



Topic-specific vs Global lexical bundles and discourse-semantic text properties in L2 argumentative writing

 SARA ALJOHANI  
King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Research on formulaic language in L2 writing has demonstrated that the use of lexical bundles is crucial to disciplinary discourse and development and that bundle use varies across proficiency levels and tasks. However, little is known about which writing-relevant textual properties these lexical bundles reflect. Do texts that employ more formulaic expressions become more lexically specific and/or more referentially or causally cohesive, or does formulaic language simply signify discourse without contributing to meaning? This study addresses these questions by investigating whether lexical-bundle coverage is associated with three Coh-Metrix composites (i.e., Specificity, Referential Cohesion and Deep Cohesion) in a corpus of 480 Arabic-L1 English argumentative essays. More specifically, the paper explores two types of lexical bundles, which are topic-specific and global ones, and seeks to examine their relationships with the three composites. To achieve this, essay-level composite scores were modelled on topic-specific and global coverage using a multivariate multiple regression with fixed effects for prompt and educational level. The results demonstrate that topic-specific lexical bundles are significantly associated with Referential Cohesion, whereas global coverage is associated with Deep Cohesion. Neither type is reliably associated with Specificity. These findings suggest that topic-specific bundles pattern with stronger local semantic alignment, whereas global bundles are more closely linked to causal/argumentative scaffolding.

Keywords: Lexical bundles, Coh-Metrix, specificity, referential cohesion, deep cohesion, arabic L1, ESL/EFL argumentative writing, phraseological competence

Introduction

Formulaic sequences are widely recognised as a hallmark of proficient writing, yet they vary considerably in their functions. Corpus-based research has demonstrated that such recurrent sequences can serve diverse communicative purposes, such as referring to entities, processes, or settings, structuring discourse, or conveying stance (Biber et al., 2004; Hyland, 2008; Römer, 2010; Gray & Biber, 2013). Formulaic competence has both semantic and rhetorical dimensions, a distinction supported by research on lexical bundles (e.g., Hyland & Jiang, 2018; Shin & Won, 2024; Samraj, 2024; Xia & Kessler, 2025). Written texts also display varying levels of Specificity, Referential Cohesion, and Deep Cohesion (Graesser et al., 2004; McNamara et al., 2014; Paivio, 1986, 2014)—three discourse-semantic properties explored in this paper in terms of whether they are associated with different types of formulaic language. The rationale is that merely knowing that formulaic language matters is not enough for pedagogy and assessment. What is equally important is identifying which specific formulaic patterns predict key features of Specificity, Referential Cohesion and Deep Cohesion, and under what contextual or methodological controls (such as prompt types and educational level) these relationships hold.

Existing literature has explored the links between lexical bundle use and register or disciplinary variations, reporting inventories and functions that differ across spoken and written registers and across fields (Biber et al., 2004; Gray & Biber, 2013; Hyland, 2008). Studies have also identified the overuse of stance or text-organising bundles in student writing, as well as the underuse of referential (research-oriented) bundles compared with expert prose (Cortes, 2004; Chen & Baker, 2010; Hyland, 2008). Furthermore, research links lexical bundle use to rated writing quality, but not simply through raw frequency. Instead, the type and contextual fit of lexical bundles have been underscored as key factors influencing writing quality in both automated and human evaluations. Two essays may use the same number of bundles, but the one that uses functionally appropriate, structurally referential, and strongly associated sequences is rated better (Bestgen & Granger, 2014; Bestgen, 2017; Senel, 2025). Li and Fang (2022) also demonstrate that raters are sensitive to the combined effects of structure, function, frequency and association strength, concluding that appropriateness matters more than mere accumulation. However, little is known about the influence of different types of lexical bundles (i.e., topic-specific and global) on Specificity, Referential Cohesion, and Deep Cohesion.

This paper addresses the aforementioned gap by examining the relationship between lexical bundle coverage and Coh-Metrix composites, namely Specificity, Referential Cohesion, and Deep Cohesion, in a corpus of 480 argumentative essays written by Arabic L1 learners of English. It distinguishes between topic-specific and global lexical bundles to answer the following two research questions:

1. To what extent is overall lexical bundle coverage associated with the three Coh-Metrix composites (Specificity, Referential Cohesion, Deep Cohesion)?
2. To what extent are each of the topic-specific and global lexical bundles associated with Specificity, Referential Cohesion, and Deep Cohesion?

The contribution of this paper lies in distinguishing global lexical bundles from topic-specific ones and examining how each type is associated with Specificity, Referential Cohesion, and Deep Cohesion of L2 argumentative essays. In addition, it outlines practical pedagogical implications, particularly for the pedagogy of lexical bundles. The remainder of the paper is structured as follows. The literature review explores related works, including research on formulaic language, cohesion- and specificity-related indices of text and L2 writing development. Next, the methods and results are presented, followed by a discussion of the findings and limitations. Finally, the implications of the study are presented.

Literature Review

Lexical Bundles

Formulaic language is a broad term that encompasses a range of multiword units, including collocations, idioms, and lexical bundles (Wray, 2002). Among these, lexical bundles, the primary focus of this paper, are defined as “extended collocations of three or more words that statistically co-occur in a register” (Cortes, 2004, p. 400). They play an important role in research on second language (L2) writing proficiency. A related concept is formulaic competence, which refers to the ability to use lexical bundles appropriately (Wray, 2002). According to Bestgen (2017), formulaic competence is the native-like use of frequent word sequences and is “a key aspect in the development of L2 writing proficiency” as well as a feature of advanced writing ability that has often been overlooked in writing assessment (p. 65). Recent studies extend this view by showing that the structure, function, and strength of associations of multiword sequences predict L2 ratings beyond raw frequency (e.g., Crossley, 2020; Li & Fang, 2022). However, not much is known about how different types of lexical bundles influence writing quality.

