1 Social Media Paralanguage and Emoji

1.1 Introduction

This book explores the kinds of meanings that can be made with emoji, small graphical icons encoded as Unicode characters. Emoji are a frequent feature of digital communication and, as we will see, can enact a wide range of functions in tandem with language and other semiotic resources. We began writing at home during lockdown in Sydney in the midst of the third wave of the COVID-19 pandemic in Australia. At 11 o'clock each morning, the New South Wales State Premier Gladys Berejiklian provided a televised update on the COVID-19 case numbers. These addresses generated large volumes of social media reaction on platforms such as Twitter. Emoji proliferated in these reactions, appearing in the ambient audience's responses to the case numbers, and in their criticism or praise of the government's policies. For example, journalists made use of emoji to encapsulate, categorise, and evaluate information provided during the premier's press conference, as in Text (1.1).

Text (1.1)
🚨 1431 new cases in NSW
💔 12 deaths (119)
😔 Woman in her 30s died (unvaccinated)
🎯 7.2 M 💉
🚑 Ambulance 160 in ICU (🅾️0)
🏥 979 in Hospital (⬇️62)
🦠 7 new cases in Wilcannia (88)
🦠 53 new cases in Western NSW (⚠️11 Burke)
💁‍♀️ @gladysB says 'The next 2 weeks will be the worst'

Not only do the emoji in this tweet[1] visually organise the message into a list, they articulate both key content (e.g. 🦠 + '*cases*') and key feelings (e.g. 💔 + '*deaths*'), together with the written verbiage in the post.

Everyday social media users also discussed their reactions to the daily updates in Twitter interactions such as Exchange 1.1.

[1] Tweets are short character-constrained messages posted to the social media service Twitter.

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

Exchange 1.1

Text (1.2) **User 1:** Anyone else feel like the day doesn't start until @gladysB tells us exactly how screwed we all are?

Text (1.3) **User 2**: Same, about to get my 11am presser coffee made☕

Within this exchange, there appears to be a relationship between '*coffee*' and the HOT BEVERAGE ☕ emoji. However, this emoji also seems to be important in terms of how interpersonal alignments are negotiated in the exchange, hinting at solidarity among the interactants and perhaps even some ironic enthusiasm. This book aims to deal with these kinds of relations – of emoji to any co-occurring text and to the interactive context – that are so central to how emoji make meaning in social media discourse.

1.2 The Semiotic Flexibility of Emoji

In order to function in a wide variety of discursive arenas, emoji need to have high semiotic flexibility. By virtue of their technical encoding, emoji are treated by social media platforms, and the devices on which they are navigated, as characters. This means that they can integrate seamlessly with the rest of the Unicode characters in the body of a social media post. This property also contributes to their semiotic capacity to make meanings relatively seamlessly with the written verbiage in the rest of the post, hereon the 'co-text'. For example, as a prelude to the discussion in the chapters that follow, let us consider the sorts of meaning emoji realise in the example tweets: Text (1.4), Text (1.5), and Text (1.6).

Text (1.4) Lovely chatting with {Name} from @{Username} about the insane endurance feats of @{Username} & @{Username} and the approach to supporting such giants.

Grab a cuppa & have a listen! ☕
👇
[tiny URL]

It comes after brill insights from {Name} on his PW record

Text (1.5) I drink so much ☕ I feel like 🤮 even thinking 🌫️ bout it

Text (1.6) Good Morning Tweepies!..💕.Monday Morning!.😝.Slept in!.😬😲.. have to say hello and Goodbye.👋 … see you this arvo! … 👋🤗☕☕☕ ☕☕☕ [GIF]

In Text (1.4), we see how the 👇 emoji is used as a visual deictic, pointing to the location of the 'tiny URL'[2] in the text. The emoji also potentially references

[2] A 'tiny URL' is short alias linking to a longer URL (Uniform Resource Locator) used to locate resources on the internet such as a web page.

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

the implicit call to action of 'clicking on' the URL since it resembles the mouse pointer icon of some operating systems. We thus have an emoji making meaning by visually depicting a material action and by referencing how a stylised rendering of that hand gesture has been adopted in another mode. In Text (1.5), the ☕ and 🤮 emoji are used to directly represent the entity (a hot beverage) and action (vomiting) they depict, thereby effectively replacing the written text for these. Noteworthy here, however, is that while the ☕ emoji depicts a category of things, when interpreted in combination with linguistic co-text a reader would supply just one example of that category, such as 'coffee' or 'tea'. Lastly in Text (1.6), the 💕, 😝, 😬, 😲, and 🤗 emoji act to imbue the text with various emotive inflections. Unlike in Text (1.5), however, we are unlikely to interpret the emoji as indexing the expressions or gestural paralanguage of the author or of other textual participants. Rather we recognise that in this context the emoji coordinate with the affective and convivial meanings made in the text, contributing to the inclusive solidarity enacted. In these small examples alone, we can see evidence of the semiotic suppleness of emoji.

