Same Spelling, Different Functions *Mah*: Evaluating Language Models' Understanding of Singlish Particles

Chan Young Jung
Department of Linguistics, Korea University
laspebro@korea.ac.kr

Abstract

Singlish discourse particles exhibit tone-sensitive polysemy, where identical orthographic forms—distinguished only by prosodic cues in speech—serve distinct pragmatic functions. This poses a fundamental challenge for unimodal language models that must infer particle meanings solely from text. We thus investigate whether contextual information enables language models to predict appropriate particles in a cloze-style task, and whether increased data exposure—through domain-specific pretraining or in-context prompting—improves performance. To enable fair evaluation, we organize particles into semantic groups that minimize intra-group functional overlap. We test three BERT variants—including a Singlish domain-specific SingBERT model—and GPT-4.1 under zero-shot, definition-prompted, and few-shot conditions. Results demonstrate that domain-specific pretraining yields consistent performance gains over general English models (56.2% vs 30.1%), yet absolute performance remains modest across all approaches. GPT-4.1 shows variable performance across semantic groups and prompting strategies (23.8%–66.4%). These findings reveal that contextual cues only partially compensate for the absence prosodic information, highlighting fundamental limitations of text-only approaches for contact languages with substrate-derived pragmatic systems and the need for prosody-aware computational methods.

1 Introduction

Colloquial Singapore English (hereafter Singlish) is an English-based contact language that draws substrate influences from Singapore's multilingual landscape, including Malay, Tamil, and Sinitic varieties such as Hokkien and Cantonese (Deterding, 2007; Leimgruber, 2011; Chow and Bond, 2022; Ningsih and Rahman, 2023). A hallmark characteristic of Singlish is its extensive use of pragmatic discourse particles—such as *lah*, *leh*, *hor*, and *sia*—which, while removable without affecting grammaticality, encode important propositional content (Ler, 2006; Chow, 2021). Crucially, these particles are seldom monosemous; their meanings and discourse functions are jointly determined by contextual and prosodic cues (Lim, 2007; Wong, 2014; Soh et al., 2022). This tone-sensitivity potentially undermines the ability of unimodal language models to process Singlish, as they operate solely on orthographic input without prosodic notation.

Intra-particle polysemy is illustrated in the following sentences from the English subset of the National University of Singapore SMS corpus (Chen and Kan, 2015), a collection of over 55K messages in Singapore English:

(1) U typing the outline into the google doc *hor*? (#15340)

(2) Drive carefully when u come back *hor*... Raining heavily... (#15123)

In (1), *hor* functions as a confirmation-seeking question marker, converting the proposition into an interrogative while presuming its truth value. In contrast, in (2) *hor* adds precautionary force to an imperative, emphasizing the warning nature of the utterance. These functional distinctions are distinguished through rising versus falling tonal contours (Gupta, 1992; Kim, 2014; Lee, 2018; Chow, 2021; Liu et al., 2022; Chow et al., 2024). These examples illustrate that while prosodic cues disambiguate particle functions in speech, particle meanings in written Singlish must be inferred from context. This raises the question of whether contextual information alone enables models to predict the appropriate particle.

Since Singlish particles are embedded within English lexical items and syntactic structures, two possibilities arise: on the one hand, English-based models might benefit from cross-lingual transfer from high-resource English data, as demonstrated by Armstrong et al. (2022) on Jamaican Patois. Conversely, tone-conditioned discourse particles—which are absent from standard English varieties—may fall outside the latent knowledge models acquire through pretraining. We investigate whether increased data exposure, either in the form of language-specific training data or particle definitions and usage examples within a prompt, can compensate for limited exposure to Singlish during initial pretraining.

This work addresses the computational challenge of Singlish particle disambiguation through the evaluation of contemporary language models using a novel semantic grouping approach. We extract particle-containing sentences from the NUS-SMS corpus and develop a manually annotated subset. Drawing on extensive prior literature, we identify the pragmatic functions of 10 common Singlish particles and organize them into semantically coherent groups that minimize functional overlap within groups, while enabling fair comparison of model performance across different functional categories. We evaluate three masked language models with varying training data exposure—from general English pretraining (BERT-base-uncased) to multilingual training (BERT-base-multilingual-uncased) to domain-specific training on Singlish and Malaysian English texts (SingBERT)—alongside a generative model (GPT-4.1) tested under different data exposure conditions: zero-shot, few-shot, and definition-prompted settings.

Our contributions are threefold: (1) We develop a semantic grouping strategy for Singlish particles that minimizes intra-group functional overlap, enabling fair evaluation of model performance on distinct pragmatic functions; (2) We evaluate how different training data exposures affect particle function recognition across masked language models and generative models; (3) We provide empirical evidence that even domain-specific pretraining achieves only modest performance on tone-sensitive particles, demonstrating fundamental limitations of text-only approaches and the need for prosody-aware methods for contact languages.

