Unpacking Ambiguity: The Interaction of Polysemous Discourse Markers and Non-DM Signals

Jingni Wu
Georgetown University
jw2175@georgetown.edu

Amir Zeldes
Georgetown University
amir.zeldes@georgetown.edu

Abstract

Discourse markers (DMs) like 'but' or 'then' are crucial for creating coherence in discourse, yet they are often replaced by or co-occur with non-DMs ('in the morning' can mean the same as 'then'), and both can be ambiguous ('since' can refer to time or cause). The interaction mechanism between such signals remains unclear but pivotal for their disambiguation. In this paper we investigate the relationship between DM polysemy and co-occurrence of non-DM signals in English, as well as the influence of genre on these patterns. Using the framework of eRST, we propose a graded definition of DM polysemy, and conduct correlation and regression analyses to examine whether polysemous DMs are accompanied by more numerous and diverse non-DM signals. Our findings reveal that while polysemous DMs do co-occur with more diverse non-DMs, the total number of co-occurring signals does not necessarily increase. Moreover, genre plays a significant role in shaping DM-signal interactions.

1 Introduction

Identifying and understanding discourse relations is fundamental to discourse comprehension. Discourse markers (DM) such as 'and', 'because', and 'however' have been widely recognized as the most typical indicator of coherence relations and are also referred to as discourse connectives or cue phrases (Forbes-Riley et al., 2006). Early research focused on DMs as the sole device indicating relations, and their presence is often used to distinguish explicit and implicit relations (Webber and Joshi, 1998; Robaldo et al., 2008). In applied Natural Language Processing (NLP) they also remain the focus of research on automatic detection of discourse relation signaling, as evidenced by the series of DISRPT (Discourse Relation Parsing and Treebanking, see Braud et al. 2024) shared tasks including DM detection as a track. Since such markers come from a closed list, systems can target only these words or phrases (Yu et al., 2019), then focus on disambiguation, with recent system scores achieving over 93% F1-scores for English (Liu et al., 2023).

However, more recent studies have shown that DMs account for only a small fraction of discourse relations, which can be signaled by *reference* (e.g. anaphora to indicate ELABORATION[1]), *semantic* (antonymy to indicate CONTRAST), *lexical* ('the next day' can indicate temporal SEQUENCE like the DM 'then'), *morphological* (past followed by present tense can also indicate SEQUENCE) and *graphical* cues (e.g. a question mark signaling a QUESTION relation). In this paper we follow the taxonomy of non-DM signal types proposed by Zeldes et al. (2025), which distinguishes eight major classes with a total of 45 subtypes, illustrated in Table 1. Such non-DM signals can be crucial for disambiguating otherwise ambiguous DMs, such as 'since', which can signal both CAUSE and temporal CIRCUMSTANCE relations. Taken together, DMs and such similar non-DM devices are referred to collectively as discourse relation *signals* (Das and Taboada, 2018a,b; Zeldes et al., 2025).

Despite extensive research on DMs and other signals individually, far less attention has been given to their interaction. Prior studies have examined the distribution of DM-signal co-occurrence and explored potential motivations from corpus-based (Das and Taboada, 2019; Crible, 2020) and experimental perspectives (Crible and Demberg, 2020; Grisot and Blochowiak, 2017). These studies have revealed that DM-signal co-occurrence is influenced by cognitive constraints and information density, and that several factors, such as the ambiguity of DMs (Crible, 2020), the semantics of discourse relations (Das and Taboada, 2019; Crible and Demberg, 2020), and genres (Crible, 2020), af-

[1] Here and below we will assume discourse relation labels commonly used in Rhetorical Structure Theory (Mann and Thompson, 1988). Our definition of what constitutes anaphoric reference aligns with (Zeldes, 2022).

Proceedings of the 6th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025), pages 14–26
November 9, 2025

signal type	subtypes	example
dm	but, then, on the other hand...	*[They wanted to] [**but** couldn't]*<adversative−contrast>
graphical	colon, dash, semicolon layout items in sequence parentheses, quotation marks question mark	*[Let me tell you a story **:**]*<organization−preparation> *[Introduction]*<organization−heading> *1. wash [2. cut]*<joint−list> *it rained [**(**and snowed a bit**)**]*<elaboration−additional> *[Did you**?**]*<topic−question> *No.*
lexical	alternate expression indicative word/phrase	*He agreed. [**That is** he said yes]*<restatement−repetition> *They planned a party! [That's **nice**/**Can't wait**!]*<evaluation−comment>
morphological	mood tense	***Go** with them [I think you should]*<explanation−motivation> *I **started** an hour ago, [now I**'m** resting]*<joint−sequence>
numerical	same count	*[**Two** reasons.]*<organization−preparation> *First...*
reference	comparative demonstrative / personal propositional	*[I don't want **it**]*<adversative−antithesis> *I want **another one**.* *They met **Kim**. [**This person** / **she** was...]*<elaboration−additional> ***They met Kim**. [**This encounter** was...]*<elaboration−additional>
semantic	antonymy attribution source lexical chain meronymy negation repetition/synonymy	*Beer is **cheap**, [wine is **expensive**]*<adversative−contrast> *[**Kim** said]*<attribution−positive> *they would* *it was **funny** [so they **laughed**]*<causal−result> ***The house** was big, [**the door** two meters tall]*<elaboration−additional> *Kim danced, [Yun did**n't** dance]*<adversative−contrast> *They met **Dr. Kim**. [**Dr. Kim**/**The surgeon** was...]*<elaboration−additional>
syntactic	infinitival/relative clause interrupted matrix clause modified head nominal modifier parallel syntactic construction past/present participial clause reported speech subject auxiliary inversion	*a plan [**to** win]*<purpose−attribute> *[I meant –]*<orgnization−phatic> *I mean,* *a **plan** [to win]*<purpose−attribute> *articles [**explaining** chess]*<elaboration−attribute> ***it's all** tasty [**it's all** pretty]*<joint−list> *Kim appeared [**dressed** in black]*<elaboration−attribute> *[Kim said]*<attribution−positive> ***that** they **would*** *I would have [**had** I known]*<contingency−condition>

Table 1: Signal types and subtypes, with examples highlighting in red the signal tokens which indicate the relation of the unit in square brackets.

fect the likelihood of co-occurrence. However, the specific mechanisms governing DM-signal interactions remain unclear. In particular, little is known about which conditions favor such co-occurrences, how different signals contribute to disambiguation and the resulting effect, what happens when conflicting signals appear, and how these patterns vary across discourse relations and genres.

