associations between interpretable indicators and expert judgments without assuming linearity or requiring distributional normality. This enables direct inspection of which structural and pragmatic cues systematically align with particular metafunctions.

Language Distribution To examine how translanguaging relates to metafunction-aligned linguistic features, we conducted a token-level language distribution analysis.

First, we identified the highest-correlating linguistic features for each metafunction based on the Spearman correlation analysis. For each metafunction, we selected the most interpretable linguistic features with the strongest correlations spanning POS, UD, and DA indicators.

Next, we applied token-level language identification using NLTK language detection. Each utterance containing one of the selected linguistic features was tokenized, and individual tokens were classified as English or Spanish. Tokens were then aggregated across utterances containing each feature.

5 Results

This section presents three complementary analyses: (1) human annotation reliability, (2) correlations between linguistic features and expert-assigned SFL metafunction labels, and (3) language distribution across features aligned with each metafunction.

5.1 Human Annotation Reliability

Inter-rater reliability between human annotators was substantial overall ($\kappa = 0.595$; exact set-level agreement $= 0.620$). Agreement was highest for interpersonal meaning ($\kappa = 0.693$), followed by ideational meaning ($\kappa = 0.551$), and lowest for textual meaning ($\kappa = 0.444$). These values indicate that metafunction identification was moderately stable across human annotators and textual meaning presented the greatest interpretive difficulty (Table 2).

Metafunction	κ
Ideational	0.551
Interpersonal	0.693
Textual	0.444
Overall micro	0.595
Exact set-level agreement	0.620

Table 2: Human–human agreement for metafunction labels per utterance.

5.2 Feature–Metafunction Alignment

We next examined whether computationally extractable linguistic features aligned with expert metafunction labels. Spearman correlations were computed separately for POS features, UD relations, and DA features.

5.2.1 Ideational Meaning

Ideational meaning showed the strongest and most consistent structural alignment. POS features, such as verb counts and participant-related categories, produced correlations in the range of $\rho \approx .30$–$.38$ (Figure 1). UD relations associated with participant–process structures, including `nsubj` ($\rho \approx .31$), `obj`, `obl`, and adverbial modifiers (`advmod`, `advcl`), also showed positive associations (Figure 2). These results suggest that grammatical patterns involving subjects, verbs, and objects are associated with ideational meaning in translanguaging small-group discourse.

5.2.2 Interpersonal Meaning

Interpersonal meaning, as expected, aligned most clearly with dialogue act features. DA categories reflecting questioning, (dis)alignment, uncertainty, and backchannel behavior produced correlations in the range of $\rho \approx .14$–$.23$ (Figure 3). In contrast, POS and UD features showed weak or negative associations with interpersonal labels (Figures 1 and 2), suggesting that interpersonal meaning is expressed more through conversational actions (e.g., questions, agreement, and disagreement) than through grammatical structure within a clause.

5.2.3 Textual Meaning

Textual meaning exhibited moderate but narrower alignment with clause linkage and cohesion features. In the UD heatmap, coordinating conjunc-

