Computational Humanities Research

www.cambridge.org/chr

Toward an ontological representation of fictional characters

Antoine Bourgois, Jean Barré, Olga Seminck and Thierry Poibeau

Lattice (CNRS & ENS-PSL & Université Sorbonne Nouvelle), Montrouge, 92120, France

Research Article

Cite this article: Bourgois Antoine, Jean Barré, Olga Seminck and Thierry Poibeau. 2026. "Toward an ontological representation of fictional characters" *Computational Humanities Research*, **2**:e6, https://doi.org/10.1017/chr.2026.10025

Received: 30 September 2025
Revised: 4 January 2026
Accepted: 3 February 2026

Keywords:
character embeddings; characterization; computational literary studies; diachrony; French novel

Corresponding author:
Antoine Bourgois;
Email: antoine.bourgois@protonmail.com

Abstract

Characters are central to narrative theory but remain under-specified in computational work, where they are often reduced to clusters of words or vectors. We propose an operationalizable ontology of characterization that bridges narratological theory and NLP. From BERT-based clustering of character descriptions, we derive 17 classes of attributes (actions, emotions, traits, relations, possessions, etc.), validated through manual annotation (k = 0.77) and automatic classification (64% accuracy vs. 12% baseline). Applied to character similarity tasks for French fiction, our framework outperforms existing models. By aligning narratological insights with computational methods, we move toward a representation of fictional characters as structured, comparable entities for large-scale literary analysis.

Plain language summary

Existing computational approaches to literary characters typically aggregate all available textual evidence into a single representation, without distinguishing between what kind of information each piece conveys. This article argues that keeping these dimensions separate matters. A character's emotional profile, social position and physical behavior are not equivalent, and collapsing them obscures both the structure of individual characters and the nature of similarity between them. The proposed ontology addresses this by organizing character attributes into 17 interpretable categories grounded in narratological theory. The key finding is that character similarity is facet-dependent: small, well-chosen combinations of ontological dimensions outperform fully aggregated representations and better align with expert literary judgments. This has a theoretical implication beyond the computational results. Beyond an intrinsic property of the characters, resemblance between fictional figures varies with the aspects of characterization a reader chooses to foreground.

Introduction

Who does not still feel a pang when thinking of Emma Bovary, Anna Karenina or Elizabeth Bennet? Long after the details of the plot have faded, characters linger in memory, embodying desires, fears and ideals. Narratologists have long debated whether these figures are mere "beings of paper" (Barthes 1970) or the illusion of persons (Jouve 1992). What remains clear is their power: characters are the hinge between text and reader. But how can this power be captured in a form that computers, too, can process and compare?

Over the last decade, computational approaches have sought to represent literary characters via roles, personas, embeddings or social networks (Bamman, Underwood, and Smith 2014; Labatut and Bost 2019; Yang and Anderson 2024). Despite the interesting findings and visualizations that led to large-scale studies, these representations also have their shortcomings: noisy representation, shallow categories, low interpretability and limited dialogue with narratology. Meanwhile, within narrative theory, *characterization* is modeled as a complex process of attributing properties to agents in a storyworld (Jannidis 2013; Margolin 1983; Ryan 1992). The challenge, then, is clear: to design an interpretable representation of fictional characters that is both theoretically grounded and computationally operational.

We propose an ontological representation of fictional characters that organizes textual evidence into 17 semantically grounded classes of attributes (e.g., actions, emotions, personality traits, relations and possessions), yielding a computational definition of character that is no longer merely syntactic but explicitly semantic.

Our framework is designed to: (i) operationalize *character-space* (Woloch 2003) via coreference chains, supporting scalable extraction through aggregation of mentions and distributional profiling of modifiers, possessives and verbs, to yield per-character attention profiles at scale (Barré et al. 2023); (ii) formalize characterization by mapping heterogeneous textual cues to 17 narratology-grounded classes; and (iii) enable downstream evaluation in French fiction through *CharaSim-fr*, a character-similarity benchmark integrating Bamman, Underwood, and Smith

This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-ShareAlike licence (https://creativecommons.org/licenses/by-sa/4.0), which permits re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited.

(2014) protocols with new French items, showing how ontology-based representations align with human judgments.

The article is organized as follows: the "Related works" section reviews related approaches to character modeling in literary theory and NLP. The "An ontology for computational characterization" section presents our ontological framework, from its narratological foundations to its automatic classification pipeline. The "Character similarity task" section introduces the *CharaSim-fr* benchmark and evaluates our ontology-based representations on character similarity tasks. Finally, the "Discussion" section reflects on the implications for narratology and computational literary studies, outlining future directions for large-scale, theory-grounded character analysis.

Related works

Narratology

Debate over the centrality of character to narrative is as old as poetics. In *Poetics* I.VI, Aristotle subordinates character to plot: "The Plot, then, is the first principle [...] Character holds the second place" (Aristotle 1996, 335 BCE). Yet modern dramaturgy and screenwriting famously invert this hierarchy: "What the reader wants is fascinating, complex characters" (McKee 1997, 100). This tension has framed subsequent theorizing.

A durable heuristic is Forster's (1988) opposition between *flat* and *round* characters (constructed around a single idea vs. capable of convincing surprise). But across the 20th century, narratology diversified the very *definition* of character. Formalist and structuralist perspectives treat the figure as a textual/functional unit: a "being of paper" (Barthes 1966) or a role in an action grammar, from Propp's (1928) *dramatis personae* to Greimas's actantial schema (Greimas 1966). Barthes's *S/Z* radicalizes this by replacing the person with a web of semes tied to a proper name (Barthes 1970). French poetics of description (Hamon) makes this operational: descriptive cues (*prosopographie*, *éthopée*, *topographie* and *chronographie*) distribute traits across a novel's *personnel*, scripting recognizability and difference (Hamon 1998, 1993).

Reception-oriented and cognitive approaches shift emphasis from *what a character is* to *how readers construct one*. For Jouve (1992), character is an *effet-personnage*: an emergent construct at the intersection of textual cues and readerly activity, addressed to an *implicit reader* and invested by a *real reader*. Margolin (1983) offers a precise account of the inferential process: *characterization* (ascription of properties from actions, descriptions and setting) builds up to *character-building* (stabilizing a constellation of traits through classification, hierarchization and temporal scope).

Possible-worlds narratology further insists that characters are processed as entities in structured storyworlds rather than mere bundles of signifiers (Ryan 1992). Syntheses, such as Jannidis (2013) (and Jannidis 2008), align these strands by defining characterization as direct/indirect attribution of properties to referential anchors, articulating mimetic (human-likeness), thematic (symbolic roles) and synthetic (constructedness) dimensions. Three knowledge sources drive reader inferences: (a) a *basic type* for sentient beings (inside/outside and theory of mind), (b) *character models/types* (e.g., *femme fatale* and *hard-boiled detective*), and (c) *encyclopedic knowledge* (world/genre conventions) (Jannidis 2013). Texts ascribe not only psychological or social attributes but also physiological and spatiotemporal properties; characterization may be explicit or implicit, local or global, and is constrained by the level of narration and the cues available to the reader (Jannidis 2013; Margolin 1983).

Computational narratology

Over the past decade, a significant share of computational narratology research has focused on the question of characters. Most notably Bamman, Underwood, and Smith (2014), who propose a *Bayesian mixed-effects model* to infer latent character types from a large corpus of English novels. Their study demonstrates that computational methods can identify underlying character personas by analyzing linguistic patterns associated with character actions, descriptions or possessions. Following this seminal approach, the computational literary studies community has developed specialized tools to automate the extraction of character information. Notable examples include BookNLP for English (Bamman, Lewke, and Mansoor 2020; Bamman 2021), as well as its counterparts adapted to other languages, such as French (Mélanie-Becquet et al. 2024), German (Ehrmanntraut, Konle, and Jannidis 2023) and Dutch (van Zundert, van Cranenburgh, and Smeets 2023).

These pipelines have significantly advanced the automatic extraction of character-related information across multiple stages. They enable more accurate identification of character mentions, offer improved coreference resolution tailored to long literary texts (Bourgois and Poibeau 2025) and enable the extraction of various character-related linguistic features, such as speech acts, verbs and adjectives. In addition, they provide computational representations of characters that can be leveraged for a wide range of applications.

