Effects of Context on the Use of Descriptive Verbs

Radina Dobreva (r.dobreva@ed.ac.uk)

Frank Keller (keller@inf.ed.ac.uk)

Alexandra Birch (a.birch@ed.ac.uk)

School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, UK

Abstract

Action descriptions can include or omit various types of information. In this paper, we are interested in the inclusion of manner in verbs. We use the concept of descriptive verbs, first introduced by Snell-Hornby (1983), and hypothesise that the use of descriptive verbs is reliant on having enough context to determine if the descriptive verb is correct and preferred as opposed to a more general non-descriptive verb. We conduct two online experiments in which participants are asked to indicate their preference for a verb after seeing varying amounts of textual and visual context. Our results show that textual context does not contribute to verb choice. However, we find evidence that videos contain information which creates agreement between participants, suggesting there are objective reasons to choose a descriptive or non-descriptive verb.

Keywords: verb descriptivity, manner, context, audio description

Introduction

Descriptions of actions and events can vary greatly, depending on what describers choose to include or omit. One source of variability is manner – we can choose to include more or less information about the way in which an action is performed. In the domain of motion verbs specifically, languages have been shown to differ in where and how much they express manner (Talmy, 2000; Slobin, 2006). English is a language of high manner salience (Slobin, 2006). It has been shown that in English manner tends to be encoded within the main verb of a clause, there is a large vocabulary of manner verbs and English speakers often include manner information in their descriptions of events (Matsumoto, 2003; Slobin, 2006; Papafragou, Massey, & Gleitman, 2006; Akita & Matsumoto, 2020). However, most experimental work studying manner of motion has focused on finding cross-linguistic differences in how and to what extent languages encode manner. There has been less attention on nuances of what manner information is expressed within specific languages (but see Akita and Matsumoto (2020) and Slobin, Ibarretxe-Antuñano, Kopecka, and Majid (2014)).

While manner is a heterogeneous concept that can encode many different types of information, one particularly relevant dimension here is specificity. That is, manner verbs differ in how specific the manner included is (Snell-Hornby, 1983; Slobin, 2000; Slobin et al., 2014; Akita & Matsumoto, 2020). In addition to less specific manner verbs such as "run" and "walk", English has a large vocabulary of more specific manner verbs such as "dash" and "stroll". In this study, we use a particular account of specificity by Snell-Hornby (1983), who first introduced the notion of descriptive verbs. These are verbs which express a core action and a modificant conveying something about the manner or circumstances around the action. In the rest of this paper, we use *descriptive verbs* to mean verbs which encode specific manner and *non-descriptive* verbs to refer to more basic categories of actions. For example, "chuckle" can be defined as "to laugh quietly". It is a descriptive verb with "laugh" as the core action and what we take to be its corresponding non-descriptive verb.

Our study deals with the question of what motivates the use of descriptive verbs vs. non-descriptive verbs. We hypothesise that the physical action itself does not always contain all necessary information and describers need a wider context to determine if a descriptive verb is correct or not. To study this, we use a genre of descriptions which have a well-defined context, namely movie audio descriptions. Audio descriptions in movies describe the events on screen to provide a complete experience of the movie, primarily aimed at increasing accessibility for blind and visually-impaired people. We conduct two online studies asking participants to indicate their preference for descriptive and non-descriptive verbs after seeing varying amounts of context. To our knowledge, no other experimental study has addressed the choice of descriptive over non-descriptive verbs. Our work is the first to attempt to discover what factors guide the lexical choice of verbs with respect to descriptivity. We find that on the whole people agree whether a descriptive or non-descriptive verb is more appropriate in each case. However, wider context does not seem to affect preferences and the video clip provides sufficient information.

Background

Verb manner and descriptivity

Talmy's (2000) influential work identified a typological distinction between what he calls satellite-framed and verb-framed languages. This distinction is based on where in a clause languages express path of motion – whether in the main verb (verb-framed) or in a separate structure (satellite-framed). English is a satellite-framed language which tends to encode path outside the main verb. For instance, in the sentence "Sally walked into the room." path is expressed using the preposition "into" and not the main verb. This leaves the

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

verb slot open for other types of information such as manner which in this example is expressed in the verb "walk". Further work by Slobin has looked at languages' "manner salience" or how much speakers of different languages include manner information in event descriptions (Slobin, 2000, 2006; Slobin et al., 2014).

