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

Anonymous ACL submission

Abstract

This paper examines how a Systemic Functional Linguistic (SFL) framework can inform the design of computational features for analyzing dialogic interactions in multilingual classrooms. Classroom discourse with multilingual language learners (MLLs) often involves fluid use of multiple linguistic resources, yet standard natural language processing (NLP) methods rely on surface linguistic features that are weakly aligned with the functional meanings through which ideas, stance, and cohesion are enacted in multilingual classroom interaction. To address this gap, we map core dimensions of SFL onto interpretable linguistic indicators, including part-of-speech patterns (POS), universal dependency (UD) relations, cohesion cues, and dialogue act (DA) categories. Correlation analyses show that POS and UD patterns align most strongly with ideational meaning, DA signals provide the clearest evidence of interpersonal meaning, and textual meaning is only weakly represented across all indicator types. These findings clarify which features most effectively capture functional aspects of language use in multilingual discourse and provide a foundation for developing more inclusive and scalable NLP approaches to analyzing classroom interaction.

1 Introduction

Classroom interaction involving multilingual language learners (MLLs) is often characterized by *translanguaging*, in which students flexibly mobilize resources from multiple languages, varieties, and registers to participate in joint activity (García and Lin, 2017). These practices are particularly evident in *small-group problem solving*, where learners collaboratively make sense of disciplinary content (e.g., scientific phenomena) by constructing explanations, negotiating epistemic stance, and coordinating roles and actions toward shared goals (Mercer, 2004; Chi, 2009; Sun et al., 2019).

Speaker	**Utterance**	**Meta-functions**
Yesenia	¿Qué pasó? ¿Qué vieron? *[What happened? What did you see?]*	Ideational, Interpersonal
Trevor	Matando a las ... um ... urchins *[Killing the ... um ... urchins]*	Ideational, Interpersonal
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.]*	Ideational

Table 1: Excerpt of Multilingual Discourse with SFL Metafunction Labels.

Growing interest in computational approaches to classroom discourse analysis reflects a broader effort to support instructional improvement and large-scale research (Rosé et al., 2019). In addition, analytic tools have been proposed to provide systematic insight into interactional practices such as questioning, uptake, and participation, and to enable comparisons of discourse processes across classrooms (Demszky et al., 2025). Extending these analytic approaches to multilingual classrooms is particularly promising for demonstrating how students flexibly coordinate multiple linguistic resources to engage in small-group discourse.

Despite this promise, recent work on AI-supported discourse analytics has identified persistent challenges in interpreting automatically derived linguistic features in ways that meaning-

fully support instructional analysis or feedback (Rodríguez-Ortiz et al., 2025). Furthermore, these analytic tools are typically trained and evaluated on monolingual or standardized corpora and treebanks, reflecting assumptions about linguistic uniformity that do not hold in multilingual interaction (Authors, 2025a; Joshi et al., 2020). While such features capture grammatical regularities, they are poorly aligned with the functional work of discourse (Blodgett et al., 2020).

To address this gap, we draw on Systemic Functional Linguistics (SFL), a theory of language explicitly designed to link linguistic form to social meaning (Halliday, 1993, 1994; Ballard, 1980). SFL conceptualizes language as a semiotic system organized around three metafunctions: ideational meaning, which construes experience and logical relations; interpersonal meaning, which enacts social roles and stance; and textual meaning, which organizes information flow and cohesion. Because these metafunctions connect linguistic choices to social activity, roles, and communicative contexts, they provide a principled framework for relating form and meaning (Table 1).

Systemic Functional Linguistics has supported fine-grained analyses of academic writing, classroom genres, and teacher feedback, including in multilingual settings (Halliday and Hasan, 1985; Martin and Rose, 2007; Troyan et al., 2022). However, prior computational work has largely focused on approximating SFL features from existing treebanks (Honnibal and Curran, 2007), leaving metafunctional patterning in naturally occurring multilingual discourse underexplored. As a result, current NLP systems lack principled mechanisms for approximating metafunctional meaning in multilingual interaction, limiting both analytic scope and model explainability.

