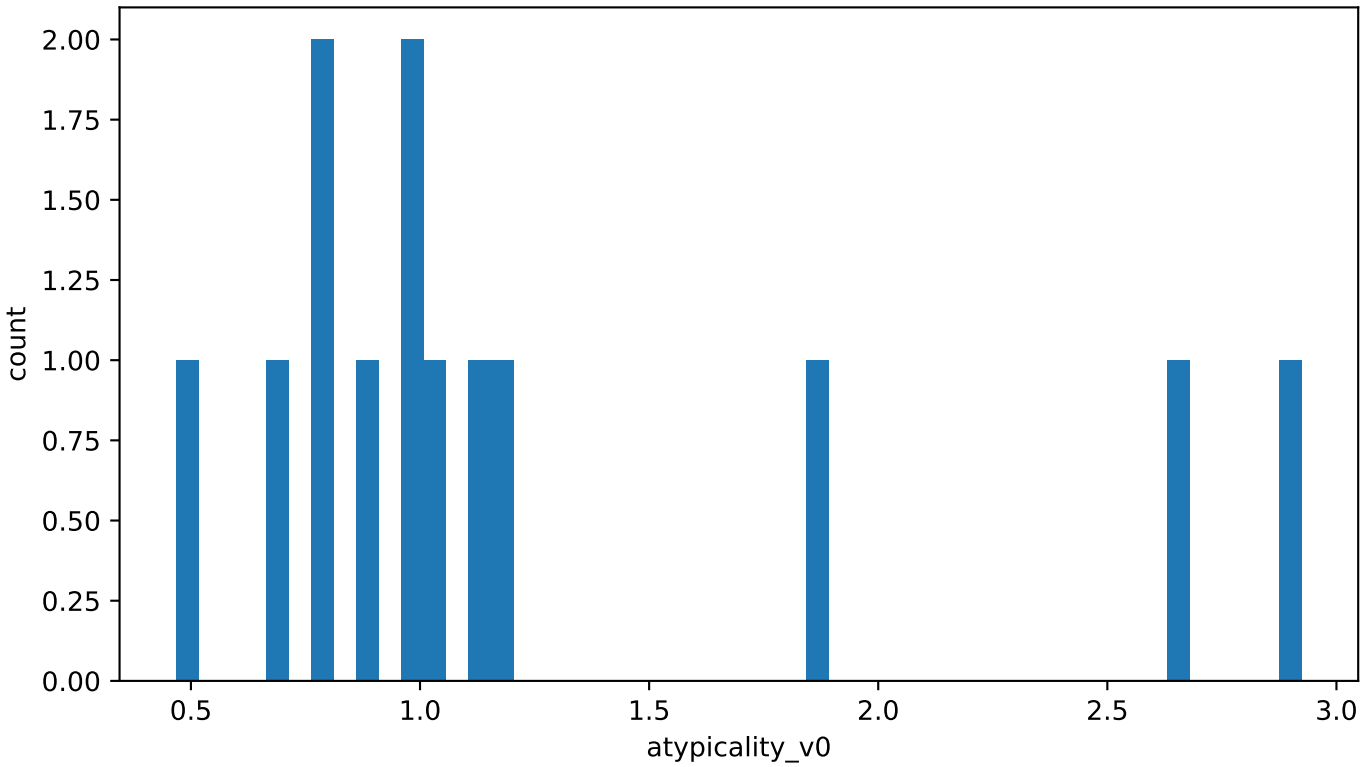
Atypicality score v0 (speaker×dataset)

