

Comparing British and American Audio Description of Movies

Igor Sterner¹, Alex Lascarides¹ and Frank Keller¹

¹*School of Informatics, University of Edinburgh, United Kingdom*

Abstract

Narrating the visual component of movies is known as audio description. It is a narrative technique designed to enable blind and visually impaired individuals to follow the story. However, it is far more constrained than most narratives: the descriptions not only need to convey the story in the movie, but they must also fit into gaps between dialogue and they need to conform to guidelines that exist in each region. In this work, we compare audio description created in the United Kingdom against audio description created in the United States. We use guidelines written for these two regions, alongside the impressions from a practitioner in the field, to motivate specific hypotheses about the differences. We test these hypotheses against our pre-existing corpus, which provides both human-authored American and British audio description for each of 207 movies. Results provide quantitative evidence to uphold all tested hypotheses, including differences in lexicon, the use of the progressive aspect, the use of passive constructions, the use of subjective adjectives and modifiers, when characters are named, how scenes are cued, and degree of overlap with movie dialogue and music. Our work offers a quantitative lens into the narrative technique of audio description.

Keywords

movies, audio description, guidelines, narrative, dialect

1. Introduction

Movies tell stories using a carefully curated mix of linguistic, auditory and visual information. Audiences extract meaning and narrative from all of this information, often to great enjoyment. However, being able to articulate that meaning-making process is a challenge.

This is a challenge tackled by accessibility services that cater to individuals for whom one or more modality is less available. The accessibility service that aims to cater to blind and visually impaired individuals is called audio description (AD). AD is a translation of the visual (and unclear aural) information into a verbal narration. The narration is constrained to fit in gaps between dialogue, and it is voiced by a neutral narrator.

The task requires an audio describer to begin with a full understanding of the narrative being told [1], which involves analysis of the full movie and potentially also a screenplay, dialogue subtitles, and the other available resources. They must then identify suitable gaps between dialogue and find ways to translate the visual part of the narrative into a verbal narrative. Every AD will be different, because these stages involve subjective interpretation and storytelling.

AD is arguably a literary and narrative art form [2]. There is debate about how much creative license should be afforded to audio describers. This debate spans a scale of strictly objective description of the visuals [3] to a full re-narrativisation of the movie that prioritises coherent storytelling [4]. For practical purposes, some parts of this debate are answered for particular regions by their AD guidelines. For instance, American guidelines [Annexe 4 of 5] stipulate that all subjective interpretation is unacceptable, while British guidelines [Annexe 6 of 5] state that when it reduces potential confusion, particular visuals can be subject to a certain amount of interpretation. Ultimately, however, all perception is interpretive and complete objectivity is simply impossible. Audio describers rely on their training, experience, and common sense in order to maximise the understanding and enjoyment (i.e., access) that they provide to their audience. Given the overarching goal of providing AD is achieving access,

CMN'26: Workshop on Computational Models of Narrative, June 8-10, 2026, Madrid, Spain.

✉ igor.sterner@ed.ac.uk (I. Sterner); alex@inf.ed.ac.uk (A. Lascarides); keller@inf.ed.ac.uk (F. Keller)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the extent to which American and British AD scripts differ in practice has never been quantitatively evaluated.

Consider the AD of the Trinity nuclear test scene in the 2023 movie *Oppenheimer*.¹

- (1) a. Kistiakowsky is open mouthed, staring through his goggles. In the car, Feynman screws up his eyes from the light. Oppenheimer takes off his goggles. An immense fireball billows into the night sky. Oppenheimer's eyelids flicker open. The explosion climbs higher and higher, white heat at its core darkening at the edges into an orange mushroom shape. Lawrence is standing outside the car with his hand on the door, watching the spreading glow on the horizon. Oppenheimer stares into the blackened orange clouds with sparks flying in their midst. On his mattress, Rabi shifts onto his side for a clearer view, keeping the tinted glass in front of his face. Other observers sit up to watch the still-blossoming explosion. Rabi starts to lower his glass. Oppenheimer watches the light fade to a red glimmer. [UK]
- b. Now, standing outside the car, Lawrence gapes as the bright line shines on his face, and Feynman squeezes his eyes shut. Teller's lips puff out with a sharp exhale as his expression remains unchanged. In the bunker, Oppenheimer removes his goggles, blinking rapidly and squinting to see the explosion growing into a column of fire that blooms high into the dark sky, taking on a mushroom shape. Seen from over Lawrence's shoulder, the explosion's brilliant light gives off an amber halo around its irregular edges, which, miles away, tinges Oppenheimer's awed face. The roiling flames billow over each other, black smoke starting to form within the inferno. Rabi finally rolls over and peers through his welder's glass at the explosion. The amber glow on Oppenheimer's face dims as the black smoke grows in the flames, which fold over on themselves. Rabi lowers the glass from his eyes as the blackness consumes the remaining flames, leaving behind a ghostly cloud of smoke in the dark. [US]

