

One Hundred Years of UFO Witness Reports as a Language Corpus

Abstract

We treat the National UFO Reporting Center archive — 111,961 first-person witness narratives spanning 1905 to 2023 — as a sociolinguistic corpus and apply computational text analysis to surface statistical structure that is interpretable regardless of one’s beliefs about the underlying phenomena. Using sentence-transformer embeddings, TF-IDF analysis, and density-based clustering, we report three findings. First, witness vocabulary tracks the cultural availability of descriptive terms, with “flying saucer,” “triangle,” “orb,” and “tic-tac” rising and falling at culturally identifiable moments. Second, independent witnesses on the same night in different states produce semantically convergent narratives at rates well above chance; our clustering surfaces 4,227 such same-night events without supervision. Third, the corpus organizes into 30 stable narrative archetypes that partially map to the Hynek and Vallée classification systems and extend them in two directions (sound-only encounters, animal-reaction reports). We validate the pipeline against 10 reference events with known causes, ranging from missile launches and meteor showers to historically prominent high-volume sightings; all 10 are surfaced by at least two of four independent detection methods. The pipeline, derived analytical datasets, and an interactive dashboard are open; the underlying narratives remain the property of NUFORC.

1. Introduction

Over 111,000 people have filed reports with the National UFO Reporting Center since its founding in 1974. Whatever these reports describe — misidentified aircraft, atmospheric phenomena, satellites, hoaxes, or something else — the reports themselves are real. They constitute a corpus of first-person observational narratives collected under consistent conditions over roughly half a century, with metadata (date, location, reported shape, duration) that enables structured analysis.

This corpus has been used before, primarily for statistical summaries: how many sightings per year, which states report most, what shapes are most common. But the narratives themselves —

the actual words witnesses use to describe what they saw — have received almost no systematic computational analysis. This is the gap we address.

Our question is not “are UFOs real?” It is: what can 112,000 witness narratives, treated as a language corpus, tell us about how people describe anomalous aerial observations, how that language changes over time, and whether independent witnesses produce consistent accounts?

We process the NUFORC archive through a ten-stage computational pipeline: acquisition, cleaning, geocoding, sentence-transformer embedding, temporal vocabulary analysis, same-night clustering, narrative archetype discovery, flap detection, signature phrase extraction, and validation against reference events. Each stage is idempotent and reproducible. The code is MIT-licensed. An interactive dashboard allows readers to explore the corpus, verify our claims, and discover patterns we did not anticipate.

1.1 Two readings, one corpus

Two interpretations are consistent with every pattern we report, and this paper does not adjudicate between them.

Under a **cultural-priming reading**, witnesses describe what they see using vocabulary their culture has made available, so convergent narratives reflect convergent language rather than convergent observations. The 1950s witnesses say “flying saucer” because the press taught them the term; the 2020s witnesses say “tic-tac” because Navy footage taught them the term. Under this reading, the corpus is a mirror of culture, not of the sky.

Under an **observational reading**, witnesses describe similar things because they observed similar things, and shared vocabulary is the medium through which that similarity becomes visible. Under this reading, the convergence of same-night narratives is noteworthy precisely because it suggests a shared stimulus.

Both readings predict the same statistical signatures in the corpus, and we have no instrument that can separate them from text alone. What we can do is characterize the structure precisely enough that future work — incorporating physical sensor data, controlled elicitation, or matched non-UFO corpora — has something concrete to test against.

We return to this tension at the end of each analytical section.

2. The Corpus

Source

The National UFO Reporting Center (NUFORC) has operated continuously since 1974, primarily through a telephone hotline and, since the late 1990s, a web submission form. Reports are logged by NUFORC’s director with minimal editorial intervention: the witness narrative is preserved largely verbatim, with occasional annotations noting credibility assessments or follow-up contact.

We work from a January 2024 scrape of the NUFORC public databank (approximately 148,000 raw reports), sourced from the `kcimc/NUFORC` dataset on HuggingFace. The data is used with explicit research permission granted by Christian Stepien, NUFORC’s CTO, in May 2026.

Cleaning

The raw corpus undergoes several cleaning steps designed to preserve the linguistic character of the data while removing records that cannot support analysis:

- **Date parsing.** NUFORC uses inconsistent date formats. We normalize all dates and apply a two-digit year pivot (00–29 → 20xx, 30–99 → 19xx) consistent with prior published analyses.
- **Narrative cleaning.** HTML entities are decoded. Editorial annotations appended by NUFORC’s director — for example, “Witness elects to remain totally anonymous; provides no contact information. PD” — are stripped via regex to prevent them from contaminating embeddings and downstream clustering.
- **US-only filter.** We restrict to US reports for this analysis (cleaner geocoding, single-language narratives, larger coherent sample). International analysis is future work.
- **Minimum narrative length.** Reports shorter than 30 characters after cleaning are dropped — these are typically metadata-only entries with no analyzable content.

After cleaning: **111,961 reports** (75.7% retention), spanning 1905 to 2023, across all 50 US states plus DC.

Geocoding

City-state pairs are geocoded against the 2024 US Census Bureau Gazetteer files (51,674 named places). 85.9% of reports match at city level; the remaining 14.1% fall back to state centroids. No external API calls are required.

Three representative narratives

The range of the corpus is best conveyed by example.

