COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13099

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13099

Are Emoji Processed Like Words? An Eye-Tracking Study

Patrizia Paggio,[a,b] Alice Ping Ping Tse[a]

[a]*Department of Nordic Studies and Linguistics, Centre for Language Technology, University of Copenhagen*
[b]*Institute of Linguistics and Language Technology, University of Malta*

Received 30 July 2020; received in revised form 12 July 2021; accepted 5 January 2022

Abstract

In this study, we investigate the processing of object-denoting emoji in sentences using eye tracking. We hypothesize that (a) such emoji are more difficult to process when used as word replacement; and (b) their processing is subject to ambiguity constraints similarly to what happens with words. We conduct two experiments in which participants have to read sentences in which an emoji either follows or replaces a word. Control stimuli not containing emoji are also tested. In the second experiment, the emoji are presented in two different disambiguating contexts. First fixation duration, total visit duration, and total revisit duration are modeled in the various conditions using linear mixed models. Both our hypotheses are supported. We observe longer total visit time for non-redundant emoji, and higher values for all three measures for ambiguous emoji. We conclude that lexical access may be more difficult for emoji especially when they are used as word replacements and their meaning is not immediately clear. Furthermore, we also conclude that non-redundant emoji are more difficult to integrate in the processing of the sentence than the equivalent words, or emoji used in a redundant way. In turn, our results indicate that emoji may not always be as immediate and easy to process for readers in spite of their popularity.

Keywords: Ambiguity; Emoji processing; Eye tracking; Lexical access

Correspondence should be sent to Patrizia Paggio, University of Copenhagen, Department of Nordic Studies and Linguistics, Centre for Language Technology, Njalsgade 136, Copenhagen S, 2300, Denmark. E-mail: paggio@hum.ku.dk

1. Introduction

Emoji have acquired great popularity as a means of expression in online messaging. Emoticons were an initial attempt to reproduce facial images in text, thereby allowing users to add the equivalent of non-verbal expressions of emotion to online text (Walther & D'Addario, 2001). Emoji have since developed into a relatively rich inventory of pictographs, including not only the original smileys, but also pictures of many different categories, from animals to objects of several kinds. Emoji have been described as ubiquitous signs that can be understood across languages and cultures (Azuma & Ebner, 2008). However, it has also been shown that there are differences in the way linguistic communities use them, both as regards frequencies of usage of different emoji or emoji category, and the meaning users assign to them. For example, the large empirical study conducted in Lu et al. (2016) found that the frequency distribution of emoji is influenced by linguistic, cultural, and geographical properties. Moreover, the results presented in Miller et al. (2016) on the basis of a survey indicate variability in sentiment as well as semantic interpretation of 25 popular emoji not only across different platforms (such as Windows vs. Apple), where the same emoji have different renderings, but also within the same platform. In turn, such differences in the way users interpret emoji may result in misconstruals and misunderstandings (Tigwell & Flatla, 2016). As much as 40% of the emoji considered by Miller and colleagues have a high misconstrual score, and very few emoji are described in the same way. A highly ambiguous example in the study is Apple's rendering of "unamused face" 😒 for which participants offered different interpretations, such as *disappointment*, *depressing*, *unimpressed*, and *suspicious* (Miller et al., 2016, p. 264).

It should be noted in this respect that the meaning assigned to an emoji by a user may not necessarily be the same as the definition given by the Unicode standard.[1] In fact, most users will probably never encounter such a definition, and will construe the meaning of emoji bottom-up based on graphical rendering and usage in a variety of contexts. This process may result in private meanings shared in microcommunities of family and friends and so-called *repurposing* (Wiseman & Gould, 2018), that is, the development of a different meaning than the one originally intended by the Unicode consortium. The semantics of emoji has been explored by applying word embedding models (Barbieri, Ronzano, & Saggion, 2016). The content similarity between different emoji, and between emoji and words, is investigated by comparing their vectorial representations built over large text corpora. These methods show that the meaning of at least some emoji changes depending on language and culture. For instance, the study in Barbieri, Kruszewski, Ronzano, and Saggion (2016) mentions "weary face," which is used in different contexts in American and British English, and emoji related to the beach and good weather, which are used in similar contexts in Italian and Spanish, but not in British English. The meaning of emoji can also change over time. In Barbieri, Marujo, Karuturi, Brendel, and Saggion (2018), it is noted that the "pine" emoji co-occurs with emoji related to camping and other outdoors activities in spring and summer, while in autumn and winter, it is associated with Christmas.

An interesting question is whether emoji should be thought of as words. It is often mentioned that in 2015, Oxford dictionaries declared the emoji "face with tears of joy" the word of the year.[2] However, strictly speaking emoji are characters, corresponding to a Unicode symbol, while emoji renderings are pictographs. For example, "face with tears of joy" has the Unicode symbol U+1F602, and can be rendered in different ways depending on the platform, for example, as 😂 on Apple products.

Semantically, they can correspond to the meaning of a word 🍕 or to a complex concept that is best expressed as a sequence of words 😂. From the point of view of the way they are used in texting, they often accompany words, thereby either providing a pictorial representation of the concept expressed by the word or adding an emotional comment to it, but they can also be used as word replacements (Donato & Paggio, 2017). Three such cases extracted from Twitter as part of the study in Donato and Paggio (2018) are shown below. In (1), the emoji is redundant and illustrates the word *home*; in (2), it adds the expression of an emotion; and in (3), it replaces the word *car* and is therefore clearly non-redundant.

1. I love my home 🏠
2. I can't 😢
3. let me go get my 🚗

The same authors also analyzed the position of the emoji in the sentence. In their collection of 2475 examples randomly sampled from Twitter, the emoji occurs at the end of the message in 51% of the cases. As for whether it replaces a word or not, it is redundant 36% of the time when it occurs in sentence final position, and only 32% of the time otherwise (Donato & Paggio, 2017). Therefore, we can assume that usages in which emoji are not sentence final, and used as word replacements, are fairly frequent. It was noted earlier that the lack of a conventionalized meaning can result in different interpretations of emoji and create misunderstanding. Using emoji as replacements for words probably increases the potential for miscommunication. Taken together, these two factors may have an effect on the ease or difficulty with which emoji are processed in sentences.

In this study, we investigate the processing of emoji by studying the eye movements of users while they read sentences containing emoji. We are interested in two questions, which give rise to two different hypotheses. The first one is that emoji are more difficult to process when used in a non-redundant way, that is, instead of a word, rather than accompanying a word as a kind of illustration (cf. (3) with (1) above). The second is that the processing of emoji is subject to constraints deriving from the context in which they occur in a way similar to what happens with words. Increased processing time should be observed, in particular, if the emoji is ambiguous and its interpretation has to be derived from the context. To investigate these hypotheses, rather than looking at facial expression emoji as many previous studies have done, we focus on emoji denoting objects.

On a more general level, we also want to discuss whether emoji, when used in text messaging, are processed in a way similar to words and are therefore subject to different levels of processing associated with lexical access and postaccess integration.

The structure of this article is as follows. In Section 2, we briefly review literature where either the processing of emoji has been addressed, or eye tracking has been used to investigate word processing in a way that is relevant to understand the processing of emoji. Then we describe our methodology in Section 3, which includes our experimental setup as well as the description of a questionnaire administered to the experiment participants. Section 4 reports results, which are discussed in Section 5, where we also draw the conclusions and illustrate ideas for further research.

2. Background

From the cognitive evidence we have on the reading of logographic scripts compared to alphabetic scripts, we know that it requires more visual processing, and involves the right hemisphere to a larger extent (Tan et al., 2001). We also know that speakers develop specialized reading strategies tailored to the writing script of their language (Tzeng & Wang, 1983). There is also evidence showing that Chinese children learn to read by writing, that is, copying the characters many times (McBride-Chang, Chung, & Tong, 2011; Tan, Spinks, Eden, Perfetti, & Siok, 2005; Wang, McBride-Chang, & Chan, 2014). This method of acquisition influences both lexical representation and lexical access of Chinese characters/words (Chinese also has a more fuzzy word boundary). The grapheme to phoneme conversion process during lexical access is not essential in the reading of Chinese. One would expect the processing of emoji to exhibit the same phenomenon, that is, to take place without access to the phonology of the symbol as a prerequisite of lexical access, since there is no standard pronunciation or sound associated with an emoji. During our experiments, when being asked to read a sentence containing emoji aloud, participants usually just read the word the emoji represented, with a few of them adding the word *emoji* afterward.

Since emoji do not have a specific associated sound, we may expect their processing to be more similar to that of characters in logographic languages than to words in alphabetic ones. However, readers do not learn to write emoji the way Chinese children learn to write Chinese characters, nor is the meaning of emoji conventionalized the way the meaning of characters is in a logographic language. Rather, readers are supposed to infer the meaning of emoji in an intuitive manner. The interesting question to us is how easy the task is, and how it affects the reading process.

Relevant evidence also comes from studies comparing the reading of Japanese hiragana (syllabic, with a specific associated syllable) and kanji (adopted logographic Chinese characters). Several studies have found differential activation in brain regions during processing of hiragana versus kanji (Coderre, Filippi, Newhouse, & Dumas, 2008; Thuy et al., 2004); for a review, see Mori (2014). A meta-analysis of neuroimaging studies of the processing of hiragana compared to Chinese characters as well as Western words found commonalities across languages in terms of gross cortical regions, but localization within those regions depending on the writing system (Bolger, Perfetti, & Schneider, 2005). The functional magnetic resonance imaging (fMRI) study in Buchweitz, Mason, Hasegawa, and Just (2009) found differences in the activated brain areas when native Japanese speakers read hiragana and kanji, respectively. The results support the idea that processing of character or language symbols that denote a specific sound or a specific meaning can be different. Brain areas associated with phonological processing are more activated during the reading of syllabic hiragana, while areas associated with visuospatial processing are more activated during the reading of kanji. Coderre et al. (2008) also found differential activation in brain regions for hiragana and kanji, which they elicited using the Stroop task. In spite of the different ways in which hiragana and kanji are processed, Japanese readers are used to processing them together in text. The question is how well readers of alphabetic words integrate the semantics of emoji in written text.