Various methods have been proposed for extracting information and constructing representations of characters. Attribute extraction methods are most commonly syntax-based, as implemented in pipelines, such as BookNLP or FanfictionNLP (Yoder et al. 2021). These approaches often rely on hand-crafted rules over part of speech tags and parsing trees to identify features, such as verbs for which a character is the subject/object, possessive dependents and predicatives.

Regarding the aggregation of extracted attributes into computational character representations, different strategies have been explored. A common approach is to merge all extracted features into a single representation for each character, either by constructing a frequency-based vector or by averaging the embeddings of individual attributes (e.g., bitstring (Bamman, Underwood, and Smith 2014), GloVe (Flekova and Gurevych 2015) or BERT-like models (Barré et al. 2025)).

While such aggregation provides a compact summary of character attributes, it also comes with important limitations. By collapsing heterogeneous information into a single unified representation, much of the nuance of individual attributes is lost, and the resulting vectors can prove difficult to interpret. Not all attribute types are equally informative, and certain syntactic classes may be more relevant depending on the downstream task.

To address this issue, Yang and Anderson (2024) propose maintaining separate representations for each class of attributes. By averaging attribute embeddings from a large encoder–decoder transformer Language Model (T5-11B), they represent each character as a set of five high-dimensional vectors. This design encompasses the intuition that some feature types may be more useful than others for specific applications, such as similarity assessment. In their work, they also suggest that concatenating these complementary vectors into a single representation could be a direction for future works. This approach represents a promising step forward in capturing richer character information. However, syntactic classes,

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

while convenient for attribute extraction, still bundle together very heterogeneous types of attributes. For instance, the possessive category may indiscriminately group clothing (hat and coat), vehicles (car and train), human relations (their children) or more abstract notions, such as dreams, power or homeland. Such diversity makes the resulting representations more difficult to interpret and limits their usefulness for fine-grained analyses. From the perspective of cultural studies or computational humanities, where research often targets specific kinds of attributes, this lack of granularity constitutes a significant limitation.

Along similar lines, Mian et al. (2026) argue that character importance itself should not be treated as a single scalar notion. Drawing on narratological theory, they model *character significance* as a composition of distinct narrative dimensions (e.g., action, speech, interiority, narration and being talked about by others), and show that different definitions of centrality yield markedly different character rankings. Their work highlights that collapsing heterogeneous narrative evidence into a single score obscures the internal structure of character prominence.

Another line of research leverages generative large language models to construct character representations, either in the form of paragraph-style summaries (Brahman et al. 2021) or as structured "character sheets." These methods typically proceed in one of two ways: either by directly prompting the LLM to produce the desired character representation, or, as proposed by Gurung and Lapata (2024), by adopting a question–answer framework that incrementally gathers information about a character across different dimensions, such as dialogue, personality and physical attributes, knowledge, plot role and motivations.

While these representations tend to be rich and readily interpretable by human readers and researchers, they are also computationally expensive to generate and do not easily scale to large-scale computational literary studies involving thousands of full-length books.

Evaluating character representations remains inherently complex. Unlike standard machine-learning evaluations that rely on held-out gold standards, defining what makes a "good" representation of a literary character is difficult because such representations are multidimensional and involve a certain degree of subjectivity. Consequently, evaluation is necessarily approximate, capturing only certain aspects of the representation. Some evaluation protocols are explicit about the dimension they target (e.g., narrative role, social rank and psychological profile), whereas others offer broader assessments against expert judgment without clearly specified criteria.

A first family of approaches relies on supervised evaluation, where representations are tested by training a downstream model to predict predefined classes. The nature of these classes varies across studies. For instance, Flekova and Gurevych (2015) frame the task as personality prediction, distinguishing extraverted from introverted characters. Srinivasan and Power (2022) propose a role-based ternary system, labeling characters as protagonist, antagonist or supporting figures. Likewise, Chen, Chen, and Chen (2019) classify characters into core, secondary and peripheral categories derived from character networks, while Jahan, Mittal, and Finlayson (2021) focus on identifying character roles from Vladimir Propp's *morphology of the folktale* (Propp 1928).

Another evaluation approach is masked-character identification, which tests whether a human reader (or a generative LLM) can recover the identity of a character from an anonymized description generated by the character representation pipeline (Brahman et al. 2021; Gurung and Lapata 2024).

Some studies adopt unsupervised similarity-based evaluation, which does not require training a classification model. One approach is triplet-based comparisons, in which literary experts annotate triples of characters indicating that character A is more similar to character B than either A or B is to C. Representations are then evaluated on how well they reflect these expert-defined similarity judgments (Bamman, Underwood, and Smith 2014).

Yang and Anderson (2024) proposed the *AustenAlike* benchmark, which evaluates character representations along three notions of similarity: structural, social and expert-defined. Characters are grouped by narrative role, social rank or literary expert judgment, and character representations are compared against these groupings. The conclusions from this work illustrate the current state of character-centered research in computational narratology. While feature-based representations have improved, they still do not fully capture the richness of literary characters: even recent models have difficulty matching expert judgments or representing the multidimensional nature of character similarity.

The gap is conceptual as much as technical. Most feature-based representations remain syntactic or latently lexical: they can say that two figures are close in a vector space, but not along which dimensions of characterization they are. Similarity between characters is not a single dimension; it is a bundle of relations that cut across actions, affect, cognition, social roles, physical description, possessions and more. Without a shared inventory of attribute types, these dimensions get muddled and similarity becomes opaque.

Our response is to move from syntactic proxies to semantic classes. We introduce an explicit ontology based on the work of Jannidis (2013) using the three theoretic axes that define characters: anthropological, biological and psychological. Each of these three hyper-class axes is further divided into fine-grained subclasses, resulting in a total of 18 interpretable categories for character description. The result is a structured, multi-granular profile for each character: comparable within and across classes, scalable over large corpora and meaningful from a narratological perspective.

An ontology for computational characterization

Ontology definition

Our ontology is informed by narratology and designed for computational literary studies. Following syntheses, such as Margolin (1983) and Jannidis (2013), we assume that characterization proceeds by attributing properties to referential anchors, and that these properties cluster along three macro-domains commonly mobilized in analysis: physical/bodily description and states (Hamon's prosopographie and descriptive scaffolding), psychological/mental life (affects, cognition and dispositions) and anthropological positioning (roles, relations and identity markers). To make these operational, we articulate this three macro-classes into 18 mutually exclusive and collectively exhaustive fine-grained categories. Two principles guide the design of our ontology: clarity and distinctness. This means that each class has a precise definition to minimize overlap and that adjacent classes are separated by concrete decision rules. For example, there is a difference between *physical objects* and *body parts*: the first is part of the character's body, whether the second is an object that can be carried by it.

Our ontology can capture a large spectrum of attribute types that covers the descriptive surface exploited by readers while remaining legible for large-scale computation.

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

Table 1. Ontology of character attributes grouped into four macro-classes

Hyper-class	Class	Definition
Anthropological	`Communication`	Actions intended to convey meaning between two or more characters (e.g., speaking, shouting, gesturing, writing and signaling).
	`Human relations`	Connections to other people across social spheres: family, work or hierarchical (e.g., brother, enemy, mentor, boss and subordinate).
	`Possession`	Non-physical ownership or abstract belongings (e.g., reputation, land rights, subscription, secret and business).
	`Sociological`	Social identity, roles or group membership (e.g., job, gender, age, social rank, marital status, profession and ethnicity).
Physical	`Action`	Any deliberate physical activity performed by the character, visible to an outside observer, excluding motion verbs (e.g., fighting, eating, reading and lifting).
	`Body parts`	Any part of the character's body, internal or external (e.g., hand, brain, leg and eye).
	`Environment`	Attributes describing the character's physical surroundings, context or immediate location (e.g., room, battlefield and forest).
	`Existence/state`	Temporal, physical or health conditions of the character, including life stage or intoxication (e.g., dead, drunk, ill and immobile).
	`Motion`	Movement of the character's body to change position or location (e.g., sit, stand, walk, run and jump).
	`Physical objects`	Physical objects associated with the character that can be carried, worn or used (e.g., sword, backpack, hat, horse, car and food).
	`Physical traits`	Observable physical characteristics (e.g., tall, blue-eyed and muscular).
Psychological	`Affect/emotion`	Feelings or emotional states experienced by the character (e.g., anger, joy and fear), visible through words, gestures or inner thought.
	`Capability/obligation`	Skills, abilities, duties or moral/legal expectations of the character (e.g., can swim, must obey, is required to fight).
	`Cognition`	Internal mental processes, thoughts or imaginative states, including dreaming, planning and remembering (e.g., thinking, dreaming and forgetting).
	`Perception`	Sensory awareness or perception involving at least one modality (e.g., sees, hears, smells and notices).
	`Personality traits`	Enduring behavioral tendencies or dispositions (e.g., brave, selfish and curious).
Other	`Idiomatic expression`	Figurative or set phrases expressing traits, states or behaviors of the character (e.g., "cool as a cucumber" and "hot-headed").
	`Not valid`	Items incorrectly extracted that do not describe a character or valid attribute (e.g., random nouns and misidentified entities).