On their own, emoji are rather under-specified and stylised representations that hint at a range of ideas. However, in combination with their co-text, they can make a range of complex meanings. Emoji can optionally involve themselves in the organisation of the text as a coherent semantic unit, can contribute to the articulation of entities and activities, and can resound with the emotional implications of the text, as we will explore in detail in Chapters 4, 5, and 6. In addition, each of the examples we have just explored offers a potential 'bond' to any interactants in the ambient social media audience through the way that emoji and language are used to share and position particular values. We will deal with such social affiliation in Chapters 7 and 8, considering both interactive exchanges and ambient communing. Emoji also enter into relationships with multimedia beyond the written co-text, such as digital stickers, memes, and simple animations. For example, Text (1.6) includes an animated image (a GIF), created from a snippet of a 1966 episode, 'Atlantic Inferno', from the British TV series *Thunderbirds*. The GIF depicts the electronic marionette puppet character, John Tracy, sitting up abruptly in bed. In terms of the meanings the tweet makes as a multimodal text, the GIF appears to coordinate intermodally with broadly shared ideas about coffee, mornings, and alertness that we will explore in Chapter 9.

1.3 The Semiotic Complexity of Encoding and Rendering 'Picture Characters'

Emoji are fascinating in their capacity to involve themselves in intricately specific meanings made within localised social media subcultures, at the same time as being malleable and under-specified enough to make meanings across

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

a vast array of social contexts. In terms of their visual appearance, they are presented to users of social media platforms such as Twitter, Instagram, Facebook, and TikTok as small, coloured glyphs. While it might be tempting to interpret the meaning-potential of emoji in terms of their visual appearance alone, they are complex semiotic resources. As we will explore in Chapter 2, emoji are 'picture characters' with some of the affordances of written characters and some of the affordances of images. One way of conceiving of this is to treat emoji as an iconographetic mode:

> … the element *icono* (from the Greek for *image*), generally refers to pictorial characters; it does not matter whether these are iconic or symbolic characters in isolation. The second part of the term, *graphetic*, is understood in terms of the Greek word *graphé* (or writing) and will hence refer to all kinds of written characters: *typed* characters in the case of digital communication. (Siever, 2019, p. 129, original formatting)

The concept of a picture icon used for interactive digital communication arose with the practice of mobile text messaging in Japan. The popular term 'emoji' is itself derived from the Japanese 絵文字 ('e' = 'picture'; 'moji' = 'written character'). However, just what constitutes a 'picture' or the property of pictoriality can be difficult to establish (Wilde, 2019). Studies have noted a range of analytical quandaries regarding how emoji, as picture icons, can represent apparently complex concepts via a visual shorthand that is both imprecise and expansive enough to range across the multitude of meanings needed in a wide range of communicative situations. Some studies appeal to a contrast with visual images to understand this semiotic potential: 'the specifics of the individual representation is often incidental to the underlying meaning of the ideogram [i.e. an emoji] – this is unlike images where the particulars of a given image are often more crucial than what it is representing generally (i.e. it is a photo of your dog, not just a photo representing the semantic notion of "dog")' (Cappallo et al., 2018, p. 2, clarification added). By this logic, there is an inherent difference between an image of a dog and the DOG 🐕 emoji.

In order to meaningfully account for how emoji make meaning as picture characters, we also need to factor in the way they operate inside 'semiotic technologies' (Zhao, Djonov, & van Leeuwen, 2014) such as social media in which they are deployed. We thus need to isolate each relevant semiotic mode and resource, as well as their particular affordances and functions. Part of this involves understanding exactly what emoji are as technical constructs and the implications this has for how they can be analysed. Emoji are 'encoded' as characters with unique code points in the Unicode Standard. This standard is widely adopted worldwide as a consistent method for encoding typed information in software, enabling cross-platform interoperability. In terms of their visual representation, emoji are 'rendered' by software as glyphs which gives them their distinctive appearance and 'colorful cartoon form' (Davis & Edberg, 2018).

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

This process is controlled by the software companies who own the platforms and is not standardised by Unicode. For instance, consider Text (1.7).

Text (1.7) Day 5 of covid. Feeling a little better. Still no taste or smell. Does anybody know how long that takes to come back I can't taste my coffee☕😭

The first emoji in the sequence ending Text (1.7) is the HOT BEVERAGE, which has the unique hexadecimal code point U+2615 and is rendered as the glyph ☕. The final emoji, LOUDLY CRYING FACE, has the code point U+1F62D and is rendered as 😭.

