Multimodality in memes. A cyberpragmatic approach

Francisco Yus
University of Alicante

Abstract
In this chapter, a corpus of 100 instances of a specific type of meme, image macro (the one typically made up of one top and one bottom stretch of text and a picture in-between), is analyzed in its multimodal quality, specifically in search of different categories that these memes might fit into depending on the relationship existing between picture and text and its impact on the quality of the eventual interpretation. An underlying assumption in the chapter, broadly within a cyberpragmatic framework (Yus, 2011), is that different text-picture combinations will have an impact on eventual relevance by yielding different balances of cognitive effects and mental effort, the latter sometimes compensated for by an offset of additional cognitive effects in the shape of implications.

1 Introduction: Image macros

In 1976, Dawkins proposed the term meme to refer to the idea of disseminated cultural unit. This meant information with the capacity of infecting people’s minds with a kind of duplicating replicability; imitation, for its part, was said to be the key to the survival of memes. Therefore, in this initial conceptualization, memes were analogous to the biological concept of gene, in the sense that they were seen as self-replicating and being communicated from person to person. More recently, memes have been defined as units of information, ideas or mental representations, cultural instructions that are not only self-replicating but also contagious (Taecharungroj & Nueangjamnong, 2014, p. 152).

From this initial idea of meme, replicating pieces of information that spread through the Net through user-to-user(s) communication have been labelled similarly: Internet memes. This is not surprising, since these memes do share some properties initially ascribed to memes in general. Among the many definitions of internet meme stand the following: (a) Any artefact that appears on the Internet and produces countless derivatives by being imitated, remixed, and rapidly diffused by countless participants in technologically mediated communication (Dynel, 2016, p. 662); (b) A group of digital items sharing common characteristics of content, form, and/or stance, which were created with awareness of each other, and were circulated, imitated, and/or transformed via the Internet by many users (Shifman, 2014, p. 41); and (c) a relatively complex, multi-layered, and intertextual combination of (moving) picture and text that is disseminated by the active agency of internet users, becoming popular among them (Laineste & Voolaid, 2016, p. 27).

Crucially, a distinctive feature of internet memes is that they are invariably ascribed to a “family” and part of their effective processing entails this initial family ascription (Nissenbaum & Shifman, 2017, p. 484; Segev et al., 2015, p. 418). In this chapter, I will focus on a particular family of memes: the image macro meme, made up of a line or two of text on top of the meme, line(s) of text at the bottom and one picture in the middle. The image macro meme presents a number of interesting text-picture combinations with interesting pragmatic implications.
2 Multimodality

Although studies on multimodality have been published for more than thirty years, multimodality is becoming increasingly important nowadays due to the pervasiveness of discourses on the Internet that combine different modes of communication (text, picture, audio, video...). A basic definition of multimodality is suggested by Stökl (2004, p. 9), for whom “multimodal refers to communicative artefacts and processes which combine various sign systems (modes) and whose production and reception calls upon the communicators to semantically and formally interrelate all sign repertoires present.” For van Leeuwen (2015, p. 447), the term “indicates that different semiotic modes (for instance language and picture) are combined and integrated in a given instance of discourse or kind of discourse.” In fact, the most interesting picture-text combination for a pragmatic (i.e. cyberpragmatic) analysis is the one in which the eventual meaning of the meme cannot be obtained from the partial meanings of text or picture taken separately, but only from their combined meanings that yield implications.

Finally, Jewitt (2014) summarizes the key assumptions in studies of modality: (a) all communication is multimodal; (b) analyses focused solely or primarily on language cannot adequately account for meaning; (c) each mode has specific affordances arising from its materiality and from its social histories which shape its resources to fulfil given communicative needs; and (d) modes concur together, each with a specialized role, to meaning-making; hence relations among modes are key to understand every instance of communication. In the specific case of meme communication, co-occurrence of text and picture (fourth assumption in Jewitt, ibid.) is especially important and ideal for a relevance-theoretic analysis, as will be commented upon in the next Section.

3 Text-picture combinations and relevance

In Yus (2016) some space is devoted to the pragmatic implications, in terms of relevance, of processing combinations of text and picture such as the ones found in the memes under analysis in this chapter. Yus (2016) argued that, in the same way as we have explicit interpretations (explicatures) and implicit or implicated interpretations (implicatures) of verbal utterances, visual content also leads to visual explicatures and visual implicatures (see Forceville, 2014; Forceville & Clark, 2014; and Wharton, 2009 for discussion). That is, when faced with a picture in a meme, the user also has to make inferential hypotheses concerning the role that it plays in the overall comprehension of the meme. For instance, the user needs to infer whether the picture has a purely denotative quality (the user is simply expected to identify the referent of the picture without any further implications), in which case its interpretation would be a visual explicate, or whether it has a wholly inferential connotative quality that can only be obtained from the combination of the picture and contextual information (as happens with implicated meanings from utterances). In this case, we would be dealing with a visual implicature. The visual explicate is easy to process, since the user simply identifies the visual information in the picture and straightforwardly matches it with the most appropriate mental referent. Visual implicatures, on the contrary, are fully inferential and have to be obtained with the aid of context.

