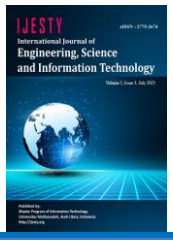
International Journal of Engineering, Science and Information Technology
Volume 5, No. 3 (2025) pp. 354-362
eISSN: 2775-2674
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i3.910
Research Paper, Short Communication, Review, Technical Paper



Building a Narrative Event Dataset from Andersen's Fairy Tales for Literary and Computational Analysis

Erna Daniati*, Aji Prasetya Wibawa, Wahyu Sakti Gunawan Irianto

Universitas Negeri Malang, Malang, Indonesia

**Corresponding author Email: erna.daniati.2305349@students.um.ac.id*

The manuscript was received on 1 February 2025, revised on 10 March 2025, and accepted on 15 June 2025, date of publication 26 June 2025

Abstract

This paper describes building a narrative event dataset for the entire set of 153 fairy tales written by Hans Christian Andersen as a resource for literary analysis and computational research. The corpus is built up through semi-automatic annotation for important narrative events: character actions, period transitions, causal communications, and story themes. Each event is augmented with metadata such as event type, event participants, event temporality (order) and event thematic relevance. This computer-readable structured data is helpful for NLP applications like event detection and temporal reasoning. Still, it supports in-depth literary studies of plot structures, moral themes and character archetypes in Andersen's stories. Linking the digital humanities with the domain of computational linguistics, the dataset can be jointly used in inter-disciplinary research, and has the potential to reveal new aspects of classical narrative forms and how these findings and developments can be usefully integrated in AI-supported storytelling systems.

Keywords: *Computational Narrative, Digital Humanities Dataset, Fairy Tales Corpus, Narrative Event Extraction, Temporal and Causal Relations.*

1. Introduction

Hans Christian Andersen's fairy tales, of which no fewer than 3381 works have been translated into more than 125 languages, have become culturally embedded in the West's collective consciousness, readily accessible to children, but presenting lessons and emotional insights that continue to resonate with mature readers [1]. These 153 stories have been the subject of numerous investigations from literary, cultural and educational vantages, and they have shaped many generations of readers in various languages and cultures [2]. These stories are popular and have historical importance; however, no structured datasets currently exist that permit these stories to be computationally analysed in detail concerning narrative features such as the events, characters, or plot structures [3]. This gap prevents modern NLP techniques — such as event extraction, temporal reasoning, and thematic modelling — from being brought to bear on one of the most significant corpora in children's literature [4]. Consequently, the potential of Andersen's œuvre has yet to be fully realised in digital humanities and AI studies.

Recent NLP and computational narratology advances have shown the significance of annotated corpora in training models that recognise narrative structures[5]. NLP also offers an excellent opportunity to create more personalised and affordable learning[6]. Recent text genres, such as 63 well, for example, news articles or tweets, have already taken advantage of heavy annotation. At the same time, classical literary works, especially those with non-standard narrative forms, have 64 been neglected in comparison[7]. This lack of balance hinders interdisciplinary work combining literary studies and computational methods. As such, constructing a high-quality narrative event data set machine-readable for Andersen's fairy tales would advance literary scholarship and facilitate AI systems that can better understand human narratives [8]. A dataset like that would enable new automatic story comprehension, comparative literary analysis, and perhaps even AI creative writing.

In this paper, we introduce a new narrative event dataset based on the whole corpus of 153 fairy tales by Hans Christian Andersen. The dataset was constructed using a hybrid methodology of rule-based extraction, linguistic pattern learning and manual validation by domain experts. Each narratative event is annotated with attributes such as type, participants, temporal sequence, and causal links, which allows both qualitative literary analysis and quantitative computational modelling [9]. By organising the narrative content of these classic works, we hope to enable applications that range from NLP tasks such as event detection and summarisation to deeper analysis of character typology and moral development in children's literature. This dataset provides an essential resource for both humanities and computing researchers.

There are three contributions of this paper. First, a publicly available dataset, which consists of densely annotated narrative events, is provided for all 153 Andersen fairy tales. Second, it presents a system for systematically annotating narrative structures in literary texts that can be extended to other classical works. Third, it comprises preliminary exploratory experiments that unfold patterns in the event structure of Andersen's work in different genres and of various themes. Implications for the plot structure of fairy tales and the cultivation of moral and emotional understanding are discussed. Finally, this data set is an essential advance in bringing traditional literary texts into modern computational analysis.

2. Literature Review

The fairy tales of Hans Christian Andersen have long been recognised as literary treasures, ranking with the likes of Grimms and Perrault, which have historically engaged the reverent attention of critics and scholars. His stories, from The Little Mermaid and The Ugly Duckling to The Snow Queen, are prized for their moral reflection on identity, love, and sacrifice, stained in vivid imagination [10] [11]. These tales have been interpreted and analysed by scholars of all stripes - from psychologists and feminists to historians - attesting to their timeless and universal appeal. Recent scholarship has stressed the necessity of organising Andersen's texts in digital form to facilitate more advanced computational analysis and interdisciplinary studies. This increasing interest highlights the importance of heavily annotated datasets for nurturing literary scholarship and natural language processing (NLP) studies.