To this end, this paper makes three main contributions: i) we introduce a dataset of 10 transcripts of small-group collaboration in multilingual classrooms that are annotated by experts for three SFL metafunctions; ii) we examine how linguistic features such as part-of-speech tags, syntactic dependencies, and dialog acts correlate with SFL metafunctions; iii) we assess the difficulty of automatically *predicting* SFL metafunctions at the utterance level, and show the potential and limitations of a baseline zero-shot prompting approach with large language models (LLMs).

2 Related Work

2.1 Systemic Functional Linguistics

Systemic Functional Linguistics (SFL), developed by 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 broader cultural and situational contexts, and much of its analytical power lies in specifying how meanings are instantiated in observable linguistic form during interaction.

A central contribution of SFL is its account of lexicogrammatical systems that realize metafunctional meanings in discourse. Harman's (2017) model of language in context foregrounds lexicogrammatical resources as a means of linking linguistic components to ideational, interpersonal, and textual meanings (see Figure 5 in Appendix A). From this perspective, metafunctions are enacted through concrete language choices, including participant and process types, mood and modality, and cohesive and thematic organization, rather than through contextual descriptors alone.

These lexicogrammatical systems have been used extensively to analyze curriculum genres, knowledge construction, and scaffolding practices in classroom discourse (e.g., Schleppegrell, 2004; Hammond and Gibbons, 2005; Jones and Lock, 2011; Humphrey and Macnaught, 2015; Khote and Tian, 2019).

2.2 Translanguaging in NLP

Recent advances in multilingual NLP have expanded model capacity for processing multiple languages and have begun to address phenomena such as code-switching and cross-lingual transfer. In sociolinguistics, code-switching refers to the alternation between languages within or across turns (García and Wei, 2015), while cross-linguistic influence describes how linguistic resources from one language shape meaning-making in another (Odlin, 2003). Although multilingual models increasingly attempt to account for both phenomena, they continue to struggle with naturalistic code-switching, particularly when switches occur mid-clause or serve discourse-pragmatic functions such as stance or alignment (Zhang et al., 2023).

More broadly, most computational approaches still treat languages as separable systems and eval-

uate performance against monolingual or standardized benchmarks (Authors, 2025a). This perspective underrepresents multilingual discourse in which speakers draw on an integrated semiotic repertoire to construct meaning (García and Lin, 2017). In educational settings, this creates challenges for accurately representing student contributions and for capturing discourse-level processes such as explanation-building, peer alignment, and sustained participation. While SFL-informed studies have documented these practices in multilingual classrooms (Khote and Tian, 2019; Troyan et al., 2022), they have rarely been connected to automated representations that support large-scale multilingual discourse analysis.

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. These transcripts came from longer classroom sessions that were part of a larger study of multilingual interaction in science classrooms; for the present analysis, we used pre-selected transcript segments capturing focal episodes of collaborative sensemaking and translanguaging. Across these segments, 123 students participated. The classrooms varied in grade level, science content, and linguistic demographics, with groups composed of MLLs with diverse language backgrounds (e.g., Spanish, Arabic) as well as monolingual English-speaking peers. The selected segments capture naturally occurring small-group problem-solving episodes in which students collaboratively interpret evidence, negotiate stance, and coordinate action using multilingual resources.

3.2 Human Annotation

Two bilingual annotators, fluent in English and Spanish, independently annotated all transcripts using a finalized annotation protocol. Both annotators had expertise in SFL and multilingual classroom discourse. The protocol provided a reference standard for evaluating whether computational linguistic features could approximate ideational, interpersonal, and textual meanings in student discourse.

The annotation task directly addressed the study's guiding research question: *How can the metafunctions of Systemic Functional Linguistics be conceptualized for analyzing collaborative multilingual classroom discourse and operationalized into automated linguistic features?*

Annotators identified all metafunctions expressed in each utterance, allowing for 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 how metafunction labels were distributed across the corpus based on adjudicated human annotations (Table 1). Multifunctional utterances were far more common than single-function realizations, with combined ideational–interpersonal labels representing 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.

This distribution highlights the multifunctional nature of students' multilingual peer discourse and underscores the analytic complexity of the task. Because utterances frequently expressed overlapping experiential, interactional, and organizational meanings, subsequent analyses of automated feature alignment must be interpreted in light of these densely layered communicative patterns.