2 Related Work

2.1 NLP for Creoles

Natural language processing research on creole languages has historically been sparse, despite creoles being spoken by hundreds of millions of people worldwide (Lent et al., 2021). This neglect stems from societal stigmatization rooted in colonial histories, their predominantly oral nature, and exclusion from major multilingual datasets and language family classifications (Lent et al., 2022b, 2024).

Creoles present computational challenges that distinguish them from typical low-resource scenarios. Unlike languages with clear genealogical lineages, creoles emerge from complex contact situations involving multiple substrate and superstrate languages. This mixed ancestry undermines standard transfer learning assumptions: Lent et al. (2022a) demonstrated that straightforward transfer from ancestor languages to creoles often fails to achieve expected performance gains, as lexical items may derive from one language while syntactic structures reflect another.

Recent efforts have expanded to include named entity recognition (Adelani et al., 2021), sentiment analysis (Muhammad et al., 2022), and comprehensive multilingual evaluation frameworks (Lent et al., 2024). However, semantic disambiguation challenges—particularly for substrate-derived features like Singlish discourse particles—remain largely unaddressed. Addressing these challenges requires approaches that account for the complex interplay between superstrate lexical foundations and substrate pragmatic systems, as we examine in the context of Singlish particle processing.

2.2 Computational Approaches to Singlish Particle Disambiguation

Computational approaches to Singlish discourse particles have emerged from diverse methodological directions, with early work focusing on syntactic representation rather than semantic disambiguation. Wang et al. (2017) conducted foundational work by creating a Universal Dependencies treebank for Singlish and training neural dependency parsers with neural stacking to integrate English syntactic knowledge. While achieving significant parsing improvements, their approach treated particles uniformly within grammatical frameworks rather than addressing their polysemous functions.

Rule-based approaches have attempted representation within formal grammar frameworks. Chow (2021) and Chow and Bond (2022) developed HPSG-based grammars representing sentence-final particles as heads selecting sentences as complements, organizing particles into hierarchical types based on positional constraints. Although structurally thorough, these approaches focus on syntactic distribution rather than semantic disambiguation.

Neural generation approaches have addressed particles within broader paraphrasing frameworks. Liu et al. (2022) integrated particle processing into Singlish-to-English translation through "semantic level rewriting," demonstrating that particles like *lah* (mood marker) and *leh* (tentative request marker) require clause-level understanding rather than word-level replacement. However, their approach did not specifically target disambiguation of particle functions based on prosodic or contextual cues.

Recent work has begun to explicitly address particle semantics. Chow et al. (2024) created SingDict, an open-source dictionary including particles with tonal annotations, while Foo and Ng (2024) specifically tackled disambiguation for three particles (*lah*, *meh*, *hor*) using task-driven representations with SingBERT, subtracting vector embeddings to isolate particle representations and performing unsupervised clustering to identify pragmatic functions. Current computational approaches to processing Singlish have also expanded to include style transfer (Liang et al., 2025), content moderation (Foo and Khoo, 2025), and multimodal understanding (He et al., 2025), reflecting growing recognition of Singlish's computational importance.

Our work differs from prior approaches in three key respects. First, rather than treating all particles uniformly (as in syntactic approaches) or focusing on individual particles in isolation (as in clustering-based methods), we explicitly address how tone-dependent polysemy creates overlapping pragmatic functions across particles. Second, we investigate whether varying levels of data exposure enable models to learn contextual patterns that compensate for missing prosodic information. Third, while previous work has primarily focused on syntactic parsing or translation, we directly evaluate models' capacity for particle prediction in authentic conversational Singlish, demonstrating that substrate-derived, tone-sensitive pragmatic features fundamentally limit models' processing of contact languages.

3 Methods

3.1 Dataset Construction and Particle Selection

We extract particle-containing sentences from the English subset of the National University of Singapore SMS corpus (Chen and Kan, 2015), which contains over 55,000 short, informal text messages from Singaporeans, primarily students at the National University of Singapore. The corpus provides naturalistic data with some chronologically ordered conversations, allowing for opportunistic retrieval of conversational context.

To identify target particles for analysis, we compiled definitions and functional descriptions from extensive prior research on Singlish discourse particles (Chow, 2021; Gupta, 1992; Kim, 2014; Khoo, 2012; Lee, 2018; Leimgruber, 2016; Leimgruber et al., 2021; Lim, 2007; Liu et al., 2022; Platt and Ho, 1989; Soh et al., 2022; Wong, 2014). We used regular expressions to extract sentences containing these particles, focusing specifically on substrate-derived particles (thus excluding *one* and *what*, which share orthographic forms with Standard English words despite having distinct pragmatic functions in Singlish).

Through manual inspection and frequency analysis, we identified 10 particles that were both frequent in the corpus and whose meanings could be reliably verified in the literature. Four particles exhibited tonal polysemy—distinct pragmatic functions associated with different tonal variants. Table 1 presents our final particle inventory with their tonal variants, pragmatic functions, syntactic environments, and number of appearances in our dataset.

We applied basic preprocessing including deduplication, removal of anonymization artifacts, filtering sentences with fewer than three words to ensure sufficient context, and exclusion of purely Mandarin or other substrate language content. For particles exhibiting tonal polysemy, we manually classified each instance into the closest functional variant based on contextual and syntactic cues, following the definitions established in our literature review.