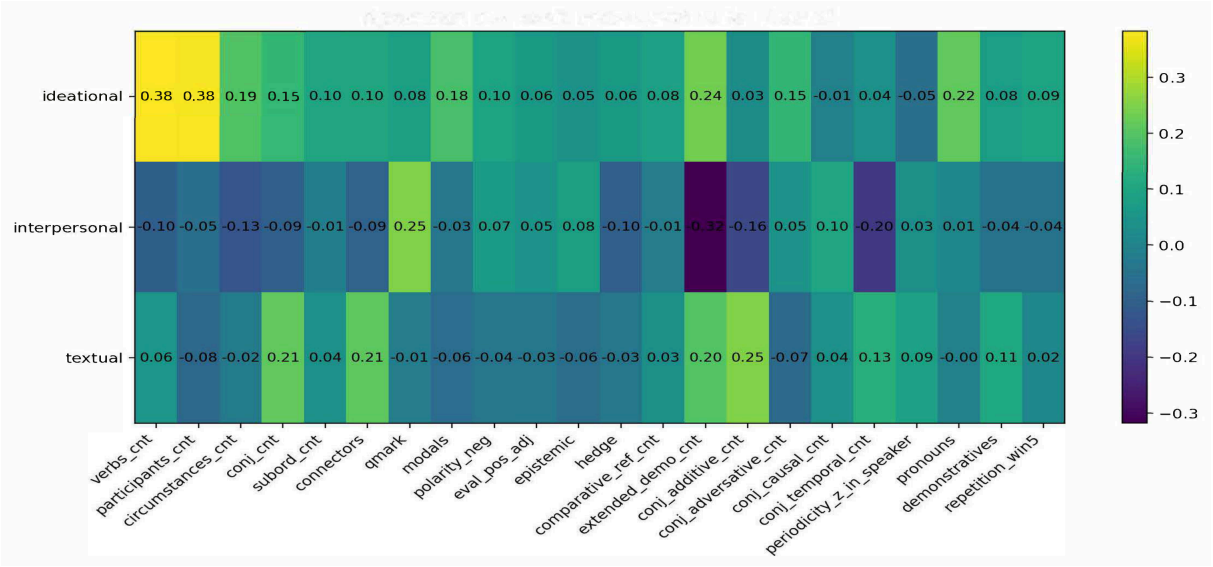

Figure 1: Spearman correlations of Parts-of-Speech features and SFL metafunction labels.

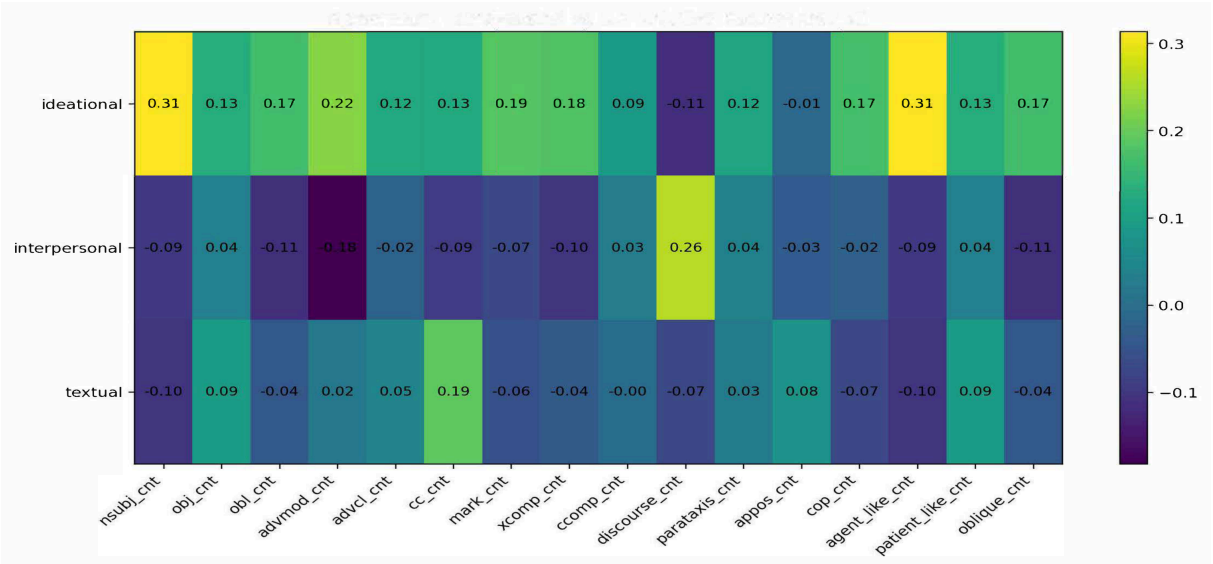

Figure 2: Spearman correlations of Universal Dependencies relations and SFL metafunction labels.

tions (cc) and parataxis relations showed correlations of $\rho \approx .19$–$.26$ (Figure 2). POS categories associated with connectors and certain demonstratives yielded smaller positive associations ($\rho \approx .03$–$.21$; Figure 1). These results indicate that features such as connectors and clause links capture some aspects of textual meaning but do not fully represent how ideas are organized in the discussion.

5.3 Language Distribution Across Metafunction-Aligned Features

To examine how translanguaging appears across metafunctions, we analyzed the language distribution of the highest-correlating linguistic features identified in the Spearman analysis. For each feature, we first identified all utterances in which the feature occurred and then counted the English and Spanish words (tokens) within those utterances. The reported counts represent the total number of tokens appearing in feature-containing utterances.