Both the British and American versions ((1-a) and (1-b), respectively) describe similar events in the scene, using language that is largely an objective description of the visuals. Nevertheless, the language is almost poetic and paints a vivid scene. There is some flexibility, especially in a short thread, as to when an action is described—e.g., compare the description of Lawrence outside a car watching the explosion (7th sentence in UK, 1st sentence in US). There is also some flexibility on which of the less salient actions to describe, as we can see from the choices to describe only one of Kistiakowsky's or Teller's expressions.

A part of the process of creating AD also relies on knowledge of who the audience will be. AD created in the United States will follow guidelines from that region; the same holds for AD created in other regions, in particular the United Kingdom. In our study, we will only consider English AD from these two regions, since they form the largest available data, but English AD is also created for other regions (e.g., Australia) and for many other languages. Each region will have dialectal and lexical preferences, and audio describers naturally take this into account in their decisions on how to express content.

Comparing the differences between versions of AD can offer a lens into AD as a narrative technique. Prior work has relied on theoretical analyses of guidelines [6] and anecdotal evidence [7]. They offer hypotheses as to the differences between American and British AD, but do not provide quantitative evidence in support of these hypotheses. In this paper, we aim to address this gap. For fair and controlled experiments, we need to compare AD of the same source material from each region. In separate work, we have collected a corpus of AD transcripts that makes this possible; see [8] for more details on the corpus and on the automatic generation of AD. For each of 207 movies, the corpus provides a professional American AD transcript and a professional British AD transcript.

We will begin by describing hypotheses about the differences between AD in the regions. These are based on lexical variation, aspect, voice, subjectivity, character naming, dialogue overlap and description of scene changes. The specific hypotheses we define are motivated by the AD guidelines and claims made by a leading practitioner in the creation of AD. We design and implement automatic methods that test these hypotheses, and apply them to our corpus. Results provide supporting evidence to uphold all of the hypotheses. We finish by discussing the implications of our work.

¹The video of this scene with American AD is available here: <https://adawardsgala.org/wp-content/uploads/2024/07/Oppenheimer.mp4>.

2. Hypotheses

We make hypotheses of the differences between American and British AD based on two sources. The first source is the AD guidelines from each region.² AD guidelines provide recommendations as to how audio describers should produce their descriptions. Our interpretation of the guidelines and the differences between them is based on Bittner's analysis [6]. The second source is Campbell's claims in [7]. Campbell works in a leading role in the creation of AD at Red Bee Media, a large broadcasting and media services company. Her claims represent the perspective of an AD practitioner.

For this section, we will provide examples from one movie, *Oppenheimer* (2023). Its American AD was a finalist in both the EGA Hermes Awards and Audio Description Awards Gala. The first unique aspect of this movie is that the officially released British AD is a post-edit of the American AD, which is uncommon (henceforth, we'll refer to this version as 'post-edit').³ The second unique aspect of this movie is that an entirely new UK AD version was created by the BBC for its broadcasting of the movie (henceforth, we'll refer to this version as 'UK').

Consider the following scene. *Oppenheimer* has been in the New Mexico desert with Lawrence and Frank. *Oppenheimer* says 'Come', and leads them off. When the **scene changes** to one later in the day in a tent, the AD versions are as follows.