As Unicode characters, emoji form part of a designed and institutionalised pictographic lexicon defined by the Unicode Consortium's bureaucratic processes. This consortium is a conglomerate of entities that controls which characters are added to the Unicode Standard. A total of 674 emoji were added to Unicode in 2010 and their numbers have increased with each new version of the standard. The consortium's voting members include technology companies such as Adobe, Apple, Facebook, Google, Microsoft, and Netflix, institutional members like the Ministry of Endowments and Religious Affairs of Oman and the University of California at Berkeley, supporting members like Emojipedia, and a variety of associate and individual members (Unicode Consortium, 2021a). The Emoji Subcommittee, a part of Unicode's Technical Committee, evaluates proposals for new emoji based on various inclusion criteria, which will be discussed in further detail in Chapter 2.

Emoji's visual presentation as coloured glyphs also depends on the communicative channel used to create or read a social media post. Unlike encoding, the visual 'rendering' of emoji is controlled by the particular vendor (operating, software system, or platform) applying the font to the Unicode characters. For instance, if I enter the HOT BEVERAGE emoji from Text (1.7) into Twitter, the emoji will be rendered as a Twemoji, the distinctive rendering style used by Twitter, and displayed as: ☕. Therefore, it will have a different visual appearance compared to the emoji typed in Microsoft Word, which would display as ☕. Twemoji were created by Twitter designers working in collaboration with the company Iconfactory (Twitter, 2020). Vendors display emoji differently to showcase the unique branding and visual style they wish to project. According to its designers, the flat, gradient-free visual design of Twemoji aims to convey 'light-hearted, fun versions of the familiar icons users around the world know and love' in a visual style 'that would be easily identified as uniquely Twitter's' (Iconfactory, 2022). Twemoji have a minimalist, 'flat' design, with rounded shapes, dots, and lines used to articulate facial expressions, and without shading or 3D effects. In terms of colour palette, Twemoji 'tend to use colors that are similar to – or at least complement – the Twitter logo' (Gray & Holmes, 2020, p. 16). Figure 1.1 shows instances of Twemoji

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

Figure 1.1 Examples of Twemoji rendering of emoji from the Unicode 'Smileys and People' category

from the 'Smileys and People' category in the current version, Twemoji 14.0 v. This release of Twemoji includes 3,245 emoji which map to Unicode 14. All emoji in this book will be presented in this style as it is open-source and our main corpora were collected from Twitter.

To refer to emoji, we will use the naming convention adopted in Unicode, the Common Locale Data Repository (CLDR) Short Name, for instance, 'HOT BEVERAGE' for '☕'. The Common Locale Data Repository (CLDR) project, run by the Unicode Consortium, aims to provide locale data (e.g. relating to different languages) in an interoperable XML format so that it can be used in a variety of computer applications. For instance, emoji will have different CLDR Short Names depending on language, and these may be provisional for new emoji and change with version releases. However, CLDR Short Names should not be confused with the meaning of an emoji, which will instead be analysed using close text analysis and corpus-based methods. As visible in Text (1.1) and Text (1.7), because emoji are technically characters, they appear in-line with the rest of the written characters in the post (e.g. letters, digits, and symbols). This also means that the user can enter emoji through the keyboard, or a palette menu, without the author leaving the post's preparation window.

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

While emoji in their rendering as glyphs are images, their visual meaning potential is constrained. They cannot incorporate free-form components and are limited to the small size of accompanying textual characters. Their visual rendering also tends to be stylised rather than realistic, as seen in Figure 1.1. Emoji tend to have a limited degree of visual specificity, in part due to their constrained size. Paradoxically, it is this under-specification that means they are open-ended enough to enter into distinct relations with their co-text, and, in effect, make more complex meanings. We will return to these ideas about emoji encoding and rendering in Chapter 2, where we consider some of the technical dimensions touched on here in more detail, as well as reflect on their important implications for creating and processing corpora containing emoji.

1.4 Emoji as a Social Media Paralanguage

Our approach in this book is driven by observation of the close relationship between emoji and the linguistic meanings in social media posts; in other words, how intertwined emoji appear to be with the meanings made in their co-text. This perspective aligns with a shift in emoji research from attributing independent linguistic meanings to emoji towards analytical frameworks that prioritise the relationship between emoji and language. Research exploring the consistency of emoji interpretation (without the provision of contextual information to interpreters) has found that only a few emoji have completely unambiguous meanings (Częstochowska et al., 2022). As emoji have proliferated and become ubiquitous across digital communication, their apparent pragmatic meanings have become diluted (Konrad, Herring, & Choi, 2020) or have undergone semantic drift (Arviv & Tsur, 2021). Accordingly, an individual emoji may be considered a resource that is 'graphematically ambiguous, as the specific linguistic unit it refers to is not fixed but variable and determined by the context' (Dürscheid & Meletis, 2019, p. 174). As such, emoji are heavily dependent on their linguistic co-text, which acts as 'a clear verbal anchorage' (Sampietro, 2016, p. 110) for the meaning made by the multimodal text as a whole.

