

Emoji Name Drift: How Unicode’s Stable Names Diverge From the Words People Actually Use

Every emoji has a formal Unicode name that never changes and a CLDR short name that can. This study compares the two for all 1,400 emoji that have a formal name, tracks how CLDR’s names changed across 21 releases, and checks a small sample of real search words against both.

MAIN FINDING

Most emoji names still match: 68% read the same in both systems. About 1 in 7 (13.6%) differ in a way that changes how you would describe the emoji, and most of that gap was there from the first CLDR release.

- Comparison complete for Unicode 18.0 and Emoji 18.0
- Severity labels not yet reviewed by people
- CLDR 49 final not yet used (beta 2 data)

PUBLISHED	VERSION	EVIDENCE
October 13, 2026	1.0	Unicode and CLDR data files, model-assigned labels, a hand-picked sample of Reddit threads

Canonical URL: <https://ultratextgen.com/research/emoji-name-drift/>
UltraTextGen Research · Version 1.0 · Published October 13, 2026

SUMMARY

What the study found

952 of 1,400

emoji have the same name in Unicode and CLDR
That is 68.0% once capital letters, hyphens and punctuation are ignored. None match letter for letter, because Unicode writes names in capitals.

● MEASURED FROM THE DATA FILES

190 (13.6%)

differ in a way that changes how you would describe the emoji

62 of them (4.4%) differ substantially. Depending on which of two reviewers you trust, the share is 11.1% to 15.2%.

◇ MODEL-ASSIGNED LABELS

74%

of today's differences were already there when CLDR first named the emoji

CLDR has renamed 171 of these emoji across 21 releases. The last rename that changed more than capital letters or an ampersand was in CLDR 40.

● MEASURED FROM RELEASE HISTORY

58%

of differing names keep a word the formal name lost as a CLDR keyword

Keywords carry words like *hocho*, *aubergine* and *pray*, so a search by the old word can still land on the emoji.

● MEASURED FROM THE DATA FILES

15 of 31

words real people typed are found in a CLDR name or keyword

The 16 misses are mostly feelings, picture descriptions, occasions and slang, not object names. This is a hand-picked sample of 9 Reddit threads, so it shows how searches fail, not how often.

□ DIAGNOSTIC SAMPLE

What this means for you

- If an emoji search fails, check the word before blaming the emoji's name. Three cases in the sample (*hatchet* for 🪓, *high* and *pray* for 🙏, *heart eyes* for 😍) involve words CLDR already lists, so the gap is not in CLDR's vocabulary. The likely cause is how the app matches words, though we did not test the apps.
- Some of the emoji people struggle with have the same name in both systems. 😓 and 🪓 do. The trouble with 😓 is what it means, and the trouble with 🪓 is how one app searches, not what either is called.
- The part of the gap that is real is vocabulary no standard lists: occasions such as St. Patrick's Day, feelings such as *sigh*, and community nicknames. Search tools have to supply those themselves.

BACKGROUND

Two names, two jobs

Every emoji that is a single character has a formal Unicode name, such as SMILING FACE WITH HEART-SHAPED EYES for 🥰. Unicode treats that name as an identifier and never changes it. It also has a CLDR short name, such as *smiling face with heart-eyes*, which screen readers speak and keyboards search. CLDR can revise that name between releases.

The two names are built for different jobs, so a difference between them is not a mistake. This study measures how far apart they are and what kind of change separates them. It does not treat distance as error. Whether the gap helps or hurts is a question the evidence below answers case by case.

A third layer sits outside both standards: the words people actually type when they look for an emoji. Neither Unicode nor CLDR promises to match it. The study keeps three questions apart. **Naming drift** is a label that changed. **Meaning drift** is people using an emoji for more than its label says. **Search-vocabulary drift** is people typing words that appear in no name.

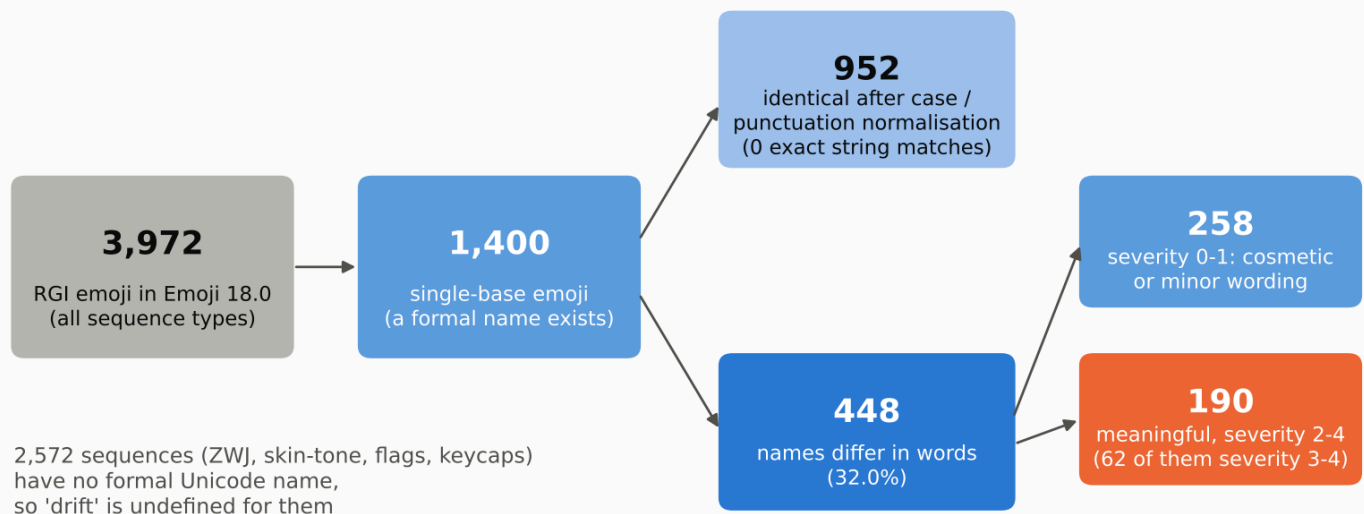
Emoji that are built from several characters, such as skin-tone, flag and family emoji, have no formal Unicode name. They only have CLDR names assembled by rule. The comparison here therefore covers the 1,400 emoji that have both names, and leaves the other 2,572 out. Skin-tone names mostly inherit the base emoji's name (655 of 665), so the gap measured here carries over to them.

RESULTS

How many names differ

Of the 3,972 emoji in Emoji 18.0, 1,400 have a formal Unicode name: 1,193 single characters and 207 written as a character plus the emoji-style selector U+FE0F (such as ❤️ and 🙄). Of those, 952 match after capitals, accents, hyphens and punctuation are ignored. The other 448 differ in at least one word.

From the full emoji universe to meaningful name drift



From the full emoji list to meaningful differences. The 2,572 emoji built from several characters have no formal name and are not compared.

A small share of the 448 differences matter. The study scored each one from 0 (same words) to 4 (major difference, or the formal name is opaque). Most are small: 258 score 0 or 1, usually a dropped word like *face* or *sign*.

