Scientometrics (2025) 130:3629–3650
https://doi.org/10.1007/s11192-025-05373-4

Positive covariation or trade-off? A cross-disciplinary investigation of shell nouns and their congruent expressions in research articles

Yunyun Wang[1] · **Guangwei Hu**[2]

Received: 15 September 2024 / Accepted: 17 June 2025 / Published online: 11 July 2025

Abstract
Despite considerable attention to their textual/semantic/cognitive functions, shell nouns have rarely been examined as linguistic resources of nominalization in relation to alternative, congruent expressions (e.g., reporting clauses and evaluative clauses) to map out how such linguistic resources are used in academic writing. This cross-disciplinary study examined the use of shell nouns in 240 research articles drawn from four disciplines (physics, chemical engineering, sociology, and education) that represent two disciplinary groupings (i.e., hard/soft disciplines and pure/applied disciplines). Statistical analyses of shell nouns and their alternative, congruent constructions revealed that the two soft disciplines used shell nouns in all functional (sub) categories and their congruent constructions significantly more frequently than the two hard disciplines did. By contrast, few significant differences were found between the pure and applied disciplines in the use of shell noun constructions and their alternative expressions. Further correlational analyses pointed to a positive covariation rather than a trade-off between shell noun constructions and their congruent expressions in the construction of disciplinary knowledge. The observed patterns of use are attributable to the prevalent knowledge legitimation codes and the distinctive textual styles characteristic of hard and soft disciplines.

Keywords Shell noun · Grammatical metaphor · Research article · Cross-disciplinary difference

✉ Guangwei Hu
guangwei.hu@polyu.edu.hk

Yunyun Wang
yunyun.wang@connect.polyu.hk

[1] HHB924, Department of English and Communication, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong, China

[2] HHB920, Department of English and Communication, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong, China

Introduction

Academic writing has long been known for its compressed and information-heavy discursive style. A related feature is its extensive use of abstract nouns that results in information-dense text, when compared to everyday spoken discourse (Biber & Gray, 2011). One type of abstract nouns, namely shell nouns, has attracted considerable research attention lately. A shell noun is an unspecific general noun that serves as a "conceptual shell" (Schmid, 2000) whose meaning is made complete by a more specific stretch of co-text. As an illustration, the highlighted word in the following example is a shell noun: "there is general **consensus** that such individuals and families have opportunities to exert power and influence unshared by other segments of society." The shell noun encapsulates the ensuing proposition presented in the *that*-clause and characterizes it as a generally accepted opinion. In English, words such as *fact, idea, problem*, and *situation* are typical shell nouns. The property of "shell-nounhood" arises from the inherent semantic gaps afforded by shell nouns and the filling of these semantic gaps by more detailed information in their vicinity (Benitez-Castro, 2015; Schmid, 2000).

Despite systematic endeavors (see Flowerdew & Forest, 2015; Schmid, 2000) to probe into the nature of such semantic gaps, categorize shell nouns and their functions, and explore their lexical-semantic potentials, there has been little research that examines shell nouns in tandem with their alternative and functionally interchangeable expressions. Consequently, it remains unclear how these shell nouns interact with their functional alternatives in disciplinary writing and what this interaction reveals about their rhetorical potential. This line of inquiry holds particular significance because, compared to their alternative expressions, shell nouns constitute linguistic resources of nominalization that contribute to constructing technicality and creating chains of reasoning in scientific discourse (Halliday & Matthiessen, 2014; Martin, 2007). As such, exploring their interplay with alternative expressions is likely to yield fresh insights into the extent of nominalization in academic writing. To bridge this gap in our knowledge, the present study set out to explore whether shell nouns are deployed in a complementary, positively covarying or tangential relationship with their alternative linguistic resources in research articles sampled across a spectrum of disciplines.

Functions of shell nouns in academic writing

The entities labelled as shell nouns in this study have been captured by a range of related terms, including carrier nouns (e.g., Ivanič, 1991), general nouns (e.g., Halliday & Hasan, 1976), metadiscursive nouns (e.g., Jiang & Hyland, 2018), signaling nouns (e.g., Flowerdew & Forest, 2015), and unspecific nouns (e.g., Winter, 1992). Schmid's (2000) seminal work provides a clear conceptualization of shell nouns, delineates their linguistic properties, and forms a conceptual basis for determining whether a noun functions as a shell noun within its co-text, hence the adoption of the term in the present study. Based on his discussion, a shell noun simultaneously fulfills the textual function of linking, the semantic function of characterizing, and the cognitive function of temporary concept-formation. For example, in the sentence "Another major **challenge** is that the microscopic mechanisms of decision-making are quite unclear from our dataset", *challenge* as a shell noun forms a temporary cognition shell whose referential meaning is variable and context-dependent

(cognitive function), creates lexical cohesion by referring cataphorically to the ensuing *that*-clause (textual function), and characterizes the proposition as a difficult and demanding task (semantic function). A growing body of research has explored shell nouns' referential and characterizing functions in various academic genres to understand how they are used rhetorically to achieve textual cohesion and construe evaluation/stance (e.g., Aktas & Cortes, 2008; Charles, 2003, 2007; Gray, 2010; Gray & Cortes, 2011; Jiang, 2015).

One essential form of textual cohesion in academic writing is the deployment of shell nouns to construct cataphoric or anaphoric references to the ensuing or foregoing chunk of information (Francis, 2002; Halliday & Hasan, 1976). The cataphoric references are often realized through the N + clause (N-cl) and N + *be* + clause (N-be-cl) constructions, whereas the anaphoric references are often seen in the *th*-demonstrative + N (th-N) and *th*-demonstrative + *be* + N (th-be-N) constructions (Schmid, 2000). These cohesion patterns of shell nouns were found in several empirical studies (e.g., Aktas & Cortes, 2008; Jiang & Hyland, 2021) to vary across disciplines, over time, and between authors with different expertise. For instance, Aktas and Cortes (2008) compared the use of six common shell nouns between published authors and ESL students, finding that the published authors tended to use *fact* for cataphoric reference and prefer the th-N construction when deploying *effect*, *result* and *problem*. In a diachronic study of metadiscursive nouns in academic writing, Jiang and Hyland (2021) showed that the *this* + N construction, the most frequently employed shell noun structure, has increased markedly in frequency over the past 50 years, due to its ability to achieve rhetorical succinctness. Examining the referential functions of *problem* and *way* as shell nouns in disciplinary student writing, Benitez-Castro (2021) found that *problem* was often used in sociology with intersentential or intrasentential anaphoric references (e.g., *his Ed becomes a simple mechanical problem*), whereas engineering and business students tended to use *way* in cataphoric realizations (e.g., *an effective way of doing …*).

A growing body of research has come to recognize shell nouns as an indispensable linguistic resource for academic authors to communicate their stances through the different choices of head nouns to characterize the propositions in question. Charles' (2003, 2007) earlier investigations into stance construction indicated that shell nouns contributed significantly to stance-marking in theses by enabling the authors to create an authorial voice and construct knowledge as members of their disciplines. The ability to express stances through nominal constructions has been linked to advanced academic literacy. For instance, Jiang (2015) reported that L1 writers made more frequent use of the N + complement construction to mark factual events, discourse acts, and cognitive beliefs, in alignment with the generic conventions of argumentative essays. Furthermore, stance expression through shell nouns is likely to vary across disciplines in response to different epistemological beliefs. Such variation has been found in several studies to be particularly salient between what Becher and Trowler (2001) classified as soft and hard disciplines. Jiang and Hyland (2015) suggested that the N + complement construction is an important rhetorical device for academic authors to front-load their stance and evaluate the ensuing proposition. They found that the construction was employed more often in the humanities and social sciences than hard disciplines to construct an interpretative writing style. In a cross-disciplinary study of the noun + *that* construction, Jiang (2017a) revealed that compared to hard disciplines, soft disciplines tended to rely on this structure to project stance and voice and to build knowledge, especially when it came to the description of mental reasoning and the judgement of (un)certainty about the status of information presented. In another study examining shell noun variation between a pure and an applied hard discipline, Dong et al. (2020) found that the former favored the use of shell noun constructions to communicate fact-like

propositions, whereas the latter tended to highlight the desirability of certain activities through shell noun constructions that construed modality and volition dynamically.

