

Emojis: A Grapholinguistic Approach

Christa Dürscheid & Dimitrios Meletis

Abstract. The present article stands at the interface of CMC research and grapholinguistics. After outlining which features are typical of the writing of private text messages, the focus of the first part of the paper (Sections 2 and 3) lies on the use of emojis. Notably, emoji use is not—as is commonly done—analyzed under a pragmatic perspective, but grapholinguistically, at the graphetic and graphematic levels: emojis are conceptualized as visual shapes that may assume graphematic functions within a given writing system. In the second part (Section 4), it is underlined that all variants of written digital communication (such as the use of emojis, but also all other characters) are made possible only due to the Unicode Consortium's decisions; this, finally, is argued to have far-reaching consequences for the future of writing.

1. Preliminary Remarks

In this paper, the use of emojis will be considered within a framework known in the German-language research area as “Schriftlinguistik” (*grapholinguistics*). As will be demonstrated, this term is not equivalent to the terms *graphemics* or *graphematics*. In a much broader sense, grapholinguistics entails different aspects of writing (among them research on scripts and writing systems, the history of writing, orthography, graphematics, the acquisition of reading and writing, text design and text-image-relations, and differences between the written and spoken modalities of language) (cf. Dürscheid 2016).¹ This paper's main

Christa Dürscheid
Department of German Studies, University of Zurich
Schönberggasse 9, 8001 Zürich, Switzerland
duerscheid@ds.uzh.ch

Dimitrios Meletis
Department of Linguistics, University of Graz
Merangasse 70/III, 8010 Graz, Austria
dimitrios.meletis@outlook.com

1. To date, this textbook is only available in German (in its 5th edition).

Y. Haralambous (Ed.), *Graphemics in the 21st Century. Brest, June 13-15, 2018. Proceedings Grapholinguistics and Its Applications* (ISSN: in progress), Vol. 1. Fluxus Editions, Brest, 2019, p. 167–183. <https://doi.org/10.36824/2018-graf-duer>
ISBN: 978-2-9570549-0-9, e-ISBN: 978-2-9570549-1-6

focus will be on a certain phenomenon within this vast field of topics—the fact that texts are increasingly being enriched by images. These include emojis,² ASCII signs, stickers, GIFs, photos, and videos, i.e., different kinds of visual elements that Herring and Dainas (2017) subsume under the umbrella term *graphicons*. Among these graphicons, emojis constitute their own inventory of visual units. Not only is their number growing annually (at this point, there exist about 3,000, see <https://unicode.org/emoji/charts/emoji-counts.html> <31.08.2019>), but their use in everyday writing, for instance in WhatsApp messaging, is also on the rise. Unlike, for example, photos or videos, emojis function as an integrated part of text messages. They are situated on the same line as the other characters and often substitute them (cf. *I'll come by car* > *I'll come by 🚗*). These features give rise to the question whether emojis may become the basis of a new way of writing (or even a new language), a question never asked with regard to the other types of graphicons. This question is also motivated by the unique technical status of emojis: among graphicons, they are the only visual elements that are included in the Unicode Standard. Notably, the inclusion of each new emoji requires a well-elaborated proposal to the Unicode Consortium. However, once such a proposal is approved, the emoji in question can be inserted into texts like any other character (see Section 4).

The theoretical framework on which this paper is based will be discussed in the next section: we will present relevant research on computer-mediated communication (CMC) on the one hand and on grapholinguistics on the other. After that, a short overview of emoji research will be given. Here, the focus will shift towards the question of how emojis may be analyzed from a grapholinguistic point of view (Section 3). In this context, data from a Swiss project empirically investigating the use of emojis will provide insight into the various functions they fulfill in WhatsApp messages (cf. Ueberwasser and Stark 2017). While these functions can be explained from a pragmatic perspective (cf. Danesi 2016; Pappert 2017; Beißwenger and Pappert 2019; Dainas and Herring in press), the present paper will instead focus on the functions emojis fulfill at the graphematic level (cf. Dürscheid and Frick 2016, Dürscheid and Siever 2017). Section 4 will then address the question of which role the Unicode Consortium plays with respect to the use of emojis. How far-reaching are the consortium's decisions and what are the consequences of the (non-)inclusion of a graphic sign in the Unicode

2. As for the plural of *emoji*, the Oxford English Dictionary states that both variants, *emoji* and *emojis*, are allowed (see <https://www.oed.com/>). Interestingly, in 2016, Emojipedia, a famous website covering the use of emojis, revealed that, based on empirical data, the use of plural-*s* is increasingly popular (see <https://blog.emojipedia.org/emojis-on-the-rise-as-plural/> <30.09.2019>).

character set? A short reflection on the future of emoji use and related open questions will conclude the paper (Section 5).