Table 1 presents the 18 classes and their working definitions used for the manual annotation.

Operationalizing

Our aim is the automatic, ontology-aware classification of character attributes. The first step toward this goal is to identify characters and extract their attributes. As noted earlier, several pipelines already exist for this purpose. For this work, we build upon the off-the-shelf Propp-fr pipeline (https://github.com/lattice-8094/propp), designed to process full-length novels. It allows us to extract character mentions (pronouns, proper nouns and noun phrases), coreference chains and attributes derived from syntactic dependency rules.

To operationalize our ontology for large-scale computational studies, we must reclassify those syntactic-based attributes into our fine-grained ontology classes. Exhaustive manual annotation is unrealistic given the scale of the task involving thousands of books and millions of attributes.

One possible approach would be to construct an exhaustive dictionary of attributes for each class (e.g., all words referring to body parts). To accelerate this process, one could leverage existing semantic resources, such as WordNet or VerbNet, as done by Flekova and Gurevych (2015), who assign the most frequent sense to each word. This approach has the limitation of being unable to account for polysemy, as each word is assigned a fixed class, whereas in reality, words often have radically different senses depending on their context, especially in fictional writing.

For example, in French, the verb *filer* can have many meanings:

- Il **file** de la laine.
  (*He **spins** wool.*)
- En voyant la police, elle **a filé**.
  (*When she saw the police, she **ran away**.*)
- Elle m'**a filé** un coup de main pour tondre la pelouse.
  (*She **gave** me a hand mowing the lawn.*)
- Ce garçon **file un mauvais coton**.
  (*That boy **is headed for trouble**.*)

To overcome this issue, models that incorporate contextual information are required. Neural classifiers based on contextual embeddings provide a particularly suitable solution (Huang et al.

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

2019). These models generate dynamic representations of words conditioned on their surrounding context, thereby capturing polysemy and subtle usage patterns that static embeddings or hand-curated lexicons cannot.

We therefore adopt a contextual neural classification approach. To enable ontology-based labeling of extracted attributes, the classifier must first be trained. This necessitates a gold-standard dataset: a manually annotated collection of attribute instances, each assigned to the appropriate class within our ontology. Such a resource provides the supervision required for the model to learn robust, context-sensitive mappings between character attributes and fine-grained ontological categories.

Gold dataset

To ensure that the classifier will be able to generalise effectively, and be useful for downstream computational literary studies, the gold dataset must contain a representative pool of character attributes drawn from diverse texts.

For this purpose, we apply the `Propp-fr` pipeline to the full *Chapitres* corpus (Leblond 2022), which contains 2,960 full-length French novels published between 1811 and 2020. This corpus provides diversity across authors, periods, literary subgenres, narrative perspectives and representation of the literary canon.

The corpus contains 234M tokens, from which 25M attributes were extracted (10.83%). (This ratio of approximately 1:10 between character attributes and total tokens further underscores the character-centric nature of narratives, providing additional justification for the approach adopted in this work). The distribution of attribute types is as follows: agents 49.5 percent, patients 12.0 percent, possessives 27.3 percent and modifiers 11.2 percent.

Annotation pool construction

We adopt a strategy specifically designed to maximize the diversity and generalization potential of the annotated attributes.

We begin with the 25M attributes extracted from the *Chapitres* corpus. Each attribute is represented as a 1,024-dimensional $\text{CamemBERT}_{\text{LARGE}}$ vector (Martin et al. 2020), obtained by embedding the full novels using overlapping rolling windows. This ensures that the embedding for each attribute captures maximum context.

To construct the *annotation pool*, we employ a maximal diversity sampling strategy:

1. *Random seed selection:* We begin by randomly selecting the first attribute from the 25M-attribute pool. This serves as the initial seed for the annotation pool.
2. *Compute cosine distances:* For each remaining attribute in the 25M pool, we compute its cosine distance to every attribute already included in the annotation pool.
3. *Determine minimum distances:* For each candidate attribute, we retain the minimum cosine distance to any attribute in the current annotation pool. This ensures that each candidate's similarity to the existing pool is quantified by its closest neighbor.
4. *Select the next attribute:* We select the attribute with the maximum of these minimum distances, that is, the attribute that is farthest from all attributes in the annotation pool.
5. *Repeat steps 2–4:* We continue iteratively computing distances and selecting the attribute with maximal minimum distance until the pool reaches the desired size (1,000 attributes).

This procedure guarantees that the final annotation pool is highly diverse, avoiding redundancy and over-representation of common attributes (e.g., to say, to have, young and to watch), which the final model is likely to already capture.

The next step is to have human annotators assign ontology labels to each attribute in this pool, providing the gold standard for supervised training of our classifier.

Manual annotation procedure

Before beginning the task, annotators are asked to carefully read the ontology definitions in order to familiarize themselves with all the classes. During the process, the definitions remain accessible at all times for reference.

For each annotation, the annotator is presented with:

- the in-context attribute to be labeled;
- the character mention serving as the head of the attribute (i.e., the entity to which the attribute refers);
- the syntactic category assigned by the `Propp-fr` pipeline.

The annotator must then select the single ontology class to which the attribute belongs. Additionally, the annotation interface provides an open comment field, where annotators may:

- justify their decision;
- indicate alternative categories that were considered;
- flag cases they find difficult or ambiguous.

The annotation interface is shown in Appendix A.

Inter-annotator agreement and adjudication

To assess the reliability of the annotation scheme, we conducted a double annotation on a subset of 406 attributes, with each attribute independently annotated by two annotators (in total four annotators participated).

Inter-annotator agreement yielded an observed accuracy of 0.79 and Cohen's $\kappa = 0.77$, which falls within the "substantial agreement" range (0.61–0.80) defined by Landis and Koch (1977). This level of agreement is particularly encouraging given the difficulty of the task and the relatively large number of categories (18 classes in total). Agreement was particularly strong for frequent and concrete categories, such as *body parts*, *physical objects*, *communication*, *human relations* and *motion*. By contrast, it decreased for less frequent or more abstract categories, including *existence state*, *possession* and *capability/obligation*.

Following this phase, annotators participated in an adjudication round, reviewing all cases of disagreement. In this process, they agreed on the final label for each contested attribute, discussed edge cases and proposed refinements to the ontology definitions. As part of this review, the category *idiomatic expression* was merged into *not a valid character attribute*, reflecting a consensus that such cases fall outside the scope of our ontology.

At the end of this process, we obtained a consolidated set of 406 gold annotations. To increase the dataset size, a single annotator then labeled an additional 194 attributes from the annotation pool. The result is a gold dataset of 600 attributes, each labeled with one of the 17 ontology categories, which serves as the basis for following training and evaluating our classifier for automatic attribute classification.

Neural classifier training and evaluation

The following section presents the training setup and evaluation of a neural classifier for automatic attribute categorization.

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

Input representation

As described in the "Annotation pool construction" section, each attribute is represented by a 1,024-dimensional $\text{CamemBERT}_{\text{LARGE}}$ embedding. To this vector, we concatenate a four-dimensional binary feature encoding the syntactic role predicted by the Propp-fr pipeline (*agent*, *patient*, *modifier* and *possessive*). The resulting 1,028-dimensional vector serves as input to the model, while the output is the ontology label, defined either at the coarse-grained (4 classes) or fine-grained (17 classes) level.

