University of Birmingham

Sunken ships and screaming banshees

Fuoli, Matteo; Littlemore, Jeannette; Turner, Sarah

[Link to publication on Research at Birmingham portal](#)

Sunken Ships and *Screaming Banshees*: Metaphor and evaluation in film reviews

MATTEO FUOLI[1], JEANNETTE LITTLEMORE[1], SARAH TURNER[2]

[1]University of Birmingham, [2]Coventry University

Abstract

It has been suggested that metaphor often performs some sort of evaluative function. However, there have been few empirical studies addressing this issue. Moreover, little is known about the extent to which a metaphor needs to be *creative* in order to perform an evaluative function, or whether there are differences according to the type of evaluation, such as its degree of explicitness and its polarity. In order to investigate these questions, 94 film reviews from the Internet Movie Database (IMDB) were annotated for creative and conventional metaphor, and for positive and negative, inscribed and invoked evaluation. Approximately half of the metaphors in our corpus were found to perform an evaluative function. Creative metaphors were significantly more likely to perform an evaluative function than conventional metaphors. Metaphorical evaluation was found to be significantly more negative than non-metaphorical evaluation. Both creative and conventional metaphors are used more frequently to perform inscribed evaluation than invoked evaluation. However, the tendency towards inscribed evaluation is stronger for conventional metaphors than for creative metaphors. From a theoretical perspective, these findings call into question fundamental assumptions about the role of metaphor in performing evaluation, such as the claim, made in the Systemic Functional Linguistics literature, that metaphor invariably 'provokes' attitudinal meanings. We have shown that it *can* do so, but that it does not *always* do so. The study also offers methodological contributions, by introducing a new protocol for the annotation of creative metaphors as well as detailed guidelines for coding evaluation at different levels of explicitness.

Keywords: creative metaphor, invoked evaluation, Appraisal framework, manual corpus annotation

1   Introduction

'Spawn' is an in-your-face, screaming banshee of a film

This quote is taken from a review of a film that appeared on a film review website. It offers a strong evaluation of the film by making an explicit, creative metaphorical comparison with a screaming banshee, a terrifying mythological creature from the Celtic tradition. It has been suggested that evaluation is often expressed by metaphor, and that metaphor nearly always performs some sort of evaluative function. As we can see in the example above, the metaphors that are used to express evaluation can be very striking and creative. However, we do not know the extent to which a metaphor *needs* to be creative in order to perform an evaluative function, or whether there are differences according to the type of evaluation, such as its degree of explicitness and its polarity, which affect the extent to which metaphor is used. Investigating these relationships is important because it helps us to understand the different communicative resources that people draw on when expressing different kinds of evaluation. Specifically, it teaches us about how metaphor functions in communication, and how and why people use language creatively in everyday contexts. In this paper, we explore the relationship between creativity, metaphor and evaluation in an intrinsically evaluative genre, that of the film review. Specifically, we investigate the extent to which evaluation is performed by metaphor, the kinds of evaluation that are most likely to be performed by metaphor, and whether there is a relationship between the type of metaphor used and the polarity and explicitness of the evaluation.

2.   Background

In this section, we begin by defining metaphor and exploring the distinction between *creative* and *conventional* metaphor. We go on to define evaluation, exploring the distinction between *inscribed* and *invoked* evaluation. We then discuss the relationship between metaphor and evaluation, in order to provide a rationale for our research questions, which are presented in Section 2.3.

2.1   What is 'metaphor'?

Simply conceived, metaphor is the device by which a concept is described in terms of another, unrelated concept (Cameron, 2003). Perhaps you've been having a rough day, for example, where *rough* in its most literal sense relates to texture and the sense of touch, not to periods of time. *Rough* can therefore be said to have an interpretation which seems incongruous with the context, thus producing a metaphor. In order to resolve this apparent incongruity, it is necessary to look for concepts that can be transferred, or *mapped*, from the incongruous domain of 'roughness' to the topic of 'a difficult day'. We might draw on our experience of hiking over literally rough ground to do this, calling to mind how difficult and exhausting the endeavour might have been. In so doing, we can understand that in talking

2

about a rough day, we refer not to literal ideas of touch, but to our experiences of rough terrain and the similarities between these experiences and the challenges of our day. These, then, are metaphors.

Historically, metaphor has been considered solely a literary device – an example of creative, deliberate language use, with little relevance to everyday communication. However, the work of Lakoff and Johnson in the 1980s (Lakoff & Johnson, 1980/2003) broadened our understanding of metaphor and consequently the scope of metaphor research. They demonstrated that much of the human conceptual system is metaphorical in nature, i.e., that we understand those more complex, abstract aspects of our experiences by relating them to more concrete, embodied tangible things. The complex emotions surrounding depression, for example, may be expressed by references to *drowning*, to being *weighed down*, or to *feeling trapped*. This leads to metaphor appearing in conventional language, and since Lakoff and Johnson, metaphor has indeed been shown to be used in all sorts of communicative contexts beyond the literary (Littlemore, 2019).

These changing approaches to the study of metaphor highlight the fact that there are different kinds of metaphor, and have led to an increased focus on the distinction between novel and conventional metaphor. The kinds of metaphor that Lakoff and Johnson discuss are, for the most part, highly conventional and would possibly not be considered metaphorical at all by the majority of language users. Others, however, like the example with which we opened the paper, are more novel. It has been shown that novel metaphors are processed in different ways from conventional metaphor; they involve processes of comparison rather than categorisation (Bowdle & Gentner, 2005) and recruit different areas of the brain when being interpreted (Cardillo et al., 2012). They are more likely than conventional metaphors to evoke an embodied simulation, which makes them more powerful and more noticeable (Cacciari et al., 2011).

At this point it is important to consider what is meant by a novel metaphor. This is a metaphor that involves drawing together previously unrelated concepts. For example, referring to a *screaming banshee of a film* is a novel metaphor because it involves a mapping that is unlikely to have been made before.

Novel metaphors such as these are somewhat rare in language. What is more common is for people to take conventional metaphors and use them in a novel way by combining or extending them in new ways. For example, consider this conversation:

[1]   'How can we reconcile these two ideas?'

        'Throw them both out of the window; they can reconcile on the way down'[1]

The idea of 'throwing ideas out of the window' is conventional, but the idea of those ideas doing anything 'on their way down' is novel. We can consider this to be an example of an elaboration of a conventional metaphor. It is not the case that the speaker is developing an

---

[1] Two of the authors in discussion as they prepared this paper.

3

entirely new mapping; instead, she extends and elaborates upon an existing one by adding more detail, personifying the 'ideas' and giving them the ability to 'reconcile' themselves.

Both of these strategies can be encapsulated in the term *creative use of metaphor* because they differ in some way from conventional language usage. The fact that creative uses of metaphor encompasses both novel metaphor per se and the creative manipulations of conventional metaphor is also discussed by Semino (2011), who argues that the juxtaposition of several related metaphors in the same part of the text can be considered creative use of metaphor even if the metaphors themselves are conventional.

However, Semino's focus is on the ways in which metaphor can be creatively used across different genres, so she does not go into detail on the myriad ways in which conventional metaphors can be creatively manipulated. In addition to extending existing mappings, as in Example [1] above, these include, for example, altering the valence, introducing a new collocation, or altering the tense or part of speech of a conventional metaphor. More examples of the ways in which conventional metaphors can be manipulated in creative ways are provided in Section 3.3. As we will see later in the paper, the creative use of metaphor is relevant to our discussion of the interplay between metaphor and evaluation.

2.2   What is 'evaluation'?

One of the most important things we do with language is express our opinions. We use words such as *influential* and *masterpiece*, for example, to praise books and works of art, or words such as *corrupt* and *unscrupulous* to criticize politicians. These expressions are examples of the linguistic phenomenon of evaluation. Evaluation is a broad functional category that groups together all the linguistic resources that speakers use to convey their subjective attitudes, feelings and stances in discourse (Hunston & Thompson, 2000). These include adjectives (e.g. *unique*), adverbs (e.g. *intelligently*), nouns (e.g. *crap*) and verbs (e.g. *outshines*). In fact, evaluative meanings often transcend the boundaries of individual lexical units and spread over longer stretches of text, as shown in Example [2][2].

> [2]   Musicals are as good as the songs and <u>there's not one you'd leave the theater humming</u>.

