Emoji-based semantic representations for abstract and concrete concepts

Published Version:
Wicke P., Bolognesi M. (2020). Emoji-based semantic representations for abstract and concrete concepts. COGNITIVE PROCESSING, 21 Special issue Eliciting Semantic Properties: Methods and Applications(4), 615-635 [10.1007/s10339-020-00971-x].

Availability:
This version is available at: https://hdl.handle.net/11585/768985 since: 2020-11-23

Published:
DOI: http://doi.org/10.1007/s10339-020-00971-x

Emoji-based semantic representations for abstract and concrete concepts

Wicke Philipp, Bolognesi Marianna

Philipp Wicke

School of Computer Science

University College Dublin

Belfield, Dublin, D4 - Ireland

philipp.wicke@ucdconnect.ie

Dr Marianna Bolognesi

Faculty of Medieval and Modern Languages

University of Oxford

47 Wellington Square

OX1 2JF, Oxford, UK

marianna.bolognesi@mod-langs.ox.ac.uk

ORCID: 0000-0002-3292-8968

ABSTRACT

An increasingly large body of converging evidence supports the idea that the semantic system is *distributed* across brain areas and that the information encoded therein is *multimodal.* Within this framework, feature norms are typically used to operationalize the various parts of meaning that contribute to define the *distributed* nature of conceptual representations. However, such features are typically collected as verbal strings, elicited from participants in experimental settings. If the semantic system is not only distributed (across features) but also *multimodal*, a cognitively sound theory of semantic representations should take into account different modalities in which feature-based representations are generated, because not all the relevant semantic information may be easily verbalized into classic feature norms, and different types of concepts (e.g., abstract vs. concrete concepts) may consist of different configurations of non-verbal features.

In this paper we acknowledge the multimodal nature of conceptual representations and we propose a novel way of collecting non-verbal semantic features. In a crowdsourcing task we asked participants to use emoji to provide semantic representations for a sample of 300 English nouns referring to abstract and concrete concepts, which account for (machine readable) visual features. In a formal content analysis with multiple annotators we then classified the cognitive strategies used by the participants to represent conceptual content through emoji.

The main results of our analyses show that abstract (vs. concrete) concepts are characterized by representations that: 1. consist of a larger number of emoji; 2. include more face emoji (expressing emotions); 3. are less stable and less shared among users; 4. use representation strategies based on figurative operations (e.g., metaphors) and strategies that exploit linguistic information (e.g. rebus); 5. correlate less well with the semantic representations emerging from classic features listed through verbal strings.

Keywords: semantic representation, feature norms, abstract concepts, concrete concepts, emoji, visual features

1. Introduction

Featural views of cognition rely on the idea that some brain regions are specialized for the processing of information encoded in different modalities (Barsalou et al. 2003; Barsalou 2008; Lynott and Connell, 2013). According to this paradigm, when we process a concept like CAR, we think about how cars look like as well as how they sound, move, and smell; when we process a concept like CAKE we think about how cakes look like, how they taste and so on. These different types of information are *distributed* across brain areas. This view of cognition is typically referred to as *multimodal* view of cognition, and it is often operationalized by means of semantic feature norms (Wu and Barsalou 2009). Semantic feature norms are therefore used as a proxy for conceptual representations, under the assumption that a concept can be represented by a set of different types of features that express information about its perceptual, functional, associative, encyclopaedic and other type of content. Such a way to operationalize conceptual content is supported by several studies showing that feature-based representations predict various psychological behaviors, well-known in cognitive psychology (see McRae and Jones 2013 and section below for a brief literature review).

Semantic feature norms are typically elicited in experimental setting or in online crowdsourcing tasks by asking participants to list the most salient descriptors that come to their mind when they think of given concepts. The raw data are then normalized by the analysts and various measures are collected, together with the features' production frequency. For example, given the concept designated by the word *car*, typically participants mention <has four wheels>, <it's a transportation> and so on.

Although the semantic feature norms paradigm is extensively used in cognitive science and cognitive psychology, two caveats at least might be identified.

First, this type of research is typically performed on concrete concepts, for which the majority of semantic feature tasks are conducted (McRae et al. 2005; Vinson and Vigliocco 2008). However, as Recchia and Jones (2012) point out, words denoting abstract concepts are more frequent than words denoting concrete concepts. Moreover, a large body of scientific literature suggests that abstract and concrete concepts differ in the information that they may contain (with abstract concepts being richer in linguistic information and concrete concepts richer in perceptual information, e.g., (Recchia and Jones 2012; Lenci 2018). This difference is supported by empirical studies showing that abstract and concrete concepts differ in the way they are learned, remembered, and processed (Binder et al. 2005; Bolognesi and Steen 2018). It therefore remains controversial to claim that the semantic feature norms paradigm can be reliably used to operationalize conceptual content, given the fact that empirical evidence is for the majority collected on concrete concepts only.

Secondly, and most importantly, the semantic feature paradigm is grounded in the idea that cognition is not only *distributed* (across features), but also *multimodal* (across modalities). Several scholars agree that knowledge in the semantic memory is subdivided by modality into, for example, visual components in one cortical area and functional components in another (Farah and McClelland 1991). Switching from one modality-specific representation to another involves important cognitive costs (Pecher et al. 2004). Within the semantic feature norms paradigm, however, semantic features in property generation tasks are always elicited as lists of *linguistic* strings. In other words, semantic features expressing, for instance, perceptual properties or taxonomic properties, are both elicited from participants by means of words. Arguably, not all conceptual content is equally easy to verbalize into words: some types of features may be better expressed by means of visual clues, for instance, rather than by linguistic strings. For example, Bolognesi (2017) has shown that different types of features are exploited to construct and express metaphors in images vs. in language: while metaphors in images typically exploit perceptual similarities based on shapes and colors between two entities, which are easily conveyed within the pictorial mode of communication, perceptual similarity does not usually play a role in metaphors used in language. The latter types of metaphors, instead, tend to be constructed on shared taxonomic features associated to the compared entities. Taxonomic features are easy to express within the verbal system by using words denoting the two different levels of the taxonomy, e.g., *monkey* and *mammal*, but less so within the pictorial one: how can one discriminate the pictorial representations of a monkey from that of a mammal?.

On a different note, the Dual Coding Theory (Paivio 1971; 1986; 2010), suggests that there are two functionally independent but interconnected systems in memory, one specialized for the representation of nonverbal information and the other specialized for the representation of verbal information. At least two types of representations, therefore, may co-exist in the semantic memory, one based on visual clues (called *imagens*) and one based on linguistic clues (called *logogens*). A number of functional imaging studies have tackled this distinction, mainly by analyzing the processing of abstract vs concrete concepts (Binder et al. 2005), supporting the claim that there is a distinction between perceptually encoded knowledge and verbally encoded knowledge.

The combination of these two points (neglecting abstract concepts, and neglecting other modes in which features can be expressed) in combination with theoretical insights such as those proposed by the Dual Coding Theory suggests that the reliability of the semantic feature norms paradigm may be restricted to some types of concepts only.

Following these observations, we hereby propose, describe, and analyze a dataset of semantic representations of concrete and abstract concepts, which have been represented by means of *visual* components of meaning. Such visual representations are operationalized through pictograms that are commonly used in instant messaging and online apps: emoji. We claim that our visual set of representations can complement the classic verbally-conveyed features, to give a fuller and richer view of conceptual representations, within a truly *multimodal* view of cognition.

The specific research questions that lead the analyses can be summarised as follows:

RQ1: what are the strategies used to represent the content of concrete vs abstract concepts through emoji?

RQ2: are abstract concepts richer in emotional content (face emoji) than concrete concepts?

RQ3: do abstract concepts need on average more emoji than concrete concepts to be expressed?

RQ4: to what extent do the emoji-based semantic representations correlate with the linguistic features semantic representations?

Based on these questions, the paper is structured as follows: in Section 2 we provide a brief literature review on the semantic feature norm paradigm, we explain how this approach falls short in accounting for the multimodal nature of conceptual representations and we introduce our dataset of visually expressed features, followed by a short introduction on emoji. Section 3 explains the methods and experimental setup of the crowdsourcing task, used for implementing the algorithm that automatically generated averaged representations of each concept, and for the content analysis through which we annotated the crowdsourced visual features. In Section 4, we provide the data analysis, then discussed in Section 5. Finally, Section 6 provides a conclusion that links our findings to the general aims of the special issue in which the current study is embedded: “Eliciting Semantic Properties: Methods and Applications”.

2. Theoretical Background

Several models of conceptual representation based on different theoretical backgrounds make use of semantic features, thus assuming that semantic representations can be formalized as sets of distributed features (Baroni et al. 2010; Bolognesi 2017; Cree et al. 2006; Sartori et al. 2005; Taylor et al, 2007; Vigliocco et al. 2009; Vigliocco et al. 2004). These models include classical and prototype-based models of semantic memory (Medin and Schaffer 1978; Minda and Smith 2002; Smith and Medin 1981) as well as network-based models (Collins and Loftus 1975).

Semantic features encode different types of information, which is often manually classified on the basis of different types of taxonomies. For example, Cree and McRae (2003) proposed a brain region taxonomy according to which different types of features are processed in different brain areas. Given for example the concept KNIFE, <has a handle> is a ‘visual–form and surface’ type of feature, while <is sharp> is a ‘tactile’ type of feature, and these two features are processed in distinct neural circuits. Wu and Barsalou (reported in McRae et al. 2005) proposed a knowledge-based taxonomy organized into 4 classes of features (Entity-related, Situation-related, Taxonomic and Introspective), each divided into various sub-types. This taxonomy was then refined in order to account for abstract concepts (Barsalou and Wiemer-Hastings 2005). Vinson and Vigliocco (2008) proposed a 5-categories taxonomy that accounts for feature types produced for nouns and verbs. Finally, Recchia and Jones (2012) proposed a 19-categories taxonomy that constitutes an adapted version of existing coding schemes, applied to semantic features of both concrete and abstract concepts.

Based on the manual annotation of feature types, observations have been made on the nature of the conceptual representations that characterize, for example, concrete and abstract concepts. For example, Barsalou and Wiemer Hastings (2005) showed that abstract concepts are characterized by more situation-related features than concrete concepts which, in turn, are characterized by more entity-related features. Moreover, it has been shown that abstract concepts involve subjective experiences and emotions more than concrete concepts (Vigliocco 2009).

