

“Cats be outside, how about meow”: Multimodal humor and creativity in an internet meme

Camilla Vásquez ^a, Erhan Aslan ^{b,*}

^a Applied Linguistics, Department of World Languages, University of South Florida, Tampa, 33620, USA

^b TESOL and Applied Linguistics, University of Reading, 210B Edith Morley, Whiteknights, Reading, RG6 6AA, UK



ARTICLE INFO

Article history:

Received 4 April 2020

Received in revised form 17 October 2020

Accepted 18 October 2020

Keywords:

Internet memes

Image macros

Multimodality

Wordplay

Paronymy

Register humor

ABSTRACT

Adding to research on internet memes and humor, this study examines a set of image macros related to a specific viral media event (i.e., “cash me ousside”). This particular meme is linked to a popular catchphrase uttered by a young teenage girl, who appeared on a 2016 episode of the U.S. television talk show, *Dr. Phil*. We compiled a dataset of 220 image macros related to this media event from three online platforms. Our analysis focuses on various forms of linguistic humor, which most often rely on multimodal interactions between textual and visual elements. Our findings reveal multiple instances of wordplay involving paronymy as well as register-based humor and humor involving the voicing of recognizable figures. In addition, multimodal play often involves semiotic blends of this particular meme with other popular memes. Overall, our study illustrates unique combinations of unexpected elements and semiotic extensions of a non-standard catchphrase with other visual images, thereby establishing incongruity and generating humorous meanings.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

One major trend in online humor is the exploitation of multimodality (Chovanec and Tsakona, 2018). Humorous internet texts very often present creative blends of linguistic as well as other semiotic elements. Among the genres that appear in digital communication, internet memes have received increasing scholarly attention due to their widespread use across many digital platforms. This paper examines multimodal texts related to the popular internet meme “Cash me ousside/how bah dah” that emerged from a segment of an American talk show that went viral in 2016. The goal of this paper is to bring to the fore the interplay between multimodal creativity and linguistic humor in the context of digital participatory culture. To shed light on this interplay, we draw on existing approaches to the study of internet humor in our analysis of the similarities and differences between individual expressions of one popular internet meme. Contributing to the growing body of research in humor and internet memes, our study provides a unique analysis of different forms of language-related humor in these creative and multimodal internet texts, as they pertain to a single viral media event, as well as the different ways in which textual and visual elements are combined to create humor. In what follows, we provide a brief review of theoretical approaches to linguistic humor followed by a discussion of potential interfaces between these approaches and the study of humor in internet memes as multimodal constructions. We draw on both traditional and more recent approaches to humor to

* Corresponding author.

E-mail addresses: cvasquez@usf.edu (C. Vásquez), erhan.aslan@reading.ac.uk (E. Aslan).

understand the multimodal compositionality in the memes under investigation as well as the inferential strategies involved in the meaning-making processes needed to interpret the text-image combinations.

1.1. Internet memes and humor

Pre-dating the internet, memes are cultural units that are both spreadable and adaptable. In recent years, however, memes have become more closely associated with digitally-mediated communication. Defined by Shifman (2014: 14) as “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,” internet memes can take numerous forms, such as videos or photoshopped images that have been derived from a social or political event, a popular culture reference, or anything else that has some kind of value or meaning to a certain culture or group. Memes appear across many social media platforms where they are circulated by internet users, and where they undergo constant reproduction, adaptation and recontextualization. Whether or not a particular meme goes viral depends on the extent to which its content is humorous or emotional, its clarity and simplicity, its positioning (where it appears), and its participatory nature (whether or not the meme is open to repackaging and appropriation) (Shifman, 2014).

In many cases, irony, humor and play are essential to meme's creation, circulation, and transformation. According to Knobel and Lankshear (2007), three patterns contribute to a meme's spreadability: humor, intertextuality (i.e., references to popular culture events, artefacts, icons, phenomena and practices) and anomalous juxtapositions, known as incongruity. Incongruity has been defined as an incompatibility in the arrangement of the constituent elements of an object, event, idea, social expectation and so forth, with the normal or expected pattern (McGhee, 1979) – and humor often emerges due to the violation of such expectations (Suls, 1972). Internet memes generally demonstrate second-order incongruity (Yus, 2020) which refers to “images taken from films or TV programs (e.g., stills) that are later re-used (and hence re-contextualised) in the meme” (p.13). Of relevance to some humor observed in memetic content – particularly the internet meme analyzed in the present study – is the theory of superiority which accounts for disparagement humor, wherein humor is derived from messages intended to “elicit amusement through the denigration, derogation, or belittlement of a given target (e.g., individuals, social groups, political ideologies, material possessions)” (Ferguson and Ford 2008: 283–284). According to Shifman (2011) most internet memes exhibit three key features of humor: *playfulness* (i.e., the audience is invited to take part in a game); *incongruity* (i.e., humor resulting from two incongruous elements); and *superiority* (i.e., the audience perceiving themselves as superior). Participating in memetic activities (i.e., in their creation, distribution, appreciation) means engaging in a playful process of collaborative interaction to generate online humor (Seiffert-Brockmann et al. 2018).

Internet memes most prototypically take the form of the so-called “image macro”: a visual image with superimposed textual elements. Dynel (2016) defines the image macro as “a captioned image that typically consists of a picture and a witty message or a catchphrase” (p. 663). A subset of these multimodal blends are referred to as “stock character macros” (Shifman, 2014: 343). In these, a visual of a strong personality serves as the central anchor in the middle of a square or rectangular space, and a linguistic setup is found at the top of the space, often followed by a linguistic punchline at the bottom. Commonly, the textual components appear in block, all-caps white Impact font outlined in black. While there is extensive research on humorous memes, research focusing on their pragmatic analysis is limited. Dynel (2016) was one of the first scholars to approach image macros from the perspective of pragmatics. She argues that this genre of memes functions as a particular type of visual-verbal joke, noting that in many cases, the verbal joke in the image macro is at least partly incomprehensible without the background knowledge of the larger meme family to which it belongs. Similarly, Dancygier and Vandelanotte (2017) used the linguistic framework of construction grammar to explain the interplay between the visual and verbal components involved in processing a prototypical image macro. They proposed a continuum of multimodal interaction, where on one end, the image is secondary, in that it supports the text, but it is not, strictly speaking, necessary for its interpretation. And on the other end, the image is essential to interpreting the larger meaning of the meme.