Regardless of how it is expressed, every act of evaluation involves a *source*, namely the person expressing the opinion, and a *target*, that is, the 'thing' being evaluated (Du Bois, 2007). The target can be either an entity, including objects, cultural products and people, or a proposition, expressed by a clause. Evaluative expressions may be used to convey either a positive or negative attitude towards the target, a property known as *evaluative polarity* (Hunston, 2011). Evaluative meanings may be further broken down into a number of more specific parameters, including, for instance *comprehensibility*, *importance*, or *expectedness*

---

[2] Throughout the article, we mark instances of evaluation with underlining and instances of metaphor with bold font.

(Bednarek, 2006). The 'good-bad' parameter, however, is the most basic one and underlies all forms of evaluative language (Hunston & Thompson, 2000: 25).

Evaluation is a highly context dependent phenomenon. Except for a limited set of expressions that tend to have a relatively 'stable' evaluative meaning (e.g. *awesome*, *terrible*), contextual cues and background assumptions, related for example to genre, play a big part in whether a stretch of text is interpreted evaluatively. Fuoli (2018) discusses *thin* and *light* as examples of adjectives that carry a neutral, descriptive meaning in most contexts, but that fulfil an evaluative function in advertising discourse, where they are often used to highlight desirable features of products. Polysemous words may carry evaluative and non-evaluative meanings. One example is the adjective *electric*, which is in most cases used as a neutral classifying adjective, but which can also be used to praise someone's artistic performance (Hunston, 2010: 14). The context-dependent nature of evaluation also affects the polarity of evaluative items. Some expressions may have negative polarity in certain contexts and positive in others. Take, for instance, the adjective *cheap*. This word can be used to positively evaluate, say, a hotel room, but also to criticize a product for its poor build quality or a person for their greed.

Evaluative meanings can be expressed more or less explicitly. A reviewer, for instance, may criticize a film overtly through lexical items that are clearly and unambiguously negative, as shown in Example [3] below.

[3] A better title for this <u>nostalgic mess</u> would be "50 missed opportunities".

Alternatively, they may convey their opinion indirectly via language that implies an evaluative stance:

[4] <u>It took me half of the movie just to figure out what was going on</u>.

Within the Appraisal framework (Martin & White, 2005), which emerged from the Systemic Functional Linguistics (SFL) tradition and which has become one of the most influential descriptive models of evaluation, wordings that convey the writer's stance explicitly are labelled *inscribed* evaluation and instances where the opinion is expressed indirectly *invoked* evaluation. This distinction is conceptualized as a continuum that reflects "the degree of freedom allowed readers in aligning with the values naturalised by the text" (Martin & White, 2005: 67). At one end of the continuum, we find linguistic expressions that denote evaluation, that is, intrinsically evaluative lexis that "tells us directly how to feel" (Martin & White, 2005: 62). At the other end, we have factual statements that, in the context in which they are used, are intended to trigger an evaluative inference without actually spelling out how the author feels. In Example [5], for instance, the reviewer's seemingly neutral description of scenes from the film suggests a negative appraisal. Crucially, the reviewer does not voice this opinion explicitly, using evaluative lexis such as *badly written* or *implausible*; these negative meanings are left for the reader to infer.

> [5] There are also a few scenes in which the killer <u>suddenly appears behind the next victim in a situation such that (s)he clearly would have been seen moving in that direction</u>.

Martin & White (2005) identify two additional sets of strategies for invoking attitudes that are more explicit than factual statements yet less overt than evaluative inscriptions, as shown in Figure 1. Writers may *flag* an evaluation by using counter-expectancy markers such as *however* or *actually*, intensified lexis, rhetorical questions and 'non-core' vocabulary. One step up the explicitness cline we find *provoked* evaluation, which is realized primarily via lexical metaphor. Thus, Martin and White (2005) consider metaphor as a device for expressing evaluation implicitly rather than explicitly, a point to which we return below.
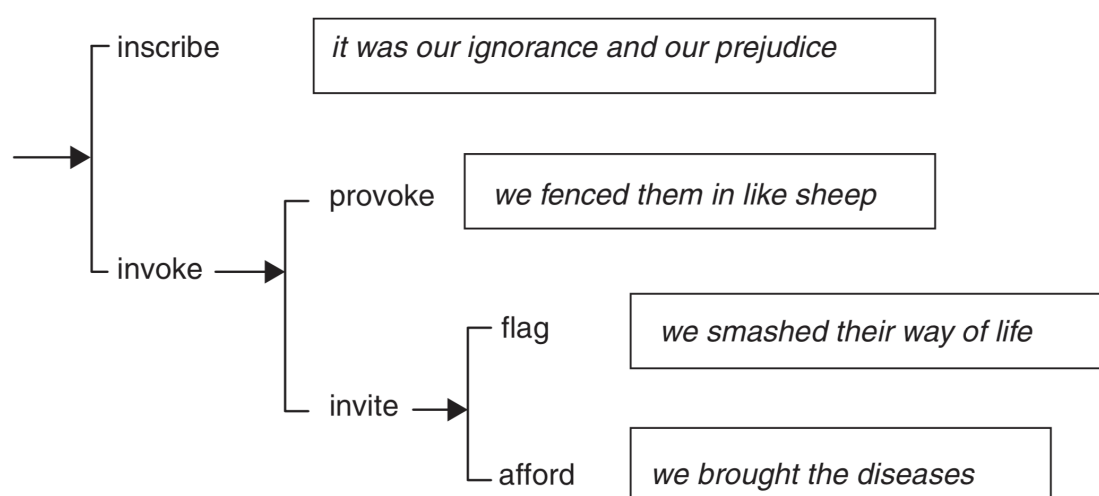
```
                ┌─ inscribe ──  it was our ignorance and our prejudice

   ──────►│
                │                      ┌─ provoke ──  we fenced them in like sheep
                └─ invoke ──►│
                                       │              ┌─ flag ──  we smashed their way of life
                                       └─ invite ──►│
                                                     └─ afford ──  we brought the diseases
```

Figure 1. Strategies for expressing evaluation at different levels of explicitness (Martin and White, 2005: 67)

2.3 The relationship between metaphor and evaluation

We saw at the beginning of the paper that metaphor is sometimes used to perform an evaluative function, and that in this paper, our aim is to investigate this in more depth. Work stemming from SFL appears to converge on the idea that evaluation is one of the main (if not the main) functions performed by metaphor in discourse. Martin (2020: 13), for example, argues that "[l]exical metaphors[3] are deployed to provoke a reaction". Along similar lines, Simon-Vandenbergen (2003) describes evaluation as a key motivating factor for most lexical metaphors. Crucially, as seen above, metaphor is considered in SFL as a resource for expressing evaluative meanings covertly rather than explicitly (e.g. Hood & Martin, 2005; Martin, 2020; Martin and White, 2005; Liu, 2018). Martin (2020: 13) summarizes the

---

[3] The term *lexical metaphor* is used in the SFL literature to distinguish metaphor involving lexical resources from phenomena classed as *grammatical metaphor*, such as nominalisation. Martin (2020: 1) presents lexical metaphor and conceptual metaphor as broadly overlapping. As we shall argue, however, there seem to be differences in the conceptual scope of these two categories.

argument for this theoretical position as follows: "unlike inscribed attitude involving explicitly attitudinal lexis, [metaphors] do not specify the precise attitude involved – leaving this for a reader to abduce based on their reading of the lexical metaphor in relation to its co-text".

However, while intuitively appealing, these proposals are largely theoretical and have not thus far been verified empirically. An additional problem is that the conceptual boundaries of metaphor are not defined clearly in the SFL literature and, as a result, it is unclear whether all types of metaphor are always considered to 'provoke' evaluation. The examples discussed in Martin and White (2005) would fall into the category of creative metaphor, as defined above. One of them is shown below.

> John Howard says he knows how vulnerable people are feeling in these times of economic change. He does not. **For they are feeling as vulnerable as a man who has already had his arm torn off by a lion, and sits in the corner holding his stump and waiting for the lion to finish eating and come for him again.** This is something more than vulnerability. It is injury and shock and fear and rage. And he does not know the carnage that is waiting for him if he calls an election. And he will be surprised. [Ellis, 1998, reproduced in Martin and White 2005: 65]

This example is from journalist Bob Ellis, criticising Australian Prime Minister John Howard's 1990s economic rationalism. Here, Ellis uses a creative metaphor to describe the experience of vulnerability in times of economic change. Martin and White (2005) argue that this utterance provokes rather than inscribes evaluation because the speaker does not explicitly condemn the economic policy or the Prime Minister. Rather, this negative judgment is implied by the analogy between being eaten by a lion and experiencing the effects of this economic policy expressed in the metaphor.