A triangle sighting (Vandalia, Ohio, September 4, 2009):

I got up, ready for school, I sat in the window with my dog. I looked up in the sky and saw a triangler figure. My eyes were wide open. So I ran right to the door. As soon as I was on the front porch, my mouth dropped. It had a white light circle in the middle and on the sides were colored lights that were red, white, and blue. It traveled from the north east to the south west. The next day I saw a very large helicopter following the same path. Passing the house about six times.

A daytime sphere (Chicago, Illinois, July 11, 2020):

I video taped for as long as I could, the UFO continued moving in its point but stayed still in location. After staring for 3 minutes, I went back to bed. It was very early in the morning. I woke up again at 5 am and nothing in the location an hour ago, not a shooting star or bright planet. The item was gone. This was neither a drone or helicopter. This was a little sphere moving on its axle.

A sound-only encounter (East Bridgewater, Massachusetts, December 14, 2012):

At approximately 2am on sunday, december 16th 2012, i was sitting down watching t.v. when i heard and felt a highly percussive, what can only be described as, horn blast, it had such resonance that i could feel my chest vibrating. It continued to sound in short blasts of first one, and then three "trumpets." Each 'blast' lasted from 3-6 seconds and could be physicaly felt as well as heard, it also caused both of my cats to run and hide under my bed for the next half hour or so.

The narratives are colloquial, observational, and specific. The median length is 692 characters — roughly a long paragraph. Spelling and grammar vary widely; we preserve them as-is.

3. The Shape of the Data

Temporal distribution

Reporting volume is not evenly distributed. The corpus contains only 644 reports from before 1960, grows steadily through the 1970s–1990s, and surges in the 2000s as internet access made the NUFORC web form available to a mass audience. The peak year is 2014 (7,439 reports); volume has modestly declined since, possibly reflecting competition from social media platforms as a reporting venue.

The 2010s alone account for 49,899 reports — 44.6% of the corpus. This recency bias is important context for all temporal analyses: we are not comparing equal-sized samples across decades. Our temporal vocabulary analysis (Section 4) addresses this by computing rates per 1,000 reports in each time bin rather than raw counts.

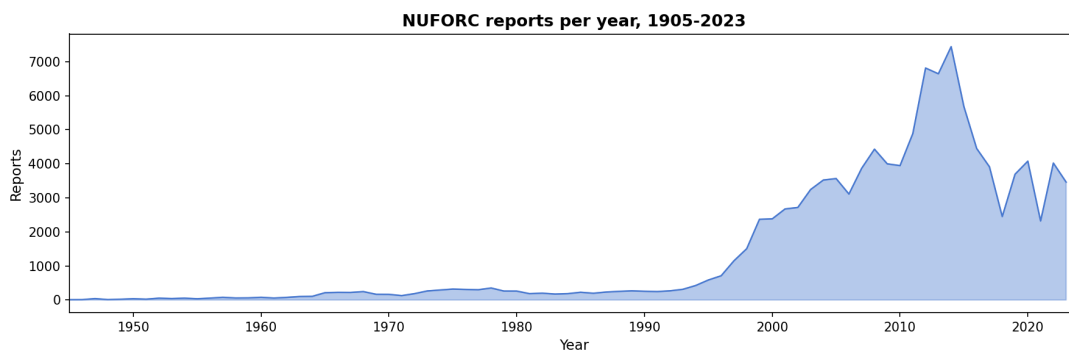


Figure 1. Reports per year, 1945–2023. Volume tracks internet adoption, peaking in 2014.

Geographic distribution

California leads with 14,224 reports (12.7%), followed by Florida (7,180), Washington (6,141), and Texas (5,543). Reports have been filed from all 50 states. The geographic distribution broadly tracks population density, with notable exceptions: Washington state reports at roughly 3× its population share, a pattern attributable to the state’s cultural association with UFO sighting (Kenneth Arnold’s 1947 report was over Washington) and NUFORC’s physical location near Seattle.

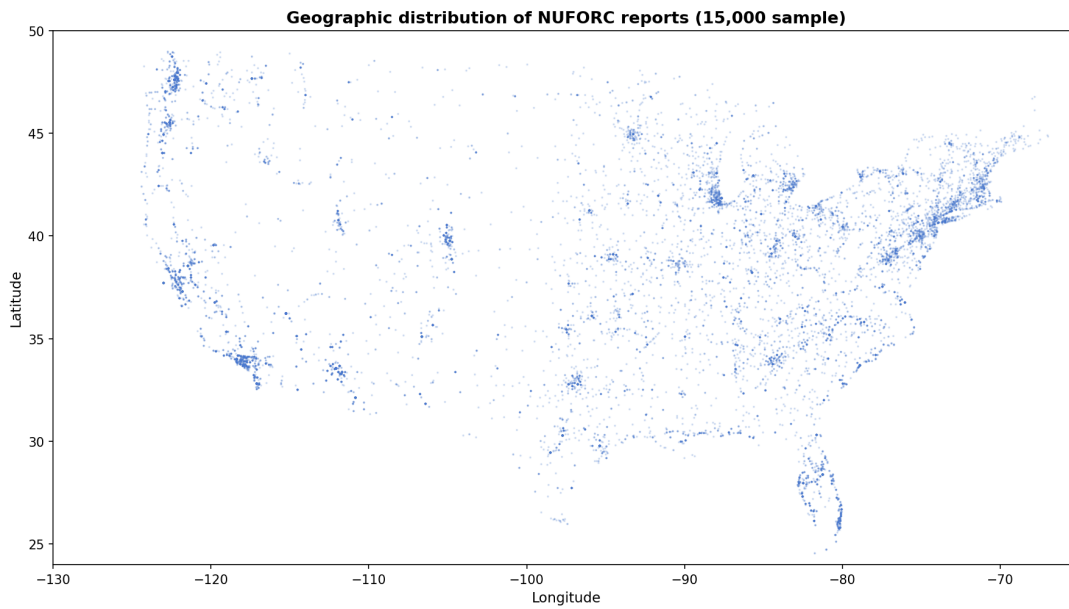


Figure 2. Geographic distribution of reports (15,000 sample, city-level geocoded).

