Chapter 7. Visual metaphor versus verbal metaphor: A unified account

Francisco Yus

Abstract

Multimodal metaphors are those “whose target and source are each represented exclusively or predominantly in different modes” (Forceville 2006: 384), mainly with a verbal-visual interface of source and/or target. When multimodality is analyzed in metaphors, the verbal and visual inputs are often treated as different phenomena demanding different interpretive strategies when searching for a metaphoric interpretation. In this chapter, on the contrary, it is claimed that the comprehension of verbal, visual and multimodal metaphors involves similar mental procedures. Although the perception of images differs from linguistic decoding, reaching an interpretation of metaphors entails similar adjustments of conceptual information of texts and images and multimodal combinations, regardless of the modal quality of the input.

Keywords: Relevance theory, conceptual upload, ad hoc concepts, ad hoc pointers, visual-conceptual interface

1. Introduction

In this chapter, visual metaphor comprehension is compared to verbal metaphor comprehension and analyzed mainly from a cognitive pragmatics point of view (specifically within “relevance theory,” as proposed by Sperber and Wilson 1995 [1986]), but also with reference to cognitive linguistics where necessary. The main claim underlying this chapter is that the comprehension of verbal and visual metaphors involves similar mental procedures. Although the perception of images is obviously different from linguistic decoding, reaching an interpretation of visual metaphors also entails an adjustment of conceptual information – a stage during compre-
hension that will be called *conceptual upload* – in the same way as verbal metaphors.

Therefore, although it is not denied that the combination of visual and verbal inputs in multimodal metaphors can indeed generate interesting interpretive outcomes, as the chapters in this book demonstrate, in many studies of multimodality visual and verbal metaphors are often treated as different phenomena. By contrast, in this chapter it will be argued that both types of metaphor (and also multimodal metaphors with combinations of text and image) are “decoded” by specialized mental modules, which deliver schematic information that has to be enriched inferentially in order to obtain the intended interpretation. Besides, as will be illustrated below with several Spanish political cartoons by El Roto, visual metaphors can also be arranged on a scale depending on the gap existing between the prototypical referent of the image and the cartoonist’s intended referent, which has to be adjusted inferentially in the same way as in verbal metaphors, whose interpretation involves the hearer’s inferential adjustment of the concept that the speaker encodes in order to obtain the speaker’s intended interpretation.

In this sense, the cartoons analyzed in this chapter contain metaphors of a pictorial/visual nature, whereas multimodal metaphors, as defined by Forceville, are “metaphors whose target and source are each represented exclusively or predominantly in different modes” (Forceville 2006: 384). However, it follows from my central thesis in this chapter that the model presented here is capable of accommodating multimodal metaphors, in the same way as metaphors with only verbal or only visual inputs.

2. **Relevance theory and ad hoc concept formation**

Sperber and Wilson’s (1995 [1986]) *relevance theory* (henceforth RT) predicts that human comprehension follows two stages:

i. Following a path of least effort, test interpretive hypotheses (disambiguations, reference resolutions, enrichments, implicatures) in order of accessibility.

ii. Stop when the interpretation satisfies the current expectation of relevance.

For instance, an advertisement by London Transport quoted in Tanaka (1994) only contained the text “Less bread, no jam.” The reader of this ad will follow a path of least effort and conclude, initially, that “bread” and
“jam” have to do with food, but faced with the incongruity between “food” and the advertised “transport company,” the reader will continue testing interpretive hypotheses concluding, at a second stage, that bread is a colloquial word for “money” and jam refers to “traffic jams.” The reader will now be satisfied at this interpretation and stop processing here.

Specifically, for RT, comprehension does not normally start in a communicative vacuum, but takes place against a context of previous utterances whose interpretation (stored in the short-term memory) works as a background against which new information is processed. Cognitive linguistics has also dawn attention to the role of context for the right comprehension of metaphors. For example, for conceptual metaphor theory, metaphors may be activated as part of the hearer’s understanding of context, and this will make the interpretation of metaphors easier at subsequent stages in discourse. Similarly, conceptual blending theory stresses the role of context in metaphor comprehension: “because cognitive activity mediates the relationship between words and the world, the study of meaning is the study of how words arise in the context of human activity, and how they are used to evoke mental representations” (Coulson, quoted in Tendahl and Gibbs, in press).

This “cumulative” background context of previous utterances in the conversation is normally absent in the processing of visual metaphors, which are inserted in media discourses such as newspapers, billboards or magazines, and hence the viewers have to interpret them from scratch, without this readily available “short-term memory store” of information. This does not mean that visual metaphors do not require a great deal of background knowledge for their satisfactory interpretation. The metaphors used in cartoons, such as the ones analyzed in this chapter, are often related to recent news-worthy events whose knowledge is essential to get the right extent of the metaphoric mappings (cf. El Refaie this volume; Schilperoord and Maes this volume; Forceville 2005).

RT predicts two clear-cut phases during interpretation: one of decoding and one of inference. The first one is in charge of the language module of the mind (Fodor 1983), which apprehends a linguistic sequence and yields a de-contextualized but grammatical “logical form” which has to be enriched in order to be meaningful. By contrast, cognitive linguistics disregards modularity in favor of what is called embodied-mind hypothesis, according to which “the same neural mechanisms used in perception and bodily movement play a role in all forms of conceptualization, including
the creation of lexical fields and abstract reasoning” (Ruiz de Mendoza, 2005: 36).

For RT, language does not encode thoughts, but only clues that help the hearer access the speaker’s thoughts, which are often more complex than the literal meaning encoded by the utterances. For example, the (b) versions of the following utterances are closer to the thoughts that the speaker intended to communicate with them than the schematic (and communicatively useless) (a) versions, the ones actually uttered:

(1)  
   a. It’s too wide.  
   b. The table that I bought yesterday is too wide to go through the door.