Unit of Analysis The annotation unit was a single speaker turn defined as one sentence or short utterance. Multi-clause turns were annotated as a whole when they maintained a unified communicative intent.

Rank-Order Labeling Utterances often express more than one metafunction (see Table 1). In such cases, annotators selected all relevant labels and ranked them by prominence, with the primary communicative function marked first. For example, the utterance *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). Annotators relied on expert intuition rather than fine-grained grammatical analysis and used a notes field to justify decisions in borderline cases. This coarse-grained approach prioritized communicative function and dialogic engagement (Lane and Kent, 2018) over strict linguistic segmentation.

Reliability Inter-rater reliability was assessed using Cohen's kappa for each metafunction, yielding substantial agreement for ideational meaning ($\kappa = 0.551$), high agreement for interpersonal meaning ($\kappa = 0.693$), and moderate agreement for textual meaning ($\kappa = 0.444$), based on conventional interpretation thresholds (Landis and Koch, 1977). Overall micro-averaged agreement was $\kappa = 0.595$, with an exact set-level agreement of 0.620 (Table 3). Following reliability assessment, the primary author adjudicated all disagreements to produce a single gold-standard set of SFL labels used in subsequent analyses.

4 Methods

Unlike prior work that uses linguistic features as inputs to downstream predictive models, our goal is analytic rather than classificatory. We examine how theoretically motivated linguistic indicators align with expert judgments of functional meaning by operationalizing SFL metafunctions in collaborative, multilingual, small-group discourse. This is achieved through automatic language identification, Stanza parsing for English and Spanish (Qi et al., 2020), Universal Dependencies (UD) (de Marneffe et al., 2021), and large language model outputs for dialogue act (DA) classification (Jurafsky and Martin, 2025; Finch et al., 2023). For each utterance, Stanza was used to perform tokenization, lemmatization, part-of-speech (POS) tagging, and UD parsing.

We extracted and organized linguistic features from parser and model outputs. This aggregated token-level annotations into utterance-level representations, computed counts and proportions of POS categories and UD relations, and grouped related features into interpretable structural and discourse-relevant sets. Dialogue act labels generated by the LLM were standardized and aligned with the interpersonal metafunction.

All extracted features were merged with expert SFL annotations at the utterance level to create a unified analysis dataset.

4.1 Linguistic Features

To build interpretable computational indicators for SFL metafunctions, we constructed a feature set spanning POS information, UD relations, and DA categories. Each captures a distinct aspect of multilingual classroom interaction and contributes complementary evidence about meaning-making processes. For the complete list of linguistic features, see Appendix B.

POS and UD structure. POS tagging and UD parsing provide widely used structural representations 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 functions (e.g., nominal subjects, objects, obliques), making observable grammatical patterns available for analysis. These patterns align with SFL constructs, such as participants, processes, and circumstances.

Harman's (2017) model of language in context characterizes metafunctional meanings through lexicogrammatical resources rather than computational features (see Figure 5 in Appendix A). Building on this framing, we identified POS and UD patterns that plausibly correspond to these resources and can be extracted reliably using existing NLP tools. Features were selected if they (a) aligned with lexicogrammatical resources identified in SFL theory and (b) were robust to English–Spanish translanguaged discourse. This mapping reflects a methodological decision to operationalize SFL constructs through available syntactic representations.

To approximate aspects of textual organization, we additionally included UD relations such as coordinating conjunctions, subordinators, discourse marker attachments, and parataxis, which capture clause linkage and information sequencing across turns (Authors, 2025b). Based on this theory-informed selection process, we extracted 22 POS-based features and 15 UD relation features to represent structural patterns relevant to ideational and textual meaning.