While previous studies have confirmed that polysemous DMs co-occur with additional signals, there has been little systematic analysis of how different types and combinations of non-DM signals help resolve ambiguity. This study seeks to bridge this gap by analyzing the distribution, number, type, and co-occurrence patterns of signals with polysemous DMs across genres. We focus on the following research questions:

1. Are polysemous DMs accompanied by more numerous or more diverse non-DMs?
2. What are the typical combination strategies for DM and non-DM signals?
3. Are strategies and distributions general, or are they genre-specific?

Because of their lower information content, we hypothesize that polysemous DMs will exhibit a stronger connection with non-DM prevalence. We also anticipate that different genres will exhibit distinct preferences for specific types of signals for polysemous DMs when resolving DM ambiguity, in part because they involve different prior likelihoods of certain relations. We therefore expect the relationship between DM polysemy and the number and diversity of co-occurring non-DMs to vary by genre.

2 Related Work

Previous studies have demonstrated that discourse relations are frequently signaled not just by DMs, with over 80% of signaled relations exhibiting some other textual cues, both with and without the presence of accompanying DMs (Taboada and Das, 2013; Das and Taboada, 2018a,b). Moreover, it has been found in many cases that multiple signals indicate discourse relations simultaneously (Das and Taboada, 2018b; Webber et al., 2019). Among these, the combined use of DMs and non-DM signals is particularly common and serves to signal a wide variety of relations (Das and Taboada, 2019). For instance, in the following example from the GUM corpus (Zeldes, 2017), 'while' functions as a typical DM for the CONCESSION relation, which is further reinforced by a lexical chain connecting existing 'studies of the psychology of art' with 'no work', creating a contrast between previous work that exists and a gap in academic literature:

(1) [**While** studies of the psychology of art have focused on ... no work has been ...] [Relation: ADVERSATIVE-CONCESSION; DM: 'While'; Signal: semantic (lexical chain)] (File: *GUM_academic_art*)

Although this pattern is very common in academic writing, little attention has been paid to the ways in which ambiguous DMs such as 'while' (which can also mean *during a time that...*) resolve to a unique intepretation thanks to co-occurring signals in this manner, and the joint use of DMs and signals remains a complex question.

Non-DM signals can 1) overlap with DMs in meaning, potentially leading to redundancy, 2) co-occur with DMs but function independently (potentially signaling multiple distinct relations), and 3) may complement DMs in specific types of relations and environments (Hoek et al., 2018). Recent studies have begun to explore the underlying triggers of the *DM + other signals* phenomenon. Das and Taboada (2019) suggested that such combinations may arise from the inherent ambiguity of certain DMs which can signal various relations. For example, the DM *and* can mark additive LIST and temporal SEQUENCE relations, among other options, as illustrated in the following examples from GUM:

(2) [I came home last night **and** told you.] [Relation: JOINT-SEQUENCE] (File: *GUM_conversation_grounded*)

(3) [... borders of our moral **and** ethical understanding.] [Relation: JOINT-LIST] (File: *GUM_essay_ghost*)

Building on this, researchers have introduced the concept of *marking strength* or *signaling strength* of DMs, which can be assessed by the number and frequency of discourse relations they can signal (Asr and Demberg, 2012). Zeldes and Liu (2020) proposed the *delta-softmax* metric, which quantifies prediction accuracy degradation for a trained neural model when a word is removed to estimate its signaling strength for a relation, providing empirical validation of an intuitive graded *signalyness* phenomenon. For instance, 'but' could be significantly less ambiguous than 'and' as a DM, in that removing 'but' would make the relation much harder to predict than removing 'and'.

This strength directly influences how DMs interact with non-DM signals: it has been suggested that DMs tend to co-occur more frequently with other signals when indicating a wide range of discourse relations (Das and Taboada, 2019). In such cases, non-DM signals can play a disambiguation role, helping to clarify the intended relation (Crible and Demberg, 2020). However, although patterns might be typical of specific genres, for example if formal texts prefer stronger and less ambiguous DMs, the association between DMs and other signals has not been found to vary significantly across genres in previous work (Crible, 2020).

In addition, combinations of DMs and non-DM signals vary across relation types, but they are not necessarily driven by inherent semantics (Das and Taboada, 2019). That is to say, certain relations tend to prefer either DM-only or DM-plus-signal combinations. This is partly influenced by the inherent semantics of the discourse relations themselves (e.g., weakly connected sentences), but also appears to reflect an independent pragmatic strategy for ensuring clarity of the writer's intention.

While prior research has qualitatively identified some factors influencing the co-occurrences of DM and non-DM signals, a systematic analysis of how specific non-DM signals interact with ambiguous DMs across relation types and genres remains underexplored. In particular, the co-occurrence patterns between ambiguous DMs and accompanying signals have not been quantitatively mapped. Using the largest sample of annotated discourse relation signals to date, this study addresses this gap by investigating 1) the types and frequencies of non-DM signals that co-occur with ambiguous DMs, 2) how these combinations vary across genres, and 3) whether certain signal combinations contribute to disambiguating the intended discourse relation.