As our brief suggestions about the meanings made in Text (1.4), Text (1.5), and Text (1.6) at the beginning of this chapter have suggested, emoji may enact a range of semiotic relations with their co-text. For example, they can serve either a referential role (replacing words) or a modal role (modifying or complementing the surrounding text) (Siever, 2019). Some studies analogise emoji with co-speech gesture and suggest that, like beat gestures accompanying speech, emoji 'are not taking on the function of grammar, but acting in relation to written text' (McCulloch & Gawne, 2018). Other studies broaden the scale of context to consider cultural meanings, and argue that interpreting emoji requires a degree of 'familiarity with the cultural conventions of various aspects of contemporary society, along with an eclectic range of knowledge from

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

Eastern and Western written and gestural languages, sign languages and even fictional communication systems' (Seargeant, 2019, p. 25).

Given their strong connection to the meanings conveyed in their written co-text, we approach emoji as a form of paralanguage. Paralanguage is semiosis, such as gesture, which is dependent on language (Abercrombie, 1968). This dependency is sometimes described as 'parasitic' since it depends 'on the fact that those who use them are articulate ("linguate") beings' (Halliday & Matthiessen, 1999/2006, p. 606) and will vary depending on the kind of expression plane involved. For instance, in the case of paralanguage where the body is used for expression, this dependency might be 'sonovergent' with spoken language, that is, in-sync or in-tune with the phonological patterns of co-speech, or 'semovergent', that is, coordinating with linguistic meanings made in the co-speech (Martin & Zappavigna, 2019). Thus, rather than attempting to catalogue emoji as a kind of visual lexicon, we focus our attention on modelling the meaning-potential that emoji realise in concert with language.

Even where emoji appear in isolation in a text, they are likely to be dependent on co-occurring language within the broader context of situation, for instance, a preceding linguistic move in an exchange, as suggested by research on the role of images as moves in social media interactions (Jovanovic & van Leeuwen, 2018). The idea that emoji serve a paralinguistic function is also supported by corpus-based studies that have observed their semantic coordination (Gawne & McCulloch, 2019) and syntagmatic alignment (McCulloch & Gawne, 2018) with language. Our approach is also compatible with experimental studies that have suggested that, while emoji have some capacity for very simple sequencing and tend to interact with the linguistic grammatical structure, they do not seem to have developed their own grammatical structural potential (Cohn, Engelen, & Schilperoord, 2019). Emoji's visually stylised under-specification is also one of the reasons that emoji tend to coordinate with more elaborated meanings construed in their written co-text.

However, while we consider emoji as a form of paralanguage, we do not follow the approach taken in some studies of directly equating emoji with gestures (Gawne & McCulloch, 2019; McCulloch & Gawne, 2018). This is because we view emoji as a distinct semiotic mode with its own particular affordances and meaning potential. These affordances are realised via the expression plane of the 'picture character'; a different expression plane to modes which realise their meaning via embodiment (e.g. gesture, posture, voice quality, etc.). As previously mentioned, emoji are a 'designed' resource with specific digital affordances, and it is crucial to isolate these affordances to understand their semiotic potential. In simple terms, rather the studying emoji as if they were images or gestures, we study them for their own distinct meaning potential, taking into account their unique design and digital functions. Studies which liken emoji to gesture appear to be motivated by

the apparent iconicity of popular emoji that depict stylised facial expressions and body gestures (e.g. CRYING FACE 😢, ROLLING ON THE FLOOR LAUGHING 🤣, THUMBS UP[3] 👍, OK HAND 👌, CLAPPING HANDS 👏, etc.). However, a direct equivalence of emoji and gesture risks proscription of emoji's meaning potential – as Albert observes, 'the formal analogy between emoji faces in general and the corresponding facial expressions provokes the misleading inference that there must also be a functional analogy' (2020, p. 68). While it may be tempting to suggest that emoji 'share various properties and characteristics with other systems, they're actually adding something quite new to the resources we use to express ourselves' (Seargeant, 2019, pp. 35–6).

This is not to say that emoji are not agnate to other kinds of paralanguage. A dimension that gesture and emoji do share in common is their general dependency on their linguistic co-text. Employing McNeill's (1992) diagnostic criteria for determining the degree to which semiotic modes can function independently of language, Gawne and McCulloch (2019) observe that 'gestures and co-speech emoji are closely integrated into meaning with the accompanying speech/text'. This study suggests that emoji may be likened to gesture since they 'do not decompose into smaller morphological units, they do not show predictable syntax, their meaning is shaped by context-specific use, and there is accepted variation in form' (Gawne & McCulloch, 2019). According to this account, unlike language, emoji are global and synthetic, non-combinatoric, context-sensitive, and do not have standards of form.

