Russian Journal of Linguistics
ISSN 2687-0088 (Print), ISSN 2686-8024 (Online)

2026 Vol. 30 No. 3 738–763
http://journals.rudn.ru/linguistics

https://doi.org/10.22363/2687-0088-46681
EDN: NZJZHD

Research article / Научная статья

Semantic compression in small-format texts: A corpus-based study of book synopses

Dmitry S. KHRAMCHENKO

MGIMO University, Moscow, Russian Federation
✉d.khramchenko@inno.mgimo.ru

Abstract
Digital book retailers often provide prospective purchasers with only a brief synopsis prior to purchase. The means by which brief texts, most often under 250 words, can simultaneously summarize, brand, promote, and persuasively present an 80.000-word novel remains under-described in the existing literature. The aim of the study is to examine how mystery/thriller synopses use semantic compression (maximising meaning per word) to meet both narrative and marketing goals. The study focuses on three main aspects: (1) the use of semantic compression to achieve persuasive effects; (2) the lexical and grammatical patterns that are characteristic of this microgenre; (3) the linguistic mechanisms behind the compression process. The research is based on a corpus of 300 (56.220 words) English Goodreads synopses (2022–2025). Phase 1 used AntConc and CLAWS-C7 tagging for length, sentence profile, lexical density, keywords, and voice/mood. Phase 2 applied double-coded discourse analysis grounded in Systemic Functional Linguistics, small-format text theory, and paratext theory, charting compression devices at macro-, meso-, and micro-levels. Synopses average 187 words, show high lexical density (52%), and prefer interrogatives and active voice. 86% follow a seven-move functional-pragmatic structure: hook, micro-portrait, temporal leap, thematic bundle, suspense question, comparative tag, sell-line. Compression is further executed via paratactic stacks, em-dash grafts, cardinality lists, modifier clusters, and systematic omission of subplots and endings. Recycled evaluatives and genre labels act as semantic shortcuts that activate readers' genre schemata. Semantic compression is theorized as an important cognitive-discursive phenomenon functioning at the junction of literature and commerce. Book synopses are hybrid literary-marketing micro-genres whose layered compression turns narrative surplus into reader engagement. The study offers a replicable template that writers, publishers, copywriters, and automated-copy systems can use to craft more persuasive small-format texts.

Keywords: *semantic compression, small-format text, paratext, book synopsis, functional linguistics, corpus linguistics, book marketing*

© Dmitry S. Khramchenko, 2026
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
https://creativecommons.org/licenses/by-nc/4.0/legalcode

For citation:
Khramchenko, Dmitry S. 2026. Semantic compression in small-format texts: A corpus-based study of book synopses. *Russian Journal of Linguistics* 30 (3). 738–763. https://doi.org/10.22363/2687-0088-46681 EDN: NZJZHD

Семантическая компрессия в малоформатных текстах: корпусное исследование книжных синопсисов

Дмитрий Сергеевич ХРАМЧЕНКО ✉

МГИМО МИД России, Москва, Российская Федерация
✉d.khramchenko@inno.mgimo.ru

Аннотация
На цифровых сервисах для любителей книг пользователям зачастую доступен лишь краткий синопсис произведения. Дискурсивные механизмы, позволяющие тексту менее 250 слов одновременно пересказать, «забрендировать», прорекламировать и продать роман гораздо большей длины, до сих пор недостаточно изучены. Цель исследования — выявить, как аннотации к детективам и триллерам реализуют семантическую компрессию (максимизацию смысла на единицу объёма) для решения литературных и маркетинговых задач. В фокусе внимания — (1) использование семантической компрессии для достижения персуазивного прагматического эффекта; (2) лексико-грамматические паттерны, характерные для данного микрожанра; (3) лингвистические механизмы компрессии. Материалом послужил корпус из 300 англоязычных аннотаций с сайта Goodreads (2022–2025 гг.; 56 220 слов). На первом этапе исследования применялись AntConc и CLAWS-C7 для определения длины и типа предложений, лексической плотности, ключевых слов и грамматических особенностей. На втором этапе проведён дискурсивный анализ в русле системно-функциональной лингвистики, теории малоформатных текстов и теории паратекста, позволивший описать приёмы компрессии на макро-, мезо- и микроуровнях. Средняя длина аннотаций составила 187 слов; лексическая плотность 52 %; преобладают вопросительные конструкции и активный залог. 86 % текстов следуют семичленной функционально-прагматической структуре: рекламный «крючок», микропортрет героя, пространственно-временной сдвиг, тематическая плотность, вопрос-саспенс, интертекстуальное уподобление, метатекстуальная реклама. Дополнительная компрессия достигается паратаксисом, парантезой с тире и двоеточием, перечислениями, цепочками прилагательных и систематическим опущением побочных линий сюжета. Оценочные эпитеты и жанровые маркеры активируют фреймы культурного опыта реципиентов. Семантическая компрессия рассматривается как важнейший когнитивно-дискурсивный феномен на стыке литературы и коммерции. Аннотация — гибридный литературно-маркетинговый микрожанр, чья многоуровневая компрессия конвертирует сюжетные тропы в читательский интерес. Статья предлагает авторам, издателям, копирайтерам и системам автоматической генерации текстов воспроизводимую функционально-прагматическую модель написания персуазивных малоформатных текстов.
Ключевые слова: *семантическая компрессия, малоформатный текст, паратекст, книжный синопсис, функциональная лингвистика, корпусная лингвистика, маркетинг*

Для цитирования:
Khramchenko D.S. Semantic compression in small-format texts: A corpus-based study of book synopses. *Russian Journal of Linguistics.* 2026. Vol. 30. № 3. P. 738–763. https://doi.org/10.22363/2687-0088-46681 EDN: NZJZHD

1. Introduction

Reading fiction may be getting less popular in today's smartphone-driven, internet-dominated cultural epoch, but the publishing business continues to make money. Digital bookselling has turned the once leisurely practice of browsing dust-jackets into a split-second decision taken on the small screen of a mobile device. In a compressed attention economy like this, the short promotional text of a couple of paragraphs that accompanies every title on Amazon, Kobo, Barnes & Noble, or Goodreads carries a disproportionate persuasive functional-pragmatic load. Publishers have long relied on jacket copy to entice readers, yet the online environment amplifies two major constraints that make modern-day blurbs a distinct linguistic object: (1) severe spatial limitations on retail platforms with the so-called "above-the-fold" character caps, and (2) algorithmically curated lists that pit thousands of book titles against each other in a scrollable feed. In these circumstances, success depends on what linguists would describe as semantic compression, i.e., the strategic packing of maximum narrative promise and attention-grabbing verbal expression into minimal textual space. Despite a growing body of scholarship on digital paratexts (e.g., Batchelor 2021, Nord 2019, Skare 2021) and on micro-genres and small-format texts such as tweets (Page 2013, Zappavigna 2017), abstracts of scholarly articles (Ayers 2008, Boginskaya 2022, Cherkunova 2021, Golubykh 2020), or titles of artworks (Kharkovskaya et al. 2019), there has been surprisingly little linguistic exploration of the digital book synopsis as a small-format text whose core communicative purpose is to sell books.

This article intends to fill the gap by combining Systemic Functional Linguistics (SFL) with corpus-based methods to examine how synopses use lexico-grammatical means to achieve semantic compression. We define the book synopsis as any publisher- or author-supplied text of ≤ 350 words (though in practice most fall within the 150-250-word range) that introduces a work of literary fiction (e.g., a novel, novella, novelette, or a collection of short stories) to prospective readers on the dedicated digital platform. Although such uploads often replicate more traditional jacket copy, publishers routinely shorten or restructure original blurbs to fit the platform's constraints, producing a special hybrid genre situated somewhere between the classical back-cover blurb and the Twitter-style teaser, existing at the intersection of literary fiction, marketing and advertising, and pop-cultural discourse. From an SFL perspective, these texts must juggle three metafunctional demands (Halliday & Matthiessen 2013): (1) ideational, as they summarize plot, setting, genre, and characters without "spoiling" the story; (2) interpersonal, by positioning the prospective reader as co-conspirator and confidant while being a target market; and (3) textual, because they orchestrate the whole discourse in

general and information flow in particular so that key pragmatic hooks occupy salient clause-initial or clause-final positions.