In his work on memes, Yus (2019) has demonstrated the applicability of relevance theory to understanding the interpretation of image macros in terms of their multimodal compositionality. Applying a relevance-theoretic perspective to the interpretation of online texts, Yus (2019) argues that “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” (p. 107). In other words, the meaning of a meme can only be interpreted through the identification of the connotative meanings of text, image, and text-image combinations.

Other studies, such as Zenner and Geeraerts' (2018), have placed special emphasis on the linguistic elements of internet memes. These authors explored different forms of wordplay commonly exploited in image macros, including homonyms, homophones, and paronyms (i.e., when two words or phrases sound similar, but have different meanings). Furthermore, Dynel (2016: 662) observed that popular memetic catchphrases often include a vernacular that is “replete with not only language mistakes but also abbreviations and acronyms as well as peculiar vocabulary,” and that this type of non-standard language “may be deemed a crucial element of Internet memes, which pivot on replication and imitation.” This observation clearly applies to the catchphrase associated with the meme that is our focus in this study.

Building on this previous research that addresses the verbal and visual elements of internet memes and how they interact with each other in creative and playful ways to generate humorous meanings, in this study we consider one specific meme i.e., “cash me ouside.” Although this meme centers on a specific catchphrase, the image macros that it inspires include many variations on both the texts and the images included. Our study also takes up Dancygier and Vandelanotte's (2017) call for

further study of the role of wordplay in internet memes, as we explore various types of linguistic humor that appear in our data. Finally, our study illustrates how the “cash me ousside” meme is primarily about an identifiable stance that is expressed via a recognizable, non-standard catchphrase, a visual reference to its speaker, or some combination of the two.

1.2. “Cash me ousside”

In 2016, on an episode of the American talk show, *Dr. Phil*, a 13-year old “troubled teen,” Danielle Bregoli, and her mother discussed their strained relationship. Early in the episode, Danielle said she had lied, cheated, and stolen cars. In response to the audience laughing at her outrageous behaviour and her comments, Danielle reacted with a tough-sounding threat, by defiantly challenging them saying, “Catch me outside, how about that!” Perhaps due to Danielle’s surprising display of audacity as well as to her idiosyncratic pronunciation of this particular utterance (i.e., /kɛʃ mi ɑːsɑː hɑ bɑ dəʔ/), the video of this media event instantly went viral, and “cash me ousside” became an overnight meme, generating a wealth of new, remixed and diverse multimodal reproductions in the form of image macros and YouTube videos. Although Danielle produced several other grammatically and phonetically “non-standard” language forms during the program, it was this single utterance – frequently accompanied by an image of her face – that became repeatedly entextualized (most typically represented as “cash me ousside, howbow dah”), reproduced, and remixed in multiple formats online. To internet users, there was something catchy and interesting about this utterance – starting with how the word *catch* was pronounced as *cash* by Danielle and extending to its larger illocutionary force – which stimulated much memetic activity.

Internet memes are ultimately about some expression of affect or emotions, such as rage, anger, joy, and excitement in reaction to real-world events (Denisova, 2019; Wiggins, 2019). Not unlike the popular Chinese meme discussed by Jiang and Vázquez (2020: 4), it is possible that “the visual representation of an extreme emotional expression” is one of the reasons for the productivity and spread of the “cash me ousside” meme as well. As will be discussed, the intensity of emotion associated with this meme is expressed both at the level of embodiment (i.e., the image of Danielle’s face alone is enough to call up associations with the memetic catchphrase) and at the level of language (i.e., the utterance clearly expresses a confrontational stance). Besides the expression of intense affect, the pronunciation of the catchphrase is also ambiguous, which makes it highly adaptable to further remixing, as well as amenable to playful transformations.

Our previous study of YouTube comments responding to the original *Dr. Phil* video found that internet users associated Danielle’s language use with a range of social categories such as race, class, as well as regions in the U.S., mainly the southern states (Aslan and Vázquez, 2018). In this study, we shift our focus to explore a set of “cash me ousside” image macros. Specifically, in order to expand on prior linguistic analyses of other categories of internet memes, we explore the different forms of linguistic humor found in our dataset of image macros. We were also interested in understanding the various ways in which verbal and visual elements are combined to create humorous meanings related to this meme. Our study is unique in its focus on a dataset of diverse image macros associated with one specific meme, which emerged from a mass media moment that went viral – featuring a young teenage girl who was unknown at the time.

1.2.1. Methodology

Our dataset comprises 220 multimodal image macros retrieved from three online platforms. Data were collected from Facebook (N = 42), Google (N = 135), and Knowyourmeme (N = 43) from May 2017 through June 2018. Collecting data within the two years after this media event went viral allowed us to gather a wide range of texts, resulting in a large and diverse dataset. To analyze the texts, we grouped together the image macros with shared linguistic characteristics. We then considered the distinctive linguistic features, visual components, as well as their interplay, in our analysis. Rather than devising a system of classification that would comprehensively account for every item in our dataset, our aim instead was to identify some of the most productive patterns or principles (i.e., “types of linguistic humor”) for creating diverse expressions of this particular meme.

2. Findings

Our analysis resulted in three categories representing different types of linguistic humor: 1) wordplay, 2) voicing, and 3) register humor. Below, we address both similarities as well as differences within each category.

2.1. Wordplay

This category comprises three subcategories: 1) wordplay exploiting the literal meaning of *cash*, 2) wordplay involving paronymy and word blends, and 3) wordplay exploiting the phonological similarity between *cash* and *cats*.

2.1.1. Literal meaning of *cash*

Paronymy is a productive resource for wordplay in internet humor (Vázquez, 2019). “Two words are paronyms when their phonemic representations are similar but not identical” (Attardo, 1994: 110–111). In the examples that exploit paronymy, the orthographic representation of *catch* as *cash* (which reflects Danielle’s pronunciation) is also represented semiotically. Therefore, these examples, which rely on the literal meaning of *cash*, build visually upon the relations of paronymy that are already embedded in the ambiguity of the original *catch/cash* utterance.

In Example 1, the *cash* from catchphrase, is interpreted literally in the top image, where the visual of a smiling cashier is reinforced with the textual offer, *I can cash you out over here*. This image macro uses a dual image format that invites the initial processing of the top half, which provides the linguistic set-up, followed by the bottom half, which includes the complete catchphrase and serves as the “punchline.” Here, the bottom half of the image macro consists of a visual of Danielle on the set of the *Dr. Phil* program, with the complete catchphrase overlayed on the image. The incongruity presented here lies in the contrast of the polite service encounter with a smiling cashier juxtaposed with the more aggressive stance presented in the original “cash me ousside” utterance.



