

Heckling at ontologies

Saul Albert, Toby Harris and Pat Healey

This research created **authorial descriptions** and **audience annotations** of the same episode of Doctor Who, then compared them as media metadata.

Media is moving online, where richer metadata can enable more flexible and inter-linked production, distribution and reception processes. However, creating TV media metadata that is relevant to both authors, broadcasters and viewers is a significant practical and conceptual challenge.

To investigate the differences between these contexts of media production and use, a formal, script-based media ontology was used to describe an episode of

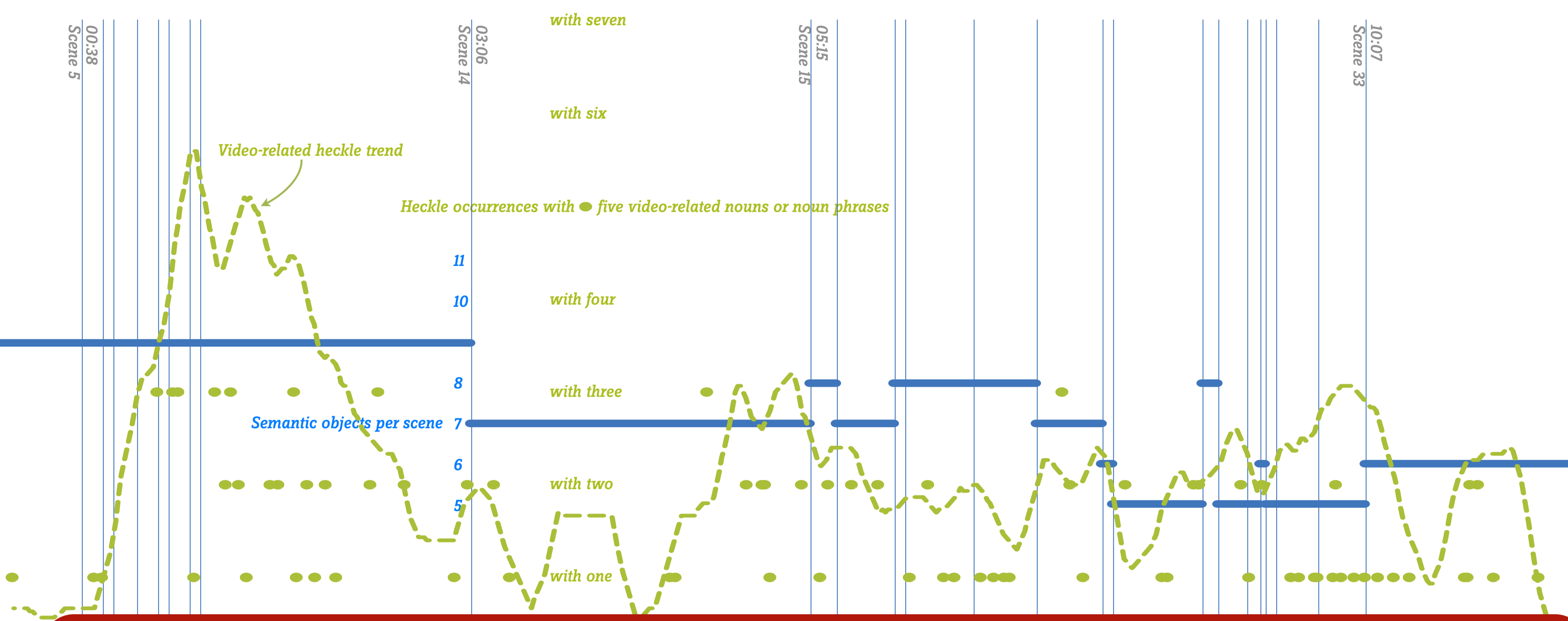
Doctor Who, which was then compared to transcripts of text-chat conversations between viewers of that episode.

Results revealed a limited but potentially useful overlap in how authors, broadcasters and viewers describe and use media. Our research suggests directions that could exploit that overlap to create and evaluate media metadata in novel ways.