Most existing experimental work in these directions has explored Talmy's typology and manner salience in terms of whether or not manner is specified at all and how it is included in the sentence. It has been shown that English speakers express manner very often in their descriptions of events (Papafragou et al., 2006). Manner is mostly expressed within the main verb although further modification using adverbs also occurs (Slobin et al., 2014; Akita & Matsumoto, 2020). English also has a large variety of manner of motion verbs, compared to other languages (Slobin et al., 2014). For German, which is typologically similar to English in this respect, it has been shown that there is within-language variation for the expression of manner, specifically between caused- and self-motion events, which suggests there may be systematic conditions under which manner is more or less likely to be included (Lewandowski, 2021). These results motivate the further study of manner within specific languages and the use of different types of manner verbs in more detail.

Beyond verbs of motion, Snell-Hornby (1983) defines the notion of descriptive verbs, which also express manner. Descriptive verbs contain an element of modification of the action they express which can also be conveyed using adjectives or adverbs. For example the verb "strut" is defined as "walk with pompous or affected stiff erect gait" (Snell-Hornby, 1983, p.25). The meaning of this verb includes the core action "walk" as well as two types of modification – physical ("stiff erect gait") and the evaluative, subjective judgement of the speaker ("pompous"). Descriptive verbs are manner verbs, but not all manner verbs are descriptive – for instance "walk" is a manner verb but is not descriptive.

Some important aspects of descriptive verbs are their focus and that they often express subjective evaluation. Snell-Hornby (1983) argues that the focus of descriptive verbs falls within the modificant rather than the core action of the verb. For example the use of the verb "dawdle" brings attention to the slowness rather than the precise nature of the action. Some descriptive verbs also express the speaker's evaluation or experience of an event and the involved participants. As the modificants of descriptive verbs are often adjectives and adverbs which can be understood by a subjective norm (e.g. something is judged as excessively slow or as breaking a social norm), a speaker using a descriptive verb expresses their subjective judgement and attitude towards the event. This characterisation of descriptive verbs raises interesting questions about the patterns of their use. Descriptive verbs contain extra information beyond the physical realisation of an action, and the expression of that information depends on subjective judgement, social norm and additional knowledge about the event being described. One question therefore is how people gain that information which would license the use of the descriptive verb as opposed to a non-descriptive alternative. In our study we use audio described scenes from movies to address this question.

Audio description

Audio description (AD) is an assistive service primarily intended for blind or visually impaired audiences which describes visual elements which are not otherwise available to them. AD can be found for example in movies, museum exhibits or live theatre performances. The data for our experiments comes from AD for movies. AD for movies describes the events happening on screen in the pauses between dialogue and sound effects and aims to recreate and preserve the experience of the movie as much as possible. Thus, individual descriptions cannot be seen as a standalone description of a video clip, but must be understood within the wider context of the movie.

Several studies have explored the contents and use of language in audio description. Some of these (Turner, 1998; Piety, 2004; Hurtado & Gallego, 2013) aim to categorise AD based on the content of the description (e.g. character description, setting, action description). Other studies (Salway, 2007; Tomadaki, 2006; Reviers, 2018) utilise corpus linguistic methods and qualitative analysis to explore lexical properties of audio description (e.g. proportion of open class words, verbs, nouns, types of verbs based on what kind of action they describe). Verbs are very common in AD as they represent actions and events which need to be described. Furthermore, due to time limitations in a movie, describers need to select expressive and concise vocabulary items. Descriptive verbs by their nature have the potential to be useful in audio description, as they convey subtleties which can evoke the impressions and interpretations suggested by the movie. However, this raises questions about the connection between the video material and the descriptions: is the choice of verb evident from the video alone, or can it only be understood within the wider context of the scene or movie? We expect that the evaluative judgements and subjective interpretations conveyed by descriptive verbs somehow arise from the context of the movie surrounding the particular action. The main objective of this paper is to test whether this is the case.