However, other studies seem to suggest that, in some cases, metaphor may also serve to inscribe evaluation. Simon-Vandenbergen (2003) brings a number of examples of conventional metaphorical expressions used for describing verbal processes that embed explicit evaluative meanings, such as *babble*, *bite someone's head off* or *jabber*. Similarly, Bednarek (2009) discusses examples of highly conventionalized metaphorical expressions which convey affect explicitly, such as *my heart sank* or *he had a broken heart*. These examples raise the question of whether the degree of explicitness of the evaluative meaning conveyed by a metaphor is a function of the type of metaphor used. In other words, do conventional metaphors tend to inscribe evaluation and creative metaphors to invoke it? As SFL does not distinguish between different types of lexical metaphors and has not addressed the relationship between metaphor and evaluation systematically, this remains an open question.

Within the metaphor literature itself, it has been argued that metaphor often performs some sort of evaluative function, but not always. For example, Semino (2008: 31) in her review of the functions of metaphor in discourse suggests that metaphor is frequently used to evaluative and to express attitudes and emotions, although she also proposes a number of other non-

evaluative functions performed by metaphor, such as persuading, reasoning, explaining, theorizing, entertaining, and organising the discourse. In her corpus-based study of fixed expressions and idioms in English, Moon (1998), found that metaphorical idioms are significantly more likely to serve an evaluative function than non-metaphorical idioms. Further evidence for a possible link between evaluation and metaphor can be found in Turner's (2014) study of French and Japanese learners of English. She found that when learners used metaphor in their written work, this was frequently to perform evaluative functions. Many of these evaluative metaphors were highly conventional, especially at the lower levels, suggesting that evaluation is 'baked' into a lot of conventional metaphor. However, this study did not examine the extent to which evaluation was performed without using metaphor, so it is not possible to draw firm conclusions as to the role of metaphor in performing evaluation relative to non-metaphorical language.

Additional theoretical support for the idea that the use of metaphor is linked to evaluation comes from the fact that metaphor is often used to express emotion. The linguistic expression of emotion, also known as *affect*, is generally considered as an integral part of the broader phenomenon of evaluation. Within the Appraisal framework, affect is considered as the most basic type of evaluative meaning, with other forms of evaluation representing "institutionalized feelings" (Martin and White, 2005: 45). A number of studies have shown that people often use metaphor when describing personal emotional experiences. In their study of women's accounts of cancer, for example, Gibbs & Franks (2002) discuss cases where the participants used highly creative, poetic metaphors to describe their experiences with the illness. Fainsilber & Ortony (1987) also found that people produced more metaphor, and particularly creative metaphor, when describing intense emotional experiences. They propose three hypotheses to explain this finding: the compactness hypothesis, the vividness hypothesis and the inexpressibility hypothesis. The compactness hypothesis refers to the idea that metaphor provides "a particularly compact means of communication" (Fainsilber & Ortony, 1987: 125), allowing a large amount of information to be conveyed in a far more compact way than literal speech does. The vividness hypothesis holds that metaphors can provide richer and more detailed accounts of experience than literal language, while the inexpressibility hypothesis holds that 'metaphors provide a way of expressing ideas that would be extremely difficult to convey using literal language' (Gibbs, 1994: 124). All of these come to the fore in the expression of intense, personal experiences. Such experiences are often difficult to express *without* recourse to metaphor.

Much of previous research on the relationship between metaphor and emotion has focused on negative experiences. One reason for this might be that in the field of metaphor studies, people have tended to research negative experiences more than positive ones. A more interesting idea is that metaphor in general and creative metaphor in particular are more likely to be triggered by negative emotional experiences than by positive ones. Studies have identified a human bias to give greater weight to negative entities (Rozin & Royzman, 2001), with people paying more attention to and remembering negative entities and events than to positive entities and events. Therefore negative experiences are more salient. One reason for this may be that negative emotions activate the sympathetic nervous system and increase arousal levels, whilst positive emotions activate the parasympathetic nervous system and

bring arousal levels down. Negative experiences are therefore more vivid which, according to the vividness hypothesis (Fainsilber and Ortony, 1987), means that they are likely to trigger more creative metaphor use.

There is some evidence from the metaphor literature to suggest this may be the case. For example, in her study of metaphorical fixed expressions introduced above, Moon (1998) found that evaluative metaphorical expressions were more likely to perform negative evaluation than positive evaluation. Further support comes from work on metaphor perception, where it has been shown that adjectival metaphors are more likely to evoke negative meanings than positive meanings, and that they are significantly more likely to do so than nominal metaphors and predicative metaphors (Sakamoto & Utsumi, 2014). Further support comes from research showing that media such as art and music provide creative outlets for negative experiences, and that people enjoy experiencing negative emotions in response to creative art and music (e.g. Bastian, 2017; Schubert, 1996). Thus the desire to produce creative metaphor may emanate in part from the need to share negative evaluation, which reflects the interpersonal function of both metaphor and evaluation.

To sum up this section, there are arguments to suggest that metaphors are often used to evaluate, and that evaluation is more likely to be performed by creative metaphor than by conventional metaphor. There is also indirect evidence to suggest that the use of creative metaphor is more likely to be triggered by negative emotional experiences than by positive ones. This leads us to hypothesise that the more creative the metaphor is the more likely it is that it will perform an evaluative function, and that creative metaphor will more likely be used to perform negative evaluation than positive evaluation. We saw above in Section 2.2 that in SFL models, the use of metaphor is most often associated with invoked evaluation than with explicit evaluation. Therefore one might hypothesise that creative metaphor is more likely than conventional metaphor to be involved in negative invoked evaluation and that both types of metaphor are more likely than non-metaphorical language to be used for this purpose. Based on this reasoning, we formulate our research questions and their associated hypotheses as follows:

RQ1:   To what extent does metaphor perform an evaluative function?

We expect a substantial amount of metaphor to perform an evaluative function.

RQ2:   Are creative metaphors more likely than conventional metaphors to perform evaluation?

We expect that creative metaphors are more likely to perform evaluation than conventional metaphors.

RQ3:   Is metaphor more likely to be used to convey negative or positive evaluation?

We expect metaphor to be used more frequently to perform negative evaluation than positive evaluation.

RQ4:   Does metaphorical creativity relate to evaluative polarity?

We expect creative metaphors to be used to perform more negative evaluation than conventional metaphors.

RQ5:   Is metaphor more likely to inscribe or invoke evaluation?

We expect metaphor to be used more often to produce invoked evaluation.

RQ6:   Does the explicitness of the evaluation differ according to whether the metaphor is creative?

We expect creative metaphor to be used more often than conventional metaphor to produce invoked evaluation.

3   Methodology

In order to explore these research questions, we chose to examine evaluation and metaphor in the genre of film reviews. Specifically, we focus on online reviews written by non-professional critics. These texts are produced by film enthusiasts for an audience of peers and are published on websites such as the Internet Movie Database (IMDb), Rotten Tomatoes or Metacritic. Online film reviews are an ideal genre for investigating both metaphor and evaluation. As the chief purpose of film reviews is to express the writer's personal views and assessment of a film in order to encourage, or discourage, prospective viewers, they tend to incorporate a wide variety of evaluative language (Taboada, 2011). The fact that reviews are written, asynchronous texts means that the authors have time to reflect on their choice of words, which is likely to result in more metaphor use (Hanks, 2006; Steen et al., 2010). Similarly, the fact that they have more time to reflect on their choice of words and to use the language playfully means that one might also expect a higher concentration of creative metaphors. The relatively familiar relationship between the author and the reader combined with the fact that a secondary purpose of the reviews is to entertain means that we are likely to see a fair degree of humour, which may also involve creative word play, often involving creative metaphor.

To answer the research questions outlined above, we annotated a corpus of film reviews for both evaluation and metaphor and examined overlaps between these two categories. We used Nvivo (QSR International, 2020) for this purpose, as it allows researchers to query the corpus for instances where a stretch of text has been coded with multiple labels. Evaluation and metaphor were annotated independently of one another to capture all cases of each phenomena, regardless of overlap. Thus to answer question 1, for example, we divided the

10

number of text spans coded as evaluative and metaphorical by the total number of text spans coded as evaluative. Example [6] below illustrates a text span annotated for both evaluation (underlined) and metaphor (in bold).

