

**STUDIES ON ELLIPSIS IN SPOKEN VS. WRITTEN DISCOURSE.**

Avazov Khamza Panjievich
Teacher of Karshi State University

<https://doi.org/10.5281/zenodo.20050665>

Abstract: ellipsis — the principled omission of linguistically recoverable elements — has been extensively documented as a key cohesive mechanism in English discourse. Yet the ways in which its frequency, type, and function differ across spoken and written modes remain incompletely understood, particularly in relation to the interplay between register, communicative purpose, and grammatical licensing. This article presents a comparative corpus-based study of ellipsis across four registers: spontaneous spoken conversation, scripted broadcast speech, literary prose fiction, and published academic writing. Drawing on a 500,000-word annotated corpus, the study analyses the distribution of VP ellipsis, clausal ellipsis, nominal ellipsis, gapping, sluicing, and stripping across these modes, and investigates the discourse functions each type serves within its host register. The results reveal that ellipsis is substantially more frequent and structurally diverse in spoken registers, where it fulfils interactional and processing functions, while written registers deploy it selectively for stylistic economy and information focus. These findings carry significant implications for discourse theory, corpus linguistics, and the teaching of both spoken and written English.

Keywords: ellipsis, spoken discourse, written discourse, corpus linguistics, register variation, cohesion, VP ellipsis, clausal ellipsis, discourse analysis, applied linguistics

NUTQIY VA YOZMA NUTQDA ELLIPSIS BO‘YICHA TADQIQOTLAR.

Avazov Khamza Panjiyevich
Qarshi davlat universiteti o‘qituvchisi

Annotatsiya: Ellipsis — lingvistik jihatdan tiklanadigan unsurlarni asosli tushirish — ingliz nutqidagi asosiy bog'lovchi mexanizm sifatida keng hujjatlashtirilgan. Biroq uning chastotasi, turi va funksiyasining nutqiy va yozma uslublarda qanday farq qilishi hali to'liq tushunilmagan, ayniqsa registr, kommunikativ maqsad va grammatik ruxsat berish o'rtasidagi o'zaro ta'sir nuqtai nazaridan. Ushbu maqola to'rtta registr bo'yicha ellipsisning qiyosiy korpusga asoslangan tadqiqotini taqdim etadi: o'z-o'zidan nutqiy suhbat, ssenariyga asoslangan efir nutqi, adabiy nasriy fantastika va nashr etilgan akademik yozuv. 500,000 so'zli annotatsiyalangan korpusga asoslanib, tadqiqot VP ellipsis, klausul ellipsis, nominal ellipsis, gapping, sluicing va strippingning ushbu uslublardagi tarqalishini tahlil qiladi va har bir tur registr ichida qaysi nutq funksiyalarini bajarishini o'rganadi. Natijalar shuni ko'rsatadiki, ellipsis nutqiy registrlarda sezilarli darajada ko'proq chastotaga ega va strukturaviy jihatdan xilma-xil bo'lib, u yerda o'zaro ta'sir va qayta ishlash funksiyalarini bajaradi, yozma registrlarda esa stilistik tejamkorlik va ma'lumotga e'tibor uchun tanlab qo'llaniladi. Bu natijalar nutq nazariyasi, korpus lingvistikasi va nutqiy hamda yozma ingliz tilini o'qitish uchun muhim ahamiyatga ega.

Kalit so'zlar: ellipsis, nutqiy nutq, yozma nutq, korpus lingvistikasi, registr o'zgarishi, bog'lanish, VP ellipsis, klausul ellipsis, nutq tahlili, amaliy lingvistika.

ИССЛЕДОВАНИЯ ЭЛЛИПСИСА В УСТНОЙ И ПИСЬМЕННОЙ РЕЧИ.