The previous research on the linking and characterizing functions of shell nouns has greatly advanced our knowledge of textual cohesion as well as attitude projection and stance construction with respect to various propositions in academic writing. A limitation of previous work, however, lies in its treatment of shell nouns merely as a unique group of nouns without duly considering their nominalization function and their relationship with other functionally equivalent constructions. A shell noun does not simply link or characterize a proposition but does so by virtue of a nominalized structure. In other words, shell nouns are linguistic resources of nominalization that are functionally coterminous with alternative linguistic constructions (e.g., verbs and adjectives) in terms of textual cohesion and proposition characterization. Conceptual work on shell nouns has generally centered on the development of a working taxonomy that captures their inherent semantic distinctions and the range of shared properties (Flowerdew & Forest, 2015; Hunston et al., 1998; Jiang & Hyland, 2018; Schmid, 2000). However, previous empirical studies employing such taxonomies as analytical frameworks have tended to examine the functions of shell nouns in isolation, despite the fact that academic authors can choose among a variety of linguistic resources to communicate similar meanings. For example, a reporting clause (i.e., *argue that…*) can be used instead of the shell noun *argument* (i.e., *make the argument that…*) to characterize a proposition. Similarly, rather than using *problem* in a shell noun construction (*The problem is that…*), academic authors may opt for its adjective form in an extraposed-*it* complement clause (*It is problematic that…*) to characterize a proposition. Given such functional interchangeability, shell nouns represent only one side of the rhetorical equation, and a comprehensive understanding of their rhetorical functions in academic discourse requires a simultaneous examination of their relationship with functionally equivalent alternative expressions. To address this knowledge gap, an analytical framework is needed that can link shell nouns as nominalizations of qualities, processes, and relations to their alternative congruent constructions (Halliday & Matthiessen, 2014; Hu & Perez, 2022).

Analytical framework

To address the rhetorical and nominalizing functions of shell nouns, we propose an analytical framework whereby shell noun constructions can be examined along with and compared against their equivalent expressions in alternative grammatical forms (e.g., verbal and adjective constructions). Following the seminal work conducted by Dong et al. (2020) and Dong and Fang (2021), we anchored our analysis of shell nouns within the framework of grammatical metaphor in Systemic Functional Linguistics (SFL) (Halliday & Matthiessen, 2014). Expanding this line of inquiry, we also incorporated Jiang and Hyland's (2018) functional taxonomy of metadiscursive nouns as a key analytical framework for our investigation into shell nouns.

Shell nouns are essentially grammatical metaphors. According to SFL, when our experiential world is encoded through congruent realizations, events or processes on the semantic level are expressed by verbal groups on the lexico-grammatical level, participants by nominal groups, qualities by adjectives, circumstances by adverbial groups, and logical relations by conjunctions (Halliday & Matthiessen, 2014). When there is a stratal tension between semantics and lexico-grammar, a grammatical metaphor is created, where a

semantic meaning is realized by an incongruent or metaphorical form (Hu & Perez, 2022). For instance, a process described congruently by a verb (e.g., *solve*) is construed by a noun (e.g., *solution*), a quality typically described by an adjective (e.g., *secure*) is realized by a nominal expression (e.g., *security*), and a logical relation normally denoted by a conjunction (e.g., *therefore*) is indicated incongruently by a verb phrase (e.g., *result in*) (Halliday & Matthiessen, 2014). Nominalization is the most common form of grammatical metaphor where processes, qualities, and logical relations are expressed by means of nominal phrases. The general drift of the metaphorical process is thus a shift in the direction of abstractness, thingness, and lexical density, whereby information is packed, and experiences with the world reconstrued, into the nominal mode, which differs from the more literal, informal and direct language of daily life.

Shell nouns exhibit a similar metaphorical process at work. The functional taxonomy developed by Jiang and Hyland (2018) to classify shell nouns into attribute, entity, and relation ones resonates with the SFL account of nominalization as a metaphorical means of construing qualities, processes, and relations in academic writing. In this sense, shell nouns can be considered as nominalizations where the co-referencing relations between them and the propositions that they refer to are re-configured and transformed from various types of congruent expressions. Our analytical framework thus operates with three broad categories to capture the deployment of shell nouns as grammatical metaphors and their alternative congruent expressions to construe research processes, project attitudinal evaluations, and develop lines of logical reasoning, respectively. It is crucial to note that these mapping relations are not established on a one-to-one, word-specific basis but maintained within each functional domain, such as cognition, discourse or status. In other words, we do not expect every individual shell noun to have a congruent counterpart or vice versa. Rather, the correspondence exists at the level of a particular category. For example, when reporting cognitive ideas, academic writers can employ a range of cognition-related shell nouns (e.g., *idea, assumption,* and *hypothesis*), which as a group are functionally equivalent to a

Table 1 Functional correspondence between shell nouns and their congruent forms

Metaphorical form	Congruent form
ENTITY SHELL NOUNS	REPORTING VERBS
Event shell nouns (e.g., *finding, observation*)	Research acts (e.g., *find, observe*)
Discourse shell nouns (e.g., *conclusion, argument*)	Discourse acts (e.g., *conclude, argue*)
Cognition shell nouns (e.g., *assumption, idea*)	Cognition acts (e.g., *assume, believe*)
ATTRIBUTE SHELL NOUNS	EVALUATIVE ADJECTIVES
Quality shell nouns (e.g., *problem, importance*)	Quality adjectives (e.g., *problematic, important*)
Status shell nouns (e.g., *possibility, uncertainty*)	Status adjectives (e.g., *possible, uncertain*)
RELATION SHELL NOUNS	LINKING ADVERBIALS
Shell nouns of causality (e.g., *result, consequence, reason, factor, effect*)	Adverbials of causality (e.g., *therefore, so, because, since*)
Shell nouns of comparison (e.g., *difference, similarity, [in]consistency, discrepancy*)	Adverbials of comparison (e.g., *similarly, likewise, in contrast, whereas*)
Shell nouns of condition (e.g., *condition, circumstance, extent*)	Adverbials of condition (e.g., *if, unless*)

Note. We exclude the categories of "Text" and "Manner" originally in Jiang and Hyland (2018) either because members of the categories cannot be readily mapped onto any congruent form or because the identification of such forms would go beyond our research capacity

range of cognition verbs used for the same purpose (e.g., *think, believe,* and *assume*). The comprehensive functional correspondence between shell nouns and their congruent forms is presented in Table 1.

The first category comprises and connects entity shell noun constructions and reporting clauses, which can express various types of research process in the formulation of propositions. An entity shell noun denotes the academic author's characterization of a proposition as an event, a speech act or mental reasoning. They are metaphorical equivalents of reporting clauses where the academic author's "meta-comment" towards a reported proposition is conveyed by a reporting verb (Charles, 2006; Thompson, 1996). In this regard, Hyland's (1999) classification of reporting verbs allows for conceptual connections to be made between shell nouns denoting events, discourse, and cognition and verbs reporting research, discourse, and cognition acts. The following sentences from our dataset illustrate how an author's judgment of a proposition can be expressed either through a shell noun construction metaphorically (Example 1) or with a reporting verb congruently (Example 2).

