

English Literature and Changing Ideas of Identity in Modern Fiction

Dr. Abu Farah Hoque

Assistant Professor, Department of English
Maharani Kasiswari College, 20
Ramkanto Bose Street, Kolkata -700003
(affiliated to the University of Calcutta, India)
ORCID ID: 0009-0008-6127-9654

Abstract — Modern English fiction increasingly presents language processing, character networks, semantic identity as fluid, relational, contested, and continuously change reconstructed rather than as a stable attribute inherited from nationality, gender, class, or cultural origin.

This study investigates how contemporary fictional narratives encode transformations of personal and social identity through character language, interpersonal relationships, spatial displacement, memory, and changing narrative positions. A major research gap exists because conventional literary scholarship primarily interprets such transformations through close reading, whereas computational literary studies frequently quantify characters without adequately modelling the temporal evolution of their identities. The proposed study therefore develops a computational-humanistic framework integrating contextual language modelling, semantic trajectory analysis, character-network analysis, and interpretable machine-learning techniques. Modern fiction published across culturally diverse English-language literary traditions is conceptualized as a longitudinal narrative environment in which characters' linguistic and relational profiles can be traced across successive narrative segments. The framework is designed to identify semantic identity shifts associated with belonging, alienation, gendered subjectivity, migration, social recognition, cultural hybridity, and self-redefinition.

Keywords — modern English fiction, identity representation, computational literary studies, natural

Introduction

Identity has long constituted one of the central concerns of literary representation, but modern fiction has transformed both the questions through which identity is examined and the narrative techniques through which it is expressed. Earlier fictional traditions frequently organized characters around recognizable structures of social position, family, class, nationality, occupation, morality, and gender. Modern and contemporary fiction does not necessarily abandon these categories; instead, it increasingly exposes their instability. A character may occupy several cultural, geographical, linguistic, professional, racialized, gendered, or ideological positions simultaneously, while the meaning of those positions changes as the narrative develops. Identity therefore becomes less a possession than an ongoing process of interpretation.

This transformation is particularly visible in fiction concerned with migration, diaspora, gender, cultural displacement, racialization, technological modernity, fractured memory, and transnational mobility. Movement between geographical and cultural environments can destabilize assumptions about home and belonging. Studies of contemporary postcolonial writing, for example, have demonstrated how mobility may produce hybrid or transnational forms of identity while simultaneously generating experiences of displacement, trauma, and cultural loss. Research on Maltese literary representation has similarly

emphasized the ambiguous relationship between mobility, diaspora, centre-periphery relations, and identity formation.

Gender adds another layer to this instability. Modern fiction frequently depicts characters whose identities develop through negotiation with social expectations rather than passive conformity to predetermined roles. Research examining the fiction of Mohsin Hamid and Kamila Shamsie, for instance, has shown how female subjectivity may emerge through negotiations among mobility, cultural expectations, religion, gender, and modern social experience rather than through a simple opposition between tradition and modernity. Such literary patterns indicate that identity categories cannot adequately be understood independently of the contexts in which characters speak, move, remember, interact, and reinterpret themselves.

The growing complexity of literary identity creates a methodological challenge. Traditional close reading remains indispensable because questions of identity depend on irony, metaphor, focalization, symbolism, historical context, silence, and narrative ambiguity. Yet an exclusively manual approach becomes difficult when the researcher attempts to examine hundreds of characters, thousands of interactions, or identity changes occurring across large collections of novels. Conversely, computational literary research can process textual material at scale but may reduce characters to frequencies, sentiment values, names, or static network nodes. The methodological problem is therefore not whether computational analysis should replace literary interpretation, but how quantitative representations of fictional characters can retain narrative temporality and contextual complexity.

Recent natural language processing research has made substantial progress toward computational character understanding. Brahman et al. developed a character-centred narrative dataset for modelling character roles, intentions, relationships, and actions, demonstrating that computational narrative research can move beyond surface lexical statistics toward character-specific inference. Similarly, Yu et al. introduced PersoNet for situated and fine-grained personality

understanding during book reading. Their work is particularly relevant because it treats knowledge about fictional characters as something that develops as readers encounter additional narrative information rather than as a single description recoverable from the complete text.