Dialogue acts. Dialogue Act (DA) analysis provides a complementary layer of discourse-pragmatic information beyond structural representations. DA labels identify the communicative function of an utterance, including statements, questions, acknowledgments, agreements, disagreements, and assessments (Jurafsky and Mar-

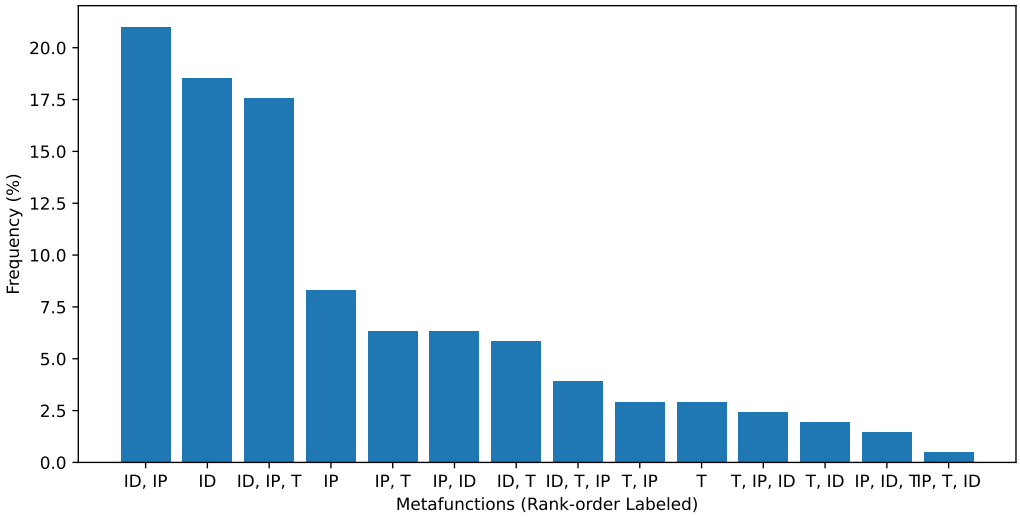

Figure 1: Distribution of adjudicated SFL tags under rank-order labeling. ID = ideational; IP = interpersonal; T = textual.

tin, 2025). To generate DA labels, we followed the LLM-based annotation approach described by Finch et al. (2023) by prompting ChatGPT-5.1 to assign dialogue acts to utterances based on the Switchboard taxonomy (Jurafsky and Martin, 2025). This approach yielded fine-grained pragmatic distinctions, including stance expressions, alignment and disalignment moves, disagreement, and questioning. Dialogue act labeling and SFL metafunction labeling were treated as separate annotation tasks. Dialogue acts were used solely as linguistic features, whereas metafunction labels served as the analytic outcome variable.

Because dialogue acts closely align with interpersonal meaning in SFL, we selected 20 theory-informed dialogue act features representing stance, negotiation, and social alignment.

Mapping to SFL metafunctions. Table 2 illustrates how selected POS, UD, and DA features align with SFL metafunctions. These features form the representational basis for the SFL-informed analysis presented in Section 4, where correlation-based heatmaps show how structural and pragmatic cues collectively contribute to identifying ideational, interpersonal, and textual meanings. Importantly, alignment between a feature and a metafunction does not imply that the feature fully realizes that metafunction; rather, it indicates a statistically detectable association between surface linguistic cues and expert interpretations of meaning.

4.2 Evaluation

After annotation, POS-based features, UD relations, and DA categories were extracted and merged with the adjudicated expert SFL labels at the utterance level. Spearman correlations were then computed to quantify associations between each linguistic feature and the presence of ideational, interpersonal, and textual meanings. This approach enabled systematic comparison of how structural indicators (POS and UD) and discourse-pragmatic cues (DA) related to the three metafunctions in the expert-annotated data.

LLM Classification of SFL Labels We assessed the potential of automatic qualitative discourse analysis tools. Following Finch et al. (2023), we prompted ChatGPT-5.1 (OpenAI, 2025) using the same finalized annotation instructions provided to human annotators. The prompt framed the model as an expert in Systemic Functional Linguistics and instructed it to assign one or more metafunction labels to each utterance based on its communicative function (see Appendix C).

