Official renamings and incidental mentions: Revision-level evidence from English Wikipedia

Max Sedlmair Independent researcher · ORCID 0009-0007-2471-5595

Preprint, version 1.0 · 19 October 2026 · A German translation is deposited alongside; the English text is authoritative.

Abstract

When a government renames a mountain, a country or a sea, the new name has to reach far more texts than the one about the place itself: every article on a storm, a fish, a climber or a railway that mentions the place in passing. This study asks how such incidental mentions change on English Wikipedia after an official renaming, and how the change relates to the encyclopedia's own decision about the title of the place's article.

Four renaming events were examined: Mount McKinley to Denali (2015), Swaziland to Eswatini (2018), Gulf of Mexico to Gulf of America (2025), and Denali back to Mount McKinley (2025). For each event, the articles that mentioned the place in running text in their last version before the act — as far as they could still be found through search and links in October 2026 — were followed revision by revision for 600 days (358 to 9,805 articles per event).

Where the place's article took the new name, the new name stood in the running text of 40.8 % (Denali) and 15.4 % (Eswatini) of these articles 600 days later; where the article kept the old name, of 0.04 % (Gulf of America) and 1.1 % (Mount McKinley). In the Gulf case, 1,617 of 1,620 revisions that inserted the new name were followed by a revision restoring the earlier count, with a median interval of about twelve minutes.

Counting a name is not the same as observing its use. In random samples of insertions, coded by two language models and by the author, many insertions turned out to mention the new name rather than use it: in the second Denali event, **24 of 36 sampled insertions placed "Mount McKinley" in glosses such as "Denali, federally designated as Mount McKinley"**. For the first Denali event, whether the counts measure adoption or only presence depends on how glossed forms are coded; the result is reported as undecided.

Across the four events, the presence of the new name covaries with the title of the place's article; the design does not separate the title decision from the act, from source usage or from later editing. The core design was frozen before data collection; all later corrections are reported alongside the frozen results. A fifth event, the US federal renaming of Lake Ontario in August 2026, will be observed on 22 December 2026 and 22 December 2027, together with a companion study on YouTube. The analysis is descriptive and takes no position on the merits of any renaming, on the governments that enacted them, or on Wikipedia's naming policies.

Keywords: place names; toponymy; renaming; Wikipedia; revision history; use and mention; naming conventions; prospective design freeze

1. Why this study

The question. A renaming is announced in one sentence. Whether the new name becomes ordinary is decided elsewhere: in the many texts that refer to the place for some other reason — the track of a hurricane, the range of a fish, the route of a railway, the career of a mountaineer. These incidental mentions are where a name becomes part of everyday reference, or does not. This paper looks at exactly those mentions, in a corpus where every change is recorded: the revision history of English Wikipedia.

Why Wikipedia. Most of its articles were written long before any given renaming and for other purposes. Every edit is stored with a timestamp. And the encyclopedia has written rules for names: the title of a place's article follows the name most commonly used in reliable English-language sources, and other articles are expected to use the name of that title. An official act is, by these rules, neither necessary nor sufficient for a change of name on Wikipedia. This makes it possible to ask a question that is hard to ask elsewhere: **does the name in incidental mentions follow the official act, or the encyclopedia's own title decision?**

What was expected. Before any data were collected, three expectations were written down with numerical thresholds (§2):

1. *Alignment with the title.* Where the place's own article takes the new name, the new name spreads into the articles that mention the place in passing; where the article keeps the old name, it does not.
2. *Removal.* Where the article keeps the old name, insertions of the new name elsewhere are mostly undone.
3. *Prediction.* Both patterns will hold for a renaming that had not yet been observed: Lake Ontario, renamed Lake America for US federal use on 27 August 2026.

What was found, in brief. The pattern of the first two expectations is visible in the data, and strongly so in the case of removal in the Gulf case. But a count of a name is not a measurement of its use. When samples of the counted insertions were read one by one, many turned out to be mentions of the name rather than uses of it — "Denali,

federally designated as Mount McKinley" — or to concern something else that bears the same name. The paper reports both the counts and what they capture, and it says where the second limits the first.

Why this paper, and why now. The author studies how newly introduced official names enter ordinary language, so far in YouTube comments (Sedlmair 2026c, 2026f). The US federal renaming of Lake Ontario in August 2026 offered an event that could be watched from its first weeks. A comment section has no written naming rule; an encyclopedia does. Observing the same event on both platforms on the same days therefore says something neither platform can say alone. This paper builds the Wikipedia side of that comparison: it develops and tests the measurement on four completed renamings, and it fixes in advance how the Lake Ontario case will be measured on 22 December 2026 and, again, on 22 December 2027.

1.1 Position of this study

This study is descriptive. It takes no position on the merits of any of the renamings examined, on the governments that enacted them, or on the naming policies of Wikipedia. The cases were selected for their structure — the same feature renamed in opposite directions by administrations of different parties, a second renaming arising from the same legal act, and a case outside the United States — not for their political content. All measurements and decision rules were fixed before the data were collected and apply identically to every case. The findings concern how a collaboratively edited encyclopedia handles place names; they should not be read as evidence for or against any government, party, or the encyclopedia itself.

1.2 Related work and what this study adds

The study builds on several lines of research and owes its tools to them. Renaming as an exercise of authority over space is a central theme of critical toponymy (Rose-Redwood, Alderman & Azaryahu 2010; Giraut & Houssay-Holzschuch 2016). Disputes over names on Wikipedia are well documented, from the Gdańsk/Danzig conflict (Jemielniak 2014) onwards. Studies of Burma/Myanmar (Luyt 2017), Ganga/Ganges (Kumar 2017), Bombay/Mumbai in the French edition (Jones 2017), the Falklands/Malvinas (Góngora-Goloubintseff 2020), the Crimea article (Gredel 2017) and the naming of COVID-19 in Chinese Wikipedia (Hu 2024) analyse how editors weigh state authority, neutrality and common usage, mostly on talk pages and in the article at the centre of the dispute. The nature and roles of written policies and rules in Wikipedia have been described in detail (Butler, Joyce & Pike 2008), and conflict on Wikipedia has been traced over time (Yasseri et al. 2012) and across space (Osborne, Graham & Dittus 2021). Recent work on the Russian–Ukrainian conflict finds that changing the spelling or form of place names is among the most frequent kinds of politically motivated edits in Wikipedia revision histories (Tamrazyan, Garambois & Boros 2026), and that forks of the encyclopedia replace place names systematically (Trokhymovych et al. 2025; Vestel & Degaetano-Ortlieb 2026).