2. Theoretical Background

Significant research on CMC is closely linked to the name of Susan Herring, Professor of Information Science and Linguistics at the Indiana University Bloomington, where she also founded and still directs the Center for Computer-Mediated Communication. Of her many works on the topic, one that is particularly worth mentioning is “Pragmatics of Computer-Mediated Communication,” a handbook she co-edited with two colleagues (cf. Herring, Stein, and Virtanen 2013). In his chapter, Markus Bieswanger compiles the most relevant features of writing in CMC and discusses them at both the grapholinguistic level and the stylistic level (cf. Bieswanger 2013). Bieswanger lists a bundle of typical writing features for CMC such as acronyms (*OMG*), letter and number homophones (*4you*), nonstandard spellings, and punctuation (*really???*). As far as the stylistic level is concerned, he describes, among other features, the accumulation of syntactic reductions and the use of colloquial expressions or dialectal elements. It is noteworthy that these features are used predominantly in private, informal everyday communication (e.g., messages in a WhatsApp family chat). Obviously, this means that not all types of texts on the internet exhibit these features. For example, to date, they hardly ever occur in texts directed at a large, anonymous readership (e.g. on university and company websites) or texts produced in the context of more formal one-to-one communication (e.g., business emails).³

While the features listed above are discussed in detail in both German and English research on CMC, a different approach is found predominantly in the German research tradition: Here, a terminological distinction is made between *medium*, *form of communication*, and *text genre* (cf. Dürscheid 2005). A letter of application, for instance, can be considered a special type of text (*text genre*) that may be sent as an email (*form of communication*) via computer or mobile phone (*medium*). However, the boundaries between these devices are increasingly blurred, as nowadays, mobile phones function almost identically to computers and can be used to write a range of significantly differing types of texts such as letters of application or Facebook postings (for example about one’s last holiday trip); these, ultimately, constitute texts from entirely different *text genres*. The term *form of communication* is used to describe the various

3. This applies to the first contact with customers. If emails are exchanged back and forth quickly, formalities may be abandoned to some extent. This is to say that the more dialogical a text becomes, the sooner the above-mentioned features occur.

communicative practices which are possible within these media. These include an email exchange, a telephone call, a text chat, or any other kind of interaction at the oral or the written level (cf. Jucker et al. 2018). *Text genre*, finally, refers to different communicative purposes that motivate these interactions and enable different types of written texts (or different types of oral conversations, respectively). Some examples for such text genres are (at the written level) business letters, love letters, letters of application, or holiday greetings. Among the given examples, it is predominantly the area of CMC research meeting the following criteria that is treated in this paper: texts which are mediated by smartphones and are part of an interpersonal exchange carried out in a private, informal setting. Consequently, *text genres* such as business letters are not taken into consideration here, and neither are more formal communications on LinkedIn or other social networks.

As mentioned above, we will concentrate on the analysis of the graphematic functions of emojis, which means that the following considerations are situated at the interface between CMC and grapholinguistics. The term *grapholinguistics* is used here instead of other alternatives such as *graphonomy* or *grammatology* which are meant to designate research on writing systems (cf. the numerous works of Peter T. Daniels and Florian Coulmas, for instance). One reason for insisting on grapholinguistics is that we need an expression that refers not exclusively to one research domain of written language but to *all* writing-related aspects (cf. Dürscheid 2016). Furthermore, the use of *grapholinguistics* is of programmatic character, highlighting that writing is by no means a secondary system subordinate to spoken language but instead a fully functional form of language in and of itself and must be examined in its own terms (cf. also Meletis 2019). Worth mentioning in this respect is a dictionary of “Schriftlinguistik” edited by Martin Neef, Said Sahel, and Rüdiger Weingarten. It is part of a series of online (and, later, printed) dictionaries covering various linguistic subfields (e.g., phonetics and phonology, word formation). While this project started out in German, the long-term plan is to also publish the dictionaries in English. The fact that grapholinguistics is a field included in this compilation of dictionaries indicates that its relevance in German linguistics has risen. This is also underlined by the fact that more and more research is being embedded in this framework (cf. Neef 2015; Meletis 2018; Dürscheid 2018).