5 Results

This section reports analyses examining the alignment between automated linguistic features and human-assigned SFL metafunction labels. We summarize patterns of metafunction use across the cor-

Feature	What It Measures	SFL Metafunction	Example
`verbs_cnt` (POS)	Verb realizations indicating processes being construed in the utterance.	Ideational (processes)	*The water heats and evaporates.* → verbs = *heats*, *evaporates*
`nsubj_cnt` (UD)	Grammatical subjects indicating entities participating in a process.	Ideational (participants)	*The salt dissolves.* → `nsubj` = *salt*
`da_qw` (DA)	Dialogue acts labeled as wh-questions.	Interpersonal (stance, inquiry)	*Why does the temperature change?* → `da_qw` = *Why*

Table 2: Mapping Between Lexico-grammatical Features and SFL Metafunctions.

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

Table 3: Human–Human Agreement for Metafunction Labels per Utterance.

Metafunction	κ
Ideational	0.372
Interpersonal	0.232
Textual	-0.010
Overall micro	0.288
Exact set-level agreement	0.249

Table 4: Human–LLM Agreement for Metafunction Labels per Utterance.

pus to characterize the multifunctional nature of the data, report agreement between human annotators and an LLM using Cohen's kappa (Table 4), and present Spearman correlation analyses visualized with heatmaps for POS features (Figure 2), UD relations (Figure 3), and dialogue act features (Figure 4).

5.1 Ideational meaning

Ideational meaning showed the strongest and most consistent alignment with structural linguistic features. As visible in the POS heatmap (Figure 2), verb counts and participant-related POS categories produced the highest positive correlations with ideational labels (e.g., $\rho \approx .30$–$.38$). The UD heatmap (Figure 3) shows a similar pattern, with subject relations (`nsubj`) reaching $\rho \approx .31$, alongside additional positive contributions from `obj`, `obl`, and adverbial modifiers (`advmod`, `advcl`). These patterns indicate that participant–process structures provide the most robust automated cues for ideational work.

Human–LLM agreement for ideational meaning was moderate ($\kappa = 0.372$; Table 4), indicating that experiential content is comparatively accessible for automated approximation, though substantial gaps remain relative to expert annotation.

5.2 Interpersonal meaning

Interpersonal meaning correlated most strongly with dialogic behavior rather than structural features. As shown in Figure 4, questions, (dis)alignment moves, expressions of uncertainty, and backchannels produced the clearest positive associations with interpersonal labels (e.g., $\rho \approx .14$–$.23$). In contrast, POS and UD features showed weak or negative correlations (Figures 2 and 3), reflecting the interactionally focused nature of interpersonal work in small-group science talk.

Human–LLM agreement for interpersonal meaning was relatively low ($\kappa = 0.232$; Table 4), suggesting that interactional nuance and stance remain challenging for automated models to approximate.

5.3 Textual meaning

Textual meaning showed moderate associations with features related to cohesion and clause linkage. In the UD heatmap (Figure 3), coordinating conjunctions (`cc`) and parataxis relations exhibited the strongest textual correlations ($\rho \approx .19$–$.26$). The POS heatmap (Figure 2) further shows that connectors, temporal conjunctions, and certain demonstratives contributed smaller positive correlations ($\rho \approx .03$–$.21$). These results suggest that local cohesion markers capture only a portion of how students sequence reasoning across turns in multilingual peer discourse.

Human–LLM agreement for textual meaning was near zero ($\kappa = -0.010$; Table 4), indicating that discourse organization and information flow were the least recoverable metafunctions for auto-

