



## Packaging Authority: A Comparative Corpus-Based Study of Passive Voice, Cleft Constructions, and Topicalization in the Conclusions of W-Category and Y-Category Linguistics and Literature Journals

Ms. Aneeqa Faraz<sup>1</sup>, Dr. Ayesha Asghar Gill<sup>2</sup>

1. University of Agriculture, Faisalabad, [aneeqafaraz2711@gmail.com](mailto:aneeqafaraz2711@gmail.com)
2. Assistant Professor, University of Agriculture, Faisalabad, [ayesha.asghar@uaf.edu.pk](mailto:ayesha.asghar@uaf.edu.pk)

### Abstract

Non-canonical structures—passive voice, cleft constructions, and topicalization—allow academic writers to manage information flow and construct disciplinary authority. While prior corpus studies have examined these structures within single genres, none have compared their deployment across institutionally stratified publication tiers, and the conclusion section remains underexamined as a site of information packaging. This study compares how these structures function in the conclusions of Higher Education Commission (HEC) Pakistan-recognized W-category (high-tier) and Y-category (mid-tier) linguistics and literature journals. A purposive corpus of 80 conclusions (N = 29,559 words) was tagged and analyzed in AntConc 3.5.9, with frequencies normalized per 1,000 words and compared via chi-square tests; qualitative analysis drew on Huddleston and Pullum's Information Packaging model, Prince's Given–New taxonomy, and Halliday's Theme–Rheme framework. W-category writers favored epistemically calibrated modal passives (26.9% vs. 18.7%),  $\chi^2(1, N = 229) = 5.41, p = .020$ , while Y-category writers favored unmodalized assertion. No genuine topicalization occurred in either corpus; both relied on thematic fronting, deployed more richly in W-category texts. Tier-based distinction lies in strategic precision rather than structural novelty.

**Keywords:** Non-canonical syntax; information packaging; passive voice; cleft constructions; topicalization; Theme–Rheme; corpus linguistics

### Introduction

Academic writing is not merely the transmission of information; it is a carefully structured form of discourse through which writers organize knowledge, establish authority, and guide readers toward particular interpretations (Lambrecht, 1994; Halliday & Matthiessen, 2014). A central mechanism in this organization is the management of information flow within and across sentences, achieved largely through syntactic choice. Among the available resources, non-canonical sentence structures, particularly passive voice, cleft constructions, and topicalization, allow writers to depart from the default Subject–Verb–Object order in order to foreground or background participants, manage given and new information, and maintain thematic continuity across stretches of discourse. Traditional grammars have often treated canonical word order as the neutral default, but discourse-oriented research demonstrates that departures from this order are purposeful communicative strategies rather than arbitrary stylistic variation (Lambrecht,

1994; Halliday & Matthiessen, 2014). Passive constructions allow writers to background an agent and foreground a process or result; cleft constructions isolate a constituent for contrastive emphasis; and topicalization (preposing) links new material to prior discourse by fronting an object, adjunct, or complement. In the high-stakes register of academic publishing, where clarity, objectivity, and disciplinary authority are paramount, the strategic deployment of these structures plausibly distinguishes more sophisticated scholarly prose from less rhetorically controlled writing. Despite sustained interest in information structure and corpus-based discourse analysis, comparatively little research has examined how non-canonical structures function across different tiers of academic publication. Existing studies have generally examined isolated constructions or general measures of syntactic complexity within single, relatively homogeneous corpora (Banks, 1995; Zhou & Chen, 2021), without asking whether information-packaging strategies systematically differ between higher- and lower-prestige publication venues. This gap is particularly salient in the Pakistani academic context, where the Higher Education Commission (HEC) Journal Recognition System (HJRS) formally stratifies journals into categories, including the high-tier W-category and the mid-tier Y-category, yet no corpus-based study has used this institutional distinction as a basis for comparing syntactic choice. A second, related gap concerns genre placement: prior corpus studies of non-canonical structures have focused on full articles, abstracts, or results/discussion sections (Banks, 1995; Zhou & Chen, 2021), leaving the conclusion, the section in which writers must synthesize findings and assert final rhetorical authority, comparatively unexamined.

This study addresses these gaps through a comparative corpus-based investigation of conclusion sections drawn from W-category and Y-category linguistics and literature journals recognized by the HEC. Specifically, the study aims to:

- (1) compare the frequency and distribution of passive voice, cleft constructions, and topicalization across W-category and Y-category conclusions;
- (2) examine how authors in each tier deploy these structures to manage information packaging and Theme–Rheme progression; and
- (3) evaluate the degree to which these structures contribute to textual cohesion and discourse quality across the two tiers.