3 Data

This study uses the Georgetown University Multilayer (GUM) Corpus which consists of 16 spoken and written, informal and formal style English text types (Zeldes, 2017) (see corpus details in Appendix A). The corpus originally contained Rhetorical Structure Theory (RST) annotations, which were recently extended based on Enhanced Rhetorical Structure Theory (eRST, Zeldes et al. 2025), adding annotated DMs and seven types of non-DM signals based on the taxonomy proposed by Das and Taboada (2018a), as well as adding multiple concurrent and tree breaking relations edges to the initial RST trees. With over 250,000 tokens, this is currently the largest dataset annotated for DMs and non-DM signals of discourse relations.

Since this study relies on accurate discourse annotations, we also report how quality was assured. Inter-annotator agreement studies on GUM showed F1 scores of 92.3 for DM identification and 90 for relation association (36 docs, 32K tokens). For non-DM signals, many types were automatically derived from gold syntax and coreference annotations, with others manually corrected or added. On a subset of documents, human–human agreement yielded an F1 of about 0.80.

4 Polysemy of DMs

The ambiguous nature of DMs arises from their one-to-many relationship with discourse relations. DMs that can signal multiple relations, such as ‘and’, are often described as *weak signals* (Asr and Demberg, 2012; Das and Taboada, 2019; Crible, 2020), as they do not map consistently to a single meaning, in contrast to unambiguous DMs such as ‘despite’, which always marks a CONCESSION.

Going beyond previous categorical approaches to such polysemy, we adopt a graded, quantifiable definition of DM polysemy by calculating the Shannon Entropy (Shannon, 1951) of DM meanings, which measures how evenly a DM is distributed across multiple discourse relations. A high entropy score indicates that a DM appears equally in multiple relations, while a lower score means that a DM is used in only one or very few types of discourse relations, or with a strong predominant sense. We expect a high entropy score here for the most polysemous DMs, for example, a high value for DMs like ‘and’ or even ‘but’, and the lowest value for DMs like ‘despite’.

Shannon Entropy is calculated by measuring the probability of the DM appearing in each discourse relation. The polysemy score is computed as follows:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) \qquad (1)$$

x_i is the possible discourse relation signaled by a DM, n is the number of distinct relations signaled by the DM, and $P(x_i)$ is the probability of the DM signaling the relation x_i.

5 DM-Signal Co-occurrences

5.1 General Distribution

Across 16 genres, 21,435 discourse relations are annotated in our data, of which 1,372 (6.4%) are indicated by both DMs and non-DM signals. This result aligns fairly closely with Das and Taboada (2018a)’s finding for Wall Street Journal news (7.55%, see also Liu and Zeldes 2019). However as suspected, we observe substantial variation across genres (see Figure 1): *essay* (8.7%), *bio* (8.6%), and *whow* (how-to guides form Wikihow, 8.5%) show a higher proportion of DM-signal co-occurrence, whereas *conversation* has the lowest proportion (3.9%).

Among the 1,372 instances of DM-signal co-occurrence, 96% are marked by DM + 1 signal or DM + 2 signals, while just 3% are marked by three to four signals. Only a handful of cases include more than five signals (see Table 2).

The most commonly used DM in co-occurrence with other signals across genres is the connective ‘and’ (36.6%), which is generally the most frequently used DM as well. Almost all genres in our corpus employ ‘and’ in DM-signal co-occurrences, except for *academic* where the conjunction ‘by’ is the most common DM favoring non-DM signal accompaniment, as in example (4), where the DM signaling the MEANS relation is accompanied by the lexical signal ‘using’:

(4) **by** using a second order Rao and Scott (1981) ... correction

The top three most frequently used signal types in co-occurrences are *semantic*, *syntactic*, and *lexical* across genres, though different genres favor different types of signals, in part due to the format of texts. In spoken genres such as *vlog*, *conversation*, and *court*, there is a large amount of *reference* signals used along with DMs to indicate relations

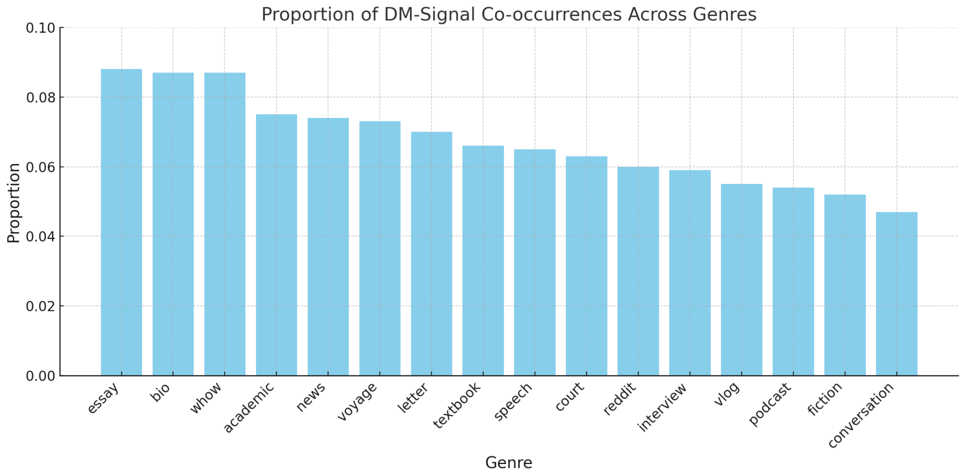

Figure 1: Proportion of DM-Signal co-occurrence across genres

	DM + 1 Signal	DM + 2 Signals	DM + 3 Signals	DM + 4 Signals	DM + 5 Signals	DM + 6 Signals	DM + 8 Signals
Total counts	1092	229	42	6	1	1	1
Proportion	79.55%	16.7%	3.1%	0.44%	0.07%	0.07%	0.07%

Table 2: Pattern of DM + signal combinations in co-occurrences

such as ELABORATION (see Figure 2). Trivially, *graphical* signals such as quotation marks to signal ATTRIBUTION cannot occur in spoken language and are restricted to written data.

5.2 Polysemous DMs and Signal Patterns

The DM 'so' has the highest polysemy score across all genres in our dataset, while the DM 'for' exhibits the most diverse range of accompanying signals (see Table 3). Here, *diversity* [2] refers to the number of distinct non-DM signal types that co-occur with a given DM, including individual signal types (e.g. *semantic*) and combinations of multiple types (e.g. *semantic* + *lexical*).