1.5 A Social Semiotic Perspective on Emoji–Text Relations

The central goal of a social semiotic approach to communication is to understand how the different resources available to language users make meaning in the contexts in which they are used. In order to achieve this aim, not only do we need a theory of meaning and tools for analysing meaning-making, but we need a principled means for exploring how communicative modalities combine. In addition, we require ways of managing this complexity so that we achieve an elegant description of such semiotic coalescence. To systematically explore the meaning made in emoji–text relations, we will draw on social semiotics and its multimodal concern with understanding the semiotic systems that operate within and across modalities. We will approach these meanings methodically as 'bundles of oppositions' (Ngo et al., 2021, p. 8), adopting the relational theory of meaning that underlies work in Systemic Functional Linguistics (SFL). This kind of approach treats semiosis as a resource rather than a collection of rules

[3] Emoji glosses are sourced from https://emojipedia.org/ (accessed 11 November 2020), an emoji dictionary developed by professional lexicographers.

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

and treats the relations between choices in meaning as key to understanding how those choices function in real-world contexts.

Our functional approach manifests as a concern with three essential functions of language, termed 'metafunctions' by Halliday and Hasan (1985): the ideational (how experience is represented), interpersonal (how relationships are enacted), and textual (how text is organised). For instance, the oppositions in meaning we touched on when considering Text (1.1) at the beginning of this chapter can be seen to span what an SFL perspective on language views as *field* (the domain of experience), *tenor* (the interpersonal construction of relationships and stances), and *mode* (the organisation of the information flow of text) (Halliday & Matthiessen, 2004). In terms of field, the emoji in Text (1.1) contribute to co-construing the kinds of topics and experiences at stake: the MICROBE 🦠, AMBULANCE 🚑, and HOSPITAL 🏥 converge with verbal meanings about a health emergency. In terms of tenor, the PENSIVE FACE 😔 and BROKEN HEART 💔 resonate with details about deaths in the written verbiage to suggest negative emotions about this emergency. In addition, in terms of mode, the emoji themselves act as visual bullet points, organising the text into a list, at the same time as thematising the key information elaborated in the co-text. It is this kind of combinatorial meaning-making that we will focus on in the chapters which follow.

Inspired by work attempting to model paralanguage using Systemic Functional Semiotics recently consolidated in Ngo et al. (2021), one of the major assumptions that we make in this book is that language and other modalities coordinate inter-semiotically to make meaning. As such, we view written language and emoji as complementary semiotic resources and are interested in how they are interwoven, or more technically '*converge*' to create meaning in social media texts. This assumption of complementarity is also in line with earlier research into how images and written language coordinate in picture books where three types of relations of convergence were described: *concurrence* in ideational meaning, *resonance* in interpersonal meaning, and *synchronicity* in textual meaning (Painter & Martin, 2012; Painter, Martin, & Unsworth, 2013). These types of relations were used by Ngo et al. (2021) to explore how gesture and co-speech interrelate in embodied semiosis, resulting in the social semiotic model of paralanguage which informs the analytical approach adopted in this book.

Parkwell's (2019) metafunctional analysis of the meaning-potential of the TOILET 🚽 emoji aligns with our approach and serves as a noteworthy example of previous social semiotic work specifically focused on emoji. The TOILET 🚽 was used by popular musical artist Cher to discuss former US President Donald Trump on Twitter without using his name. The study draws on the perspective of multimodality (as outlined by Kress and van Leeuwen, 2001; O'Halloran, 2004) and Zappavigna's (2018) metafunctional analysis of

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

hashtags, to demonstrate how a single emoji can express experiential, interpersonal, and textual functions. The conclusion of the study highlights the highly contextual and flexible nature of emoji as a modality that is 'likely to continue to shift and morph with the changing needs and contexts of social media users' (Parkwell, 2019, p. 9). Another social semiotic study, conducted by He (2022), analysed the use of emoji in news story comments on the Chinese social media platform Weibo. This study adopted the 'intermodal coupling' of semiotic resources as its analytical unit, building upon the notion that meanings created through different modes can be complementary, as proposed by Painter et al. (2013). It found that emoji realise two distinct interpersonal functions: construing attitude targeting linguistic co-text and enacting social bonds with interactants around shared attitudes. These functions encompass emoji's capacity to 'not only directly reflect a commenter's attitude through the depiction of facial expression and gesture, but … to guide readers to detect the buried implications in a text' (He, 2022, p. 12). Other social semiotic studies of Weibo have also identified that emoji offer expanded pragmatic potential in relation to the co-text, serving as 'a multimodal layer of meaning in which emojis may not only substitute, reinforce, or complement text, but also perform speech acts, highlight subjective interpretations, and convey higher degrees of informality and/or casualness' (Yang & Liu, 2021, p. 166).