The measurement tools are equally established. Token-level provenance across revisions, including deletion and reinsertion, is available for the whole of English Wikipedia (Flöck & Acosta 2014; Flöck, Erdogan & Acosta 2017), and revert detection has been refined beyond identical restorations (Flöck et al. 2012). Simons et al. (2024) count personal names revision by revision to trace their addition, removal and endurance in two articles; the counting used here follows the same idea. Changes of entity names have been extracted from Wikipedia text (Kanhabua & Nørvåg 2010; Holzmann & Risse 2014), and word substitutions mined from pairs of revisions (Yatskar et al. 2010; Recasens, Danescu-Niculescu-Mizil & Jurafsky 2013). The distinction between using a word and mentioning it has been annotated in Wikipedia text before (Wilson 2010, 2012); for the Gulf case, Teixeira da Silva & Nazarovets (2025) screened scientific publications for uses of the new name as a statement of fact. Outside Wikipedia, the uptake of "Gulf of America" has been followed in news media (Benton 2025, 2026), and the handling of disputed names on digital maps has been described (Quinn & Tucker 2017; Carrera Hurtado 2025 for the Gulf case).

In the work identified and checked for this study, no prior study was found that compares, across official renaming events in opposite directions, the revision-by-revision addition, removal and retention of a new place name in article groups defined by a fixed rule from their state immediately before each event and by incidental mention of the place; relates these trajectories to the title history of the place's own article; checks the counts with a use/mention validation; and tests a written prediction on a later renaming, in coordination with a second platform. That combination is the contribution of this paper. None of its components is new. If relevant work has been missed, the author apologises to its authors and would be grateful to be told (see *Note on sources and scope*).

2. Hypotheses and prior knowledge

The design was fixed on 17 October 2026, before any data on the four events were collected (Freeze v0.2; later addenda A1–A9 and a correction register are deposited, see *Data availability*). Three hypotheses were stated with numerical decision rules:

- **H1 (alignment with the title).** If, at the end of the observation window, the place's own article carries the new name, at least 25 % of the fixed articles carry the new name in a standing way (more occurrences than before T0); below 10 % refutes. If the article keeps the old name, at most 5 % do; above 15 % refutes. Values in between are undecided.
- **H2 (removal).** Where the title keeps the old name, at least 50 % of the articles into which the new name was

inserted are not standing at the end of the window, that is, carry no more occurrences of it than before T0.

- **H3 (prospective).** H1 and H2 hold for Lake Ontario on day 117 after the act.

Each event needed at least 30 fixed articles for H1 and at least 10 articles with insertions for H2. A rule for the instrument was added before any coding (Addendum A4): if fewer than 80 % of the coded insertions use the new name for the place, the counts for that event measure occurrences of the name, not its adoption, and H1 with a new title is not counted as supported; if more than 20 % concern no running text or something else, the event is not measurable under the frozen definition and the strict recount decides. This gate applies to every count-based measure of the event, including H2 (Addendum A9).

Prior knowledge, disclosed. Before the design was fixed, an unsystematic pilot on the Gulf case (100 random articles) was run to decide whether the project was worth pursuing. The research question was reformulated after that pilot. The Gulf case is therefore treated as a development case throughout. Confirmatory claims rest on the other three events and on the prospective test (cf. Weston et al. 2019 on preregistering analyses of existing data; Grieve 2021 on the place of observation in linguistics).

3. Data and method

3.1 Events and start dates

Event	Act	Start date (T0)	Role
Denali 2015	Secretarial Order 3337, US Department of the Interior (signed 28 Aug 2015, made public 30 Aug 2015)	30 Aug 2015	validation
Eswatini 2018	Royal declaration, 19 Apr 2018	19 Apr 2018	validation
Gulf 2025	Executive Order 14172	20 Jan 2025	development
Denali 2025	Executive Order 14172; Secretarial Order 3424 (14 Feb 2025)	20 Jan 2025	validation
Lake Ontario 2026	Executive Order 14422 (US federal usage)	27 Aug 2026	prospective test

T0 is the calendar day on which the act was first made public. All windows are date-based and start at 00:00 UTC on that day; the baseline is the last version of each article before 00:00 UTC. The analysis does not use the hour of the announcement. Two consequences are reported. First, the intention to rename the Gulf and the mountain had been stated publicly before the act, on 7 January 2025; insertions made between that statement and T0 belong to the baseline, not to the insertions (in the Gulf case this concerns one article with insertions). Second, the first two Gulf insertions counted after T0 (03:06 and 04:05 UTC on 20 January, the evening of 19 January in US Eastern time) precede the act on that day. Leaving out all insertions of the first 24 hours does not change any H2 result (§4.4). The same date-based rule, with 00:00 UTC on 27 August 2026, applies to Lake Ontario.

3.2 Article groups and their limits

Candidates were all articles returned by a full-text search of the article source for each name form, together with all articles linking to the place's article or its redirects. A candidate was kept if its last version before T0 contained the old name at least once in running text. The place's own article, articles whose title contains either name or refers to naming, renaming or executive orders, and disambiguation pages were excluded. These **legacy articles** were all written before the act and for other purposes.

The candidate search was run against the Wikipedia index of 17–18 October 2026, and the history of each candidate was then read back to T0. The groups are therefore reconstructed from the present, not enumerated from a historical snapshot. Articles that existed at T0 but were later deleted, merged, or edited so that they no longer contain either name or a link to the place cannot be found this way. How many such articles exist, and in which direction their absence biases the results, cannot be determined from the data. **The population of this study is therefore the articles that could still be found in October 2026 and whose earlier versions mentioned the place in running text**, not every article that mentioned the place at T0. The rule that defines the groups was fixed in advance; the membership lists were not.