3.2 Semantic Grouping Strategy

Rather than attempting simultaneous classification across all particle functions, which would unfairly penalize models due to substantial semantic overlap and an excessively wide range of candidate labels,

Particle	Tone	Pragmatic Function	DEC	IMP	INT	Count
ah	rising	confirms understanding/acknowledgement	+	+		425
	low	tag question/echo-question marker			+	367
bah	low	hedge; uncertainty/lack of commitment	+	+	+	316
hor	rising	question marker; speaker believes proposition true			+	107
	low falling	warning/disclaimer marker	+	+		50
lah	low	persuades acceptance of proposition	+	+		511
	falling	presents solutions; conveys annoyance	+	+		88
leh	mid-level	persuades action/belief acceptance	+	+		146
	low	marks new information/counters assumptions	+			672
liao	falling	past tense/perfective aspect marker	+	+	+	624
lor	mid-level	marks obviousness/resignation; agreements	+	+		957
mah	high	marks information as obvious	+			254
meh	high	question marker; skepticism/doubt			+	181
sia	rising	reduces distance; surprise/admiration	+		+	212
Total						**4,910**

Table 1: Singlish sentence-final particles with tonal variants, pragmatic functions, syntactic environments, and number of occurrences in our NUS-SMS subset. DEC = declarative, IMP = imperative, INT = interrogative. A "+" indicates the particle can occur in the sentence type. Particles with multiple rows show tone-sensitive polysemy.

Group	Particles	Count
1	low *ah*, rising *hor*, high *meh*	655
2	low *leh*, falling *liao*, high *mah*, rising *sia*	1,762
3	falling *lah*, mid-level *leh*, low falling *hor*, low *bah*	600
4	low *lah*, rising *ah*, mid-level *lor*	1,893

Table 2: Semantic grouping of Singlish particles to minimize functional overlap within groups.

Paradigm	Model	Data Exposure
BERT-based	BERT-base-uncased	English
	BERT-base-multilingual-uncased	Top 102 languages (Wikipedia)
	SingBERT	Singlish/Manglish (subreddits, forums)
GPT-4.1	Zero-shot	Baseline
	Few-shot	NUS-SMS sentences with particles
	Definition-prompted	Particle definitions

Table 3: Model configurations with their respective data exposure. BERT models are pretrained on the indicated corpora; GPT-4.1 variants receive information through in-context prompting.

we organized particles into four semantic groups that minimize functional overlap within groups while allowing fair evaluation across distinct pragmatic domains. Table 2 summarizes our semantic grouping with the distribution of instances across groups.

Our grouping strategy prioritizes syntactic and functional similarity while maintaining relatively balanced instance counts across groups. Group 1 comprises question markers that occur exclusively in interrogative contexts. Group 2 includes particles that function as non-imperative markers. Groups 3 and 4 organize the remaining particles using complementary pragmatic functions while avoiding intra-group semantic overlap.

3.3 Model Selection and Implementation

Table 3 summarizes our experimental setup across two paradigms: BERT-based models and GPT-4.1, selected to capture different aspects of particle understanding and exposure to training data.

Our evaluation task requires models to predict the correct particle for masked positions in authentic Singlish sentences. Within each semantic group, models select from 3–4 candidate particles (e.g., Group 1 candidates are low *ah*, rising *hor*, and high *meh*). The models predict which particle should appear in context, and we compare this against the actual particle found in the corpus.

BERT-based models. For masked language models, we implement a probabilistic scoring approach to handle particles that tokenize into multiple subwords. Given a sentence with [MASK] in the particle position, we expand the mask to accommodate the number of subwords in each candidate particle. For each candidate, we compute the cumulative log-probability by iteratively predicting each subword position, conditioning subsequent predictions on previously selected tokens (greedy left-toright decoding). The candidate with the highest cumulative log-probability is selected as the model's prediction. We use BERT-base-uncased, BERT-base-multilingual-uncased, and SingBERT in their released forms with default tokenizers, without additional fine-tuning on our task.

GPT-4.1. For the generative model, we use structured prompts that present the cloze sentence and explicitly list the candidate particles for the given semantic group. We set `temperature=0.0` to ensure deterministic predictions and `max_tokens=5` to enforce concise responses containing only the predicted particle. For the few-shot condition, we provide 3 example sentences per particle drawn from the NUS-SMS corpus; these examples were held out from the test set to prevent data leakage. For the definition-prompted condition, we include functional descriptions of each candidate particle based on our literature review. Complete prompt templates and particle definitions are provided in Appendix A.

3.4 Evaluation Metrics

We report accuracy as our primary metric: the proportion of correct predictions out of all test instances. Given the substantial frequency imbalance across particles within groups (Table 1), we report both micro-averaged and macro-averaged metrics. Micro-averaged accuracy weights predictions by instance count, reflecting overall performance as weighted by the natural distribution of particles in conversational Singlish. Macro-averaged metrics compute unweighted averages across particles, revealing whether models perform consistently across all particle types regardless of frequency. For detailed group-level analysis (Tables 4–7), we report micro-averaged accuracy alongside macro-averaged precision and F1 scores to assess per-particle performance.