(1) Of the literature examining women's journey on the tenure track, there was even greater support for my **argument** that the pipeline is designed to exclude nonideal workers from accessing the tenure track. (16-SA-J1-08)

(2) He **argued** that the Sun was located close to the center of the Galaxy, and that the motion of stars could be described as that of a gas in a quiescent atmosphere. (16-HP-J1-01)

The second category in the framework projects an academic author's evaluation of and attitude toward the quality and status of propositions. It offers academic authors the means of evaluating a proposition metaphorically through a shell noun or congruently through an evaluative adjective. Academic authors' evaluative stance can be further distinguished between quality and status evaluation. Quality evaluation concerns an academic author's judgement on the values (e.g., *advantage* and *limitation*) of a proposition in terms of positive and negative attributes. Status evaluation, on the other hand, encodes an academic author's assessment of a proposition in terms of deontic, dynamic, and epistemic modality (Palmer, 2001). Examples 3 and 4 illustrate how assessments of the possibility or certainty of propositions are conveyed through metaphorical and congruent expressions, respectively.

(3) The results also point to the **possibility** that for adults, the acquirement of specific knowledge and ability is less laborious and amenable than is the acquirement of general abilities. (16-SA-J5-07)

(4) It is **possible** that schools in poorer neighborhoods or with a larger proportion of students from lower SES backgrounds may systematically use different disciplinary practices. (16-SA-J1-01)

The last category describes "the circumstances and formation of actions and states of affairs" or encodes "how a writer understands the connection or relationship to information in a proposition" (Jiang & Hyland, 2018, p.517). The circumstances and connections indicated by relation/manner shell nouns can be mapped onto those construed by linking adverbials to convey various logical relations (e.g., place, time, reason, manner, purpose, condition, and degree). Thus, for example, a causal relation can be signaled incongruently (i.e., through a grammatical metaphor) by the shell noun construction *for this reason* (Example 5) or congruently by the logical conjunction *therefore* (Example 6).

(5) While the solution-diffusion model is effective for single component systems, it ignores coupling effects between penetrants that occur in binary mixtures. For this **reason**, it fails to predict the negative solute rejection which is observed in some studies. (16-HA-J5-05)

(6) The present review seeks to supplement the current state of knowledge by examining the gaming and educational design of the software used in these interventions. **Therefore**, it focuses on the notion of "serious games". (16-SA-J4-06)

As can be noted from the above examples, two different devices are useful to determine whether a given expression is metaphorical or not: Morphological derivation (e.g., *observe* → *observation* and *uncertain* → *uncertainty*) and agnation or syntactic derivation (e.g., *therefore/so* → *consequence*) (Ravelli, 2003). In this regard, though shell nouns are not necessarily morphologically derived from their corresponding verbal or adjective forms, they can be viewed as grammatically metaphorical linguistic forms, namely nominalizations of processes, qualities, and logical and coherence relations that arise from stratal remapping between semantics and grammar, and serve to reconstrue human experience into knowledge (Dong et al., 2020; Flowerdew & Forest, 2015).

Unlike the previous analytical frameworks that focus solely on the semantic and rhetorical potential of shell nouns, our analytical framework seeks to extend the existing theoretical discussion by including their nominalizing potential and their congruent constructions—functionally equivalent but common-sense and dynamic representations. To acquire a comprehensive understanding of shell noun constructions in academic writing, one needs to look into nominalized entities (i.e., shell nouns) typical of scientific writing in tandem with the concomitant use of the corresponding congruent resources characteristic of everyday language (Biber et al., 1999; Halliday & Matthiessen, 2014). By recognizing shell nouns as both semantic and nominalized structures, the analytical framework presented above is expected to generate new insights into the deployment of shell nouns in academic discourse. Given their inherent complementarity, it is reasonable to expect the distributions of shell nouns and their congruent equivalents to covary in academic writing. A positive covariation suggests that shell nouns and their congruent forms function side by side to project research processes, attitudinal evaluations or logical relations, potentially providing evidence for disciplinary variations in these projections. Conversely, a negative covariation implies a likely disciplinary preference for either shell nouns or their congruent forms, pointing to the possibility that different disciplines exhibit unique discursive patterns that mirror the register variation across levels of formality and persuasiveness (Fang & Dong, 2021).

Previous research found that diachronic changes in the incidence of shell nouns in research articles were attributable to corresponding changes in the use of their congruent expressions (e.g., reporting clauses) and the evolution of disciplinary knowledge-making practices over time (Y. Wang & Hu, 2023). However, it remains to be known whether synchronic cross-disciplinary differences exist in the use of shell nouns and their alternative expressions. As made clear earlier, disciplinarity has been found to influence the use of shell nouns (Charles, 2007; Dong et al., 2020; Jiang, 2022). Therefore, it should be examined as a key variable in a study of the possible covariation between shell nouns and their congruent constructions in academic discourse. Furthermore, previous research (e.g., Y. Wang & Hu, 2023) focused on only a subset of shell nouns and their congruent forms and, consequently, did not address the metaphorical-congruent relations in other types of shell noun (e.g., attribute shell nouns). To bridge the research gaps discussed here and above, the present study aimed to answer the following three research questions:

Table 2 Disciplinary divisions and chosen disciplines

	Hard	Soft
Pure	Physics	Sociology
Applied	Chemical engineering	Education

Table 3 Selected journals

Discipline	Selected journal
Physics	*Reviews of Modern Physics* *Physical Review X* *Physical Review Letters* *Contemporary Physics* *European Physical Journal Plus*
Chemical engineering	*Energy & Environmental Science* *Applied Catalysis B—Environmental* *Chemical Engineering Journal* *Journal of Catalysis* *Journal of Membrane Science*
Sociology	*Annual Review of Sociology* *American Sociological Review* *Sociological Methods & Research* *American Journal of Sociology* *Sociology-The Journal of The British Sociological Association*
Education	*Review of Educational Research* *Educational Psychologist* *Internet and Higher Education* *Computers & Education* *Learning and Instruction*

1 Do research articles from disciplines distinguished along common faultlines differ markedly in the frequency with which shell noun constructions are used? If so, which functional categories and subcategories of shell nouns see significant cross-disciplinary variations?
2 Are there significant cross-disciplinary differences in the incidence of congruent constructions?
3 How do cross-disciplinary differences in the use of shell nouns, if any, relate to concomitant variations in the incidence of their congruent linguistic forms?

Method

Corpus construction

To answer our research questions, we compiled a corpus of research articles from four disciplines: Physics, chemical engineering, sociology, and education. These disciplines were chosen not only because they differed from those examined in previous studies but also because they represent prototypical examples of the disciplinary groupings defined by Becher and Trowler (2001). Their selection allowed us to operationalize two common distinctions of knowledge fields adopted in the literature: Hard versus soft disciplines, and

pure versus applied disciplines (Becher & Trowler, 2001). Thus, physics and chemical engineering represent hard disciplines, whereas sociology and education typify soft disciplines. Along another faultline, physics and sociology are pure disciplines in contrast to chemical engineering and education as applied disciplines. The two disciplinary divisions examined in this study, along with the chosen disciplines serving as representative examples, are shown in Table 2.

To obtain a representative sample of research articles, we identified the top five research journals for each discipline based on the impact factors provided by the Journal Citation Reports (see Table 3). Specifically, we computed the average two-year impact factors of the journals in each sampled discipline for the 5-year period of 2016–2020 and chose the five journals with the highest average impact factors. The choice of this timeframe struck a balance between recency, sufficiency, and relevance of the data to current academic discourse. The strategy of sampling journals based on impact factors, as widely adopted in academic discourse research (e.g., Cao & Hu, 2014; Hu & Cao, 2015; Hyland & Jiang, 2018; Jiang & Hyland, 2021), aimed to capture a journal's visibility and influence within its discipline. Consequently, the sampled journals were believed to reflect the highest standard and valued practice of their respective disciplines. As probability sampling is the best strategy to obtain a representative sample (Beins & McCarthy, 2012), we employed stratified random sampling by journal, a probability sampling method frequently used when the population consists of strata (e.g., journals), to construct the sub-corpora of research articles for the four disciplines. Specifically, we accessed all the research articles published in each sampled journal within the target five-year period in the Web of Science database. Random sampling was then conducted to draw 12 research articles from the list, yielding a total of 60 research articles for each discipline and 240 for the four disciplines combined.