Interestingly, since 2009, there has even been an entry on grapholinguistics in the German Wikipedia (see <https://de.wikipedia.org/wiki/Schriftlinguistik>, <30.09.2019>). The English Wikipedia, on the other hand, only includes an entry on *graphemics* but not *grapholinguistics*. It states that “graphemics or graphematics is the linguistic study of writing systems and their basic components, i.e., graphemes” (<https://en>.

wikipedia.org/wiki/Graphemics, <30.09.2019>).⁴ This gives readers the impression that *graphematics* encompasses all aspects of the study of writing systems which is, however, inaccurate: writing systems research deals with many more topics than graphematics—and grapholinguistics is still broader (cf. Meletis 2019, Chapter 2). Meletis distinguishes between *graphetics*, the study of the visual resources used in writing, and *graphematics*, the study of the relation between visual units (so-called “basic shapes”) and corresponding linguistic units (such as phonemes, syllables, morphemes). While graphetics treats all aspects of the materiality of writing (as, for example, the choice of typeface or the effect its appearance has on its processing by humans), its “main object of study is *scripts*, defined as inventories of discrete visuo-graphic basic shapes such as the Roman script, the Chinese script, and the Japanese inventories hiragana and katakana” (Meletis, 2018, p. 62). These scripts and the basic shapes they consist of—in the case of Roman script often referred to as ‘letters,’ in Chinese script as ‘characters,’ but cf. Meletis (in press)—are studied for their materiality alone, i.e., dissociated from any linguistic function they might assume in a given context. They are not bound to a given language and its respective writing system, which becomes obvious when considering that many of them—such as the Roman script and the Cyrillic script—are commonly used for more than one writing system (e.g., English, German, Dutch, Italian, and many more for Roman, and Russian, Ukrainian, Serbian, etc. for Cyrillic).

Following this view, a writing system, as the main object of study of graphematics, is the combination of a script and a language (cf. Weingarten 2011). Thus, for instance, the German writing system employs Roman script for the German language, the English writing system Roman for English, the Ukrainian writing system Cyrillic for Ukrainian. The inventory of punctuation signs could also be seen as a script, as could the inventory of digits. Both of these inventories are employed across an even wider range of writing systems than scripts such as Roman or Cyrillic; consider, for example, the comma which appears in very similar functions in many typologically diverse writing systems. Similarly, in our grapholinguistic approach, emojis constitute their own inventory of basic shapes and are used as communicative and sometimes genuinely graphematic resources whose functions are not specific to a given language or writing system, although this would have to be tested in a comparative typological study. The different facets of emoji use will be explored in the following section.

4. Note that on the website of the conference at which a part of this paper was presented, both terms are also used as synonyms: while the conference title was “/grafe-matik/,” its subtitle was “Graphemics in the 21st century” (see <http://conferences.telecom-bretagne.eu> <25.09.2019>).

3. Emojis and Their Use

In the last years, we witnessed a rise in works on emojis specifically from linguistic and semiotic perspectives. The following selection of titles is thus only supposed to give a first idea of the current state of research in this field: Marcel Danesi's book "The Semiotics of Emoji" (2016) distinguishes between emoji semantics, emoji grammar, and emoji pragmatics, demonstrating the use of emojis from these different perspectives. Susan Herring and Ashley Dainas examine different types of graphicons (among them emojis) sampled from public Facebook groups and analyze their frequency as well as their pragmatic functions (cf. Herring and Dainas 2017). According to their findings, emojis may serve, for instance, to express feelings or to clarify the communicative intention of an utterance (as a kind of "tone modification"). Another article (cf. Dainas and Herring in press) presents an emoji survey "administered online in early 2018 to determine how social media users interpret the pragmatic functions of popular emoji types". The abstract from which this quote is taken concludes with the assertion of "the importance of analyzing emoji meaning from the perspective of pragmatics". A concise monograph that strongly emphasizes this aspect has just been published in German and is titled "Handeln mit Emojis" (Beißwenger and Pappert, 2019).⁵ In it, the authors distinguish two main strategies of emoji use: *making readable* ("Lesbarmachen") and *making visible* ("Sichtbarmachen"). *Making readable* refers to using emojis in order to provide readers with information on how to interpret an utterance, while the goal of *making visible* is visually framing an utterance (cf. *ibid.*, pp. 71–73).

While all of the above-mentioned works are grounded in semiotic or pragmatic approaches, Dürscheid and Siever (2017) focus on the grapholinguistic functions that emojis fulfill. Graphetically, they can be used as visual units to separate sentences from each other (instead of a period or a comma) or to indicate the end of the message, and graphematically, they can be functionalized in order to substitute a single grapheme or a sequence of graphemes. Note that this structural analysis of emojis does not compete with the determination of their communicative functions but instead complements the pragmatic approach with a different perspective. This is illustrated in Fig. 1, a text message sent along with a photo.

