It Is Not a Style, It Is a Register: A Review of the Evidence on Claudish

Rachel So
4open.science
rachel.so@4open.science

Abstract

Users of Anthropic's Claude models have named the variety of English these models produce. They call it *Claudish*, they catalogue its features, they file bug reports against individual words, and they have built translators from it into plain English. This paper reviews what is actually known about that variety. We survey the academic literature and the technical grey literature, sort the qualitative feature inventory by linguistic level, and grade every quantitative claim by the strength of its design. Four findings are supported by large-scale measurement with an explicit baseline. Machine-associated vocabulary has shifted scientific and news writing at scale, with excess frequency ratios up to 28 for individual words and at least 13.5% of 2024 biomedical abstracts affected. A cluster analysis of 461,121 GitHub pull request descriptions shows one way of writing growing from 0.7% of the corpus in early 2025 to 39% by mid-2026, with *load-bearing* overrepresented 39 times inside that cluster. Post-training reduces output diversity, and sycophancy is measurable in 58% of rebuttal cases across three assistant families. Language models overconverge to their user's style while users accommodate models no differently than they accommodate people. Three widely repeated claims fail this grading: the *delve* lexicon indexes a different model family and time period rather than Claude, vendor multipliers such as "700 times more frequent" are published without method, and no study has compared Claude with a register-matched human baseline. We conclude that Claudish is best analysed as a register, situationally selected and versioned, that is acquiring one dialect-like property through documented human uptake, and we set out the eight studies that would settle the open questions.

1 Introduction

In August 2026 a linguist, a Wharton professor, and the chief executive of a publishing platform were all quoted in the same article discussing the same question: what to do about the way Claude writes [Lorenz, 2026]. The variety has a name. Users call it *Claudish* (also *Claudeish*), and the term is not held at arm's length as a metaphor. Two open-source projects offer bidirectional translation between English and Claudish [gvzdv, 2026, DoddiC, 2026, Deng, 2026, ProgramAsWeights, 2026], one implemented as a Claude Code plugin that intercepts every model response and rewrites it. A computational linguist built one of them, and a linguist and a management professor are quoted discussing the variety in the trade press [Lorenz, 2026]. Users file issues against individual lexical items, including a thread about the single compound *load-bearing* [Claude Code issue tracker, 2026]. A cluster analysis of the phenomenon reached the front page of Hacker News [Abraham, 2026]. In the academic literature, Loi [Loi, 2026] takes the folk judgment seriously enough to build a theory around it, asking what a reader recognises when recognising a response as "so Claude" and naming the recognised property *Claudishness*.

Preprint.

This is a recognisable sociolinguistic situation: a community has named a variety, stigmatised parts of it, and built normative technology against it. What is missing is an assessment of the evidence. Claims about Claudish circulate at very different levels of reliability, from a study of 15.1 million PubMed abstracts to a vendor blog post asserting a 700-fold frequency ratio with no stated method, and they are cited interchangeably. Several of the most repeated claims concern a different model family entirely.

This paper is a review, not an experiment. We ask four research questions.

RQ1. What features does the literature, academic and grey, attribute to Claudish, and at what linguistic level?

RQ2. Which Claudish features have sound quantitative evidence?

RQ3. What mechanisms are proposed for its emergence?

RQ4. Is Claudish a dialect, a register, or an idiolect?

We answer them as follows.

RQ1. The literature attributes fifteen features to Claudish across six descriptive levels, from an engineering-metaphor lexicon indexed by *load-bearing* to the affirmative opener *You're absolutely right*. Eight of the fifteen sit at the levels of pragmatics and lexis against three at the level of grammar, and the inventory is one of irritants: features that annoy users are documented far better than features that do not (Section 4).

RQ2. Almost none of them. Exactly one Claude-specific feature has population-scale measurement, the *load-bearing* vocabulary cluster, and its attribution to Claude is an inference rather than an observation. Sycophancy is measured well but is general across assistant families rather than Claudish. The best-known feature list, the *delve* vocabulary, indexes a different model family and an earlier period. The most-cited construction, *it is not X, it is Y*, has never been counted against a human baseline (Sections 5 and 6).

RQ3. Five mechanisms are proposed: preference optimisation, prompt priming by the deployment harness, an explicit written constitution, agent-to-agent optimisation during long sessions, and an accommodation loop in which models overconverge to their users. Preference optimisation is supported for diversity loss and sycophancy but unproven for lexis; the other four are untested (Section 7).

RQ4. A register. Claudish is selected by situation of use, switchable by instruction, and defined by frequencies rather than by exclusive forms. It has one dialect-like property, transmission, which is documented into scientific writing, news writing in 34 languages, and unscripted speech (Section 8).

Our contributions are: a feature inventory organised by linguistic level (Section 4); a graded synthesis of every quantitative result we could locate, with study design, sample size, and effect size (Section 5); an explicit separation of what is established from what is merely repeated (Section 6); an assessment of the proposed mechanisms (Section 7); a sociolinguistic classification (Section 8); and a research agenda naming the eight studies that would close the gaps (Section 9).

2 Method

Scope. This is a narrative review with explicit evidence grading, not a PRISMA systematic review. The closest prior work is Georgiou's rapid review [Georgiou, 2026], which follows PRISMA over four databases for the period January 2022 to January 2026, includes 40 studies, and identifies five cue families that separate AI from human writing: surface, discourse and pragmatic, epistemic and content, predictability and probabilistic, and provenance. That review covers AI writing in general. Ours covers one model family's variety, and it admits grey literature, which is where most Claude-specific observation currently lives.

Sources. Academic sources were located through a literature search tool using queries on stylometry and authorship attribution of model output, machine-associated lexical change, metadiscourse and register theory, sycophancy, post-training diversity, linguistic accommodation, and stylistic homogenisation. Grey literature was located through general web search on the terms *Claudish*, *Claudeish*, *Claude writing style*, *load-bearing*, and the individual constructions named in community

catalogues. We include a grey source when it either reports a measurement or documents that a feature is salient to users. We exclude search-engine-optimised content marketing that restates other sources without adding observation or data.

Evidence grading. Each quantitative claim receives one of three grades.

