

Hanmoji: What Chinese Characters and Emoji Reveal About Each Other

Jennifer 8. Lee
Emojination
San Francisco, California, USA
jenny@emojination.org

Jason Li
Visual Culture Studies, The Civic Beat
Hong Kong, Hong Kong
jason@thecivicbeat.com

An Xiao Mina
Visual Culture Studies, The Civic Beat
Hong Kong, Hong Kong
an@thecivicbeat.com

ABSTRACT

Chinese characters predate the introduction of digital emoji by approximately 3,000 years. Despite the temporal gap, there are striking parallels between the canonical set of Chinese radicals and the set of emoji that have currently been approved by the Unicode Consortium. Comparing the 214 Kangxi Chinese radicals with the 3,019 emojis in the Unicode 12.0 set can reveal semantic gaps and provide directions for new emoji. Our analysis found that 72.4% of radicals have reasonable emoji equivalents, while only 17.8% of radicals lack any emoji equivalent that we could determine.

CCS CONCEPTS

• Applied computing → Language translation; Media arts.

KEYWORDS

Emoji; Chinese; Chinese Radicals; Kanji; Kangxi; Linguistics

ACM Reference Format:

Jennifer 8. Lee, Jason Li, and An Xiao Mina. 2019. Hanmoji: What Chinese Characters and Emoji Reveal About Each Other. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion)*, May 13–17, 2019, San Francisco, CA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3308560.3316543>

1 BACKGROUND

The earliest known forms of Chinese writing is found on oracle bones, dating to the Shang Dynasty (18th–12th Century BC). Chinese scripts evolved through various stages: *guwen* (“ancient figures”) found in inscriptions from the period around the late Shang dynasty (c. 1123 BC), the *dazhuan* (“Great Seal”) scripts of the Zhou dynasty (1046 to 256 BC), and the modern day forms, which were largely established in the Qin period (221–207 BC). Many of the original oracle writing were based on visual representations of natural objects, such as “horse”, “dog”, and “mouse”.

The origin of modern digital emoji is now commonly cited to be 1999, when Japanese designer Shigetaka Kurita created what has become the first popular and widely used set, consisting of 176 emoji, for DoCoMo. The Unicode Consortium became the body tasked to unify the disparate sets from various carriers. Starting with the 625 emoji characters in the 2010 Unicode Standard, the set now numbers 3,019 emojis as of the Unicode 12.0 release, a count

which includes the gender and skintone variations. Since emoji stem from Japanese culture, and aspects of Japanese writing, specifically the kanji characters, are derived from Chinese characters, there is a natural cultural link from Chinese characters to emoji.

2 METHOD

We assembled the set of 214 Kangxi Chinese radicals which have become the de facto ordering standard for Chinese dictionaries since their original introduction in the 1,615 *Zihui* dictionary. We then aggregated the English definitions for the radicals from yellowbridge.com [1], Wikipedia, and Wiktionary, with a special effort to take a “broad” definition when the term was translated into English. While some concepts were direct and straightforward – sun, water, fire, pig – others were more conceptual, such as (mi4), which variously could be “over”, “crown”, “3-side enclosure from the top”, “cloth cover”.

We then went through the 214 radicals and tried to match the radicals with 1) direct emoji equivalents, where the definition is essentially equivalent, 2) indirect emoji equivalents, where the meaning can be understood even if the definition is not equivalent, and 3) associative emoji meaning, where if forced to find an emoji that represents an idea, we have one that contains it as a quality of a component. Figure 1 shows the first 55 Chinese radical characters we explored, while the full list can be accessed at <http://bit.ly/hanmoji>

We made a special categorization of radicals which represented visual stroke, without a specific meaning – such as (slash, radical 4) and 冫 (fang1, radical 22). For those without direct or indirect emoji parallels, we then methodically asked whether an emoji could reasonably be added, either because it represented a notable semantic gap, and if such an image could reasonably be expected to pass the specific criteria laid out by the Unicode Consortium (<http://unicode.org/emoji/proposals.html>), which include whether a proposed emoji is visually distinct and is not transient (e.g. has demonstrated longevity).

3 RESULTS AND DISCUSSION

Of the 214 Kangxi radicals, we count that 84, or 39.3%, have reasonably direct emoji equivalents as of the March 2019 Unicode 12.0 release, which had 3,019 emojis, including all the gender and skintone variations [2]. In addition, an additional non-overlapping 71, or 33.2%, have reasonable indirect emoji equivalents for at least one of their meanings. And third, we found 21, or 9.8%, emoji that

had associative qualities such as 🍯 (HONEY POT) for 甘 (sweet). So 72.4% of radicals have reasonable equivalents, while only 17.8% of radicals lacked any of the three.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

© 2019 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-6675-5/19/05.