The relevance of a semantic feature approach to conceptual representation is empirically supported by a large body of literature. In particular, it has been shown that semantic features can explain the semantic priming effect (concepts that share semantic features prime one another, while concepts that share loose word associations do not (McRae and Boisvert 1998; Cree et al. 1999). McRae, Cree, Westmacott and de Sa (1999) showed that participants are faster to verify that a feature is part of a concept if it is strongly rather than weakly correlated with other features of that concept. For example, they are faster to verify that <has a long tail> is a feature of RAT than of PIG because such feature is highly correlated with other features of RAT, but not of PIG. Moreover, the number of features associated with given concepts correlates with their processing time (decision times and errors in lexical decision tasks are lower for concepts with many features (Pexman et al. 2002; Pexman et al. 2003). This suggests that concepts with many features, which tend to be recognized more quickly than concepts with a few features, have richer representations, and this peculiarity facilitates their access and retrieval. Finally, pairs of concepts sharing distinctive features are typically judged to be more similar than concepts sharing an equal number of relatively frequent features (Mirman and Magnuson 2005). In other words, features shared by only a few concepts (i.e., distinctive features) hold a special status, and therefore pairs of concepts sharing distinctive features are judged to be more similar to one another than concepts sharing an equal number of frequent features.

In support of the distributed and multimodal nature of cognition, empirical evidence collected from patients with category-specific semantic deficits supports the multimodal nature of semantic representations in memory. For example, in a pioneering study Warrington and McCarthy (1987) reported category-specific impairments observed in brain-damaged patients, showing that the category of living things and that of artifacts depend on, respectively, visual and functional information. This has been supported and extended by empirical analyses conducted within the semantic feature norms paradigm, in which it has been shown that these two categories there are characterized by different types of feature configurations (Garrard et al. 2001; Cree and McRae 2003). Finally, a recent feature-based neurocomputational model of semantic memory showed that the role of different kinds of features in the representation of concepts, both in normal and neurodegenerative conditions, has important implications for the study of normal vs. impaired cognition: the network model proposed by the authors solved naming tasks and word recognition tasks very well, exploiting the different role of salient versus marginal features in concept identification, and showed that in the case of damage, superordinate concepts were preserved better than the subordinate ones. Interestingly, the degradation of salient features, but not of marginal ones, prevented object identification. (Ursino et al. 2018).

Overall, the distributed and multimodal nature of conceptual representations is well documented in cognitive science and studies based on semantic features have repeatedly and consistently demonstrated that the semantic feature norms paradigm is a reliable way to operationalize the content of conceptual representations. Nonetheless, semantic features are typically collected as verbal strings elicited from participants. This could potentially limit the type of information that they can capture: arguably, the information that participants can easily verbalize when they perform the task is easily captured, while the information that cannot be easily verbalized is not. Conceptual content that cannot be easily verbalized may remain unexpressed and thus undocumented within the semantic feature norms paradigm. It follows that the multimodal nature of conceptual representations may actually be biased toward the expression of those features that can easily be put into words, thus leaving aside modality-specific information that cannot be easily put into words.

One of the main strategies in which humans express themselves, besides words, is through images. Research shows that not all the information encoded in visual representations of concepts can be encoded in language. For example, functional neuroimaging studies show different patterns of activation during matched word and picture recognition tasks (Gorno-Tempini et al. 1998; Moore and Price 1999; Chee et al. 2000; Hasson et al. 2002; Bright et al. 2004; Gates and Yoon 2005; Reinholz and Pollmann 2005). As indicated in Binder et al.'s (2009) meta-analysis, different studies argue against a complete overlap between the knowledge systems underlying word and object recognition, based on the existence of patients with profound visual object recognition disorders but relatively intact word comprehension (Warrington 1985; Farah 1990; Davidoff and De Bleser 1994). Therefore, a richer and more exhaustive set of semantic features that aims at providing a reliable proxy for conceptual content should in principle include both verbally-expressed and visually-expressed features.

2.1. Emoji: a pictorial mode of communication

The rise of social media and virtual communication has recently enabled new ways to encode meaning: The Unicode Standard emoji code, a set of now more than 3000 pictograms, which became available in most of our online messaging services. In their emoji Report 2015[1] the company Emogi stated that about 92% of internet and instant messaging users utilizes emoji. The word *emoji* is derived from the Japanese *e* for "picture" and *moji* for "word" and the resemblance to the English word "emotional" is coincidental (Taggart 2015). Nonetheless, many emoji are being used to express emotions and therefore all major social media platforms use sentiment analysis to understand the content of the users. This creates a great interest in research on semantic evaluation and analysis of emoji (Barbieri et al. 2018; Barbieri et al. 2016; Barbieri et al. 2016).

Compared to words, emoji are characterized by a stronger iconic relationship with the referents that they designate. While the relationship between a word and its referent is mostly arbitrary, symbolic, and based on social conventions, the relationship between an emoji and the designated referent is less arbitrary, because it is based on iconicity. For example, the iconic nature of the relationship between an actual dog and □ is not preserved in the sign *dog*.

In this regard, the reader may think that emoji are comparable to Egyptian hieroglyphs. However, the pictograms that constitute the Egyptian ancient language, unlike emoji, are a sound-based phonetic system (Jespersen and Reintges 2008). For example, a sequence of hieroglyphs can be composed of the pictorams that provide the phonetics of the word BELIEF. The first hieroglyph would be a foot, the Egyptian pronunciation of the word foot would sound like "B", the next hieroglyph would depict a reed, which would be pronounced with an "E". This is followed by the lion hieroglyph, which in Egyptian is *Leo* and provides the "L". This can be called a *rebus principle*, which in principle can also be applied to emoji use, but we believe (and test in Section 4.4) that this might be a quite uncommon communicative strategy among emoji users. See Figure 1 for an example of the rebus principle in both hieroglyphs and emoji.

[1] Emogi.com's Emoji Report 2015, retrieved 2018-11-02 from www.emogi.com

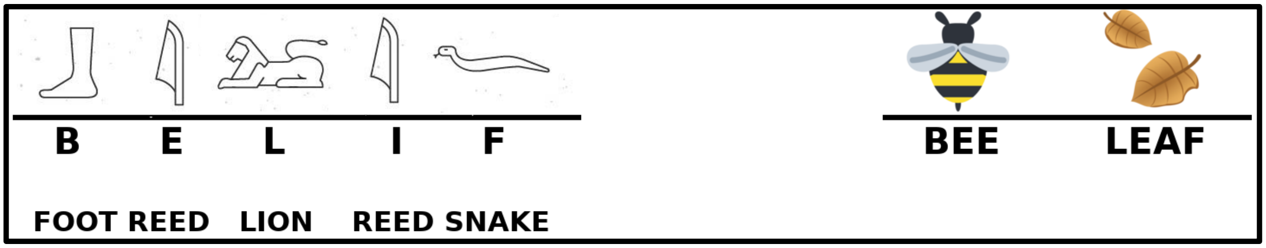

Figure 1: Rebus translation of the word *BELIEF* in both Egyptian hieroglyph and emoji. Left: Sequence of hieroglyphs providing the phonetics. The first hieroglyph is a foot, the Egyptian pronunciation of the word foot would sound like "B", the next hieroglyph would depict a reed, which would be pronounced with an "E". This is followed by the lion hieroglyph, which in Egyptian is *Leo* and provides the "L". Right: Two emoji that can be read as "Bee" and "Leaf" and compose the word "Belief" if the rebus principle is applied.

Another interesting difference between the verbal system and the emoji system is that while the first developed, changed, and adapted through the centuries, the latter evolves in a radically different and significantly faster way. The drafts of emoji sets are based on proposals received by the Unicode Consortium, reviewed by the Unicode Emoji Subcommittee, and selected on the basis of the Emoji Selection Factors[2]. The selection factors for inclusion of a new emoji are: Compatibility, expected usage level, image distinctiveness, completeness (does the emoji fill a gap in the set of existing emoji) and frequently requested emoji. Thanks to these adaptations, the original set of 176 12-by-12-pixel characters created by Shigetaka Kurita for the telecommunication company DOCOMO in 1998/99, has now extended to a set of more than 2800 pictograms. The latest Unicode draft proposal[3], for example, includes wheelchairs (manual, motorized, with a woman or a man), guide dogs, prosthetics but also rather arbitrarily specialized new emoji such as a falafel and a skunk.

The rising demand in emoji can therefore lead to the addition of more and more specific emoji, to foster inclusion and political correctness. An example is the collaboration between Apple and the Unicode Consortium to implement a new technical solution that would include a skin tone modifier, which has been introduced to Unicode 8.0 (Stark and Crawford 2015). The political and societal principles that drive the inclusion of new emoji, lead to a peculiarity of the emoji vocabulary: some domains display a wide variety of options to differentiate between concepts at a very specific level, such as the difference between a red-hair and a blond-hair woman, while in other domains the availability of emoji is quite limited, for example there is a very limited variety of emoji describing clothing, and virtually no emoji representing the variety of musical instruments, or furniture. Because these domains, even though populated by concrete instances, do not have a variety of emoji from which users can draw to express meaning, are also subject to creative uses of existing emoji.

Another peculiar example is the lack of graphic and sexual content in the set of available emoji. This gap gave rise to a series of creative uses of existing emoji, to express sexual content. For example, characteristic use of the □ emoji. The □ is widely known to be understood as a metaphor for male genitalia (Highfield 2018). This connotation was so strong that Instagram temporarily banned its hashtag use #□.

The variety of methods that users can resort to, in order to express a concept through an emoji, vary, and arguably not all of these methods hold the same status, are used with the same frequency, and are equally appreciated by emoji users. With our research questions we aim to explore the degree by which various communicative strategies (e.g., based on metaphors, based on rebus, etc.) are used by emoji users to represent abstract and concrete concepts.

---

[2] Unicode-Consortium. Draft Emoji Candidates: http://www.unicode.org/emoji/future/emoji-candidates.html

[3] Unicode-Consortium. Submitting Emoji Proposals: https://www.unicode.org/emoji/proposals.html

We hypothesize that mechanisms based on figurative operations known in the scientific literature as metaphors, idioms, and metonymies, are preferred, compared to artificially created methods that use lots of emojis, portray pictures or short stories. Our hypothesis builds upon a preliminary study conducted by (Wicke 2017). Wicke proposes to investigate emoji as semantic primitive building blocks (Wierzbicka 1996) and analyses methods preferred by users to translate action verbs into emoji sequences. The results show that the most effective methods are the *Rebus*, *Literal* and Metaphor methods. Moreover, the results suggest that the shorter the sequence, the better is the interpretability. The rendering of those actions found its application in a computational storytelling system (Veale 2017).