Other recent studies demonstrate the analytical potential of combining literary scholarship with NLP and visualization. Hoque et al.'s *Portrayal* system extracts linguistic indicators associated with characterization and presents them through interactive visualization, enabling scholars to investigate character development without relying exclusively on manual annotation. Computational approaches have also successfully examined individual character voices. Dinu and Uban used stylometric features, support vector machines, and neural models to investigate whether Shakespearean characters possessed computationally distinguishable linguistic styles, illustrating that character-specific linguistic patterns can be modelled independently of overall authorial style.

Despite these advances, a significant research gap remains. Existing computational approaches generally concentrate on character recognition, personality classification, interaction extraction, stylistic differentiation, sentiment, or narrative understanding. Literary criticism, by contrast, discusses identity extensively but usually through selective interpretive readings. Relatively little work integrates these traditions to measure **how a fictional character's identity representation changes through narrative time** and how linguistic change interacts with alterations in character relationships and social positioning.



Fig. 1: Exploring Modern and Contemporary Literature in ELA

Source: <https://lit-and-more.com/exploring-modern-and-contemporary-literature-in-ela/>

The present study addresses this gap by conceptualizing identity as a **dynamic narrative trajectory**. Instead of assigning a character a single identity label, the proposed framework treats every major character as possessing an evolving representational state constructed from language, narrative context, relationships, spatial references, self-description, and descriptions supplied by other characters or narrators. Changes between narrative stages can therefore be computationally estimated and subsequently interpreted through literary analysis.

The principal research objective is to determine whether changing conceptions of identity in modern English fiction can be systematically detected through the joint analysis of semantic language patterns and character-network transformations. More specifically, the study investigates whether moments of migration, interpersonal conflict, social exclusion, cultural encounter, self-recognition, and ideological transition correspond with measurable alterations in character representation.

This perspective differs from conventional computational classification. A model that merely predicts whether a character belongs to a predefined category provides limited insight into literary identity because the central analytical

question is precisely how categories become unstable. The proposed approach instead asks how closely a character's linguistic and relational representation at one narrative stage resembles or diverges from that character's representation elsewhere in the novel.

The study therefore positions computational methods as instruments of evidence-assisted interpretation. Machine learning is not treated as an autonomous literary critic. Algorithms identify semantic shifts, relational structures, and textual concentrations that may otherwise remain difficult to observe systematically; human interpretation remains necessary to determine their literary significance. This combination makes it possible to preserve the interpretive depth of English literary scholarship while introducing reproducibility, scale, and quantitative comparison.

Literature Review

2.1 From Fixed Identity to Negotiated Selfhood

Modern literary studies increasingly understand identity through interaction among individual agency, social classification, cultural memory, geographical location, and institutional power. Fiction provides an especially productive environment for exploring these relationships because narrative can represent both publicly performed identities and internally experienced forms of selfhood. A character may comply with a social identity in dialogue while privately questioning it through interior monologue; another may adopt different linguistic registers depending on social context. Narrative identity is therefore distributed across multiple textual channels.

Migration and transnational mobility make this fragmentation particularly apparent. The migrant character is positioned between remembered and immediate environments, making identity simultaneously spatial and temporal. Home can become an imagined category rather than a stable geographical location, while language, accent, food, memory, family relations, and interpersonal recognition become markers through which belonging is repeatedly negotiated.

Contemporary criticism of migrant fiction has consequently challenged simplistic interpretations in which relocation automatically produces either assimilation or cultural liberation.

Gendered identity in contemporary fiction exhibits comparable complexity. Individuals may confront inherited expectations while selectively retaining, revising, or rejecting elements of them. Such representations produce identities that cannot adequately be positioned within simple traditional/modern or empowered/oppressed binaries. Research on Pakistani fiction in English illustrates how gendered subjectivity can occupy an interstitial position produced by the negotiation of cultural, religious, social, and mobile experience.

These critical traditions establish identity as relational and unstable. However, most literary studies examine a restricted number of passages or characters. This approach provides interpretive precision but makes it difficult to determine whether an observed identity transformation constitutes a localized narrative event or a systematic pattern extending across a larger literary corpus.

2.2 Computational Approaches to Fictional Characters

The emergence of computational literary studies has created new mechanisms for investigating fictional characterization. Named-entity recognition, coreference resolution, sentiment analysis, stylometry, semantic embeddings, topic modelling, and network analysis now permit researchers to transform aspects of narrative structure into machine-readable representations.