These objectives are pursued through three research questions: (RQ1) In what ways do passive voice, cleft constructions, and topicalization vary in frequency and distribution across W-category and Y-category conclusions? (RQ2) How do authors in each tier deploy these structures to manage information packaging and Theme–Rheme progression? (RQ3) To what degree do non-canonical structures shape textual cohesion and discourse quality differently across the two tiers? The study adopts a corpus-based comparative design combining quantitative and qualitative analysis. A purposively sampled corpus of 80 conclusion sections (40 from W-category and 40 from Y-category journals in linguistics and literature, published 2019–2026;  $N = 29,559$  words) was compiled, part-of-speech tagged using the Penn Treebank tagset, and analyzed in AntConc 3.5.9 through regular-expression concordance queries and manual verification. Raw frequencies were normalized per 1,000 words to control for the unequal sub-corpus sizes (17,287 words in the W-category corpus, hereafter WCC; 12,272 words in the Y-category corpus, hereafter YCC), and Pearson chi-square tests of independence assessed the statistical significance of tier-based distributional differences. Qualitative analysis interpreted representative concordance lines through three complementary theoretical lenses: Huddleston and Pullum's (2002, 2005) model of Information Packaging, Prince's (1978, 1981) Given–New taxonomy of information status, and Halliday's Systemic Functional Linguistics (Theme–Rheme organization). Full methodological detail is provided in Section 3. The remainder of the article is

organized as follows. Section 2 reviews relevant literature and identifies the research gap. Section 3 details the corpus, data collection, and analytical procedures. Section 4 presents the quantitative and qualitative findings together with discussion. Section 5 concludes, outlines the study's contributions and limitations, and proposes directions for future research.

### **Literature Review and Theoretical Framework**

Information structure emerged as a distinct theoretical domain in the mid-to-late twentieth century, beginning with Chafe's (1976) argument that speakers organize utterances according to assumptions about what is already known versus cognitively active in discourse. Lambrecht (1994) later integrated syntactic, pragmatic, and prosodic dimensions into a unified framework distinguishing topic (what an utterance is about) from focus (its new or contrastive contribution). Huddleston and Pullum (2002, 2005) provided the structural foundation adopted in the present study, distinguishing canonical clauses, which follow unmarked Subject–Verb–Object order, from non-canonical clauses, in which constituents are reordered to control information flow while preserving truth-conditional equivalence. Passive voice promotes a patient into subject position while backgrounding the agent; cleft constructions split a single proposition into two clausal components to create informational focus; and topicalization fronts a constituent that would otherwise occur post-verbally, linking new material to prior discourse. Complementing this structural account, Prince (1978, 1981) argued that sentence construction reflects writers' assumptions about readers' knowledge states rather than purely grammatical preference. Her Given–New taxonomy classifies constituents according to whether they are already accessible in discourse (given) or newly introduced (new), and her concept of informative-presupposition explains how *it*-clefts can present new or contestable claims within a structure conventionally associated with presupposed content, thereby lending claims an appearance of epistemic authority (Prince, 1978, p. 884). Halliday's Systemic Functional Linguistics (Halliday & Matthiessen, 2014) offers a third lens, modeling the clause-initial Theme as the point of departure for a message and the Rheme as its informational development; thematic progression across clauses, whether linear (Rheme becoming the next Theme) or constant (a single Theme sustained across several clauses), has long been associated with textual coherence (Daneš, 1974). Empirical work on these structures in professional registers shows that they are systematically motivated rather than stylistically incidental. Banks (1995) found that, although *it*-clefts and pseudo-clefts occur relatively infrequently across eleven scientific articles, related structures such as extraposition and existential constructions perform hedging functions that contribute to an impersonal, objectified academic register. Zhou and Chen (2021), analyzing 227 applied linguistics articles within an Appraisal-theory framework, found that pseudo-clefts cluster in Literature Review and Results/Discussion sections, where they serve evaluative and highlighting functions tied to authorial stance. Both studies, however, treat their corpora as relatively homogeneous, leaving open whether information-packaging strategies vary systematically with publication prestige. Topicalization has been examined extensively in World Englishes scholarship. Leuckert (2017a) found that Indian English speakers front constituents considerably more often than British English speakers, while Hong Kong English speakers do so less often despite Cantonese being a topic-prominent language, suggesting that first-language transfer into English topicalization is neither direct nor automatic. A companion study (Leuckert, 2017b) showed that second-language speakers often front entirely new information to sustain conversational flow, challenging the assumption that topicalization strictly follows given-before-new sequencing. Yixuan (2023) applied Prince's taxonomy to locative inversion, there-insertion, and left dislocation in fairy tales, concluding that such structures function as coherence-building devices that reduce processing effort, and recommended their extension to formal academic

