

# Atypicality score v0 (speaker×dataset)

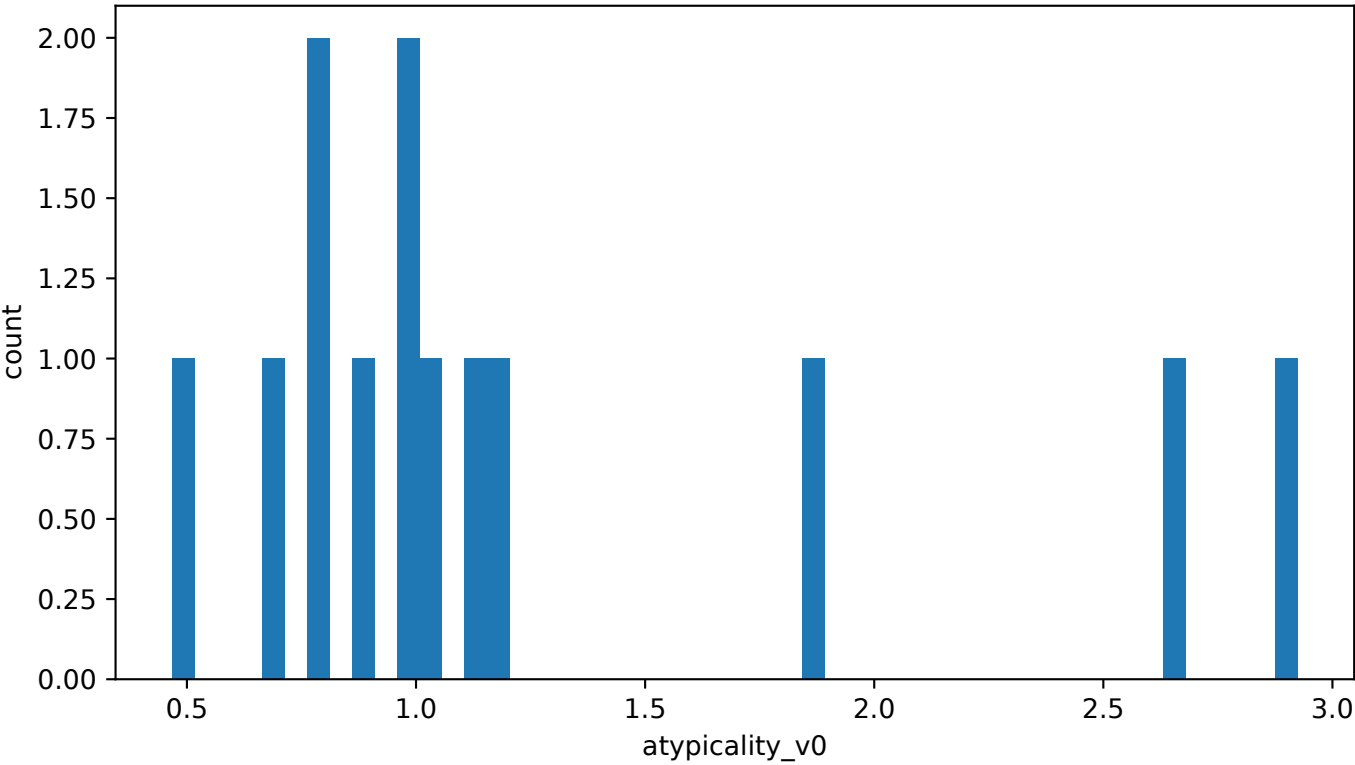
- rows: 15
- datasets: cejc, csj
- score: RMS(z) within dataset
- scaler: robust\_z (median/MAD) if generated as default

Output columns:

dataset, speaker\_id, role, n\_rows, atypicality\_v0, top\_contrib\_json, is\_outlier\_p99

Note: v0 is a deviation score for pragmatic feature distribution, not a diagnosis.