As is evident from this example, emojis are positioned on the same line as characters and are approximately equal in size. The photo, on the contrary, is presented separately. Although it is semantically connected with the text, it is not positioned within the text, but on top of it. The text itself consists of a short sentence followed by five sun emo-

5. The English translation is (the authors' own suggestion): "How to do things with emojis".



FIGURE 1. Text message with emojis

jis which likely imply that the sun is shining wherever the message was composed. It is also possible, however, that the writer used the sun emoji only in order to render the message a little more cheerful and colorful (cf. Dürscheid and Frick 2016), i.e., with no intention at all of making a statement about the current weather situation. Of course, it is also possible that the writer wanted to combine these two functions. Irrespective of these considerations, it must be noted that the five emojis in 1 do by no means stand for the word *sun*, which is to say that they are not used logographically. If this were the case, the text would have to be read as *The beach says hi sun sun sun sun sun*, and it is highly unlikely that this was the writer's intention. Thus, in this example, the sun emoji is used merely as a graphetic resource that does not assume any linguistic function, i.e., it does not refer to any specific linguistic unit. Irrespective of this, it does of course have a context-sensitive communicative function.

In the following, however, we will show that emojis, similar to other basic shapes, can be graphematically functionalized in order to refer to different linguistic levels: In the word *month*, for example, the sun emoji may replace the <o>, i.e., be used as an allograph of the letter <o>. For the word *frontdoor*, an emoji representing a door can be used to replace <door> (*front* 🚪). The emoji in the sentence *Shall we build a 🧑‍🌾 today?* in which it substitutes the word *snowman* functions similarly. If the writer were to also omit the article in this sentence, the emoji would even substitute an entire noun phrase (i.e., *a snowman* or *the snowman* or *our snowman*). As this example shows, interpreting sentences in which an emoji substitutes a noun phrase might produce a number of different readings. Technically, in the last two examples, emojis function as ideograms.⁶

6. In this vein, emojis are similar not only to digits but also to other special characters such as <%> or <&>.

The concept of ideography—at least the question of whether it constitutes writing—has been under a lot of scrutiny (cf. Unger 1990), and today, it is common consensus that ideograms are not considered writing since in its narrow definition, writing is interpreted only as the graphic representation of *specific* linguistic units (cf. Daniels 2018, p. 157). Following this, only *glottography*, i.e., ‘language writing,’ is considered writing, contrary to what is known as *semasiography*, referring to visual units that represent concepts or ideas (cf. *ibid.*, p. 126). While glottography can be read, i.e., decoded directly, semasiography can only be interpreted but never read since no specific linguistic units are associated with the visual shapes.

Returning to our example from above, it is not clear how the snowman emoji would be spelled out. This means that strictly speaking, emojis are ideograms, visual resources used with a communicative function and a meaning, but they are not writing *proper*. This, however, would restrain us from analyzing them in a grapholinguistic approach given that grapholinguistics is only invested in the study of writing. The solution is that while at the formal level, emojis are special characters, graphematically they may be ideograms, and in some uses even logograms: when decoding a given written utterance in which an emoji substitutes a phrase, a word, or a morpheme, the reader commonly decides for one *specific* phrase, word, or morpheme, respectively, in order to read the utterance. This is, for instance, the case in the example given in the preliminary remarks above, *I’ll come by car > I’ll come by 🚗*. Here, the emoji is associated with a specific linguistic unit—in our terms, it is used graphematically. Note that such an association is often fluid and not only different individuals might associate emojis with different linguistic units, but also the same individual might read the same emoji in different ways depending on various contextual factors.

Authentic examples of the different uses of emojis both at the segmental and suprasegmental graphematic levels are presented in Dürscheid and Siever (2017).⁷ This paper also provides data on how often emojis are used in WhatsApp chats and which of them are the most popular. This analysis is based on a research project on WhatsApp communication in Switzerland (see <http://www.whatsup-switzerland.ch> <30.09.2019>). The data were collected in 2014, and the text corpus consists of around 750,000 messages for which sociodemographic information is also available (age, gender, mother tongue, etc.). In Fig 2, one example chosen from this corpus will be presented.

The text in the example approximately translates to “I already went __ today and the weather is nice for once,” where the underscore indi-

7. A short English version of this paper, titled “Beyond the alphabet—communication with emojis,” can be found on www.academia.edu.

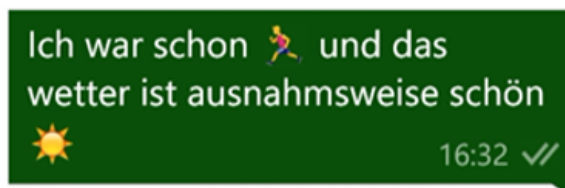


FIGURE 2. Graphematic use of the ‘Person Running’ emoji

cates the first emoji⁸ which replaces a graphematic word, i.e., a sequence of graphemes—the question, now, is which one, as this emoji (its name on Emojipedia is given as “Person Running,” its Unicode code point is U+1F3C3) can represent a variety of verbal expressions: *running*, *on the run*, *walking*, or *jogging*, to name a few. Thus, as illustrated in Fig. 3, the emoji as a visual unit is the *signifier* of different morphemes which are themselves, in the sense of Saussure, bilateral signs. This renders the emoji graphematically ambiguous, as the specific linguistic unit it refers to is not fixed but variable and determined by the context or the reader’s interpretation of a given text in which it is used. Note that the global concept the emoji represents—in the case of Fig. 3 the common concept underlying the words *run*, *walk*, *jog*, and others—is relatively constant and allows the emoji to be interpreted irrespective of the given context. When used to substitute morphemes, words, phrases, etc. within written sentences, emojis become graphematic units as they are treated by readers as sequences which are *read* instead of being only *interpreted*.