Therefore, interpreting a meme entails a “division of labor” between the processing of the text, the processing of the picture, and the identification of possible connotative meanings.
for text, picture, and text-picture combinations. Specifically, upon finding a meme in a Facebook profile, or within a WhatsApp message, the user will have to engage in the following inferential strategies, which are not meant to be successive but performed according to the user’s expectations of, and search for, relevance in the meme:

**Strategy 1.** To decode and inferentially enrich the verbal content of the meme (top and bottom lines of text) in order to obtain the explicit interpretation of the text or explicature.

**Strategy 2.** To derive implicatures from verbal content, if these are necessary to reach a relevant interpretation of the verbal content of the meme.

**Strategy 3.** To decode and inferentially enrich the picture to yield a visual explicature.

**Strategy 4.** To derive implicatures or implications from the picture in the meme, if these are necessary to reach an adequate interpretation of the meme as a whole.

**Strategy 5.** To infer possible combinations of text and picture to yield interpretations (typically implicated ones or verbal-visual implicatures) that are only possible from the combination of these sources of information (text and picture) and not from either of them taken separately (Tsakona, 2009, p. 1172). Very often, the information obtained from the picture will lead to an inferential backtracking (and re-interpretation) after the accompanying text has been processed, in the light of the information provided by visual content. On other occasions, the text will force a new interpretation of the picture, resulting in either a new visual explicature or leading to the derivation of a previously unforeseeable visual implicature. In my opinion, the effects of this combination of text and picture are the ones that are more interesting for a cyberpragmatic analysis of why memes turn out effective (e.g. humorous), since the eventual interpretation demands the user’s active participation in combining sources of information for the sake of an eventual satisfactory (i.e. relevant) interpretation.

Combining interpretations from text and picture entails a kind of iconic literacy in order to process adequately the linear and time-demanding word-by-word processing of the text and the instantaneous visual impact of the picture in the meme. El Refaie & Hörschelmann (2010, p. 200) write about types of literacy concerning the interpretation of cartoons, which are also applicable to meme interpretation: “Interpreting cartoons is a matter of drawing on many different types of literacy, which form the necessary preconditions for readers to be able to discover relevant connections between the fictional scene of a cartoon and a political argument. In this particular case, multimodal literacy included the ability (1) to establish the referents of a cartoon both on the level of the make-believe world and of the real-life world of current political events, (2) to impose a narrative on the cartoon picture, and (3) to draw on intertextual references.”

**Strategy 6.** To access as much contextual information as is necessary to obtain interpretations out of strategies 1-5 above. In the case of some memes, the reader’s background knowledge on current affairs, newsworthy events, political issues, etc. (his/her literacy) is crucial to understanding the meme properly, to the extent that, very often, the meme makes little sense if it is separated from the specific time-frame and pieces of news that justified its publication (Kardaş, 2012, p. 208; El Refaie & Hörschelmann, 2010, p. 197). This is so to the extent that, as Conradie et al. (2012, p. 41) remark for cartoons, and equally applicable to memes, we can differentiate two frameworks of analysis, one related to the reader’s background knowledge (reader-orientated strategy) and one focused on the semiotic qualities of the discourse (text-orientated strategy): the first strategy involves looking at how the reader interprets the discourse (based on his/her subjective conceptual framework), while
text-orientated studies involve examining the discourse (visual and verbal) per se, focusing on its form and style.

Several qualities of the multimodal meme will determine which of the strategies (1-5) listed above (aided by contextualization, strategy 6), will be most likely to take place first and in which order the other strategies will be performed. Among others, the salience of some discursive element in the meme will lead the user to focus his/her attention there before other parts of the meme are processed. This may have implications for how the eventual relevance is assessed; for instance, for how the intended implications from text-picture combinations are derived. In theory, the reading path for the meme should start at the top (text processing), continue with the picture in the middle (visual processing) and then finish at the bottom of the meme (text processing). However, the picture may draw the user’s attention before the lines of text are interpreted for relevance and this salience will lead to an alteration of the eventual cognitive effects if the order of processing does not match the initially intended one.

