Voices and Echoes in Fictional Dialogue: A Study of Linguistic Coordination in Literary Texts

Ioana-Roxana Boriceanu[1], Alina Iacob[1], Liviu P. Dinu[1,2]
[1]Faculty of Mathematics and Computer Science
[2]Human Language Technologies Research Center
University of Bucharest, Romania
{ioana-roxana.boriceanu, ioana.iacob}@s.unibuc.ro, ldinu@fmi.unibuc.ro

Abstract

This study investigates linguistic coordination in fictional dialogue, examining whether the phenomenon typically observed in natural conversation also appears in imagined exchanges created by authors. We analyse dialogues from ten English novels by Jane Austen and E. M. Forster using the Project Dialogism Novel Corpus (PDNC) to measure linguistic convergence across nine function word categories from the Linguistic Inquiry and Word Count (LIWC) lexicon, complemented by network based measures that capture how linguistic adaptation shapes interactions among characters. The results provide evidence of convergence in both authors, confirming that linguistic coordination extends to literary dialogue. The network analysis supports these findings, revealing that alignment is generally reciprocal, unevenly distributed but widespread, and often crosses social and narrative boundaries. Taken together, these results suggest that linguistic coordination in fiction does not depend on deliberate stylistic planning, but reflects underlying cognitive mechanisms involved in language processing and social interaction.

Keywords: linguistic coordination, fictional dialogue, stylistic accommodation, network analysis

1. Introduction and Related Work

Linguistic coordination is a long-studied phenomenon in linguistics, cognitive science, and social psychology, in an attempt to grasp the elusive nature of language processing during instances of dialogue. Different denominations have been used to describe the phenomenon in which interlocutors adapt different aspects of their speech (syntactic patterns, vocabulary choices, phonetics) to match patterns of previous utterances: alignment (Pickering and Garrod, 2004), accommodation (Giles et al., 1991), entrainment (Levitan and Hirschberg, 2011), coordination (Clark, 1996), convergence (Giles et al., 1991), linguistic style matching (Gonzales et al., 2010; Ireland and Henderson, 2014), correspondence effect (Heyes, 2011), and chameleon effect (Chartrand and Bargh, 1999). Throughout this paper, these terms are used interchangeably to describe the broader phenomenon of linguistic coordination. Despite terminological differences, all terms point to the same fundamental tendency: in conversation, the language of interlocutors becomes similar. Giles et al. (1991) defines "convergence" as "a strategy whereby individuals adapt to each other's communicative behaviours in terms of a wide range of linguistic–prosodic–nonverbal features."

Research has identified two main explanations for linguistic coordination. On the one hand, it is studied as a behaviour aimed at increasing social approval, as described by Communication Accommodation Theory (Giles et al., 1991); on the other hand, it is considered an automatic cognitive mechanism that facilitates the brain's processing of language during dialogue, as part of what Pickering and Garrod (2004) call the Interactive Alignment Model, according to which interlocutors mimic each other's speech patterns because they share a common mental representation during instances of speech production and comprehension. In other words, speakers tend to mimic one another's linguistic patterns not only for social reasons but also because this simplifies the cognitive load involved in real-time language processing. The fact that linguistic coordination arises during natural language use is well established (Litterer et al., 2025), yet the question remains as to what extent this process is a social mechanism and to what extent it is a cognitive one.

Computational studies have approached the social dimension of linguistic coordination from various angles, consistently showing that alignment reflects interpersonal dynamics. Higher linguistic similarity has been linked to greater engagement and group cohesion (Babcock et al., 2014; Gonzales et al., 2010), as well as to differences in power, with lower-status speakers aligning more to those of higher status (Danescu-Niculescu-Mizil et al., 2012). Other work associates language style matching with relational outcomes such as negotiation success, open-mindedness, and long-term compatibility (Ireland et al., 2011; Ireland and Henderson, 2014; Ben-Haim and Tsur, 2021). In contrast, research on the cognitive mechanisms of alignment suggests that convergence is not purely

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 2581–2593
11–16 May 2026.

automatic. Doyle and Frank (2016) showed that coordination occurs mainly at the lexical level and depends on conversational goals, indicating that speakers use alignment strategically as well as unconsciously.

A particularly compelling case for the automatic component of linguistic coordination comes from Danescu-Niculescu-Mizil and Lee (2011), who showed that linguistic coordination of function-words occurs even in "imagined conversations", namely in dialogues extracted from movie scripts. As the authors pointed out, "this provides evidence that coordination, assumed to be driven by social motivations, has become so deeply embedded into our ideas of what conversations 'sound like' that the phenomenon occurs even when the person generating the dialogue is not the recipient of the social benefits." While this does not deny the social purposes of linguistic coordination, it stands as evidence for its deeply automatic and unconscious nature and the role of priming.

While the linguistic coordination effect has been shown to be present in screenplay-extracted dialogues, it is not clear whether the same would hold true for novel-extracted dialogues. Screenplay dialogues are meant to be spoken, and they are often judged by how convincingly they reflect natural speech. Consequently, it is expected that they retain many patterns of naturally occurring dialogues. Literature, on the other hand, does not primarily aim for believability. According to Gérard Genette's foundational theory, literary works "are supposed to pretend to reproduce" a certain "model" (Genette, 1988). This double distancing from any pretence of reproducing reality suggests that novels constitute a highly stylised and consciously crafted linguistic construct. As a consequence, detecting the phenomenon of linguistic coordination in fictional dialogues would provide further proof that it is deeply embedded in an individual's cognitive processes, as part of the processes of language representation and production.

Building on these insights, we take a further step into the realm of imagined conversation by applying and extending the computational framework proposed by (Danescu-Niculescu-Mizil and Lee, 2011) for measuring linguistic convergence in the domain of fictional dialogue. To probe more deeply into the unconscious nature of linguistic coordination, we analyse dialogues extracted from novels and examine how linguistic adaptation operates in stylised, author created exchanges. The present study focuses on two corpora of English novels by Jane Austen and E. M. Forster and pursues three main goals:

1. Determine whether linguistic coordination can be detected in fictional dialogues extracted from novels.
2. Examine how patterns of linguistic coordination vary across different writers.
3. Explore how linguistic coordination structures shape characters' interactions, using network based measures.

By applying existing linguistic resources to a new domain of fictional dialogue, this study contributes to ongoing work exploring how language resources support diverse forms of linguistic analysis.