Recent research has attempted to address this gap by examining the use of formulaic sequences in learner texts and linking them to writing quality (Bestgen & Granger, 2014; Li & Fang, 2022). For example, Bestgen and Granger (2014) illustrated that texts with highly associated lexical bundles receive higher scores than those with weaker pairings, regardless of frequency. Similarly, Bestgen (2017) demonstrated “a strong link between formulaic competence [operationalised via collocational indices] and raters’ judgements of text quality” (p. 74). Li and Fang (2022) reported that writing quality is influenced by the type and contextual fit of formulaic language. Overall, the literature identifies variables, such as frequency, strength of association, discourse function, and lexical bundle type, that influence writing quality. Such literature motivates the need to explore the association between topic-specific and global lexical bundles, on the one hand, and L2-writing-relevant text properties captured by Coh-Metrix composites, on the other.

In fact, the use of lexical bundles not only improves writing quality but also allows L2 learners to produce texts more efficiently. This is supported by theoretical frameworks, such as Dual-Coding Theory (Paivio, 1986, 2014) and Levels-of-Processing Theory (Craik & Lockhart, 1972), which propose that formulaic phrases are retrieved holistically and more quickly than novel combinations (Wray, 2002). This eases processing load for L2 learners and allows them to dedicate more attention to higher-level concerns, such as content and organisation. This is further supported by research cited above (e.g., Bestgen, 2017; Bestgen & Granger, 2014), which indicates that a greater use of formulaic expressions is associated with higher perceived fluency and is seen as more cohesive.

Furthermore, research shows that formulaic sequences serve several functions (Biber et al., 2004; Hyland, 2008; Gray & Biber, 2013). These can be categorised into referential bundles, which name entities, processes and settings; discourse-organising bundles, which are text-oriented sequences that structure arguments; and stance bundles, which help express evaluation or commitment. The use of such bundles by L2 learners has also been explored. Studies show that students tend to overuse clausal and stance bundles while proficient academic writers favour referential and text-organising bundles (Cortes, 2004; Hyland, 2008; Chen & Baker, 2010). Moreover, L2 writers typically use fewer and less varied lexical bundles.

Coh-Metrix Composites: Specificity, Referential Cohesion, and Deep Cohesion

Another strand of research concerns the discourse-semantic properties of texts, with a focus on Coh-Metrix composites used in this paper. Coh-Metrix is an automated text analysis tool that

computes theoretically motivated indices of cohesion, lexical sophistication/specificity, syntax, and discourse semantics from running text (Graesser et al., 2004; McNamara et al., 2014). Three Coh-Metrix composites are of particular importance here. The first, Specificity, indexes lexical granularity through inverse noun/verb hypernymy and polysemy, where higher values indicate more precise and less generic lexis that reduces interpretive ambiguity (Graesser et al., 2004; McNamara et al., 2014). The second, Referential Cohesion, supports the maintenance of a stable discourse representation. It measures how propositions are tied across adjacent sentences and larger spans through overlap in content words, nouns, and stems, as well as semantic similarity (McNamara et al., 2010). The third, Deep Cohesion, estimates the explicit signalling of underlying relations, such as cause, goal, and enablement, through causal and intentional connectives and verb structures, helping readers track relational structure beyond surface adjacency (Graesser et al., 2004; McNamara et al., 2014).

These composites target functions that readers rely on when constructing situation models, as emphasised by the Level-of-Processing theory, which suggests that deeper semantic processing can improve comprehension and memory (Craik & Lockhart, 1972). In addition, the Dual Coding theoretical framework can be relevant here. Although developed for imagery and concreteness, it also suggests that richer and more discriminable representations can aid processing, which seems compatible with gains in lexical specificity (Paivio, 1986, 2014). Empirical evidence further demonstrates that greater lexical specificity and cohesion are associated with higher-quality L1 and L2 writing (Crossley & McNamara, 2012; Crossley, 2020; McNamara et al., 2010). Other studies indicate that referential lexical bundles and causal signalling correlate with expert judgements of coherence and overall quality, whereas indiscriminate use of connectives can be neutral or even detrimental, especially for skilled readers (Crossley & McNamara, 2012; Graesser et al., 2011; McNamara et al., 2010). Similar investigations have revealed that lexical sophistication, specificity, and global and referential cohesion explain variance in essay scores across tasks (Crossley, 2020; Guo et al., 2013; Crossley & McNamara, 2016). It is also important to note that referential and deep (causal) cohesion are considered key dimensions along which texts vary (Graesser et al., 2004). Overall, these composites offer an automated measurement of lexical precision (Specificity), integration of ideas (Referential Cohesion) and relational signalling (Deep Cohesion). They are central to coherence assessments in L2 argumentative writing, and exploring their association with different types of lexical bundles holds importance for L2 writing pedagogy and assessment.

Arabic L1 ESL Learners

The corpus used in this paper is a collection of essays written by L1 Arabic learners of English. Therefore, it is useful to review existing research on the use of lexical bundles by Arabic L1 ESL/EFL learners. The literature suggests that Arab learners tend to overuse additive connectives (e.g., and, also, moreover) and rely on listing or additive progression in their writing, while experiencing difficulties with subordinating or contrastive relations, such as *because* and *although* (Al-Jarf, 2001). This is often attributed to transfer from Arabic discourse traditions, in which coordination and reiteration are valued for coherence (Connor, 1996). It should be noted, however, that evidence specific to Arabic-L1 bundle use remains limited and mixed, warranting caution in generalising findings beyond task and genre.

