Identifying Visual Depictions of Animate Entities in Narrative Comics: An Annotation Study

Lauren Edlin and **Joshua Reiss**
School of Electronic Engineering and Computer Science
Queen Mary University of London
L.Edlin@qmul.ac.uk, Joshua.Reiss@qmul.ac.uk

Abstract

Animate entities in narrative comics stories are expressed through a number of visual representations across panels. Identifying these entities is necessary for recognizing characters and analysing narrative affordances unique to comics, and integrating these with linguistic reference annotation, however an annotation process for animate entity identification has not received adequate attention. This research explores methods for identifying animate entities visually in comics using annotation experiments. Two rounds of inter-annotator agreement experiments are run: the first asks annotators to outline areas on comic pages using a Polygon segmentation tool, and the second prompts annotators to assign each outlined entity's animacy type to derive a quantitative measure of agreement. The first experiment results show that Polygon-based outlines successfully produce a qualitative measure of agreement; the second experiment supports that animacy status is best conceptualised as a graded, rather than binary, concept.

1 Introduction

Comics are a rich multi-modal medium for automatic discourse processing, yet empirical work investigating their narrative structures is still a nascent research area. Current approaches include computational descriptions of narrative structures that are used to automatically generate comics from chat scripts (Kurlander et al., 1996), video game logs (Shamir et al., 2006; Thawonmas and Shuda, 2008), or video (Yang et al., 2021). Approaches applying linguistic methods includes Visual Narrative Grammar, which categorizes panels based on their narrative function and describes grammar-like constraints distinguishing valid from invalid panel sequences (Cohn, 2013, 2020), and deriving relationships between high-level narrative structures and patterns of low-level text and image parts (Bateman et al., 2018).

This research is part of a larger project that seeks to identify narrative affordances in comics by examining compositions of units such as panels, text, symbols, characters, backgrounds, etc. Taking inspiration from annotation schemes for text narrative, we explore a method of annotation and assess the reliability with inter-annotator agreement experiments. This paper focuses on identifying animate entities in images. Our previous work examined coreference agreement of characters, and we hope to link image and discourse referents in the text in future work. We hope this work contributes to a full-fledged annotation scheme that can be applied to future corpora which would contain both annotations in the non-text areas of comics, as we look at here, and annotation of the textual areas of comic pages, for a truly multi-modal approach to discourse referents.

1.1 Identifying animate entities in comics

Identifying animate entities in comics is important for narrative analyses, as animate entities give rise to unique narrative affordances. A distinct feature of comics is that unlike other media such as film and literature, readers are prompted to infer an entity's movements, actions or intentions from static images and text. Information given in one panel primes the readers expectations for the next panel. A comic creator will therefore compose panels in a way that distinguishes entities that are not expected to move and think from animate entities, the latter of which structures events that progress the plot.

One narrative element of which animacy is foundational component is the concept of *character*. Successful annotation schemes used for corpus analyses of narrative in text have defined characters as "an animate being that is important

Proceedings of the The 5th Workshop on Narrative Understanding, pages 82–91
July 14, 2023 ©2023 Association for Computational Linguistics

Figure 1: An example of the character Ms. Marvel changing her appearance, and arguably her animacy status, over several panels (Wilson et al., 2015, p. 1)

to the plot" (Jahan and Finlayson, 2019, p. 13). Binary (Moore et al., 2013) and hierarchical (Zaenen et al., 2004) animacy annotation schemes have been developed to describe animate entities types that constitute characters. While work on character tracking for cohesion analysis is being applied to comics (Tseng et al., 2018; Tseng and Bateman, 2018), adding animacy as a criteria of character could aid in character identification (Jahan et al., 2018), as well as recording non-character animate entities and potential characters.

Determining whether a drawing depicts an animate entity, however, is not straightforward. Cases where animacy is ambiguous or uncertain regularly appear in comics as they often tell narratives about fantastical scenarios; many science fiction and fairy tale stories include things like talking animals, zombies, aliens and robots, which may or may not meet a threshold for animacy. Furthermore, an entity's animacy status may change or be hidden from the reader. An example of this is depicted in Figure 1 where Ms. Marvel (Kamala Kahn) is a superhero with shape-shifting powers. In these panels, Ms. Marvel is disguised as a sofa before transforming back to her typical appearance. Ideally, an annotated corpus or computational model would track these two depictions as the same referent while also accounting for this change of animacy status.

