Impact of linearity on emoji sequences

Limitations of emoji

Author: Lena Thamsen

ANR: u236425

SNR: 2016596

Master Thesis

Study: Communication and Information Sciences

Track: Business Communication and Digital Media

Faculty: Tilburg School of Humanities

University: Tilburg University

Supervisor: Dr. N.T. Cohn

Second Reader: Dr. J.A.A. Engelen

Date: January 2019

Acknowledgements

This thesis would not have been possible without the encouragement and patience of my supervisors, friends and family. I would like to take the opportunity to thank my supervisor Neil Cohn for guiding and helping me through the process of planning, performing and writing this Master thesis, as well as for his valuable feedback. Furthermore, I would like to thank my second reader Jan Engelen for his comments on my Go-No-Go document and for giving me the chance to conduct this research. Another thank you goes to my family who supported me during the process of my thesis and the entire Master studies. In addition, I would like to thank all respondents for their participation in the experiment, as well as for their encouraging comments and their cooperation.

Abstract

Emoji are visuals, which are forced to behave like written text. They are proclaimed to emerge into a new language (Danesi, 2016; Lu, Ai, Liu, Li, Wang & Mei., 2016), generating the question whether emoji possess the ability to form as rich and complex expressions as either the written or visual language. This thesis explores the limitations of emoji-only expression due to their artificial linearity by looking at the affordances of the written and visual modalities and their required adjustments. Emoji are demanded to act according to the affordances of the written modality, forcing them to occur in linear sequences, which in return obliges them to lose some of their visual characteristics. The present study investigates whether the linearity of emoji-only expressions limits how people can process and understand emoji sequences. The results show that linear emoji sequences require significantly longer reaction times and cause considerably more errors than visual expressions using analog images. Hence, emoji sequences are strongly limited by their artificial linearity.

Keywords: emoji, linearity, affordances, written language, visuals, emoji sequencing

Table of Contents

1. Introduction

The increasing importance of social media and the expansion of computer-mediated conversations have caused emoji to become a ubiquitous aspect within our daily mobile communication (Hakim, 2017). Emoji are graphic symbols carrying meanings (Lu, Ai, Liu, Li, Wang & Mei., 2016), which are widely used to embellish and humanize text messages (Hakami, 2017). They function in similar ways as gestures along speech: Supplementing and enhancing writing (Cohn, 2015; Hakami, 2017). Our natural language is rarely ever limited to speech alone (Cohn, 2015). Instead, 90% of spoken utterances in descriptive discourse are accompanied by gestures (Kelly & Watts, 2015) in order to clarify meanings and to transmit non-verbal cues (Riva, 2002). For this reason, language is claimed to be “multi-modal” (Cohn, 2015). In strictly written communication, however, non-verbal cues, like facial expressions, thoughts and feelings, cannot be transmitted (Dresner & Herring, 2010; Kelly & Watts, 2015). Thus, emoji allow for reincorporation of non-verbal information in text-based conversations. They possess the ability to substitute nouns and actions, as many emoji embody objects, events or activities (Dresner & Herring, 2010; Pavalanathan & Eisenstein, 2016). Emoji are used alongside with or substitutional for written text and function as cues for controlling the emotional valence of textual communication (Walther & D’Addario, 2001). For instance, adding a laughing face (😀) can disambiguate whether a statement is meant to be sarcastic/ironic, or if it is meant in a mean or funny manner.

The way emoji appear alongside with written language obliges them to behave like written text, even though emoji actually fall into the category of graphics (Danesi, 2016). However, the affordances of visuals and written language differ in several aspects, generating some limitations on how emoji can be employed (Cohn, 2015). Forcing a modality into the affordances of another modality requires adaptation (Cohn, 2013), which often leads to constraints in how specific modality-characteristics interact and how they can or should be used (Gibson, 1979). The way emoji are utilized in text-based conversations, forces them into artificial linear sequences (Cohn, Engelen & Schilperoord, 2018). Although the combinational properties of visuals naturally allow graphics to occur as analog images, which are displayed as one spatial unit (Cohn, 2012a), emoji can only be combined linearly, which limits their ability to form complex expression (Cohn et al., 2018). Nevertheless, many scholars still proclaim emoji as an emerging language (Danesi, 2016; Lu et al., 2016). Hence, the question arises whether emoji really possess the ability to become their own new language, or if their unnatural linearity limits how fast and how accurately emoji meanings can be processed.

2. Theoretical Framework

2.1 Affordances of Writing and Drawing

When writing a text, people use building blocks (letters) to create units (words) (Cohn, 2015), which can later be combined to whole sentences. Therefore, written language allows for complex utterances with many different options of variation (Klein, 1985). Alphabetic writing matches the syntax of the language it transcribes, allowing for the conjunction of single ideas into complex interconnected ones in some rule-based arrangement (Danesi, 2016).

Furthermore, alphabetic written scripts involve laying out texts in some linear form, such as from left-to-right, right-to-left, up-down, or down-up (Danesi, 2016). The way alphabetic written language is structured forces the visual building blocks (letters) and units (words) to appear in a linear way - one letter after the other, and one word after the other (Klein, 1985). Hence, when looking at Figure 1 it can be seen that letters (building blocks) of a single word (unit) are structured linearly. In other words, every unit consists of linearly composed building blocks. For instance, when looking at the word *sun*, the letter *s* is used first, then *u*, and then *n* to form the unit *sun*. Rearranging the order of the building blocks would change the meaning of the unit. Likewise, single units also appear linearly in a sentence setting (Klein, 1985). The units *the*, *sun*, *is* and *shining* are laid out in a linear way in order to compose a more complex sentence structure.

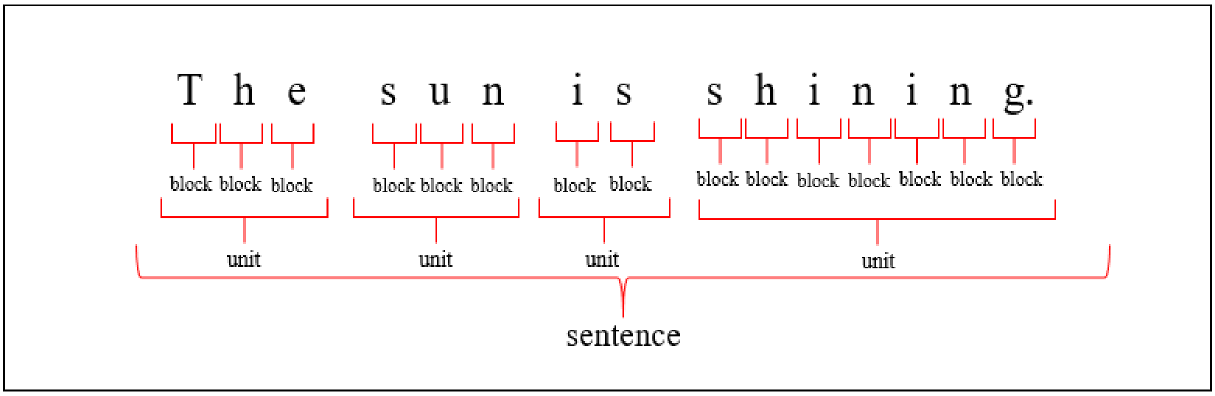


Figure 1. Linear composition of building blocks and units in a sentence setting

Furthermore, written language allows for applications of building blocks in various ways (Klein, 1985). The English alphabet offers 26 buildings blocks, which can be arranged and rearranged to form new units (Wordcounter, 2015). While several languages like Spanish and Polish consist of more than 26 letters, languages like Italian and Portuguese involve less than 26 building blocks (Wordcounter, 2015). However, in every alphabetic writing system building blocks allow for complex and various combinations without searching through a list of every whole word in the language (Cohn, 2015). Figure 2 displays how new words can be created by combining different building blocks. By adding letters to the unit *sun*, it is possible to compose a variety of new words (e.g. forming the word *sunny*). Furthermore, when combining the unit *sun* with another unit like *shine* in a linear way, a whole new unit can be created.

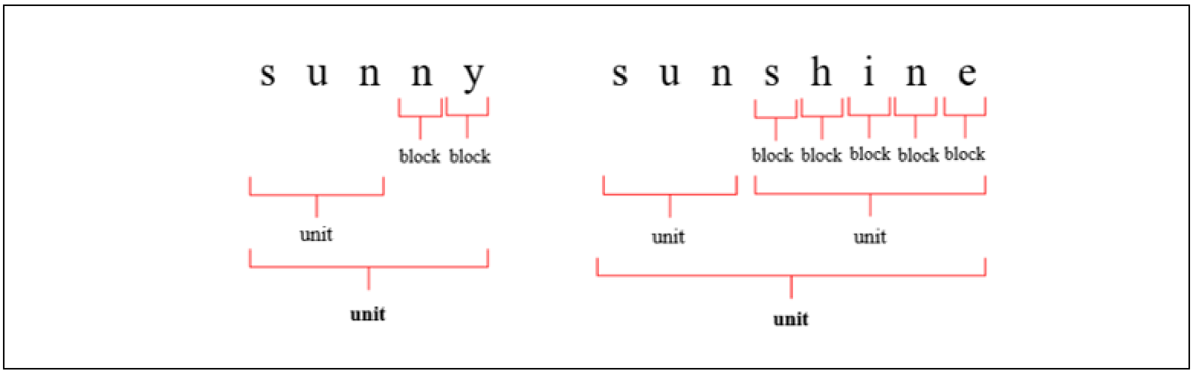


Figure 2. Combinational properties of building blocks and units in written texts

Linearity is a universal feature and one of the main characteristics of writing (Puranik & Lonigan, 2009). Puranik and Lonigan (2009) found that children can demonstrate universal features of writing before they are able to develop knowledge of the language-specific characteristics of a writing system. Hence, young children do not demonstrate knowledge of letters as they are usually constructed or show any correspondence to the length of the spoken word, but rather correspond to features common to all languages such as linearity and segmentation of letters and words (Puranik & Lonigan, 2009). Before being able to write conventionally, children try to express meaning through scribbles, such as by using dots, circles, and shapes, which are arranged linearly and segmented by small spaces (Puranik & Lonigan, 2009). These findings support the Linearity Hypothesis stating that universal features of writing, such as linearity and discreteness, are learned before language-specific features, such as spelling and directionality (Tolchinsky, 2003). This means that although writing cannot be understood by young children in terms of language-specific features before they receive formal writing instructions in school, they still possess the ability to reproduce common characteristics of writing. Hence, they are able to detect and reflect that a word consists of building blocks and that these blocks are composed linearly. Moreover, they perceive that buildings blocks, as well as units are separated by spaces, although they do not comprehend that this is to convey meaning (Puranik & Lonigan, 2009).

To conclude, linearity is such a strong characteristic of writing that it can be identified as one of the main affordances of this modality, even by people who do not comprehend all properties of the writing system. Linearity is therefore in the nature of the modality of written language (Danesi, 2016).

A similar building pattern as in writing is also visible in drawings: Simple building blocks, such as lines and shapes, are combined to create larger units, like representational drawings (Cohn, 2012a). However, the production and perception of a spatially perceived whole image does not happen linearly. In many cases, the perceiver of a drawing does not see how the building blocks of an image are composed temporally, which means that production and comprehension of a drawing usually remain separated (Cohn, 2012a). In written language linearity occurs because time is converted into space. In drawings, however, time is only needed to add the spatial properties together, while linearity and the conversion from time into space is not required (Cohn, 2012a). Hence, written language possesses temporality and linearity, whereas the temporality of drawings disappears once the production has ended (Cohn, 2012a).