Classifier architecture

To evaluate whether contextual embeddings can be automatically mapped to our ontological categories, we trained a feed-forward neural network on the manually annotated dataset. The architecture consisted of two hidden layers (512 and 64 units, ReLU activation and dropout 0.6), followed by a softmax classification head. The model was trained with the Adam optimizer (learning rate 10^{-3}, weight decay 10^{-4}), using cross-entropy loss and early stopping based on validation loss (patience = 5, maximum 100 epochs).

All experiments were conducted with stratified k-fold cross-validation ($k = 25$) to reduce variance due to dataset imbalance.

Classification results

Performance was evaluated both at the level of the full ontology (17 fine-grained classes) and at the higher level of abstraction (coarse-grained superclasses). *Fine-grained classification:* On the 600 annotated attributes, 25-fold cross-validationyielded an overall accuracy of 0.647. In the 17-class setting, random baseline accuracy is below 0.06, while the majority-class baseline (70/600 instances) is 0.116. The classifier therefore achieves a clear improvement over both baselines, indicating that contextual embeddings provide effective discriminative information.

Coarse-grained classification: Collapsing the ontology into four superclasses (*anthropological*, *physical*, *psychological* and *not-valid*) further improved performance: 20-fold cross-validation yielded an accuracy of 0.737. This gain is consistent with the reduced label space (4 vs. 17) and should primarily be attributed to task difficulty rather than stronger intrinsic separability of the categories.

Interpretation

These first results highlight both the feasibility and the challenges of automatically mapping character attributes to ontology classes. The classifier performs best for clear and concrete categories, while abstract or underrepresented classes remain more difficult to predict. These trends closely mirror inter-annotator agreement, suggesting that categories which are inherently harder for humans to classify also present greater challenges for the model.

Expanding the gold dataset with additional manually annotated examples could be explored in future work to potentially improve performance on the more difficult, low-frequency categories. However, this may not be critical for downstream applications, as the majority of classification errors occur in rare attributes. In the remainder of this study, we focus on evaluating the usefulness of our ontological classification for practical tasks in the field of computational literary studies.

Next, we leverage these annotated attributes to construct semantically informed character representations.

Character representation vectors

Following Margolin's (1983) approach to character conceptualization, we aggregate classified attributes into unified, vectorial representations of characters. For each ontological category (17 fine-grained and 4 coarse-grained), in-class attribute embeddings are aggregated and averaged to produce a single 1,024-dimensional vector per category.

The average vectors are then processed through the following steps:

1. **Normalization:** Each mean vector is L2-normalized to unit length. This ensures that subsequent operations, such as PCA, are insensitive to the overall magnitude of the embeddings, while preserving directional information on a unit hypersphere.
2. **Principal component removal:** To reduce dominant, non-informative directions shared across characters, the first principal component (PC) of all non-zero normalized vectors for a given attribute class is computed. Each vector is then projected orthogonally to this component, highlighting unique, discriminative characteristics for each character.
3. **PCA-based dimensionality reduction:** For each attribute class, PCA is applied to all non-zero vectors after PC removal. The top K PCs (e.g., $K = 15$) are retained, and each vector is projected onto this reduced subspace.

The resulting centered and reduced vectors provide a compact, discriminative representation of a character for each attribute class, suitable for downstream tasks, such as clustering, similarity analysis or classification.

To enable comparison, we also consider two simpler baselines. The first, following Yang and Anderson (2024), computes separate average embeddings for each syntactic category extracted by the `Propp-fr` pipeline (agent, patient, modifier and possessive), yielding four additional vectors per character. The second baseline collapses all attributes into a single undifferentiated vector by averaging embeddings across all attribute instances, regardless of category. For both baselines, the resulting vectors are centered and reduced following the same post-processing steps described above, ensuring comparability across methods.

Character similarity task

To evaluate the quality of our character representations, we rely on character similarity benchmarks inspired by prior work mentioned in the "Computational narratology" section. We use two different benchmarks:

1. Bamman, Underwood, and Smith (2014)-derived triplets: We used the French translations of the literary works present in this dataset and tested the same hypothesis.
2. `CharaSim-fr`: We propose a new benchmark, following the same principles as Bamman, Underwood, and Smith (2014) but this time built upon well-known characters from French literature.

Each triplet (A, B, C) expresses a similarity hypothesis: character A is more similar to B than either A or B is to C. For example:

Julien Sorel (A) from Stendhal's *Le Rouge et le Noir* resembles Lucien de Rubempré (B) from Honoré de Balzac's *Illusions perdues* more than either character resembles Gilliatt (C) From Victor Hugo's *Les Travailleurs de la mer*.

To evaluate a model, we compute the cosine similarity between each pair of characters (AB, BC, AC) and check whether

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

$\text{sim}(A,B) > \text{sim}(A,C)$ and $\text{sim}(A,B) > \text{sim}(B,C)$. A triplet is considered correctly predicted if both conditions hold, meaning the model aligns with the gold-standard similarity judgment.

Following Bamman, Underwood, and Smith (2014), triplets are organized by difficulty into four classes:

- *Class A:* Sanity checks. Characters A and B are more similar to each other in every respect than to character C.
- *Class B:* Intra-author similarity. Two characters from the same author are more similar to each other than to a closely related character from a different author.
- *Class C:* Cross-author similarity. Characters A and B from two different authors are more similar to each other than to a third character C from the same author as A.
- *Class D:* Exploratory/challenging hypotheses. These triplets test more subtle distinctions, such as point-of-view differences or temporal distance.

For both benchmarks, we perform the triplet similarity evaluation using each of the 26 character vectors (4 syntactic, 1 overall average, 4 coarse-grained ontology and 17 fine-grained ontology) to systematically compare the performance of the different representation types.

Bamman, Underwood, and Smith (2014) French benchmark

The Bamman, Underwood, and Smith (2014) baseline performance steadily decreases from Class A to Class D, reflecting the increasing difficulty of the triplets. Despite this decline, overall performance remains strong, with an overall accuracy of 64 percent (where random choice would yield ≈0.33). This provides a meaningful reference for evaluating our character representations (see Table 2).

Turning to the syntactic attribute classes, performance is mixed. The *modifiers* class performs particularly well (77% accuracy), while patient lags behind (42%). Interestingly, when all attributes from Propp-fr are aggregated into a single vector (all syntactic), performance rises to 65 percent, surpassing most individual syntactic classes and slightly above the baseline.

For the coarse-grained ontological classes, overall accuracy is relatively consistent (50%–58%) but remains lower than for some of the syntactic-based categories, suggesting that broad semantic categories do not capture character similarity as effectively. Notably, the non-informative class behaves as intended: aggregating attributes labeled as non-informative yields very low accuracy (27%), even below random chance (≈0.33). This confirms that these attributes contribute little to character discrimination, and removing them helps focus on more meaningful semantic signals.

In contrast, the fine-grained ontological classes exhibit a highly diverse pattern. Some classes, such as *possession* (12%) and *communication* (23%), perform poorly, while others, like *motion*, *personality traits* and *affect emotion*, achieve high overall accuracy (69%–73%), in some cases surpassing the baseline, but not quite reaching the best performing syntactic-based vector from modifiers attributes.

Combining fine-grained ontological vectors

Individual fine-grained ontological vectors capture specific dimensions of character similarity, but human judgments often integrate multiple, distinct aspects simultaneously. A human annotator might consider personality, actions, social relations and other traits together when assessing which characters are most alike.

To model this multi-dimensional perspective, we exhaustively explored all combinations of the 17 fine-grained ontological classes (from pairs to the full set). The total number of combinations is

$$\sum_{k=2}^{17} \binom{17}{k} = 131{,}054.$$

For each combination, we concatenate all vectors and similarity scores were recomputed on our benchmark triplets. We then grouped the results by the number of classes included and computed the distribution of accuracy scores, as well as average, minimum and maximum accuracy for each number of classes (1–17).

The full results of this experiment are displayed in Appendix B. Mean accuracy steadily increases as more classes are combined, rising from 0.49 for single-class vectors to a peak of approximately 0.646 around 9–10 classes. Beyond this point, adding additional classes yields diminishing returns and a slight decrease in mean accuracy, reflecting the inclusion of lower-performing or noisy attributes.