Consequently, a satisfactory annotation scheme should include relevant animate entities beyond the notion of character. This research proposes and tests an initial annotation scheme for identifying areas on images visually representing animate entities on comic pages through two annotation experiments. Experiment 1 asks annotators to identify animate entities by outlining them on a digital comic page, and experiment 2 prompts annotators to select the type of animacy for each annotated entity from experiment 1. The levels of inter-annotator agreement are measured for each, with conclusions drawn about the causes of disagreement for each task to inform future work.

2 Experiment 1: Identifying animate entities using outlines on comic pages

This first experiment tests a method for delineating animate entities according to reader judgments. Annotators are prompted to outline areas directly onto comics pages where they believe shows a depiction of an animate being. Having annotators draw on the page circumvents the researcher's assumptions about what should be considered animate, since the researcher is not pre-selecting potential candidates for annotators to judge. The areas outlined by each annotator are compared against one another to produce a qualitative measure of agreement.

2.1 Methods

2.1.1 Annotation scheme and implementation

Annotators are given an annotation scheme and a digital comic within a responsive browser-based tool to outline areas directly onto a given comic page. The Comics Annotation Tool (CAT) is an online interface that facilitates remote annotation. The main CAT interface is shown in Figure 2 - comic pages are given one at a time on the left, and annotation prompts are given on the right. Annotators use their keyboard and mouse to create closed polygons on the digital canvas. Panels, shown in red, are pre-segmented to guide annotators when making their outlines. The panels are pre-segmented because panel identification has very high inter-annotator agreement according to previous work (Edlin and Reiss, 2021). Animate entities are outlined in purple, and each closed

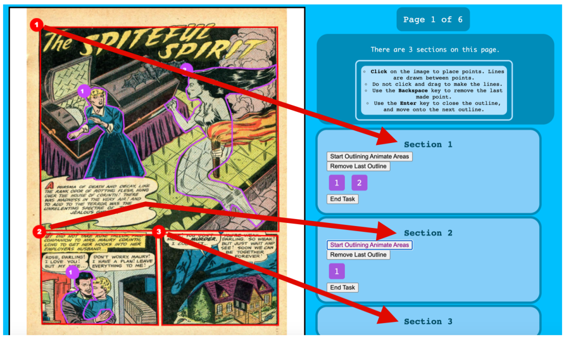

Figure 2: The main interface of the CAT. The red arrows point the panel ID number to the corresponding section.

outline generates a corresponding purple number in the panel/section where it was drawn.

The intent of this annotation project is to capture the build-up of information in the reader's mental model, rather than the comic authors perspective. Annotators are therefore instructed to try their best to segment areas within each panel in the order they would normally read, and make outlines based on the information they have up to the current point that they've read; they are not to go back and revise their outlines based on information unavailable at that point in the narrative, as this would not accurately capture how their mental model developed. Annotators are given whole pages to facilitate a natural reading experience, and to allow the use of all information, including visual cues and text, to make their annotations in both experiments.

The full annotation scheme and instructions can be downloaded directly from the CAT. The definition of an animate being or thing given to annotators is: *a depiction of an entity that displays human or higher animal-like behaviours, and/or can communicate autonomously and move intentionally.*[1] Some examples of animate versus inanimate representations from the annotation scheme are given in Table 1.

Four comics were selected from the *Alarming Tales*, which is a comics magazine that ran for six issues between 1957 and 1958. These comics were chosen as they are out of copyright, and exhibit a common artistic style of illustration that

[1]All supplementary material is available at `https://github.com/le300/CAT_Annotation_Experiment_3`, including the full annotation schemes for both experiments with examples and more detailed explanations, the comics used for annotation, and all code implemented in the evaluations.

Animate	Inanimate
Talking tree	Tree blowing in the wind
Wolf	A dead wolf
Sentient A.I	Supercomputer
Sleeping teacup	Self-playing piano

Table 1: Examples of animate and inanimate entities from the annotation scheme.