      [6] The actors are <u>mostly **mobile wooden statues**</u>.

In the sections below, we give more detail about the corpus and the annotation protocols we used.

3.1   The corpus

We compiled our corpus by down sampling a large, publicly available[4] collection of IMDb reviews collected by Pang & Lee (2004). The original corpus includes 1000 positive and 1000 negative film reviews. From these, we randomly selected 94 texts equally subdivided between positive and negative reviews. The total corpus size is approximately 60,000 words, which represents a rich, yet manageable, sample for manual annotation.
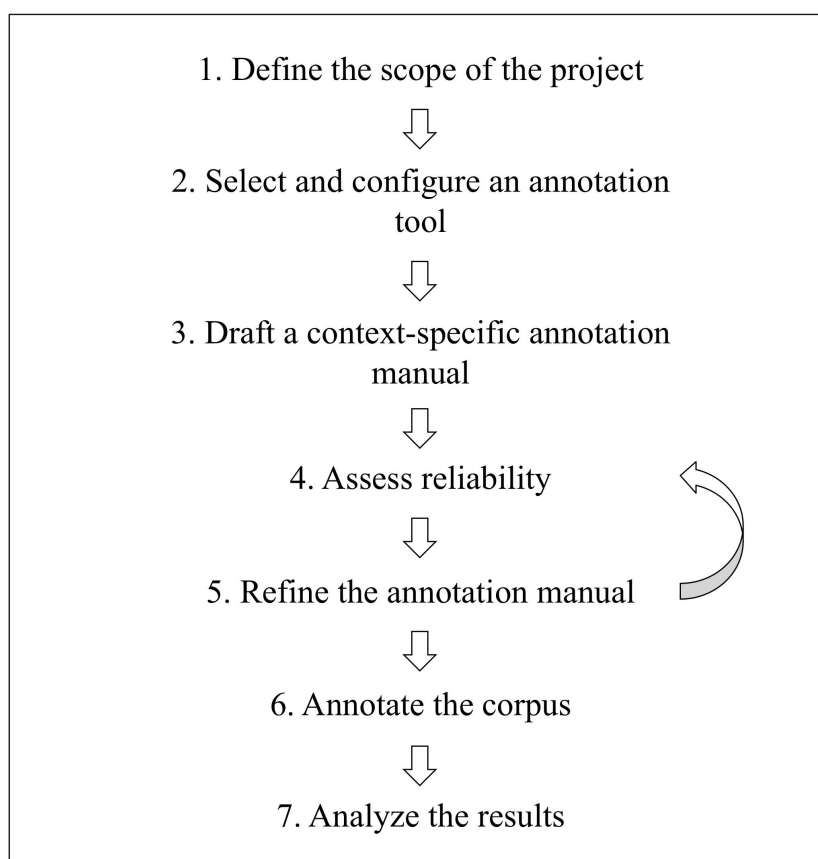
3.2 Corpus annotation

Annotating metaphor and evaluation is an inherently subjective process as both are context-dependent discursive phenomena with fuzzy conceptual and lexical boundaries. To address these methodological concerns, we followed the stepwise annotation procedure proposed by Fuoli (2018), which is shown in Figure 2. A key feature of this approach is that it incorporates an iterative process for optimizing the transparency and replicability of the annotation guidelines. Before coding the corpus, we developed detailed annotation manuals for both metaphor and evaluation (step 3). The manuals, which can be found in the Supplementary Materials, include operational definitions of our categories and a detailed description of the protocols we used to identify and categorize instances of metaphor and evaluation. Next, we carried out three rounds of inter-coder agreement testing in order to assess the reliability of the coding protocols and identify areas for improvement (steps 4 and 5). The results of the inter-coder agreement tests are presented in Section 3.5. After we determined that reliability had reached a ceiling, we moved on to annotate the rest of the corpus. Jeannette Littlemore and Sarah Turner annotated half of the remaining portion of the corpus for metaphor each (consulting with one another on all ambiguous cases) and Matteo Fuoli annotated the whole of the remaining sample for evaluation. Whenever any of the annotators encountered ambiguous instances that they were not able to resolve on their own, they consulted the rest of the team to help determine the most adequate coding. In the interest of transparency and reproducibility, we have made the fully annotated corpus available via the Open Science Framework repository at this URL: https://osf.io/y7v54/?view_only=4cb57e05fc344a29bf9322009ada2e5f.

---

[4] The corpus can be downloaded here: http://www.cs.cornell.edu/people/pabo/movie-review-data/

Fig 2. The step-wise corpus annotation procedure

1. Define the scope of the project

⇩

2. Select and configure an annotation tool

⇩

3. Draft a context-specific annotation manual

⇩

4. Assess reliability

⇩

5. Refine the annotation manual

⇩

6. Annotate the corpus

⇩

7. Analyze the results

3.3   Annotation protocol for metaphor

In this study, we define a metaphorical expression in the following way:

A string of one or more words that describes one entity in terms of another unrelated entity by means of comparison.

Under this definition, the highlighted text span in Example [7] below would be an example of a metaphorical expression.

[7] It's pretty much a **sunken ship** of a movie.

Here, the words *sunken ship* are being used to describe the movie. In order to understand how the metaphor is functioning in this example, the reader needs to identify elements of 'sunken ships' that can be applied to 'movies', i.e. that it is a wreck with no hope of salvage or rescue. This enables the movie to be negatively evaluated in a marked way.

3.3.1 Procedure for identifying metaphor

In order to identify metaphors we employed a procedure that drew on two previously attested approaches: Cameron's (2003) vehicle identification procedure and the PRAGGLEJAZ (2007) metaphor identification procedure (MIP), combining elements of each. Our reason for doing this was that we wanted to combine the best elements of each, allowing us to focus on metaphor at the level of the phrase (which is a more natural way of looking at metaphor) with a robust technique for ensuring that we were definitely dealing with metaphor and not other related tropes such as metonymy.

We began by reading the entire text to establish a general understanding of the meaning. We then identified meaning units at the level of phrase following Cameron's (2003) vehicle identification procedure. For each meaning unit, we established its meaning in context, (i.e. its contextual meaning, taking into account what comes before and after the meaning unit). Having done so, we determined whether or not the phrase had a more basic contemporary meaning in other contexts than the one in the given context. For our purposes, basic meanings tend to be

—More concrete [what they evoke is easier to imagine, see, hear, feel, smell, and taste];

—Related to bodily action;

—More precise (as opposed to vague);

However, unlike the PRAGGLEJAZ (2007) MIP, we did not consider historically older meanings to be more basic. We also included metaphors that crossed word-class boundaries, as this is often a central characteristic of metaphor. For example, *staggering* is an adjective in its metaphorical sense but a verb in its literal sense. Strict adherence to the MIP would not code the adjective *staggering* as a metaphor as it does not share the same word class as its literal meaning. However, we coded it as metaphor because its meaning could be understood in comparison to the verb. In our analysis, we only considered open-class lexical units, excluding closed-class items and de-lexicalised verbs (*make, do, put, take, give, have,* and *get*). It should also be noted that basic meanings are not necessarily the most frequent meanings of a particular word or phrase.

If the meaning unit had a more basic current–contemporary meaning in other contexts than the given context, we decided whether the contextual meaning contrasted with the basic meaning but could be understood in comparison with it. If the meaning unit met all of these criteria, it was marked as metaphorical.

In some cases, metaphors were identified at the level of the single word. However, a single metaphor often extended beyond single words. This could occur when:

i) The expression was a conventional idiom, such as *have your cake and eat it*. In cases such as this the whole idiom was coded as a span of text that conveys metaphorical meaning.

ii) There were hyphenated words which form a single lexical unit e.g. *tough-as-nails* Salander.

iii) There was an adjectival entailment of a metaphorically-used noun (or an adverbial entailment of a metaphorically-used verb) that was internally semantically coherent with the *literal* sense of the noun or verb, as in Example [8] below.

> [8] It's pretty much a **sunken ship** of a movie (Ships can sink in the 'literal' world and 'sunken' is serving as a premodifier of 'ship' in this sentence.)

Phrases that were internally coherent were marked as a single metaphor, even when there was a non-metaphorical stretch of texts separating them. For instance, in Example [9] below, the word *depth* and the phrase *skin deep* both belong to the same overall idea, so the whole phrase is marked as a single metaphor.

> [9] The real **depth** of his character is only **skin deep**

In some cases, the focus on internal coherence meant that whole grammatical phrases could be coded as a single metaphor, as in Example [10].