2. Data

In our study, we used the Project Dialogism Novel Corpus (PDNC) Vishnubhotla et al. (2022), a large, manually annotated collection of English novels designed for quotation attribution research. The corpus provides, for each quotation, the identified speaker and addressee(s), its quotation type (explicit, anaphoric, or implicit), and links to any characters mentioned within the utterance. Each novel also includes metadata describing the canonical names of characters, aliases, gender, and narrative importance.

As a first preprocessing step, we cleaned the raw quotations by removing punctuation and special symbols, and converted all text to lowercase. We retained the speaker, addressee, gender, importance, and cleaned text for each quote, and removed empty entries.

Since most authors in the PDNC are represented by only a single annotated novel, we focus on the six novels by Jane Austen and the four novels by E. M. Forster, the two authors with the largest number of works in the corpus. Although this selection was based on data availability, it is particularly appropriate from a literary standpoint, as Forster himself acknowledged Austen as his model, admiring her mastery of domestic comedy while extending it towards social critique and reflection on personal relations (Colmer, 1982). This connection adds contextual depth to the analysis and enriches the interpretation of linguistic patterns observed in their works.

Table 1 presents an overview of the data used in this study. The first column lists the two authors, followed by their individual works. For each novel, the table reports the total number of annotated quotations, the number of tokens contained in quoted dialogue, and the number of distinct speaking characters (Chars) identified in the corpus metadata. Subtotals are provided per author, along with overall corpus totals at the bottom.

3. Method

Our approach builds on the convergence framework proposed by Danescu-Niculescu-Mizil and

Author	Novel	Quotes	Tokens	Chars
Jane Austen	*Emma*	1,593	76,141	18
	Mansfield Park	1,157	58,121	36
	Northanger Abbey	842	28,227	20
	Persuasion	488	26,252	35
	Pride and Prejudice	1,270	48,132	74
	Sense and Sensibility	1,085	51,184	23
	Subtotal	**6,435**	**288,057**	**206**
E. M. Forster	*A Passage to India*	2,123	37,088	42
	A Room with a View	1,634	29,694	63
	Howards End	2,606	46,709	51
	Where Angels Fear to Tread	1,000	18,560	18
	Subtotal	**7,363**	**132,051**	**174**
Total		**13,798**	**420,108**	**380**

Table 1: Corpus summary

Lee (2011), which measures linguistic coordination between fictional characters in movie dialogue. Since one of our main questions was whether such coordination also occurs in literary dialogue, we applied the same method to our corpus of novels. In addition, we constructed a convergence network for each novel and analysed several network-level metrics to describe interaction patterns among characters.

3.1. Data Preprocessing

For each novel, we used the annotated quotations provided in the PDNC dataset. For consistency with Danescu-Niculescu-Mizil and Lee (2011), we analysed nine LIWC-derived linguistic categories (Pennebaker et al. (2015), Ireland et al. (2011)) that are generally known to be processed non-consciously by humans (Chung and Pennebaker, 2011). Table 2 lists these nine categories, together with the number of lexemes in the LIWC dictionary and some representative examples.

Furthermore, consecutive quotes by the same speaker and addressed to the same character were merged into single dialogue turns. From these turns, we built pairs of adjacent utterances representing exchanges between two speakers, where one turn by speaker A ($A \rightarrow B$) is followed by a reply from speaker B ($B \rightarrow A$). Two control sets were also created: (1) non-adjacent pairs, where the reply of speaker B skips one turn, and (2) randomised pairs, where replies were permuted within each A, B dyad. For the randomised condition, we generated a single permutation per dyad without replacement, so that each reply appeared exactly once, and used a fixed random seed (set to 123) to ensure reproducibility. These control pairs allow us to test whether any observed convergence reflects genuine local adaptation rather than general similarities in language use.

Family	LIWC Count	Examples
articles	3	a, an, the
auxiliary verbs	144	am, have, do
conjunctions	28	and, or, but
adverbs	69	very, really
impersonal pronouns	46	it, those
negations	57	no, not, never
personal pronouns	70	I, you, she
prepositions	60	in, on, at
quantifiers	89	many, few, some
Total lexemes	**566**	–