For ideational meaning, the highest-correlating features included participant structure and verb processes. Participant structure combines the POS participant feature (`participants_cnt`) with the UD noun subject feature (`nsubj_cnt`), which produced identical token counts in the dataset and therefore appear jointly in Table 3. The next highest-correlating feature was verb processes (`verbs_cnt`), which capture material and relational processes expressed through verbs. Across utter-

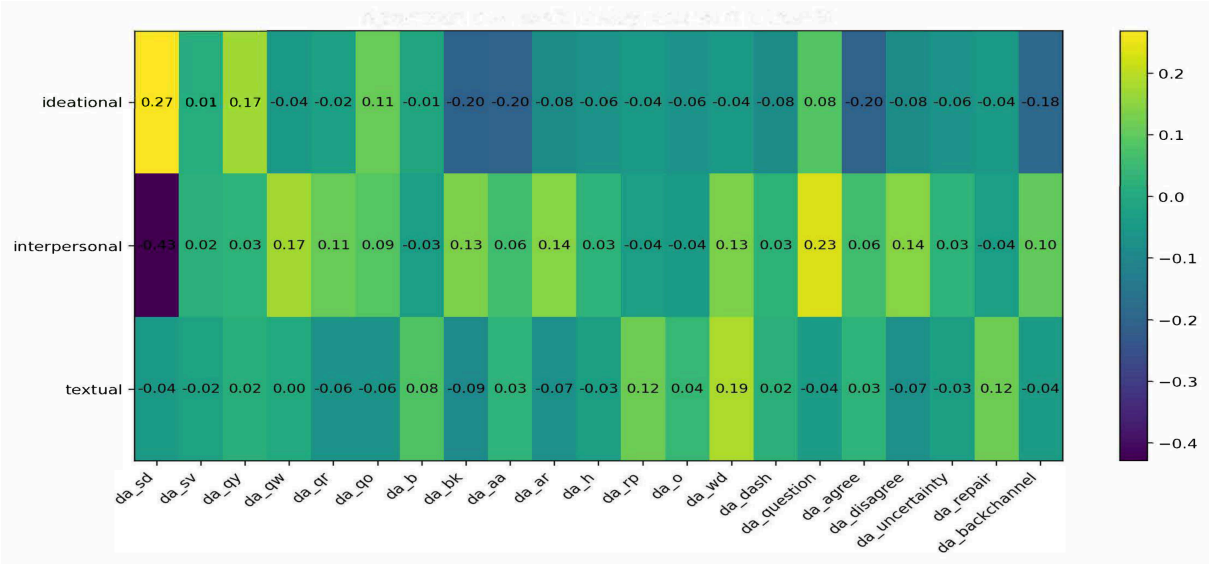

Figure 3: Spearman correlations of Dialogue Act features and SFL metafunction labels.

Meta	Feature	EN	ES	EN%	ES%	Example
–	Corpus	454	81	84.9	15.1	–
ID	PS	334	60	84.8	15.2	"**Se comen** la kelpo." [They **eat** the kelp.]
ID	VB	360	69	83.9	16.1	"We **put** the aluminum..."
IP	QM	156	30	83.9	16.1	"**¿Qué pasó? ¿Qué vieron?**" [**What happened? What did you see?**]
IP	DIS	52	7	88.1	11.9	"Pero, **no creo** que..." [But **I don't think** that...]
TX	CC	277	46	85.8	14.2	"**Entonces**, ¿hay muchos erizos...?" [**So**, are there many urchins...?]

Table 3: Language distribution across linguistic features aligned with each SFL metafunction. Bold text highlights the lexical element realizing the feature. PS = participant structure; VB = verb processes; QM = question markers; DIS = disagreement; CC = clause conjunctions.

ances containing these ideational structures, English accounted for 334 tokens and Spanish for 60 tokens (84.8% English). Utterances containing verb processes included 360 English tokens and 69 Spanish tokens (83.9% English).

Interpersonal features showed slightly greater multilingual variation. Question markers (`qmark`) occurred in utterances containing 156 English tokens and 30 Spanish tokens (83.9% English). Disagreement markers (`da_disagree`) appeared less frequently but were similarly English-dominant, occurring in utterances containing 52 English tokens and 7 Spanish tokens (88.1% English). These results indicate that interpersonal meaning in the dataset is commonly instantiated through questioning and occasional stance-taking moves during collaborative interaction.

