Manipulating texture and cohesion in academic writing: A keystroke logging study

Neil Evan Jon Anthony Bowen [a],[*], Nathan Thomas [b]

[a] *Dept. of English & Linguistics, Thammasat University, Klongluang, Pathumthani, Bangkok, 12121, Thailand*
[b] *Centre for Applied Linguistics, UCL Institute of Education, University College London, 20 Bedford Way, Bloomsbury, London, WC1H 0AL, United Kingdom*

A B S T R A C T

Research has repeatedly shown that problems arise when students are asked to link information co-textually and contextually across larger phases of discourse. Within Systemic Functional Linguistics (SFL), a text-oriented theory of language, co-textual and contextual links are analyzed and operationalized in terms of textual and logical metafunctions, both of which work together to connect and enable experiential and interpersonal metafunctions. While most writing studies to date have investigated text as product (synoptic approach), there has been increasing interest in studying text as an evolving process (dynamic approach). The current study contributes to this emerging research by examining the real-time choices made by six student writers. Drawing on keystroke logging software (Inputlog), it explores writers' revision choices within the systems of THEME, INFORMATION, and IDENTIFICATION, in conjunction with the logical metafunction. Results indicate that complex choices contribute to unfolding cohesiveness and information flow, where choices in specificity and congruency are key contributors to managing texture while also manipulating complexity and context-dependency. Overall findings suggest that students may benefit from an explicit focus on the nominal group as a means to create and maintain texture and cohesion through over-specification, classification (pre-modifiers), and qualification (post-modifiers).

1. Introduction

In moving from everyday, common sense discourse to academic discourse, students are required to communicate in more specialized and discipline-specific ways. This includes learning how to condense and expand upon meanings while simultaneously categorizing, taxonomizing, and presenting information to varying levels of abstractness (Halliday, 2004; Maton, 2020). In academic writing, students are also expected to organize and link this discourse co-textually and contextually so that text progresses logically from one topic to the next. Suitable links between topics rely on resources such as anaphora (words that refer backward in the text), cataphora (words that refer forward in the text), exophora (words such as deictics that point outside the text), and homophora (words that have shared cultural meanings). Links may also be forged by phenomena such as grammatical cohesion (Halliday & Hasan, 1976), lexical organization (Hasan, 1985), thematic patterning (Davies, 1997), and participant tracking (Martin, 1992). It is this collection of resources that when combined with other meanings create a synthesis called *texture*. This notion of texture is often described

---

metaphorically, where text is likened to a fabric composed of interlocking threads of meaning. Within Systemic Functional Linguistics (SFL), there has been a wealth of studies on how writers forge connections between words, groups of words, and clauses to create texture (see below). These studies have repeatedly shown that students have little trouble deploying such resources at the clausal level yet struggle when it comes to linking ideas together in longer text.

As part of these on-going investigations into the linking of co-text and context, research has moved away from isolated clauses and on to patterns of language that are both constitutive of and activated by context of culture (language as system/potential) and context of situation (language as instance) (see Halliday & Matthiessen, 2013). Accordingly, SFL studies have frequently looked at shared meanings and expectations in a discourse community reflected in, for example, the "Sydney School" approach to genre (toward the culture side of the cline), as well as language patterns in texts reflected in studies of register and the variables field, tenor, and mode (toward the situational side of the cline). Researchers may also choose to position themselves with respect to development of language in a culture (phylogenesis), development of language in an individual (ontogenesis), and development of language in a situation/text (logogenesis).[1]

However, despite the potential for focusing on the development of language (logogenesis) in text (see Coffin & Donohue, 2012), to date, the majority of SFL-inspired research in this timeframe has concentrated almost exclusively on speech (Ventola, 1987; Yang, 2010). And while some studies have looked at logogenesis in written text (e.g., Klein & Unsworth, 2014), they have done so in relation to writing being a finished, instantial product rather than a series of unfolding choices (many of which are absent in the final product). Overall, in terms of the logogenesis of written text (studying meaning as it unfolds during the writing process), there are just a few studies that examine this in first language writing (see Bowen, 2019; O'Donnell, 2013 for studies on English texts) and no studies—as far as we know—on logogenesis in second language writing. Therefore, there is a lack of information on how writers bring together meanings in real time to form a harmonious and cohesive unit during the writing process.

In this study, we address these issues by drawing upon keystroke logging (KSL) and SFL to examine the revisions of three master's students with English as a first language (L1) and three with English as a second language (L2). We included L1 and L2 groups that were similar in age, education level, and major of study so that we could make direct comparisons on their unfolding choices (there is no existing data that we know of on how L1 writers manipulate texture and cohesion as they write). Theoretically, the study investigates how revisions involving overlapping choices in the systems of THEME, INFORMATION, and IDENTIFICATION can contribute to an understanding of how texture and cohesion develop in real time in each of these groups of writers. Methodologically, the study seeks to explore further the potential of KSL to facilitate inquiry into the emerging and dynamic nature of writing, particularly with regard to tracing revision functions. The implication is that such research can support L2 writing development in specific domains by exploring how enculturated L1 writers use increasingly complex language within those domains—in this instance, language choices that contribute to the organization and integration of information in higher-level L1 academic writing.

1.1. Studying text as product

In analyzing information flow, linguists can explore any number of language features or systems that contribute to texture. Recently, one of the most popular focal points has been Theme (see He, 2020, for a review). In the Hallidayan tradition, Themes are labelled as topical, interpersonal, or textual. Topical Theme is obligatory and occurs only once; it alone controls the thematic potential and markedness of a clause and marks the Theme–Rheme boundary (Halliday & Matthiessen, 2013). In declarative clauses, the unmarked (topical) Theme conflates with the Subject; marked Themes are experiential elements other than the Subject (e.g., circumstantial adjuncts). If interpersonal and/or textual Themes occur, they precede topical Theme, typically following a textualˆinterpersonalˆtopical realization,[2] as shown in Fig. 1.

Theme thus represents essential choices when orienting propositions, foregrounding new information, and establishing a context for interpreting and developing Rheme. Consequently, the pedagogical value of teaching students to make appropriate thematic choices—particularly for academic writing—is well documented (Alexander, 2019; Wei, 2016a).

In terms of how Theme contributes to texture, a number of studies have shown how thematic choice varies between L1 and L2 writers (He, 2020; Herriman, 2011; Wei, 2016b). In general, findings have shown that L2 writers tend to overuse textual Themes (particularly conjunctive adjuncts) and interpersonal Themes (notably modal adjuncts),[3] which contributes to a more dialogic (oral) style of progression in their writing. By contrast, L1 writers incorporate more informationally dense Themes in their writing, which is often instantiated in complex nominal groups (NGrp). L1 writers also tend to make more strategic use of unmarked topical Themes (what we call Subject-Theme in this study) to advance their texts. This advancement can effectively be done via nominalization within Theme (e.g., when verbs or adjectives become nouns, as in "distribute→distribution" or "unstable→instability," respectively). Nominalization is just one form of what SFL terms *grammatical metaphor* (GM), and it is a powerful means by which a clause can be condensed into a NGrp. For example, "cars emit more and more CO2 gas" can be reconstrued as "increasing CO2 emissions." This reconstrual can then be used as a referent in the unmarked Theme "slot" in the example clause "increasing CO2 emissions are problematic." Overall, these elements reflect the well-cited remark that advanced academic writing makes heavy use of complex language realized in NGrps (Musgrave & Parkinson, 2014) and through GM (Halliday, 2004).

---

[1] For useful elaborations on these concepts, see Halliday and Matthiessen (2013, pp. 27−29).

[2] The symbol ˆ denotes order of elements in SFL.

[3] This is often attributed to explicit teaching of conjunctions and modals, L1 transfer, and lack of exposure to specific genres/disciplinary expectations.

However,	more realistically,	the increasing cost of petrol	will force more people to switch to electric cars.
textual	interpersonal	topical	
Theme			Rheme

Fig. 1. Unmarked Multiple Theme.

Complexity in NGrps reflects a distinction between "the potential to refer, and the potential to expand" (Ravelli, 2004, p. 132) and is exemplified in a study by Whittaker, Llinares, and McCabe (2011). In their four-year study of 12 to 16-year-old Spanish L2 English learners, the researchers focused on the NGrp's potential to act as a referent (information management via recoverability) and a carrier of *semantic density* (register appropriateness through pre-/post-modification). Their results showed greater textual cohesion in later texts (15/16-year-olds) due to a slight increase in co-textual referents (endophora) and NGrp complexity, and a decrease in contextual reference (exophora). Despite the task being restricted to 20 minutes, resulting in relatively short texts, their results illustrated how more advanced writers can exploit the NGrp to fine-tune how they introduce (presenting reference) or track (presuming reference) an entity in a text (Martin, 1992).