Table 2: The LIWC families used in the analysis

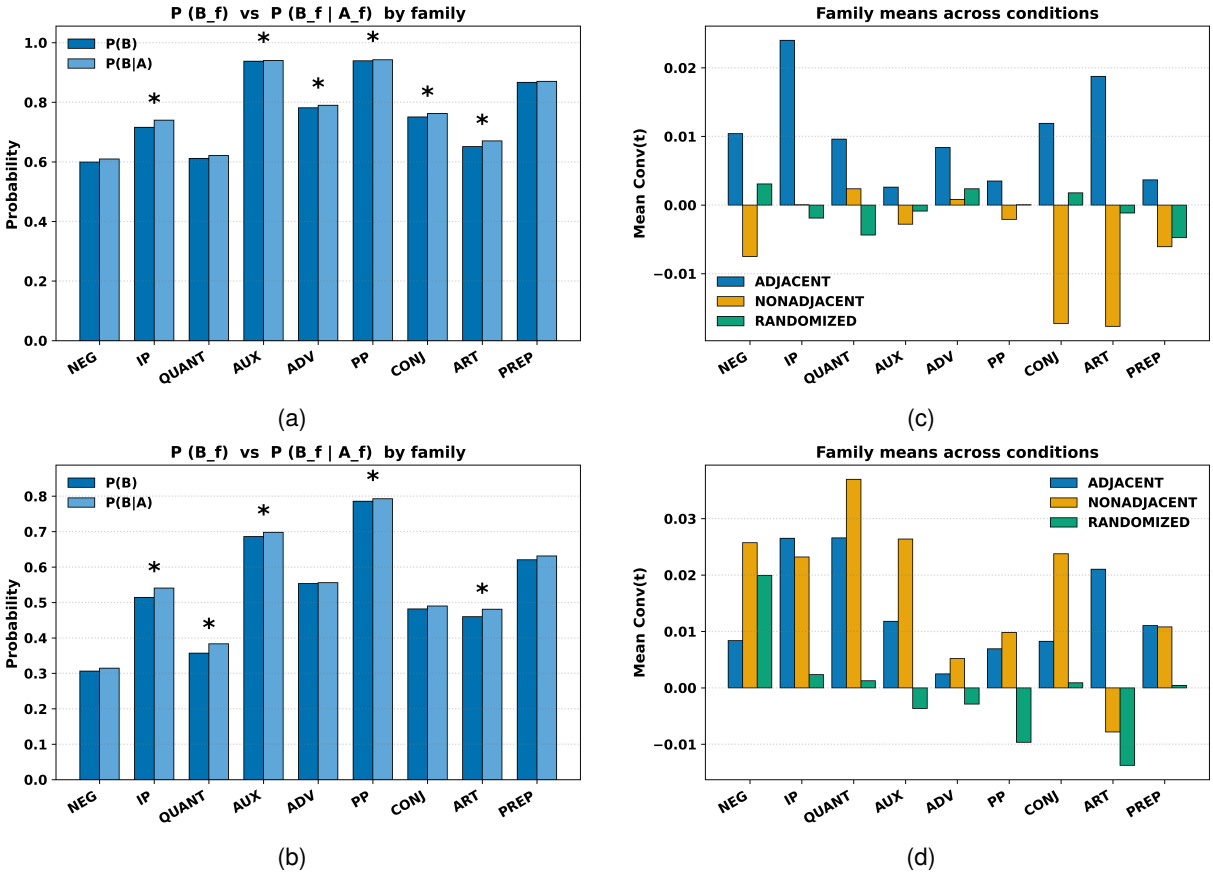
3.2. Convergence Computation

To measure linguistic coordination, we followed the method introduced by Danescu-Niculescu-Mizil and Lee (2011). For each dialogue pair in which speaker A addresses speaker B ($A \rightarrow B$), we examined whether B's utterance contained any words from a given LIWC family after A had used that same family in the preceding turn. Convergence for each linguistic family f was computed as the difference between the conditional probability that B uses f given that A used it, and the baseline probability that B uses f in general:

$$\text{Conv}_{A,B}(f) = P(B_f \mid A_f) - P(B_f) \qquad (1)$$

A positive value of Conv indicates that B is more likely to use a given linguistic feature after A has used it, signalling convergence. Negative values suggest divergence, while values around zero imply no systematic adaptation.

The computation was performed separately for each A–B dyad and each linguistic family. To ensure reliable estimates, we retained a dyad–family (A, B, f) only if speaker A produced at least k triggers for f (at least k turns with $A_f = 1$), we set $k = 3$. In preliminary checks, thresholds $k < 3$ produced unstable estimates of $P(B_f \mid A_f)$, reflected in wide bootstrap confidence intervals for low-frequency families, whereas $k > 3$ substantially reduced coverage by eliminating many dyads involving minor characters and yielding sparse networks. With $k = 3$, we retain a broad set of interacting pairs while filtering extremely low-frequency



Baseline probability $P(B_f)$ vs. conditional probability $P(B_f \mid A_f)$ by LIWC family. Asterisks mark families where the difference is statistically significant ($p < 0.05$).

Mean convergence scores Conv by LIWC family across three conditions: adjacent (true dialogue pairs), non-adjacent (skip-one), and randomised controls.

Figure 1: Linguistic convergence patterns in the Austen ((a), (c)) and Forster ((b), (d)) corpora.

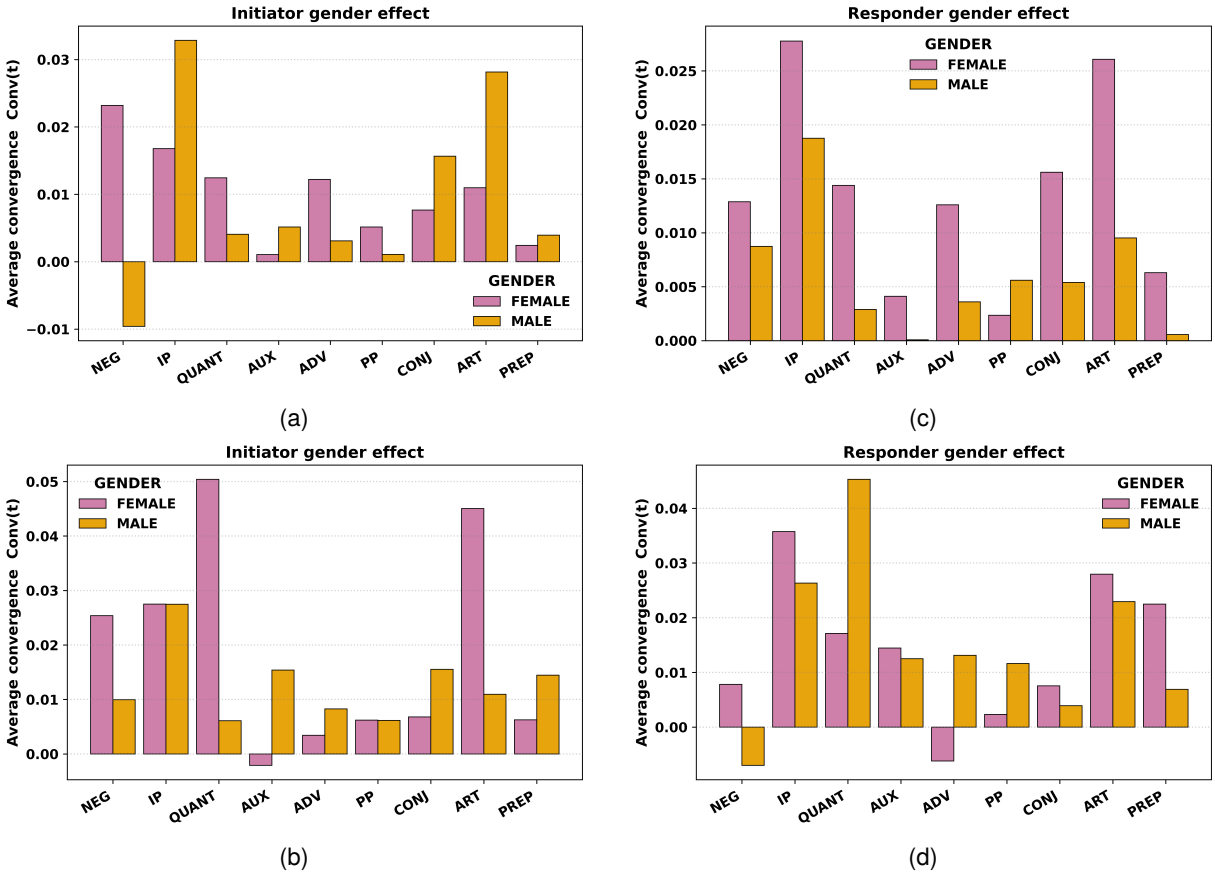
cases that would otherwise distort conditional probability estimates.

3.3. Convergence Network

Beyond pairwise convergence, we examined the structure of accommodation and interaction patterns at the character level within each novel as a directed, weighted network. In this network, nodes correspond to speaking characters, and a directed edge from A to B indicates that speaker A addressed speaker B at least once in the dialogue. The weight of each edge represents the mean convergence value between the two speakers, averaged across the nine LIWC families.