At the same time, synopsis writers pursue a clear commercial telos, that is to swiftly move readers down the purchase funnel — quick reading, clicking, shelving, buying. Drawing on compression theory (Johnson et al. 2003) and relevance theory (Sperber & Wilson 1995, Wilson & Sperber 2002), in this article, we conceptualize semantic compression as the maximization of implicature per discursive unit. Thus, every adjective, descriptive phrase, deixis, or evaluative token must earn its keep by multiplying narrative affordances in the reader's mind. A compressed synopsis is therefore not merely shorter. It is semantically and pragmatically denser, exploiting metafunctional synergies to the full extent to trigger specific genre expectations and emotional resonance, staying within strict word count limitations.

The study addresses three research questions:

RQ1: How do synopses for mystery and thriller novels use the principle of semantic compression to achieve the pragmatic effect of persuasiveness and fulfil their commercial function?

RQ2: What are the characteristic linguistic patterns of book synopses in the mystery and thriller genre, establishing them as a distinct genre of small-format texts?

RQ3: Through which specific linguistic mechanisms and operators is semantic compression realized at the macro-, meso-, and micro-levels of the synopsis?

2. Theoretical background

This study combines perspectives from digital discourse analysis, paratextual theory, functional linguistics, pragmatics, and information theory in order to explain why digital book synopses look and function the way they do. The section proceeds in three steps. First, it sketches the notion of a small-format text as a product of the platform economy. Second, it treats semantic compression as the discursive mechanism that allows such texts to carry narrative, evaluative, and commercial information simultaneously. Third, it describes the book synopsis through Genette's paratextual framework, showing how it mediates between the full literary novel and the pop-cultural marketplace.

2.1. Small-format texts in digital culture

Ubiquitous mobile screens and feed-based interfaces, as well as a chronic surplus of competing messages, have normalised genres that are measured not in pages but in centimetres of scroll (Miller & Shepherd 2009, Zappavigna 2021). Following Kubryakova (2001), we use the term small-format text for any self-contained verbal creation whose primary design constraint is extreme brevity.

Recent research isolates four recurrent functional-linguistic properties of such texts: their conciseness, secondarity, pragmatic intentionality, and expressive

density. Word count limits for small-format texts differ based on their particular subgenre. For instance, news headlines may average 6 to 10 tokens, whereas research abstracts stretch to 250–300, but the organising principle is a narrow channel and a single, high-stakes reading pass (Poulimenou et al. 2016). Small-format texts are typically dependent on a more expansive primary text: tweets to linked articles, titles to stories, headlines to news coverage, abstracts to scholarly journal papers (Ayers 2008, Remchukova & Apostolidi 2018, Khramchenko 2023, Kharkovskaya & Cherkunova 2020, Zibin et al. 2024). Their overriding communicative goal is quick uptake, accomplished by semantically and pragmatically saturated vocabulary, syntactic formulae, incorporated intertextuality, and genre-specific clichés (Aijmer 2005, Starostina et al. 2021). Because they occupy the intermediary position between browsing and deeper engagement, small-format texts concentrate their rhetorical expressiveness and persuasiveness, often mixing informational and attitudinal semantic elements in the same clause, which is an optimal strategy in the modern-day environment of mosaic thinking, multitasking, and a diminished capacity for deep reflection among regular Internet users (Cherkunova & Ponomarenko 2021, Carr 2020, Sdobnikov 2021).

2.2. Semantic compression

The efficiency of small-format texts is secured by semantic compression, which is a process more conceptual than mere syntactic shortening. Drawing on mathematical information theory and Relevance Theory (Sperber & Wilson 1995, Wilson & Sperber 2002), semantic compression can be described as the reduction of linguistic length by means of the activation of inferential paths already stored in the recipients' cognitive repertoire. Fillmore's (2006) frame semantics and Swales's (1990) genre schemata provide the descriptive apparatus for how those inferential shortcuts work.

Although it is frequently mentioned in small-format texts research works as one of their defining qualities, semantic compression as an important linguistic phenomenon remains underexplored. Cherkunova (2021) notes that compression quality is measured not by the amount of text deleted but rather by the accuracy with which the elided meaning can be recovered. As for digital platforms' book synopses, success is assessed empirically and often remains elusive: higher click-through and "Want-to-Read" rates indicate that readers have indeed reconstructed the intended scenario and affect. The functional-pragmatic and cognitive-discursive mechanisms of semantic compression are yet to be theorized and analyzed to get a better understanding of how linguistic properties of these texts serve commercial purposes. This study will try to bridge this gap.

2.3. Book synopses as hybrid paratexts

In the broader metadiscourse that now surrounds contemporary fiction (e.g., author interviews, jacket design, influencer "BookTok" reviews, and algorithm-generated recommendation feeds), the book synopsis occupies an

important paratextual slot. In Genette's (1997) terms, such a synopsis is an epitext, i.e., a verbal threshold that mediates between the primary text (the novel) and its potential readership, modifying first reception without belonging to the narrative proper. Functioning at the juncture of platform browsing and literary engagement, the small-format paratextual synopsis links three poles: the book's title, the substantial narrative and plot twists it promises, and the expectations of a digitally distracted audience.

Because it inhabits this threshold, the synopsis becomes a pivotal marketing instrument. Publishers use it for genre signalling, brand voice, pop-cultural triggering, and commercial persuasion (cf. Anisimov 2024, Kavgić & Kavgić 2018, Mahlknecht 2015, 2016, Molodychenko 2020, for analogous findings in film industry paratexts of posters and movie taglines). Yet, the book synopsis does more than advertise. It also frames interpretation by hinting at plot composition and affective tone, thereby guiding readers' sense-making before page one of the novel is even turned. Operating entirely outside the diegesis, it nonetheless exerts a powerful pragmatic impact on how that diegesis will be entered and whether it will be entered at all (a crucial moment for the sales). The book synopsis stands as a key paratextual area where marketing imperatives and narrative framing coalesce through the linguistic mechanisms of semantic compression.

Three discourse traditions merge in this micro-genre of the small-format paratext: literary discourse, marketing discourse, and pop-cultural discourse. The synopsis must remain recognisably tied to the novel's dramatis personae, setting, genre tropes, and inciting incident, but crucially withholding spoilers. It must also function as copywriting, using evaluative superlatives ("*propulsive page-turner*") and calls to action ("*Prepare yourself…*") to push the reader closer to a purchase (Baverstock & Bowen 2019, Gray 2010, Squires 2007). Finally, the synopsis relies on intertextual elements ("*for fans of Knives Out*") to position the book within an existing cultural background, borrowing prestige and instantly communicating genre specifics (Fiske 2010).

Systemic Functional Linguistics (Halliday & Matthiessen 2013) helps clarify this multifunctionality. In small-format synopsis texts, the ideational metafunction packs plot and theme, the interpersonal metafunction commands or intrigues, and the textual metafunction marshals these moves into a scan-friendly layout. Semantic compression is what keeps all three metafunctions operational in under 350 words and makes the synopsis a vital mediator between author, publisher, market, and reader.

3. Materials and methods

To find out how semantic compression is created in contemporary digital book synopses, corpus statistics are combined with functional-linguistic discourse analysis. The study is carried out in three stages: (1) building a genre-controlled corpus of synopses; (2) extracting quantitative patterns in length, vocabulary, grammar, and syntax; and (3) interpreting these patterns through qualitative, function-oriented coding.

3.1. Corpus design

Different literary fields naturally adhere to sharply different paratextual conventions. For instance, a romance synopsis must telegraph chemistry and emotional payoff, whereas an epic-fantasy synopsis must gesture toward world-building, with literary fiction, on the other hand, often foregrounding theme over plot. Mixing all of them in one study would introduce genre noise that could mask the rhetorical devices we wanted to isolate. By holding genre constant, we ensure that recurring linguistic patterns reflect true strategies of semantic compression rather than genre-specific mandates.

The genre of mystery/thriller was selected for the corpus because suspense fiction is structurally obliged to balance revelation and concealment, thus making semantic compression strategies especially visible. The market prominence of this genre and its popularity also guarantee professionally produced synopses rather than hastily written self-descriptions.

Inside the single marketing category "Mystery/Thriller," multiple sub-genres co-exist, e.g., domestic suspense, legal procedural, cozy mystery, spy fiction. This internal range gives the study variation in length and tone while also retaining a shared core of suspense conventions. The corpus thus supplies breadth for quantitative claims without sacrificing the comparability that mixed-genre sampling would undermine.