(2)  
   a. The cinema is some distance from here.  
   b. The cinema is too far to go walking.

Secondly, and following a relevance-seeking criterion, the hearer’s mind undertakes an inferential process of mutual parallel adjustment of explicit content, implicatures and context (including the information from preceding utterances) until a satisfactory interpretation is achieved, at which point processing stops. In my opinion, despite the apparent differences (see El Refaie 2003: 85-90), this model of utterance interpretation is applicable to visual and multimodal metaphor comprehension in the same way as to verbal metaphor comprehension.

As pointed out above, there is always a greater or lesser informational gap between what the speaker says (what is encoded) and what the speaker intends to communicate with the utterance. This task often involves an adjustment of the conceptual information encoded, that is, interpretation involves the creation of ad hoc concepts during interpretation (see Carston 2002; Pilkington 2000). Since we store many more concepts in our mind than words to encode them, inevitably there is a greater or lesser amount of adjustment of encoded concepts needed in order to grasp the speaker’s intended interpretation. This is applicable to almost every concept, not just the adjustment of concepts regarding metaphor comprehension. Consider, for instance, the examples provided in (3) below (Vega-Moreno 2004: 317):

(3)  
   a. The sofa is soft.  
   b. Baby skin is soft.  
   c. The cat is soft.
The hearer of (1a-c) is expected to adjust the encoded concept “soft” into a more appropriate and contextualized type of softness that specifically applies to sofas, skins and cats respectively, that is, adjust into more relevant *ad hoc* concepts SOFT*, SOFT**, and SOFT***.²

*Ad hoc* concepts can be formed either by a process of narrowing of the encoded concept (what the hearer interprets is more specific than the encoded concept) or a process of broadening of the encoded concept (the hearer interprets more general or broader information than that encoded), all of them resulting from a relevance-seeking procedure.³ These processes are illustrated in (4) and (5) respectively:

| (4) | a. The fish attacked some people near here (FISH* = dangerous fish, e.g., sharks).  
    | b. The boy has a temperature (TEMPERATURE* = higher temperature than normal).  
    | c. It will take some time to fix the car (SOME TIME* = longer than it usually takes). |
| (5) | a. The steak is raw (RAW* = undercooked).  
    | b. She is a genius! (GENIUS* = not literally a genius, but having some of his/her qualities).  
    | c. It was quiet in the street last night (QUIET* = with very little noise). |

Within RT, verbal metaphor comprehension also involves a pragmatic adjustment (broadening, narrowing or both simultaneously), and RT suggests a different approach to metaphor analysis than do cognitive linguistics. In short, the former is more interested in the role of metaphor for communication and hence in the role of context favoring a certain adjustment of concepts, while the latter mainly focuses on the cognitive motivation for certain metaphors, their conceptual organization and the inference patterns involved in their creation. But, as such, both accounts are not mutually exclusive (see Ruiz de Mendoza and Pérez Hernández, 2003; Ruiz de Mendoza, 2005; Tendahl and Gibbs, in press).

In Vega-Moreno (2004: 208), three main types of *ad hoc* concepts are proposed for verbal metaphors:

(a) *Ad hoc* concepts which contain qualities which are applicable to *all* the prototypical referents of the encoded concepts and also to a range of other referents, as in (6) below:

| (6) | A. Why does your boyfriend want you to go with him everywhere? |
B. Because he is a baby.

(BABY* denotes a person who cannot be independent, cannot look after himself, can’t do things alone, etc. These are qualities applicable to all babies (as prototypical referents) and also to some adults such as the speaker’s boyfriend).

(b) Ad hoc concepts which contain qualities which are applicable to some of the prototypical referents of the encoded concepts and also to a range of other referents, as in (7) below:

(7) Being the only boy, Dave has always been the prince of the house.

(PRINCE* denotes a subset of princes who are spoilt and do as they please, as well as a set of young boys who are not princes but are spoilt and do as they please).

(c) Ad hoc concepts which contain qualities which are applicable to none of the prototypical referents of the encoded concepts but are applied to other referents, as in the utterance quoted in (8):

(8) I tried to persuade him to change the essay topic but there was no way. He is an iron bar.

(IRON BAR* denotes people who are difficult to convince, persuade, etc., qualities which are not found in iron bars as prototypical referents).

In my opinion, these three cases are not only inherent to verbal metaphor comprehension, but are also found in the processing of visual metaphors. In this sense, case (c) is interesting because it gives rise to the so-called emergent features or emergent properties which apparently do not belong to the target domain of the metaphor but seem to emerge during comprehension (Gineste, Indurkhya and Scart 2000; Wilson and Carston 2006). These emergent properties might appear to be found only in the interpretation of verbal metaphors but, as will be argued below, they are also frequent in visual metaphor comprehension (cf. Yus 2003a) and in any multimodal combination of text and image.

Many explanations have been suggested for the creation of these emergent properties. I will follow an interesting proposal by Vega-Moreno (2004) within a relevance-theoretic point of view, and I will argue that this proposal is applicable, in a similar way and with the necessary adjustments, to the processing of emergent properties in visual or multimodal metaphors. More generally, I will show to what extent conceptual assessment is in-
involved in visual metaphor comprehension, basically through what will be called stable versus innovative conceptual upload. The analysis will be divided into several steps that the reader is expected to go through during the interpretation of a visual metaphor. Comparisons with verbal metaphor comprehension will be made where necessary, and there is an inherent claim in this proposal: that combinations of text and image in multimodal metaphors demand similar interpretive procedures involving conceptual adjustment of encoded information. Variations are only found in the way schematic information is obtained by mental modules (in a more holistic way in pictures, in a more linear way in texts).