In this sense, van Leeuwen (2015, p. 457) reminds us that texts are not really linear in their processing. As has been suggested, a top-bottom linear reading of the meme is expected on many occasions. But reading paths are mainly created by differences in salience, and depend on the textual or visual element that attracts the reader’s or viewer’s attention over and above other elements (Kress & van Leeuwen, 2006, p. 218). Differences in salience can be realized by foregrounding or by differences in size, boldness, tonal contrast, or color. In this way, visual compositions can set up particular hierarchies between the elements to attract the attention and guide the movement of the hypothetical Internet user’s eyes within and across the different discursive elements of the meme. Such reading path will begin with the most salient element, from there it will move on to the next most salient element, and so on, in a trajectory that need not be similar to the top-down order of multimodal discourses such as memes. Besides, salience may vary enormously from meme to meme even if the memes neatly belong to the same family. And finally, the actual salience of elements in the meme will vary for different users; different areas of the meme will draw the attention of a variety of users in different directions, so there is no guarantee that the same reading paths will be followed across users.

4 Methodology

In this chapter, an analysis is carried out of a corpus of 100 memes randomly selected from a Google search with the text query “image macro meme”. The main purpose is to determine what category of text-picture combination is more frequent and why, together with predictions of interpretive relevance. Along these same lines, several publications have addressed text-picture interfaces (see for instance Trifonas, 2015; van Leeuwen, 2006; Tsakona, 2009; Sarapik, 2009). However in this chapter, McCloud’s (1994) taxonomy of different categories for multimodal combinations in comics will be applied to the analysis, since memes share their textual-visual quality with comics. Actually, only one of McCloud’s categories cannot be found in the corpus of memes (see below). Along with McCloud’s taxonomy, other sets of categories for text-picture relationships of semiotic (Barthes, 1977) and social-semiotic (Gill, 2002; Salway & Martinec, 2002; Martinec & Salway, 2005; Chan, 2011) orientations will also be applied to the data. The interest lies, of course, in the text-picture relationships that contribute to specific interpretive outcomes, and not so much in the purely semiotic interrelation of modes. For example, Cohn (2013) proposed four ways of
connecting text and pictures: (1) inherent (text and picture are part of each other’s structures); (2) emergent (text and picture are directly interfaced with each other); (3) adjoined (text and picture are integrated but not interfaced directly); and (4) independent (text and picture are fully separate). This taxonomy, while interesting from a purely semiotic point of view, is not useful for a cyberpragmatic analysis of memes, since memes exhibit a rigid placement and quality of both modes. A more fruitful approach, in my view, is to work out the inferential implications of the interrelations of both modes and how their combinations yield relevant interpretive outcomes.

The next step is to analyze the corpus of 100 memes and determine into which category they may be ascribed, assessing the role that text and picture play in the eventual interpretation while, at the same time, looking for specific text-picture combinations that may be interesting for a cyberpragmatic account of their relevance for the user who is processing the meme. From this analysis, some conclusions are drawn.

Throughout the analysis, special attention is devoted to determining (a) which contribution each mode (text/picture) makes to the overall global meaning of the meme; (b) whether there are cases in which the text makes little or no contribution to the interpretation of the meme, which is mainly provided by the picture, or it is the opposite relationship: the picture making little or no contribution to the interpretation of the meme, that is, cases where the picture add little or no meaning to the meme; (c) in which cases text and picture combine to generate implicated meanings that may only be obtained from the combination of the information from both modes; and (d) whether there is some specificity to memes that does not neatly correspond to any of the categories suggested by McCloud (1994) for comics.

5 Text-picture combinations in image macro memes

5.1 Word specific, where pictures illustrate but do not significantly add to a largely complete text

This category is similar to Barthes’s (1977) concept of illustration, in which the text is the primary mode and the picture plays no substantial role in altering the meaning conveyed verbally. It also suits Martinec and Salway’s (2005) exemplification, in which the text takes prominence over the picture.

25 instances out of the corpus of 100 memes can be ascribed to this category. Sometimes, the picture is useful in narrowing the scope of the meaning of the words (the picture also prevents a potential incongruous interpretation of the text). Consider (1-2):

(1) Top text: If tomorrow isn’t the due day  
Picture: Picture of a university student drinking a beer in a pub.  
Bottom text: then today isn’t the do day.
(2) Top text: Why make college the 4 best years of your life  
Picture: Picture of a university student drinking a beer in a pub.  
Bottom text: if you can make it 6.

In (1-2), the text is more or less autonomously interpreted, but it is more adequately processed next to the picture of the person who might possibly utter these words, i.e. a university student. In both memes, the picture is the same and it exemplifies the kind of
person that would typically produce these utterances. The eventual balance of cognitive
effects and mental effort, as envisaged by relevance theory (Sperber & Wilson, 1995), is
optimized thanks to the contribution of the picture in narrowing down the potential range of
speakers of such utterances; the picture also helps the user in determining the exact meaning
of, for example, do in (1) and make in (2). Furthermore, the picture helps in the derivation of
a number of implications on a stereotypical university student’s attitude to homework and
commitments to university duties in general.