4 Results and Discussion

Figure 1 presents micro-averaged accuracy across all groups and models, while Figure 2 shows overall performance averaged across semantic groups. Detailed results for each semantic group are shown in Tables 4–7.

4.1 Domain-Specific Training Effects

SingBERT consistently outperforms both general English and multilingual BERT models across all semantic groups, achieving an overall micro-averaged accuracy of 56.2% compared to 30.1% for BERT-base and 34.1% for BERT-multi. However, these modest absolute performance levels highlight the fundamental difficulty of the particle disambiguation task when prosodic information is unavailable. The performance advantage is most pronounced in Group 2 (non-imperative markers), where SingBERT achieves 62.4% accuracy compared to 15.0% and 18.9% for the baseline models respectively, demonstrating the critical importance of domain-specific exposure to Singlish linguistic patterns.

Notably, the performance gap narrows in Group 3, where BERT-multi outperforms SingBERT (40.2% vs 36.7%). This suggests that cross-lingual transfer from substrate languages may benefit certain pragmatic functions that bridge multiple linguistic systems within the multilingual architecture.

Model	**Acc**	**Prec**	**F1**
BERT-base	52.4	39.4	32.5
BERT-multi	51.9	35.9	30.8
SingBERT	65.5	63.4	59.1
GPT-4.1 (0-shot)	63.5	58.7	58.4
GPT-4.1 (def)	66.4	62.5	58.2
GPT-4.1 (few)	65.8	45.0	43.2

Table 4: Group 1 results: Question markers (low *ah*, rising *hor*, high *meh*). All models show relatively strong performance, with GPT-4.1 definition-prompted achieving highest accuracy. For all groups (Tables 4–7), accuracy is micro-averaged while precision and F1 are macro-averaged.

Model	**Acc**	**Prec**	**F1**
BERT-base	15.0	20.7	11.9
BERT-multi	18.9	20.7	17.5
SingBERT	62.4	58.9	52.3
GPT-4.1 (0-shot)	52.1	19.2	17.9
GPT-4.1 (def)	54.8	29.2	27.8
GPT-4.1 (few)	55.9	25.1	25.1

Table 5: Group 2 results: Non-imperative markers (low *leh*, falling *liao*, high *mah*, rising *sia*). Largest performance gap between SingBERT and other models, highlighting domain expertise importance.

4.2 Generative Model Performance

GPT-4.1 demonstrates variable performance across prompting strategies and semantic groups. Zero-shot micro-averaged performance ranges from 23.8% (Group 3) to 63.5% (Group 1), indicating substantial but uneven knowledge of Singlish particle functions. Definition prompting consistently improves performance across all groups, with the