Complexity through GM is also a powerful means to manipulate referents. Indeed, GM has been one of SFL's key contributions to studying advanced writing abilities, and its potential benefits are well known (Xuan & Chen, 2019). In one study, for example, Schleppegrell (2004) found that minimal use of GM by L2 writers resulted in texts that instructors found wordy, informal, and at times difficult to follow. However, increased use of GM allowed one L1 writer to elaborate on a topic while achieving a level of technicality, authoritativeness, and texture that was more highly valued. Thompson (2010) has also argued, as have Ryshina-Pankova and Byrnes (2013), that an increase in the variety of nominalizations (particularly those serving a cohesive function as anaphoric referents) can result in texts that are better received.

Similar studies have shown how writers can construct complex meanings through GM that contribute to thematic progressions and impersonal stances (Baratta, 2010; Schleppegrell, 2004). However, GM often promotes a loss of information, increased ambiguity, uncertainty, and abstraction (Halliday, 2004). Subsequently, extracting the implicit meanings and relations embedded within GMs often requires repeated exposure to the specific discourse community from whence that GM evolved. In one chapter connecting SFL with the sociology of education, for example, Martin (2007) suggested that GM is the fundamental "social semiotic nub of institutionalized learning" (p. 55). Consequently, although some L2 writers make minimal use of GM but still pass English university courses (Baratta, 2010; Liardét, 2013), the agnation of GM throughout the entire semiotic system makes it an ideal yet challenging candidate for researching/teaching advanced writing skills.

Overall, judgement values for higher-level L1 and L2 thematic choices in academic writing indicate a proclivity for texts that exhibit high levels of cohesion and coherence while also displaying movement toward a stronger *rhizomatic code*. A rhizomatic code is a *semantic code* that reflects increasing context-independence and complexity in meanings; it is a concept from Legitimation Code Theory (LCT) that construes "social fields of practice as *semantic structures* […] that comprise *semantic gravity* and *semantic density*" (Maton, 2020, p. 62, emphasis in original). In terms of texture, NGrps and GMs afford writers a semantic space to move referents from specific to general, concrete to abstract, condensed to expanded, and thus enable writers to manipulate context-dependency (±semantic gravity) and complexity (±semantic density). In this respect, it is perhaps unsurprising that "research is showing that high-achieving student work is typically characterized by semantic waves or recurrent shifts in context-dependence and complexity that weave together different forms of knowledge" (Maton, 2020, p. 60).

1.2. Studying text as process

A popular means to study written language as it unfolds is with KSL, computer software that records inputs made by users (e.g., key presses) and logs these activities against a time stamp (Lindgren & Sullivan, 2019). Using KSL has several advantages: it is unobtrusive, detailed, non-reactive, and because it does not require an observer, enjoys a certain ecological validity. Methodologically, KSL has primarily been used to focus on cognitive processes that are linked to revision or pause-related activity (Lindgren & Sullivan, 2019). For example, using revision algorithms and notation systems, researchers have been able to explore specific aspects of online revision (Guo, Deane, Rijn, Zhang, & Bennett, 2018), including whether the type of revision depends on the writing phase/stage in which it occurs (Xu, 2018) and how revision strategies seem to be consistent across writing sessions and projects (Bowen & Van Waes, 2020). Other studies have used KSL to examine aspects such as writing from sources (Leijten, Van Waes, Schriver, & Hayes, 2014) and written discourse as a dynamic process (Bowen, 2019; O'Donnell, 2013).

The works by Bowen (2019) and O'Donnell (2013) are particularly relevant here, as they used KSL to highlight how meaning (or choice) emerged in the unfolding of written texts. O'Donnell (2013) combined KSL with three popular discourse structure theories—generic structure potential, rhetorical structure theory, and thematic progression—to account for the ways in which text could be organized in real time. His findings illustrated how written text is indeed a gradually evolving process rather than a single instance of creativity. Bowen (2019), meanwhile, deployed KSL to examine how the language choices of two L1 English undergraduates unfolded as they composed academic essays. Bowen used foundational SFL concepts to show how certain structures, functions, and systemic choices played a key role in shaping these students' essays, and further highlighted the importance of the NGrp in academic writing.

Embracing such a dynamic view of written text aligns with Vygotsky's interest in the formation of knowledge in a person (process) over the knowledge itself (product). To use Byrnes's (2006) words, "the *process* was likely to be more revealing of the organization of mental activity than merely observing the *product*" (p. 8, emphasis in original). The current study builds on such initiatives; it seeks to demonstrate how a focus on text as process enables us to look at the evolving generation of meaning in text (instantial evolution). Of note in this regard is Larsson's (2018) exemplification of how students need to be able to unpack and repack negotiated meaning, how this can be observed in academic classroom discourse, and how teachers can be instrumental in scaffolding this process, essentially "teaching to wave" (p. 63; see Maton, 2020 above). Insights into the logogenesis of texture can serve to enhance such forward-thinking pedagogy. Instructors can establish more nuanced feedback loops that attend to both the product and process of student writing. Taken together, interest is thus not only methodological and theoretical, but also pedagogical in that it has the potential to inform well-founded approaches to writing pedagogies, particularly at the intermediate and advanced levels.

2. Theoretical underpinning: the structure of information flow

This section introduces three distinct yet overlapping systems that contribute to texture.

2.1. THEME

Within SFL, the system of THEME is the means by which thematic function is realized in English.[4] However, one key theoretical issue is whether the boundary of Theme should be extended to include multiple experiential (topical) Themes, up to and including the Subject (for more information, see Davies, 2004). This study adopts an extended view of Theme as per a number of scholars (e.g., Davies, 1997; Downing, 1991); we also adopt the term *additional Theme* to refer to "non-Subject" experiential Themes and *Subject-Theme* to refer to "Subject" *experiential Themes*. This is primarily because the text-type we are studying (problem-solution) calls for more complex relations to be set up across clauses. For instance, when grammatical Subject is afforded thematic status, it typically functions to maintain the topic, while other experiential elements in Theme serve as "circumstantial frameworks" (Downing, 1991) or "contextual Frames" (Davies, 1997) in conjunction with REFERENCE.

A simple schematic of our view of THEMEˆRHEME can be seen in Fig. 2 (↘ indicates realization statements, which are represented here with examples from our dataset):

2.2. INFORMATION

Whereas THEME is a system of the clause, INFORMATION is a system of the information unit. Fundamentally, when presenting information, discourse is segmented into manageable units: one piece of information equals one unit. In spoken discourse, this is encapsulated inside the tone unit, where the speaker chooses what information to present as the focus of their message. The culmination of this focus is signaled by the tonic, which is the most prominent syllable holding a major pitch change; the constituent holding the tonic is deemed "New" and everything before or after is assumed by the speaker as "Given." Thus, Given and New do not refer to literal gradations of information as old or new (or to notions of recoverability). However, since there are no pitch movements in writing, how can researchers transpose this to written language?

Although Halliday has (rightly) stated that THEME and INFORMATION are not the same thing, other scholars have justly argued that the two clause functions of ThemeˆRheme and GivenˆNew are closely related. Specifically, Theme usually contains Given (recoverable) information—what is predictable or known—while Rheme (that which is not labelled Theme) usually contains New (non-recoverable) information—what is unpredictable or unknown. Fries (1992), for example, has supported this notion with two useful propositions: (a) written information is typically sequenced to take into account the same cognitive resources as speech; and (b) to highlight New, it is best placed at the end of a clause (the last item being most salient to the reader–the "end-weight" principle). To this end, Fries has suggested that in apposition to unmarked Theme, analysts should adopt the term N-Rheme—that which would most likely take the tonic in speech.[5] Furthermore, Fries (2002) saw the N-Rheme's patterning as representative of "the goals of the text" (p. 126), whereas Subject-Theme reflects a text's method of development. In other words, the contextual prominence provided by N-Rhemes is more varied, as they provide an elaborating function, reflecting the point(s) of a text (Martin, 1992). Subject-Themes, on the other hand, orient the message to the co-text. Therefore, this study assumes the notion that "one cannot explore the semantic contribution of Theme in the written language separate from the contribution of Rheme and N-Rheme" (Fries, 2009, p. 23).