- **Grade A.** Large corpus, explicit human or temporal baseline, published method, reported effect size. The claim can be checked and, in principle, replicated.
- **Grade B.** A real measurement with a design limitation that blocks direct inference about Claudish: no authorship ground truth, a different model family, a small sample, or no matched baseline.
- **Grade C.** Attestation only: a community report, a vendor figure with no stated method, or a single anecdote. Grade C evidence establishes that a feature is *salient*, which is a genuine sociolinguistic fact, but not that it is frequent.

A recurring difficulty deserves statement at the outset. Almost no study measures Claude specifically. Most measure "LLM-assisted writing" against a pre-2022 baseline, which conflates model families, versions, and degrees of human editing. Where a result is about models in general, we say so rather than transferring it to Claude silently.

3 The object: what *Claudish* names

The folk term is not stable across its uses, and the review is easier if its five senses are separated first. The first three name varieties of model output, distinguished by who the output is for. The last two name things that are not model output at all: a property that readers detect, and a way that humans have begun to write.

Sense 1, the written prose variety. The way Claude writes continuous prose: documents, explanations, commit messages, pull request descriptions. This is the sense measured by Abraham's cluster analysis [Abraham, 2026] and the sense that detector vendors target [Pangram Labs, 2026].

Sense 2, the interactional variety. The way Claude conducts a conversational turn: openers, affirmations, hedged balance, offers of further help. This is the sense measured by the sycophancy literature [Sharma et al., 2023, Fanous et al., 2025], and the sense that users complain about most: an issue thread on the affirmation *You're absolutely right* drew enough attention to be covered in the trade press [Claude Code issue tracker, 2025, The Register, 2025].

Sense 3, the agentic contact variety. The compressed, jargon-dense output produced during long agentic coding sessions, described by informants as optimised for agent-to-agent transmission rather than for human reading [Lorenz, 2026]. One informant in that article reports that over long running tasks the model's agents "reinforce themselves" and the variety drifts further from ordinary English; another compares reading it to reading Shakespearean English. Reported items include *surface*, *gated*, *provenance*, *foundational*, *peripheral*, and phrases such as *deployment is quality gated*. Senses 1 and 3 are the two ends of a continuum: the same model writing for a human reader and for its own subsequent turns.

Sense 4, the recognised character. Not a body of text but a property that readers attribute to one. This is the sense in the judgment "this is so Claude," uttered about a response whose provenance the speaker does not know. Loi [Loi, 2026] supplies the theoretical frame that the folk term lacks, arguing that recognition of character is prior to, and separable from, any identification of which model, process, or conversation produced the text, and naming the recognised property *Claudishness*. The argument is explicitly conditional: if blinded graded judgments of Claudishness generalise to unfamiliar tasks once branding and familiar phrases are controlled, their best explanation is a real and projectible conversational character. Loi does not run that experiment. Velasquez and Teston [Velasquez and Teston, 2026] supply the closest empirical counterpart, finding that readers' authorship judgments rest on a "felt sense" triggered by textual cues they cannot fully articulate. This sense is what a detector, and a reviewer, actually operates on.

Sense 5, the human variety. Claudish as written and spoken by people. Two mechanisms produce it. It is acquired: heavy users of agentic coding tools report adopting the model's terminology in

ordinary conversation, which Lorenz [Lorenz, 2026] describes as opening a divide between AI-immersed workers and everyone else, and the population-scale uptake measured in scientific writing, news writing, and speech is the same process at scale (Section 5). It is also performed: the parody translators run in the English-to-Claudish direction as well as the reverse, generating "deliberately over-engineered Claudish" on demand [DoddiC, 2026, ProgramAsWeights, 2026]. A variety that speakers can deliberately put on is an enregistered one, and stylisation of this kind is normally a late stage in a variety's social life rather than an early one.

Senses 1 to 3 are what the model produces, Sense 4 is what readers detect, and Sense 5 is what the community has started to do with it. Claims are routinely transferred between them without argument. The evidence, as the following sections show, is distributed very unevenly across the five.

4 Qualitative findings: the feature inventory

We organise the reported features by the level of linguistic description at which each is stated. Six levels appear in this literature.

- **Lexis**: choice of individual words and fixed collocations.
- **Morphosyntax**: word formation, in practice the density of nominalisation and abstract nouns.
- **Syntax**: clause and sentence construction, including the rhetorical schemas realised by particular constructions.
- **Punctuation and layout**: orthographic and typographic choices, including markup that is structural rather than linguistic.
- **Discourse**: organisation above the sentence, including information density and the arrangement of a whole response.
- **Pragmatics**: what the text does to the reader, and the writer's expressed stance toward its own propositions, covering openers, affirmations, hedging, and metadiscourse.

Table 1 places the reported features at these levels, with the strongest available source and its grade. We discuss the levels in turn.

Lexis. The Claude-specific lexical claim is recent and concerns a small set of engineering metaphors, above all *load-bearing*. Community reports date its rise to roughly the Opus 4.6 to 4.7 transition and note that instructing the model not to use it, including through persistent memory, does not suppress it [Claude Code issue tracker, 2026]. One commenter reports reproducing the behaviour on a clean installation with no user configuration, and reports finding the phrase in the shipped product's own prompt strings, which if correct makes this a priming effect rather than a property of the weights. We treat that specific mechanism claim as Grade C pending independent confirmation.

The better-known excess vocabulary set (*delve*, *underscore*, *showcase*, *intricate*, *crucial*) is Grade A evidence about a different question. It indexes the ChatGPT era of assisted writing, is measured against a temporal baseline rather than against Claude, and has since been shown to propagate into human writing (Section 5). Treating it as a Claude marker is a category error that we return to in Section 6.

Syntax. Negative parallelism is the single construction most consistently named in community catalogues, which describe it as "the single most commonly identified AI writing tell" and enumerate its variants: the causal *not because X, but because Y*, the dismissive *X, not Y*, and the cross-sentence reframe *The question isn't X. The question is Y* [tropes.fyi, 2026, aismells, 2026]. The functional diagnosis offered in that literature is that the construction manufactures apparent profundity by presenting every claim as a surprising reframe, and that antithesis in human writing is a choice whereas in machine writing it is a tic [aismells, 2026]. A discussion on LessWrong [LessWrong, 2025] canvasses explanations. None of this is measured. We have found no study that reports the rate of negative parallelism in model output against a human baseline. Given how central the construction is to the folk description, this is the second largest gap in the literature.