Like language, drawings are highly conventionalized to individuals and groups, which results in the development of graphical schematic patterns (Wilkins, 1997). It was found that when people are asked to very quickly draw a house and a person, they consistently draw the same simple conventionalized images (Kistler, 1988) (Figure 3). Interestingly, nobody's

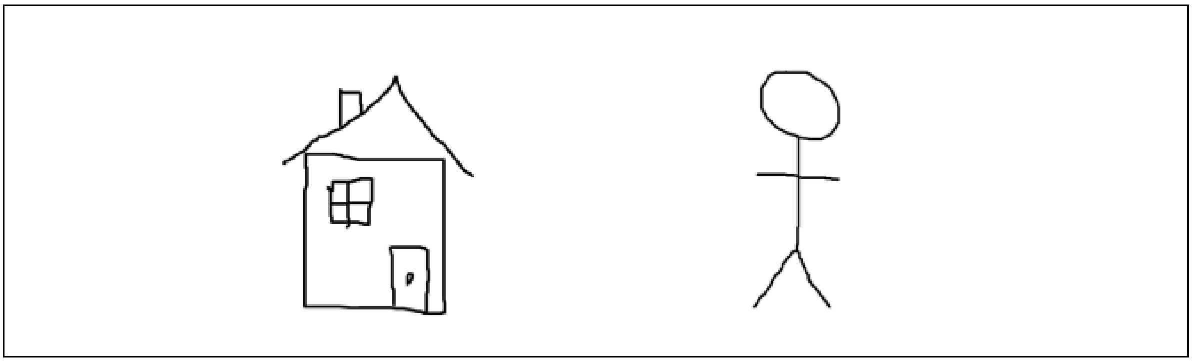


Figure 3. Conventionalized represenations of a house and a person

house actually looks like the common representation, and a stick figure only resembles an actual person in very basic features. However, these schematic representations reflect the conventionalized patterns of drawing, although they might slightly differ culturally (Cox, 1998).

Drawings allow to connect graphic lines and shapes in perceptible ways (Cohn, 2012a), and therefore offer various ways of combinations. People store schematic parts of images, which can then be combined in a larger picture (Wilson, 1988). Hence, it is possible to express more complex situations in drawings by simply combining building blocks spatially (Cohn, 2012a). For instance, when we want to draw a grandma, we can add specific features, such as glasses or curly hair to the image to make it look more resembling to the image of the grandma, which we have memorized in our head (Figure 4).

Figure 4. Options of spatial combinations of building blocks

Our graphical lexicon stores visual vocabulary used for drawings and allows us to encode various schemas during the drawing process (Cohn, 2012a). “Graphical lexical items” range in size, which means that the lexicon must involve individual graphemes that form the basic graphic parts of a representation, such as dots, lines, curves, circles, or squares (Cohn, 2012a). The foundation for grapheme production already begins when children start drawing by scribbling in early age (Golomb, 1992). This usually happens alongside with children's first attempts of speaking by playing with the phonemes of speech (Golomb, 1992; Matthews, 1983). Furthermore, the graphic lexicon also stores schematic elements of images, such as the way hands or feet are visually expressed (Wilson & Wilson, 1977; Wilson, 1988), or even full schematic representations like stick figures (Cohn, 2012a). These stored parts can be used combinatorically in order to create a larger picture.

While written language connects elements across a linear sequence, drawings combine elements in a single spatial display, showing that the combinatorial properties of these two modalities differ. This means, that instead of drawing single elements next to each other, the nature of drawings allows us to integrate parts spatially (Cohn, 2012a). For instance, the written utterance *My grandma has a big nose* combines the single building blocks (letters) and full units (words) in a linear sequence. When expressing the same utterance visually, we would not draw the *grandma* first and a *big nose* next to her (Figure 5). Instead, we would combine the elements *grandma* and *big nose* as one analog image (see Figure 6).

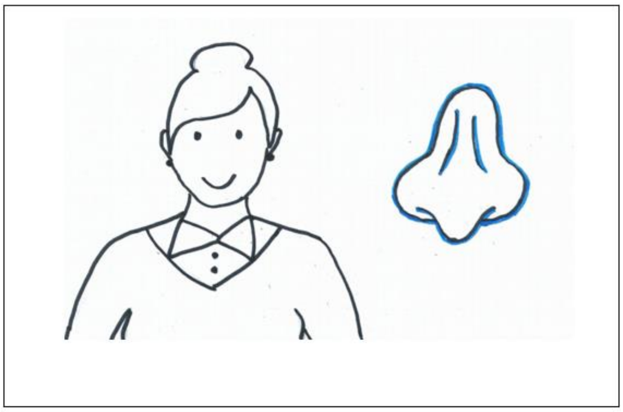


Figure 5. Linear drawing of the phrase: "My grandma has a big nose."

Figure 6. Analog drawing of the phrase: "My grandma has a big nose."

Hence, the main characteristics of drawings, and visuals in general, are their flexibility in how simple building blocks and schematic representations can be combined and that elements can appear in various types of spatialities. The nature of visuals requests graphics to occur as analog displays, making it possible to express complexity in one spatial unit.

2.2 Affordances of Emoji

Emoji are used to supplement, enhance and humanize written text (Cohn, 2015; Duerscheid & Siever, 2017; Hakami, 2017), allowing for a reincorporation of non-verbal cues into written conversations (Kelly & Watts, 2015). However, emoji are demanded to act according to the affordances of the written modality, forcing them to occur in linear sequences, which in return obliges them to lose some of their visual characteristics. As mentioned above, one of the main affordances of the written modality is its linearity (Klein, 1985), while the affordances of visuals allow for combinations in analog spatial displays (Cohn, 2012a). Hence, there is a significant difference of the combinational properties of both modalities, which puts several limitations on how emoji can be processed (Cohn, 2013). As a result, emoji's main affordances are their artificial linearity and low combinational flexibility, which reflect on several aspects regarding emoji sequencing, grammar and visual displays. Since scholars have proclaimed emoji to emerge into a new language (Danesi, 2016; Lu et al., 2016), the question arises whether emoji show the same characteristics of other communicative systems and actual languages (Cohn, 2015). To really be able to become their own language, they would require one key component: grammar (Cohn, 2015). However, emoji's artificial linearity not only influences how emoji-only expressions can be visually displayed, but also what kind of grammar-like structures can be conveyed by them and how complex they can be clustered in sequences.

2.2.1 Emoji Sequencing and Grammar. In text-based conversations, people often use emoji sequences, which occur primarily in repetition, or reduplication, of the same emoji (McCulloch & Gawne, 2018). Furthermore, emoji are frequently clustered with others that fit thematically, or are part of a common semantic field (McCulloch & Gawne, 2018). Danesi (2016), additionally, found that emoji possess the ability to substitute whole words that sound like the terms of the images portrayed, but without conveying the same meaning. This process of converting or transliterating words into emoji is called calquing (Danesi, 2016). For instance, when expressing the utterance *bombshell bikini* the following emoji could be used: 💣🐚👙 (Danesi, 2016). In regard to emoji's affordances, this emoji sequence is composed linearly, while each emoji substitutes one part of the written utterance. However, the meaning of each emoji does not convey the actual meaning of the utterance it is supposed to express (an actual bombshell bikini does not consist of a bomb and shell). When emoji are used in this substitutional way, it is reminiscent of a picture puzzle (Danesi, 2016). Thus, in order for the receiver to understand calquing, they must possess an advanced knowledge of a specific natural language (Danesi, 2016).

In order to emerge into their own language, emoji would require some sort of grammar-like structures in the way they are sequenced (Cohn, 2015). Grammatical systems are based on a set of constraints, controlling how the meaning of an expression can be presented in a coherent way (Cohn, Jackendoff, Holcomb, & Kuperberg, 2014). Jackendoff and Wittenberg (2014) characterized the range of linguistic structures and utterances from words to phrases as "the hierarchy of grammars". Emoji are often used at the lowest stage of the hierarchy, the level of "one-unit grammars", meaning that single emoji units stand alone without any sequential properties (Cohn et al., 2018). Using "one-unit grammars" in emoji-only expressions requires us to choose one single emoji character from a pre-existing list of emoji. Instead, the nature of visuals would actually allow us to encapsulate multiple units of emoji in one analog unit. Moreover, emoji sequences primarily appear at the level of “linear grammars”, with a reduced use of categorical roles and embeddings (Cohn et al., 2018). Linear grammars entail a series of juxtaposed units, with no internal phrasal connections (Jackendoff & Wittenberg, 2014), and therefore remain in fairly low levels of the grammatical hierarchy (Cohn et al., 2018). Hence, it is fairly hard to convey more complex grammatical structures by emoji-only expressions. In natural language, grammar and meaning differ from each other. Therefore, an active sentence roughly conveys the same meaning as its passive version (Cohn, 2015). For instance, the active sentence *Farah kissed Tommy* and the passive version *Tommy was kissed by Farah* both transmit the same meaning but differ in the sequencing of their grammatical structure. When expressing the same utterance by using only emoji, the restructuring from active to passive would influence the sequencing of their grammatical structure, as well as their conveyed meaning: Active sentence: 👧💋👦; passive sentence: 👦💋👧. When interpreting the passive version, one could think that *Tommy kissed Farah* rather than *Tommy was kissed by Farah*. Hence, the artificial linearity of emoji limits the possibility to differentiate between active and passive sentence structures.

Moreover, emoji frequently appear in lists, which can either be ordered or unordered (Cohn et al., 2018; McCulloch & Gawne, 2018). When these lists are unordered, they could be rearranged and would still keep the same meaning, as for instance: 🏖👙🌊🌅☀ (listing some random emoji regarding beach). This shows that unordered lists impart no internal connections (Cohn, 2015), and are a conventionalized way of communicating without involving a grammar the way that sentences do (Schnoebelen, 2012). When looking at ordered lists, like 👧👙 🏖☀ (meaning that the girl wore a bikini on the beach in the sun), rearrangement would generate a new order, which may end up becoming an unordered list again, conveying no internal connections other than their semantic field connected with a beach (Cohn, 2015). This shows that rearranging the structure of an ordered list would not keep a consistent meaning, while in natural language, rearranging the grammatical structure of a sentence would have a strong impact on the perception of the meaning (Cohn, 2015).

Grammatical structures have not only been observed in speech and written language. Instead, visual narratives used in comics have been shown to possess a grammatical structure as well (Cohn, 2013). However, the grammar of sequential images does not consist of nouns and verbs but can rather be considered a narrative structure (Cohn, 2013). Nevertheless, the way sequential images act, is based on principles of combination like any other grammar, involving groupings of images, hierarchical embeddings and roles played by images (Cohn, 2013). Cohn et al. (2018) found that when people communicate using emoji-only expressions, they tend to employ fairly simple grammatical constructions like linear grammars, since emoji are forced to always appear in linear sequences. In our natural language, grammar involves categorical roles, which allow for semantic distinctions between agents, patients, objects and actions (Cohn et al., 2018). One important characteristic of linear grammar, however, is placing the agent (or doer) before the patient or object of that action (Chan, Lieven, & Tomasello, 2009). Hence, emoji's artificial linearity may force us to place the agent first, followed by the action. For instance, "🎅🎁" would mean that Santa is the doer (agent), which precedes an action (bringing the present). This pattern is commonly represented in both full languages and simple communication systems (Cohn, 2015). Interestingly, the same agent-action order appears when people arrange pictures to describe events from an animated cartoon, or when people communicate using only gestures (Gershkoff-Stowe & Goldin-Meadow, 2002). Yet, these systems are not considered to be a language, which shows that the agent-action order seems to not appear due to a grammar, but rather due to basic heuristics, based on meaning alone (Cohn, 2015). Emoji, as shown in the Santa-example, seem to fall into the same system.

Furthermore, natural language grammars consist of groupings of units (Cohn, 2015), such as embedded clauses as for example: *Farah, who has curly hair, kissed Tommy*. This means, that our natural language grammar allows us to include clauses into a coherent sentence. In this example, the clause *Farah has curly hair* was embedded into the original sentence *Farah kissed Tommy*. Similarly, visual narratives possess the ability to use center-embedded phases, resembling embedded clauses in our grammatical structures of speech. Figure 7 shows a short comic strip displaying that panel 1, 2 and 5 could be separated from panel 3 and 4 and would still obtain an internal connection and coherent meaning. Panel 3 and 4 could also stand on their own without any connection to the surrounding panels (Cohn, 2012b).