discourse. Psycholinguistic research clarifies why deployment of these structures, rather than mere familiarity with their form, distinguishes more from less skilled writers. Gattei, París, and Shalom (2021) showed via eye-tracking that supportive discourse contexts improve reading speed for non-canonical structures but do not eliminate processing cost, while Lubis and Miranti's (2024) systematic review of nine studies found that passive-heavy texts generally produce slower reading and weaker coherence judgments, implying that effective use of the passive requires careful contextual support rather than indiscriminate application. The literature establishes that non-canonical structures perform systematic discourse functions in professional and academic writing, and that their effective use depends on pragmatic competence rather than grammatical knowledge alone. However, two specific gaps remain. First, no corpus-based study has stratified an academic corpus according to institutional publication tier as the primary comparative variable; existing studies (Banks, 1995; Zhou & Chen, 2021) examine expert writing as a relatively undifferentiated category, despite the Pakistani HEC system offering an established, formal basis for such a comparison. Second, prior research has concentrated on full articles or on results/discussion and literature review sections, leaving the conclusion, where writers must compress findings and assert final disciplinary authority, largely unexamined as a discrete site of information packaging. The present study addresses both gaps directly by comparing the frequency, distribution, and discourse functions of passive voice, cleft constructions, and topicalization in the conclusion sections of HEC W-category and Y-category linguistics and literature journals.

## **Methodology**

This study employs a corpus-based comparative design combining quantitative frequency analysis with qualitative discourse interpretation. The quantitative strand establishes the distributional profile of three non-canonical structures, passive voice, cleft constructions, and topicalization, across two institutionally defined sub-corpora; the qualitative strand interprets representative concordance lines through the Information Packaging, Given–New, and Theme–Rheme frameworks outlined in Section 2. Following Creswell and Creswell (2018), the two strands are treated as complementary rather than merged, each addressing a distinct facet of the same textual data. The target population comprised research articles published between 2019 and 2026 in HEC-recognized W-category (high-tier) and Y-category (mid-tier) journals in linguistics and literature. Eighty articles were selected through purposive sampling (Patton, 2015; Palinkas et al., 2015): 40 from W-category and 40 from Y-category journals, restricted to these two disciplines to control for cross-disciplinary variation in rhetorical convention. Only conclusion sections (including those titled Conclusion, Discussion and Conclusion, or Implications and Conclusion) were extracted, as these sections concentrate the discourse-organizing and evaluative work most relevant to information packaging. W-category texts were drawn from SAGE Journals and comparable international platforms; Y-category texts were drawn from the official hosting sites of HEC-recognized national journals. The resulting corpus totaled 29,559 words: 17,287 in the W-category corpus (WCC) and 12,272 in the Y-category corpus (YCC). This word-count asymmetry, arising naturally from longer average conclusions in W-category texts rather than from sampling design, was addressed through frequency normalization (Section 3.4). Extracted conclusions were compiled into plain-text sub-corpus files and part-of-speech tagged using the Penn Treebank tagset. Tagged files were imported into AntConc 3.5.9, the primary instrument for concordance extraction, frequency counting, and Key-Word-in-Context (KWIC) verification. Passive voice was identified through regular-expression queries targeting auxiliary-plus-past-participle sequences (e.g., `(is|are|was|were|be|been|being)_VB[A-Z]*\s+\w+_VBN`), with each hit manually verified against Huddleston and Pullum's (2002)