- (2) a. Inside a tent, Lawrence drinks from a bottle and passes it to Frank. [UK]
- b. After nightfall, warm lantern light glows on the men's faces inside a tent. Lawrence passes a liquor bottle to Frank, who takes a swig. [US]
- c. ~~After nightfall~~ **In a tent at night**, warm lantern light glows on the men's faces ~~inside a tent~~. Lawrence passes a ~~liquor~~ bottle **of booze** to Frank, who takes a swig. [post-edit]

All three versions cue the scene-change with a fronted adverbial ('inside a tent'; 'after nightfall'; 'in a tent at night'). Both British and American AD guidelines state that scene changes should be cued, in particular when it is difficult to infer from the soundtrack. The American guidelines are more explicit in this recommendation, articulating that markers like 'now' can be used. In *Oppenheimer*, 'now' is used 16 times in the American version, almost all of which are removed in the post-edit (leaving 3), and also rarely used in the UK version (used 3 times). In contrast, the British guidelines do not provide any explicit suggestions as to how signalling a change in scene should be achieved. We can also see **lexical variation** in how the alcoholic beverage is described (UK and post-edit: 'a bottle' or 'a bottle of booze'; US: 'liquor bottle'). This difference is raised by Campbell, but not by any of the region-specific guidelines.

Consider now the following description, which occurs at the end of a scene when *Oppenheimer* is giving a lecture in Holland.

- (3) a. *Oppenheimer* is standing at the lectern. [UK]
- b. *Oppenheimer* stands at a lectern. [US]
- c. **At the front**, *Oppenheimer* stands at a lectern **looking slightly nervous**. [post-edit]

The British and American descriptions are near-identical, except for the fact that they differ in **aspect**. Both are in the present tense, but the British description employs the progressive aspect, while the American description is simple. Campbell claims that American descriptions employ almost exclusively the simple present tense, while in British descriptions there is the option to switch to the present progressive and simple present for variation. This claim is supported by the guidelines, which all state that the present tense should be used, with the guidelines from the UK and from Ireland additionally allowing progressive aspect to be used.

In the post-edited version of the American description, the describer has also chosen to incorporate a more detailed description of *Oppenheimer*'s body language ('looking slightly nervous'). Note that the

²Rai [5] appends a snapshot of guidelines at the time of their writing to their appendix (the American guidelines we use are in Annexe 4 and British guidelines in Annexe 6).

³This is a presumption, based on observing the data and both stating that they are created by the same company, Deluxe. We presume that the American AD version of *Oppenheimer* was provided as a starting point to a team that scripted and voiced the British AD version.

describer has chosen not to describe the visuals directly (e.g. Oppenheimer’s pursed lips or avoidant eyes). Instead the adjective ‘nervous’, additionally modified by the adverb ‘slightly’, is used to make a judgment about Oppenheimer’s emotions. In this case, this decision was presumably made due to the very short gap between dialogue into which this description has to fit. The British guidelines explicitly afford flexibility with regard to **subjective interpretation**, in particular when this reduces potential confusion. However, the American guidelines are explicit in discouraging this. Campbell acknowledges the assumption that British AD tends to do more work in interpreting emotions than American AD. However, she argues that it is impossible to be entirely objective and there is not a claim of any systematic difference.

Consider now a scene when Oppenheimer is walking down a dark street to an event and an FBI agent is taking notes of the license-plates of attendees.

- (4) a. Oppenheimer sees a torch being shone at cars parked on a suburban street. [UK]
 b. Oppenheimer notices someone with a flashlight checking out a car and taking a note. [US]
 c. Oppenheimer notices someone with **shining** a flashlight, ~~checking out torch by~~ a car and ~~taking a note.~~ **making notes.** [post-edit]

The British description employs the **passive voice**, while the American description stays in the **active voice**. Campbell argues that in British AD the passive voice is used as a tool to direct the audience’s focus, while American AD typically avoids it. The guidelines do not suggest any major difference in syntactic construction between American and British AD.

Consider now how AD **names characters**. In Oppenheimer, there are three board members in the hearing for his security clearance. We see the third board member, Ward Evans, at the back of the hearing from the beginning of the movie, but he is not named in the dialogue until towards the end of the movie (one of the other board members says ‘A full written opinion, with a dissent from Mr. Evans, will be issued to the AEC in the coming days.’). A full 8 minutes before this utterance occurs in the movie, the British description opts to name him (‘A smile from Evans, the oldest board member.’). In contrast, at this point in the movie, the American description does not name him (‘An elderly member of the board smiles.’). In contrast again, in the post-edit, the name is introduced (‘An older member of the board, Ward Evans, smiles.’). The American version names him for the first time directly after the dialogue that introduces him (‘Robb shakes hands with the chairman, the elderly Evans and the third board member.’). The guidelines differ in whether they permit characters to be named before the sighted audience finds out their name: while the American guidelines stipulate that the description should always wait to name the character after the sighted audience finds out their name, British guidelines state that it is also acceptable to name a character earlier, unless there is mystery attached to the name. Campbell argues that human describers use their own common sense to prioritise proportionate access to that of a sighted person.