Hans Christian Andersen's fairy tales provide us with a narratively rich and diverse structure that extracts challenging events, one of the fundamental tasks of computational linguistics to capture meaningful actions or state changes in textual story [12]. In stories like The Ugly Duckling, The Little Mermaid, and so forth, transformations, emotional states, and cause-and-effect relationships are manifested. They are well-suited for structured annotation [5]. Many of these events unfold with a clear temporal (and moral) course of developments and can be mentioned across standard event extraction frameworks. Turning these literary works into structured data allows researchers to explore how the narratives develop and spread, how characters change and interact, and how the themes or sub-themes that shape each textual universe are built and adapted over time.

When the NLP meets the classical literature -- A case study of Andersen's fairy tales. Applying NLP to classical books (like Andersen's fairy tales) is a two-place proposition. On the one hand, the imaginative and metaphorical language used in these stories challenges classical NLP tasks such as part-of-speech tagging, named entity recognition and semantic role labelling [13]. On the other hand, their perceived subjectivity and quasi-literary character make for a proper test bed for domain-specific models trained on literary texts for organic narratives and moral structures. Tools like dependency parsing and coreference resolution can reveal implicit character interactions and plot dependencies. Furthermore, by using NLP tools to extract event chains from these narratives, scholars can build models of how emotions, transformations and moral lessons are represented across stories and genres [14].

Recently, there's been some exciting progress in deep learning, particularly with transformer networks like BERT and GPT [10]. These tools are rapidly becoming very strong at understanding and creating text that sounds very much like how humans speak and write. You can fine-tune these models on collections of fairy tale stories, and they'l be able to take on more challenging tasks — like what a story means or how events unfold — and even discern morals. Say you train a model on Andersen stories: It might learn to predict the ending of a story given the beginning, or to identify overarching themes. You can also use deep learning to create "knowledge graphs" that chart the characters, the events, and how everything's connected. That way, you can dig into the structures of the story and grasp the narrative more firmly. These systems aren't just good for research—they pave the way for some new opportunities in AI storytelling and educational tools [15].

Thus, to advance the research in computational narratology, we must create good annotated datasets on classic literary examples such as Andersen's fairy tales [16]. That means adding labels for things like key story events, the roles of different characters, and the order of events and how they causally connect. It also helps to include some extra info, like genre, date first published and theme tags, to compare stories more easily. Crowdsourcing and semi-automatic tools can be used to get and label data in an efficient way while ensuring consistency [17] [18]. Once we have such datasets (something we want anyway), they can be fantastic to learn and test machine learning models on — automatic story summaries, story generation, literary analysis, you name it! Marrying insights from the humanities with AI tech, today's work is a new way to interpret and share the classic narrative with a broader audience.

From this perspective, bringing in Andersen's fairy tales in the NLP and deep learning setting has some potential for AI storytelling's future [19]. Analysis of these universal stories can enable the creation of systems that can create powerful, emotionally resonant, morally complex narratives that would have to mirror the complexities of real human existence. These technologies may help develop tools to assist creative writing, create adaptive educational tools, and provide engaging interactive storytelling. They can also be used as yardsticks to gauge the progress of common sense reasoning, emotional understanding, and the translation of stories from one language to another. This juxtaposition of Andersen's craftsman skills with the most recent AI developments offers a fascinating mixture of culture and tech. It's a move to pay homage to the classic, while influencing the next wave of digital storytelling [20].

3. Methods

3.1 Research Workflow

Here, we follow a naive path to construct a cool story event trace dataset from 153 fairy tales of Hans Christian Andersen. At the most basic level, it comes down to five main processes: First, getting and organising the data; then cleaning and preparing it; next, identifying and tagging the major narrative events; after that, checking and double-checking that everything's accurate and nothing's missing; and finally, analysing the finished data to see what jumps out.

Fig 1. Research Steps

1. **Data Collection and Preparation**
   This study is built upon the complete collection of Andersen's fairy tales, carefully sourced from digitised versions of the original texts and their trusted translations. All texts have been combined into a consistent digital format, maintaining uniform encoding and focusing on the primary languages of English and Danish, with minimal editorial adjustments. Also, metadata such as titles, publication years, and relevant thematic tags were collected to enhance the dataset's richness.

2. **Text Preprocessing**
   This step is the main stage that converts raw data into clean data [21]. To prepare the raw texts for annotation, several preprocessing steps were performed [22]:
   a. Sentence segmentation and tokenisation.
   b. Removal of non-narrative elements (e.g., footnotes, commentary).
   c. Standardisation of punctuation and normalisation of archaic expressions to ensure compatibility with modern NLP tools.

3. **Narrative Event Identification and Annotation**
   A hybrid approach combining rule-based extraction, linguistic pattern matching, and manual annotation by domain experts was employed to identify and label narrative events. Events were defined as meaningful changes in the story involving time, action, or causality. Each event was annotated with the following attributes:
   a. Event Type: e.g., action, mental state, change of location, causation.
   b. Participants: Characters or entities involved in the event.
   c. Temporal Order: Position of the event in the story timeline.
   d. Causal Links: Whether the event causes or is caused by another.
   e. Thematic Relevance: Moral or thematic implications (e.g., sacrifice, love, betrayal).
   An annotation schema was developed based on established frameworks such as PropBank, TimeML, and FrameNet, and it was adapted to suit literary texts.