3. Methods

3.1. Stimuli and Materials

A sample of 300 nouns was extracted from the tails of the distribution of concreteness ratings reported by Brysbaert et al. (2014). The entire Brysbaert dataset is composed of 40,000 generally known English word lemmas with concreteness ratings obtained from 4,000 participants using crowdsourcing for data collection. In this large dataset the words were presented in isolation and participants were asked to rate the degree of concreteness of the related concept, on a scale from 1 to 5. The individual ratings were then averaged. For instance, a word like *dandelion* is associated to an average score of 5 (extremely concrete), while a word like *belief* is associated with an average score of 1.19 (extremely abstract). In order to select the 300 words we took into account a few factors. First of all, we considered only nouns, and therefore dropped other words that were present in Brysbaert's et al. database (e.g., *eh*). Secondly, we took into account the word frequency, and dropped those words that displayed very low frequency values according to the "Word Frequencies in Written and Spoken English: based on the British National Corpus" (WFWSE) (Leech 2014). We used nouns with a frequency of at least 10 in a million of words. Thirdly, we took into account the fact that the distribution of concreteness ratings in Brysbaert's database is normal but slightly skewed toward the extremely concrete pole: the average concreteness of the 14,593 nouns reported in the database is 3.53 on a 5-point scale. We therefore selected concrete nouns with concreteness score above 4.0 and abstract nouns with concreteness score below 2.0. The final factor that we took into account is the fact that many words in Brysbaert's database are associated with average concreteness measures that involve high standard deviation scores. This phenomenon may be problematic because it suggests that given the same word, some participants rated the concept as very abstract, some rated it as very concrete, because they may have had different senses in mind, associated to the same polysemous word. In fact, as Reijnierse, Burgers, Bolognesi and Krennmayr (2019) suggest, when words that have a literal a metaphorical meaning are disambiguated by the context, the two senses (e.g., *support*: 1. a board or scaffolding that holds something up; 2. psychological help) tend to be rated respectively as very concrete and very abstract.

To summarize, the 300 concepts in our dataset have the following properties:

- They were extracted from the Brysbaert dataset
- They had a frequency of at least 10 times per million words, according to the "WFWSE"
- They were either very concrete with a concreteness score above 4.0 or very abstract with a concreteness score below 2.0.
- They had an average standard deviation of >0.6, whereas the average standard deviation of all nouns in Brysbaert's database is >1.

The set of emoji used for this task has been taken from the official Unicode version 11 (for an overview, see https://unicode.org/emoji/charts/emoji-versions.html). This version is the latest version published in 2018, when the data collection was conducted. The Unicode Consortium updates the set of emoji each year with several version releases. However, we believe that the slight changes introduced in the new releases do not affect the overall findings of our study, for various reasons. First, newly proposed emoji take several months (and sometimes years) before they are included in a new version of the dataset. Second, whenever new emoji enter the Unicode it is the responsibility of each platform provider to implement a specific rendering and therefore the pace with which the new emoji become available on various platforms is slow and irregular. Users sometimes are not even aware of about the latest emoji accepted by Unicode, because their device does not support them yet.

The challenge of an ever-increasing set of emoji poses questions for a scientific study: How can we offer all emoji to users without overwhelming them with the more than 2,800 emoji to choose from? How do we account for different renderings of the same emoji, on different devices (Android, Windows, etc)? And finally, how can we present the emoji set in a clear and easily browsable way? To answer these questions, we decided to use www.emojicopy.com (see Figure 2). This web application provides a very simple copy and paste interface, which clusters the emoji according to categories common on most service providers, hence answering all of our demands. A possible drawback of this solution is that the clustering of emoji imposes a certain order that may affect the choice and hence our results. Nonetheless, as mentioned above, the clustering is the one common to most service providers, whilst not depicting emoji rendered for a specific provider.

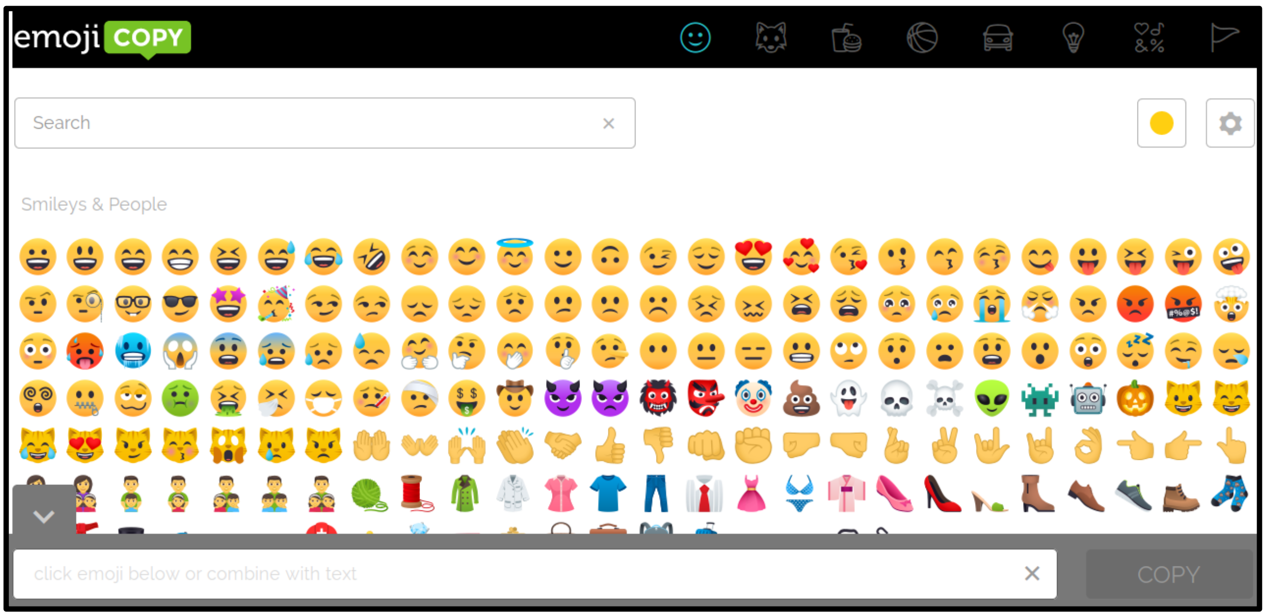

Figure 2: Edited screenshot of the online platform emojicopy.com, providing the latest set of emoji for an easy copy-paste access system.

3.2. Data collection: setting up the crowdsourcing task

The data collection has been designed as a crowdsourcing task, conducted on the platform *Figure Eight* (www.figure-eight.com, formerly *Crowdflower*), which provides a reliable interface for human micro-tasking, used by a variety of companies and researchers. Workers were asked to use a computer rather than their phone, and were asked to declare that they were older than 18 years old and English native speakers.

Crowdsourcing is a useful and efficient tool to obtain large amounts of human annotated data with little effort and has shown to be comparable to university participant samples (Behrend et al. 2011). However, the monetary compensation is often the only motivation for participating in the task (Brabham 2010). Therefore we must take into account that participants will be prone to abuse the setup to maximize their rewards whilst minimizing their efforts. Fortunately, *Figure Eight* provides methods to detect cheaters and irregular behaviors. On submission of the task, the analyst can provide a gold standard of questions with a given answer, and crowdworkers who fail too many of the gold standard questions will be marked as cheaters. Besides these potential drawbacks involved in using a crowdsourcing task to collect our data, we strongly believe that presenting the emoji task on the web is a more natural way to investigate the use of emoji than in an offline experimental design, because emoji are used solely online.

Before we conducted the actual study, we ran a pilot study with 20 words. Each word was represented by emoji by nine participants. At this stage we asked participants for feedback about the experimental design and the use of the online keyboard to select the emoji. The data collected in this pilot study was not included in the final dataset of experimental data.

The experimental study featured a list of 300 words (150 concrete, 150 abstract). Each word was represented by emoji by 11 participants for a total of 3300 unique responses, which were then manually annotated by two independent annotators in order to investigate the different methods that participants used to perform the task (see Section 3.3).

In the instructions we explained the nature of the task and provided examples of possible representations, indicating to the participants that the optimal sequence would consist of a number between 1 and 4 emoji, that would describe the concept in the most precise, yet non-redundant, way. Additionally, we provided a step-by-step guide with screenshots on how the participants are supposed to use the *emojicopy* website to create their responses.

The full instructions are reported in the appendix to this paper.

Crowdsourcing settings were configured so that only the online workers with highest reliability history could take part in our data collection. In *Figure Eight* these are participants who received the best feedback for accuracy in performing previous tasks. Moreover, participants were selected from 6 English speaking countries (USA, New Zealand, Ireland, UK, Canada and Australia). Furthermore, we tracked the subject ID in order to evaluate cheating behavior. For each item participants were rewarded with 4 cents, which adds up to 4 euros to complete a list of 100 items.

Participants were then presented with the first word, and asked to follow the link to *emojicopy.com* in order to create a sequence of emoji that best describes the concept. The chosen sequence had then to be pasted into the task form. Finally, participants were asked to provide optional written feedback for each word, as well as for the task in general. This optional written feedback was useful to give the participants the chance to explain any irregularities, problems or difficulties with the task or word. Figure 3 displays a screenshot of the task, for the concept TIME BOMB.

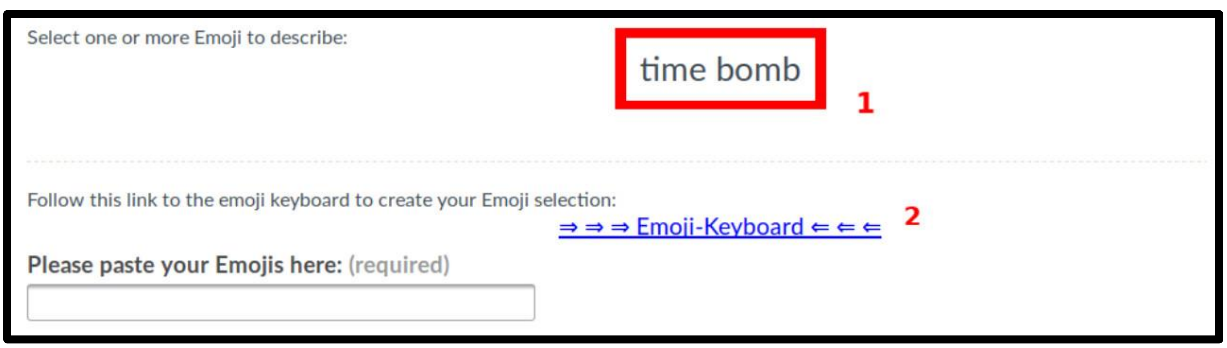

Figure 3: This image shows one of the 300 tasks that have been posted to the online workers. In this example, the worker needs to describe the concept TIME BOMB (red box, indicated by 1), using emoji. The worker can do so by following the provided link (indicated by 2) to the emoji keyboard.

3.3. Content analysis: Coding scheme for the manual annotation

The investigation of representational strategies is the main tool that can leverage insight into the question of how abstract versus concrete concepts are represented by means of visual features We therefore focused on an accurate, manual classification of the representational strategies used by online workers to describe the stimuli. The classification system was initially proposed in Wicke (2017), and later refined and further described by the authors of the present paper.

Table 1 reports the final version of the coding scheme used to annotate the representational strategy types.