Story no.	**Age** (mean/range)	**Gender**
1	31.4 (21-37)	1F/4M
2	36.4 (22-50)	1NB/2F/4M
3	35.2 (21-51)	4F/1M
4	29.2 (22-38)	4F/1M

F=female, M=male, NB=non-binary

Table 2: Participant demographics for experiment 1.

persisted throughout the silver age of comics, and are all of the same sci-fi fantasy genre. Except for one four-page comic, all other stories are five pages which gives a total of 19 pages for annotation. Three of these comics were used in previous annotation experiments, allowing for comparison between studies. Finally, all stories appear to exhibit entities with unclear animacy according to the lead author's judgment. The comics were downloaded from Comic Book Plus,[2] which is an internet archive of open source and copy right free comics.

2.1.2 Participants

Five participants produced annotations per story for a total of 20 annotators. All participants were required to be fluent in English and have UK or US nationality. No participants annotated more than one story to prevent some annotators becoming familiar with using the CAT than others. An overview of participant demographics are given in Table 2.

All participants were recruited on the online crowd-sourcing platform Prolific. Crowd-sourcing has been shown to be an efficient method for comics annotation (Tufis and Ganascia, 2018), and annotation experiments in our previous work found that word-of-mouth and crowd-sourcing recruitment produced similar results on similar tasks (forthcoming). Participants were compensated £11/hour through the Prolific platform.

	Animacy mean IOU scores			Character mean IOU scores		
Story no.	Mapped-only	Unmapped-included	Diff.	Mapped-only	Unmapped-included	Diff.
1	0.725	0.649	0.076	0.725	0.694	0.031
2	0.704	0.695	0.009	0.802	0.795	0.007
3	0.693	0.629	0.064	0.603	0.538	0.065
4	0.716	0.641	0.075	-	-	-

Table 3: Results for animacy outline agreement compared with results from previous work on character outline agreement.

2.1.3 Inter-annotator agreement metrics

A rough estimate of annotator agreement is counting the number of overlapping outlined areas through the researcher's judgment - the more annotators that outlined the same areas, the higher the agreement. However, the precision of outlines between annotators will differ, as some annotators may leave a larger gap between the boundaries of the illustration part they intend to indicate. A qualitative judgment is therefore insufficient to count outline overlaps, especially if a panel is crowded with lots of outlines.

To confirm whether two outlines sufficiently overlap, we use the quantitative metric of *Intersection over Union* (IOU, or Jaccard Index). IOU is a similarity metric of two sets - more precisely, the size of the intersection divided by the size of the union of given sets A and B: $IOU(A, B) = |A \cap B|/|A \cup B|$. In this case, set elements are the pixels within an outline. IOU scores are calculated using the Python Shapely library, which defines the annotator's outlines as closed polygon objects. We use in-built intersection and union functions to determine the IOU score between two polygons.

An IOU score is between $[0, 1]$. The threshold for sufficient overlap between two outlines is not established as there is no ground truth for comparison. A similar experiment in our previous work, with tested agreement for identifying characters, found an overlap threshold of 0.6 was adequate for rectangular bounding boxes, although the agreement threshold between stories ranged from 0.6 to 0.8. We use the same technique here to determine an overlap threshold for polygon outlines.

Overall IOU agreement between a given pair of annotators is calculated by iterating over every panel in a story, and trying every possible mapping from one annotator's outlines to the other's in each panel. The permutation of outline pairs for each panel that yields the highest IOU score, is stored as the inter-annotator mapping for that pair of annotators, and those pairs are considered "mapped' segments. All the IOU agreement scores for that panel are summed for use in an overall mean IOU score.

Naturally, one annotator may make more outlines than the other on a given page. If there is a mismatch in the number of outlines, an "empty" outline (with polygon area 0 and panel intersection 0) is added accordingly. Empty outlines always receive an IOU score of 0 when compared to all outlines from another annotator, hence penalizing the overall IOU score. Both *mapped-only* and *non-mapped-included* pair-wise IOU score distributions are calculated. The differences between mean pair-wise IOU scores is consequently a relative measure between stories.

Finally, additional qualitative data is collected to help interpret disagreements. The CAT includes a text area per page where annotators can write to express uncertainty regarding the animacy of identified entities. The reasons given by annotators can aid explanations of disagreement and guide the next experiment.