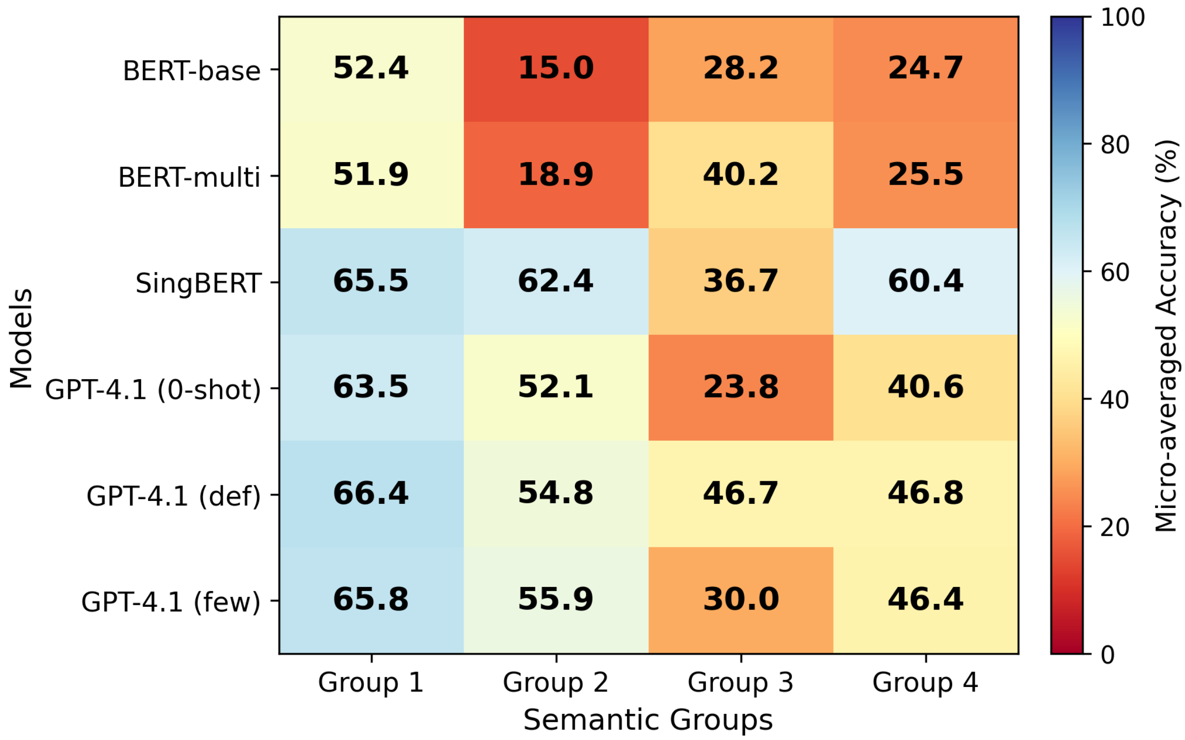

Figure 1: Model performance across semantic groups showing micro-averaged accuracy percentages. SingBERT demonstrates consistent performance across groups, while other models show variable effectiveness depending on pragmatic domain.

Model	Acc	Prec	F1
BERT-base	28.2	21.2	20.8
BERT-multi	40.2	26.3	25.3
SingBERT	36.7	46.3	37.4
GPT-4.1 (0-shot)	23.8	20.7	10.6
GPT-4.1 (def)	46.7	44.3	31.9
GPT-4.1 (few)	30.0	40.1	23.7

Table 6: Group 3 results: Mixed pragmatic functions (falling *lah*, mid-level *leh*, low falling *hor*, low *bah*). Most challenging group across all models, with definition prompting showing greatest improvement.

Model	Acc	Prec	F1
BERT-base	24.7	39.1	19.7
BERT-multi	25.5	36.3	19.8
SingBERT	60.4	57.6	55.0
GPT-4.1 (0-shot)	40.6	28.3	23.8
GPT-4.1 (def)	46.8	47.6	45.1
GPT-4.1 (few)	46.4	49.6	46.1

Table 7: Group 4 results: Persuasive/confirmatory markers (low *lah*, rising *ah*, mid-level *lor*). SingBERT shows strong performance, with consistent improvement across GPT-4.1 prompting strategies.

most substantial gains in Group 3 (46.7% vs 23.8% zero-shot). This improvement pattern indicates that GPT-4.1 possesses latent knowledge about Singlish pragmatics that can be activated through appropriate metalinguistic scaffolding.

Few-shot prompting shows mixed results, sometimes underperforming definition prompting (Group 3: 30.0% vs 46.7%), suggesting that metalinguistic guidance may be more effective than exemplar-based learning for this task.

4.3 Semantic Group Analysis

Group 1 (question markers) shows the most consistent performance across models, with all models except BERT-base achieving above 50% accuracy. This suggests that interrogative particles may be more learnable due to their clearer syntactic constraints.

Group 2 exhibits the largest performance disparity between domain-specific and general models, highlighting the importance of exposure to Singlish-specific pragmatic patterns. The poor performance of general English models (15.0% and 18.9%) indicates that these particles encode discourse functions not readily transferable from standard English patterns.

Group 3 proves most challenging across all models, with no model achieving above 47% accuracy.

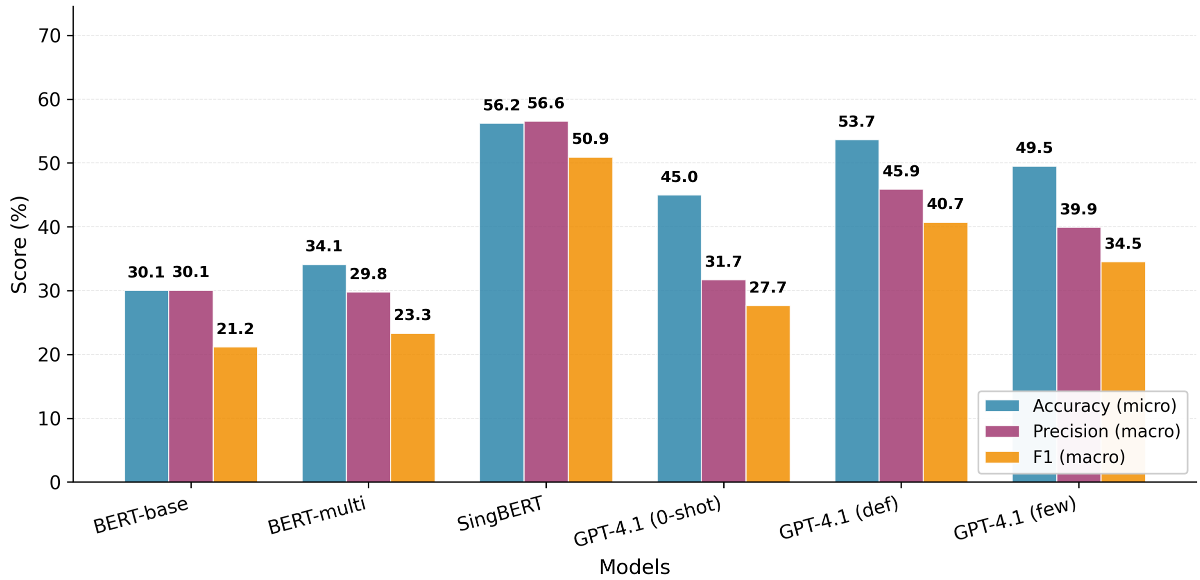

Figure 2: Overall performance comparison showing micro-averaged accuracy, macro-averaged precision, and macro-averaged F1 scores, each averaged across the four semantic groups.