RESULT	EMOJI	SHARE OF 1,400
Letter-for-letter match	0	0.0%
Same after normalization	952	68.0%
Differ in words	448	32.0% (95% interval 29.6 to 34.5)
Meaningful, score 2 or more	190	13.6% (11.9 to 15.5); 11.1 to 15.2 by reviewer
Substantial, score 3 or more	62	4.4% (3.5 to 5.6); 3.9 to 5.4 by reviewer
Major, score 4	18	1.3% (0.8 to 2.0)

The 1,400 are the whole set, not a sample, so the intervals only describe how firm the labels are. The spread between the two reviewers is the better guide to uncertainty. Treat any count as roughly two percentage points either way.

EXAMPLES

Twenty examples

The table shows the range from a formatting change to a change of subject. Every row was checked against the Unicode and CLDR files themselves. The last two rows are controls: their names did not change at all.

EMOJI	UNICODE NAME	CLDR NAME	KIND, SCORE	WHAT HAPPENED
	FACE WITH LOOK OF TRIUMPH	face with steam from nose	F, 3	The formal name records an interpretation (triumph); CLDR describes the picture (steam from the nose), so the emotion label that gave the emoji its meaning is gone from the name.
	HUGGING FACE	smiling face with open hands	E, 3	CLDR 40 replaced the colloquial 'hugging face' with the literal 'smiling face with open hands', a rename that moved away from ordinary speech (the word 'hug' survives only as a keyword).
	GRINNING FACE WITH SMILING EYES	beaming face with smiling eyes	D, 3	The formal name is word-for-word the CLDR name of a different emoji () , so searching by formal name finds the wrong face.
	HEAVY BLACK HEART	red heart	N, 3	'Black heart' is now the name of a separate emoji (); the legacy black = filled convention would send a formal-name searcher to the wrong one.
	RAISED HAND WITH PART BETWEEN MIDDLE AND RING FINGERS	vulcan salute	I, 4	An anatomical description became a fandom reference; nobody could derive 'Vulcan' from the formal name.
	HOCHO	kitchen knife	I, 4	A Japanese loanword is replaced by plain English; 'hocho' survives only as a keyword.

EMOJI	UNICODE NAME	CLDR NAME	KIND, SCORE	WHAT HAPPENED
	PERMANENT PAPER SIGN	infinity	K, 4	A legacy glyph-source name ('permanent paper sign') versus the concept everyone calls it (infinity); the two names share no word.
	PISTOL	water pistol	N, 3	A neutral 'pistol' becomes 'water pistol' (CLDR 38), tracking how vendors draw it; a rename that narrowed the object.
	BLACK RIGHT-POINTING TRIANGLE	play button	M, 3	A geometric description ('black right-pointing triangle') becomes a function ('play button').
	TWISTED RIGHTWARDS ARROWS	shuffle tracks button	M, 3	Twisted arrows become 'shuffle tracks button': what the symbol does replaces what it looks like.
	DANCER	woman dancing	G, 3	Formal DANCER has no gender; CLDR 30 renamed it 'woman dancing', adding a gender that the name never had.
	PERSON WITH HEADSCARF	woman with headscarf	G, 3	One of three moves in the opposite direction from gender-neutralisation (with  and ): 'person with headscarf' becomes 'woman with headscarf'.
	WEIGHT LIFTER	person lifting weights	G, 2	'Weight lifter' becomes the gender-neutral 'person lifting weights' (CLDR 30), the pattern repeated across the sports people.
	AUBERGINE	eggplant	D, 2	British 'aubergine' becomes American 'eggplant'; the old word is kept as a keyword, so both dialects can search it.
	FATHER CHRISTMAS	Santa Claus	I, 2	British 'Father Christmas' becomes 'Santa Claus' (a regional-name change that CLDR keywords bridge with 'Christmas' and 'claus').

EMOJI	UNICODE NAME	CLDR NAME	KIND, SCORE	WHAT HAPPENED
	RECREATIONAL VEHICLE	sport utility vehicle	E, 3	'Recreational vehicle' (an RV) became 'sport utility vehicle': two different vehicles for one glyph.
	EARTH GLOBE EUROPE-AFRICA	globe showing Europe-Africa	C, 1	A pure formatting change: 'earth globe europe-africa' becomes 'globe showing Europe-Africa' (severity 1).
	WAVING WHITE FLAG	white flag	B, 1	Control for a trivial rename: 'waving white flag' becomes 'white flag' (the word 'waving' is dropped; severity 1).
	SLEEPY FACE	sleepy face	none, 0	Control: formal and CLDR names are identical ('sleepy face'), yet Reddit readers disagree on what it means (snoring, sighing, exhaustion): meaning drift without name drift.
	AXE	axe	none, 0	Control: formal and CLDR names are identical ('axe') and CLDR keywords already include 'hatchet'; the reported search failure cannot be a vocabulary gap there.

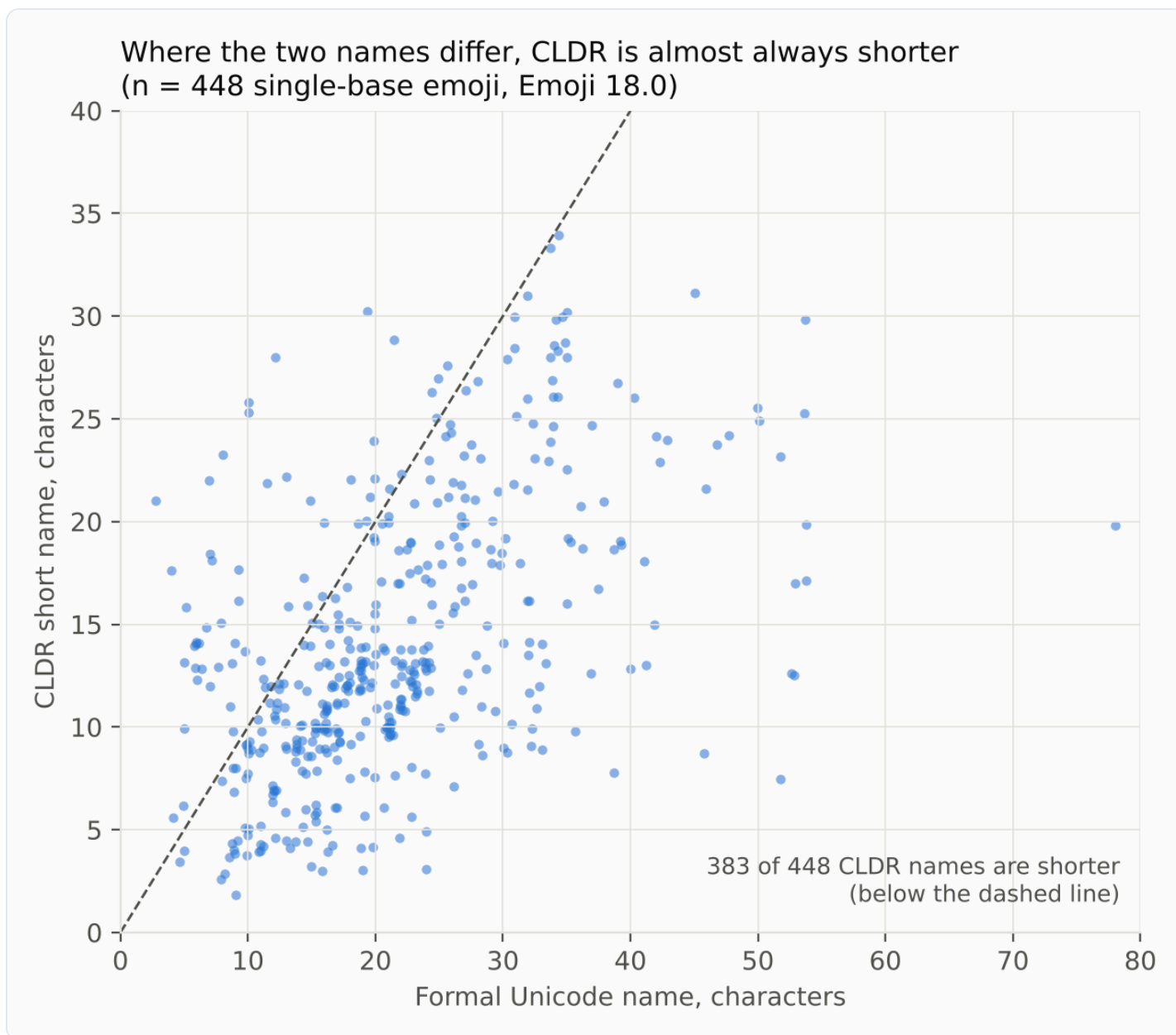
Kinds: A cosmetic, B simplification, C expansion, D synonym, E different main noun, F emotion reworded, G person or role wording, I cultural or regional, K substantial change, M function relabel, N colour or rendering legacy. Scores run from 0 (same words) to 4.