There is not a lot of research dealing specifically with the way emoji are processed to integrate them in the understanding of textual stimuli, and these few studies mostly focus on emoticons or facial expression emoji. An example is Ousterhout (2017), which used electroencephalogram (EEG) to investigate whether the meaning of emoji is clear enough to give rise to incongruity effects, in particular N400. The study was inconclusive, and the author hypothesized that the failure in eliciting an N400 effect may be due to the participants showing varying interpretations of the emoji used in the experiment. In Weissman (2019), “wink” emoji were found to elicit the same ERP effects as ironic texts, that is, a P200 followed by a P600 effect. In the second experiment reported in the same paper, the N400 effect was observed in contexts where emoji replaced words in sentence final position. Both experiments thus suggest that the content of the emoji is somehow integrated into the meaning of sentences during reading. A study that investigates the processing of emoji using the eye-tracking paradigm is Robus, Hand, Filik, and Pitchford (2020), which found longer fixation durations (both first fixation and later fixations) on the emoji (either smiling or frowning emoji) when used in sentence final position compared to sentence initially. The authors claimed that the effect was possibly due to the semantic integration processes that usually occur in sentence final position. Therefore, since our aim is to study lexical access and processing of emoji in sentence reading, we have constructed our stimuli in such a way that the emoji and target words are never in sentence final or initial position.

Howman and Filik (2020) applied eye-tracking analysis to the processing of emoticons in sarcastic comments. They found longer first-pass and total reading times on a word when it was followed by a wink emoticon rather than a full stop. In contrast, shorter first fixation duration (FFD), first-pass reading times, and regression reading times were observed in the regions preceding the same word. The study showed that the presence of an emoticon in a sentence can affect the processing of various parts of a sentence.

Given the scarcity of eye-tracking studies targeting emoji, it is relevant to see what we can learn from eye-tracking studies of word reading processes when investigating the way emoji are read. A large number of previous studies (e.g., Armstrong & Olatunji, 2012; Godfroid, Boers, & Housen, 2013; Guillon, Hadjikhani, Baduel, & Rogé, 2014; Rayner, 2009; Rayner, Sereno, Morris, Schmauder, & Clifton Jr, 1989; Rehder & Hoffman, 2005) suggest that eye tracking is a useful paradigm to investigate cognitive processes in different fields like reading and attention, and particularly sentence comprehension processes and parsing phenomena. Though the measurement is just a reflection, that is, not a direct measure, of the cognitive processes involved or the cognitive load, the results can provide objective and quantifiable insight into what is happening in the brain, especially subconscious behaviors, under natural reading setting. Eye-tracking measures, especially those concerning fixations and gaze duration, can indicate the position where the reader encounters difficulties or undergoes time-consuming cognitive processes. Fixations can also show where the reader is paying more attention. Eye-tracking studies on the scan path of reading, that is, the regression or rereading phenomenon, can reveal where, when and how we reanalyze the text during ambiguity resolution (von der Malsburg & Vasishth, 2013).

A wealth of studies has used eye tracking to investigate reading patterns involving words. Eye tracking provides a record of the processing of text down to the level of individual words in millisecond scale, and it is generally recognized that measures of eye movements during reading tell us how words are processed, and that this processing is affected by a number of different factors. Word length and frequency as well as low-level processing (Kennedy, Pynte, Murray, & Paul, 2013; Kliegl, Grabner, Rolfs, & Engbert, 2004) explain much of the variance, but effects are also due to higher level processing, for example, lexical surprisal (Hale, 2001; Smith & Levy, 2013), syntactic complexity (Demberg & Keller, 2008; Staub, 2011), and the plausibility of the context in which the target word appears (Abbott & Staub, 2015; Rayner, Warren, Juhasz, & Liversedge, 2004). It has been shown, in fact, that there is a complex interplay between lower lexical and higher level contextual factors. For instance, word frequency and contextual predictability may give rise to additive or interactive effects (Hand, Miellet, O'Donnell, & Sereno, 2010; Sereno, Hand, Shahid, Yao, & O'Donnell, 2018).

It has been claimed that different eye-tracking measures may reflect the different stages involved in word processing. Thus, it has been suggested that *first fixation duration*, that is, the time spent by the eyes fixating on the target word area the first time, may be closely related to the initial stages of the processing, particularly lexical access (Inhoff, 1984). In contrast, other measures seem related to the integration of words in the surrounding syntactic and semantic context. Examples are *gaze duration*, a measure that aggregates consecutive fixations on the same word before the next word is fixated (Just & Carpenter, 1980), *go-past time*, which measures the duration of all fixations until the eyes go past the target word including time spent looking again at earlier parts of the sentence (Rayner et al., 2004), or number of *regressions* in and out of the target word (Rayner et al., 1989). The different effects of properties such as frequency and length, on the one hand, and syntactic and semantic contextual clues, on the other hand, on different eye tracking measures, have been explained by models of word processing according to which lexical access and integration are two distinct processing steps, particularly the E–Z model (Reichle, Rayner, & Pollatsek, 2003). In contrast, reading is modeled in terms of parallel processing in the SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005), according to which several words are processed at the same time and lexical completion is not strictly tied to the order in which words occur in sentences.

A word property that is particularly relevant for our study is lexical ambiguity, in which several eye-tracking studies have shown to affect reading processes (Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Rayner & Morris, 1991; Sereno, Pacht, & Rayner, 1992). In Rayner and Duffy (1986), participants were asked to read for comprehension sentence stimuli containing ambiguous words. First fixation duration, gaze duration during first visit, and total gaze duration including regression were compared to those obtained on control stimuli with non-ambiguous words. The results show different patterns depending on whether the words are balanced ambiguous (*equi-biased* in the authors' terminology), that is, whether both senses are equally likely, or not. Balanced ambiguous words require more first fixation time, probably due to the fact that both readings are present in memory at first glance. Ambiguous words for which one sense is dominant, on the contrary, display longer gaze duration measures than the control words, probably because only the most likely reading is accessed to start with, but postaccess processing may be more complex if the context enforces the less likely interpretation. The study in Duffy et al. (1988) establishes that gaze duration on the target word is affected by the type of target word (ambiguous or control), the type of ambiguity (balanced or not), and the location of the disambiguating clause (preceding or following the target word). The authors claimed that their results are supported by a model of lexical access in which several senses of an ambiguous word are activated, but the degree of activation is influenced by frequency and prior context. Which model best captures the facts related to the processing of ambiguous words is discussed in Sereno et al. (1992), who compare the processing of subordinate senses of ambiguous words to that of high-frequent and low-frequent control words. This study finds subordinate senses of ambiguous words to be significantly harder to process than both control types across three of the measures being investigated, that is, total fixation time, fixation of the word following the target, and the next two words (but not total gaze duration). The authors conclude that their results are compatible with lexical access models in which both senses are activated, either being ordered by frequency and bias (the *reordered access* model), or with the dominant sense accessed first but not being successfully integrated because of the context (the *integration* model).

It is an open question in what way the eye-tracking measures discussed above would reflect the processing of emoji in text. Furthermore, it is unclear in what sense we can apply the concept of lexical access to emoji, since it cannot be claimed that language users possess a vocabulary of emoji in the same way as they do for words. If talking about a mental lexicon of emoji makes sense, we do not know what emoji properties are represented in such a lexicon, and thus what cognitive processes are involved during lexical access or how readers integrate the semantic information provided by the emoji with the rest of the sentence.

In this study, we look at two main aspects of emoji processing. The first one is whether the way emoji are used, either following a word or non-redundantly as a word replacement, affects processing. The second one is whether ambiguity affects their processing similarly to what has been observed for words.

3. Methodology

This study consists of two distinct experiments which were, however, run at the same time. Thus, the stimuli of the two experiments were mixed and randomized such that the sentences from each context set were at least five trials apart. A python program was used to generate eight randomized versions that fulfill the aforementioned constraints with the chosen contents for a specific set. There are four different experiment–content sets in total as we counterbalanced the two contents in each of the two experiments (2×2), see Section 3.4 and Section 3.5 for details. We ran the two experiments together so that the sentences from the two experiments could act as the fillers of each other. The participants, procedure, and apparatus were the same. They are described below followed by separate descriptions of purpose and materials for each of the two experiments.

Table 1
Statistical power of three models predicting FFD, TVD, and TRD from the effect of condition with varying number of participants (29, 40, and 60). Effects are rounded up to three decimals

	Power with Varying Number of Participants					
	Exp 1			Exp 2		
Measure	$n = 40$	$n = 50$	$n = 60$	$n = 40$	$n = 50$	$n = 60$
FFD (effect ≈ 0.02)	0.065	0.069	0.073	0.211	0.255	0.298
TVD (effect ≈ 0.14)	0.697	0.795	0.864	0.999	1	1
TRD (effect ≈ 0.1)	0.428	0.516	0.595	0.992	0.998	1

3.1. Power analysis

An initial number of stimuli and participants were considered based on previous works, particularly the second experiment in the study discussed in Duffy et al. (1988), where 32 participants were exposed to 20 sentence stimuli each. Data were therefore collected and analyzed from an initial sample of 29 participants in a pilot project for the current study. Several sources have pointed to the difficulty of assessing statistical power for linear mixed models (Brysbaert & Stevens, 2018; Kumle, Vo, & Draschkow, 2020). Nevertheless, we calculated the approximate statistical power of the models we built on this initial sample using the web interface described in Judd, Westfall, and Kenny (2017).[3] Given the different effects established for the eye-tracking measures under examination, the counterbalanced design adopted in our study, and the participant and item variability observed in this initial sample,[4] the calculation showed that further expanding the number of participants would yield the power levels displayed in Table 1.

As can be seen, the small effects established by the model on first fixation duration (FFD) could not be predicted reliably even by doubling the initial number of participants. For total visit duration (TVD) and total revisit duration (TRD), on the contrary, power reaches a level ranging from 0.5 to 1 given a sample of 50 participants. It was therefore decided to set the final sample size to 50.

3.2. Participants

The experiment participants were 53 students (21 males, 32 females; mean age: 26.42 years with SD 3.71 years) from the University of Copenhagen. None of them has any reading difficulties. They did the experiment voluntarily without any reward. All of them are either native speakers or proficient users of English who met the English language proficiency requirements for their study program at the University of Copenhagen.[5] They have normal or corrected vision. All of them signed an informed consent statement allowing us to collect their eye tracking and survey data and use them for the present study. Six pilot participants were also recruited to test our materials and experimental design.