DM	non-DM signal diversity
for	29.50
and	26.64
if	25.00
by	20.80
when	19.00

Table 3: Top 5 DMs with the highest signal diversity

[2]Since DM frequency varies across genres, we normalized diversity by dividing the number of unique co-occurring signal types by the square root of total DM occurrences. This accounts for diminishing returns and prevents frequent DMs from being unfairly penalized.

This raises the question of whether more polysemous DMs tend to co-occur with a greater number of non-DM signals and exhibit more diverse signal patterns, on account of the less consistent mapping of their form to a specific meaning. To answer these questions, we employed fitted regression models to examine the relationship between DM polysemy (independent variable) and two dependent variables: (1) the total number of co-occurring non-DM signals and (2) the diversity of signal types associated with each DM.

Our results, based on both Pearson correlation and regression analyses (see details in Appendix C), suggest that polysemous DMs are more strongly associated with the diversity, rather than the quantity, of accompanying non-DM signals. While we observe a weak but statistically significant correlation between entropy and the total number of co-occurring signals ($r = 0.248, p < 0.05$), this association does not hold in a multiple regression model where both entropy score and total co-occurring signals are included as predictors of normalized signal diversity. In contrast, entropy remains a significant predictor of normalized diversity, even after controlling for signal quantity ($p < 0.001$). This supports the view that more polysemous DMs require more diverse signal patterns rather than just more signals to clarify their discourse functions.

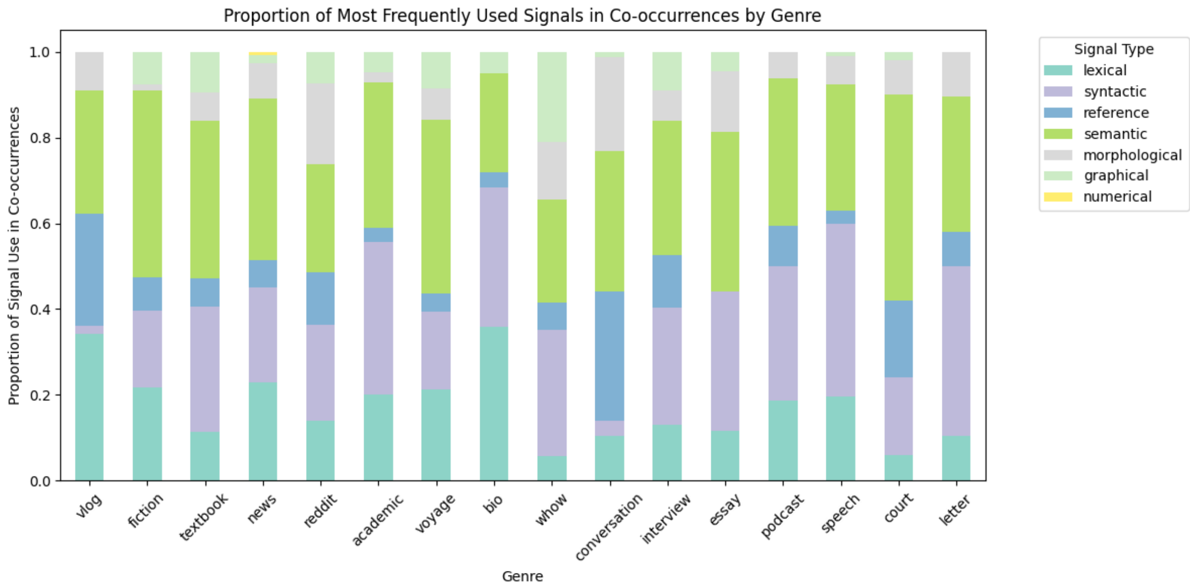

Figure 2: Proportion of most frequently co-occurring signals by genre

However, the overall explanatory power of entropy score alone is modest (adjusted $R^2 = 0.071$), suggesting that other factors may influence the relationship between DM polysemy and signal diversity. To further explore this, we considered genre as a variable. The regression model (see details in Appendix C) that includes genre and its interaction with entropy score significantly improved model fit ($p < 0.000001$, adjusted $R^2 = 0.090$), suggesting that the effect of DM polysemy on co-occurring signal patterns varies across genres. Notably, genres such as *vlogs* exhibited a significantly stronger positive relationship between DM polysemy and signal diversity, while others like *letter* showed a weaker or even negative trend. This variation highlights that the need for signal diversity in disambiguating polysemous DMs is not uniform, but shaped by genre-specific discourse norms. These genre-specific effects raise the question of what kinds of non-DM signal patterns are employed in each genre, which we address in the next section.

Looking at patterns rather than counts of signals in more detail, certain signal types consistently co-occur with highly polysemous DMs, suggesting that these signals play a crucial role in disambiguating them. For example, *lexical* and *syntactic* signals frequently appear across multiple cases and are more likely to be combined with other signal types, reinforcing their role in guiding interpretation (see Table 4).

In summary, our hypothesis is partially supported: polysemous DMs are more likely to exhibit diverse combinations of non-DM signal, possibly due to their less stable mapping of form to meaning, but they do not consistently co-occur with a greater number of signals. Given prior evidence that signal co-occurrences vary in quantity across genres (Figure 1), we now turn to investigate the impact of genre variation, and examine the hypotheses within individual genres in the following section.

5.3 Signal Combinations and Genre Effects

According to the entropy scores, the most polysemous DMs within each genre are presented in Table 5[3]. Notably, the most ambiguous DMs within each genre differs from those identified as globally most ambiguous. The DM ‘and’ is the most polysemous in six genres, and DM ‘so’ and ‘as’ are the second most ambiguous DMs in eight genres. By contrast, ‘also’ is the most polysemous DM in only one genre.