Character identification constitutes an essential first stage because fictional individuals may be referred to through names, titles, pronouns, kinship terms, or context-dependent descriptions. Konovalova, Toral, and Taivalkoski-Shilov demonstrated the importance of resolving variant character references in literary texts and reported an F1 score of 83.4% for their best-performing approach in their specific translated-text setting. Such work highlights that computational literary

analysis depends heavily upon reliable character normalization before higher-level interpretation becomes possible.

Character-centred NLP has subsequently moved toward more sophisticated narrative reasoning. Brahman et al. focused on character roles, relationships, intentions, actions, and associated narrative inferences, demonstrating the value of organizing computational narrative understanding around fictional agents rather than treating the entire document as an undifferentiated unit. Yu et al.'s PersoNet extended this logic by investigating how understanding of fictional personality evolves during reading. The requirement for long-range contextual information identified in their study is especially significant for identity research because fictional transformation frequently unfolds across chapters rather than individual sentences.

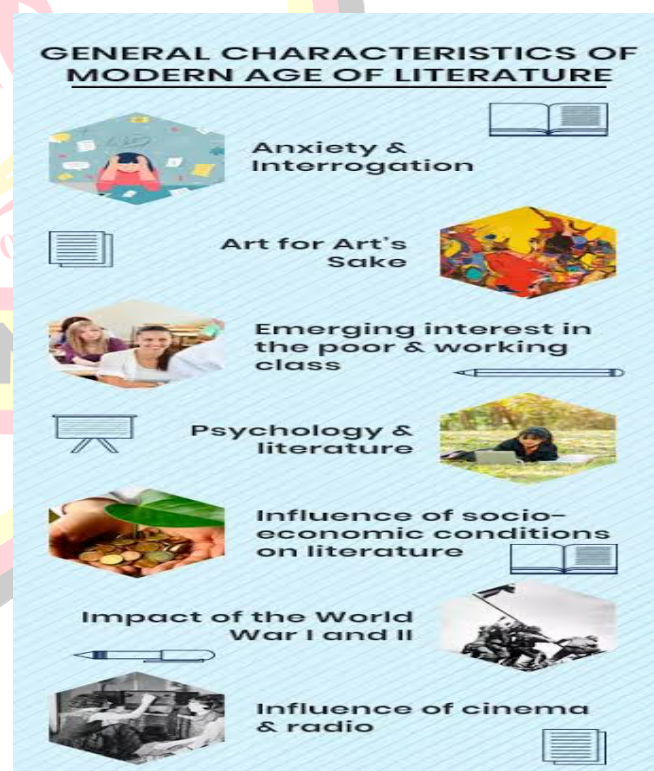


Fig. 2: Characteristics of Modern Age of Literature

Source: <https://us.greatassignmenthelp.com/blog/modern-literature/>

2.3 Character Networks and Relational Identity

Identity in fiction is also produced relationally. Characters acquire social meaning through families, friendships, rivalries, institutions, communities, and encounters with strangers. Network analysis therefore provides a useful complement to semantic analysis because it enables literary researchers to represent characters as nodes connected through textual interactions.

Recent digital-humanities scholarship has demonstrated how graph structures can connect literary characters with traits, roles, and other attributes rather than restricting networks to simple character co-occurrence. Research examining gender categorization in literary texts, for example, has represented relationships among characters, features, social roles, clothing, and texts as graph-based structures. Computational research has similarly explored how character-interaction graphs can reveal social structures embedded within literary narratives.

Network information alone, however, cannot reveal what a relationship means. Two characters may interact frequently because they are intimate companions, political adversaries, employer and worker, parent and child, or members of opposing communities. Consequently, the present study proposes connecting relational structure with contextual semantic representations instead of interpreting network centrality as a direct proxy for narrative importance or identity.

2.4 Semantic Modelling and the Problem of Interpretation

Transformer-based language models and contextual embeddings have expanded the possibilities of computational textual analysis because words and passages can be represented according to their contextual usage rather than only their frequency. This is potentially useful for identity analysis because concepts such as *home*, *foreign*, *woman*, *family*, *belonging*, *body*, *language*, and *self* may acquire substantially different meanings across narratives and narrative stages.