This group contains particles with the most complex pragmatic functions and highest degree of contextual sensitivity, suggesting that current approaches struggle with highly context-dependent semantic disambiguation.

Group 4 shows strong performance for SingBERT (60.4%) and moderate but consistent improvement for GPT-4.1 across prompting strategies, indicating that persuasive and confirmatory functions may be more accessible to computational models.

4.4 Implications for Contact Language Processing

Our findings reveal several key implications for computational approaches to contact languages. The consistent benefits of domain-specific pretraining underscore the necessity of specialized training data for creole language processing. However, the universally modest absolute performance levels—even SingBERT achieves only 56.2% overall accuracy—point to fundamental limitations in current text-only approaches when pragmatic meaning is prosodically encoded. The variable effectiveness of different prompting strategies suggests that large generative models possess relevant but unevenly accessible knowledge about contact language features, with definition prompting (53.7%) substantially outperforming zero-shot approaches (45.0%) while few-shot prompting shows inconsistent benefits (49.5%).

5 Conclusion

This study demonstrates that tone-sensitive pragmatic phenomena in Singlish discourse particles expose fundamental limitations of contemporary language models operating solely on orthographic input. Even with domain-specific pretraining, performance remains far from human-level, underscoring the difficulty of capturing prosodically encoded meaning from text alone. These findings highlight that pragmatic interpretation in contact languages cannot be reduced to surface form recognition: it requires sensitivity to prosody, stance, and interactional context.

Our semantic grouping framework provides a principled evaluation methodology that mitigates functional overlap between particles, offering an approach generalizable to other contact varieties with complex pragmatic systems. The framework reveals systematic performance patterns: interrogative markers (Group 1) achieve relatively consistent results across models, while non-imperative markers (Group 2) and context-dependent functions (Group 3) prove substantially more challenging, particularly for models lacking Singlish-specific training. The variable success of definition prompting—with gains of over 20 percentage points in Group 3—further indicates that large generative models contain latent knowledge of such systems, but that this knowledge requires explicit scaffolding to be reliably accessed.

Taken together, these results argue for expanding computational approaches to under-resourced contact languages beyond text-only evaluation toward multimodal, prosody-aware methods that recognize the interplay of substrate-derived pragmatic systems with lexifier structures. Beyond Singlish, this work illustrates how contact languages can serve as critical testing grounds for theories of meaning in NLP, revealing where current models succeed, where they fail, and what linguistic knowledge remains inaccessible through distributional learning alone.

Limitations

While our evaluation provides valuable insights into computational particle disambiguation, certain limitations must be acknowledged. First, our manual annotation of tonal variants introduces potential subjectivity, particularly for ambiguous cases where contextual cues are insufficient to determine intended prosodic realization. Given the complex sociolinguistic nature of Singlish and the inherent difficulty of defining "native speaker" status in a contact language context, some degree of interpretive judgment is unavoidable. We mitigated this through extensive consultation of established literature and consistent application of documented functional criteria.

Second, the NUS-SMS corpus, while representing a specific demographic (primarily university students), constitutes one of the few available naturalistic Singlish corpora with substantial particle usage. Despite its demographic constraints, this corpus provides valuable authentic data. Our semantic grouping strategy, while involving theoretical judgment about functional similarity, is grounded in established pragmatic distinctions documented in extensive prior literature on Singlish discourse particles.

Finally, our evaluation framework focuses on Singlish and textual particle prediction, which may limit generalizability to other contact languages or multimodal contexts. Future work incorporating prosodic information and expanding to additional creole varieties could enhance our understanding of computational challenges across diverse contact language phenomena.

Acknowledgments

This research was supported by the BK21 FOUR (Fostering Outstanding Universities for Research) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF). We acknowledge the creators of the NUS-SMS corpus for making their data available for research purposes.

References

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, and 1 others. 2021. Masakhaner: Named entity recognition for african languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131.

Ruth-Ann Armstrong, John Hewitt, and Christopher Manning. 2022. Jampatoisnli: A jamaican patois natural language inference dataset. ArXiv preprint arXiv:2212.03419.

Tao Chen and Min-Yen Kan. 2015. The national university of singapore sms corpus. ScholarBank@NUS Repository, https://doi.org/10.25540/WVM0-4RNX.

Siew Yeng Chow. 2021. *A computational grammar of Singlish using HPSG*. Ph.D. thesis, Nanyang Technological University.

Siew Yeng Chow and Francis Bond. 2022. Singlish where got rules one? constructing a computational grammar for singlish. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5243–5250.