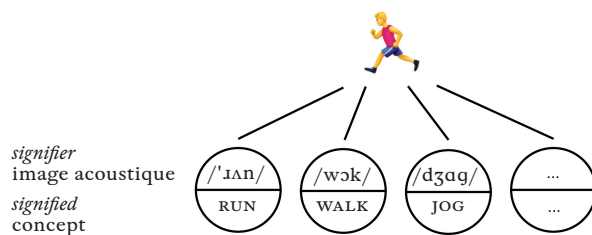


FIGURE 3. Example of the representation of an emoji within Saussure’s sign model

Another interesting example of emoji use can be found on the Twitter page of a Police Department located in the heart of St. Louis County

8. The second emoji is used probably in the same way as the sun emojis in Fig. 1.

(see <https://twitter.com/CreveCoeurPD>).⁹ On a regular basis, Creve Cœur Police post tweets in which security announcements are given concerning residents' properties and which have the goal of increasing public safety. These tweets imply that the use of emojis is no longer restricted to private everyday communication. However, it can still be assumed that the respective form of communication (in this case Twitter) continues to play an important role in how a text is structured; it is less likely that texts such as these tweets appear in print flyers, for example.

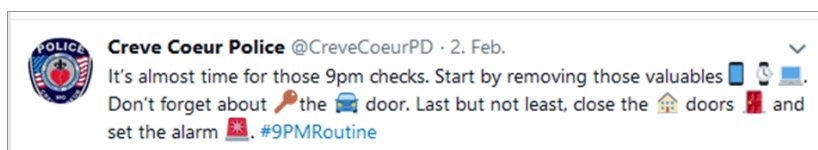


FIGURE 4. Emoji use in a non-private setting

In Fig. 4, the key emoji can be read as the verb *lock*, the car emoji as the noun *car*. The question arises as to how often emojis are used with such a logographic function. The data from the Swiss WhatsApp project suggest that this is actually only the case for a small number of all instances of emoji use. However, the text messages included in the corpus were collected in 2014; at that time, *Emoji Prediction* was not yet available to most writers. This software feature, introduced in 2014 on iOS and in 2016 on Windows Phone 8, facilitates the inclusion of emojis. The writer no longer needs to scroll through the list of emojis in order to find a suitable emoji, as context-related image suggestions are presented analogously to word suggestions. This also serves to highlight the substantial impact that technology exerts on writing.¹⁰ It is only due to the fact that emojis are available in the Unicode character set that we write with them today, and it is only because software presents emoji suggestions that they are increasingly functionalized logographically. This brings us to the next topic—the relation between Unicode and grapholinguistics.

9. Thanks to Marc Wilhelm Küster for bringing this to our attention. Cf. also Küster (in this volume).

10. The same was true for former SMS communication (cf. Bieswanger 2013). Certain writing strategies such as word junctions without spaces (*ShallWeMeetThisWeek-end?*) were established because of the 160 character limit of text messages.

4. Unicode and Grapholinguistics

In this section, the influence of the Unicode Consortium on writing will be discussed. In this context, it will also be argued that grapholinguists should have a say in the Unicode Consortium. Currently, the Consortium consists of about 20 people; Mark Davis co-founded it in 1991 and is its long-time president. All of the major IT companies (Apple, Microsoft, IBM, Facebook, Adobe, etc.) are “full members”. Additionally, there are three “institutional members” (e.g., the University of California in Berkeley) and two supporting members who also have the right to vote.¹¹ The members’ main task is to check applications for the admission of new characters and to make a preselection of these characters on which they vote in an annual meeting. The Consortium thus functions as a kind of gatekeeper (cf. Dürscheid 2018).

In the following, a short passage of the Unicode website is presented. This quote stresses the relevance of having a character coding system that facilitates the smooth exchange of data:

The Unicode Standard provides a unique number for every character, no matter what platform, device, application or language. It has been adopted by all modern software providers and now allows data to be transported through many different platforms, devices and applications without corruption. Support of Unicode forms the foundation for the representation of languages and symbols in all major operating systems, search engines, browsers, laptops, and smart phones—plus the Internet and World Wide Web [...].