3.3 Counting

Every revision of every legacy article within 600 days after T0 was retrieved through the public MediaWiki API, and the occurrences of the old and the new name were counted (patterns in the deposited `cases.py`). Full texts were not stored. An **insertion** is a revision that increases the count of the new name over the previous revision; a **removal** is a later revision in the window that brings the count back to or below its level before the insertion. An article is **standing** on day d if its last version up to that day carries more occurrences of the new name than before T0. "Not standing" therefore means "no more occurrences than before T0", which is not the same as "no occurrence at all"; both are reported where they differ. Counts are net per revision: a revision that removes one occurrence and adds another remains invisible.

Two definitions of running text are reported. The **frozen definition** removes comments, references and citation templates. The **strict definition**, added after the data showed that category links, flag templates, infoboxes and tables had been counted (§6), also removes categories, file names, all templates and tables. The decision rules in §2 are applied to both. The strict recount covers the baseline version, the versions on days 117 and 600, and every revision that inserted the new name under the frozen definition, together with its predecessor; a revision that would count as an insertion only under the strict definition is therefore not covered.

Units and intervals. H1 and H2 are shares of articles in the legacy groups; within the reconstructed groups these are complete counts, not samples. Their Clopper–Pearson 95 % intervals are given as a conventional guide to how much the shares could move under small changes in editing; they rest on a simple binomial model and ignore that articles share editors and coordinated edits. The use/mention shares in §4.2 are shares of insertions in random samples of 50 per event, drawn without replacement from a finite number of insertions; their intervals are binomial and therefore somewhat conservative. For zero counts the one-sided upper bound $1 - 0.05^{(1/n)}$ is given.

3.4 Title of the place's article and the naming rules in force

The title of the place's article on day *d* was read from the page-move log, as the source title of the first logged move after *d*, or the current title if no move followed (correction 15, §6). This rule was checked against the complete move history of each article, followed through all source and target titles (Table A3).

The versions of the two relevant Wikipedia guidelines in force at each T0 were retrieved from their revision histories: *Naming conventions (geographic names)* (oldids 677136588, 830771987, 1259288986) and *Article titles* (oldids 676233668, 833979064, 1270397965); verbatim excerpts are deposited. All three versions of the geographic naming conventions state that the name of the title "should be used consistently throughout the article", that this "applies to all articles using the name in question", and that "places should generally be referred to by the same name as is used in their article title", with an exception for historical contexts. All three state that the choice of title does not automatically follow an official name change, and that for places in the United States actual usage is preferred where it differs from the official federal name. All three versions of *Article titles* give more weight, after a name change, to the name used in reliable sources published after the change; in 2018 and 2025 this provision has its own section under the shortcut WP:NAMECHANGES, in 2015 it is part of the section on common names.

3.5 Use/mention check

For each event, 50 insertions were drawn at random. Each insertion was classified as a whole: **R** if at least one inserted passage of running text uses the new name to refer to the place ("…flows into the Gulf of America"); **M** if the new name appears only as a name being talked about (renaming, "also known as", "formerly", "federally designated as"); **N** if the new name is not in running text at all (only a table, template, category or file) or refers to something else (the national park, a vehicle model, a person); **X** if the extract shown was insufficient for a decision. All coders saw the same mechanically prepared extract: up to three inserted sentences of running text and a count of occurrences outside it.

Two language models coded all 200 sampled insertions independently, with identical instructions, in sessions without memory, history or web access: Claude Opus 5.5 (Anthropic) and GPT-6 (OpenAI), both by self-report on 18 October 2026. The author coded a random subset of 40 (10 per event) without seeing the model codes, after a practice block with invented examples. By the rule fixed before coding (Addendum A4), the model code was used where the two models agreed and the author's code where they disagreed, if available; otherwise the case was left undecided.

Because the decision for the two Denali events turned on this check, the author afterwards coded all remaining sampled Denali insertions with running text (61 further insertions, again without seeing the model codes), so that all 73 sampled Denali insertions with running text carry a human code. These codes are reported as a sensitivity analysis. They are not a verified gold standard — there was no second human coder — and the decision to give them weight was taken after they were known; where the two codings lead to different decisions, the result is reported as undecided (Addendum A9).

3.6 Robustness

The 20 accounts with most insertions were left out one at a time; any single account with more than 20 % of insertions is reported. Account names are pseudonymised with a keyed hash on the author's computer and are not published.

4. Results

4.1 Articles and titles

Event	Legacy articles (frozen / strict)	Revisions in window	Title of the place's article in the window
Denali 2015	358 / 292	15,181	new: Denali (moved on day 0)
Eswatini 2018	4,460 / 2,918	172,410	old until day 186, then new: Eswatini
Gulf 2025	9,805 / 8,471	195,071	old: Gulf of Mexico (never moved)
Denali 2025	1,131 / 872	32,697	old: Denali (two moves on day 64, reverted within 33 minutes)

The independent audit of the move history (Table A3) confirms the titles read by the corrected rule. In the strict recount, one article each in Denali 2015, Eswatini and the Gulf was excluded because its version on day 600 can no longer be retrieved; in the frozen analysis it was counted as not standing.

4.2 What the counts capture

Agreement between the coders on all 200 sampled insertions was high: 95.0 % between the two models ($\kappa = 0.92$, n = 200), 97.5 % between Claude and the author ($\kappa = 0.96$, n = 40) and 95.0 % between GPT and the author ($\kappa = 0.92$, n = 40). On the 73 sampled Denali insertions with running text, which include the harder cases, agreement was lower: $\kappa = 0.75$ between Claude and the author, 0.69 between GPT and the author, 0.86 between the two models. Where they differed, the models more often read a name with a gloss ("Denali (formerly Mount McKinley)") as plain use.

Table 1. Classification of 50 random insertions per event under the frozen extraction (rule of Addendum A4: model consensus, author's code on disagreement). Unit: sampled insertions.

Event	R	M	N	X / undecided	R share	N share
Denali 2015	29	1	17	2 / 1	61.7 % [46.4; 75.5]	36.2 %
Eswatini 2018	14	0	33	0 / 3	29.8 % [17.3; 44.9]	70.2 %
Gulf 2025	45	0	3	1 / 1	93.8 % [82.8; 98.7]	6.2 %
Denali 2025	11	20	12	3 / 4	25.6 % [13.5; 41.2]	27.9 %

Under the frozen extraction, only the Gulf case cleared both instrument thresholds. In the three validation events, more than a fifth of the counted insertions were in tables, templates or categories, or referred to something else, so by the rule fixed before coding the strict definition decides (Addendum A4). The strict definition itself, however, was introduced after the data had shown the category problem (correction 16). The validation results are therefore measurement-repaired, not fully confirmatory.