1.6 Using Corpora to Understand Emoji

The majority of corpus-based studies on emoji have been undertaken within the realm of computational science, utilising a corpus-driven methodology and incorporating machine learning techniques. These studies frequently aim to leverage emoji to enhance sentiment analysis (Kralj Novak et al., 2015; Shiha & Ayvaz, 2017) and typically view emoji as 'emotion tokens' for monitoring sentiment polarity (Wolny, 2016). Some studies aim to create emoji sentiment lexicons in an effort to surpass classification methods that are based on manual annotation or CLDR Short Names (Fernández-Gavilanes et al., 2018; Kimura & Katsurai, 2017; Kralj Novak et al., 2015), while others utilise the Unicode description as a means of classifying emoji (Eisner et al., 2016). A number of studies have centred on emoji sense prediction and disambiguation (Barbieri et al., 2018; Guibon, Ochs, & Bellot, 2018; Shardlow, Gerber, & Nawaz, 2022), and have monitored longitudinal changes in emoji semantics (Robertson et al., 2021).

This methodological context has proven fertile for research into how emoji have been used during the COVID-19 pandemic, primarily through the lens of quantitative studies using corpus-driven or sentiment analysis techniques to analyse emoji frequency and density in social media discourse (Das, 2021). This line of inquiry holds promise for yielding valuable insights that can

benefit domains such as public health initiatives and finance. For example, some studies have proposed new methods for understanding the gender-based disparities in the effects of COVID-19 (Al-Rawi et al., 2020) and for charting the correlation between emotional uncertainty and market volatility (Lazzini et al., 2021). Especially germane to this book is the vein of research examining the role of emoji in discourse related to remote work, including studies that examine emoji usage in videoconferencing chat (Dürscheid & Haralambous, 2021) and closed captions (Oomori et al., 2020).

Quantitative studies across various domains have shown a general interest in determining the most commonly used emoji. Unicode releases up-to-date information on emoji usage patterns, including the Unicode Emoji Subcommittee Chair's report on the most frequently used emoji in 2021 (Daniel, 2021). Additionally, various tools such as 'Emoji Tracker' (Rothenberg, 2013) aim to monitor emoji uptake in real-time, offering a dynamic insight into emoji trends and usage patterns. The top ten emoji used worldwide in 2021, according to Unicode (Daniel, 2021), were the following.

1 FACE WITH TEARS OF JOY 😂
2 RED HEART ❤️
3 ROLLING ON THE FLOOR LAUGHING 🤣
4 THUMBS UP 👍
5 LOUDLY CRYING FACE 😭
6 FOLDED HANDS 🙏
7 FACE BLOWING A KISS 😘
8 SMILING FACE WITH HEARTS 🥰
9 SMILING FACE WITH HEART EYES 😍
10 SMILING FACE WITH SMILING EYES 😊

While this kind of frequency list cannot tell us the 'meaning' of these emoji, it does suggest that they are most likely involved with construing broadly positive meanings, depending on how they interact with their co-text.

Efforts within corpus linguistics to study emoji usage are relatively new, likely due to the technical challenges posed by the unique features of emoji that can complicate the use of traditional corpus analysis tools such as concordance software. Chapter 2 will delve into the specific challenges posed by emoji as special characters in corpus processing, exploring both their encoding and rendering. These issues require the analyst to pay close attention to what is actually being counted. While emoji might be roughly interpreted as a 'lexical unit', they are in fact often composed of Unicode character sequences. This means that corpus linguistic software will not necessarily be able to capture all emoji unless it has the capability for recognising these sequences as a single unit. Thus some kind of work-around for concatenating relevant emoji sequences will be required to meaningfully process emoji

(Zappavigna & Logi, 2021). Accordingly, this book relied on a custom script, along with a python library, to accurately count and inspect emoji concordance lines. This approach ensured that Unicode emoji sequences, which are not readily recognisable by standard concordance systems, were properly accounted for.

Corpus-based studies of emoji in linguistics have nevertheless attempted to draw on standard corpus methods such as analysis of frequency lists, concordance lines, and n-grams. Some corpus-based studies have combined pragmatics and corpus analysis methods (Li & Yang, 2018; Pérez-Sabater, 2019; Sampietro, 2019). For example, Kehoe and Gee (2019) undertook a large-scale data-driven corpus pragmatic analysis of emoji use on Twitter, using a corpus of 40 million English and German language tweets. Replicating the results of previous research, the FACE WITH TEARS OF JOY 😂 was the most frequent emoji in this dataset, followed by LOUDLY CRYING FACE 😭 (in English) and the RED HEART ❤️ (in German). The study employed collocational analysis to disambiguate different emoji uses. For instance, it distinguished multiple meanings for FOLDED HANDS 🙏, including 'thanking, pleading, praying or giving a high five' (Kehoe & Gee, 2019, p. 2). The study noted that collocational span, as well as the frequent repetition of emoji within tweets, posed challenges for corpus analysis of emoji patterning.