Авазов Хамза Панжиевич

Преподаватель Каршинского государственного университета

Аннотация: Эллипсис — принципиальное опускание лингвистически восстанавливаемых элементов — широко задокументирован как ключевой когезивный механизм в английском дискурсе. Однако способы, которыми различаются его частотность, тип и функции в устных и письменных модусах, остаются недостаточно понятными, особенно в отношении взаимодействия между регистром, коммуникативной целью и грамматическим лицензированием. Данная статья представляет сравнительное корпусное исследование эллипсиса по четырем регистрам: спонтанному устному разговору, сценаризованной вещательной речи, литературной художественной прозе и опубликованным академическим текстам. Опираясь на аннотированный корпус объемом 500 000 слов, исследование анализирует распределение VP-эллипсиса, клаузуального эллипсиса, номинального эллипсиса, гэппинга, слуайсинга и стриппинга в этих модусах, а также изучает дискурсивные функции, которые каждый тип выполняет в своем регистре. Результаты показывают, что эллипсис существенно чаще встречается и структурно разнообразнее в устных регистрах, где он выполняет интеракционные и обработочные функции, в то время как письменные регистры используют его избирательно для стилистической экономии и фокусировки информации. Эти выводы имеют важное значение для теории дискурса, корпусной лингвистики и преподавания устного и письменного английского языка.

Ключевые слова: эллипсис, устный дискурс, письменный дискурс, корпусная лингвистика, вариация регистров, когезия, VP-эллипсис, клаузуальный эллипсис, дискурсивный анализ, прикладная лингвистика

Introduction

Language does not operate identically across all modes of communication. The spoken word and the written text are distinct semiotic systems shaped by different conditions of production, different audience relationships, and different cognitive demands. Among the many grammatical features that are distributed unevenly across these modes, ellipsis — the omission of elements whose content can be recovered from context — occupies a particularly revealing position. It sits at the intersection of syntax, pragmatics, and discourse structure, and its distribution therefore reflects both the formal properties of the mode and the communicative needs it serves.

The contrast between spoken and written ellipsis is not merely quantitative. Spoken conversation operates under real-time processing pressures, within frameworks of shared situational knowledge, and across adjacency pairs whose tight structural coupling allows omission to proceed almost automatically. Written discourse, by contrast, is produced and processed in conditions of displacement and relative anonymity, placing a premium on explicitness and textual self-sufficiency. These different conditions predict not only that ellipsis will be more frequent in speech, but that its types and functions will differ systematically across modes.

Despite a rich theoretical literature on ellipsis (Merchant, 2001; Lobeck, 1995; van Craenenbroeck & Merchant, 2013) and an established tradition of corpus-based register research (Biber, 1988; Biber et al., 1999), comparative studies that specifically examine ellipsis

across the spoken–written continuum using large annotated corpora remain relatively sparse. This study addresses that gap by providing a systematic, corpus-driven comparison of ellipsis across four registers that span the spoken–written mode distinction. The research questions guiding the study are: (1) How does the frequency and type of ellipsis vary across spoken and written registers? (2) What discourse functions does ellipsis serve in each register, and how do these differ? (3) What theoretical and pedagogical conclusions can be drawn from these cross-register patterns?

Literature Review

Ellipsis has been a central topic in syntactic theory since at least the mid-twentieth century. Within the generative tradition, the key analytical question has been whether ellipsis involves phonological deletion of a syntactically present but phonetically null structure, or whether the elided position contains no syntactic content at all (Merchant, 2001; Lobeck, 1995). The former position, associated with the ellipsis-as-deletion approach, holds that the surface gap corresponds to a full syntactic representation that is simply not pronounced; the latter, associated with pro-form or null-anaphora approaches, treats the gap as semantically interpreted but syntactically empty.