Amazon's Goodreads, a popular social platform for book readers with retail functionality, has been chosen as a source for the empirical material of the study. The sampling frame was defined by the platform's "Mystery/Thriller" genre category. The sample was then filtered for original English-language publications released between January 2022 and October 2025 to ensure that the synopses reflect current marketing strategies. Titles were sampled from the platform's paginated search results and curated lists (e.g., "New Releases" and "Most Read"). A digital random number generator was applied to the retrieved list to select the final sample; however, the randomization state was not archived. An initial pool of 428 candidate titles was identified. From this pool, texts were excluded if they exceeded the 350-word upper threshold (to prevent contamination by long-form paratexts, e.g., entire book prologues, chapter excerpts, author interviews, or extended editorial reviews; 11 excluded), lacked a designated publisher or author description (61 excluded), or were identified as translations or non-English originals (18 excluded). Only the publisher- or author-supplied "Book description" appearing in the first Goodreads panel was harvested. User reviews and editorial endorsements, which can also be found on the Goodreads website, were excluded. The corpus includes original English publications only, eliminating compression artefacts caused by translation or localization. After automatic de-duplication, which removed 31 records with exact title/author matches or identical text blocks, and manual vetting, which removed 7 further records with incomplete or clearly non-publisher metadata, 300 synopses met all criteria. HTML tags, "Read more" truncators, and any ancillary metadata were stripped. As a result, a plain-text corpus of 56,220 words was compiled.

3.2. Phase 1: quantitative profiling

The corpus was tokenized using AntConc 4.2.4, which was used exclusively for concordance search, frequency profiling, and collocational analysis. Part-of-speech tagging was performed separately with the CLAWS C7 tagger, accessed via the UCREL CLAWS web interface, using the default C7 tagset and no genre-specific parameter adjustments. To verify tagging quality on this specific genre, a random subsample of 30 synopses (10% of the corpus) was manually inspected. Tag errors were corrected before the final counts were extracted.

For the purposes of the study, we defined a "token" strictly as an individual word form. Punctuation marks and special symbols were excluded from the token count. To calculate the type-token ratio (TTR) and lexical density, raw texts were lemmatized and stripped of punctuation using standard Python NLP libraries. As the synopses in the corpus are uniformly short (mean = 187.4 words; range 103-338), standard TTR was treated as a descriptive indicator of lexical diversity rather than a robust comparative metric. TTR was calculated on the aggregated corpus. This value should therefore be interpreted cautiously, as standard TTR is sensitive to corpus size. Alternative length-resistant measures such as MATTR or MTLD were not computed in the present study and are reserved for future work.

The following metrics were gathered: (1) structural, including tokens per synopsis, sentences per synopsis, and mean sentence length; (2) lexical frequency and keyness (calculated against the BNC Sampler comprising approximately 1.000.000 words; log-likelihood scores exceeding 6.63 ($p < .01$) were retained, a Bonferroni correction applied to control for multiple comparisons); (3) POS distribution, active/passive voice ratios, interrogative frequency for grammatical data; and (4) TTR and lexical density to evaluate the information load. For the analysis of the category of voice specifically, voice was coded at the level of finite clauses with verbal predicates capable of active/passive alternation. Agentless passives were counted as passive. Ambiguous middle constructions and non-finite forms were excluded from the voice count.

The obtained data clarified the answer to RQ2 and generated a shortlist of salient devices for manual inspection.

3.3. Phase 2: functional–discursive coding

The qualitative stage targeted three nested scales: macro-compression moves, meso-strategies of semantic compression, and micro-compression devices.

The coding protocol included three steps:

• Step 1. An initial scheme was induced from 30 synopses (10% of the corpus).

• Step 2. A second analyst independently applied the draft codes to the same sample. Inter-coder agreement reached $\kappa = .86$ (Cohen's kappa). Discrepancies were resolved, and the manual was refined to 14 category labels.

• Step 3. The author then coded the remaining 270 synopses. 10% were double-checked, yielding a final $\kappa = .92$. NVivo 14 was used for annotation and frequency counts.

The unit of coding was the synopsis as a whole for the presence of macro-level moves. Meso- and micro-level devices were coded at clause, phrase, or lexeme level where relevant. The initial coding scheme was developed by the author and refined through discussion with a second analyst. Cohen's kappa was calculated as $\kappa = (po - pe)/(1 - pe)$, where po is observed agreement and pe is expected agreement by chance. Category-level kappa values were not reported due to the low frequency of some categories.

To analyze the emotional tone of the lexical composition, a sentiment analysis was also conducted. Sentiment coding was supplementary and descriptive. The unit of analysis was the lexeme in context. Each synopsis was classified according to its dominant functional polarity: negative, positive, or neutral/mixed.

Lexeme polarity, i.e., positive vs. negative, was determined using manual coding (as automated tools would likely mistakenly categorize many domestic thrillers as neutral, with out-of-context words like *dream*, *family*, *home*, or *husband* being typically scored as neutral by standard NLP dictionaries, though in a thriller synopsis these lexemes can be actually used to set up a positive premise as a contrast to the negative threat — the idyllic "facade" that is about to be destroyed). Contextual adjustments were then made for genre-specific tropes. Because this coding was exploratory, separate intercoder reliability for sentiment was not calculated — this is acknowledged as a limitation.

3.4. Mixed-methods integration

Quantitative outputs flagged where semantic compression might reside (e.g., unusually high lexical density, spikes in interrogatives). Qualitative coding then explained how exactly those formal characteristics deliver narrative and commercial value, answering RQ1 and RQ3. Triangulating quantitative data with close reading strengthened internal validity and mitigated the limitations of either approach taken in isolation.

The data were manually collected in October 2025 from publicly accessible Goodreads book pages. Only publisher- or author-supplied promotional descriptions were used. No user accounts, private user data, or non-public information was collected. The full synopses are not redistributed because they are copyrighted promotional texts. However, the metadata and coding framework can be shared upon reasonable request.

The resulting multilevel analysis provides a robust picture of the linguistic mechanisms behind book synopses and, by extension, of semantic compression in the twenty-first-century book marketing.

4. Results

The quantitative analysis of the corpus of 300 synopses of books in the genre of mystery and thriller reveals distinct structural, lexical, grammatical, and syntactic patterns that cooperatively contribute to their function as highly engineered, persuasive small-format texts. Their slightly more expansive format, compared to other types of small-format texts like newspaper headlines or short slogans, allows for different strategies to achieve semantic compression to produce a desired pragmatic effect.

4.1. Length and structural features

As small-format texts, book synopses have to balance the need for narrative setup with the unavoidable constraints of the modern digital attention economy. This characteristic is reflected in their length and structural composition.

The average word count for a synopsis in the corpus is 187.4 words, with a range from 103 to 338 words. It is longer than the average for other types of small-format texts and reflects the synopsis's additional functional burden of establishing genre, character, setting, and a clear inciting incident. 78% of the synopses fall within the 150–250-word range, which points to a stable length range for this platform-specific microgenre.

Virtually all synopses are multi-sentence prose paragraphs. However, a technique of fragmentation and rhythmic variation is achieved through staccato lists or fragments, found in 16% of the corpus. They serve to break the prose rhythm and deliver key information with punchy immediacy, consistent with the stylistic features of slogans and advertising texts, e.g.:

(1) "*A stay-at-home mom with a past. A has-been rock star with a habit. A reality TV producer with a debt. Three disparate lives. One deadly secret.*" ("What Have We Done" by Alex Finlay);
(2) "*Six episodes. One killer*." ("Murder in the Family" by Cara Hunter).

A very common structural formula of digital book synopses in the corpus involves establishing a scene of normalcy or perfection, then followed by a pivotal conjunction (most often "but," "until," "while," or "when") that shatters the previously established facade. This structure mirrors the inciting incident of the novel's plot itself. For example:

(3) "*Childhood sweethearts Nicole and Tom are a normal, loving couple—until a massive lottery win changes their lives overnight*" (from "The Manor House" by Gilly Macmillan);
(4) "*William Wooler is a family man, on the surface. But he's been having an affair, an affair that ended horribly this afternoon at a motel up the road*" (from "Everyone Here Is Lying" by Shari Lapena).