3.3. Procedure

The procedure was explained to the participants. The eye-tracking system was then calibrated for each of them right before the start of the experiment. The calibration is a five-point calibration, with one point at each of the four corners and one point at the center of the screen. The experimental trials were preceded by three practice trials. At the beginning of each trial, a left fixation cross was displayed for 1.5 s. The participant was asked to look at the fixation cross that marked the position of the first letter of the sentence. The sentence was then displayed after a blank page of 0.5 s. Participants were told to read the sentences silently and press the spacebar to indicate that they had finished reading the sentence. A cross would then appear at the sentence initial position of the next target sentence: Participants were asked to look at the cross and wait for the onset of the next sentence.

After 25% of the trials, they were asked to repeat aloud the sentence they had just read. This was done to ensure that they were paying attention to what was shown to them on the screen. In general, participants had no difficulty in repeating the sentences to the experimenter. Occasionally, however, they could not recall the name used for the sentence subject.

Participants were asked to keep their head still throughout the whole experiment, which took about 5 min. There was no recalibration after the read-aloud as the experiment was relatively short and a recalibration after each read-aloud trial (25% of the trials) could affect its fluency. The experimenter did ask the participants to try their best to keep their head still. It must be noted that the Tobii software can tolerate small head movements and validates the gaze points acquired. In addition, an area of interest (AoI) 1 cm bigger than the cross was drawn around the fixation cross at which the participants were asked to look right before each sentence and the proportions of fixations within the AoI did not vary according to the presentation sequence. Moreover, there is no increase in missing data of FFD or TVD according to the presentation sequence either. The validity and accuracy of the data do not seem to deteriorate over the course of the experiment.

After the experiment, participants were asked to fill in a questionnaire on their language proficiency and use, as well as their use of emoji in texting. They were also asked to judge, for each of the ambiguous emoji, the acceptability of the two interpretations used in the experiment. The interpretation is presented in the survey as a single word (e.g., *cereals*, or *soup*). The survey and a summary of its results are available in the Appendix.

3.4. Apparatus

Binocular eye movements were recorded using a Tobii T120 eye tracker, with 120 Hz sampling frequency, 0.5 degrees accuracy, 0.3 degrees spatial resolution, 0.2 degrees head movement error, and 0.1 degrees drift. The screen resolution was 1280 × 1024 pixels. The distance between the participant's eyes to the eye-tracker was kept at about 64 cm. The data were validated and averaged from both eyes. The eye movement data were processed by means of Tobii Studio 3.4.8. The identification by velocity threshold (I-VT) fixation filter was set to be 30 degree/s, with minimum fixation duration of 60 ms. Adjacent fixations with maximum duration of 75 ms and angle of 0.5 degree between them were merged. We extracted

Table 2
Emoji used in the first experiment

Emoji	Name	Unicode	Freq	Word
🍕	pizza	1F355	27.1M	pizza
⚽	soccer ball	26BD	32.8M	football
✈️	airplane	2708	34.2M	plane
🎈	balloon	1F388	28.7M	balloon
📷	camera	1F4F7	108.5M	camera
🎁	wrapped gift	1F381	20.4M	gift
🌸	cherry blossom	1F338	97.7M	flower
🌈	rainbow	1F308	22.7M	rainbow
💍	ring	1F48D	14.2M	ring
🎂	birthday cake	1F382	20.3M	cake

the data preprocessed by the Tobii Studio, that is, FFD, TVD, and TRD values, for further data analysis in R.

3.5. Experiment 1

In the first experiment, we investigated the effect of emoji in sentence processing through eye movement patterns. Sentences with and without emoji were presented to the participants. The goal was to test if it makes a difference in terms of processing whether emoji are used redundantly after a word with the same meaning, or whether they occur as word replacements. Sentences with no emoji were used as controls. If the lexical access of emoji is more difficult than for words, longer first fixation is expected when emoji are used as word replacements compared to word targets. In contrast, for the condition where the emoji are used redundantly, that is, following the word they denote, we predict that the FFD for the emoji will be shorter due to semantic priming across the preceding word and the emoji, since lexical access in this case is facilitated. If the inclusion of an emoji in a sentence causes more demanding post-access integration, longer TVD and revisit duration are expected at the emoji target.

Ten frequent emoji were chosen from the 250 most frequent emoji on http://www.emojitracker.com. We selected the most frequent ones with unambiguous meaning while avoiding smileys since we were not interested in emotion-oriented emoji. To confirm the preliminary judgement of the two authors, we asked the pilot participants for their opinions and they all agreed that the emoji we selected were unambiguous. They also all agreed on the interpretation of the emoji in the contexts provided (see below). The selected emoji are shown in Table 2 with their frequencies (extracted on 27/11/2018), their official names, their Unicode symbols as reported in https://emojipedia.org, and the target words they are associated with in the experiment. We controlled the number of letters for the target words to be within four and eight letters.

Ten different neutral contexts were constructed. The sentences start with a common English name as the subject (10 different names) followed by a verb selected from a list of frequent verbs with relatively open arguments to eliminate the predictability factor. The objects of the sentence are our target words, emoji, or combinations of the two. All sentences thus consist of subject, verb, object (word, emoji, or word + emoji), and an adjunct. The adjuncts were added so that the target area (the emoji or target word) was not in sentence final position. This was done to avoid the occurrence of additional viewing (visit) time due to processes that typically happen at the end of the sentence, for example, integration. In addition, we deliberately inserted two spaces between each word to isolate the words/emoji better, and thus, ensure a more robust drawing of the AoI since the accuracy of the eye tracker is typically 0.5 degrees. The length of the sentences is up to 42 characters including spaces. The sentences were presented in the middle of the screen, which is a 17 inches TFT monitor. All the sentences took up only one line. The font used was Cambria, with size 22.

Each of the emoji is used in two contexts resulting in 20 sets of sentences. For each set, there are three conditions: (a) word only, (b) emoji only (the word is replaced by the emoji), and (c) word + emoji (the word is followed by the corresponding emoji). The complete list of stimuli is provided in the Appendix, but an example is shown below. For each linguistic context, the conditions are shown one after the other.

(a) 1. John brought a pizza yesterday.
    2. John brought a 🍕 yesterday.
    3. John brought a pizza 🍕 yesterday.
(b) 1. Ann bought a pizza from the shop.
    2. Ann bought a 🍕 from the shop.
    3. Ann bought a pizza 🍕 from the shop.

Each participant had to read 30 sentences relating to the first experiment, that is, 10 different contexts $*$ 3 conditions. The three conditions were shown as far as possible in different contexts. For example, the same participant read word only condition in context (a), emoji only condition in context (b), and word + emoji condition in context (a).

3.6. Experiment 2

In the second experiment, we investigated once again the processing of emoji in sentences. This time, however, we chose emoji that could receive different interpretations depending on the context to see how ambiguity affects the processing in terms of eye movement. The expectation here was that ambiguous emoji would receive longer FFD, TVD and TRD than non-ambiguous ones. In other words, we expected the emoji in this experiment to be more difficult both in terms of lexical access and post-access integration. We also expected the processing of the non-redundant emoji in some contexts to be more difficult than in others, the assumption being that at least for some of the potentially ambiguous emoji, one of the contexts may imply a more obvious interpretation than the other. As in the first experiment, the FFD for the redundant emoji was expected to be shorter if there was semantic priming from the preceding word.

Table 3
Emoji used in the second experiment

Emoji	Name	Unicode	Frequency	Words
🥣	bowl with spoon	U+1F963	unav.	soup/cereals
☕	hot beverage	U+2615	21M	coffee/tea
🍟	French fries	U+1F35F	9M	McDonald’s/chips
👑	crown	U+1F46B	14M	crown/queen
👫	woman and man holding hands	U+1F4BC	1M	friends/lovers
💼	briefcase	U+26FA	1M	job/briefcase
⛺	tent	U+1F384	22M	tent/camping
🎄	Christmas tree	U+2614	11M	Xmas tree/Xmas
☔	umbrella with rain drops	U+1F64F	205M	umbrella/rain
🙏	folded hands	U+1F451	56M	thanks/please

Ten of the 250 most frequent ambiguous emoji from http://www.emojitracker.com were selected, such that two contexts could be constructed for each emoji, corresponding to two different interpretations. Table 3 shows the selected emoji together with their names, codes, frequencies (extracted on 27/11/2018), and the words associated with each emoji in the experiment. Note that a frequency count was not available at the time for the “bowl with spoon” emoji. The emoji was nevertheless selected since the image seems very easy to understand. According to Google Trends, in February 2021, the “bowl with spoon” emoji was about 10% less frequent than “hot beverage.” The emoji were chosen by the authors, who are from very different cultures, in such a way that agreement could easily be reached about their meaning in the various contexts. Although this is no guarantee that the experiment participants would share the same view, it does ensure that the choices made are not totally monocultural. In addition, and in order to confirm our preliminary judgement, we asked the pilot participants (one from Finland, one from Spain, one from the Netherlands, one from Australia, and two from Denmark, aged between 30 and 61 years) for their opinions. They all agreed that the emoji we selected were understandable and the denotations suggested by the associated words possible. As we did for the first set of materials, we also controlled the number of letters for each of the target words to be within four to eight letters, except for the “French fries” emoji, where *McDonald’s*, which has nine letters, is used as one of the target words.

Ten sets of sentence stimuli were constructed, one for each emoji. The sentence structure was the same as in the first experiment. Within each set, two contexts were constructed such that the context enforces a specific interpretation of the emoji. There are two conditions: (a) emoji only (the emoji replaces the word) and (b) word + emoji (the emoji is preceded by a word corresponding to the specific sense implied in the sentence). Therefore, for each emoji, four sentences were constructed, that is, 2 contexts $*$ 2 conditions. The complete list of stimuli is provided in the Appendix, but an example is shown below.

(a) 1. Jack ate 🥣 for dinner.
    2. Jack ate soup 🥣 for dinner.
(b) 1. Jack ate 🥣 for breakfast.
    2. Jack ate cereals 🥣 for breakfast.

Depending on the condition, context is provided to interpret the emoji either by the final adjunct (1), or when the word preceding the emoji is read (2).

In addition to the stimuli just described, five filler sets were constructed, with four examples in each (2 contexts $*$ 2 conditions). These emoji are also ambiguous, but the sentence structure is slightly different. The complete list is included in the Appendix.