2.2 Results

Table 3 gives both the mean IOU scores for the distributions of annotator-pair scores, both mapped-only and unmapped-included outlines, per story. The results from previous work testing the annotation scheme for character using bounding box annotation (Edlin and Reiss, 2021) are also included for comparison, as the same stories were used except for story 4. Figure 3 shows the animacy distributions as boxplots; the median, 1st and 3rd quartiles, and min to max pairwise IOU scores per distribution are emphasised. Overall, the IOU threshold for overlapping outlines appears to be 0.7. The mean overlap IOU per story is more consistent compared to the results for character, indicating that polygon outlines are more reliable than bounding boxes.

[2] http://www.https://comicbookplus.com/

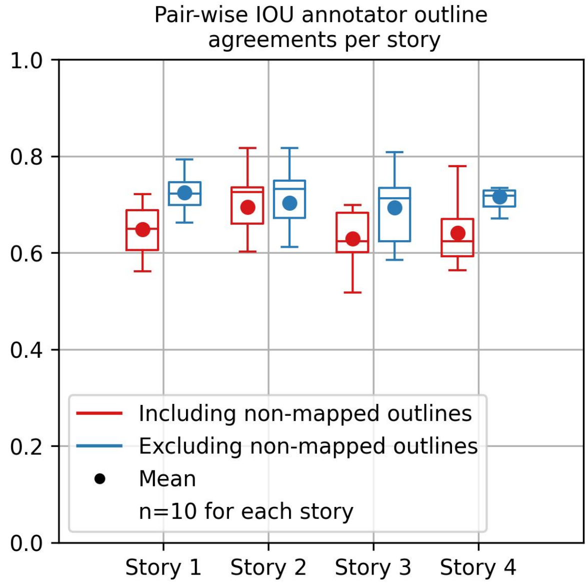

Figure 3: Boxplots showing both mapped-only and non-mapped-included distributions of IOU annotator pair agreements, per story.

Figure 4: Each panel shows an example of the most disagreed upon entity per story. Starting in the top-left and going clock-wise: story 1 features plant-men, story 2 has a man entering the fourth dimension, story 3 features a robot-plant, and story 4 shows a blob-like creature being shot and killed.

The difference between the two mean IOU scores per story provides a measure of agreement relative to the other stories. For the animacy results, Story 2 has the lowest difference between mean IOU scores, while Story 1 and 4 exhibit the greatest amount of disagreement. Disagreement between stories is due to both reader interpretation of entities as well as the number entity instances - a frequently appearing entity with uncertain animacy will pull the IOU score lower. Since reference labels for entities were not annotated as in previous work, the exact number of an individual entity's instances cannot be objectively verified. However, possible explanations for disagreement by taking instance frequency into account may be derived using annotator's written feedback.

Entities that are clearly human have negligible disagreement, as any disagreement can be explained as errors with using the CAT. Each story has a particular entity that elicits the most disagreement. Story 1 has high disagreement because it features "plant-men" being grown on mass as shown in Figure 4. Individual annotator's outline counts are between 64-112, which is the widest range of all stories; the high frequency of plant-men causes a double IOU penalty.

While the plant-men elicit disagreement on pages 1 and 2, they become unanimously agreed upon as animate by page 3-panel 4, and remain so the rest of the story. This panel appears to depict the plant-men moving on their own rather than planted in the ground. Annotator 5 states: "(It's) still unclear if the plant men are animate in section 1 and 2, but by section 4 and 5 it looks like they are displaying higher animal-like behavior (choosing to fight)." This suggests the their status changes due to new evidence of intentional movements, implying that intentional movement is more indicative of animacy than simply a human-like appearance.

Story 2 has the lowest disagreement. Annotators outline counts range from 43-47, suggesting fewer potential entity instances than in story 1. This story is about a man who enters the "fourth dimension" where his appearance changes into abstract shapes. Significant disagreement occurs in the very first panel on page 1 where the man is entering the fourth dimension for the first time, as shown in Figure 4. Some annotators outlined the whole entity, others outlined two separate parts divided by the plane, and others did not make any outlines. Several annotators expressed uncertainty about the fourth dimension. For example, annotator 1 states "... I'm not given enough information as to whether the in the fourth dimension, humans can communicate and move on their own volition".