Table 1

Key metrics related to book synopsis structure

Metric	Average	Highest (sub-genre)	Lowest (sub-genre)
Average word count	187.4 words	212.1 words (Procedural/Legal)	168.5 words (Cozy mystery)
Final sentence length (words)	14.8 words	19.3 words (Gothic mystery)	11.2 words (Domestic thriller)
Use of fragments/lists (%)	16%	24.1% (Domestic thriller)	5.8% (Procedural/Legal)

The genre-specific data (Table 1) show that the observed text-creation strategies are deliberate. Procedural and legal thrillers exhibit the highest word count, which may be explained by the need to establish more complex plot mechanics and character relationships. On the other hand, cozy mysteries, which rely on a familiar and comforting formula, seem to require less setup. Domestic thrillers demonstrate the highest use of fragments/lists and the shortest final sentences (a communicative strategy designed to create a sense of breathlessness) and end on a sharp, unsettling hook which mirrors the genre's focus on immediate, more personal peril readers can encounter in their day-to-day lives.

4.2. Syntactic features

The syntactic choices made by the authors of the synopses are tailored to engage the reader directly and frame the narrative as an urgent puzzle.

Interrogative utterances are a dominant feature. They appear in 38.3% of the synopses, almost always as the final sentence.

(5) "*And what happens if they are wrong?*" ("All That Is Mine I Carry With Me: A Novel" by William Landay);
(6) "*Would she?*" ("The Only Survivors" by Megan Miranda).

Questions like (5) or (6) serve the function of creating a hermeneutic gap, thus directly inviting the reader to find the answer by reading the book.

(7) "*Prepare yourself for a thrilling, addictive novel...*" ("The Soulmate" by Sally Hepworth)

Imperative structures are used more sparingly but are still present in approximately 9.3% of small-format texts. Phrases like (7) act as direct commands, presenting the book as a compelling experience to be undertaken.

The functional-linguistic analysis of the grammatical category of voice shows a clear preference for the active voice (72%), as it provides dynamism and clarity. However, the passive voice (28%) is used with deliberate precision to create mystery or emphasize an event over its agent.

(8) "*a body has been found*" ("Mastering the Art of French Murder" by Colleen Cambridge);
(9) "*a baby is stolen*" ("Good Bad Girl" by Alice Feeney).

Constructions like (8) or (9) aim to obscure the perpetrator and put the focus squarely on the crime to enhance the pragmatic effect of mysteriousness.

4.3. Lexical features

The lexical composition of the synopses is highly functional. It prioritizes words carrying a significant thematic, emotional, and pragmatic load.

The analysis of the distribution of parts of speech (Figure 1) confirms a focus on narrative substance. Nouns (32.1%) and verbs (24.8%) form the core of the synopses and establish the key agents and actions of the plot. Adjectives (19.5%) are crucial for establishing tone and atmosphere, with lexemes like "*dark*", "*chilling*", "*deadly*", and "*secret*" appearing frequently in the small-format texts.

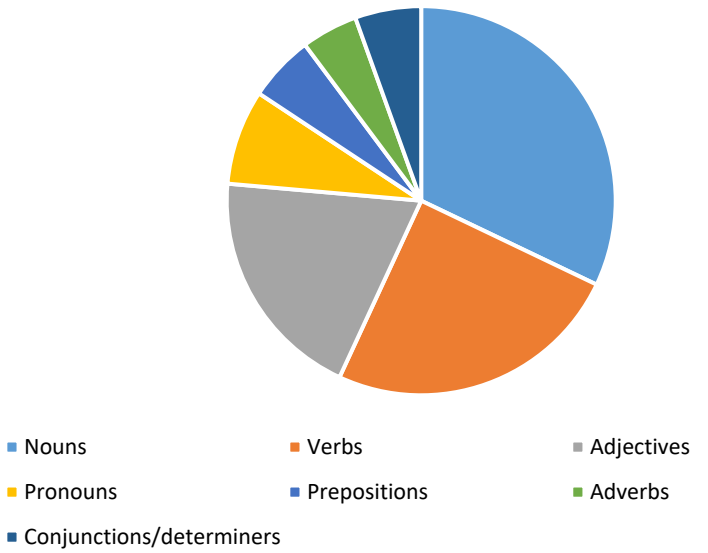


Figure 1. Frequency of word classes in the book synopsis corpus, %

Word frequency analysis of the synopses reveals an arsenal of lexical means centered on the core elements of the genre. The most common content words cluster into clear semantic fields that function as genre signifiers (Table 2).

The dominance of the "Deception/concealment" semantic field establishes the central hermeneutic puzzle for the reader. Its lexemes compress complex themes of trauma, unknowability, betrayal, and moral ambiguity into single, evocative terms that are easy for the reader to perceive and immediately framing the narrative as a mystery to be solved — you only need to buy the book to learn the real truth. This is complemented by the "Violence/peril/death" semantics, with its visceral lexemes of *murder*, *dead*, *blood*, and *killer* establishing the high stakes and physical threat typical of the genre. These words serve a crucial pragmatic function of promising the reader an experience of conflict and mortality. Considering that the reader specifically opened the Goodreads page with a thriller on it, the lexical composition of the synopses can serve as an extra hook and stimulus to find the book itself. The

"Investigation/revelation" semantic field linguistically mirrors the plot's central narrative drive and places the reader alongside the main characters in their search for epistemological certainty.

Table 2

Semantic categories of frequent words in book synopses

Semantic field	Frequent words	Associated themes
Deception/ concealment	Nouns*: secret, lie, past, shadow, mystery, truth, history, facade, deception, web, puzzle, secrets, secrets, secrets (the word "secret" appears in dozens of synopses).* Adjectives*: hidden, dark, unknown, guarded, buried, cryptic, clandestine, sinister, sordid, unspoken, mysterious.* Verbs*: hides, conceals, lying, keeping, unearth, uncover, reveal, haunts.*	Unknowability, history & trauma, facade vs. reality, moral ambiguity, betrayal
Violence/ peril/death	Nouns*: murder, killer, victim, death, crime, danger, threat, body, violence, tragedy, revenge, nightmare, blood.* Adjectives*: deadly, brutal, shocking, devastating, dangerous, chilling, harrowing, bloodthirsty, terrifying.* Verbs*: killed, dead, vanish, disappear, threatens, stalks, haunts, destroy.*	High stakes, overt conflict, mortality, physical threat, the macabre
Urgency/pace	Nouns*: race against time, ticking clock, page-turner, cat-and-mouse game, twists.* Adjectives*: gripping, propulsive, fast-paced, heart-pounding, breathtaking, addictive, edge-of-your-seat, twisty.*	meta-textual function of describing the reading experience itself
Investigation/ revelation	*found, discover, uncover, mystery, question, case, suspect*	The central narrative drive, puzzle-solving, the hermeneutic quest, epistemological doubt, unearthing the buried
Relationships/ family	*family, mother, sister, husband, woman*	The domestic sphere, personal stakes, dysfunctional dynamics, interpersonal trust, gendered roles

Crucially for successful promotion, the "Urgency/pace" semantic field works on an important meta-textual level of discourse. In these small-format secondary texts, lexemes such as *propulsive*, *page-turner*, *gripping*, and *binge-worthy* do not describe the plot per se, but they perform the function of providing direct instructions to the reader on how to consume the primary text of the book. This vocabulary is clearly borrowed from marketing and media discourse. It presents the novel as a fast-paced, highly addictive consumer experience. At the same time, the "Relationships/family" semantic field grounds the abstract threats of terrifying plot events into the personal and much more relatable context of a domestic sphere. Words like *family*, *mother*, *sister*, and *husband* situate the fictional conflict within the intimate space of human connection — a key characteristic of the popular domestic thriller sub-genre. Synergetically, all these semantic fields form a

cohesive discursive functional-pragmatic system which efficiently communicates the genre of the primary text and promises a specific emotional and intellectual experience, thereby persuading potential readers to engage with and, ultimately, purchase the novel.

The emotional tone of the lexical composition (Figure 2) is predominantly negative, as is intrinsic to the genre. Lexemes with negative connotations (e.g., *fear, dead, kill, sinister, secret, lie*) were found in 68.7% of the synopses. Positive lexemes (e.g., *love, perfect, family, dream*) appeared in 23.1% of small-format texts, often used to establish the "perfect" facade which is about to be shattered. Neutral/mixed connotations are found in 8.2% of texts. It is important to note that although such lexemes as *family* or *dream* are semantically neutral when taken in isolation and out of context, in the specific contextual constraints of the domestic-thriller subgenre, they are functionally used with positive polarity specifically to establish the initial impression of normalcy which the plot will subsequently destroy.