The creation of an AD script also involves selecting **suitable moments to voice the description**. The British guidelines stipulate a strict rule to never voice description over dialogue. Meanwhile, the American guidelines are somewhat more permissive. As one example, when Oppenheimer is left behind in a lab and his classmates leave, one of them shouts back “Don’t forget to clean up!”. The American version voices description over this entirely, while the British version does not. The American description opts to talk about a green apple which soon becomes salient to the narrative.

In sum, claims of the differences between American and British audio description are as follows.

1. **American terms**: American AD uses a higher frequency of American English lexical items.
2. **British terms**: British AD uses a higher frequency of British English lexical items.
3. **Present progressive**: British AD uses the present progressive aspect more frequently.
4. **Subjectivity**: British AD exhibits greater subjectivity in interpreting visual information and character emotions.
5. **Passive**: British AD uses the passive voice more frequently.
6. **Naming in advance**: British AD names characters in advance more frequently.
7. **Dialogue overlap**: American AD voices over dialogue (and music) more frequently.

3. Methods

3.1. Materials

We now test these claims using our pre-existing corpus of AD [8]. For each of 207 movies, the corpus provides professional American and British AD, transcribed from their audio tracks with automatic methods.⁴ Automatic signal processing steps were taken to ensure that both tracks describe identical versions of the movie and to align them with each other. For every movie, the corpus also provides human-transcribed and automatically aligned dialogue subtitles.

There also exists an accessibility service designed to cater to individuals who are deaf and hard of hearing, which is often called SDH (Subtitles for the Deaf and Hard-of-Hearing). These subtitles contain salient non-verbal audio (e.g., Lawrence exhaling during the Trinity explosion scene), unclear speaker identification (e.g., that Oppenheimer is narrating immediately after), and text simplification when dialogue is spoken too quickly to read (from a different movie, ‘I just thought that maybe this was gonna be more effective.’ becomes ‘I just thought this would be more effective.’). All 207 movies are also provided with these subtitles. For each movie, the corpus additionally provides videos of a number of scenes released publicly for the movie. For 84 out of the 207 movies, the corpus also includes a screenplay aligned at the scene level.

All automatic tools used in the construction of this dataset were validated against a separate set of ten full movies, for which we gathered professional transcription, manually annotated scene boundaries and manually annotated screenplay alignment. Scene boundary annotation was performed first based on the movie visuals, and then the British and American AD corresponding to each scene was determined. For our experiments, we will only use the corpus American AD, British AD, and the dialogue subtitles.⁵ More details about the corpus and its construction are in [8].

3.2. Automatic Pre-processing

We process each AD and dialogue sentence with automatic NLP tools from the Stanza toolkit [9]. This provides syntactic parses in which the output features Universal Dependencies [10], Penn Treebank POS tags [11] and named entity recognition according to [12]. Details on how we use the parsed output of the different versions of AD to estimate the frequency of the linguistic features of note are given below.

3.3. Significance Testing

We test differences between American and British AD for statistical significance with paired, two-tailed permutation tests, approximated by $N = 10,000$ and with threshold $\alpha = 0.05$. The intuition for this statistical test is that if there was not a systematic difference between American and British AD, we could randomly swap the ‘American’ and ‘British’ labels for any given movie AD without changing the average results. The test estimates the probability that the observed difference between the two regions could have occurred due to random variation.

3.4. Statistics

Table 1 gives some statistics of the AD and subtitle portions of the corpus. Some differences between the British and American versions are already evident. On average, the AD of a single movie contains 589 and 833 sentences in the two versions, respectively. We observe significantly higher lexical diversity⁶ and mean dependency distance in British AD, as compared to American AD. Both are significantly lower than in the subtitles. The speech rate is statistically indistinguishable between the two AD

⁴There is an automatic step that supports the correct spelling of character names. Example (1) is derived from the same automatic transcription, which resulted in the correct spellings of the characters.