While story 3 also appears to have fewer unique entities than story 1 with a range of 48-58 individual annotator outline counts, disagreement occurs when a "robot-plant" is introduced on page 3. Annotators disagreed due to uncertainty as to whether it's movements are intentional. The agreement increased on page 4 as information about the robot plant's intentions are described in a speech by a human agent. Annotator 4 states "I thought the plant only displayed animate features when the text stated it was trying to water the men. Before that point, it was just growing as is usual." Unlike other stories where annotators only cited visual cues pointing to animacy status, in this case the text information was a significant factor in judging animacy status.

Finally, story 4 has a difference score similar to story 1, however fewer outlines were made overall with 35-44 range individual outline counts between annotators. This story is about a hunter on an alien planet. Two aliens are featured: one a blob-like creature, and the other dinosaur-like creature. The blob creature elicited more disagreement overall. However, both creatures were shown as being shot and killed, which elicited disagreement on whether these creatures remain animate while in the process of being shot. While the annotation scheme explicitly states that dead entities should not be outlined, some annotators continued to outline instances after the shooting depictions of shooting. This instruction to stop outlining killed entities therefore appears to be unintuitive.

2.3 Discussion

These findings point to several indicators of animacy, including: (a) being or having been shown to have been a human, (b) showing evidence of autonomous movement and speaking, (c) having the appearance of an animal, and (d) not having the appearance of a plant. Evidence of movement with intent appears to be a foundational facet of animacy, as the plant-men in story 1 and the robot-plant in story 4 gained higher agreement once they were considered to be moving on their own volition with clarified intentions.

These indicators could be interpreted as a coarse hierarchy that roughly reflects the one described by Zaenen et al. (2004). Humans are at the hierarchy's apex, with entities shown speaking in language just beneath. Entities with autonomous movements are next, however the threshold between human-like and animal-like animacy cannot be distinguished based on autonomous movement alone. For example, the human-like appearance of the plant-men in story 1 may suggest a higher level of animacy than the robot-plant in Story 4, even though they are both understood to be animate once they show intentional movements. Since animal-like refers to animals such as mammals, birds and reptiles, considering lower-animal level of consciousness may better describe certain cases where voluntary movement is not an animacy requirement. Lastly, the lower levels in the hierarchy include being a robot, followed by appearing as a plant. A dead entity would be at the bottom of the hierarchy if included.

Lastly, it appears that this method is successful in capturing a reader's updating mental model. This also suggests that animacy ambiguity in itself is a compelling narrative technique - some entities produced high disagreement in the beginning of the story only to increase in agreement as the story progresses. This uncertainty about an entity's animacy attentuates the ability of a reader to infer what comes next, which prompts the reader to continue on to resolve the tension and subsequently progresses the plot.

3 Experiment 2: introducing an animacy hierarchy

This experiment builds on the previous one by asking annotators to assign a hierarchical animacy type to pre-outlined entities on comic pages. Experiment 1 implemented animacy identification through a coarse binary choice - outlined areas contain an animate entity according to the annotators judgment, while everything outside the outline does not. While the results of this technique gives initial insight into animacy indicators, the nature of disagreements can be further parsed using a quantitative metric. A simple hierarchy of animacy status is devised to assess whether agreement can be achieved by offering annotators several options.

3.1 Methods

3.1.1 Annotation scheme and implementation

The new annotation scheme asks participants to judge the animacy status of a pre-outlined entity by selecting the best description from an animacy hierarchy list. The hierarchy broadly reflects potential animacy thresholds suggested

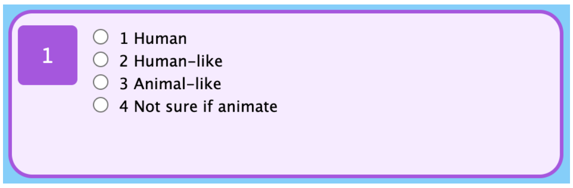

Figure 5: An example of animacy type choices as displayed in the CAT.