3. Stages in visual metaphor comprehension: A proposal

3.1. Perception: Visual versus verbal

My claim is that processing visual metaphors does not differ substantially from processing verbal metaphors, the main difference being the way in which the input is transferred to the central inferential processor. RT predicts, following Fodor’s (1983) theory of the modularity of mind, a context-free decoding of a linguistic string by the language module, which sends a de-contextualized string of linguistic information to the central processor in order to be enriched inferentially into a fully contextualized (and optimally relevant) interpretation that supposedly matches the speaker’s intended one.

Visual information, on the other hand, is decoded by another module: the perceptual module. The language module and the perceptual module share similar properties: (a) they are fast and automatic (i.e., they are capable of a high-speed transference of information, and they are automatically activated by the appropriate type of input: “linguistic” and “visual” respectively); (b) they are domain-specific (both modules are only activated by a specific type of input); (c) they are part of our genetic endowment (i.e., they are not “learned” and possess an evolutionary quality); and (d) they have a uniform path of development (unfolding) across individuals and cultures. In short, these mental modules get activated automatically when the appropriate type of input reaches them, and both yield de-contextualized pieces of evidence of the sender’s intention to communicate some information. This context-free information is then enriched inferentially in order to obtain a fully satisfactory (i.e., relevant) interpretation of the verbal or visual input.
However, these qualities of modules do not entail that no choices are ever made during this phase of verbal or visual decoding. For instance, the language module often has to choose between two possible logical forms for the same linguistic string. In the same way, it has been demonstrated that although the perceptual module seems to engage in a one-to-one matching between object and referent, it also has to make choices as to what visual information it is actually processing. Specifically, when readers interpret a visual metaphor, they start by perceiving the image, that is, by identifying the visual input. This is done through a subconscious or subattentive comparison with previously stored information on the visual attributes of the object or objects depicted (see McMahon 2003; Kriegel 2004). When the image is supposed to be intentionally communicated to the readers beyond a simple perceptual recognition, processing moves one step beyond into a more conscious stage of interpretation, loaded with inferential activity (see below). This mental storage of prototypical referents that we possess is made up of two basic types of information which undergo a constant process of updating and stabilization through subsequent visual perceptions:

(a) Prototypical visual referent: encyclopedic entry containing visual elements and attributes that an item depicted in an image is typically made of. For instance, the prototypical referent of an image of a cat would contain visual attributes that are stored as typical of cats (type of hair, colors, ears, whiskers, paws…) and which allows for an easy visual identification.

(b) Prototypical visual syntax: other items typically associated with another object depicted in an image. In general, processing is faster if the visual arrangement of objects in the image fits our storage of prototypical visual syntax for these objects, a sort of “visual schema” that precedes and influences actual perception (cf. McMahon 2003: 266). In the above “cat example,” we will expect to find visual representations of this animal in specific scenarios with objects forming a prototypical visual syntax regarding its representation (e.g., cat on a branch, on a mat, playing with wool…).

In general, as the number of visual features of the image which belong to the prototypical visual referent increases and its prototypical visual syntax fits stored schemas, the effort involved in its processing will decrease accordingly. Highly iconic images are normally filled with features fitting the prototypical visual referent of the image that the reader possesses, but there can be other images containing less prototypical features, generating so-called scales of iconicity.
Besides, visual perception shares a bottom-up and a top-down quality. It is bottom-up because the reader constructs and integrates the prototypical visual referents from the available visual elements (as claimed by Gestalt theory). But at the same time it is top-down because readers test the visual input against their mental storage of prototypical visual referents, anticipating, as it were, and even influencing the recognition of the object depicted in the image.\footnote{8} The perceptual module draws on a conceptual repertoire that contains a range of visual referents and is subject to constant revision and updating through subsequent visual perceptions of similar images. Perception is never isolated, and past exposure to objects constrains future perceptions (Villafañe and Mínguez 1996: 100). Hence, each perception of the physical object or picture to which the referent is associated helps the reader to update the prototypical referent that was created when the reader first perceived it.

3.2. Ad hoc pointers

The previous section on visual perception is important to understand why certain images are interpreted metaphorically. In my opinion, the key to a shift from a purely denotative interpretation of the image, often sub-attentive, to a connotative metaphoric interpretation, loaded with inferential processing, lies in the detection of an incongruity that turns up between the activation of the stored prototypical visual referents during perception and the actual visual configuration of the image or images making up the visual metaphor (Forceville 1996: 115). This kind of incongruity has been labeled ad hoc pointer (Yus 2005), in the sense that an ad hoc visual arrangement or configuration created by the author with specific communicative purposes points towards a connotative interpretation, alerts the reader towards a connotative interpretation, often a metaphorical one. This idea entails an increased mental effort in moving beyond a sub-attentive visual perception into an effort-demanding inferential activity in search of the right features that can be transferred metaphorically from what we can label “the source image” to what we can call “the target image.”

Of course, as we enter this inferential phase, the reader takes the responsibility of grasping the intended metaphoric interpretation (or his/her own personal interpretation) and the author of the image can only hope that the reader will be able to select the appropriate encyclopedic features associ-
ated with the visual referents of the images and infer which are the ones involved in the metaphoric interpretation (El Refaie 2003: 81; this volume).

Incidentally, there may be visual ambiguity, in which a metaphoric interpretation of the image is intended but a purely denotative interpretation is also valid, that is, occasions on which there is no apparent “ad hoc pointer” and hence the metaphoric interpretation may not be accessed. In these cases, it is the readers’ search for an optimally relevant interpretation that will guide them beyond a purely denotative interpretation. In general, visual metaphors are integrated in other discourses (images in advertisements, cartoons in the press) and the readers know that these images are intended to communicate specific, non-denotative information, and hence they will not be cognitively satisfied at a purely denotative level.