Experimental design

Task

In both experiments participants were asked to indicate their preference for one of two verbs, one non-descriptive and the other descriptive, after viewing a varying amount and type of context. The types of context were a short video clip containing the action, the current sentence AD containing the verb of interest and preceding context in the form of previous AD and dialogue within a minute of the current AD. See Figure 1 for an example of the different types of textual context. In Experiment 1, we always included the video and varied the textual context between conditions. In Experiment 2 we only used

Preceding context	A wave knocks them over, and rushing water carries them away. Water pours around them as lights flicker. Jack struggles against the current toward a stairwell. - This way! They pull themselves along a wall toward the stairwell. - Give me your hand! Jack pushes Rose through the shoulder-high water toward the railing. Rose grabs the railings and climbs to another locked gate - Oh, God! Jack shakes the gate as the water rises below them. - Help! The water reaches their feet.
Current AD	A steward [runs] past them.
Verb options	runs, sprints

Figure 1: An example of the textual context for a stimulus. The lines marked with "-" in the preceding context denote dialogue. The original verb in the current AD is in brackets and was replaced with "..." in the experiment.

the text-based context types. Otherwise the stimuli and setup were identical between the two experiments. The indication of preference was done using a five-point Likert scale ranging between the two verb options, with the labels "Strongly prefer [non-descriptive]", "Prefer [non-descriptive]", "Both are equally preferred", "Prefer [descriptive]", "Strongly prefer [descriptive]".

Hypothesis

Our main hypothesis is that the presence of textual context will influence the choice of non-descriptive or descriptive verb, where the original is descriptive. In other words, participants will be more likely to choose a descriptive verb when they are given textual context and not only video. We do not expect to see the same effect for examples where the original verb is non-descriptive. This is because we assume that the most informative verb was used by the original describer and there is no reason to use a descriptive verb in those examples. We also assume our participants to be cooperative and to follow Grice's principles of informativity (Grice, 1975).

We are also interested to see if people on the whole choose the verb from the original description, as this would confirm that there is some agreement between speakers about the situations where each is appropriate. Alternatively, there may be a bias towards either non-descriptive or descriptive verbs, for example due to the experiment setup or just a general preference for one or the other.

Materials

Verb list We selected 15 verbs covering a variety of actions. The verbs were selected in triples with one non-descriptive verb and two corresponding descriptive verbs: (1) run, jog, sprint; (2) walk, stroll, stride; (3) look, stare, glance; (4) smile, beam, grin; (5) hold, clutch, grip. Most of the descriptive verbs can be found in (Snell-Hornby, 1983) and their dictionary definitions include a reference to their non-descriptive counterpart.

Dataset The data for the stimuli comes from the LSMDC dataset (Rohrbach et al., 2017). LSMDC is a freely available dataset initially intended for automatic generation of audio description. It consists of short video clips (a few seconds each) and corresponding audio descriptions transcribed from the original audio tracks composed by professional audio describers. AD was aligned to the video manually to reflect the described event more closely, rather than using the timestamp of the original AD which may come before or after the described event depending on how much time there is between the dialogue and sound effects. We chose this dataset because of its availability and quality relative to other datasets, which have not been manually checked. Of crucial importance is the alignment between AD and video. There are, however, several caveats. Firstly, the manual alignment depends on human judgement and it may not always be accurate. Second, LSMDC only releases short video clips of a few seconds, so we are limited to using only a few seconds of video context. Third, no dialogue is provided with LSMDC, so we had to manually add this to our experiment (described below).

Data selection The data for the experiment was selected based on two broad factors: the textual context and the video content. To prepare the textual context, the following steps were taken:

- Using spaCy, we found instances of each verb in its third person singular form and removed instances that were preceded or followed by an adverb.
- Out of these, 100 were selected randomly for the non-descriptive verbs and 50 for the descriptive verbs (or fewer where the total number of occurrences was less than 50).
- Then instances were manually filtered to remove those which had phrases modifying the verb, contained errors, where the verb had a different sense or there was more than one verb from the group in the same sentence.

- Finally, 20/10 examples were selected for the non-descriptive and descriptive verbs, mostly at random, but aiming to have different surrounding structure (e.g. prepositions) and avoiding very long sentences.