Strategy type	
Literal	Literal representation of the referent as a strategy that is based on using the emoji which represents the actual referent. An example for a concrete concept representation would be: BIKE = □ and an example for an abstract concept representation would be: GENERATION = □□□□□□. Notably, the former usually is depicted with one emoji alone, whereas the example of an abstract concept representation is depicted by the sequence of emoji: from baby to grandfather is one generation.
Rebus	Rebus is a linguistic strategy based on word form: the form of the word (typically its phonetic pronunciation) is constructed via emojis denoting words that read aloud compose the sound of the target word. An example for a concrete concept representation would be: CARPET = □□ (read: car + pet) and an example for an abstract concept representation would be: CAREER =□□ (read: car + ear).
Phonetic similarity	Phonetic similarity is a linguistic strategy based on which they use an emoji that displays an emoji with a phonetically similar name. SHELF □ (uses 'shell' which is phonetically similar), or CHEST ♟ (chess) and an example for an abstract concept representation would be: PATIENCE □ (shows a patient).
Figurative Construction	Figurative Constructions, including symbols, idioms, metaphors and metonymies. An example for a concrete concept representation would be: CAGE □ (cages usually have lockers, it is a component) or FLOUR □ (cakes contain flour) STAIRS □ (a graph with bars that go up, like stairs do) and an example for an abstract concept representation would be: LUCK □ (there is no contiguity, nor similarity relation between these two, it is conventionalized and arbitrary that luck is this plant); COMPARISON □□ (based on the idiom comparing apples to oranges); VALUE □ (dollar bill is a type of value, which is monetary value); COURAGE□ (psychological courage is physical strength) or OBJECTIVE □ (the abstract objective becomes the concrete target).

Table 1: Coding scheme of strategy types.

On a second level of analysis, we investigated the structure and relations between individual emojis used in sequences of more than one emoji. On a preliminary inspection of the data, the two analysts agreed on the data-driven coding scheme of sequence types reported in Table 2:

Sequence type	
OP (OR and POLYSEMY)	The sequence consists of emoji in which each emoji contributes a separate representation and therefore the emoji can be read as connected through an OR relation (e.g. this or this or this). If those representations also denote different meanings, this signals polysemy and therefore this label is denoted OP. An example would be TOAST = □□, where each emoji represents one of the

	meanings of toast (bread sandwich OR cheers). Contrary, TOAST = □□ would only represent one meaning with (bread AND sandwich).
OS (OR and SIMILAR)	The sequence consists of emoji in which each emoji contributes a separate representation with a similar meaning. An example would be FLOWER = □□□, where we see sunflower OR rose, OR tulip etc.
AT (AND and TEMPORAL)	The sequence shows a sequence of emojis that has to be taken all together to construct the meaning of the concept. You need the first AND the second AND the third emoji to construct the meaning of the concept. If this relationship relies on a time-relatedness or cause-effect relation, this will be annotated with AT. An example would be: □□✈PASSPORT (little book AND you first show to the border police AND then you can fly). Here, neither of the emoji works by itself: □ book; □ control; ✈plane; the cause of the first emoji is explained by the order of the following emoji.
AS (AND and SPATIAL)	Similar to the AT method, but here it denotes a spatial-relation of the emoji in the sequence. An example would be: □□ GARAGE, because it requires a car AND a house, none of these two symbols alone is sufficient to represent garage, but the combination of them.

Table 2: Coding scheme for annotating the emoji sequence types.

3.4. Procedure for the manual annotation of the raw responses

The 3300 responses collected with the crowdsourcing task were manually annotated by two independent raters in six incremental batches of data. After each round of annotations the two analysts met and discussed the agreements and disagreements. Inter-annotator agreement scores were calculated on the independent analyses, but the discussions that followed each round of analysis helped improve the annotation scheme and also the interrater reliability scores (as reported in Section 4). The first batch included 55 items, which were annotated according to the strategy (Literal, Rebus, Phonetic Similarity or Figurative Construction) and the sequence type (OP, OS, AT and AS). The six batch sizes included an increasing number of concepts, resulting in batches of, respectively, 55, 253, 363, 825, 913 and 891 individual responses. Interrater reliability was calculated according to standard procedures used in formal content analyses (Bolognesi et al. 2017), using the Krippendorff α (1970; 2004) metrics, which corrects the raw percentage of agreement by the probability of chance agreement, and takes into account potential missing values. In particular, after each annotating session (performed by the two independent coders on a batch of data), we calculated the interrater reliability scores. Then, we discussed the disagreements and solved them in a discussion. Simultaneously we refined the coding scheme, adding examples and clarifying the description of the categories. At the end of the annotation process, when all the batches of data were annotated and the reliability was calculated on each batch, one of the annotators revised the annotations of the whole dataset, in accordance with the latest version of the coding scheme. Therefore, the very final annotations of all the 3300 items result from the final version of the coding scheme and the final and agreed revision of the individual annotations.

3.5. Measures to evaluate homogeneity among participants' responses

Evaluating the homogeneity of participants' responses toward the same stimuli presented various technical difficulties. First, different responses displayed different sequence lengths (different numbers of emoji). Second, on a semantic level we have to take into account that □ is similar to □, even though these are two different emoji. Whereas, □ is somewhat similar to □, but not as much similar as the two apples variants. The two standard distance measures we have preliminarily explored to operationalize the similarity between participants' responses are the Dice coefficient and the Jaccard Coefficient. However, both measures neglect the internal semantic similarity, e.g. □ being similar to □. We therefore opted for cosine similarity, and used the word embedding *emoji2vec* (Eisner et al. 2016[4]). A similar discussion of emoji similarity measures can be found in (Pohl et al. 2017). The word embedding is able to operate with the internal semantic similarity. For example, *emoji2vec* is capable of producing the same simple algebraic operations on word vectors as demonstrated in *word2vec* (Mikolov et al. 2013): vector("King") - vector("Man") + vector("Woman") = vector("Queen") in *emoji2vec:* □ - □ + □ = □.

There are a few issues with this embedding. First, not all current emoji have been used to train this model. Nonetheless, only a few emoji are not included and this should therefore not strongly impact the similarity measure. Second, the embedding might not capture fine-grained semantic distinctions, such as the fact that □ is similar to □, but not completely. The word embedding recognizes that these are similar as "faces", but does not distinguish the similarity in the shape of the eyes. Furthermore, we can only evaluate the similarity between two emoji at the time. Therefore, for comparing sequences we need to compare all combinations of individual emoji and then calculate an average similarity.

For example, in a pilot study we observed that the nine individual responses provided for the concept WATER (□□□,□□□,□□☔,□,□,□□,□,□□,□□) were more homogeneous (i.e., more similar to one another) than the nine responses provided for the concept STORY (□□□□□, □□, □, □□□□, □□□□□□□, □□□□□□□, □□□, □□□, □□♀□□♀).

Based on the participants' responses, we constructed a homogeneity matrix for each concept, as those displayed in Figure 4, which show all cosines similarities between each pair of participants' responses collected in the pilot study, for the concepts STORY and WATER. The average cosine of these values, calculated on the whole concept matrix, gives us the overall homogeneity among participants' responses, for that given concept. As this figure shows, the shades in the matrix of the concept WATER are brighter than those in me matrix of the concept STORY, suggesting that the emoji-based representations provided by participants are more similar to one another in the case of WATER, than in the case of STORY.

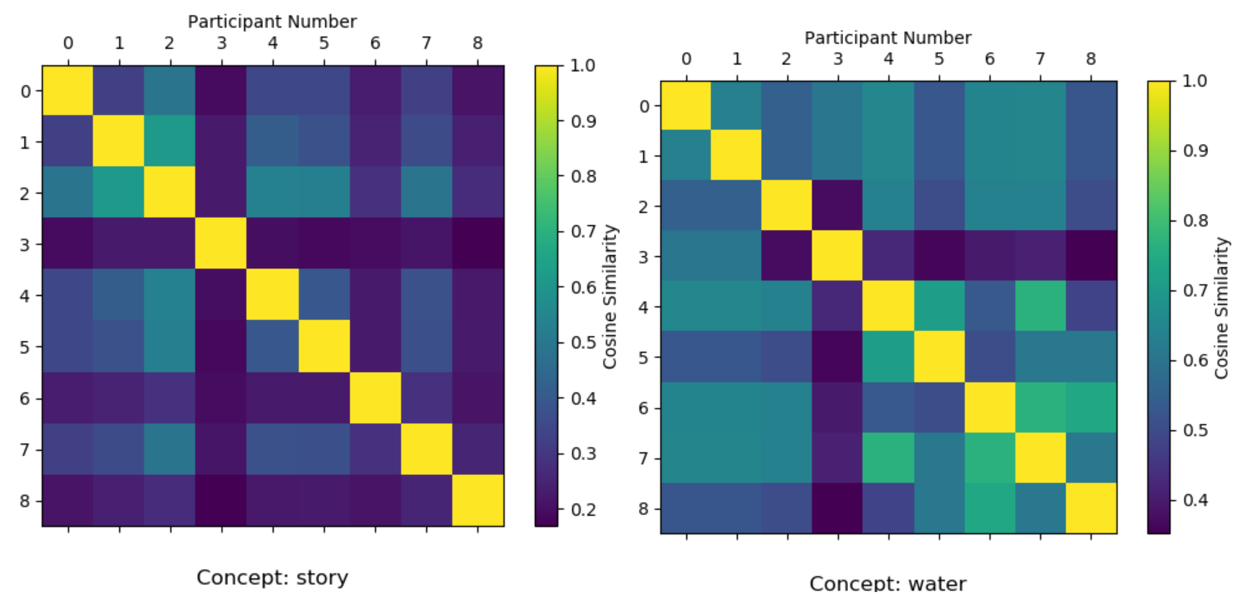

Figure 4: Cosine matrices showing the similarities between the representation provided by each pair of participants (numbered with a number from 0 to 8) The brighter (yellow) the color, the higher the averaged

[4] The pre-trained emoji-embedding is provided at: github.com/uclmr/emoji2vec

cosine similarity between the representations provided by two participants, for the same concept. As each participant has 100% similarity to his/her own response, the diagonal is filled with the highest similarity value, which is 1. If the sequences are completely different and the respective emojis are not similar with respect to the cosine similarity, the value approaches 0 (dark blue).

As a matter of fact, the resulting average homogeneity obtained for WATER is 0.570 while the resulting average homogeneity obtained for STORY is 0.307.

3.6. Automatic generation of average representations

We hereby explain how we pursued one possible way of automatic generation for an average, standardized emoji-based representation for all the 300 concepts in our sample. Our approach creates high quality data that can be used for related and further studies. The provided dataset of standardized visual representations can be used to support machine translation systems (mostly using word embeddings to translate words into emoji) with complex, sequential representations of both concrete and abstract concepts. It should be noted that the presented approach is by far not the only solution to generate an average representation, e.g. one could combine the vector representations to obtain complex average representations. For example, one could experiment with vector addition or multiplication, which are common strategies in these vector spaces (Lenci 2018). Our dataset encourages us or other researchers to investigate different strategies for such automatic generation.

