1. Background to the study

Recent years have seen a growing interest in visual or verbo-pictorial metaphor, particularly in advertising and cartoons (cf. Forceville and Uriós-Aparisi 2009, Hidalgo and Kraljevic 2011, Pérez-Sobrino 2016). This increased interest is largely the result of a growing awareness of the powerful persuasive effects of metaphor, particularly when it occurs in visual modes of expression (McQuarrie and Phillips, 2005). In order to analyse the role played by metaphor in visual or verbo-pictorial contexts, some researchers have adapted frameworks and methodologies that are used to analyse metaphor in linguistic contexts. These include, for example, Šorm and Steen’s (forthcoming) development of a method for visual metaphor identification, VISMIP, which is based on Steen et al.’s (2010) method for verbal metaphor identification (MIPVU). Others have proposed new analytical frameworks that take the visual or verbo-pictorial context as their starting point (cf. Forceville 1996, Forceville and Urios-Aparisi 2009, Hidalgo and Kraljevic 2013). However, the field is still largely embryonic and researchers are still struggling to find ways to improve the robustness of their frameworks in order to make their findings more generalizable. As Forceville (2009, 22) points out, there is a greater need for empirical validation across a greater number of examples in order

---

1 This research was funded by a Marie Curie Individual Fellowship (Project ref. EMMA-658079) and by the Ministry of Innovation and Competitiveness, Spain (national project FFI2013-43593-P).
to test the robustness of their hypotheses.

In this paper we tentatively propose a number of principles on which to build a sound methodology for metaphor research in non-verbal contexts. This is an initial set of recommendations, which we hope will be refined as work in this area progresses. In order to formulate these principles, we draw on our experience with three corpus-based studies that we conducted into multimodal metaphor and other figurative operations in advertising, reflecting on the operational decisions that we made during these studies. Our aim is to suggest ways in which metaphor scholars interested in multimodality can better deal with (a) the establishment of a protocol for the identification of metaphor and other figurative operations, (b) the compilation of a representative and diverse corpus of real examples, (c) the issue of inter-rater reliability, and (d) the potential contributions of manual annotation software programs. Although verbal and multimodal metaphor do resemble each other to some extent, we hold (in line with Forceville 1996, 2009) that the different affordances offered by the visual mode mean that a different set of analytical resources and methodological tools is required to study non-verbal metaphor.

This paper consists of four sections. In section 2, we provide an overview of three different studies that we conducted, outlining the challenges that we faced in the four areas listed above. In section 3 we show how these challenges were met, and in Section 4 we highlight how our solutions might enhance the current body of knowledge on multimodal metaphor.

2. Overview of the three studies

In this section we briefly overview the nature and goals of three different studies that we have conducted involving the use of real advertisements containing metaphor and other types of figurative operation (metonymy, hyperbole, etc.). The studies presented below have in common that the fact that they are all based on medium or large corpora of real advertisements, and that at least two researchers were involved in the identification of the multimodal figurative operations under investigation. For each study, we outline the major challenges that we had to address in order to reach the research goal. In Section 3 we explain in more detail the decisions that were taken in each case.

Study 1: Exploring the efficacy of multimodal metaphor and metonymy in advertising
**Goal:** The overall aim of this study was to test whether figurative complexity has an effect on the speed of processing, understanding, and perceived appeal of advertisements in English, Spanish and Chinese (the results from the pilot study for this project can be found in Littlemore and Pérez-Sobrino 2017). This required the compilation of a corpus of 30 advertisements to use as stimuli for the experiments. The study was conducted as part of the Marie Curie fellowship “Exploring multimodal metaphor in advertising” (EMMA 658079).

**Challenges:** It was necessary to compile a balanced corpus of advertisements representing the five types of figurative operations under investigation. These were (in order of increasing complexity): metonymy, metonymic chain, metaphor, metaphtonymy, and metaphoric complex. We then had to design a protocol to identify and characterise the five figurative operations under investigation, and to test inter-rater reliability using the protocol. Advertisements were chosen from the three linguistic backgrounds under investigation (English, Spanish and Chinese).