Cohesion vs. Coherence

Halliday and Hasan (1976) distinguish between cohesion and coherence, with the former referring to textual elements such as reference, repetition, and conjunction, and the latter denoting meaningful and interpretable discourse. However, subsequent research suggests a

cohesion–coherence paradox. While frequent conjunctive ties may support comprehension for less experienced readers, they can be redundant or even counterproductive for skilled readers, whose coherence-building depends more on discourse structure and lexical cohesion (McNamara et al., 1996; McNamara et al., 2010; Crossley & McNamara, 2012). As a result, coherence cannot be reduced to just connective density. While organising discourse and signalling stance are crucial, they do not necessarily enhance a text's specificity or referential integration. Therefore, it is essential to examine the relationship between topic-specific and global lexical bundles and Coh-Metrix composites of Specificity, Referential Cohesion, and Deep Cohesion of texts.

From the literature, it is evident that only a few studies have used a design that differentiates between topic-specific and global lexical bundles, controls for task and educational level, and examines how they both relate to Coh-Metrix composites at the essay level among Arabic L1 learners. The present study addresses this gap by using such a design, as detailed in the following section.

Methodology

This paper employed a cross-sectional, corpus-based design to investigate whether lexical-bundle coverage is associated with Coh-Metrix discourse-semantic composites in L2 argumentative writing (Biber et al., 2004; Hyland, 2008; Graesser et al., 2004; McNamara et al., 2014). Specifically, it investigates whether overall lexical-bundle coverage correlates with three Coh-Metrix composites (Specificity, Referential Cohesion, and Deep Cohesion) and whether topic-specific and global lexical bundles show distinct patterns of association with these outcomes. Lexical-bundle coverage is defined here following established practice in bundle research, i.e., quantifying the extent to which a text consists of recurrent multi-word sequences. The dependent variables are operationalised as three essay-level Coh-Metrix composites constructed from published indices.

Data Collection

The dataset comprised 480 essays retrieved from ARAB-CC: Learner Corpus of English Essays, an existing learner corpus of 957 essays written by Arabic L1 learners of English enrolled in secondary or tertiary education. The selection criteria included all argumentative prompts and their related essays, resulting in 17 prompts with 480 essays written by students across four educational levels (HS12, BA2, BA3, MA1). These prompts, along with their ARAB-CC prompt IDs and pertinent frequencies, are presented in Table 1.

The ARAB-CC is available on Sketch Engine and is licensed under the MIT License (Kilgarrieff et al., 2014). It is tagged by TreeTagger using the Penn Treebank tag set with modifications from Sketch Engine (Schmid, 1994; Kilgarrieff et al., 2014). A metadata list is also available, containing information such as educational level, essay ID, publication date, student IDs, and prompt IDs. This metadata enables handling individual essays as the unit of analysis. Student and prompt IDs were retained to model repeated production and topic effects using fixed effects. Because multiple essays were contributed by the same writers in our selected-prompt subset (140 writers; 263 essays; 51.4% with repeats; max = 5), inference was clustered by writer (Student ID). Since repeated authorship per prompt was sparse, there was no need to model author-specific effects. To ensure validity, a mixed-effects sensitivity check with random intercepts for both writer and prompt was conducted (see Appendix A, Table A5). The results confirmed that writer-level variance did not affect the main inferences.

Additionally, the n-gram tool in Sketch Engine was used to compile lists of both topic-specific and global lexical bundles. For each of the 17 prompts, the 100 most frequent trigrams and

Table 1. *Argumentative prompts and essay counts*

Prompt ID	Prompt	N
1	Do you think single-sex schools are a good idea? Why or why not	56
2	Does extending the school year sound like a good idea to you? Why or why not?	28
4	At what age should people be allowed to drive a car, and why?	16
6	Do you believe that homework is necessary? How much homework is needed?	34
7	Should students evaluate their teachers? Do you agree or disagree?	36
8	Should school start times be adjusted so that students can get more sleep in the morning?	32
10	Why do you think more women and girls are not pursuing careers in computing and tech fields?	49
11	Are paper books better than e-books?	18
14	Should tablet computers replace textbooks in schools?	34
15	Do people overuse air conditioning?	13
22	Young people enjoy life more than older people do	39
24	"A teacher's ability to relate well with students is more important than excellent knowledge of the subject being taught"	41
30	Do you think banning cars from entering the city centre is a good idea?	17
33	Fast food restaurants should be banned in the vicinity of schools. Do you agree or disagree with this suggestion?	19
35	Can Moroccan Arabic and Maltese language be considered a dialect of Arabic and explain why	27
36	Is computer communication good or bad for developing interpersonal skills?	2
39	How do you feel about the use of cellphones as educational tools?	19
	Total	480

the 100 most frequent four-grams (minimum token frequency ≥ 5) were extracted from the target-prompt sub-corpus. Keyness was computed using Dunning's log-likelihood against a reference of the remaining prompts (leave-one-prompt-out). While two prompts did not yield any topic-specific bundles at these thresholds, they were retained as fixed-effect levels to preserve task-related variance. To assess whether results were driven by low-count prompts, models were re-estimated after excluding prompt categories with fewer than 10 essays (see Appendix A, Table A6). The main effects for topic-specific and global coverage were unchanged, indicating that the findings are not driven by low-count prompt categories.

To distinguish topic-specific from global bundles, a two-step procedure combining keyness and independent annotation was employed. Accordingly, to establish the topic-specific lexical bundle lists, bundles with significant positive keyness for the held-out prompt ($G^2 \geq 10.83$, $p < .001$) were classified as topic-specific. At the same time, high-frequency bundles per prompt were treated as topic-specific if they contained prompt-specific domain lexicon, ensuring they carried content tied to that prompt. Regarding global lexical bundles, they constituted the top n -grams by overall frequency after excluding those already classified as topic-specific for any prompt. In parallel, a global bundle could not contain domain-specific tokens and had to recur across multiple prompts, a requirement tested under several reasonable cut-offs (e.g., $\geq 8/17$, $\geq 10/17$, $\geq 12/17$ prompts, and a $\geq 60\%$ rule). Two independent annotators applied a topic and global-labelling scheme to the resulting lists. Disagreements were resolved through adjudication to a single gold label used in all analyses. Inter-annotator agreement is provided in Appendix A, Table A7. As for coverage, it was computed per essay as the proportion of word tokens belonging to matched bundles (case-folded; punctuation-stripped; whitespace tokenisation). Longest-match priority was applied when a sequence overlapped multiple phrases. Finally, function words in the denominator were retained to ensure comparability with the Coh-Metrix scores.