Shape distribution

Witnesses describe 25 normalized shape categories. “Light” is the most common (28,459 reports, 25.4%), followed by “circle” (10,815), “triangle” (8,337), and “fireball” (6,935). The dominance of “light” reflects a reporting reality: at night, most witnesses can describe luminosity and movement but not solid geometry.

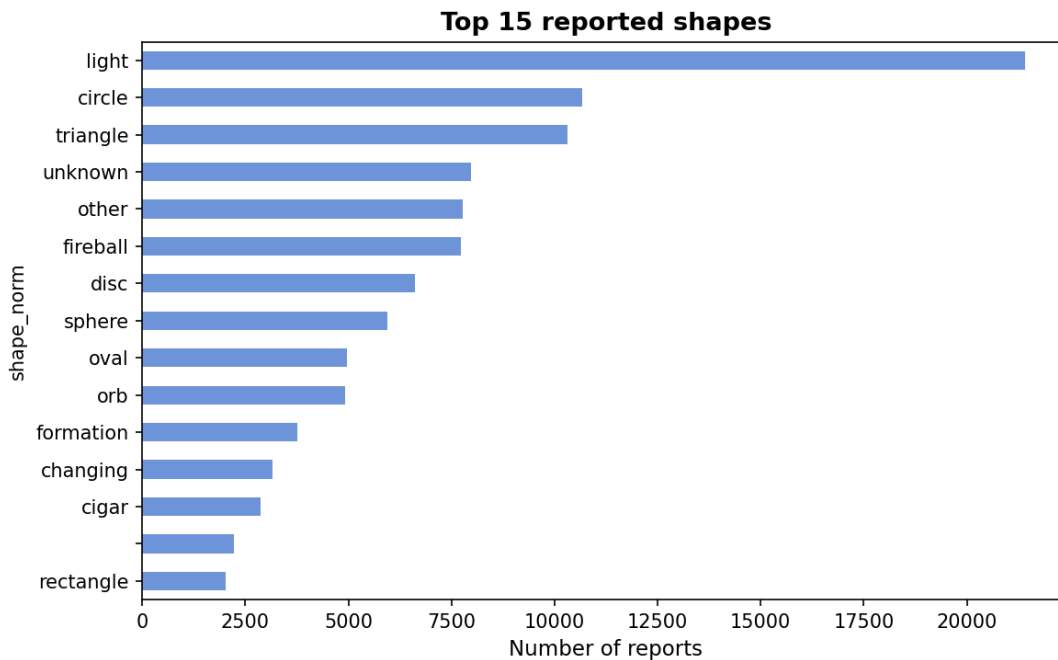


Figure 3. Top 15 reported shapes. “Light” dominates — at night, witnesses describe luminosity, not geometry.

4. Vocabulary Changes Over Time

The most striking finding in the corpus is how precisely witness vocabulary tracks cultural availability. We define a vocabulary of 55 culturally significant UFO-related terms — shape descriptors, behavior descriptors, color terms, and cultural-era markers — and compute their frequency per 1,000 narratives in 5-year bins from 1940 onward.