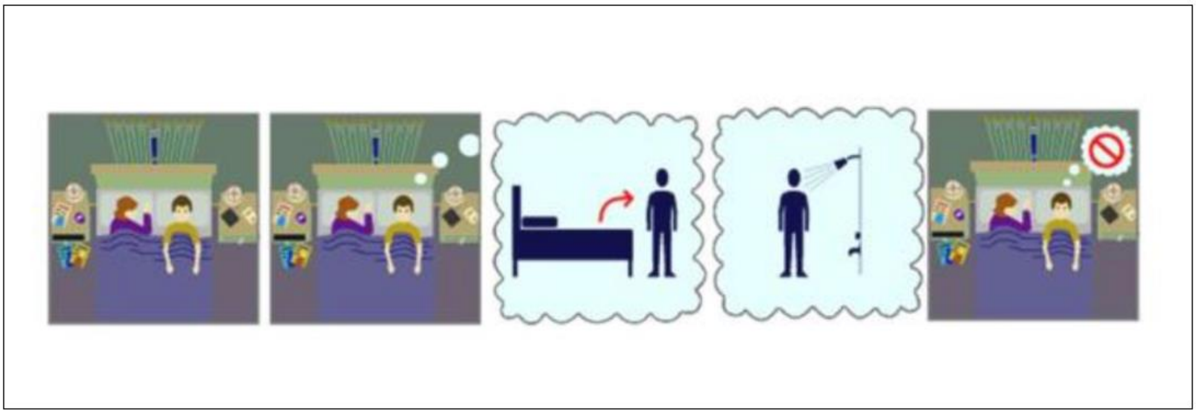


Figure 7. Embedded clauses in comic strips. *One Night art © 2006 Tym Godek*

Even though, there is evidence that the grammar of visual narratives acts in similar ways as sentence grammar, scholars like Cohn (2015) and Schnoebelen (2012) still doubt that emoji possess the same characteristics and features to transmit such complex grammatical structures. When trying to build a sentence with an embedded clause, as explained above, by using only emoji, one is immediately faced by grammatical limitations due to emoji's artificial linearity. For example, transforming the sentence *The boy is eating a lollipop* into an emoji sequence, one is forced to stick to an agent-action sequence, rather than an agent-patient-action sequence (Cohn, 2015): 👦🍭. However, this fairly simple grammatical construction can somewhat be represented by emoji. Yet, when looking at just slightly more complex sentences like *The boy, who likes soccer, is eating a lollipop* the differences between language grammar and emoji become clearer. Transforming this sentence into a linear emoji sequence produces an incoherent unstructured sequence: 👦♡⚽🍭. It is not possible to differentiate the embedded clause from the original clause, which shows that due to the linearity and the strict agent-action pattern, emoji lack the flexibility to form more complex grammatical structures.

To conclude, since emoji are obliged to appear in linear sequences, grammatical limitations occur. It was found that grammatical constraints and violations, as well as semantic and sequential incongruities cause problems in obtaining sentence comprehension and sometimes even require a revision of how words within a sentence are related to each other (Friederici, & Meyer, 2004). Hence, emoji's grammatical constraints suggest that they are not using complex aspects of sequencing and grammar, and that, other than in visual narratives like comics, emoji sequencing remains in fairly low levels of the grammatical hierarchy (Cohn et al., 2018).

2.2.2 Visual Displays of Emoji. In written language we can recombine around 26 letters to form new words, requiring us to only store building blocks, instead of whole units. In contrast to that every emoji in the constantly growing list of emoji characters has to be stored as a whole. The emoji keyboard consists of 2823 icons in the Unicode Standard, which exist in various genres, including facial expressions, common objects, places and types of weather, and animals (Emojipedia, 2018). As the list of emoji grows constantly, it is becoming harder to find the right emoji, which fits best to the expressed statement. Using emoji requires a conscious search process through an external list, which differs from the way we naturally speak or draw (Cohn, 2015). As mentioned before, new words and even drawings can be composed by combining building blocks into larger units (Cohn, 2012a; Cohn, 2015). This means, that instead of searching through an external list of “vocabulary”, we can easily generate new units from our own mental vocabulary during writing and drawing (Cohn, 2015). Emoji do not offer the option to build new units by combining building blocks, which results in a limited flexibility in emoji use.

In 2015, Apple introduced the option to choose between emoji’s skin and hair colors to add more diversity to what emoji can express (NRZ, 2015). However, this change only regarded “people emoji”, while the emoji “face” (smileys) are still only displayed in yellow. Although, this was an important step into the direction of a more flexible use of emoji, more complex combinations are still not possible. Hence, when expressing *my grandma*, we can simply choose the “grandma emoji” (👵) with different skin colors, but we are unable to add any specific features, like length of hair, size of eyes or shape of head.

Furthermore, since emoji are used to substitute or enhance text (Cohn, 2015; Duerscheid & Siever, 2017; Hakami, 2017) they must act according to the affordances of written language. As mentioned above, in written texts building blocks, as well as whole units, are combined in linear sequences (Danesi, 2016; Klein, 1985), whereas in drawings simple building blocks are connected in analog spatial displays (Cohn, 2012). Since emoji do not consist of individual building blocks but are rather presented to the users as whole units (Cohn, 2015), these units can only be combined as linear sequences and not as one analog image. Hence, when coming back to the earlier example *My grandma has a big nose*, it is not possible to express this statement in an analog form (combined in one emoji). Instead, we would have to use the grandma emoji (👵) and the nose emoji (👃) and combine them linearly: 👵👃.

Moreover, due to the strict agent-action pattern of emoji sequencing (Cohn, 2015), it is almost impossible to involve predicates when communicating with only emoji. For instance, when expressing the utterance *The man is taking the dog for a walk*, we would have to use the following emoji in a linear sequence: 🚶🐕. There is no way to express that the man is actually walking the dog. Instead, we can only assume what the man might be doing with the dog, or how these two icons are related (he could also be petting or feeding the dog, or the dog following the man). Drawings, in comparison, would allow to express the same utterance in a spatial display as one unit, involving the actual action of "taking the dog for a walk" (Figure 8).

Another conspicuousness of how emoji are displayed is that it is not possible to adjust sizes of single emoji characters. Every emoji is displayed in the same size (Danesi, 2016), which again is related to the fact that emoji are presented as whole units, rather than as single building blocks (Cohn, 2015). Figure 9 shows how the sentence *The man is taking the dog for a walk* could be expressed in a linear emoji sequence. The way the emoji are presented, make it seem as if the dog is as tall as the man. Furthermore, due to the agent-action pattern in emoji sequencing, we need to display the man (agent) before the dog (action). However, in an analog image the dog would in most cases be portrayed in front of the man to indicate that the man is walking the dog (Figure 8).

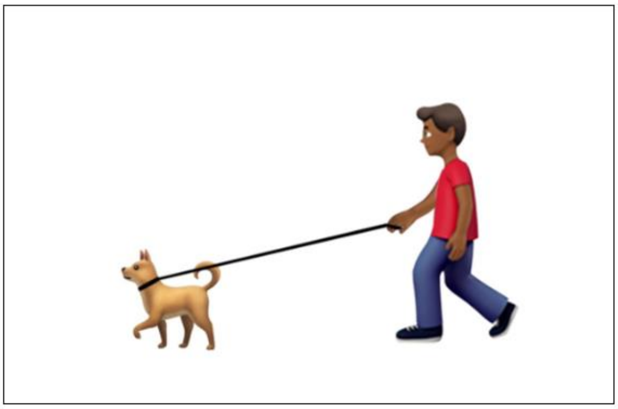


Figure 8. Analog emoji image of the phrase: "The man is taking the dog for a walk."

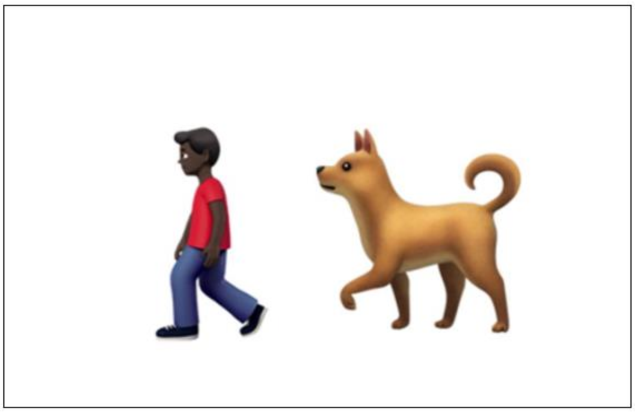


Figure 9. Linear emoji sequence of the phrase: "The man is taking the dog for a walk."

To conclude, emoji do not only lack the flexibility to form complex sequences and convey grammatical structures, but also show several limitations in how they can visually display utterances. Their linearity, as well as the agent-action pattern, force emoji to be used in ways that are usually unnatural for the visual modality.

2.3 Adjusting affordances crossmodally

In our everyday life we process information from different modalities simultaneously (Kircher, Sass, Sachs, & Krach, 2009). In order to properly comprehend and encode meaning, the sensory input is converted into concepts linked to the distributed semantic associations (Mesulam, 1998). The process of comprehension happens automatically and is usually very fast (Kircher et al., 2009).

The so called “Principle of Equivalence” is a proposal for the relationship between modalities in cognition, across structure, development, and processing, and it posits that “We should expect that the mind/brain treats all expressive capacities in similar ways, given modality-specific constraints.” (Cohn, 2013). Thus, the principle indicates that different human behaviors emerge from the same brain, using the same cognitive features, and should therefore all be developed, structured and processed in analogous ways (Cohn, 2013).

However, the Principle of Equivalence also indicates that different human behaviors have some cognitive differences (Cohn, 2013). Thus, while various aspects of cognition overlap across domains, modality-related constraints, like forcing emoji to act like text, need some sort of adaptation in the way they are processed. This means that sometimes the properties between domains differ, because different modalities request different inquiries from their expressive systems (Cohn, 2013), such combinational properties (Cohn, 2012a).

As mentioned before, the way emoji are used in our digital conversations forces them to act like written text. Hence, they must behave according to the characteristics of the written modality, which leads to modality-related constraints. These constraints need, in regard to the Principle of Equivalence, some sort of adaptation in the way they are processed (Cohn, 2013).

When forcing modalities into affordances which seem unnatural, limitations on perception and cognition are observed. As an example, gestures used in sign languages rely on spatial contrasting, while speech is linear and non-spatial (Goldin-Meadow, 1999). Manually Coded English (MCE), however, unnaturally imitates the properties of spoken English in the manual modality (Supalla & McKee, 1991), because it evades the natural properties and demands of an actual sign language (Supalla, 1986; Supalla & McKee, 1991). When forcing sign language to act like spoken language, it was found that MCE violates constraints on the perception and processing of a signed language (Supalla & McKee, 1991). Since sign language uses several forms of expression simultaneously, when forming an utterance (e.g. hands and facial expression), which are all part of the sign language structure, the linearity of MCE would have an impact on how complex people can express themselves. Hence, this would result in a structural limitation of the modality and causes delays and cognitive complications in how the sign language can be processed (Supalla & McKee, 1991).

The example of how MCE affects the structure and cognition of sign language (Supalla & McKee, 1991) shows that when a modality is forced to act according to the affordances of another modality, it influences how this modality is perceived and processed. According to the Principle of Equivalence, it can be assumed that modality-specific constraints somewhat show an impact on how the brain treats a certain expressive capacity (Cohn, 2013).