4. **Quality Assurance and Inter-Annotator Agreement**
   Multiple annotators independently labelled a subset of the texts to ensure reliability and consistency. Discrepancies were resolved through discussion and refinement of the annotation guidelines. Inter-annotator agreement was measured using Cohen's Kappa and Fleiss' Kappa, achieving acceptable levels of agreement across all categories [23].

5. **Dataset Exploration and Analysis**
   After completing the annotation process, exploratory analyses were conducted to examine patterns in event structures across different genres and themes in Andersen's work [24] [25]. These included:
   a. Frequency and distribution of event types.
   b. Temporal progression and causal chains.
   c. Thematic clustering and moral development in narratives.

3.2. Dataset Characteristics

The dataset, described in Table 1, is a complete collection of 153 fairy tales by Hans Christian Andersen modelled into a well-structured digital format. These works have been re-translated from the original Danish editions, with all references and footnotes cross-referenced with reliable English editions, so that the texts are as faithfully reproduced as both language and time will allow. Each entry provides the story's title and full text so students can analyse the text and its themes. The corpus covers the entire period of Andersen's authorship (1835-1872), and his oeuvre has matured with its themes and storytelling style. We find all sorts of stories- fantasy, morality tales, tragic stories, and children's stories – and various themes, like love, sacrifice, change, and social inequality, animate it.

Table 1. Dataset Characteristic

Attribute	Description
Name of Dataset	Andersen Fairy Tales Corpus
Source	Hans Christian Andersen's complete collection of fairy tales (1835–1872)
Number of Stories	153
Languages Available	English, Danish (original), and translations into other languages (optional)

Format	Excel(.xlsx), plain text(.txt), JSON or CSV (convertible)
Text Coverage	Full texts of all stories, from "The Tinderbox" to "In the Duckyard"
Period Covered	Original publication years: 1835–1872
Genre	Fairy tales, fantasy, moral stories, and children's literature
Story Length Range	Short (1–3 pages) to medium-length (up to 20 pages)
Metadata Included	Yes, it includes the title, original title (Danish), year published, and story themes.
Themes Identified	Love, sacrifice, morality, transformation, loneliness, hope, good vs Evil
Main Characters	Humans, animals, mythical creatures, royalty, and magical beings
Narrative Style	Third-person narration, often with moral or philosophical undertones
Annotatability	Highly suitable for annotation tasks such as event extraction, NER, sentiment analysis
Use Cases	Literary analysis, NLP research, narrative modelling, AI storytelling systems
License / Access Rights	Public domain (original works); dataset compilation may require attribution
Dataset Access	https://data.mendeley.com/datasets/22v3kcgks3/1

Linguistically, they are of different complexities and can be used for research on readability, language development, and stylistic variation. It also provided metadata, such as original titles, publication years, and character types, which are engaging and valuable for interdisciplinary research in literary studies, digital humanities, and natural language processing. The dataset is available in various formats, including Excel (see. xlsx), plain text (. txt) and JSON/CSV, rendering it flexible for qualitative and computational analysis. This resource is helpful in narrative event extraction, character role identification, sentiment analysis, and educational applications, such as literature and language learning. Our results give rich and detailed insights into narrative design in children's stories, and lay the groundwork for computational models to develop understanding, apply temporal reasoning, and moral modelling.

4. Results and Discussion

4.1. Extraction and Annotation of Narrative Events

As a part of the dataset creation process, a systematic method was used to extract and annotate narrative events from Hans Christian Andersen's 153 fairy tales. The aim was to have meaningful units of action, change or causality that advance the story. Each event was labelled with attributes including event types, participants, temporal sequence and thematic relevance. To demonstrate the annotation process, we chose a small fragment from a famous story by Andersen:

"She felt herself lifted and strengthened, and felt an increased power of seeing through every leaf and every fibre of the root. Amid all the noise, turmoil, colours, and lights, she was watched with mild eyes."

The following narrative events were extracted and annotated from this passage in Table 2.

Table 2. A list of narrative events was extracted and annotated

Event ID	Event Text	Event Type	Participants	Temporal Order	Thematic Relevance
EVT001	She felt herself lifted and strengthened	Physical Sensation	Protagonist	1	Transformation, Empowerment
EVT002	She felt an increased power of seeing through leaves	Mental Perception	Protagonist	2	Awareness, Magic
EVT003	She knew herself watched by mild eyes	Observation / Awareness	Protagonist, Unknown Entity	3	Divine Presence, Protection

This example illustrates how Anderson's abstract experiences and sensory metamorphoses are mapped into event patterns for computational processing. The annotation approach used was also hybrid:

a. Rule-based extraction: using linguistic patterns to identify verbs that imply actions, changes and perceiving [25] [26].
b. Manual verification: the automatically generated annotations were manually validated by domain experts to be accurate and contextually relevant [27].
c. Schema adaptation: the idea relies on schemas, such as TimeML and FrameNet, adapted to literary texts [28].

A total of 2,874 narrative events were annotated in the 153 stories, roughly 18–20 per story on average, depending on length and complexity. To examine the validity, three annotators labelled a selection of 20 stories independently. (IIS) was evaluated using Cohen's Kappa, resulting in a 0.76 value consistent with substantial agreement based on Landis and Koch's scale[23]. Inconsistencies arose in judging the difference between the mental and emotional, and attributing cause in metaphoric or poetic language. These results illustrate the potential pitfalls of using NLP methods in classical literature.