RESULTS

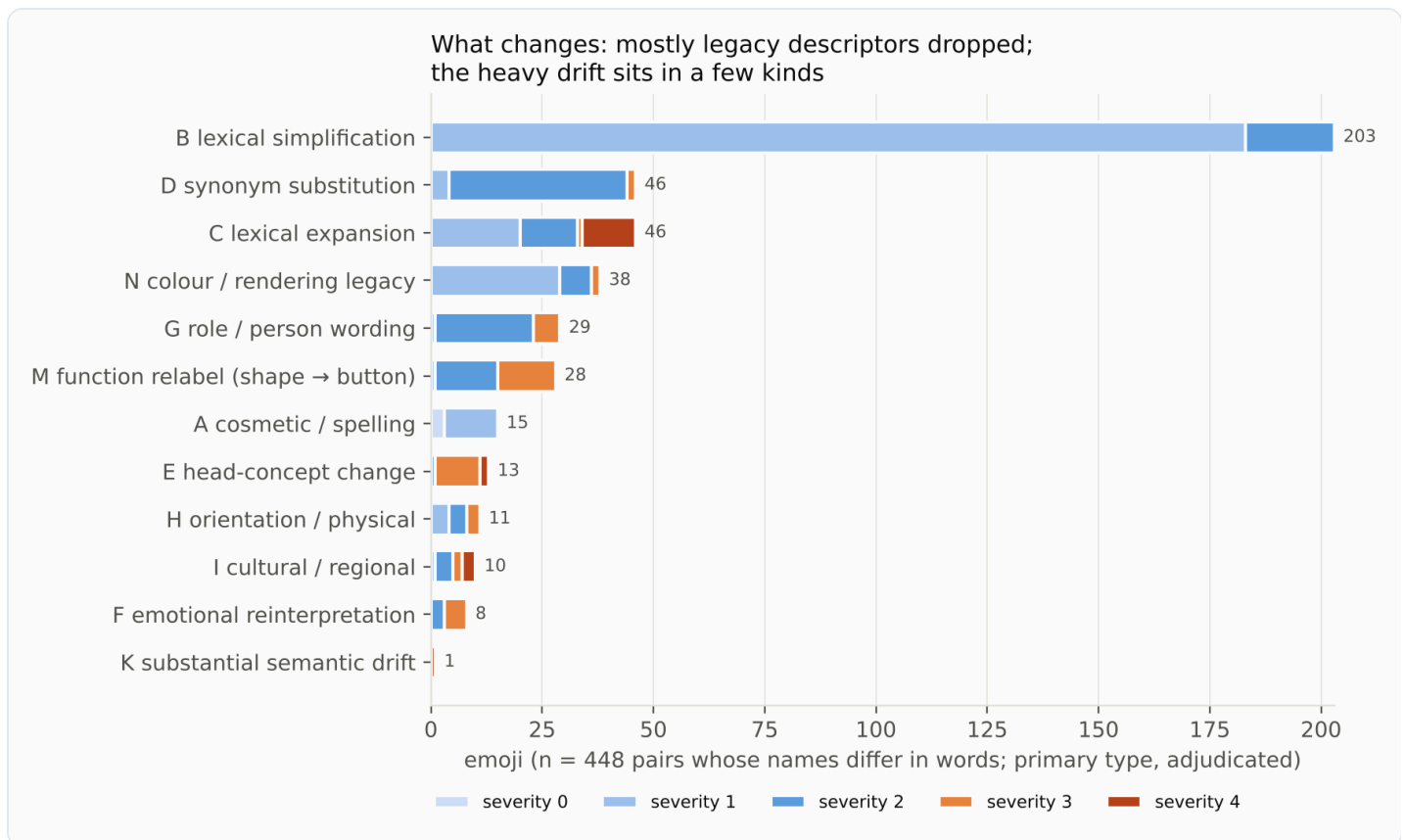
What changes

Most differences are Unicode's older vocabulary being dropped. The words that disappear most often are *face* (51 emoji), *symbol* (34), *squared* (29), *black* (28), *sign* (27) and *clock* (24). The words that appear most often are *button* (55), *person* (28) and *Japanese* (17).

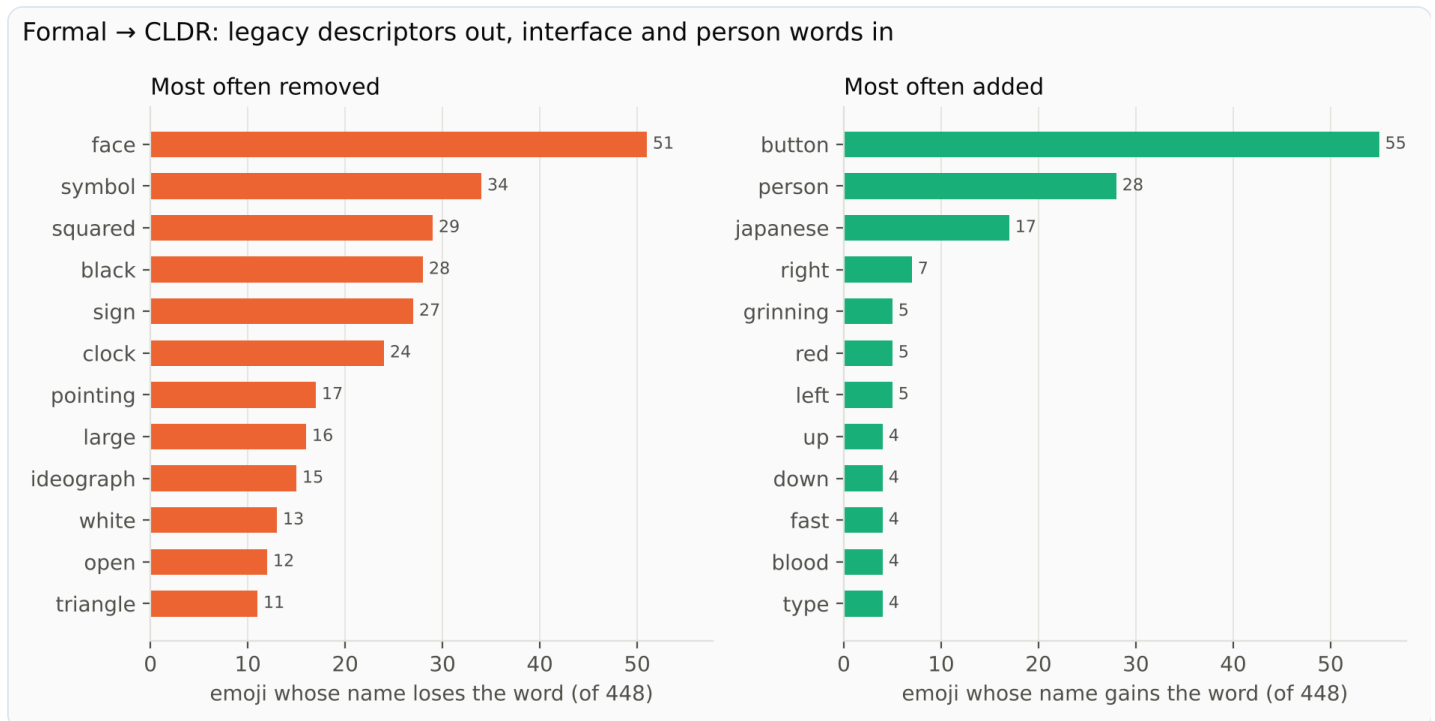
CLDR names are almost always the shorter ones. Where the names differ, the CLDR name is shorter in 383 of 448 pairs (median 13 characters against 20) and shorter in words in 321 (median 2 words against 3).