Abstracts, titles, footnotes, tables, figures, reference lists, and appendices in the selected research articles were removed; equations and symbols were replaced with "EQ" and "SYM", respectively. The cleaned research articles were then imported to the corpus query software, Sketch Engine, followed by the annotation and compilation procedures (Kilgarriff et al., 2014). This process yielded a corpus of 2,241,997 tokens. The meta-information of its sub-corpora is displayed in Table 4.

Data coding and analysis

Given the large sample of research articles included and the close textual analysis required, we were unable to examine all three categories of shell nouns and their congruent counterparts in our analytical framework (see Table 1). Because relation shell nouns and linking adverbials constituted particularly large categories and required analytical

Table 4 Profile of the corpus

Sub-corpus	Discipline	No. of articles	Token
1	Physics (hard/pure)	60	639,505
2	Chemical engineering (hard/applied)	60	382,105
3	Sociology (soft/pure)	60	609,303
4	Education (soft/applied)	60	611,084
Total		240	2,241,997

work beyond our existing capacity, the present study focused only on entity and attribute shell nouns and their alternative congruent constructions, namely the first two categories in Table 1. To extract relevant shell nouns from the corpus, we interrogated the Sketch Engine concordance with Corpus Query Language (CQL) for lexico-grammatical constructions containing clear slots to be occupied by shell nouns. The CQL queries, which can be found in Appendix A, were developed on the basis of the shell noun structures listed in Schmid (2000), Flowerdew and Forest (2015), and Jiang (2017b). The identified shell noun constructions were then checked manually by consulting their concordance lines to determine if the noun slots were occupied by genuine shell nouns. Subsequently, all identified shell nouns were coded with the analytical framework presented earlier (see the first column of Table 1). Similarly, the congruent forms of entity and attribute shell noun constructions were identified in the corpus using the CQL queries presented in Appendix B. Manual concordance checking was conducted to ensure their eligibility. All the verified congruent constructions were coded according to the categories listed in the second column of Table 1.

To ensure the reliability of the coding, the first author and a doctoral student with a background in applied linguistics first coded 10% randomly selected instances of shell nouns and their alternative constructions independently. The inter-coder reliability was excellent (Cohen's $k = 0.87$). Coding discrepancies, which mostly concerned borderline shell noun subcategories, were resolved through discussions that converged on a consensus. Given the established coding reliability, the first author then coded the remaining 90% of the shell nouns and their alternative expressions. The raw frequencies of the coded constructions were counted by (sub)category and normalized to incidence per 10,000 words to enable comparisons among texts of different lengths and across the four disciplines.

To address our first research question, a series of independent-samples *t*-tests were run to assess the differences between the hard and soft disciplines and between the pure and applied disciplines for the two semantic categories of shell nouns and their subcategories. To answer our second research question, parallel independent-samples *t*-tests were performed on the corresponding categories of congruent forms (i.e., reporting clauses and evaluative clauses) and their subcategories. To address the last research question, Pearson's correlational analyses were run on the normalized frequencies of shell nouns and their corresponding congruent forms to ascertain the strength and direction of the associations. The alpha was set at .05 (two-tailed) for all the statistical analyses.

Table 5 Descriptive and inferential statistics for (sub)categories of shell nouns: Hard vs. soft disciplines (frequency per 10,000 words)

Category	Hard discipline	Soft discipline	*t*	*df*	*p*	*d*
ENTITY	9.096	16.37	− 7.47	238	< 0.001	− 0.964
Event	5.661	7.19	− 2.59	238	0.010	− 0.335
Cognition	2.478	7.08	− 8.71	238	< 0.001	− 1.124
Discourse	0.957	2.09	− 4.76	238	< 0.001	− 0.615
ATTRIBUTE	4.073	7.79	− 6.06	238	< 0.001	− 0.782
Quality	1.559	2.15	− 2.00	238	0.047	− 0.258
Status	2.513	5.64	− 6.16	238	< 0.001	− 0.795

Results

Cross-disciplinary variations in the use of shell nouns

Table 5 presents the normalized frequencies of shell nouns in the different (sub)categories by the hard/soft disciplinary grouping and the results of the *t*-tests. Specifically, the two soft disciplines combined used significantly more entity and attribute shell nouns than the two hard disciplines combined did. The effect sizes, indicated by Cohen's *d* values, were large in magnitude. Significant cross-disciplinary differences were also found for all the subcategories of shell nouns, with a medium effect size for discourse shell nouns and large effect sizes for cognition and status shell nouns.

Entity shell nouns were found in our dataset to encode a variety of research activities, including factual observations (e.g., *finding, phenomenon,* and *evidence*), analytical procedures (e.g., *attempt, process,* and *step*), verbal statements (e.g., *argument, explanation,* and *conclusion*), and cognitive judgements (e.g., *hypothesis, belief,* and *awareness*). Example 7 illustrates how a specific proposition is encapsulated by a discourse shell noun *claim* to signify an assertion to be verified.

(7) There is some support in the literature for the **claim** that encouraging students to explicitly consider audience and purpose will enhance the quality of written argument. (16-SA-J2-12)

Attribute shell nouns, favored by the soft disciplines, were used to mark authors' direct intrusion into the propositions with their evaluations of qualities (e.g., *importance, advantage,* and *difficulty*) and states of affairs (e.g., *possibility, need,* and *potential*). Example 8 exemplifies how a quality shell noun *challenge* initiates an evaluation and includes the highlighted clause within its evaluative scope.

(8) Another foreseeable **challenge** is that a respondent could simply opt out of wearing a sensor, or as reported above, the group may interact virtually with someone who is not physically present and not wearing a sensor. (16-SP-J3-01)

The effect sizes reported in Table 4 indicated that the cross-disciplinary differences observed in the use of cognition and status shell nouns were markedly more pronounced, when compared with event, discourse, and quality shell nouns. The comparable effect sizes of these two categories of shell nouns might have stemmed from their functional overlap in the expression of modality. Both cognition and status shell nouns often express authors' modal judgments in terms of probability (e.g., *assumption* and *likelihood*, *belief* and *certainty*), obligation (e.g., *expectation* and *need*), and inclination (*hope* and *tendency*). The modality resonance between them, together with their respective congruent forms, is addressed by Halliday and Matthiessen (2014) through the notion of interpersonal grammatical metaphor. They are regarded as explicit or metaphorical expressions of modality where authors can overtly convey their modal judgments on the status and validity of the projected information with a subjective orientation (e.g., by means of cognition shell nouns and cognition reporting verbs) or an objective orientation (e.g., by means of status shell nouns and adjectives). In comparison with simple modal verbs and adverbs (e.g., *may, must, possibly,* and *certainly*), explicit modal expressions give an objective flavor to the propositions in question by enabling authors to distance themselves from the projected information and are thus favored in academic registers (Biber & Finegan, 1989). Examples

Table 6 Descriptive and inferential statistics for (sub)categories of shell nouns: Applied vs. pure disciplines (frequency per 10,000 words)

Category	Applied discipline	Pure discipline	*t*	*df*	*p*	*d*
ENTITY	12.40	13.06	− 0.616	238	0.539	− 0.080
Event	6.52	6.33	0.320	238	0.749	0.041
Cognition	4.48	5.08	− 0.993	238	0.322	− 0.128
Discourse	1.40	1.65	− 1.024	238	0.307	− 0.132
ATTRIBUTE	6.00	5.86	0.208	238	0.835	0.027
Quality	1.79	1.93	− 0.467	238	0.641	− 0.060
Status	4.21	3.94	0.509	238	0.611	0.066

9 and 10 illustrate the explicit projections of epistemic modality by means of cognition and status shell nouns, respectively.