The same applies to the adjustment of the concept coded by some specific word in the
text of the meme (more on concept adjustment below). In (3), for instance, with the same
picture as (1-2) above, the concept coded by attendance is narrowed probably to “attending
bars and cafes in college” but not classes; another possible narrowing of attend could render
“being present in the classroom without paying any attention to what is being said by the
teacher”. Similarly, the concept coded by the word disturbing in (4) acquires more specific
meanings (even if maintaining its basic denotation) when placed next to a picture of Darth
Vader from Star Wars:

(3) Top text: I have 100% college attendance
Picture: Picture of a university student drinking a beer in a pub.
Bottom text: but 0% class attendance.

(4) Top text: I find your lack of documentation
Picture: Picture of Darth Vader.
Bottom text: disturbing.

Concept adjustment is one of the inferential strategies that users deploy when turning the
schematic meaning of the words into meaningful and contextualized interpretations. Another
inferential strategy is reference assignment. In this case, the picture in the meme may also be
useful when identifying who the text is referring to, as in (5):

(5) Top text: Maggie Smith battled cancer while filming the last Harry Potter movie.
Picture: Photograph of Maggie Smith.
Bottom text: She carried on, not wanting to disappoint the fans.

The picture in the meme may also play a part in the eventual success of an incongruity-
resolution pattern (see Yus, 1997; 2016; in press) in the text of the meme. Incongruity-
resolution is one of the most typical joke structures, in which the addressees of the joke face
an incongruity due to the manipulation of their inferential strategies, and they have to find a
resolution that reconciles the incongruous parts of the joke.

However, in this meme pattern the picture is not essential for the text to be perceived as
humorous (i.e. the text itself follows the incongruity-resolution pattern without needing the
picture for the derivation of humorous effects). Sometimes, the role that the picture plays is
simply that of an illustration, and therefore it suits this first category. Take, for instance, the
meme in (6). It may well be told or read without an accompanying picture. But in the meme,
the picture helps in associating black with back race, as if a racist (white) joker were telling
the joke, and enhances the subsequent incongruity provided by the bottom text of the meme.

(6) Top text: What’s black and never works?
Picture: White man smiling.
Bottom text: Decaffeinated coffee, you racist bastard.

On other occasions, though, the incongruity-resolution pattern is constructed out of text interpretation only, and the picture could well be suppressed without significant informative loss, as in (7-8):

(7) Top text: Do you like Big Bang?
   Picture: Smiling girl.
   Bottom text: ‘Cause I’m sure we’d make fantastic babies.

(8) Top text: Why do people say the sky’s the limit
   Picture: Drawing of a dinosaur.
   Bottom text: when there are footprints on the moon?

In (7), the meme plays with the initial accessibility to the name of the famous TV series, but the word Bang acquires a different meaning after reading the bottom text, which becomes incongruous regarding the initial interpretation. This is a joke pattern that could well be told orally without needing an accompanying picture. Something similar occurs with (8), in which the picture of a dinosaur has no connection whatsoever with the text of the meme.

5.2 Picture specific, where the picture dominates and words do not add significantly to the meaning of the picture

This is similar to the previous category (word-specific and pictures illustrating), but in the one here under scrutiny it is the picture that is central to the meme: the words are supplementary and play an exemplification role. That is, the words exemplify situations in which the gestures provided by the picture might be produced (Wu, 2014, p. 1417). Only 7 instances of the corpus of 100 memes fit into this category. This indicates that pictures generally play either the role of exemplifying, emphasizing or amplifying the meaning of the text, or are combined with the text in order to generate interpretations which can only be obtained from this combination. However, instances in which the picture is the main source of information (and eventual relevance) in the meme are far less frequent.

One of the cases in which the picture does dominate occurs when the meme shows a person who is famous for saying something. The picture itself is worth the user’s attention and the text merely emphasizes the attributes of the person depicted in the picture. In this case, the user is expected to be able to retrieve from background knowledge the specific information which justifies the appearance of the famous person in the meme (e.g. what he/she typically said or did that leads to the creation of the meme). An example is (9):

(9) Top text: Either you’re naive or have scruples.
   Picture: Picture of actor Hugh Laurie in his role as Dr. House with an angry look on his face.
   Bottom text: I’m not sure which is worse.

5.3 Duo specific, where words and pictures send essentially the same message

This category is homologous to the social semiotics category of equivalence (Gill, 2002), since the content is communicated both by the picture and its accompanying text, resulting in
some degree of redundancy in meaning (Chan, 2011, p. 149). Predictably, almost no instance of the meme corpus fits this category, since one of the inherent features of this kind of meme is the overlap of meaning between text and picture. Only meme (10) below might be close to the attribute of this category. The text of the meme refers to happiness, and this feeling is paired with the picture of happiness on a famous actor’s face:

(10) Top text:  You find it offensive. I find it funny.
   Picture:  Picture of actor Leonardo di Caprio smiling and proposing a toast.
   Bottom text:  That’s why I’m happier than you.