**Study 2: Statistical study of a larger corpus of advertisements**

**Goal:** The overall aim of this study was to investigate whether different combinations of figurative operations (metaphor, metonymy, and combinations of the two) are favoured in advertisements for different types of products. It also explored whether or not the figurative operations were more likely to occur in the written or pictorial mode in different types of products. It required the compilation of a corpus of 210 authentic advertisements across a range of product types (Pérez-Sobrino 2016).

**Challenges:** We had to find a way to ensure that our corpus was representative and sufficiently diverse to make valid qualitative and quantitative generalisations about metaphor and metonymy and their patterns of use in different kinds of advertisements. Hence, as well as focusing on figurative operations, we also needed to annotate additional features at a more discursive level, such as the type of marketing strategy behind each advertisement, and the use of the different modes of expression in which the figurative operations occurred.

**Study 3: The role of figurative operations in the popularisation of viral marketing campaigns**

**Goal:** The overall aim of this study was to explore the role played by a wide range of figurative operations (correlational and resemblance metaphor, metonymy, verbal, situational and dramatic irony, hyperbole,
and understatement) in the popularisation of viral marketing campaigns. We were interested in establishing whether or not there is a relationship between a video’s tendency to go viral and three factors: the nature of the figurative operations it employs; the figurative density; and the positioning of the figurative operations within the video. We were also interested in exploring whether it made a difference if the figurative operations were contained in the words, the images or the music. This required the compilation of a corpus of 30 audiovisual advertisements, which were then coded for the different figurative operations they contained, the mode in which they occurred, and the point at which they occurred during the video.

Challenges: As well as addressing the methodological issues mentioned above, this study of audiovisual advertisements introduced three new variables: other types of figurative expression, “figurative density” (the number of figurative operations that co-occur), and “positioning” of the identified figurative operations along the advertising timeline. We employed NVivo qualitative analysis software to code and visualise the variables in each advertisement and to extract these data for statistical analysis.

3. Aspects of operationalization in multimodal metaphor research

By way of interim summary, we have claimed that the current challenge in the field of multimodal metaphor is to move away from a qualitative analysis of individual advertisements towards the statistical analysis of larger data sets. In order to do this, we need to identify robust data collection techniques that eliminate researcher bias. The figurative operations within the data sets then need to be identified and coded using reliable and replicable protocols. Finally, the analysis of audiovisual data presents additional challenges, as patterns of figurative operations need to be tracked over time. Due to their complexity, the examination of video advertisements requires a more sophisticated set of analytical tools (similar to those employed in gesture studies, cf. Brenger and Mittelberg, 2015).

a) Developing a Protocol for metaphor identification

Steen et al.’s (2010) MIPVU (Metaphor Identification Procedure Vrije Universiteit Amsterdam) is, to date, the only available set of instructions to identify metaphor-related words in verbal discourse. In order for a word to be defined as ‘potentially metaphorical’, there must be a difference between the contextual meaning of a word and its most basic meaning as
found in the dictionary, and this relationship must be one of comparison. At first sight, it appears to be difficult to transfer this approach to visual metaphor. Even though there are visual dictionaries available in the market (e.g. http://www.visualdictionaryonline.com/), they present decontextualized visual elements (see Figure 1), thereby making it impossible to discern whether a representation of a multimodal element in certain scenario is literal or figurative.

![Figure 1. Entry for “bird” in The Visual Dictionary Online (retrieved on 23rd August 2016)](image)

Nonetheless, the emphasis on the juxtaposition of contextual vs. basic meaning can be transposed to multimodal settings, albeit with some adjustments. Šorm and Steen (forthcoming) establish that in visual metaphors at least one of the two domains is expressed or cued by visual means that show perceptual incongruities. They suggest that this incongruity can be analysed on three different levels: expression, conceptualization and communication. At the level of expression, the analyst must look at entire image (including words and pictures) and describe what it is being represented; at the level of conceptualisation, he or she must decide whether or not there is a metaphor at the intersection of the different modes, and if so, decides whether it is novel or conventional; at the level of communication, he or she must determine whether metaphor is used deliberately or not (a summary of this protocol can be found in www.vismet.org). An alternative approach to the identification and classification of visual metaphor can be found in Forceville (2009: 30) who argues that incongruous meaning in pictures can take three forms: (a) where the image exhibits perceptual similarity with the product it represents, (b) where an unusual placement of an image within a scene prompts the reader to think of it in terms of something else, and (c) where
the simultaneous signalling of two elements prompts new cross-domain mappings. These forms of incongruity are not mutually exclusive.