Textual features showed similar distributions. Clause conjunctions (combining `conj_additive_cnt` and `cc_cnt`) occurred primarily in English tokens (277 English vs. 46 Spanish; 85.8% English). These conjunctions represent clause-linking devices used to organize and connect ideas across utterances.

Overall, the distribution of English and Spanish tokens across metafunction-aligned features closely mirrors the corpus baseline (84.9% English, 15.1% Spanish). Spanish lexical material appears across ideational, interpersonal, and textual contexts at similar rates.

6 Discussion

This study examined whether interpretable computational features align with expert-annotated SFL metafunctions in multilingual classroom discourse. The results clarify which aspects of classroom meaning are detectable through structural and dialogue-based indicators and which remain difficult to approximate computationally.

Ideational meaning showed the strongest alignment with POS and UD features, particularly with grammatical configurations involving subjects, verbs, and objects. This pattern is consistent with SFL theory, where ideational meaning is realized through participant and process structures expressed in clause grammar. The moderate correlations suggest that structural representations

capture part of ideational meaning in collaborative discourse but do not fully represent the reasoning processes students articulate during group sense-making.

Interpersonal meaning aligned most clearly with dialogue-act indicators, particularly question types and disagreement markers. Questioning (e.g., wh- and yes/no questions) showed the strongest associations, while POS and UD features were weak or inconsistent. Importantly, this shows that interpersonal meaning is primarily realized through interactional moves rather than clause-level structure. Questioning, in particular, also plays a central role in coordinating peer reasoning, probing ideas, and negotiating explanations.

Textual meaning was the most difficult dimension to approximate computationally. Cohesion-related UD relations such as coordinating conjunctions and parataxis showed moderate associations, while many discourse-organizing functions operate across turns. This suggests that modeling discourse organization may require context windows spanning multiple turns.

The findings also show that multilingual classroom discourse can be analyzed computationally without assuming strict language boundaries. POS, UD patterns, and DA indicators aligned with metafunctional meanings even when utterances reflected translanguaging across English and Spanish resources. While SFL analyses have traditionally emphasized lexico-grammatical cues (Harman, 2017), these results suggest that automatically detectable pragmatic indicators (e.g., dialogue acts) can also support computational analysis of metafunctional meaning.

The language distribution analysis further clarifies translanguaging in the dataset. Although English accounted for most lexical tokens overall, Spanish lexical material appeared at similar rates across the ideational, interpersonal, and textual. MLs systematically draw on Spanish resources throughout collaborative interaction and do not confine them to specific communicative functions, while English more consistently provided the grammatical structure of disciplinary explanations. Such patterns align with translanguaging perspectives in which MLs draw flexibly from a unified linguistic repertoire (Otheguy et al., 2015).

Feature–metafunction alignment suggests applications for scalable educational discourse analytics. POS and UD patterns of ideational meaning could support automated detection of explanation and causal reasoning during collaborative problem solving (Sun and Rosé, 2020), while dialogue-act features associated with interpersonal meaning could be used to track questioning, agreement, and stance-taking during group interaction (Wise and Chiu, 2014). Because textual meaning showed weaker alignment with clause-level features, these results highlight the need for discourse-level models that capture how students organize and connect ideas across turns in extended collaborative dialogue (Worsley et al., 2021).

7 Conclusion

This study examined how Systemic Functional Linguistics (SFL) metafunctions can be approximated through interpretable computational features in multilingual small-group science discourse. Structural POS and UD features aligned most strongly with ideational meaning, dialogue act indicators aligned with interpersonal meaning, and cohesion-related features showed limited alignment with textual meaning. Human annotators demonstrated moderate to substantial reliability across metafunction categories.

These results suggest that several SFL constructs are partially observable through automatically extractable linguistic features. Grammatical configurations captured cues for ideational meaning, while dialogue-based indicators reflected interactional patterns associated with interpersonal meaning. Textual meaning proved more difficult to detect with clause-level representations, indicating that modeling discourse organization may require broader conversational context.

Overall, theory-informed feature design supports interpretable analysis of translanguaging classroom discourse and provides a foundation for future computational work connecting linguistic theory with educational discourse analytics.

8 Limitations

This study analyzed a relatively small dataset consisting of 204 utterances from ten classroom transcripts. The results therefore reflect exploratory patterns within naturally occurring multilingual classroom interaction and should not be interpreted as broadly generalizable findings.