To better understand how linguistic accommodation shapes social interaction, we examined the resulting networks to identify patterns of adaptation between characters. We computed several graph-based metrics to capture different aspects of interaction and coordination.

To determine how much characters interact in each novel, we measured network density, which reflects how many character pairs interact relative to all possible pairs. To assess whether accommodation tends to be mutual, we examined reciprocity, indicating how often linguistic influence flows in both directions rather than one-sidedly. The mean edge weight was used to estimate the overall strength of adaptation between speakers, while the balance between positive and negative edges distinguishes novels where linguistic alignment dominates from those characterised by stylistic divergence. Measures of gender and category assortativity were used to test whether accommodation tends to occur within socially or narratively similar groups. We also analysed modularity and community structure using greedy modularity optimisation on an undirected projection of the dialogue network, in order to examine whether conversations form distinct clusters of closely interacting characters rather than a single integrated dialogue space. Finally, inequality measures such as the edge-weight Gini coefficient and PageRank entropy show how evenly linguistic influence is distributed, distinguishing networks in which stylistic coordination is shared among many speakers from those dominated by a few central figures.



Initiator gender effect - how convergence toward female or male initiators varies across families.

(e) Responder gender effect - how female and male responders differ in adaptation strength across families.

Figure 2: Gender effects on linguistic convergence in the Austen ((a), (c)) and Forster ((b), (d)) corpora.

4. Experiments and Results

We applied the convergence analysis and network based measures described in Section 3 to the ten selected works. For each author, we first computed the metrics across all texts to obtain an overall view of convergence patterns, and then carried out a more detailed analysis for each individual work, computing linguistic convergence for all speaker pairs and using the resulting values to construct the corresponding convergence network.

For consistency across figures, the x-axis labels in the bar plots use abbreviated forms of the nine linguistic families: negations (NEG), impersonal pronouns (IP), quantifiers (QUANT), auxiliary verbs (AUX), adverbs (ADV), personal pronouns (PP), conjunctions (CONJ), articles (ART), and prepositions (PREP).

4.1. Convergence in Literary Texts

Figure 1 presents convergence patterns in the Austen and Forster corpora. Statistical significance in Figures 1a and 1b was assessed using paired-sample t-tests on per-pair differences $(P(B_f \mid A_f) - P(B_f))$. Confidence intervals were estimated by nonparametric bootstrap resampling.

In the Austen corpus, Figure 1a shows that the conditional probability that a speaker uses a linguistic feature after the interlocutor has used it $(P(B_f \mid A_f))$ is higher than the baseline probability $(P(B_f))$ for most LIWC families. The difference is significant for almost all categories except negations, quantifiers, and prepositions, indicating a consistent tendency towards linguistic adaptation. Figure 1c compares convergence across control conditions and shows that only adjacent dialogue pairs, corresponding to true conversational exchanges, display positive convergence. Non-adjacent and randomised pairs remain close to zero. These results confirm that the observed alignment reflects immediate conversational adaptation rather than general stylistic similarity.

As in the case of Jane Austen, the fictional dialogues of E. M. Forster exhibit measurable linguistic accommodation, as shown in Figure 1b. The conditional probabilities $(P(B_f \mid A_f))$ are generally higher than the baselines $(P(B_f))$ for most function-word families, indicating a tendency towards alignment. However, compared with the Austen corpus,

the Forster data reveal a more selective pattern. Figure 1d shows that only three out of nine feature families display higher convergence for adjacent dialogue pairs than for the control conditions, while in most cases the values are close to zero. This suggests that linguistic adaptation in the works of Forster occurs less consistently.

4.1.1. Gender Influence on Convergence

Similar to Danescu-Niculescu-Mizil and Lee (2011), we examined the effect of gender on linguistic convergence across the ten novels. We focused on two aspects: first, whether characters adapt differently depending on the gender of the initiator, and second, whether one gender tends to converge more than the other. The results for the two analysed authors are shown in Figure 2. The comparisons presented here are descriptive and exploratory; the values represent mean convergence differences, and no formal significance testing was performed.

One interesting observation is that, for both authors (Figures 2c and 2d), female characters generally adapt more than male characters. Regarding the effect of initiator gender, no clear pattern emerges, as the way characters adapt across the nine LIWC families differs for each author. We also found that, in most novels by Jane Austen, characters adapt less in same-gender interactions, with *Pride and Prejudice* being the main exception, where women show stronger adaptation towards each other. In the novels of E. M. Forster, by contrast, male characters tend to accommodate more when speaking with other men than with women.

4.1.2. Role Influence on Convergence

In the PDNC corpus, the characters are divided into three categories: major, intermediate, and minor, according to their narrative importance. We analysed whether narrative importance influences linguistic adaptation, following the same approach used for gender. The results for both authors are shown in Figure 4.

The results show that, in the Austen corpus, characters tend to adapt more when interacting with major or intermediate speakers, while minor characters show a more variable pattern across families. When responding, major and intermediate characters also exhibit stronger convergence, suggesting that more central characters are both more influential and more responsive in dialogue.

In the works of Forster, the pattern is less consistent. Minor initiators occasionally elicit higher convergence, but overall differences across importance levels are smaller. This suggests that narrative importance has a weaker impact on linguistic alignment in the dialogues of Forster.

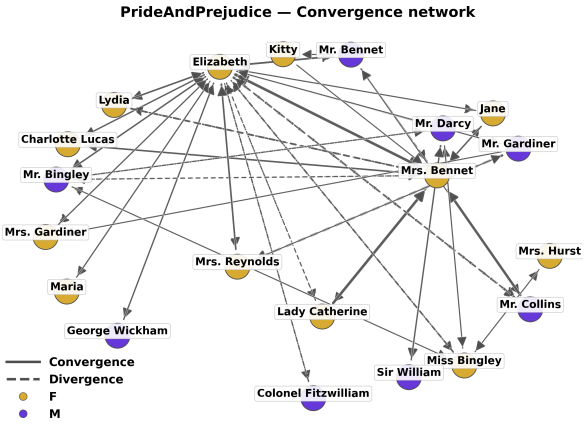


Figure 3: Example of a convergence network

4.2. Network Analysis

Metric	Austen	Forster
Nodes	95	74
Edges	312	237
Density	0.035	0.044
Reciprocity	0.88	0.84
Positive edges	0.48	0.54
Negative edges	0.31	0.26
Edge-weight Gini	0.57	0.54
PageRank entropy	0.92	0.92
Gender assortativity	-0.16	0.03
Category assortativity	-0.21	-0.09

Table 3: Summary of convergence network metrics for the aggregate Austen and Forster corpora.