(9) Our **guess** is that sociometers affect group dynamics less in natural groups because members have real tasks that must be accomplished despite their participation in the study. (16-SP-J3-01)

(10) Moreover, the inclusion of teacher reports reduces the **likelihood** that results are influenced by reporting bias, and it provides further evidence that paternal incarceration may be harmful for parental involvement in schooling. (16-SP-J2-07)

Unlike the comparisons between the soft and hard disciplines, the differences in the use of shell nouns between the applied and pure disciplines were not statistically significant (see Table 6 for the descriptive and inferential statistics). The effect sizes were also very small. The applied and pure disciplines both preferred event, cognition, and status shell nouns. Examples 11 and 12 illustrate, respectively, how event shell nouns *finding* and *observation* were used by authors from the applied and pure disciplines to encode the process of scientific inquiry in a similar way:

(11) Variation in the reactant concentration and synthesis duration did not change the crystalline phase in the UiO-66 particles. This **finding** is similar to the finding of Cavka et al. who suggested that prolonged synthesis time did not affect the crystallization structures of UiO-66. (16-HA-J5-08)

Table 7 Descriptive and inferential statistics for (sub)categories of congruent constructions: Hard vs. soft disciplines (frequency per 10,000 words)

Category	Hard discipline	Soft discipline	*t*	*df*	*p*	*d*
REPORTING CLAUSE	46.07	51.8	− 2.29	238	0.023	− 0.296
Research act	30.32	27.0	1.65	238	0.100	0.213
Cognition act	10.14	14.2	− 4.17	238	< 0.001	− 0.538
Discourse act	5.62	10.7	− 6.31	238	< 0.001	− 0.815
EVALUATIVE CLAUSE	12.47	18.67	− 5.022	238	< 0.001	− 0.648
Quality adjective	7.22	7.45	− 0.332	238	0.740	− 0.043
Status adjective	5.26	11.22	− 6.392	238	< 0.001	− 0.825

(12) In particular, the measurements revealed spectral features that survive up to timescales far outliving the coherent polaritons. This **observation** sheds light on the possible existence of an incoherent reservoir which contributes to the blueshift of the polaron-LP resonance that exists for timescales > 10 ps. (16-HP-J2-03)

Cross-disciplinary variations in the use of congruent constructions

Table 7 presents the descriptive and inferential statistics for the different (sub)categories of congruent expressions when the hard and soft disciplines were compared. The research articles from the two soft disciplines used significantly more reporting clauses introduced by reporting verbs than those from the two hard disciplines did, though the effect size was small. The former also used markedly more evaluative clauses introduced by evaluative adjectives than the latter did, with a medium effect size indicating a substantial difference. Of the subcategories of congruent constructions, the soft-discipline research articles also used cognition reporting clauses, discourse reporting clauses, and status evaluative clauses significantly more frequently than the hard-discipline research articles did. For all these comparisons, the effect sizes were either medium (cognition reporting clauses) or large (discourse reporting clauses and status evaluative clauses).

As in the case of shell nouns, comparisons between the applied and pure disciplines yielded non-significant results except for reporting clauses as a main category and research reporting clauses as a subcategory. As shown in Table 8, reporting clauses occurred significantly more often in the applied-discipline research articles than in the pure-discipline ones, with a medium effect size. This difference was clearly due to the difference in research reporting clauses, which were markedly more frequent in the applied disciplines than the pure disciplines, with an effect size approaching the threshold for a large effect (i.e., $d = 0.8$) (Cohen, 2013).

These results complemented the cross-disciplinary differences observed in the use of shell nouns. Compared with the hard disciplines, the soft disciplines not only showed a stronger preference for cognition and discourse shell nouns but also demonstrated a greater tendency to use their congruent constructions (i.e., cognition and discourse reporting clauses). Examples 13 and 14 illustrate how academic authors introduced voices and viewpoints of various sources into the discourse through cognition and discourse reporting clauses.

Table 8 Descriptive and inferential statistics for (sub)categories of congruent constructions: Applied vs. pure disciplines (frequency per 10,000 words)

Category	Applied discipline	Pure discipline	*t*	*df*	*p*	*d*
REPORTING CLAUSE	54.20	43.67	4.335	238	< 0.001	0.560
Research act	33.86	23.45	5.454	238	< 0.001	0.704
Cognition act	12.10	12.19	− 0.097	238	0.922	− 0.013
Discourse act	8.24	8.03	0.238	238	0.812	0.031
EVALUATIVE CLAUSE	14.99	16.15	− 0.893	238	0.373	− 0.115
Quality evaluation	7.36	7.31	0.070	238	0.944	0.009
Status evaluation	7.64	8.84	− 1.197	238	0.233	− 0.155

(13) After all, structuralists **believe** that "impulses and emotions explain nothing," as they are "always the results, either of the power of the body or the impotence of the mind". (16-SP-J2-05)

(14) Besides, we need to **reiterate** that the association meta-analyzed in this study is correlational and therefore no causality can be deduced. (16-SA-J1-01)

In a similar vein, the variation in the use of status shell nouns observed between the hard and soft disciplines was paralleled by a similar difference in the use of status adjectives. As pointed out earlier, status shell nouns and adjectives together constitute what Halliday and Matthiessen (2014) refer to as explicit objective expressions of modality. They are distinguished from other types of modality by their greater potential of depersonalizing and construing formality. Status shell nouns and adjectives often appear in constructions that delay or extrapose the notional subject with a dummy pronoun *it* due to structural considerations (Rodman, 1991). Such "anticipatory *it*" constructions offer writers additional means of concealing the source of an attitude with an impersonal subject *it* and increasing the facticity of a statement (Herriman, 2000; Hyland & Tse, 2005), in contrast to explicit subjective modal expressions (e.g., *I think* and *I believe*) (Halliday & Matthiessen, 2014). Compared with implicit objective modal expressions (e.g., *possibly* and *probably*), they allow academic authors to extrapose their modal judgement at the beginning of the clause and present the proposition as an explicit and non-negotiable one (Collins, 1994; Hyland & Tse, 2005). Example 15 illustrates how a soft-discipline author explicitly marked epistemic and deontic modality by status adjectives *true* and *necessary,* respectively, while keeping the original knowledge propositions intact.

(15) Although it is largely **true** that women with children suffer from the pipeline, it is **necessary** to explain why academic men have benefitted from being married and having children. (16-SA-J1-08)

Covariations between shell nouns and their congruent constructions

To determine how the use of shell nouns related to their alternative expressions, Pearson's correlational analyses were performed on the normalized frequencies of shell nouns and their corresponding congruent forms. Table 9 summarizes the correlational coefficients and the levels of statistical significance. Entity shell nouns as a group showed a significant but weak positive correlation with reporting clauses. Thus, the more frequently entity shell nouns were used, the more frequently reporting clauses were deployed. Stronger positive correlations were observed for the subcategories of cognition shell nouns/reporting clauses

Table 9 Correlations between frequencies of shell nouns and their congruent constructions

Shell noun	Congruent construction	*r*	*p*
ENTITY	REPORTING CLAUSE	0.240	< 0.001
Event	Research act	0.111	0.085
Cognition	Cognition act	0.376	< 0.001
Discourse	Discourse act	0.252	< 0.001
ATTRIBUTE	EVALUATIVE CLAUSE	0.335	< 0.001
Quality	Quality adjective	0.100	0.123
Status	Status adjective	0.360	< 0.001

and discourse shell nouns/reporting clauses. In both cases, the *r* values either exceeded the threshold for a medium effect size (i.e., $r = 0.30$) or approached it. Similarly, the occurrences of attribute shell nouns and evaluative clauses were significantly and positively correlated, with a medium effect size. Of their two subcategories, a significant positive correlation was found between the frequencies of status shell nouns and adjectives, with a medium effect size. These results evidenced a positive covariation, rather than a trade-off, between shell noun constructions and their congruent forms in the 240 research articles from the four sampled disciplines examined in the present study.