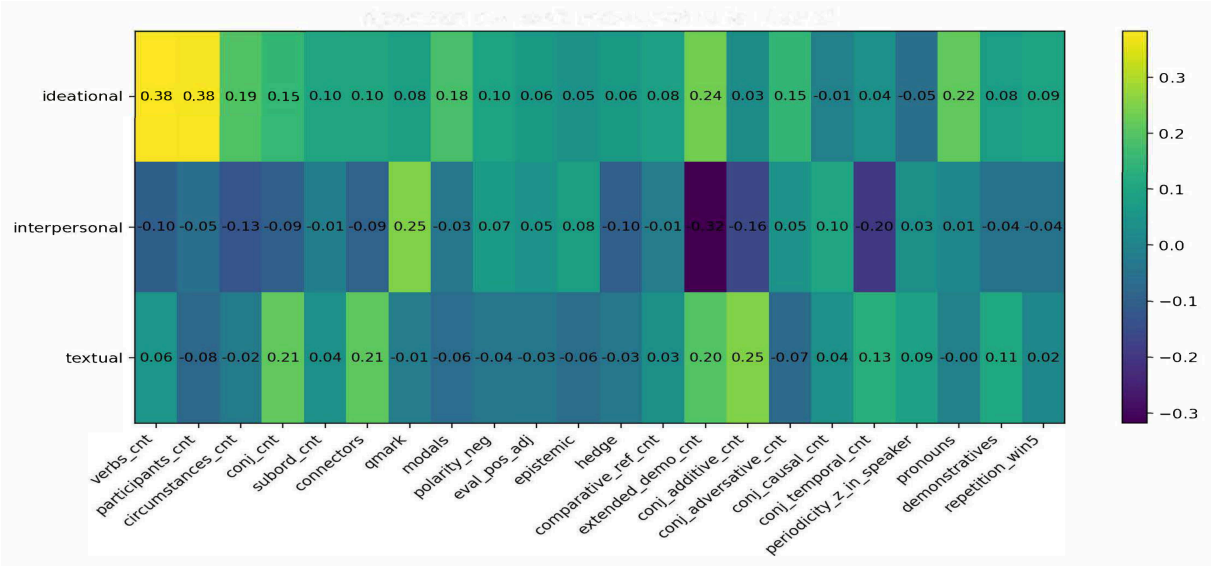

Figure 2: Spearman correlations of Parts-of-Speech features and SFL metafunction labels.

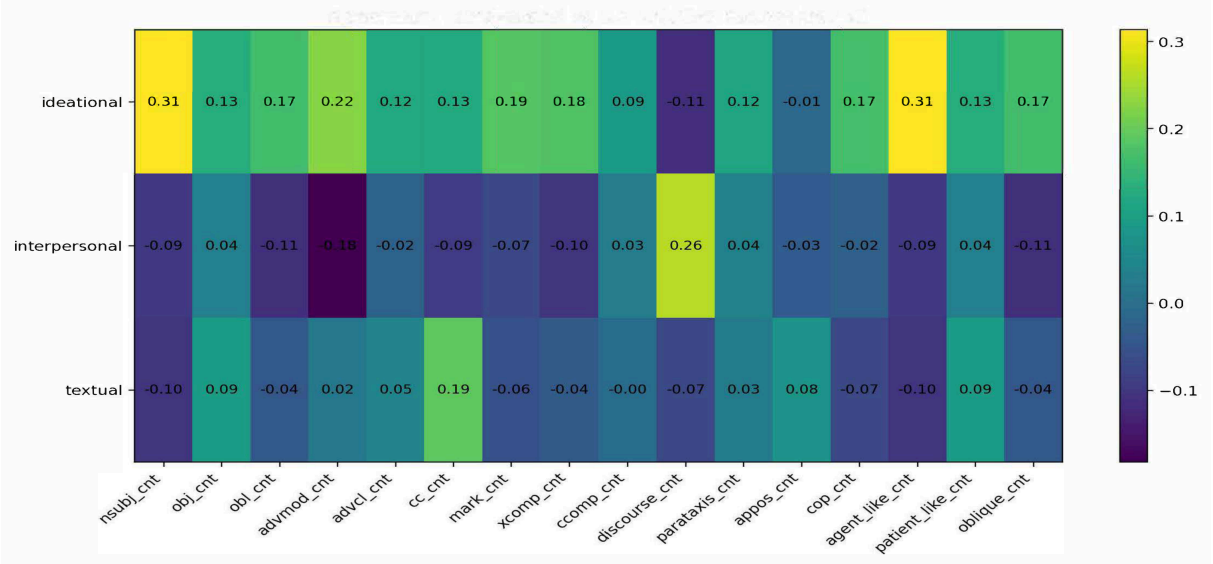

Figure 3: Spearman correlations of Universal Dependencies relations and SFL metafunction labels.

mated approximation.