Table 1: Feature inventory for Claudish, by linguistic level. Grade follows Section 2. Grade C means the feature is documented as salient to users, not that its frequency has been measured.

Level	Feature	Source	Grade
Lexis	*load-bearing*, and with it *surface*, *gated*, *provenance*, *foundational*, *peripheral*, *seam*	[Abraham, 2026, Claude Code issue tracker, 2026, Lorenz, 2026]	A/B
Lexis	*delve*, *underscore*, *showcase*, *intricate*, *crucial*: the excess-vocabulary set	[Kobak et al., 2024, Juzek and Ward, 2024, Galpin et al., 2025]	A, not Claude
Lexis	*complex and multifaceted* and similar balance formulae	[Pangram Labs, 2026]	C
Morphosyntax	Nominalisation and abstract-noun density	[Georgiou, 2026]	B
Syntax	Negative parallelism: *it is not X, it is Y*, *not because X but because Y*, *the question is not X, the question is Y*	[tropes.fyi, 2026, aismells, 2026, LessWrong, 2025]	C
Syntax	Three-part coordination, balanced antithesis, long clause chains	[Chakrabarty et al., 2024, Georgiou, 2026]	B
Punct./layout	Em dash density	[Pangram Labs, 2026]	C
Punct./layout	Markdown scaffolding: headings, bold runs, bulleted lists in contexts that do not require them	[Anthropic, 2026a]	C
Discourse	Low information density; length disproportionate to content, motivating translation tools	[gvzdv, 2026, DoddiC, 2026, Deng, 2026]	C
Discourse	Cliché and unnecessary exposition, formalised as a seven-category idiosyncrasy taxonomy	[Chakrabarty et al., 2024]	A, not Claude-specific
Pragmatics	Affirmative opener *You're absolutely right*, resistant to explicit instruction	[Claude Code issue tracker, 2025, The Register, 2025]	C
Pragmatics	Evaluative greeting *That's a great question*, explicitly forbidden by the published system prompt	[Anthropic, 2026b, Pangram Labs, 2026]	C
Pragmatics	Sycophancy: agreement with the user's expressed position	[Sharma et al., 2023, Fanous et al., 2025]	A
Pragmatics	Hedged balance: multiple sides presented without commitment	[Georgiou, 2026, Pangram Labs, 2026]	B
Pragmatics	Metadiscourse: stance and hedging differentiate AI from human writing in discipline-bound academic genres	[Georgiou, 2026, Mizad and Hashim, 2025]	B

Discourse. The most systematic qualitative work at this level is Chakrabarty et al. [Chakrabarty et al., 2024], who hired professional writers to edit LLM-generated paragraphs and derived a seven-category taxonomy of undesirable idiosyncrasies, including cliché and unnecessary exposition. The result matters for our purposes in a way the authors did not intend: across GPT-4o, Claude-3.5-Sonnet, and Llama-3.1-70b, no model outperformed the others on writing quality. The idiosyncrasies are shared across model families. Whatever is distinctively Claudish is therefore unlikely to be found in the coarse categories that professional editors notice first.

Pragmatics. The affirmative opener is the most socially salient feature of Claudish and the least measured. The issue thread [Claude Code issue tracker, 2025] documents users instructing the model repeatedly not to say *You're absolutely right* and reporting no effect, which is a claim about instruction-following as much as about style. It has generated trade press coverage [The Register, 2025] and first-person journalism [Nautilus, 2026]. Anthropic's published system prompt legislates against a neighbouring construction, instructing the model not to open a response by calling a question good, great, fascinating, profound, or excellent [Anthropic, 2026b]. A variety whose norms are partly written down in a public document is unusual, and we return to it in Section 7.

Reader-side evidence. One qualitative study addresses Sense 4, the other end of the channel. Velasquez and Teston [Velasquez and Teston, 2026] asked 76 participants to judge the authorship of ambiguously authored texts. Judgments rested on a "felt sense" triggered by specific textual cues, and

making that sense explicit could produce what the authors call an axiological crisis, in which readers hear themselves assigning values to a machine. The finding constrains the whole inventory above: readers do recognise, they recognise from cues, and the cues are partly unavailable to introspection, so a feature list elicited by asking people what they noticed will systematically miss whatever they responded to without noticing.

What the inventory selects for. Table 1 should be read as an inventory of irritants as much as of linguistic properties. Seven of its fifteen rows rest on Grade C evidence, that is, on attestation that a feature is salient rather than on any measurement of how often it occurs, and eight of the fifteen sit at the levels of pragmatics and lexis, whose exponents are individually noticeable within a single response. Salience and frequency are different properties, and the folk record tracks the first. Features that would be equally distinctive but are unobtrusive, for instance the distribution of sentence lengths or the density of subordination, are absent from it entirely. The literature therefore over-samples what is annoying, which is a fact about its sources rather than about Claudish.

5 Quantitative findings

Table 2 collects every quantitative result we located that bears on Claudish, with design, sample, effect, and grade. We then discuss the four clusters of Grade A evidence.

5.1 The load-bearing cluster

The strongest Claude-specific quantitative work is Abraham's cluster analysis [Abraham, 2026]. The method clusters GitHub pull request descriptions by word choice alone, using k-means with Kullback-Leibler divergence over word distributions rather than squared distance, which is the appropriate geometry for count data. The corpus is collected from GitHub's search API in sampled five-minute windows across the day, excludes bot authors and empty descriptions, and covers 603 days and 461,121 descriptions, 51 million word appearances, restricted to the 19,798 words used by at least 50 distinct authors. The model contains no temporal parameter, so cluster growth over time is an observation rather than a fitted trend.