Maximum accuracy is achieved with carefully selected small combinations of 3–5 classes, reaching up to 0.96 on the French adaptation of the Bamman, Underwood, and Smith (2014) similarity benchmark.

As shown in Table 2, the optimal combinations of fine-grained ontological classes confirm the value of selective mixing. For example, pairing *affect emotion* + *personality traits* already yields 85 percent accuracy, surpassing any individual class. Adding a third class (*motion*) increases accuracy to 88 percent, while carefully chosen four- and five-class mixes (*affect emotion* + *personality traits* + *action* + *physical objects* and *affect emotion* + *personality traits* + *action* + *motion* + *physical objects*) reach 96 percent.

This result highlights that a judicious mixture of complementary ontological dimensions can strongly enhance character similarity predictions, approaching multi-dimensional human-like judgment. In contrast, including all possible classes does not further improve performance, as lower-quality attributes can dilute the signal.

Finally, the correlation between attribute usage ratio (frequency of appearance in character vectors) and accuracy is weak. Some infrequent attributes, such as *personality traits* (4.52%), are highly informative, achieving 73 percent accuracy, whereas more common attributes, like *agent* (54.9%), perform relatively poorly; the same pattern is observed for coarse-grained ontological classes. This indicates that informativeness is not proportional to frequency and underscores the importance of careful attribute selection when constructing effective character representations.

CharaSim-fr benchmark

Benchmark dataset construction

Our newly introduced `CharaSim-fr` benchmark comprises 31 gold-standard triplets created by a panel of French literature experts. Each triplet follows the same formulation used above (X is more similar to Y than X and Y are to Z) with examples sampled at varying degrees of difficulty across the four classes (10 Class A, 5 Class B, 5 Class C and 11 Class D). In total, the benchmark covers 63 distinct characters drawn from 165 books from the *Chapitres* corpus, published between 1830 and 1988.

A key design choice that distinguishes CharaSim-fr from the first benchmark is how repeated appearances of the same

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

Table 2. Accuracy results on the French adaptation of the Bamman, Underwood, and Smith (2014) similarity benchmark

Attribute category	Att. %	A	B	C	D	Acc.
Baseline						
Bamman, Underwood, and Smith (2014)	-	100%	75%	69%	50%	64%
Syntactic classes						
Patient	11.10%	5/5	3/4	0/6	3/11	42%
Possessive	27.81%	4/5	4/4	4/6	2/11	54%
Agent	54.90%	4/5	3/4	2/6	6/11	58%
Modifier	6.19%	5/5	3/4	3/6	9/11	77%
All syntactic	100.00%	4/5	4/4	3/6	6/11	65%
Coarse-grained ontological classes						
Non-informative	2.38%	4/5	2/4	1/6	0/11	27%
Anthropological	21.13%	4/5	2/4	4/6	3/11	50%
Physical	28.63%	4/5	4/4	2/6	5/11	58%
Psychological	47.86%	3/5	3/4	2/6	7/11	58%
Fine-grained ontological classes						
Possession	0.13%	1/5	0/4	0/6	2/11	12%
Communication	15.49%	3/5	1/4	0/6	2/11	23%
Environment	0.67%	2/5	2/4	2/6	4/11	38%
Sociological	1.79%	3/5	1/4	1/6	6/11	42%
Body parts	3.64%	3/5	3/4	2/6	3/11	42%
Non-informative	5.16%	4/5	2/4	1/6	4/11	42%
Existence state	2.24%	2/5	2/4	3/6	4/11	42%
Capability obligation	0.42%	4/5	2/4	2/6	4/11	46%
Perception	7.94%	4/5	3/4	3/6	3/11	50%
Action	4.52%	3/5	2/4	3/6	5/11	50%
Human relations	5.72%	4/5	1/4	5/6	3/11	50%
Physical traits	1.63%	4/5	3/4	2/6	6/11	58%
Cognition	17.29%	3/5	3/4	2/6	7/11	58%
Physical objects	2.83%	4/5	3/4	3/6	7/11	65%
Motion	12.59%	4/5	4/4	4/6	6/11	69%
Personality traits	4.52%	5/5	4/4	5/6	5/11	73%
Affect emotion	13.46%	4/5	3/4	3/6	9/11	73%
Optimal fine-grained ontological classes mix						
(2) Affect emotion + personality traits	17.98%	5/5	4/4	4/6	9/11	85%
(3) Affect emotion + personality traits + motion	30.56%	5/5	4/4	5/6	9/11	88%
(4) Affect emotion + personality traits + action + physical objects	25.33%	5/5	4/4	5/6	11/11	96%
(5) Affect emotion + personality traits + action + physical objects + motion	37.91%	5/5	4/4	5/6	11/11	96%

Note: We display the original baseline results alongside our evaluations. The reported baseline is computed as the weighted average of the best-performing model for each class, selected from the three models presented in Bamman, Underwood, and Smith (2014). For each attribute group, we also report the average ratio of all character attributes used in constructing the corresponding vector. Per-class performance (A–D) and overall accuracy are shown.

fictional persona are handled. Whereas they evaluate the similarity of the same character sampled from two different books (Hypothesis 5: "Sherlock Holmes, in *The Adventures of Sherlock Holmes*, resembles Sherlock Holmes in *The Sign of Four* (Conan Doyle) more than either version of Holmes resembles Watson in *The Adventures*"(Bamman, Underwood, and Smith 2013)), CharaSim-fr builds a single representation across multiple works: for any character that appears in several books, we aggregate all extracted attributes from every occurrence to produce that character's vector. For instance, Rocambole (Pierre Ponson du Terrail) and Arsène Lupin (Maurice Leblanc) each appear in eight different books in our dataset.

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

Table 3. CharaSim-fr benchmark results summary

Attribute category	Att. %	Acc.
All syntactic – baseline	100.00%	74%
Fine-grained ontological classes		
Possession	0.13%	29%
Capability obligation	0.28%	35%
Physical objects	4.27%	50%
Human relations	3.96%	50%
Perception	8.35%	50%
Sociological	1.90%	53%
Environment	0.63%	53%
Physical traits	1.63%	59%
Personality traits	3.62%	59%
Cognition	14.39%	62%
Motion	15.52%	62%
Existence state	2.27%	62%
Body parts	4.02%	62%
Communication	14.47%	62%
Affect emotion	12.08%	65%
Non-informative	5.50%	68%
Action	7.00%	68%
Ontological classes mix		
Action + personality traits	10.62%	79%
Action + personality traits + physical traits	12.24%	76%
Action + personality traits + physical traits + human relations	16.20%	82%

Note: Full results in Appendix C.

Evaluation and results

The evaluation on `CharaSim-fr` broadly confirms the tendencies observed in the first benchmark, while also revealing notable differences.

Syntactic-based vectors again provide a strong signal, with accuracies ranging from 53 percent to 71 percent. As shown in Table 3, the aggregated syntactic vector achieves 74 percent, outperforming all individual categories. Among the fine-grained ontological classes, results are more heterogeneous: *possession* (29%) and *capability/obligation* (35%) perform poorly, while *action*, *affect emotion*, *communication* and *body parts* reach 62 percent to 68 percent. Interestingly, the strongest ontological classes here are not the same as in the previous benchmark. Moreover, the vector built from attributes labeled as *non-informative* unexpectedly reaches 68 percent, suggesting that some of these features may still capture relevant signals in this context.

Combining ontological classes again improves performance, though the optimal mixes differ from those observed previously. In CharaSim-fr, the best combination (*action* + *personality traits* + *physical traits* + *human relations*) yields 82 percent accuracy, whereas in Bamman, the most effective mixes were centered on emotional attributes. This divergence may reflect cultural or corpus-specific differences, or differences in how literary experts conceive the notion of similarity between characters.