We first explain the algorithm, then provide pseudo-code, which is followed by an explanation of an example. The algorithm iterates through all concepts and their replies by the contributors. Each reply is a sequence of at least one emoji and is hereafter referred to as *sequence*. The algorithm orders the sequences for each concept by length, i.e. the number of emoji from each individual contributor. From the list of ordered sequence lengths, the median (*Md*) is determined. We decided to use the median as opposed to the mean, because it is less skewed by small proportions of extreme values, which we could observe in our data. From the most common emoji, the *Md* number of emoji are chosen as average representation. A few exceptions have been implemented to account for edge-cases: (I) if the selected emoji occurs only once, the algorithm does not select this emoji. (II) if the selected emoji only occurs within context of other emoji, which do not occur more often, the algorithm does not select this emoji. (III) if multiple emoji occur the same amount of times, the algorithm chooses those which rely on the annotated technique, which has been used most often for representing the concept. This can exceed the median length. The exceptions for (II) and (III) did occur less than 10 times and have been corrected manually. The pseudo-code of the algorithm is displayed below:

```
for each concept in all_concepts
      define all_response_lengths = 0
      define emoji_count = dict #dictionary key: emoji/value: occurrence

      for each response in participant_responses
            all_response_lengths add length(response)
            for each emoji in response
                  if emoji_count[emoji] > 0           (I)
                        emoji_count[emoji] add 1

      md = median(all_response_lengths)
      sort(emoji_count) by value: occurrence
      representation = emoji_count[0 to md] select emoji by key index
```

A practical example of the algorithm at work for the concept GLASS can be seen in Table 3.

Concept	Glass
Answers (Sequence length)	□□ (2), □ (1), □□ (2), □(1), □□ (2), □□ (2),□□(2), □□ (2), □□ (2), □□□(3), □□□□(4)
Sequence lengths (ordered)	1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 4
Median - *Md*	2
Most common emoji	□ (8x), □ (5x), □ (4x), □ (2x), □ (2x), □ (1x), □ (1x)
Average representation (#*Md* most common emoji)	□□

Table 3: Average representation obtained automatically for the concept GLASS.

In the example in Table 3, the algorithm is depicted for the concept GLASS. Here, the algorithm first lists the sequences (answers) with their lengths. That is, □□ was one of 11 replies and has the sequence length: 2. All sequence lengths are being ordered and the median of 2 is determined. After ordering the occurrences of emoji in all sequences, the top 2 (median) most common emoji are being chosen as average representation.

3.7. Comparing and correlating the vector representations of concrete and abstract concepts obtained with verbal and visual features.

In order to better understand to what extent the semantic representations based on visual features (emoji) compare to the classic semantic representations, based on linguistically expressed semantic features, we compared and contrasted the semantic representations emerging from the emoji dataset with those emerging from traditional datasets of semantic features, for both abstract and concrete concepts.

While most of the literature provides datasets of features elicited on lists of concrete concepts (e.g., McRae et al. 2005), Recchia and Jones (2012) collected semantic features for a sample of concrete and abstract concepts. After a quick comparison to measure the number of overlapping concepts between our dataset and existing datasets of linguistically encoded semantic features, we opted for comparing our data to the semantic features collected by McRae and colleagues (2005) (hereinafter MR-Set), often used as a benchmark to evaluate new datasets, as well as Recchia and Jones (2012) (hereinafter RJ-Set), because they include abstract concepts as well.

Table 4 provides an overview of the specifications of both datasets.

Dataset	**RJ Set**	**MR Set**
Type	abstract / concrete	concrete
Number of concepts	551	541
Total number of features	2098	2526
Overlapping concepts with our dataset	70	59
Number of features related to the overlapping concepts	528	573

Table 4: Specifications of Recchia and Jones (2012) dataset on the left and McRae and colleagues (2005) dataset on the right, in relation to the dataset of emoji.

In order to compare the featural representations obtained from the verbally expressed semantic features in the RJ and MR sets with the representations obtained from the emoji-based semantic features we proceeded as follows. For each comparison (i.e., RJ compared with the emoji set and MR compared with the emoji set) first we vectorized each overlapping concept using all the features as vector coordinates. This operation allowed us to compile two concept-by-feature contingency matrices that included all the overlapping concepts between RJ and our dataset and MR and our dataset, respectively. In the intersecting cells of each matrix we reported the production frequency for each concept/feature combination, based on the RJ and on the MR sets respectively. Consequently, the tables of cosine similarities between concept pairs were created. We therefore obtained a 70 by 70 table of cosine measures displaying the proximities between the 70 concepts shared by the RJ set and our emoji dataset, and a 59 by 59 table of cosine measures displaying the proximities between the 59 concepts shared by the MR set and our emoji dataset. The cosine similarities in these tables were calculated on the basis of feature production frequencies observed in the RJ and in the MR sets.

In the same way, we created the 70 by 70 and 59 by 59 tables of cosine similarities using our dataset of visual features for the creation of the contingency matrices. Finally, we correlated the 70 by 70 cosine table based on RJ with our emoji set, and the 59 by 59 cosine table based on MR with our emoji set.

4. Analysis

4.1. Cleaning the data

A total of 3300 responses were collected from 124 participants. During the data collection period, the crowdsourcing platform allows for a review of the responses. During this review we were able to monitor the incoming responses and reject hundreds of those that were not related to the task (scam responses) and replace them with new, legitimate responses. New participants were invited to replace the unreliable ones, until a total of 3300 responses was obtained. Nonetheless, some unrelated responses were not correctly detected until a manual classification of each of the 3300 responses was performed by two analysts (see Section 3.4). These were classified as null responses in the manual classification (6.58%) and they include different types of issues. For example, Example 1 shows an example of unrelated response that was eventually classified as null: this participant repeatedly responded with the same sequence of emoji for a variety of different concepts. Example 2 shows another type of issue classified as null: this participant has used an emoji to depict his/her own inability to find a satisfactory representation for the concept. This answer is irrelevant to the task, but provides an explanation for the failed response, i.e. the concept was too difficult to represent. In this regard, the written feedback that some participants provided was useful to help the analysts understand and discriminate between types of responses. Example 3 shows a response that at first sight might be classified as null (scam), while the written explanation helps the analysts to understand the participant's rationale for such choice.

Example 1:
Concept → doll
Reply →□♂□□□□□□●
Feedback →

Example 2:
Concept → liability

Reply → □♀
Feedback → Cannot show abstract concepts except for love I guess

Example 3:
Concept → expectation
Reply → □
Feedback → I think most people have expectations of winning money.

4.2 The dataset of visual features (emoji): a general overview.

The dataset of emoji collected through the crowdsourcing task for the 300 abstract and concrete concepts encompassed a total of 9089 instances of individual emoji including modifiers[5], and a total of 1012 unique emoji. Of these, a total of 5529 emoji (707 unique ones) were produced to represent abstract concepts, and a total of 3560 emoji (697 unique ones) were produced to represent concrete concepts.

The data is released online in a repository, available on the Open Science Framework (OSF) platform, at the following url: https://osf.io/qcn6h/

The raw data was then used to construct a contingency matrix in which we collected the production frequencies of each concept with each individual emoji. This matrix is also available in the online repository. Finally, based on the vectorized representations of each of the 300 concepts we constructed the square and symmetric table of cosine similarities. Looking at the number of face emoji used to represent respectively the concrete and the abstract concepts, we observed the following frequencies of use:

Number of faces used for concrete concepts: 153 / 3570 = 4.29 %

Number of faces used for abstract concepts: 1127 / 5571 = 20.23 %

There seems, therefore, to be a much higher use of face emoji to represent abstract concepts, compared to concrete ones.

4.3 Content Analysis: Manual annotations

As described in Section 3.4, the 3300 raw responses were divided into incremental batches and analyzed in a formal content analysis by two independent coders, based on the coding scheme described in Section 3.3. The inter-annotator agreement was calculated for each batch, before discussing the individual analyses, and measured by Krippendorff's alpha scores, reported in Table 5 and they show overall an increase in reliability for both, the annotation of strategy types and the annotation of sequence types.

Batch No.	1	2	3	4	5	6
# items	55	253	362	825	913	891

[5] Emoji are defined by the Unicode Standard using code-points to implement emoji as characters. There are certain modifiers which allow to transform the skin tone of emoji, gender or nationality of flag emoji. These modifiers are relevant semantic units, but cannot be displayed by themselves.

Strategy Type						
Alpha (Krippendorff)	**0.756**	**0.64**	**0.865**	**0.743**	**0.846**	**0.872**
Sequence Type						
Alpha (Krippendorff)	**0.535**	**0.346**	**0.619**	**0.476**	**0.671**	**0.865**

Table 5: Interrater reliability scores for the annotation of Strategy type and Sequence type across the 6 annotation sessions. The alpha score ranges from 0 (no agreement) to 1 (full agreement). The range between 0.5 and 0.7 is interpretable as medium agreement, whereas, above 0.7 is considered to be high reliability.

As described also in Bolognesi, Pilgram and van den Heerik (2017), content analyses organized with multiple annotating sessions have the advantage of keeping track of the improvement in reliability scores throughout the coding sessions, and enable the analysts to increasingly refine and clarify the coding scheme and test the improvements in the next annotating session.

Overall, the most frequent strategy used to represent a concept is the *Figurative Construction* with 1947 occurrences (59% of the 3300 responses). With 1119 occurrences, the *Literal* was the second-most used strategy (33.91%). The *Phonetic* and *Rebus* are the least-common strategy methods with 31 and 42 occurrences respectively. For the sequence types, the most frequent sequences are And-Spatial (AS) type (710 occurrences) and Or-Similar (OS) type (537) occurrences. The And-Temporal (AT) type occurred 249 times and the Or-Polysemy type occurred only 17 times.

Table 6 provides examples for each of the strategies in both categories, abstract and concrete, whereas Table 7 shows the occurrences of strategies adopted to represent, respectively, concrete and abstract concepts.

Examples	Literal strategy	Rebus strategy	Phonetic strategy	Figurative strategy
Abstract	□□□□ (GENERATION)	□□ (BELIEF)	□□□ (PATIENCE)	⛪□ (FAITH)
Concrete	□ (OWL)	□□□ (CARPET)	□ (SOAP)	□□□□ (PASSPORT)

Table 6: The rows define the type of concept and the columns the type of strategy. Therefore, each cell contains an example of the strategy applied to the type of concept.

The abstract concept GENERATION is being represented by the literal depiction of a generation of humans, i.e. baby to grandfather. Whereas the concrete concept OWL is being represented by a single emoji depicting an owl. In case of the rebus strategy, we can see the example of *bee* + *leaf* = BELIEF and the example of *car* + *pet* = CARPET. In the phonetic strategy example, we see that PATIENCE is depicted with *patients* and SOAP with *soup*. Interestingly, during the assessment of the experiment the soap emoji □[6] with Unicode U+1F9FC was already available but not used by one out of eleven participants. As addressed in Section 3.1, we can assume that novel emoji are too unfamiliar to be found as representatives. For the figurative strategy, the abstract example represents faith using the church and praying hands as metonymy. The concrete concept *passport* is being depicted using several metaphor and metonymy.