Likewise, it can be conjectured that because emoji are forced to act against the nature of the visual modality, it will also have an impact on how they can be processed. Visuals usually occur in an analog display (Cohn, 2012a). According to the modality-specific affordances of visuals, processing of analog images should be natural to most people, and it happens instinctively without delays or other constraints. When communicating with emoji, however, one would have to process the images linearly to relate it to the meaning. For instance, the earlier example of *the boy eating a lollipop*, shows how linearity decelerates the processing of the emoji sequence: 👦🍭. We would first need to process the image of *the boy* and then the image of *the lollipop*. Afterwards our minds would need to make a connection between these two images in order to make sense of the emoji sequence, which might be delayed due to the agent-action pattern. If this sequence was presented as an analog emoji image, we could process and understand the image, using aspects of cognition which are natural for the visual domain, facilitating the cognition process (Cohn, 2013). Figure 10 shows the same example utterance *The boy is eating a lollipop*, but this time the boy and the lollipop are combined in one single spatial display. Hence, we can now process the image as one unit, which would accelerate how our brain makes the connection between the image and the meaning.

Figure 10. Analog emoji image of the phrase: "The boy is eating a lollipop."

The same constraints can be observed for compound words. Emoji substitute each part of a compound word, which is reminiscent of a picture puzzle (Danesi, 2016). The compound word *dumpster fire*, for example, could be expressed with the following emoji sequence: 🗑🔥. We would now first have to process the image of the *dumpster* and then the image of the *fire* in order to generate both meanings. Next, we would need to combine both meanings to relate it back to the implied meaning of the compound word. This process indeed reminds of a picture puzzle (Danesi, 2016), or of playing picture charade. Yet, when presenting the same compound word as an analog emoji image, this process would be simplified. As shown in Figure 11, the analog image version displays the *dumpster fire* as one whole image. This does not require us to first process both individual meanings separately from each other to then later combine them again in order to make sense of the meaning. Instead, we can process the meaning in once.

Figure 11. Analog emoji image of the compound word: "dumspter fire"

The affordance of linearity has an impact on the grammatical and sequential complexity of emoji expressions and also influences how intricately emoji utterances can be visually displayed. Therefore, this research aims to investigate how the affordance of linearity decelerates our ability to understand emoji sequences in comparison to analog images. It is assumed that emoji are limited by their constraint to act like text, which forces fairly unnatural linear sequences for graphic communication (Cohn et al., 2018) and therefore impedes the cognition process. Hence, this study focuses on answering the following research question:

RQ: *To what extent is the process of understanding emoji sequences limited by their artificial linearity?*

3. Relevance of the Study

In order to investigate this research question, this study presented people with written utterances preceded by emoji that either matched or mismatched the words. The emoji were either displayed as a linear sequence or as an analog image. To determine to what extent the process of understanding emoji sequences is limited by their artificial linearity, response accuracy and reaction time were measured during the experiment.

Previous studies have not directly investigated how presentation type of images (analog or linear) influences people's ability to process the meaning of an image. Nevertheless, it was found that people's mental imagery appears to influence how pictures are processed in general. An experiment by Shepard (1978) revealed that people are able to make the same quick judgements about images in their absence as in their presence. This means, that they imagined a mental picture of the stimuli object, which helped them to respond to the stimuli. Since people imagine a visual image of what will be presented to them, they are able to respond faster to the stimuli, which accelerates their reaction time. This mental image also plays a role during the process of drawing. The so called 'Image' Image is the mental image of what a person intends to draw (Cohn, 2012a). Hence, when a person is presented with a word like *house*, this house is expected to look somewhat similar to this person's mental picture of a house. This process of imagining a mental picture is common in the visual modality and is part of the visual cognition process (Shepard, 1978). However, as mentioned before, emoji must act against the affordances of the visual modality and instead occur in linear sequences. When forcing a modality to appear in a linear way, if this is against its nature, constraints occur (Supalla & McKee, 1991). People are less familiar with graphics which are displayed in linear sequences, than with analog graphics. Therefore, it can be assumed that since linear sequences are unnatural in the visual modality, people would take longer to respond to linear emoji images than to analog emoji images (Cohn, 2013). Furthermore, people probably already have a mental image in mind, when being presented with a word prime, which further accelerates the process of responding to analog images. Hence, this leads to the following hypothesis:

H1: *People's reaction time to respond to analog emoji images is faster than their reaction time to respond to linear emoji sequences.*

Several studies have found an effect of congruence and incongruence on people's reaction time, when being faced with a lexical decision task. A study by Holcomb and Neville (1990), for instance, compared semantic priming effects in the visual and auditory modalities using accuracy and participant's reaction time as their measures. Participants had to fulfill a lexical decision task where stimuli were word pairs consisting of "prime" words followed by an equal number of words, which were semantically related to the primes (congruent) and words, which were unrelated to the primes (incongruent). The experiment revealed that people responded significantly slower to incongruent words, than to congruent words. Moreover, the participants were also fundamentally less accurate in the incongruent condition than in the congruent condition. Similarly, semantics strongly affect people's reaction time in lexical decisions (word vs. non-word) tasks (Becker, 1979), and words preceded by semantically related words are more quickly recognized as words than if they are preceded by unrelated words (Bentin, McCarthy, & Wood, 1984; Meyer & Schvaneveldt, 1971). Although, these studies mostly compare reaction times of congruence and incongruence within one modality, similar results can be expected when priming crossmodally (Holcomb & Anderson, 1993). Hence, this current study aims to investigate whether this effect also holds true in an emoji-related lexical-decision task. According to the results of former studies it can be assumed that the response process of matching written utterances with semantically related emoji images will be faster than matching written utterances with semantically unrelated emoji images. This leads to the following hypothesis:

H2: *People's reaction time to respond to congruent emoji images is faster than their reaction time to respond to incongruent emoji images*

In addition, it was found that semantic context and frequency interact with each other, when asking people to fulfill a lexical decision task (Becker, 1979). This means that people need more semantic context, when responding to an unfamiliar and less frequently used stimulus than to a familiar stimulus. Since the analog presentation type is natural to the visual modality, people should be able to respond quickly to the analog stimuli. Linear image sequences, however, are less commonly used and people would probably need more semantic context to be able to respond quickly. In this regard, this research aims to find an interaction between semantics (congruent or incongruent emoji images) and presentation type (analog or linear). Hence, the following prediction was made:

H3: *There is an interaction between semantics (congruent/incongruent) and presentation type (analog/linear) of emoji sequences.*

4. Method

4.1 Study Design

In this 2 (semantics) x 2 (presentation type) experiment, it was measured if people understand a written utterance when it is expressed using an analog congruent or analog incongruent emoji image and when it is expressed with a linear congruent or linear incongruent emoji sequence. This was done according to the principles of crossmodal semantic priming, where people are, for instance, presented with a written stimulus preceded by a visual target, which may be semantically related (or unrelated) to the prime (Tabossi, 1996). In this experiment the prime was a written utterance like *The man eats pizza* and the visual target was the semantically matching (or semantically mismatching) analog emoji image or linear emoji sequence. Hence, the written modality and the visual modality were semantically primed, which makes this process crossmodal. Participants were asked to fulfill a lexical decision task, which is often used to indicate if words can be faster and more easily recognized than non-words (Lucas, 1999). Often these lexical decision tasks are combined with priming. Therefore, in this experiment, participants were asked to indicate, with a button-press, whether the presented visual target matches the written prime.

In order to determine how fast people were able to recognize and respond to the visual target and match it correctly to the prime, participants' response time and matching accuracy were recorded.

4.2 Participants

The experiment involved 156 participants (63 male, 87 female, 6 other). Participants' ages varied between 16 and 77 years, with an average of 29.3 years (*SD* = 8.85). The participants came from 19 different nationalities: 47.7 % of the respondents resided in the Netherlands, 27.6% came from Germany, and 7.1% lived in either the USA or the UK.

In order to measure emoji fluency and expertise among participants a generalized score was created based on a 7-point Likert scale. This Emoji Language Fluency Score (ELF-score) consisted of frequency of sending emoji (*M* = 5.7, *SD* = 1.09), the frequency of emoji-only texts (*M* = 3.81, *SD* = 1.33), emoji expertise (*M* = 4.8, *SD* = 1.06), emoji enjoyment (*M* = 5.4, *SD* = 1.02), and emoji efficiency (*M* = 3.03, *SD* = 1.42). The respondents were also asked to indicate how frequently they received messages involving emoji. However, these responses were not tracked correctly and had to be excluded from the data analysis. Next, the measurements were turned into a generalized percentage score. The average ELF-score of the respondents within this study was 0.446 (*SD* = 0.14), showing that overall participants' emoji language fluency scores were fairly high.

4.3 Stimuli

In order to test the three hypotheses, 20 written utterances were created, consisting of three compound words and 17 activity phrases (e.g. "the man eats pizza."). The experiment implied four conditions: (1) analog congruent, (2) linear congruent, (3) analog incongruent, and (4) linear incongruent. A Latin Square Design was used to counter balance the 2 (semantics) x 2 (presentation type) design applied in this experiment (see Figure 12).

The stimuli used in the condition 'linear congruent' were original emoji characters, which were not manipulated, but only composed in a linear sequence, like for instance: 🗑🔥 (dumpster fire). The images in the condition 'analog congruent', however, were manipulated in a way that the same common emoji icons appeared in an analog way and belong together as a consistently composed image.

In the conditions ‘analog incongruent’ and ‘linear incongruent’ the same images and sequences as created for the congruent conditions were used. In these cases, the images and sequences were simply presented together with a mismatching written prime. All manipulations were created with Photoshop (see Appendix).

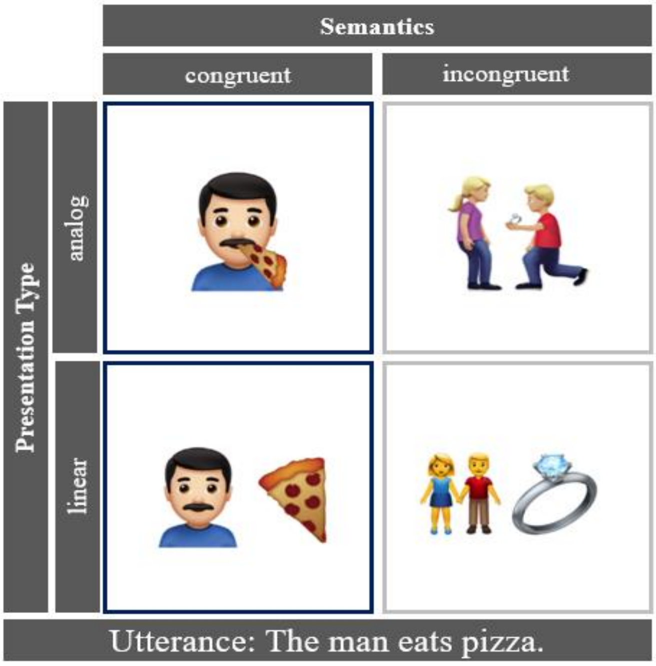


Figure 12. Example of the experimental layout

4.4 Procedure

The experiment was created with the survey-software Qualtrics and implied an introduction, a demographic section, several questions about participants’ emoji language fluency and the actual experiment. The decision task was programmed using lab.js, a JavaScript plugin allowing for the recording of response times of button presses.

First, participants were presented with a short introduction to explain what will be asked in the following sections. Furthermore, they were promised that the survey data will be collected in a completely confidential manner. In order to continue with the experiment, participants had to consent their participation.

Secondly, the survey implied four demographic questions regarding participants’ age, gender, current country of residence and additional language skills next to English.

Following, participants were asked to rate their level of emoji language fluency (ELF score). They were presented with seven questions regarding their frequency of emoji usage, their emoji expertise, and their operating system.

On the next screen, the experiment officially starts, and it was emphasized that participants are requested to use the “1-button” to indicate that the written utterance and the emoji match and to use the “0-button” to indicate that the written utterance and the presented emoji do not match semantically. Afterwards, they were presented with a random order of the 20 trials of text-emoji pairings. Written utterances were preceded by an emoji image on a next screen. This image was either analog congruent, linear congruent, analog incongruent or linear incongruent. As stated in the introduction, participants were asked to indicate whether the emoji image (analog or linear) matches the written utterance that was presented to them on the previous screen.