Example 1. Cash You Out.

Example 2 is identical in format, and also relies on the literal meaning of *cash* as money. This time, the service interaction setting represented in the top panel is that of a fast-food restaurant drive-through, with the employee explaining that the credit card machine is not working, so *we can only get cash*. The image of Danielle with her catchphrase again appears in the lower panel.



Example 2. Credit Card Machine is Broke.

Although it too uses a format of two stacked horizontal panels and an image of a service encounter, Example 3 deviates from the two previous examples in that the textual elements appear within speech bubbles rather than in the all caps letters in Impact font (the font most typically associated with internet memes). Only Danielle's decontextualized face appears here, photoshopped onto an image of a check in the lower of the two panels. This visual decontextualization means that, in order to understand the reference, viewers need to rely much more on the textual components in the third example than they did in the previous two examples, where more complete images of Danielle's torso appeared shown in the meme's originating context: the set of the *Dr. Phil* show.



Example 3. You Can't Cash This Here.

These examples all share semantic and semiotic representations of the literal meaning of *cash* as money. In them, Danielle's face and her signature catchphrase appear in the bottom panel and are juxtaposed with a customer service representative making some unmarked statement involving the word *cash* in the top panel. The humorous effect in these examples is conveyed through depictions of unexpected interactions involving Danielle and different staff in polite service encounter contexts focusing on the paronymic word “cash.”

2.1.2. Blends with cash

In the next examples, Danielle's unconventional pronunciation of *catch* as *cash* once again receives focus and becomes the primary source of wordplay. However, examples of wordplay in this category feature larger words which contain the element *cash* in them. Example 4 is more cognitively demanding and structurally distinct from the previous examples, both in terms of the use of the catchphrase and its visual format. This example relies to a far greater extent on viewers' existing knowledge of the meme in order to be meaningful in any sense. In addition, the visual format also differs from most of the other image macros in the dataset: it is comprised of four panels with no superimposed text. The catchphrase is only indirectly referenced here through two images of Danielle's face and a bit of text that appears on a product package (i.e., *cash*). This example requires the audience's participation in terms of accessing the necessary background knowledge as well as more inferencing since the humor is not explicitly conveyed. Denisova (2019: 31) attributes this complexity to the “coded nature” of memes, meaning that in order to unpack memes, one usually needs to be aware of some cultural, social, political and/or internet-specific references. Since the catchphrase itself is not presented at all, viewers must be able to identify the face of Danielle and already know that she is the originator of the memetic “cash me ousside” utterance. Only after that connection has been made, viewers can then make the additional sound-based connection between *cash* and *cashew*, the latter of which is emphasized in the cropped, close-up of the product label in the lower right panel. This is what Dancygier and Vandelanotte (2017: 567) refer to as frame metonymy “by which characteristic parts of a frame are sufficient to call up whole frames.” This example presents an extreme case of incongruity, where there is no ideational connection whatsoever between Danielle and the bag of cashews. Whereas the top left panel shows Danielle with a neutral expression, the bottom left panel shows Danielle with an angry or defiant expression, which is then echoed in the close up of the product label in the bottom right panel, into which the word “triggered” has been inserted.

Example 5. Cashmere.



Example 4. Cashews.

Example 5 is more conventional in its presentation, in terms of its dual image format, similar to Examples 1–3. However, it departs from those examples in that the image of Danielle, along with a modified version of the catchphrase, appears in the top half rather than the bottom half. Here, “cash me outside” has been humorously transformed into the paronymic blend, *cashmere outside*, echoed in the bottom half with an image of what is presumably a cashmere goat. The wordplay continues in the bottom half, where the second part of the represented catchphrase includes the onomatopoetic representation of a goat bleating (with vowel repetition in the final word – i.e., *daaaah*, reminiscent of a goat’s *baaah*). Example 5 is unusual in its integration of wordplay in both halves of the image macro (*cashmere* + [goat] + *daaaah*/[*baaah*]). The image of Danielle that appears in the top portion is one in which she has an angry expression, similar to the one on the bottom left panel in Example 4. Whereas Examples 1–3 seemed to highlight the contrast between a polite, or at least ordinary, service encounter utterance with Danielle’s more aggressive or challenging stance, the relative placement of the two text-image components in Example 5 suggest instead a relation of comparison – i.e., Danielle’s speech is analogous to the sounds produced by an animal.

Unlike the examples in the previous subcategory, examples in this category do not rely on the literal meaning of *cash* as *money*. Instead, in these examples, *cash* appears as a sub-component of some larger word (e.g., *cashew*, *cashmere*), and it is this resulting new word that is reinforced semiotically, as images of it appear in the image macro, along with some representation of Danielle. Examples in this category are highly incongruous in terms of the ideational meanings of the visual elements included (i.e., there is no obvious relationship between Danielle and cashew nuts or cashmere goats); rather, the emphasis seems primarily to be on the wordplay itself and an accompanying image in which this resulting new word is represented along with Danielle’s defiant facial expression that contributes to the humorous effect.

These two examples show that even when the same linguistic principle is involved in generating humor (i.e., wordplay based on word blending), there may be differences in the ways the same affective meanings are conveyed. Both Examples 4 and 5 include an image of Danielle with her signature defiant facial expression along with some unrelated visual element related to the resulting word blend (i.e., *cashew* and *cashmere*). However, Example 4 includes no explicit reference to the catchphrase, yet the embodied affective meaning of aggression is reinforced with a new textual element (“triggered”), a label that accounts for Danielle’s angry facial expression. Moreover, the word *cashew* is completely random, and is included presumably only for its phonological relationship with *cash*. In contrast, Example 5 does include an adapted version of the complete catchphrase, as well as the same angry facial expression. Yet at the same time, due to the ordering of visual elements, it is possible Example 5 is also ridiculing Danielle’s verbal expression – an interpretation that would be supported by superiority theories of humor. These examples demonstrate that while user creativity is evident here in the various associations made with incongruous elements external to the meme itself, these need to be accompanied by more stable and fixed visual and textual elements (e.g. Danielle’s face [parts of] the catchphrase – or both) to generate humorous meanings.



Example 5. Cashmere.