structural criteria to exclude stative pseudo-passives. Cleft constructions were identified through targeted searches for it-cleft and wh-cleft frames and verified against the canonical [It + BE + focused constituent + that/who-clause] and [wh-clause + BE + focused constituent] templates. Because topicalization is highly context-dependent, candidate instances were located through contextual search strings (e.g., In this article, Taken together) and confirmed through manual inspection for genuine non-subject constituent fronting, as distinct from Hallidayan thematic fronting. Raw frequencies were normalized per 1,000 words (Normalized Frequency = Raw Frequency / Total Words × 1000) to control for the unequal sub-corpus sizes (McEnery & Hardie, 2011; Brezina, 2018). Pearson chi-square ( $\chi^2$ ) tests of independence, computed in IBM SPSS Statistics at  $\alpha = .05$ , assessed whether tier-based distributional differences were statistically reliable; effect sizes were reported as Cramér's V (tables larger than  $2 \times 2$ ) or the phi ( $\phi$ ) coefficient ( $2 \times 2$  tables). Given the very low raw counts for cleft constructions (4 in WCC, 1 in YCC), normalized cleft frequencies are interpreted as indicative rather than statistically conclusive (Brezina, 2018), with qualitative analysis providing the primary evidential basis for cleft-related claims. A randomly selected 20% sample of concordance lines, proportionally drawn from both sub-corpora, was independently reviewed by a second coder against the structural criteria of Huddleston and Pullum (2002) and Prince (1978)/Halliday and Matthiessen (2014). Inter-rater agreement reached 85%, within the range considered acceptable for manual corpus coding (Krippendorff, 2004); disagreements, concentrated in structurally ambiguous stative pseudo-passives and clause-initial adjuncts, were resolved through discussion and consensus before finalizing the dataset.

## Findings and Discussion

### Corpus Profile

Table 1 summarizes the structural profile of the compiled corpus.

**Table 1.** *Structural profile of the conclusion corpus*

Sub-Corpus	Journal Tier	Articles / Words
WCC	High-tier (W-category)	40 / 17,287
YCC	Mid-tier (Y-category)	40 / 12,272
Total	—	80 / 29,559

Although the two sub-corpora are balanced in article count, WCC conclusions are, on average, longer than YCC conclusions, a pattern consistent with Biber et al.'s (1999) observation that higher-register academic texts tend to deploy more elaborated discourse structures. All frequency comparisons below are normalized per 1,000 words to control for this asymmetry.

### Passive Voice: A Shared Resource, Differently Calibrated

Passive voice was the most frequent non-canonical structure in both sub-corpora, with agentless passives exceeding 86% of all passive tokens in each, confirming that impersonal, process-oriented expression is a stable register feature of academic writing rather than a marker of publication tier. However, the two corpora diverged sharply in passive sub-type distribution (Table 2).

**Table 2.** *Distribution of passive voice sub-types by journal tier*

Passive Sub-type	WCC (%)	YCC (%)	$\Delta$ (pp)
Simple present	37.1	49.7	+12.6 (Y)
Modal	26.9	18.7	+8.2 (W)
Perfective	20.0	19.3	+0.7 (W)
Simple past	14.3	9.4	+4.9 (W)
Progressive	1.7	2.9	+1.2 (Y)

A chi-square test across the full five-category distribution approached but did not reach significance,  $\chi^2(4) = 7.98$ ,  $p = .092$ , Cramér's  $V = .152$ , indicating a discernible but modest overall trend. When the analysis was focused on the theoretically central contrast, modal versus simple present passives (the two categories driving the omnibus trend), the association with publication tier was statistically significant,  $\chi^2(1, N = 229) = 5.41$ ,  $p = .020$ , Fisher's exact  $p = .026$ ,  $\phi = .15$ . This confirms that WCC writers allocate a significantly greater share of their passive resources to modal passives, while YCC writers favor unmodalized present-tense assertion. Qualitatively, this distinction reflects more than grammatical preference. As Hyland (1998, p. 1) observes, “the ability to present claims with appropriate certainty is perhaps one of the most important skills a writer can develop.” A simple present passive (e.g., *differences are found*) presents a claim as categorical, whereas a modal passive (e.g., *differences might be found*) signals epistemic caution and acknowledges the provisional nature of empirical claims. The higher modal-passive rate in WCC therefore suggests that W-category writers more consistently calibrate certainty to evidential strength, a hallmark of rhetorical maturity in scholarly argumentation (Hyland, 2005). This addresses RQ1 and RQ3 directly: tier-based variation in passive voice is not about whether the passive is used, but about which sub-type is favored and what stance it encodes.