Each participant was presented with 20 written utterances and was therefore asked to match/mismatch 20 emoji images. The process in which utterances and image conditions were displayed to each individual participant was randomized. Yet, every participant saw images from each of the four conditions.

4.5 Data Analysis

In order for the data to be statistically valid, an outlier removal process on the basis of individual trials was conducted. Trials with an overly fast reaction time (below 300 milliseconds) were excluded, because it was impossible for the participants to have seen the emoji properly. Next to this, an upper threshold of 5844 milliseconds was identified, which was calculated as 2.5 times the Standard Deviation of the mean above the mean. Afterwards, all trials above that threshold were excluded. Next, the measured reactions times per participant were averaged across emoji items, generating an average reaction time and an average accuracy score for each of the four conditions (analog congruent, analog incongruent, linear congruent, linear incongruent).

To test the effect of presentation type and semantics on reaction time a 2 x 2 repeated-measures ANOVA was performed. Another 2 x 2 repeated-measures ANOVA was conducted to examine the effect of presentation type and semantics on response accuracy. Responses to congruent images were identified as accurate, when the written prime and the visual target were correctly recognized as matching. In the incongruent conditions, on the other hand, responses were considered as accurate, when the participants correctly detected that the written prime and the visual target were mismatching. To confirm that people were not responding to the stimuli at random and that respondents gave carefully considered answers, a one sample t-test against the frequency rate of .5 was implemented. Lastly, a correlation analysis was performed to examine if participants' overall emoji language fluency (ELF) scores relate to their average reaction times and response accuracy. Since this study focused primarily on emoji-only expressions, another correlation analysis was conducted to investigate if the ELF score's subcomponent "expertise of emoji-only sequences" specifically correlates with people's reaction times and response accuracy.

5. Results

5.1 Effect of Presentation Type and Semantics on Reaction Time

To test whether people's reaction time to respond to analog emoji images is faster than their reaction time to respond to linear emoji sequences, a 2 x 2 repeated-measures ANOVA was performed. There was a main effect for presentation type (F (1, 155) = 46.62, $p < .001$), indicating that processing of analog images is faster than processing of linear images. Hence, H1 is supported by the data. A main effect for semantics (F (1,155) = 23.26, $p < .001$) arose because reaction times to respond to congruent emoji were slower than reaction times to respond to incongruent emoji. Hence, H2 is not supported by the data, which showed the exact opposite result as predicted. Finally, an interaction between presentation type and semantics (F (1, 155) = 37.3, $p < .001$) arose because reaction times to linear congruent emoji sequences were significantly slower than reaction times to analog congruent, analog incongruent and linear incongruent emoji. Hence, H3 is supported by the data.

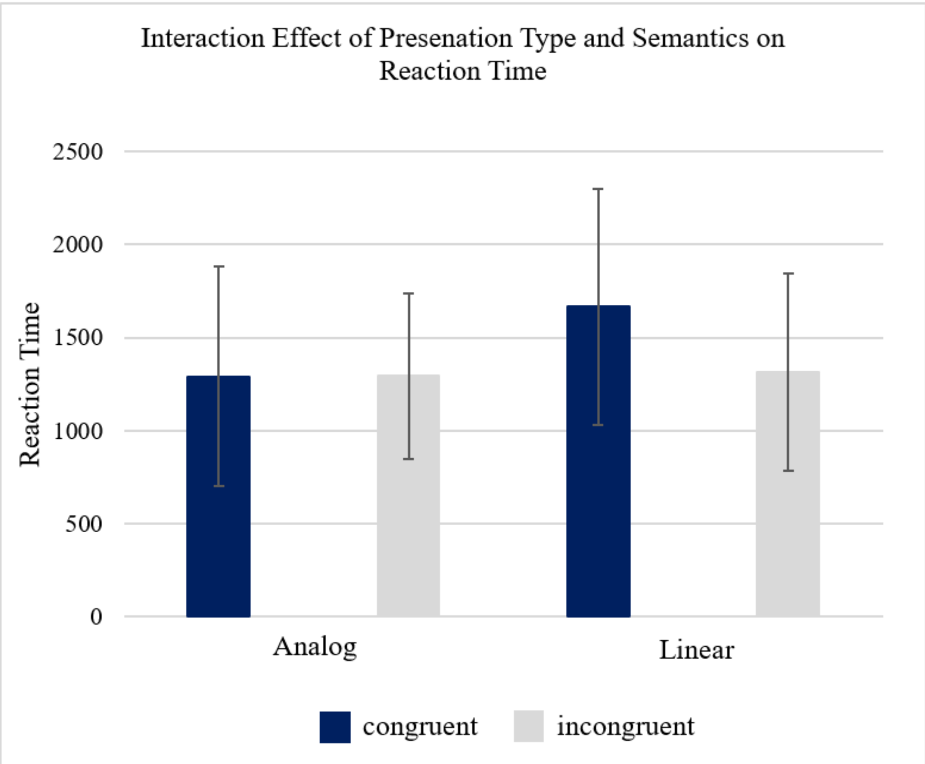


Figure 13. Reaction times for emoji following text across presentation type (analog, linear) and semantics (conruent, incongruent)

5.2 Effect of Presentation Type and Semantics on Response Accuracy

A one-sample t-test was conducted to test whether the chance to randomly match or mismatch images across all conditions is higher than 50%. The analysis showed that the level of accuracy for all four conditions was higher than chance: analog congruent (*Mdif* = .47, $t(155) = 63.34$, $p < .001$), linear congruent (*Mdif* = .36, $t(155) = 19.48$, $p < .001$), analog incongruent (*Mdif* = .48, $t(155) = 94.48$, $p < .001$), linear incongruent (*Mdif* = .48, $t(155) = 104.2$, $p < .001$). This confirms that people were not pressing at random and that respondents indeed gave carefully considered answers.

To test whether people's responses to analog emoji images are more accurate than their responses to linear emoji sequences, a 2 x 2 repeated-measures ANOVA was performed. A main effect for presentation type (F (1, 155) = 27.28, $p < .001$) arose because responses to analog images were more accurate than responses to linear images. A main effect for semantics (F (1,155) = 41.88, $p < .001$) indicated that responses to congruent emoji were less accurate than responses to incongruent emoji. An interaction (F (1,155) = 28.01, $p < .001$) between presentation type and semantics arose because responses to incongruence were significantly more accurate for both analog and linear images, and semantics only influenced response accuracy for linear congruent images.

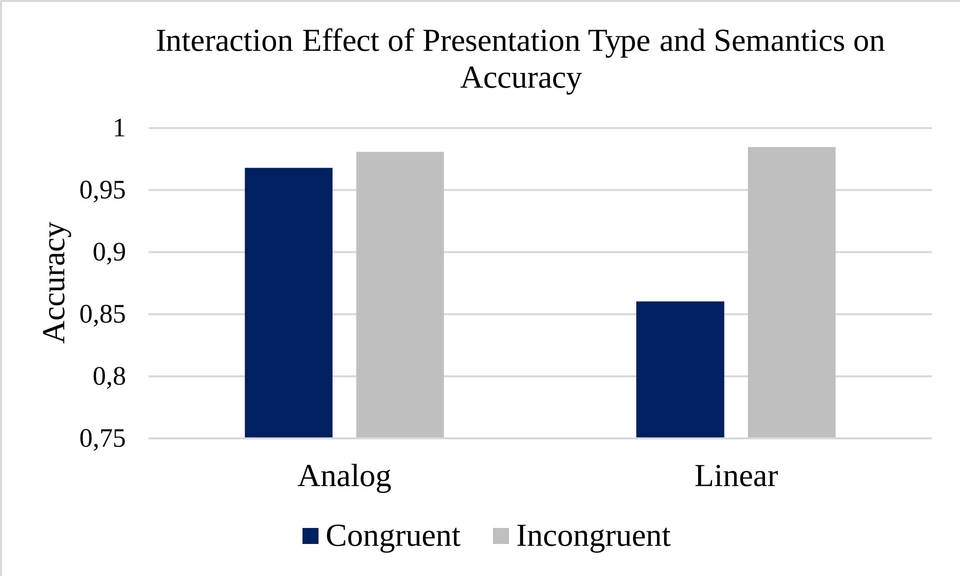


Figure 14. Recognition of a matching relationship between emoji and preceding text across presentation type (analog, linear) and semantics (congruent, incongruent)

5.3 Effect of individual Differences on Reaction Time and Response Accuracy

The analysis showed negative correlations between respondents' emoji language fluency (ELF) scores and their reaction time to analog congruent (r = -.17, p = .036) and linear congruent (r = -.17, p = .034) emoji images, indicating that the higher people's ELF scores, the shorter their reaction time to respond to congruent emoji. No correlation arose between ELF scores and analog incongruent (r = -.11, p = .19) and linear incongruent (r = -.03, p = .73) emoji. The analysis also showed no correlations between people's ELF scores and their response accuracy.

Additionally, special attention was paid to participants expertise in using "emoji-only sequences", which was a subcomponent of the overall ELF score. Since this experiment focused specifically on emoji-only expressions, the "emoji-only" subcomponent of the ELF score targets the particular manipulations employed in this study in an interesting way. Hence, when testing solely people's expertise of emoji-only sequences, the analysis showed a correlation between people's emoji-only scores and analog congruent (r = -.20, p = .01), as well as linear congruent (r = -.17, p = .032) emoji. This correlation indicates that the more frequent people send emoji-only messages, the faster their reaction times to respond to congruent emoji images. No correlation was found between people's emoji-only expertise and analog incongruent (r = -.11, p = .16) and linear incongruent (r = -.11, p = .18) images.

6. Discussion

This research investigated to what extent processing of emoji expressions is limited by their artificial linearity. The main findings of the experiment revealed that linearity slows people's reaction time to decisions about whether emoji reflect the meaning given in prior text, and this linearity causes significantly more inaccurate answers. Hence, participants made faster and more accurate responses to analog emoji images than to linear emoji sequences. In addition, the analysis also showed an interaction between semantics and presentation type of the emoji sequences.

The experiment supported the assumption that presentation type influences how fast emoji can be processed. Emoji must act like written text and are therefore forced into artificial linear sequences (Cohn et al., 2018), affecting the cognition process of these images (Cohn, 2013). The modality of written language puts different requests on its expressive system than the visual modality, which required respondents to make adjustments in how emoji-only expression could be processed (Cohn, 2013). The difference in combinational properties between the analog and linear image condition had a strong impact on how people were able to deal with the presented stimuli (Cohn, 2012a; Cohn, 2015). In this experiment, the analog condition showed emoji images combining two emoji characters in an analog spatial display, which resembles the nature of visuals, while in the linear condition emoji were displayed next to each other. Hence, analog images could be processed as one whole unit (Cohn, 2013), whereas linear sequences required respondents to first process one image and then the other image of the sequence, which demands more time. Furthermore, respondents might have had a mental picture in mind, when they were presented with the written utterance (Shepard, 1978), which more resembled the analog images than the linear sequences. This again might have facilitated people's response process and accelerated their reaction time.

Despite these limitations of affordances, emoji are still a very efficient way to humanize and enhance our written texts (Duerscheid & Siever, 2017; Hakami, 2017). While computer-mediated communication lacked the ability to transmit non-verbal cues in the past (Kelly & Watts, 2015), emoji offer new ways to control the emotional valence of textual communication (Walther & D'Addario, 2001). Hence, the transformation from speech into text forced our text-based communication to adjust their modality-specific affordances. By involving visuals like emoji into our written texts, CMC was faced with the need to adjust modality-related characteristics. The affordances of written texts are transferred to the visual modality, forcing emoji images to act like text. This transfer, in return, obliges emoji to lose some of their visual characteristics, like their combinational flexibility. This means that when we want to adapt the characteristics of one modality to make it act according to the affordances of another modality, several affordances might get lost (Cohn, 2013). A similar limitation of affordance transfer between modalities happened in the example of sign language and MCE (Supalla & McKee, 1991). When trying to adapt the characteristics of sign language to make it behave according to the affordances of speech, the affordance of simultaneity in the signed modality could not be transferred and got lost (Supalla, 1986; Supalla & McKee, 1991).