3.3. Visual-conceptual interface

Upon detecting the *ad hoc pointer*, the reader of the image(s) enters another stage in interpretation, which I will call **visual-conceptual interface**, in-between a sub-attentive perception of the images and a fully inferential extraction of a relevant connotative (i.e., metaphoric) interpretation of the image. At this stage, the reader aiming at an optimally relevant interpretation has to raise a number of preliminary hypotheses concerning the intended relationship that holds between the depicted images and the encyclopedic (conceptual) information stored about the referents of these images, mostly of a stereotypical quality. In short, the readers would ask themselves questions such as the following:

1. Which are the two images related metaphorically? Are both present in the picture? Visual metaphor involves a mapping of information transferred from one image to another, which we have called *source image* and *target image* respectively. Often both images are present in the picture (either fused together or separated) but sometimes one of them – normally the source image – is absent. Therefore, there are different degrees of mental effort involved in processing visual metaphors depending on whether both the *source image* and the *target image* are depicted in a metaphoric visual configuration, or one of the images is absent and is only accessible through an inferential operation regarding the encyclopedic information on its prototypical referent. At the same time, some mental effort has to be devoted to identifying the *source* and *target images* in the first place, which are not always clearly distinguishable, even when both images are present.
2. What kind of visual arrangement is there between the images? The reader is also expected to infer what relationship holds between the previously identified source image and target image. Is the target image supposed to be like the source image? Is it opposed to the source image? (cf. Phillips and McQuarrie 2004.)

3. Are the prototypical encyclopedic referents of the images themselves the ones that are going to undergo inferential adjustment in order to obtain a metaphoric interpretation or do the images stand for a different encyclopedic referent? I believe that visual metaphor comprehension, in a similar way to verbal metaphor comprehension, also involves an access to and adjustment of conceptual information stored in or attached to the encyclopedic prototypical referents of the image or images depicted. This implies that it is of utmost importance to determine whether the author intends the most accessible referents of the images to undergo metaphoric processing or whether the intended sources of metaphor have to be found elsewhere. This is the case of images that stand metonymically for other referents. For instance, in the visual metaphor described in (9), it is the prototypical referent of the image “buildings” that works as target domain for metaphoric mapping, whereas in (10), the visual referent of “book page” is not expected as target domain, since it stands for “culture in general,” which is the intended target domain.


Similarly, in a cartoon by El Roto (figure 1), a syringe is depicted with a television tower instead of the needle. An incongruity in the visual syntax of the image works as an ad hoc pointer alerting the readers to a metaphoric interpretation. The readers’ search for relevance will lead them to dismiss the conceptual features of the prototypical referents of “syringe” and “television tower” as the ones undergoing metaphoric assessment, and they will probably infer, instead, that “syringe” stands metonymically for drugs in general, and that “television tower” stands metonymically for television in general as a mass medium, and the conceptual features of “drugs” and “television” are the intended source and target of this visual metaphor. The cartoon also includes the text “la gran droga” (the big drug), which works
as an anchorage (in Barthes’ 1977 sense) of the image facilitating the metonymic relationship between syringes and drugs in general. Using the cognitive linguistics terminology that distinguishes source-in-target metonymies and target-in-source metonymies, in this case we would, instead, encounter examples of source-in-source metonymy and/or target-in-target metonymy between prototypical referents depicted and the intended referents. Specifically in figure 1, both the source image (tower) and the target image (syringe) stand metonymically for the actual source and target referents undergoing metaphoric interpretation (television/drugs).

Figure 1. Cartoon by El Roto, El País, 6 June 2003.

3.4. Conceptual upload

The ad hoc pointer and the preparatory phase of visual-conceptual interface lead to a fully inferential stage in the processing of the image or images, which will be called conceptual upload. Since this stage is centered upon the inferential assessment and adjustment of conceptual information attached to the prototypical encyclopedic referents of the images (or the referents intended through metonymy), the distinction between visual and verbal input to metaphoric interpretation no longer matters (we are now at a cognitive, fully inferential phase of interpretation). The reader has now entered a fully inferential stage that takes either the information from the verbal utterance or the identified visual images as blueprints or clues for an optimal metaphoric interpretation.

At this stage of conceptual upload, and following a relevance-oriented path of accessibility, the reader will compute conceptual features stored in
the encyclopedic information of the intended referents of the images (either the prototypical referents of these images themselves, or the referents to which these images point metonymically, as commented upon above) and will try to find the ones that can be applied to the other image, a mental procedure which can be called ad hoc choice of image-associated conceptual features. This is a similar inferential activity to the one intended to obtain metaphoric mappings in verbal and multimodal metaphors.

Sometimes this assessment of possible ways in which the images can be related does not result in any metaphorical outcome despite the visual incongruity because the reader is unable or unwilling to find any metaphoric connections between these images. On other occasions, though, the reader’s search for relevance will lead to a metaphoric interpretation and to a selection of features which can be transferred from the source image to the target image. In this sense, two possible types of conceptual upload can be identified:

(1) Stable conceptual upload of image-associated conceptual features takes place when interpreting the visual metaphor involves an adjustment of one or several features belonging to the prototypical encyclopedic referent of the image depicted. The maker of a visual metaphor fitting this type will expect all the readers to have a similar store of conceptual information filling up the prototypical referent of the image or images depicted. But, crucially, in this type of conceptual upload the encyclopedic feature of the referent associated with the source image maintains its conceptual stability, and the reader only has to broaden, as it were, its denotation to include the referent associated with the target image.