***Cleft Constructions: Rare, but More Elaborated in W-Category Texts***

Cleft constructions were rare overall, and *it*-clefts were entirely absent from both corpora (Table 3), despite being among the most productive focus-marking devices documented in formal English prose (Biber et al., 1999; Delin, 1995). Given the very small raw counts (4 in WCC, 1 in YCC), the normalized density figures (3.44 vs. 1.01 per 1,000 words) are best treated as indicative of a directional tier-based pattern rather than as statistically robust evidence (Brezina, 2018).

**Table 3.** *Cleft constructions and topicalization by journal tier*

Structure	WCC	YCC
<i>it</i> -clefts	0	0
<i>wh</i> -clefts / reversed <i>wh</i> -clefts	4 (3.44 per 1,000 words)	1 (1.01 per 1,000 words)
Genuine topicalization	0	0

Qualitative inspection shows that the four WCC instances split evenly between *wh*-clefts and reversed *wh*-clefts, used at key argumentative junctures to categorize findings and mark logical transitions; the single YCC instance shows no comparable structural elaboration. Following Prince's (1978, p. 884) account of informative-presupposition, *it*-clefts allow writers to present new or contestable claims within a structure conventionally associated with presupposed content, lending the claim an appearance of established acceptance. The complete absence of this device from both corpora, and especially from YCC, may constitute a structural correlate of the comparatively lower rhetorical assertiveness associated with mid-tier academic prose (Delin, 1995; Swales, 1990), bearing directly on RQ3.

***Topicalization: A Null Finding and the Centrality of Thematic Fronting***

The most striking finding of the study is that no genuine instances of syntactic topicalization, true constituent fronting from canonical post-verbal position, were identified in either corpus, despite 786 clause-initial structures being retrieved and manually examined through AntConc concordance analysis. This null result is theoretically unsurprising: Prince (1978) specifies that topicalization requires discourse-given information and contrastive focus within an interactive context, conditions rarely met in the formal, monological register of academic conclusions;

Huddleston and Pullum (2002) similarly characterize topicalization as a marked construction used sparingly even where licensed. In place of syntactic topicalization, both corpora relied extensively on Hallidayan thematic fronting, textual themes (however, therefore, thus), marked topical themes (in this study, in this article), and interpersonal themes (notably, significantly, importantly), as the primary mechanism for organizing clause-initial position (Table 4).

**Table 4.** *Thematic fronting resources by journal tier*

Thematic Resource	WCC (instances)	YCC (instances)
Marked topical themes (e.g., “In this study”)	15	1
Interpersonal themes (e.g., notably, significantly)	7	2
Textual themes (however, therefore, thus)	majority category, ~73%	majority category, ~93%

WCC shows substantially higher frequencies across all thematic categories, particularly marked topical themes (15 instances vs. 1) and interpersonal themes (7 vs. 2), indicating both a stronger tendency to anchor conclusions within an explicit research frame and a wider range of evaluative resources for signaling authorial stance. This pattern directly answers RQ2: rather than displacing constituents syntactically, W-category writers achieve discourse organization and reader guidance through a richer and more frequent deployment of textual-metafunction resources, a distinction that is functional and frequential rather than structural.

### **Synthesis**

Taken together, the findings indicate that the linguistic distinction between high-tier and mid-tier academic writing does not lie primarily in which non-canonical structures are available or even used, both corpora draw on the same shared structural repertoire, agentless passives, sparse clefts, and thematic fronting in place of topicalization, but in how strategically and precisely that repertoire is deployed. W-category writers more consistently calibrate epistemic stance through modal passives, sustain a wider and more structurally elaborated use of cleft constructions, and anchor their conclusions more explicitly and evaluatively through thematic fronting. These differences, individually modest but cumulatively consistent (small-to-medium effect sizes, directionally aligned findings across three independent structures), support the conclusion that information-packaging sophistication, rather than syntactic novelty, is what differentiates high-tier from mid-tier academic conclusions. This synthesis directly answers all three research questions and substantiates the study's central claim regarding the rhetorical and pragmatic, rather than purely structural, basis of publication-tier variation.