After that, examples were further filtered using the video, excluding any where the action under question was not easily visible, if the context was in some way malformed (e.g. missing character names) or if the AD was at or just after the start of a scene (based on temporal keywords such as "later", "now", etc.). After the filtering, 5 examples per verb were selected at random. The context for each example was then collected (the preceding ADs within a one minute window and the corresponding dialogue). Some examples turned out to have no preceding AD context within that time frame, so those were manually substituted with other appropriate examples.

Dialogue alignment Since dialogue was not available with the original LSMDC dataset, we added it manually using subtitles from OpenSubtitles and audio recordings from AudioVault. For each movie and instance in our trial list, we downloaded the corresponding subtitle file (the one with the most downloads from OpenSubtitles) and the audio track containing audio descriptions. For some instances in our list the audio track on AudioVault did not correspond to the descriptions from the LSMDC dataset. Those instances were excluded and substituted with other ones for which the audio track is available and appropriate. We also excluded any movies which had an 18+ rating. We then annotated each instance to contain the dialogue as well as previous AD. The process was as follows:

- Locate the first context AD and current AD in the audio track and record their timing; also record the timing of the first instance of dialogue in the minute preceding the current AD.
- Listen for all instances of dialogue within the minute preceding the current AD, then copy them with timestamps from the subtitles file.
- Re-order the ADs and subtitles to reflect the true ordering from the audio file.
- Format dialogue to reflect speech said by different characters (i.e. putting consecutive utterances by the same character together and splitting lines containing utterances from different characters).

Technical details and participant recruitment

The experiment was created using jsPsych v7.3 (de Leeuw, Gilbert, & Luchterhandt, 2023). Participant recruitment was through Prolific. For both experiments we recruited participants who had listed both their primary language and their first language as English. The experiments were set up with a between-subjects design with each participant only seeing one condition and a small subset of the total data. Stimuli were presented in pre-defined groups of 10. Each participant was also shown a practice trial and two extra stimuli in the middle and end as attention checks, totalling 13 stimuli. In Experiment 1 participants were also given a brief physical description of the character whose action is the focus of each trial. Statistical analysis was done using R (R Core Team, 2020) and the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

Experiment 1

Setup

In this experiment we aimed to find out if extra context surrounding the action in the video has an effect on the choice of verb. Therefore, we had the following 4 conditions: (1) video only; (2) video and current AD (without the verb); (3) video and preceding context; (4) video, current AD and preceding context. In addition, in each trial we asked participants to indicate whether they have seen the movie that the clip is from. We had a total of 261 participants. Participants were filtered using two attention check questions. For conditions with preceding context, we also filtered out all participants who spent less than 5 seconds viewing the text screen. Some participant data was discarded due to technical issues such as restarting the experiment or incomplete responses. Overall, 51 participants were discarded, resulting in data from 210 participants remaining available for analysis. Each group of stimuli in each condition had at least 5 participant responses post filtering.

Results and analysis

Figure 2 shows the distribution of responses grouped by whether the original verb was descriptive or not and by condition. As we can see from the figure, there is no big difference between the four conditions for either descriptive or non-descriptive verbs. There is a noticeable difference in score distribution between descriptive and non-descriptive verbs, indicating that participant preference largely corresponds to the verb from the original AD, i.e. where the original is non-descriptive, more participants preferred non-descriptive, and where it is descriptive, more participants preferred descriptive. It is also interesting to note that the non-descriptive verb was more often acceptable when the original is descriptive than the other way around. Finally, participants overwhelmingly had a preference towards one of the options – the middle option ("Both are okay") was not used very often.

We fitted a linear mixed effects model with fixed effects for current AD, preceding context, whether the participant has seen the movie for a particular clip and verb type, and interactions between current AD and verb type, preceding context and verb type, and whether the movie has been seen and verb type:

score $\sim$ *(current AD + preceding context + seen movie) * verb type + (verb type|participant ID) + (1|video ID)*

We chose this model since we hypothesise an effect of context which differs between originally descriptive and non-