At the heart of both these approaches, is an understanding that advertisers need to suggest that there is a connection between the product and another entity, and consumers need to work out what this connection is. As Pérez-Sobrino points out (forthcoming), it is the (absence) of internal consistency between the represented element and the context in which it occurs that makes it possible to detect figurativeness in multimodal settings (much in tune with MIPVU for verbal discourse).

The protocols we have discussed so far are designed for the identification of metaphor, but not for other figurative operations, such as metonymy. All three of our studies required an expanded protocol that would encompass additional figurative operations. Here we illustrate the protocol that we developed in Study 1, for the identification and characterisation of metaphor and metonymy, and the interactions between them. In our account, we refer in detail to Figure 2, which is an advert for ice cream produced by a Chinese company. For the sake of clarity, we have translated the advertisement into English.

![Figure 2. Kaku. Pure ice cream](image)

1. Identification of possible target domains. If a consumer saw this advertisement from a distance, and were unable to see the writing in the
bottom right hand corner, they could still infer that the advertisement is for ice cream because the lolly stick helps to select “milk” as a salient component of “cows” over other more prototypical features of cows, such as “meat” (contrast this, for example, with the meaning effects that would be triggered by a barbecue skewer).

2. Identification of possible source domains in the surrounding pictorial context. In this case, our encyclopaedic knowledge of ice cream helps us infer that the ice cream is not made of beef, but rather of the milk produced by cows. In addition, the green and clean natural setting conveys that the milk use comes from healthy and well-fed animals.

3. Metaphoric or metonymic relationship? We formulated a simple set of instructions to characterise the domains identified in the previous steps as metaphoric or metonymic. If the relationship between the image and its meaning could be phrased in an “A IS LIKE B” fashion, we coded it as metaphor. If the relationship could be phrased as “A IS RELATED TO B”, we annotated it as metonymy. This allowed us to label the ice-cream-cow connection as ‘metaphor’, and the stick-ice cream and the milk-cow relationships as ‘metonymy’.

4. Patterns of interaction. If one of the metonymies identified in the previous step was related to another and then to another, we labeled it as a ‘chain of metonymies’, whereas if a “RELATED TO” relationship supported an “IS LIKE” relationship, we labelled it as ‘metaphtonymy’. This was the case in the advertisement under discussion: the COW FOR MILK and the STICK FOR LOLLY metonymies need to be integrated in the metaphorical source and target of the metaphor ICE CREAM IS COW respectively in order to convey the idea that the ice cream is made of pure milk (as the logo says). We also bore in mind the possibility of additional combinations, such as metaphorical complexes, which came to play when more than one metaphor was present in the advertisement.

This is by no means intended to serve as an exhaustive list of metaphor-metonymy combinations. For instance, in Study 2 we found up to 10 distinguishable figurative operations featuring metaphor, metonymy, or a combination of both: metonymy, metonymic chain, multiple-source in target metonymy, multiple-source in target metonymic chain, metaphor, metaphtonymy (featuring variants such as metaphtonymic chain), multiple-source in target metaphtonymy, and metaphorical complexes such as single-source metaphoric amalgams, double source/target metaphoric amalgams, and metaphoric chains (also featuring metonymies within one of the metaphors involved in the chain). Please see Ruiz de Mendoza and Galera (2014) for a full account of the meanings of these terms.
In Study 3 we expanded this annotation protocol to identify alternative figurative operations, as we were interested in operations based on contrast (such as irony) and intensification (such as hyperbole). In order to make room for these alternative operations, we formulated the instructions as follows: if the incongruity between two units, or between a represented object and the one suggested by the pictorial context involved contrast and could be formulated as “A IS OPPOSITE TO B”, we marked it as irony. If the incongruity could be explained in terms of “A IS MUCH MORE THAN B” relationship, we labeled it as hyperbole. In contrast, if the incongruity relied on an “A IS MUCH LESS THAN B” relationship, we marked it as understatement.

b) Compiling a corpus of authentic examples that reflect real-world use

Unlike linguistic metaphor scholars, who are able to use data from language corpora in their analyses, multimodal metaphor scholars have very few corpora of visual metaphors available to them. An important exception is VisMet (www.vismet.org), an online resource of annotated images, that can be used by students and researchers from different academic fields interested in visual rhetoric. The current version, VisMet Baby, contains around 350 images (advertising, art, and political cartoons). VisMet fills an important gap in multimodal metaphor studies, as it is the first large-scale accessible corpus of annotated visual metaphors.