2.3. IDENTIFICATION

Referential expressions (*participants* in SFL terms) are realized via choices in the system of REFERENCE (Halliday & Matthiessen, 2013). Their contribution to texture is primarily realized in and controlled by the NGrp. However, the relationship between referring expressions is a semantic one, where the properties between them are "in no way constrained to match the grammatical class of the items it refers to" (Halliday & Hasan, 1976, p. 32). Consequently, this study draws on Martin's (1992) work on reference as a discourse semantics system, realized via choices in IDENTIFICATION, as per Fig. 3:

---

[4] Small capitals denote the name of a system in SFL.
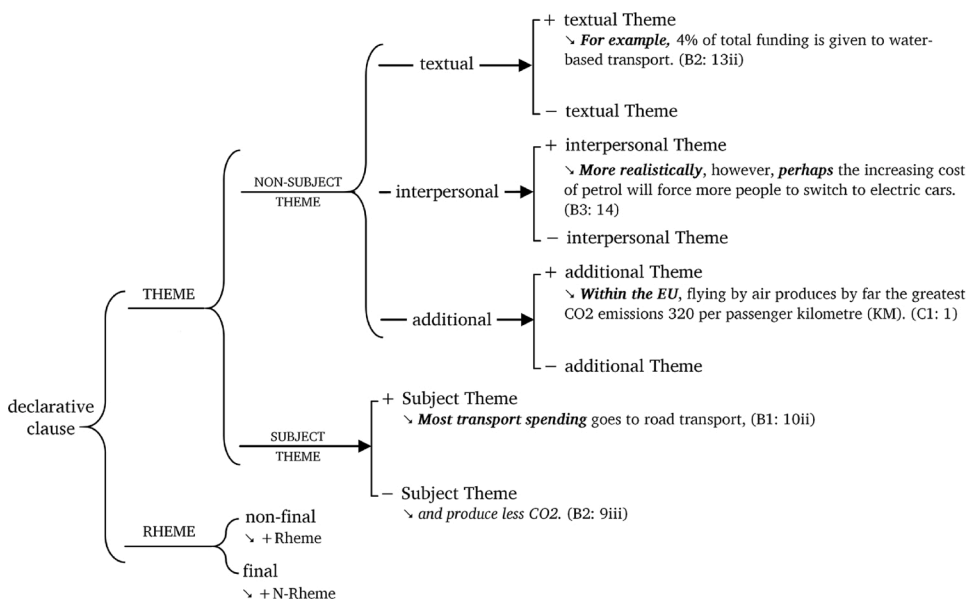[5] N-Rheme approximates to Martin's (1992) "New" and Matthiessen's (1995) "Culmination."
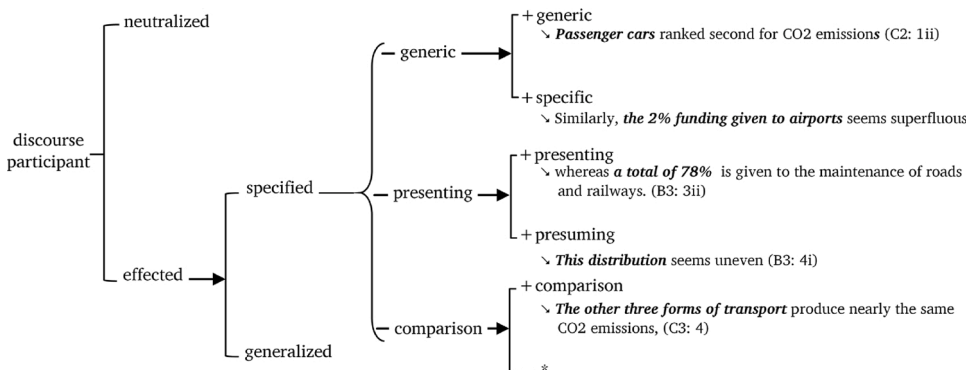
Fig. 2. System of THEME.

Fig. 3. IDENTIFICATION (Adapted from Martin, 1992; examples from our data).

Choices in IDENTIFICATION indicate if a participant is recoverable (presuming) or non-recoverable (presenting). Whenever a participant is presumed, the expectation is that it can be tracked to its original source via shared knowledge (homophora), source(s) within text (anaphora/cataphora), or sources outside text (exophora).

2.4. Conflation of THEME | INFORMATION | IDENTIFICATION

Although we present the above systems as if they exist on separate planes, they are in fact symbiotic; THEME, INFORMATION, and IDENTIFICATION can align in many ways while maintaining considerable overlap. In the unmarked form, where unmarked realizations in all three systems conflate, information flow can be conceptualized as per Fig. 4:

Although Fig. 4 is an idealization, a deviation from this unmarked configuration can be linked to a specific purpose: Rheme & Given & presuming,[6] for example, can function as a cohesive device, whereas Theme & New & presenting scaffolds discontinuity, as shown in the transition from T-unit[7] 3ii to 3iii in Fig. 5:

Moore (2012) set out to evaluate the combination of these three systems when examining the flow of information in written text. By manually coding approximately 10,000 words from nine engineering textbooks, he found that the unmarked (most frequent) combination of choices in his data was Theme & presuming, followed by Rheme & presenting ($p < .001$). Moore went on to suggest that when writers deviate from this unmarked combination, the readability of their text often suffers. Our study also looks at how these

---

[6] The symbol & denotes a systemic combination in SFL.

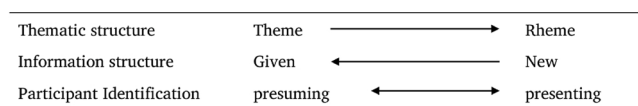[7] A T-unit is an independent clause plus any dependent clauses that may be attached to it.

Thematic structure	Theme	———————→	Rheme
Information structure	Given	←———————	New
Participant Identification	presuming	←——————→	presenting

Fig. 4. Unmarked Model of Information Flow in English.

T-unit	Theme	Rheme
3i	Even though public transport such as buses, coaches and railways *(unmarked: Theme & Given & presuming)*	produce a combined total of 145 grams per passenger kilometre *(unmarked: Rheme & New & presenting)*
3ii	they *(unmarked: Theme & Given & presuming)*	receive just 15% (New & presenting) of total EU funds for transport *(marked: Rheme & Given & presuming)*
3iii	whereas a total of 78% *(marked: Theme & New & presenting)*	is given to the maintenance of roads and railways *(New & presuming)*

Fig. 5. Manipulating Information flow through THEME, INFORMATION, and IDENTIFICATION.

three systems can aid in the readability of text, but it does so by examining the language of revisions. We use the following questions to organize and guide our discussion:

1 How does thematic revision attend to texture in L1 and L2 student writing?
2 How do L1 and L2 writers attend to texture when integrating new knowledge into their text?
3 What role do nominal group revisions play in L1 and L2 student writing with regard to maintaining thematic development, texture, and cohesion?

3. Methodology

3.1. Participants

Participants were three British L1 English users (L1 group) and three Chinese L2 English users (L2 group). We assigned pseudonyms to the participants: B1–3 for the British participants and C1–3 for the Chinese students. All participants were aged 21–23 and were enrolled full-time in a one-year Master's program in Applied Linguistics or Teaching English as a Foreign Language (TEFL) at a UK university. All had successfully completed their first semester of study (three modules over a four-month period). All had graduated with Bachelor's degrees in the Humanities and had spent the previous two years studying in English. None of the L1 participants had any specialized instruction on academic writing. The L2 group had scores of 7.0+ on the International English Language Testing System (IELTS) and had received instruction on academic writing as part of their preparation for studying on their programs.

3.2. Design and procedure

As aspiring social scientists, this study's participants are engaged in an increasingly abstract and technical discipline, characterized by increasing epistemic and social relations (*élite code*) and a rhizomatic code of varying strengths (Maton, 2020). Consequently, an ability to recontextualize and integrate knowledge in increasingly complex ways is a fundamental skill in this field. However, as subject knowledge (epistemic relation) is a known affective variable in text quality, we designed a task to minimize its effects while also requiring participants to recontextualize existing knowledge and integrate new knowledge in a relatively complex way. We piloted this task on three undergraduate L2 English users preparing for their IELTS exams and assigned a short problem-solution essay. We chose this text-type as it is a common yet challenging type of essay that calls for a presentation of a topic (field), an evaluation (tenor), and a high level of integration of these elements (mode). The task also included just one problem and one solution, which limited participants to a chain like structure: intro^problem^solution^conclusion.