### **Conclusion**

This study compared passive voice, cleft constructions, and topicalization in the conclusion sections of HEC W-category and Y-category linguistics and literature journals. Passive voice dominated both corpora, but W-category writers favored epistemically calibrated modal passives, while Y-category writers relied on unmodalized declarative assertion. Cleft constructions were rare overall, with *it*-clefts entirely absent, though W-category writers used a more varied and elaborated range of *wh*-clefts. No genuine syntactic topicalization occurred in either corpus; both relied instead on Hallidayan thematic fronting, deployed more frequently and with a wider evaluative range by W-category writers. These results indicate that the syntactic signature of high-tier academic writing lies not in structurally novel or rare constructions, but in the precision, epistemic calibration, and thematic continuity with which a shared repertoire is deployed. This

refines rather than contradicts prior assumptions: non-canonical structures remain purposeful discourse strategies, but their tier-distinguishing function operates at the level of sub-type selection and frequency rather than structural presence or absence. The study offers the first corpus-based comparison of non-canonical syntax across HEC publication tiers and the first to isolate the conclusion as a discrete analytical site. Methodologically, it demonstrates the value of combining AntConc-driven extraction with multi-theoretical qualitative interpretation. Pedagogically, the findings suggest that novice and second-language writers may develop rhetorical authority less by acquiring new syntactic forms than by calibrating epistemic stance through modal passives and deploying thematic resources more deliberately within their conclusions.

## References

- Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies*. Routledge. <https://doi.org/10.4324/9780203087701>
- Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.10>
- Ariel, M. (1990). *Assessing noun-phrase antecedents*. Routledge.
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum. <https://doi.org/10.5040/9781350933996>
- Banks, D. (1995a). There is a cleft in your sentence: Less common clause structures in scientific writing. *ASp. La revue du GERAS*, 7–10, 3–11. <https://doi.org/10.4000/asp.3702>
- Banks, D. (1995b). There is a systematic minority: Non-canonical structures in scientific English. *ASp. La revue du GERAS*, 7–10, 303–314. <https://doi.org/10.4000/asp.4034>
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). John Wiley & Sons.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511804489>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.
- Birner, B. J., & Ward, G. (1998). *Information status and non-canonical word order in English*. John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.40>
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Caldecott, M. (1999). *It-clefts and wh-clefts: Their functions in English discourse* (Master's thesis, University of British Columbia). <https://open.library.ubc.ca/collections/ubctheses/831/items/1.0089307>
- Callies, M. (2009). *Information highlighting in advanced learner English: The syntax-pragmatics interface in second language acquisition*. John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.186>
- Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide*. Cambridge University Press.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). Academic Press.
- Chocano, G., Jiménez, R., Lozano, C., Mendikoetxea, A., Murcia, S., O'Donnell, M., Rollinson, P., & Teomiro, I. (2007). An exploration into word order in learner corpora: The WOSLAC project. In M. Davies, P. Rayson, S. Hunston, & P. Danielsson (Eds.),

- Proceedings of Corpus Linguistics 2007 (Article 113). University of Birmingham.  
<http://ucrel.lancs.ac.uk/publications/CL2007/>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Ablex.
- Collins, P. C. (1991). *Cleft and pseudo-cleft constructions in English*. Routledge.  
<https://doi.org/10.4324/9780203321584>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F. Daneš (Ed.), *Papers on functional sentence perspective* (pp. 106–128). Academia.
- Declerck, R. (1988). Restrictive when-clauses. *Linguistics and Philosophy*, 11(2), 131–168.
- Delin, J. (1992). Properties of it-cleft presupposition. *Journal of Semantics*, 9(4), 289–306.  
<https://doi.org/10.1093/jos/9.4.289>
- Delin, J. (1995). Presupposition and shared knowledge in it-clefts. *Language and Cognitive Processes*, 10(2), 97–120.
- Doherty, M. (2001). Discourse relators and the beginnings of sentences in English and German. *Languages in Contrast*, 3(2), 223–251.
- Eggs, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). Continuum.
- Ferreira, F. (2003). The misinterpretation of non-canonical sentences. *Cognitive Psychology*, 47(2), 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7)
- Flowerdew, J. (2008). Corpus-based analyses of the problem-solution pattern in a genre of professional discourse. *Corpus Linguistics and Linguistic Theory*, 4(1), 55–78.
- Flowerdew, J. (2012). *Discourse in English for academic purposes*. Routledge.  
<https://doi.org/10.4324/9780203073711>
- Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 11–33). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.16.02flo>
- Fries, P. H. (1981). On the status of theme in English: Arguments from discourse. *Forum Linguisticum*, 6(1), 1–38.
- Gattei, C. A., París, L. A., & Shalom, D. E. (2021a). Information structure and word order canonicity in the comprehension of Spanish texts: An eye-tracking study. *Frontiers in Psychology*, 12, 629724. <https://doi.org/10.3389/fpsyg.2021.629724>
- Gattei, C. A., París, L. A., & Shalom, D. E. (2021b). The processing of non-canonical word orders: An eye-tracking study on the effects of context and focus. *Journal of Psycholinguistic Research*, 50(4), 843–865. <https://doi.org/10.1007/s10936-020-09751-2>
- Gómez-González, M. A. (2001). *The theme-topic interface: Evidence from English*. John Benjamins Publishing Company.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.
- Hyland, K. (1998). *Hedging in scientific research articles*. John Benjamins Publishing Company.
- Jamshaid, S. (2024). Pakistani English: An analysis of grammatical-syntactical structures of Pakistani English novels. *Journal of Development and Social Sciences*, 5(2), 335–352.  
[https://doi.org/10.47205/jdss.2024\(5-11-S\)33](https://doi.org/10.47205/jdss.2024(5-11-S)33)
- Jespersen, O. (1924). *The philosophy of grammar*. George Allen & Unwin.
- Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*. MIT Press.