- 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

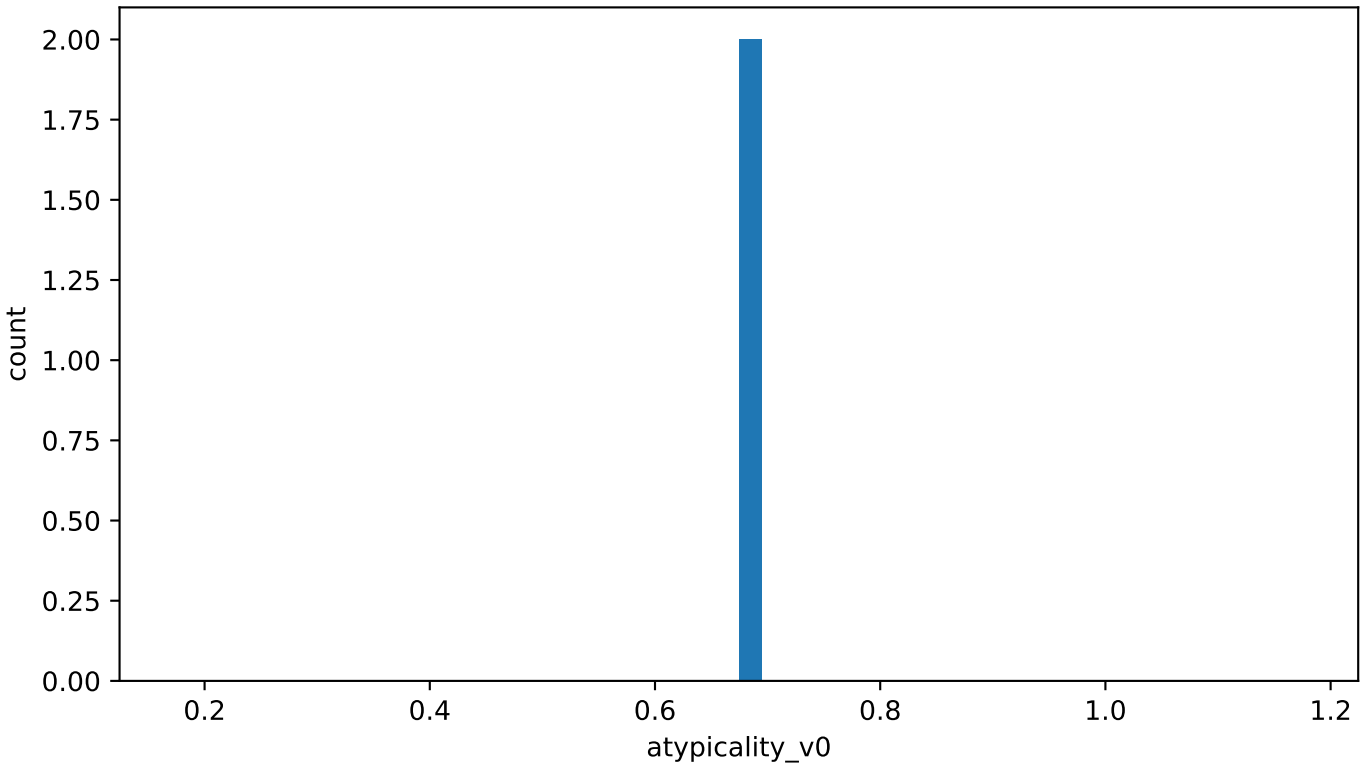



```
{
  "n": 13,
  "mean": 1.2596314545608143,
  "std": 0.7279004318259444,
  "p50": 0.9683098899735689,
  "p90": 2.5041913484748486,
  "p95": 2.7666662369763702,
  "p99": 2.893013574344766,
  "max": 2.9246004086868882
}
```


Top 30 speakers (highest scores)

speaker_id		n_rows	atypicality_v0	is_outlier_p99
Z10A		96	2.924600408686882	True
Z10B		40	2.661376789169249	False
Z101		38	1.875449585697242	False
IC01		1040	1.177494520807374	False
IC07		46	1.149438304757736	False
N20A		38	1.014713364243442	False
IC03		632	0.968309889973689	False
IC02		952	0.967694771455605	False
IC08		30	0.906072766450474	False
IC04		408	0.782244216462994	False
N10A		126	0.775075566805002	False
IC06		150	0.705605165719365	False
IC05		222	0.467133559061528	False

Distribution: csj




```
{
  "n": 2,
  "mean": 0.6744907594765952,
  "std": 0.0,
  "p50": 0.6744907594765952,
  "p90": 0.6744907594765952,
  "p95": 0.6744907594765952,
  "p99": 0.6744907594765952,
  "max": 0.6744907594765952
}
```


Top 30 speakers (highest scores)

speaker_id	n_rows	atypicality_v0	is_outlier_p99
R	36	0.674490759476595	True
L	219	0.674490759476595	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