In terms of the task, we asked participants to write a short essay in response to the prompt: "Discuss how governments could reduce the amount of CO_2 emissions (bar chart in Supplement A) by altering their investment in transport infrastructure (pie chart in Supplement A)." We divided the task into three parts to mimic components of the IELTS academic writing test: part 1 asked participants to describe the data presented in the two charts (introducing new knowledge); part 2 asked them to address the prompt (contextualizing new knowledge); and upon completion of parts 1 and 2, the participants received part 3, which gave them additional information to incorporate into their essay (recontextualizing old knowledge and integrating new knowledge). In other words, in part 3, we asked participants to revise earlier text in light of new information while maintaining the structure and cohesion of the essay (see Supplement A for the full task).

Before commencing the task, participants were given an information sheet on Inputlog (Leijten & Van Waes, 2013). This enabled data collection without direct observation. We told participants not to use pen and paper when composing text, but we repeatedly stressed that they could revise on the computer as much as they wanted and that they should concentrate on composing a well-structured essay. They completed all three parts of the task in a single sitting, which lasted an average of 44 min.

3.3. Data preparation and coding

We divided finished essays into T-units and numbered them. We then identified the main verbal group (Process), and delimited participants (see Supplement B for the full procedure). Group-level constituents were split into component functions according to the systems of THEME and RHEME in Fig. 2 (see above). This resulted in each clause being split into four component parts. KSL data was then used to (re)create the process aspect of the texts—i.e., we reinserted into each text previously hidden deletions (indicated through strikethrough and red font) and insertions (italic blue font) (see Supplement B).

The resultant (fully coded) texts took the form of the matrix (typos, spelling mistakes, and false starts were omitted from analysis). In Fig. 6, the writer began by typing "The next closest," but upon reaching {1} went back and made revision number 1: [1]{~~The next closest~~}FP. We classified this revision as a *Forward Progression* (FP), as it remained within the functional component currently being realized (i.e., the writer did not move outside of Theme). The writer then typed "In comparison, the next highest…," but upon reaching {2} went back and made revision number 2: [2]{~~next~~ | *second*}CP. We classified this revision as a *Commutative Progression* (CP) because the writer had crossed over component boundaries but remained within the clause currently being realized. Two other types of revision movement were *Forward Revisions* (FWDRs) and *Backward Revisions* (BWDRs). For full details of these revision types, see Supplement C (or Bowen & Van Waes, 2020).

As well as their position, revisions were also categorized for function. For example, a revision that increased the evaluative stance of a message in terms of judgement, appreciation, or affect (Martin & Rose, 2007) was labelled "+interpersonal"–denoting a choice within the interpersonal system. Similarly, a revision that modified a participant, as in "The rest of the [28]{*EU transport*} infrastructure systems," where "*EU transport*" was added as a Classifier, was labelled + EXPANSION.

4. Results and discussion

We begin our reporting of the results with an overview of the students' writing activity.

4.1. Overview of writing activity

Table 1 shows that the L1 group typed, on average, 2102 characters more than the L2 group, which is reflected in their higher average word count (619.67 vs. 375). Table 1 also shows that the L1 group spent more time writing/typing ($M = 33.13$, $SD = 10.51$) than the L2 group ($M = 22.32$, $SD = 2.76$). Table 2 shows overall revision activity.

Table 2 shows overall revision activity. The bottom row displays the ratio of all revisions to functional revisions (see Supplement D for the breakdown of functional revisions) and enables a broad comparison of functional revisions across the cohort, from 6.20 to 15.82 ($M = 9.97$, $SD = 3.45$).

4.2. Thematic revision and texture

We now move on to explore the function of revisions with regard to texture and cohesion, beginning with RQ 1: How does thematic revision attend to texture in L1 and L2 student writing?

4.2.1. Revision placement: Theme^Rheme
This section begins with a consideration of revision placement with respect to each functional component in THEME, as shown in Table 3.

Table 3 shows the main difference between groups is the level of revision within Theme.[8] Proportionally, the L2 group paid more attention to the development of the clause (RHEME) than the point of departure (THEME). However, these figures may be misleading because Rheme incorporates the main Process, and as discussed below, this area was problematic for two L2 writers.

4.2.2. Revisions within theme
With regard to revisions within Theme, the L1 group often placed experiential meanings before the Subject to contextualize clauses (additional Themes = 20.4%), which accords with research into the increased use of complex Themes by advanced L1 writers (He, 2020). An example of a complex Theme is shown in Example 1:

---

[8] There was also a notable difference in that the L1 group inserted 17 clauses and the L2 group inserted three.

T-unit	Point of departure: THEME ➡➡➡		Development of clause: RHEME ⬅⬅⬅	
	Theme	Subject-Theme	Rheme	N-Rheme
2i	[1]{~~The next closest~~}FP {1} In comparison,	the [2]{~~next~~ | *second*}CP highest amount of emissions	stem from {2}	[44]{*petrol or diesel powered*}BWDR {45} passenger cars at 130 grams per passenger KM,

Fig. 6. Example of Coding Matrix for Writing Activity.

Table 1
Summary of Writing Activity.

	L1 group				L2 group			
	B1	B2	B3	Mean	C1	C2	C3	Mean
Characters in Process	5448	5432	3524	4801.33	2532	2999	2568	2699.67
Characters in Product	3380	4785	3304	3823	1971	2526	2447	2314.67
Final word count	517	778	564	619.67	327	405	393	375
Writing time (min)	37:19	41:00	21:10	33.13	19:52	25:03	22:41	22.32
Time on task (min)	52:26	48:06	36:39	45.57	36:43	56:44	32:45	41.77

Table 2
Summary of Revision Activity.

	L1 group				L2 group			
	B1	B2	B3	Total	C1	C2	C3	Total
All revisions	617	404	148	1169	217	254	109	580
Functional revisions	39	48	18	105	24	41	12	77
Ratio	15.82	8.42	8.22	11.13	9.04	6.20	9.08	7.34

Table 3
Breakdown of Revisions in Each Functional Slot.

	Theme	Subject-Theme	Rheme	N-Rheme	Total
L1 group	32	18	13	25	88
L2 group	12	11	19	31	73

Example 1 (B1)

T-unit		Theme	Subject-Theme	Rheme
N-Rheme				
9	In what follows I will argue that [14]{~~the comparatively low~~}FP {14} because [15]{*carbon dioxide*}FP emission {15} by air transport is so high compared to [16]{*most*}FP other forms {16} of transport,	it	is the key target for reducing	Its [17] {~~combined~~}FP {17} gross carbon footprint.

In this example, revision 15 adds a Classifier ("*carbon dioxide*") to disambiguate a head noun (emission), while revision 16 adds an indefinite quantifying Numerative ("*most*"). Though revision 15 is redundant, as only one type of emissions is mentioned throughout, this T-unit constitutes a hyper-Theme (similar to what is traditionally known as a topic sentence); thus, the addition enhances the point of departure for a larger phase of discourse. Consider also Example 2:

Example 2 (B1)

T – unit	Theme	Subject-Theme	Rheme	N-Rheme
11i		Most [21]{*transport*}FP spending {21}	goes to	road transport,

Classifiers, such as those in Example 1 and revision 21 above, organize a referent into a mutually exclusive set, where "nouns serving as Classifiers have no determination system, so they do not refer to particular participants only to generalized ones" (Halliday & Matthiessen, 2013, p. 378). In Example 2, for instance, the thing (spending) is labelled as belonging to the subset "transport." This extended categorization (or taxonomizing) is frequent in academic English (Halliday, 2004) and was common in the L1 group's revisions (B1: "[15]{*carbon dioxide*} emission," "Most [21]{*transport*} spending," "[34]{*New*} infrastructure;" B2: the [28]{*EU transport*} infrastructure systems," "[44]{*petrol or diesel powered*} passenger cars;" B3: "total [2]{*EU*} funds;" "expenditure on [3]{*existing*} transport infrastructure"). Overall, Classifiers allow a writer to create generic referents—which are inherently presuming—and, thus easier to present as Given within thematic peaks.