- Lambrech, K. (1994). Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620607>
- Lambrech, K. (2001). A framework for the analysis of cleft constructions. *Linguistics*, 39(3), 463–516. <https://doi.org/10.1515/ling.2001.021>
- Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 105–122). Mouton de Gruyter. <https://doi.org/10.1515/9783110867275.105>
- Leuckert, S. (2017a). How interactional needs shape information structure: An analysis of the discourse functions of topicalization in three L2 varieties of English. *Yearbook of the German Cognitive Linguistics Association*, 5(1), 25–42. <https://doi.org/10.1515/gcla-2017-0003>
- Leuckert, S. (2017b). Structures at the syntax-discourse interface: The use of non-canonical word orders in varieties of English. Language Science Press. <https://doi.org/10.5281/zenodo.1042455>
- Leuckert, S. (2017c). Typological interference in information structure: The case of topicalization in Asia. *Zeitschrift für Anglistik und Amerikanistik*, 65(3), 283–302. <https://doi.org/10.1515/zaa-2017-0029>
- Li, C. N., & Thompson, S. A. (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 457–489). Academic Press.
- Lidzba, K., de Marco, D., Wilke, M., Krägeloh-Mann, I., & Hauser, T.-K. (2013). Late-mastered peaks of grammatical difficulty: An fMRI study on the processing of object topicalization in developing populations. *NeuroImage*, 78, 442–451. <https://doi.org/10.1016/j.neuroimage.2013.04.032>
- Lidzba, K., Konietzko, A., Schwilling, E., Krägeloh-Mann, I., & Winkler, S. (2013). Processing of non-canonical word-order: A case-series on lesion-induced reorganized language and age-effects in typical development. *Brain and Language*, 127(3), 377–387. <https://doi.org/10.1016/j.bandl.2013.09.004>
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.
- Liu, L., & Ambridge, B. (2021). Balancing information-structure and semantic constraints on construction choice: Building a computational model of passive and passive-like constructions in Mandarin Chinese. *Cognitive Linguistics*, 32(3), 349–388. <https://doi.org/10.1515/cog-2019-0100>
- Lobo, M., Santos, A. L., Soares-Jesel, C., & Vaz, S. (2019). Effects of syntactic structure on the comprehension of clefts. *Glossa: A Journal of General Linguistics*, 4(1), 74. <https://doi.org/10.5334/gjgl.645>
- López-Beltrán Forcada, P. (2014). Information structure violations in passive constructions at the syntax-discourse interface by advanced L2 English learners (Master's thesis, Universidad de Granada).
- Los, B. (2009). The consequences of the loss of verb-second in English: Information structure and syntax in interaction. *English Language and Linguistics*, 13(1), 97–125. <https://doi.org/10.1017/S1360674308002876>
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157–176). John Benjamins Publishing Company. <https://doi.org/10.1075/z.64.11lou>