The non-DM signals that co-occur with polysemous DMs exhibit diverse combination patterns, which vary across genres. A single DM may be more likely to be paired with entirely different signals depending on the genre. For example, the DM ‘and’ is most frequently used with *lexical_chain* signals (a subtype of *semantic*) signals, see example (5)) in nearly all genres, except for *vlog*, *bio*, *whow*, *conversation*, and *podcast* (see Appendix Table 7). Here the lexical relation between the related items

[3]When comparing the polysemy across genres, we normalized the entropy score by dividing the raw entropy score by the maximum possible entropy for each DM in each genre.

DM	Top 3 co-occurring Types	Top 3 most frequent combinations
so	morphological, lexical, syntactic	(lexical + reference), (syntactic + reference + graphical)
in	syntactic, lexical	(syntactic + syntactic), (lexical + syntactic + syntactic)
with	semantic, graphical, syntactic	(reference + semantic), (syntactic + syntactic), (numerical + semantic + semantic)
as	syntactic, lexical, morphological	(lexical + semantic)
and	reference, lexical, semantic	(reference + graphical), (semantic + semantic), (lexical + syntactic)

Table 4: Top 5 Polysemous Discourse Markers and Co-occurring Signal Patterns

Genre	DM	Raw entropy	Normalized entropy
Court	and	2.85	0.61
Reddit	so	2.59	0.60
Conversation	and	2.63	0.58
News	as	2.55	0.57
Fiction	so	2.35	0.53
Voyage	as	2.33	0.53
Interview	and	2.34	0.52
Vlog	and	2.27	0.52
Speech	so	2.20	0.51
Wikihow	so	2.25	0.50
Textbook	so	2.16	0.48
Podcast	and	2.15	0.48
Biography	also	1.90	0.44
Academic	as	1.92	0.44
Letter	as	1.84	0.42
Essay	and	1.64	0.40

Table 5: Entropy score of DMs per genre

'information' and 'content' forms a semantic signal next to 'and' to indicate that the two clauses are part of a list.

(5) [The Penn State wiki was never proposed as a source of official <u>information</u>, **and** the university already hosts non-official <u>content</u> ...] [Relation: JOINT-LIST; DM: 'and'; signal: semantic (lexical chain)](File: *GUM_letter_wiki)*

To further assess whether genre systematically affects the distribution of non-DM signals for polysemous DMs, we conducted Chi-Squared Goodness of Fit. For each genre, we compared the signal-type distribution to the global (genre-agnostic) distribution for the same set of DMs. After applying False Discovery Rate (FDR) correction, we found that all 16 genres show statistically significant deviations ($p_{\text{corrected}} < 0.05$), confirming that genre has a strong effect on the signaling strategies used to support polysemous discourse markers. Genres such as *vlog* and *conversation* exhibited the largest deviations, suggesting that signal use in these genres is especially distinct from the overall norm.

This variation can be attributed to the nature of spoken genres such as *vlogs* and *conversations*, which emphasize audience interaction and shared common ground. In these contexts, indicative words and personal references are more commonly used to enhance engagement and coherence. Similarly, other spoken genres tend to favor *reference* signals, particularly *personal references*, using chains of pronouns to help the audience recall previously mentioned content. Semantic signals in the genre *podcast* show a particularly strong use of *meronymy*, using words in a part-whole relationship alongside the polysemous 'and' to indicate elaborations on complex information.

In addition, 'and' tends to use combined signals more frequently than other polysemous DMs, dovetailing with our initial hypothesis about non-DMs compensating for ambiguous DMs. Notably, in almost all genres where 'and' is the most polysemous DM, it co-occurs with multiple signals, except for the genre *essay*, where it primarily appears by itself or with a single signal type. Among all signal combinations, the most frequent combined signal set for 'and' is *reference* + *semantic*, i.e. anaphora and lexical relations between words in the units joined by 'and'. Interestingly, *letter* is the only genre where the most polysemous DM is 'as', yet it does not co-occur with any additional non-DM signals. Looking at its instances, nearly 65% are used to indicate MODE relations (manner/means), as opposed to only 32.2% in the rest of the corpus, suggesting that this usage may simply be more predictable as a default in *letters* – the most common sense in the remaining genres is indicating a temporal CIRCUMSTANCE, similarly to 'when'.

Many DMs exhibit reduced polysemy within individual genres compared to their global scores, suggesting that their meaning is more specialized and thus less ambiguous in certain contexts. However, some DMs show substantial variation across genres, potentially requiring a greater variety or higher number of non-DM signals to aid interpretation in specific genres (see Figure 3).

To identify DMs whose polysemy varies the most across genres, we compared their normalized within-genre polysemy scores with their global polysemy scores. The top five discourse markers with the largest shifts are 'so', 'in', 'with', 'given', and 'indeed', which align with the overall polysemy ranking observed earlier. Highly polysemous DMs exhibit greater variance across genres, likely because their multiple meanings make them more adaptable to different discourse needs, which can be disambiguated either by non-DM signals, or simply by their use in a genre with strong priors on expected senses. In contrast to DMs with lower polysemy, which may serve more stable functions, highly polysemous DMs can shift more dramatically depending on genre-specific discourse structures, discourse relation compatibility, communicative conventions, and signaling strategies.

Among the genres, *academic*, *reddit*, and *court* seem to have larger variance, indicating that DMs used in these genres experience the most sizable shifts in polysemy compared to their global usage. These genres may have DMs that behave very differently in terms of polysemy compared to their global usage. In contrast, DMs in *fiction*, *podcast*, and *letter* appear to behave similarly locally and globally.

In addition, we examined the relationship between the number of co-occurring non-DM signals, the diversity of those signals, and the DM polysemy within each genre. Our global analysis confirms that polysemous DMs tend to co-occur with more diverse signal patterns, however, since the frequency and variety of co-occurring signals differ across genres, we extended this investigation within individual genres to determine whether genre influences this phenomenon. The results indicate that spoken genres such as *court*, *podcast*, and *vlog*, DM polysemy strongly correlates with both the number and diversity of co-occurring non-DM signals, while other genres' results almost align with our findings across genres, that the higher a DM's polysemy score, the more diverse these signal combinations tend to be. This supports our hypothesis that specific genres, particularly spoken contexts and formal or unusual settings (e.g. courtroom transcripts or academic writing), adopt distinct non-DM signaling strategies which help in the disambiguation of polysemous DMs.