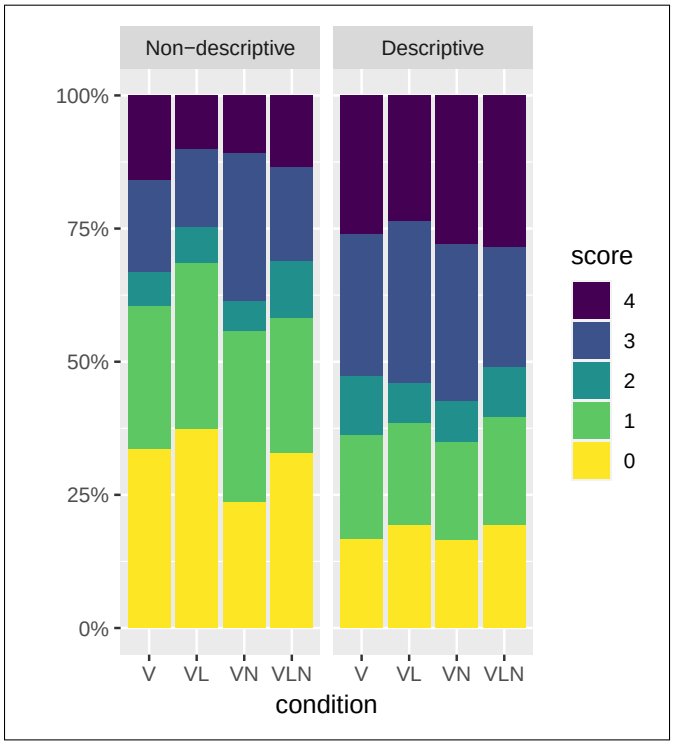

Figure 2: The percentage of each score given, grouped by whether the verb in the original description was descriptive or not and shown for each condition. Condition labels: V – video only, VL – video and current AD, VN – video and preceding context, VLN – video, current AD and preceding context. Score labels: 0 – strongly prefers non-descriptive verb, 1 – prefers non-descriptive verb, 2 – both are okay, 3 – prefers descriptive verb, 4 – strongly prefers descriptive verb

descriptive verbs. However, we do not expect an interaction between the different types of context. We found that verb type had a significant effect on score ($\beta = -0.75$, SE = 0.2, $p < 0.001$), which confirms our earlier observation that participants tend to agree with the original verb. However, neither current AD nor preceding context had a significant effect. There were also no significant interactions between verb type and context. Whether the participant had seen the movie also does not seem to affect score significantly.

Discussion

Overall, we found no evidence to suggest that the presence of context affects the choice of descriptive verbs more than non-descriptive verbs. Therefore, we have no support for our hypothesis that descriptive verbs will be chosen more often with context than without. In fact, we do not find any context effect whether for descriptive or non-descriptive verbs.

We do find that overall participants tend to choose the same verb as in the original AD, which suggests that there are objective reasons for the selection, rather than simply personal preference or style. From Figure 2, we also see that there is a more even spread of preferences in the descriptive group than the non-descriptive group. Our assumption when collecting the data was that the describers who created the original AD used the most informative verb available. However, there may be other factors influencing the choice of descriptive or non-descriptive verb. We do not know whether there are videos in our data where a descriptive verb was available, but the original describer chose to use a non-descriptive one. This would explain why participants still chose descriptive verbs in the non-descriptive condition. On the other hand, where the original describer chose a descriptive verb, its non-descriptive counterpart is still true, which could explain why we see a more even spread in that condition.

Several factors may contribute to the results we observe. Firstly, while efforts were made to control for reading time in the conditions with preceding context, we cannot be sure that the participants really read and engaged with the text. Second, the specific context that we included may not be sufficiently informative. We only included one minute of previous context, which in some cases simply described a previous scene or parts of the current scene. It could be that a lot of information is lost when transcribing to text or some information is learned throughout the plot of the movie (e.g. characters' personalities). It appears that the information in the video is already sufficient to create a bias towards the correct verb and textual context does not add any information. In the next experiment we further test if this is the case.

Experiment 2

Setup

Following the results from Experiment 1, we considered that the absence of context effect may be due to the video already containing sufficient information to make a choice and wanted to find out whether the context may still have an effect without the video. We conducted a follow-up experiment in which participants were not shown the videos and were instead given only the current AD sentence or the preceding context with the current AD sentence.

As in Experiment 1, participants were filtered using two attention check questions and only data from participants who spent at least 5 seconds viewing the text in the condition with preceding context were used. Again, some participant data were discarded due to technical issues. The total number of responses was 112 of which the data from 100 participants remained after filtering. Once again, each group of stimuli in each condition had at least 5 participant responses.