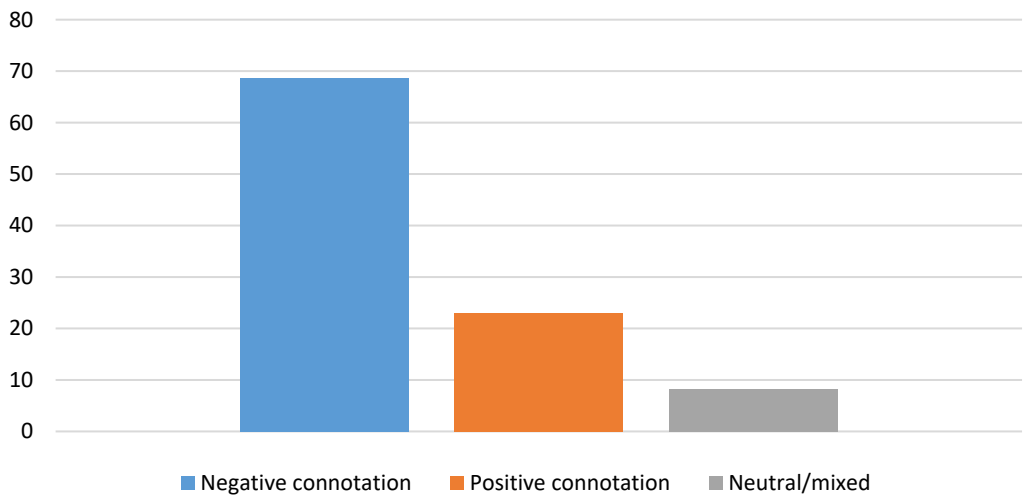


Figure 2. Occurrence of connotative lexemes in book synopses, %

The type-token ratio (TTR), calculated as the number of unique words (types) divided by the total number of words (tokens) in the corpus, is 0.4215. It indicates moderate lexical diversity, appropriate for the longer text format of synopses. Nevertheless, it is still constrained by the need to reuse a core vocabulary of genre-specific terms (*murder, secret, past, dark, dead*) to effectively and repeatedly signal the book's narrative content to potential readers, triggering their associations and expectations from the Mystery/Thriller genre.

The average lexical density of the corpus is 0.52 (or 52%). This finding is logical and supports the theory of format-dependent semantic compression. Due to the nature and positioning between literary fiction and advertising, book synopses must construct coherent narrative paragraphs. This is why they require a higher

proportion of function words (pronouns, determiners, prepositions, conjunctions, articles) to create a grammatical support structure. Nonetheless, a density of 52% still indicates a strong preference for content-rich vocabulary — a fact confirming that semantic compression remains a primary authorial goal. The small-format text is lean and purposeful, from the communicative point of view, with over half its words dedicated to carrying the core thematic and narrative semantics.

4.4. The operators of semantic compression in book synopses

Functional-linguistic analysis of the corpus allows us to identify the most common operators of semantic compression in small-format synopsis texts. They can all be classified according to the textual level at which they function.

4.4.1. Macro-compression moves

The most obvious layer of semantic compression operates at the scale of the whole small-format text. An analysis of the corpus shows that nearly every synopsis is constructed from the same arsenal of stackable modules: inciting-incident hooks (10), character micro-portraits (11), temporal and spatial leaps (12), thematic bundles (13), suspense questions (14), comparative tagging and intertextual elements (15), and metatextual sell lines (16), e.g.:

(10) "*At a busy festival site on a warm spring night, a baby lies alone in her pram, her mother vanishing into the crowds*" ("Exiles" by Jane Harper);
(11) "*Ajay is the watchful servant… Sunny is the playboy heir… And Neda is the curious journalist…*" ("Age of Vice" by Deepti Kapoor);
(12) *"…one bloody night in 1929. …It's now 1983… At seventeen, Lenora Hope hung her sister with a rope…*" ("The Only One Left" by Riley Sager);
(13) "*…double agents, blackmailed CEOs, illegal arms transfers, yachting oligarchs, and more*." ("The Helsinki Affair" by Anna Pitoniak);
(14) "*Was it a tragic accident, or something far more sinister?*" ("The Death of Us" by Lori Rader-Day);
(15) "*The Martian meets 127 Hours in this "astoundingly great" (Gillian Flynn, #1 New York Times bestselling author) and scientifically accurate thriller…*" ("Whalefall" by Daniel Kraus);
(16) "*What Have We Done is both an edge-of-your-seat thriller and a gut-wrenching coming-of-age story*" ("What Have We Done" by Alex Finlay).

Each of these modules (or macro-compression moves) carries a distinct functional-pragmatic load, e.g., hook, orientation, genre coding, or sales pitch. All of them together function as the "zip algorithm" that collapses a full literary narrative into a single screen-length paratext.

Macro-compression in book synopses may be regarded as an architectural template, which is flexible enough to fit any sub-genre and rigid enough to guarantee that each desired persuasive pragmatic effect (e.g., shock, orientation,

intrigue, branding) is achieved in the smallest possible verbal form. A synopsis was classified as following the full macro-structure if it contained all seven macro-compression moves presented in Table 3. The complete structure appears in 86% of the small-format texts in the corpus. The remaining 14% omit only one or two moves presumably due to length constraints.

Table 3

The macro-compression moves

Move	Communicative goal	Typical linguistic realisation
1. Inciting-incident hook	Jolt the reader, stake the crime/conflict in under 25 words	Stand-alone sentence; medial or initial placement
2. Character micro-portrait	Encode key protagonists in under 30 words each	Parallel noun phrases; heavy adjectival front-loading; triadic rhythm
3. Temporal / spatial leap	Compress multi-decade or multi-locale plots	Deictic adverbs (“then”, “now”), ellipsis, past vs. present split paragraphs
4. Thematic bundling	Signal genre and stakes via a list of abstract nouns	Asyndetic or polysyndetic lists, sometimes alliterative
5. Suspense question	Transfer the epistemic burden to the reader	Rhetorical interrogatives, modal verbs
6. Comparative tagging	Borrow semantics from known texts and import ready-made frames from pop culture	“x meets y”, “for fans of…”, intertextual name-drops
7. Metatextual sell line	State reading experience in one breath	Hyperbolic adjectives + evaluative noun (e.g., “thriller”, “saga”)

4.4.2. Meso-level compression strategies

At the sentence level, five recurrent compression techniques operate to allow a several-paragraph small-format secondary text to function both as an effective plot summary of a long-form literary text and a successful sales copy. They are paratactic clause stacking (17), high lexical density (18), em-dash and/or colon insertion (19), cardinality lists (20), and elliptical mini-paragraphs (21), e.g.:

(17) “*Prepare for an education you’ll never forget. A “fiendishly funny” (Booklist) mix of witty wordplay, breathtaking twists and genuine intrigue...*” (“Murder Your Employer” by Rupert Holmes);

(18) “*Compulsively readable, provocative, and disturbing, Penance is a cleverly nuanced, unflinching exploration of gender, class, and power*” (“Penance” by Eliza Clar);

(19) “*Like any enterprising woman, Bea knows what she’s worth and is determined to get all she deserves—it just so happens that what she deserves is to marry rich*.” (“Stone Cold Fox” by Rachel Koller Croft);

(20) “*Ten days, eight suspects, six cities, five authors, three bodies . . . one trip to die for*” (“Every Time I Go on Vacation, Someone Dies” by Catherine Mack);

(21) “*To keep one another safe.*
To hold one another accountable.
Or both.” (“The Only Survivors” by Megan Miranda).

Each of these techniques parcels meaning into tight, highly recoverable discursive units (see Table 4).

Table 4

Sentence-level semantic compression devices

Device	Typical linguistic realisation	Communicative purpose
Paratactic clause stacking	Full stop in lieu of subordination	stringing independent clauses together and avoiding subordination to slow the reader's eye and pad the word count; supplying command, genre, tone, and affect; the paratactic rhythm mimics spoken pitch and anchors attention
High lexical density	Nominalisations, adjective piles	compressing events into single nouns; compressing all the merits of a novel into a single sentence; preloading connotations
Em-dash or colon insertion	Parenthetical span	Punctuation is exploited to graft micro-synopses onto macro-sentences; the colon tends to front-load an inciting incident
Cardinality list	Digit + noun	Signaling stakes
Elliptical mini-paragraph	Isolated textual fragments, often in a vertical column	Dramatisation of the novel's conflict; attention-grabbing

Together, these meso-strategies transform a synopsis into a discursive unit, resembling a high-pressure linguistic chamber, where every punctuation mark, choice of nouns, adjectives, and numerals, as well as a line break and paragraph layout, carries a disproportionate pragma-semantic load. The abstract principle of semantic compression is realized precisely through this set of practical means, ensuring that a five-second scan imparts a sixty-second pitch and, ideally, leads to a purchase.