Although overall revision was much lower in the L2 group, similar alterations to NGrps occurred within thematic components. For

instance, in Example 3, revision 22 deletes a head noun ("f") and adds an inexact Numerative ("amount of"). In referential terms, this shifts the Subject-Theme toward non-specific.

Example 3 (C1)

T- N- 14	unit Rheme Although it [21]{does not}FP {21} did not cost much for airport development	Theme the [22]{CO2 emission figure}FP {22} amount of CO2 emission	Subject- Theme will result in	Rheme very serious Impact on our environment. {23}.

Similarly, in Example 4, C2 decreases the specificity of the NGrp within Theme via revision 8, which replaces a definite determiner with an indefinite comparative: [8]{these | *different*}:

Example 4 (C2)

T- N- 9	unit Rheme With regards to [8]{these / *different*}FWDR {9} transportation forms,	Theme EU funds	Subject-Theme [10]{*was*}CP {11} distributed	Rheme [9]{funds}CP {10} differently [3]{among various}FWDR {4}[17] {*in 2002*}INS {18}.

Similar changes are made in C2's other revisions: "the bar [31]{chart} and the pie chart" and "a [37]{new form | *brand new*} transportation tool."

However, whereas the L1 group actively sought to maintain unmarked conflations, the L2 group was not so consistent. In Example 5 below, for instance, C2 makes a revision that results in a marked conflation of Rheme & New & presuming:

Example 5 (C2)

T-unit 5ii	Theme	Subject-Theme Which	Rheme ranked	N-Rheme [15]{*the*}FWDR{16} second place for transportation CO2 emissions.

While this revision may well stem from the difficulty of using the definite article in English, it transforms an initially well-written clause into a somewhat awkward one.

Although revisions to experiential meanings within Theme were common in the L1 group (30.8 %), of revisions), revisions to interpersonal meanings within Theme were much rarer (8.2 %); moreover, revisions to interpersonal Themes (e.g., thematized comments) never occurred. Although the sample-size limits conclusions, one hypothesis could be that through repeated use in this locale, these "objective" frames have become general purposive and are reminiscent of formulaic language use. For example, the following were inserted as whole constructions with relatively little pause time between the typing of each word: "[32]{*It should however be noted that*}" (B1); "[7]{*It is possible that*}" (B2); "[18]{*It is worth noting that*}" (B3).

4.2.3. Development of the clause

Here the data reveal a symbiosis between THEME and RHEME, where changes in one system can open up options in the other, as shown in Example 6:

Example 6 (B2)

T- N- 10	unit Rheme [5] In what follows I will argue that [33]{*if CO2 emissions are to be reduced*}CP{34}	Theme [32]{money / *the current distribution of EU funds*}{33}FWDR	Subject- Theme need to be	Rheme redistribution in [34]{order to cause a change in the current system / *a more efficient manner*}CP {35}

In Example 6, B2 transforms a relatively simple proposition (redistributing money will cause change) into a well-developed and circumspect hyper-theme. Firstly, revision 32 (money "| *the current distribution of EU funds*") replaces a simple Subject-Theme (money) with a complex NGrp that includes a GM (nominalization). B2 then deletes the proposition in the N-Rheme and moves to the Theme to add "*if CO2 emissions […]*." Following this, B2 makes revision 34: [34]{*a more efficient manner*}CP. This final revision adds implicit appraisal that pairs the N-Rheme to the Subject-Theme, creating a hyper-Theme that is one-half problem, one-half solution; i.e., the problem (the distribution of EU funds) will be discussed in relation to the solution (need to be efficiently redistributed). This example highlights how revisions that affect global-level concerns can call for complex choices in multiple systems and at multiple levels. However, such activity was notably absent amongst the L2 group, which accords with the oft-cited finding that less experienced revisers often focus more on localized revisions, choosing to focus predominantly on form.

4.3. Incorporating new knowledge into text

We now look at how the writers incorporated new knowledge (part 3 of the task), starting with Example 7:

Example 7 (B2)

T-unit	unit	Theme	Subject-Theme	Rheme
N-	Rheme			
2i	[1]{The next closest}FP{1} In comparison,	the[2]{next /second}CP highest amount of emissions	stem from {2}	[44]{petrol or diesel powered}CP{45} passenger cars at 130 g per passenger KM,
3	[46]{On the other hand,}CP{47}	[45]{electric cars, which are relatively new to the market,	produce	only 20 g CO2 per passenger KM,}FP{46} [47]{making it by far the lowest CO2 emitting mode of transport currently available.}CP{48}

In this example, revision 44 (¨*petrol or diesel powered*¨) adds an Epithet (a function in SFL, typically realized as an adjective) to a noun ("cars") that was already pre-modified by a Classifier ("passenger") and post-modified by a Qualifier (prepositional phrase). While this revision limits the interpretation of the N-Rheme, it also reveals how old text can be reworked to incorporate new text. Looking for a place to insert text, B2 seems to realize that the original formulation of "passenger cars" could also be construed as categorizing "electric cars" (new knowledge about to be introduced via revision 45). Consequently, the original referent is made more specific by taxonomizing it further via an Epithet. This allows T-unit 3 to be integrated via a textual Theme^Subject-Theme realization. As a result, revision 44 allows "electric cars" to be presented as Given (directly comparable to "petrol or diesel powered cars"). Similarly, consider Example 8:

Example 8 (B2)

T-unit	Theme	Subject-Theme	Rheme	N-Rheme
12	[50]{However, even with the shift to electric powered cars	*this figure*	*seems*	*slightly high*}FWDR{51}

This time a textual Theme^additional Theme^ Subject-Theme combination is used to create a constant Theme progression. However, due to the length of the fronted dependent clause acting as additional Theme, it appears to take its own tone unit when read back. Such a choice enables a writer to maintain an unmarked conflation of Theme & Given & presuming and Rheme & New & presenting amongst the major participants.[9] Similar combinations are seen in Examples 9 and 10, only with textual Theme^interpersonal Theme^Subject-Theme combinations:

Example 9 (B2)

T-unit	Theme	Subject-Theme	Rheme	N-Rheme
14	[51]{More realistically, however,	*the increasing cost of petrol*	*will force*	*more people to switch to electric cars.*}FWDR

Example 10 (B2)

T-unit	Theme	Subject-Theme	Rheme	N-Rheme
21	[52]{However, I would suggest that	*the majority of redistributed funds*	*be taken from*	*the railways, which currently stamds at 26%.*}FWDR

The Subject-Themes in these examples are both experiential metaphors. However, in this instance there is a double layering of GM, as metaphorical figures are embedded as participants within other metaphorical configurations (i.e., incongruent material processes in Rheme). This enables B2 to condense a lot of information into these clauses while simultaneously constructing participants that are Given & presuming.[10]

Overall, all writers used GM to some extent: L1 group ($M = 3.66$, $SD = 1.25$); L2 group ($M = 2.33$; $SD = 0.47$). This is unsurprising given the potential that GM affords writers, as noted above. However, there was a difference between the groups in terms of how GMs were used in combination with textual devices, namely conjunctive relations. For instance, in Example 11, B1 uses an experiential metaphor as part of an embedded clause to create a rather lengthy additional Theme. However, this may not have been possible without the inclusion of the conjunctive relations inserted through revision 6 and the noun phrase inserted during normal production:

Example 11 (B

T-	unit		Theme	Subject-Theme	Rheme

(continued on next page)

---

[9] Given is placed within Subject-Theme—a textual metaphor acting as an anaphoric referent (this figure).

[10] "the increasing cost of petrol" presumes homophora (common knowledge), and "the majority of redistributed funds" presumes and summarizes something already mentioned.

(*continued*)

N-11	Rheme			
	[6]{~~Although~~ / *Firstly, because*}BWDR{17} cars, buses and coaches [35]{~~use the roads~~ / *make up the majority of peoples transport needs,*}FP{36}	the current expenditure of 52% [36]{*on roads*}CP{37}	[49]{~~is~~ / *seems*}	~~slightly high~~{28}/ *appropriate.*}FWDR{50}

Here, revision 6 ("¨~~Although~~ | *Firstly, because*") changes a projected relationship from concessive to expectant. Later on, revision 35 changes what is presented as Given within Theme (the focal point of the contextual frame) from "the roads" (a referent of a generic inanimate class) to "the majority of people's transport needs." This results in a complex NGrp presupposing that there is such a thing as transport needs and that these needs can be quantified and evaluated. Such complex manipulations highlight how the two groups of writers differed with respect to a focus on macro and micro-level revisions. Specifically, the L1 group seemingly attended more to global-level concerns by revising across functional components and clauses, often utilizing conjunctive relations to help them achieve their aims ($n = 26$). The L2 group mainly revised within functional components rather than across them or across clauses, with little deployment of conjunctive relations ($n = 8$). This amounts to a difference in the deployment of FPs, CPs, FWDRs, and BWDRs, as shown in Table 4:

Table 4 shows that both groups show a preference for FPs, which represent revisions that occur close to the point of inscription. The main difference is in the deployment of FWDRs (35 to 14), which mainly reflect revising while proofreading (see Bowen, 2019), and CPs (21 to 12), which reflect revising across functional components. Despite the small sample size, Table 4 provides further evidence for the well-known finding that less-experienced writers tend to focus on local-level revisions, while experienced writers often attend more to global-level revisions.