2.1.3. Phonological similarity between cash/cats

Cats have been a productive source of internet humor since the emergence of *LOLCats* in the mid 2000s. Miltner (2014) contends that “*LOLCats* are a key example of the types of content that are changing the way people engage in cultural participation, creative engagement, community interaction, and identity construction.” The basis of lolspeak – an internet neologism referring to the language associated with *LOLCats* – as a creative source of humor are anthropomorphized cat images with non-standard language superimposed upon them. Our dataset included numerous instances of image macros featuring visuals of cats in combination with some variation of the catchphrase as the textual component – typically with no visual references to Danielle. Once again, these types of image macros can be understood as incongruous, since these animals have no immediate association with Danielle, nor with the meaning of the catchphrase.

Thibault and Marino (2018) argue that cats always have the potential to become memes and that “any other meme can be catified featuring countless variations, iterations, and proliferations” (p. 487). The authors attribute cats’ popularity as internet memes to two features: cats’ anthropomorphic nature which allows them to be used to convey human feelings, emotions and behaviors and cats’ naivety and spontaneous silliness which makes them endearing and encourages people to share content that depicts them. Cat images appear especially productive in this context, given the phonological similarity between *cats* and *catch*. In the textual components in all of the cat-featuring examples, the *catch* or *cash* from Danielle’s original catchphrase is further reduced to *cat* or *cats*. Very often, this is then followed by an orthographic representation of the (phonologically reduced) elements “me” + “ou [w]” + “[tside],” which creates another paronymic form: the onomatopoetic *meow* appearing in Examples 6–8. This morphophonemic wordplay is paralleled visually through the accompanying images of various cats.



Example 6. Cats Meowside.

In Example 6, the cat in the image stands upright on its hind legs and adopts a possibly confrontational physical posture, which may serve as a visual reference to Danielle confronting the audience in the original show. Similarly, the cat's sideways-turned baseball cap and athletic wear in Example 7 suggests an element of toughness. The humor effect these two examples convey has to do with the aggressive and tough-sounding nature of the catchphrase (i.e., a threat or a challenge), juxtaposed with images of small, adorable, and decidedly non-threatening-looking cats.



Example 7. Cat Meow Side.

Example 8 is similar to the previous two examples, with its variation on the catchphrase including both *cat* and *meow*. However, this is the only example in the cat category that includes a visual reference to Danielle; the partial wig the cat is wearing is undoubtedly a reference to Danielle's hair. This is another example of frame metonymy (Dancygier and Vandelanotte, 2017) – in this case, semiotic frame metonymy – where only one part (i.e., a few wisps of hair) is used to index the larger frame of Danielle's head.



Example 8. Cat Meow Outside.



Example 9. Cats Be Outside.

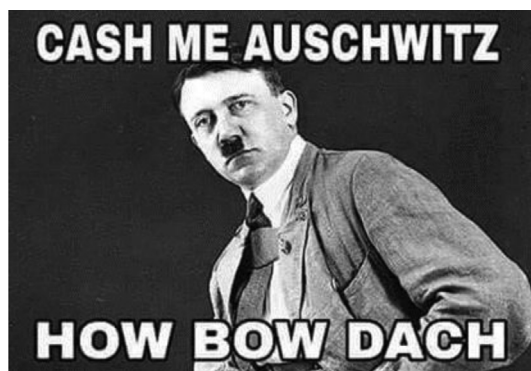
Example 9 is slightly different, because rather than appearing in the top part of the catchphrase (where it would have a paronymic relationship with the original utterance), the *meow* appears in the bottom element instead, in the second half of the catchphrase, where it is used to simply substitute the word *that*. And while the wordplay in the top segment does include the transformation of *catch* to *cats*, it then continues in a different vein, with what might be a reference to African American Vernacular English (AAVE) in the use of invariant *be*: i.e., *Cats be outside*. This interpretation is consistent with our previous study of YouTube metacommentary on “cash me ousside” (Aslan and Vázquez, 2018), where many users’ comments interpreted Danielle’s idiosyncratic pronunciation as sociolinguistic crossing into AAVE. The image of two cats in an outdoor setting reinforces the literal meaning of *outside* in the catchphrase, and again, any visual connotations of toughness in this image are humorously ironic.

The examples discussed in this cat meme category show variation in the relationship between the image depicted and the illocutionary force of the transformed catchphrase. In several instances, the ironically depicted toughness associated with the cats’ position or stance may serve as visual allusions to Danielle’s tough sounding threat in the original show. The threat is conveyed through visual elements, such as a confrontational physical posture, or athletic wear possibly denoting toughness, and is juxtaposed with images of small, non-threatening looking cats. However, the focus of the humor in these examples seems to reside primarily in the paronymic wordplay involving the catchphrase. Additionally, the lack of visual references to Danielle and the similarity of pronunciation between “cash me ousside” and “cat meow side”/“cats be outside” in Examples 6–9 suggests that background knowledge about the catchphrase is essential in humor resolution.

2.2. Voicing

In our next category, the catchphrase was transformed into a linguistic style associated with a specific individual, fictional character, or historical figure. Imitation of others’ language is often a key element of memes (Milner, 2016). In these cases, the

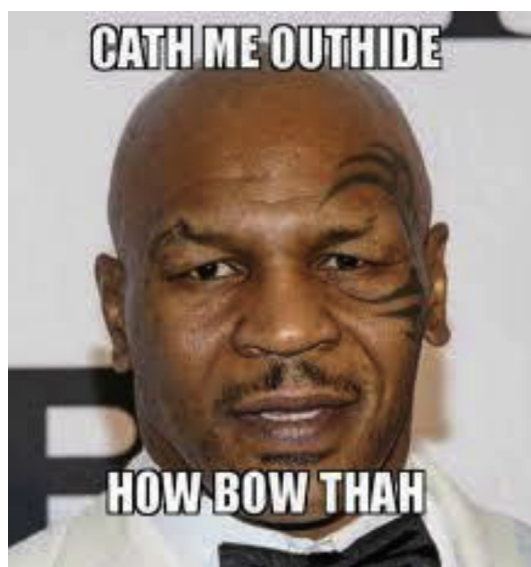
juxtaposition of an image of some recognizable figure with the catchphrase – expressed using a very particular orthographic or syntactic style – makes it appear as though the individual depicted in the image is “voicing” the catchphrase. This type of voicing is a productive strategy in many genres of internet humor found across multiple platforms (Vázquez, 2019).