<http://www.unicode.org/standard/WhatIsUnicode.html> <29.08.2019>

All Unicode characters (currently approx. 139,000) have a specific name (e.g., GREEK SMALL LETTER A) and are encoded with a numerical value. However, the concrete graphic realization that is finally assigned to a given Unicode code point and that appears on the device that is used to display the character depends on the specific font that is being used. Note, for example, how the respective sun emojis in Fig. 1 and 2 differ—even if just in details—with respect to their form and their color. They are concrete visual instantiations of the same basic shape. In this vein, from a grapholinguistic perspective, the vast majority¹² of Unicode characters are basic shapes that may be “embodied as graphs (sometimes referred to as *glyphs*), concrete physical instantiations” (Meletis, 2018, p. 63).

11. A complete list of the members is available at <http://www.unicode.org/consortium/members.html> <29.08.2019>.

12. There exist some Unicode characters which do not have a visual representation such as the soft hyphen character which marks boundaries between written syllables (cf. Haralambous and Dürst in this volume).

Unicode's predecessor was the ASCII character set ("American Standard Code for Information Interchange"). It originated in the 1960s, initially comprised only 64 characters and was eventually extended by one bit to 128. This, of course, led to various problems such as the faulty representation of characters from non-Roman scripts (such as Cyrillic). Moreover, many special characters from the Roman script could also not be represented correctly.¹³ Thus, for example, when sending emails including German umlauts or the sharp s (i.e., <ß>) in international correspondence, the German closing formula for *Best regards* could become *Sch%ne Gr&ße* (instead of *Schöne Grüße*) or instead of the <é> in the word *variété*, only an empty box could appear.

These times, however, belong to the past. In the long term, the goal of the Unicode Consortium is to integrate all scripts from the past and the present into the Unicode Standard. While the decision to include scripts currently in use appears self-evident, the question of why historical scripts should have a place in Unicode is justified. Consider, for example, a person wanting to write an article about cuneiform characters and to then publish it on the internet; without respective Unicode values, that person would have the option of inserting the cuneiform characters into the respective document as images. This is complicated and cumbersome; it is much easier and straightforward to type in the Unicode value of respective characters. Moreover, if image files were used, people who search the internet for articles on cuneiform characters would not find them with the aid of search engines. This goes to show that there are good reasons to include old scripts as well, which is how in Unicode, Egyptian hieroglyphics stand next to Germanic runes—to name just two examples. However, many historical scripts (e.g., Rongorongo) are still missing, as well as some scripts that are currently being used only by a small minority. These are listed on the website of the Script Encoding Initiative (SEI), a research project at the University of California at Berkeley (see below).

Obviously, every decision to include a new character in the Unicode Standard needs to be carefully examined since once a given character is added, it cannot be removed. This leads to the issue of the inclusion of emojis in Unicode and specifically the following question: Which criteria are crucial for an emoji coding proposal to be accepted or rejected? A page titled "How to Submit Proposal Documents" contains detailed information on this topic and lists points for and against accepting emoji proposals.¹⁴ For example, an emoji for a local food that is unknown in

13. The following passages are taken partially from an article that appeared in German under the title "Bild, Schrift, Unicode" (cf. Dürscheid 2018).

14. See http://unicode.org/emoji/proposals.html#selection_factors <30.09.2019>. Note that in July 2019, the Unicode Consortium launched a new website in celebration of the world emoji day (see <https://home.unicode.org/>

other regions (e.g., Swiss “Käsespätzle”) is rather unlikely to be included. Another important criterion is the assumed frequency of a prospective emoji: the emoji must represent something that is either in use world-wide or which is at least particularly frequent in a certain population group. Furthermore, an important criterion is whether it can be used in a sentence. This also explains why the emoji inventory contains so many characters that represent concrete things, such as sports equipment, means of transport, animals, and plants. If integrated into sentences (e.g., *I am 🐶, I love 🌻*), these for the most part culturally unspecific emojis are easy to decode for any reader.

An application for the inclusion of a new emoji can be submitted at any time. However, from application to final decision, up to two years can pass. Unsurprisingly, such a long-awaited decision is thus always expected with great excitement. Every year in June, the Unicode Consortium becomes the center of attention when it finally announces the new emojis to be introduced. Consider a small selection of headlines from the first half of 2019 (all accessed on 28.08.2019) which highlights the media’s and public’s interest in the introduction of new emojis:

New Emojis Are Coming: Interracial Couples, Guide Dogs, Falafel and More

<https://www.nytimes.com/2019/02/06/technology/new-emoji.html>

Disability emojis: Guide dog and wheelchair user revealed

<https://www.bbc.com/news/newsbeat-48989950>

One Woman Wants To Create This: *Insert Afro Emoji Here*

<https://www.npr.org/2019/03/31/708537582/one-woman-wants-to-create-this-insert-afro-emoji-here?t=1567008113435>