Dataset 1: Semantic, authorial representation modelling content and narrative.

We worked with BBC information architects and production staff¹ to develop the 'BBC Stories' ontology². Capable of modelling fact or fiction, we took a time-bending episode of Doctor Who and created the meta-data to represent the people, places and things, the events that tie them together, and the interpretations that create stories from them. Data preparation: the dataset was loaded into an RDF triple-store and queried to produce a list of objects referenced in the scenes corresponding to the first 15 minutes of the programme.

1. http://www.bbc.co.uk/rd/projects/mythology_engine.shtml
2. <http://toby.net/projects/bbcstories>



Comparison 1: Temporal distribution of authorial and audience datasets

This graphs the first 15 minutes of the programme from left to right. The stepped blue line shows the quantity of objects referenced per scene in the authorial data, along with the scene boundaries modelled in that data. The dashed green line shows the trend of heckles referencing the programme content, constructed as a rolling average of the individual heckle occurrences weighted for the quantity of references in each.

Observations and analysis:

Comparison 1: neither the count of **semantic objects per scene** nor the position of scene boundaries correlate to the count or distribution of **audience mentions of programme-related things**, suggesting very different intentions and imperatives in how the datasets were produced.

Discussion and conclusions:

Attempts to draw comparable interpretations from the datasets showed that despite ostensibly representing the same media object, the two datasets are very different.

More data and more complex analyses than were used here may help to align production-oriented media metadata with audience response^{1,2}.

However, this research suggests that media objects are used and interpreted in such different ways by producers, broadcasters and viewers that these hard-won alignments are unlikely to produce media metadata that is meaningful for all three.

[1] Shamma, D., & Kennedy, L. (2010). Tweetgeist: Can the twitter timeline reveal the structure of broadcast events. CSCW Horizons.

Further work:

Asking what kinds of **user experiences** could help bring these two descriptions together, we have built a time-based **visualisation** of both datasets as an annotation layer on top of the video, and are working on **interfaces** to enhance audience chat using minimal but sufficient semantic metadata eg. 'conceptual auto-complete'.