6 Conclusion

This study investigates the relationship between DM polysemy, the number and diversity of co-occurring non-DM signals, and the role of genre in these interactions. Our findings partially support the hypothesis that polysemous DMs exhibit more diverse non-DM signal patterns but do not necessarily co-occur with a greater number of non-DM signals. Moreover, genre greatly shapes DM polysemy, with significant variations in DM entropy and signal usage. Spoken genres (e.g. *court*, *podcast*, *vlog*) show a stronger dependence on non-DM signals to disambiguate polysemous DMs, while some written genres exhibit little or no correlation. This pattern likely reflects cognitive and interactional pressures in speech, where speakers must maintain fluency under real-time constraints (Clark, 2002) and often deploy additional cues to support coherence. Moreover, DMs in spoken discourse frequently serve interactive functions, such as managing turn-taking or structuring talk (Clark and Tree, 2002), which may further increase their co-occurrence with diverse signals.

These findings challenge theoretical views of DMs in frameworks that assume that DM marking means we do not need to consider other types of signals, such as in the Penn Discourse Treebank framework, where alternative lexicalizations marking a relation are generally only considered if a DM is absent (Prasad et al., 2018). They also suggest a consequence for treating DMs and other types of markers as categorically consistent across different

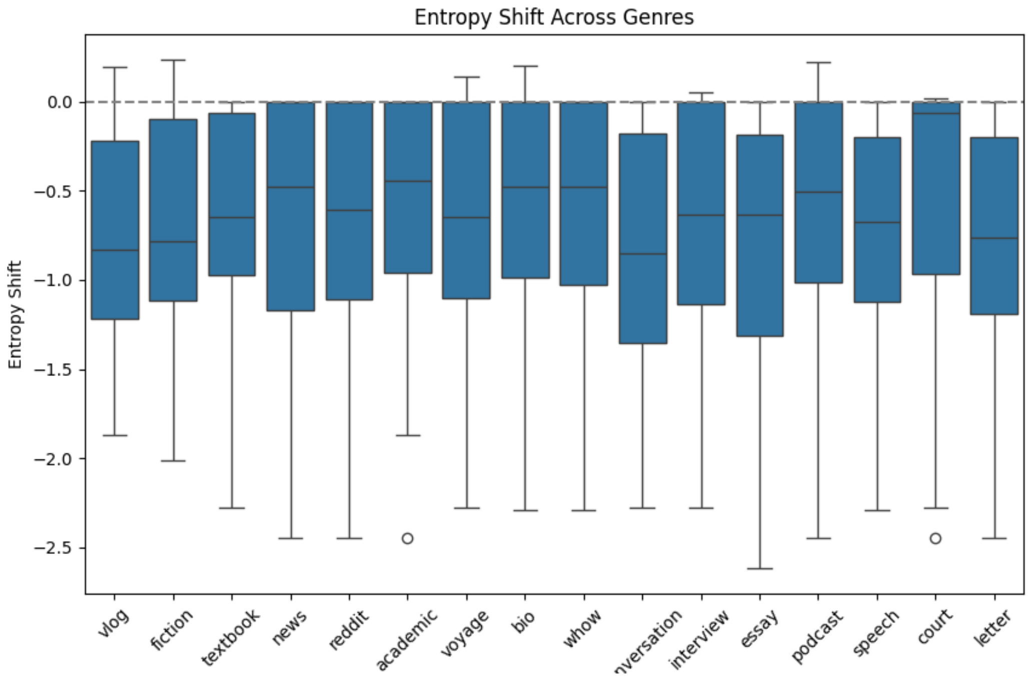

Figure 3: Entropy shift across genres

types of text: in practice, we find great variation in the extent and types of signaling present based on genre.

On the other hand, some genres exhibit little to no significant correlation among DM polysemy, the number and the diversity of non-DM signals, suggesting that different discourse contexts may impose different constraints on how DMs interact with non-DM signals. Additionally, we identified DMs whose polysemy scores are highly shifted across genres, such as, 'so', 'in', 'with', 'given', 'indeed', and 'while'. This finding suggests that certain polysemous DMs are more sensitive to contextual variation, whereas others maintain stable meanings across different discourse settings. Further research is needed to understand the extent to which the picture of genre variation presented here is comprehensive, which could be carried out with new eRST data on unusual genres that has recently become available (for example in the GENTLE corpus, Aoyama et al. 2023, which includes annotations for poetry, legal writing, and more).

7 Discussion

This study does not fully account for the distribution of different discourse relations, which can further shape the observed patterns of polysemy and signal co-occurrence. Prior research has demonstrated that certain non-DMs are more commonly used to disambiguate DMs in specific relations, such as *contrast* and *consequence* (Crible and Demberg, 2020), and different relations may vary in their sensitivity to signals, with some relations being more reliant on co-occurring non-DM cues for disambiguation. Moreover, the compatibility between DMs and specific signals may play a greater role in guiding interpretation than sheer signal frequency. Future work should therefore examine how relation type conditions the use of non-DM signals with polysemous DMs, and expand analysis to larger silver-standard multilayer corpora such as AMALGUM (A Machine Annotated Lookalike of GUM, Gessler et al. 2020), enriched with automatic annotation of discourse relations and signals, which would be less accurate, but mitigate the problem of data sparseness.

Beyond theoretical implications, these findings also have practical relevance for NLP. Current discourse parsers often treat explicit relations with DMs as straightforward, yet our results show that polysemous markers frequently rely on co-occurring signals. Incorporating such cues could improve discourse relation classification and domain adaptation, while also enhancing explainability in downstream tasks such as summarization or dialogue systems.