5.4 Distinctive metafunction profiles

Across the three heatmaps, the metafunctions displayed clear separability. Ideational and interpersonal meaning showed consistent negative or near-zero correlations, indicating that content-heavy turns rarely coincide with interactional negotiation. Ideational and textual features exhibited correlations near zero ($|\rho| < .10$), reflecting independence between generating scientific ideas and structuring discourse flow. Textual and interpersonal meaning also remained largely distinct, with small negative or near-zero correlations except in cases where dialogic cues (e.g., clarification questions) simultaneously functioned as transitional markers. Taken together, these patterns support the SFL view that metafunctions represent complementary but distinct dimensions of meaning.

6 Discussion

The results illustrate both the potential and the limits of approximating SFL metafunctions through computational features in multilingual small-group science discourse. Overall, the findings clarify how different metafunctions are realized in peer interaction and which aspects of meaning are most directly reflected in available linguistic representations.

Ideational meaning aligned most strongly with structural features such as subjects, verbs, and

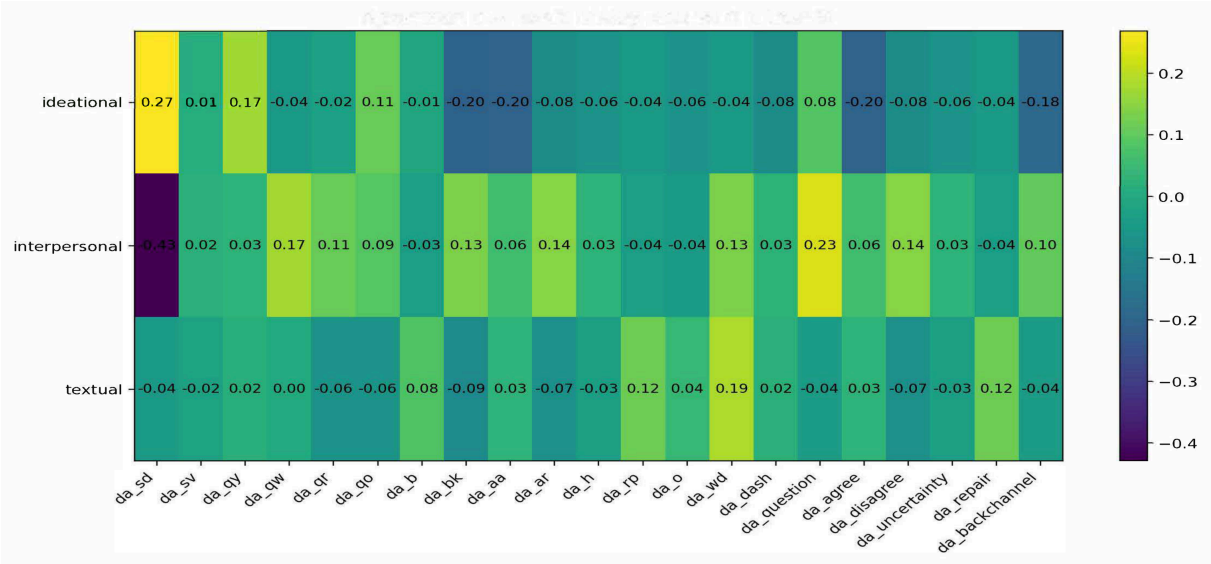

Figure 4: Spearman correlations of Diagnostic Acts and SFL metafunction labels.

objects, reflecting the central role of participant–process relations in construing experience. POS and UD features captured core experiential content but did not fully represent the layered meanings often compressed into minimal multilingual utterances. This pattern explains why ideational meaning showed the strongest alignment between automated features and expert labels, as it is typically expressed through overt grammatical roles explicitly encoded in syntactic representations.

Interpersonal meaning aligned primarily with dialogic behaviors rather than structural density. Questioning, alignment and disalignment moves, and backchannels showed the strongest associations, while POS and UD features contributed little. This reflects the pragmatic, context-sensitive nature of interpersonal work in collaborative discourse, which depends on sequential positioning and local interactional framing that are difficult to infer from isolated utterances. Correspondingly, automated dialogic labeling showed limited alignment with expert judgments.