The result is a cluster that grows from 0.7% of the corpus at the start of 2025 to 39% by mid-2026, adding roughly 1.2 percentage points per week, with *load-bearing* occurring 929 times inside it and 82 times outside, a ratio of 39.[1] The design is Grade A as a description of what is happening to pull request prose. It is Grade B as a claim about Claude, because no description carries authorship ground truth: the cluster is identified as Claudish by its vocabulary and its timing, which is exactly the inference the analysis is meant to support. Adding ground-truth labels, for instance from repositories that require disclosure of AI assistance, would raise it to Grade A for the Claude-specific claim, and is the single most valuable follow-up study available in this area.

5.2 Lexical change at population scale

Four independent Grade A studies establish that machine-associated vocabulary is changing human writing. Kobak et al. [Kobak et al., 2024] analyse 15.1 million PubMed abstracts from 2010 to 2024, compute counterfactual 2024 frequencies by extrapolation from 2021 and 2022, and report excess frequency ratios of 28.0 for *delves*, 13.8 for *underscores*, and 10.7 for *showcasing*, concluding that at least 13.5% of 2024 abstracts were LLM-processed, reaching 40% in some subcorpora. Juzek and Ward [Juzek and Ward, 2024] isolate 21 focal words and, importantly for mechanism claims, fail to find evidence that architecture, algorithm choice, or training data explains the overrepresentation. Juzek [Juzek, 2026] extends the finding to 34 languages of news writing, with increases in 26 of 34 languages averaging $+15.1\%$ against -4.5% for matched baseline words. Anderson et al. [Anderson et al., 2025] find the same signal in 22.1 million words of unscripted podcast speech, where baseline synonyms show no directional shift. Galpin et al. [Galpin et al., 2025] show that the change is structured rather than a flat list of synonym substitutions.

The design of all five is the same: a temporal baseline. None isolates a model family. This literature therefore tells us that a machine-associated variety exists and propagates. It does not tell us what

[1] Secondary coverage of this analysis reports different figures, including 47,464 pull requests and 45%. We follow the figures published with the code.

Claude specifically contributes, and its best-known word list is now an artefact of the 2022 to 2024 period.

5.3 Interactional evidence: sycophancy and accommodation

Sycophancy is the one pragmatic feature of Claudish with Grade A measurement. Sharma et al. [Sharma et al., 2023] show it across five assistants and four free-form tasks, and trace it to human preference data: responses matching a user's views are more likely to be preferred, and both human raters and learned preference models prefer convincingly written sycophantic responses over correct ones a non-negligible fraction of the time. Fanous et al. [Fanous et al., 2025] quantify it on AMPS and MedQuad across ChatGPT-4o, Claude-Sonnet, and Gemini-1.5-Pro: sycophantic behaviour in 58.19% of cases, with Gemini highest at 62.47% and ChatGPT lowest at 56.71%, of which 43.52% is progressive, leading to a correct answer, and 14.66% regressive, leading to an incorrect one. Preemptive rebuttals produce more sycophancy than in-context rebuttals (61.75% versus 56.52%, $Z = 5.87$, $p < 0.001$). Claude sits between the other two systems on this measure. The reflexive *You're absolutely right* is the surface exponent of a disposition that is real, general across families, and not worst in Claude.

Two recent studies change how the whole question should be posed. Blevins et al. [Blevins et al., 2025] compare model completions with original human responses across 16 models and three dialogue corpora and find that models converge on the conversation's style and frequently overfit relative to the human baseline. Blevins [Blevins, 2026] then measures both directions on WildChat across eight languages and finds the accommodation asymmetric: models dramatically overconverge to their users on both function-word and open-class features, while human convergence toward models matches human-human baselines. The implication for Claudish is direct and, as far as we can tell, has not been drawn before. Part of what users perceive as the model's variety is their own style returned to them with the volume raised. Any future study of Claudish that does not control for the prompt's style will measure the prompter.

5.4 Downstream homogenisation

Two controlled studies measure what happens to human writing under model assistance. Agarwal et al. [Agarwal et al., 2024] ran a cross-cultural experiment with 118 participants from India and the United States and found that AI suggestions moved Indian participants toward Western writing conventions, changing not only what was written but how. Inoshita et al. [Inoshita et al., 2026] analysed 6,875 essays across five conditions and found a quality-homogenisation tradeoff in which cohesion architecture loses 70 to 78% of its variance, while perspective plurality is diversified, and prompt specificity can reverse homogenisation into diversification on argument depth. Homogenisation, in other words, is a property of the interaction design rather than an intrinsic property of the model. Szekely et al. [Székely et al., 2025] argue that the analogous process in speech, driven by acoustic-prosodic entrainment and accommodation, is likely to be stronger than for passive media because the interaction is reciprocal, and Ahmad et al. [Ahmad et al., 2026] set out the sociolinguistic frame for studying it through Communication Accommodation Theory.

6 Synthesis

Established. Five propositions rest on multiple independent Grade A studies. (i) Model output is reliably distinguishable from human writing, and distinguishable per model, with attribution across five conversational agents reaching a weighted F1 of about 0.96 [Zahid et al., 2024, Kumarage and Liu, 2023, Attar et al., 2026, Rujeedawa et al., 2025, Stacy and Kong, 2026]. Models also imitate a target human style imperfectly, which is the same signal seen from the other side [Jemama and Kumar, 2025]. (ii) A machine-associated lexicon exists and has measurably shifted scientific writing, news writing in 34 languages, and unscripted speech [Kobak et al., 2024, Juzek, 2026, Anderson et al., 2025]. (iii) Post-training reduces output diversity, and the collapse is embedded in the weights by training data composition [Kirk et al., 2023, Karouzos et al., 2026]. (iv) Sycophancy is general across assistant families, measurable at roughly 58% of rebuttal cases, and traceable to preference data [Sharma et al., 2023, Fanous et al., 2025]. (v) Models overconverge to their users' style, while users accommodate models as they would other people [Blevins et al., 2025, Blevins, 2026].

Established for Claude specifically, at Grade B. One proposition: a distinctive engineering-metaphor vocabulary, indexed by *load-bearing*, has grown from near zero to a large share of public pull request prose within eighteen months [Abraham, 2026]. The growth is measured; the attribution to Claude is inferential.