The corpus of advertisements provided by VisMet was very useful in the selection of English examples for Study 1. We respected the annotation provided by the authors when possible; however, we refined their annotation whenever we detected alternative operations according to the protocol outlined in the previous section. These usually took the form of figurative complexes (that is to say, if we detected a metonymy involved in the metaphor annotated by VisMet coders, we annotated it as metaphtonymy in our study). For the collection of Spanish and Chinese advertisements, we established the following set of instructions: (a) they

---

2 The reader can find a refined set of instructions in Pérez-Sobrino, Littlemore and Houghton (forthcoming) where we show how irony was further subdivided into “verbal”, “situational”, and “dramatic”. We also identified cases of contrast without pragmatic effects (as opposed to contrast leading to irony). These were labelled as “unmarked contrast”. Finally, cases of metaphor were divided into ‘resemblance metaphor’ and ‘correlational metaphor’, following Ruiz de Mendoza and Pérez-Hernández (2011).
had to contain text and images, (b) the text had to be originally written in Chinese and Spanish, respectively, (c) there must be some incongruity in the picture, or between the picture and the text, and (d) it must be possible to understand the advertisement without a high degree of specific cultural knowledge. This led us to exclude advertisements for products that do not exist in other countries, ones that contained obscure cultural references, and ones that contained word puns. All adverts were subject to a pre-test with native speakers from each country. We continued to collect advertisements until the three native speakers participating in the selection (one from China, one from Spain, and one from UK) agreed they were perfectly understandable. We collected 60 adverts (20 per nationality) that met these criteria, but eventually we only used 30 in the experiment due to time constraints.

When it comes to making generalizations about metaphor in real discourse, the researcher must ensure that the corpus is fully representative of the type of advertisements about which he or she wishes to make the claim. This was the aim of Study 2, as here we were interested in finding relationships between metaphor and metonymy and other advertising variables such as the type of product that was being promoted. In order to minimise the analyst’s impact in the collection of 210 advertisements, we focused on two factors that were deemed to play a relevant role in the sampling of a corpus of study: diversity, i.e., how faithfully it represents all the possibilities present in real discourse, and representativity, i.e., how reliably the selection avoids analyst bias during the sampling. Let us illustrate this with some examples from the study.

Product type has been found to be an influential factor in determining the effectiveness of metaphor advertising (Ang and Lim 2006, Chang and Yen 2013). Following Kotler and Armstrong (1997), we placed the products into seven categories: convenience (e.g. bread, cereal, and magazines), shopping (e.g. appliances, furniture, and clothing), and specialty (e.g. diamonds, luxury cars) (Copeland 1924), unsought goods (encyclopedias of fire extinguishers, Perreault and McCarthy 2002), “tangible” (e.g. fitness centers, laundry) and “intangible” (e.g. education, broadcasting) services (Bhattacharya 2006: 83). We regarded NGOs and charities as a separate category from intangible services as they do not aim to sell an idea, but rather to denounce a situation. Each of these categories appeared to have a different marketing strategy, and this was the focus of our study. In order to make the selection more independent from the researcher, we selected only every the third advertisement found per product type.

Finally, for the selection of videos for Study 3, we took into account
two factors: the number of views and the type of product. In order to study the effect of figurative language on the popularisation of audiovisual advertisements, we had to match a popular video with a non-popular video of the same product. However, marketing experts have not yet identified the number of views from which a video can be considered popular or viral, as this highly varies across countries and years. For the sake of practicality, we established a fictional division between non-popular (those who reached less than 500,000 views) and popular advertisements (from a 1,000,000 views onwards), and collected pairs of popular and non-popular videos of the same brands and products. Our final corpus contained a total of 30 advertisements (15 advertisements in each category), with a number of views ranging from 1728 to 115,904,727.

c) Reliability and introspection

When dealing with multimodal metaphor identification, it is natural to cast doubt on the potential of one analyst’s conscious processing to accurately reflect unconscious mental processes distributed across the population as a whole. One way of addressing this problem is to have more than one rater with a reasonable degree of inter-rater reliability.