Example	Animacy type	Reason
Talking tree	Human-like	Speaks language through its own volition
Wolf	Animal-like	Displays animal-like behaviours
Zombie	Not sure if animate	Depends on the rules of the world established in the story - if the state of mind is unclear, it is best to put *not sure if animate*
A Sentient AI	Human-like	Speaks, thinks, or shows other higher-level behaviours
A Supercomputer	Not sure if animate	If an AI is not sentient, select the lowest animacy option of *not sure if animate*
Dragon	Animal-like	Displays animal-like behaviours

Table 4: Examples of entities, their animacy type, and the reason for the type assignment from the annotation scheme.

in experiment 1. *Human* is ascribed to clearly human entities. *human-like* refers to entities with sentience and intentions, primarily indicated through movement, speaking language, or showing other behaviours like deliberate humor or planning. Entities with sentience that cannot speak but show behaviours displayed by animals such as mammals, birds, reptiles, etc. are assigned *animal-like* animacy. Lastly, *not sure if animate* is chosen when there is uncertain, ambiguous, or no animacy detected. A pilot study was run that included an *inanimate* option at the bottom of the hierarchy, however this seemed to confuse annotators as an outline entity already suggests at least a slight potential for animacy. Overlapping outlines made by only one or two annotators out of five in experiment 1 are expected to be assigned *not sure if animate*.

The CAT is updated to deploy these tasks for remote annotation. Figure 5 shows an example form with animacy type choices. Each pre-made outline has a corresponding form, where the options are placed from top to bottom according to the highest-level animacy to lowest. Table 4 gives some examples entities of various animacy types.

Story no.	**Age** (mean/range)	**Gender**
1	39.8(21-62)	2F/3M
2	29.2(21-37)	1F/4M
3	31.2(23-39)	2F/3M
4	29.4(21-43)	3F/2M

F=female, M=male, NB=non-binary

Table 5: Participant demographics for experiment 2.

The same comics from experiment 1 are used to compare results. Outlines are placed where at least one annotator made an outline in the previous experiment. These are considered areas of potential animacy, as non-outlined areas indicate that all annotators agreed that there is no animate entities present.

3.1.2 Participants

Participants were again recruited from Prolific, with 5 unique participants allocated per story for a total of 20 annotators. The same criteria of English fluency and UK or US nationality were required, and each annotator was compensated £11/hour. An overview of participant demographics is given in Table 5.

3.1.3 Inter-annotator agreement metrics

Krippendorff's α (KA) is a standard inter-annotator reliability measure (Artstein and Poesio, 2008; Krippendorff, 2011). A KA score is between $[-1, 1]$; -1 indicates complete disagreement, 1 indicates complete agreement, and 0 indicates chance agreement. A score of 0.8 is considered a threshold for excellent agreement, while 0.68 is considered sufficient agreement (Artstein and Poesio, 2008, p. 591). Annotator's ratings are tested all-against-all to provide an overall KA score per story.

Ratings between annotators are weighted in the KA calculation according to the data measurement scale. The annotation scheme describes animacy types by name, as well as a numbered top to bottom hierarchy in the CAT itself. Therefore, KA scores for both ordinal and nominal weightings are calculated. The KA scores are calculated using the Krippendorf python package (Castro, 2017).

	All-against-all animacy KA scores	
Story no.	Nominal	Ordinal
1	0.441	0.541
2	0.551	0.751
3	0.388	0.558
4	0.72	0.787

Table 6: Results for Experiment 2, including both the all-against-all KA scores for the animacy hierarchy task and the mean pair-wise reference KA scores, per story.

3.2 Results

All-against-all KA scores per story are shown in Table 6. The nominal weighted scores consistently produced lower agreement than the ordinal weighted scores, with only story 4 reaching the threshold for adequate agreement. Both stories 2 and 4 reach adequate agreement with the ordinal scale weighting, while stories 1 and 3 only achieved middling agreement. Overall, adequate to high agreement was not universally achieved across all stories according to either scale. Nevertheless, these results support further development of an animacy hierarchy assignment using an ordinal-type ranking.

Stories 1 and 3 exhibit low agreement. As in experiment 1, story 1 elicits high disagreement due to the plant-men - two annotators primarily assigned them *human-like* animacy, while the other three assigned *not sure if animate*. Unlike the findings from experiment 1, only one annotator indicated a change in animacy status as the story progressed by updating the plant-men from *not sure if animate* to *human-like* on page 3. Story 3's low agreement is also again due to the robot-plant instances, where some annotators primarily categorized the robot-plant as *not sure if animate*, while others assigned *animal-like* animacy.