⁵The exception to this will be our analysis of scene changes, for which we will only use the ten manually processed movies.

⁶We used the length-controlled variant of type-token ratio from McKee [13].

Table 1

Corpus statistics. The subtitles are not provided with sentence segmentation, nor do we rely on such segmentation for any analysis on the subtitles in this paper. *This number is artificially c. 4% lower for some of the movies, due to a step in our alignment that corrects the different speed at which movies are shown in the two regions.

	UK	US	Subtitles
# movies	207	207	207
# tokens	1 173 935	1 460 267	2 121 054
# sentences	121 831	172 442	
# tokens per movie	5671 \pm 2123	7054 \pm 2399	10 247 \pm 3829
# sentences per movie	589 \pm 216	833 \pm 238	
Lexical diversity	89.2 \pm 10.4	86.1 \pm 9.9	117.0 \pm 14.6
Mean dependency distance	2.5 \pm 0.1	2.4 \pm 0.1	
Speech rate (tokens per minute)	196.3 \pm 13.8*	197.8 \pm 12.6	189.0 \pm 24.6
Number of character names per movie	340.8 \pm 135.8	449.0 \pm 161.2	215.3 \pm 101.7

versions (196 and 198 tokens per minute, respectively, $p = 0.21$). This is higher than the 160 words per minute suggested in the American guidelines. Finally, we observe that in AD, characters are mentioned significantly more than in dialogue ($p < 0.01$ for AD variants). This aligns with the statement in the British guidelines that names tend to be used more frequently in AD than in normal speech.

3.5. Experimental Tests

Lexical variation. We quantify lexical difference by measuring the frequency of American- and British-specific terminology. Lists of British and American terms are taken from online resources.⁷ There are 314 American terms and 677 British terms. Some of these terms are only more frequent in one region for a particular sense, and our subsequent analysis will be sense-independent. As one step to address this, we remove from the American list words also in the top 5,000 most frequent words in the British National Corpus [14]. Similarly, we remove from the British list the top 5,000 most frequent words in the Corpus of Contemporary American English [15]. This results in removing 3 and 14 words, respectively, leading to lists of length 311 and 663 words. Furthermore, for each movie we exclude all terms matching or contained in character names in the movies (for instance in one movie, a character is called ‘ladybug’, which is removed from the American list for that movie). We compute the frequency of these terms per sentence within the AD transcripts and report a macro-average over movies.

Aspect and voice. For differences in aspect, we identify instances of the present progressive by detecting when all of the following conditions are met in a sentence: (1) a gerund or present participle (2) has an auxiliary dependent form of ‘to be’ (3) which is in the present tense. We compute the proportion of such description sentences in each movie and report the macro-average. We identify sentences that use the passive voice by labelling sentences that contain any dependency relation marking a passive subject or passive auxiliary. We compute the proportion of description sentences containing passive constructions in each movie and report the macro-average.

Subjective judgments. For our automatic analysis of subjective judgments in AD, we focus our analysis on the use of adjectives and how they are modified. De Smedt [16] annotates English adjectives with percentage subjectivity scores and provides an algorithm for incorporating scores from adverbial modifiers. We apply their approach and determine the subjectivity of each description sentence and report a macro-average over movies.

Naming in advance. We identify whether AD introduces a character’s name before they are named in the dialogue. We use character lists from IMDb (modulo uncredited characters and enumerated

⁷American terms: https://en.wikipedia.org/wiki/Glossary_of_American_terms_not_widely_used_in_the_United_Kingdom, British terms: https://en.wikipedia.org/wiki/Glossary_of_British_terms_not_widely_used_in_the_United_States.

Table 2

Results based on a macro average over 207 movies (*only 184 movies; for the remainder, zero valid characters were detected for the analysis). In each row, the underlined result indicates the larger number of the two.