4.4.3. Micro-level compression devices

Beneath the sentence-level mechanism of semantic compression, there is an even tighter stratum, where single lexemes and micro-constructions are pressed into service to carry discourse functions that would otherwise require clauses or sentences. Seven devices occur with striking regularity across the corpus: modifier clusters (e.g., *heart-pounding thriller; spellbinding; page-turning; binge-worthy; mesmerizing; muscular; distinctive; grab-you-by-both-ears*), genre markers (e.g., *locked-box mystery; gothic suspense; full-fledged spy fiction; bloody slasher*), the so-called early proper nouns (e.g., *Titus Crown is the first Black sheriff*), negative polarity teasers (e.g., "*Nobody ever goes to Hartwood Hall*"; "*Nothing will prepare you for the truth*"), zero-state verbs (e.g., "*A speeding Mercedes jumps the curb*"; "*A baby lies alone in her pram*"; "*Secrets always fester under the surface*"), interrogative zoom-cuts (e.g., "*Did the victim jump? Was she pushed?*"; "*Who took Avery Wooler?*"), and hyphenated metaphors (e.g., *Poison-Ivy League; edge-of-your-seat*). Together, they explain why a small-format synopsis of around 150 words can feel as information-rich as several full pages (see Table 5).

Table 5

Word-level economy in book synopses

Device	Linguistic form	Typical token cost	Functional-pragmatic effect generated
Modifier cluster	Intensifying adjective / stack	1–3	blending positive stance and intensity, promising bodily response
Genre marker	Noun phrase with specifier	2–4	Activating full genre frame; obviating trope listing. By naming the frame, the synopsis spares itself the cost of enumerating the genre conventions
Early proper noun	Capitalised noun phrases	1–3	Pre-empting later referential ambiguity. Once the protagonist is locked in by name, pronouns can be safely used, reducing repetition and saving syntactic re-anchoring to minimize the word count.
Negative-polarity teaser	Quantifier + verb	4–8	Compressing foreboding and curiosity into a small number of tokens; setting a universal expectation, and hinting at the riveting plot
Zero-state verb	Present simple	3–6	Keeping the line lean and casting the scene in vivid, filmic real-time
Interrogative zoom-cut	Wh-/aux inversion	4–6	Implying a dilemma (e.g., suicide vs. murder; kidnapper vs. taken by a parent); projecting the reader into the role of an investigator
Hyphenated metaphor	Compound	1–3	Compressing satire or bodily response into one word, conserving both space and processing time

The micro-level compression devices show how digital-platform paratexts function as meticulously crafted capsules of meaning. Each lexical choice is a multiplier, which expands outward in the reader's mind and cultural background to reclaim the narrative space that the actual physical text must relinquish. This is how semantic compression is realised not only at the level of sentences and paragraphs but also in the very morphology and punctuation of the small-format texts.

4.4.4. Semantic compression by omission

The functional-linguistic analysis of the corpus reveals that the most economical technique, paradoxically, is silence. The authors of book synopses shed every discursive element that does not move units, trusting the reader's imagination and genre literacy to supply what is missing. Three kinds of omission dominate the corpus: (a) sub-plot omissions, as complex romantic triangles and subplot arcs that animate the primary text of the novel, usually vanish from the synopsis; (b) functional anonymity, when details that cannot be cut, get anonymized (e.g., a nameless "*mysterious institute*", "*clandestine college*") standing in for pages of world-building; and (c) hidden resolutions, whereby synopses will never answer their own questions like "*Who killed J.D. Grimthorpe?*" to convert readers' ignorance into purchase motivation.

5. Discussion

5.1. The essence of semantic compression

Based on the analysis of small-format texts, semantic compression can be defined as the systematic reduction of overt linguistic material while preserving (or intentionally guiding) the recovery of the underlying propositional content. It is a meaning-preserving many-to-one mapping, where a rich semantic representation (e.g., the event structure, participant roles, temporal setting, and causal relations of a narrative) is encoded in a shorter expression, whose interpretation depends on the recipients' ability to retrieve the elided information through shared world knowledge, cultural background, genre frames, and pragmatic inference.

Unlike syntactic compression, which shortens the surface structure of discourse (e.g., deleting function words, using contractions), semantic compression operates at the level of conceptual representation. The speaker/writer deliberately chooses high-information lexical anchors ("*playboy heir*"), schematic genre labels ("*gothic suspense*"), intertextual trigger-elements ("*in the style of Daisy Darker and Rock Paper Scissors*"), or culturally familiar scripts ("*road-trip gone wrong*") that activate large, pre-stored cognitive frames in the addressee's memory. As these frames already contain participants, typical sequences of events, potential plot twists, and evaluative stances, the small-format secondary text can safely omit them. For example, instead of expanding precious textual space to describe a classical Agatha Christie-style murder that happened in a closed physical setting with a limited pool of potential suspects, a single lexical anchor "*locked-room mystery*" activates a vast pre-existing cultural frame in the memory of an experienced reader, who is familiar with popular genre tropes, thus achieving maximum narrative orientation with minimal token expenditure. Participating in such discourse, the listener/reader is meant to reconstruct the missing details through pragmatic enrichment (Relevance Theory), frame-based inference (Fillmore 2006), and genre schemata (Swales 1990).

It is important to emphasize that semantic and syntactic kinds of compression are by no means mutually exclusive. On the contrary, they are highly integrated and function synergistically. For example, syntactic reduction (e.g., the omission of conjunctions or the use of paratactic phrasing) serves as the necessary structural mechanism that enables semantic compression to manifest effectively under strict word-count limitations. For the purposes of this study, the distinction between semantic and syntactic compression is treated heuristically. Semantic compression is identified through frame-activating lexical and discursive units. Syntactic compression is associated with structural reduction. The formal quantification of this ratio remains a task for future research.

From an information-theoretic perspective, semantic compression increases the ratio of conveyed meaning to the length of the textual message by exploiting redundancy in the shared common ground. In discourse, it manifests in small-format texts. For example, the synopsis of "All the Sinners Bleed" by S.A. Cosby:

(22) "*A Black sheriff. A serial killer. A small town ready to combust*."

Three noun phrases supply setting, protagonist, villain, and central conflict. Everything else is compressed into the culturally shared "small-town crime thriller" genre frame.

Research on semantic compression explains how speakers balance informativity against processing effort, sheds light on genre-specific micro-conventions (e.g., list syntax, rhetorical questions), and provides empirical ground for modelling the role of stored schematic knowledge in real-time comprehension.

Small-format texts (whether headlines, tweets, advertising taglines, slogans, academic abstracts, or any other type) are purposefully created for reading environments where space is scarce and attention fleeting. In such circumstances, semantic compression is not just a stylistic choice. It is the mechanism that lets a message survive the "squeeze." By trading many vital details for inference and substituting expansive descriptions with frame-activating discursive elements (e.g., "*road-trip gone wrong*," "*locked-room mystery*"), the author of the small-format text maximizes the ratio of meaning conveyed to the number of characters used. Such economy has two immediate advantages. First, it lowers the cognitive load as the reader who scans the text on a phone screen can grasp the gist in a single fixation. Second, it lowers the production cost because the writer can get a complex proposition across without exceeding the limits imposed by a template or layout (e.g., the 140 characters visible in many notification previews).

The online book synopses, as a special genre of small-format texts, distil these constraints to their purest form. A browsing user (a potential buyer and reader of the novel) spends mere seconds on a title before deciding whether to click "Want to Read." The website displays only the first ±200 characters of any blurb or synopsis before a "More" break. It means that the copy must (a) signal the genre for recommendation algorithms as well as for human sorting, (b) ignite curiosity strong enough to survive countless distractions, and (c) incorporate enough thematic and tonal markers to fulfil the publishers' marketing goals. Semantic compression makes all these three tasks compatible. A phrase such as "*An Agatha Christie-style locked-box thriller set on a storm-lashed island*" activates users' background knowledge and triggers an entire library of reader expectations (e.g., setting, cast size, narrative tempo, possible plot twists) without explicitly nominating any of them. The synopses, therefore, fit comfortably inside the digital platform's visual limits and tell prospective readers everything they need to know to make a choice to buy the book: what kind of journey this will be, why it feels fresh, how it correlates with the previously read pieces of fiction, and what pleasure it promises.