For the second experiment, participants had to read 20 experimental sentences (10 emoji $*$ 2 conditions) as well as 10 filler sentences for a total of 30 trials. The presentation of the experimental sentences was counterbalanced such that each context was presented only once to a participant, for example, emoji only condition in context (a) and emoji + word condition in context (b) or vice versa. Layout and mode of presentation were the same as in the first experiment.

To sum up, each participant had to read 60 sentences in total, 30 for the first experiment, and 30 for the second experiment including 10 fillers. The order of the stimuli was randomized.

3.7. Analysis methods

The analysis is based on three eye-tracking measures, that is, *FFD*, *TVD*, and *TRD*. We take the first measure mostly to reflect lexical access processes, while the other two reflect post-lexical integration. FFD is the total time measured when the gaze first keeps a relatively still central foveal vision at a small area within an AoI during the first pass through the AoI. The duration is usually about 150–300 ms. TVD is the total time the eyes spend on the AoI. There can be multiple visits. For each visit, the duration is the time from the start of the first fixation until the end of the last fixation within the AoI, before the exit saccade. In other words, TVD is the sum of the durations of all the visits within the AoI including both fixations and all the saccades between the fixations. TRD was calculated by subtracting first fixation as well as immediate refixation time from the TVD on a given AoI, and thus, corresponds to time spent regressing into the AoI. Fixations shorter than 100 ms were disregarded.

In the drawing of AoIs, we tried to balance sensitivity and selectivity by adding two spaces between each word or emoji. The left and right boundaries of an AoI fall at around the middle of the space between two adjacent tokens, whether words or emoji. As for height, the boundaries are about 1 cm from the top and bottom of the word or emoji. The size of the emoji targets in visual angle was between 0.08×0.15 and 0.12×0.15 degrees. For the word targets, it was between 0.15×0.09 and 0.55×0.1 degrees.

In the analysis of the first experiment results, we distinguish between four different conditions, that is, *word-only*, *emoji-only*, *word+emoji* with *emoji* target, and *word+emoji* with *word* target. In other words, in *word+emoji* stimuli, AOIs were set on the emoji or the preceding word. For the second experiment, we distinguish between three conditions, that is, *emoji-only*, *word+emoji* with *emoji* target, and *word+emoji* with *word* target.

Table 4
Skipping probabilities (%) in different conditions in the two experiments

Condition	Target	Exp1	Exp2
both	emoji	31.3	7.4
both	word	10	3
e_only	emoji	11.1	3.2
w_only	word	10	NA

The raw data consisted of observations for 53 participants × 10 different sentences × 4 conditions in the first experiment, and × 3 conditions in the second one. Data from two participants had to be removed, in one case because of too few valid fixations within the AoIs, that is, very high skipping probability, and in the other because the emoji seemed to be skipped systematically.

We used R (R Core Team, 2013) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) to create a number of linear mixed effect (LME) models to test the effect of the independent condition variable on each of the measures, where condition was treated as a fixed effect, while subject, item, and emoji frequency were treated as random effects. We also tested the effect of ambiguity introduced in the second dataset through an additional independent variable called "experiment." Random intercepts for all the variables were included in all the models. Before creating the models, the values of the three dependent measures were first visualized using Q-Q plots, and then log-transformed to diminish the positive skew and improve model fit (Baayen & Milin, 2010; Ingram & Hand, 2020). Maximum likelihood was used to fit the models. Significance of a main effect on the measures was established by likelihood ratio tests where the full model with the effect in question was compared against the model without it (Gerstenberg, 2021; Winter, 2013). In the cases in which a significant overall effect was established, post-hoc pairwise comparisons between the various conditions were run using the emmeans package (Lenth, 2021) and applying Tukey adjustment.

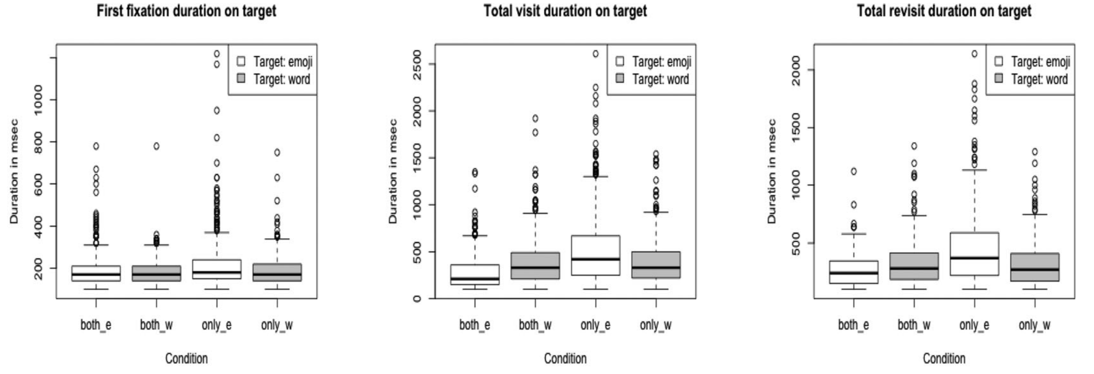
4. Results

In this section, we report results obtained on the two sets of stimuli first separately, and then in comparison to one another. We then focus on the processing of stimuli containing ambiguous emoji.

After having removed data from the two problematic participants, the remaining total observations were 2040 and 1530 for the two experiments, respectively. Skipping probabilities based on TVD in the different conditions are shown in Table 4. The empty values were omitted in the analysis. Values smaller than 100 ms for the three measures under consideration, which were 295 (4.8% of the total observations for the three measures) in the first experiment and 201 (4.4%) in the second one, were also disregarded.

Table 5
Mean, standard deviation, and 95% confidence interval values of first fixation duration (FFD), total visit duration (TVD), and total revisit duration (TRD) for different conditions and targets in the first experiment expressed in msec

Measure	Condition	Target	M	SD	CI
	both	emoji	190.27	88.38	183.08, 197.46
FFD	both	word	181.01	56.93	176.37, 185.65
	e_only	emoji	216.12	126.68	205.81, 226.43
	w_only	word	187.27	68.76	181.67, 192.87
	both	emoji	282.85	194.18	267.05, 298.65
TVD	both	word	390.63	250.54	370.22, 411.04
	e_only	emoji	523.52	389.03	491.86, 555.18
	w_only	word	385.17	241.86	365.49, 404.85
	both	emoji	278.74	174.92	264.50, 292.97
TRD	both	word	335.25	210.72	318.09, 352.41
	e_only	emoji	474.51	370.78	444.33, 504.69
	w_only	word	324.90	212.73	307.59, 342.21



(a) FFD distribution (b) TVD distribution (c) TRD distribution

Fig. 1. First fixation duration (left), total visit duration (center), and total revisit duration (right) distribution in different conditions expressed in msec. In the *word+emoji* condition, the measures are given separately for the emoji and the word ("both_e" and "both_w" in the figure, respectively).

4.1. Experiment 1

In Table 5, we show summary and dispersion counts for FFD, TVD, and TRD in different conditions and targets in the first experiment. All measures are given in ms. Fig. 1 displays the distributions of the same three measures in the various conditions. As can be seen, there are outliers in the higher end of the distribution. In other words, all distributions have a positive skew. Outliers in the right tail constitute 4.8% of the overall distribution for FFD, 4.7% for TVD, and 6% for TRD.

The values of the three dependent measures were log-transformed to diminish the positive skew. Then a number of LME models were created to test the effect of the four conditions

Table 6
Summary of effects established by linear mixed effects models predicting the effect of condition on first fixation duration (FFD), total visit duration (TVD), and total revisit duration (TRD). Significance established by comparing the full models with models without the condition predictor

	FFD	TVD	TRD
Fixed effects			
intercept	5.192	5.739	5.634
condition	0.010	0.180	0.052
Random effects			
item intercept	0.003	0.047	0.026
subj intercept	0.013	0.052	0.041
emoji freq. intercept	4.649e–4	3.881e–10	0.005
Residual	0.105	0.309	0.319
χ^2	0.970	52.341	3.257
df	1	1	1
p	.325	<.001	.071

on each of the measures, where condition was treated as a fixed effect, while subject, item, and emoji frequency were treated as random effects. Random intercepts for all three variables were included in all the models. The models were compared to reduced versions without the effect in question to establish significance, and in the cases in which a significant main effect of condition was found, pairwise differences were tested. The results are shown in Table 6 and discussed below for each eye tracking measure separately.

4.1.1. First fixation duration results

As can be seen from Table 5, the mean FFD in the *emoji-only* condition ("e_only" in the table) is slightly longer than in all other conditions irrespective of the target, indicating that initial processing of an emoji used instead of a word may be relatively difficult. It may be the case that lexical access of emoji takes longer in general, while in the redundant *word+emoji* condition ("both" in the table), access was facilitated by the priming effect of the preceding word. However, the standard deviation is higher in the *emoji-only* condition, and there are many outliers in the higher end of the distribution for this condition as shown in the plot in Fig. 1(a). In fact, the linear model predicting FFD from condition only found a very small effect of $\approx$0.01, and the comparison of the model with an equivalent one without the condition predictor did not reach significance (Table 6).

4.1.2. Total visit duration results

Here we consider TVD, which is the total amount of time given by all visits within an AoI of the target in each condition. This measure may contain immediate refixations as well as time from revisits to the target.

From the distribution in Fig. 1(b), we see that the TVD on a word target seems almost the same irrespective of whether the word is used alone or followed by an emoji. The TVD distribution on an emoji target, on the contrary, is longer and varies considerably more when

Table 7
Effects of condition on total visit duration (TVD) in the first experiment: pairwise comparisons for different conditions

Measure	Pair	Effect	Diff (%)	*p*-Value
TVD	both_e – both_w	−0.351	14.2	<.0001***
	both_e – e_only	−0.572	17.7	<.0001***
	both_e – w_only	−0.346	14.1	<.0001***
	both_w – e_only	−0.221	12.5	.0011**
	both_w – w_only	0.005	10	.9998
	e_only – w_only	0.226	12.5	.0008**

the emoji is used alone compared to when it follows a word. The linear model predicting TVD from condition found an effect of 0.18, and the comparison of the model with an equivalent one without the condition effect reached significance ($\chi^2 = 52.341$, $p < .001$, see Table 6).