Example 10. Cash me Auschwitz.

In Example 10, Adolph Hitler is depicted as the central figure and the catchphrase is transformed to be congruent with this image: *outside* is changed to *Auschwitz* and the final word (i.e., *that*, represented most commonly in our dataset as *dah*) is substituted here by the German word, *dach*. Beyond the modified catchphrase, there is no visual representation of Danielle here; however, the figure of Hitler is obviously associated with aggression and violent threats. In order for these types of voicing examples to be effective, there must be some correspondence between the individual depicted and a variation of the catchphrase that in some way references this particular individual's manner of speaking. In Example 10, the use of German words connects the text to Hitler.

Similarly, in Example 11, it is the orthographic representation of Mike Tyson's characteristic lisp (i.e., *cath*, *outhide*) that links the catchphrase to the famous boxer. However, if the image of Mike Tyson had been substituted with an image of Danielle on the set of the *Dr. Phil* program (as appears, for instance, in Examples 1 and 2), it would probably not be possible for viewers to make the association with Tyson's speech specifically, simply from this particular orthographic transformation of the catchphrase (In that case, it would most likely appear to viewers as “funny spelling,” with no immediate associations with any specific speaker.) For these examples of voicing to be effective then, there must be some correspondence with the linguistic transformation of the catchphrase and the individual depicted. It is also worth pointing out that Tyson is a media figure associated with aggression and physical violence. Yet, not all images in this category are as menacing, as can be seen in the next example.



Example 11. Cath me outhide.

Like images of cats, the enigmatic figure of Yoda (from the *Star Wars* films), with his distinctive syntax, is also a productive element found in many internet memes. The site *knowyourmeme* lists several variations of Yoda memes and thus, Example 12 may be considered a “meme blend” (Dancygier and Vandelanotte, 2017) in terms of both its visual and textual components. Specifically, Yoda’s face, ears, and light saber have been photoshopped onto Danielle’s hair and body, with the *Dr. Phil* set in the background. Moreover, the textual component, the catchphrase, has been syntactically inverted, according to the image macro creator’s understanding of how “Yoda syntax” works. Furthermore, the selection of the character of Yoda here is not coincidental, and is a key element in creating humor in this example. As is explained on *knowyourmeme*: “The character Yoda is known for his wisdom, thus this specific way of talk is associated with wise and sagacious thoughts. Using yodaspeak to say something obvious, vulgar or common creates a comic and absurd situation.”¹ Similar to the additional layer of meaning we suggested in Example 4 – as mocking or condescending with respect to Danielle’s speech – in Example 12, selecting the character of Yoda to voice the catchphrase may serve as a way of highlighting the “vulgarity” or “absurdity” of Danielle’s utterance. Milner (2016: 62) has pointed out that, especially in meme blends, “intertextual references layer one after another as content spreads.” In this instance, which relies on the intertextual insertion of another figure from a different popular culture source, the image of the wise character of Yoda uttering Danielle’s colloquial catchphrase is what creates multimodal incongruity. Furthermore, the addition of Danielle’s coral lipstick onto Yoda’s face represents a form of visual incongruity and also, no doubt, contributes to the humor in this multimodal ensemble.



Example 12. Outside cash me.

In examples of this category, the individuals depicted are represented and imagined as voicing the memetic catchphrase. Furthermore, in these examples, the visual and verbal components are highly interdependent in that the visual reference to a particular speaker is necessary to interpret the specific, idiosyncratic linguistic transformation of the catchphrase. Once again, while some of the individuals depicted are associated with a stance of toughness and aggression (e.g., Hitler, Tyson), which corresponds to the original illocutionary force of the catchphrase, in other cases (i.e., Yoda) a sense of multimodal playfulness – and possibly also condescension – is highlighted instead.

2.3. Register humor

The third and final category of examples exploits register humor in various ways. Venour et al. (2011) describe register humor as a situation where most of the language in a text appears in a particular style or tone, except for a few words which are expressed in a register that is somehow different from the rest. While register is not a precisely defined concept, it is related to factors such as formality, literariness, archaicism, and so on. Alexander (1984) views register humor as the “comical confusion” of two registers that usually involves the selection of “a lexeme or phraseological unit from a different style level than the context would predict” (Alexander 1984, p. 60). Similarly, Piata, 2020 points out that register humor arises from the incongruent mixing of different stylistic varieties. In other words, register humor relies on subverting expectations about a stylistic variety that is considered natural and/or appropriate in a particular situation. This register incongruity creates a clash of styles – and in several of these examples of the “cash me outside” meme, the clash of styles is achieved either linguistically, or visually, or through a combination of both.

¹ <https://knowyourmeme.com/memes/yodaspeak>.

One subset of these examples features Danielle's face along with a lengthy text, in which her original highly-colloquial and very concise utterance has been transformed into a much more formal and verbose paraphrase. Example 13, which features a recognizable image of Danielle with an angry expression from the set of the *Dr. Phil* program, presents a clear instance of register incongruity in that the first part of the text appears in a hyperformal register, perhaps associated with legal (e.g., *adjourn*, *reconvene*) or educational (e.g., *dissertation*) contexts, which then contrasts sharply with the hyper-colloquial and orthographically non-standard *Howbow dah*, which appears below.



Example 13. Adjourn this discussion

Similarly, the catchphrase in Example 14 appears transformed as a hyper-standard and much more verbose reformulation, while retaining the illocutionary force of “cash me ousside, how about that.” In this example, the word *fisticuffs* is not merely more formal but is also an archaicism indexing a more historical register. In addition, in Example 14 a pair of nerdy glasses has been photoshopped onto Danielle's face, as though this were the type of thing one might expect to hear from a more intellectual version of Danielle. This added semiotic element serves to further emphasize the incongruity between the more formal textual message presented here and the original colloquial utterance from which it was derived.

Example 15 offers a multimodal variation on this types of register humor. It consists of a sequence of three, horizontally-arranged text-image combinations which, when read from top to bottom, increase in formality, verbosity, syntactic complexity, as well as indirectness. The confrontational or aggressive illocutionary force of the utterance remains constant, but its linguistic realization varies – similar to Examples 13–14. Furthermore, as the language in Example 15 increases progressively from the highest level of directness and informality to the highest level of indirectness and formality, the corresponding visual representation of Danielle becomes increasingly more and more abstracted as well.