Table 2. Sampled insertions with running text only. Unit: sampled insertions.

Event	n (of 50 sampled)	R	M	N	X	R share (excl. X and undecided)	Coding
Denali 2015	37	29	1	6	0	80.6 % (29/36) [64.0; 91.8]	rule of A4 (one undecided)
		23	5	7	2	65.7 % (23/35) [47.8; 80.9]	author, all 37
Eswatini 2018	14	14	0	0	0	100 % [76.8; 100]	rule of A4
Gulf 2025	45	45	0	0	0	100 % [92.1; 100]	rule of A4
Denali 2025	36	11	20	1	0	34.4 % (11/32) [18.6; 53.2]	rule of A4 (four undecided)
		10	24	2	0	27.8 % (10/36) [14.2; 45.2]	author, all 36

For Denali 2015, the R share among sampled insertions in running text is 80.6 % under the rule fixed in advance — above the threshold of 80 % by one insertion — and 65.7 % on the author's codes. The difference lies almost entirely in glossed forms: the author coded "Denali (formerly Mount McKinley)" and "Denali (Mount McKinley)" as talk about the name, as the codebook prescribes for "formerly", while the models mostly read them as use. Seven of the 35 insertions the author could classify concern a person named Denali, a vehicle model or the national park. Both intervals are wide and include or touch the threshold. **For Denali 2015, whether the counts measure adoption or only presence of the new name therefore depends on the coding; the result is reported as undecided.**

In Denali 2025, both codings agree that the new name was mostly not used as the name of the mountain. In the author's coding, 24 of the 36 sampled insertions with running text (66.7 % [49.0; 81.4]) placed "Mount McKinley" in phrases such as "Denali, federally designated as Mount McKinley", "Denali (also called Mount McKinley)" or "then known as Mount McKinley", and ten used it as the name of the mountain. This is an observation about the sample; the interval includes 50 %.

4.3 Standing presence (H1)

Table 3. Legacy articles in which the new name is standing on day 600 (H1). Unit: articles.

Event	Title	Frozen definition	Strict definition	Instrument (R share, running text)	H1
Denali 2015	new	36.9 % (132/358)	**40.8 %** (119/292) [35.1; 46.6]	80.6 % (A4 rule) / 65.7 % (author)	presence ≥ 25 %; as adoption undecided (coding-dependent)
Eswatini 2018	new	30.0 % (1,339/4,460)	**15.4 %** (450/2,918) [14.1; 16.8]	100 %	undecided
Gulf 2025	old	0.03 % (3/9,805)	**0.04 %** (3/8,471) [0.01; 0.10]	100 %	supported (development case)
Denali 2025	old	1.6 % (18/1,131)	**1.1 %** (10/872) [0.6; 2.1]	34.4 % (A4 rule) / 27.8 % (author)	supported

Where the title kept the old name, the new name was standing in 3 of 8,471 legacy articles in the Gulf case and in 10 of 872 in Denali 2025 (strict definition).

Where the title took the new name, the picture is less clear than the counts suggest. In Denali 2015, the new name stood in 40.8 % of the legacy articles, well above the 25 % threshold, but whether this is adoption depends on the coding (§4.2). In Eswatini, the strict rate is 15.4 %, in the undecided range; it is roughly half the frozen rate, because flag templates, infobox fields, tables and category links are no longer counted. On day 117, before the title of the country's article changed, standing presence in running text was 0.4 %; on day 600 it was 15.4 %.

The contrast holds within one object. The same mountain, mentioned in running text, carried its new name in a standing way in 40.8 % of legacy articles 600 days after the 2015 renaming and in 1.1 % 600 days after the 2025 renaming. This is a contrast, not a mechanism: title status, the direction of the official change, source usage and historical context are not separately identified.

4.4 Removal (H2)

Gulf 2025. 586 of 9,805 legacy articles (6.0 %) received at least one insertion. On day 600, 583 of them were not standing (99.5 % [98.5; 99.9]), and in all 583 the new name no longer occurred at all. Under the strict definition the figures are 537 of 540 (99.4 % [98.4; 99.9]). Of 1,620 inserting revisions, 1,617 were followed by a revision that restored the count to its earlier level (a count-defined removal, not the tracked survival of individual tokens), with a median interval of 0.2 hours. Insertions were made mostly from unregistered or temporary accounts (1,250 of 1,620); removals almost exclusively from registered accounts (1,563 of 1,617). Edit summaries of removals most often contained revert vocabulary (848) or references to Wikipedia policy or consensus (460); references to the act itself were rare in removals (21) and in insertions (67). No single account made more than 2.5 % of insertions.

Denali 2025. Insertions were rarer: 37 of 1,131 legacy articles received any. Here, "not standing" and "no longer present" differ markedly, because "Mount McKinley" is also the historical name of the mountain: 20 of the 37 articles contained it before T0. Under the frozen definition, 19 of 37 (51.4 % [34.4; 68.1]) were not standing on day 600, but only 12 of 37 (32.4 % [18.0; 49.8]) no longer contained the name at all. Because the instrument check showed more than 20 % of insertions outside running text (Table 1), the strict recount decides for this event (§2): under the strict definition, 16 of 26 articles with insertions in running text were not standing on day 600 (61.5 % [40.6; 79.8]), and 11 of 26 (42.3 %) no longer contained the name in running text. **The point estimate lies above the threshold of H2; the interval includes 50 %, so this is support by the point estimate, not a robust confirmation.** One account made 27 of the 73 inserting revisions (37 %); without it, 16 of 27 articles (59.3 %) were not standing (frozen definition). Since most sampled insertions in this event were mentions of the name rather than uses (§4.2), H2 says little here about the fate of the name in use.

Day 0. Leaving out all insertions made in the first 24 hours after 00:00 UTC on T0 changes no H2 figure: in the Gulf and Denali 2025 cases, the 20 and two insertions of that day all fell into articles that also received later insertions. Without the two Gulf revisions made before the act on 20 January, 1,615 of 1,618 inserting revisions were followed by a count-defined removal.