	Comparison	Unit	UK	US	Significance
(1)	American terms	(terms per 1,000 sentences)	5.0	<u>16.8</u>	$p < 0.01$
(2)	British terms	(terms per 1,000 sentences)	<u>15.8</u>	4.6	$p < 0.01$
(3)	Present progressive	(proportion of sentences, %)	<u>3.0</u>	0.3	$p < 0.01$
(4)	Subjectivity	(mean sentence score, %)	<u>16.2</u>	13.3	$p < 0.01$
(5)	Passive	(proportion of sentences, %)	<u>4.6</u>	1.3	$p < 0.01$
(6)	Naming in advance*	(proportion of characters, %)	<u>31.4</u>	9.0	$p < 0.01$
(7)	Dialogue overlap	(proportion of AD time, %)	8.5	<u>14.1</u>	$p < 0.01$

characters) and define a mapping based on all unique permuted variants of the name. The first named entity that maps to each character is then identified in each AD script and the dialogue subtitles. We keep occurrences that match both of the following conditions: (1) the character is identified as occurring in both AD versions earlier than 60 seconds after its occurrence in the dialogue subtitles, (2) at least one of the AD versions names the character between the time the character is named in the dialogue subtitles and 60 seconds later. After these conditions, there are three cases: (1) neither AD version names in advance, (2) the American version names in advance, (3) the British version names in advance. We compute counts of cases $(2)/(1)+(2)+(3)$ and $(3)/(1)+(2)+(3)$ and report a macro average over the movies with at least one detected character matching the conditions (184/207 movies).

Dialogue overlap and scene changes. Our measure of dialogue overlap is the proportion of duration of AD that overlaps with the dialogue subtitles. Dialogue subtitles also include music lyrics. For our experiments on how scene changes are articulated, we only use the separate set of ten movies with manual scene annotation. Based on this annotation, we train a BERT language model [17] to identify scene boundaries based on the AD text alone. The model takes as input six sentences, and a separator token is placed between sentences three and four. The training task is binary classification, with labels of whether the separator token represents a scene boundary or not. We perform cross-validation on the ten movies, and report results based on training on the American, British or both AD versions. For fair comparisons, we need to control the amount of data available to each trained model. Earlier we found that American AD is typically longer than British AD; training on both AD versions will always result in more data than training on only one. We address this by subsampling the data from each movie such that the number of positive and negative samples is equal to the number from the movie with the fewest samples. This equates to sampling 74 descriptions that start new scenes and 474 that do not.

4. Results and Discussion

Results of the application of the tests to the corpus are given in Table 2.

Lexical variation. Rows 1 and 2 provide supporting evidence of a difference in lexicon. There are an average of 16.8 American terms per 1,000 sentences of American description, significantly more than the 5.0 in British description sentences. The British terms are used on average 15.8 times per 1,000 sentences of British description, significantly more than the 4.6 times in American description sentences. These frequencies are still low, in part because our lists are not comprehensive and in part also demonstrating that lexical variation only plays a small role in differences between the regions (around 1% of sentences).

The most frequent American terms in American AD are ‘SUV’ (414 US counts vs. 134 UK counts; British equivalent might be ‘4x4’), ‘flashlight’ (337 US counts vs. 40 UK counts; ‘torch’), and ‘sidewalk’ (307 US counts vs. 19 UK counts; ‘pavement’). The most frequent British terms in British AD are ‘amongst’ (184 UK counts vs. 16 US; American equivalent might be ‘among’), ‘car park’ (152 UK counts

vs. 2 US; ‘parking lot’), and ‘windscreen’ (138 UK counts vs. 6 US; ‘windshield’). These are unsurprising results, and a clear demonstration of why AD needs to be localized for these two dialects of English.

Aspect and voice. Results provide clear evidence that the present progressive is rarely used in American description (0.3% of sentences) but used significantly more frequently in UK descriptions (3.0% of sentences). Passive constructions are used more than three times as frequently in British AD compared to American AD (4.6% vs. 1.3%, $p < 0.01$). Evidently these are both only tools in the toolkit of a British audio describer, albeit even then not a frequently used tool. Both the progressive aspect and the passive voice are often considered to be more complex syntax than simple aspect and active voice; algorithms that are designed to simplify English often change passive sentences to active ones. Meanwhile the American guidelines are more explicit than the British ones in their recommendation to use simple language — our results here can be seen as a consequence of this.

Subjective judgments. Based on this analysis, we find a significantly higher average subjectivity score in British description sentences compared to American sentences (16.2% vs. 13.4%, $p < 0.01$). For instance, the word ‘beautiful’ (which has a score of 100% subjective) occurs almost three times more frequently in British description than in American description (87 vs. 35 occurrences). Indeed the American guidelines directly address the use of this word, negatively stating that it is vague and evaluative.