Similar to experiment 1, story 2 shows higher agreement than stories 1 and 3. Story 2 features the man entering the fourth dimension; this character is shown both as a normal human and then as a series of abstract shapes that can still walk and talk in the fourth dimension. Annotators mainly ascribed *human-like* animacy to instances where the man is fully in the fourth dimension, demonstrating a step down the hierarchy from *human* to *human-like*. Unlike the first experiment, disagreement actually occurred for another entity that first presents as a human-shaped shadow or a silhouette. Annotators assigned *human*, *human-like* and *animal-like* to this entity.

Finally, story 4 exhibits the highest agreement overall. Besides several instances of the blob and dinosaur-like the creatures, the story only appears to have humans which contributes to the high agreement. The blob-like alien again produced the most disagreement with annotators either choosing *not sure if animate* or *animal-like* animacy. The dinosaur-like alien was consistently assigned *animal-like* animacy.

3.3 Discussion

While providing a useful quantitative metric of disagreement, the hierarchy does not accurately capture judgments of animacy status for entities across these comics, as only two of the stories achieved adequate agreement using ordinal KA weightings. Examples of where this hierarchy fails can be seen in story 1 and story 3 which had the lowest scores. Both stories feature plant appearing entities - namely the plant-men from story 1 and the robot-plant from story 3. Since these entities caused significant disagreement in the previous experiment, the expectation in this experiment was for annotators assign them *not sure if animate*. However, if an annotator did detect animacy, then neither *human-like* nor *animal-like* animacy intuitively describes their status. An added lower-level category on the hierarchy that describes non-human and non-animal-like entities would be a beneficial addition.

The all-against-all KA scores were nonetheless consistently higher using ordinal weightings. Further development of an animacy type classification should therefore explicitly use an ordinal scale. Getting rid of category names to instead use a numbered scale may circumvent potential confusion due to the animacy category names themselves - naming a category *animal-like* may exclude some relevant entities like the blob-like creature from story 4, for instance. A numbered scale is also a less coarse and may more easily include lower levels of animacy without having to actually name them.

Finally, recall that the results from experiment 1 suggest that animacy ambiguity can be a purposely used narrative device. Capturing this type of ambiguity within the annotation scheme is therefore especially important for future work on narrative understanding. The *not sure if animate* category implemented in this experiment does not clearly achieve this. In future work, an ambiguity measure could be derived from disagreements themselves, or measured by developing an annotation scheme from the intentions of the author rather than the perspective of the reader.

4 Conclusion

Animate entities are an important component of defining characters and understanding broader narrative structures and affordances in comics. We explored methods of identifying animate entities visually through two rounds of annotation experiments: the first asked annotators to outline areas on comic pages with a polygon outlining tool, and the second attempted to quantify agreement by having annotators assign a hierarchical animacy category to each entity.

Results from the first experiment show that an outline-based agreement method, which imposes a binary concept of animacy, can qualitatively measure agreement between stories. In experiment 2, we develop a hierarchy of animacy types based on disagreements from the first experiment, with the results showing that a graded rather than categorical concept of animacy performs adequately for some stories. An ordinal scale may better capture edge cases where the line between human-like and animal-like behaviours; for instance, a crow shows human-like behaviours such as strategical planning. Additionally, both studies suggested that purposeful animacy ambiguity is a valuable narrative tool, and should be accounted for in future developments of the annotation scheme.

4.1 Future work

In future work we will refine the annotation scheme for implementation for corpus analyses. The resulting corpora, along with other annotations, can be used to derive narrative structures from lower-level units. The method from Experiment 1 for visual discourse referent annotation can be incorporated with linguistic reference annotation in the comic's text for multi-modal discourse processing. Additionally, both studies suggested that purposeful animacy ambiguity is a valuable narrative tool, and should be accounted for in future developments of the annotation scheme.

Limitations

Several limitations relate to the study setups. One shortcoming is the small number of comics used in the experiments. While this is adequate for initial exploration into animate entity identification, these results cannot be generalised - comics in particular have an incredible number of potential types of animate entities and beings, and four comic stories do not touch on most of them. Another limitation is the reliance on crowd-sourced recruitment and remote annotation. The researcher is not able to instruct the annotators in person and check their understanding of the annotation scheme. Although the annotation task are seemingly relatively simple at this point, word-of-mouth recruitment of annotators who are more familiar with annotation processes are likely a better choice in future work, especially as the annotation scheme develops to include more complicated concepts.