Table 7 shows pairwise comparisons between each pair of conditions. All comparison but one resulted in a significant difference. To interpret the effect in each comparison, we transform it via its exponentiated value to get the proportional difference in geometric means between the two conditions.[6] In general, we see that in all the comparisons involving the *emoji only* condition, the emoji target is gazed at more than the target it is being compared with—whether a word or an emoji in the *word+emoji* condition. The difference in attention received by a word target preceding an emoji compared to when it occurs in isolation is not significant. Finally, in the *word+emoji* condition, the emoji is looked at less than the word.

4.1.3. Total revisit duration results

The distribution of TRD (Table 5 and Fig. 1c) seems quite similar to the one observed for TVD, with roughly the same amount of regression being directed toward a target word no matter whether the word is followed by an emoji or not, and more regression time spent on a target emoji when the emoji replaces rather than follows a word. The linear model predicting TRD from the condition variable, however, only found a small effect of ≈0.05, and approached but did not reach significance (see again Table 6).

4.1.4. Conclusion: Experiment 1

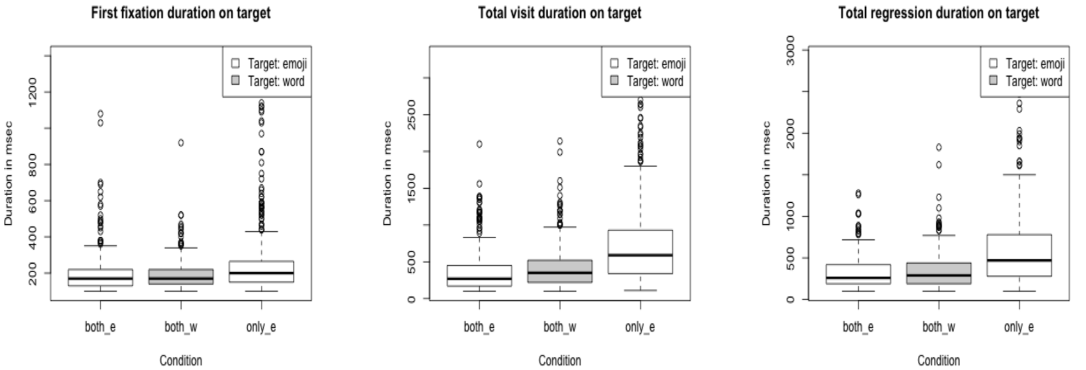
To conclude, a significant effect of condition was established for TVD but not TRD nor FFD. In accordance with our hypotheses, the emoji receives significantly less visual attention in terms of TVD than the other targets examined when it follows a related word. Conversely, it receives significantly more visual attention when it replaces a word.

4.2. Experiment 2

In Table 8, we show summary and dispersion counts for FFD, TVD, and TRD for different conditions and targets in the second experiment. All measures are given in msec as in the first experiment.

Table 8
Mean, standard deviation, and 95% confidence interval values of first fixation duration (FFD), total visit duration (TVD), and total revisit duration (TRD) for different conditions and targets in the second experiment expressed in msec

Measure	Condition	Target	M	SD	CI
FFD	both	emoji	198.80	111.37	189.74, 207.86
	both	word	189.54	77.65	183.22, 195.86
	e_only	emoji	248.47	177.96	233.99, 262.95
TVD	both	emoji	369.47	289.50	345.91, 393.03
	both	word	418.28	274.10	395.97, 440.59
	e_only	emoji	748.37	579.85	701.18, 795.56
TRD	both	emoji	347.14	239.51	327.65, 366.63
	both	word	358.98	247.70	338.82, 379.14
	e_only	emoji	644.13	540.04	600.18, 688.08



(a) FFD (b) TVD (c) TRD

Fig. 2. First fixation duration (left), total visit duration (center), and total regression duration (right) distribution in different conditions expressed in msec. In the *word+emoji* condition, the measures are shown separately for the emoji and the word ("both_e" and "both_w" in the table, respectively).

The distributions of the three measures are visualized in boxplots in Fig. 2. As was the case for the measures obtained in the first experiment, there are outliers in the higher end of the distribution. They constitute 7.6% of the overall distribution for FFD, 6% for TVD, and 7.2% for TRD. All distributions are slightly more right-skewed than in the first experiment.

As we did in the first experiment, the values of the three measures were log-transformed to diminish the effect of the outliers, and LME models were created to predict the effect of condition on each measure. This time, there are three conditions rather than four. As was done in the first experiment, condition was treated as a fixed effect while subject, item, and emoji frequency were treated as random effects, and random intercepts for all three variables were included in all the models. Significance of the main effect of condition on the measures was established by comparing the full model with the effect in question against a model without it. The effects predicted, and the results of significance testing for each of the models compared

Table 9
Summary of effects established by linear mixed effects models predicting the effect of condition on first fixation duration (FFD), total visit duration (TVD), and total revisit duration (TRD) in the second experiment. Significance was established by comparing the full models with models without the predictor

	FFD	TVD	TRD
Fixed effects			
intercept	5.247	5.953	5.802
condition	0.071	0.290	0.274
Random effects			
subj intercept	0.015	0.063	0.036
item intercept	0.004	0.025	0.012
emoji freq. intercept	0.002	0.005	0.004
Residual	0.154	0.367	0.413
χ^2	10.132	55.798	26.155
df	1	1	1
p	<.01	<.001	<.001

Table 10
Effects of condition on first fixation duration (FFD), total visit duration (TVD), and total visit duration (TRD) in the second experiment: pairwise comparisons for different conditions

Measure	Pair	Effect	Diff (%)	*p*-Value
FFD	both_e – both_w	0.016	10.16	.836
	both_e – e_only	−0.167	11.81	<.0001***
	both_w – e_only	−0.183	12	<.0001***
TVD	both_e – both_w	−0.189	12.08	<.0001***
	both_e – e_only	−0.677	19.67	<.0001***
	both_w – e_only	−0.488	16.29	<.0001***
TRD	both_e – both_w	−0.040	10.41	.820
	both_e – e_only	−0.540	17.15	<.0001***
	both_w – e_only	−0.499	16.47	<.0001***

to its counterpart without the effect of condition, are shown in Table 9. Post-hoc pairwise comparisons between the various conditions were run as done for the first experiment. The results can be inspected in Table 10.

4.2.1. *First fixation duration results*

First fixation measures for the second experiment are shown in Table 8 and Fig. 2(a). The tendency we see is similar to the one observed in the first experiment, with emoji receiving longer FFD than other targets in the *emoji-only* condition although variance is again quite high. The linear model predicting FFD from condition only found a small effect of 0.07 (Table 9). Contrary to what we saw in the first experiment, the comparison of the model with an equivalent one without the condition effect reached significance ($\chi^2 = 10.132$, $p < .01$). Two significant pairwise differences were found (Table 10) both involving the *emoji-only* condition, where FFD is consistently higher than in the other cases.

4.2.2. Total visit duration results

The results for total visit time in the second experiment show that the emoji receives lower visit time on average when it follows a related word, and considerably more when it occurs alone as can be seen in Table 8 and Fig. 2(b). Comparing a full LMEs model and a reduced version without the fixed effect in question shows a significant effect of condition on TVD ($\chi^2 = 55.798$, $p < .001$, see Table 9). Table 10 illustrates that all pairwise differences are significant, with emoji targets showing higher effects for TVD when occurring alone, and smaller than the word in the *word+emoji* condition.

4.2.3. Total revisit duration results

The distribution of TRD for the second experiment (Table 8 and Fig. 2c) shows a similar pattern to the one observed for TVD. An emoji target is revisited for a slightly shorter time on average when it follows a related word (and less than the word itself), and for considerably longer when it occurs alone. The linear model predicting TRD from condition found an effect of 0.27, and the comparison of the model with an equivalent one without the condition effect reached significance ($\chi^2 = 26.155$, $p = < .001$, see Table 6). As is the case for FFD, the two pairwise significant comparisons involving the *emoji-only* condition reach significance. The effects go in the same direction, with non-redundant emoji being gazed at more than both words and emoji in the *word+emoji* condition.

4.2.4. Conclusions: Experiment 2

To sum up the results of the second experiment, we found significant effects of condition on FFD, TVD, as well as TRD. In all cases, the emoji receives less visual attention than the other targets examined when it follows a related word, possibly due to semantic priming from the word itself. Conversely, it receives significantly more visual attention when it replaces a word. These results go in the same direction as those obtained in the first experiment for the conditions that were tested in both experiments, and both confirm our first hypothesis that emoji may be more difficult to integrate when they replace a word. In the second experiment, however, integration difficulty is also reflected in increased time spent regressing to the emoji targets. Moreover, the effect found for FFD points to the fact that lexical access may have been harder for the emoji targets used in this dataset.

In the section below, we look more closely at the differences between the results of the two experiments.

4.3. Comparison between the two experiments

When we compare mean and standard deviation values for the various eye tracking measures associated with the conditions common to the two experiments, that is, *word+emoji* condition with either emoji or word target, and *emoji-only* condition, we observe that all the measures display higher means and larger standard deviation on the emoji target in the second experiment. The difference is especially remarkable for TVD and TRD in the *emoji-only* condition. In general, the increased difficulty of processing emoji in non-redundant use is consistent with the way the participants described their preferences for emoji usage in the post-experiment survey, where 43 of them (81.1%) reported using emoji in sentence final position in a redundant way as in (a) below, whereas five (0.9%) responded that they insert the emoji in the middle of the sentence replacing the corresponding word as in (b), and 5 (0.9%) following the word as in (c).