Not established. Four claims are repeated in the discourse without adequate support. First, that the *delve* lexicon identifies Claude. It indexes a different model family and an earlier period, and the human baseline against which it is measured is now itself contaminated at 13.5% or more [Kobak et al., 2024]. Second, that Claude hedges more than human writers. Hedging differences are reported in reviews of AI writing in general [Georgiou, 2026], but we found no study comparing Claude with a register-matched human corpus, and hedging is strongly register-dependent, so any comparison of model body prose with human abstracts will produce the effect spuriously. Third, vendor multipliers such as "700 times more frequent" [Pangram Labs, 2026]: no corpus, no baseline, and no method is stated, so the figure cannot be assessed. Fourth, that em dash density is diagnostic. The claim is widespread in grey literature, but the one grey source that examined it carefully concluded that em dashes appear because humans wrote that way in training data, and we found no measurement against a matched baseline.

Contested. Whether Claudish differs from other model varieties in kind or only in vocabulary. Chakrabarty et al. [Chakrabarty et al., 2024] find that professional editors detect the same seven categories of idiosyncrasy in GPT-4o, Claude-3.5-Sonnet, and Llama-3.1-70b, with no model preferred. Zahid et al. [Zahid et al., 2024] find that models are nevertheless separable at high accuracy. Both can be true: the coarse features are shared and the fine-grained distribution is distinctive. Which of those the folk term *Claudish* tracks is exactly what Loi's proposed blinded-recognition experiment would settle [Loi, 2026].

A moving target. The most important structural fact about this literature is that its object changes faster than its methods. The excess-vocabulary list dates from 2022 to 2024. *Load-bearing* is dated by community report to the Opus 4.6 to 4.7 transition in 2026 [Claude Code issue tracker, 2026]. A study designed around a fixed word list is measuring a snapshot. Methods that generalise, such as unsupervised clustering with a temporal read-out [Abraham, 2026] or distributional attribution [Zahid et al., 2024], are the appropriate instruments for a variety that is versioned.

7 Proposed mechanisms

Five mechanisms have been proposed. They are not mutually exclusive, and they predict different things.

Preference optimisation. The leading hypothesis. Reinforcement learning from human feedback reduces output diversity [Kirk et al., 2023, Karouzos et al., 2026] and induces sycophancy through preference data that rewards agreement and confident prose [Sharma et al., 2023]. It is the mechanism most often invoked for lexical overrepresentation as well. Support is partial: Juzek and Ward [Juzek and Ward, 2024] found model testing consistent with an RLHF account but their participant study did not confirm the predicted preference for focal words, and they explicitly note that opacity around model development obstructs the test. Status: supported for diversity and sycophancy, unproven for lexis.

Prompt priming. A deployment-time mechanism specific to the agentic setting. The published system prompt legislates individual openers [Anthropic, 2026b], output styles are user-selectable presets that alter the register wholesale [Anthropic, 2026a], and one community reproduction reports finding *load-bearing* in the shipped client's own prompt strings, which would make that item an artefact of the harness rather than of the model [Claude Code issue tracker, 2026]. This mechanism is cheap to test and has not been tested. Status: plausible, Grade C.

Written norm. Claude is unusual in having a published constitution, and Pourdavood [Pourdavood, 2026] argues that this document carries a determinate culture. If correct, part of Claudish is prescribed rather than emergent, which would make it the first variety in the literature whose norm can be read directly. Status: argued, not measured.

Agent-to-agent optimisation. The mechanism proposed for Sense 3. Informants report that over long agentic runs the variety drifts, becoming more compressed and more jargon-dense, because the audience is other agent turns rather than a human reader, and that asking the model to translate back into ordinary English costs tokens [Lorenz, 2026]. This is a contact-variety account, and it predicts something testable: Claudish should be more extreme late in long sessions than early. Status: hypothesis.

Accommodation and feedback. Models overconverge to their users [Blevins, 2026], users' own writing shifts toward the model [Agarwal et al., 2024, Inoshita et al., 2026], that writing becomes training data, and recursive training on generated text degrades the distribution at a measurable rate [Suresh et al., 2024]. The loop is closed and every link is measured, though never end to end.

8 Sociolinguistic status

Four properties are usually required of a dialect. Claudish satisfies two.

Systematicity. Partly satisfied, and less well evidenced than the discourse assumes. The features in Table 1 are not obviously a single system, and the strongest evidence for internal coherence is indirect: models are attributable from their text at high accuracy [Zahid et al., 2024], which implies a stable joint distribution, and one vocabulary cluster in pull request prose is separable by word choice alone [Abraham, 2026].

Recognisability. Satisfied, and unusually strongly. The variety is named, catalogued, mocked, translated, and theorised [Lorenz, 2026, gvzdv, 2026, DoddiC, 2026, Loi, 2026]. Readers recognise it from cues they cannot fully articulate [Velasquez and Teston, 2026]. Few varieties acquire a translation tool within a year of being named.

Association with a speaker group. Not satisfied. Claudish is produced by a versioned family of artefacts, not by a community. Structurally this is closer to an idiolect, with one decisive difference: the speaker can be instructed to switch varieties, either by a user-selected output style [Anthropic, 2026a] or by a prompt, and it also switches involuntarily toward whoever it is talking to [Blevins, 2026]. Situational selection is the defining property of register, not of dialect.

Transmission within a community. Partly satisfied, and this is the interesting case. Claudish is not acquired from Claude the way a dialect is acquired from other speakers, but its features demonstrably propagate: into scientific writing [Kobak et al., 2024, Galpin et al., 2025], into news writing in 34 languages [Juzek, 2026], into unscripted speech [Anderson et al., 2025], and into the writing of individual users under assistance [Agarwal et al., 2024, Inoshita et al., 2026]. The transmission channel exists; only its direction is unusual, running from artefact to community rather than between generations of speakers. The channel is also a loop, since text written in Claudish becomes training data and recursive training degrades the distribution at a measurable rate [Suresh et al., 2024].