# Distribution: cejc



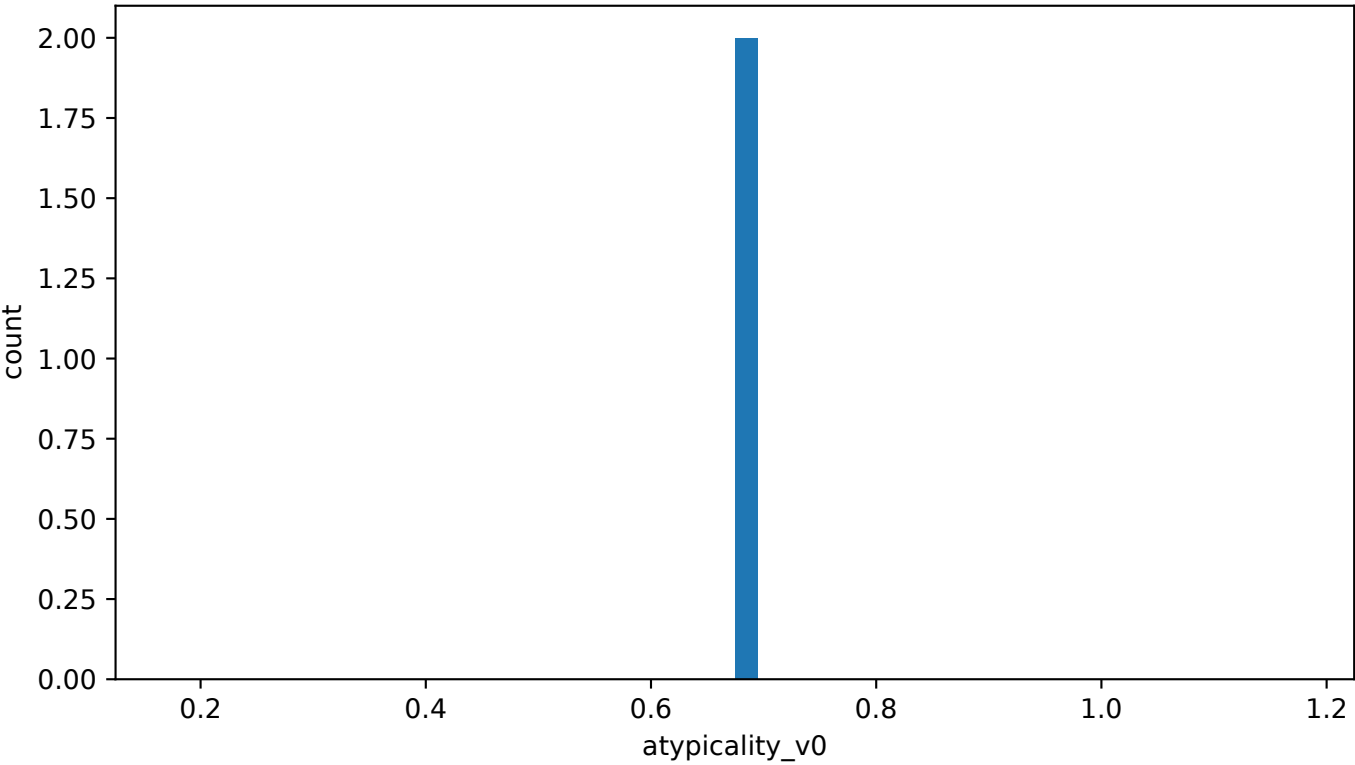
```
{
  "n": 13,
  "mean": 1.25963,
  "std": 0.7279,
  "p50": 0.96831,
  "p90": 2.50419,
  "p95": 2.76667,
  "p99": 2.89301,
  "max": 2.9246
}
```

Next: see the next page for Top-N speakers table.

## Top 30 speakers: cejc

speaker_id	role	n_rows	atypicality_v0	is_outlier_p99
Z10A		96	2.9246	True
Z10B		40	2.66138	False
Z101		38	1.87545	False
IC01		1040	1.17749	False
IC07		46	1.14944	False
N20A		38	1.01471	False
IC03		632	0.96831	False
IC02		952	0.967695	False
IC08		30	0.906073	False
IC04		408	0.782244	False
N10A		126	0.775076	False
IC06		150	0.705605	False
IC05		222	0.467134	False

# Distribution: csj



```
{
  "n": 2,
  "mean": 0.674491,
  "std": 0,
  "p50": 0.674491,
  "p90": 0.674491,
  "p95": 0.674491,
  "p99": 0.674491,
  "max": 0.674491
}
```

Next: see the next page for Top-N speakers table.

Top 30 speakers: csj

speaker_id	role	n_rows	atypicality_v0	is_outlier_p99
R		36	0.674491	True
L		219	0.674491	True

## How to use v0 (next)

- 1) Pick outliers (p99 or top-N) per dataset.
- 2) For each outlier, inspect `top_contrib_json` (features with largest  $|z|$ ).
- 3) Then go to `analysis/v1/gold=v13/examples` and sample representative turns.
- 4) (Phase3-2) LLM labeling: add functional labels (repair/question/backchannel/topic-shift e