Discussion

The results reported in the preceding section indicated that irrespective of metaphorical or congruent forms, the soft-discipline research articles employed cognition, discourse, and status constructions much more frequently than the hard-discipline ones did. These observed patterns of use add to our knowledge that there are cross-disciplinary variations in the discursive deployment of lexico-grammatical resources with distinctive semantic potential to characterize propositions. As revealed in this study, the soft disciplines had a markedly stronger tendency than the hard disciplines to project verbal statements, cognitive beliefs, and explicit objective expressions of modality. These tendencies were attributable to the knowledge-knower structures dominating the disciplines concerned.

Expanding Bernstein's (1999) conceptualizations of knowledge structures by incorporating a sociological perspective, Maton (2000, 2014) posits that disciplines adopt different knowledge legitimation codes, which are prevalent configurations of dominant knowledge and knower structures. Hard disciplines typically subscribe to a knowledge code which combines a horizontal knower structure with a hierarchical knowledge structure. In these disciplines, knowledge is constructed by hierarchically integrating new knowledge into ever general theories and by verifying new propositions against well-established scientific principles and widely-accepted disciplinary procedures, independent of the personal attributes and social backgrounds of researchers as knowers. By contrast, soft disciplines usually operate with a knower code that integrates a hierarchical knower structure with a horizontal knowledge structure. New knowledge in such disciplines is legitimatized by appeals to academic authors' expertise, personal voice, authority and experience, and is represented in distinctive and specialized languages, which "make different and often opposing assumptions, with each language having its own criteria for legitimate texts, what counts as evidence, and what counts as legitimate questions, or a legitimate problematic" (Bernstein, 1999, p. 163).

The different knowledge legitimation codes prevailing in soft and hard disciplines could explain why research articles from the soft disciplines used more cognition and discourse shell nouns/reporting clauses than their counterparts from the hard disciplines did. Given the dominance of a knower code, soft disciplines are by nature more dialogic and tend to project various voices in the process of argumentation and persuasion. Cognition and discourse shell nouns and their congruent constructions serve this purpose well by allowing academic authors to give voice to individuals and project their arguments and viewpoints. Specifically, cognition shell nouns/reporting clauses communicate conceptualizations, mental reasoning, and logical thinking that can legitimate knowledge through knowers' personal dispositions and authority. Likewise, discourse shell nouns/reporting clauses can convey verbal arguments emanating from different viewpoints to generate new knowledge.

In contrast, knowledge legitimation in hard disciplines depends more on empirical evidence obtained through disciplinarily accepted principles and procedures and less on the projection of personal voices and viewpoints of individual researchers, hence a reduced need for cognition and discourse shell nouns or their congruent constructions.

The knowledge legitimation codes dominant in hard and soft disciplines could also provide a plausible explanation of the cross-disciplinary differences observed in the use of status shell nouns/evaluative clauses. As pointed out earlier, modality expressions can be distinguished in terms of explicitness: Implicit (e.g., *may* and *will*) and explicit (e.g., *I think* and *it is possible*) (Halliday & Matthiessen, 2014). Cognition and status expressions (e.g., *assume* and *possible*) function similarly to constitute explicit expressions of modality. As reported earlier, the soft-discipline research articles deployed explicit modality more often than their counterparts from the hard disciplines did. Such expressions included cognition shell nouns (e.g., *the conjecture that*), cognition reporting clauses (e.g., *assume that*), status shell nouns (e.g., *the possibility that*), and status evaluative clauses (e.g., *possible that*). In contrast to implicit modality expressions where modality is integrated into a proposition, an explicit modality expression extraposes the modality from the modalized proposition; as a result, the extraposed element is linguistically marked and receives much emphasis (Rodman, 1991). This helps to amplify or boost an academic author's personal voice and assert his/her authority. Thus, explicit modality expressions, which overtly index academic authors' professional judgements on the probability/plausibility/necessity of propositions, cater well to the knowledge-making practices of soft disciplines, where "knowledge … is personal, subjectively meaningful, holistic, value-laden, subject to contextual dynamics, and contingent on argumentation rather than universally shared criteria for verification" (Hu, 2018, p.565). By contrast, because of their strong commitment to disciplinarily established scientific principles, criteria, instrumentation, and procedures, hard disciplines tend to distance from subjective judgements and personal interpretations, hence a lesser need for explicit modality expressions. The varying commitment to authorial intervention observed in this study resonated with Hu and Cao's (2015) finding that linguistic resources for representing strong voices (e.g., boosters) reinforce authorial commitment to knowledge claims and contribute to the projection of a privileged knower for effective persuasion in soft disciplines.

Finally, the dominant knowledge legitimation codes that disciplines operate with could also explain why more pronounced cross-disciplinary differences were observed in the incidence of shell nouns than that of their congruent forms. Despite the positive covariations found between shell nouns and their congruent constructions, the soft disciplines used significantly more event and quality shell nouns than the hard disciplines did, but no such differences were found for their congruent forms. Furthermore, the effect sizes for the cross-disciplinary variations in the occurrences of shell noun constructions were mostly larger than those for their congruent forms. Thus, shell nouns distinguished the soft-discipline research articles from the hard-discipline ones more clearly than the congruent constructions did. These tendencies would be attributable to the nature of disciplinary knowledge. As Martin (1993a, p. 241) points out, hard disciplines often aim to construct taxonomic knowledge by distilling and reclassifying experience via technical terms, whereas soft disciplines pursue "interpretations coded in the discourse patterns of the texts", which "interpret the world from a nominal point of view." As nominalized entities, shell nouns are often welded to a larger grammatical environment and are amenable to further meaning expansion, thus enabling authors to create chains of reasoning and construct lines of argument across discourse (Halliday & Matthiessen, 2014). In this way, shell noun constructions scaffold text by "realizing events [and qualities] as participants so that

logical connections can be realized inside the clause" (Martin, 1993b, p. 292), and layers of information structure can be constructed textually.

Contrary to our expectations, this study revealed few significant differences in the use of shell nouns and their congruent constructions between the applied and pure disciplines. In contrast, differences in the use of these constructions were more pronounced between the hard and soft disciplines. These patterns of use might have stemmed from the epistemological distance between the disciplinary groupings. While pure disciplines are primarily concerned with theoretical understanding, applied disciplines value the practical application of knowledge to address real-world problems (Nesi & Gardner, 2006). Thus, an applied discipline is related to and dependent on a corresponding pure discipline. This epistemological distance is much smaller than the great divide between hard and soft disciplines, which are not only independent of each other but also adopt diametrically contrasting epistemologies and knowledge-making practices (Becher & Trowler, 2001). Linguistic practices, shaped by disciplinary epistemologies (Hu, 2018), are likely to reflect the magnitude of epistemological distance.

Importantly, this study revealed positive covariations, rather than trade-offs, between the incidence of shell noun constructions and that of their congruent forms in disciplinary writing. The positive covariations existed for both entity shell nouns/reporting clauses and attribute shell nouns/evaluative clauses, and were more pronounced for the cognition and status subcategories. These patterns suggest that the frequent projection of arguments/statements, cognitive beliefs, and explicit modalities is consistent across the disciplines and the grammatical realizations (i.e., shell noun constructions and their verbal or adjective counterparts). These findings could be plausibly explained by the increasing pressures on academics to promote their research, through various linguistic and rhetorical resources, and secure publication in an attention economy (Hyland, 2023). Cognition shell nouns/reporting clauses, discourse shell nouns/reporting clauses, and status shell nouns/adjectives are linguistic and rhetorical resources that can be strategically drawn on to project distinct authorial voices (e.g., *Our guess is that…*; *we need to reiterate that…*) and construct value arguments (e.g., *This observation sheds light on…*; *there was even greater support for my argument…)* to highlight the significance and contribution of one's own research (Hu & Bonsu, 2025). Such rhetorical constructions help authors position their work in disciplinary contexts and demonstrate its potential impact and value in response to the increasing competition and marketisation of scholarly publishing (Q. Wang & Hu, 2024).