Where the title changed, removals were rare under the frozen definition: 6.4 % of articles with insertions in Denali 2015 and 2.9 % in Eswatini; for Eswatini this figure includes template and category insertions and says little about running text.

4.5 Timing (exploratory)

In Denali 2015, the article was moved on the day of the announcement, and two thirds of all insertions (206 of 311) fell into the first quarter. In Eswatini, act and title change lie six months apart. In the 186 days between them, 144 inserting revisions were recorded in 76 articles; in the 30 days after the title change, 1,025, of which 583 came from one account implementing a category decision (frozen definition). This comparison was not part of the frozen analysis and is not evidence of an effect of the title change.

4.6 Prospective test and companion study

Reference values for Lake Ontario on day 117 (Gulf and Denali 2025 pooled, frozen definition): insertions in 3.9 % [3.6;

4.3] of legacy articles, standing presence 0.2 % [0.1; 0.3], not standing in 95.4 % [92.9; 97.1] of articles with insertions. Data collection is blocked by the collection script until 22 December 2026. The result will be added to this preprint as a new version, without changes to the rules. For Lake Ontario, every mention of either name in the legacy articles on day 117 will be coded individually, all mentions of “Lake America” by the author (Addenda A5, A6). The executive order concerns the name used by the US federal government; Lake Ontario is shared with Canada, and the study treats the renaming as a US federal designation, not as a change of name elsewhere.

The same renaming is the subject of a parallel study by the author on YouTube comments, which follows the old and the new name in comments under videos uploaded before the act (Sedlmair 2026f). Its second observation is scheduled for the same day, 22 December 2026. Both studies will then be set side by side: one event, one date, two platforms with different editorial regimes — an encyclopedia with a written naming rule and a comment section without one. The two measures are not identical (a stock of mentions in articles here, a flow of new comments there), so the comparison will be a coordinated description with raw counts and intervals, without a formal test of difference; the measures were fixed on 18 October 2026, before either dataset for that date was collected (Addenda A5, A6). A nationwide US election day, 3 November 2026 (day 68), falls inside the window. Insertions and comments with the new name will be reported separately for days 0–67 and 68–117, without a hypothesis and without political interpretation.

Beyond the observation already fixed for 22 December 2026, an additional long-horizon observation is prospectively fixed for 22 December 2027 (Day 482). It will be conducted for both the Wikipedia and YouTube studies under their respective frozen samples and measurement rules, irrespective of the 2026 findings, and will be reported as a subsequent longitudinal extension (Addendum A7). Articles or videos that are deleted or merged in the meantime will be reported as attrition, not replaced.

5. Discussion

The presence of the new name covaries with the title. Across the retrospective cases, standing prevalence of the newly introduced name covaries with article-title status; the observational design does not identify title status as the cause and does not separate it from official naming acts, source usage, editorial attention or other contemporaneous changes. Where the place’s article kept the old name, the new name remained rare as a standing end state, and in the Gulf case it was removed almost everywhere within minutes. Where the article took the new name, the name became common in the legacy articles, but in neither event can this be called adoption without qualification: in Denali 2015 the answer depends on how glossed forms are coded, in Eswatini the rate lies in the undecided range. With four events, of which one is a development case, this is an observed association within the reconstructed article groups, not a demonstrated effect of the title decision.

The pattern is what Wikipedia’s own rules prescribe. The geographic naming conventions in force at each T0 asked all articles to use the name of the title, with exceptions for historical contexts (§3.4). The finding is therefore not that such a rule exists, but how it is carried out: in the Gulf case by many hands within minutes, with removal summaries frequently containing revert or policy vocabulary; after a title change by follow-up editing and, partly, by automated category maintenance. The data do not show whether removals reflect individual editors applying the rule or coordinated attention, such as consistency campaigns or watchlist activity around a contested name. The title decision itself follows the use of the name in reliable sources and is not independent of the act or of media practice; the study cannot separate these influences.

Use and mention matter for the count. The clearest linguistic observation of this study concerns glosses, and it differs between the two Denali events. In 2025, when “Mount McKinley” was the newly introduced official name but not the title, the author coded 24 of 36 sampled insertions in running text as talk about the name, mostly in forms such as “Denali, federally designated as Mount McKinley”: the official name is acknowledged, the title name remains the referring expression. In 2015, when “Denali” was both the new name and the title, the author coded 23 of 35 classifiable sampled insertions as plain use, five as talk about the name — typically “Denali (formerly Mount McKinley)”, where the old name serves as a bridge — and seven as concerning something else. A simple count of the name would overstate its adoption in 2025 and, to a lesser degree, in 2015. The naming conventions themselves warn that search engines “do not normally distinguish consistent *use* of a name from a single *mention*”. This parallels the first snapshot of the companion study on YouTube, where ten of the eleven comments containing “Lake America” in the first weeks after the act discussed the name rather than used it (Sedlmair 2026f; on referential uptake of institutional names more generally, Sedlmair 2026c). With 73 coded insertions from samples of two events, this observation is preliminary.

What the next version can show. If the pattern holds, the newly introduced name “Lake America” should appear in the Wikipedia legacy articles mainly as a gloss beside “Lake Ontario”, not as the referring name, as long as the lake’s article keeps its title. This is stated here as a descriptive expectation, not as an additional hypothesis and not as a test; it will be checked against the per-mention coding of December 2026 on both platforms (Addenda A5, A6), with raw counts. If the expectation does not hold, that will be reported in the same way.

Limitations. One language edition and four events. The article groups are reconstructed from the index of October 2026 (§3.2); pages deleted, merged or emptied of both names since T0 are missing, and the direction of the resulting bias is unknown. T0 is a calendar day in UTC, not the hour of the act (§3.1). The name patterns were not validated for

recall: variant spellings, misspellings or paraphrases of either name are not counted. Counts are net per revision. Titles are endogenous to source usage. The title rule had to be corrected after data collection (§6), and the frozen counting definition did not exclude templates and categories; both corrections change decisions and are reported in full. The use/mention check rests on random samples of 50 insertions per event; for the Denali events, the additional codes come from one human coder, the author, with no second coder, and were given weight only after they were known. The strict running-text definition removes all templates and has not itself been validated against rendered text. The intervals for article shares do not model dependence between articles. Pseudonymisation of accounts is not anonymity: revision identifiers lead back to public accounts.