What is not in dispute across theoretical approaches is the recoverability condition: the omitted material must be uniquely identifiable from linguistic or situational context (Hankamer & Sag, 1976). This condition, variously formalised as e-givenness (Merchant, 2001), discourse-linking, or semantic identity, ensures that ellipsis cannot produce genuine ambiguity about propositional content. The major structural types — VP ellipsis, NP ellipsis, clausal ellipsis, gapping, sluicing, and stripping — differ in what can be omitted, what licensing heads are required, and how the antecedent is identified.

Register and the Spoken–Written Distinction

Biber's (1988) seminal multi-dimensional analysis of register variation established the empirical foundation for understanding how grammatical features are distributed across modes. Biber identified a dimension of "involved versus informational production" on which spoken conversation and academic writing sit at opposite poles, with spoken conversation characterised by high interactiveness, real-time processing, and reliance on shared context, and academic writing by high informational density, explicitness, and hierarchical syntactic integration.

Biber et al. (1999) extended this analysis in the Longman Grammar of Spoken and Written English, documenting the dramatically higher frequency of ellipsis in conversation compared with academic prose and news reportage. Carter and McCarthy (2006), drawing on the CANCODE corpus of naturally occurring spoken English, argued that many features of spoken grammar — including ellipsis, deixis, and vague language — are systematically absent from traditional pedagogical grammars, leaving learners unprepared for authentic spoken interaction. Their work established the concept of "spoken grammar" as a distinct grammatical system rather than a deficient version of written grammar.

Discourse-Functional Studies of Ellipsis

Beyond its grammatical properties, ellipsis has been studied as a discourse-functional phenomenon. Halliday and Hasan (1976) classified ellipsis alongside reference, substitution, conjunction, and lexical cohesion as a primary cohesive device, noting that it creates textual coherence by establishing presuppositional ties between clauses. Nominal ellipsis creates cohesion by omitting a noun phrase whose referent is already active in the discourse; verbal and clausal ellipsis create it by omitting predicate or sentential material tracked from a preceding or following utterance.

More recent work in interactional linguistics (Couper-Kuhlen & Selting, 2018) has emphasised the action-sequential dimension of ellipsis in talk: in adjacency pair sequences (question–answer, invitation–acceptance, offer–response), the second pair part is frequently elliptical because its propositional content is entirely determined by the first pair part. From this perspective, ellipsis is not just a cohesive device but a resource for interactional alignment and recipient design.

Studies examining written ellipsis have focused on its stylistic functions in literary prose and journalism (Leech & Short, 2007) and on its role in signalling information structure in academic argument (Swales, 1990). These accounts converge on the view that written ellipsis is a deliberate stylistic choice exploiting the reader's ability to recover antecedents from the immediately preceding text, and that it is most frequent in contexts where parallel or contrastive information structure creates recoverable omissions naturally.

Research Methodology

The study is based on a 500,000-word corpus comprising four equal sub corpora of 125,000 words each. The four registers were selected to represent points along the spoken–written continuum: (1) spontaneous spoken conversation, drawn from the British National Corpus (BNC) demographic sub corpus; (2) scripted broadcast speech, drawn from BBC radio and television transcripts in the BNC; (3) literary prose fiction, comprising contemporary published novels; and (4) published academic writing, drawn from peer-reviewed linguistics and education journals. This design allows comparison not only between prototypically spoken and written registers, but also between the two intermediate registers of scripted speech and literary prose, which share features of both modes.

Annotation Procedure

Elliptical constructions were identified and annotated through a two-stage procedure. In the first stage, automated part-of-speech tagging using the CLAWS7 tagset was combined with a purpose-built regex parser to flag likely ellipsis sites — specifically, auxiliary verbs without following lexical verb phrases, bare NPs in contexts licensing nominal ellipsis, and coordination structures with apparent gapping. In the second stage, a trained human annotator reviewed all flagged instances and a 20% random sample of non-flagged text to identify any instances missed by the parser. Each confirmed ellipsis instance was annotated for: type (VP ellipsis, nominal ellipsis, clausal ellipsis, gapping, sluicing, stripping), directionality (anaphoric or cataphoric), antecedent distance (within clause, within sentence, across sentences), and primary discourse function (cohesion, information focus, interactional efficiency, stylistic economy, register marking).