> [10] you can't help **going in with the baggage** of good reviews

However, if there were two distinct ideas in the same metaphorical phrase, these were marked as separate metaphors. For instance in Example [11] below, 'the one-two punch' and 'derailing itself' are different metaphorical ideas, one from the domain of fighting and one from the domain of rail travel, even though they work together in the sentence.

> [11] The actors, and their relationship together, present the **one-two punch** that prevents Double Jeopardy from **derailing itself** entirely.

iv) There is an adjectival entailment of a metaphorically-used noun (or an adverbial entailment of a metaphorically-used verb) that that is internally semantically coherent with the *metaphorical* sense of the noun or verb but which would not occur in literal language:

> [12] which is in contrast to the **negative baggage** that the reviewers were likely to
have

In the physical world, baggage cannot be 'positive' or 'negative'. This expression is only ever used in its metaphorical sense (unlike the phrases *bee stings* and *sunken ship*, which can exist in the physical world).

Phrases were coded as metaphor even when they were signalled with tuning devices such as *like* or *as*. Individual words were not broken down into their metaphorical components. We

followed an overarching principle where we kept the length of the annotated text spans to a minimum.

3.3.2 Procedure for identifying creatively-used metaphor

Having identified all examples of metaphor in our corpus, we then determined whether these metaphors were being creatively or conventionally used.

Metaphors were coded as creatively-used under the following conditions:

1. When they introduced a completely new metaphorical mapping drawing together previously unrelated elements, as in Example [13].

> [13] These guys know how to **graft a comic book onto celluloid**

2. when they used a conventional metaphorical mapping in a new way, playing with the meaning or the form or both.

This could be achieved in one or more of the following ways:

a) Altering the valence of a metaphor (positive and negative)

> [14] Actually, Robin Williams does a lot of shouting. He shouts a lot about helping people, and a lot of people cry because they are **moved** by his words. I won't tell you that you can't be **moved** by his words, because I too, was **moved** by his words. I was **moved** in such a profoundly negative way that I was reminded of how cheap and phony a cinematic experience can be.

Usually when we are *moved* by something, it has positive connotations, but here the reviewer is evaluating Robin Williams in an overtly negative way by using *moved* creatively and imbuing it with negative connotations.

b) Introducing a new collocation

This occurred in cases where conventional collocational patterns involving metaphor were flouted:

> [15] steal clout from (one might 'have' clout, but one would rarely 'steal' it)

> [16] delicate power (near oxymoron)

> [17] Christina Ricci, **hot off her shoulda-been-nominated turn** in "the opposite of sex" (creative extension of 'hot off the press')

c) Introducing more detail into a conventional mapping, or extending it in a novel way (often evoking hyperbole or litotes)

16

> [18] James Cameron took the big-budget action film with aliens , which featured multiple aliens doing basically the same thing , although on a much-larger scale, and boy , did he take that route ! **I'd say at about 165 mph** or so . . .

d) Altering the tense or part of speech of a conventional metaphor

> [19] A **sunken ship** of a movie (it is more conventional to metaphorically refer to a 'sinking ship', rather than a 'sunken' one)

e) Using a metaphor in a new context where it is not usually used, or to talk about something that it's not usually used to talk about

> [20] There is not **an original or inventive bone in its [the film's] body** (this expression is usually used about a person, not a film)

f) Using a 'twice true' metaphor

Twice true metaphors are metaphors which work on two levels; they have a literal meaning that is relevant to the context of the film they are being used to describe.

> [21] Once 'Jaws' has attacked, **it never relinquishes its grip** (Here, 'it' refers to both the film and the shark).

g) Combining metaphor with metonymy in a novel way

> [22] It's typical of unimaginative cinema to **wrap things up with a bullet** (Here, the 'bullet' refers metonymically to the act of killing someone off at the end of the film)

h) Combining two conventional metaphors in a novel way

> [23] A **big helping** of **whoop-ass** behaviour

Here there are two conventional metaphors: *big helping* and *whoop-ass*. Juxtaposing them is creative, and construes 'whoop-ass behaviour' as something that might be served up in a restaurant.

i) Using strong and unlikely or unexpected personification

> [24] The decor possibilities are endless - **disco balls had yet to migrate into the dark corners of the attic**, big hair was worth its weight in Aquanet, and the louder the fashion, the better the look.

j) Introducing dramatic contrast

> [25] The great master shows his hand there as the tensions **build** as rapidly in the second part as they **lay fallow** in the first.

k) Using recontextualisation and appropriation

In Example [26] below, the whole phrase is coded as creative metaphor, as the creativity comes from the appropriation of a well-known phrase, even though the only metaphor here is *fishy*)

[26] **Something is fishy in the state of Universal**

3.4 Annotation protocol for evaluation

We developed a set of explicit criteria for identifying units of evaluation in our corpus and for categorizing them based on their polarity and explicitness. For the purpose of this study, a unit of evaluation is defined as follows:

A string of one or more words that conveys the writer's positive or negative emotions, attitudes or judgments towards someone or something.

In line with previous work (see Section 2.2), this definition covers an open-ended range of expressions of any length and belonging to any word class. For a stretch of text to be considered an instance of evaluation, it had to involve a discernible evaluative target. Thus, words that are used to describe positive or negative phenomena, such as *success* or *crime*, were not coded as evaluative unless they were included in text spans that convey the writer's opinion of someone or something.

To help achieve consistency in our annotations, we took a conservative approach to the identification of the textual boundaries of evaluative units. Accordingly, we kept the length of annotated text spans to a minimum, leaving out all lexical items that did not directly contribute to the evaluative meaning of the expression, such as the subject of the clause or words referring to the evaluative target. Examples [27] and [28] below illustrate the difference between our approach and a less conservative approach, respectively.

[27] She's an <u>ass-kicking</u> cybertech warrior who rights the wrongs of men.

[28] <u>She's an ass-kicking cybertech warrior</u> who rights the wrongs of men.

In line with the Appraisal framework, expressions relating to the writer's emotions (i.e. affect) were included in the analysis. However, as we were mainly interested in how metaphor is used by speakers to perform evaluation, we only coded instances of *authorial* affect, that is, expressions that convey the reviewers' own emotions. Expressions describing emotions attributed to other people, such as a character in the movie, were not coded. Thus, for instance, we annotated the word *loved* in Example [29] below but ignored the expression *unhappy* in Example [30].

[29] And Judd Hirsch steals the film by actually acting great (he's a stereotype, but I just <u>loved</u> the man anyway).

[30] Rosalba (Licia Maglietta), an *unhappy* housewife from Pescara, finds herself - and love - in Venice.

Evaluative expressions can, in some cases, be nested inside one another. This phenomenon occurs when an expression evaluating a given target is embedded within a wider stretch of text which, in turn, serves to convey evaluation of a different target. Nested evaluative expressions thus typically involve two evaluative targets: an *immediate* target and a *contextual* target. The immediate target is the object or person that is directly modified by the embedded evaluative expression. The contextual target is the object or person that is assessed by the embedding unit of evaluation. In Example [31], for instance, the evaluative adjective *nice* modifies the immediate targets *hair* and *costumes*. In turn, the phrase *complete with nice hair and costumes* serves as a positive evaluation of the contextual target *The Mod Squad*.

[31] The Mod Squad is certainly a slick looking production, complete with *nice* hair and costumes, but that simply isn't enough.

Where we encountered nested evaluative expressions, we annotated both the embedded and embedding units.

All units of evaluation were coded as either positive or negative. When markers of negation reversed the polarity of an evaluative expression, they were incorporated into the annotated text span, as in Example [32].

[32] The characters and acting is nothing spectacular, sometimes even bordering on wooden.

When this was not possible because the negation marker was too far from the evaluative expression it modified, we annotated the evaluative expression only but with the polarity reversed.

In some cases, negative evaluations are used to invoke a positive appraisal of the movie. This is common, for instance, in reviews of horror films, where negative qualities such as *creepy*, *terrifying, ominous* are sought after and appreciated as key elements of the genre. Example [33], taken from a review of Spielberg's *Jaws*, illustrates this occurrence. In cases like this, the evaluative expression was coded as both inscribed negative - the 'face value' polarity - and as invoked positive.

[33] He's building the tension bit by bit, so when it comes time for the climax, the shark's arrival is truly terrifying.