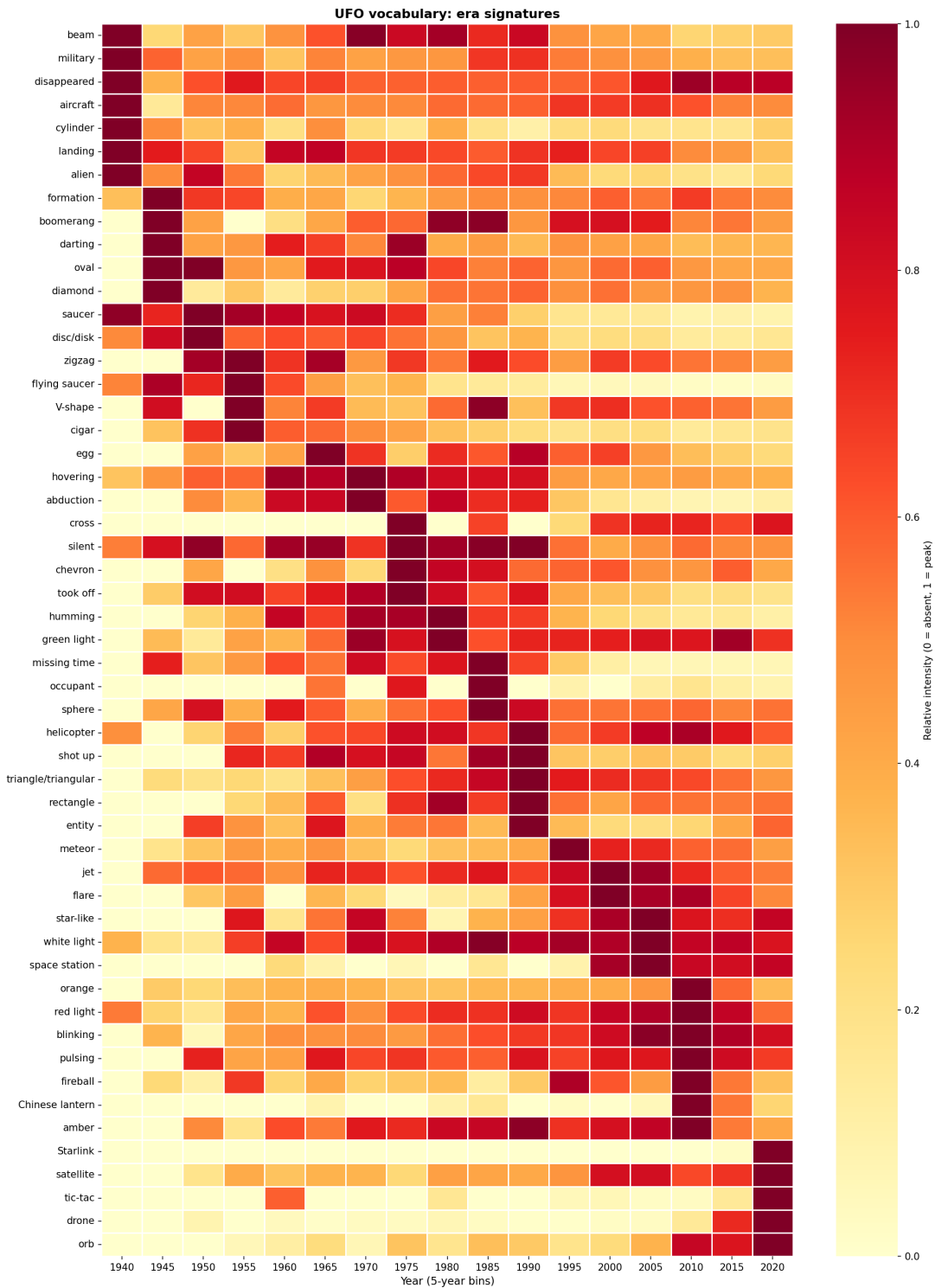


Figure 4. Era signature heatmap. Each row is a UFO-related term; each column is a 5-year bin. Color intensity shows relative frequency (normalized per term). Read left to right for a compressed cultural history of UFO reporting.

The heatmap reads as a compressed cultural history of UFO reporting in America.

The saucer era (1947–1965)

“Flying saucer” appears in the corpus precisely when the term enters American English, following Kenneth Arnold’s June 1947 report and the associated press coverage. “Disc” and “saucer” peak at 179 and 141 per thousand reports respectively in the early 1950s, then decline steadily. “Cigar” — the other canonical 1950s shape — tracks the same arc, peaking at 105/1k in the late 1950s. By the 1990s, “flying saucer” has declined to under 20/1k. The term did not disappear because saucer-shaped objects stopped being reported; it disappeared because witnesses stopped using that word.

The abduction era (1965–1985)

“Hovering” peaks at 287/1k in the 1970s — the decade of *Close Encounters of the Third Kind* and the Betty and Barney Hill case entering popular culture. “Abduction” peaks at 29/1k in the 1970s. “Missing time” — the specific phrase that became diagnostic of abduction narratives after Budd Hopkins’s 1981 book — spikes to 15/1k in the mid-1980s, exactly when Hopkins and Whitley Strieber were publishing. “Silent” peaks at 86/1k in the mid-1970s, reflecting a new emphasis on the absence of engine noise as a distinguishing feature.

The triangle era (1982–2000)

“Triangle” and “triangular” together peak at 196/1k in the 1990s, coinciding with the Belgian wave (1989–90), the Hudson Valley wave (1982–86), and the Phoenix Lights (1997). “Chevron” and “boomerang” — related shape terms — peak in the same window. This is the era when the “black triangle” entered the UFO lexicon as a primary shape category.

The orange-orb era (2005–2020)

“Orange” explodes to 269/1k in the 2010s, up from near zero before 2000. “Fireball” peaks at 47/1k in the same decade. “Chinese lantern” appears for the first time at 11/1k — a term that literally did not exist in UFO reporting before sky lanterns became commercially available in the US around 2008. The co-occurrence of “orange,” “fireball,” and “Chinese lantern” strongly suggests that a significant fraction of 2010s reports describe sky lanterns, though not all orange-orb reports can be so explained.

The post-disclosure era (2017–present)

“Drone” reaches 61/1k in the 2020s. “Tic-tac” — a term coined by Navy pilots in the 2004 USS Nimitz encounter but not public until the 2017 *New York Times* disclosure — appears at 12/1k. “Starlink” appears at 19/1k. “Orb” — the current preferred term, notably used in congressional testimony — reaches 61/1k. “UAP” (the government’s replacement term for “UFO”) is beginning to appear in civilian reports.

What the eras imply

The pattern is unambiguous: witnesses use the words their culture provides. This is not surprising — it would be remarkable if they did not — but the precision of the tracking, and the ability to date cultural shifts to within a few years using only the witness narratives, demonstrates that the corpus has genuine sociolinguistic structure.

The vocabulary tracking we observe is jointly produced by what witnesses saw, what they could name, and which witnesses chose to file. NUFORC reporters are self-selected toward people who believe their sighting was anomalous and who know NUFORC exists. The 2017 emergence of “tic-tac” in civilian reports therefore reflects both the diffusion of the term into general awareness and the NUFORC reporting population being unusually attentive to UAP news. We cannot separate these mechanisms from the corpus alone. The pattern is real; the causal interpretation is constrained. This is the central tension articulated in Section 1.1: the cultural-priming reading and the observational reading predict identical vocabulary curves.