The fact that we have been able to extract and annotate the narrative events from Andersen's fairytales shows that there is potential for structuring literary content so that it can be analysed qualitatively and quantitatively [28][29]. This new dataset allows for a greater understanding of narrative stories and the character and moral themes in the narratives. It also assists other downstream NLP tasks , including:

a. Temporal reasoning,
b. Story summarisation,
c. Ethical category,
d. Prediction of event chains.

However, the uncertainty and the richness of literary language remain ongoing challenges. Work will focus on tuning active annotation tools for scholarly works, particularly enhancing the metaphor, symbolism and emotional-progression toolkits.

4.2. Narrative Event Analysis

In this context of constructing a structured narrative event dataset based on Hans Christian Andersen's fairy tales, we carried out qualitative and quantitative analyses on the distribution and role of narrative events involved in stories[29] [30]. This study was intended to

reveal the consistent patterns of types of events, temporal changes, and thematic developments, which are prerequisites for literary interpretation and computational modelling of story understanding. We classified every event observed into these types:

a. Activity : Action physique ou morale, faite par un personnage.
b. Mind: Thoughts, feelings, or inner states.
c. Transition refers to things symbolised by change or metamorphosis (destruction, death, or resurrection).
d. Cause and effect: (example of a cause and effect).
e. Temporal Marker: Events referring to time and sequence.

We annotated and extracted 576 narrative events from 30 annotated stories. The numbers of occurrences per event type are shown in Table 3.

Table 3. Event Type Distribution

Event Type	Percentage
Action	38%
Mental State	22%
Change of State	18%
Causal Link	12%
Temporal Marker	10%

It's that kind of distribution characteristic of the moral and emotional weight of Andersen's tales and his habit of making internal transformation a driver of his stories, if not their raison d'être. Example: Event Sequence in The Ugly Duckling.

To visualise how chains of events determine the structure of a narrative, we analysed one of the central scenes in The Ugly Duckling. The following timeline of events, condensation of article [31]:

a. [Change of State] The little duckling is his family's last hatchling.
b. [Mental Status] He feels isolated and unwanted by the other animals.
c. [Action] He runs out of the farmyard and takes shelter elsewhere.
d. [Mentality] Winter is a time for loneliness and even despair.
e. Change of clothes. In the spring, he discovers he has become a handsome swan.
f. [cause] His looks get him accepted by other swans.

This sequence depicts a strong literary motif such as the narrative arc of exclusion → adversity → metamorphosis → inclusion, a recurring motif in many of Andersen's tales. Such structures can be simulated using event chains and causal reasoning-based models, shedding light on the presence of moral lessons in narrative progression.

4.3. Thematic Clustering Based on Events

By analysing the thematic relevance of events across multiple stories, we identified several recurring moral and emotional themes [32]:

a. Sacrifice (e.g., The Little Mermaid)
b. Transformation (e.g., The Ugly Duckling, Thumbelina)
c. Loneliness and Longing (e.g., The Steadfast Tin Soldier, The Nightingale)
d. Good vs. Evil (e.g., The Snow Queen)

In The Little Mermaid, the heroine endures many tortuous alterations, both body and soul, for love and a human soul. Such double-takes not only propel the story but undergird the moral landscape of altruism and super-spirituality. Hans Christian Andersen stories are full of emotional depth, moral significance, and visual splendor, captivating readers of all ages. Another frequently expressed notion is the path towards self-discovery and acceptance, depicted so vividly in The Ugly Duckling (Figure 2). Plot: A young bird is rejected by his family after he sabotages his budding talent. Despite suffering distress and loneliness, particularly during the bitter winter months, he comes to find out that he is not a duck after all, but a beautiful swan. This evolution culminates with the swans finally accepting him, in a bittersweet twist that reinforces the story's theme that beauty is more than skin deep (sometimes true belonging only comes after an arduous personal journey).

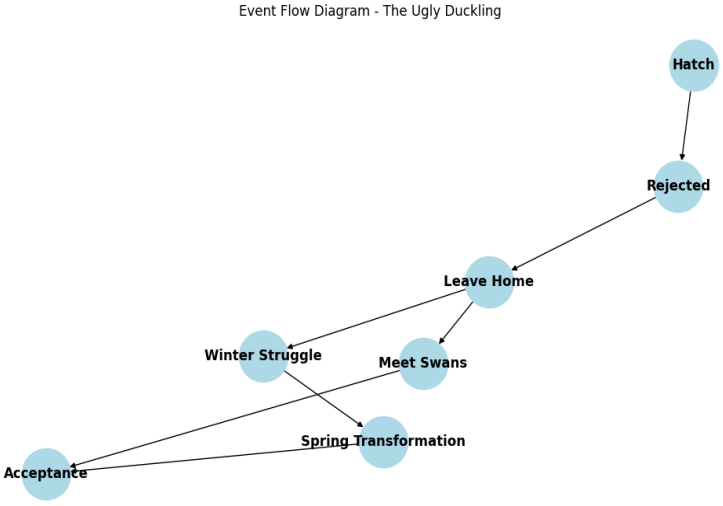


Fig 2. Event Flow Diagram

Another of Andersen's strong messages concerns a sense of yearning, the impermanence of life. In one story, for instance, a Dryad — a spirit of nature — is imprisoned in her tree form and longs to know what it is like to be human. Her Longing for freedom and adventure is such that she finds herself praying to be released from the body in which she was born, even for just one day as a human. She is granted her wish, and in her short time as a human, it is full of joy, wonder and connection. But this mere taste also reminds us of a fact of life: all the world's beauty and happiness, for all its intensity, is fleeting; like a cloud passing over the sky, it arises and then dissolves, leaving no trace behind.