Results and analysis

In Figure 3 we compare the results from this experiment with two of the conditions in Experiment 1 (video + current AD and video + current AD + preceding context). We can see from the figure that for the two conditions without video, there is not a strong preference for either the descriptive or non-descriptive verb. The scores in these conditions are more evenly distributed than in the conditions with video where we see that there is stronger preference for the original verb

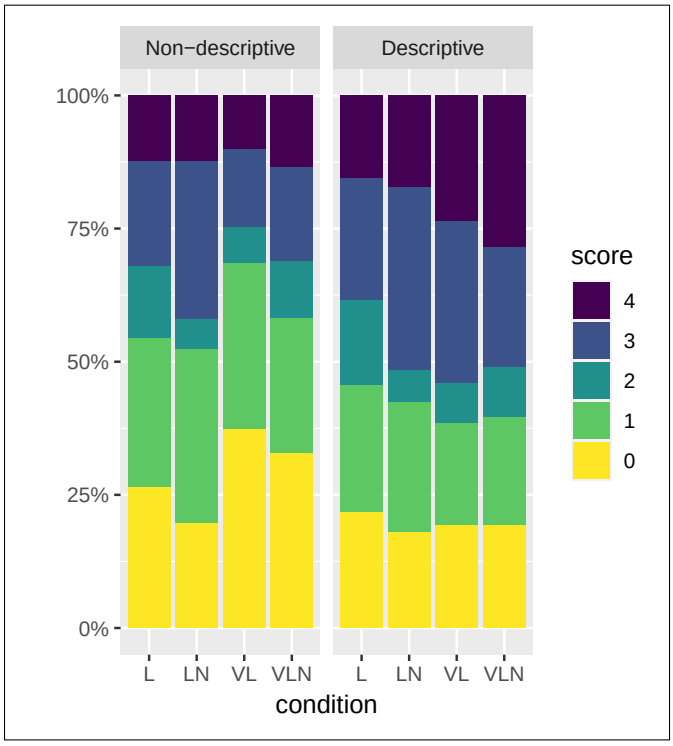

Figure 3: The percentage of each score given, grouped by whether the verb in the original description was descriptive or not and shown for each condition. Condition labels: L – current AD only, LN – current AD and preceding context, VL – video and current AD, VLN – video, current AD and preceding context. Score labels: 0 – strongly prefers non-descriptive verb, 1 – prefers non-descriptive verb, 2 – both are okay, 3 – prefers descriptive verb, 4 – strongly prefers descriptive verb

type. Additionally, in the condition with only current AD, the "Both are equally preferred" option is chosen more than in other conditions. These results provide additional evidence against our original hypothesis, showing that textual context does not seem to have much of an influence on participants' choices. The presence of the video makes participants more certain about either the descriptive or non-descriptive option and they tend to choose the verb in agreement with the original description.

To test the significance of these findings, we fitted a linear mixed effects model on the data from this experiment and a subset of the data from Experiment 1 which includes the two conditions above (video + current AD and video + current AD + preceding context). Our model had the presence of a video, preceding context and verb type as fixed effects, with all interactions between them:

score $\sim$ *video * preceding context * verb type + (1|participant ID) + (1|video ID)*

We find a significant effect of video ($\beta = 0.33$, SE = 0.13, $p < 0.05$) which suggests that the presence of the video affects participants' choices compared to only textual context. We also find a significant interaction between video and verb type ($\beta = -0.68$, SE = 0.16, $p < 0.001$) which suggests that the effect of video is not the same between descriptive and non-descriptive verbs. It seems that the presence of the video might make participants more likely to choose the original verb. Interestingly, the effect of verb type on its own is not significant which may suggest that without the video participants do not make significantly different choices in the two groups of stimuli. As in Experiment 1, there is no significant effect of preceding context.

Discussion

The results from this experiment confirm our suspicion that the video already contains a strong enough signal for participants to overall prefer the verb from the original description. Additionally, there is no evidence that the textual context contains any of that necessary information.