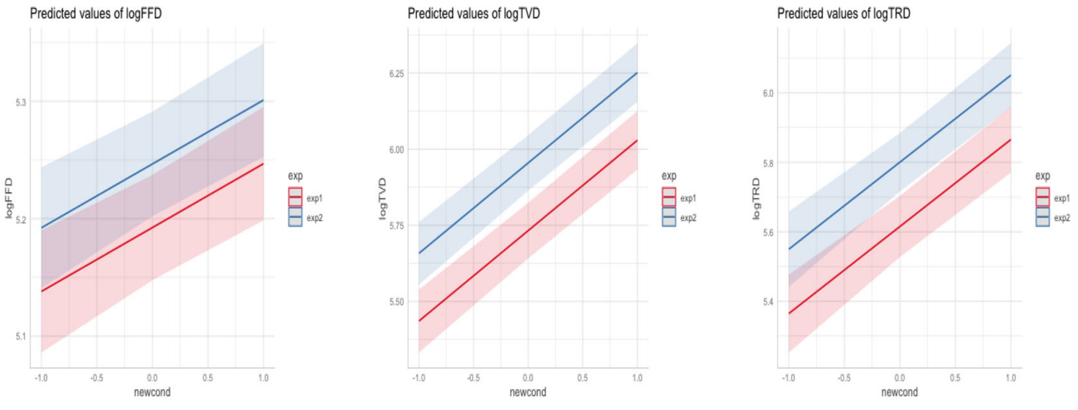
(a) She drove the car to work 🚗
(b) She drove the 🚗 to work.
(c) She drove the car 🚗 to work.

However, this difficulty seems more serious in the second experiment. If it is true that first fixation reflects the difficulty of lexical access, the target seems slightly more difficult to access in the second set of stimuli, particularly in the *emoji-only* condition. The increase in TVD and TRD in the second experiment, especially again in the *emoji-only* condition, may be due, in contrast, to the fact that the emoji chosen for the second experiment are more ambiguous and therefore more difficult to integrate. They are, however, also less frequent than in the first experiment. The mean average frequency of the emoji used is, in fact, 40.66M in the first experiment (SD = 33.56M), while in the second, it goes down to a mean of 37.78M (SD = 64.85M).

We fitted LME models for FFD, TVD, and TRD on a combined dataset with results from both experiments after having removed the data from the *word-only* condition. This time we treated condition as well as experiment as fixed effects, and subject, item, and emoji frequency as random effects. Random intercepts for all three variables were included in all the models. Maximum likelihood was used to fit the models, and significance of the effect of experiment was tested by comparison with a reduced version of the model without the experiment condition. As usual, the significance of post-hoc pairwise differences was obtained by applying Tukey adjustment.

The effects established by the models are visualized in Fig. 3 and summarized in Table 11 together with the relevant significance results. Significant differences between the model with the experiment predictor and the one without are achieved for TVD ($\chi^2 = 13.434$, $p < .01$) and TRD ($\chi^2 = 8.7834$, $p < .05$). The model for FFD approaches significance ($\chi^2 = 5.9012$, $p = .05231$). We note that in all models, the variance introduced by emoji frequency is very small, much smaller than subject and item variance. All pairwise comparisons for TVD and TRD but one reach significance (Table 12). The exception is the difference in the effect on TRD between emoji and word targets in the *word+emoji* condition where, although the word receives slightly more regressions, significance is not reached.

In conclusion, there was increased visual attention in the second experiment in all the conditions represented in both experiments, and especially in the *emoji-only* condition. Introducing an experiment variable as a fixed effect in the models yielded significant results for TVD and TRD, indicating increased integration difficulty for non-redundant emoji in the second experiment. It was also noted that the effect of frequency on the results is very small. Therefore, in the next section, we explore ambiguity as a possible explanation for the increased visit duration on emoji targets in the second experiment.



(a) FFD (b) TVD (c) TRD

Fig. 3. Predicted effects of condition on first fixation duration (FFD), total visit duration (TVD), and total revisit duration (TRD) in the two experiments. All three dependent variables were log-transformed, and the condition variable ("newcond" in the figure) was scaled and centered before fitting the models.

Table 11
Summary of effects established by linear mixed effects models predicting the effect of condition and experiment on first fixation duration (FFD), total visit duration (TVD), and total revisit duration (TRD) in the entire dataset (first and second experiment combined). Significance established by comparing the full models with models without the experiment predictor

	FFD	TVD	TRD
Fixed effects			
intercept	5.192	5.730	5.619
condition	0.048	0.309	0.230
experiment	0.054	0.225	0.167
cond*exp	0.025	−0.022	0.051
Random effects			
item intercept	0.002	0.023	0.009
subj intercept	0.015	0.063	0.038
emoji freq. intercept	0.001	0.003	0.007
Residual	0.134	0.341	0.372
χ^2	5.901	13.434	8.783
df	2	2	2
p	.052	<.01	<.05

4.4. *Ambiguous emoji*

As mentioned earlier, in a post-experiment survey, participants were asked to score on a Likert scale from 1 (non acceptable) to 5 (highly acceptable) the acceptability of the two interpretations of the emoji used in the second experiment. For each emoji, the two senses were shown one after the other in random order on the same page.

Table 13 shows the acceptability scores. For each emoji interpretation, the table reports median score, sum, and the difference in sum score with the alternative interpretation. For

Table 12
Effects of condition on first fixation duration (FFD), total visit duration (TVD), and total visit duration (TRD) in the combined models: pairwise comparisons for different conditions

Measure	Pair	Effect	Diff (%)	*p*-Value
TVD	both_e – both_w	−0.272	13.13	<.0001***
	both_e – e_only	−0.624	18.66	<.0001***
	both_w – e_only	−0.353	14.23	<.0001***
TRD	both_e – both_w	−0.127	11.35	0.029
	both_e – e_only	−0.501	16.50	<.0001***
	both_w – e_only	−0.374	14.53	<.0001***

Table 13
Average acceptability scores assigned by participants to different emoji meanings

Emoji	Meaning	Median Score	Sum	Sum Diff
☕	coffee	5	267	131
☕	tea	2	136	131
💼	briefcase	5	256	94
💼	job	3	162	94
🍟	chips	5	254	81
🍟	McDonald’s	3	173	81
👑	crown	5	255	78
👑	queen	3	177	78
⛺	tent	5	261	66
⛺	camping	4	195	66
🎄	Xmas tree	5	263	64
🎄	Xmas	4	199	64
👫	lovers	4	223	54
👫	friends	3	169	54
🥣	soup	4	195	24
🥣	cereals	3	171	24
🙏	please	5	230	21
🙏	thanks	4	209	21
☔	umbrella	5	224	18
☔	rain	4	206	18

each pair of meanings associated with the single emoji, the meaning that was judged the more acceptable is listed first. Emoji pairs are ordered based on the difference in their sum scores. The difference in total score between the preferred and the dispreferred interpretation is statistically significant both in terms of a Welch two-sample *t*-test ($t = 6.9023$, p-value $< .001$), and a Kruskal–Wallis rank sum test ($\chi^2 = 14.153$, p-value <0.001).

Table 14
Mean and (SD) values for the eye-tracking measures depending on acceptability in pairs of emoji interpretations

Measure	Preferred	Dispreferred
FFD	206.49 (22.10)	204.65 (10.06)
TVD	471.83 (95.51)	480.09 (71.35)
TRD	424.04 (96.86)	433.64 (70.40)

As can be seen from Table 13, the difference in acceptability between the two senses of the emoji varies from 131 in the case of the coffee–tea pair, to 18 in the umbrella–rain pair. As a consequence, it is not so clear-cut whether the interpretations are balanced ambiguous, or whether one of the senses is dominant. The difference in acceptability seems best thought of as a continuum rather than a binary distinction.

Table 14 shows mean and standard deviation values for FFD, TVD, and TRD in the two groups. In general, the dispreferred group shows less variation in all measures, and for TVD and TRD slightly higher means, but the differences are small. In fact, the difference in acceptability between the two groups does not correspond in our data to any significant effect on how much visual attention emoji in each group receive. LMEs models were created for FFD, TVD, and TRD obtained for the second experiment stimuli, where condition and a binary bias factor were entered as fixed effects and subject, item, and emoji frequency as random effects. For none of the measures could a significant effect of bias be established.

Correlation tests were also run between the three eye-tracking measures averaged over all trials for each emoji and (a) the sum acceptability scores the same emoji were given by survey participants; and (b) the differences in sum scores between competing emoji interpretations. Bearing in mind the results discussed in Rayner and Duffy (1986), where it is claimed that balanced ambiguous words may require more initial visual attention (if both readings are accessed from the start), whereas non-balanced ambiguous words may require more post-lexical processing time, we were expecting to observe similar tendencies in our correlation tests. No significant correlation was found, however, between sum scores and eye-tracking measures (Spearman's ρ between -0.14 and 0.14, all p-values between 0.5 and 0.8). Correlation was also weak between score differences and the same measures (Spearman's ρ between -0.23 and 0.15, all p-values between 0.3 and 0.6). In sum, although the emoji used in the second experiment as a whole presented a more serious challenge to the participants, no systematic effect of the difference between the two interpretations could be observed. This result seems to point to the fact that the meaning of an ambiguous emoji may be generally unclear or vague when the emoji does not follow an associated word.

5. Discussion and conclusion

In this paper, we have investigated how emoji are processed by users during sentence reading by focusing on three eye-tracking measures: FFD, TVD, and TRD.

In general, the results we have obtained from the experiments and analyses reported in the preceding sections for total visit and partly for revisit duration confirm both our hypotheses. We see that an object-denoting emoji, when used as a replacement for a word, is gazed at for longer time in total than a word in the same context. Moreover, a redundant object-denoting emoji that follows a word as a kind of illustration of it receives less total visit time than the word. We interpret these results as evidence of the fact that emoji used as word tokens are more difficult to integrate in the processing of the sentence than the word tokens themselves, or emoji used in a redundant way. This is the result we were expecting given our first hypothesis. A possible explanation may be the fact that this usage is not so common, as also pointed out by the experiment participants in their survey answers. In other words, regardless of how well-known the emoji itself may be, it requires more attention when it replaces a word because of the low predictability of this usage (Hale, 2001; Smith & Levy, 2013).

Another possibility could be that the longer visit duration was due to the lack of parafoveal preview benefits. Object denoting emoji and words have different properties, in fact, and the emoji could be too visually complex to be perceived during the preview processes. However, since the participants must gradually have become aware that some of the sentences contained emoji, it cannot be excluded that the emoji were within the expected "word" pool that could be benefitted from the parafoveal preview (Schotter, Lee, Reiderman, & Rayner, 2015; Veldre & Andrews, 2017). The study in Howman and Filik (2020) showed that emoticons could be previewed, and that the presence of emoticons can affect eye movement patterns while reading a sarcastic comment. Emoticons and emoji are, of course, visually quite different, so the issue of how emoji behave concerning parafoveal vision merits further investigation.