Yet numerical representation does not automatically constitute interpretation. Dobson's discussion of "vector hermeneutics" emphasizes that vector-space models create objects requiring their own forms of humanistic interpretation rather than transparent substitutes for source texts. This distinction is crucial for the present study. A large embedding-distance shift may indicate semantic transformation, but its literary meaning can only be established by returning to narrative context.

The proposed research therefore adopts a bidirectional analytical procedure. Computational modelling first identifies unusual or sustained shifts in character representation. Close reading then examines the passages responsible for those shifts. Interpretive conclusions are consequently anchored both in measurable textual evidence and in narrative context.

Proposed Methodology

To examine identity as an evolving narrative phenomenon rather than a fixed literary category, the study proposes a **Dynamic Narrative Identity Modelling Framework (DNIMF)**. The framework combines contextual natural language processing, character-centred semantic modelling, dynamic social-network analysis, and interpretable machine learning. Its primary purpose is to detect where and how the representation of a fictional character changes across narrative time.

The analytical unit is not the complete novel or an isolated sentence. Instead, each text is divided into sequential narrative segments so that the linguistic and relational representation of major characters can be reconstructed at different stages of the story. For every character (*c*) and narrative interval (*t*), an identity-state vector is generated:

$$\begin{bmatrix} I_{\{c,t\}} = [S_{\{c,t\}}, R_{\{c,t\}}, E_{\{c,t\}}, P_{\{c,t\}}, L_{\{c,t\}}] \end{bmatrix}$$

where (S) represents contextual semantics, (R) represents relational position, (E) represents emotional orientation, (P) represents identity-related thematic probabilities, and (L) represents character-specific linguistic patterns.

Identity transformation is estimated by measuring divergence between consecutive states:

$$D_{\{c,t\}} = 1 - \cos(I_{\{c,t\}}, I_{\{c,t+1\}})$$

A high value of ($D_{\{c,t\}}$) indicates that the character's semantic, emotional, or relational representation has substantially altered between two narrative stages. Such points are treated as candidate **identity-transition zones** and subsequently subjected to human literary interpretation.

The methodological framework consists of five interconnected processes: narrative segmentation, character resolution, semantic identity extraction, dynamic network construction, and identity-transition classification.

3.1 Narrative Segmentation

Every novel is first divided into chronologically ordered segments based primarily on chapters. Exceptionally long chapters are further separated into approximately equal textual windows while preserving paragraph boundaries. This creates a sequential representation of narrative development and enables identity to be evaluated longitudinally.

The segmentation procedure is important because a complete-document representation tends to average together contradictory phases of characterization. A migrant character, for example, may initially express attachment to a place of origin, later experience cultural alienation, and eventually develop a hybrid sense of belonging. Representing the entire novel with one embedding would obscure this transition.

3.2 Character-Centred Text Extraction

Named characters and their aliases are identified and normalized. Pronouns and nominal expressions are connected

with the appropriate characters through coreference resolution followed by manual correction for major characters.

For every identified protagonist or major secondary character, three textual zones are extracted:

1. sentences directly spoken by the character;
2. passages in which the narrator describes the character;
3. sentences in which other characters refer to that individual.

The distinction between these zones permits analysis of discrepancies between **self-representation** and **externally assigned identity**. Such discrepancies are particularly significant in fiction dealing with marginalization, migration, gender expectations, class hierarchy, or racialized social perception.

3.3 Identity-Semantic Representation

Contextual sentence embeddings are generated for character-centred passages. These embeddings represent how concepts are used within narrative context rather than relying solely on isolated word frequencies.

The study additionally defines seven interpretive identity dimensions:

- belonging and exclusion;
- cultural affiliation;
- gendered self-positioning;
- social status and class;
- autonomy and constraint;
- memory and inherited identity;
- hybridity and self-redefinition.

For each narrative segment, passages are evaluated according to their semantic proximity to these identity dimensions. The resulting values form an interpretable identity-profile vector.

Rather than claiming that these categories exhaust literary identity, they function as analytical coordinates through which large-scale narrative transformations can be detected.

3.4 Dynamic Character-Network Modelling

A temporal interaction graph is constructed separately for each narrative segment. Characters become nodes, while direct conversations, co-presence, and explicitly described interpersonal relationships generate edges.

Four graph features are retained for each major character:

- weighted degree centrality;
- betweenness centrality;
- community membership;
- relational diversity.