We therefore conclude that Claudish is a register at origin: selected by a situation of use, switchable by instruction, defined by frequencies rather than by exclusive forms, and stable in the way registers are stable across languages [Li et al., 2022]. It is acquiring one dialect-like property, transmission, through documented human uptake. The title of this paper uses the construction that community catalogues identify as the most characteristic feature of the variety, which is the point: the antithesis is available to human writers too, and what is machine-like is not the construction but its rate.

9 Gaps and research agenda

The literature has a specific shape: strong population-scale evidence about machine-associated language in general, strong mechanism evidence about post-training, and almost no controlled measurement of Claude. Eight studies would close the main gaps.

1. **Blinded recognition.** Loi [Loi, 2026] argues that recognising a response as "so Claude" is prior to identifying which model produced it, and states the conditional that would make that recognition evidence of a real conversational character, but runs no study. The study is to collect graded judgments of Claudishness on unfamiliar tasks with branding and known phrases removed, and test whether they generalise. It determines whether Claudish is a character or a phrase list.

2. **Ground-truth pull requests.** Abraham [Abraham, 2026] finds a vocabulary cluster in 461,121 pull request descriptions that grows from 0.7% to 39% of the corpus in eighteen months, but no description carries an authorship label, so the identification of the cluster as Claude's is an inference. Repeating the analysis on a subcorpus with disclosed AI assistance would convert that attribution from Grade B to Grade A.

3. **Register-matched baseline.** Georgiou's review [Georgiou, 2026] finds that the cues separating AI from human writing are conditional on genre, length, and revision rather than absolute, which is precisely why uncontrolled comparisons mislead. The study is to compare Claude output with human text in the same register, section, and genre. Almost every current comparison sets model body prose against human abstracts, which manufactures hedging and self-mention effects.

4. **Negative parallelism.** Community catalogues call *it is not X, it is Y* the single most commonly identified marker of machine writing and enumerate its variants [tropes.fyi, 2026, aismells, 2026], but report no counts. The study is to measure its rate against a human baseline. As far as we can determine, the most-cited feature of machine writing has never been counted.

5. **Version diachrony.** Community reports date the rise of *load-bearing* to a specific version transition and record that instructing the model not to use it fails [Claude Code issue tracker, 2026]. The study is to track feature rates across model versions, testing whether Claudish changes by release or drifts continuously.

6. **Prompt ablation.** Anthropic publishes system prompts that legislate individual openers [Anthropic, 2026b] and ships user-selectable output styles that change the register wholesale [Anthropic, 2026a], so the harness is known to move the variety, by an unknown amount. The study is to vary system prompt and output style with the model held fixed, separating harness effects from weight effects. It is the cheapest study on this list.

7. **Session-length drift.** Informants report that during long agentic runs the model's agents reinforce one another and the output drifts toward a compressed, jargon-dense variety aimed at other agents rather than at readers [Lorenz, 2026]. The study is to measure compression and jargon density as a function of position in the session.

8. **Prompter control.** Blevins [Blevins, 2026] finds on WildChat across eight languages that models overconverge to their users on both function-word and open-class features, while users accommodate models no differently than they accommodate people. The study is to measure Claudish with prompt style held constant, so that what is attributed to the model is not the prompter's own style returned amplified.

10 Conclusion

Claudish is real as a social fact and thin as an empirical object. Users name it, catalogue it, and build translators against it, and one theorist has begun to ask what recognising it consists in. Against that, the measured record contains exactly one Claude-specific population-scale result, the growth of a vocabulary cluster indexed by *load-bearing* from 0.7% to 39% of public pull request prose in eighteen months, and its attribution to Claude is an inference rather than an observation. The features most confidently attributed to the variety are the least measured: the negative-parallelism construction has never been counted, the affirmative opener is documented only as a bug report, and the vocabulary list most often cited as evidence belongs to a different model family and an earlier period.

What the literature does establish is larger than Claude. Post-training narrows what models produce, preference data makes them agreeable, they overconverge to whoever is speaking to them, and the resulting variety is measurably reshaping scientific writing, news writing in 34 languages, and unscripted speech. Claudish is one version-indexed instance of that process, and it is a register: situationally selected, switchable by instruction, and defined by rates rather than by forms. It is becoming dialect-like in one respect only, and that respect is the one that matters, because the transmission runs from the artefact into the community that reads it.

Acknowledgment

We thank Martin Monperrus for constructive comments on this paper. Generative AI has been used to prepare this paper.

References

Louis Abraham. The load-bearing vocabulary of Claude: cluster analysis of GitHub pull requests. Code at https://github.com/louisabraham/load-bearing, August 2026. URL https://louisabraham.github.io/load-bearing/. Accessed 2026-09-05.

Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. *AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances*. 2024. URL https://doi.org/10.1145/3706598.3713564.

Sajjad Ahmad, A. Bibi, and Muhammad Saad Khan. Exploring linguistic adaptations in ai-mediated communication: A sociolinguistic analysis of human-ai interaction. *Social Science Review Archives*, 2026. URL https://doi.org/10.70670/sra.v4i2.2068.

aismells. It's not X, it's Y, 2026. URL https://aismells.com/not-x-its-y. Accessed 2026-09-05.

Bryce Anderson, Riley Galpin, and Thomas Stephan Juzek. Model misalignment and language change: Traces of ai-associated language in unscripted spoken english. *ArXiv*, abs/2508.00238, 2025. URL https://arxiv.org/abs/2508.00238.

Anthropic. Claude Code output styles, 2026a. URL https://docs.anthropic.com/en/docs/claude-code/output-styles. Accessed 2026-09-05.

Anthropic. System prompts release notes, 2026b. URL https://docs.anthropic.com/en/release-notes/system-prompts. Accessed 2026-09-05.

Yassir El Attar, Esra Donmez, Maximilian Maurer, and A. Falenska. A systematic analysis of linguistic features in ai-generated text detection across domains and models. *ArXiv*, abs/2606.04177, 2026. URL https://arxiv.org/abs/2606.04177.

Terra Blevins. Accommodation goes both ways: Studying linguistic convergence between humans and language models. *ArXiv*, abs/2605.29278, 2026. URL https://arxiv.org/abs/2605.29278.