The types 1 and 2 of ad hoc concept formation that were introduced in section 1 above for verbal metaphors would belong to this type of conceptual upload:

Type 1. When one or several conceptual features of the prototypical encyclopedic referent associated with the source image that the author intends the reader to apply to the target image can be found in all the prototypical referents represented in the source image. In this case, by means of a process of conceptual broadening, an ad hoc CONCEPT* is created and applied to the target image. For instance, in the aforementioned cartoon by El Roto depicting a lorry throwing buildings into a huge garbage dump, one or several of the features of the prototypical encyclopedic referent “rubbish” (for example “being useless,” “having no quality,” etc.) are mapped onto the encyclopedic referent “buildings” associated with the target image. These conceptual features form an ad hoc concept RUBBISH*, which is the result
of broadening the prototypical concept “rubbish” in order to cover the unusual referent “buildings” of the target image.

Type 2. One or several conceptual features of the prototypical encyclopedic referent of the source image that the author intends the reader to apply to the target image can be found in some but not all of the prototypical encyclopedic referents represented in the source image (or referred to by this image). Again, a process of conceptual broadening is required so that an ad hoc CONCEPT* is created.

For example, in another cartoon by El Roto (figure 2) the reader can see a man on a surfboard sliding on a huge wave, but the wave is made of buildings, instead of water, and this unusual wave looks as if it is about to break on the surfer. The author probably intends to communicate metaphorically that the urge to build houses in Spain (the construction wave or construction bubble) will eventually break (or burst) and harm us in the same way as the huge wave is about to break on the surfer and probably harm him as well. The harming quality of breaking waves can be found in some (but not all) of the prototypical encyclopedic referents of the image depicted, specifically only those waves which are big enough to break onto surfers and harm them, and a new ad hoc concept WAVE* is created as a result of an adjustment (broadening) in order to fit the new encyclopedic referent “construction bubble” of the target image.

Figure 2. Cartoon by El Roto, El País, 19 October 2003.
Innovative conceptual upload of image-associated conceptual features. No conceptual features that the author intends the reader to apply to the referent associated with the target image seem to be found in the prototypical encyclopedic referent associated with the source image. These conceptual features are not stabilized in the prototypical referents but arise, in the same way as do emergent properties in verbal metaphors, as part of the reader’s relevance-seeking interpretation procedure. This kind of conceptual upload fits the third type of ad hoc concept construction introduced in section 1 above for verbal metaphors.

The reader will consider encyclopedic features not directly applicable to the referent of the target image. Inevitably, this mental operation entails the adjustment of one or several of these features and, as a consequence, they will be deprived of their conceptual stability when attributed to the referent associated with the target image. Unlike cases 1 and 2 above, in which the features were minimally adjusted in order to include the new referent but maintained their conceptual quality, in this third case a substantial adjustment of the features is required and only by losing their stability can they be applied to the new referent. In fact, the resulting emergent properties are not directly applicable to the target, and hence a deep process of adjustment is required, in the same way as are emergent properties that arise in case 3 of conceptual adjustment during verbal metaphor comprehension.

For example, another cartoon by El Roto (El País, 9 June 2002) depicts a goal keeper clearing a book (instead of the expected ball) with his fists. In the metaphor THE BOOK IS A BALL, none of the features of the prototypical referent “ball” seems to be applicable to the referent “book” of the target image. Whatever metaphoric interpretation the reader chooses (e.g., football makes people reject books” or “football stops people from getting real culture” – if “book” stands metonymically for “culture”) will emerge during the processing of the referents of the images, and will involve more inferential effort than a simple broadening of one or several features of the prototypical encyclopedic referent of the source image in order to obtain an ad hoc concept BALL*. More inferential activity will have to be devoted to adjusting some feature(s) that belong to the prototypical encyclopedic referent.

Vega-Moreno (2004: 318f) exemplifies this possible explanation of emergent properties with the verbal metaphor communicated in (11) below:
In this metaphor, our knowledge of bulldozers does not include information about them being stubborn or disrespectful, and hence this metaphoric quality of the boss emerges during interpretation. In a nutshell, the hearer can select, as a starting point, the assumptions that bulldozers are machines and are used to remove obstacles in their way. These are not directly applicable to the boss (as it would be in stability-preserving cases 1 and 2 of ad hoc concept formation). Cognitive linguistics would explain this metaphor by claiming that there is an underlying conceptual metaphor THE MIND IS A MACHINE. For RT, though, the hearer creates an ad hoc concept BULLDOZER* which involves a radical adjustment of its denotation so that it also includes a kind of removal, a type of obstacle, and a range of situations that warrant the derivation of a relevant metaphoric interpretation. In other words, an ad hoc concept [REMOVE OBSTACLES IN THE WAY]* is created with an adjustment which not only applies to machines, but also includes the act of despising, rejecting, undermining people’s feelings and people’s thoughts. As will be seen in section 4.2 below, there is a similar conceptual adjustment in the case of type-three visual metaphors and, in principle, in any multimodal combination of text and image.

4. Visual metaphor comprehension: Some examples

4.1 Examples of visual metaphor comprehension involving stable conceptual upload

(a) THE EARTH IS A SAUCEPAN (figure 3)

1. The reader finds figure 3 in a newspaper and infers that its author intends to communicate some information by means of this wordless cartoon.

2. He perceives the iconic signs of the cartoon by a bottom-up and top-down matching with previously stored prototypical visual referents of the item(s) depicted. A saucepan is identified. The continents of the Earth are also identified. These are superimposed on the saucepan.
3. An incongruity arises during the perception of the elements of the image. The Earth and a saucepan cannot be fused in one image. There is an anomalous visual arrangement regarding the mental storage prototypical combinations of objects depicted together (what above was labeled “visual syntax” of the image) that works as an *ad hoc pointer* that alerts the reader to an intended metaphoric interpretation beyond the simple depiction and perception of the drawing in the cartoon.