Figure 3 illustrates an example of a convergence network from the Austen corpus. Each node represents a character, coloured based on gender, and directed edges indicate the strength and direction of linguistic adaptation between them. Solid edges correspond to positive convergence, showing instances where speakers align linguistically with their interlocutors, while dashed edges indicate divergence, representing cases where the linguistic behaviour of one speaker moves away from that of the other. Thicker edges denote stronger effects, and colour intensity reflects the magnitude of adaptation.

We remind that characters are included in the convergence network only if they participate in at least one dyad–family (A, B, f) satisfying the trigger threshold $k \geq 3$; speakers who do not meet this criterion for any family are excluded from the network representation. This filtering reduces the set of speakers from 206 to 95 in Austen and from 174 to 74 in Forster, primarily affecting peripheral characters with few dialogue turns.

Table 3 summarises the main structural properties of the aggregate convergence networks for the

two authors, facilitating direct comparison across metrics. For completeness, the full set of network metrics computed for each individual novel is provided in Table 4.

In the Austen corpus, the convergence network, constructed from all six novels, comprises 95 characters connected by 312 directed edges, resulting in a low density (0.035) and high reciprocity (0.88). The low density is expected in novels, where only a subset of characters interact directly within the story world, while the high reciprocity indicates that most of these interactions are mutual rather than one-sided. This reflects the turn-taking structure of natural dialogue, where conversational exchanges typically involve responses in both directions.

The proportion of positive edges (48%) exceeds that of negative ones (31%), confirming that alignment is more frequent than divergence and supporting the earlier evidence for consistent convergence across linguistic families. Variation in adaptation strength is moderate (edge-weight Gini $= 0.57$), and measures of influence distribution such as PageRank entropy (0.92) and the top 10% strength share (0.37) indicate that linguistic influence is uneven but broadly shared among characters. This suggests that coordination in Austen's fiction is a collective feature of the dialogue network rather than the outcome of a single dominant speaker. Both gender and category assortativity are slightly negative ($r = -0.16$ and $r = -0.21$), indicating that convergence edges are more likely to connect characters across social and narrative categories than within homogeneous groups.

In the Forster corpus, the convergence network comprises 74 characters connected by 237 directed edges, yielding a slightly higher density (0.044) and a reciprocity of 0.84. The reciprocity indicates that conversations are largely bidirectional, though somewhat less so than in Austen's works, suggesting a modestly more asymmetrical flow of adaptation.

The proportion of positive edges (54%) exceeds that of negative ones (26%), confirming that linguistic coordination is again the dominant tendency. Variation in adaptation strength is moderate (edge-weight Gini $= 0.54$), while measures of influence distribution such as PageRank entropy (0.92) and the top 10% strength share (0.35) show that stylistic influence is uneven but not concentrated in a few individuals. This points to a conversational structure where linguistic coordination is shared among several key figures rather than centralised. Gender assortativity is slightly positive ($r = 0.03$), and category assortativity is weakly negative ($r = -0.09$), implying that accommodation occurs both within and across social or narrative boundaries.

Overall, the network results are consistent with the convergence patterns described earlier. In Austen's novels, linguistic adaptation is frequent, balanced, and broadly shared among characters, reflecting a cohesive conversational structure in which alignment contributes to social integration. In Forster's works, convergence is also present but less evenly distributed, suggesting that adaptation depends more on context and character relations.

The community analysis offers further insight into how linguistic coordination structures dialogue in fiction. A full listing of detected communities and their representative members for each novel is provided in Table 5. In Austen's novels, the resulting networks show socially coherent groupings. Across the six works, characters cluster into three or four well-defined communities that broadly mirror family ties, recurring interaction spheres, and differences in social position. In *Pride and Prejudice*, for instance, one cluster centres on the Bennet household and their extended connections, another groups Elizabeth's wider social contacts such as the Gardiners and other key intermediaries, while a separate cluster brings together Darcy, Bingley, and their immediate circle. In *Sense and Sensibility*, the Dashwood-centred group forms one major cluster, while the Steele sisters and the surrounding Barton Park society constitute another prominent conversational sphere. The moderately high modularity values indicate that convergence occurs most strongly within these interaction circles rather than being evenly distributed across the entire narrative network.

Forster's novels display a similar but often more sharply segmented structure. In *A Passage to India*, Indian and British characters largely occupy separate clusters, with a smaller mixed community reflecting bridging figures such as Fielding. In *Howards End*, three principal groups emerge around the Wilcox family, the Schlegels, and Leonard Bast, corresponding closely to the novel's central social divisions. *A Room with a View* shows communities structured around romantic and familial ties, while *Where Angels Fear to Tread* separates a Herriton-centred English circle from a cluster anchored in the Italian setting, with limited overlap. Modularity scores again point to a moderately strong partitioning, suggesting that in Forster's fiction linguistic adaptation tends to follow existing social and cultural lines rather than dissolving them.

Taken together, the network and community analyses reveal how linguistic convergence shapes the social texture of fictional worlds. They show that alignment in dialogue is not random but reflects the social hierarchies, cultural boundaries, and interpersonal ties that structure each narrative. This approach offers a way to trace how linguistic coordination contributes to characterisation and social dynamics, highlighting differences between authors

in how they construct interaction and depict community. By mapping convergence across characters, we gain a clearer view of how writers use language to represent social cohesion, distance, and conflict within their fictional societies.

4.3. Discussion

The results show that linguistic convergence, originally described as a feature of natural dialogue, can also be observed in fictional conversations extracted from novels. Both Austen and Forster display measurable patterns of linguistic alignment across several function-word families, confirming that convergence emerges even in imagined exchanges written by authors rather than produced in real interaction.

Our findings are broadly consistent with the work of Danescu-Niculescu-Mizil and Lee (2011) on imagined dialogue, which likewise reports function-word convergence across interlocutors. In this respect, literary dialogue aligns with earlier studies in showing that convergence remains detectable even outside face to face interaction. At the same time, the effects observed here are less uniform across families and works, especially in the Forster corpus, suggesting that convergence in novel dialogue may be weaker or more context dependent than in less mediated conversational settings. Since the corpora, preprocessing decisions, and interaction types differ across studies, we treat this comparison as strictly qualitative.

Unlike the overall convergence effect, gender related patterns are less directly comparable across studies, since they are likely shaped by corpus specific narrative and social factors. In our data, these effects vary across novels and are therefore treated as exploratory. Taken together with earlier findings on Twitter and movie dialogue, our results suggest that linguistic convergence persists across increasingly mediated forms of interaction, although its strength appears less consistent in highly stylised literary dialogue.