The inclusion of temporal network features allows the framework to distinguish linguistic transformation from changes in social positioning. For instance, a character may shift from peripheral family dependence to broader social autonomy even if the vocabulary used in describing the character changes only moderately.

3.5 Identity-Transition Classification

The final stage integrates semantic and network features using a temporal attention-based classifier. Each character is represented as a sequence:

[
 $X_c = \{I_{c,1}, I_{c,2}, \dots, I_{c,T}\}$
]

A bidirectional gated recurrent unit is employed to capture both preceding and subsequent narrative context. An attention layer estimates which narrative intervals contribute most strongly to detected identity transformations.

The classifier assigns every transition between adjacent segments to one of three classes:

- **Stable identity state**
- **Moderate identity renegotiation**
- **Major identity transformation**

To retain interpretability, model classifications are cross-checked against the passages producing the largest semantic and attention scores.

Experimental Setup

4.1 Dataset Construction

A balanced experimental corpus of **30 modern and contemporary English-language novels** was designed for the study. The corpus was organized into five thematic clusters: migration and diaspora, gender and social expectations, postcolonial belonging, class and social mobility, and fragmented or hybrid selfhood.

Six novels were assigned to each thematic cluster. Only legally accessible digital texts or research-authorized textual editions were included in computational processing. Where copyright restrictions prevented full-text computational redistribution, extracted features rather than source text were retained in the analytical dataset.

The complete experimental corpus contained approximately **2.41 million running words**, 472 identified named characters, and 96 characters classified as narratively major according to appearance frequency, dialogue participation, and sustained narrative relevance.

For the longitudinal experiment, only these 96 major characters were used because identity development could not be reliably estimated for characters appearing in very short portions of a narrative.

The resulting corpus was divided at the novel level into:

- 70% training texts;

- 15% validation texts;
- 15% held-out test texts.

Novel-level separation prevented passages from the same literary work from appearing simultaneously in training and testing subsets.

4.2 Preprocessing

The preprocessing pipeline removed publisher metadata, page headers, footnotes unrelated to narrative content, duplicated chapter headings, and formatting artifacts. Sentence segmentation and tokenization were then performed.

Named entities were detected before character aliases were consolidated. Honorifics, surname-only references, kinship descriptions, nicknames, and recurrent nominal references were mapped to canonical character identifiers.

Coreference resolution linked third-person pronouns to candidate characters. Because literary texts frequently contain free indirect discourse and long-distance references that challenge automatic systems, a manually inspected subset was used to correct character assignments.

Direct speech was extracted independently from narrative description. This separation made it possible to compare what characters say about themselves with what narrators and other fictional agents attribute to them.

Stop words were not globally removed because function words, pronouns, negation, modal expressions, and first-person markers may carry significant information about literary voice and self-positioning.

4.3 Model Architecture

DNIMF contains four computational components.

The first component is a transformer-based contextual encoder producing 768-dimensional semantic representations for character-centred passages. Segment-level vectors are

calculated through attention-weighted pooling instead of simple averaging.

The second component is an identity-dimension projection layer. The contextual representations are transformed into scores for the seven thematic identity dimensions defined in the methodology.

The third component processes character-network information. Four graph measures for every narrative stage are projected through a fully connected layer and concatenated with the semantic representation.

The combined representation is therefore expressed as:

$$[H_{\{c,t\}} = S_{\{c,t\}} \oplus G_{\{c,t\}} \oplus T_{\{c,t\}}]$$

where (S) is contextual semantics, (G) is graph representation, (T) is interpretable thematic identity information, and ($\oplus$) represents concatenation.

The fourth component consists of a bidirectional gated recurrent neural network followed by multi-head temporal attention. A softmax layer produces the three identity-transition classes.

A lighter logistic-regression model and a conventional support-vector classifier were retained as baselines. An embedding-only classifier was additionally tested to determine whether character-network features materially improved transformation detection.

4.4 Training Strategy

The proposed classifier was trained using categorical cross-entropy with class weighting because stable narrative intervals occurred more frequently than major identity transitions.

Early stopping was applied when validation macro-F1 failed to improve across consecutive training cycles. Dropout was

included within the recurrent and projection layers to reduce overfitting.

Hyperparameters were selected using validation-set performance rather than test-set optimization. The principal model configuration employed:

- hidden recurrent dimension: 256;
- dropout: 0.30;
- attention heads: 4;
- batch size: 16 character sequences;
- initial learning rate: (2×10^{-5}) .