5.4 Additive, where words amplify or elaborate on a picture or vice versa

By contrast, this category is frequent in the corpus here under analysis. Up to 37 instances, one way or another, exhibit this text-picture relationship in which there is amplification or elaboration of one mode regarding the other.

This category reminds us of Barthes’ notion of anchorage, in which words help the user reach a more fine-grained interpretation of the related picture, as intended by the creators of the meme, or vice versa. The category is also similar to social semiotics notions such as: (1) extension, in which one mode provides new information on the information from the other mode; (2) enhancement, where the text adds an informative element to the picture or vice versa; (3) exposition, where the picture elaborates on aspects of the text or vice versa; and (4) augmentation, involving a picture extending or adding new meanings to the text or the text extending the picture by providing an additional element (Chan, 2011, p. 154).

A frequent text-picture relationship in this category is that in which the picture amplifies or elaborates on the meaning of the adjacent text. Consider these memes:

(11) Top text:  I think we have to fix this problem
   Bottom text:  the old-fashioned way.
(12) Top text:  I changed all my passwords to “incorrect”
   Picture:  Famous actor playing a dumb character.
   Bottom text:  so, whenever I forget, it says, “your password is incorrect”.
(13) Top text:  I want to go to Taco Bell
   Picture:  Woman in tears.
   Bottom text:  but I’m on an all-carb diet.

In (11), the user’s background knowledge of Brando and of his role in The Godfather makes it easier to adjust the concept encoded by the word fix (see Section 6 below for a description of concept adjustment under relevance theory). Besides, the user is guided by the picture when trying to narrow and connote the meaning of the phrase the old-fashioned way, initially broad and little specific. In (12), the picture of a famous actor playing a dumb character enhances the absurd idea that he came up with as described in the text of the meme. Finally, in (13) the picture of a woman in tears helps enhance the user’s inference of the woman’s urge to go to this restaurant and in obtaining a more vivid picture of her addiction to fast food. In all of these memes, the eventual relevance is optimized because the picture is useful in obtaining a more accurate interpretation of the accompanying text.

A frequent type of meme which may also be ascribed to this category is the abundant
series of memes which share the same picture but contain different texts (this text-picture configuration is also compatible with the meme-specific inferential strategy of *ad hoc visual referent adjustment*, see Section 6 below). In this case, the main interest lies in a specific kind of gesture that users find funny (often a gesture made by an actor/actress and reproduced as a film frame). The different versions of text are descriptions of situations in which this gesture might be produced. However, this also involves some kind of amplification or elaboration of the interpretation of the picture via adjustment (see Section 6). An example is provided in (14) below and further examples are reproduced in Figure 1. All the memes depicted in this Figure are single instances of a whole series in which the picture remains the same but the text is changed from meme to meme.

(14) Top text: When somebody adds another plate to the sink  
Picture: Famous actor with a hateful look on his face.  
Bottom text: while I’m washing the dishes.

![Memes with different texts](image)

**Figure 1.**

Finally, another role of the picture is to facilitate the successful outcome of the incongruity-resolution humorous strategy in the meme. In the first category above (Section 5.1), an example was provided where the picture of the meme played a minor role in the unfolding of an incongruity-resolution schema. In this category, by contrast, pictures play more significant roles in the humorous outcome of this schema. Firstly, the picture may help trigger an incongruity during the construction of an appropriate scenario for the interpretation of the meme; this is what Yus (2013a; 2013b) labelled *make-sense frame*. In a nutshell, inferring the intended interpretation of the meme also involves the extraction of general information about the world and everyday situations that is stored as accessible chunks of encyclopedic information (specifically stored as “I conceptualize X as p” or as a more factual “I believe
that p”). This information is often retrieved almost unconsciously in order to make sense of the intended “scenario” for the comprehension of the meme, and may be exploited for humorous purposes. Consider the meme in (15):

(15) Top text: I feel like nobody loves me
Picture: Woman with tears running down her face.
Bottom text: especially when I eat bananas with my butt.

In this meme, the picture of a woman in tears makes it easier to build up a make-sense frame of sheer sadness. The picture amplifies the inference of the negative feelings associated with the fact that nobody loves the woman. The user is led to believe that the woman is so sad at not being loved that she finally bursts into tears. However, this make-sense frame just constructed is radically invalidated by the bottom line of text.

Secondly, the picture may facilitate the successful outcome of a text-based incongruity-resolution strategy. For example, if this strategy is based on making one interpretation of an initial part of a joke more likely to be selected only to invalidate it in a subsequent part of the joke (thus generating incongruity), the picture may be useful in leading the user to choose that very sense of the word, as intended. In (16), for instance, there is an incongruity centered upon the user’s disambiguation of the verb *beat*, which encodes several possible senses. The user will inevitably select the sense that is most accessible, the one demanding least mental effort, in this case “to strike violently or forcefully and repeatedly”, and the scary look of the man in (16) facilitates this almost unconscious, initially relevant, choice of one of these senses of *beat*, which is again invalidated by the bottom text of the meme and replaced with a more unlikely (initially less relevant but eventually correct) interpretation.