Name length in characters for the 448 differing pairs. Points below the dashed line have a shorter CLDR name.



The 448 differing names by main kind of change and severity. Simplification dominates, and it is almost always severity 1 or 2.



Words that leave the Unicode name and words that enter the CLDR name, counted over the 448 differing pairs.

The heavy changes fall in a few kinds. Function relabels turn a shape into a job: *black right-pointing triangle* becomes *play button* (28 emoji). Colour and rendering legacy removes Unicode's black-and-white convention, which describes filled versus outline and not hue (38 emoji). Person wording follows a Unicode policy of using "person" for human-form emoji (29 emoji).

Eighteen emoji score 4. Eleven of them have a formal name that is a code point, such as SQUARED CJK UNIFIED IDEOGRAPH-6708, which CLDR turns into *Japanese "monthly amount" button*. The rest are cases where the formal name is opaque or points at something else, such as *permanent paper sign* for *infinity* and *hocho* for *kitchen knife*.

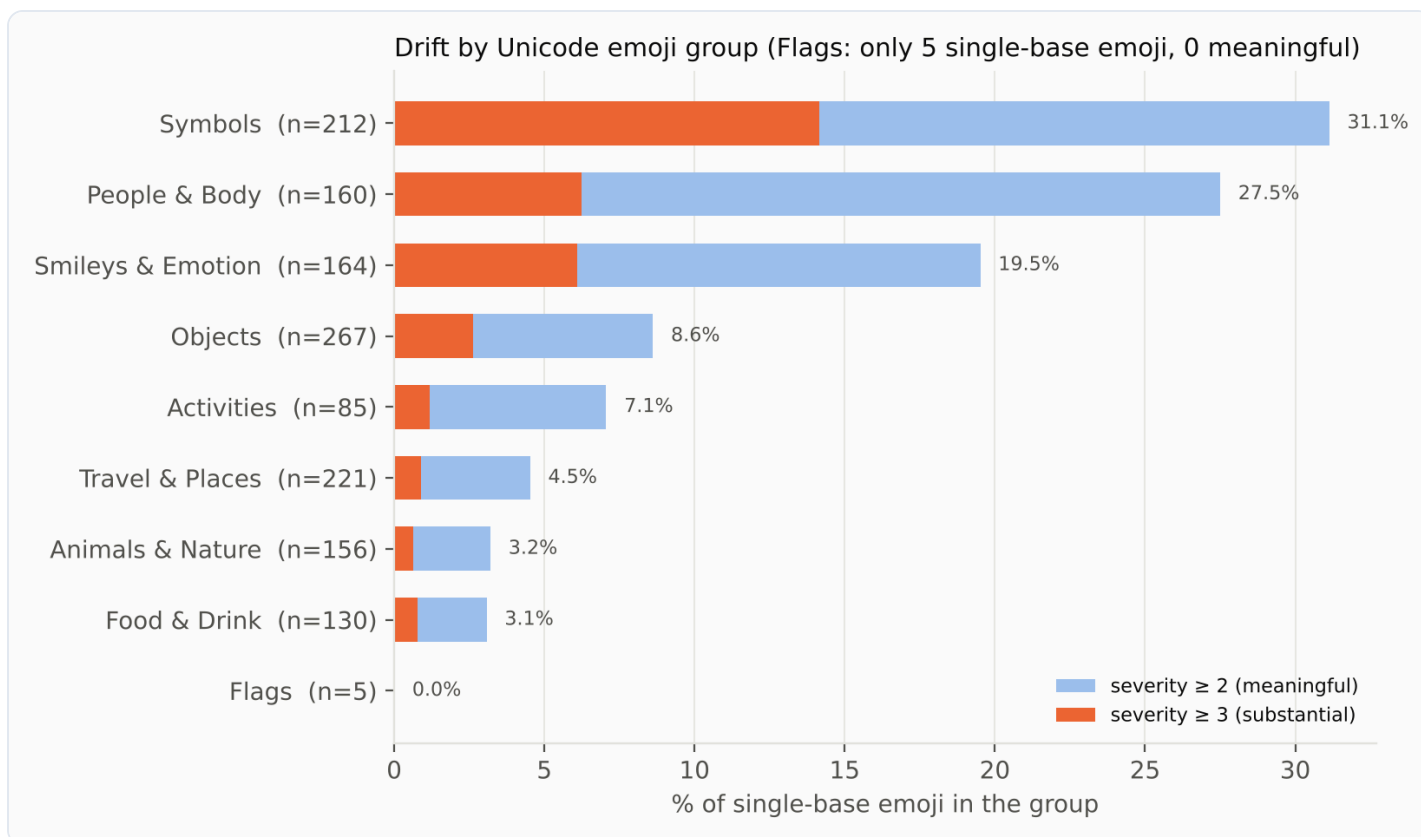
The main noun changes in 75 of the 448 pairs (16.7%), judged by hand. A simple rule that guesses the head noun from word order flagged 258, so it is not used here. Counting words alone also misleads: edit distance, which punishes long names, was the weakest guide to how much a change matters, and a sentence-meaning score was among the best (0.90 area under the curve for scores of 3 or more).

Two formal names are now another emoji's CLDR name, which is the one place where the gap has a practical cost. GRINNING FACE WITH SMILING EYES is the CLDR name of 😊, and searching it finds the wrong face. HEAVY BLACK HEART contains "black heart", which is now the name of 🖤.

RESULTS

Which emoji differ most

Symbols differ most, then people, then faces. Animals, food and travel hardly differ at all. Group sizes are shown because the flag group has only 5 emoji with a formal name and should be ignored.



Share of single-character emoji in each Unicode group whose names differ meaningfully (score 2 or more, light bars) or substantially (score 3 or more, orange bars).