Unicode emoji 12.0: Waffles, otters and period positivity

<https://www.livemint.com/mint-lounge/features/unicode-emoji-12--0-waffles-otters-and-period-positivity-1550208051560.html>

For a long time, the work of the Consortium did not receive this kind of attention. It was, in fact, in 2010, precisely when emojis were included in the Unicode Standard, that this suddenly changed, which is also acknowledged on the Unicode website: “Emoji were adopted into the Unicode Standard in 2010 in a move that made the characters available everywhere. Today, emoji have been used by 92% of the world’s online population. And while emoji encoding and standardization make up just one small part of the Consortium’s text standards work, the growing popularity and demand for emoji have put the organization in the international spotlight.”¹⁵ This underscores not only Unicode’s importance

the-unicode-consortium-launches-new-website-in-celebration-of-world-emoji-day-2/<30.09.2019>). As noted in the press release, this website “will make information about the emoji proposal process more easily accessible while encouraging public participation and engagement in all Unicode initiatives”.

15. See <https://home.unicode.org/the-unicode-consortium-launches-new-website-in-celebration-of-world-emoji-day-2/<08.10.2019>>.

for global data exchange, but also that grapholinguists need to consider the consortium's work when investigating the impact of emojis on communication.

It is also worth noting that the decision to include emojis in the Unicode character set in the first place was certainly not an easy one. On the one hand, there were practical reasons in favor of their inclusion: emojis had already been used millions of times on Japanese mobile phones and large IT companies insisted on the need for globally uniform coding. On the other hand, the question arose whether emojis might only be a trend that would subside in a few years from then. Another question was whether images should be included in Unicode at all. And these questions only raise additional questions, including: Which criteria should be used in decision-making; which proposals should be accepted and which should be rejected? These are exactly the questions that Mark Davis, co-founder of the Unicode Consortium, addressed when he gave an interview in the Swiss newspaper *NZZ am Sonntag*.¹⁶ Every year, far more coding proposals are submitted than can be accepted, making a strict selection crucial. However, it is doubtful whether enough linguistic expertise is consulted when the consortium discusses these decisions.

This brings us to the point that is also advocated in Dürscheid's German publications. As mentioned above, the Unicode Consortium includes the representatives of all major internet companies (e.g., Adobe, Apple, Microsoft, Google) as full members and some additional institutional and supporting members. Among these, there is currently merely one researcher in linguistics: Dr. Deborah Anderson from the University of California. This should definitely change; linguists should have a lively interest in working on the future of the Unicode character set and the question of which basic shapes should be added (and which not). As for Deborah Anderson's background, she is a member of the Script Encoding Initiative (SEI), established in 2002, which is devoted to the preparation of proposals for the encoding of scripts in Unicode. As pointed out on its website, the SEI advocates the inclusion of minority and historic scripts into Unicode:

For a minority language, having its script included in the universal character set will help to promote native-language education, universal literacy, cultural preservation, and remove the linguistic barriers to participation in the technological advancements of computing. For historic scripts, it will serve to make communication easier, opening up the possibilities of online education, research, and publication.

<http://www.linguistics.berkeley.edu/sei/index.html> <30.09.2019>

16. See <https://nzzas.nzz.ch/gesellschaft/emojis-nachricht-mit-gefuehl-ld.1336511> <08.10.2019>.

Since today, in many literate societies, almost all reading and writing occurs digitally, it is essential to pay attention to the work of the Unicode Consortium, or even better: to participate in it. Linguists, and especially grapholinguists, must be actively involved in deciding what direction this process takes in the future. Not only do linguists have valuable insight into questions concerning the use of written language, but specialists in the field are also aware of the far-reaching sociolinguistic consequences of the introduction of digital writing in a given community.

5. Outlook

At the end of this paper, many questions remain unanswered: What will be the future role of the Unicode Consortium regarding the adoption of new characters? And what will be the future of emojis? Are they just a trend that will eventually disappear? In this vein, it must be noted that thanks to Unicode, it is to be expected that the number of emojis will increase continuously. However, it is also possible that new technologies will emerge that could make emojis obsolete. For example, voice messages might replace text messages and thus make emojis irrelevant. In any case, it will be interesting to observe how the relationship between image and writing will develop further and which graphics will still be used in the communication practices of the future. Moreover, it would be interesting to carry out another data collection of WhatsApp messages, comparing the new results with formerly described emoji practices in WhatsApp. Such a comparison of how writers employ emojis at various points in time might indicate that the frequency of emoji use is diachronically growing. And given the optimization of *Emoji Prediction*, emojis might also be increasingly used as logograms. Finally, it would be interesting to investigate whether emojis are on the rise also in contexts in which they were formerly not commonly used, for example in text genres such as business letters or on social media channels of universities, churches, museums, etc., i.e. in non-private settings. When considering the respective Instagram, Twitter and Facebook pages of such institutions, this already seems to be the case (see for the University of Zurich, for instance, <https://www.facebook.com/uzh.ch/>).