More significantly, the outlining method for identifying animate entities does not capture inanimate entities that become animate, as discussed in Section 1.1. An annotator using the scheme tested here would not outline the sofa in Figure 1, although the sofa should be included in an annotated corpus to capture an important update to the reader's mental model. While these experiments show that these updates are somewhat obtained through interpreting reader feedback about their outlines, the limitations in developing an annotation scheme solely from the reader's perspective are apparent. Developing a comparable annotation scheme from the creator's perspective may facilitate fuller analyses of narrative structures. Since a creator knows that the sofa and Kamala Kahn are linked through coreference, both would be outlined and given the same reference label. Integrating these two perspectives into one corpus could give insights into how creator's intentions to communicate larger narrative structures are expressed in lower-level configurations of image and text.

Acknowledgements

We offer a sincere thank you to all the annotators who took part in this study, and to the two anonymous reviewers who provided valuable and detailed feedback. This research was funded by the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

References

Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

John A Bateman, Annika Beckmann, and Rocío Inés Varela. 2018. From empirical studies to visual narrative organization: Exploring page composition. In *Empirical Comics Research*, pages 127–153. Routledge.

Santiago Castro. 2017. Fast Krippendorff: Fast computation of Krippendorff's alpha agreement measure. https://github.com/pln-fing-udelar/fast-krippendorff.

Neil Cohn. 2013. *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images.* A&C Black.

Neil Cohn. 2020. Your brain on comics: A cognitive model of visual narrative comprehension. *Topics in cognitive science*, 12(1):352–386.

Lauren Edlin and Joshua Reiss. 2021. An empirically-based spatial segmentation and coreference annotation scheme for comics. In *Proceedings of the 14th International Symposium on Visual Information Communication and Interaction*, pages 1–8.

Labiba Jahan, Geeticka Chauhan, and Mark A Finlayson. 2018. A new approach to animacy detection. In *Proceedings of the 27th International Conference on Computational Linguistics*.

Labiba Jahan and Mark Finlayson. 2019. Character identification refined: A proposal. In *Proceedings of the First Workshop on Narrative Understanding*.

Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability.

David Kurlander, Tim Skelly, and David Salesin. 1996. Comic chat. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 225–236.

Joshua Moore, Christopher JC Burges, Erin Renshaw, and Wen-tau Yih. 2013. Animacy detection with voting models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 55–60.

Ariel Shamir, Michael Rubinstein, and Tomer Levinboim. 2006. Generating comics from 3d interactive computer graphics. *IEEE computer graphics and Applications*, 26(3):53–61.

Ruck Thawonmas and Tomonori Shuda. 2008. Comic layout for automatic comic generation from game log. In *New Frontiers for Entertainment Computing*, pages 105–115. Springer.

Chiao-I Tseng and John A Bateman. 2018. Cohesion in comics and graphic novels: An empirical comparative approach to transmedia adaptation in city of glass. *Adaptation*, 11(2):122–143.

Chiao-I Tseng, Jochen Laubrock, and Jana Pflaeging. 2018. Character developments in comics and graphic novels: A systematic analytical scheme. In *Empirical Comics Research*, pages 154–175. Routledge.

Mihnea Tufis and Jean-Gabriel Ganascia. 2018. Crowdsourcing comics annotations. In *Empirical Comics Research*, pages 85–103. Routledge.

G. Willow Wilson, Wyatt Jacob, and Adrian Alphona. 2015. Ms.marvel: Generation y.

Xin Yang, Zongliang Ma, Letian Yu, Ying Cao, Baocai Yin, Xiaopeng Wei, Qiang Zhang, and Rynson WH Lau. 2021. Automatic comic generation with stylistic multi-page layouts and emotion-driven text balloon generation. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(2):1–19.

Annie Zaenen, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M Catherine O'Connor, and Thomas Wasow. 2004. Animacy encoding in english: Why and how. In *Proceedings of the workshop on discourse annotation*, pages 118–125.