Another difference between the groups in revisions to Theme involved the insertion of thematized comments, as exemplified in Example 12:

Example 12 (B3)

T-N-9i	unit Rheme	Theme	Subject-Theme	Rheme
	[18]{*It is worth noting that*}FWDR{19}.	there	is also a growing swell of vocal support for	the use of hybrid/electric cars [19]{*as a solution to this,*}CP{20}

In Example 12, the thematized comment ("*It is worth noting that*") is inserted to contextualize the upcoming clause in light of a depersonalized evaluative stance. Moreover, in this instance, the depersonalized alignment is further enhanced by the addition of the causal consequence: [19]{*as a solution to this*}. However, thematized comments were absent in the L2 data.

4.4. Nominal group revisions and texture

When focusing on NGrp specificity, there are four key structural elements: (Determiner)*ˆ(Modifier)*ˆThing ˆ(Qualifier)*,[11] where specifying potential decreases from Determiner to Qualifier. In terms of these elements, we coded revisions as a change in EXPANSION, representing a choice to elaborate, enhance, or extend upon the head noun (Thing), as shown in Table 5.

Table 5 shows that NGrp revision was a key focus for both groups. The L1 group made a total of 105 functional revisions, and the L2 group made 76. This means that 31.43 % of the L1 group's revisions, and 44.74 % of the L2 group's revisions involved NGrp modifiers. This finding confirms the overall importance of NGrps in written academic English and in revising academic text in general (Bowen, 2019).

In terms of contributing to texture, many of the revisions involving modifiers represented important choices when introducing (presenting) or tracking (presuming) referential entities and, thus directly affected the texture and cohesiveness of a text. For instance, in addition to the examples already discussed (particularly those in the previous section), revision 34 in Example 13 illustrates how adding an Epithet and deleting a Qualifier ("*New* | infrastructure | ~~for these vehicles~~") creates a generic referent ("New infrastructure") that also increases presuppositional content (intensity of presuming reference).

Example 13 (B1)

T-N-13i	unit Rheme	Theme	Subject-Theme	Rheme
	[39]{~~However,~~}FP{40}	[34]{*New* / infrastructure / ~~for these vehicles~~}CP{35}	is [35]{*needed*/ ~~significant and will become more~~ / *for these vehicles*}CP{34}	([36] {*for example,*} FP {37}charging points in service ststions,research and development to improve the technology), {36}

Conversely, revision 35 ("*needed for these vehicles* | ~~significant and will become more~~") adds a specific referent while decreasing presuppositional content; i.e., it intensifies the gradation of presenting reference within Rheme. The origin of these kinds of contextualizing Qualifiers ("*for these vehicles*" in Example13) is said to be typically cataphoric (Matthiessen, 1995). However, in many cases in

---

[11] The asterisk* denotes the element can be repeated; parenthesis indicates optionality.

Table 4
Deployment of Revision Types.

	L1 group				L2 group			
	B1	B2	B3	Total	C1	C2	C3	Total
FPs	25	10	7	42	19	16	7	42
CPs	7	11	3	21	2	9	1	12
FWDRs	4	24	7	35	2	12	0	14
BWDRs	3	3	1	7	1	6	1	8
				105				76

Table 5
Frequency of NGrp Modifications (±EXPANSION) in Revision Activity.

	L1 group				L2 group			
	B1	B2	B3	Total	C1	C2	C3	Total
±EXPANSION	11	15	7	33	10	17	7	34

the corpus, Qualifiers added through revision were anaphoric. Consequently, when text is read back, many appear unnecessary—that is, they created over-specified referents. There were 14 such instances of Qualifiers in our data from the L1 (10) and L2 group (4), including B1: "EU spending [4]{*on transport infrastructure*}," "emerging technologies [44]{*in all types of transport,*}"; B2: "major transport [27]{*within the EU*}", "52 % [36]{*on roads*}", C1: "sea ports [13]{*of EU transport funds.*}"; and C2: "differently [17]{*in 2002*}", and "the most environmental friendly choice [32]{*in daily life,*}". Redundancy occurred as well in revisions involving pre-modifiers (e.g., Classifiers as discussed in Examples 1, 2, and 7 above), though on a reduced scale ($n = 10$). However, compared to the L1 group, the L2 group shows reduced revision activity (9:1 ratio).

Over-specified referents include non-unique identifying information, or information not relevant to recoverability. While research has revealed that most over-specifications do not alter recoverability (Arts, Maes, Noordman, & Jansen, 2011), there has been no research into the processability of over-specification in academic texts. Nevertheless, whenever revision led to over-specification, for these writers it usually involved exophoric reference (entities originating outside the text). Consequently, in terms of texture and cohesion, these revisions directly contributed to the creation of text rather than integration of text (Halliday & Hasan, 1976). However, taking into consideration that these revisions took place after initial text construction, and by examining their effect on surrounding text, they often reveal a secondary purpose that does contribute to texture; namely, that over-specifications within marked Theme served to elongate this component with regard to the T-unit as a whole. As Halliday (2004) noted, such complex constructions within Theme often attract their own tone groups, making the information contained within them more prominent when read back. In the end, a renegotiation of tenor, with writers becoming readers, may serve to demarcate what was one tone unit into two, allowing new information to be placed apart from the main clause. This in turn allows an unmarked conflation of Theme & Given & presuming to be maintained.

Moreover, as the specifying potential of a modifier increased, its frequency of occurrence in revision activity amongst the writers decreased; i.e., modifications involving Determiners (the constituent with the most specifying potential) were almost non-existent, whereas moving from left to right in terms of structural ordering, there were increasing frequencies: Epithets (5), Classifiers (8), and Qualifiers (14).

In other cases, revisions to participants were clearly centered on semantic considerations, where "logical sequences of figures are reconstrued as experiential configurations of elements" (Halliday & Matthiessen, 2013, p. 715), freeing up referential potential in the form of down-ranked figures or phrases utilizing GM, as shown in Examples 14 and 15:

Example 14 (C1)

T-unit		Theme	Subject-Theme	Rheme
N-Rheme				
11i	[27]{*Moreover,* [28]{hybrid or electric cars}FP {28} *owing to the amount of CO2 emission produced by hybrid/ electric cars,*	*the government*	*May need to* [29]{reinforce / encourage}FP {29}	*population to choose* [30]{these / them}FP {30}FWDR

In Example 14, a simple noun phrase is entered ("hybrid or electric cars") and then replaced with a down-ranked figure. As Halliday and Matthiessen (2013) noted, "a figure, realized metaphorically by a NGrp rather than congruently by a clause, gains access to the textual systems of the NGrp—most significantly, the system of DETERMINATION" (p. 716).[12]

Similarly, in Example 15, B1 initially enters "the redistribution" as Subject-Theme. However, after a pause of 10.2 s, revision 28 changes the nucleus of the Subject-Theme from an experiential metaphor ("redistribution") to a logical metaphor ("*first stage*") and adds an adverbial appraisal ("*meaningful*").

---

[12] DETERMINATION is the grammatical correlate of IDENTIFICATION.

Example 15 (B1)

T-unit	Theme	Subject-Theme	Rheme	N-Rheme
18ii	[…]	the [28]{*first stage in meaningful*}FP redistribution {28}	is	[…]

While the relatively small data set precludes definitive conclusions, in revising thematic components, writers can take advantage of syntax (NGrp modifiers), semantics (GMs), or both, as a powerful means to construe/reconstrue and expand/condense key elements, and thus build upon epistemic relations within the text while also moving the text forward cohesively.