Inter-rater reliability was assessed using a second annotator who independently coded a 10% subsample. Cohen's kappa for ellipsis-type classification was 0.86, and for discourse-function classification was 0.79, both indicating strong agreement. Disagreements were resolved through discussion and adjudication against the theoretical criteria established in the annotation manual.

Analysis and Results

The aggregate frequency counts confirm the predicted mode difference with considerable clarity. Spontaneous conversation yielded 97.3 ellipsis instances per 10,000 words — the highest of all four registers. Broadcast speech showed a markedly lower rate of 61.8 per 10,000 words, reflecting the more planned and formal character of scripted speech despite its spoken mode. Literary fiction produced 48.4 instances per 10,000 words, while academic prose showed the lowest frequency at 19.7 per 10,000 words — approximately one-fifth the rate of spontaneous conversation.

These figures confirm and extend the findings of Biber et al. (1999) and Carter and McCarthy (2006), situating ellipsis firmly as a feature of involved, context-dependent, interactionally dense registers. The steep gradient from conversation to academic prose corresponds directly to the shift from high shared context and real-time processing pressure to low shared context and conditions favouring explicitness.

Type Distribution Across Registers

The distribution of ellipsis types varied substantially across registers. VP ellipsis was dominant across all four, accounting for 41% of all instances in conversation, 37% in broadcast speech, 31% in fiction, and 44% in academic prose. Its prevalence in academic prose despite the overall low ellipsis frequency reflects the register's heavy use of comparative and contrastive structures ("Experiment A produced stronger effects than Experiment B did"), which license VP ellipsis almost automatically.

Clausal ellipsis — the omission of entire subject-verb sequences — was most frequent in conversation (29% of conversational ellipsis instances) and fiction (24%), and much rarer in academic prose (8%). This reflects the central role of clausal ellipsis in spoken adjacency pair responses and in narrative dialogue. The following attested examples illustrate its operation across registers:

(1) [Conversation] A: "Have you eaten yet?" / B: "No, not yet." [Clausal ellipsis: full response would be "No, I have not eaten yet"]

(2) [Fiction] She wanted to leave immediately. Could not, of course. [Clausal ellipsis: subject-auxiliary omitted]

(3) [Academic] Results were positive across all three conditions, though not uniformly so. [Clausal ellipsis in concessive clause]

Gapping — the omission of a repeated verb in coordinate structures — was proportionally highest in academic prose (18%) and fiction (17%), registers in which parallel syntactic

structures are a common rhetorical device. Sluicing (e.g., "She left, though I don't know why") was most frequent in fiction (14%), consistent with the register's use of free indirect thought and narratorial uncertainty. Nominal ellipsis and stripping were relatively infrequent across all registers.

Discourse Functions

Qualitative analysis of a stratified 250-instance sample (62–63 instances per register) revealed clear register-specific functional profiles. In conversation, the dominant function was interactional efficiency (48% of instances): ellipsis accelerated turn-taking, reduced processing load, and projected responsive alignment between interlocutors. Cohesion maintenance was secondary (31%), reflecting the tight clausal linkage between adjacency pair parts.

In broadcast speech, cohesion maintenance was most prominent (44%), as scripted speech relies on ellipsis primarily to avoid lexical repetition across sentences while maintaining the appearance of spontaneity. In fiction, stylistic economy and information focus were equally important (34% and 33% respectively), consistent with literary prose's use of ellipsis to create rhythmic compression and to foreground new information by omitting given background.

In academic prose, information focus dominated (51%), reflecting the register's use of ellipsis almost exclusively in comparative, contrastive, and parallel structures where the foregrounding of differing results or claims is the primary communicative aim. Register marking — the use of ellipsis to signal the register or style of a text — was a secondary function across all registers, accounting for 10–14% of instances in each.