The differences between the two authors suggest that convergence is also influenced by stylistic and narrative factors. Austen's dialogues show stronger and more consistent adaptation, higher reciprocity, and more cross-group accommodation, indicating that her characters are portrayed as socially responsive and linguistically attuned to one another. In contrast, Forster's networks reveal weaker and more uneven convergence, with alignment varying across contexts and character roles. This may reflect his focus on social distance and tension, where dialogue often highlights difference rather than agreement.

Another interesting observation emerged when examining the novels individually. The degree of linguistic convergence and the corresponding network properties vary across works, suggesting that these measures may also capture aspects of each author's stylistic profile. For Austen, *Mansfield Park* (1814) shows the strongest evidence of convergence, with frequent and balanced adaptation among characters, reflecting the novel's emphasis on extended social interactions within a stable community. In contrast, *Northanger Abbey* (completed in 1799, published in 1817) displays no significant convergence effect, possibly due to its lighter, more parodic tone and smaller range of speakers. *Emma* (1815) and *Persuasion* (1817) exhibit similar moderate convergence and cohesive dialogue networks that mirror their focus on nuanced, character-driven exchanges.

For Forster, *A Room with a View* (1908) and *Howards End* (1910) show stronger and more consistent convergence, with cohesive and interconnected character networks that reflect socially harmonious worlds. In contrast, *A Passage to India* (1924) and *Where Angels Fear to Tread* (1905) display weaker convergence and more fragmented interaction structures. The compact and highly reciprocal network of *Where Angels Fear to Tread* reflects its limited set of speakers and its focus on a divided community, while the fragmented network of *A Passage to India* mirrors themes of cultural distance and misunderstanding. These differences may indicate variation in the author's style over time; however, additional analysis would be needed to confirm this interpretation.

Beyond these authorial contrasts, the emergence of convergence in these fictional dialogues indicates that the phenomenon is not dependent on direct conversational exchange but arises from cognitive processes inherent in language use and the mental modelling of interaction. Our analysis focuses on function-word categories precisely because these elements are highly frequent, carry little independent semantic content, and are typically produced without deliberate stylistic control. Unlike lexical choices, which authors may manipulate consciously for characterisation or rhetorical effect, function words largely reflect grammatical structure and discourse organisation. For this reason, systematic conditional dependencies in these categories are less likely to result from explicit stylistic planning and more plausibly reflect automatic priming mechanisms involved in the mental simulation of dialogue. The fact that such alignment emerges even in fully imagined exchanges suggests the mechanisms driving linguistic alignment extend beyond immediate communicative goals, reflecting how language naturally encodes patterns of responsiveness and reciprocity. At the same time, the structure of the convergence networks, which reproduce distinctions of class, gender, and hier-

archy, highlights the social dimension of the phenomenon. This evidence supports accounts that view alignment as rooted in general mechanisms of language processing and social coordination.

While these findings provide insight into the dynamics of linguistic coordination in fiction, several limitations should be noted. First, the analysis covers a limited number of authors and novels, which restricts how far the results can be generalised. It would be interesting to examine how the coordination mechanism manifests across different authors, genres, and literary movements. Second, our focus on LIWC-based function-word categories captures linguistic coordination at a subtle, non-conscious level. A more extensive analysis could incorporate higher-level linguistic features such as syntax or semantics to provide a more comprehensive view of alignment. Finally, the network measures capture structural tendencies that offer a useful abstraction of interaction, though they cannot fully represent the contextual and narrative complexity of character relations.

5. Conclusions

This study examined linguistic convergence in fictional dialogue, using novels by Jane Austen and E. M. Forster as case studies. By applying established measures of linguistic coordination and combining them with network based analyses, we showed that convergence, typically associated with natural conversation, also appears in written, imagined exchanges. This supports the view that linguistic adaptation reflects underlying cognitive mechanisms that influence how dialogue is represented, even outside real social interaction.

At the same time, differences between authors and between individual works indicate that convergence is also shaped by narrative structure and stylistic intent. Austen's dialogues show stronger and more consistent alignment, while Forster's display greater variability, often reflecting themes of distance and misunderstanding. The presence of convergence in consciously crafted texts suggests that even stylised representations of speech preserve traces of natural conversational behaviour. These patterns also indicate that linguistic coordination may serve as a quantitative marker of stylistic, thematic, and social contrasts in literature.

The results also have implications for both cognitive and literary research. From a cognitive perspective, they provide additional evidence that linguistic coordination is deeply rooted in the mechanisms of language processing and representation, suggesting that future studies should give greater attention to cognitive factors alongside social explanations. From a literary perspective, they highlight the hybrid nature of fiction as a linguistic and cultural artefact. Contemporary literary theory increasingly emphasises the connections between literature and extra literary domains such as cognitive science, sociology, and economics, moving beyond the traditional view of art for art's sake. The presence of cognitive mechanisms such as linguistic coordination, modulated by authors' stylistic and discursive choices as our results show through variation across characters and writers, offers new insights into the interplay between cognition, style, genre, and literary design.

Future work will extend this approach to a broader range of authors and genres to explore whether linguistic coordination can serve as a stylistic marker in literary writing and how it contributes to the social and aesthetic dimensions of fictional dialogue. It would also be informative to examine texts by amateur or less experienced writers to test whether similar coordination patterns emerge and whether they correlate with perceived quality, credibility, or readability. In addition, we plan to apply this framework to other languages to examine the extent to which convergence patterns generalise across linguistic systems. While the network based analysis can be transferred directly, the convergence measures based on function-word categories may require language specific adaptation.

6. Acknowledgements

This work was supported by Ministry of Research, Innovation and Digitization, CNCS-UEFISCDI, project SIROLA, number PN-IV-P1- PCE-20231701, within PNCDI IV and by the project "Romanian Hub for Artificial Intelligence - HRIA", Smart Growth, Digitization, and Financial Instruments Program, 2021-2027, MySMIS no. 334906.

Ethical Considerations

This study uses only publicly available literary texts contained in the Project Dialogism Novel Corpus (PDNC), which is distributed for research and educational purposes under appropriate licenses. No personally identifiable or sensitive information is included in the dataset. All analyses were performed on textual materials originally published in the public domain. The research does not involve human participants, experiments, or any data that would raise ethical or privacy concerns. The authors affirm that the study complies with the ethical standards and with good research practices regarding transparency, reproducibility, and respect for intellectual property. The scripts used for preprocessing, convergence computation, and network construction will be made available upon request to facilitate reproducibility.