Observed production	Literal	Rebus	Phonetic	Figurative	Total

[6] As this emoji is a recent addition to the Unicode not all systems do support the rendering of this emoji. Different renderings can be found at: https://emojipedia.org/bar-of-soap/

frequencies	strategy	strategy	strategy	strategy	
# abstract concepts	23	21	16	**1429**	1489
# concrete concepts	**1096**	21	15	518	1650
Total	1119	42	31	1947	3139[7]

Table 7: Frequencies by which the different strategies were adopted to represent abstract and concrete concepts respectively.

A Pearson's Chi² test showed that there is a significant dependence between the two variables (type of strategy and concreteness): Chi-square = 1450.738, df = 3, Cramer's V = 0.680, p < 0.001.

Figure 5 displays a mosaic plot in which the residuals are visualized. As the plot shows, there is a strong relation between the category of concrete concepts and the literal strategy of representation, and between the category of abstract concepts and the figurative strategy of representation (residuals >4).

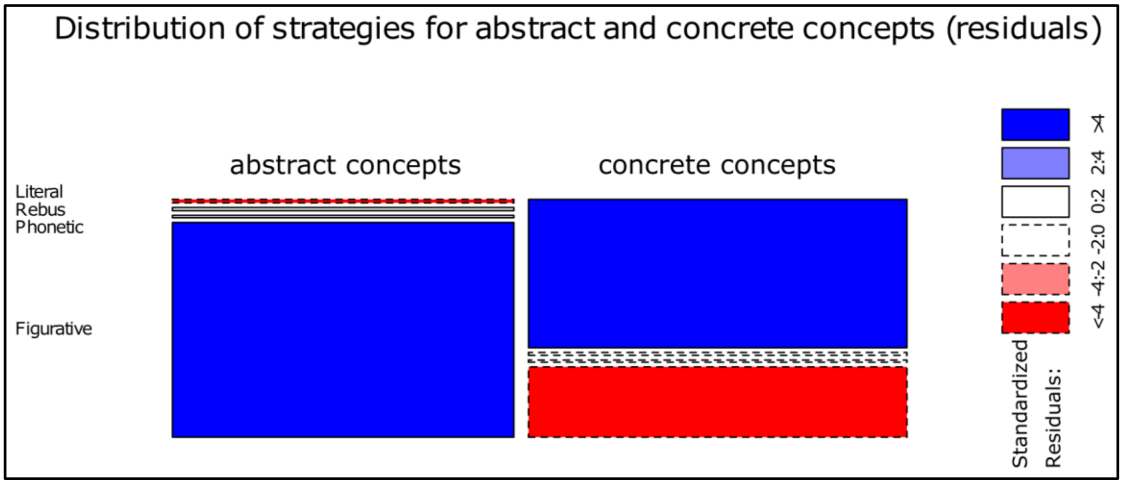

Figure 5. Mosaic plot displaying the analysis of the residuals for the type of strategy and the two types of concepts (abstract and concrete). The analysis and the plot were performed in R (version 3.5.1). The table of residuals is stored in the online repository on OSF.

In relation to the average number of emoji used to represent concrete and abstract concepts, abstract concepts required a significantly higher number of emoji than concrete ones (t= -9.506, p<0.001).

Finally, in relation to the type of sequence used to represent concrete and abstract concepts, Table 8 shows the frequencies by which the sequence types were produced for abstract and concrete concepts. A Pearson's Chi² test showed that there is a significant dependence between the two variables (type of sequence and concreteness): Chi-square = 189.020, df = 3, Cramer's V = 0.354, p < 0.001. Figure 6 displays the analysis of the residuals for the type of sequence and the type of concept.

Observed production frequencies	OP	OS	AT	AS	Total

[7] The total number of responses does not add up to 3300, because some of the responses were annotated as null (scam) responses and had to be discarded.

# abstract concepts	1	213	189	**517**	920
# concrete concepts	16	**324**	60	193	593
Total	17	537	249	710	1513

Table 8: Frequencies by which the different types of sequences were adopted to represent abstract and concrete concepts respectively.

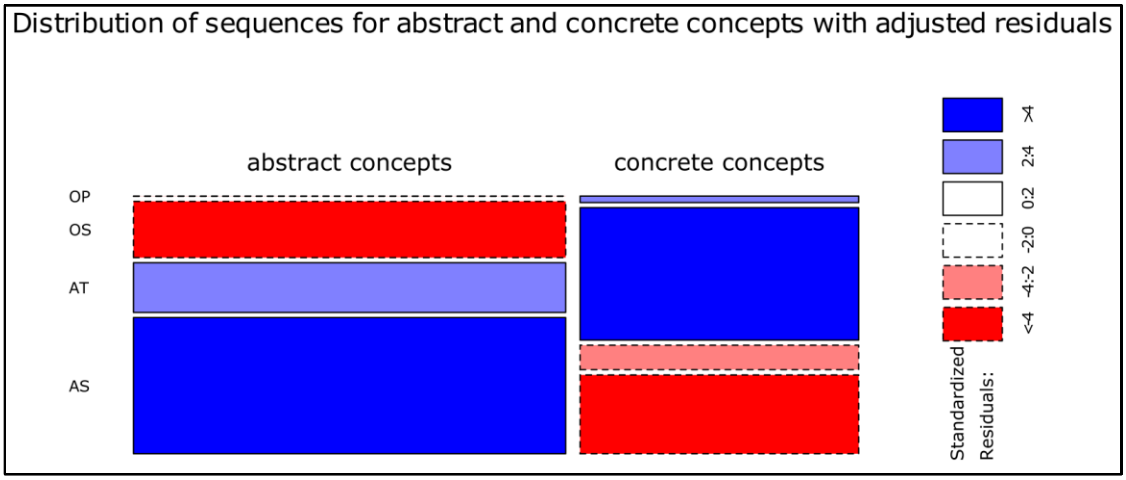

Figure 6. Mosaic plot displaying the analysis of the residuals for the type of sequence and the two types of concepts (abstract and concrete). The analysis and the plot were performed in R (version 3.5.1). The table of residuals is stored in the online repository on OSF.

As Figure 6 shows, abstract concepts are positively associated with AND sequence types (i.e., all the elements of the sequence contribute to a comprehensive representation), whereas concrete concepts are positively associated with OR sequence types (i.e., each element of a sequence of emoji represents the concept by itself, and various alternative options are provided).

Finally, we compared the homogeneity of the responses provided the participants, in relation to abstract and concrete concepts respectively, as described in Section 3.5. To do so, we compared the average cosine similarity scores between representations provided in response to concrete concepts and representations provided in response to abstract concepts (see Section 3.5 for the methods). An unpaired t-test showed that concrete concepts have on average a significantly higher similarity score than abstract concepts ($t = 15.619$, $df = 298$, $p = 1.361*10^{-40}$, Cohen's $D = 1.804$). In other words, the emoji-based representations provided in response to concrete concepts are on average more homogeneous than those provided for abstract concepts.

To summarize, our analyses show that:

- The annotators achieved high agreement scores in the annotation of both representational strategies and sequence types;
- Figurative and literal strategies are the most frequently used strategies;
- And-Spatial and Or-Similar are the most frequently used sequence types;
- Figurative constructs are the preferred strategy for abstract concepts;
- Literal representations are the preferred strategy for concrete concepts;
- Concrete concept representations show greater homogeneity among participants;
- Abstract concept representations show less homogeneity among participants.

4.4. Automatic generation and analysis of averaged representations

Table 3 in Section 3.6 showed how our algorithm averaged the representations provided by various participants into a single representation for the concrete concept GLASS. Table 9, instead, shows how our algorithm averaged the representations provided by various participants into a single representation for the abstract concept COMPARISON. As this example shows, in this case the resulting averaged representation is composed by an apple and an orange, which stand for the idiomatic expression *comparing apples to oranges.*

Concept	Comparison
Answers (Sequence length)	□□ (2), □□ (2), □□ (2), □□ (2), □ (1), □□ (2), □□ (2), □□ (2), □↔□ (3), □♀□ (2), □□ (2)
Sequence lengths (ordered)	1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3
Median - *Md*	2
Most common emoji	□ (2x),□ (2x), □ (2x), □ (1x), □ (1x), □ (1x), □ (1x), □ (1x), □ (1x), □ (1x), □ (1x), □ (1x), □ (1x), ↔ (1x), □ (1x), □♀ (1x), □ (1x), □ (1x), □ (1x)
Average representation (#*Md* most common emoji)	□□

Table 9: Example of the averaged representations algorithm applied to the collected data.

The similarity score of responses for the concept COMPARISON is rather low (0.29), whereas the average representation of the concept MEAT (Table 10), has a similarity score of 0.78.

Concept	Meat
Answers (Sequence length)	□□□□□□ (6), □ (1), □□□□□□ (6), □ (1), □□□□ (4), □□ (2), □ (1), □□ (2), □□□□□ (5), □□□□ (4), □□□ (3)
Sequence lengths (ordered)	1, 1, 1, 2, 2, 3, 4, 4, 5, 6, 6
Median - *Md*	3
Most common emoji	□ (11x),□ (7x), □ (6x), □ (5x), □ (2x), □ (1x), □ (1x), □ (1x), □ (1x)
Average representation (#*Md* most common emoji)	□□□

Table 10: Example of the averaged representations algorithm applied to the collected data. The concept MEAT is represented, on average, with different food/meat emoji.

Further examples show that the algorithm generalizes fairly well: On the one hand, very homogeneous responses such as BOOT, represented by every participant with □, has a similarity score of 1. On the other hand, rebus representations, such as ❌⛺ for EXTENT has a score of 0.15. These examples show that the algorithm is capable of creating average representations of the provided answers and it is capable to transfer the strategy type into the averaged representation. An assessment of the quantitative and qualitative strength of this transfer can be investigated in future work.

The algorithm did not generate an average representation for 14 of the 300 concepts, because the individual responses for these 14 concepts did not share a single emoji, even though they shared *similar* emoji. As an example, the responses for the concept IRONY, represented as follows by the participants: □□, □□, □, □□, □□□♀, □□, ⛈☀, ⁇, □♀□♀, □.

4.5. Comparing the semantic representations of concrete and abstract concepts emerging from classic semantic feature data, with those emerging from visual features (emoji).

Table 11 shows the average correlations obtained when we compared the semantic representations of the 70 overlapping concepts between our dataset and RJ Set, and the 59 overlapping concepts between our dataset and MR Set (as described in detail in Section 3.7).

	Correlation MR Set with emoji set (N=59, all concrete concepts)	Correlation RJ Set with emoji set (N=70, abstract and concrete concepts)
Average Pearson correlation coefficients	0.808	0.769

Table 11: The average Pearson correlation coefficients between the semantic representations of the 59 concepts shared by our dataset and MR set, as well as the 70 concepts shared by our dataset and RJ set.

Both sets show medium and high correlation coefficients with the emoji dataset (between 0.58 and 0.97 with MR set and between 0.43 and 0.99 with RJ set), and on average high correlations between the emoji dataset and the dataset of semantic features expressed by words. This suggests that overall, the semantic representations of these concepts, and the relations between these concepts, are somehow preserved in the two types of spaces, i.e., the semantic space constructed on the basis of visual features and the semantic space constructed on the basis of verbally-produced semantic features.