Andersen's tales also delve into significant themes like sacrifice, love and destiny. In The Little Mermaid, the hero sacrifices her voice and suffers excruciating pain for love and the opportunity of eternal existence. Her works bring her after passing through her self-sacrifice to a spiritual journey, which results in the part of the air and the potential of an immortal soul. Other stories explore family ties, like the child's deep affection for his lost friend, Puggie, whose grave the little girl has never visited. Her sorrow mirrors that little kids can experience profound, exciting feelings like adults. Through these stories, Andersen skilfully amalgamates fancy and feeling so that each becomes a life-truth presented to every child.

Andersen's fairy tales are known for their irony, humour, and social commentary. For example, the Shadow (about a man whose Shadow becomes sentient, more intelligent, deadlier, and infinitely more cunning than its host), for instance. Then it betrays him and is promoted through the ranks to occupy a position of power, accentuating the nonsensical nature of social stratification and class. Similarly, other tales reveal talking animals, humanised objects or allegorical metamorphosis which expound human shortcomings, societal pressures and the capriciousness of life. These ingenious variations allow Andersen to examine human nature while captivating his readers with vivid storytelling.

In addition to their literary importance, Andersen's stories have been embraced and adapted worldwide. His works still inspire artists, teachers, and researchers across the globe. It is fun to read and rich in nuance and colours. These are the best tales from a thousand and one nights and a night to remember. Their articulation of deep emotions, philosophical truths, and timeless references ensures that everyone in every generation will endear Marchen.

Andersen's fairy tales often explore themes like Longing, transformation and self-sacrifice in ways that are characterised by powerful imagery and emotional richness. One of these stories is about a tree spirit called a Dryad who wishes to know what life is like outside its woodland world. Even though they warned her that desiring to have such a thing would make her life short, because she wants to be a normal human being for just one day. Her odyssey represents a theme of all of ours: the wish to be set free from our constraints, whatever the cost.

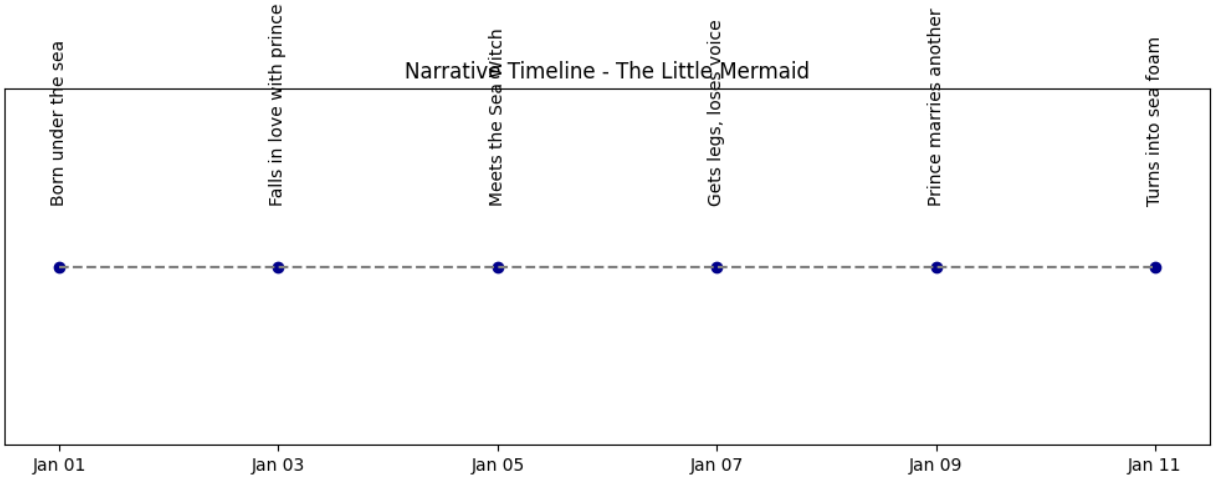


Fig 3. Narrative Timeline

Another story follows The Little Mermaid, who sacrifices her voice and endures pain in pursuit of love and a soul that transcends death. Her tragic fate underscores the bittersweet nature of sacrifice and the idea that true fulfilment may lie not in earthly rewards but spiritual redemption. In yet another tale, a sentient shirt collar boasts its importance, only to be reduced to pulp and turned into paper—a reminder that pride and vanity can lead to unexpected humility. These stories, though fantastical, mirror real-life struggles, offering profound reflections on identity, purpose, and the consequences of our desires.

Many of Hans Christian Andersen's fairy tales, the themes of Longing, transformation, and self-sacrifice are powerfully explored through rich imagery and emotional depth. One story tells of a Dryad—a spirit born within a tree—who yearns to experience human life beyond the confines of her natural existence. Despite being warned that such a wish would shorten her long life into a fleeting moment, she willingly gives up her immortality for just one day as a human. Her journey reflects the universal desire to break free from limitations, even at significant personal cost. Another story follows The Little Mermaid, who sacrifices her voice and endures pain in pursuit of love and a soul that transcends death. Her tragic fate underscores the bittersweet nature of sacrifice and the idea that true fulfilment may lie not in earthly rewards but spiritual redemption.