(16) Top text: I love beating women
Picture: Man with a very scary look on his face.
Bottom text: to the door so I can hold it open for them.

5.5 Parallel, where words and picture follow different courses without intersecting

This category resembles the social semiotics label of *exposition*, in which picture and text are at the same level of generality (van Leeuwen, 2011, p. 676). Predictably, the quality of this category is absent in the corpus; no meme seems to contain this text-picture relationship, since the eventual interpretation of memes normally relies on the combination of text and picture, and it makes little sense to put both modes together in the meme if these are not somehow related.

5.6 Montage, where words are treated as integral parts of the picture

This is a typical category in comics, since artists often skillfully connote the texts with iconic connotations (dripping words, creative use of bold letters, or letters with a visually-connoted shape). In a way, then, both modes become semiotically integrated. Consequently, it is similar to the social semiotics term *homospatiality*, where different semiotic modes co-occur in one spatially bonded homogenous entity (Lim Fei, 2004). Chan (2011, p. 152) provides the example of the words of a poem, *Stingray*, which were themselves arranged in the shape of a stingray. It also suits Cohn’s (2013) aforementioned semiotic term *inherent* (text and picture
are part of each other’s structures).

This category is absent in the corpus of memes, since text type (font, size...) is fixed and imposed upon the user by the available software, and therefore creative iconization of the text is impossible.

5.7 Interdependent, where picture or words together convey an idea that neither could convey alone

In contrast, this category is frequent in the corpus with up to 30 instances. Besides, it is the category that raises more interest for a relevance-centered cyberpragmatic analysis concerned with how users obtain interpretations from the available information through interfaces. The fact that interpretations from memes in this category cannot be obtained from the partial meanings of text or picture taken separately makes the meme utterly context-dependent and very significant for the kind of analysis proposed in this chapter.

This category also suits Barthes’ (1977) term relay, according to which “text and picture do not ‘say the same thing’ but convey different, complementary content” (van Leeuwen, 2011, p. 657). To some extent, this category exhibits the qualities of the social semiotics term distribution (juxtaposed pictures and text jointly constructing information) and divergence (the meanings of text and picture contradict each other and convey new information out of this contradiction). Besides, the category is related to what Jewitt (2016) calls multimodal ensemble, where all the modes combine to convey a message’s meaning. The information is distributed across modes, and “any one mode in that ensemble is carrying part of the message only: each mode is therefore partial in relation to the whole of the meaning” (p. 73).

The most frequent text-picture interaction in this category of memes is that in which the information from the picture invalidates, to a greater or lesser extent, the information provided by the text. In (17), for example, access to the intended ironical-critical remark is facilitated by the picture of Keanu Reeves with a lunatic look on his face:

(17) Top text: What if Obama needs all of our personal information
Picture: Lunatic-looking picture of actor Keanu Reeves.
Bottom text: because he’s setting up a matchmaking service and wants us all to find love?

Besides, a typical function of text-picture combination is to trigger the derivation of implications from text, with the help of a connoted picture. Consider the following memes (18-20):

(18) Top text: Join the marines they said
Picture: Picture of a marine holding an umbrella over president Obama.
Bottom text: You’ll be a hero they said.
(19) Top text: Wants a girl who sees him for his personality
Picture: Face of a very ugly man.
Bottom text: if she’s hot.
(20) Top text: Ok class, you have exactly 50 minutes to complete this exam
Picture: Picture of a beautiful young female teacher in front of blackboard and with a wide smile on her face.
Bottom text: but before you begin, let’s all take five minutes to correct the mistakes I made while writing the test.

In (18), the activity depicted in the picture contradicts the presumption of honor and pride that the soldier should feel for being a marine. In (19), the picture of an utterly ugly man contradicts the extent of his demands for a specific type of woman made in the text of the meme. Finally, in (20) the picture of an apparently happy teacher who seems to love her profession is at odds with the information of the text concerning her (in)ability to prepare well-written exams.

Needless to say, the most radical case of text-picture relationship yielding a different meaning for the text of the meme from the one obtained from the partial meanings of text and picture occurs when the picture forces the opposite interpretation to the one obtained from the text. A good example is the series of memes sharing the same picture of two girls laughing. This picture triggers a radically different interpretive course for the texts, invariably related to sex role-connoted information, to the extent that their interpretation is reversed (Figure 2).