The only precedent in reliability testing in metaphor identification is to be found in the work carried out by the Pragglejaz group (2007). They examined reliability in metaphor identification in a separate series of specially conducted methodological studies using Cohen’s and Fleiss’s Kappa (Dunn 1989). Because we were focusing on the number of figurative operations identified and the degree of figurative complexity, we employed Cronbach’s Alpha, which tests the reliability of numerical scales, rather than the Kappa coefficient, which is used for categorical data.

We annotated the examples individually in two rounds. We calculated the inter-coder reliability for metaphor, metonymy and figurative complexity. We found a fair degree of agreement for the identification of metaphor (0.716), which is commonly agreed to be an acceptable degree of internal consistency in studies of this kind (DeVellis 2012). This parallels the result obtained by the Pragglejaz Group when their results were observed in pairs (0.7 for the news texts). Even though we are aware that there is plenty of room for improvement, this finding suggests that our protocol has proven useful way of identifying figurative operations in advertisements.

In contrast, the degree of inter-rater agreement for the annotation of metonymy was significantly lower, 0.496. The fact that metonymy relates
two inter-connected elements that belong to the same domain means that it is often difficult to spot the figurative connection, as it sometimes overlaps with the literal language (Littlemore 2015, 126-127). Whereas Cohen’s Kappa for assessing the reliability has been reported in verbal metonymy identification (Markert and Nissim 2003), it still remains a challenge for non-verbal metonymy, and metonymy-based figurative operations (such as metonymic chains and metaphonymy). After this first check, we identified the cases of disagreement and attempted to resolve these through discussion.

The higher complexity of Study 2 (in terms of the annotation scheme and the larger amount of advertisements in the corpus) made it difficult to expect an acceptable degree of inter-coder reliability. One rater annotated the whole corpus individually and two external expert researchers supervised the annotation on a case-by-case basis. Although this system relies more on introspection by a single researcher, the live discussion of each of the cases between the three researchers involved helped to refine incorrect accounts, and disregard controversial cases for which full agreement between the three was not reached. This system of external refereeing may help to endorse the manual annotation of multimodal metaphor (and other figures) in large corpora with greater consistency, which might sometimes be at stake in introspection-based analysis.

We opted for a similar decision in the annotation of the examples in Study 3. The audiovisual nature of the examples studied made it very complicated to compare two individual annotations of the examples, especially when deciding upon the positioning of the figurative operations throughout the advertisement. We decided that the best way to analyse the advertisements in a consistent way was through joint annotation on a case-by-case basis with a third analyst being brought in to help decide in ambiguous cases.

d) Reliance on specific software (NVivo)

The analysis of videos brings two new variables into play: sound and time. Sound and music can operate on their own as modes of representation (see Pérez-Sobrino 2014 for an account of musical metonymy) or appear in conjunction with other texts or images to represent a certain idea. The increasing number of factors at work in an audiovisual advertisement makes it difficult for the analyst to access them online. Some of them might be elusive, or happen in very short period of time, or co-occur to the extent that may escape the analyst’s awareness. Nvivo is a convenient tool that can be used to reduce the difficulties in annotating multimodal, audiovisual
data, which also allowed for the extraction of the data for more sophisticated statistical analyses.

We first transcribed each of the 40 videos selected for Study 3 and imported both the advertisement and the transcription into Nvivo (Figure 3). In order to ease the annotation task, we split each video into smaller units at the sentence level, or in places when visible changes in the visual content occurred. We also added the exact time span for each unit (in milliseconds) to ensure the exact coverage of each unit with respect to the total duration of the advertisement. Each scene was later coded for figurative operations and use of modes.

Figure 3. Annotation of figurative operations in “Old Spice. The man your man could smell like” advertisement with Nvivo.