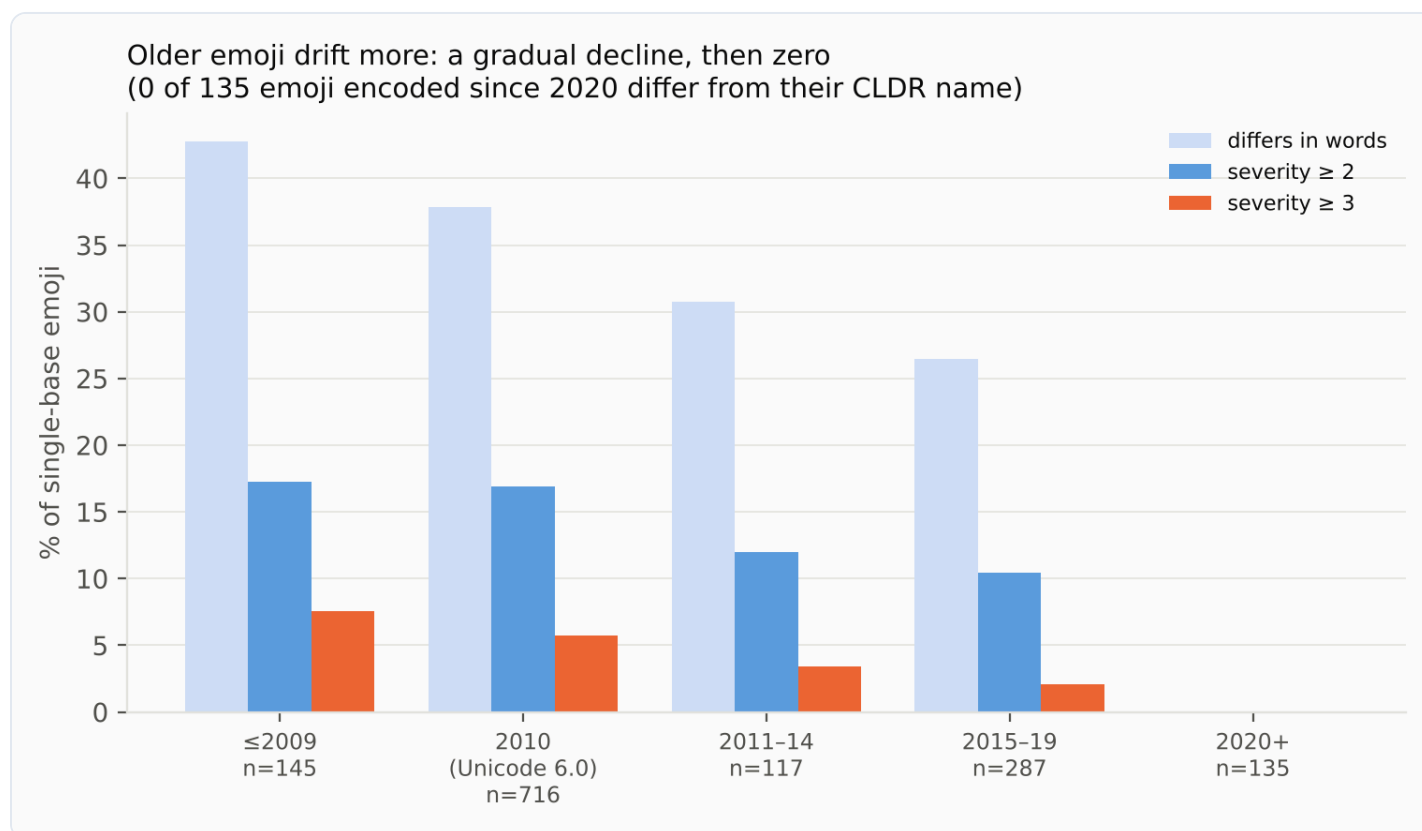
GROUP	EMOJI	DIFFER IN WORDS	SCORE 2 OR MORE	SCORE 3 OR MORE
Symbols	212	137 (64.6%)	31.1% (25.3 to 37.7)	14.2%
People & Body	160	71 (44.4%)	27.5% (21.2 to 34.9)	6.2%
Smileys & Emotion	164	57 (34.8%)	19.5% (14.2 to 26.2)	6.1%
Objects	267	57 (21.3%)	8.6% (5.8 to 12.6)	2.6%
Activities	85	31 (36.5%)	7.1% (3.3 to 14.6)	1.2%
Travel & Places	221	66 (29.9%)	4.5% (2.5 to 8.1)	0.9%
Animals & Nature	156	15 (9.6%)	3.2% (1.4 to 7.3)	0.6%
Food & Drink	130	11 (8.5%)	3.1% (1.2 to 7.6)	0.8%
Flags	5	3 (60.0%)	0.0%	0.0%

The people-versus-objects gap needs a caveat. People & Body sits 15.7 points above the rest of the set (interval 8.6 to 22.5), but 28 of its 44 meaningful differences are one rule: formal names such as *weight lifter* became *person lifting weights*. Without that rule the group is at 10.0%, level with Objects. The extra distance for people is mostly one documented policy, not a general habit of ambiguity. Symbols is highest for a different reason: 28 function relabels and 11 Japanese button names that are code points.

Emoji written as a character plus the emoji-style selector differ more: 43.5% differ in words against 30.0% for the others, and 18.8% differ meaningfully against 12.7%. The old text-style symbols in that group carry more of Unicode's black, white and heavy vocabulary (38 of 207 names, against 33 of 1,193).

Does age predict it?

Older emoji differ more, but the pattern is a cliff, not a slope. Of the 135 emoji encoded in 2020 or later, none differs from its CLDR name. Among older emoji, age adds almost nothing once the group is accounted for (0.05 severity points per decade, interval -0.04 to 0.14).



Share of emoji whose names differ, by the year Unicode encoded the base character. None of the 135 encoded since 2020 differs.

INTERPRETATION The cliff is consistent with Unicode naming new characters in CLDR's style from about 2020. We found no policy statement saying so, so treat it as a reading of the data. It does not support the