Finally, some pictures also trigger incongruity when the user reads the text of the meme. The role of the picture is essential here in provoking the entertainment of two competing interpretations of some ambiguous part of the text, thus fitting into the category of text-picture combination (the eventual parallel entertainment of two interpretations is not possible from text only, but the information from the picture is also required for this to happen). In relevance-theoretic terms, this simultaneity of interpretations goes against the usual processing strategy, which would normally be directed at the most accessible-relevant-interpretation of that bit of text while dismissing, often in an unconscious way, other alternative interpretations. In fact, when we engage in an inferential choice of an interpretation, we are generally not aware of other less accessible interpretations once the relevant one has been selected.

By contrast, in some of the memes fitting into this category, the reader is forced to entertain both senses of some portion of text simultaneously. The additional mental effort involved in this dual processing has to be compensated for by extra (or different) cognitive effects: an offset of humorous effects, or amusement at acknowledging the cunning play with words, for instance. Meme (19) is an example which plays with two simultaneous senses of credit:

(19) Top text: Gets one of only 12 perfect scores in the world on macroeconomics.
In other memes in this category, the text is altered in order to create an ambiguity that mirrors the accompanying picture and, as a consequence, the text and the ambiguity arising from it are utterly connected to the picture, as in (20-21):

(20) Top text: Bitches.
     Picture: Portrait of musician Handel.
     Bottom text: Can’t Handel my oratorios.
(21) Top text: France adopted a new
     Picture: Picture of France’s President Macron.
     Bottom text: “Macron” economic policy.

6 Meme specificity: Ad hoc visual referent adjustment

For comics, McCloud (1994) proposed a category, montage, which is absent in memes. Similarly, there is an inferential strategy that seems to be inherent to memes, which will be labelled ad hoc visual referent adjustment, due to its resemblance with verbal concept adjustment. Through this inferential strategy, the different texts used for the very same picture constrain the meanings of the picture, a kind of modal affordance (Jewitt, 2016, p. 72), in which the final interpretation of one of the modes is constrained by the other mode.

Concept adjustment, as proposed by relevance theory (Sperber & Wilson, 1995; Wilson & Carston, 2006), has already been mentioned in this chapter in passing; but so far, the term has only been applied to verbal communication in the relevance-theoretic bibliography. As summarized in Yus (2010), during the interpretation process, the prototypical concept encoded by a word is adjusted by the hearers so that it meets their expectations of relevance. The outcome of this adjustment is an ad hoc concept which is similar, but not identical, to the stabilized concept coded by the word. They are ad hoc “because they are not linguistically given, but are constructed online in response to specific expectations of relevance raised in specific contexts. There is a difference then between ad hoc concepts, accessed by a spontaneous process of pragmatic inference, and lexicalized concepts, which are context-invariant” (Carston, 2002, p. 322).

In certain contexts, the concept that the speaker intends to communicate is broader (i.e. less exact) than the concept that the word he/she has chosen literally communicates (or codes), as in (22a-e):

(22) a. There is a rectangle of lawn in the shed.
    [not an exact rectangle].
    b. We entered a pub, but we left since it was empty.
    [there were people in the pub -e.g. the waiter- but not interesting people]
    c. I’ve got a thousand things to do this morning.
    [many things, but not a thousand].
    d. Don’t worry. I’ll be ready in two minutes.
    [in a while, surely longer than two minutes].
    e. This steak is raw.
On other occasions, the concept that the speaker intends to communicate is narrower (i.e., more exact) than the concept that the word he/she has chosen literally communicates (or codes), as in (23a-e):

(23) a. I’ve got nothing to wear for the party.
   [specifically, nothing appropriate, nothing classy, etc.].

   b. María has a brain.
   [specifically, an outstanding brain, not simply a brain; she is very intelligent].

   c. This boy has a temperature.
   [specifically, more temperature than he should].

   d. It will take some time to fix this car.
   [specifically, longer than you imagine; longer than it would normally take].

   e. Antonio drinks too much.
   [specifically, he drinks too much alcohol].

The innovative inferential concept-related strategy found in memes is that, besides the typical concept adjustment required for verbal content (involving narrowing or broadening), sometimes the user also has to engage in an inferential concept adjustment but this time applied to the referent of the picture in the meme. This is what happens in several series of memes in which the picture is the same across memes and the users change the accompanying text in humorous ways. These texts trigger slightly different interpretations of the prototypical referent of the picture, and therefore the inferential strategy leading to these slightly differing interpretations may be called \textit{ad hoc visual referent adjustment}.

Users are aware that there are many different texts for the same picture in a series of memes and know that sometimes the relationship between picture and text in the meme is not straightforward. A frequent example is the memes with a picture of a person making a certain gesture. Upon finding an instance of this series of memes, the user will infer the gesture portrayed in the picture, but not simply as the typical (i.e., default) interpretation of the gesture, but with interpretive variations depending on the text that is above and below the picture, a proper case of adjustment but this time applied to the referent of the gesture depicted. On some occasions, the resulting \textit{ad hoc visual referent} will be narrower than the prototypical, default one provided by the picture (as happens with verbal adjustment); on other occasions, the user will infer a broader referent than the default one in the picture. The outcome will be, in both cases, an \textit{ad hoc visual referent} that meets the user’s expectation of relevance upon interpreting the meme, and which works specifically for this meme and not for other picture-text relationships in the other memes which belong to the same series (repeated picture with different texts).