In short, semantic compression is the core functional-pragmatic cognitive-discursive mechanism of small-format paratextual discourse, and book synopses are its showcase: a few high-density lines that captivate, persuade, inform, and brand — all before the user's thumb finishes its next swipe.

Considering the results of the quantitative and qualitative analysis of the main linguistic features of small-format texts, the data point toward a single organising principle that underlies the synopsis: every structural, syntactic, grammatical, and lexical decision of the author functions as a miniature act of semantic compression. The averages we have just examined (i.e., the narrow 150-to-250-word band, the fragmentary list inserts, the preference for interrogative closure, the disciplined recycling of a core vocabulary of secrets, lies, murders, and revelations) are not just random stylistic traits. They are the visible traces of an overarching pressure to fit a novel-length narrative as well as an advertising pitch into a constrained space that will fit on a phone screen and be absorbed in a five-second scroll.

Semantic compression begins with length. A small-format text of roughly 187 words must achieve what the whole novel does in 150,000: establish setting, sketch characters, announce genre, ignite suspense, compel the reader. Such an immense functional-pragmatic load explains the rhythmic alternation of full sentences and staccato fragments. The former supply cohesion, whereas the latter inject important information bursts which would otherwise occupy numerous paragraphs in the primary text. Syntax follows suit. An interrogative final line is both a cliff-hanger and an invitation to immediately buy the book to learn more. For instance, a two-word question such as "*Would she?*" compresses an entire web of narrative possibilities into nine characters. Even voice alternations serve a similar communicative purpose: passive constructions like "*a body has been found*" instantly foreground the crime, at the same time erasing the agent and telescoping plot and mystery into just a single clause.

The obtained lexical data reinforce this pattern. High-frequency words in the "Deception/concealment" and "Violence/peril/death" semantic fields play a pivotal role in allowing the synopses to forgo lengthy exposition. Lexemes "*secret*," "*killer*," and "*dark*" each trigger larger thriller schemata stored in the reader's memory and cultural background. Marketing-inflected adjectives (e.g., "*propulsive*," "*binge-worthy*") compress a promised reading experience of excitement and addictiveness into a handful of syllables. Even the mildly elevated TTR and the 52% lexical density demonstrate a small-format-related delicate balance between necessary grammatical means and semantically rich vocabulary, ensuring that every function word earns its keep and every content word contributes to a larger narrative frame.

In other words, the synopsis is a clear example of how language can be densified until each token fires off a chain of inferences in the reader's mind. The following subsection, therefore, turns from quantitative surfaces to the functional-pragmatic cognitive-discursive mechanism itself to examine in detail how semantic compression operates as the hidden engine that makes the paratexts in question not just intelligible and memorable but also simultaneously persuasive.

5.2. Semantic compression as a discourse-level phenomenon

The primary text of the novel (literary fiction), its synopsis, which blends marketing, advertising and storytelling in a paratextual small format, and the wider pop-cultural mediascape constitute a unique functional-pragmatic discursive space (see Figure 3), where semantic compression holds a significant place.

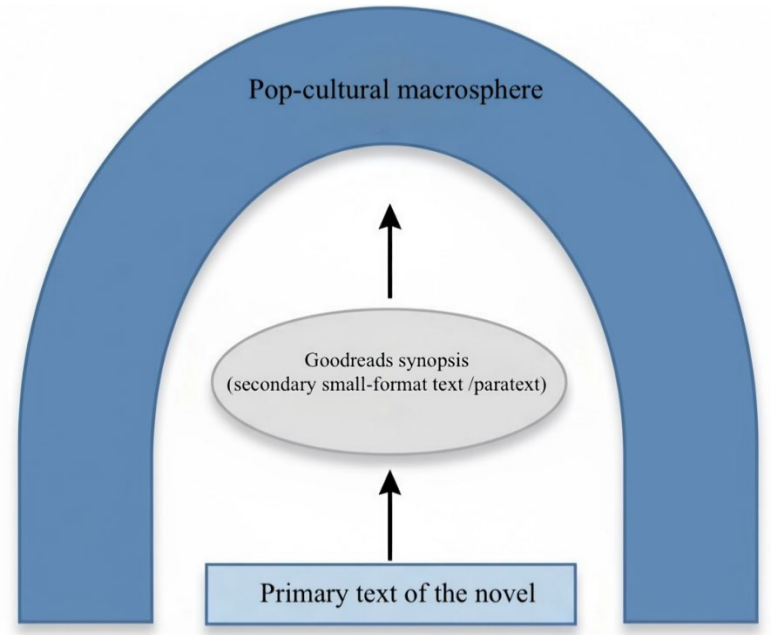


Figure 3. The macro-discursive space of the book synopsis

Although semantic compression occupies the middle tier (see Figure 3), it is meaningful only because it reaches simultaneously downward to the primary text of the novel for narrative raw material and upward to the pop-cultural macrosphere for the shared cognitive frames that let that material be miniaturized.

The synopses must encode the novel's irreducible specifics, i.e., its inciting crime, its distinctive protagonists and villains, its emotional resonance, and its ethical stakes. These elements are non-negotiable and can't be omitted, otherwise the synopsis ceases to be about a particular book. Semantic compression, therefore, acts as a reducing valve: it siphons just enough propositional content from a hundreds-of-thousands-word manuscript to preserve recognizability, then strips away everything non-essential (sub-plots, minor settings, secondary villains, delayed twists).

The reader who encounters the synopsis is supposed to be already fluent in genre tropes and memes, cinematic clichés, streaming-era buzzwords, and popular writers. Compression relies on that fluency to expand micro-elements into macro-semantics. For instance, the phrase "*locked-box mystery*" instantly invokes Agatha Christie and the Knives Out movie franchise; "*propulsive page-turner*" taps a decade of jacket copy telling the reader what a thriller is supposed to feel like. These

ready-made schemas reside in the pop-cultural layer of the discursive space, and the small-format text of the synopsis merely activates them.

Because the synopsis is the only part of the discursive space that must fulfil both functions at once, i.e., be faithful to the book and at the same time be legible and enticing inside a saturated media feed, semantic compression functions primarily in the paratextual zone. It is the synopsis's operative principle to (a) extract core narrative elements from the primary text, (b) then match each extracted element with a high-density sign from pop culture (e.g., a genre label, a buzz adjective, a reference title, a writer's branded name), and finally (c) release the hybrid signal in under 350 words.

All in all, semantic compression is not an optional feature added by marketers. It is an important cognitive-discursive phenomenon that allows the book synopsis to function at the junction of literature and commerce, translating the deep semantics of the novel into the shallow and quickly processed code of platform reading by exploiting the cultural capital stored one level above and the narrative capital stored one level below. The result is a miniature discourse which is simultaneously derivative, autonomous, and, from the viewpoint of marketing and the contemporary book economy in the digital epoch, indispensable.

6. Conclusion

This study set out to explain the commercially significant persuasive pragmatic effect of semantic compression which is meant to assist in book sales and to do so through a functional-linguistic analysis of digital book synopses as prototypical small-format texts. The findings demonstrate that these unique paratexts are not a mere loose blend of book summary and advertising but a complex, multi-tiered pragma-semantic system for converting narrative material into potential market value.

The study offers a replicable descriptive model for analyzing persuasive-oriented paratextual features in digital epitexts, enriching paratext and small-format text theories. It provides empirical evidence for relevance-theoretic pragmatics, showing how radical underspecification maximizes cognitive and fiscal effects through cultural frames. Finally, it frames suspense fiction as a high-contrast space that exposes semantic compression strategies in genre linguistics. Regarding the practical implications, the three-tier coding scheme may assist publishers and copywriters in drafting more conventionally effective promotional synopses, although actual effects on sales or click-through behaviour were not empirically tested. Additionally, these parameters can refine automated systems to generate human-competitive copy.

The limitations of this study include a lack of empirical sales data, leaving the causal link between semantic compression and commercial success theoretical. Furthermore, while the macro-framework is universal, the specific lexical triggers identified are optimized for Mystery/Thrillers; applying them to other genres via automated generation could be ineffective or even detrimental. Future work should

test whether identical compression strategies apply in the synopses for books of other genres, like romance or comedy, or in translated paratexts where cultural frames unavoidably differ. Experiments with eye-tracking and click-through metrics could correlate specific semantic compression devices with measurable sales upticks, further strengthening our understanding of the interdependence between linguistic form and market performance.