As explained in section 2.2, we operationalized evaluative explicitness as a binary distinction between inscribed and invoked instances. We define inscribed evaluation as feelings and evaluations that are explicitly conveyed by expressions that are manifestly positive or negative in the context in which they are used. With inscribed evaluation, the exclusive function of the expression is to evaluate something or someone:

[34] The special effects in Mary Poppins were <u>groundbreaking</u>.

We operationalized invoked evaluation as an assessment of someone or something which is not expressed overtly, but is implied by what the reviewer is saying. Their evaluative stance can be inferred from the context, based on implicit assumptions about what counts as good or bad in a given situation. Typically, with invoked evaluation the text span does not exclusively serve an evaluative function, but also conveys factual information. In Example [35], for instance, the reviewer critiques the movie by describing aspects that do not receive enough attention. The phrase *there's no attention given* conveys factual information about the content of the movie, but is also interpreted evaluatively as indicating a flaw in the way given historical circumstances are depicted in the film.

[35] The sequel really dumbs down the social context of the originals. It takes place during "The Great Slump" but <u>there's no attention given to what was causing the Depression</u>.

With invoked evaluation, the whole action, event or proposition that suggests a positive or negative opinion was annotated, as shown in Examples [35].

Sarcasm was treated as a case of invoked evaluation. In Example [36], for instance, the underlined expression is used ironically to emphasize the predictability of the movie's plot.

[36] What does she do? She invents a fiance! Then when everyone wants to meet him, she tells some poor schmoe she met at a wedding that she will pay him $1000 to pretend to be in love with her for a company dinner, and pick a fight with her at the end, thus breaking the engagement but still being able to keep her job, since the guy ends up looking like a jerk and she is the poor, defenceless female. He, of course, goes along with it. Gee, <u>I wonder if they get together in the end</u>.

When sarcasm reversed the polarity of the evaluation, we double coded the evaluative expression for both the 'face-value' polarity and the invoked, sarcastic negative polarity. For example, the expressions *benevolent studio gods*, *delighted* and *thrilled* in Example [37] were coded both as explicitly positive and as invoked negative. The negative meaning is inferred from a sarcastic reading of the sentence which is warranted by the wider context in which it appears.

[37] Last year, the <u>benevolent studio gods</u> gave us Digimon, and this year, they bestow Max Keeble's big move on <u>delighted</u> moviegoers across the country. Parents will be <u>thrilled</u> because they'll finally have something to drag little Austin and Kayla to see.

In addition to the criteria outlined in this section, we made a number of detailed choices and rules, all of which are described in full in the complete annotation manual, which is given in Supplementary Materials.

3.5   Inter-coder agreement

Table 1 shows the results of the three rounds of inter-coder agreement testing we carried out for each category in our coding scheme. We report the average values of three inter-coder agreement measures: observed agreement, chance-corrected kappa and prevalence-adjusted bias-adjusted kappa (PABAK). PABAK is a measure of inter-coder agreement developed by Byrt et al. (1993) as an alternative to kappa to address situations where the distribution of categories in a corpus is highly skewed. A well-documented problem with kappa is that in cases where one category is substantially over-represented compared to another, high levels of observed agreement can yield very low or even negative kappa scores (Artstein & Poesio, 2008). This issue arises because, in cases of strongly unbalanced distribution, the amount of agreement that would occur by chance is inherently high (Eugenio & Glass, 2004; Feinstein & Cicchetti, 1990). PABAK corrects kappa for prevalence by assuming equal distribution of the categories in the corpus. In our case, inter-coder agreement was calculated separately for each category based on the number of characters in the corpus that were coded for a given category versus the number of characters that were left uncoded. Given that, taken individually, the features we annotated are relatively rare, uncoded characters vastly outnumbered coded ones, in many cases exceeding a 9:1 ratio. We therefore decided to report PABAK in addition to observed agreement and kappa scores in order to provide a more accurate picture of the levels of agreement reached in our tests.

Table 1. Intercoder agreement results

	Mean observed agreement (%)			Mean kappa			Mean PABAK		
	R1	R2	R3	R1	R2	R3	R1	R2	R3
Inscribed evaluation	92.29	94.69	92.66	0.39	0.62	0.53	0.85	0.89	0.85
Invoked evaluation	88.90	84.26	86.08	0.24	0.36	0.28	0.78	0.69	0.72
Positive evaluation	95.04	93.10	95.07	0.50	0.49	0.67	0.90	0.86	0.90
Negative evaluation	90.40	88.75	89.60	0.51	0.49	0.47	0.81	0.77	0.79
All metaphor	99.13	98	97.71	0.94	0.78	0.83	0.98	0.96	0.95
Creative use of metaphor	99.75	99.48	99.08	0.83	0.55	0.42	1	0.99	0.98

As Table 1 shows, PABAK scores were 0.69 or higher, indicating substantial agreement between annotators for all the coded categories (Landis & Koch, 1977). Overall, these results thus suggest that the guidelines for annotating evaluation and metaphor developed for this study are well defined and reliable. Levels of agreement were especially high in the case of metaphor. Perhaps unsurprisingly, agreement was lowest in the case of invoked evaluation. This result reflects the inherently subjective and context-dependent nature of this type of evaluation.

4   Findings

We used the coding query functionality in Nvivo to cross tabulate categories and quantify overlaps between metaphor and evaluation. At this point, it is worth briefly addressing the way in which NVivo reports its coding counts. In some cases, there is no one-to-one mapping between stretches of text coded for metaphor and for evaluation. In some cases, the overlap was only partial, meaning that a single stretch of text coded for evaluation could be counted as both metaphorical and non-metaphorical. For example, the sentence 'these awkward subplots pad out the running time to adequate feature length' was coded as negative evaluation, whereas *pad out* was coded as metaphor. NVivo would therefore count this as an example of evaluation both containing, and not containing, metaphor. In addition, as discussed above, some instances of evaluation were double-coded as both positive and negative and as both inscribed and invoked (e.g. instances of sarcasm). These aspects of the coding approach we adopted mean that some of the sum figures across sets of comparisons do not match. For example, if we add up the number of positive and negative evaluative items involving metaphor presented in table 6, we obtain 1341 instances. This number is higher than the number of items coded as both metaphorical and evaluative reported in Table 2 (1299). This discrepancy can be explained by the fact that the counts in table 6 necessarily incorporate double-coded items, whereas those in Table 2 include any item coded as evaluative, regardless of its polarity. These inconsistencies do not affect the validity of our conclusions, however, as each research question is dealt with separately and the calculations performed to answer it are based on internally-consistent counting criteria.

4.1   To what extent does metaphor perform an evaluative function?

The percentage of metaphorical items that served an evaluative function is shown in Table 2. These findings indicate that there was a roughly equal split between metaphorical items that convey evaluation, such as *[the film has] the sweetness of a candy apple* and metaphorical items that do not perform evaluation, such as *somewhere along the way*. Therefore, in contrast to previous work, we found that the majority of metaphor is not, in fact, used to perform evaluation.

Number of items that are both metaphorical and evaluative	1299
Total number of metaphorical items	2599
Percentage of metaphorical items that are also evaluative	**49.98%**

Table 2. Percentage of metaphorical items that served an evaluative function

4.2   Are creative metaphors more likely than conventional metaphors to perform evaluation?

We were interested in investigating whether creative or conventional metaphor would be more likely to perform evaluation. In order to do this, we performed a chi-square test comparing the proportion of creative and conventional metaphors that performed an evaluative function.

Table 3 shows the extent to which creative metaphor performed evaluation. We see that approximately three quarters of the creative metaphors were evaluative. These findings suggest that creative metaphors that performed an evaluative function were much more common than creative metaphors that did not perform any sort of evaluation.

Number of creative metaphors that are evaluative	140
Total number of creative metaphors	190
Percentage of creative metaphors that are evaluative	**73.68%**

Table 3. Percentage of creative metaphors that performed an evaluative function

Table 4 shows the extent to which conventional metaphor performed evaluation. These findings indicate that conventional metaphors that did not perform an evaluative function were slightly more common than conventional metaphors that did perform some sort of evaluative function.

Number of conventional metaphors that are evaluative	1160
Total number of conventional metaphors	2410
Percentage of conventional metaphors that are evaluative	**48.13%**

Table 4. Percentage of conventional metaphors that performed an evaluative function

Finally, we conducted a chi square test using the raw figures in the tables above to establish whether the difference between these two distributions was significant. The difference was indeed significant with creative metaphors performing more evaluation than conventional metaphors (χ^2(1) 13.4072 p < .001). Table 5 gives examples of metaphors that were coded in each category. As in the examples above, text spans coded as expressing evaluation are underlined while text spans expressing metaphor are in bold.