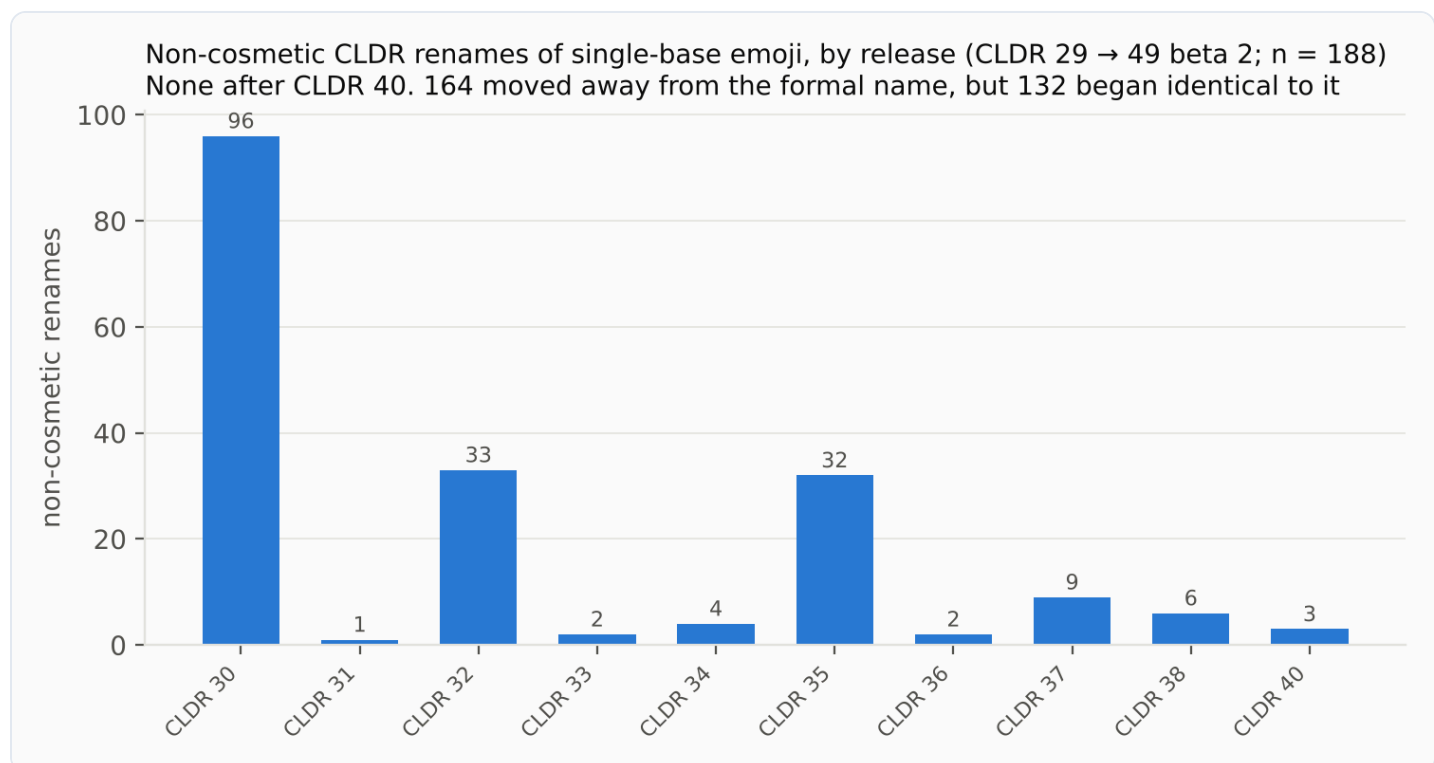
idea that formal names stay still while language moves: most of the gap was in place from CLDR's first release, as the next section shows.

RESULTS

Is it drift? What CLDR's rename history shows

A gap measured on one day is not change over time. To test for drift, the study read the English short names in every CLDR release from 29 to 49 beta 2 (21 releases that carry names). We use **divergence** for the gap on one date and **drift** only for CLDR renames between releases.

CLDR renamed 171 of the 1,400 emoji in ways that go beyond case or an ampersand, in 188 separate changes. Almost all fell in the early releases: 96 in CLDR 30, 33 in CLDR 32 and 32 in CLDR 35. The last non-cosmetic rename was in CLDR 40.



Non-cosmetic CLDR renames of single-character emoji, by release. Releases 41 to 49 beta 2 changed none.

Most of the gap predates the history. When CLDR first named these emoji, 337 already differed from the formal name. Today 448 do. Of those 448, 331 (73.9%) differed from the start, and 117 departed later, only 8 of them after CLDR 35.

Renames mostly moved names away from the formal one: 164 of 188 did, 14 moved closer and 10 did not change the overlap. That headline needs care. 132 of the 188 started from a name identical to the formal

one, so they could only move away, and 96 are the single step from CLDR 29 to 30. Among renames that did not start identical, 38 moved away and 12 moved closer.

Some renames moved away from everyday speech. CLDR 40 replaced *hugging face* with *smiling face with open hands*. CLDR 38 added a colour to *question mark* (now *red question mark*) and narrowed *pistol* to *water pistol*. Others went the other way: *dancer* and *weight lifter* became *woman dancing* and *person lifting weights*.

CLDR's release notes record English name updates from the Emoji Subcommittee (CLDR 32), a review and revision of names and keywords (CLDR 34), refinements for Emoji 12 (CLDR 36) and the renames of *dizzy face* and *hugging face* (CLDR 40). They give a reason for almost none of the individual names, so this study does not claim a reason for any single rename.

RESULTS

Keywords close part of the gap

A short name is not supposed to hold every synonym. CLDR keeps those in a separate keyword list, and it is that list that bridges the formal name to everyday words. Of the 424 differing pairs where a word left the name, 245 (57.8%) keep at least one of those words as a keyword, and 146 (34.4%) keep all of them.

In 27 pairs the whole formal name is itself a keyword, for example *hocho* on 🔪 and *aubergine* on 🍆. In only 12 of the 448 pairs does no word of the formal name survive anywhere in the name or keywords. A comparison of names alone would have overstated what a searcher loses. A second tokenizer gave 244 of 427 for the same count, so the safe statement is about 58%.

RESULTS

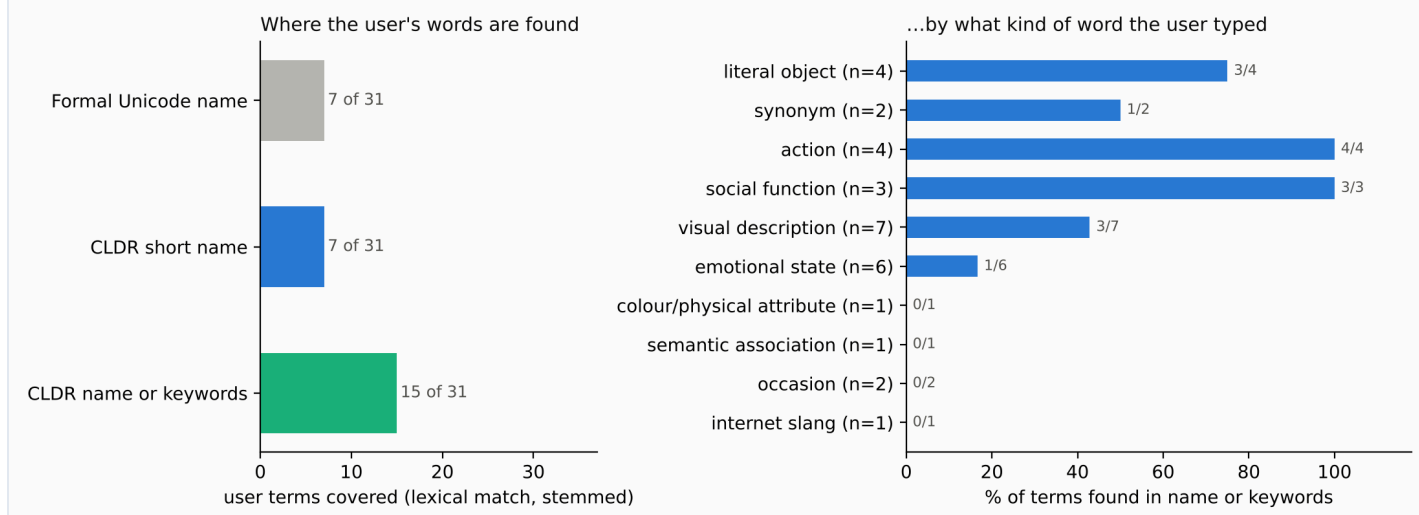
The words people use

☐ DIAGNOSTIC SAMPLE **Read this first**

This section rests on 9 Reddit threads (2017 to 2026) that were picked by hand because they show people struggling to find or name an emoji. Threads where nobody complained are invisible, and one is a post by someone promoting their own emoji search tool. It shows which kinds of failure exist, not how common they are.