6. Deviations from the frozen design

All deviations are listed in the correction register (Appendix A); the main ones:

1. **Export (correction 14).** The export step of the collection script did not finish for the largest events and was replaced by a streaming exporter with identical output columns.
2. **Title rule (correction 15).** The page-move log was queried only for moves away from the place's titles, so moves back were missing. The frozen rule (target of the last move before *d*) returned vandalistic or short-lived titles, for example "Mount McKinley:Redirect" for Denali 2015 and "Denali/Mount McKinley" for Denali
2025. Under the frozen rule, H1 would have been recorded as refuted for both Denali events. The error was found in the move log before the outcome report was read; both reports are deposited. An independent audit of the complete move history confirms the corrected titles (Table A3).
3. **Running text (correction 16).** The frozen definition did not remove category links, templates and tables. The strict definition (§3.3) was added; by the rule for the use/mention check fixed before coding (Addendum A4), it decides for the three validation events.
4. **Use/mention check (correction 17).** A first attempt at manual coding of all 200 insertions was abandoned after three cases, because the tool showed raw source text without a clear unit. It was replaced, before any further coding, by the procedure in §3.5.
5. **Recount error (corrections 21, 27).** In the strict recount, 55 revisions that belong to articles in both the Gulf and the Denali 2025 cohorts were counted for only one of the two events, which produced 7 instead of 3 standing Gulf articles under the strict definition and 9 instead of 3 under the frozen definition. The revisions were recounted for both events; all strict figures were regenerated by one script (`auswertung_streng.py`). The frozen recount now reproduces the frozen analysis exactly. No decision changed.
6. **Author coding for Denali (corrections 28, 31).** All sampled Denali insertions with running text were coded by the author, as requested by a first external review. A second review pointed out that giving these codes precedence was decided after they were known; they are now reported as a sensitivity analysis, and Denali 2015 is reported as undecided.
7. **Wording and scope (corrections 30, 32–34).** Following the second review, the population is described as reconstructed from the 2026 index (§3.2); H2 is reported with its frozen meaning, "not standing", alongside "no longer present"; the instrument gate is applied to H2, so the strict recount decides H2 for Denali 2025; the date-based UTC definition of T0 is stated with a day-0 sensitivity analysis; and the description of the guideline versions is corrected.
8. **Status of the retrospective results.** Because of corrections 15, 16 and 28–31, the four retrospective events are reported as measurement-repaired validation, not as fully confirmatory tests. Table A2 shows each event under the frozen output and each correction. The only fully prospective test is Lake Ontario.

Before release, the manuscript was submitted twice to a critical review by a second language model (ChatGPT, OpenAI), instructed to act as a demanding peer reviewer, the second time on the full deposit. All points of both reviews were adopted, in part with a narrower formulation; a table with each point, the source checked, the decision and the recomputed value is deposited.

Data availability

Zenodo (this record): frozen design and addenda A1–A9, correction register, code (`cases.py`, `pull_wikipedia.py`, `export_neu.py`, `analyse.py`, `analyse_v0_2_1.py`, `schritt2.py`, `schritt3.py`, `schritt3c.py`, `auswertung_streng.py`, `h2_check.py`, `h2_streng.py`, `build_coding.py`), SHA-256 checksums, revision-level counts with pseudonymised accounts and edit summaries reduced to their category, the strict recount, the move-history audit, verbatim excerpts of the guideline versions, the 200 coded insertions with all codings, both analysis reports, and the response table to the second review. No account names are published.

Note on sources and scope

This paper draws on work from several fields: toponymy, Wikipedia research, computational linguistics and philosophy of language. Reference entries were checked against publisher or index records; that check establishes that each work

exists as described, not that it was read in full. The literature search was extensive but not exhaustive, and several recent works came to light only through citation chaining. If relevant work has been overlooked or misrepresented, the author apologises to its authors and would be grateful to be told; it will be acknowledged in the next version. Refutations of the findings are equally welcome; the data and code needed for them are deposited.

About the author and contributions

Max Sedlmair holds the Staatsexamen in English philology from the University of Regensburg and works as an independent researcher without institutional affiliation. He has no formal training in statistics, computer science or corpus linguistics; his knowledge of corpus-linguistic method is self-taught, built up over a series of studies of YouTube comment data and this one.

The study is the author’s in its conception and in its decisions. He set the research question and reformulated it after the pilot; conceived the link to his YouTube study of the same renaming, which turns one event into a comparison of two platforms with different editorial regimes on the same days; chose the events, including the two renamings of the same mountain in opposite directions so that the design is balanced across administrations; decided which name forms are counted and how the article groups are defined; insisted that the counts be checked against what they capture rather than reported at face value, and coded 101 insertions himself; decided that the Lake Ontario mentions be coded one by one and that a second observation be fixed one year later; set the neutrality rules for the text; and submitted the draft twice to adversarial review before release.

Provenance and AI disclosure

The design documents, collection and analysis code, literature screening and the drafts of this text were produced with Claude (Anthropic) in dialogue with the author, who reviewed and decided each step. The literature search was extended with OpenAI deep research; the critical reviews were carried out with ChatGPT (OpenAI). Two language models served as coders (§3.5). All figures in this paper can be recomputed from the deposited material.