Furthermore, the experiment showed a significant effect of semantics on people's reaction time. This effect of context would also be supported by former studies, which discovered a strong difference in reaction time when comparing congruent stimuli to incongruent stimuli (Bentin et al., 1984; Meyer & Schvaneveldt, 1971). However, the effect found in this research is exactly opponent to that in prior experiments. While previous studies revealed that people respond significantly slower to incongruent stimuli than to congruent stimuli, the current study found faster reaction times for incongruence than for congruence. Yet, most of the former studies compared differences in reaction times between the recognition of words and non-words, or words preceded by semantically related or unrelated words (e.g., Becker 1979, 1980; Meyer & Schvaneveldt, 1971). This study, however, compared a written prime followed by a semantically related or semantically unrelated visual target, meaning that the lexical decision task had to be made crossmodally. Moreover, the scores for accuracy in both incongruent conditions were also very high, showing that people correctly recognized that the written prime and the visual target were mismatching to a high proportion. Interestingly, respondents were able to correctly detect incongruence that much faster in both presentation types (analog and linear) that this might be evidence that the incongruent stimuli were very obvious. For instance, the written utterance *dumpster fire* was preceded by either an analog or linear image of an *apple tree* in the incongruent conditions. Since *dumpster fire* and *apple tree* are strongly semantically unrelated, people were able to detect the incongruence immediately. As people's mental image after being presented with the written prime probably did not match the incongruent image at all (Shepard, 1978), the mismatch was fairly easy to detect.

Moreover, the experiment found an interaction between semantics and presentation type because reaction times to linear congruent emoji sequences were significantly slower than reaction times to analog congruent, analog incongruent and linear incongruent emoji. Hence, people were able to quickly process incongruence in both presentation types and had no difficulties to quickly understand analog emoji images, whereas linear congruent emoji sequences were significantly harder to process. This supports the argumentation that forcing emoji into artificial linear sequences impinges on people's ability to process these visuals. In addition, people are not familiar with visuals appearing in a strict agent-action pattern (Cohn et al., 2018) and without any combinational flexibility (Cohn, 2012a), which impedes the process of understanding. Similarly, a study by Becker (1979) showed that semantic context and word frequency do interact during a lexical decision task. The effect of context is larger for low-frequency words than for high-frequency words, which means that in order to respond quickly, people require more context when the stimulus (word) is less familiar, than when the stimulus (word) is frequently used. This effect might be relatable to the current study. Since people are generally less familiar with linear images, because they are unnatural for the visual modality, more semantic context is required to be able to respond to the lexical decision task. Hence, the cognition process is more demanding, requiring more time to make the contextual connection to the word prime. Analog images, however, are more familiar and natural to the respondents, and therefore do not need as much semantic context. Nevertheless, this would not explain why people in general react significantly faster to congruence than to incongruence in the linear condition. However, this might again be related to the obviousness of the stimuli material.

This obviousness of the used emoji images might also support why people were generally very accurate in responding to the lexical decision task. Former studies found that words followed by a semantically related word were identified more accurately than words followed by semantically unrelated words or nonwords (Bentin et al., 1984). Interestingly, in regard to these results it would be expected that identifying incongruent emoji images entails more errors. However, in this study, respondents on average correctly identified incongruent stimuli as mismatching in almost 100% of their responses, generating very high accuracy scores for the incongruent conditions. Moreover, there were also on average almost no inaccurate responses made for analog congruent emoji, supporting that displaying visuals as one whole unit (Cohn, 2012a) helps to quickly and accurately understand the meaning. Slightly more incorrect responses occurred for linear congruent images, generating a lower average accuracy score than for the other three conditions. Although the written prime and the linear congruent emoji were actually semantically matching, respondents indicated several emoji sequences to have no semantic relation to the written prime. This supports the argumentation that people have generally more difficulties in processing and understanding visual linear sequences compared to analog images. While the data already supported H1, showing that processing of linear images takes significantly longer than processing analog images, the accuracy rates suggest that it is also significantly harder to understand the meaning of these linear sequences. This might be due to the grammatical limitations (Cohn et al., 2018), the strict agent-action pattern (Cohn, 2015), the fact that all emoji are displayed in the same size (Danesi, 2016), or the lack of combinational flexibility (Cohn, 2015) of emoji-only sequences. All these factors are against the nature of the visual modality and therefore impede the understanding process of linear emoji sequences.

Finally, the experiment demonstrated that people's emoji language fluency scores do correlate with their ability to quickly respond to the lexical decision task. This means that the higher people ranked their own emoji skills, the faster they were able to respond to analog and linear congruent emoji. Since participants were familiar with the emoji characters used in both conditions (analog congruent and linear congruent), less time was required to process the meaning (Becker, 1979). However, no correlation was found for the incongruent conditions, since on average respondents made fairly quick responses to all incongruent stimuli. This shows that incongruence could easily be detected by people with high ELF scores, as well as by people with low ELF scores. Similar results were found when looking solely on respondent's expertise in using emoji-only sequences. People with a more frequent use of emoji-only expressions responded faster to analog and linear congruent emoji than people with a less frequent use of emoji-only sequences. This suggests that people commonly using emoji-only sequences, might have generated a better understanding of linear emoji sequences. They might be more familiar with repeatedly occurring limitations of emoji due to the strict agent-action pattern (Cohn, 2015; Cohn et al., 2018) or the fact that these sequences often resemble a picture puzzle (Danesi, 2016), and might have slightly adapted their way of processing these sequences. According to Becker (1979) being familiar with the stimuli requires less context and impedes the process of understanding. Lastly, ELF scores did not correlate with people’s rates of response accuracy because overall almost no matching errors were made during the experiment.

6.1 Limitations and suggestion for future research

This research had two main limitations. The first constraint is that incongruent stimuli were too easy to detect and therefore simplified the match/mismatching process. Since the congruent and the incongruent stimuli did not relate to each other at all, respondents had no difficulties ascertaining incongruence and were able to respond to mismatching images immediately. Future studies might therefore use incongruent stimuli that are less obviously unrelated to the written utterance. For instance, the word *dumpster fire* () could be followed by either an analog or a linear image of *car fire* (). Hence, the *fire* emoji () would be used in the images of both, the congruent and the incongruent condition, which makes the mismatch less apparent.

Furthermore, the written utterances used in this study consisted of short sentences, as well as of compound words and therefore tested these two different types of written primes in all four conditions. However, literature showed that the difficulty in expressing sentences with only emoji lays within the strict agent-action pattern (Cohn, 2015; Cohn et al. 2018), whereas the difficulty of expressing compound words with solely emoji might be more related to the fact that it resembles a picture puzzle (Danesi, 2016). When looking deeper into the current stimuli, people's reaction times to match "sentence primes" to congruent analog (M = 1592.34) and congruent linear (M = 1999.60) emoji were on average slower than their reaction times to match "compound word primes" to either congruent analog (M = 1221.05) or congruent linear (M = 1682.75) emoji. This would give an indication that expressing a sentence with linear emoji-only sequences is even harder to process than expressing a compound word in the same way. However, this experiment only employed three compound word primes and 17 sentence primes, meaning that this assumption should be further examined in future research. Thus, future studies might be interested in testing compound words and sentences separately from each other to investigate if these two types of written primes have different impacts on and interactions with presentation type and semantics.

7. Conclusion

To conclude, this research supported the assumption that the artificial linearity of emoji-only expressions limits people's ability to respond to emoji images. Linearity influences the response duration, as well as the response accuracy and affects how emoji expressions are processed and understood. Thus, forcing one modality to behave in ways that are only natural to another modality has an impact on processing. Pushing emoji away from the affordances of the visual modality to instead make them act like text results in constraints, which in return impair the cognition process.

However, emoji are still an omnipresent part of our daily text-based conversations and opened new ways to integrate more emotions into our texts. Without them computer-mediated communication would be restricted to only texts with less options for emotional complexity. Emoji are very useful for enhancing and humanizing digital conversations

(Hakami, 2017), giving the possibility of involving humor, affection and even sarcasm into short messages (Cohn, 2015). Since different types of communication are called a language (spoken, written, visual or sign language), emoji communication is proclaimed as an emerging language as well (Danesi, 2016; Lu et al., 2016). Nevertheless, emoji possess less richness and complexity as either the written or visual language, limiting them in making natural expressions (Cohn, 2015). The linearity of emoji-only sequences causes many constraints on a grammatical, sequential and visual level, which make these expressions less flexible and elaborate. Thus, the proclamation of emoji becoming an emerging language is very doubtful, even if it is a useful communicative system.

Graphic References

Godek, T. 2006. One Night. Originally posted on March 20, 2006. Retrieved from: http://www.yellowlight.scratchspace.net/comics/onenight/onenight.html

References

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of experimental Psychology: Human Perception and Performance, 5*(2), 252–259. doi:10.1037/0096-1523.5.2.252

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8*(6), 493–512. doi:10.3758/bf03213769

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343–355. doi:10.1016/0013-4694(85)90008-2

Chan, A., Lieven, E., & Tomasello, M. (2009). Children's understanding of the agent-patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English. *Cognitive Linguistics, 20*(2). doi:10.1515/cogl.2009.015

Cohn, N. (2012a). Explaining 'I Can't Draw': Parallels between the structure and development of language and drawing. *Human Development, 55*(4), 167–192. doi:10.1159/000341842doi:10.1145/3178876.3186157

Cohn, N. (2012b). Visual Narrative Structure. *Cognitive Science, 37*(3), 413–452. doi:10.1111/cogs.12016

Cohn, N. (2013). *The visual language of comics: Introduction on the structure and cognition of sequential images.* London, UK: Bloomsbury.

Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia, 64*, 63-70. doi:10.1016/j.neuropsychologia.2014.09.018

Cohn, N. (2015). *Will emoji become a new language?* Retrieved from: http://www.bbc.com/future/story/20151012-will-emoji-become-a-new-language

Cohn, N., Engelen, J., & Schilperoord, J. (2018). *The grammar of emoji? Patterning in communicative pictorial sequencing.* Tilburg Center of Cognition and Communication (TiCC).

Cox, M. V. (1998). Drawings of people by Australian Aboriginal children: The inter-mixing of cultural styles. *Journal of Art & Design Education*, *17*(1), 71–79. doi:10.1111/1468-5949.00107

Danesi, M. (2016). *The semiotics of emoji: The rise of visual language in the age of the internet*. Bloomsbury Publishing.

Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, *20*(3), 249–268. doi:10.1111/j.1468-2885.2010.01362.x

Duerscheid, C., & Siever, C. M. (2017). Beyond the alphabet-communication of emojis. *Kurzfassung eines (auf Deutsch) zur Publikation eingereichten Manuskripts*.

Emojipedia. (2018). Retrieved from: http://www.emojipedia.org.

Friederici, A. D., & Meyer, M. (2004). *The brain knows the difference: two types of grammatical violations. Brain Research, 1000(1-2), 72–77.* doi:10.1016/j.brainres.2003.10.057

Gershkoff-Stowe, L., & Goldin-Medow, S. (2002). Is there a natural order for expressing semantic relations? *Cognitive Psychology*, *45*(3), 375–412. doi:10.1016/s0010-0285(02)00502-9

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429. doi:10.1016/s1364-6613(99)01397-2

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in cognitive sciences*, *3*(11), 419-429.

Golomb, C. (1992). *The child's creation of a pictorial world. Berkeley:* University of California Press.

Hakami, S.A.A (2017). *The importance of understanding emoji: An investigative study*. University of Birmingham.