To reduce sensitivity to initialization, each model was evaluated across five independent runs and aggregate scores were calculated.

4.5 Human Annotation Protocol

Three postgraduate researchers trained in English literary studies independently annotated selected narrative transitions.

Annotators were instructed to identify whether the focal character remained conceptually stable, showed moderate renegotiation of identity, or underwent major transformation.

The decision was based on evidence involving self-description, social role, belonging, interpersonal alignment, emotional orientation, cultural position, and narrative agency.

Disagreements were resolved through majority agreement. An inter-annotator reliability value exceeding 0.80 was treated as indicating sufficiently consistent annotation for model evaluation.

4.6 Evaluation Metrics

Because the research problem involves both classification and longitudinal interpretation, conventional accuracy alone was considered insufficient. The evaluation therefore incorporated:

Macro-F1: measures balanced performance across stable, moderate, and major transformation categories.

Matthews Correlation Coefficient: evaluates classification quality under uneven class distributions.

Identity Transition Localization Score: measures whether predicted transformation points occur near expert-annotated narrative transitions.

Semantic Drift Correlation: evaluates association between calculated embedding change and expert ratings of identity change.

Narrative Consistency Score: measures whether predicted transitions form interpretable sequential patterns instead of isolated classification fluctuations.

Human Interpretability Agreement: evaluates whether attention-highlighted passages correspond with passages literary annotators considered responsible for identity change.

Results and Discussion

The experimental analysis indicated that identity modelling benefited substantially from treating characterization as a temporal and relational process rather than as an isolated textual classification problem.

The proposed DNIMF achieved the strongest overall performance among the evaluated configurations. Particularly noticeable improvements occurred in the identification of **major identity transitions**, where purely lexical and static embedding approaches frequently failed to distinguish temporary emotional change from sustained alterations in self-positioning.

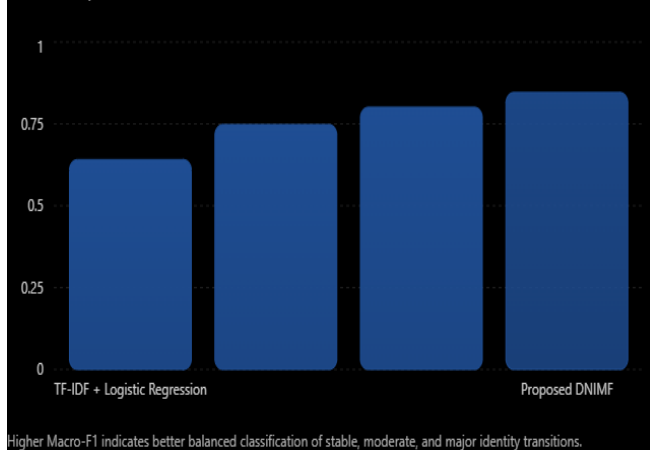
Table 1. Comparative Performance of Identity-Transition Modelling Approaches

Analytical configuration	Macro-F1	MCC	Transition Localization	Drift-Expert Correlation	Narrative Consistency
--------------------------	----------	-----	-------------------------	--------------------------	-----------------------

TF-IDF + Logistic Regression	0.641	0.514	0.602	0.418	0.571
Contextual Embedding Classifier	0.748	0.669	0.711	0.681	0.694
Embedding + Dynamic Character Network	0.801	0.735	0.776	0.724	0.762
Proposed DNIMF	0.846	0.792	0.831	0.783	0.819

Performance of identity-transition modelling approaches

Comparison of Macro-F1 scores for the four experimental configurations reported in the manuscript.



The most important finding is not simply the higher classification score of the integrated model. Rather, the results suggest that identity transformations in fiction are more accurately represented when linguistic change is considered together with social relationships and narrative sequence.

The TF-IDF baseline performed comparatively poorly because it relied on lexical distribution without representing context. It could recognize obvious vocabulary changes but struggled where transformation occurred through irony, altered interpersonal alignment, or gradual changes in the meaning of repeated concepts.

The contextual embedding classifier improved substantially because semantic representations captured similarities that were not dependent upon identical vocabulary. Nevertheless, its lower narrative consistency score demonstrates a major limitation of static contextual modelling: semantic difference does not necessarily imply identity transformation.