	Evaluative	Non-evaluative
Creative	(n=140) **[not] throwing any concrete plot details across the table**	(n=50) my **internal way-back machine swept me back to the mid-1960s** (here, the metaphor is deemed to be descriptive and possibly humorous/entertaining)
Conventional	(n=1160) the relationship between Howie and Big John is **evenly paced**	(n=1250) L.I.E. stands for Long Island Expressway, which **slices through** the strip malls What plot there is **hinges on** who has Enola (Here, the metaphors are deemed to be purely descriptive)

Table 5. Examples of evaluative and non-evaluative creative and conventional metaphors

4.3   Is metaphor more likely to be used to convey negative or positive evaluation?

We were interested in investigating whether evaluation that involved metaphor would be more positive or more negative than evaluation that did not involve metaphor. In order to do this, we performed a chi-square test comparing the number of positive and negative evaluative expressions involving metaphor with the number of positive and negative evaluative expressions not involving metaphor.

Table 6 shows the polarity of evaluation involving metaphor. We see that of all instances of evaluation involving metaphor, the majority were negative.

	Raw count	% of total
Positive evaluation involving metaphor	559	41.69
Negative evaluation involving metaphor	782	58.31

Table 6. Number of cases of positive and negative evaluation involving metaphor

Table 7 shows the polarity of evaluation not involving metaphor. We see that of all the instances of non-metaphorical evaluation, just over half were negative.

	Raw count	% of total
Positive evaluation not involving metaphor	1757	45.90
Negative evaluation not involving metaphor	2071	54.10

Table 7. Number of cases of positive and negative evaluation not involving metaphor

The results of the chi square test show that the difference between these two distributions was statistically reliable. Metaphorical evaluation was found to be significantly more negative than non-metaphorical evaluation ($\chi^2(1)$ 7.1288 p < .01). Table 8 includes examples of each case considered in this test.

	Involving metaphor	Not involving metaphor
Positive evaluation	(n=559) This film is filled with many little scenes which are <u>absolute **gems**</u>	(n=1757) Cameron directs them <u>so skillfully</u>, and <u>so suspensefully</u>

Negative evaluation	(n=782)	(n=2071)
	A <u>relative</u> **flop** at the cinema	It's <u>somewhat silly</u>

Table 8. Examples of positive and negative metaphorical and non-metaphorical evaluation.

4.4 Does metaphorical creativity relate to evaluative polarity?

We have seen above that metaphor, when used evaluatively, was significantly more likely to perform negative evaluation than positive evaluation. However, we were also interested in ascertaining the extent to which metaphorical creativity related to the polarity of the evaluation it is being used to perform.

Table 9 below shows the percentage of evaluative creative metaphor used for positive and for negative evaluation. We see that creative metaphors were used more often to perform negative evaluation, with approximately two thirds of evaluative creative metaphors being used negatively.

	Raw count	% of total
Number of creative metaphors used for positive evaluation	58	**39.73**
Number of creative metaphors used for negative evaluation	88	**60.27**

Table 9. Number of creative metaphors that were used for positive and negative evaluation

Table 10 shows the percentage of evaluative conventional metaphor used for positive and for negative evaluation. The results for conventional metaphor paint a similar picture to those for creative metaphor, with approximately two thirds of the evaluative conventional metaphors being used negatively.

	Raw count	% of total
Number of conventional metaphors used for positive evaluation	501	**41.89**
Number of conventional metaphors used for negative evaluation	695	**58.11**

Table 10. Number conventional metaphors that were used for positive and negative evaluation

The difference in distribution between positive and negative evaluation within creative and conventional metaphor was not significant (χ^2(1) 0.2506 p= .617). Creative metaphor and conventional metaphor behave similarly when performing evaluative functions, with both performing slightly more negative than positive evaluation. Table 11 shows examples of these four scenarios.

	Positive evaluation	Negative evaluation
Creative metaphor	(n=58) **perking up the movie like an injection of anti-depressant**	(n=88) **it plods back and forth, up and down a long and winding road before it ends up nowhere**
Conventional metaphor	(n=501) Even the acting in From Hell is **solid**, with the dreamy Depp turning in a typically **strong** performance and **deftly handling** a British accent	(n=695) It's pretty **run of the mill**

Table 11. Examples of creative and conventional metaphorical language serving positive and negative evaluative functions

4.5   Is metaphor more likely to inscribe or invoke evaluation?

Having established that just under half the metaphors in our corpus were used to perform an evaluative function, and that these were significantly more likely to perform negative evaluation, we now turn to investigate the relationship between explicitness of evaluation (i.e. inscribed or invoked) and metaphor use.

Martin and White's (2005) Appraisal framework places metaphor within the invoked evaluation category, with no mention of metaphor in any other evaluation type. However, we found that metaphor actually serves more often to convey evaluation explicitly than implicitly. Table 12 below shows the percentage of metaphorical evaluative expressions used to perform inscribed and invoked evaluation. We see that approximately two thirds of metaphorical evaluative expressions performed inscribed evaluation, with approximately one third performing invoked evaluation.

	Raw count	% of total
Number of metaphorical inscribed evaluative items	872	**64.02**
Number of metaphorical invoked evaluative items	490	**35.98**

Table 12. Number of metaphorical evaluative items used for inscribed and invoked evaluation

4.6 Does the explicitness of the evaluation differ according to whether the metaphor is creative?

As seen above, metaphor was more likely to be used to perform inscribed rather than invoked evaluation. However, we were also interested in investigating whether the creativity or conventionality of the metaphor had an effect on the explicitness of the evaluation it was used to perform. To answer this question, we compared the number of instances of inscribed and invoked evaluation across the two metaphor types by means of a chi-square test.

Table 13 shows the types of evaluation performed by creative metaphor. We see that creative metaphor is used to perform inscribed and invoked evaluation equally, with just over half of evaluative creative metaphors being used for inscribed evaluation and just under half of creative metaphors being used for invoked evaluation.

	Raw count	% of total
Number of creative metaphors used for inscribed evaluation	76	**51.01**
Number of creative metaphors used for invoked evaluation	73	**48.99**

Table 13. Number of creative metaphors used for inscribed and invoked evaluation

Table 14 shows the types of evaluation performed by conventional metaphor. Unlike creative metaphor, there is a far more noticeable difference between the types of evaluation. When conventional metaphor performed an evaluative function, approximately two thirds of these were inscribed evaluation, while approximately one third were invoked evaluation.

	Raw count	% of total
Number of conventional metaphors coded as inscribed evaluative items	797	**65.65**

Number of conventional metaphors coded as invoked evaluative items	417	**34.35**

Table 14. Number of conventional metaphors used for inscribed and invoked evaluation

The results of a chi-square test show that, even though both creative and conventional metaphors are used more frequently to perform inscribed evaluation, the tendency towards inscribed evaluation is significantly stronger for conventional metaphors than for creative metaphors (χ^2(1) 12.3598 p < .001). For creative metaphors, the behaviour is more balanced. In other words, creative metaphors are equally likely to be used for inscribed or invoked evaluation but conventional metaphors are more likely to be associated with inscribed evaluation. Table 15 gives examples of metaphors that were coded in each category.

	Inscribed	Invoked
Creative	(n=76) Most of it **fell flatter than a cartoon character that drops off a cliff**	(n=73) Burns is content to allow the film to **ramble aimlessly towards** its irritatingly predictable conclusion
Conventional	(n=797) The predictable ending that **shattered** our hopes	(n=417) This is what **binds the movie together**

Table 15. Examples of creative and conventional metaphors performing invoked and inscribed evaluation

5   Conclusion

In this paper we have explored the relationship between metaphor and evaluation in the context of film reviews. We were interested in establishing whether the use of metaphor was driven by different types of evaluation (positive or negative, inscribed or invoked), and whether different types of evaluation were related to the tendency to use creative or conventional metaphor. We found that metaphor was only used evaluatively in roughly half of cases, which means that they are not as tightly related as some of the previous literature has suggested. Creative metaphors were more likely to perform an evaluative function than

conventional metaphors, which may relate to the ability of creative metaphor in particular to express evaluation in a vivid and compact fashion (Fainsilber & Ortony, 1987).