Besides using Sketch Engine, essays were run through Coh-Metrix to generate a list of indices for each essay. Selected indices were aggregated into three composites, with all indices z-scored within the corpus and averaged so that higher values indicate a greater presence of the construct. Table 2 summarises indices and transformations.

Table 2. Coh-Metrix composites and component indices

Composite	Coh-Metrix indices	Transformation	Direction (higher =)
Specificity	WRDHYPn, WRDHYPv, WRDPOLc	z score, invert hypernymy & polysemy to reflect specificity	More lexically specific wording
Referential Cohesion	CRFCWO1, CRFCWOa, CRFNO1, CRFSO1, LSASS1, LSAPP1	z score, average	Greater local semantic/ lexical overlap
Deep Cohesion	CNCCaus, SMCAUSv, SMCAUSr	z score, average	More causal/ argumentative linkage

Data Analysis

To answer the research questions, the association between lexical-bundle coverage and the three Coh-Metrix composites was analysed at the individual essay level. This was achieved through a *multivariate* multiple regression that simultaneously regressed the three composites on topic-specific and global lexical bundle coverage and essay length, with fixed effects for prompt and educational level. The independent variables (bundle coverage and length) were z-scored to ensure comparability with the Coh-Metrix composites. As noted earlier, standard errors were clustered by writer to account for repeated essays by the same student.

Before moving to the results, one last point regarding model diagnostics is worth mentioning. Several assumptions were inspected to evaluate linearity and homoscedasticity and to identify influential observations. These include residual-versus-fitted and scale–location plots, Q–Q

plots of residuals, and leverage/Cook's distance. Additionally, variance-inflation factors (all <5) were examined to assess potential collinearity. Standardised coefficients (β) with 95% confidence intervals and two-sided p-values are reported in the Results section.

Results

The study examined the association between topic-specific and global lexical bundles and the three Coh-Metrix composites (Specificity, Referential Cohesion, Deep Cohesion) in L2 argumentative essays by Arabic L1 ESL learners. This section first presents descriptive statistics for the whole corpus (480 essays), followed by results addressing each research question.

Before presenting the results, it is important to provide an overview of how lexical-bundle use and the three composites vary across essays (summarised in Appendix A, Tables A1 and A2). These tables present descriptive statistics for the whole corpus, along with bivariate correlations. The data revealed that students produced relatively short texts (215 words on average), with substantial variation in length. Topic-specific bundles accounted for about 9.8% of the words in a typical essay, while global bundles accounted for roughly 2%. When combined, total bundle coverage reaches approximately 12%. At the essay level, the Specificity, Referential Cohesion, and Deep Cohesion composites show substantial dispersion, indicating measurable variation in discourse-semantic features across texts. Among the three composites, topic-specific coverage aligned most strongly with Referential Cohesion in simple correlations, whereas global coverage showed modest alignment with Deep Cohesion. Length did not correlate with higher cohesion or specificity. R^2 (adjusted) values are reported per composite.

Following the descriptive statistics, the analysis addresses the two research questions. As illustrated in Table 3, total coverage was found to be significantly associated with Referential Cohesion but showed no reliable association with Deep Cohesion or Specificity.

As for the second research question, the analysis distinguishes between topic-specific and global lexical bundle coverage. As shown in Table 3, topic-specific coverage is strongly associated with higher Referential Cohesion, but not with Deep Cohesion or Specificity. By contrast, global coverage is associated with Deep Cohesion, shows no reliable association with Referential Cohesion, and exhibits a small negative trend for Specificity. Essay length is unrelated to Referential Cohesion and Specificity, and shows a small negative association with Deep Cohesion. Overall, the results suggest that the three composites are primarily driven by topic-specific lexical bundles for Referential Cohesion, whereas global bundles mainly support Deep Cohesion.

Besides highlighting the relationships between bundle types and the three composites, the results emphasise the importance of prompts and educational level as factors shaping these relationships. Variations in prompts and educational levels are naturally reflected in the lexical affordances writers engage with. There are prompts, for example, that invite writers to use concrete, spatial, or experience-based language, which tends to evoke higher topic-specific coverage and stronger Referential Cohesion. Prompt-level analysis (see Appendix A, Table A3) revealed that prompts involving manipulable entities and spatial settings (e.g., transport, print vs. digital books) showed greater topic coverage and higher Referential Cohesion. Conversely, more abstract prompts relying on evaluation and stance tend to elicit global organisers without necessarily raising referential alignment. This suggests that abstract prompts, which emphasise ideas over tangible experiences, may lead to lower lexical specificity and weaker local referential alignment. Consistent with this pattern, a small number of prompts display zero mean topic coverage, likely because few topic bundles were identified for those essays. However, the general trend persists, with higher topic-specific coverage linked to higher Referential Cohesion.