4. The reader enters a visual-conceptual interface, in which a number of hypotheses are made and the so-called *prototypical visual referents* of the images (already dealt with by the perceptual module) are contrasted with the parallel *prototypical encyclopedic referents* of these images.

   The reader’s hypotheses at this stage will prepare the ground for a fully inferential stage of visual metaphor comprehension, and should include the following conclusions: (a) the saucepan is the source image; (b) the Earth is the target image (a likely conclusion obtained by the reader in a relevance-seeking procedure); (c) the encyclopedic referent EARTH has qualities of the encyclopedic referent SAUCEPAN; and (d) these are the prototypical referents intended by the author; the images of the Earth and the saucepan do not stand metonymically for other referents.

5. The reader starts computing assumptions in order of accessibility following a relevance-guided procedure, beginning with the ones which belong to the prototypical encyclopedic referent depicted in the source image: SAUCEPAN.
6. One of the encyclopedic features of the prototypical encyclopedic referent SAUCEPAN seems to be directly applicable to the Earth: “heats up gradually,” since the Earth, due to global warming and the so-called “greenhouse effect” is also heating up gradually. This involves the creation of an ad hoc concept SAUCEPAN* whose quality “heats up gradually” remains relatively stable in the metaphoric process (i.e., undergoes a minimal adjustment via broadening). This new concept is applicable to all saucepans and also metaphorically to the new encyclopedic referent EARTH. In this sense, this could perhaps be a case of ontological metaphor in Lakoff and Johnson’s (1980) terminology, since in this case an inherent quality of saucepans is attributed to the target image. It would also fit Ruiz de Mendoza’s (1998) one-correspondence metaphor, since in this case only one correspondence between the source and the target is exploited.

This visual metaphor would fit case 1 of ad hoc concept formation, in which the intended conceptual feature is found in all the prototypical referents depicted by the image plus a number of other entities included through broadening.

7. The presence of SAUCEPAN* in what can be called the explicit content of the referents attached to the items depicted in the cartoon warrants the derivation of a number of possible implicated conclusions. Cartoons are a good example of a medium in which current news-worthy events play a part in the generation of implicated conclusions. In this case, if the reader knows about the fact that, at the time the cartoon was published, there was a debate on the Kyoto protocol and whether Japan and Australia would sign it (i.e., if this information is manifest to him/her, in RT terms), this information will influence both the accessibility to the visual metaphor and the mental effort devoted to its processing.

(b) THE BALLOT BOX IS A DICE (figure 4)
Steps 1. and 2. as above.

3. The reader enters a visual-conceptual interface, in which a number of hypotheses are made concerning the encyclopedic referents of the images, once the prototypical visual referents have been perceived, again preparing the ground for a fully inferential stage. Some conclusions should be derived: (a) the dice is the source image; (b) the dice stands metonymically for “gambling with dice” and more generally for “all types of gambling” (and hence the encyclopedic referent intended by the author to undergo metaphoric transference is not DICE, but GAMBLING); (c) the ballot box is the target image (facilitated by the reader’s background knowledge about the fact that the cartoon was published in a time of political elections); (d) the ballot box is in a metonymic relationship to “political elections” (and hence the referent intended by the author to undergo metaphoric transference is not BALLOT BOX, but POLITICAL ELECTIONS in general or more specifically the ones taking place in the near future); (e) the political elections have qualities of gambling.

4. The reader starts computing assumptions in order of accessibility following a relevance-guided procedure, beginning with the ones which belong to the prototypical encyclopedic metonymic referent of the source image: GAMBLING.

5. One of the encyclopedic features of the prototypical referent “gambling” seems to be directly applicable to political elections: “involves an unpredictable outcome.” Strictly speaking, this cartoon should be included in type 1, if we consider that all dice involve unpredictability. However, we may also hypothesize that since there is also a kind of gambling that gener-
ates a wholly predictable outcome (for instance gambling involving the use of loaded dice which always produce the desired result) in this case not all the referents for image of a dice would contain this quality. In this more unlikely case, the reader would be expected to create an *ad hoc* concept GAMBLING* whose feature “would involve an unpredictable outcome” applicable to most (but not all) kinds of gambling and also to other unpredictability-loaded events such as political elections. Consequently, this visual metaphor would now fit case 2 of *ad hoc* concept formation during visual metaphor comprehension.

6. The presence of GAMBLING* in what can be called the *explicit content* of the image depicted in the cartoon warrants the derivation of a number of possible implicated conclusions, some strongly implicated (e.g., the unpredictability of Spanish elections) and other weakly implicated (e.g., the more emotion-connoted implication that no Spanish party is trustworthy).

4.2. Example of visual metaphor comprehension involving innovative conceptual upload

(a) THE TENNIS RACKET IS A SHARK’S JAW
Phillips and McQuarrie (2004: 123) reproduce an advertisement for a tennis racket in which a shark’s jaw has been superimposed on a racket frame. How would the audience process this ad? Basically steps 1 to 3, as proposed for the previous examples, would also apply here. The unusual fusion of images and the anomalous visual syntax (i.e., bringing together visual elements – racket and jaw – which are not stored as part of the prototypical arrangement of objects such image can normally be made of or surrounded by) work as an *ad hoc* pointer to a non-denotative interpretation of the images.