In addition to this general lack of preference for the use of a logograph as a word replacement in English, we also see that even longer TVD is required to integrate ambiguous emoji in more demanding contexts. Thus, in the second experiment, where an ambiguous emoji must be assigned an interpretation to make sense of the rest of the sentence, we found a significant increase in total visit and revisit durations on emoji, particularly in the *emoji-only* condition. This allows us to conclude that, as stated by our second hypothesis, ambiguous emoji are subject to increased difficulty of post-access integration similarly to what was found for ambiguous words (Rayner & Duffy, 1986).

The results we obtained for FFD point in the same direction. No significant difference in FFD between the conditions in the first experiment could be established. However, participants give emoji replacing words slightly longer FFD, indicating that the initial processing of emoji may be slightly more difficult than it is for word tokens unless the emoji follow words that may provide a priming effect. Note, however, that Robus et al. (2020) found a shorter FFD on emoji (smiling or frowning face emoji) than words when such an emoji is in sentence initial as opposed to sentence final position. There are two differences between our emoji targets and those used in this study. First, our emoji are not facial expression emoji, and they denote objects. Semantically, they can be regarded as concrete nouns when used non-redundantly. The recognition and lexical access processes may be more complex in this case since such a usage is less common (only 9.62% in our survey), and thus, it may be both less frequent in normal reading and more surprising. The effect of surprisal has been shown to be a factor in fixation duration. Second, the emoji in Robus et al. (2020) were in sentence initial vis-à-vis final position, while ours are in the middle of the sentence. The FFD in our study might include the integration process for preceding words.

The difference in FFD between the different conditions is significant in the second experiment, where it is highest in the *emoji-only* condition. We know that frequency has a strong effect on the first fixation of words. Although we made sure to include very frequent emoji in both experiments to control for frequency, we did not manage to find usable examples within an overall narrower frequency range. As a result, the emoji used in the second set of stimuli are in general less frequent. However, including emoji frequency as a random effect in our models showed a very small effect of frequency in all cases. Therefore, it is reasonable to assume that the ambiguity of the emoji might be responsible for the increased first fixation time. In fact, participants have to choose one out of at least two possible interpretations to understand the succeeding context. In Rayner and Duffy (1986), it is argued for ambiguous words that only the preferred reading is probably accessed initially, so that increased first fixation time should not be expected. This is contrary to a different view of lexical access defended by other researchers, according to which all meanings of a word are at least momentarily accessed (Onifer & Swinney, 1981). In the case of emoji replacing words, participants may not have a clear meaning in mind when they see these logographs simply because their meaning is not totally conventionalized or clearly derivable from their iconic properties (Miller et al., 2016). Our results seem to support this explanation.

It is not possible based on our results to answer in a conclusive manner the general question we posed at the beginning, as to whether emoji are processed like words during reading. We can, however, make a few tentative observations. One is that using object-denoting emoji as word replacements makes it more difficult for readers to process sentences irrespective of how frequent, well-known, or immediately interpretable the emoji is. English readers are not used to mixing alphabetic words and logographs in the way Japanese readers mix syllabic and logographic characters. From this perspective, it can be said that emoji are not processed in the same way as words are. It has been shown in Buchweitz, Mason, Hasegawa, and Just (2009) that reading of the syllabic Japanese hiragana and logographic kanji activated different brain areas. We may need to employ brain imaging techniques like fMRI or paradigms that measure brain responses like ERP to further investigate our question of whether emoji are processed like words and to study the way emoji are processed by people who use different writing systems for their native languages.

The other observation is that the processing of emoji is subject to semantic and contextual constraints similarly to what happens to words. In particular, our study is the first to our knowledge to have shown, based on eye tracking results, that emoji are subject to ambiguity effects that make it harder for readers to integrate them in the overall processing of a sentence when their meaning is not immediately clear. From this perspective, emoji seem to behave like balanced ambiguous (also called equi-biased) words that, as several studies have demonstrated (Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989), are harder to process because readers have to access several meanings and integrate the correct one in subsequent processing.

Our study has several limitations. One is the fact that our sample can be considered relatively small in spite of the fact that we increased the number of participants after a pilot study and the power calculations based on that study. Another limitation is due to the second-language background of the participants, which may have influenced their interpretations and responses in spite of their proficient command of English and the fact that the words we used are all frequent and easy ones. A third issue is the fact that we did not attempt to account for the different visual properties of the emoji involved in the stimuli, which may indeed have a strong influence on the ease with which they are processed. Yet, another issue concerns the eye-tracker used: Tobii T120 is not an optimal eye-tracker for reading research due to its low sampling rate. It would be interesting to replicate the experiments with better equipment, and maybe using a chinrest, to further minimize the possibility of inaccuracy and imprecision in the eye-tracking data.

Finally, more eye-tracking measures should be studied to get a more precise understanding of how the regions around the target are processed. In particular, looking at gaze time spent on the disambiguating context that follows the target emoji in the second set of stimuli, would be an important follow-up to our work. If both senses of an ambiguous target are activated as predicted by the lexical access model, there should be more processing time independent of whether the sense being used is the dominant one. If the dominant sense is more active, as implied by the reordering model, only the sentence in which the less dominant reading is used should receive higher processing time. According to Sereno et al. (1992), if a sentence contains biased ambiguous words, readers spend more time on the post-target region when it is consistent with the subordinate reading. An examination of the time spent on the T+1 region may shed light on which of these two models is the more adequate to deal with ambiguous emoji.

In conclusion, further research is needed to cast more light on the processing of emoji and the way they are represented in the mental lexicon, including investigating their use in different positions and in sequences; studying their interaction with other words in the same sentence or the text; applying different techniques such as fMRI and EEG; and looking at effects of language background, age, and culture from different user populations. We hope to have provided an initial contribution casting light on the fact that emoji, although apparently very popular with users of online messaging platforms, may not always be as immediate and easy to process for readers as one may think.

Acknowledgments

We would like to thank Albert Gatt and Sarah Grech from the University of Malta for their feedback. We would also like to acknowledge the reviewers for their useful comments and suggestions. Finally, we are indebted to the participants, who took part in the experiments on a voluntary basis.

Funding

Department of Nordic Studies and Linguistics, University of Copenhagen

Notes

1 https://www.unicode.org/reports/tr51/
2 https://languages.oup.com/word-of-the-year/2015/
3 The web interface is available at https://jakewestfall.shinyapps.io/two_factor_power/.
4 The calculator only allows for two random effects and a two-level main condition. Therefore, the power calculations are only indicative.
5 For example, a minimum total score of 83 in the TOEFL iBT test.
6 See https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqhow-do-i-interpret-a-regression-model-when-some-variables-are-log-transformed/ foradetailedexplanationofthisprocedure.

References

Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E-Z reader's null predictions. *Journal of Memory and Language*, *85*, 76–87.

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, *32*(8), 704–723.

Azuma, J., & Ebner, M. (2008). A stylistic analysis of graphic emoticons: Can they be candidates for a universal visual language of the future? In Luca, J., & Weippl, E. R. (Eds.), *Proceedings of ED-MEDIA 2008–World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 972–979). Vienna, Austria: Association for the Advancement of Computing in Education (AACE).

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Barbieri, F., Kruszewski, G., Ronzano, F., & Saggion, H. (2016). How cosmopolitan are emojis? Exploring emojis usage and meaning over different languages with distributional semantics. In *Proceedings of the 24th ACM International Conference on Multimedia*, MM '16 (pp. 531–535). New York, NY: Association for Computing Machinery.

Barbieri, F., Marujo, L., Karuturi, P., Brendel, W., & Saggion, H. (2018). Exploring Emoji usage and prediction through a temporal variation lens. *arXiv e-prints*, arXiv:1805.00731.

Barbieri, F., Ronzano, F., & Saggion, H. (2016). What does this emoji mean? A vector space skip-gram model for twitter emojis. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 3967–3972). ELRA (European Language Resources Association).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bolger, D. J., Perfetti, C. A., & Schneider, W. (2005). Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping*, *25*(1), 92–104.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, *1*(1).

Buchweitz, A., Mason, R. A., Hasegawa, M., & Just, M. A. (2009). Japanese and English sentence reading comprehension and writing systems: An fMRI study of first and second language effects on brain activation. *Biling (Camb Engl)*, *12*(2), 141–151.

Coderre, E., Filippi, C., Newhouse, P., & Dumas, J. (2008). The stroop effect in kana and kanji scripts in native Japanese speakers: An fMRI study. *Brain and Language*, *107*, 124–132.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210.

Donato, G., & Paggio, P. (2017). Investigating redundancy in emoji use: Study on a twitter based corpus. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 118–126).

Donato, G., & Paggio, P. (2018). Classifying the informative behaviour of emoji in microblogs. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018).*

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621–636.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). Swift: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813.

Gerstenberg, T. (2021). *Psych 252: Statistical methods for behavioral and social sciences*. Stanford University. https://psych252.github.io/psych252book/(visited 2021-02-10).

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental l2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483–517.

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, *42*, 279–297.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology. Human Perception and Performance*, *36*, 1294–1313.

Howman, H. E., & Filik, R. (2020). The role of emoticons in sarcasm comprehension in younger and older adults: Evidence from an eye-tracking experiment. *Quarterly Journal of Experimental Psychology*, *73*(11), 1729–1744. PMID: 32338575.

Ingram, J., & Hand, C. J. (2020). Words from the wizarding world: Fictional words, context, and domain knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2179–2192.

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, *23*(5), 612–624.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601–625.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kennedy, A., Pynte, J., Murray, W. S., & Paul, S.-A. (2013). Frequency and predictability effects in the Dundee corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, *66*(3), 601–618. PMID: 22643118.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262–284.

Kumle, L., Vo, M. L.-H., & Draschkow, D. (2020). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*, 2528–2543.

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.6.2-1.

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. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 770–780).

McBride-Chang, C., Chung, K. K., & Tong, X. (2011). Copying skills in relation to word reading and writing in Chinese children with and without dyslexia. *Journal of Experimental Child Psychology*, *110*(3), 422–433.

Miller, H. J., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). "Blissfully happy" or "ready to fight": Varying interpretations of emoji. In *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM 2016)*.

Mori, Y. (2014). Review of recent research on kanji processing, learning, and instruction. *Japanese Language and Literature*, *48*(2), 403–430.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*(3), 225–236.