The feature set focused primarily on clause-level structural indicators and dialogue act signals. Many aspects of discourse organization operate across multiple turns (e.g., idea development during col-

laborative reasoning) which were not directly modeled in the current analysis.

Correlation analyses measured statistical associations between linguistic features and metafunction labels. These analyses identify systematic relationships but do not establish causal explanations for how linguistic forms realize metafunctional meanings in classroom discourse.

9 Ethical considerations

This study uses classroom transcripts containing children's speechollected under IRB approval. All data were preprocessed to remove personally identifying information, and only anonymized transcripts are included in the dataset. All researchers involved in data handling and analysis completed appropriate human subjects training (e.g., CITI). The study follows established ethical guidelines for research involving minors and educational data.

References

W. L. Ballard. 1980. M. a. k. halliday, language as social semiotic: The social interpretation of language and meaning. *Language in Society*, 9(1):84–89.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Michelene T. H. Chi. 2009. Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1):73–105.

Lee J. Cronbach and Paul E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52(4):281–302.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Dorottya Demszky, Jing Liu, Heather C. Hill, Shyamoli Sanghi, and Ariel Chung. 2025. Automated feedback improves teachers' questioning quality in brick-and-mortar classrooms: Opportunities for further enhancement. *Computers & Education*, 227:105183.

Sarah E. Finch, Eun Sook Paek, and Jinho D. Choi. 2023. Leveraging large language models for automated dialogue analysis. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 202–215. Association for Computational Linguistics.

Ofelia García and Angel M. Y. Lin. 2017. Translanguaging in bilingual education. In Ofelia García, Angel M. Y. Lin, and Stephen May, editors, *Bilingual and Multilingual Education*, 2 edition. Springer.

Ofelia García and Li Wei. 2015. Translanguaging, bilingualism, and bilingual education. In Wayne E. Wright, Suhan Boun, and Ofelia García, editors, *The Handbook of Bilingual and Multilingual Education*. Wiley.

M. A. K. Halliday. 1993. Towards a language-based theory of learning. *Linguistics and Education*, 5(2):93–116.

M. A. K. Halliday. 1994. *An Introduction to Functional Grammar*. Edward Arnold.

M. A. K. Halliday and Ruqaiya Hasan. 1985. *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. Deakin University Press.

Jennifer Mary Hammond and Pauline Gibbons. 2005. Putting scaffolding to work: The contribution of scaffolding in articulating esl education.

Ruth Harman. 2017. Bringing it all together: Critical take(s) on systemic functional linguistics. In Ruth Harman, editor, *Bilingual Learners and Social Equity*. Springer.

Matthew Honnibal and James Curran. 2007. Creating a systemic functional grammar corpus from the penn treebank. In *Proceedings of the Workshop on Deep Linguistic Processing*, pages 89–96. Association for Computational Linguistics.

Sally Humphrey and Lucy Macnaught. 2015. Functional language instruction and the writing growth of english language learners in the middle years. *TESOL Quarterly*, 50(4):792–816.

Rodney Jones and Graham Lock. 2011. *Functional Grammar in the ESL Classroom: Noticing, Exploring and Practicing*. Palgrave Macmillan.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Daniel Jurafsky and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*, 3 edition.

Nihal Khote and Zheng Tian. 2019. Translanguaging in culturally sustaining systemic functional linguistics. *Linguistics and Education*.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Anne Lane and Michael L. Kent. 2018. Dialogic engagement. In Kevin A. Johnston and Maureen Taylor, editors, *The Handbook of Communication Engagement*.

J. R. Martin and David Rose. 2007. *Working with Discourse: Meaning Beyond the Clause*, 2 edition. Continuum.

Andrea Martinenghi, Gregor Donabauer, Simona Amenta, Sathya Bursic, Mathyas Giudici, Udo Kruschwitz, Franca Garzotto, and Dimitri Ognibene. 2024. LLMs of catan: Exploring pragmatic capabilities of generative chatbots through prediction and classification of dialogue acts in boardgames' multiparty dialogues. In *Proceedings of the 10th Workshop on Games and Natural Language Processing @ LREC-COLING 2024*, pages 107–118, Torino, Italia. ELRA and ICCL.

Neil Mercer. 2004. Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, 1(2):137–168.