The output of the annotation can be seen in Figure 4. In this dynamic representation of figurative operations, we can visualise three new aspects of analysis: coverage, density, and positioning. Coverage refers to the percentage of the advertisement that a figurative operation covers. For example, situational verbal irony (in pink) covers 76.01% of the advertisement, whereas visual hyperbole (in blue) covers over 10.9%. Coverage was a crucial aspect of our study, as we hypothesised that operations based on contrast (alone or in conjunction with hyperbole) determined the popularisation of advertisements.
Figure 4. Dynamic representation of figurative operations in “Old Spice. The man your man could smell like” advertisement with Nvivo

Another novel element that emerged in our analysis of the videos is **figurative density** (after Müller 2008). This concept describes the number of different figurative operations occurring at the same time. It can be calculated by counting the number of segments of different colour that are represented one below the other for each unit of analysis (represented by short white breaks; see more on the rationale behind the division of the advertisement in narrative units below). Figure 4 shows that figurative density is much higher around the middle of the advertisement, where situational verbal irony (pink), verbal metonymy (yellow), situational verbo-pictorial irony (green), and unmarked verbal contrast (purple) are annotated simultaneously, than towards the end of the timeline, where only two operations are annotated (visual hyperbole, in blue; and verbal metonymy, in yellow). Finally, this kind of dynamic visualisation allowed us to determine whether the positioning of figurative operations also played a role in the popularisation of a video. We anticipated that the accumulation of figurative operations towards the end of the advertisement would contribute to the creation of a feeling of surprise, thus making the advert more likely to be perceived as innovative and worth viewing. In order to compare the weight and position of figurative operations across the 40 timelines we retrieved from Nvivo, we divided each timeline into four quartiles, and divided each quartile into four parts. We then counted the number of figurative segments per quartile in order to quantify the amount of figurative density in each part of the advertisement. For example, as Figure 5 shows, the quartile with higher accumulation of figurative units is the second, with 14 figurative segments.
This system allows the analyst to draw comparisons concerning the weight of figurative density within each quartile within and across advertisements. The results of Study 3 (Pérez-Sobrino, Littlemore and Houghton, forthcoming) demonstrated that the presence of figurative operations in either the first or the fourth quartile was significantly related to the popularity of the advertisement ($p < 0.01$), whereas the presence of figurative operations in the second and third quartiles bore no relationship to an advertisement’s popularity.

5. Final remarks

Over the course of this paper we have shown different ways to increase the validity and reliability of studies of multimodal metaphor, based on introspection. We summarise below three lessons learned from our experience in conducting multimodal metaphor research:

1. Our protocol for identification ensured a reasonable level of reliability for multimodal metaphor (0.716), but not metonymy (0.496). The individual annotation by two or more researchers must be complemented by live discussion of the examples or the joint annotation by the researchers in order to achieve greater granularity in the analysis.

2. The compilation of a real corpus of examples has to be diverse and representative. In order to make claims concerning the behaviour of metaphor in real discourse, researchers must ensure that the corpus closely resembles the type of data about which claims are going to be made, and that it reflects all the different facets that can be encountered.

3. The use of specific software (such as Nvivo) is suitable (and necessary) for the analysis of audiovisual data. It allows for the annotation of...
figurative operations in the advertisement timeline (in milliseconds) and allows the analyst include new variables in the analysis, such as coverage, density, and the positioning of figurative operations.

Future studies could usefully investigate the issue of inter-rater reliability in multimodal metaphor research, not as an intermediate step to validate analyses, but as a research subject in its own right. It would be interesting to discover why, for example, multimodal metonymy attracts lower inter-rate reliability scores than metaphor. One could also compare the reliability scores received by different items within a single study and seek to identify features of those items that may explain why they provide so much disagreement. It is also worth examining whether metaphorical complexity (in terms of more complex metaphors, or images rendering metaphor) correlates with inter-rater reliability difficulty\(^3\). A more serious consideration of inter-rater reliability is required in order to improve and refine identification protocols, although the complexity of naturally-occurring multimodal data means that this can be a very difficult challenge.

References


\(^3\) Bodo Winter, p/c.


Pérez-Sobrino, Paula, Jeannette Littlemore, and David Houghton. (forthcoming) The role of figurative operations in the popularisation of advertising campaigns.

**Secondary references**


2) *Koku. Pure ice cream*
Advertising Agency: Brand 8, Beijing, China
Published: May 2012
Source: https://adsoftheworld.com/media/print/kaku_cow

3) *Old Spice. The man your man could smell like*
Advertising agency: MJZ
Date of release: February 2010
Source: [http://es.coloribus.com/archivo-de-publicidad-y-anuncios/tv-](http://es.coloribus.com/archivo-de-publicidad-y-anuncios/tv-)