However, a closer look showed that of the 70 abstract and concrete concepts overlapping with RJ set, 52 are concrete, and only 18 are abstract. We therefore compared the correlation coefficients between the emoji-based featural representations (our dataset) and the word-based featural representations (RJ dataset) for the group of concrete concepts and the group of abstract concepts separately. Through a t-test we observed that the average correlation coefficients between the emoji space and the RJ space, differs significantly for abstract vs. concrete concepts ($t = -3.723$, $p < 0.001$): the average correlation coefficient is higher for the concrete concepts ($M = 0.804$, $SD = 0.115$) than for the abstract concepts ($M = 0.669$, $SD = 0.174$). In other words, the correlation between the semantic representations based on visual features and those based on linguistically generated features (RJ dataset) is particularly high only for concrete concepts, and less so for abstract ones.

5. Discussion

In this paper we present an innovative repository of semantic features, collected for a dataset of 300 abstract and concrete concepts. The repository hereby described and analyzed consists of *visual* representations of the concepts, operationalized by means of emoji, elicited from online workers in a large scale crowdsourcing task. Together with the dataset of emoji based representations produced by the online workers, we also release a dataset of machine-generated averaged representations, based on the emoji that were produced most frequently by the online workers to represent the same concept. This dataset can be used for chatbots, Twitterbots, visual and/or emoji-based storytelling algorithms and other types of AI implementations, within the growing field of computational creativity. Furthermore, it can be regarded as a database of concepts grounded by human knowledge to be compared against machine learning attempts which aim to capture this knowledge.

Overall, our dataset displays various peculiarities, such as the fact that some concepts tend to be uniquely represented by one emoji while others tend to be represented by sequences of emoji. In this section we summarize our specific findings and discuss them within relevant theoretical frameworks addressing the nature of abstract and concrete concepts respectively.

The research questions investigated in the current paper were stated as follows:

RQ1: what are the strategies used to represent the content of concrete vs abstract concepts through emoji?

RQ2: are abstract concepts richer in emotional content (face emoji) than concrete concepts?

RQ3: do abstract concepts need on average more emoji than concrete concepts to be expressed?

RQ4: to what extent do the emoji-based semantic representations correlate with the linguistic features semantic representations?

Through our analysis we demonstrate that, overall, the most common strategy used to represent a concept is the *Figurative Construction*, which covers 59% of the total responses, followed by the *Literal* strategy (33.91%). The *Phonetic* and *Rebus* are the least-common strategies. In relation to the type of concept, figurative strategies are particularly strongly associated to abstract concepts, while literal strategies to concrete ones. Moreover, phonetic and rebus strategies are positively associated to abstract concepts, but not to concrete ones.

We interpret these results in the following way. First of all, figurative strategies, based on figurative operations such as metaphors and metonymies, appear to the most preferred strategy. This suggests that when we do not have at our disposal a visual element (an emoji) that literally represents the referent designated by the concept (e.g., an image of a boot to represent the concept BOOT) we resort to figurative constructions, by means of metaphorical and metonymic associations. We prefer this strategy over more linguistic strategies, such as the phonological strategy and the rebus strategy. This suggests that when asked to represent a concept by visual features (emoji) we focus on the conceptual content, rather than on the linguistic form (i.e., the word) that expresses such concept. The fact, however, that linguistic strategies (phonological and rebus) are significantly more related to the representation of abstract concepts, compared to concrete ones, suggests that indeed linguistic information, and the attention to the signifier rather than the signified, is stronger for abstract concepts, than for concrete ones. This idea is supported by various accounts of cognition (Paivio 1971; 1986; 2010; Borghi and Binkofski 2014) suggesting that linguistic information holds a privileged status in the representation and in the grounding of abstract concepts, compared to concrete ones.

Our results also show that, on average, abstract concepts require a higher number of emoji to be represented, compared to concrete ones. This seems to be in line with theoretical accounts suggesting that abstract concepts are grounded in whole situations, which constitute instances of experiential contexts in which such concepts are experiences (Barsalou and Wiemer-Hastings 2005). In other words, while for concrete concepts people tend (when possible) to represent the entity designated by the concept, and focus on that, for abstract concepts people tend to represent a whole situation in which the abstract concept can be experienced. This finding is also supported by our analysis of the *types* of emoji sequences that characterize the representation of concrete and abstract concepts, respectively. In particular, we observed that while for concrete concepts whenever participants used more than one emoji, these were typically alternative individual representations of the concept, such as various versions of a page to represent the concept page: □□ (sequence labelled as O, which stands for OR), for abstract concepts they tended to use sequences in which each individual emoji contributed to shape an aspect of the concept, and had to be taken together with the other emoji of that sequence, such as □□□ to represent the concept PATIENCE: it requires time, it implies behaving well, and it is usually evaluated as a positive virtue. All these elements contribute to represent the concept of patience (sequence labelled as A, which stands for AND).

Moreover, we showed that the representation of concrete concepts is on average more homogeneous (there is more agreement among online workers on using the same emoji to represent concrete concepts) compared to abstract concepts. This is also intriguing, because it suggests that abstract concepts are more subject to individual variation, due to personal experiences in which such concepts are grounded. Besides being more homogeneous among participants, the semantic representations of concrete vs abstract concepts appear to be more closely related to one another. In other words, the neighbourhood of the concrete concepts.

Finally, we showed that face emoji, which constitute the first type of emoji ever implemented, and from which the very term *emoji* generates, because such facial expressions symbolize the related emotions, are on average more frequently used to represent abstract concepts, compared to concrete concepts. This finding is in line with existing theories aimed at explaining the cognitive grounding of abstract concepts in affective and emotional states (Kousta et al. 2011; Vigliocco et al. 2014). As a matter of fact, our data shows that effective information plays a greater role in the representation of abstract vs concrete concepts.

With our last research question we investigated and compared the semantic representations emerging from classic semantic features expressed through words, to the semantic representations emerging from visual features expressed through emoji. Overall, we reported averaged high degrees of correlations between the semantic representations of concrete and abstract concepts represented by the two types of features i.e., visually and linguistically conveyed. In particular, we found significantly higher correlations between the representations of concrete concepts across the two spaces, compared to the representations of abstract concepts. In other words, it is for abstract concepts that the two different types of features (expressed by words or expressed visually, by emoji) construct slightly different semantic representations. The type of information used to represent (visually) abstract concepts, does not fully overlap with the type of information that can be conveyed through words, to represent the same concepts. This point remains open to further investigation, for the time being.

6. Conclusions

The current article is part of the special issue “Eliciting Semantic Properties: Methods and Applications”. The goal of this special issue is to investigate and extend the use of property listing tasks and semantic feature norming studies, by illustrating their breadth of applications, as well as highlighting the theoretical and methodological issues faced by this paradigm, which may limit its usefulness.

In the current paper we questioned whether the same elicitation procedure (e.g., asking for features expressed by verbal strings) is equally valid for any kind of concept, specifically for concrete vs. abstract ones, and how do elicitation procedures in general affect results. We also questioned whether by looking at visual features (operationalized through emoji) we may observe differences between concrete and abstract concepts that may not be captured when we look at verbally-produced semantic features.

Method-wise, we showed that the manual coding of the raw features into feature types is usually a crucial part of the process, that enables researchers to mine their data in a structured way, by means of content analysis (Bolognesi et al. 2016). However, we should observe that the process of manual annotation becomes somehow ‘immoral’ when the amount of manually annotated data points reaches several thousands items. We therefore suggest that, depending on the aims of the study and on the research questions to be tackled, researchers might need to recast their priorities: in studies for which the manual annotation is crucial, then the number of data points may be limited. In our study, for example, we limited the number of participants to 11 online workers per concept. Notably, we took measures to carefully select the most reliable online workers (this is often crucial in crowdsourcing tasks) and filtered only those that matched our required demographic specifications e.g., regarding their native language. This allowed us to collect and manually annotate 3300 data points, compared to the 9000 data points we would have had to manually annotate, had we opted for 30 participants per concept. Our results show, for example, that even with a limited number of participants judging each concept, there is a clear and significant difference in homogeneity among the responses provided for concrete and abstract concepts respectively: participants tend to agree on the same types of representations for concrete concepts, and to disagree on the representations of abstract concepts, although for the latter category they tend to deploy similar representational strategies. This finding emerges from our data despites the limited number of participants. Conversely, in studies with different aims, where a manual annotation process is not envisioned, researchers might prioritize the number of participants who are submitted to the property generation task, opting for a larger sample of online workers.

To conclude, with our study we discussed theoretical and methodological issues involved in the property generation tasks and semantic feature norms paradigm; we investigated the differences in representing abstract and concrete concepts; we highlighted the need to extend this experimental paradigm beyond the current borders, and we described and released an innovative dataset of visual features, based on emoji, which can be used in many fields, ranging from classic cognitive scientific research, to the most recent developments in AI and computational creativity.

7. Acknowledgements

The research has been sponsored by the Creative Language Systems Group (afflatus.ucd.ie). The authors would like to acknowledge the financial and scientific support on Prof. Tony Veale, director of the Creative Language Systems Group, for the realization of this project.

This article is the result of the close collaboration of both authors, who equally contributed to it. In general, Philipp Wicke was responsible for the data collection, crowdsourcing has been conducted by the Creative Language Systems Group under the supervision of Prof. Tony Veale; the analysis was performed by both authors, and the results discussed together. Both authors equally contributed to write the paper and edit each other's' text. For the specific concerns of the Italian academic attribution system, Marianna Bolognesi is responsible for writing sections 1, 2, 3.3, 3.7, 4.5, 5, 6. Philipp Wicke is responsible for writing sections 2.1, 3.1, 3.2, 3.4, 3.5, 3.6, 4.1, 4.2, 4.3, 4.4.

7.1. Exemption from ethical approval

The study of crowdsourcing emoji sequences on *figure-8* has been granted exemption from requiring ethics approval by the Ethics Committee of the first author's home university, University College Dublin, Ireland, under the protocol number UCD HREC-LS, Ref.-No.: LS-E-19-7-Wicke-Veale. The study has been granted exemption as it included an anonymous survey that did not involve identifiable data, nor any vulnerable groups. All participants took part voluntarily to the study, thus agreeing with the terms and condition of the platform *figure-8*. All procedures performed were in accordance with the ethical standards of the institutional and/or national research committee (UCD HREC-LS, Ref.-No.: LS-E-19-7-Wicke-Veale) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

8. References

Barbieri F, Camacho-Collados J, Ronzano F, Anke LE, Ballesteros M, Basile V, Patti V, Saggion H (2018) SemEval 2018 Task 2: Multilingual Emoji Prediction. Proceedings of The 12th International Workshop on Semantic Evaluation 24–33.