Samuel Messick. 1995. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9):741–749.

Ricardo Otheguy, Ofelia García, and Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review*, 6(3):281–307.

Ayesha Qamar, Jonathan Tong, and Ruihong Huang. 2025. Do LLMs understand dialogues? a case study on dialogue acts. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 26219–26237, Vienna, Austria. Association for Computational Linguistics.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python nlp package for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108.

Carolyn P. Rosé, Oliver Ferschke, Varun Dutt, Yohan Jo, Yifei Cui, Rui Wang, Rohan Kumar, and Emily Fancher. 2019. Analyzing collaborative learning processes automatically: Exploiting the advances of natural language processing. *International Journal of Artificial Intelligence in Education*, 29(3):305–346.

Mary J. Schleppegrell. 2004. Review of classroom discourse analysis: A functional perspective by f. christie. *Language in Society*, 33(3):443–446.

Chen Sun and Carolyn P. Rosé. 2020. Measuring collaborative problem solving using computational linguistic analysis. In *Proceedings of the International Conference on Learning Analytics & Knowledge (LAK)*, pages 263–272.

Chen Sun, Valerie J. Shute, Angela Stewart, Jade Yonehiro, Nicholas Duran, and Sidney D'Mello. 2019. Towards a generalized competency model of collaborative problem solving. *Computers & Education*, 143:103672.

Francis J. Troyan, Jose D. Herazo, and Marina Ryshina-Pankova. 2022. Sfl pedagogies in language education: Special issue introduction. *System*, 104:102694.

Alyssa F. Wise and Ming Ming Chiu. 2014. Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics*, 1(2):48–71.

Marcelo Worsley and 1 others. 2021. Multimodal learning analytics: A review and future directions. *Journal of Learning Analytics*, 8(3):1–32.

Ruochen Zhang, Jan Christian Blaise Cruz, Genta Winata, and Alham Fikri Aji. 2023. Multilingual large language models are not (yet) code-switchers. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12567–12582.

A Language in Context

Cultural context			
Genre (purpose and audience)			
Register	*Field*	*Tenor*	*Mode*
Discourse semantics (Martin 2014)	***Ideational***	***Interpersonal***	***Textual***
	Experiential	*Negotiation*	*Identification*
	Logical	*Appraisal*	*Periodicity*
Lexico grammatical resources	***Participants*** *(nominal phrases/ groups)*	***Mood*** *in clause (declarative, interrogative, imperative).*	***Cohesive devices*** *(reference, repetition, ellipsis)*
	Processes *(verbs)*	***Modality*** *(type of modal verbs and adjuncts to express degrees of obligation, certainty)*	***Theme and Rheme sequencing*** *(point of departure in clauses, linking among themes in subsequent clauses)*
	Circumstances *(prepositional phrases, adverbials)*	***Polarity*** *(continuum of positive to negative)*	***Clause combining*** *(e.g., conjunction)*
	Logical relations among clauses and sentences *(e.g., hypotaxis or parataxis)*		

Figure 4: Language in context, adapted from (Harman, 2017).

B Linguistic Feature Labels

Table 4: Universal Dependencies (UD) feature labels.

Linguistic Feature	Category	Description
nsubj_cnt	UD	Grammatical subjects indicating who acts or experiences the process.
obj_cnt	UD	Direct objects representing affected or impacted participants.
obl_cnt	UD	Oblique dependents expressing time, place, or manner.
advmod_cnt	UD	Adverbial modifiers adding manner, time, degree, or stance.
advcl_cnt	UD	Adverbial subordinate clauses (e.g., because, if, when).
cc_cnt	UD	Coordinating conjunctions joining clauses or ideas.
mark_cnt	UD	Subordination markers signaling logical relations.
xcomp_cnt	UD	Open clausal complements extending processes.
ccomp_cnt	UD	Finite clause complements that project ideas (e.g., think, say).
discourse_cnt	UD	Discourse markers managing interaction and phase shifts.
parataxis_cnt	UD	Loosely attached clauses expressing sequencing or elaboration.
appos_cnt	UD	Appositions reclassifying or identifying participants.
cop_cnt	UD	Copular clauses expressing relational processes.
agent_like_cnt	UD	Semantic-role approximation of agents (derived from `nsubj`).
patient_like_cnt	UD	Semantic-role approximation of patients (derived from `obj`).
oblique_cnt	UD	Semantic obliques indicating circumstances.