In terms of polarity, metaphorical evaluation was significantly more negative than non-metaphorical evaluation, with creative and conventional metaphors behaving in the same way in this respect. This finding confirms previous work on the negative nature of metaphorical fixed expressions (Moon, 1998), but it extends this existing work to metaphor more generally, regardless of whether it occurs in a fixed expression. The fact that both creative *and* conventional metaphors are used in a similar way with respect to polarity is somewhat surprising given previous work showing a link between creativity and descriptions of negative experiences. However, this could be partly due to the nature of the events being evaluated. In order for negative evaluation to have an impact on creativity, it seems that the events being evaluated should be emotionally impactful and personal, whereas in our study the review writers are evaluating more external elements, e.g. plot, cinematography, artistry, and acting. Another explanation for this finding could relate to the modality of the communication. Previous work on the relationship between creative metaphor and affect has focused on corpora of spoken testimonies and interviews, where participants may be expressing emotion that is rather less 'processed' than what may be expressed in writing. This could give rise to a clearer link between negative affect and creative metaphor. For this reason it would be worth investigating the relationship between metaphor and evaluation in other genres and modalities.

We also found that metaphor was more likely to perform inscribed evaluation than invoked evaluation but when we looked individually at the two types of metaphor (i.e. creative and conventional), we saw that they followed different patterns. Conventional metaphor was more likely to perform inscribed evaluation whereas creative metaphor was equally likely to perform both kinds of evaluation. This may be because inscribed evaluation involves cases where the evaluation is encoded within the word or phrase. This is more likely to be the case for conventional metaphors that have developed to assume a conventional evaluative function, such as the metaphorical use of the word *shattered* shown in Table 15 above. In contrast, invoked evaluation is more implicit, relying on interpretation of the double meanings and entailments in a metaphor, that is, underspecified meanings where the interpretative work needs to be done by the reader. Creative metaphor allows the writer to create their own images and to throw out the meaning in a non-directive way, leaving it to the reader to find their own interpretation, without being constrained by conventional metaphorical mappings.

The results of our analysis call into question the claim made in the SFL literature that metaphor invariably 'provokes' attitudinal meanings. As suggested above, one reason why SFL researchers make this claim may be that they are thinking mainly in terms of creative metaphor. However, in our study we found that even creative metaphors did not only invoke evaluation. Metaphors were involved in a range of evaluative expressions, ranging from very implicit to very explicit. This finding suggests that the four levels in which the evaluative explicitness cline is subdivided in Appraisal theory may need to be rethought, at least for

what concerns metaphor. Metaphor should not be confined to the category of *provoked* attitude. Its function should be interpreted more flexibly and less deterministically, taking into account both the co-text and context in which metaphorical expressions occur. The distinction between conventional and creative metaphor could also be usefully incorporated into the Appraisal framework and used as the basis for a more nuanced account of its evaluative functions.

To sum up, metaphor is an important resource for expressing evaluation. However, our research has shown that the relationship between metaphor and evaluation is complex. It is therefore advisable to consider different types of both metaphor and evaluation when exploring this relationship, as our study has shown that different types of evaluation (i.e. polarity and explicitness) and different types of metaphor (in terms of creativity) may relate to each other differently.

References

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4), 555–596.

Bastian, B. (2017). A social dimension to enjoyment of negative emotion in art reception. *Behavioural and Brain Sciences*.

Bednarek, M. (2006). *Evaluation in media discourse: Analysis of a newspaper corpus*. A&C Black.

Bednarek, M. (2008). *Emotion talk across corpora*. Palgrave.

Bednarek, M. (2009). Emotion talk and emotional talk: Cognitive and discursive perspectives. In H. Prishwa (ed.), Language and Social Cognition: Expression of the Social Mind. Mouton De Gruyter.

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193–216.

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, *46*(5), 423–429.

Cacciari, C., Bolonini, N., Senna, I., Pellicciari, M.C., Miniussi, C. and Papagno, C. (2011). Literal, fictive and metaphorical motion sentences preserve the motion component of the verb: a TMS study, *Brain and Language*, 119: 149-157.

Cameron, L. (2003). *Metaphor in Educational Discourse*. London: Continuum.

Cardillo, E.R., Watson, C.E., Schmidt, G., Kranjec, A. and Chatterjee, A. (2012). From novel to familiar: tuning the brain for metaphors, *Neuroimage*, 59: 3212-3221.

Carter, R. (2004). *Language and Creativity: The Art of Common Talk*. London: Routledge.

Cook, G. (2000). *Language Play, Language Learning*. OUP Oxford.

Deignan, A. (2005). *Metaphor and Corpus Linguistics*. John Benjamins Publishing.

Du Bois, J. W. (2007). The stance triangle. *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, *164*(3), 139–182.

Ellis, B. (1998, June 6). Opinion: What's race got to do with it? *The Sydney Morning Herald*.

Eugenio, B. D., & Glass, M. (2004). The kappa statistic: A second look. *Computational Linguistics*, *30*(1), 95–101.

Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbol*, *2*(4), 239–250.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543–549.

Fuoli, M. (2018). A stepwise method for annotating APPRAISAL. *Functions of Language*, *25*(2), 229–258.

Gibbs Jr, R. W., & Franks, H. (2002). Embodied metaphor in women's narratives about their experiences with cancer. *Health Communication*, *14*(2), 139–165.

Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge University Press.

Goatly, A. (2011). *The Language of Metaphors*. London: Routledge.

Hanks, P. (2006). Metaphoricity is Gradable. In A. Stefanowitsch & S. Gries (Eds.), *Corpora in Cognitive Linguistics vol.1: Metaphor and Metonymy*. Berlin / New York: Mouton de Gruyter.

Hood, S., & Martin, J. (2005). Invoking attitude: The play of graduation in appraising discourse. *Revista Signos*.

Hunston, S. (2010). *Corpus approaches to evaluation: Phraseology and evaluative language* (Vol. 13). Routledge.

Hunston, S., & Thompson, G. (2000). *Evaluation in text: Authorial stance and the construction of discourse: Authorial stance and the construction of discourse*. Oxford University Press, UK.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By* (Edición: 1st). Chicago: University of Chicago Press.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.

Littlemore, J. (2019). *Metaphors in the Mind: Sources of Variation in Embodied Metaphor*. New York, NY: Cambridge University Press.

Liu, F. (2018). Lexical metaphor as affiliative bond in newspaper editorials: A systemic functional linguistics perspective. *Functional Linguistics*, *5*(1), 2.

Lubart, T. I., & Getz, I. (1997). Emotion, Metaphor, and the Creative Process. *Creativity Research Journal*, *10*(4), 285–301. https://doi.org/10.1207/s15326934crj1004_1

Martin, J. R, & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. London & New York: Palgrave Macmillan.

Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Clarendon Press.

Müller, C. (2009). *Metaphors Dead and Alive, Sleeping and Waking, A Dynamic View*. Chicago: University of Chicago Press. Retrieved from https://www.degruyter.com/view/product/535157

Pang, B., & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *ArXiv:Cs/0409058*. Retrieved from http://arxiv.org/abs/cs/0409058

Pragglejaz, G. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, *22*(1), 1–39. https://doi.org/10.1080/10926480709336752

This is the accepted version of an article published in the journal *English Language and Linguistics*. For verbatim quotations and page numbers, please refer to the published version.

QSR International Pty Ltd. (2020). NVivo Qualitative Data Analysis Software [Software], https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Sakamoto, M., & Utsumi, A. (2014). Adjective Metaphors Evoke Negative Meanings. *PLOS ONE*, *9*(2), e89008. https://doi.org/10.1371/journal.pone.0089008

Schubert, E. (1996). Enjoyment of negative emotions in music: An associative network explanation. *Psychology of Music*, *24*(1), 18–28.

Semino, E. (2008). *Metaphor in Discourse* (Edición: 1). Cambridge, UK ; New York: Cambridge University Press.

Simon-Vandenbergen, A.-M., Taverniers, M., & Ravelli, L. J. (2003). *Grammatical metaphor: Views from systemic functional linguistics* (Vol. 236). John Benjamins Publishing.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification. From MIP to MIPVU*. Amsterdam: John Benjamins.

Taboada, Maite. (2011). Stages in an online review genre. *Text & Talk*, *31*(2), 247–269. https://doi.org/10.1515/text.2011.011

Turner, S. L. (2014). The development of metaphoric competence in French and Japanese learners of English. *PhD Thesis*.

Williams-Whitney, D., Mio, J. S., & Whitney, P. (1992). Metaphor production in creative writing. *Journal of Psycholinguistic Research*, *21*(6).