5. Conclusion

Overall, we have attempted to explore how revisions involving THEME, INFORMATION, and IDENTIFICATION can contribute to an understanding of how texture is created and maintained in real time in writing. Consequently, and by bringing together KSL and SFL, it is hoped that we have made several contributions to L2 writing research.

First, by coding revisions in terms of language functions and structures, we have attempted to fill a void in L2 writing research while also forwarding an emerging dialogue between KSL and a text linguistic approach (Bowen & Van Waes, 2020;). Accordingly, the study showed one possible way in which KSL can be used to reveal hitherto unseen language choices in written text while also showing that exploring a phenomenon as complex as texture may benefit from a more consolidated view of the language systems involved. In this light, this study is the first to explore multiple overlapping language choices in written text as it unfolds, and the first to explore how both L1 and L2 writers integrate new knowledge while attempting to maintain textual unity in digitally constructed texts.

Second, by examining revisions to thematic elements, the findings indicated that a key textual resource for both groups of writers was the ability to create more informationally dense, interconnected, and coherent text through revisions involving THEME. The L2 writers, for instance, made good use of coordinating conjunctions and demonstrative pronouns in this part of the clause. In contrast, the L1 group made good use of substitution, complex NGrp modifications, and downranked (embedded) units, which they often used to pack more information (often presupposed) into Theme.

Third, by challenging writers to incorporate new information, we showed how Themes became complex sites of logogenetic potential, where writers attended to an unmarked correlational model of information flow in two ways, through: (a) elongating Themes (often via over-specification of NGrps), which led to additional "tone" units that could, theoretically, accommodate new content within the departure of the clause; and (b) choices in EXPANSION (pre- and post-modification of NGrps) that renegotiated participants with respect to field (descriptions/categorizations), tenor (increased formality), and mode (presuming/presenting identities).

Fourth, we provided further evidence of the importance of the NGrp to academic writing. Namely, by examining revisions to NGrps in light of their contribution to unfolding texture, the results suggest that Classifiers (pre-modifiers) and Qualifiers (post-modifiers) seem to have the greatest potential for fine-tuning a referent's complexity and identity.

Therefore, although further research is needed, overall results suggest that NGrps and prepositional phrases could be key areas for explicit instruction when revising academic texts for texture. Specifically, teaching students how to take advantage of NGrp elements (Determiner*ˆModifier*ˆThingˆQualifier*) may help them manipulate referents for complexity (semantic density) through EXPANSION while simultaneously renegotiating co-textual/contextual dependency (semantic gravity) through IDENTIFICATION. This, in turn, can increase the potentiality of their choices in INFORMATION (Given/New). For example, teachers can make students aware of the potential of Classifiers to not only elaborate upon the Thing, but also as a means to tie it to its co-text (through anaphora)[13] or context (through homophora), establishing its NGrp as presumed and Given. Similarly, an explicit focus on prepositional phrases may help writers construct complex Qualifiers (most Qualifiers being prepositional phrases) and additional (marked) Themes. Qualifiers, for example, while being the modifier with the least specifying potential, and thus having generally less effect on semantic gravity, arguably hold the greatest potential when it comes to increasing semantic density. Specifically, teachers may find value in highlighting the fact that Qualifiers are almost always downranked units (a form or recursive embedding), which means they have increased potential for construing complex instantial meanings that are not inherently tied to the Thing they modify. Prepositional phrases can also serve as additional logical connectors (providing a complimentary strategy to conjunctive adjuncts) and assist in creating additional tone units (via over-specification leading to elongated themes).

Moreover, although specific issues will vary depending on the context, group of writers, and text-type, there are additional potential pedagogical benefits from taking a dynamic approach to L2 writing research. Most notably, having such insight into students' *languaging*—language play, indecision, and eventual resolution of their authorial decision-making—can enable teachers to forestall unhelpful and/or inefficient writing practices. For instance, the L2 writers in the current study were repeatedly indecisive about what tense/aspect to use, as evidenced by their revisions. By having such information, a practitioner working with this group, or a larger one similar in demographics, could intervene and shift students' attention to larger, more important considerations with respect to the task. In our problem-solution task, for example, time may have been better spent clarifying the problem or fine-tuning warrants.

Conversely, through KSL data, languaging can be brought to the forefront by the teacher and be used to stimulate collaborative dialogue as to why a writer is manipulating (play) and/or deliberating (indecision) over certain choices. In our dataset, for instance, L2 writers showed uncertainty with regard to DETERMINATION (a textual system that deals with nominal deixis; i.e., determiners). A teacher working with this data could use such languaging information to provide instruction on why these elements can be difficult for L2 users

---

[13] A previously introduced class or property relating to the Thing can be presented as a Classifier.

while also legitimizing their (previously hidden) attempts at such complex language choices. Moreover, previously unseen languaging can also be incorporated into lessons based on authentic examples from students' writing. This, then, can lead to "exploration of scientific [wave-like] discourse in a broader sense" (Larsson, 2018, p. 74), where students, with the guidance of a teacher, can explore the various ways in which discourse can be packed and unpacked through the language structures they are manipulating (as per Musgrave & Parkinson, 2014). One instance of this can be seen in Example 6 above. Prior to any revision, the participant wrote: "In what follows I will argue that money needs to be redistributed […]". After multiple revisions, this part of the clause evolved into: "In what follows I will argue that *if CO_2 emissions are to be reduced, the current distribution of EU funds* needs to be redistributed […]". A teacher could break down what these revisions have contributed to the text and highlight how the writer has packed more meaning into the clause through the embedding of a conditional clause and inclusion of a complex NGrp. As Maton (2020) noted, "teaching students how to master semantic waves is an urgent task" which requires "bespoke means for knowledge-building" (pp. 79–80). In this way, a dynamic approach lends itself to realizing more nuanced learning opportunities, especially for learners at intermediate and advanced levels of proficiency.

6. Limitations and future research

Firstly, the study focused only on externalized processes, yet many more distributed processes occur internally, including language choices not directly realized on the screen. Future studies could make use of think-aloud protocols, video observations, and post-task interviews to further contextualize language choices.

The second limitation is the sample size. Larger samples would lend themselves to a more delicate categorization system. For example, instead of labelling EXPANSION as increasing (+) or decreasing (−), choices might be coded in terms of elaboration, extension, or enhancement and further linked to semantic ranges in terms of semantic gravity and semantic density (see Maton, 2020).

It may also be advantageous to give participants the first two parts of the task beforehand, so that they may familiarize themselves with it and make an outline. Alternatively, to increase ecological validity, future studies could sample graded assignments. Moreover, the study only examined one text-type: problem^solution. While this was based on the assumption that a more challenging text-type would require higher levels of knowledge integration, other text-types clearly call on different meanings and mechanisms to create texture. For instance, narratives typically rely on prepositions of time and place for much of their cohesive patterning.

In terms of future research, the coding system for revision movement and revision placement in functional terms opens up an interesting space for future "process" analysis. Moreover, by using KSL and SFL, and incorporating developments from LCT, it is possible to profile writing as it happens in terms of semantic profiles (K. Maton, personal communication, June 22, 2020). One could, for example, use the translation devices for semantic density (Maton & Doran, 2017) and semantic gravity to chart the semantic ranges of writers to further research-based pedagogy. We are hopeful that future research could also move to larger scale studies, involving multiple writers across multiple sessions. This is important from a theoretical point of view, as process research shows that it is sometimes hard to interpret the indirect relation between process indicators and underlying cognitive processes. Overall, we hope that this study will inspire other L2 writing scholars to explore further possibilities to combine different research methods, both from a process and product perspective.

Declaration of Competing Interest

We hereby declare that this work has not been previously published, and is not currently being considered for publication in any other venues. Furthermore, we acknowledge that we will not submit this work to any other venue until we have written notification of a decision from the editorial board of the Journal. The authors have no conflicts of interest to disclose. No special funding was received for this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jslw.2020.100773.