Table 5: Parts-of-Speech (POS) feature labels.

Linguistic Feature	Category	Description
verbs_cnt	POS	Verb count indicating process density in the utterance.
participants_cnt	POS	Pronouns and articles indicating participant tracking.
circumstances_cnt	POS	Prepositions and adverbs expressing circumstances.
conj_cnt	POS	Basic coordinators creating cohesion.
subord_cnt	POS	Subordinators expressing logical relations.
connectors	POS	Discourse connectors organizing idea flow.
qmark	POS	Interrogative punctuation indicating question mood.
modals	POS	Modal verbs expressing obligation, possibility, or ability.
polarity_neg	POS	Negation markers.
eval_pos_adj	POS	Positive evaluative adjectives.
eval_neg_adj	POS	Negative evaluative adjectives.
pronouns	POS	Pronouns signaling engagement and reference chains.
epistemic	POS	Epistemic stance markers.
hedge	POS	Softening or mitigating expressions.
comparative_ref_cnt	POS	Comparative reference forms relating entities across discourse.
extended_demo_cnt	POS	Extended demonstratives indicating spatial, temporal, or discourse reference beyond the clause.
conj_additive_cnt	POS	Additive conjunctions extending or elaborating information.
conj_adversative_cnt	POS	Adversative conjunctions signaling contrast or counter-expectation.
conj_causal_cnt	POS	Causal conjunctions expressing cause–effect relations.
conj_temporal_cnt	POS	Temporal conjunctions sequencing events.
repetition_win5	POS	Lexical repetition within a five-sentence window.
demonstratives	POS	Demonstratives (this, that, these).
periodicity_z_in_speaker	POS	Normalized sentence index within a speaker's sequence.

Table 6: Dialogue Act (DA) feature labels.

Linguistic Feature	Category	Description
da_sd	DA	Statement–Non-Opinion (factual).
da_sv	DA	Statement–Opinion.
da_qy	DA	Yes/No question.
da_qw	DA	Wh-question.
da_qr	DA	Alternative question.
da_qo	DA	Other or rhetorical question.
da_b	DA	Backchannel minimal response.
da_bk	DA	Backchannel-as-answer.
da_aa	DA	Agreement or acceptance.
da_ar	DA	Disagreement or rejection.
da_h	DA	Hedging or uncertainty.
da_rp	DA	Repair, repetition, or clarification.
da_o	DA	Other or unclassified dialogue act.
da_wd	DA	Word-dependent or fragment.
da_dash	DA	Missing or undefined dialogue act label.
da_question	DA	Grouped category: any question dialogue act.
da_disagree	DA	Grouped category: disagreement.
da_uncertainty	DA	Grouped category: hedging or uncertainty.
da_agree	DA	Grouped category: agreement.
da_repair	DA	Grouped category: repair or clarification.
da_backchannel	DA	Grouped category: backchannels.

C Metafunction Label Distribution

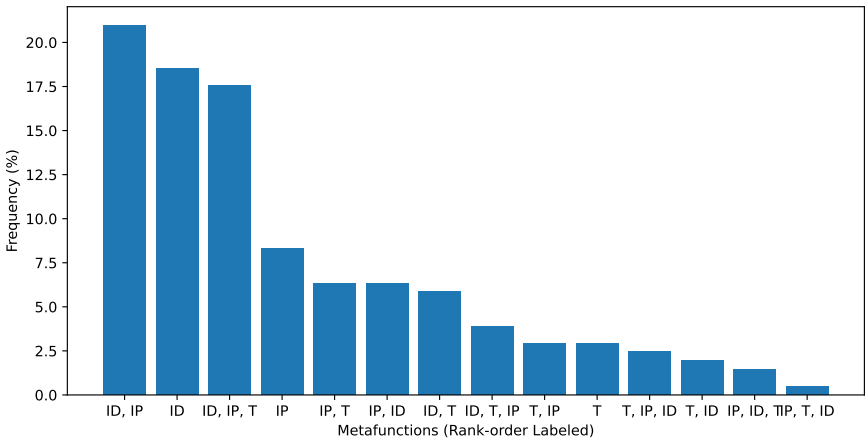

Figure 5: Distribution of adjudicated SFL tags under rank-order labeling. ID = ideational; IP = interpersonal; T = textual.