5. Independent Corroboration: Same-Night Clustering

If a single witness reports a strange light, that is an anecdote. If twelve witnesses in eight states report the same thing on the same evening, independently and without coordination, that is a dataset.

We test for this by grouping reports by date, computing pairwise cosine similarity of their narrative embeddings within each date, and running agglomerative clustering with a cosine distance threshold of 0.35. Reports that cluster together on the same night — meaning their narratives are semantically similar despite being filed independently — constitute a “same-night event.”

The algorithm finds **4,227 same-night clusters** comprising 16,720 reports (14.9% of the corpus). The median cluster contains 3 reports; the largest contains 96. **96% of clusters span multiple states**, meaning the witnesses were geographically dispersed.

Known events surfaced blind

The pipeline has no knowledge of UFO history. It has never heard of the Phoenix Lights. Yet:

September 19, 2009. A cluster of 37 reports across 11 northeastern states (CT, MA, MD, NC, NJ, NY, OH, PA, RI, VA, VT) describes a “cone-shaped light” or “white light shining down in a cone shape.” Mean cosine similarity: 0.71.

November 7, 2015. Two clusters emerge. One contains 89 reports from 7 western states (AZ, CA, CO, ID, NM, NV, UT) describing “a broad point of white light with a long contrail.” This was the US Navy’s Trident missile test off the California coast. A second, smaller cluster on the same date captures 15 reports with higher narrative similarity (0.92) – closer witnesses or a different vantage of the same event.

July 4th, every year. Clusters of “orange orbs” and “red lights in formation” appear reliably on Independence Day, concentrated around the hours when fireworks displays are ending. These are almost certainly Chinese lanterns released during celebrations.

March 26, 2020. Ten reports across 7 states describe “at least 36 spherical objects in a straight line.” This is a Starlink satellite train, visible during the early weeks of the COVID-19 pandemic when more people were outdoors in the evening.

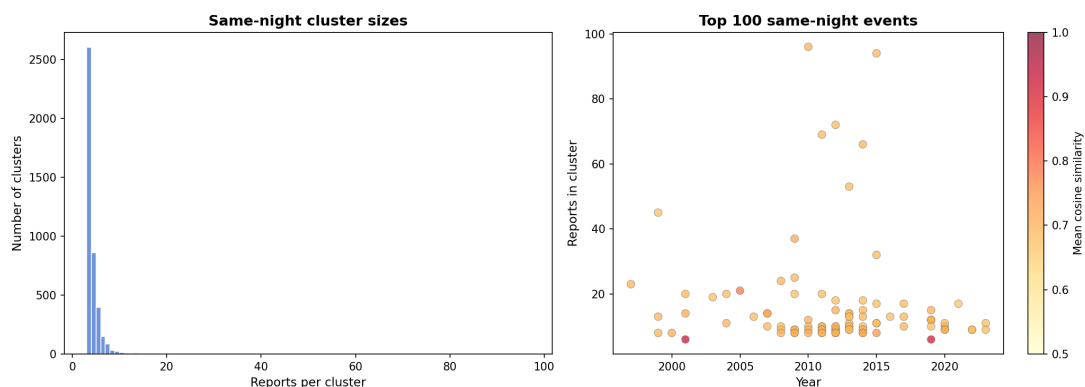


Figure 5. Same-night cluster analysis. Left: distribution of cluster sizes. Right: top 100 clusters plotted by year and size, colored by narrative coherence.

Limits of “independent”

The “independent” in “independent witnesses” requires qualification. Witnesses on the same night may share exposure to local news, social media coverage, or weather reports, any of which can introduce correlated vocabulary without correlated observation. For events with a clear physical trigger (Trident missile test, Starlink train), the convergence is unsurprising. For events without an identified trigger, the convergence is suggestive but not probative. The clustering algorithm cannot distinguish between witnesses who saw the same thing and witnesses who heard about the same thing.

6. Narrative Archetypes

Beyond individual events, do witness narratives fall into recurring patterns — archetypes — that transcend specific dates and locations?

We reduce the 384-dimensional embedding space to 25 dimensions with UMAP, then cluster with HDBSCAN (min_cluster_size=200, min_samples=10). This discovers **30 narrative archetypes** covering 40,572 reports (36.2% of the corpus). The remaining 63.8% are unclustered — narratives too unique or hybrid to fit a single pattern.

Selected archetypes, characterized by distinctive TF-IDF terms and dominant shapes:

ID	Reports	Top shape	Distinctive terms	Description
A0	3,303	orb	orb, orbs, orange, moving	Orange orb formations
A24	2,370	fireball	fireball, orange, appeared	Fireball sightings
A10	1,939	triangle	craft, triangle, sound, lights	Black triangle encounters
A18	1,591	fireball	meteor, bright, atmosphere	Meteor/bolide observations
A11	1,120	light	helicopter, craft, red	Helicopter-like craft
A14	607	unknown	sound, humming, loud, vibration	Sound-only encounters
A3	517	cigar	cigar shaped, metallic, trail	Cigar-shaped objects
A12	381	other	drone, lights, flashing	Drone sightings
A1	323	light	missile, launch, vanderberg	Launch observations
A19	308	light	dog, barking, dogs	Animal-reaction reports

ID	Reports	Top shape	Distinctive terms	Description
A6	258	oval	blimp, thought, balloon	Blimp/balloon misidentifications

The archetypes share categorical structure with existing classification systems. Hynek’s “nocturnal light,” “daylight disc,” and “close encounter” types are recognizable. The triangle archetype (A10, 1,939 reports) is the single largest non-generic cluster, reflecting the dominance of the black triangle as a reported form since the 1980s.