From the nine threads the study took 31 words people typed or proposed and checked each one against the Unicode name, the CLDR name and the CLDR keywords. Seven appear in the formal name and seven in the CLDR short name. Fifteen appear somewhere in the CLDR name or keywords, eight of them in keywords only. Sixteen appear nowhere.

DIAGNOSTIC SAMPLE, not a benchmark: 31 terms from 9 Reddit threads, hand-picked

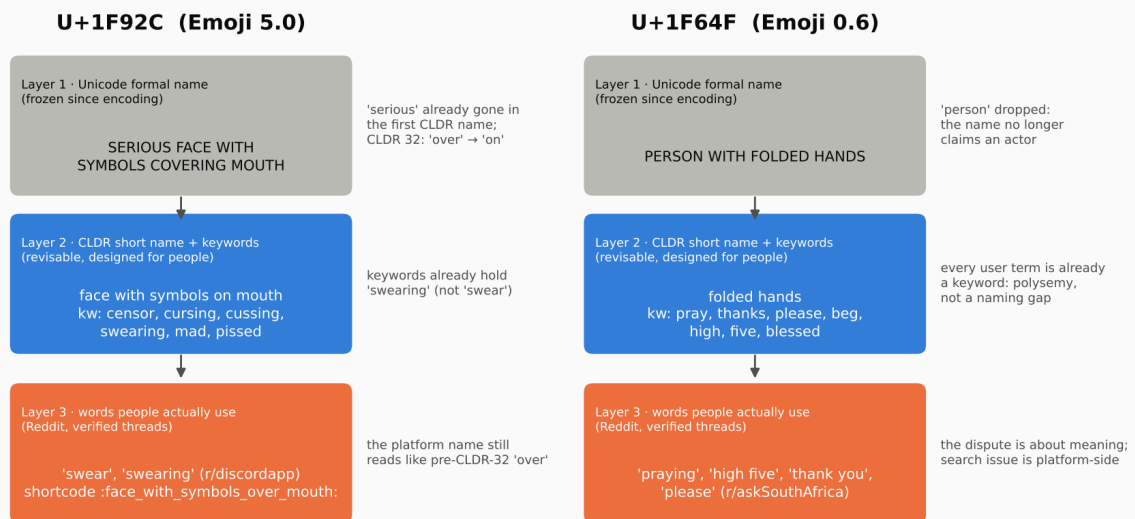


Diagnostic sample of 31 words from 9 Reddit threads. Hand-picked threads, not a benchmark.

The pattern is in the kind of word. Actions and social functions were found 7 times out of 7, and literal objects 3 out of 4. Feelings were found 1 time out of 6, and picture descriptions 3 out of 7. Occasion, slang and association words (St. Patrick's Day, *bottom emoji*, *think for* 🧠) were found 0 times out of 4. Of the 16 misses, only two are object or synonym words.

Across all 1,400 emoji, *Patrick* and *Nevada* appear in no name or keyword. *Shore* and *snot* appear nowhere, and *sigh* only on 😞. The occasion vocabulary people type is simply not in the data for any emoji.

Three layers, one emoji: stable identity, a revisable name, and the words people use



Three layers for two emoji: the frozen Unicode name, the revisable CLDR name and keywords, and the words people typed.

Six cases

- 🔪 **axe, hatchet, tomahawk.** The poster could use the emoji elsewhere but not find it in a chat app's search. The CLDR keywords for 🔪 are *ax, axe, chop, hatchet, split, wood*, and they were the same in CLDR 36 and 39. Only *tomahawk* was missing, so this is not a vocabulary gap. We could not test the app, so the cause is its own matching or data.
- 🙏 **praying, high five, thank you.** The name *folded hands* dropped the word *person*, and every word the thread offered is already a keyword. The argument is about meaning, which no name can settle. This is the cleanest case of meaning drift being a different thing from naming drift.
- 😴 **snoring, sighing, exhaustion.** Both names are *sleepy face*. Readers disagree about what it shows, and none of their words (*snore, sigh, tear, sweat*) is a keyword, though *sleep, tired* and *crying* are.
- 🗿 **swear.** One chat app's emoji code reads `face_with_symbols_over_mouth`, which equals the CLDR name before CLDR 32 changed *over* to *on*. That match is our inference, because the app does not document where its names come from. *Swearing* is a keyword.
- 🙄 **bottom emoji.** A joke nickname in one reply. No standard would adopt it, and it is a fair search alias for a product to add.
- 🧠 **think.** *Thinking* matches 🙄 by name. 🧠 is linked to the idea only by association (*intelligent, smart*), which is what the person asking for semantic search wanted.

READING

What this means for emoji search

▶ PRODUCT IMPLICATION **□ EDUCATED GUESS** The leverage for better emoji search is in aliases and meaning, not in new names. The basis: names are identifiers and CLDR keywords already carry much of the old vocabulary (58% of differing pairs), while the 16 misses in the sample are occasions, feelings, picture descriptions and slang that no standard lists.

Before fixing a failed search, sort it into one of three causes. A **vocabulary gap** means CLDR lacks the word (*tomahawk, snore, St. Patrick's Day*). A **platform failure** means CLDR has the word and the app still misses it (*hatchet, high*). A **meaning dispute** means people agree on the word and disagree on the emoji (🙏). The fixes differ, and looking the word up in the CLDR keyword list takes a minute.

Language adds a further layer this study did not measure. Unicode says annotations are language-specific, and one thread describes a German speaker who could search in English but not German on a phone; commenters blame the active keyboard language, with no vendor documentation to confirm it. CLDR's German data had names and keywords for every emoji we spot-checked, and gaps for 🍀 and 🍀. A German, French or Japanese version of this study is the natural next step.

We found no vendor documentation of semantic, meaning-based emoji search from Apple, Google or Microsoft. Absence in our search is not proof that none exists.

LIMITS

Limitations

- **The severity labels come from language models.** Two models from the same family labeled all 448 differing pairs without seeing each other, and a third pass resolved the 166 where they split or either scored 3 or more. They agreed on exact score 71.7% of the time and within one point 98.4% of the time. No person has reviewed the labels, so every label and its one-line reason is in the data file for you to check or overrule.
- **Score 4 leans on one rule.** We scored code-point names as 4 because they are opaque. That rule accounts for 11 of the 18.
- **The CLDR 49 data is beta 2.** The names Unicode published for Emoji 18.0 equal it for all 1,400 emoji, so we expect the final release to change little. That is an educated guess based on that match, not a check against the final files. Later versions keep this address and say what changed.
- **Only the renames are measured over time.** The Unicode-versus-CLDR gap is a snapshot. CLDR 29 to 30 mixes first-draft names with first-release names, and the history does not separate them.
- **English only.** The names, keywords and Reddit sample are all English.

- **The Reddit sample is small and hand-picked.** We read the threads through a public archive because Reddit blocks automated reads, so vote counts are snapshots. Two secondary threads could not be found and are not used.
- **“Found” means present in CLDR’s data.** It does not mean an app will return the emoji. Apps may stem or prefix-match words differently, and we tested no app.
- **Drawings are not compared.** Studies by others attribute much emoji misreading to how a platform draws the emoji. This study measures names only.