Consider the following two uses of ‘beautiful’:

- (5) a. Glancing around the room, he turns back to find a beautiful woman glaring at him. [UK]
b. He looks away, then looks back and finds a dark-haired young woman standing before him. [US]
- (6) a. Bridget notices the church is mainly filled with beautiful young women. [UK]
b. Bridget notices many of the pews filled with fashionable young women. [US]

In example (5), the British description opts for a more detailed description of how the character scans the room (‘Glancing around the room’ vs. ‘He looks away’), and recovers from this use of narration time by describing the woman as ‘beautiful’ (vs. ‘dark-haired’ and ‘young’). In example (6) we can see that the two versions are near-identical, except for the choice to use ‘beautiful’ (British description) as compared one syllable longer ‘fashionable’ (American description).

Naming in advance. Results provide evidence that in British AD characters are consistently named in advance more frequently (31.4% vs. 9.0% of characters identified in our analysis, $p < 0.01$). When a character is named in advance in British AD, they are named on average 11 minutes early, compared to seven minutes in American AD. At the extremes, on one occasion a character is named in British AD over 100 minutes before being named in the dialogue, and on another occasion a character is named over 60 minutes early in American AD. Both cases concern supporting characters that appear only briefly in the opening quarter of the movie but are more fully developed in the final quarter of the movie. Our findings show that some American audio describers do opt to name in advance, even if this goes against their guidelines. Note that the cases under consideration here only concern when one, and not both, of the versions opted to name in advance. The 9.0% of characters named in advance in American AD thus represents occasions when the British describer did not name in advance. We suspect that this number represents roughly how often describers from the same region will differ in their choices to name in advance. This leads to the suggestion that in 22.4% of cases the British describer made a regionality-influenced decision to name in advance. In most cases (59.6%), both versions are naming the character immediately after they are named in the dialogue.

Dialogue overlap. Results show that 14.3% of American description time is overlapping with the subtitles, compared to the significantly lower 8.4% of British description time. This supports the claim that British AD is stricter about not overlapping with character dialogue and music. This result is, however, unsurprising given our earlier statistic that British AD uses on average 29.3% fewer sentences.

Table 3

Cross-validation results based on training a model to predict movie scene boundaries from AD transcripts.

	UK			US		
	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>P</i>	<i>R</i>	<i>F</i> ₁
Trained on UK	62.6	55.4	58.4	61.9	72.4	65.8
Trained on US	73.6	42.9	53.7	70.6	66.0	67.0
Trained on US+UK	66.1	55.1	59.6	65.1	69.0	65.9

Scene changes. Results are shown in Table 3. We can see that recall is significantly lower for the task of identifying scene boundaries from British AD transcripts, as compared to from American AD transcripts (trained on UK: 55.4% vs. 72.4%, $p = 0.003$; trained on US: 42.9% vs. 66.0%, $p = 0.003$; trained on UK+US: 55.1% vs. 69.0%, $p = 0.013$). Training only on American transcripts and testing on British transcripts leads to the worst performance ($F_1 = 53.7\%$, differences to other F_1 scores all have $p \leq 0.05$). This suggests that this model learnt to rely on cues (most likely explicit cues like ‘now’) in American transcripts that are not present as frequently in the British transcripts. In sum, these results provide supporting evidence that scene boundaries are verbalized less explicitly in British AD than in American AD.

5. Conclusion

AD is a narrative technique. We have shown at scale the existence of small but statistically distinguishable differences between how the narrative is verbalized in American vs. British descriptions. All of the differences are justified based on differences between the regions’ guidelines, except for lexical variation, which has not received prior theoretical treatment.

1. American AD uses American terms roughly three times more frequently than British terms; the same holds vice versa.
2. American AD rarely uses the progressive aspect or passive voice, while British AD uses these in 3% and 4% of sentences, respectively.
3. Both American and British AD use subjective adjectives and modifiers, but American AD uses them more sparsely.
4. Of characters named in at least one version immediately after being named in the dialogue, American AD names that character in advance one third of the frequency in British AD.
5. Both American and British AD, on occasion, narrate over dialogue, but American AD overlaps more frequently (and is also longer in general).