The archetypes also surface categories the traditional systems miss. Sound-only encounters (A14, 607 reports) — where witnesses hear a low humming or vibration but see nothing — have no Hynek category but form a tight, coherent cluster. Animal-reaction reports (A19, 308 reports) are another emergent category: narratives where the primary evidence is behavioral (dogs barking, animals agitated), clustering together because of shared vocabulary rather than shared visual descriptions.

Prosaic archetypes separate cleanly. Meteor observations (A18), drone sightings (A12), missile launches (A1), and balloon misidentifications (A6) each form their own cluster rather than mixing with the non-prosaic archetypes.

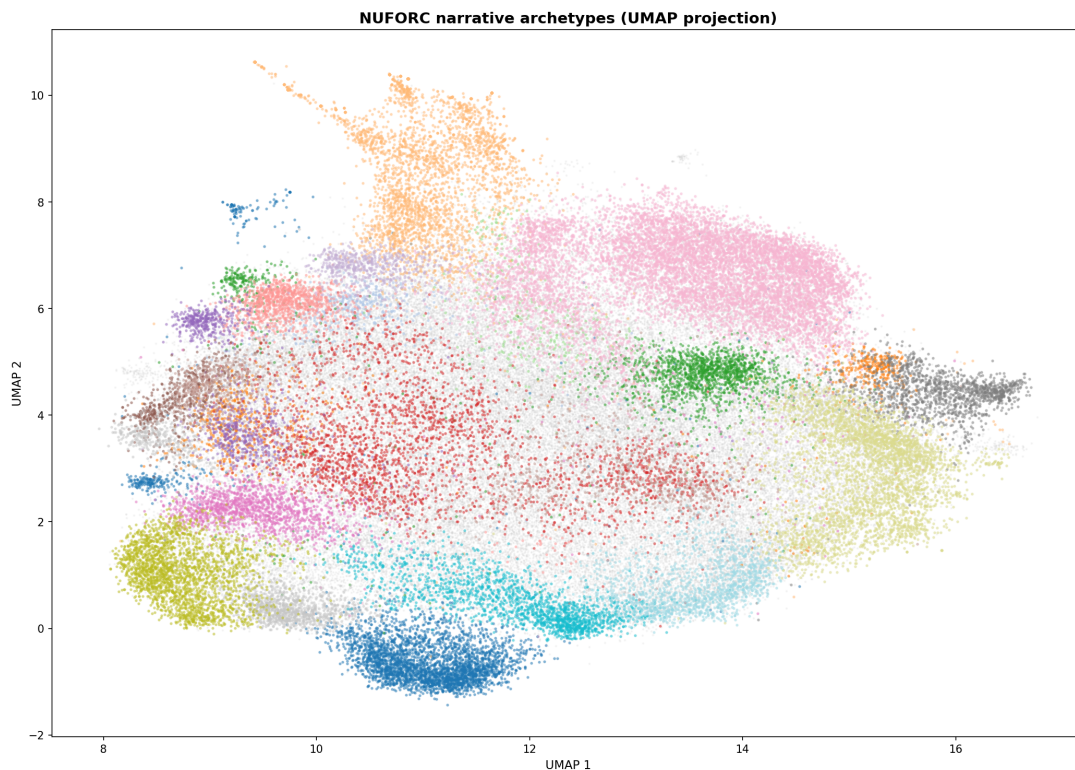


Figure 6. UMAP 2D projection of 112,000 narrative embeddings, colored by archetype.

Archetype stability

We test sensitivity to clustering hyperparameters by varying `min_cluster_size` from 100 to 500 and `min_samples` from 5 to 20 across a 20-point grid, holding the UMAP reduction fixed. The number of discovered archetypes ranges from 2 to 62, reflecting a bimodal structure in the embedding space: at low `min_samples` (5–10), HDBSCAN finds 2–3 dominant density peaks that absorb nearly all reports; at higher `min_samples` (15–20), it resolves these peaks into 13–62 substructures. The 30 archetypes we report at our reference setting (`min_cluster_size=200`, `min_samples=10`) represent one resolution of this hierarchy. The key qualitative finding — that prosaic archetypes (meteors, drones, launches, balloons) separate cleanly from non-prosaic archetypes, and that sound-only and animal-reaction clusters emerge as novel categories — holds across the `min_samples=15–20` range where the finer structure is resolved.

7. Flaps: When and Where Reports Surge

A “flap” in UFO parlance is a period of unusually concentrated reporting in a geographic area. We formalize this: for each state, we compute the annual baseline reporting rate (reports per week) and flag weeks that exceed $3\times$ the baseline. Overlapping detections within 14 days are merged.

The algorithm detects **530 flaps** in the 1990–2023 period.

Flaps are short. The median duration is 7 days. Most are single-week events centered on a single trigger: a dramatic sighting, a missile launch, a holiday.

Flaps decay fast. The median half-life — time from peak reporting to half the peak rate — is **1 day**. This is consistent with a model where a triggering event generates immediate reports, a brief secondary wave as news spreads, and then rapid return to baseline. Flaps do not self-sustain.

Most flaps are regional. 75% of detected flaps have concurrent elevated reporting in at least one adjacent state, suggesting that whatever triggers a flap is often visible across state lines.