Take, for instance, the memes depicted in Figure 3. The prototypical gestures depicted in the picture (the little child’s gesture and hand movement) are interpreted, in general, as the child being proud of having achieved some goal (except for the bottom right meme, which conveys a totally different feeling). However, this default, prototypical interpretation has to be adjusted (narrowed or broadened) so as to infer correctly the kind of feeling enacted by the actions described in the text that are placed next to the same picture in those memes. Hence, the \textit{ad hoc visual referent} inferred for the gesture will be different when accompanied by the text \textit{Character in Hamlet. Survived to the end} (happiness mixed up with relief), by the text
Finished work in class (more like “sheer pride”), or by the text Divorced parents. Twice the Xmas presents (intense joy). In all of these instances, the user has to adjust the referent of the gesture accordingly.

Figure 3.

However, there is a limit to this inferential strategy of visual referent adjustment. In the same series, there are certainly memes that fit into the category of ad hoc visual referents from the coded prototypical gesture, because they resemble the gesture but are not exactly the same as that prototypical one coded by the gesture. However, it can be stated that there is a point beyond which the text no longer aids in visual referent adjustment, but triggers a different referent altogether. An example is the series of memes sharing the picture of a woman in tears (some instances of this series have already been quoted in this chapter). In one of the instances, (24), the user will have no problem in broadening the scope of the visual referent for the woman in tears so as to include an unusual source of the tears due to the information communicated by the text:

(24) Top text: I’m so tired of eating
Picture: Tears running down a woman’s face.
Bottom text: at all of the restaurants near work.

In other memes from this series, the picture is the same but the extent of the pain suffered by the woman has to be adjusted depending on the accompanying text. In (24), the user can indeed imagine a situation in which the woman is on the verge of a breakdown due to her being forced to eat at those locations. However, there is a limit to adjustment by resemblance between the default visual referent of the coded nonverbal action and the ad hoc visual
referent, beyond which, what the user ends up inferring, is a humorous clash between a serious nonverbal action (crying) and the text that, in theory, should provide the user with the reason for supporting its generation, as is the case with memes (25-26):

(25) Top text: I’m so thirsty
Picture: Tears running down a woman’s face.
Bottom text: but if I drink, I’ll have to get up to go pee.
(26) Top text: I’m hungry
Picture: Tears running down a woman’s face.
Bottom text: but I already brushed my teeth.

7. Concluding remarks

In this chapter, a corpus of 100 memes of the “image macro family” has been analyzed in their text-picture combinations and especially concerning the pragmatic implications of these combinations. The default taxonomy of categories has followed McCloud’s (1994) proposal for comics (a medium very close to memes in its verbal-visual semiotic quality), yielding the following instances of memes: (a) word specific (25 memes), (b) picture specific (7), (c) duo specific (1), (d) additive (37), (e) parallel (0), (f) montage (0), and (g) interdependent (30).

It comes as no surprise that those categories for comics that “separate” the meanings conveyed by text and picture are not present in the corpus. Image macro memes such as the ones analyzed here are meant to communicate information by combining the partial meanings provided by text and picture into a slightly or radically different interpretation. In general, the analysis revealed that texts acquire prominence in the eventual interpretation and pictures either illustrate what the text communicates (first category, 25 instances), amplify or elaborate the textual meaning (fourth category, 37 instances), or aid in altering the textual meaning, often radically (seventh category, 30 instances).

Furthermore, although most of the interpretive procedures are easy to explain from a pragmatic point of view, a new inferential strategy has been found to be inherent to meme communication: ad hoc visual referent adjustment, similar to the relevance-theoretic proposal of ad hoc concept adjustment for verbal communication but innovative, at the same time, considering that this is the first time that such inferential strategy is applied to visual inputs.

References


Stöckl, Hartmut. (2004). In between modes: Language and image in printed media. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (Eds.), *Perspectives on multimodality* (pp. 9–30). Amsterdam: John Benjamins.


Bionote
Francisco Yus teaches pragmatics at the University of Alicante, Spain. He has a PhD in linguistics and has specialized in the application of pragmatics (especially relevance theory) to media discourses and several communicative issues such as misunderstandings, irony and humor. He has also developed his own pragmatic approach to Internet-mediated communication (Ciberpragmática, Ariel, 2001; Ciberpragmática 2.0, Ariel, 2010; Cyberpragmatics, John Benjamins, 2011) and he is the editor (with Chaoqun Xie) of the journal Internet Pragmatics (John Benjamins).