Our work contributes to a better understanding of the creation of AD. It can serve to help practitioners better analyse their own habits and encourage the field to make more active decisions in guidelines. AD also serves as an attractive domain for research in narratology, because it is a highly constrained form of storytelling that makes the underlying processes of narrative construction in movies explicit. Using our corpus of British and American AD, we have here provided empirical evidence that AD is constrained both by dialect differences and by guidelines.

Acknowledgements

IS is funded by an EPSRC PhD studentship (project reference 2923920). We thank Nina Gregorio and Marie Campbell for comments.

References

- [1] G. Vercauteren, A narratological approach to content selection in audio description, Ph.D. thesis, Antwerp University, Belgium, 2016. URL: <https://repository.uantwerpen.be/docman/irua/4a8d3c/11347.pdf>.
- [2] J. Snyder, Audio description: The visual made verbal, in: *The didactics of audiovisual translation*, John Benjamins, 2008. doi:10.1075/btl.77.18sny.
- [3] Audio Description Coalition, Standards for Audio Description and Code of Professional Conduct for Describers based on the training and experience of audio describers and trainers from across the United States, 3rd ed., 2009. URL: https://audiodescriptionsolutions.com/wp-content/uploads/2020/04/ad_standards_090615.pdf.
- [4] J.-L. Kruger, Audio narration: re-narrativising film, *Perspectives* (2010). doi:10.1080/0907676X.2010.485686.
- [5] S. Rai, J. Greening, L. Petré, A comparative study of audio description guidelines prevalent in different countries, Londra: Royal National Institute of Blind People (2010). URL: <https://unidescription.org/storage/app/uploads/public/5f1/a3e/bb1/5f1a3ebb17896460620035.pdf>.
- [6] H. Bittner, Audio description guidelines: A comparison, *New Perspectives in Translation* 20 (2012) 41–61. URL: https://www.uni-hildesheim.de/media/_migrated/content_uploads/AD_Guidelines_Comparison_-_Read.pdf.
- [7] M. Campbell, Crossing the pond: A deep dive into UK and US audio description, Red Bee Media Blog, 2025. URL: <https://www.redbeemedia.com/blog/crossing-the-pond-a-deep-dive-into-uk-and-us-audio-description/>.
- [8] I. Sterner, M. Lapata, A. Lascarides, F. Keller, REFRAMED: Towards realistic audio description generation for movies, 2026. In submission.
- [9] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A python natural language processing toolkit for many human languages, in: A. Celikyilmaz, T.-H. Wen (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, Online, 2020, pp. 101–108. URL: <https://aclanthology.org/2020.acl-demos.14/>. doi:10.18653/v1/2020.acl-demos.14.
- [10] M.-C. de Marneffe, C. D. Manning, J. Nivre, D. Zeman, Universal dependencies, *Computational Linguistics* 47 (2021) 255–308. URL: https://doi.org/10.1162/coli_a_00402. doi:10.1162/coli_a_00402.
- [11] M. P. Marcus, B. Santorini, M. A. Marcinkiewicz, Building a large annotated corpus of English: The Penn Treebank, *Computational Linguistics* 19 (1993) 313–330. URL: <https://aclanthology.org/J93-2004/>.
- [12] R. Weischedel, M. Palmer, M. Marcus, E. Hovy, S. Pradhan, L. Ramshaw, N. Xue, A. Taylor, J. Kaufman, M. Franchini, M. El-Bachouti, R. Belvin, A. Houston, OntoNotes release 5.0, 2013. doi:10.35111/xmhb-2b84.
- [13] G. McKee, D. Malvern, B. Richards, Measuring vocabulary diversity using dedicated software, *Literary and linguistic computing* 15 (2000) 323–338.
- [14] G. Leech, P. Rayson, et al., *Word frequencies in written and spoken English: Based on the British National Corpus*, Routledge, 2014.
- [15] M. Davies, *The corpus of contemporary American english (COCA), 2008–.* URL: <https://www.english-corpora.org/coca/>.
- [16] T. De Smedt, W. Daelemans, “vreselijk mooi!” (terribly beautiful): A subjectivity lexicon for Dutch adjectives., in: N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 3568–3572. URL: <https://aclanthology.org/L12-1145/>.
- [17] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:*

Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: <https://aclanthology.org/N19-1423/>. doi:10.18653/v1/N19-1423.