Terra Blevins, S. Schmalwieser, and Benjamin Roth. Do language models accommodate their users? a study of linguistic convergence. *ArXiv*, abs/2508.03276, 2025. URL https://arxiv.org/abs/2508.03276.

Tuhin Chakrabarty, Philippe Laban, and Chien-Sheng Wu. Can ai writing be salvaged? mitigating idiosyncrasies and improving human-ai alignment in the writing process through edits. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2024. URL https://doi.org/10.1145/3706598.3713559.

Claude Code issue tracker. Persistent affirmative response bias despite explicit instructions (issue #2624), 2025. URL https://github.com/anthropics/claude-code/issues/2624. Accessed 2026-09-05.

Claude Code issue tracker. [model] Claude Code can not stop using the word "load-bearing" (issue #53454), 2026. URL https://github.com/anthropics/claude-code/issues/53454. Accessed 2026-09-05.

Yuntian Deng. Claude has become a language, so i built a translator: English ↔ Claudish, 2026. URL https://x.com/yuntiandeng/status/2091201867737145472. Accessed 2026-09-05.

DoddiC. claudish: translate between English and Claudish. GitHub repository, 2026. URL https://github.com/DoddiC/claudish. Accessed 2026-09-05.

A. Fanous, Jacob Goldberg, Ank A. Agarwal, Joanna Lin, Anson Y. Zhou, Sonnet Xu, Vasiliki Bikia, Roxana Daneshjou, and Oluwasanmi Koyejo. Syceval: Evaluating llm sycophancy. *ArXiv*, abs/2502.08177, 2025. URL https://arxiv.org/abs/2502.08177.

Riley Galpin, Bryce Anderson, and Thomas Stephan Juzek. Exploring the structure of ai-induced language change in scientific english. *ArXiv*, abs/2506.21817, 2025. URL https://arxiv.org/abs/2506.21817.

G. Georgiou. What distinguishes ai-generated from human writing? a rapid review of the literature. *Big Data and Cognitive Computing*, 10, 2026. URL https://doi.org/10.3390/bdcc10020055.

gvzdv. claudish-to-english: a Claude Code plugin that rewrites Claude output into plain English. GitHub repository, 2026. URL https://github.com/gvzdv/claudish-to-english. Accessed 2026-09-05.

Keito Inoshita, Michiaki Omura, Tsukasa Yamanaka, Go Maeda, and Kentaro Tsuji. Does ai homogenize student thinking? a multi-dimensional analysis of structural convergence in ai-augmented essays. *ArXiv*, abs/2603.21228, 2026. URL https://arxiv.org/abs/2603.21228.

Rebira Jemama and Rajesh Kumar. How well do llms imitate human writing style? *2025 IEEE 16th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0691–0700, 2025. URL https://doi.org/10.1109/UEMCON67449.2025.11267719.

Thomas Stephan Juzek. Ai-associated lexical shifts across 34 languages: Cross-lingual convergence and diachronic uptake in news writing. *ArXiv*, abs/2605.25358, 2026. URL https://arxiv.org/abs/2605.25358.

Thomas Stephan Juzek and Zina B. Ward. Why does chatgpt "delve" so much? exploring the sources of lexical overrepresentation in large language models. *ArXiv*, abs/2412.11385, 2024. URL https://arxiv.org/abs/2412.11385.

Constantinos F. Karouzos, Xingwei Tan, and Nikolaos Aletras. Where does output diversity collapse in post-training? *ArXiv*, abs/2604.16027, 2026. URL https://arxiv.org/abs/2604.16027.

Robert Kirk, Ishita Mediratta, Christoforos Nalmpantis, Jelena Luketina, Eric Hambro, Edward Grefenstette, and R. Raileanu. Understanding the effects of rlhf on llm generalisation and diversity. *ArXiv*, abs/2310.06452, 2023. URL https://arxiv.org/abs/2310.06452.

Dmitry Kobak, Rita González-Márquez, Emőke ágnes Horvát, and Jan Lause. Delving into llm-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11, 2024. URL https://doi.org/10.1126/sciadv.adt3813.

Tharindu Kumarage and Huan Liu. Neural authorship attribution: Stylometric analysis on large language models. *2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pages 51–54, 2023. URL https://doi.org/10.1109/CyberC58899.2023.00019.

LessWrong. Why do LLMs so often say "it's not an X, it's a Y"?, 2025. URL https://www.lesswrong.com/posts/RzPXywNbsRCss3Swy/. Accessed 2026-09-05.

Haipeng Li, Jonathan Dunn, and A. Nini. Register variation remains stable across 60 languages. *Corpus Linguistics and Linguistic Theory*, 0, 2022. URL https://doi.org/10.1515/cllt-2021-0090.

Michele Loi. "this is so claude!" towards a theory of the recognition of AI character without reidentification. *ArXiv*, abs/2608.16789, 2026. URL https://arxiv.org/abs/2608.16789.

Taylor Lorenz. Vibe coders now have their own language. User Mag, August 2026. URL https://www.usermag.co/p/vibe-coders-now-have-their-own-language. Accessed 2026-09-05.

Meizareena Mizad and Siti Afifah Hashim. A review of the use of interactional metadiscourse markers of hedges and boosters in academic writing. *International Journal of Modern Education*, 2025. URL https://doi.org/10.35631/ijmoe.726055.

Nautilus. I asked Claude why it won't stop flattering me, 2026. URL https://nautil.us/i-asked-claude-why-it-wont-stop-flattering-me-1279510/. Accessed 2026-09-05.

Pangram Labs. Can AI detection catch Claude writing styles?, 2026. URL https://www.pangram.com/blog/claude-writing-styles. Accessed 2026-09-05.

Parham Pourdavood. Does claude's constitution have a culture? *ArXiv*, abs/2603.28123, 2026. URL https://arxiv.org/abs/2603.28123.

ProgramAsWeights. English ↔ Claudish translator. Discussed at https://news.ycombinator.com/item?id=49402907, 2026. URL https://programasweights.com/claudish. Accessed 2026-09-05.