Furthermore, the results reveal an interaction between prompt type and educational level and how this interaction influences the relationship between lexical bundle coverage and the

Table 3. *Multivariate multiple regression: effects of bundle coverage and length*

Specification	Independent variables (z)	Specificity β [95% CI] p	Referential cohesion β [95% CI] p	Deep cohesion β [95% CI] p	R ² (adj.)
Model 1: Total coverage model	Total bundle coverage	−0.005 [−0.091, +0.080], p = 0.907	+0.231 [+0.124, +0.338], p < 0.001	+0.059 [−0.063, +0.180], p = 0.344	0.40/ 0.32/ 0.09
	Length (DESWC)	+0.016 [−0.142, +0.173], p = 0.845	−0.011 [−0.168, +0.146], p = 0.895	−0.167 [−0.335, +0.000], p = 0.050	
Model 2: Topic & global coverage	Topic-specific coverage	+0.039 [−0.047, +0.125], p = 0.374	+0.239 [+0.131, +0.347], p < 0.001	+0.006 [−0.110, +0.122], p = 0.918	0.41/ 0.32/ 0.09
	Global coverage	−0.085 [−0.177, +0.006], p = 0.068	+0.012 [−0.086, +0.111], p = 0.808	+0.108 [+0.013, +0.204], p = 0.026	
	Length (DESWC)	+0.019 [−0.136, +0.175], p = 0.809	−0.009 [−0.166, +0.149], p = 0.915	−0.171 [−0.336, −0.006], p = 0.043	

three composites (see Appendix A, Table A4). For example, essays by HS12 students are shorter on average yet show the highest topic-bundle coverage and the strongest Referential Cohesion. In contrast, essays by university-level writers (BA2–MA1) exhibit lower topic coverage scores for Referential Cohesion, possibly reflecting a developmental shift toward more abstract argumentative writing, where global scaffolding increases without parallel gains in Referential Cohesion. These differences underscore the importance of using these variables (prompts and educational level) as fixed effects in the modelling and offer pedagogical implications which are discussed below.

Overall, the results suggest a differentiated role for the two bundle types. Topic-specific bundles are more strongly associated with local referential alignment (Referential Cohesion), whereas global coverage is more clearly linked to causal/argumentative linkage (Deep Cohesion). While variations in tasks and educational level may have influenced these results, the coverage–composite relationship remains consistent, and essay length did not account for these effects.

Discussion

This section interprets the findings in relation to broader linguistic and theoretical frameworks. Generally, the findings suggest that topic-specific and global lexical bundles do not contribute

equally to the Coh-Metrix discourse–semantic composites. Whereas topic-specific bundles, used to verbalise entities, settings, processes, and attributes, are strongly associated with higher Referential Cohesion, global bundles are linked to Deep Cohesion (causal/argumentative scaffolding). By contrast, neither topic-specific nor global bundle coverage showed a reliable association with Specificity, suggesting that bundle use in this dataset was more closely linked to discourse connectivity than to lexical precision. These findings carry both theoretical and practical implications (Biber et al., 2004; Hyland, 2008; McNamara et al., 2010; McNamara et al., 2014). The composites align with common L2-writing rubric dimensions. Referential Cohesion corresponds to idea development and support (local alignment), while Deep Cohesion corresponds to argumentative organisation and logic.

First, the results align with models of text processing that distinguish between Referential and Deep Cohesion. Topic-specific bundles appear to pattern with lexical and semantic overlap and are associated with stronger local situation-model continuity (Graesser et al., 2004; McNamara et al., 2010, 2014; Zwaan & Radvansky, 1998). By contrast, global bundles (e.g., on the other hand, it is important to) are more closely associated with features captured by Deep Cohesion and may function as cues to argumentative scaffolding rather than as direct contributors to lexical specificity or referential continuity (Halliday & Hasan, 1976; McNamara et al., 2010). One possible interpretation is that content-bearing language encourages repeated use of phrases referring to domain entities and actions, which can facilitate deeper encoding and richer associative links (Craik & Lockhart, 1972). This is reflected in higher Referential Cohesion rather than shifts in Specificity in the present findings.

Furthermore, these findings deepen our understanding of what constitutes development in L2 writing. Traditionally, research has emphasised syntactic complexity (Lu, 2010) and lexical sophistication (Kyle & Crossley, 2015). However, they do not measure whether a text has a clear situation model, which involves using words that effectively represent spatial, temporal, causal, and motivational relationships related to a given topic, allowing readers to easily follow the ideas. The results suggest that topic-specific bundle coverage is closely associated with Referential Cohesion and may therefore be interpreted as a textual correlate of stronger situation-model continuity. Thus, development appears to be captured by a writer's ability to sustain referential links through the use of domain-appropriate lexical bundles. This should complement, but not replace, syntactic complexity and single-word sophistication (Crossley & McNamara, 2012; Crossley, 2020; Li & Fang, 2022).

Similarly, the link between formulaic use and text quality warrants discussion. While this link is well-established, its underlying mechanisms remain understudied. Bestgen and Granger (2014) found that the overall density of automatically identified formulaic sequences correlates with graders' scores, while Bestgen (2017) demonstrated that such phraseological measures account for variation in quality beyond single-word richness indices. Li and Fang (2022) further extended this line of research, demonstrating that raters respond to the structure, function, frequency, and association strength of multi-word sequences. In other words, raters consider the type of phraseological material used, not just the quantity. The results of the present study offer an automated measurement that aligns with these findings. Topic-specific, content-bearing bundles are associated with higher Referential Cohesion, whereas global bundles are associated with Deep Cohesion but show no reliable gain in Referential Cohesion and a small negative trend for Specificity. This suggests that frequent bundle use, even when rhetorically effective, does not necessarily entail more lexically precise or less generic wording.

These findings suggest that teachers should provide learners with opportunities to notice topic-specific lexical bundles. The level-wise patterns suggest that earlier-stage writers tend to rely more on rehearsed, content-specific phrases, which serve as processing shortcuts for accessing

domain-relevant vocabulary and ideas. As learners advance, the challenge increases, as they need to maintain both the semantic richness and global coherence in their essays. Pedagogically, tasks focused on concrete lexical fields can therefore scaffold both conceptual access and writing fluency. This allows students to draw on common formulaic expressions while extending their semantic range. Consequently, targeted topic-bundle scaffolding may be a promising pedagogical direction, although the present study does not directly test whether it improves explicitness or imageability in student writing.