- Lubis, M. Y. A., & Miranti, R. (2024). Passive voice and active voice. *Cemara Education and Science*, 2(2). <https://doi.org/10.62145/ces.v2i2.52>
- Maamujav, U., Olson, C. B., & Chung, H. (2021). Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of Second Language Writing*, 53, 100822. <https://doi.org/10.1016/j.jslw.2021.100822>
- MacWhinney, B. (1977). Starting points. *Language*, 53(1), 152–168. <https://doi.org/10.2307/412011>
- Martin, J. R. (1992). *English text: System and structure*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.59>
- Martin, J. R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. Continuum.
- Mauranen, A. (1993). *Cultural differences in academic rhetoric: A textlinguistic study*. Peter Lang.
- McCabe, A. (1999). *Theme and thematic patterns in Spanish and English history texts* (Doctoral dissertation, Aston University). <https://publications.aston.ac.uk/id/eprint/10842/>
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395>
- Neeleman, A., & van de Koot, H. (2016). Word order and information structure. In C. Féry & S. Ishihara (Eds.), *The Oxford handbook of information structure* (pp. 358–379). Oxford University Press.
- Nwogu, K. N. (1991). Structure of science popularization articles: A discourse analysis of Rice's four-part macro-structure. *English for Specific Purposes*, 10(2), 111–123. [https://doi.org/10.1016/0889-4906\(91\)90004-G](https://doi.org/10.1016/0889-4906(91)90004-G)
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Park, K.-S. (2011). Information structure and dative word order in adult L2 learners. In J. Herschensohn & D. Tanner (Eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)* (pp. 101–109). Cascadilla Proceedings Project.
- Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131–156. <https://doi.org/10.1075/ijcl.9.1.07par>
- Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Sage Publications.
- Prince, E. F. (1978). A comparison of wh-clefts and it-clefts in discourse. *Language*, 54(4), 883–906. <https://doi.org/10.2307/413238>
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). Academic Press.
- Prince, E. F. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W. C. Mann & S. A. Thompson (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295–325). John Benjamins Publishing Company. <https://doi.org/10.1075/pdt.16.13pri>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Rankin, T. (2009). Verb second in advanced L2 English: A learner corpus study. In M. Bowles, T. Ionin, S. Montrul, & A. Tremblay (Eds.), *Proceedings of the 10th Generative*

- Approaches to Second Language Acquisition Conference (GASLA 2009) (pp. 46–59). Cascadilla Proceedings Project.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge. <https://doi.org/10.4324/9780203594070>
- Slioussar, N., & Harchevnik, M. (2024). Word order and context in sentence processing: Evidence from L1 and L2 Russian. *Frontiers in Psychology*, 15, 1344366. <https://doi.org/10.3389/fpsyg.2024.1344366>
- Sorace, A. (2011). Pinning down the expression of interface vulnerability. *Linguistic Approaches to Bilingualism*, 1(1), 1–33. <https://doi.org/10.1075/lab.1.1.01sor>
- Stefanowitsch, A., & Gries, S. Th. (2003). Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Struik, T., & van Kemenade, A. (2020). On the givenness of OV word order: A (re)examination of OV/VO variation in Old English. *English Language and Linguistics*, 24(1), 1–22. <https://doi.org/10.1017/S1360674318000187>
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell Publishers.
- Svartvik, J. (1966). *On voice in the English verb*. Mouton.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Thompson, G. (2004). *Introducing functional grammar* (2nd ed.). Arnold.
- Thompson, G. (2014). *Introducing functional grammar* (3rd ed.). Routledge. <https://doi.org/10.4324/9780203431474>
- Thompson, G., & Hunston, S. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 1–27). Oxford University Press.
- Thu, P. D. A. (2023). Exploring English information packaging constructions in written and spoken language. *Journal of Knowledge Learning and Science Technology*, 2(2), 68–74. <https://doi.org/10.60087/jklst.vol2.n2.p.74>
- Ward, G., & Birner, B. J. (2004). Information structure and non-canonical syntax. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 153–174). Blackwell Publishing. <https://doi.org/10.1002/9780470756959.ch7>
- Zaidi, S., Yasmin, F., Nimehchisalem, V., Kasim, Z. M., & Ali, A. M. (2023). Grammatical knowledge of English cleft constructions among Pakistani ESL learners across L2 proficiency levels and learning styles. *GEMA Online Journal of Language Studies*, 23(3), 76–94. <https://doi.org/10.17576/gema-2023-2303-05>
- Zhang, Y. (2022). Syntactic and semantic properties of on the contrary in British university student essay writing: A corpus-based systemic functional analysis. *Applied Sciences*, 12(20), 10635. <https://doi.org/10.3390/app122010635>
- Zhou, H., & Chen, M. (2021). What still needs to be noted: Pseudo-clefts in the academic discourse of applied linguistics. *Frontiers in Psychology*, 12, 672349. <https://doi.org/10.3389/fpsyg.2021.672349>