Siew Yeng Chow, Chang-Uk Shin, and Francis Bond. 2024. This word mean what: Constructing a singlish dictionary with chatgpt. In *Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI)@ LREC-COLING 2024*, pages 41–50.

David Deterding. 2007. *Singapore English*. Edinburgh University Press.

Jessica Foo and Shaun Khoo. 2025. Lionguard: A contextualized moderation classifier to tackle localized unsafe content. In *Proceedings of the 31st International Conference on Computational Linguistics: Industry Track*, pages 707–731.

Linus Tze En Foo and Lynnette Hui Xian Ng. 2024. Disentangling singlish discourse particles with task-driven representation. In *Proceedings of the 6th ACM International Conference on Multimedia in Asia Workshops*, pages 1–6.

Anthea Fraser Gupta. 1992. The pragmatic particles of singapore colloquial english. *Journal of Pragmatics*, 18(1):31–57.

Yingxu He, Zhuohan Liu, Geyu Lin, Shuo Sun, Bin Wang, Wenyu Zhang, Xunlong Zou, Nancy Chen, and Aiti Aw. 2025. Meralion-audiollm: Advancing speech and language understanding for singapore. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 22–30.

Velda Yuan Ling Khoo. 2012. The sia particle in colloquial singapore english.

Chong-Hyuck Kim. 2014. Discourse particles: Focusing on colloquial singapore english. *The New Studies of English Language and Literature*, 59:225–248.

Si Kai Lee. 2018. *A Nanosyntactic approach to sentence-final particles in Singlish: A cartographic perspective*. Ph.D. thesis, National University of Singapore.

Jakob RE Leimgruber. 2011. Singapore english. *Language and Linguistics Compass*, 5(1):47–62.

Jakob RE Leimgruber. 2016. Bah in singapore english. *World Englishes*, 35(1):78–97.

Jakob RE Leimgruber, Jun Jie Lim, Wilkinson Daniel Wong Gonzales, and Mie Hiramoto. 2021. Ethnic and gender variation in the use of colloquial singapore english discourse particles. *English Language & Linguistics*, 25(3):601–620.

Heather Lent, Emanuele Bugliarello, Miryam de Lhoneux, Chen Qiu, and Anders Søgaard. 2021. On language models for creoles. *arXiv preprint arXiv:2109.06074*.

Heather Lent, Emanuele Bugliarello, and Anders Søgaard. 2022a. Ancestor-to-creole transfer is not a walk in the park. *arXiv preprint arXiv:2206.04371*.

Heather Lent, Kelechi Ogueji, Miryam de Lhoneux, Orevaoghene Ahia, and Anders Søgaard. 2022b. What a creole wants, what a creole needs. *arXiv preprint arXiv:2206.00437*.

Heather Lent, Kushal Tatariya, Raj Dabre, Yiyi Chen, Marcell Fekete, Esther Ploeger, Li Zhou, Ruth-Ann Armstrong, Abee Eijansantos, Catriona Malau, and 1 others. 2024. Creoleval: Multilingual multitask benchmarks for creoles. *Transactions of the Association for Computational Linguistics*, 12:950–978.

Vivien Soon Lay Ler. 2006. A relevance-theoretic approach to discourse particles in singapore english. In *Approaches to discourse particles*, pages 149–166. Brill.

Jinggui Liang, Dung Vo, Yap Hong Xian, Hai Leong Chieu, Kian Ming A Chai, Jing Jiang, and Lizi Liao. 2025. Colloquial singaporean english style transfer with fine-grained explainable control. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 26962–26983.

Lisa Lim. 2007. Mergers and acquisitions: on the ages and origins of singapore english particles 1. *World Englishes*, 26(4):446–473.

Zhengyuan Liu, Shikang Ni, Aiti Aw, and Nancy Chen. 2022. Singlish message paraphrasing: A joint task of creole translation and text normalization. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3924–3936.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Sebastian Ruder, Ibrahim Said Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahudeen Abdullahi, Anuoluwapo Aremu, and 1 others. 2022. Naijasenti: A nigerian twitter sentiment corpus for multilingual sentiment analysis. *arXiv preprint arXiv:2201.08277*.

Nourma Silvia Ningsih and Fadhlur Rahman. 2023. Exploring the unique morphological and syntactic features of singlish (singapore english). *Journal of English in Academic and Professional Communication*, 9(2):72–80.

John T Platt and Mian Lian Ho. 1989. Discourse particles in singaporean english: Substratum influences and universals. *World Englishes*, 8(2):215–221.

Ying Qi Soh, Junwen Lee, and Ying-Ying Tan. 2022. Ethnicity and tone production on singlish particles. *Languages*, 7(3):243.

Hongmin Wang, Yue Zhang, GuangYong Leonard Chan, Jie Yang, and Hai Leong Chieu. 2017. Universal dependencies parsing for colloquial singaporean english. *arXiv preprint arXiv:1705.06463*.

Jock O Wong. 2014. *The culture of Singapore English*. Cambridge University Press.