Accordingly, a practical implication related to the nature of the tasks emerges from these results. Some prompts can either encourage or limit the use of topic-specific bundles, regardless of the writer's proficiency. Prompts with concrete settings and entities (e.g., banning cars from the city centre, legal driving age) tend to elicit higher topic coverage and higher Referential Cohesion. In contrast, more abstract, evaluative prompts (e.g., comparing print and e-books, digital communication and interpersonal skills) yield lower Referential Cohesion unless writers import referential framing. In addition, a small number of prompts yielded few topic bundles. This likely reflects within-prompt lexical diversity and limited stable 3–4-gram recurrence rather than the absence of content lexis. Another interpretation is that some domains are realised through bigrams or numbers (e.g., temperature values) or various paraphrases, depressing stable 3–4-gram counts. Furthermore, abstract prompts elicit stance or organisational bundles more than referential ones, boosting Deep Cohesion rather than Referential Cohesion.

Finally, topic-specific bundles are the phraseological means through which L2 argumentative essays acquire referential texture, while global bundles contribute deep/causal texture. Phraseological competence should therefore be viewed as the ability to sustain referential chains and deploy causal/stance cues appropriately, complementing syntactic complexity and lexical sophistication. Therefore, teachers should focus on both types throughout instruction and assessment. They can supply referential scaffolding when topics are abstract and complement this with causal language.

Limitations

Although this paper offers several key insights, it has limitations that need to be acknowledged. Firstly, because of the inclusion criteria applied to the corpus (only a subset of prompts), a limited number of essays (480) were selected. Moreover, since essays were written by Arabic L1 ESL students, the generalisability of the results is yet to be demonstrated. Future research can investigate other learner categories or academic registers (e.g., lab reports, literature reviews) in which topic framing may differ in density and function, allowing broader comparison. Second, Coh-Metrix composites, although valid and offering an automated approach to quantifying these discourse semantic constructs, might fail to capture more localised dimensions of coherence, such as topic development within a paragraph and rhetorical relations beyond causal ones.

Third, because the research is cross-sectional, it was not possible to measure within-learner differences. Such variations can be better evidenced by longitudinal studies that track the same learners across various educational levels to observe how the use of lexical bundles changes over time. In addition, the research does not triangulate findings with human or rubric-based ratings. Since the composites index reader-oriented texture, future research should model the joint predictive effect of topic coverage and Specificity, Referential Cohesion, and Deep Cohesion on human judgements of clarity, coherence, and quality. Moreover, although the bundle was identified using clear and reproducible measures, construct validity depends on category reliability; in this regard, manual annotations indicated a high inter-annotator agreement for n-gram classification. Methodologically, even though this was done using multivariate multiple

regression (with controls for prompt and educational level and writer-clustered SEs), the role of unobserved writer traits cannot be fully ruled out. Despite these limitations, sensitivity analyses corroborated the findings (see Appendix A, Tables A5 and A6).

Implications

The pedagogical implications proposed here should be interpreted with caution. Because the study is cross-sectional and corpus-based, it does not directly test instructional effects. Nevertheless, the findings suggest several pedagogically relevant distinctions regarding the functions of topic-specific and global lexical bundles in argumentative writing. The primary finding is that topic-specific bundles are associated with Referential Cohesion, which relates to idea development and local alignment (Halliday & Hasan, 1976; McNamara et al., 2014). Deep Cohesion, on the other hand, is linked to global lexical bundles.

Hence, pedagogy ought to be discipline-conscious. To maintain referential chains and keep participants, processes, and settings salient throughout an essay, the findings suggest that instruction may usefully focus more explicitly on topic-specific lexical bundles (Hyland, 2008; Hoey, 2005). For example, lexical fields and bundle families can be used to organise units and to provide spaced and varied recycling (Römer, 2010). Teachers can provide students with immediately useful phrase banks, though decontextualised memorisation should be avoided. To this end, instruction may benefit from using authentic reading or listening texts that expose students to these phrases. Another example involves the use of spaced-repetition flashcards to support lexical bundle acquisition (Barghamadi et al., 2022). Moreover, educators can cluster topic-related phrases around head nouns (e.g., public transportation system, reduce traffic congestion, vehicle access restrictions), allowing learners to change modifiers and complements without losing semantic focus. Such practical, task-based, formulaic-language instruction may be a promising pedagogy direction, although the present study does not directly test its instructional effectiveness. (Ogawa, 2020). It should also be noted that teachers need to align phrase banks with corpus-based research to prevent learners from using unusual sequences. (Northbrook & Conklin, 2021). In so doing, the instructors can assist L2 writers in developing lexical resources for elaborating processes and making the situation model available to readers by foregrounding topic-specific bundles. The findings also suggest that explicit attention to causal and reasoning language may be useful, particularly because global bundles are more strongly associated with Deep Cohesion. This can be facilitated by helping students recognise how global bundles are associated with Deep Cohesion and argumentative signalling, without assuming that these bundles also strengthen Referential Cohesion (Hyland & Jiang, 2018).

Moreover, topic-specific lexical bundles should be made more visible through task design and teacher guidance so that learners can better recognise not only rhetorical fluency, but also the contribution of phraseology to semantic development. In addition to the strategies mentioned above, teachers can use feedback to raise learners' awareness of their use of both topic-specific and global lexical bundles. In this regard, the feedback must be detailed with specific comments, not a standardised comment like 'add more details.' The extent to which students use topic-specific bundles to serve readers is one element that should be addressed in that feedback. More broadly, the findings suggest that phraseological development in argumentative writing should not be treated as a single dimension. Topic-specific and global bundles appear to be associated with different discourse-semantic functions, and this distinction may help instructors provide more targeted feedback. Rather than relying solely on broad comments, teachers may find it useful to distinguish between weaknesses in referential development and those in argumentative signalling. Future research could explore whether feedback tools or automated writing-support systems can operationalise these distinctions in pedagogically useful ways.