Threshold sensitivity. The $3\times$ baseline threshold is conventional but arbitrary. At $2\times$, the algorithm detects 4,737 raw weekly exceedances; at $4\times$, 1,014. The top 20 flaps by intensity ratio

are identical across all three thresholds, and the Phoenix Lights flap at 28× is the highest-intensity event in the corpus regardless of threshold choice.

The highest-intensity flaps correspond to reference events:

Flap	State	Date	Reports	Ratio	Shape	Event
#2	AZ	Mar 1997	70	28.0×	triangle	Phoenix Lights
#1	IL	Sep 2005	72	18.8×	light	Tinley Park (late wave)
#12	OH	Nov 1999	40	18.9×	fireball	Leonid meteor shower
#3	IL	Oct 2004	64	17.1×	light	Tinley Park
#0	CA	Nov 2015	138	10.8×	light	Trident missile test

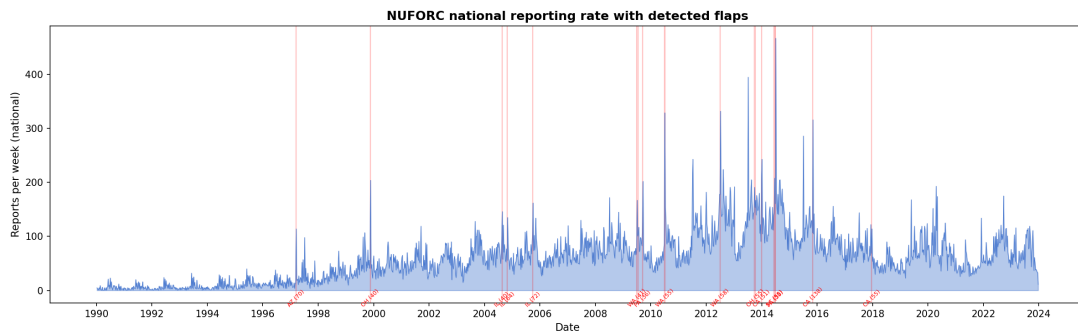


Figure 7. National weekly reporting rate with detected flaps (red bands).

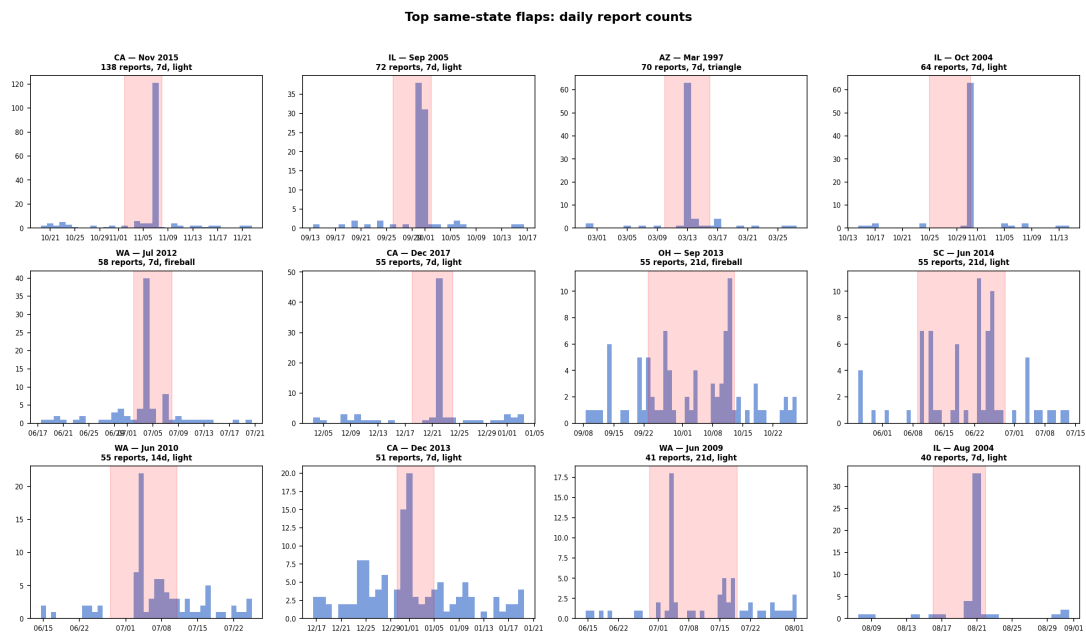


Figure 8. Daily report counts for the 12 largest flaps. Note the 1-day half-life: flaps do not self-sustain.

The Phoenix Lights flap has the highest intensity ratio in the dataset: 28× Arizona’s normal weekly reporting rate. No other event in 33 years produced that degree of anomalous reporting relative to baseline.

8. Signature Phrases

Each well-documented event in the corpus produces a distinctive linguistic fingerprint — bigrams and trigrams that appear at elevated rates in event reports compared to the corpus baseline. We express distinctiveness as a **lift ratio**: the phrase’s rate within the event divided by its corpus-wide baseline rate.

Event	Phrase	Event rate	Baseline	Lift
Tinley Park (2004)	“tinley park”	26.9%	0.1%	391×
SpaceX Launch (2017)	“vandenberg afb”	47.5%	0.2%	283×
Trident Missile (2015)	“missile launch”	83.2%	0.4%	201×
Phoenix Lights (1997)	“phoenix lights”	26.2%	0.2%	142×
Stephenville (2008)	“fort worth”	6.8%	0.1%	99×
July 4th Lanterns (2012)	“watching fireworks”	7.3%	0.2%	42×
Tinley Park (2004)	“triangle formation”	19.4%	0.8%	23×
Leonid Meteors (1999)	“meteor shower”	17.1%	0.9%	19×
Tinley Park (2004)	“red lights”	70.1%	4.8%	15×
Starlink (2020)	“single file”	2.4%	0.2%	10×
Starlink (2020)	“evenly spaced”	4.1%	0.5%	8×

Signature phrases by known UFO event



Figure 9. Signature phrases for each reference event, ranked by distinctiveness score.