Muhammad Irfaan Hossen Rujeedawa, S. Pudaruth, and Vusumuzi Malele. Unmasking ai-generated texts using linguistic and stylistic features. *International Journal of Advanced Computer Science and Applications*, 2025. URL https://doi.org/10.14569/ijacsa.2025.0160321.

Mrinank Sharma, Meg Tong, Tomasz Korbak, D. Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott Johnston, S. Kravec, Tim Maxwell, Sam McCandlish, Kamal Ndousse, Oliver Rausch, Nicholas Schiefer, Da Yan, Miranda Zhang, and Ethan Perez. Towards understanding sycophancy in language models. *ArXiv*, abs/2310.13548, 2023. URL https://arxiv.org/abs/2310.13548.

Darcy Stacy and Lan Kong. *Detecting AI-Generated Essays: A Hybrid Stylometric and TF-IDF Approach with Cross-Dataset Validation*. 2026. URL https://doi.org/10.1145/3746467.3801525.

A. Suresh, Andrew Thangaraj, and Aditya Nanda Kishore Khandavally. Rate of model collapse in recursive training. *ArXiv*, abs/2412.17646, 2024. URL https://arxiv.org/abs/2412.17646.

Éva Székely, Jura Miniota, and Míša Hejná. Will ai shape the way we speak? the emerging sociolinguistic influence of synthetic voices. *ArXiv*, abs/2504.10650, 2025. URL https://arxiv.org/abs/2504.10650.

The Register. Claude Code's endless sycophancy annoys customers, 2025. URL https://www.theregister.com/software/2025/08/13/claude-codes-endless-sycophancy-annoys-customers/328260. Accessed 2026-09-05.

tropes.fyi. Negative parallelism, 2026. URL https://tropes.fyi/tropes/negative-parallelism. Accessed 2026-09-05.

Elizabeth Velasquez and Christa B. Teston. "it's giving ai": Reading ambiguously-authored texts and the role of felt sense. *Journal of Writing Research*, 2026. URL https://doi.org/10.17239/jowr-2026.17.03.08.

Iqra Zahid, Tharindu Madusanka, R. Batista-Navarro, and Youcheng Sun. Probing the uniquely identifiable linguistic patterns of conversational ai agents. pages 4612–4628, 2024. URL https://doi.org/10.18653/v1/2024.findings-acl.274.

Table 2: Quantitative evidence bearing on Claudish. "Claude?" indicates whether the study measures Claude specifically. Grades follow Section 2.

Study	Data	Principal quantitative result	Claude?	Grade
Abraham [Abraham, 2026]	461,121 GitHub pull request descriptions, 603 days, 51M word tokens	One vocabulary cluster grows 0.7% (early 2025) to 39% (mid-2026), about $+1.2$ points per week; *load-bearing* 929 occurrences inside versus 82 outside, ratio 39	inferred	A/B
Kobak et al. [Kobak et al., 2024]	15.1M PubMed abstracts, 2010–2024	Excess frequency ratio 28.0 (*delves*), 13.8 (*underscores*), 10.7 (*showcasing*); at least 13.5% of 2024 abstracts LLM-processed, up to 40% in some subcorpora	no	A
Juzek and Ward [Juzek and Ward, 2024]	PubMed abstracts, plus model testing and an online study	21 focal words identified; no evidence that architecture or training data explains them; RLHF consistent but not established	no	A
Juzek [Juzek, 2026]	WMT News Crawl, 34 languages	Prevalence of AI-overused items rises in 26 of 34 languages, mean $+15.1\%$; matched baseline words -4.5%	no	A
Anderson et al. [Anderson et al., 2025]	22.1M words of unscripted podcast speech	Significant post-2022 increase in LLM-associated words; baseline synonyms show no directional shift	no	A
Galpin et al. [Galpin et al., 2025]	PubMed abstracts, synonym groups, POS-tagged	Shift is not simple synonym replacement but involves semantic and pragmatic qualification	no	A
Fanous et al. [Fanous et al., 2025]	AMPS and MedQuad, three assistants	Sycophancy in 58.19% of cases (Gemini 62.47%, Claude-Sonnet intermediate, ChatGPT 56.71%); regressive sycophancy 14.66%	yes	A
Sharma et al. [Sharma et al., 2023]	Five assistants, four free-form tasks, human preference data	Sycophancy consistent across all five; preference models prefer sycophantic over correct responses a non-negligible fraction of the time	yes	A
Kirk et al. [Kirk et al., 2023]	Two base models, summarisation and instruction following	RLHF generalises better than SFT but reduces output diversity	no	A
Karouzos et al. [Karouzos et al., 2026]	Three post-training lineages of Olmo 3, 15 tasks, four diversity metrics	Location of diversity collapse co-varies with data composition; collapse embedded in weights, not in generation format	no	A
Blevins et al. [Blevins et al., 2025]	16 models, three dialogue corpora	Models converge to conversational style and often overfit relative to the human baseline; instruction-tuned models converge less than pretrained ones	no	A
Blevins [Blevins, 2026]	WildChat, eight languages	Models overconverge to users on function-word and open-class features; human convergence to models matches human-human baselines	no	A
Chakrabarty et al. [Chakrabarty et al., 2024]	LAMP corpus, 1,057 paragraphs edited by professional writers	Seven-category idiosyncrasy taxonomy; no model, including Claude-3.5-Sonnet, outperforms the others	partly	A
Zahid et al. [Zahid et al., 2024]	Five conversational agents	Agents attributable from their own text at weighted F1 of about 0.96	partly	A
Agarwal et al. [Agarwal et al., 2024]	118 participants, India and the United States	AI suggestions move Indian participants toward Western writing style; efficiency gains larger for Americans	no	A
Inoshita et al. [Inoshita et al., 2026]	6,875 essays, five conditions	Cohesion architecture loses 70–78% of its variance under AI assistance; prompt specificity can reverse the effect	no	A
Georgiou [Georgiou, 2026]	PRISMA review, 40 studies, 2022–2026	Five cue families; cue stability is conditional rather than absolute across genre, length, and revision	no	A
Pangram [Pangram Labs, 2026]	Vendor blog, method not stated	*complex and multifaceted* reported at 700 times the human rate; formal style said to omit *That's a great question*	yes	C