4. The reader then enters the so-called visual-conceptual interface, in which a number of questions are asked regarding the relationship between the images depicted (already obtained via perception) and the encyclopedic information attached to them. Among others, some conclusions that the reader would be expected to derive are these: (a) the jaw is the source image; (b) the racket is the target image (a conclusion facilitated by the context of the advertisement, in which it is made clear that the author of the ad intends to characterize the racket in some way, so that it is eventually purchased); (c) the jaw stands metonymically for SHARKS in general; (d) the
racket stands metonymically for THE PLAYER’S SKILL; (e) the player’s skill acquires qualities of a shark (aggressiveness, fearful attack…).

5. The reader then starts computing assumptions in order of accessibility following a relevance-guided procedure, starting with the ones listed in the prototypical encyclopedic referent made accessible by the source image: SHARK.

6. Again, although intuitively the reader can feel that the intended interpretation has to do with some form of aggressiveness, there is nothing in the behavior of sharks that can be directly applicable to a tennis player’s skill (although, as a stereotypical feature of sharks, it would also be labeled as ontological metaphor, in Lakoff and Johnson’s 1980 terminology), and therefore all the eventual transference of information will inevitably involve “emergent properties” arising in the relevance-guided comprehension procedure. Whatever conceptual information ends up being applied to the encyclopedic referent associated with the target image will involve a loss in the stability of the storage of this information.

7. The readers then assess qualities of the encyclopedic referent associated with the source image and will adjust their denotation drastically in order to fit not only the aggressiveness of an animal, but also the aggressiveness of a tennis player. These qualities include the information that “sharks are aggressive animals.” These are possible constituents of the ad hoc concept SHARK* but not directly applicable to playing skill, so the reader has to adjust (broaden) the quality AGGRESSIVE* that belongs to SHARK* so that it also covers the tennis players’ skill. Other features such as the way sharks “chase” and “attack” their preys may also be adjusted (depending on the reader’s willingness to devote additional mental resources to this task), leading to the ad hoc concepts CHASE* and ATTACK*.

8. These adjustments are necessary in order to draw the implicated conclusion that the user of this racket will play with an unusual degree of aggressiveness. As above, the adjusted concepts warrant the derivation of a number of possible implicated conclusions.

5. Verbal before visual or vice versa?

From the examples analyzed in the previous section, we can conclude that interpreting visual metaphors also involves a great deal of conceptual upload and adjustment following a criterion guided by a search for relevance. In this sense, it is worth commenting that many visual metaphors are origi-
nal in the way they create a metaphoric link through an anomalous visual arrangement, while others seem to include an anchorage of previously used verbal metaphors which are simply transferred to a visual medium and were probably stored previously as conventionalized metaphors. In these cases, the metaphor-seeking conceptual assessment can indeed be speeded up by the fact that a particular feature of the prototypical encyclopedic referent of the source image has been made prominent by previous use through verbal means, or even facilitated by the fact that the visual metaphor only exists because there is an underlying verbal one. In a way, this is the counterpart of the well-studied fact that verbal metaphors also draw on a conceptual repertoire of visual sensory schemas which aid in the metaphoric attribution, to the extent that these images often end up becoming conventionalized in the language and deprived of their sensory metaphoric power.\(^{11}\)

Therefore, it comes as no surprise that so many metaphors involve the mediation of visual sensory information contained in image schemas (see Lakoff 1987), or involve what can be called re-visualization of conventionalized verbal metaphors, which is extensively used by cartoonists. Several steps are involved in the comprehension of the cartoons that fit this quality: (a) initially, an image is much more effective (i.e., vivid) than the range of coded options available to communicate a thought. (b) A metaphor is created that contains a schema as a referent. This schema contains visual sensory information.\(^ {12}\) (c) Repeated use of the metaphor makes it lose its sensory vividness and it ends up becoming conventionalized and hence people stop regarding it as a metaphor. (d) The cartoonist takes this conventionalized metaphor and re-visualizes it, as it were, forcing the reader to reincorporate into its processing all the sensory vividness that the metaphor had already lost.

An example is a cartoon drawn by El Roto (El País, 15 April 1996) depicting the Earth split into two parts, and with a big gap between the North and the South hemispheres. There are people trying to jump from the Southern hemisphere onto the North one but they inevitably fall into the huge gap. This metaphor, which can be described as “there is an abyss between the North and the South,” reproduces the aforementioned steps: (a) the sensory information of an abyss is more vivid than other coded options to communicate depth and distance between A and B; (b) a metaphor THE DIFFERENCE BETWEEN NORTH AND SOUTH IS AN ABYSS is created containing visual sensory information; (c) the metaphor ends up becoming conventionalized and loses its sensory power (people stop seeing an abyss when uttering the metaphor); (d) the cartoonist re-visualizes the information con-
tained in the metaphor, forcing the reader to see the sensory qualities of the image schema that had been lost due to conventionalization.

6. Concluding remarks

Interpreting visual metaphors does not differ substantially from verbal metaphor comprehension. Both kinds of metaphor are “decoded” by a specialized mental module (Fodor 1983) which delivers schematic information that has to be enriched inferentially in order to obtain the intended interpretation (an optimally relevant one). The de-contextualized perception of images is not relevant enough (some incongruity in the image or images works as an ad hoc pointer directing the viewer towards a metaphorical interpretation) and the reader has to engage in subsequent interpretive steps involving the access to encyclopedic information either directly related to the referent depicted in the image, or made prominent by metonymic relationship to other encyclopedic referents. At this stage, the reader will adjust the conceptual information in his or her search for relevance and will generate appropriate ad hoc CONCEPTS*. This is the same kind of conceptual adjustment that takes place in the interpretation of verbal metaphors. In short, conceptual information has to be accessed and adjusted in any type of metaphor. It is only the type of decoded input feeding the inferential processor that makes a difference. Indeed, the mode in which the reader is presented with the coded information that has to be adjusted inferentially plays a major role in the quantity and quality of metaphoric conclusions derived. Normally, pictures have a more powerful impact on the reader due to their holistic gestalt-like processing and are good for “visualizing” conventionalized concepts such as “abyss” in the example above. Utterances, on the other hand, are linear, and readers make interpretive hypotheses as text is processed in a word-by-word integration into phrases and sentences, which entails differences in the way literal and implicated meanings are generated.