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*(4), 379–411. doi:10.1080/01690969308407583

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281–312. doi:10.1080/01690969008407065

Jackendoff, R., & Wittenberg, E. (2014). What you can say without syntax: A hierarchy of grammatical complexity. In F. Newmeyer & L. Preston (Eds.), *Measuring Linguistic Complexity* (pp. 65-82). Oxford:

Kelly, R., & Watts, L. (2015). Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*, 1-5. Opus: University of Bath Online Publication Store, Retrieved from: http://opus.bath.ac.uk/46780/1/emoji_relational_value.pdf

Kircher, T., Sass, K., Sachs, O., & Krach, S. (2009). Priming words with pictures: Neural correlates of semantic associations in a cross-modal priming task using fMRI. *Human Brain Mapping, 30*(12), 4116–4128. doi:10.1002/hbm.20833

Kistler, M. (1988). *Mark Kistler's draw squad*. New York: Simon and Schuster Inc.

Klein, W. (1985). Schriftlichkeit: Gesprochene Sprache - geschriebene Sprache. *Zeitschrift für Literaturwissenschaft und Linguistik, 59*, 9-35.

Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., & Mei, Q. (2016). *Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users.* Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Heidelberg, Germany: ACM.

Lucas, M. (1999). Context effects in lexical access: A meta-analysis. *Memory & Cognition, 27*(3), 385–398. doi:10.3758/bf03211535

Matthews, J. (1983). *Children drawing: Are young children really scribbling?* Paper presented at the British Psychological Society International Conference on Psychology and the Arts, Cardiff.

Mesulam, M. (1998). From sensation to cognition. *Brain 121*(6), 1013–1052. doi: http://dx.doi.org/10.1093/brain/121.6.1013

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227–234. doi:10.1037/h0031564

NRZ (2015). *Neue Emojis nach iOS-Update mit verschiedenen Hautfarben.* Retrieved from: https://www.nrz.de/leben/digital/neue-emojis-nach-ios-update-mit-verschiedenen-hautfarben-id10542554.html

McCulloch, G., & Gawne, L. (2018). Emoji grammar as beat gestures. In s. Wijeratne, E. Kiciman, H. Saggion, & S.A. (Eds.), *Proceedings of the First International Workshop on Emoji Understanding and Applications in Social Media* (Emoji2018) (pp.1-4). Stanford, CA.

Pavalanathan, U., & Eisenstein, J. (2016). More emojis, less:) The competition for paralinguistic function in microblog writing. *First Monday, 21*(11). doi: https://doi.org/10.5210/fm.v21i11.6879

Puranik, C. S., & Lonigan, C. J. (2009). From scribbles to scrabble: preschool children's developing knowledge of written language. *Reading and Writing*, *24*(5), 567–589. doi:10.1007/s11145-009-9220-8

Riva, G. (2002). The sociocognitive psychology of computer-mediated communication: The present and future of technology-based interactions. *CyberPsychology & Behavior, 5*(6), 581–598. doi:10.1089/109493102321018222

Schnoebelen, T. (2012). Do you smile with your nose? Stylistic variation in twitter emoticons. *University of Pennsylvania Working Papers in Linguistics*, *18*(2): 14.

Shepard, R. N. (1978). The mental image. *American Psychologist*, *33*(2), 125–137. doi:10.1037/0003-066x.33.2.125

Supalla, T. (1986). The classifier system in American sign language. *Noun classes and categorization, 7*,, 181-214. doi: https://doi.org/10.1075/tsl.7.13sup

Supalla, S. & McKee, C. (1991). Manually coded English: The modality question in signed language development. In S. D. Fischer & P. Siple (Eds.), *Theoretical issues in sign language research, 2*, 85-109. Chicago: University of Chicago Press.

Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes, 11*(6), 569–576. doi:10.1080/016909696386953

Tolchinsky, L. (2003). *The cradle of culture and what children know about writing and numbers before being taught*. Mahwah, NJ: Lawrence Erlbaum Associates.

Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review*, *19*(3), 324-347. doi:10.1177/089443930101900307

Wilkins, D.P. (1997). Alternative representations of space: Arrernte narratives in sand. In M. Biemans & J. van de Weijer (Eds.), *Proceedings of the CLS Opening Academic Year '97 '98* (pp. 133–164). Nijmegen: Nijmegen/Tilburg Center for Language Studies.

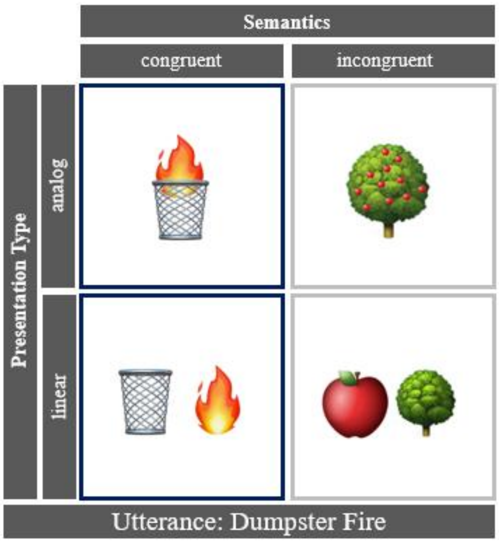
Wilson, B. (1988). The artistic tower of babel: Inextricable links between culture and graphic development. In G.W. Hardiman & T. Zernich (Eds.), *Discerning art: Concepts and issues.* Champaign: Stipes Publishing Company.

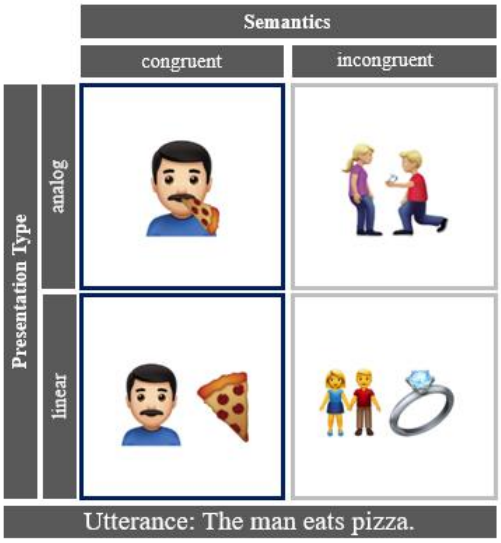
Wilson, B., & Wilson, M. (1977). An iconoclastic view of the imagery sources in the drawings of young people. *Art Education*, *30*(1), 4. doi:10.2307/3192209

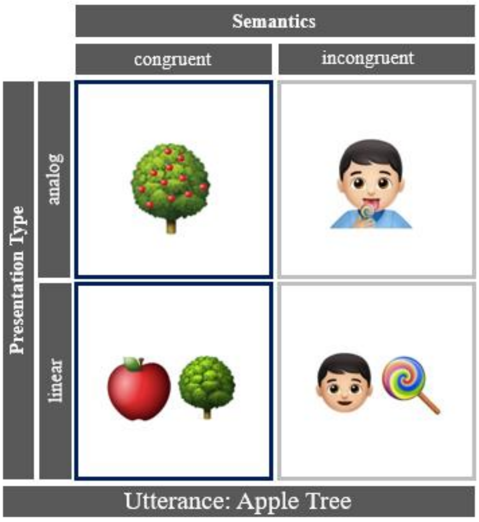
Wordcounter.net (2018). *How many letters are there in the alphabet?* Retrieved from: https://wordcounter.net/blog/2015/11/24/10950_how-many-letters-alphabet.html

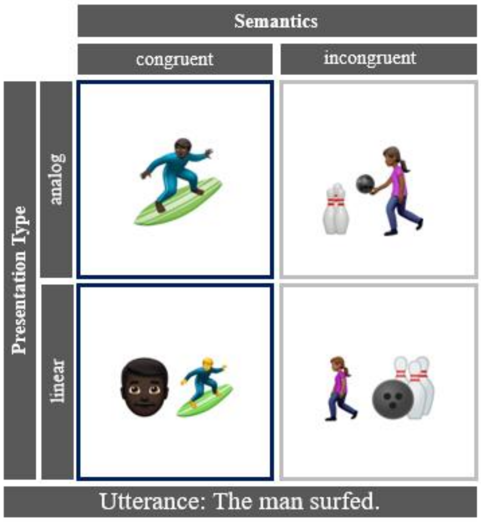
Appendix

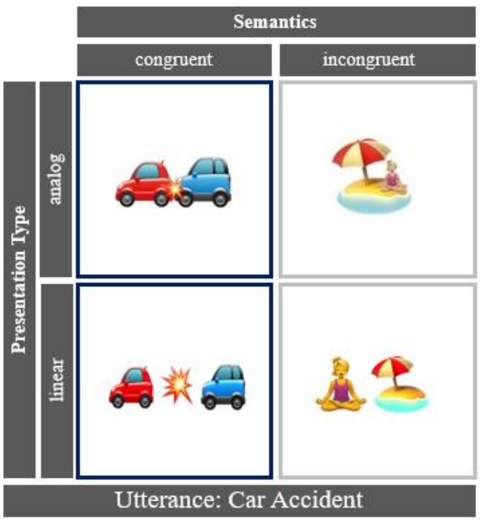
Overview of stimuli used in the experiment