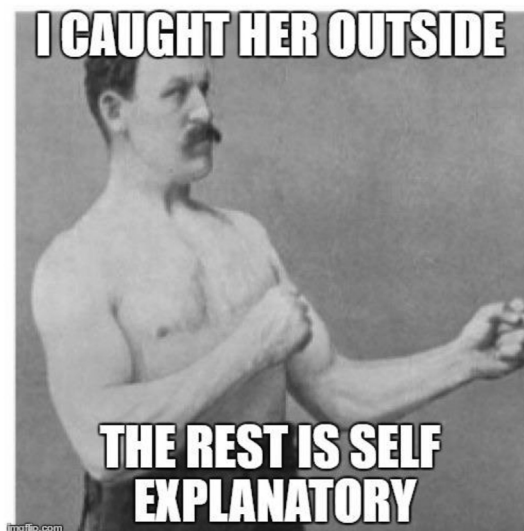


Example 14. Confront me Outdoors.



Example 15. Accost my acquaintance.

Other examples in this category also play with the trope of juxtaposing the archaic with the contemporary, yet they rely on other types of images. Example 16, for instance, is a meme blend. The image of this 19th century boxer in so-called “fisticuffs stance” appears in a meme known as “overly-manly-man,” which is characterized by this image combined with, according to knowyourmeme.com, “captions portraying the subject as an overtly masculine alpha male with misogynistic attitudes.”² This image macro includes a very indirect reference to “cash me ouside” in the syntactically-transformed element at the top (i.e., *I caught her outside*). In this image macro, Danielle’s threat of a physical confrontation to the studio audience is also invoked, but in a joke-is-on-her manner, meaning that the overly-manly-man, in fact, catches her outside and confronts her physically instead, which is implied by the bottom caption: *the rest is self-explanatory*. It is possible, in examples such as this one, that when there are additional levels of inferential processing required for their interpretation of a meme, viewers derive greater satisfaction as they attempt to decipher the joke or humor, compared to those cases involving more straightforward interpretations. Furthermore, it may also be the case that, when decoding such a text, the less explicit a connection is between the original catchphrase and its memetic transformation, the more this contributes to feelings of “in-groupness” among viewers, who may be aware that they share a sensibility as well as an understanding of something akin to an “inside joke.”



Example 16. I caught her outside.

² <https://knowyourmeme.com/memes/overly-manly-man>.



Example 17. Mayhaps we shall meet outdoors.

The black and white image in Example 17 also does not incorporate any visual references to Danielle. In it, the text that appears along with the image of the Victorian woman depicted shares some similarities with the text in Example 14. However, in Example 17, register humor is exploited further, as the textual component is additionally archaized not only with *fisticuffs* but also with lexical forms such as *mayhaps* and *shall*, associated with more historical speech. In this example, both the text and the image work together to communicate in the same historical register – and it is only through the presupposed background familiarity with Danielle's colloquial and non-standard catchphrase that the original utterance comes across as stylistically incongruous. In other words, the original source is only indirectly alluded to through the verbose, “historicizing” paraphrase of Danielle's utterance, thereby requiring the audience to be aware of the original wording of the meme's catchphrase in order to understand and appreciate the incongruity.

In these black and white “historical variations” (Examples 16–17), the reference to this meme is only textual, made through some variation of the catchphrase that is presented with no visual reference to Danielle. Conversely, other examples in this category (Examples 13–15) do include images of Danielle. It is interesting to note that, in some of these examples, Danielle's image has also been modified (i.e., the inclusion of glasses in Example 14 to create a more intellectual look; or the increasing abstraction of Danielle's face in Example 15). These findings indicate that incongruity in register humor may occur not only at textual level but also at the visual level. More specifically, the iconic and unexpected images from history (Examples 16–17) appear in juxtaposition with a contemporary vernacular text (albeit, a highly transformed one that is only indirectly alluded to), creating humorous multimodal ensembles. In some instances of this category, the meme's overall stance of confrontation is reinforced visually (Example 16), whereas in other examples (Example 17), the primary focus seems to be more on the playful linguistic transformations of the catchphrase itself.

3. Discussion

Our analysis of a related set of image macros, which referenced a specific internet meme (“cash me ousside”), has addressed the particular linguistic and semiotic resources through which humorous meanings were generated. As our analysis revealed, the internet memes in our study are complex artifacts and they exhibit unique multimodal juxtapositions, many of which involve far more variation than the prototypical image macro structure (i.e., simple top/bottom texts superimposed onto a single image), which has been the focus of most previous research on the topic (e.g., [Dancygier and Vandelanotte, 2017](#); [Dynel, 2016](#); [Zenner and Geeraerts, 2018](#)).

The most common forms of linguistic humor we identified included wordplay as well as humor involving shifts in linguistic style or register. With respect to wordplay, our dataset included several instances of paronymy related to the *catch/cash* ambiguity of the original utterance, including a reliance on the literal meaning of the word *cash*, or the incorporation of other, unrelated, words subsuming the element *cash* (e.g. *cashews*, *cashmere*). In addition, drawing on a popular source of internet humor, other examples further reduced the *catch/cash* lexical element to *cats*, which enabled meme producers to combine feline-focused transformations of the catchphrase with visuals of cats. Aligning with [Dynel \(2016\)](#), we speculate that the non-standard pronunciation of the original utterance (i.e., *cash me ousside*) was not only what drew viewers' attention to it in the first place, but was perhaps also what opened up possibilities allowing for further creative verbal transformations. As far as

register humor is concerned, most transformations of the catchphrase involved establishing a contrast between the brevity and the highly vernacular character of the original utterance with some alternatively-formulated version of the same threat, involving greater verbosity, greater syntactic complexity as well as lexical items associated with other “higher” registers (e.g., legal, educational, historical).

Next, we found that both textual and visual elements showed great variation – and interacted with one another in diverse ways. Despite this variation, our study shows that it is the catchphrase and/or some visual referent to Danielle (which functions as a visual index of the catchphrase) that unifies all these very diverse memes. Visual elements included not only those extracted from the original memetic video (i.e., Danielle’s face), but also included public figures, characters from popular culture sources, animals, historical images as well as images from other everyday contexts, leading to vernacular creativity (Burgess, 2008). Especially in instances of voicing, there was a tight integration of text and image, as the transformed catchphrase was linked to the idiosyncratic speaking style of the figure depicted. Overall, the memetic humor we observed in our dataset exemplifies transformative imitation, meaning that there is a playful interaction between some essential and fixed component in combination with some novel, creative, and unexpected user-generated elements. The essential components – whether visual or verbal – are what makes the meme recognizable as such. Like other meme researchers, we also identified numerous meme blends in our dataset, in which some element(s) of “cash me outside” were playfully merged with other popular internet memes (e.g., overly-manly-man, yoda, cats), highlighting the multi-layered intertextual nature of memetic texts.