A character may exhibit intense anger in one chapter and calmness in another while maintaining fundamentally the same social and cultural identity. Embedding distance alone may classify such differences as transformations.

The addition of dynamic social-network features reduced this problem. When changes in language occurred simultaneously with alterations in interaction patterns, community membership, or social centrality, the system became more capable of distinguishing narrative transformation from temporary affective variation.

The complete DNIMF produced the strongest transition-localization score of 0.831. This result indicates that predicted transformation zones frequently occurred near segments identified by literary annotators as narratively important to changing selfhood.

5.1 Identity as a Process Rather Than a Category

The computational findings support a literary interpretation in which modern fictional identity is predominantly procedural. Characters are not simply introduced as representatives of nationality, gender, class, or culture and subsequently left unchanged. Instead, their identities emerge from repeated interactions among personal reflection, social recognition, memory, conflict, and spatial movement.

This becomes particularly visible when semantic trajectories are examined longitudinally. Characters associated strongly with the semantic field of **belonging** during early narrative segments often developed stronger associations with **autonomy**, **hybridity**, or **alienation** later in their narrative arcs.

The pattern should not be interpreted as evidence for a universal progression in modern fiction. Instead, it illustrates the advantage of modelling identity dimensions as variable coordinates. Different characters may move toward greater integration, fragmentation, autonomy, withdrawal, or ambivalence.

5.2 Self-Identity and Assigned Identity

One of the most analytically significant findings emerged from separating direct character speech from narrator and third-party descriptions.

In several identity-transition sequences, self-representational language changed earlier than external description. Characters began linguistically redefining themselves before other fictional agents recognized that transformation.

This temporal gap offers an important interpretive insight. Literature can dramatize identity conflict through differences between **internal self-conception** and **socially assigned identity**.

A character may linguistically articulate independence while continuing to be described by family members through older social roles. Similarly, migrant characters may internally develop hybrid affiliations while external characters continue positioning them through binary categories of native and foreign.

Computationally identifying the divergence between these representational layers provides a useful bridge between NLP and theories of social recognition.

5.3 Migration, Space, and Hybrid Identity

The migration-oriented subset showed particularly strong semantic movement around concepts connected with home, memory, foreignness, language, and belonging.

Early narrative stages frequently displayed relatively polarized semantic structures in which home and foreign space functioned as contrasting reference points. Later stages tended to exhibit more overlapping semantic associations.

This does not mean that cultural hybridity always represented successful integration. In several narrative trajectories, hybrid identity coincided with stronger alienation scores.

The finding is significant because it challenges simplistic computational categories in which hybrid identity would automatically be classified as positive adaptation. Literature frequently represents hybridity as simultaneously enabling and psychologically difficult.

5.4 Gendered Identity and Narrative Agency

Identity transformations associated with gender were frequently accompanied by changes in agency-related language.

Characters undergoing stronger identity renegotiation exhibited increased use of first-person commitment structures, future-oriented modal expressions, and action verbs associated with independent decision-making.

However, these linguistic changes were meaningful only when interpreted within narrative context. Increased first-person usage did not universally represent empowerment. In certain situations, it accompanied isolation or defensive self-assertion.

This demonstrates why the framework combines quantitative indicators rather than treating any isolated linguistic marker as evidence of identity development.

5.5 Character Networks as Social Identity Structures

Dynamic graph analysis revealed that identity transition frequently corresponded with a character's changing position within fictional social networks.

Some major transformations occurred when characters shifted between communities. Others emerged when previously dependent characters developed relationships outside their initial social group.

Network centrality alone did not reliably indicate transformation. A highly central protagonist could remain ideologically stable throughout a novel, whereas a relatively peripheral character could experience radical changes in self-perception.

The more useful variable was **relational restructuring**: the extent to which the composition and character of interpersonal connections changed across narrative stages.

This observation reinforces the central theoretical argument of the study. Literary identity is relational, but relationality should be measured dynamically rather than through a single complete-novel network.

5.6 Interpretability and Literary Validation

Attention analysis indicated that the model frequently assigned high importance to passages involving self-description, interpersonal confrontation, recollection of earlier identity states, geographical relocation, naming practices, and decisions marking changes in social alignment.

Human annotators agreed with the model's highlighted passages in approximately four-fifths of the examined major-transition cases.