Conclusion

This study developed an analytical framework in which shell noun constructions can be examined along with and compared against their congruent expressions in alternative grammatical forms (i.e., reporting and evaluative clauses). Drawing on the framework, it examined how shell nouns and their congruent constructions were used in the research articles of four disciplines that represent two common distinctions between knowledge fields (i.e., hard/soft disciplines and pure/applied disciplines). Statistical analyses of the identified shell nouns and their alternative, congruent constructions revealed that the two soft disciplines used shell nouns in all functional (sub)categories and their congruent constructions (i.e., reporting and evaluative clauses) significantly more frequently than the two hard disciplines did. By contrast, few significant differences were found in the use of shell nouns and their alternative expressions between the pure and applied disciplines. Further

correlational analyses pointed to positive covariations rather than trade-offs between shell noun constructions and their congruent expressions in the construction of disciplinary knowledge. These findings, when taken together, revealed the soft-discipline research articles' greater preference for the mental and verbal projection of authorial voices, explicit modality expressions, and shell noun constructions, when compared with their hard-discipline counterparts. The observed patterns of use are attributable to the prevalent knowledge legitimation codes, the distinctive textual styles favored by hard and soft disciplines, and the growing pressures on academics to promote their research.

Several pedagogical and theoretical implications can be derived from these findings. First, as this study found that academic authors may draw on both shell noun constructions and their congruent expressions to fulfil similar functions of characterizing and evaluating propositions, it is important that EAP/ESP instructors bring their functional interrelationships and differing rhetorical effects to students' attention. In addition to recognizing the functional connections between shell noun constructions and their congruent forms, students should also be made aware that choosing one type of linguistic resource over the other type may lead to different ways of organizing propositions and constructing logical relationships in academic writing. Pedagogically, instructors can present authentic texts containing examples of shell noun constructions and their congruent expressions to guide students in noticing and understanding their functional interchangeability and differing rhetorical effects. Learning activities can be designed to make students practice using both types of resource appropriately, reinforcing the functional connection between these linguistic choices and fostering an appreciation of their distinct rhetorical effects. Second, the cross-disciplinary comparisons undertaken in this study uncovered systematic variations between the soft and hard disciplines in the use of various linguistic resources, such as mental and verbal projection clauses and explicit modality expressions. It is thus crucial to teach these lexico-grammatical resources in a discipline-sensitive manner to facilitate the effective socialization of students into the discursive and knowledge-making practices of their chosen disciplines. In practice, classroom activities that involve comparing texts from hard and soft disciplines could be beneficial for students' understanding of the unique linguistic conventions valued in their disciplinary communities. Third, the study revealed the distinctive modal orientations encoded by cognition and status shell nouns as well as their congruent forms. Together, they constitute resources for explicitly signaling modality; separately, they represent subjective and objective modality. The incorporation of these modal resources into EAP/ESP instruction can contribute to students' awareness and mastery of them to construct and disseminate disciplinary knowledge competently. Finally, the present study has advanced the theoretical discussion on shell nouns by highlighting their nominalizing potential and introducing a novel analytical framework. This framework allows for the examination and comparison of shell nouns alongside their congruent expressions to develop a fuller understanding.

This study has several limitations that should be addressed in future research. First, while it aimed to provide a comprehensive conceptualization of the functional interrelationships between shell nouns and their congruent expressions, there is room for improvement of the analytical framework to capture the full range of linguistic choices across various dimensions and categories. Consequently, further studies should seek to test and refine this framework in a broader context to advance our theoretical understanding of the functional interdependence between shell nouns and their congruent expressions. Second, similar to most previous research on shell nouns, this study was quantitatively oriented and limited the analysis to a set of constructions traditionally associated with shell nouns. The reliance on typical shell noun constructions for the automatic retrieval of target features means that the conclusions should be interpreted with an understanding that not all instances of shell nouns in the corpus might have been identified. Future research can combine automatic searches with manual perusal to include less common shell nouns. Finally, the findings were interpreted from our etic perspective as text analysts. Further studies can draw on interviews with disciplinary informants to take an emic perspective on the factors shaping the use of shell nouns in academic writing.

Appendix A

CQL queries for identifying shell noun constructions

Construction	Variant	CQL query	Query explanation
SN + *complement clause* (N-cl)	N-*that* clause	[tag = "N.*"][tag = "IN/that"]	All nouns followed by a "that" as a subordinator
	N-*to* infinitive	[tag = "N.*"][tag = "TO"]	All nouns followed by *to*-infinitive
	N-*of*/*for*	[tag = "N.*"][word = "of|for" & tag = "IN"][tag = "DT"]?[tag = "N.*"]?[tag = "VVG"]	All nouns followed by a word "*of*" or "*for*" as a preposition and a gerund
SN + *be* + *nominalization* (N-be-cl)	N-*be-that* N-*be-to* N-*be-wh*	[tag = "N.*"][tag = "MD"]?[tag = "VH.*"]?[tag = "VB.*"][tag = "RB.?"]?[tag = "IN/that|TO|W.*"]	Noun + (modal) + (have) + be + (adverb) + that/to/wh-
Demonstratives + *SN* (th-N)	*this*-N *these*-N *such*-N	[lemma = "this|these|such"][tag = "RB.?"]?[tag = "J.*"]?[tag = "N.*"]	this/these/such + (adverb) + (adjective) + Noun
Demonstratives + *be* + *SN* (th-be-N)	*this-be*-N *these-be*-N *which-be*-N	[lemma = "this|these|which"][tag = "MD"]?[tag = "VH.*"]?[tag = "VB.*"][tag = "DT"]?[tag = "RB.?"]?[tag = "J.*"]?[tag = "N.*"]	this/these/which + (modal) + (have) + be + (the/a/an) + (adverb) + (adjective) + noun

Appendix B

CQL queries for identifying the congruent expressions

Construction	Variant	CQL query	Query explanation
Cataphoric reporting clause	V-*that/to/wh*	[tag = "V.*"][tag = "IN/that|TO|W.*"]	Verb + that/to/wh-
Anaphoric reporting clause	*as/which*-(*be*)-V*ed*	[word = "as|which"][tag = "MD"]?[tag = "VH.*"]?[tag = "VB.*"]?[tag = "RB.?"]?[tag = "VVN"]	as/which + (modal) + (have) + (be) + (adverb) + verb participle
Cataphoric evaluating clause	ADJ-*that/to/wh*	[tag = "J.*"][tag = "IN/that|TO|W.*"]	Adjective + that/to/wh-
Anaphoric evaluating clause	*This/These/which-be*-ADJ	[word = "This|These|which"][tag = "MD"]?[tag = "VH.*"]?[tag = "VB.*"][tag = "RB.?"]?[tag = "J.*"]	This/These/which + (modal) + (have) + (be) + (adverb) + adjective

Funding Open access funding provided by The Hong Kong Polytechnic University. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors report there are no competing interests to declare.

References

Aktas, R. N., & Cortes, V. (2008). Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes, 7*, 3–14. https://doi.org/10.1016/j.jeap.2008.02.002

Becher, T., Trowler, P. (2001). *Academic tribes and territories* (2nd). The Society for Research into Higher Education & Open University Press.

Beins, B. C., & McCarthy, M. A. (2012). *Research methods and statistics*. Pearson.