References

Alexander, O. (2019). The contribution of Halliday to EAP writing instruction: A personal journey. *Journal of English for Academic Purposes, 41*, 1–9. https://doi.org/10.1016/j.jeap.2019.100769

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics, 43*(1), 361–374. https://doi.org/10.1016/j.pragma.2010.07.013

Baratta, A. M. (2010). Nominalization development across an undergraduate academic degree program. *Journal of Pragmatics, 42*(4), 1017–1036. https://doi.org/10.1016/j.pragma.2009.08.007

Bowen, N. (2019). Unfolding choices in digital writing: A functional perspective on the language of academic revisions. *Journal of Writing Research, 10*(3), 465–498. https://doi.org/10.17239/jowr-2019.10.03.03

Bowen, N., & Van Waes, L. (2020). Exploring revisions in academic text: Closing the gap between process and product approaches in digital writing. *Written Communication, 37*(3), 322–364. https://doi.org/10.1177/0741088320916508

Byrnes, H. (2006). What kind of resource is language and why does it matter for advanced language learning? An introduction. In H. Byrnes (Ed.), *Advanced language learning: The contribution of halliday and vygotsky* (pp. 1–28). Continuum.

Coffin, C., & Donohue, J. P. (2012). Academic Literacies and Systemic Functional Linguistics: How do they relate? *Journal of English for Academic Purposes, 11*(1), 64–75. https://doi.org/10.1016/j.jeap.2011.11.004

Davies, F. (1997). Marked theme as a heuristic for analysing text-type, text, and genre. In J. Pique, & D. J. Viera (Eds.), *Applied languages: Theory and practice in ESP* (pp. 45–80). Universitat de Valencia.

Davies, M. (2004). Two truths…of the imperial theme. In D. Banks (Ed.), *text and texture: Systemic functional viewpoints on the nature and structure of text* (pp. 51–107). L'Harmattan.

Downing, A. (1991). An alternative approach to Theme: A systemic-functional perspective. *WORD, 42*(2), 119–143. https://doi.org/10.1080/00437956.1991.11435835

Fries, P. H. (1992). The structuring of information in written English text. *Language Sciences, 14*(4), 461–488. https://doi.org/10.1016/0388-0001(92)90026-B

Fries, P. H. (2002). The flow of information in a written text. In P. H. Fries, M. Cummings, D. Lockwood, & W. Spruiell (Eds.), *Relations and functions within and around language* (pp. 117–155). Continuum.

Fries, P. H. (2009). The textual metafunction as a site for a discussion of the goals of linguistics and techniques of linguistic analysis. In G. Forey, & G. Thompson (Eds.), *Text type and texture* (pp. 8–44). Equinox.

Guo, H., Deane, P. D., Rijn, P. W., Zhang, M., & Bennett, R. E. (2018). Modeling basic writing processes from keystroke logs. *Journal of Educational Measurement, 55*(2), 194–216. https://doi.org/10.1111/jedm.12172

Halliday, M. A. K. (2004). *The language of science.* Continuum.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English* (Vol. 9). Longman.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2013). *An introduction to functional grammar* (4th ed.). Routledge.

Hasan, R. (1985). The texture of a text. In M. A. K. Halliday, & R. Hasan (Eds.), *Language, context, and text: Aspects of language in a social-semiotic perspective* (pp. 70–96). Deakin University Press.

He, M. (2020). A review on studies of theme in academic writing. *Journal of Contemporary Educational Research, 4*(3), 25–35. https://doi.org/10.26689/jcer.v4i3.1116

Herriman, J. (2011). Themes and theme progression in Swedish advanced learners' writing in English. *Nordic Journal of English Studies, 10*(1), 1–28. https://doi.org/10.35360/njes.240

Klein, P. D., & Unsworth, L. (2014). The logogenesis of writing to learn: A systemic functional perspective. *Linguistics and Education, 26*, 1–17. https://doi.org/10.1016/j.linged.2013.12.003

Larsson, P. N. (2018). "We're talking about mobility:" Discourse strategies for promoting disciplinary knowledge and language in educational contexts. *Linguistics and Education, 48*, 61–75. https://doi.org/10.1016/j.linged.2018.10.001

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research using Inputlog to analyze and visualize writing processes. *Written Communication, 30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research, 5*(3), 285–337. https://doi.org/10.17239/jowr-2014.05.03.3

Liardét, C. L. (2013). An exploration of Chinese EFL learner's deployment of grammatical metaphor: Learning to make academically valued meanings. *Journal of Second Language Writing, 22*(2), 161–178. https://doi.org/10.1016/j.jslw.2013.03.008

Lindgren, E., & Sullivan, K. P. H. (Eds.). (2019). *Observing writing: Insights from keystroke logging and handwriting.* Brill. https://doi.org/10.1163/9789004392526.

Martin, J. R. (1992). *English text: System and structure.* John Benjamins. https://doi.org/10.1075/z.59

Martin, J. R. (2007). Construing knowledge: A functional linguistic perspective. In F. Christie, & J. R. Martin (Eds.), *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives* (pp. 34–64). Continuum.

Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed.). Continuum.

Maton, K. (2020). Semantic waves: Context, complexity and academic discourse. In J. R. Martin, K. Maton, & Y. Doran (Eds.), *Academic discourse: An inter-disciplinary dialogue* (pp. 59–85). Routledge.

Maton, K., & Doran, Y. J. (2017). Condensation: A translation device for revealing complexity of knowledge practices in discourse, part 2 - clausing and sequencing. *Onomázein*, 77–110. https://doi.org/10.7764/onomazein.sfl.04

Matthiessen, C. M. I. M. (1995). *Lexicogrammatical cartography: English systems.* International Language Science Publishers.

Moore, N. (2012). Modelling the flow of discourse in a corpus of written academic English [paper presentation]. In *Proceedings of Corpus Linguistics 2011 - "Discourse and Corpus Linguistics".*

Musgrave, J., & Parkinson, J. (2014). Getting to grips with noun groups. *ELT Journal, 68*(2), 145–154. https://doi.org/10.1093/elt/cct078

O'Donnell, M. (2013). A dynamic view of choice in writing: Composition as text evolution. In L. Fontaine, T. Bartlett, & G. O'Grady (Eds.), *Systemic functional linguistics: Exploring choice* (pp. 247–266). Cambridge University Press. https://doi.org/10.1017/CBO9781139583077.016.

Ravelli, L. J. (2004). Signalling the organization of written texts: Hyper-themes in management and history essays. In L. J. Ravelli, & R. A. Ellis (Eds.), *Analysing academic writing: Contextualized frameworks* (pp. 131–152). Continuum.

Ryshina-Pankova, M., & Byrnes, H. (2013). Writing as learning to know: Tracing knowledge construction in L2 German compositions. *Journal of Second Language Writing, 22*(2), 179–197. https://doi.org/10.1016/j.jslw.2013.03.009

Schleppegrell, M. J. (2004). Technical writing in a second language: The role of grammatical metaphor. In L. J. Ravelli, & R. A. Ellis (Eds.), *Analysing academic writing: Contextualized framework* (pp. 172–189). Continuum.

Thompson, G. (2010). Grammatical metaphor and success in academic writing. In S. Hunston, & D. Oakey (Eds.), *Introducing applied linguistics: Concepts and skills* (pp. 27–34). Routledge.

Ventola, E. (1987). *The structure of social interaction: A systemic approach to the semiotics of service encounters.* Pinter.

Wei, J. (2016a). Effects of instruction on Chinese college students' thematic choice in academic writing. *Journal of Education and Practice, 7*(14), 60–68.

Wei, J. (2016b). Thematic choice in Chinese college students' English essays. *English for Specific Purposes, 41*, 50–67. https://doi.org/10.1016/j.esp.2015.09.003

Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research, 15*(3), 343–362. https://doi.org/10.1177/1362168811401154

Xu, C. (2018). Understanding online revisions in L2 writing: A computer keystroke-log perspective. *System, 78*, 104–114. https://doi.org/10.1016/j.system.2018.08.007

Xuan, W. W., & Chen, S. (2019). A synthesis of research on grammatical metaphor: Meta-data and content analysis. *WORD, 65*(4), 213–233. https://doi.org/10.1080/00437956.2019.1670911

Yang, X. (2010). *Modelling text as process: A dynamic approach to EFL classroom discourse.* Continuum.

Neil Bowen is a Research Specialist and Lecturer. He has a PhD in Language and Communication from Cardiff University, and is interested in writing research, text linguistics, and functional grammar. His most recent work can be seen in *WORD*, *Journal of Writing Research*, and *Written Communication*.

Nathan Thomas is a postgraduate researcher at the UCL Institute of Education. He is interested in the task-specific strategic behavior of English language learners, self-/other-regulation, and English medium instruction. His work has been published in journals such as *Applied Linguistics*, *Applied Linguistics Review*, *ELT Journal*, *Language Teaching*, *System*, and *TESOL Quarterly*.