Table 4. Cosine similarities over fine-grained ontological classes

Attribute class	AB	AC	BC	Out.
Affect emotion	0.72	0.13	−0.39	C
Communication	0.92	0.14	−0.07	C
Cognition	0.91	0.04	−0.18	C
Sociological	0.82	−0.48	−0.39	C
Personality traits	0.66	0.49	0.51	C
Action	0.75	−0.55	−0.44	C
Human relations	0.62	−0.42	−0.29	C
Perception	0.91	−0.53	−0.61	C
Motion	0.82	−0.41	−0.33	C
Existence state	0.75	−0.04	−0.31	C
Body parts	0.48	−0.23	−0.28	C
Physical traits	0.86	−0.37	−0.52	C
Physical objects	0.51	−0.32	−0.47	C
Environment	0.60	0.72	0.34	B
Possession	0.39	−0.13	−0.12	C
Capability obligation	0.62	0.00	0.00	C

Note: A: Julien Sorel from Stendhal — B: Lucien Leuwen from Stendhal — C: Emma Bovary from Gustave Flaubert.

Qualitative analysis

Unpacking the model behavior, we look closely at per-dimension similarities for each triplet. For each ontological class k, we retrieve pairwise similarities $s^k_{AB}, s^k_{AC}, s^k_{BC}$ (cosine on k-restricted vectors). We call the *dimension-wise outsider* the character whose average similarity to the other two is minimal in class k.

Easy case

Triplet: A = Julien Sorel (Stendhal), B = Lucien Leuwen (Stendhal), C = Emma Bovary (Flaubert)

Predicted outsider: C = 16, B = 1

Table 4 shows that s^k_{AB} dominates across 16/17 classes, while both s^k_{AC} and s^k_{BC} are markedly lower or negative-leaning, yielding C as the outsider in almost every dimension. The strongest margins occur in *sociological*, *action*, *perception*, *motion* and *physical traits*, which coheres with narratological expectations: Julien and Lucien are Stendhalian *ambitieux* (male aspirants in Restoration/July Monarchy France) with comparable social roles and activity patterns, whereas Emma's profile diverges along gendered social constraints and affective-aesthetic desire.

The lone dimension favoring B (environment) is driven almost entirely by the feature *salon*: Lucien is the only one of the three who regularly attends "the salon," which situates him sociologically within the Parisian grand monde (upper bourgeoisie). By contrast, Emma Bovary is excluded from that world (and suffers for it), and Julien Sorel remains a provincial outsider.

Harder case

Triplet: A = Rosanette (Flaubert), B = Mme Arnoux (Flaubert), C = Mme Dambreuse (Flaubert)

Predicted outsider: A = 7, B = 5, C = 3

Table 5 shows that the harder triplet exhibits a fractured signal across classes, mirroring the interpretive subtlety of *L'Éducation*

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

Table 5. Cosine similarities over fine-grained ontological classes

Attribute class	AB	AC	BC	Out.
Affect emotion	0.38	0.62	0.27	B
Communication	0.82	0.90	0.82	B
Cognition	0.20	0.55	0.59	A
Sociological	0.24	0.33	0.40	A
Personality traits	0.01	0.38	0.04	B
Action	0.67	0.21	0.31	C
Human relations	0.26	0.15	0.54	A
Perception	0.23	0.74	0.40	B
Motion	0.75	0.41	0.66	C
Existence state	0.60	0.25	0.68	A
Body parts	0.42	0.66	0.66	A
Physical traits	0.48	0.70	0.56	B
Physical objects	0.75	0.32	0.65	C
Environment	−0.08	0.11	0.12	A
Possession	0.00	0.00	0.00	-
Capability obligation	0.00	0.00	0.00	-

Note: A: Rosanette from Gustave Flaubert | B: Mme Arnoux from Gustave Flaubert | C: Mme Dambreuse from Gustave Flaubert.

sentimentale. Dimension-wise outsiders alternate: some classes push A (Rosanette) out (*cognition*, *sociological*, *existence state* and *body parts*), others isolate B (Mme Arnoux) (*affect emotion*, *communication*, *perception* and *physical traits*) and others C (Mme Dambreuse) (*action*, *motion* and *physical objects*). This non-convergence is *informative*: Flaubert distributes Frédéric's field of desire across three poles, *sensuality* (Rosanette), *ideal* (Mme Arnoux) and *ambition/social power* (Dambreuse), so that similarity depends on which facet is foregrounded. Our ontology's facets capture these shifting alignments, producing a pattern of partial agreements rather than a single majority outsider.

The dispersion of votes reflects genuinely competing interpretive frames in *L'Éducation sentimentale*: each woman anchors a distinct pole of Frédéric's sentimental and social field, so the "outsider" depends on which facet of characterization the reader foregrounds. Expert feedback articulates three defensible choices for the outsider:

- **Rosanette** is less refined, more eroticized and far more present in dialogue. In *human relations/sociological*, her lexicon skews to demi-monde ties (*friend*, *lover* and *miss*), vs. Mme Arnoux's domestic roles and Mme Dambreuse's status/kinship network. Weighting *communication+affect/body* cleanly isolates A.
- On *affect emotion/perception/traits*, **Mme Arnoux** concentrates idealized feeling and inwardness (peaks on *to love*, *fear*, with evaluatives like *good*, *heart* and *poor*). Rosanette reads more performative/interactional; Mme Dambreuse cooler and milieu-coded. If one foregrounds *affect/ideality*, B emerges as the outsider.
- **Mme Dambreuse** is more effaced, enters late and functions mainly as access to a milieu. Empirically, she is less mobile (**0** occurrence on *to go*/*to come*/*to do*) and clusters on postural/possessional verbs and salon props, while the others show higher locomotion/transactions. Emphasizing *action+motion+physical objects* makes C the outsider.

Read this way, the dispersion of expert judgments is structured: distinct facet mixes, *sociological+body* (Rosanette), *affect+personality* (Mme Arnoux) and *action+motion+objects* (Mme Dambreuse), each license a defensible outsider. This also explains our results: selective bundles beat using *all* classes (weak facets dilute signal), while targeted mixes approximate scholars' multi-dimensional reasoning and argue for expert-aligned facet weighting.

Discussion

Our permutations procedure, steered by researchers' choices, foregrounds the salient ontological classes for a given similarity judgment and shows that a well-curated subset of attributes can outperform undifferentiated concatenation. In practice, selecting the "right" classes is task-dependent: hypotheses about what drives a comparison (e.g., affect vs. action vs. sociological milieu) should guide facet inclusion, which we then validate empirically. This facet selection is not merely a tuning trick; it renders explicit the interpretive lens through which similarity is constructed and explains why targeted bundles approach multi-dimensional, human-like judgments while lower-quality attributes dilute the signal.

Conceptually, the framework does not "solve" character similarity; rather, it operationalizes the idea that characters do not exist as fixed entities with stable, intrinsic properties, but as interpretive constructs whose salience depends on which dimensions are foregrounded. In this respect, the approach aligns with Margolin's view of characterization as property ascription and with Jannidis's emphasis on the interplay between textual cues and readerly inference: profiles correspond to distributions of cues, and facet weights to interpretive emphases. More broadly, the facet-dependent nature of similarity suggests that resemblance is not a stable relation, but varies with the aspects of characterization prioritized by a reader or critic. Our results may therefore tell us at least as much about readers' and scholars' representations as about the characters themselves. By making the dimensions of character explicit and manipulable, the ontology shifts the analytical focus from uncovering an essential "true" character profile to examining how different regimes of attention construct character resemblance. Beyond comparison, computational similarity becomes a tool for probing the cognitive and cultural lenses through which fictional figures are perceived, not merely for comparing them.

Limitations

While the ontology itself is grounded in broad narratological theory rather than corpus-specific regularities, its current operationalization and evaluation are limited to French-language fiction. The experiments reported here therefore do not assess how well the proposed categories transfer across languages, literary traditions or genres. Evaluating the robustness and portability of the ontology in such settings remains an important direction for future research.

A second important limitation concerns the static nature of the proposed character representations. In the current framework, attributes are aggregated over an entire text into a single profile per character and per ontological class. This abstraction is deliberately adopted as a first step beyond purely syntactic representations, but it necessarily overlooks the dynamic and sometimes contradictory process of characterization emphasized in narratology, whereby

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

readers continuously revise their understanding of a character as the narrative unfolds. The development of dynamic character representations and appropriate evaluation protocols for them is a promising and exciting avenue for future work, but is beyond the scope of this study.