Textual meaning exhibited moderate associations with coordination and cohesion markers but remained the weakest metafunction for automated approximation. Much discourse organization in inquiry dialogue unfolds across turns through shared reference and thematic continuity, which sentence-level structural features cannot adequately capture.

Methodologically, these findings demonstrate that an SFL-guided feature design can approximate metafunctional distinctions when features reflect relevant linguistic resources. Structural features provide the clearest access to ideational meaning, pragmatic indicators support interpersonal analysis, and textual meaning remains challenging due to its distributed nature. Together, the results highlight the need for approaches that integrate structural, pragmatic, and discourse-level information to more fully represent how multilingual learners organize meaning during collaborative problem solving.

7 Conclusion

The study provides an exploratory approximation of SFL metafunctions in multilingual small-group science discourse. Ideational meaning aligned most strongly with UD and POS structural features, interpersonal meaning with LLM-based dialogue act indicators, and textual meaning with cohesion-related cues. Human–human reliability was strongest for interpersonal and ideational meanings, whereas human–LLM agreement was highest for ideational meaning. Together, these findings establish a foundation for linking SFL constructs to automated feature extraction at scale and clarify which aspects of metafunctional meaning are most directly reflected in structural linguistic representations.

8 Limitations

The feature sets do not yet fully capture discourse-level organization, culturally shaped stance expressions, or multilingual cohesion patterns central to translanguaged discourse. Dialogue act patterns are sensitive to LLM prompting and may shift across model versions. Human–LLM agreement patterns

show that models remain far less sensitive than humans to interpersonal nuance and cross-turn cohesion. The dataset remains relatively small and limits broad generalization. Future work should incorporate larger multilingual corpora, expanded cohesion modeling, and open-source dialogic resources that support reproducible analysis.

References

Authors. 2025a.

Authors. 2025b.

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.

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.

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

Terence Odlin. 2003. Cross-linguistic influence. In Catherine J. Doughty and Michael H. Long, editors, *The Handbook of Second Language Acquisition*, pages 436–486. Blackwell.

OpenAI. 2025. Chatgpt-5.1 model release notes.

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.

Miguel Ángel Rodríguez-Ortiz, Pedro C. Santana-Mancilla, and Luis E. Anido-Rifón. 2025. Machine learning and generative ai in learning analytics for higher education: A systematic review of models, trends, and challenges. *Applied Sciences*, 15(15):8679.

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, 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.

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

B Linguistic Feature Labels

Table 5: 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 6: 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 7: 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 LLM Prompt for SFL Metafunction Annotation

Prompt

You are an expert in *Systemic Functional Linguistics* with experience analyzing multilingual classroom discourse.

You will be given a transcript consisting of short speaker turns from small-group classroom interaction. Your task is to assign *Systemic Functional Linguistics (SFL) metafunction labels* to each utterance based on its communicative function.

1. Metafunction Labels

For each utterance, identify which metafunction(s) it expresses. You may choose one, two, or all three, ranked from primary to secondary (and tertiary, if applicable):

Ideational: describing an object, process, idea, or event; explaining how something works; evaluating feasibility or accuracy; or providing information about what is happening.

Interpersonal: expressing stance, feelings, uncertainty, or attitude; negotiating agreement or disagreement; persuading; softening; or positioning oneself socially with peers or the teacher.

Textual: organizing or sequencing ideas; referencing prior talk (e.g., *because, so, then*); managing cohesion; or connecting parts of the conversation.

2. Unit of Analysis

Each speaker turn should be treated as one unit of analysis, defined as a single sentence or short utterance. If a turn contains multiple clauses but expresses one continuous communicative intent, label the entire turn rather than segmenting it.

3. Multiple Labels and Ranking

Some utterances express more than one metafunction. When this occurs, rank the metafunctions from strongest to weakest alignment, with the primary communicative purpose listed first.

Example

"I don't think that will work . . . probably actually . . . yeah it probably would."

Primary: Interpersonal (expressing uncertainty, revising stance)
Secondary: Ideational (evaluating feasibility)

4. Decision Rules

This is a coarse-grained interpretive task. Use your expert intuition and do not overanalyze. If an utterance is ambiguous, select the label(s) that best capture its overall communicative function.