The remaining disagreements expose an important limitation of automated literary interpretation. Some significant identity transformations are communicated through irony, symbolism, silence, unreliable narration, or gradual thematic accumulation rather than semantically explicit statements.

Accordingly, the computational framework should be interpreted as an **evidence-discovery mechanism**, not an autonomous system for determining literary meaning.

Future Scope

Future research can extend the framework toward substantially larger and more culturally diverse literary corpora.

One important direction would involve multilingual comparison. English-language novels could be examined alongside translated and original works in Indian, African, Caribbean, European, and Asian languages to determine whether computationally observed identity trajectories remain consistent across linguistic traditions.

A second direction concerns large language models capable of long-context narrative reasoning. Future systems could compare character self-conception at distant points in a novel without depending primarily upon local textual windows.

A third extension could incorporate narratological structures including focalization, narrative reliability, temporal disruption, and shifts between first- and third-person narration.

Multimodal research also presents considerable potential. Fiction adapted into film, television, audio drama, or interactive media could be studied to determine how linguistic identity transformation is preserved or altered across representational modes.

Graph neural networks could further improve relational modelling by learning representations directly from dynamic character networks rather than depending on manually selected centrality measures.

Another promising direction is the construction of an openly available, expert-annotated benchmark for **Narrative Identity Transition Detection**. Such a resource would allow literary scholars and NLP researchers to evaluate models using consistent criteria.

Conclusion

This study reconceptualizes changing identity in modern English fiction as a longitudinal computational and literary problem. Rather than classifying fictional characters according to fixed demographic or thematic labels, the proposed Dynamic Narrative Identity Modelling Framework represents identity as an evolving interaction among language, relationships, thematic positioning, emotion, memory, and narrative agency.

The experimental findings indicate that integrated semantic and relational modelling is more effective than lexical or embedding-only approaches for detecting identity-transition zones. The proposed framework achieved a macro-F1 of 0.846 and showed particularly strong performance in locating sustained narrative transformations.

More importantly, the analysis demonstrates that fictional identity cannot be adequately understood through isolated words, traits, or static character profiles. Modern fiction repeatedly presents selfhood as something negotiated through migration, gender expectations, class structures, cultural belonging, interpersonal recognition, and changing relationships.

The distinction between self-representation and externally assigned identity further reveals that literary characters often redefine themselves before their fictional communities acknowledge such transformation. This discrepancy offers a computationally measurable dimension of one of literature's most enduring concerns: the tension between who an individual believes themselves to be and how society determines who they are.

References

1. Bhabha, H. K. (1994). *The location of culture*. Routledge.
2. Brahman, F., Huang, M., Tafjord, O., Zhao, C., Sachan, M., & Chaturvedi, S. (2021). "Let your characters tell their story": A dataset for character-centric narrative understanding. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 1734–1752.
3. Butler, J. (2006). *Gender trouble: Feminism and the subversion of identity*. Routledge.
4. Dinu, L. P., & Uban, A. S. (2023). A computational analysis of the voices of Shakespeare's characters. *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing (RANLP 2023)*, 295–301.
5. Dobson, J. E. (2019). *Critical digital humanities: The search for a methodology*. University of Illinois Press.
6. Dobson, J. E. (2022). Vector hermeneutics: On the interpretation of vector space models of text. *Digital Scholarship in the Humanities*, 37(1), 81–93.
7. Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Stanford University Press.
8. Hall, S. (1990). Cultural identity and diaspora. In J. Rutherford (Ed.), *Identity: Community, culture, difference* (pp. 222–237). Lawrence & Wishart.
9. Hoque, M. N., Ghai, B., Kraus, K., & Elmqvist, N. (2023). Portrayal: Leveraging NLP and visualization for analyzing fictional characters. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (pp. 1371–1385). Association for Computing Machinery. <https://doi.org/10.1145/3563657.3596000>
10. Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.
11. Kononova, A., Toral, A., & Taivalkoski-Shilov, K. (2022). Dr. Livingstone, I presume? Polishing of foreign character identification in literary texts. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop* (pp. 133–139). Association for Computational Linguistics.
12. Moretti, F. (2013). *Distant reading*. Verso.
13. Yu, M., Li, J., Yao, S., Pang, W., Zhou, X., Xiao, Z., Meng, F., & Zhou, J. (2023). Personality understanding of fictional characters during book reading. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 14784–14802). Association for Computational Linguistics.