These findings further contradict our original hypothesis which was that descriptive verbs cannot be reliably chosen without wider context. One reason for this may be the type of data used in our experiments. We use clips from movies, which are carefully designed to create strong impressions on viewers. Therefore, it would not be surprising if they simply contain all the necessary information to determine which verb is more appropriate. Another reason could be, as mentioned previously, that the provision of context in text form is not informative enough and using audio or a longer video sample would have a stronger effect. Using the text format also means we have no guarantee that participants fully read the given context.

Conclusion

Overall, our findings suggest that at least in this genre of video and descriptions, video clips alone are sufficient and more informative than textual context for people to choose a descriptive or non-descriptive verb. Our study shows that there is a general agreement on which option should be used, which suggests that there may be objective reasons for that choice, strengthening the distinction drawn by Snell-Hornby (1983). However, across all conditions there was still a lot of variation in preference between descriptive and non-descriptive verbs. While on the whole participants agreed with the original description, the alternative was also chosen. It could be that some of the videos in the non-descriptive category warranted the use of a descriptive verb, but the original describer chose not to use that. Future work may explore what pragmatic factors determine which verb should be used in each scenario. It would also be useful to use more strictly controlled stimuli to confirm these findings. Additionally, previous work has shown that in free production non-descriptive verbs are often favoured. It would be interesting to find out if this holds for our type of data.

Acknowledgements

This work was supported in part by the UKRI Centre for Doctoral Training in Natural Language Processing, funded by the UKRI (grant EP/S022481/1) and the University of Edinburgh, School of Informatics and School of Philosophy, Psychology & Language Sciences.

References

Akita, K., & Matsumoto, Y. (2020). A fine-grained analysis of manner salience: Experimental evidence from japanese and english. In Y. Matsumoto & K. Kawachi (Eds.), *Broader perspectives on motion event descriptions* (p. 143-180).

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *Journal of Open Source Software*, *8*(85), 5351.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3: Speech acts* (p. 41-58). New York: Academic Press.

Hurtado, C. J., & Gallego, S. S. (2013). Multimodality, translation and accessibility: a corpus-based study of audio description. *Perspectives: Studies in Translation Theory and Practice*, *21*(4), 577-594.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.

Lewandowski, W. (2021). Variable motion event encoding within languages and language types: a usage-based perspective. *Language and Cognition*, *13*(1), 34–65.

Matsumoto, Y. (2003, 01). Typologies of lexicalization patterns and event integration: Clarifications and reformulations. In S. Chiba (Ed.), *Empirical and theoretical investigations into language: a festschrift for masaru kajita* (p. 403-418).

Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, *98*(3), B75-B87.

Piety, P. J. (2004). The Language System of Audio Description: An Investigation as a Discursive Process. *Journal of Visual Impairment & Blindness*, *98*(8), 453-469.

R Core Team. (2020). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Reviers, N. (2018). Studying the language of Dutch audio description: An example of a corpus-based analysis. *Translation and Translanguaging in Multilingual Contexts*, *4*(1), 178-202.

Rohrbach, A., Torabi, A., Rohrbach, M., Tandon, N., Pal, C., Larochelle, H., ... Schiele, B. (2017). Movie Description. , *123*(1), 94–120.

Salway, A. (2007). A Corpus-based Analysis of Audio Description. *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*.

Slobin, D. (2000). Verbalized events: A dynamic approach to linguistic relativity and determinism. In S. Niemeier & R. Dirven (Eds.), *Evidence for linguistic relativity* (p. 107-138).

Slobin, D. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (p. 59-81).

Slobin, D., Ibarretxe-Antuñano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: a crosslinguistic event-naming study. *Cognitive Linguistics*, *25*(4), 701–741.

Snell-Hornby, M. (1983). *Verb-descriptivity in German and English: A Contrastive Study in Semantic Fields*. C. Winter Universitätsverlag.

Talmy, L. (2000). *Toward a Cognitive Semantics. Volume 2: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.

Tomadaki, E. (2006). *Cross-document coreference between different types of collateral texts for films.* Unpublished doctoral dissertation, University of Surrey (United Kingdom).

Turner, J. M. (1998). Some Characteristics of Audio Description and the Corresponding Moving Image. In *Proceedings of the ASIS Annual Meeting* (Vol. 35, pp. 108–17).