Barbieri F, Kruszewski G, Ronzano F, Saggion H (2016a) How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics. Proceedings of the 2016 ACM on Multimedia Conference. ACM, 531–535.

Barbieri F, Ronzano F, Saggion H (2016b) What does this Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis. In LREC.

Baroni M, Barbu E, Murphy B, Poesio M (2010) Strudel: A distributional semantic model based on properties and types. Cogn. Sci. 34:222–254.

Barsalou LW, Kyle Simmons W, Barbey AK, Wilson CD (2003) Grounding conceptual knowledge in modality-specific systems.Trends Cogn Sci 7(2):84-91.

Barsalou LW (2008) Grounding symbolic operations in the brain's modal systems. In: Semin & Smith (eds.), Embodied grounding: Social, cognitive, affective, and neuroscientific approaches. New York: Cambridge, pp. 9–42.

Barsalou LW, Wiemer-Hastings K (2005) Situating abstract concepts. In: Pecher and Zwaan (eds.) Grounding cognition: The role of perception and action in memory, language, and thought. Cambridge: University Press, pp. 129-163.

Behrend TS, Sharek DJ, Meade AW, Wiebe EN (2011) The viability of crowdsourcing for survey research. Behavior research methods 43(3): 800-813.

Binder JR, Westbury CF, Possing ET, McKiernan KA, Medler DA (2005) Distinct brain systems for processing concrete and abstract concepts. Journal of Cognitive Neuroscience 17: 905–917.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex 19(12): 2767–2796.

Bolognesi M, Pilgram R, van den Heerik R (2017) Reliability in content analysis: The case of semantic feature norms classification. Behavior research methods 49(6): 1984-2001.

Bolognesi M, Steen G (eds.) (2018) Abstract concepts: Structure, processing and modeling. Special Issue of TopiCS in Cognitive Science, 10 (3).

Bolognesi M (2017) Using semantic features norms to investigate how the visual and verbal modes afford metaphor construction and expression. Language and Cognition 9(3): 525-552.

Brabham DC (2010) Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application. Information, Communication & Society 13(8): 1122-1145.

Bright P, Moss H, Tyler L (2004) Unitary vs. multiple semantics: PET studies of word and picture processing. Brain and Language 89:417–432.

Brysbaert M, Warriner AB, Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. Behavior research methods 46(3): 904-911.

Chee M, Weekes B, Lee K, Soon C, Schreiber A, Hoon I, Chee M (2000) Overlap and dissociation of semantic processing of Chinese characters, English words, and pictures . Neuroimage 12:392–403.

Collins AM, Loftus EF (1975) A spreading-activation theory of semantic processing. Psychological Review 82: 407-428.

Cree GS, McRae K, McNorgan C. (1999) An attractor model of lexical conceptual processing: Simulating semantic priming.Cognitive Science 23:371–414.

Cree GS, McRae K (2003) Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). Journal of Experimental Psychology: General 132:163-201.

Cree, GS, McNorgan, C, McRae, K (2006) Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. J Exp Psychol Learn Mem Cogn 32(4):643-658.

Davidoff J, De Bleser R (1994) Impaired picture recognition with preserved object naming and reading . Brain and Cognition 24:1–23.

Eisner B, Rocktäschel T, Augenstein I, Bošnjak M, & Riedel S (2016) emoji2vec: Learning emoji representations from their description. arXiv preprint arXiv:1609.08359.

Farah M (1990) Visual agnosia: disorders of object recognition and what they tell us about normal vision. Cambridge, MA: MIT Press.

Farah MJ, McClelland JL (1991) A computational model of semantic memory impairment: Modality specificity and emergent category specificity. Journal of Experimental Psychology: General 120(4):339-357.

Garrard P, Lambon Ralph MA, Watson PC, Powis J, Patterson K, Hodges JR. (2001) Longitudinal profiles of semantic impairment for living and nonliving concepts in dementia of Alzheimer's type. Journal of Cognitive Neuroscience 13:892–909.

Gates L, Yoon M (2005) Distinct and shared cortical regions of the human brain activated by pictorial depictions versus verbal descriptions: an fMRI study. Neuroimage 24:473–486.

Gorno-Tempini M, Price C, Josephs O, Vandenberghe R, Cappa S, Kapur N, Frackowiak R (1998) The neural systems sustaining face and proper-name processing. Brain 121:2103–2118.

Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34:490–497.

Henning P, Domin C, Rohs M (2017) Beyond just text: semantic emoji similarity modeling to support expressive communication□□□. ACM Transactions on Computer-Human Interaction (TOCHI) 24(1):6.

Highfield T (2018) Emoji hashtags//hashtag emoji: Of platforms, visual affect, and discursive flexibility. First Monday 23, 9.

Jespersen B, Reintges C (2008) Tractarian Sätze, Egyptian hieroglyphs, and the very idea of script as picture. In The Philosophical Forum Vol. 39. Wiley Online Library, 1–19.

Kousta ST, Vigliocco G, Vinson DP, Andrews M, Del Campo E (2011) The representation of abstract words: why emotion matters. Journal of Experimental Psychology: General 140(1):14-34.

Leech G, Rayson P (2014) Word frequencies in written and spoken English: Based on the British National Corpus. London: Routledge.

Lenci A, (2018) "Distributional models of word meaning." Annual review of Linguistics 4: 151-171.

Lenci A, Lebani G, Passaro L (2018) The Emotions of Abstract Words: A Distributional Semantic Analysis. Topics in Cognitive Science 10(3):550-572.

Lynott D, Connell L (2013) Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. Behavior Research methods 45:516–526.

McRae K, Boisvert S (1998) Automatic semantic similarity priming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(3), 558-572.

McRae K, Cree GS, Seidenberg MS, McNorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. Behavior Research methods 37(4):547-59.

McRae K, Cree GS, Westmacott R, Sa VRD (1999) Further evidence for feature correlations in semantic memory. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 53(4), 360.

McRae K, Jones MN (2013) Semantic memory, in Reisberg (ed.) The Oxford Handbook of Cognitive Psychology. New York, NY: Oxford University Press, pp. 206–219.

Medin DL, Schaffer MM (1978) Context theory of classification learning. Psychological Review 85, 207-238.

Minda JP, Smith JD (2002) Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. Journal of Experimental Psychology: Learning, Memory, & Cognition 28, 275-292.

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Mirman D, Magnuson G. (2009) The effect of frequency of shared features on judgments of semantic similarity. Psychonomic Bulletin & Review 16, 671–677.

Moore C, Price C (1999) Three distinct posterior basal temporal lobe regions for reading and object naming. Neuroimage 10:181–192 .

Paivio A (1971) Imagery and verbal processes. New York: Holt, Rinehart, and Winston.

Paivio A (1986) Mental representations: a dual coding approach. Oxford. England: Oxford University Press.

Paivio A (2010) Dual Coding Theory and the mental lexicon . The Mental Lexicon 5:205-230.

Pecher D, Zeelenberg R, Barsalou LW (2004) Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. Psychonomic Bulletin & Review, 11:164-167.

Pexman PM, Holyk GG, Monfils MH (2003) Number-of features effects and semantic processing. Mem. Cognit. 31: 842–855.

Pexman PM, Lupker SJ, Hino Y (2002) The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. Psychon. Bull. Rev. 9:542–549.

Recchia G, Jones MN (2012) The semantic richness of abstract concepts. Frontiers in Human Neuroscience 41(3):647–656.

Reijnierse WG, Burgers C, Bolognesi M, Krennmayr T (2019) How polysemy affects concreteness ratings: The case of metaphor. Cognitive Science 43(8) e12779.

Reinholz J, Pollmann S (2005). Differential activation of object-selective visual areas by passive viewing of pictures and words. Cognitive Brain Research 24: 702–714 .

Sartori G, Polezzi D, Mamelia F, Lombardi L (2005) Feature type effects in semantic memory: An event related potentials study. Neurosci. Lett. 390:139–144.

Smith EE, Medin DL (1981) Categories and concepts. Cambridge, MA: Harvard University Press.

Stark L, Crawford K (2015) The conservatism of emoji: Work, affect, and communication. Social Media + Society 1, 2.

Taggart C (2015) New words for old: Recycling our language for the modern world. Michael O'Mara Books.

Taylor KI, Moss HE, Tyler LK (2007) The conceptual structure account: a cognitive model of semantic memory and its neural instantiatios. In: Hart, Kraut (eds.) Neural basis of semantic memory. Cambridge: University Press. pp. 265–301.

Ursino M, Cuppini C, Cappa SF, Catricalà E (2018) A feature-based neurocomputational model of semantic memory. Cognitive Neurodynamics 12(6):525-547.

Veale T (2017) Déja vu all over again. In Proceedings of the 8th International Conference on Computational Creativity.

Vigliocco G, Vinson DP, Lewis W, Garrett MF (2004) Representing the meanings of object and action words: The featural and unitary semantic system hypothesis.Cogn. Psychol, 48(4):422–488.

Vigliocco G, Meteyard L, Andrews M, Kousta S (2009) Toward a theory of semantic representation. Lang Cogn.1:219–247.

Vigliocco G, Kousta ST, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, Cappa SF (2014) The neural representation of abstract words: the role of emotion. Cereb Cortex 24(7):1767-1777.

Vinson DP, Vigliocco G (2008) Semantic feature production norms for a large set of objects and events. Behavior Research methods 40(1):183-190.

Warrington E (1985) Agnosia: the impairment of object recognition . In Frederiks (ed.), Handbook of clinical neurology. New York: Elsevier (pp. 333–349).

Warrington E, McCarthy RA (1987) Categories of knowledge further fractionations and an attempted integration. Brain 110: 1273-1296.

Wicke P (2017) Ideograms as semantic primes: Emoji in computational linguistic creativity. Unpublished dissertation. DOI: 10.13140/RG. 2.2. 21344.89609.

Wiemer-Hastings K, Xu X (2005) Content differences for abstract and concrete concepts. Cognitive Science 29(5): 719-736.

Wierzbicka A (1996) Semantics: Primes and universals: Primes and universals. Oxford University Press, UK.

Wu LL, Barsalou LW (2009) Perceptual simulation in conceptual combination: Evidence from property generation. Acta Psychologica 132: 173-189.

Appendix

Instructions used for the Crowdsourcing task:

> *In this job you are asked to provide an Emoji selection for a given word. For this you will be directed to a website that allows you to select Emoji in order to copy and paste them into this job. The first question will ask you whether you are a native English speaker. You only need to answer this question once and can skip after you have answered it once. For each task you will be presented with the word you need to describe. Here is an example: You have to represent the word in the red box ("Time Bomb") with a sequence of Emoji. The second step is to click on the link that directs you to the Emoji keyboard. The Emoji keyboard presents you with a variety of Emojis you can pick. Scroll down to see the entire range of categories of Emoji. You can choose from any of the available Emoji on this website. Once you found an appropriate Emoji click on it. This will add it to the text box on the bottom. After you have selected the Emoji that you need, you click on the COPY button at the end of the page and return to this task. When you return back to the task paste the sequence to finish the task.*