References

Tatsuya Aoyama, Shabnam Behzad, Luke Gessler, Lauren Levine, Jessica Lin, Yang Janet Liu, Siyao Peng, Yilun Zhu, and Amir Zeldes. 2023. GENTLE: A genre-diverse multilayer challenge set for English NLP and linguistic evaluation. In *Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII)*, pages 166–178, Toronto, Canada. Association for Computational Linguistics.

Fatemeh Torabi Asr and Vera Demberg. 2012. Measuring the strength of linguistic cues for discourse relations. *Proceedings of the Workshop on Advances in Discourse Analysis and Its Computational Aspects*, pages 33–42.

Chloé Braud, Amir Zeldes, Laura Rivière, Yang Janet Liu, Philippe Muller, Damien Sileo, and Tatsuya Aoyama. 2024. DISRPT: A multilingual, multi-domain, cross-framework benchmark for discourse processing. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4990–5005, Torino, Italia. ELRA and ICCL.

Herbert H Clark. 2002. Speaking in time. *Speech Communication*, 36(1-2):5–13.

Herbert H Clark and Jean E Fox Tree. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84(1):73–111.

Ludivine Crible. 2020. Weak and strong discourse markers in speech, chat, and writing: Do signals compensate for ambiguity in explicit relations? *Discourse Processes*, 57(9):793–807.

Ludivine Crible and Vera Demberg. 2020. The role of non-connective discourse cues and their interaction with connectives. *Pragmatics & Cognition*, 27(2):313–338.

Debopam Das and Maite Taboada. 2018a. RST Signalling Corpus: A corpus of signals of coherence relations. *Language Resources and Evaluation*, 52:149–184.

Debopam Das and Maite Taboada. 2018b. Signalling of coherence relations in discourse, beyond discourse markers. *Discourse Processes*, 55(8):743–770.

Debopam Das and Maite Taboada. 2019. Multiple signals of coherence relations. *Discours*, (24).

Katherine Forbes-Riley, Bonnie Webber, and Aravind Joshi. 2006. Computing discourse semantics: The predicate-argument semantics of discourse connectives in D-LTAG. *Journal of Semantics*, 23(1):55–106.

Luke Gessler, Siyao Peng, Yang Liu, Yilun Zhu, Shabnam Behzad, and Amir Zeldes. 2020. AMALGUM – a free, balanced, multilayer English web corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5267–5275, Marseille, France. European Language Resources Association.

Cristina Grisot and Joanna Blochowiak. 2017. Temporal connectives and verbal tenses as processing instructions: Evidence from French. *Pragmatics & Cognition*, 24:404–440.

Jet Hoek, Sandrine Zufferey, Jacqueline Evers-Vermeul, and Ted JM Sanders. 2018. The linguistic marking of coherence relations: Interactions between connectives and segment-internal elements. *Pragmatics & Cognition*, 25(2):276–309.

Wei Liu, Yi Fan, and Michael Strube. 2023. HITS at DISRPT 2023: Discourse segmentation, connective detection, and relation classification. In *Proceedings of the 3rd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2023)*, pages 43–49, Toronto, Canada. The Association for Computational Linguistics.

Yang Liu and Amir Zeldes. 2019. Discourse relations and signaling information: Anchoring discourse signals in RST-DT. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, volume 2, pages 314–317.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Rashmi Prasad, Bonnie Webber, and Alan Lee. 2018. Discourse annotation in the PDTB: The next generation. In *Proceedings of the 14th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, pages 87–97, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Alan Robaldo, Eleni Miltsakaki, Alan Lee, Rashmi Prasad, Nikhil Dinesh, Bonnie Webber, and Aravind Joshi. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. ELRA.

Claude E Shannon. 1951. Prediction and entropy of printed English. *Bell System Technical Journal*, 30(1):50–64.

Maite Taboada and Debopam Das. 2013. Annotation upon annotation: Adding signalling information to a corpus of discourse relations. *Dialogue & Discourse*, 4(2):249–281.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. The Penn Discourse Treebank 3.0 annotation manual. *University of Pennsylvania*, 35:108.

Bonnie Lynn Webber and Aravind K Joshi. 1998. Anchoring a lexicalized tree-adjoining grammar for discourse. *arXiv preprint cmp-lg/9806017*.

Yue Yu, Yilun Zhu, Yang Liu, Yan Liu, Siyao Peng, Mackenzie Gong, and Amir Zeldes. 2019. GumDrop at the DISRPT2019 shared task: A model stacking approach to discourse unit segmentation and connective detection. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 133–143, Minneapolis, MN. Association for Computational Linguistics.

Amir Zeldes. 2017. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.

Amir Zeldes. 2022. Can we fix the scope for coreference? problems and solutions for benchmarks beyond OntoNotes. *Dialogue & Discourse*, 13(1):41–62.

Amir Zeldes, Tatsuya Aoyama, Yang Janet Liu, Siyao Peng, Debopam Das, and Luke Gessler. 2025. eRST: A signaled graph theory of discourse relations and organization. *Computational Linguistics*, 51(1):23–72.

Amir Zeldes and Yang Janet Liu. 2020. A neural approach to discourse relation signal detection. *Dialogue and Discourse*, 11(2):74–99.

Text type	**Source**	**Docs**	**Tokens**
Academic writing	Various	18	17,169
Biographies	Wikipedia	20	18,213
CC Vlogs	YouTube	15	16,864
Conversations	UCSB Corpus	15	17,932
Courtroom transcripts	Various	9	11,148
Essays	Various	9	10,842
Fiction	Various	19	17,511
Forum	reddit	18	16,364
How–to guides	wikiHow	19	17,081
Interviews	Wikinews	19	18,196
Letters	Various	12	9,989
News stories	Wikinews	24	17,186
Podcasts	Various	10	11,986
Political speeches	Various	15	16,720
Textbooks	OpenStax	15	16,693
Travel guides	Wikivoyage	18	16,515
Total GUM		255	250,409

Table 6: Overview of GUM corpus by text type.