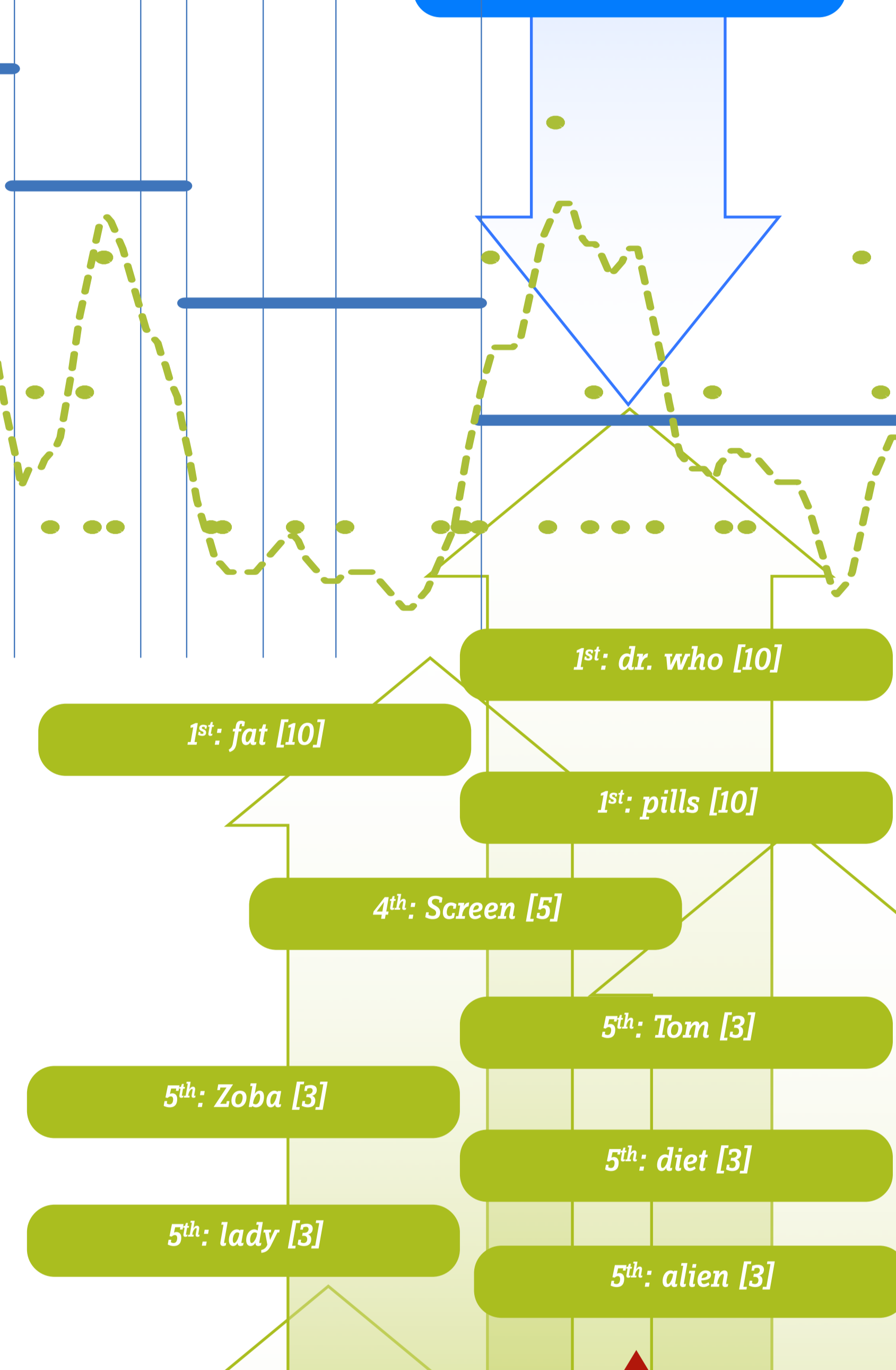
Comparison 2: no terms are common to both datasets, although there is some **conceptual overlap**. Some terms represent the same concept: "The Doctor" vs "dr who". Some terms represent different concepts that the semantic model asserts are related: ie. the "Adipose" are "aliens".

- > The authorial model's detailed allocation of descriptive resource does not correspond to our audience's interest as manifest in their chat.
- > Lexical overlaps are subject to constant shifts in the contingencies of audience communication.
- > The semantic model offers terms for concepts and their relationships that could provide a resource for audience communication, and in doing so aid analysis of that communication.
- > We suggest that conceptual overlaps may be reasoned about if both the semantic model and interactional structure are harnessed.

[2] Inches, G., Basso, A., & Crestani, F. (2011). On the generation of rich content metadata from social media. SMUC '11

- > Could audience transcripts **guide and verify** the authoring of ontologies?
- > The **economics of metadata production** could be transformed by determining what minimums of authored metadata and interactional devices are needed to make audience transcripts tractable for harvesting metadata.

- 5th: Miss Foster [21]
- 6th: Psychic Paper [21]
- 5th: The Noble home [25]
- 4th: Adipose Industries [37]
- 3rd: Donna Noble [49]
- 2nd: The Doctor [51]
- 1st: Adipose [55]



Comparison 2: Most used terms in each set

A ranked list of the most frequently occurring terms from each dataset, with their count.

Dataset 2: Free-text aggregation of mediated conversation about the programme by viewers.

We held two live screenings³ for groups of 15 people during which they used a custom 'heckle' multimedia chat system⁴ that captured their tweets and interactions in time with the video.

Data Preparation: nouns and noun phrases within the text corpus corresponding to the first 15 minutes of the programme were coded and counted.

3. <http://qmat.net/2011/10/heckle-at-social-tv/>
4. <http://thepeoplespeak.org.uk/blog/2011/10/04/heckle/>