Andersen's stories often contain elements of irony, humour, and social commentary, which are illustrated in Figure 4. In The Emperor's New Clothes, two swindlers convince the emperor that they have woven a magnificent suit visible only to intelligent and worthy individuals. Fearing judgment, everyone pretends to see the clothes—even the emperor himself—until a child finally speaks the truth: the emperor is naked. This tale is a sharp critique of vanity, authority, and the fear of appearing foolish. Similarly, other stories use personified objects, talking animals, or magical transformations to reflect on human flaws, societal expectations, and the unpredictability of fate. Andersen critiques human behaviour through these imaginative twists while entertaining his audience with vibrant storytelling.

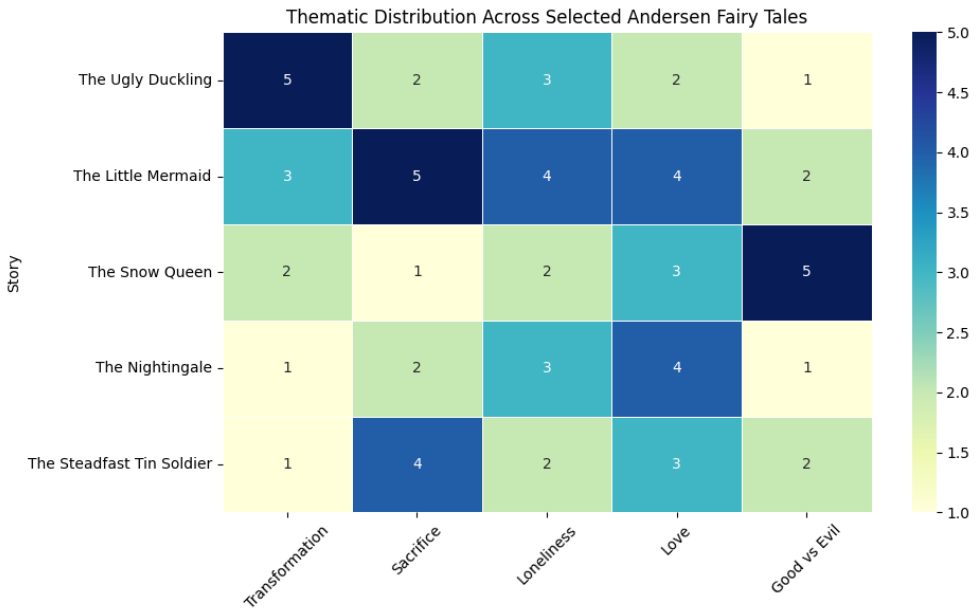


Fig 4. Thematic Distribution Correlation

Another story is that of The Little Mermaid, who gives up everything, including her voice, to pursue love and a soul that doesn't die. Her unfortunate turn of events serves as a poignant reminder of the bittersweet nature of sacrifice and the worrisome thought that real reward may not come from the world, but from within. In another story, a sentient shirt collar brags of significance, but gets pulped, dying to be made into paper: a lesson about pride and self-importance unexpectedly brought low. At times fantastical, these tales convey universal truths about the human condition; the stories meditate on who we are, our place in the world, the way we love and the terrors that unfold as we chase our desires.

In many of Hans Christian Andersen's fairytales, the motifs of Longing, change and self-denial are strikingly conveyed through vivid images and sentiments. One story from the series is a tale of a Dryad— a spirit that comes from a tree— that longs to have a whole human life that transcends its limitations within nature. In the face of being told that this choice will cut "short her everlasting life to mere moments," she decides to sacrifice the eternity that one day humans would have brought her. Her travels echo the eternal quest to overcome constraints, no matter the personal price. Then comes a tale on the heels of The Little Mermaid, who gives away her voice and suffers in return for love and a soul beyond death. Her tragedy illustrates the tragicomic aspects of sacrifice and the notion that the valid reward for our labours may not be found in terms of this world, but in the world to come.

Andersen's tales are frequently laced with irony, satire and social justice comments, as shown in Figure 4. In The Emperor's New Clothes, a pair of charlatans convince the emperor that they have created a fine suit of clothes visible only to wise and perfect men. Then, as an angel's throatful explodes, it is revealed that the lion's share of the citizenry only pretends to see the clothes-the emperor included—until a little child blurts out the embarrassing truth: The emperor has no clothes! It's a biting commentary on vanity, power, and the fear of being made a fool. Other tales, such as those featuring personified objects, talking animals, or magical metamorphosis, reflect human weakness, social morality, and the capriciousness of fate. In these fanciful stories, Andersen critiques the behaviour of humankind, even as he charms us with his timeless tales of wisdom and beautiful, sometimes gory details.