A Prompt Templates

A.1 Zero-shot Prompt

```
You are given a Singlish sentence with
a missing word marked as [MASK]. Fill
in the [MASK] with exactly one of the
following particles: {candidates}. Do
not output anything else.

Sentence: {cloze_text}
```

A.2 Definition-prompted Template

```
You are given a Singlish sentence with
a missing word marked as [MASK]. Fill
in the [MASK] with exactly one of the
following particles: {candidates}. Use
the following definitions to guide your
choice:

[Particle definitions inserted here
based on semantic group]

Do not output anything else.

Sentence: {cloze_text}
```

A.3 Few-shot Template

You are given a Singlish sentence with a missing word marked as [MASK]. Fill in the [MASK] with exactly one of the following particles: {candidates}. Below are example usages for each particle:

[Examples for each particle inserted here]

Do not output anything else.

Sentence: {cloze_text}

A.4 Particle Definitions by Semantic Group

A.4.1 Group 1: Question Markers

ah: question marker (tag question or echo-question), elicits affirmation or confirmation
hor: question marker; indicates that the speaker believes the proposition to be true
meh: question marker; indicates skepticism or doubt

A.4.2 Group 2: Non-imperative Markers

leh: marks new information or re-asserts old information, possibly to counter the addressee's assumptions; used for declaratives
liao: past tense/perfective aspect marker; used for declaratives, imperatives, interrogatives
mah: marks information as obvious; used for declaratives
sia: reduces the distance between interlocutors and marks coarseness, surprise, or admiration; used for declaratives and interrogatives

A.4.3 Group 3: Mixed Pragmatic Functions

lah: presents answers or solutions to questions or situations; conveys annoyance and unfriendliness towards the addressee; used for declaratives and imperatives
leh: persuades the addressee to take action or accept a belief; used for declaratives and imperatives
hor: indicates a warning or disclaimer; used for declaratives and imperatives
bah: hedge; marks uncertainty and lack of commitment about a proposition; used for declaratives, imperatives, interrogatives

A.4.4 Group 4: Persuasive/Confirmatory Markers

lah: persuades the addressee to accept a proposition the speaker believes to be true; used for declaratives and imperatives
ah: confirms the addressee's acknowledgement or understanding of a proposition; used for declaratives and imperatives
lor: marks obviousness and resignation; often used for agreements; used for declaratives and imperatives

A.5 Few-shot Examples by Semantic Group

A.5.1 Group 1 Examples

ah:
1. K.:)you are the only girl waiting in reception ah?
2. Oh all have to come ah?
3. K. Did you call me just now ah?

hor:
1. Hey, u haven't upload the latest copy hor?
2. U typing the outline into the google doc hor?
3. Dear so we waiting u at orchard hor? Head of train k

meh:
1. Can meh? Thgt some will clash... Really ah, i dun mind...
2. Now got tv 2 watch meh? U no work today?
3. Huh... U serious of poning ah... Deepavali not nxt wk meh?

A.5.2 Group 2 Examples

leh:
1. haha but no money leh... Later got to go for tuition...
2. Tmr v crowded leh, weekday go la...
3. Huh... I mean e orientation in e first wk leh... Not majors...

liao:
1. Juz now havent woke up so a bit blur blur... Dad went out liao...
2. not goin 4 any camps... My faculty camp oso over liao...
3. oredi on my way to e class liao...

mah:
1. C movie is juz last minute decision mah. Juz watch 2 lar...
2. U must key in the amount on top first mah
3. Lol because if im not there and you kena caught, it will be very awkward mah lol.

sia:
1. guess wad sia? i won preview tickets to this korean show!
2. Haha. Good what. Can earn another 1k plus. Rich sia u.
3. U so serious till hallucinate?! Serious sia! u better stop training

A.5.3 Group 3 Examples

lah:
1. Then give mine to the person who doesnt have it lah.
2. Okayokay but what's done is done lah.
3. Borrow the one at home lah. Also, camera i think no choice...

leh:
1. Let me know asap leh
2. What u all buying? Help me to buy leh. I go join u all now.
3. It's ok. I'm already at your place. Open the door for me leh.

hor:
1. U dun say so early hor... U c already then say...
2. Drive carefully when u come back hor... Raining heavily...
3. u so naughty!!!! dun sleep so late hor. hug you tight tight.

bah:
1. Shld be ard 4 to 5 bah. What time e thing starts ar..
2. Don't know leh. Maybe his office bah.
3. Okie... scarly u arrive first arh lol... I think he shld be going bah

A.5.4 Group 4 Examples

lah:
1. Ur haircut not bad lah, quite nice and dun really look gong.
2. i think no need lah..i go borrows from steve

3. Ohh.. I heard australia is good lah. Haha. I don’t intend to travel.

ah:
1. dun forget u still owe me a treat ah..haha
2. 630. Today ah! Later on... Dont be late... And dont gelek
3. fetch me at 6 ah.. arts there, e place u always pick me up one..

lor:
1. Anything lor up 2 u... Dun buy anything too expensive...
2. ya lor. as my friends doing agency job then many of them got more tutor than student.
3. No. They bound to tease at us, so just let them tease lor.