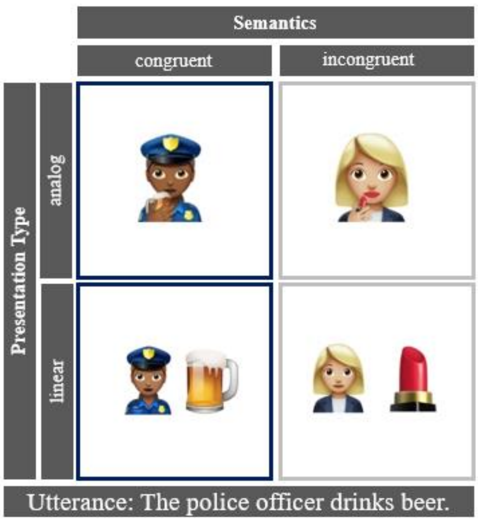


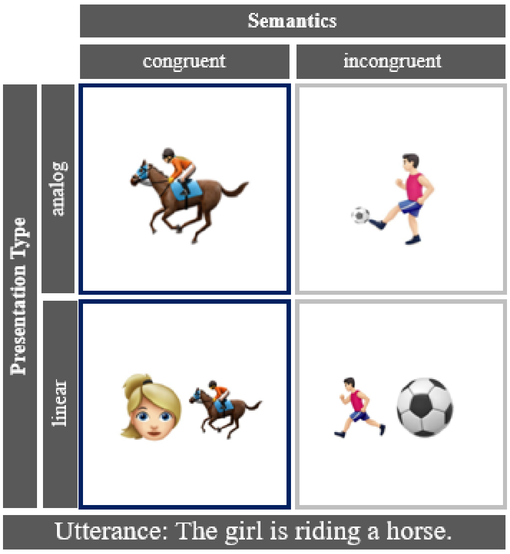


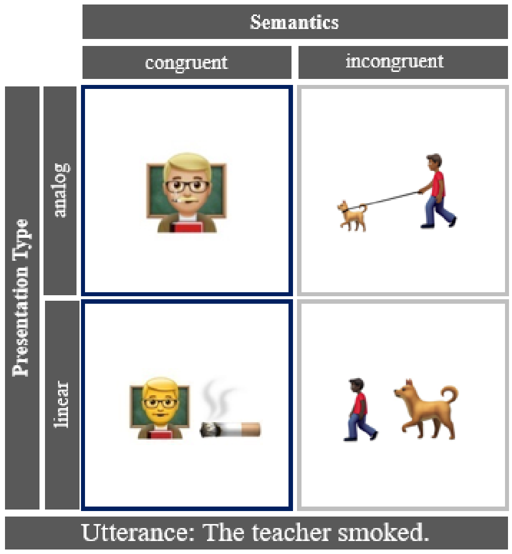


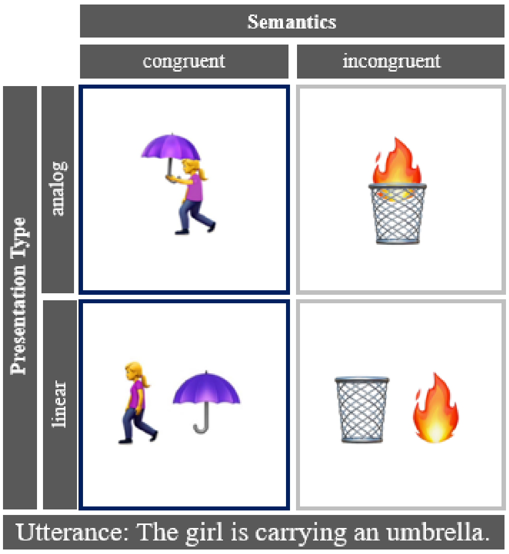


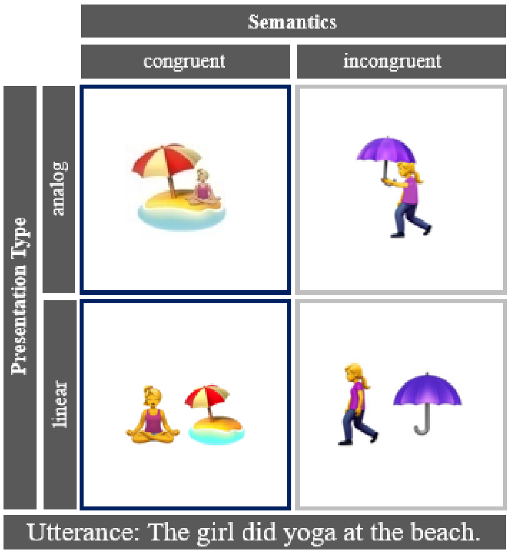


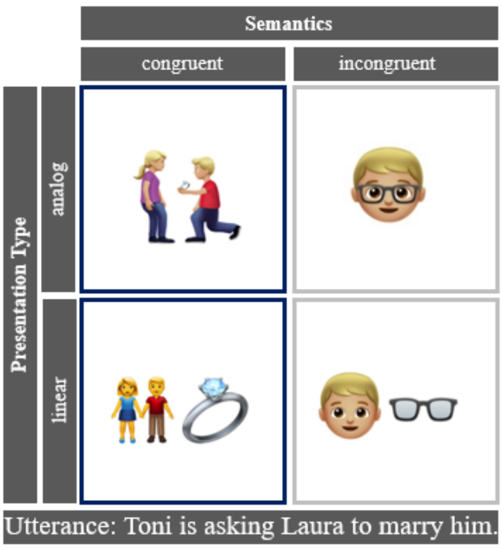


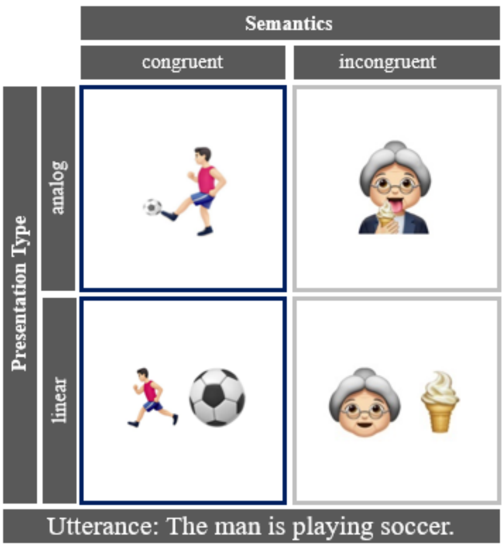

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The girl is riding a horse.

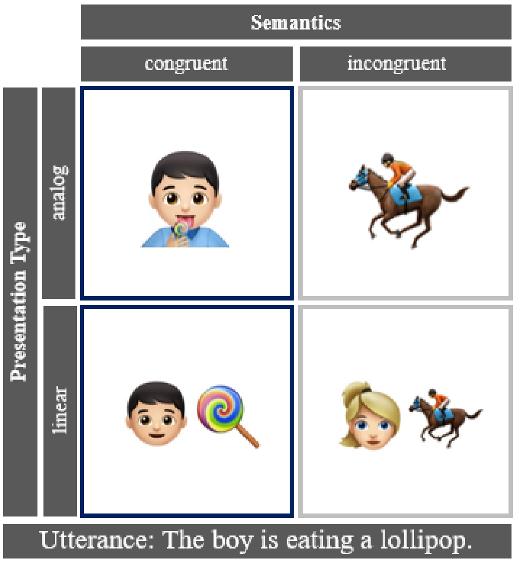

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The teacher smoked.

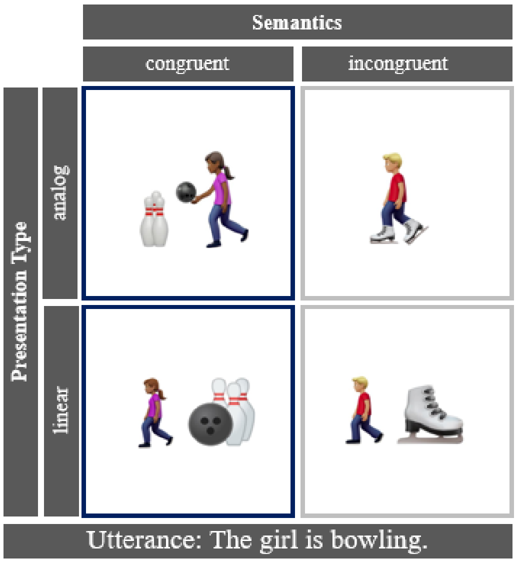

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The girl is carrying an umbrella.

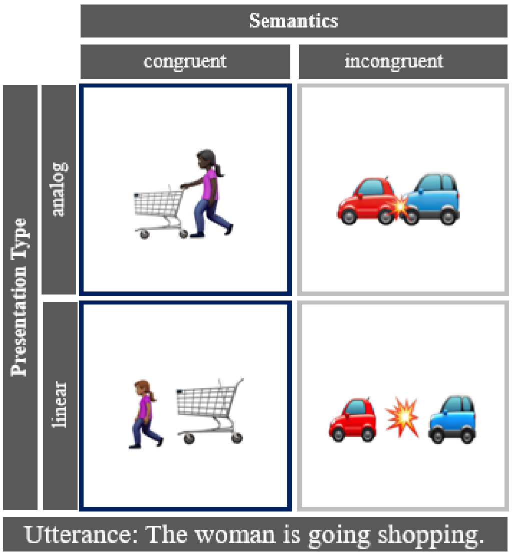

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The girl did yoga at the beach.

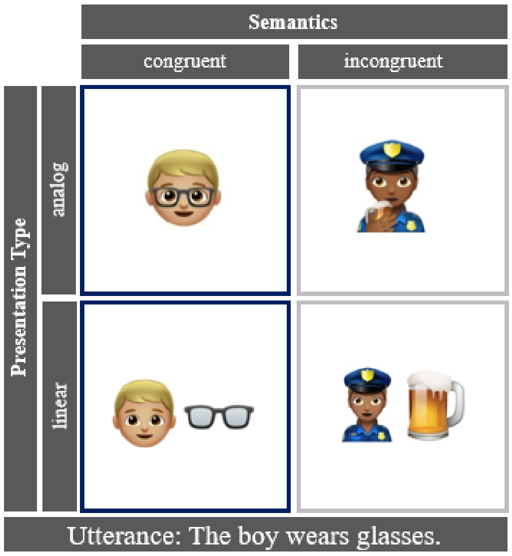

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: Toni is asking Laura to marry him.

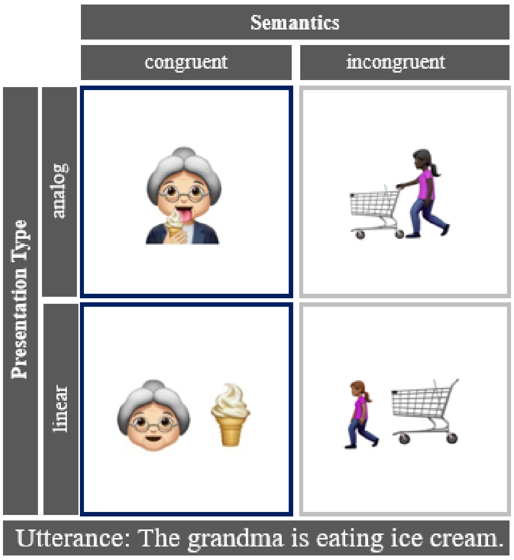

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The man is playing soccer.

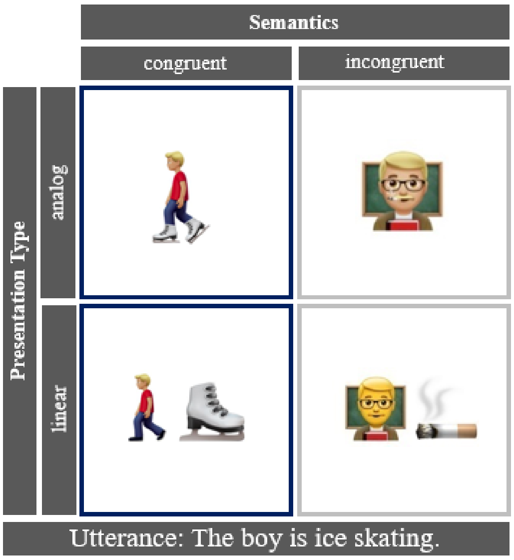

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The boy is eating a lollipop.

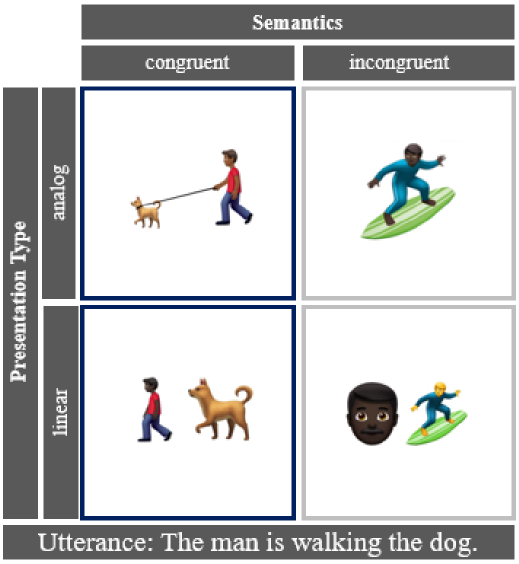

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The girl is bowling.

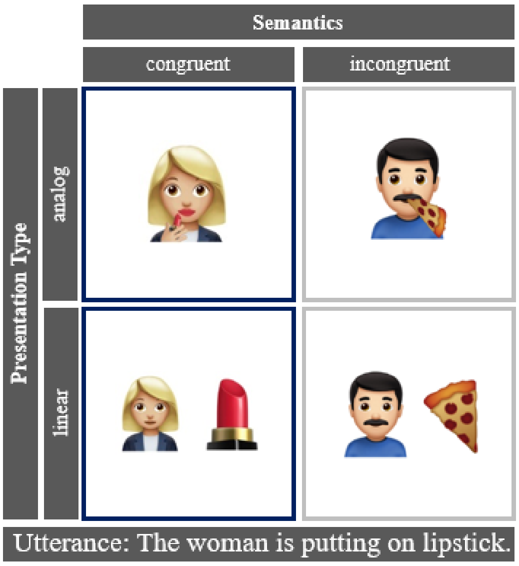

Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The woman is going shopping.


Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The boy wears glasses.


Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The grandma is eating ice cream.


Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The boy is ice skating.


Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The man is walking the dog.


Semantics
congruent
incongruent
Presentation Type
analog
linear
Utterance: The woman is putting on lipstick.