Another relevant study adopting a corpus analytical approach is a multimodal analysis of Facebook posts incorporating emoji and annotated for images (Collins, 2020). This study also found FACE WITH TEARS OF JOY 😂 to be the most frequent emoji. Echoing the challenges noted by Kehoe and Gee (2019), the study suggested that the traditional concept of a collocational span, established as useful for work on written text, was problematic for exploring emoji and for relations of images to text. These issues were somewhat ameliorated when dealing with images in the corpus by employing a large collocational span of 365 tokens (which was determined from the longest text in the corpus), while keeping a keen eye on confidence measures. The author decided, drawing on Spina's (2019) work on emoticons, that emoji 'should also be investigated within a narrower collocational span (at the "type" unit level), since research has shown that there are conventions for the position of emoji, which interact with the syntax and punctuation of written material' (Collins, 2020, p. 190). In our own work we adopt a similar approach, drawing on collocational patterns and n-grams where possible to both motivate and buttress our more qualitative discourse semantic analysis.

1.7 Corpora Analysed in This Book

In this book we employ a corpus-based approach to understanding the use of emoji in social media posts, informed by a social semiotic orientation to discourse analysis. This kind of methodological position is sometimes termed

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

'corpus-based discourse analysis' (Baker, 2006; Bednarek, 2009) since the close textual analysis is informed by patterns detected in corpora. This means that we use evidence drawn from corpora created according to specific selection criteria. It also means that, while we ground our analysis in this evidence, we draw on theoretical tools for interpreting the patterns of meaning we observe: in this case, a combination of Systemic Functional Linguistics and Multimodal Discourse Analysis.

A significant limitation of the scope of the corpus data that we consider is that we only analyse emoji used in English language posts. This choice was largely a pragmatic one, based on the linguistic proficiency of the authors and their areas of expertise relative to the complexity of the multimodal meaning-making encountered in the social media texts collected. Future work will be needed to explore emoji used with a wider range of languages, each of which might involve different systems of meaning-making. Such studies are needed to address popular claims that emoji offer a kind of 'universal' language that bridges cultural contexts (Abel, 2019). Some corpus-based cross-cultural work is emerging in this area, often tracking how a particular emoji is used across multilingual contexts. For instance, kissing face emoji have been shown to exhibit cultural variation within Europe (Sampietro, Felder, & Siebenhaar, 2022), and predominantly Western users of Twitter associate different emoji to feelings of surprise and trust than users of Chinese social media platform Weibo do (Li et al., 2019).

Four corpora are used in this book. All posts are anonymised for identifying information such as names and locations. For instance, usernames are replaced with '@user'. The first is the Sydney Emoji Corpus, which was primarily used for exploratory analysis, drawing on work presented in Logi and Zappavigna (2021b) for mapping out emoji meaning potential. This is a relatively small corpus of approximately 1,000 social media posts and digital messages collected from undergraduate university students across a range of platforms and services.[4] Its constrained size made it suitable for close, explorative, and detailed discourse analysis. The general aim of this initial work was to identify the various ways in which emoji can coordinate with language in order to make meaning. The functions and patterns of meaning identified by this initial analysis were then explored in a larger corpus, the Hot Beverage Emoji Corpus, approximately 2.5 million tweets containing the HOT BEVERAGE ☕ emoji. This corpus was created by removing retweets and tweets by bots from a larger 12 million tweet corpus collected using the Twitter Application Programming Interface (API). It was also processed to include only English language tweets. The most frequent emoji in the Sydney Emoji Corpus and Hot Beverage Emoji Corpus are shown in Table 1.1.

[4] Demographic information was collected in this study but are not reported here since sociolinguistic variables are not our current focus.

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

Table 1.1 *Emoji frequency list for the Sydney Emoji Corpus and the Hot Beverage Emoji Corpus*

	Sydney Emoji Corpus			Hot Beverage Emoji Corpus		
N	CLDR Short Name	Emoji	Frequency	CLDR Short Name	Emoji	Frequency
1	RED HEART	❤️	195	HOT BEVERAGE	☕	13,169
2	FACE WITH TEARS OF JOY	😂	118	RED HEART	❤️	1,320
3	SMILING FACE WITH HEART-EYES	😍	114	SUN	☀️	1,082
4	SMILING FACE WITH SMILING EYES	😊	91	SMILING FACE WITH SMILING EYES	😊	750
5	LOUDLY CRYING FACE	😭	88	HUGGING FACE	🤗	494
6	TWO HEARTS	💕	75	CHERRY BLOSSOM	🌸	490
7	FACE BLOWING A KISS	😘	71	SUNFLOWER	🌻	485
8	PARTY POPPER	🎉	65	COOKING	🍳	453
9	SPARKLES	✨	54	FACE BLOWING A KISS	😘	422
10	SLIGHTLY SMILING FACE	🙂	54	FACE WITH TEARS OF JOY	😂	389

The Sydney Emoji Corpus spanned posts across a range of social media platforms (e.g. Twitter, Facebook, Instagram, etc.); however, the Hot Beverage Emoji Corpus contained only tweets. While this poses a potential limitation for the work presented in this book, we did not notice a large variation in the functional potential of emoji across platforms. However, readers are advised that particular platform constraints and affordances may influence the ways in which emoji can be used. As we will explore in Chapter 2, in terms of the rendering of emoji as colourful glyphs (as opposed to their encoding as Unicode characters), emoji will have a different visual appearance that varies with platform, software, and operating system.