Benitez-Castro, M.-A. (2015). Coming to grips with shell-nounhood: A critical review of insights into the meaning, function and form of shell-noun phrases. *Australian Journal of Linguistics, 35*, 168–194. https://doi.org/10.1080/07268602.2015.1005001

Benitez-Castro, M.-A. (2021). Shell-noun use in disciplinary student writing: A multifaceted analysis of problem and way in third-year undergraduate writing across three disciplines. *English for Specific Purposes, 61*, 132–149. https://doi.org/10.1016/j.esp.2020.10.004

Bernstein, B. (1999). Vertical and horizontal discourse: An essay. *British Journal of Sociology of Education, 20*, 157–173. https://doi.org/10.1080/01425699995380
Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text, 9*, 93–124. https://doi.org/10.1515/text.1.1989.9.1.93
Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics, 15*, 223–250. https://doi.org/10.1017/s1360674311000025
Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.
Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics, 66*, 15–31. https://doi.org/10.1016/j.pragma.2014.02.007
Charles, M. (2003). 'This mystery…': A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes, 2*, 313–326. https://doi.org/10.1016/S1475-1585(03)00048-1
Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*, 310–331. https://doi.org/10.1016/j.esp.2005.05.003
Charles, M. (2007). Argument or evidence? Disciplinary variation in the use of the Noun that pattern in stance construction. *English for Specific Purposes, 26*, 203–218. https://doi.org/10.1016/j.esp.2006.08.004
Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
Collins, P. (1994). Extraposition in English. *Functions of Language, 1*, 7–24. https://doi.org/10.1075/fol.1.1.03col
Dong, M., & Fang, A. C. (2021). Shell nouns as grammatical metaphor revealing disparate construals: Investigating the differences between British English and China English based on a comparable corpus. *Corpus Linguistics and Linguistic Theory, 17*, 743–779. https://doi.org/10.1515/cllt-2018-0047
Dong, M., Fang, A. C., & Qiu, X. (2020). Shell nouns as grammatical metaphor in knowledge construal: Variation across science and engineering discourse. *Lingua, 248*, 1–25. https://doi.org/10.1016/j.lingua.2020.102946
Fang, A. C., & Dong, M. (2021). Shell nouns as a register-specific device in speech and writing. *International Journal of Corpus Linguistics, 26*, 219–247. https://doi.org/10.1075/ijcl.19059.fan
Flowerdew, J., & Forest, R. W. (2015). *Signalling nouns in English: A corpus-based discourse approach*. Cambridge University Press.
Francis, G. (2002). Labelling discourse: An aspect of nominal-group lexical cohesion. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 83–101). Routledge.
Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices: A focus on sentence-initial *this/these* in academic prose. *Journal of English for Academic Purposes, 9*, 167–183. https://doi.org/10.1016/j.jeap.2009.11.003
Gray, B., & Cortes, V. (2011). Perception vs. evidence An analysis of this and these in academic prose. *Journal for Specific Purposes, 30*(1), 31–43. https://doi.org/10.1016/j.esp.2010.06.004
Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar* (4th ed.). Routledge.
Herriman, J. (2000). Extraposition in English: A study of the interaction between the matrix predicate and the type of extraposed clause. *English Studies, 81*, 582–599. https://doi.org/10.1076/enst.81.6.582.9180
Hu, G. (2018). Disciplinary knowledge making and academic discourse. In Y. Leung, J. Katchen, S. Hwang, & Y. Chen (Eds.), *Reconceptualizing English language teaching and learning in the 21st century* (pp. 553–573). Crane Publishing.
Hu, G., & Bonsu, E. M. (2025). A cross-disciplinary study of value arguments in doctoral theses submitted to universities in Hong Kong. *English for Specific Purposes, 79*, 1–16. https://doi.org/10.1016/j.esp.2025.02.002
Hu, G., & Cao, F. (2015). Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for Specific Purposes, 39*, 12–25. https://doi.org/10.1016/j.esp.2015.03.002
Hu, G., & Perez, M. R. (2022). Effects of explicit instruction in nominalisation on ESL learners' academic writing. *Iberica, 43*, 27–54. https://doi.org/10.17398/2340-2784.43.27
Hunston, S., Francis, G., & Manning, E. (1998). *Collins Cobuild grammar patterns 2: Nouns and adjectives*. Harper Collins Publisher.
Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*, 501–530. https://doi.org/10.1093/applin/20.3.341
Hyland, K. (2023). Academic publishing and the attention economy. *Journal of English for Academic Purposes*. https://doi.org/10.1016/j.jeap.2023.101253
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
Hyland, K., & Tse, P. (2005). Evaluative that constructions: Signalling stance in research abstracts. *Functions of Language, 12*, 39–63. https://doi.org/10.1075/fol.12.1.03hyl

Ivanič, R. (1991). Nouns in search of a context: A study of nouns with both open- and closed-system characteristics. *International Review of Applied Linguistics in Language Teaching, 29*, 93–114. https://doi.org/10.1515/iral.1991.29.2.93

Jiang, F. (2015). Nominal stance construction in L1 and L2 students' writing. *Journal of English for Academic Purposes, 20*, 90–102. https://doi.org/10.1016/j.jeap.2015.07.002

Jiang, F. (2017a). Stance and voice in academic writing. *International Journal of Corpus Linguistics, 22*, 85–106. https://doi.org/10.1075/ijcl.22.1.04jia

Jiang, F. (2017b). *Metadiscursive nouns in disciplinary writing [Doctoral thesis, The University of Hong Kong]*. Hong Kong SAR.

Jiang, F. (2022). *Metadiscursive nouns: Interaction and persuasion in disciplinary writing*. Routledge.

Jiang, F., & Hyland, K. (2015). 'The fact that': Stance nouns in disciplinary writing. *Discourse Studies, 17*, 529–550. https://doi.org/10.1177/1461445615590719

Jiang, F., & Hyland, K. (2018). Nouns and academic interactions: A neglected feature of metadiscourse. *Applied Linguistics, 39*, 508–531. https://doi.org/10.1093/applin/amw023

Jiang, F., & Hyland, K. (2021). 'The goal of this analysis …': Changing patterns of metadiscursive nouns in disciplinary writing. *Lingua, 252*, 1–13. https://doi.org/10.1016/j.lingua.2020.103017

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*, 7–36. https://doi.org/10.1007/s40607-014-0009-9

Martin, J. R. (1993a). Techicality and abstraction: Language for the creation of specialized texts. In A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 223–241). The Falmer Press.

Martin, J. R. (1993b). Life as a noun: Arresting the universe in science and humanities. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 242–293). The Falmer Press.

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.

Maton, K. (2000). Languages of legitimation: The structuring significance for intellectual fields of strategic knowledge claims. *British Journal of Sociology of Education, 21*, 147–167. https://doi.org/10.1080/713655351

Maton, K. (2014). *Knowledge and knowers: Towards a realist sociology of education*. Routledge.

Nesi, H., & Gardner, S. (2006). Variation in disciplinary culture: University tutors' views on assessed writing tasks. In R. Kiely, P. Rea-Dickens, H. Woodfield, & G. Clibbon (Eds.), *Language, culture and identity in applied linguistics* (pp. 99–117). Equinox.

Palmer, F. (2001). *Mood and modality* (2nd ed.). Cambridge University Press.

Ravelli, L. J. (2003). Renewal of connection: Integrating theory and practice in an understanding of grammatical metaphor. In A.-M. Simon-Vandenbergen & M. Taverniers (Eds.), *Grammatical metaphor: Views from systemic functional linguistics* (pp. 37–64). John Benjamins.

Rodman, L. (1991). Anticipatory it in scientific discourse. *Journal of Technical Writing and Communication, 21*, 17–27. https://doi.org/10.2190/pwj6-ag95-mcqd-rg1w

Schmid, H.-J. (2000). *English abstract nouns as conceptual shells: From corpus to cognition*. Mouton de Gruyter.

Thompson, G. (1996). Voices in the text: Discourse perspectives on language reports. *Applied Linguistics, 17*, 501–530. https://doi.org/10.1093/applin/17.4.501

Wang, Y., & Hu, G. (2023). Shell noun phrases in scientific writing: A diachronic corpus-based study on research articles in chemical engineering. *English for Specific Purposes, 71*, 178–190. https://doi.org/10.1016/j.esp.2023.05.001

Wang, Q., & Hu, G. (2024). Expressions of confusion in research articles: A diachronic cross-disciplinary investigation. *Scientometrics, 129*(1), 445–471. https://doi.org/10.1007/s11192-023-04914-z

Winter, E. O. (1992). The notion of unspecific versus specific as one way of analysing the information of a fund-raising letter. In W. C. Mann & S. A. Thompson (Eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text* (pp. 131–170). John Benjamins.