References

Aijmer, Karin. 2005. Evaluation and pragmatic markers. *Strategies in Academic Discourse*. 83–96. Amsterdam: John Benjamins.

Anisimov, Vladislav E. 2024. The external hierarchy of French film discourse. *Professional Discourse & Communication* 6 (4). 34–64. https://doi.org/10.24833/2687-0126-2024-6-4-34-64 (In Russ.).

Ayers, Gael. 2008. The evolutionary nature of genre: An investigation of the short texts accompanying research articles in the scientific journal *Nature*. *English for Specific Purposes* 27 (1). 22–41. https://doi.org/10.1016/j.esp.2007.06.002

Batchelor, Kathryn. 2021. Translation, media and paratexts. *The Routledge Handbook of Translation and Media*. 122–135. London: Routledge.

Baverstock, Alison & Susannah Bowen. 2019. *How to Market Books*. London: Routledge.

Boginskaya, Olga A. 2022. Functional categories of hedges: A diachronic study of Russian research article abstracts. *Russian Journal of Linguistics* 26 (3). 645–667. https://doi.org/10.22363/2687-0088-30017

Carr, Nicholas. 2020. *The Shallows: What the Internet Is Doing to Our Brains*. New York: W.W. Norton & Company.

Cherkunova, Marina V. 2021. Means of semantic compression in modern English scientific discourse (based on abstracts to the articles from international scientific citation databases). *Professional Discourse & Communication* 3 (3). 28–38. https://doi.org/10.24833/2687-0126-2021-3-3-28-38

Cherkunova, Marina V. & Eugenia V. Ponomarenko. 2021. Mini-texts of English web-literature in the context of contemporary digital discourse: System-dynamic approach. *Vestnik of Samara University. History, Pedagogics, Philology* 27 (4). 168–175. https://doi.org/10.18287/2542-0445-2021-27-4-168-175

Fillmore, Charles J. 2006. Frame semantics. *Cognitive linguistics: Basic readings*. 373–400. Berlin: Mouton de Gruyter.

Fiske, John. 2010. *Understanding Popular Culture*. London: Routledge.

Genette, Gérard. 1997. *Paratexts: Thresholds of Interpretation*. Cambridge: Cambridge University Press.

Gray, Jonathan. 2010. *Show Sold Separately: Promos, Spoilers, and Other Media Paratexts*. New York: New York University Press.

Halliday, Michael A. K. & Christian M. I. M. Matthiessen. 2013. *Halliday's Introduction to Functional Grammar*. 4th edn. London: Routledge.

Johnson, Peter D. Jr., Greg A. Harris & Darrel C. Hankerson. 2003. *Introduction to Information Theory and Data Compression*. 2nd edn. Boca Raton: Chapman and Hall/CRC. https://doi.org/10.1201/9781420035278

Kavgić, Olga Vladimir Panić & Aleksandar Kavgić. 2018. Creativity in film taglines: Extra linguistic, textual, and linguistic analysis. *Annual Review of the Faculty of Philosophy, Novi Sad* 43 (1). 101–125. https://doi.org/10.19090/gff.2018.1.101-125

Kharkovskaya, Antonina A. & Marina V. Cherkunova. 2020. Minitexts in American media-discourse (based on the titles of Youtube videos featuring Donald Trump). *Current Issues in*

Philology and Pedagogical Linguistics 4. 44–51. https://doi.org/10.29025/2079-6021-2020-4-44-51

Khramchenko, Dmitry S. 2023. How headlines communicate: A functional-pragmatic analysis of small-format texts in English-language mass media. *Training, Language and Culture* 7 (2). 30–38. https://doi.org/10.22363/2521-442X-2023-7-2-30-38

Kubryakova, Elena S. 2001. O tekste i kriteriyakh ego opredeleniya (On text and criteria for its definition). *Text. Structure and Semantics* vol. 1. 72–81. Moscow: MGU. http://www.philology.ru/linguistics1/kubryakova-01.html (In Russ.).

Lee, Heejun & Chang-Hoan Cho. 2019. Digital advertising: Present and future prospects. *International Journal of Advertising* 39 (3). 332–341. https://doi.org/10.1080/02650487.2019.1642015

Mahlknecht, Johannes. 2015. Three words to tell a story: The movie poster tagline. *Word & Image* 31 (4). 414–424.

Mahlknecht, Johannes. 2016. *Writing on the Edge: Paratexts in Narrative Cinema*. Heidelberg: Universitätsverlag Winter.

Miller, Carolyn R. & Dawn Shepherd. 2009. Questions for genre theory from the blogosphere. *Genres in the Internet*. 263–290. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/pbns.188.11mil

Molodychenko, Evgeni N. 2020. Metasemiotic projects and lifestyle media: Formulating commodities as resources for identity enactment. *Russian Journal of Linguistics* 24 (1). 117–136. https://doi.org/10.22363/2687-0088-2020-24-1-117-136

Nord, Christiane. 2019. Paving the way to the text: Forms and functions of book titles in translation. *Russian Journal of Linguistics* 23 (2). 328–343. https://doi.org/10.22363/2312-9182-2019-23-2-328-343

Page, Ruth E. 2013. *Stories and Social Media: Identities and Interaction*. London: Routledge.

Poulimenou, Sylvia, Sofia Stamou, Sozon Papavlasopoulos & Marios Poulos. 2016. Short text coherence hypothesis. *Journal of Quantitative Linguistics* 23. 191–210. https://doi.org/10.1080/09296174.2016.1142328

Remchukova, Elena N. & Anna A. Apostolidi. 2018. Small-format 'news flow' media texts in the focus of linguistics: Genre and language peculiarities. *RUDN Journal of Language Studies, Semiotics and Semantics* 9 (3). 651–668. https://doi.org/10.22363/2313-2299-2018-9-3-651-668

Sdobnikov, Vadim V. 2021. New tasks of translation teachers: The challenge of mosaic thinking. *Professional Discourse & Communication* 3 (2). 43–54. https://doi.org/10.24833/2687-0126-2021-3-2-43-54.

Skare, Roswitha. 2021. The paratext of digital documents. *Journal of Documentation* 77 (2). 449–460.

Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. 2nd edn. Oxford: Blackwell Publishing.

Squires, Claire. 2007. *Marketing Literature: The Making of Contemporary Writing in Britain*. Basingstoke: Palgrave Macmillan.

Starostina, Julia S., Marina V. Cherkunova & Antonina A. Kharkovskaya. 2021. Education and city in the English-language small-format texts: Axiological approach. *SHS Web of Conferences* 98. Article 04013. https://doi.org/10.1051/SHSCONF/20219804013

Swales, John M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Wilson, Deirdre & Dan Sperber. 2002. Relevance theory. *The Handbook of Pragmatics*. Oxford: Blackwell.

Zappavigna, Michele. 2017. Twitter. In Christian Hoffmann & Wolfram Bublitz (eds.), *Pragmatics of social media*, 201–224. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110431070-008

Zibin, Aseel, Abdel Rahman Altakhaineh & Marwan Jarrah. 2024. Compound nouns as linguistic framing devices in Arabic news headlines in the context of the Israel-Gaza conflict. *Russian Journal of Linguistics* 28 (3). 535–558. https://doi.org/10.22363/2687-0088-39562

Article history:
Received: 28 October 2025
Accepted: 15 April 2026

Bionote:

Dmitry S. KHRAMCHENKO is Dr. Habil., Professor in the Department of English Language No. 4 at MGIMO University, Moscow, Russia. His research interests cover functional linguistics, pragmatics, discourse studies, business rhetoric and corpus linguistics.
e-mail: d.khramchenko@inno.mgimo.ru
https://orcid.org/0000-0003-3038-8459

Сведения об авторе:

Дмитрий Сергеевич ХРАМЧЕНКО — доктор филологических наук, профессор кафедры английского языка № 4 Московского государственного института международных отношений (Университета) Министерства иностранных дел России. Сфера научных интересов: функциональная лингвистика, лингвопрагматика, теория дискурса, риторика делового и профессионального общения, корпусная лингвистика.
e-mail: d.khramchenko@inno.mgimo.ru
https://orcid.org/0000-0003-3038-8459