Ousterhout, T. (2017). Investigation of the semantic priming effect with the N400 using symbolic pictures in text. In *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016), Copenhagen, 29–30 September 2016* (pp. 74–84). Linköping University Electronic Press, Linköpings Universitet.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K. (2009). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 779–790.

Rayner, K., & Morris, R. K. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. B. Simpson (Ed.), *Understanding Word and Sentence*, volume 77 of *Advances in Psychology* (pp. 175–198). Amsterdam: North-Holland.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton Jr, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*(3–4), SI21–SI49.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–476.

Robus, C. M., Hand, C. J., Filik, R., & Pitchford, M. (2020). Investigating effects of emoji on neutral narrative text: Evidence from eye movements and perceived emotional valence. *Computers in Human Behavior*, *109*, 106361.

Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language*, *83*, 118–139.

Sereno, S., Hand, C., Shahid, A., Yao, B., & O'Donnell, P. (2018). Testing the limits of contextual constraint: Interactions with word frequency and parafoveal preview during fluent reading. *Quarterly Journal of Experimental Psychology*, *71*(1), 302–313.

Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science*, *3*(5), 296–301.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General*, *140*(3), 407–433.

Tan, L. H., Liu, H.-L., Perfetti, C. A., Spinks, J. A., Fox, P. T., & Gao, J.-H. (2001). The neural system underlying Chinese logograph reading. *Neuroimage*, *13*(5), 836–846.

Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., & Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proceedings of the National Academy of Sciences*, *102*(24), 8781–8785.

Thuy, D., Matsuo, K., Nakamura, K., Toma, K., Oga, T., Nakai, T., Shibasaki, H., & Fukuyama, H. (2004). Implicit and explicit processing of kanji and kana words and non-words studied with fMRI. *NeuroImage*, *23*, 878–89.

Tigwell, G. W., & Flatla, D. R. (2016). Oh that's what you meant! Reducing emoji misunderstanding. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 859–866).

Tzeng, O. J., & Wang, W. S.-Y. (1983). The first two R's: The way different languages reduce speech to script affects how visual information is processed in the brain. *American Scientist*, *71*(3), 238–243.

Veldre, A., & Andrews, S. (2017). Parafoveal preview benefit in sentence reading: Independent effects of plausibility and orthographic relatedness. *Psychonomic Bulletin & Review*, *24*(2), 519–528.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578.

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.

Wang, Y., McBride-Chang, C., & Chan, S. F. (2014). Correlates of Chinese kindergarteners' word reading and writing: The unique role of copying skills? *Reading and Writing*, *27*(7), 1281–1302.

Weissman, B. (2019). Emojis in sentence processing: An electrophysiological approach. In *Companion Proceedings of the 2019 World Wide Web Conference* (pp. 478–479).

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv preprint arXiv:1308.5499.

Wiseman, S., & Gould, S. J. J. (2018). Repurposing emoji for personalised communication: Why 🍕 means "I love you". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18 (pp. 1–10), New York, NY: Association for Computing Machinery.

Appendix A: Stimuli experiment 1

1. a. John brought a pizza yesterday
   John brought a 🍕 yesterday
   John brought a pizza 🍕 yesterday
   b. Ann bought a pizza from the shop
   Ann bought a 🍕 from the shop
   Ann bought a pizza 🍕 from the shop
2. a. Ann bought a football from the shop
   Ann bought a ⚽ from the shop
   Ann bought a football ⚽ from the shop
   b. Marc sold a football last week
   Marc sold a ⚽ last week
   Marc sold a football ⚽ last week
3. a. Marc sold a plane last week
   Marc sold a ✈️ last week
   Marc sold a plane ✈️ last week
   b. Kim lost a plane in the street
   Kim lost a ✈️ in the street
   Kim lost a plane ✈️ in the street
4. a. Kim lost a balloon in the street
   Kim lost a 🎈 in the street
   Kim lost a balloon 🎈 in the street
   b. Sue got a balloon the other day

Sue got a 🎈 the other day
Sue got a balloon 🎈 the other day

5. a. Sue got a camera the other day
Sue got a 📷 the other day
Sue got a camera 📷 the other day
b. Tom wants a camera for Christmas
Tom wants a 📷 for Christmas
Tom wants a camera 📷 for Christmas
6. a. Tom wants a gift for Christmas
Tom wants a 🎁 for Christmas
Tom wants a gift 🎁 for Christmas
b. Eve found a gift this morning
Eve found a 🎁 this morning
Eve found a gift 🎁 this morning
7. a. Eve found a flower this morning
Eve found a 🌸 this morning
Eve found a flower 🌸 this morning
b. Jeff drew a flower in class
Jeff drew a 🌸 in class
Jeff drew a flower 🌸 in class
8. a. Jeff drew a rainbow in class
Jeff drew a 🌈 in class
Jeff drew a rainbow 🌈 in class
b. Liz saw a rainbow an hour ago
Liz saw a 🌈 an hour ago
Liz saw a rainbow 🌈 an hour ago
9. a. Liz saw a ring an hour ago
Liz saw a 💍 an hour ago
Liz saw a ring 💍 an hour ago
b. Joey put a ring on the table
Joey put a 💍 on the table
Joey put a ring 💍 on the table
10. a. Joey put a cake on the table
Joey put a 🎂 on the table
Joey put a cake 🎂 on the table
b. John brought a cake yesterday
John brought a 🎂 yesterday
John brought a cake 🎂 yesterday

Appendix B: Stimuli experiment 2

1. a. Jack ate 🥣 for dinner
Jack ate soup 🥣 for dinner

b. Jack ate 🥣 for breakfast
Jack ate cereals 🥣 for breakfast

2. a. Mia makes ☕ with organic beans
Mia makes coffee ☕ with organic beans

b. Mia makes ☕ with green leaves
Mia makes tea ☕ with green leaves

3. a. Roy loves 🍟 and eats there
Roy loves McDonald's 👑 and eats there

b. Roy loves 🍟 and eats them
Roy loves chips 🍟 and eats them

4. a. Jean knows about the 👑 jewels
Jean knows about the crown 👑 jewels

b. Jean knows about the 👑 of Denmark
Jean knows about the queen 👑 of Denmark

5. a. They were 👫 and played together
They were friends 👫 and played together

b. They were 👫 and lived together
They were lovers 👫 and lived together

6. a. Liam got a 💼 as a consultant
Liam got a job 💼 as a consultant

b. Liam got a 💼 made of leather
Liam got a briefcase 💼 made of leather

7. a. Sam bought a ⛺ for two people
Sam bought a tent ⛺ for two people

b. Sam bought a ⛺ holiday online
Sam bought a camping ⛺ holiday online

8. a. Ivy wants her 🎄 to be tall
Ivy wants her Xmas tree 🎄 to be tall

b. Ivy wants her 🎄 to be quiet
Ivy wants her Xmas 🎄 to be quiet

9. a. Lyn looked at the ☔ in the shop
Lyn looked at the umbrella ☔ in the shop

b. Lyn looked at the ☔ fall
Lyn looked at the rain ☔ fall

10. a. James said 🙏 for the gift
James said thanks 🙏 for the gift

b. James said 🙏 as he needed help
James said please 🙏 as he needed help

Appendix C: Fillers experiment 2

1. a. Lucy is 😊 today
Lucy is happy 😊 today

b. It is ☀️ today
It is sunny ☀️ today

2. a. The weather was ☃️ last winter
The weather was freezing ☃️ last winter

b. Sally made a ☃️ just then
Sally made a snowman ☃️ just then

3. a. Tim got a 🧲 at the shop
Tim got a magnet 🧲 at the shop

b. Tim is 🧲 by Lily
Tim is attracted 🧲 by Lily

4. a. The sea was very 🌊 last night
The sea was very stormy 🌊 last night

b. There was a 🌊 last night
There was a tsunami 🌊 last night

5. a. Ken tore his 👕 accidentally
Ken tore his t-shirt 👕 accidentally

b. Pat washed her 👕 finally
Pat washed her t-shirt 👕 finally

Appendix D: Survey

1. Year of birth (short text answer)
2. Your native language (short text answer)
3. Your preferred language when messaging (short text answer)
4. Your knowledge of English (none/beginner/intermediate/advanced/native)
5. How often do you encounter emojis in the messages you receive? (never/occasionally/often)
6. How often do you use emojis when you write? (never/occasionally/often)
7. Which mobile operating system are you most familiar with? (iOS (Apple)/Android/Others/I don't know)
8. Which of these usages do you prefer for how an emoji is inserted in a message?
   a. Instead of a word, as in "She drove the 🚗 to work".
   b. Together with a word as in "She drove the car 🚗 to work"
   c. At the end of the message as in "She drove the car to work 🚗"
9. Please judge for each emoji meaning listed below how acceptable it is for you on a scale from 1 (not acceptable) to 5 (highly acceptable).
   🥣 = cereals
   🥣 = soup
   … (all meanings of ambiguous emoji used in experiment 2 shown one by one)
10. Do you have any comments you would like to share with us? (text answer)

In the survey, 29 participants gave Danish as their native language, while five gave English, four Chinese, three Russian, two German or Icelandic, and one Bulgarian, Croatian, Hebrew, Norwegian, Polish, Romanian, or Spanish. The last participant is native in German and English. Regarding the language they use when messaging, 24 of them stated Danish, 18 of them use English, four responded “Danish or English,” one uses Icelandic, Chinese, Norwegian, Hebrew or Spanish, and one “English or German,” 1 “English or Chinese.”

Concerning emoji use, 79.2% of the participants stated that they often encounter emoji in the messages they receive, 18.9% of the participants responded “occasionally,” and only one of them responded “never;” 61.5% of the participants said that they often use emoji when they write, 32.1% of the participants said that they “occasionally” use emoji when they write, and only four of them (7.5%) said that they never use emoji when they write. Regarding the emoji usage practice, 43 of the participants (81.1%) prefer to use the emoji in sentence final position and as a redundant emoji. The remainder prefer to insert the emoji in the middle of the sentence, five (9.4%) replacing and five (9.4%) following the word.

Regarding the mobile system the participants use, 22 participants use Android and the rest (31 participants) use iOS. We asked this question since the rendering of the emoji is different from system to system. We use the iOS emoji in this study as the appearance of a particular emoji is the same across all iOS devices, while it can differ a bit across Android devices.