Conclusion

The paper examined the association between topic-specific and global lexical bundle coverage and three Coh-Metrix composites (Specificity, Referential Cohesion, and Deep Cohesion) across 480 essays spanning diverse prompts and educational levels. The findings show that topic-specific bundles are strongly related to high levels of Referential Cohesion, whereas global bundles are strongly associated with high levels of Deep Cohesion; neither bundle type is reliably linked to Specificity. The results point to a theoretical rationale where topic-specific lexical bundles structure conceptual information by preserving referential chains for readers to follow, whereas global bundles structure causal/argumentative relations without densifying lexical content (Halliday & Hasan, 1976; McNamara et al., 2014).

The paper also demonstrates how per-essay bundle coverage can be integrated with Coh-Metrix to quantify discourse-semantic properties most likely to be influenced by different bundle types. Also, the findings indicate the applicability of referential scaffolding in phraseological development. This type of scaffolding must be task-oriented, which implies that formulaic language is not merely a list of expressions but a system for generating conceptual material that can be learned and evaluated. Future research could study the integration of corpus analysis, cognitive measures and AI-supported feedback. This may be applicable for tracking development over time, triangulating findings with human choices, and implementing computer-generated cues that prompt writers to use topic-specific lexical packages (in the case of Referential Cohesion) and scaffolded causal language (in the case of Deep Cohesion). Through this approach, L2 writers can develop their ideas and arguments more effectively, and better instruments can be designed to measure that change.

References

- Al-Jarf, R. S. (2001). Processing of cohesive ties by EFL Arab college students. *Foreign Language Annals*, 34(2), 141–151. <https://doi.org/10.1111/j.1944-9720.2001.tb02819.x>
- Barghamadi, M., Rogers, J., & Muller, A. (2022). On the learning of multi-word units via flashcard applications. *Australian Journal of Applied Linguistics*, 5(1), 1–18. <https://doi.org/10.29140/ajal.v5n1.643>
- Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, 69, 65–78. <https://doi.org/10.1016/j.system.2017.08.004>
- Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, 28–41. <https://doi.org/10.1016/j.jslw.2014.09.004>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49. <https://doi.org/10.64152/10125/44213>
- Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second-language writing*. Cambridge University Press. <https://www.cambridge.org/core/books/contrastive-rhetoric/FC623B3C4EF3C0E9C1D6E0D35C5E3B6B>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415–443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115–135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>

- Crossley, S. A., & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, 7(3), 351–370. <https://doi.org/10.17239/jowr-2016.07.03.02>
- Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371–398. <https://doi.org/10.1111/j.1756-8765.2010.01081.x>
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193–202. <https://doi.org/10.3758/BF03195564>
- Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18(1), 109–136. <https://doi.org/10.1075/ijcl.18.1.08gra>
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18(3), 218–238. <https://doi.org/10.1016/j.asw.2013.05.002>
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman. <https://www.routledge.com/Cohesion-in-English/Halliday-Hasan/p/book/9780582550414>
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge. <https://doi.org/10.4324/9780203327630>
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum. <https://www.bloomsbury.com/uk/metadiscourse-9781350063594/>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K., & Jiang, F. K. (2018). “In this paper we suggest”: Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, 51, 18–30. <https://doi.org/10.1016/j.esp.2018.02.001>
- Khuwailah, A. A., & Shoumali, A. A. (2000). Writing errors: A study of the writing ability of Arab learners of academic English and Arabic at university. *Language, Culture and Curriculum*, 13(2), 174–183. <https://doi.org/10.1080/07908310008666597>
- Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757–786. <https://doi.org/10.1002/tesq.194>
- Li, Y. K., & Fang, A. B. (2022). Synergistic effects of multiword sequences’ structure, function, frequency and association on raters’ evaluations of essay quality. *Frontiers in Psychology*, 13, 1026658. <https://doi.org/10.3389/fpsyg.2022.1026658>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511894664>
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding. *Cognition and Instruction*, 14(1), 1–43. https://doi.org/10.1207/s1532690xcil401_1
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing the linguistic features of cohesion. *Discourse Processes*, 47(4), 292–330. <https://doi.org/10.1080/01638530902959943>
- Northbrook, J., & Conklin, K. (2021). “That sounds about right”—Lexical bundle naturalness intuitions in Japanese learners of English. *Vocabulary Learning and Instruction*, 10(1), 49–63. <https://doi.org/10.7820/vli.v10.1.northbrook.conklin>
- Ogawa, C. (2020). Teaching ideas for improving oral performance through formulaic language instruction. *Vocabulary Learning and Instruction*, 9(2), 48–54. <https://doi.org/10.7820/vli.v09.2.ogawa>
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press. <https://global.oup.com/academic/product/mental-representations-9780195066661>
- Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press. <https://doi.org/10.4324/9781315785233>

- Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction*, 3(1), 95–119. <https://doi.org/10.1075/etc.3.1.06rom>
- Samraj, B. (2024). Disciplinary differences in lexical bundle use: A cautionary tale from methodological variations. *Journal of English for Academic Purposes*, 70, 101399. <https://doi.org/10.1016/j.jeap.2024.101399>
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing* (pp. 44–49). Manchester, UK. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Senel, B. (2025). *Lexical bundle frequency as a construct-relevant candidate feature in automated scoring of L2 academic writing* (arXiv:2504.08537). arXiv. <https://doi.org/10.48550/arXiv.2504.08537>
- Shin, Y. K., & Won, D.-O. (2024). To what extent do L2 learners produce genre-appropriate language? A comparative analysis of lexical bundles in argumentative essays and speeches. *Journal of English for Academic Purposes*, 69, 101389. <https://doi.org/10.1016/j.jeap.2024.101389>
- Siu, B. W. Y. (2024). A corpus-based analysis of lexical bundles in L2 English engineering reports. *SAGE Open*, 14(4), 21582440241299997. <https://doi.org/10.1177/21582440241299997>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519772>
- Xia, D., & Kessler, M. (2025). Making arguments at lower proficiencies: Move–bundle connection in Korean university students’ English argumentative writing. *International Journal of Applied Linguistics*, 35(3), 1219–1241. <https://doi.org/10.1111/ijal.12693>
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. <https://doi.org/10.1037/0033-2909.123.2.162>

Appendix A

Table A1. Descriptive statistics of key variables (N = 480 essays)

Variable	Mean	SD	Min	Max
Topic specific coverage (proportion)	0.098	0.071	0.000	0.359
Global coverage (proportion)	0.020	0.021	0.000	0.133
Total coverage (proportion)	0.118	0.078	0.000	0.427
Length (words)	215.27	113.56	43	689
Specificity (z)	0.020	0.980	−2.490	+2.510
Referential Cohesion (z)	−0.010	1.010	−2.610	+2.640
Deep Cohesion (z)	0.000	1.020	−2.570	+2.550

Table A2. Pearson correlations among z-scored variables (N = 480)

	1	2	3	4	5	6
1. Topic specific coverage (z)	1.000	0.420***	−0.080	0.040	0.270*	0.020
2. Global coverage (z)		1.000	−0.190***	−0.070†	0.020	0.120
3. Length (DESWC, z)			1.000	0.010	−0.030	−0.100*
4. Specificity (z)				1.000	0.680***	0.240***
5. Referential Cohesion (z)					1.000	0.310***
6. Deep Cohesion (z)						1.000

Table A3. *Prompt-level topic coverage and composite means*

Prompt ID	N	Topic specific coverage	Specificity (z)	Referential cohesion (z)	Deep cohesion (z)	N < 10
1	56	0.059	+0.05	−0.07	+0.02	
2	28	0.127	−0.18	+0.22	−0.04	
4	16	0.112	−0.41	−0.36	−0.21	
6	34	0.077	−0.21	−0.19	+0.05	
7	36	0.076	−0.03	−0.01	−0.03	
8	32	0.095	−0.09	+0.14	+0.09	
10	49	0.082	+0.01	+0.11	+0.13	
11	18	0.145	+0.37	+0.31	+0.08	
14	34	0.107	+0.06	+0.18	+0.12	
15	13	0.000	−0.04	−0.09	−0.06	✓
22	39	0.099	−0.28	+0.06	+0.03	
24	41	0.095	−0.15	−0.11	+0.07	
30	17	0.213	+0.41	+0.45	+0.28	
33	19	0.178	+0.38	+0.39	+0.19	
35	27	0.107	−0.72	−0.66	−0.39	
36	2	0.000	−0.30	−0.22	−0.08	✓
39	19	0.110	+0.16	+0.09	−0.01	

Table A4. *Educational-level means (topic coverage, composites, length)*

Level	N	Topic specific coverage	Specificity (z)	Referential cohesion (z)	Deep cohesion (z)	Length (words)
HS12	36	0.195	+0.12	+0.45	+0.18	153.69
BA2	80	0.097	−0.18	−0.06	−0.04	274.09
BA3	167	0.083	−0.05	−0.12	−0.07	244.83
MA1	197	0.093	+0.07	+0.09	+0.02	177.57

Table A5. *Mixed-effects sensitivity (writer & prompt random intercepts)*

Composite	Independent variables (z)	β (standardised)	95% CI	p
Specificity	Topic specific coverage	+0.023	[−0.031, +0.077]	.402
	Global coverage	−0.051	[−0.108, +0.006]	.081
	Length (DESWC)	+0.017	[−0.069, +0.103]	.690
Referential Cohesion	Topic specific coverage	+0.201	[+0.113, +0.289]	< .001
	Global coverage	+0.015	[−0.071, +0.102]	.729
	Length (DESWC)	−0.012	[−0.091, +0.068]	.767
Deep Cohesion	Topic specific coverage	+0.004	[−0.066, +0.073]	.913
	Global coverage	+0.061	[+0.008, +0.114]	.025
	Length (DESWC)	−0.089	[−0.175, −0.003]	.043

Table A6. *Prompt sensitivity (excluding prompts with N < 10)*

Composite	Independent variables (z)	β (standardised)	95% CI	p	N
Specificity	Topic specific coverage	+0.027	[−0.026, +0.081]	.316	445
	Global coverage	−0.050	[−0.106, +0.007]	.087	
	Length (DESWC)	+0.015	[−0.078, +0.109]	.744	
Referential Cohesion	Topic specific coverage	+0.197	[+0.109, +0.285]	< .001	445
	Global coverage	+0.011	[−0.072, +0.094]	.801	
	Length (DESWC)	−0.008	[−0.095, +0.079]	.853	
Deep Cohesion	Topic specific coverage	+0.002	[−0.066, +0.070]	.950	445
	Global coverage	+0.058	[+0.007, +0.110]	.027	
	Length (DESWC)	−0.090	[−0.176, −0.004]	.042	

Table A7. *Inter-annotator agreement for bundle labelling (Topic/Global and Function)*

n-gram length	No of items	Topic/Global % agree	Topic/Global κ	Function % agree	Function κ
3-gram	1,012	95.9	0.731	97.0	0.618
4-gram	445	95.6	0.731	96.5	0.617