7. Bibliographical References

Meghan J Babcock, Vivian P Ta, and William Ickes. 2014. Latent semantic similarity and language style matching in initial dyadic interactions. *Journal of Language and Social Psychology*, 33(1):78–88.

Aviv Ben-Haim and Oren Tsur. 2021. Open-mindedness and style coordination in argumentative discussions. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1876–1886, Online. Association for Computational Linguistics.

Tanya L Chartrand and John A Bargh. 1999. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893.

Cindy Chung and James Pennebaker. 2011. The psychological functions of function words. In *Social communication*, pages 343–359. Psychology Press.

Herbert H Clark. 1996. *Using language*. Cambridge university press.

John Colmer. 1982. Marriage and personal relations in forster's fiction. In *EM Forster: Centenary Revaluations*, pages 113–123. Springer.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. *arXiv preprint arXiv:1106.3077*.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, pages 699–708.

Gabriel Doyle and Michael C Frank. 2016. Investigating the sources of linguistic alignment in conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 526–536.

Gérard Genette. 1988. *Narrative discourse revisited*. Cornell University Press.

Howard Giles, Nikolas Coupland, and Justine Coupland. 1991. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, 1:1–68.

Amy L Gonzales, Jeffrey T Hancock, and James W Pennebaker. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37(1):3–19.

Cecilia Heyes. 2011. Automatic imitation. *Psychological bulletin*, 137(3):463.

Molly E Ireland and Marlone D Henderson. 2014. Language style matching, engagement, and impasse in negotiations. *Negotiation and conflict management research*, 7(1):1–16.

Molly E Ireland, Richard B Slatcher, Paul W Eastwick, Lauren E Scissors, Eli J Finkel, and James W Pennebaker. 2011. Language style matching predicts relationship initiation and stability. *Psychological science*, 22(1):39–44.

Rivka Levitan and Julia Bell Hirschberg. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.

Benjamin Roger Litterer, David Jurgens, and Dallas Card. 2025. Coordinating chaos: A structured review of linguistic coordination methodologies. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 24984–24999.

Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2):169–190.

8. Language Resource References

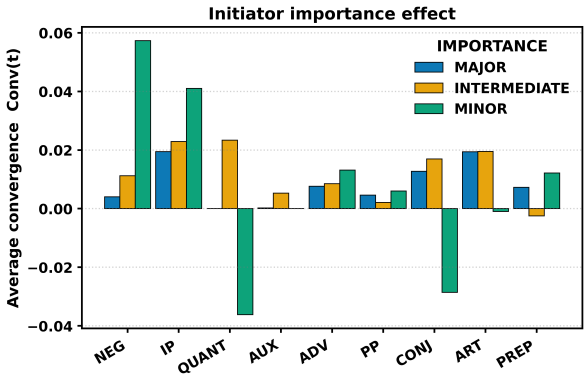
Pennebaker, James W and Boyd, Ryan L and Jordan, Kayla and Blackburn, Kate. 2015. *The development and psychometric properties of LIWC2015*.

Vishnubhotla, Krishnapriya and Hammond, Adam and Hirst, Graeme. 2022. *The project dialogism novel corpus: A dataset for quotation attribution in literary texts*.

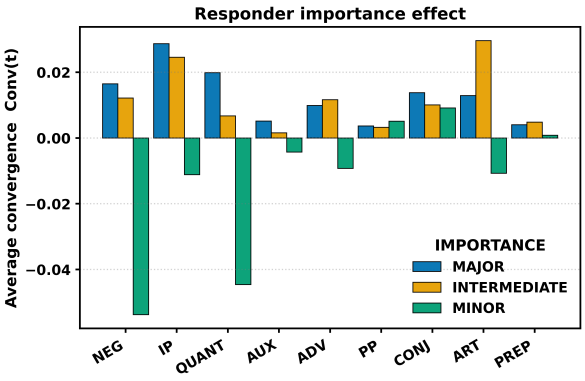
Appendix

A. Role Influence on Convergence

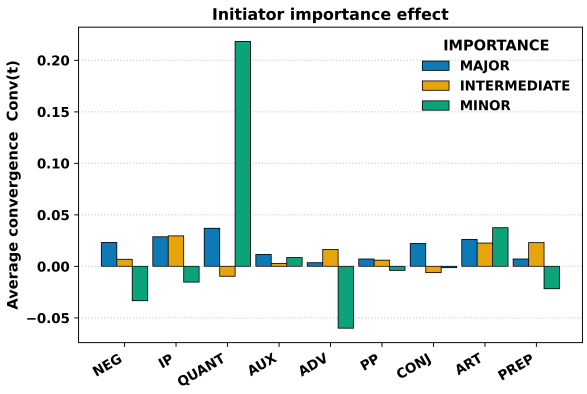
Figure 4 presents convergence patterns as a function of narrative importance, distinguishing major, intermediate, and minor characters as initiators and responders. The comparisons are descriptive and exploratory, and values represent mean convergence differences across dyads within each role category.



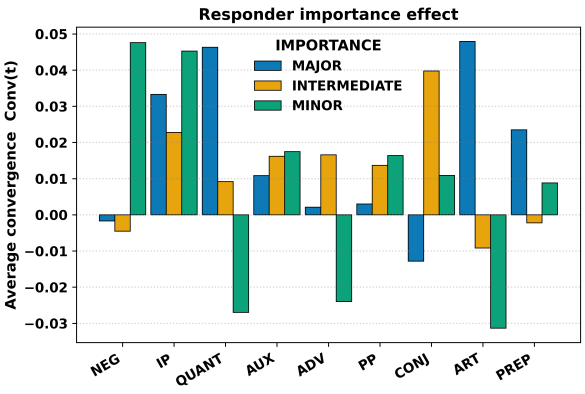
(a)



(c)



(b)



(d)

Initiator role effect - how convergence toward major, intermediate, or minor initiators varies across families.

(e) Responder role effect - how major, intermediate, and minor responders differ in adaptation strength across families.

Figure 4: Role effects on linguistic convergence in the Austen ((a), (c)) and Forster ((b), (d)) corpora.