Antecedent Distance and Directionality

Analysis of antecedent distance revealed a further mode-specific pattern. In conversation, 73% of ellipsis instances had their antecedent within the immediately preceding turn or clause — a reflection of the tight local structure of conversational sequence. In academic prose, antecedents were more frequently located within the same sentence (58%), consistent with the register's preference for clause-internal parallel structures. In fiction, cross-sentence anaphora accounted for 41% of instances, indicating that literary ellipsis often requires the reader to track antecedents across longer stretches of text. Cataphoric ellipsis — where the antecedent follows the ellipsis site — was rare across all registers (under 6%), with the highest incidence in fiction.

Conclusion

This study has provided systematic corpus-based evidence for the claim that ellipsis, while a unified grammatical phenomenon governed by the recoverability constraint, is realised very differently across spoken and written registers. Spontaneous conversation deploys ellipsis at nearly five times the rate of academic prose, with a distinctive profile dominated by clausal and VP ellipsis performing interactional functions within tight adjacency sequences. Written registers, particularly academic prose, use ellipsis sparingly and almost exclusively in syntactically predictable environments where comparative and contrastive structures automatically create recoverable omissions.

These findings have implications for discourse theory: they suggest that register-specific grammars of ellipsis are needed, and that a single account of ellipsis licensing cannot capture the full range of spoken and written deployment without reference to communicative context

and interactional structure. For corpus linguistics, the study demonstrates the value of fine-grained, multi-register annotation in revealing patterns obscured by aggregate frequency counts. For applied linguistics, it underlines the necessity of treating spoken and written grammar as distinct but related systems, each requiring targeted description and pedagogical attention.

Recommendations

On the basis of the findings reported above, the following recommendations are made for researchers, practitioners, and curriculum designers:

1. Linguistics researchers should develop annotated spoken corpora with fine-grained ellipsis coding schemes that capture both structural type and discourse function. Current corpora remain under-annotated for functional categories, limiting cross-register comparison.

2. Language teachers should treat spoken and written ellipsis as distinct, register-sensitive phenomena rather than as a single grammatical rule. Spoken grammar instruction should focus on clausal ellipsis in adjacency pairs and VP ellipsis in responses; writing instruction should address gapping and VP ellipsis in parallel and comparative structures.

3. Materials developers should include corpus-derived examples of authentic ellipsis in teaching materials, replacing invented decontextualised examples with real spoken and written instances that illustrate the discourse environment in which ellipsis naturally arises.

4. Academic writing instructors should explicitly teach the use of VP ellipsis and gapping in comparative and contrastive academic structures, as these are the ellipsis types most characteristic of the register and most likely to be underused by non-native-speaker writers.

5. Spoken language teachers and IELTS/TOEFL preparation instructors should incorporate ellipsis recognition and production tasks drawn from authentic conversational data, including adjacency pair completion tasks and spoken corpus listening activities.

6. Future research should extend the cross-register comparison to other languages and to learner corpora, in order to assess whether L1-transfer and developmental factors affect the acquisition and deployment of register-appropriate ellipsis in second language production.

References

- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.
- Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide to spoken and written English grammar and usage*. Cambridge University Press.
- Couper-Kuhlen, E., & Selting, M. (2018). *Interactional linguistics: Studying language in social interaction*. Cambridge University Press.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.



Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7(3), 391–426.

Leech, G., & Short, M. (2007). *Style in fiction: A linguistic introduction to English fictional prose* (2nd ed.). Pearson Longman.

Lobeck, A. (1995). *Ellipsis: Functional heads, licensing, and identification*. Oxford University Press.

Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

van Craenenbroeck, J., & Merchant, J. (2013). Ellipsis phenomena. In M. den Dikken (Ed.), *The Cambridge handbook of generative syntax* (pp. 701–745). Cambridge University Press