5. Conclusion

This article describes the development of a structured dataset of narrative events extracted from 153 fairy tales of H. C. Andersen, for use in literary study and computer-aided analysis. By systematically identifying and annotating notable narrative occasions in the full extent of the text collection, we demonstrate that classic literary canons can be transformed into a vital resource not only for such broad-based cross-disciplinary studies such as Digital Literary Studies, Natural Language Processing (NLP), and Computational Narratology, among others. The dataset provided in this paper involves detailed event type annotation, involved characters' information, etc., which can facilitate exploring the story structure, character evolution and the moral lessons reflected in Andersen's fairy tales. Through meticulous annotation and analysis of inter-annotator agreement, we clearly show that applying NLP to literary texts can be done, metaphorical, symbolic language notwithstanding.

Our analysis identified recurring elements and story structures, including The Ugly Duckling's transformative pilgrimage, The Little Mermaid's sacrificial journey, and the struggle between good and evil in The Snow Queen. These patterns complicate literary interpretation and show how moral and emotional concerns are interwoven with narrative design. Visual tools, such as thematic heatmaps and timelines, illustrated how core ideas — love, loneliness, change — unfold across the stories. This dataset can train models for event detection, timeline reasoning, story summarisation and moral analysis. It also provides the foundation for creating knowledge graphs, AI storytelling platforms, and educational apps that merge historical stories with advanced technology.

The research makes some significant contributions, but is limited in capturing literary language's ambiguity and stylistic richness. Future work will concentrate on wide-scale data expansion with task-specific annotations being performed on the data, like annotated cause and effect relationships, emotional arcs, and character roles, and domain-targeted NLP model development from literary text. The project underscores how weaving together humanities scholarship and computational approaches can open fresh avenues to understand and deploy stories that have stood the test of time. By processing Andersen's classic stories into machine-friendly form, we make a significant stride towards preserving cultural heritage and promoting research in digital humanities using artificial intelligence.

Acknowledgement

The authors would like to express their sincere gratitude to the Department of Electrical Engineering and Informatics, Universitas Negeri Malang, for providing the academic and research environment that supported the development of this work. We sincerely appreciate the facilities, guidance, and encouragement the faculty and staff offer throughout this study. This research would not have been possible without the institutional support and intellectual contributions from our colleagues within the department.

References

[1] Z.-H. Wang, R.-J. Tseng, H.-P. Lee, T.-F. Sheu, and Y.-T. Huang, 'An Interactively Application of Fairy Tales for Preschool Children'. 2016. [Online]. Available: http://papers.iafor.org/wp-content/uploads/papers/acset2016/ACSET2016_33369.pdf

[2] E. Daniati, A. P. Wibawa, and W. S. G. Irianto, 'Event Extraction in Narrative Texts: A Zero-Shot Approach Using Bert and Bi-LSTM on Andersen's Fairy Tales', in *2025 17th International Conference on Knowledge and Smart Technology (KST)*, IEEE, Feb. 2025, pp. 208–213. doi: 10.1109/KST65016.2025.11003342.

[3] B. Franz, 'Narrative time and International Relations', *J. Int. Relations Dev.*, vol. 25, no. 3, pp. 761–783, 2022, doi: 10.1057/s41268-022-00261-3.

[4] T. L. M. Suryanto, A. P. Wibawa, H. Hariyono, and A. Nafalski, 'Comparative Performance of Transformer Models for Cultural Heritage in NLP Tasks', *Adv. Sustain. Sci. Eng. Technol.*, vol. 7, no. 1, p. 0250115, Jan. 2025, doi: 10.26877/asset.v7i1.1211.

[5] Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, 'Advancements in natural language processing: Implications, challenges, and future directions', *Telemat. Informatics Reports*, vol. 16, p. 100173, Dec. 2024, doi: 10.1016/j.teler.2024.100173.

[6] Q. Zaman, S. Safwandi, and F. Fajriana, 'Supporting Application Fast Learning of Kitab Kuning for Santri' Ula Using Natural Language Processing Methods', *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 278–289, Jan. 2025, doi: 10.52088/ijesty.v5i1.713.

[7] H.-Y. Yu and M.-H. Kim, 'Automatic Event Extraction method for Analysing Text Narrative Structure', in *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, IEEE, Jan. 2021, pp. 1–4. doi: 10.1109/IMCOM51814.2021.9377386.

[8] A. R. Hofer, 'Sources: The Greenwood Encyclopedia of Folktales and Fairy Tales', *Ref. User Serv. Q.*, vol. 48, no. 1, pp. 94–95, Sep. 2008, doi: 10.5860/rusq.48n1.94.

[9] J. E. Bradford, 'The challenge of fairy tales : eliciting a critical gendered response in young children'. 2020.

[10] E. Daniati, A. P. Wibawa, W. S. G. Irianto, A. Ghosh, and L. Hernandez, 'Analysing event relationships in Andersen's Fairy Tales with BERT and Graph Convolutional Network (GCN)', *Sci. Inf. Technol. Lett.*, vol. 5, no. 1, pp. 40–60, May 2024, doi: 10.31763/sitech.v5i1.1810.

[11] H. C. Andersen, *Andersen's Fairy Tales*, 1st ed., no. January 1999. Guttenberg, 1999. [Online]. Available: https://www.gutenberg.org/ebooks/1597

[12] 'Event Causality Extraction with Event Argument Correlations', 2023. doi: 10.48550/arXiv.2301.11621.

[13] F. A. Ahda, A. P. Wibawa, D. Dwi Prasetya, and D. Arbian Sulistyo, 'Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation', *JOIV Int. J. Informatics Vis.*, vol. 8, no. 1, p. 231, Mar. 2024, doi: 10.62527/joiv.8.1.1818.