In line with the discussion found in Yus (2019), our findings confirm that multiple inferential strategies must be employed in the interpretation of image macros, which require taking into account the combination of both textual and visual sources of information to yield humorous meanings. As in most cases of memes, viewers’ background knowledge about current affairs, political matters and (in our case) viral media events is crucial to interpreting the meme correctly. While the catchphrase (“cash me outside/how bah dah”) itself is clearly salient, one also needs to be at least somewhat familiar with the context in which this phrase was uttered, particularly the speaker, the addressees, the communicative situation, the physical setting, other interlocutors involved, and the general affective stance being communicated, in order to establish connections between the visual and verbal elements that comprise this set of image macros. For example, the decoding of the creative wordplay examples derived from the reduction of the word *catch/cash* to *cat* serves as a prime example of establishing verbal-visual implicatures and humorous meanings. In other words, if a viewer is not familiar with the catchphrase of this meme, they will neither be able to correctly interpret a verbal joke such as *cats be outside/how about meow* (Example 10), nor will they be able to understand it solely from the visual referent, the cat image. Similarly, interpreting implicatures and humor poses more of a challenge in image macros that do not include Danielle as a visual referent (Examples, 10, 11, 16, and 17). These findings echo Yus’s (2020) observation about incongruity resolution involving the convergence of partial meanings from both visual and textual components, which results in humorous meanings.

4. Conclusions

Internet phenomena, such as the one described in our study, often go viral when they have salient, catchy, memorable or dramatic features that are open to recontextualization and creative exploitation (Rymes, 2012). In the case of “cash me outside,” the implied (though dubious) physical confrontation from an unlikely source – a young teenage girl threatening an audience full of adults – and the associated stance of aggression and toughness forms the premise for much of the resulting memetic humor. It is this extreme emotion conveyed by the speaker in expressing an obviously confrontational stance (in the original video), combined with an especially “catchy” catchphrase – characterized by an ambiguous and non-standard pronunciation – that made this meme especially open to recontextualization and creative adaptation. When this confrontational stance is juxtaposed with other elements (which are sometimes related, and sometimes completely unrelated to that stance), a shared cycle of humorous communication is created and sustained.

As other scholars have noted, internet memes are not only shared but they also invite participation, encouraging others to contribute to a form of textual and visual entertainment (Börzsei, 2013). Users must possess the requisite background knowledge needed to interpret their meanings (Denisova, 2019). The virality of this meme’s catchphrase stems as much from the distinctive way it was uttered (i.e., its phonological expression), as it does from its non-propositional content or meaning. Specifically, Danielle’s use of a markedly “non-standard” English pronunciation (albeit one that is difficult to ascribe to any specific social, regional, or ethnic variety) seems to have captured the attention of many internet users. Furthermore, its linguistic ambiguity, which, in conjunction with the strong affective meaning conveyed through Danielle’s facial expressions, is what allows for such a wide array of creative recontextualizations. In sum, it appears that the confrontational stance combined with the linguistically ambiguous and idiosyncratic expression of the catchphrase is what makes this meme particularly productive for different types of linguistic humor.

Internet memes rely on two processes, virality and adaptation. To be recognizable as a new expression of an existing meme requires the inclusion of some “fixed” or clearly “identifiable” element(s), which link it to prior expressions of that meme. As we have shown in our analysis, these links can be relatively more explicit (Examples 1 and 2) or far less explicit (Examples 4 and 16). Rather than slavishly copying existing tropes or patterns, truly creative adaptations require the introduction of increasingly novel, unexpected, or not-yet-seen-before-in-this-context types of content. This means that intra-memetic variation can be vast and unpredictable in terms of generating unique, individual expressions – and this unpredictability poses clear challenges for analysis. Nevertheless, our study has demonstrated that recurring patterns associated with specific

types of humor can also be detected across multiple instances of one meme. While, in theory, there is an infinite set of content that can be remixed together, in actual practice, most internet memes draw from a relatively restricted range of content such as recognizable popular culture references (e.g., yoda or cat images), as discussed in Vázquez (2019). One explanation for this tendency is the desire to communicate some larger stance, sensibility or emotion, in a way that will be somehow legible to, and/or resonate with, others.

A final factor may help explain why a trivial media event involving a teenage girl confronting an audience led to humorous uptake in the form of internet memes. As mentioned in our introduction, superiority is sometimes a feature of internet memes, meaning that “some people enjoy not only watching videos of others whom they perceive to be inferior, but also take pleasure in scornfully imitating them, thus publicly demonstrating their own superiority” (Shifman, 2014, p. 81). As we discussed in Aslan and Vázquez (2018), Danielle was ridiculed and mocked for her non-standard speech in the original talk show by both the program's host and audience. Taking into account how this meme originated, it is possible that some of the humorous uptake here may be linked to Danielle being regarded as inferior, uneducated, and troubled – indexed by her idiosyncratic way of speaking – which was ridiculed and mocked by both the program's host and audience, and later by internet users (Aslan and Vasquez, 2018). That being said, while the superiority feature may account for some of the memes in our data (e.g., Examples 5 and 12), we do not claim that it is a salient feature in all of the instantiations of the meme, especially not for those that include no visual references to Danielle.

Building on our findings of our study, future research could shed light on the humorous mechanisms involved in other, related digital genres, such as Graphic Interchange Format (GIFs), which also comprise a few seconds of an emotionally-charged media moment, and which also involve multimodal play. Our study shows how memes are complex multimodal texts, composed of multiple layers of meaning, and how, as such, their interpretation depends on advanced inferential strategies. By combining, and intertextually layering, various linguistic and semiotic resources, meme creators generate meanings based on emotions, daily experiences and everyday situations. Internet memes allow not just sharing, but also the co-creation of economically expressed, clever cultural artifacts, which rely on ideas and emotions assumed to be shared (or at least known) by peers within a given discourse community.

Funding

This project is not funded by any funding agencies.