<https://doi.org/10.1145/3308560.3316543>

Radical	Meaning	Direct	Indirect	Associative	Radical	Meaning	Direct	Indirect	Associative	Radical	Meaning	Direct	Indirect	Associative
1 一	one	1	👉		18 刀	knife	🔪			38 女	woman	👩		
2 丨	line		—		19 力	strength, power	💪			39 子	child	👶		
3 丶	dot, ancient punctuation mark				20 ㇏	wrap	📦			40 宀	roof		🏠	
4 丿	slash				21 匕	spoon, ladle	🥄			41 寸	inch	📏		
5 乙	second		👉		22 匚	right open box				42 小	small		🐼, 🐻	
6 乚	hook		👉		23 匚	hiding enclo	🏠			43 尢	lame	👴		
7 二	two	2	👉		24 十	ten	10			44 尸	corpse	💀	🥩	
8 宀	lid, cover			🏠	25 乚	mysti	👤			45 屮	sprout	🌱		
9 人	man	👤			26 卩	seal, kneeling man (?)	👤			46 山	moun	🏔		
10 儿	legs	👤			27 厂	cliff			🏔	47 川	river		🌊	
11 入	enter				28 厶	private	🔒		🔑	48 工	work		🔨	
12 八	eight	8		🔍	29 又	again	🔄			49 己	oneself			
13 冂	upside down box (also, city outskirts; suburbs)				30 口	mouth	👄			50 巾	turban, towel	👤		
14 冂	over, crown, 3-side enclosure from the top, cloth cover		👑		31 匚	enclo			📦	51 干	dry		👤	
15 冫	ice	❄			32 土	dirt, earth	🌍			52 幺	short thread, thread, last, small, minute		📦	
16 几	table				33 士	scholar	🎓			53 广	dotted cliff, house on a cliff		🏔	
17 匚	open box, receptacle	📦			34 攴	go	👤		🚶	54 辶	long stride		👤	
					35 夕	go slowly	👤		🕒	55 升	two hands, joined hands	👤	👤	
					36 夕	evening	🌃							
					37 大	big			👤					

Figure 1: A comparison of the Chinese Kangxi Radicals and direct, indirect and associative emoji equivalents. Only the first 55 radicals are shown. The full list of the 214, along with an expanded chart, can be accessed at - <http://bit.ly/hanmoji>

We also identified 14 Kangxi radicals, or 6.5%, that had reasonably strong chance of having either direct or indirect emoji equivalents based on the current Unicode selection factors, which include visually representable and distinct, longevity, significant semantic meaning, and popular demand. These include 龠(FLUTE), 豆(BEAN), 支(BRANCH), 穴(CAVE), 几(TABLE), and 矛(SPEAR). Next, we identified 15 Kangxi radicals, or 7.0%, that could possibly have emoji equivalents passed, but we categorize them as weaker than the previous set because of difficulty of visual representation or contemporary relevance. These include radicals such as 玉(jade), 戈(spear-axe), 臼(mortar), and 鼎(tripod).

Perhaps most interesting is that we identified a few clusters of other Kangxi radicals that fairly atomic and have important semantic meaning, but would face an uphill battle to have direct or indirect emoji equivalents passed because of the difficulty of visual representation. Some of them do have associative emoji meanings, however. These are:

- (1) Adjectives: 大(big), 小(small), 長(long), 高(tall or high).
- (2) Conveying “selfness” or privacy: 己(self), (private), 自(oneself).
- (3) General concepts that may be too broad to tie to one distinct visual representation: 生(life), 用(to use), and 辛(bitter).
- (4) Last we identified a cluster of radicals that have decent semantic meaning, but also face an uphill battle because of the difficult of representation; Textures: 皮(SKIN), 髟(HAIR), 革(LEATHER), 毛(FUR), (BRISTLE).

But the Chinese radicals are also interesting for which semantic areas are nearly completely absent, yet have significance presence in emoji, and often at an “atomic” level. For example, there are almost no emotions contained in the Kangxi Chinese radical set. In contrast, the most popular, emoji are almost always the yellow “smiley” faces which show concepts 🥲 (Crying Face), 😡 (Angry Face), 😄 (Grinning Face).

4 CONCLUSION

Notably 170, or 79.4%, of the 214 Kangxi radicals have a direct or indirect emoji equivalent, or a reasonable chance of getting one. That speaks to the universality and longevity of certain concepts in a human existence, which include animals, body parts, colors, natural elements, foods, and tools. The analysis also helped identify certain semantic gaps, including basic adjectives, general concepts, and items which should give the Unicode Consortium a basic roadmap. These include 龠(FLUTE), 豆(BEAN), 支(BRANCH), 穴(CAVE), 几(TABLE), and 矛(SPEAR).

REFERENCES

- [1] Yellow Bridge. 2019. List of Chinese Radicals. (2019). <https://www.yellowbridge.com/chinese/radicals.php>
- [2] Unicode Consortium. 2019. Emoji List, v12.0. (2019). <https://unicode.org/emoji/charts/emoji-list.html>