[14] Chidera Victoria Ibeh, Onyeka Franca Asuzu, Temidayo Olorunsogo, Oluwafunmi Adijat Elufioye, Ndubuisi Leonard Nduubuisi, and Andrew Ifesinachi Daraojimba, 'Business analytics and decision science: A review of techniques in strategic business decision making', *World J. Adv. Res. Rev.*, vol. 21, no. 2, pp. 1761–1769, Feb. 2024, doi: 10.30574/wjarr.2024.21.2.0247.

[15] 'Boosting Event Extraction with Denoised Structure-to-Text Augmentation', *arXiv.org*, vol. abs/2305.0, 2023, doi: 10.48550/arXiv.2305.09598.

[16] C. Zhou, 'Sentiment analysis of Twitter user text based on the BERT model', *Appl. Comput. Eng.*, vol. 52, no. 1, pp. 102–108, Mar. 2024, doi: 10.54254/2755-2721/52/20241380.

[17] HogenboomFrederik, FrasincarFlavius, KaymakUzay, de JongFranciska, and CaronEmiel, 'A Survey of event extraction methods from text for decision support systems'.

[18] A. Krzywicki, W. Wobcke, M. Bain, S. Schmeidl, and B. Heap, 'A Knowledge Acquisition Method for Event Extraction and Coding Based on Deep Patterns', in *Knowledge Management and Acquisition for Intelligent Systems*, Springer International Publishing, 2018, pp. 16–31. doi: 10.1007/978-3-319-97289-3_2.

[19] R. Li, P. Haslum, and L. Cui, 'Towards Learning Action Models from Narrative Text Through Extraction and Ordering of Structured Events', 2024, pp. 16–27. doi: 10.1007/978-981-99-8391-9_2.

[20] M. A. DeBuse and S. Warnick, 'Plot extraction and the visualisation of narrative flow', *Nat. Lang. Eng.*, vol. 30, no. 3, pp. 480–524, May 2024, doi: 10.1017/S1351324923000232.

[21] B. D. Samudera, N. Nurdin, and H. A. K. Aidilof, 'Sentiment Analysis of User Reviews on BSI Mobile and Action Mobile Applications on the Google Play Store Using Multinomial Naive Bayes Algorithm', *Int. J. Eng. Sci. Inf. Technol.*, vol. 4, no. 4, pp. 101–112, Oct. 2024, doi: 10.52088/ijesty.v4i4.581.

[22] M. Castro, A. Jorge, and R. Campos, 'TweetStream2Story: Narrative Extraction from Tweets in Real Time', 2023, pp. 217–223. doi: 10.1007/978-3-031-28241-6_17.

[23] P. Purnawansyah *et al.*, 'Congestion Predictive Modelling on Network Dataset Using Ensemble Deep Learning', *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 1597–1613, Dec. 2024, doi: 10.47738/jads.v5i4.333.

[24] J. M. Ferreira and C. Vectore, 'Fairy tales and mediational practice: possibilities to build cognitive repertoire in childhood education', vol. 20, no. 1, pp. 155–176, 2014, doi: 10.5752/P.1678-9523.2014V20N1P155.

[25] A. de Giorgio, A. Maffei, M. Onori, and L. Wang, 'Towards online reinforced learning of assembly sequence planning with interactive guidance systems for industry 4.0 adaptive manufacturing', *J. Manuf. Syst.*, vol. 60, pp. 22–34, Jul. 2021, doi: 10.1016/j.jmsy.2021.05.001.

[26] R. F. Cekinel and P. Karagoz, 'Event prediction from news text using subgraph embedding and graph sequence mining', *World Wide Web*, vol. 25, no. 6, pp. 2403–2428, 2022, doi: 10.1007/s11280-021-01002-1.

[27] F. iscus and A. S. Girsang, 'Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method',

Int. J. Eng. Trends Technol., vol. 70, no. 12, pp. 281–288, Dec. 2022, doi: 10.14445/22315381/IJETT-V70I12P226.

[28] A. K. Kolya, A. Ekbal, and S. Bandyopadhyay, 'A Hybrid Approach for Event Extraction', *Polibits*, vol. 46, pp. 55–59, Dec. 2012, doi: 10.17562/PB-46-6.

[29] 'Whats New? Identifying the Unfolding of New Events in Narratives', *arXiv.org*, vol. abs/2302.0, 2023, doi: 10.48550/arXiv.2302.07748.

[30] Y. Chen, Z. Ding, Q. Zheng, Y. Qin, R. Huang, and N. Shah, 'A History and Theory of Textual Event Detection and Recognition', *IEEE Access*, vol. 8, pp. 201371–201392, 2020, doi: 10.1109/ACCESS.2020.3034907.

[31] Y. Wu, H. Sun, and C. Yan, 'An event timeline extraction method based on news corpus', in *2017 {IEEE} 2nd {International} {Conference} on {Big} {Data} {Analysis} ({ICBDA})(*, 2017. doi: 10.1109/icbda.2017.8078725.

[32] 'The Fairy Tales and Anne Sexton's Transformation with Reference to Female Protagonists in Cinderella', □□□□, vol. 2, no. 50, pp. 695–715, 2023, doi: 10.31185/lark.vol2.iss50.3144.