We do not yet model negation, gradation or modality with sufficient precision, which can attenuate or invert traits and distort character profiles. Contextual embeddings should encode these shifts in polarity and degree, but in practice, it is hard to know how the model takes this into account.

Attributions about a character's properties may come from the narrator or from other characters (including antagonists, unreliable voices or situationally biased observers). In our current setup, we take such statements for granted, without systematically modeling who speaks and how credible they are, even though narratology warns that explicit ascriptions can be invalid or purely subjective when sourced to a fellow-character or unreliable narrator.

Perspectives

From a methodological perspective, an immediate extension of this work consists of expanding the manually annotated dataset and iteratively retraining the attribute classifier as new data become available. Increasing the size and diversity of the annotated pool would improve robustness, particularly for rare or abstract ontological classes. To facilitate reuse and replication, trained models will be released as part of a Python library for computational narrative analysis. Beyond French, we plan to investigate cross-linguistic extensions through annotation transfer and natively multilingual classification models, enabling tests of whether similar dimensions of characterization operate across literary traditions.

By aligning narratological insights with computational methods, we move toward representing fictional characters as structured, comparable entities that can be analyzed across large corpora. Building on this foundation, our next step is a diachronic exploration of characterization over two centuries of French fiction. We will trace the semantic trajectory of the primary figure in each novel, asking which pathways of our ontology are most activated at different historical moments and how these correlate with the rise and decline of subgenres.

This perspective opens broader possibilities. On the narratological side, it allows us to test theories of character-space and character-systems at scale, observing how the balance of roles and voices evolves over time. On the computational side, it provides a benchmark for refining character representations, moving from high-dimensional blur vectors toward semantically interpretable dimensions.

Acknowledgements. The authors would like to thank the anonymous reviewers who helped them improve this article.

Data availability statement. All code and data used in this study are publicly available at github.com/lattice-8094/fictional-character-ontology.

The repository includes the complete codebase required to reproduce the experiments reported in the article, as well as the gold-standard annotated dataset of character attributes and the two character similarity triplet datasets. The trained model for semantic classification of character attributes is readily usable in downstream applications via the open-source propp Python library.

Author contributions. Conceptualization: A.B. and J.B.; Data curation: A.B., J.B., O.S. and T.P.; Funding acquisition: T.P.; Software: A.B.; Supervision: T.P.; Writing – original draft: A.B. and J.B.; Writing – review and editing: O.S. Writing - Review & Editing,

Funding statement. This research was funded in part by PRAIRIE-PSAI (Paris Artificial Intelligence Research Institute–Paris School of Artificial Intelligence, reference ANR-22-CMAS-0008). This work has received support under the Major Research Program of PSL Research University "CultureLab" launched by PSL Research University and implemented by ANR with the reference ANR-10-IDEX-0001.

Competing interests. The authors declare no competing interests regarding this article.

References

Aristotle. 1996. *Poetics*, Vol. 335 BCE. London: Penguin Classics. Republished in 1996.

Bamman, David. 2021. "BookNLP." https://github.com/booknlp/booknlp.

Bamman, David, Olivia Lewke, and Anya Mansoor. 2020. "An Annotated Dataset of Coreference in English Literature." In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 44–54. Marseille: European Language Resources Association. https://aclanthology.org/2020.lrec-1.6.

Bamman, David, Ted Underwood, and Noah A. Smith. 2013. "Appendix to 'A Bayesian Mixed Effects Model of Literary Character'." Technical report, Carnegie Mellon University and University of Illinois Urbana-Champaign. https://www.ideals.illinois.edu/items/50007.

Bamman, David, Ted Underwood, and Noah A. Smith. 2014. "A Bayesian Mixed Effects Model of Literary Character." In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 370–9. Baltimore, MD: Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-1035.

Barré, Jean, Pedro Cabrera Ramírez, Frédérique Mélanie, and Ioanna Galleron. 2023. "Pour Une détection Automatique de L'espace Textuel Des Personnages Romanesques." In *Humanistica 2023*. Corpus: Association francophone des humanités numériques. https://hal.science/hal-04105537.

Barré, Jean, Olga Seminck, Antoine Bourgois, and Thierry Poibeau. 2025. "Modeling the Construction of a Literary Archetype: The Case of the Detective Figure in French Literature." *Anthology of Computers and the Humanities* 3: 983–99. https://doi.org/10.63744/SMbYIWcHZj87.

Barthes, Roland. 1966. "Introduction à L'analyse Structurale Des récits." *Communications* 8, no. 1: 1–27. https://doi.org/10.3406/comm.1966.1113.

Barthes, Roland. 1970. *S/Z. Number 70 in Points Essais*. Paris: Éditions du Seuil.

Bourgois, Antoine, and Thierry Poibeau. 2025. "The Elephant in the Coreference Room: Resolving Coreference in Full-Length French Fiction Works." In *Proceedings of the Eighth Workshop on Computational Models of Reference, Anaphora and Coreference*, 55–69. Suzhou: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.crac-1.5.

Brahman, Faeze, Meng Huang, Oyvind Tafjord, Chao Zhao, Mrinmaya Sachan, and Snigdha Chaturvedi. 2021. ""Let your Characters Tell their Story": A Dataset for Character-Centric Narrative Understanding." In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 1734–52. Punta Cana: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-emnlp.150.

Chen, R. H.-G., C.-C. Chen, and C.-M. Chen. 2019. "Unsupervised Cluster Analyses of Character Networks in Fiction: Community Structure and Centrality." *Knowledge-Based Systems* 163: 800–10. https://doi.org/10.1016/j.knosys.2018.10.005.

Ehrmanntraut, Anton, Leonard Konle, and Fotis Jannidis. 2023. "LLpro: A Literary Language Processing Pipeline for German Narrative Texts." In *Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023)*, 28–39, Ingolstadt: Association for Computational Linguistics. https://aclanthology.org/2023.konvens-main.3/.

Flekova, Lucie, and Iryna Gurevych. 2015. "Personality Profiling of Fictional Characters Using Sense-Level Links between Lexical Resources." In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1805–16. Lisbon: Association for Computational Linguistics. https://doi.org/10.18653/v1/D15-1208.

Forster, E. M. 1988. *Aspects of the Novel*. London: Pelican Books, Penguin Books. Reprint edition.

Greimas, Algirdas Julien. 1966. *Sémantique Structurale: Recherche de méthode. Formes sémiotiques*, 3e ed. Paris: Presses Universitaires de France.

Gurung, Alexander, and Mirella Lapata. 2024. "CHIRON: Rich Character Representations in Long-Form Narratives." In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 8523–47. Miami, FL: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.499.

Hamon, Philippe. 1998. *Le Personnel du Roman: le système Des Personnages Dans les Rougon-Macquart D'emile Zola, Number 12 in Titre courant*, 2nd éd. corr edition. Genève: Droz.

Hamon, Philippe. 1993. *Du Descriptif*. Hachette Éducation, Paris.

Huang, Luyao, Chi Sun, Xipeng Qiu, and Xuanjing Huang. 2019. "GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3509–14. Hong Kong: Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1355.

Jahan, Labiba, Rahul Mittal, and Mark Finlayson. 2021. "Inducing Stereotypical Character Roles from Plot Structure." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 492–7. Online and Punta Cana: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.39.

Jannidis, Fotis. 2008. *Figur Und Person: Beitrag Zu Einer Historischen Narratologie*. De Gruyter. https://doi.org/10.1515/9783110201697.

Jannidis, Fotis. 2013. "Character." In *The Living Handbook of Narratology*, edited by **Peter Hühn** et al. Hamburg: Hamburg University. Edition. https://www-archiv.fdm.uni-hamburg.de/lhn/node/41.html.

Jouve, Vincent. 1992. "Pour Une Analyse De L'effet-Personnage." *Littérature*, no. 85. 103–111. https://www.jstor.org/stable/41713196. Publisher: Armand Colin.

Labatut, Vincent, and Xavier Bost. 2019. "Extraction and Analysis of Fictional Character Networks: A Survey." *ACM Computing Surveys* 52, no. 5. https://doi.org/10.1145/3344548.

Landis J., Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33, no. 1: 159–74. http://www.jstor.org/stable/2529310.