METHOD

How the comparison was made

Names come from Unicode 18.0.0 (`UnicodeData.txt` for formal names, `emoji-test.txt` for CLDR short names as published) and from the CLDR English annotation files for releases 29 to 48 and 49 beta 2 (keywords and rename history). Every source file was fetched on October 12, 2026 and its SHA-256 hash is in `source-files.csv`.

Which emoji count. An emoji is in the comparison if it is fully qualified and is one base character, optionally followed by U+FE0F, with a name in `UnicodeData.txt`. That gives 1,400. Nine building-block components and 2,563 sequences are left out because they have no formal name. The 12 keycap sequences have formal names as named sequences but are not compared.

Normalization. Names are lower-cased, accents are removed, hyphens and punctuation become spaces, apostrophes are dropped and & is read as *and*. Anything that still differs after that is a difference in words.

Severity. Each of the 448 differing pairs got a score from 0 to 4 and a main kind of change (A to N, listed under the examples). The rules were written down before labeling. Two language models labeled every pair separately, a third pass applied written tie-break rules to the 166 pairs where the two split or either scored 3 or more, and we report the spread between the two models as the uncertainty. The data dictionary explains each column.

Text measures. Character similarity, word overlap, which words left and entered each name, whether a word that left is still a keyword, and a sentence-embedding similarity (all-MiniLM-L6-v2). None is treated as ground truth. Severity is the measure of how much a change matters.

Rename history. The English short name of each emoji was read in every CLDR release and compared with the release before. A change that is only case or & versus *and* is cosmetic and not counted as a rename.

Reddit sample. We took the words people typed or proposed from 9 threads, matched each against the formal name, the CLDR name and the keywords after stemming, and added the closest CLDR word and a likely cause by hand. Usernames are not recorded.

You can reproduce the 1,400 and the 952 and 448 split from two public files with reproduce-denominator.py.

RECORD

Corrections

Version 1.0, October 13, 2026. No corrections yet. A corrected version keeps this address, gets a new version number and says what changed.

DATA

Data and files

Full study, version 1.0

PDF · 1.8 MB

The same text as this page, with title, version and canonical URL on every page.

emoji-name-comparison.csv

CSV · 318 KB

Start here: all 1,400 emoji with both names, keywords, severity, kind of change and the reason for each label.

FILE	CONTENTS	SIZE
emoji-name-comparison.csv	The 1,400 emoji with a formal name: both names, keywords, group, Unicode and Emoji version, severity, kind of change, reason, text measures	318 KB
emoji-universe.csv	All 3,972 emoji in Emoji 18.0 with type, CLDR name and keywords	880 KB
cldr-name-changes.csv	Every English short-name change from CLDR 29 to 49 beta 2, for all emoji types	82 KB
single-base-renames.csv	The renames of the 1,400 emoji, with whether each began identical to the formal name and which way it moved	24 KB
reddit-vocabulary-sample.csv	The 31 words from Reddit threads, where each was found, closest CLDR word and likely cause. Thread, month and link only	9 KB
summary-statistics.json	Every figure on this page as machine-readable numbers	12 KB
source-files.csv	Address, size and SHA-256 hash of every Unicode and CLDR file used	12 KB

FILE	CONTENTS	SIZE
data-dictionary.csv	What the main columns mean	2 KB
reproduce-denominator.py	Standard-library script that rebuilds the 1,400 and the 952 and 448 split from two Unicode files	2 KB

CITE

How to cite this study

UltraTextGen Research. "Emoji Name Drift: How Unicode's Stable Names Diverge From the Words People Actually Use." Version 1.0, October 13, 2026.
<https://ultratextgen.com/research/emoji-name-drift/>

Please cite the version number and say that the severity labels are model-assigned.

SOURCES

External sources

That a formal Unicode name is an identifier that is never changed once assigned comes from the Unicode Standard's chapter on character names (chapter 4, section 4.8 (<https://www.unicode.org/versions/Unicode17.0.0/core-spec/chapter-4/>)) and the Consortium's Stability Policy (https://www.unicode.org/policies/stability_policy.html), which also says names do not always express a character's full meaning and that outright errors get a formal alias.

That a CLDR short name can change, that sequences have only CLDR names, that gender-neutral "person" names were adopted for human-form emoji, and that annotations are language-specific come from the Consortium's Unicode Emoji report (UTS #51) (<https://www.unicode.org/reports/tr51/>); how short names and keywords are defined comes from UTS #35, Annotations (<https://www.unicode.org/reports/tr35/tr35-general.html#Annotations>).

CLDR's goals for short names (unique, short, descriptive, not literal) and its guidance that keywords are what people might search for in addition to the name are in the CLDR project's short names and keywords guide (<https://cldr.unicode.org/translation/characters/short-names-and-keywords>).

The data files are UnicodeData.txt 18.0.0 (<https://www.unicode.org/Public/18.0.0/ucd/UnicodeData.txt>), emoji-test.txt 18.0 (<https://www.unicode.org/Public/18.0.0/emoji/emoji-test.txt>) and the English annotation files in the CLDR

repository (<https://github.com/unicode-org/cldr/tree/release-48/common/annotations>), with the full list of files and hashes in the data section.

That CLDR revised English names in specific releases is recorded in the release notes for CLDR 32 (<https://cldr.unicode.org/downloads/cldr-32>), CLDR 34 (<https://cldr.unicode.org/downloads/cldr-34>), CLDR 36 (<https://cldr.unicode.org/downloads/cldr-36>) and CLDR 40 (<https://cldr.unicode.org/downloads/cldr-40>).

That two people rarely pick the same word for the same thing comes from Furnas, Landauer, Gomez and Dumais, "The vocabulary problem in human-system communication" (<https://doi.org/10.1145/32206.32212>) (1987), where two people chose the same term with probability under 0.20. That people read the same emoji rendering differently comes from Miller and colleagues, "Blissfully Happy" or "Ready to Fight" (<https://doi.org/10.1609/icwsm.v10i1.14757>) (2016), and that emoji meaning shifts in use over time comes from Robertson and colleagues, "Semantic Journeys" (<https://arxiv.org/abs/2105.00846>) (2021).

The words people typed come from these threads: r/Emoji, March 2021

(<https://www.reddit.com/r/Emoji/comments/m4gmn7/>) (👉), r/askSouthAfrica, July 2026

(<https://www.reddit.com/r/askSouthAfrica/comments/1uorpv/>) (🙏), r/Emoji, March 2026

(<https://www.reddit.com/r/Emoji/comments/1s1z82m/>) (😭), r/discordapp, April 2020

(<https://www.reddit.com/r/discordapp/comments/g7mz0y/>) (🤖), r/MisreadSprites, October 2026

(<https://www.reddit.com/r/MisreadSprites/comments/1wyge2n/>) (🐱), r/monarchapp, April 2026

(<https://www.reddit.com/r/monarchapp/comments/1stwqrq/>) (🧠), r/Emoji, March 2026

(<https://www.reddit.com/r/Emoji/comments/1rv7ooe/>) (🍀), r/apple, January 2017

(<https://www.reddit.com/r/apple/comments/5nukgw/>) (😞), r/ios, January 2020

(<https://www.reddit.com/r/ios/comments/eta6lb/>) (😍), and, for the German search report, r/ios, December 2024

(<https://www.reddit.com/r/ios/comments/1hjtjy/>).