References

- Aslan, Erhan, Vázquez, Camilla, 2018. “Cash me ouside”: a citizen sociolinguistic analysis of online metalinguistic commentary. *J. Sociolinguistics* 22 (4), 406–431.
- Alexander, Richard, 1984. Verbal humor and variation in English: sociolinguistic notes on a variety of jokes. *Beitriige. Fremdsprachenvermittlung. Konstanzer Sprachlerinstitut (SLI)* 14, 53–63.
- Börzsei, Linda K., 2013. Makes a meme instead: a concise history of Internet memes. *New Media Stud. Mag.* 7, 1–28.
- Burgess, Jean, 2008. “All your chocolate rain are belong to us?”: viral video, YouTube and the dynamics of participatory culture. In: Lovink, G., Niederer, S. (Eds.), *Video Vortex Reader: Responses to YouTube*. Institute of Network Cultures, Amsterdam, pp. 101–109.
- Chovanec, Jan, Tsakona, Villy, 2018. Investigating the dynamics of humor: towards a theory of interactional humor. In: Tsakona, V., Chovanec, J. (Eds.), *The Dynamics of Interactional Humor*. Benjamins, Amsterdam, pp. 1–28.
- Dancygier, Barbara, Vandelanotte, Lieven, 2017. Internet memes as multimodal constructions. *Cognit. Ling.* 28 (3), 565–598.
- Denisova, Anastasia, 2019. *Internet Memes and Society: Social, Cultural, and Political Contexts*. Routledge, New York.
- Dynel, Marta, 2016. “I has seen image macros!” Advice animal memes as visual-verbal jokes. *Int. J. Commun.* 10, 660–688.
- Ferguson, Mark, Ford, Thomas, 2008. Disparagement humour: A theoretical and empirical review of psychoanalytic, superiority, and social identity theories. *Humour: Int. J. Humour Res.* 21, 283–312.
- Jiang, Yaqian, Vázquez, Camilla, 2020. Exploring local meaning-making resources: a case study of a popular Chinese internet meme (*biaoqingbao*). *Internet Pragmatics* 3 (2), 260–282. In press.
- Knobel, Michele, Lankshear, Colin, 2007. Online memes, affinities, and cultural production. In: Knobel, M., Lankshear, C. (Eds.), *A New Literacies Sampler*. Peter Lang, New York, pp. 199–229.
- McGhee, Paul E., 1979. *Humor: its Origin and Development*. Freeman, San Francisco-Reading.
- Milner, Ryan, 2016. *The World Made Meme: Public Conversations and Participatory Media*. MIT Press, Cambridge, MA.
- Miltner, Kate M., 2014. “There's no place for lulz on LOLCats”: the role of genre, gender, and group identity in the interpretation and enjoyment of an Internet meme. *First Monday* 19 (8). <https://doi.org/10.5210/fm.v19i8.5391>.
- Piata, Anna, 2020. Stylistic humor across modalities: the case of Classical Art Memes. *Internet Pragmatics* 3 (2), 174–201. In press.
- Rymes, Betsy, 2012. Recontextualizing YouTube: from macro–micro to mass-mediated communicative repertoires. *Anthropol. Educ. Q.* 43 (2), 214–227.
- Shifman, Limor, 2011. An anatomy of a YouTube meme. *New Media & Soci.* 14 (2), 187–203.
- Shifman, Limor, 2014. *Memes in Digital Culture*. MIT Press, Cambridge, MA.
- Seiffert-Brockmann, Jens, Diehl, Trevor, Dobusch, Leonhard, 2018. Memes as games: the evolution of a digital discourse online. *New Media & Society* 20 (8), 2862–2879.
- Suls, Jerry, 1972. A two-stage model for the appreciation of jokes and cartoons: an information processing analysis. In: Goldstein, J., Paul, McGhee (Eds.), *The Psychology of Humor*. Academic Press, New York, pp. 81–100.
- Thibault, Mattia, Marino, Gabriele, 2018. Who run the world? Cats: cat lovers, cat memes, and cat languages across the web. *Int. J. Semiotics. Law-Revue internationale. Sémiotique juridique* 31 (3), 473–490.
- Vázquez, Camilla, 2019. *Language, Creativity and Humour Online*. Routledge, London.
- Venour, Chris, Ritchie, Graeme, Mellish, Chris, 2011. Dimensions of incongruity in register humour. In: Dynel, M. (Ed.), *The Pragmatics of Humour across Discourse Domains*. John Benjamins, Amsterdam, pp. 125–144.
- Wiggins, Bradley E., 2019. *The Discursive Power of Memes in Digital Culture: Ideology, Semiotics, and Intertextuality*. Routledge, New York.
- Yus, Francisco, 2019. Multimodality in memes. In: Bou-Franch, P., Garcés-Conejos Blitvich, P. (Eds.), *Analyzing Digital Discourse: New Insights and Future Directions*. Palgrave Macmillan, Cham-Switzerland, pp. 105–131.
- Yus, Francisco, 2020. Incongruity-resolution humorous strategies in image macro memes. *Internet Pragmatics*. <https://doi.org/10.1075/ip.00058.yus>.

Zenner, Eline, Geeraerts, Dirk, 2018. One does not simply process memes: image macros as multimodal constructions. In: Winter-Froemel, E., Thaler, V. (Eds.), *Cultures and Traditions of Wordplay and Wordplay Research*. Mouton de Gruyter, Berlin, pp. 167–194.

Camilla Vázquez is Professor of Applied Linguistics at the University of South Florida, where she teaches courses in Discourse Analysis, Pragmatics and Sociolinguistics. She is the author of *The Discourse of Online Reviews* (Bloomsbury, 2014) and *Language, Creativity and Humour Online* (Routledge, 2019). Her research about online discourse has also appeared in journals such as *Current Issues in Tourism*, *Discourse Context & Media*, *Food & Foodways*, *Internet Pragmatics*, *Journal of Sociolinguistics*, *Language@Internet*, *Narrative Inquiry*, and *Visual Communication*.

Erhan Aslan is Lecturer in TESOL and Applied Linguistics at the University of Reading. He received his Ph.D. in Second Language Acquisition and Instructional Technology from the University of South Florida. His current research interests include language in the digital media, particularly internet memes and online metacommentary. His work appeared in journals, such as *Journal of Sociolinguistics*, *Pragmatics and Society*, and *CALICO Journal*. He is also one of the co-authors of *Language and Media* (2nd edition, Routledge, 2020).