References

- Beißwenger, Michael and Steffen Pappert (2019). *Handeln mit Emojis. Grundriss einer Linguistik kleiner Bildzeichen in der WhatsApp-Kommunikation*. Duisburg, Essen: Universitätsverlag Rhein Ruhr.

- Bieswanger, Markus (2013). "Micro-Linguistic Structural Features of Computer-Mediated Communication". In: *Pragmatics of Computer-Mediated Communication*. Ed. by Susan C. Herring, Dieter Stein, and Tuija Virtanen. Berlin, Boston: de Gruyter, pp. 463–485.
- Dainas, Ashley R. and Susan C. Herring (in press). "Interpreting Emoji Pragmatics". In: *Internet Pragmatics: Theory and Practice*. Ed. by C. Xie, F. Yus, and H. Haberland. Amsterdam: John Benjamins.
- Danesi, Marcel (2016). *The Semiotics of Emoji. The Rise of Visual Language in the Age of the Internet*. London: Bloomsbury Publishing.
- Daniels, Peter T. (2018). *An Exploration of Writing*. Sheffield: Equinox.
- Dürscheid, Christa (2005). "Medien, Kommunikationsformen, kommunikative Gattungen". In: *Linguistik online* 22/1.
- (2016). *Einführung in die Schriftlinguistik*. 5th ed. Göttingen: Vandenhoeck & Ruprecht.
- (2018). "Bild, Schrift, Unicode". In: *Sprache—Mensch—Maschine. Beiträge zu Sprache und Sprachwissenschaft, Computerlinguistik und Informationstechnologie für Jürgen Rolsboven aus Anlass seines sechsundsechzigsten Geburtstages*. Ed. by Guido Mensching et al. Cologne: KUPS, pp. 269–285.
- Dürscheid, Christa and Karina Frick (2016). *Schreiben digital. Wie das Internet unsere Alltagskommunikation verändert*. Stuttgart: Kröner.
- Dürscheid, Christa and Christina M. Siever (2017). "Jenseits des Alphabets. Kommunikation mit Emojis". In: *Zeitschrift für Germanistische Linguistik* 45.2, pp. 256–285.
- Haralambous, Yannis and Martin Dürst (in this volume). "Unicode from a Linguistic Point of View".
- Herring, Susan C. and Ashley R. Dainas (2017). "Nice Picture Comment! Graphicons in Facebook Comment Threads". In: *Proceedings of the Fiftieth Hawaii International Conference on System Sciences (HICSS-50)*. Los Alamitos: IEEE Press, pp. 2185–2194.
- Herring, Susan C., Dieter Stein, and Tuija Virtanen, eds. (2013). *Pragmatics of Computer-Mediated Communication*. Berlin, Boston: de Gruyter.
- Jucker, Andreas H., Heiko Hausendorf, Christa Dürscheid, et al. (2018). "Doing Space in Face-to-Face-Interaction and on Interactive Multimodal Platforms". In: *Journal of Pragmatics* 134, pp. 85–101.
- Küster, Marc Wilhelm (in this volume). "Open and Closed Writing Systems. Some Reflections".
- Meletis, Dimitrios (in press). "The Grapheme as a Universal Basic Unit of Writing". In: *Writing Systems Research*.
- (2018). "What is Natural in Writing? Prolegomena to a Natural Grapholinguistics". In: *Written Language and Literacy* 21.1, pp. 52–88.
- (2019). "Naturalness in Scripts and Writing Systems: Outlining a Natural Grapholinguistics". PhD thesis. University of Graz.

- Neef, Martin (2015). "Writing Systems as Modular Objects: Proposals for Theory Design in Grapholinguistics". In: *Open Linguistics* 1, pp. 708–721.
- Pappert, Steffen (2017). "Zu kommunikativen Funktionen von Emojis in der WhatsApp-Kommunikation". In: *Empirische Erforschung internet-basierter Kommunikation*. Ed. by Michael Beißwenger. Berlin, Boston: de Gruyter, pp. 175–211.
- Ueberwasser, Simone and Elisabeth Stark (2017). "What's Up, Switzerland? A Corpus-Based Research Project in Multilingual Switzerland". In: *Linguistik Online* 84/5. <https://bop.unibe.ch/linguistik-online/article/view/3849/5833> <06.09.2019>.
- Unger, J. Marshall (1990). "The Very Idea: The Notion of Ideogram in China and Japan". In: *Monumenta Nipponica* 45.4, pp. 391–411.
- Weingarten, Rüdiger (2011). "Comparative Graphematics". In: *Written Language and Literacy* 14.1, pp. 12–38.