Appendix A. Correction register

No.	Date	Change	Reason
1	17 Oct 2026	Research question reformulated (removal and title alignment instead of spread)	Pilot; Gulf therefore development case only
2	17 Oct 2026	Case selection by contrast logic	The new question needs variation in title status
3	17 Oct 2026	Content coding of talk pages dropped	Not needed for the question
4	17 Oct 2026	Pilot false positive (a Russian bay) documented	The legacy rule measures only increases
5	17 Oct 2026	Pilot: two Lake article titles seen, no content	Contamination minimal, documented
6	17 Oct 2026	Benton (Nieman Lab) reclassified as comparison finding	Pilot
7	17 Oct 2026	Novelty claim narrowed	Literature test: components known
8	17 Oct 2026	Instrument check (use/mention) and rule for H1 with a new title added	Objection: string ≠ adoption
9	17 Oct 2026	Terms “removal”, “revert”, “covaries with” fixed	Objections on reverts and endogenous titles
10	17 Oct 2026	Objection “Lake needs 600 days” rejected	Lake has its own 117-day window
11	17 Oct 2026	Working title changed; neutrality statement and writing rules fixed	No change to design or analysis
12	18 Oct 2026	Novelty claim restated after three literature searches	Full texts read
13	18 Oct 2026	Objection “Lake needs 600 days” rejected again	As no. 10
14	18 Oct 2026	Streaming exporter replaces export step	Export did not finish; output unchanged
15	18 Oct 2026	Title rule corrected	Move log lacked moves back
16	18 Oct 2026	Strict running-text definition added	Categories and templates were counted
17	18 Oct 2026	Use/mention check by two models + 40 human codes	First tool unsuitable
18	18 Oct 2026	Lake Ontario platform comparison and election-day windows fixed	Before data collection
19	18 Oct 2026	Comparison measures V1/V2 specified	Before data collection
20	18 Oct 2026	Lake: coding per mention	Same unit as YouTube; before data collection
21	18 Oct 2026	Recount error found (55 shared revisions)	Review; own check
22	18 Oct 2026	Interval-overlap rule for the platform comparison dropped	Review
23	18 Oct 2026	Lake: all “Lake America” mentions coded by the author	Review
24	18 Oct 2026	Retrospective results labelled measurement-repaired validation	Review
25	19 Oct 2026	Second Lake observation on 22 Dec 2027 (day 482), both platforms	Before either snapshot
26	19 Oct 2026	Outdated interval-overlap sentence removed from §4.6	Left over after no. 22
27	19 Oct 2026	Shared revisions recounted; Gulf strict 3/8,471 (was 7/8,470); three articles without retrievable day-600 version excluded from the recount	Correction of no. 21
28	19 Oct 2026	All 73 sampled Denali insertions with running text coded by the author	First review
29	19 Oct 2026	Guideline versions at each T0 checked; three references corrected	First review
30	19 Oct 2026	Population described as reconstructed from the October 2026 index; novelty claim reworded accordingly	Second review
31	19 Oct 2026	Author codes reported as sensitivity analysis, not as deciding; Denali 2015 H1 reported as undecided (coding-dependent)	Second review: precedence set after codes were known
32	19 Oct 2026	H2 reported with its frozen meaning (“not standing”) and as “no longer present”; instrument gate applied to H2, strict H2 for Denali 2025 (16/26)	Second review
33	19 Oct 2026	T0 stated as date-based UTC window; day-0 sensitivity; two Gulf revisions before the act documented	Second review
34	19 Oct 2026	Description of *Article titles* versions corrected (provision present in all three, shortcut from 2018); excerpts deposited; units and intervals separated; two wrong sentences corrected	Second review
35	19 Oct 2026	Recall of the name patterns named as a limitation; scope of the strict recount stated; parallel to the YouTube study given with its figures	Own check after the second review

Table A2. Each event under the frozen output and the corrections

Event	Frozen output (title rule as frozen)	Title rule corrected	Strict running text	Author coding of the sample (Denali)	Interpretation
Denali 2015	title read as “Mount McKinley:Redirect” → H1 (old title) refuted, 36.9 %	title new → H1 supported, 36.9 %	40.8 %; R share 80.6 % (A4 rule)	R share 65.7 %	presence above threshold; adoption undecided (coding-dependent)
Eswatini 2018	title “ESwatini” → H1 supported, 30.0 %	unchanged, 30.0 %	15.4 %	–	undecided
Gulf 2025	title unreadable; H1 (old) supported, 0.03 %	title old → supported, 0.03 %	0.04 %; H2 99.4 %	–	supported (development case)
Denali 2025	title read as “Denali/Mount McKinley” → H1 (new) refuted, 1.6 %	title old → supported, 1.6 %	1.1 %; H2 61.5 % [40.6; 79.8]	R share 27.8 %	H1 supported; H2 supported by point estimate only; new name mostly mentioned, not used

Table A3. Title history from the complete move log

Article	Logged moves (UTC)	Title in the observation window
Gulf of Mexico	13 Jan 2009: moved to a vandal title and back within 71 minutes	Gulf of Mexico throughout; “Gulf of America” redirects to the article on the naming controversy
Denali (2015 event)	15 Oct 2011: moved and back on the same day; 30 Aug 2015: Mount McKinley → Mount Denali → Denali (with one move back and forth within the hour); 1 Sep 2015: the redirect “Mount McKinley” moved and back	Denali from day 0
Denali (2025 event)	25 Mar 2025: two moves to “Denali/Mount McKinley”, moved back after 33 minutes; 16 Oct 2026 (after the window): two moves, reverted within five minutes	Denali, except 33 minutes on day 64
Eswatini	22 Oct 2018: Swaziland → ESwatini; 31 Oct 2018: ESwatini → Eswatini	Swaziland until day 186, then ESwatini / Eswatini

References

Benton, J. (2025). Nine months later, that body of water down south is still the “Gulf of Mexico” to news outlets. *Nieman Lab*, October 2025.

Benton, J. (2026). Trump can’t get the media to use “Gulf of America.” Will “Lake America” be any different? *Nieman Lab*, 3 September 2026.

Butler, B., Joyce, E., & Pike, J. (2008). Don’t look now, but we’ve created a bureaucracy: The nature and roles of policies and rules in Wikipedia. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, 1101–1110. https://doi.org/10.1145/1357054.1357227

Carrera Hurtado, G. (2025). Las interrogantes detrás de la designación del topónimo “Golfo de América”. *Anuario Mexicano de Derecho Internacional* 26, e20058. https://doi.org/10.22201/iij.24487872e.2026.26.20058

Flöck, F., & Acosta, M. (2014). WikiWho: Precise and efficient attribution of authorship of revisioned content. *Proceedings of WWW 2014*, 843–853. https://doi.org/10.1145/2566486.2568026

Flöck, F., Erdogan, K., & Acosta, M. (2017). TokTrack: A complete token provenance and change tracking dataset for the English Wikipedia. *Proceedings of ICWSM* 11(1), 408–417. https://doi.org/10.1609/icwsm.v11i1.14860

Flöck, F., et al. (2012). Revisiting reverts: Accurate revert detection in Wikipedia. *Proceedings of HT 2012*. https://doi.org/10.1145/2309996.2310000