Novel	Nod	Edg.	Dens.	Rec.	Pos.	Neg.	Gini	PR Ent.	Mod.	GenAs.	RoAs.
Emma	13	62	0.40	0.94	0.47	0.40	0.58	0.92	0.25	-0.10	-0.18
Mansfield Park	17	62	0.23	0.87	0.50	0.32	0.60	0.89	0.46	-0.22	-0.19
Northanger Abbey	12	32	0.24	0.81	0.34	0.34	0.68	0.87	0.26	-0.29	-0.38
Persuasion	17	37	0.14	0.81	0.30	0.40	0.64	0.85	0.42	-0.20	-0.25
Pride and Prejudice	20	63	0.17	0.95	0.43	0.36	0.60	0.80	0.17	-0.05	-0.38
Sense and Sensibility	17	58	0.21	0.86	0.41	0.38	0.52	0.90	0.35	-0.27	-0.23
A Passage to India	26	82	0.13	0.83	0.52	0.22	0.59	0.88	0.42	0.03	-0.08
A Room with a View	15	61	0.29	0.82	0.51	0.28	0.64	0.85	0.27	-0.07	0.01
Howards End	22	71	0.15	0.90	0.52	0.35	0.50	0.88	0.37	-0.07	-0.16
Where Angels Fear to Tread	9	24	0.33	0.92	0.54	0.33	0.57	0.86	0.29	-0.10	-0.24

Table 4: Per-novel convergence network metrics.

B. Per-novel Network Metrics and Communities

Table 4 reports the full set of network metrics for each individual novel, complementing the aggregate results presented in Section 4.2.

Nodes denote the number of characters retained in the convergence network after applying the trigger threshold $k \geq 3$. Edges (Edg.) represent the number of directed character pairs for which a convergence score is defined. Density (Dens.) is the proportion of observed directed edges relative to all possible directed pairs. Reciprocity (Rec.) indicates the proportion of edges that are bidirectional, reflecting how often convergence is mutual between two characters.

Pos. denotes the proportion of edges with positive mean convergence, indicating alignment rather

than divergence. Neg. denotes the proportion of edges with negative mean convergence, indicating divergence rather than alignment. Accordingly, $1 - \text{Pos.} - \text{Neg.}$ corresponds to edges with exactly zero mean convergence, if any occur under the averaging and rounding used. Gini refers to the edge weight Gini coefficient, measuring inequality in adaptation strength across edges. PageRank entropy (PR Ent.) quantifies how evenly influence is distributed across characters based on weighted PageRank centrality. Modularity (Mod.) captures the strength of community structure detected via modularity optimisation. GenAs. and RoAs. denote gender and category assortativity coefficients, respectively, indicating the tendency of characters to converge preferentially within the same gender or narrative category.

Table 5 lists the communities detected in each novel via greedy modularity optimisation applied to the undirected projection of the convergence networks. For each novel, all characters retained after applying the trigger threshold are grouped into clusters, corresponding to densely connected interaction spheres. These communities typically reflect recurring conversational circles structured around family ties, social affiliation, or shared narrative position, and complement the modularity values reported in Table 4.

Novel	Community	Characters
Jane Austen		
Emma	C1	Mr. John Knightley, Mrs. Weston, Harriet Smith, Isabella, Mr. Elton, Emma
	C2	Miss Bates, Mr. Knightley, Mrs. Elton, Jane Fairfax
	C3	Mr. Woodhouse, Mr. Weston, Mr. Frank Churchill
Mansfield Park	C1	Maria Bertram, Mr. Yates, Henry Crawford, Tom Bertram, Mary Crawford, Edmund, Mrs. Grant, Julia Bertram
	C2	Lady Bertram, Mr. Rushworth, Sir Thomas, Fanny Price, Mr. Price, William Price
	C3	Mrs. Rushworth, Mrs. Norris
Northanger Abbey	C1	Catherine Morland, Mrs. Morland, General Tilney, Eleanor Tilney, Henry Tilney
	C2	Isabella Thorpe, John Thorpe, James Morland, Frederick Tilney
	C3	Mrs. Thorpe, Mrs. Allen, Mr. Allen
Persuasion	C1	Mary, Captain Harville, Charles Musgrove, William Walter Elliot, Lady Russell, Mrs Smith, Mary Musgrove, Anne Elliot, Admiral Croft
	C2	Elizabeth, Mrs Clay, Sir Walter Elliot, Mr Shepherd
	C3	Mrs Musgrove, Mrs Croft, Captain Wentworth, Louisa Musgrove
Pride and Prejudice	C1	Kitty, Jane, Mrs. Bennet, Charlotte Lucas, Mr. Collins, Mr. Bennet, Lydia, Lady Catherine De Bourgh
	C2	Mr. Gardiner, Mrs. Gardiner, Mrs. Reynolds, Elizabeth, Colonel Fitzwilliam, George Wickham, Maria
	C3	Mr. Darcy, Miss Bingley, Sir William Lucas, Mr. Bingley, Mrs. Hurst
Sense and Sensibility	C1	Mr. Willoughby, Marianne, Thomas, Edward Ferrars, Elinor, Margaret, Mrs. Dashwood
	C2	Lucy Steele, Colonel Brandon, Anne Steele, John Middleton, Mrs. Jennings, Lady Middleton
	C3	Charlotte, Mr. Palmer
	C4	John Dashwood, Fanny
E. M. Forster		
A Passage to India	C1	Hassan, Dr. Aziz, Mohammed Latif, Rafi, Ralph, Ram Chand, Mahmoud Ali, Dr. Panna Lal, Hamidullah
	C2	Adela Quested, The Nawab Badahur, Ronny Heaslop, Mrs. Turton, Mrs. Moore, Miss Derek, Mr. Turton
	C3	Mr. Mcbryde, Major Callendar, Mr. Das, Mr. Amritrao, Mrs. Lesley
	C4	Mr. Haq, Cyril Fielding, Professor Godbole, A Stray Subaltern
A Room with a View	C1	Mr. Beebe, George Emerson, Miss Catherine Alan, Mr. Vyse, Mr. Emerson, Freddy
	C2	Rev. Cuthbert Eager, Miss Bartlett, Lucy, Powell, Miss Lavish
	C3	Cecil, Mrs. Vyse, Sir Harry Otway, Mrs. Honeychurch
Howards End	C1	Henry, Mr. Wilcox, Charles Wilcox, His Father'S Chauffeur, Mrs. Wilcox, Dolly, Paul, Evie
	C2	Madge, Fraulein Mosebach, Tom, Margaret, A Young Man Low Down In The Education Office, Miss Avery
	C3	Tibby, Frieda, Helen, Aunt Juley
	C4	Leonard, The Parlourmaid, Mr. Cunningham, Mrs. Bast
Where Angels Fear to Tread	C1	Lilia, The Landlady, Mr. Herriton, Spiridione, Signor Carella
	C2	Irma, Caroline Abbott, Harriet, Mrs. Herriton

Table 5: Communities detected via modularity optimisation for each novel. All characters retained in the convergence networks are listed under their respective clusters.