Two specialised corpora are used in this book to explore how emoji function to support ambient affiliation, in terms of both dialogic affiliation (where users interact with each other) and communing affiliation (where there is little direct interaction). The corpus used for the first scenario is the Quarantine Hotel Food Review TikTok Comment Corpus, employed in Chapter 7 to explore the role of emoji in negotiating affiliation in interactions between TikTok creators and commenters. This corpus was created by extracting TikTok comments from the comment feed of a TikTok video series about a Creator's experience of the food delivered to their room during a fourteen-day isolation in a quarantine

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

Table 1.2 *Emoji frequency list for the Hotel Quarantine TikTok Comment Corpus and the #Domicron Corpus*

	Quarantine Hotel Food Review TikTok Comment Corpus			#Domicron Corpus		
N	CLDR Short Name	Emoji	Frequency	CLDR Short Name	Emoji	Frequency
1	FACE WITH TEARS OF JOY	😂	700	FACE WITH SYMBOLS ON MOUTH	🤬	1,196
2	SMILING FACE WITH HEARTS	🥰	660	POUTING FACE	😡	477
3	PLEADING FACE	🥺	309	BACKHAND INDEX POINTING DOWN	👇	431
4	GRINNING FACE WITH SWEAT	😅	283	FACE WITH ROLLING EYES	🙄	350
5	ROLLING ON THE FLOOR LAUGHING	🤣	204	FACE WITH MEDICAL MASK	😷	247
6	FLUSHED FACE	😳	177	ROLLING ON THE FLOOR LAUGHING	🤣	234
7	HEART SUIT	♥️	166	THINKING FACE	🤔	233
8	SMILING FACE WITH HEART-EYES	😍	133	FACE WITH TEARS OF JOY	😂	218
9	LOUDLY CRYING FACE	😭	128	CLAPPING HANDS	👏	205
10	RED HEART	❤️	107	PILE OF POO	💩	184

hotel. This corpus included all exchanges between users in these comment feeds (5,386 posts, 60,496 words). The second corpus is the #Domicron Twemoji Corpus, collected during a period in which the Australian state of New South Wales was emerging from a COVID zero policy to an approach politicians termed 'living with COVID'. 'Domicron' was a pun amalgamating 'Dominic' and 'Omicron' (a COVID-19 variant), employed as a way of attacking the performance of the Australian state of NSW's Premier (head of government) Dominic Perrottet. This corpus contained 4,275 tweets containing one or more emoji together with the hashtags #Domicron or #LetItRipDon (117,666 words) and is used to explore how users commune around or contest social bonds tabled in tweets. The most frequently used emoji in the two specialised corpora are shown in Table 1.2.

1.8 Structure of This Book

This book is structured so as to consider the relationship of emoji to their written co-text in a systematic way. We will begin in Chapter 2 by addressing the difficult and interesting problems that emoji pose for corpus analysis.

https://doi.org/10.1017/9781009179829.001 Published online by Cambridge University Press

This chapter will explore the technical dimensions of emoji, considering their affordances as 'picture characters' that are encoded as Unicode and visually rendered as glyphs. Chapter 3 then articulates the theoretical approach to emoji–text relations that undergirds the corpus-based discourse analysis undertaken in the rest of the book, focusing in particular on the logic underlying our position that emoji function as paralanguage. The chapter will then present our framework for analysing emoji–text convergence that incorporates three simultaneous dimensions: textual synchronicity, ideational concurrence, and interpersonal resonance (as flagged in the relations we discussed for Text (1.1) at the start of this chapter). The three chapters that follow pick up each of these types of relations in turn, considering emoji one metafunction at a time at the level of discourse semantics.

Following this metafunctional exploration, Chapters 7 and 8 consider how these kinds of relations can act in the service of affiliation in terms of how social media users commune around or negotiate social bonds. We then consider in Chapter 9 the impetus towards customisation and personalisation in visual glyphs, as social media users seek to make meanings beyond Unicode character constraints with the use of digital stickers, GIFs, memes, and other kinds of visual phenomena. The book concludes by reviewing the meaning potential of emoji and visual social media paralanguages in terms of what they tell us about the development of digital communication, intermodal semiosis, and society.