Giraut, F., & Houssay-Holzschuch, M. (2016). Place naming as dispositif: Toward a theoretical framework. *Geopolitics* 21(1), 1–21. https://doi.org/10.1080/14650045.2015.1134493

Góngora-Goloubintseff, J. G. (2020). The Falklands/Malvinas war taken to the Wikipedia realm: A multimodal discourse analysis of cross-lingual violations of the Neutral Point of View. *Palgrave Communications* 6, 59. https://doi.org/10.1057/s41599-020-0435-2

Gredel, E. (2017). Digital discourse analysis and Wikipedia. *Journal of Pragmatics* 115, 99–114. https://doi.org/10.1016/j.pragma.2017.02.010

Grieve, J. (2021). Observation, experimentation, and replication in linguistics. *Linguistics* 59(5), 1343–1356. https://doi.org/10.1515/ling-2021-0094

Holzmann, H., & Risse, T. (2014). Insights into entity name evolution on Wikipedia. *WISE 2014*, LNCS 8787, 47–61. https://doi.org/10.1007/978-3-319-11746-1_4

Hu, B. (2024). Negotiation, power and ethics in online collaborative translation: Translation of “COVID-19” by Wikipedia translator-editors. *The Translator* 30(1), 78–95. https://doi.org/10.1080/13556509.2023.2278224

Jemielniak, D. (2014). *Common Knowledge? An Ethnography of Wikipedia*. Stanford, CA: Stanford University Press (ch. 3 on the Gdańsk/Danzig dispute).

Jones, H. A. (2017). *Multilingual knowledge production and dissemination in Wikipedia*. PhD thesis, University of Manchester.

Kanhabua, N., & Nørvåg, K. (2010). Exploiting time-based synonyms in searching document archives. *Proceedings of JCDL 2010*, 79–88. https://doi.org/10.1145/1816123.1816135

Kumar, S. (2017). A river by any other name: Ganga/Ganges and the postcolonial politics of knowledge on Wikipedia. *Information, Communication & Society* 20(6), 809–824. https://doi.org/10.1080/1369118X.2017.1293709

Luyt, B. (2017). "A wound that has been festering since 2007": The Burma/Myanmar naming controversy and the problem of rarely challenged assumptions on Wikipedia. *Journal of Documentation* 73(4), 689–699. https://doi.org/10.1108/JD-09-2016-0109

Osborne, C., Graham, M., & Dittus, M. (2021). Edit wars in a contested digital city: Mapping Wikipedia's uneven augmentations of Berlin. *The Professional Geographer* 73(1), 85–95. https://doi.org/10.1080/00330124.2020.1800493

Quinn, S. D., & Tucker, D. A. (2017). How geopolitical conflict shapes the mass-produced online map. *First Monday* 22(11). https://doi.org/10.5210/fm.v22i11.7922

Recasens, M., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2013). Linguistic models for analyzing and detecting biased language. *Proceedings of ACL 2013*, 1650–1659.

Rose-Redwood, R., Alderman, D., & Azaryahu, M. (2010). Geographies of toponymic inscription. *Progress in Human Geography* 34(4), 453–470. https://doi.org/10.1177/0309132509351042

Sedlmair, M. (2026c). Storm names as comparative benchmarks: Lay referential uptake of institutional weather names in German YouTube comments (Preprint v3.4). Zenodo. https://doi.org/10.5281/zenodo.22431589

Sedlmair, M. (2026f). Used or discussed? Lake Ontario and Lake America in YouTube comments: Separating name use from name talk in a frozen first snapshot of pre-existing comment sections (Preprint v1.0). Zenodo.

Simons, A., Kircheis, W., Schmidt, M., Potthast, M., & Stein, B. (2024). Who are the "Heroes of CRISPR"? Public science communication on Wikipedia and the challenge of micro-notability. *Public Understanding of Science* 33(7), 918–934. https://doi.org/10.1177/09636625241229923

Tamrazyan, H., Garambois, M., & Boros, E. (2026). Linguistic mechanisms of cultural heritage weaponization in digital encyclopedic discourse: Evidence from Wikipedia revision corpora. *Scientific Journal of Dragomanov Ukrainian State University, Series 9*, 31, 127–136. https://doi.org/10.31392/UDU-nc.series9.2026.31.09

Teixeira da Silva, J. A., & Nazarovets, S. (2025). Gulf of Mexico or Gulf of America? *Journal of Geography, Politics and Society* 15(4), 41–47. https://doi.org/10.26881/jpgs.2025.4.05

Trokhymovych, M., Kosovan, O., Forrester, N., Aragón, P., Saez-Trumper, D., & Baeza-Yates, R. (2025). Characterizing knowledge manipulation in a Russian Wikipedia fork. *Proceedings of ICWSM* 19(1), 1924–1936. https://doi.org/10.1609/icwsm.v19i1.35910

Vestel, A., & Degaetano-Ortlieb, S. (2026). Modeling linguistic imprints of war propaganda in a Russian Wikipedia fork. *Proceedings of LaTeCH-CLfL 2026*, 212–220. https://doi.org/10.18653/v1/2026.latechclfl-1.20

Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science* 2(3), 214–227. https://doi.org/10.1177/2515245919848684

Wikipedia contributors. Wikipedia:Naming conventions (geographic names), revisions 677136588 (21 Aug 2015), 830771987 (16 Mar 2018), 1259288986 (24 Nov 2024). Wikipedia:Article titles, revisions 676233668 (15 Aug 2015), 833979064 (3 Apr 2018), 1270397965 (19 Jan 2025). English Wikipedia, CC BY-SA 4.0.

Wilson, S. (2010). Distinguishing use and mention in natural language. *Proceedings of the NAACL HLT 2010 Student Research Workshop*, 29–33.

Wilson, S. (2012). The creation of a corpus of English metalanguage. *Proceedings of ACL 2012*, 638–646.

Yasseri, T., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2012). Dynamics of conflicts in Wikipedia. *PLoS ONE* 7(6), e38869. https://doi.org/10.1371/journal.pone.0038869

Yatskar, M., Pang, B., Danescu-Niculescu-Mizil, C., & Lee, L. (2010). For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia. *Proceedings of NAACL HLT 2010*, 365–368.