Leblond, Aude. 2022. *Corpus Chapitres*. v1.0.0, Zenodo. France. https://doi.org/10.5281/zenodo.7446728.

Margolin, Uri. 1983. "Characterization in Narrative: Some Theoretical Prolegomena." *Neophilologus* 67, no. 1: 1–14. https://doi.org/10.1007/BF01956983.

Martin, Louis, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. "CamemBERT: A Tasty French Language Model." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7203–19. Association for Computational Linguistics, Online. https://doi.org/10.18653/v1/2020.acl-main.645.

McKee, Robert. 1997. *Story: Substance, Structure, Style, and the Principles of Screenwriting*. New York: ReganBooks.

Mélanie-Becquet, Frédérique, Jean Barré, Olga Seminck, Clément Plancq, Marco Naguib, Martial Pastor, and Thierry Poibeau. 2024. "BookNLP-fr, the French Versant of BookNLP. A Tailored Pipeline for 19th and 20th Century French Literature." *Journal of Computational Literary Studies* 3: 1–34. https://doi.org/10.48694/jcls.3924.

Mian, Haaris, Melanie Subbiah, Sharon Marcus, Nora Shaalan, and Kathleen McKeown. 2026. "Computational Representations of Character Significance in Novels." Preprint. https://arxiv.org/abs/2601.15508.

Propp, Vladimir Iakovlevitch. 1928. *Morphologie du Conte*. Paris: Éditions Points.

Ryan, Marie-Laure. 1992. *Possible Worlds, Artificial Intelligence and Narrative Theory*. Bloomington, IN: Indiana University Press.

Srinivasan, Vardhini, and Aurelia Power. 2022. "Character Extraction and Character Type Identification from Summarised Story Plots." *Journal of Computer-Assisted Linguistic Research* 6: 19–41. https://doi.org/10.4995/jclr.2022.17835.

van Zundert, Joris, Andreas van Cranenburgh, and Roel Smeets. 2023. "Putting Dutchcoref to the Test: Character Detection and Gender Dynamics in Contemporary Dutch Novels." In *Computational Humanities Research Conference*, 757–71. CEUR Workshop Proceedings (CEUR-WS. org), France, Paris.

Woloch, Alex. 2003. *The One vs. the Many: Minor Characters and the Space of the Protagonist in the Novel*. Princeton: Princeton University Press.

Yang, Funing, and Carolyn Jane Anderson. 2024. "Evaluating Computational Representations of Character: An Austen Character Similarity Benchmark." In *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, 17–30. Miami: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.nlp4dh-1.3.

Yoder, Michael, Sopan Khosla, Qinlan Shen, Aakanksha Naik, Huiming Jin, Hariharan Muralidharan, and Carolyn Rosé. June 2021. "FanfictionNLP: A Text Processing Pipeline for Fanfiction." In *Proceedings of the Third Workshop on Narrative Understanding*, 13–23. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.nuse-1.2.

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

A. Annotation Interface

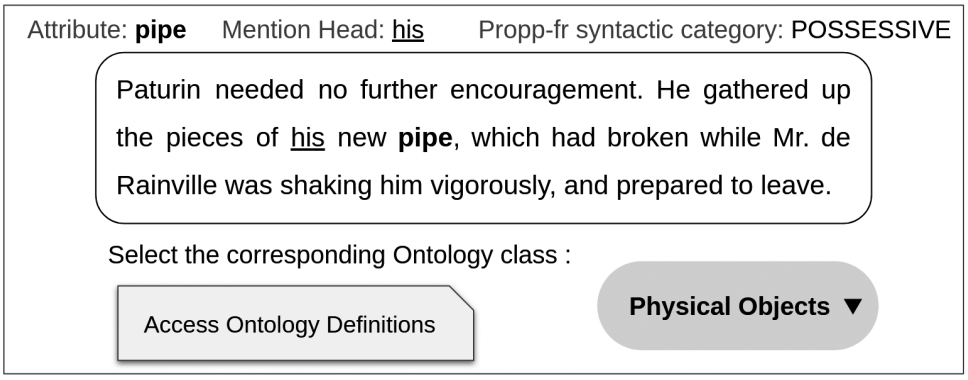


Figure A.1 Character attributes annotation interface.

B. Optimization algorithm

In order to assess the impact of the number of attributes classes considered on the overall similarity prediction task, we exhaustively explored all combinations of the 17 fine-grained ontological classes (from pairs to the full set). Maximum accuracy is achieved with carefully selected small combinations of 4–5 classes, reaching up to 0.96.

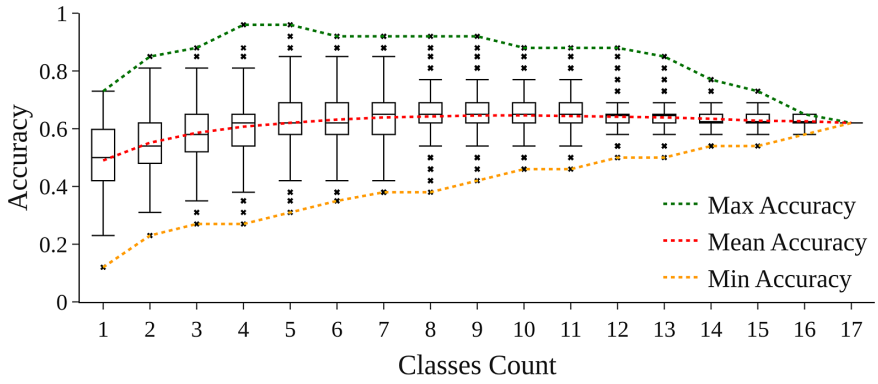


Figure A.2 Accuracy by classes count for our exhaustive combination search. Results on the French adaptation of the Bamman et al. 2014; similarity benchmark.

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press

C. CharaSim complete results

Table C.1. CharaSim-fr benchmark detailed results

Attribute category	Att. %	A	B	C	D	Acc.
Syntactic classes						
Modifier	5.85%	7/10	5/5	1/5	4/11	53%
Patient	10.76%	9/10	5/5	1/5	8/11	68%
Possessive	25.40%	8/10	5/5	3/5	6/11	68%
Agent	57.99%	10/10	5/5	3/5	6/11	71%
All syntactic	100.00%	10/10	5/5	4/5	5/11	74%
Coarse-grained ontological classes						
Non-informative	2.48%	9/10	5/5	2/5	1/11	56%
Anthropological	19.15%	7/10	5/5	3/5	5/11	59%
Psychological	42.65%	9/10	4/5	3/5	4/11	59%
Physical	35.72%	10/10	5/5	2/5	4/11	65%
Fine-grained ontological classes						
Possession	0.13%	4/10	1/5	1/5	4/11	29%
Capability obligation	0.28%	4/10	3/5	2/5	2/11	35%
Physical objects	4.27%	7/10	5/5	2/5	2/11	50%
Human relations	3.96%	6/10	3/5	3/5	5/11	50%
Perception	8.35%	7/10	4/5	2/5	4/11	50%
Sociological	1.90%	9/10	3/5	2/5	4/11	53%
Environment	0.63%	4/10	5/5	2/5	7/11	53%
Physical traits	1.63%	7/10	4/5	2/5	5/11	59%
Personality traits	3.62%	6/10	5/5	2/5	5/11	59%
Cognition	14.39%	9/10	4/5	2/5	6/11	62%
Motion	15.52%	10/10	5/5	1/5	5/11	62%
Existence state	2.27%	8/10	5/5	1/5	6/11	62%
Body parts	4.02%	7/10	4/5	3/5	6/11	62%
Communication	14.47%	9/10	5/5	3/5	4/11	62%
Affect emotion	12.08%	9/10	5/5	2/5	4/11	65%
Non-informative	5.50%	10/10	5/5	2/5	4/11	68%
Action	7.00%	9/10	5/5	3/5	6/11	68%
Optimal fine-grained ontological classes mix						
2) Action + personality traits	10.62%	10/10	5/5	3/5	7/11	79%
3) Action + personality traits + physical traits	12.24%	10/10	5/5	2/5	7/11	76%
4) Action + personality traits + physical traits + human relations	16.20%	10/10	5/5	3/5	8/11	82%

https://doi.org/10.1017/chr.2026.10025 Published online by Cambridge University Press