Notes

1. Since all the visual metaphors analyzed or referred to in this article are taken from newspapers, from now on the viewer of the visual metaphor will be referred to as “reader”.
2. I will use the convention of adding asterisks to stress the fact that the accompanying word is an ad hoc concept and not an encoded concept.
3. In fact, context plays an important role aiding the addressee in determining the kind of concept* that the speaker or the author intends. Normally, in the course of a conversation, previous utterances and background knowledge about the speaker work as an important “short-term-memory storage” of information against which new utterances are interpreted. Vega-Moreno (2004: 317) explains this with the metaphor “my boss is a shark.” If it is clear from previous turns of the conversation or from general encyclopedic information about the speaker that he is happy with his boss, the concept AGGRESSIVE may be adjusted to denote a kind of (positive) aggressiveness that involves energy and assertiveness (represented as AGGRESSIVE*). However, processing the metaphor on the assumption that the speaker is afraid of his boss’s tactics and techniques, the concept AGGRESSIVE would be adjusted to denote a kind and level of (negative) aggressiveness (AGGRESSIVE**).

4. One famous explanation is provided by the so-called conceptual blending theory (Fauconnier and Turner 1998). It suggests that a subset of the attributes and relational structure from the source and target domains are imported into a blended space where they can be combined and supplemented with information from encyclopedic knowledge. These hybrid models, or blends, are useful in explaining emergent properties.

5. The existence of a choice of two possible logical forms for the same linguistic input is often exploited by humorists in some of their jokes. In Yus (2003b: 1304), it is claimed that the source of humor in some jokes lies in the fact that the language module of the addressee has to choose between two possible de-contextualized logical forms extracted from the humorist’s utterance, as in the following example: Postmaster: “Here’s your five-cent stamp.” Shopper (with arms full of bundles): “Do I have to stick it on myself?” Postmaster: “Nope. On the envelope.”

6. It is utterly important to identify the image as intentionally communicated (ostensive in RT terms) and not simply as visual information unintentionally exuded, as it were, from the environment. The former carries a presumption of eventual relevance which the latter lacks. This stage of intention ascription is important because it constrains all the subsequent inferential activity devoted to the processing of the visual stimulus and the amount of effort that the reader will be willing to devote to this inference beyond a purely sub-attentive identification of the image.

7. In fact, there is experimental evidence that the visual system uses principles of coherence to detect whether the visual information corresponds to a unique object or belongs to separate, interrelated objects, and different specialized brain cells are devoted to these tasks (see Humphreys and Heinke 1998).

8. The fact that we normally perceive only one interpretation very rapidly indicates that we see far more than the immediate information falling on our retina. The highly accurate guesses and inferences that we make rapidly and unconsciously are based on a wealth of knowledge of the world and our expectations for the particular scene we are seeing (Cavanagh 1998).

9. The change in mode from drawing to photographic style in this example may indicate that the visual syntax not only points towards the metaphor but also the “humor” of the vignette by the unexpected association between the building and the syringe. Perhaps
the decoding of the photographic mode versus the drawing modes connected in this picture provoke the search for extra implicit meanings besides the anomalous visual arrangement. I would like to thank Eduardo Uriós for pointing this out to me.

10. Specifically, “what guarantees the interpretation is the existence of an underlying conceptual mapping from ‘bulldozer’ to ‘human being’ whereby we understand [the boss’s] behavior in terms of the figurative behavior (i.e. the way the machine functions) that we attribute to a bulldozer” (Ruiz de Mendoza and Pérez Hernández, 2003: 29).

11. Tendahl and Gibbs (in press) assert that “the motivation for metaphorical language is found in recurring sensorimotor patterns of experience that are continually enacted as neural processes in the moment of thinking, speaking, and understanding. Such recurring sensorimotor patterns at least motivate the existence and continued use of many conventional metaphors and some novel extensions or elaborations of these in creative metaphorical language”.

12. As correctly argued within cognitive linguistics, this is an example of the typical mental operation that helps people conceptualize vague or abstract domains of knowledge in terms of more specific and familiar knowledge such as the one provided by sensory input.

References

Barthes, Roland

Carston, Robyn

Cavanagh, Patrick

El Refaie, Elisabeth

Fauconnier, Gilles, and Mark Turner

Fodor, Jerry

Forceville, Charles


Ruiz de Mendoza, Francisco, and Lorena Pérez Hernández

Schilperoord, Joost, and Alfons Maes
This vol. Visual metaphorical conceptualization in editorial cartoons.

Sperber, Dan, and Deirdre Wilson

Tanaka, Keiko

Tendahl, Markus, and Raymond W. Gibbs

Vega-Moreno, Rosa

Villafañe, Justo, and Norberto Minguez

Wilson, Deirdre, and Robyn Carston

Yus, Francisco
2003a Conceptos ad hoc en el procesamiento de la metáfora. El caso de las viñetas humorísticas en la prensa, paper delivered at *Jornadas sobre Texto/Imagen*, Ciudad Real (Spain): University of Castilla-La Mancha, December.


2005 Ad hoc concepts and visual metaphor? Towards relevant ad hoc pointers, paper delivered at the 9th International Pragmatics Conference, Riva del Garda (Italy), July.