A GUM Information

B Signal Patterns for "and" by Genre

C Regression Results

Table 7: Signal Patterns for "and" by Genre

Genre	Top 3 "and" + 1 signal		Top 3 "and" + multiple signals	
	Signal Type	**Signal Subtype**	**Signal Type**	**Signal Subtype**
	lexical	indicative_word	reference + semantic	personal_reference + lexical_chain
vlog	semantic	lexical_chain	reference + reference	oral_reference + propositional_reference
	reference	personal_reference	reference + semantic	personal_reference + synonymy
	semantic	lexical_chain	graphical + graphical	items_in_sequence + semicolon
textbook	semantic	meronymy	graphical + reference	parentheses + personal_reference
	graphical	semicolon	semantic + semantic + semantic + semantic	lexical_chain + lexical_chain + lexical_chain + lexical_chain
	semantic	lexical_chain	semantic + semantic	lexical_chain + meronymy
reddit	lexical	indicative_word	reference + reference + semantic	personal_reference + propositional_reference + synonymy
	reference	personal_reference	reference + semantic + semantic	personal_reference + lexical_chain + repetition
	semantic	lexical_chain	reference + semantic	personal_reference + meronymy
academic	lexical	indicative_word	lexical + lexical	indicative_word + indicative_word
	semantic	meronymy	graphical + graphical + semantic	items_in_sequence + semicolon + meronymy
	semantic	lexical_chain	semantic + semantic	lexical_chain + lexical_chain
voyage	semantic	meronymy	lexical + lexical	indicative_phrase + indicative_word
	lexical	indicative_word	semantic + semantic	lexical_chain + meronymy
	lexical	indicative_word	lexical + lexical	indicative_word + indicative_word
bio	semantic	lexical_chain	lexical + lexical	indicative_phrase + indicative_word
	lexical	indicative_phrase	semantic + semantic	lexical_chain + meronymy
	graphical	items_in_sequence	semantic + semantic	lexical_chain + meronymy
whow	semantic	lexical_chain	graphical + semantic	items_in_sequence + lexical_chain
	reference	personal_reference	semantic + semantic	lexical_chain + lexical_chain
	reference	personal_reference	reference + reference	personal_reference + personal_reference
conversation	morphological	tense	reference + semantic	personal_reference + synonymy
	semantic	lexical_chain	reference + semantic	personal_reference + lexical_chain
	semantic	lexical_chain	semantic + semantic	lexical_chain + meronymy
fiction	semantic	meronymy	graphical + lexical	semicolon + indicative_word
	lexical	indicative_word	semantic + semantic	lexical_chain + lexical_chain
	semantic	lexical_chain	semantic + semantic	lexical_chain + meronymy
news	semantic	meronymy	lexical + morphological	indicative_word + tense
	lexical	indicative_phrase	semantic + semantic	lexical_chain + lexical_chain
	semantic	lexical_chain	lexical + lexical +lexical	indicative_word + indicative_word + indicative_word
interview	reference	personal_reference	semantic + smeantic + semantic	lexical_chain + lexical_chain + meronymy
	graphical	semicolon	semantic + semantic	lexical_chain + synonymy
essay	semantic	lexical_chain	lexical + lexical + lexical	indicative_word + indicative_word + indicative_word
	semantic	meronymy		
	lexical	alternate_expression		
	semantic	meronymy	reference + reference + semantic	demonstrative_reference + personal_reference + meronymy
podcast	reference	personal_reference	semantic + semantic	lexical_chain + synonymy
	lexical	indicative_word	lexical + lexical	indicative_word + indicative_word
	semantic	lexical_chain	lexical + lexical	indicative_word + indicative_word
speech	syntactic	parallel_synatactic_construction	reference + reference	personal_reference + personal_reference
	semantic	meronymy		
	semantic	lexical_chain	reference + semantic	demonstrative_reference + synonymy
court	reference	personal_reference	reference + reference + semantic	personal_reference + personal_reference + synonymy
	semantic	negation	reference + semantic	personal_reference + lexical_chain
	semantic	lexical_chain	reference + reference	personal_reference + personal_reference
letter	reference	personal_reference		
	semantic	meronymy		

Table 8: Pearson Correlation: Entropy Score and Total Co-occurred Signals

Variable Pair	**Correlation (r)**	**p-value**
Entropy Score - Total Co-occurred Signals	0.248	0.0137

Table 9: Model 1: Regression of Entropy Score on Normalized Signal Diversity

	Coefficient	**Std. Error**	**p-value**
Intercept	0.850	0.040	<0.001
Entropy Score	0.112	0.039	0.005
R^2		0.081	
Adjusted R^2		0.071	
F-statistic	8.44	($p = 0.0046$)	

Table 10: Model 2: Regression of Entropy Score and Total Signals on Normalized Signal Diversity

	Coefficient	**Std. Error**	**p-value**
Intercept	0.851	0.040	$<$0.001
Entropy Score	0.123	0.040	0.003
Total Co-occurred Signals	-0.0005	0.0005	0.302
R^2		0.091	
Adjusted R^2		0.072	
F-statistic	4.76	($p = 0.0107$)	

Table 11: Model 3: Regression of Within-Genre Entropy and Genre Interaction on Normalized Signal Diversity

Coefficient	**Coef.**	**Std. Error**	**p-value**
Intercept	0.866	0.034	<0.001
Within-Genre Entropy	0.055	0.033	0.098
Entropy × Genre[T.vlog]	0.132	0.050	0.009
Entropy × Genre[T.letter]	-0.119	0.061	0.053
Entropy × Genre[T.conversation]	0.075	0.047	0.116
Entropy × Genre[T.reddit]	0.073	0.047	0.123
Entropy × Genre[T.fiction]	0.077	0.055	0.162
Entropy × Genre[T.speech]	-0.066	0.049	0.178
R^2		0.121	
Adjusted R^2		0.090	
F-statistic		3.97 ($p < 0.000001$)	