The Tinley Park fingerprint produces the highest lifts: “tinley park” at 391× and “red lights” at 15×. Witnesses converge on nearly identical language independently.

The Starlink signature is linguistically noteworthy because it contains phrases that witnesses improvised to describe a phenomenon no one had words for: “evenly spaced,” “single file,” “straight line.” These are not UFO terms — they are descriptive inventions for a novel visual experience (a satellite train). Their lift ratios are modest (8–10×) because the phrases are generic, but their co-occurrence within Starlink reports is distinctive.

9. Validation Against Reference Events

We assemble a reference set of 10 events with established contexts: five high-volume historically prominent sightings (Phoenix Lights, Tinley Park, Stephenville, O’Hare, Hudson Valley) and five events with known physical triggers (Trident missile test, SpaceX Vandenberg launch, Leonid meteor shower, Starlink trains, July 4th Chinese lanterns). For each event, we ask whether four independent detection methods — same-night clustering, flap detection, signature phrase analysis, and shape distribution — surface the event without guidance.

All 10 events are detected by at least two methods, and 7 of 10 are detected by three or more. The detection misses are individually interpretable: Stephenville’s reports were distributed across multiple weeks rather than concentrated into a single flap; O’Hare’s witnesses described a single hovering disc, which the shape filter could not distinguish from the “circle” baseline; Hudson Valley’s wave spanned several years, too diffuse for the weekly flap detector.

Known event validation: detection matrix

event	Cluster	Flap	Signature	Shape
Phoenix Lights	1	1	1	1
Tinley Park IL Lights	1	1	1	1
Stephenville TX	1	0	1	1
O’Hare Airport Disc	1	1	1	0
Hudson Valley Wave	1	0	1	1
Trident Missile Launch	1	1	1	1
SpaceX Vandenberg Launch	1	1	1	1
Leonid Meteor Shower 1999	1	1	1	1
Starlink Satellite Trains	1	1	1	0
July 4th Chinese Lanterns 2012	1	1	1	1

Figure 10. Validation matrix. Green = detected by that method. All 10 reference events pass ($\geq 2/4$ methods).

The pipeline’s ability to surface these events without prior knowledge — and its ability to correctly segregate known-cause events from historically prominent sightings into separate clusters and archetypes — constitutes the primary evidence that the methodology is sound.

10. What This Is and Isn't

This is a corpus-linguistics analysis. We treat witness narratives as text and look for statistical structure: vocabulary trends, clustering patterns, temporal dynamics. The methodology is deliberately agnostic about what witnesses actually saw.

This is not an attempt to prove or debunk the reality of UFOs, UAP, or any specific sighting. The analysis cannot distinguish between “witnesses describe the same thing because they all saw the same thing” and “witnesses describe the same thing because they share a cultural vocabulary for anomalous aerial events.” Both explanations are consistent with the data. We return to the framing articulated in Section 1.1: both the cultural-priming reading and the observational reading predict identical statistical signatures. We do not resolve this ambiguity; we document the structure precisely enough that future work can test against it.

Known limitations:

- **Selection bias.** NUFORC reports are self-selected. The corpus over-represents people motivated enough to find and fill out the NUFORC form, likely skewing toward those who believe their sighting was genuinely anomalous. This bias is structural and affects every finding: the vocabulary curves reflect what *NUFORC reporters* said, not what all observers said.
- **Temporal bias.** 89% of the corpus post-dates 1990. Pre-internet reports are dramatically underrepresented.
- **Embedding model.** We use `all-MiniLM-L6-v2`, a general-purpose sentence transformer. A model fine-tuned on observational or phenomenological language might produce different clustering results.
- **Geocoding.** 14.1% of reports use state centroid coordinates. These are included in temporal and textual analyses but may introduce noise in geographic analyses.

11. Open Data and Reproducibility

The analysis code is MIT-licensed and available at [repository URL]. The pipeline is fully reproducible from a single `make all` command.

What we release: all analysis code (scripts 00–10, dashboard, tests), derived analytical datasets (embeddings, cluster labels, archetype assignments, flap detections, signature phrases, validation results), and the interactive dashboard (Streamlit).

What we do not release: the underlying NUFORC narrative text. This remains the property of NUFORC. Readers wishing to reproduce the analysis must obtain the source data independently from nuforc.org or a public mirror.

Future work. The May 2026 Pentagon PURSUE Release 01 contains 162 declassified records with parseable dates for 66 of them. Date-level cross-referencing against NUFORC produces 40 temporal overlaps, but date co-occurrence alone is uninformative given baseline NUFORC reporting volume. A meaningful cross-corpus analysis requires matching on location, narrative content, and reported phenomenology, and is left for future work pending fuller access to the PURSUE document set.

The interactive dashboard at [dashboard URL] allows readers to search the corpus, explore archetypes, map flaps, and read the same-night cluster narratives that support our findings.

Acknowledgments

Data sourced from the National UFO Reporting Center (nuforc.org), used with research permission. Any errors of methodology, interpretation, or framing are the author's alone.