

使用说明

脚本

- `code/parse_clec_records.py`：把 `ST3.txt` 解析成 `ST3_records.json`（只有在文件不存在或你想重建时才需要跑）
- `code/clec_dependency_analysis.py`：主流程，依存解析 + 统计，输出核心结果 JSON
- `code/clec_dependency_plots.py`：可选，把 JSON 生成 CSV 和 PNG
- `code/clean_clec.py`：可选，清理原始文本里的纠错标签，生成 `ST3_clean.txt`（不影响主流程）

怎么跑

如果 `data/.../CLEC/ST3_records.json` 已经有了，可以从第 2 行开始跑。

```
python .\code\parse_clec_records.py
python .\code\clec_dependency_analysis.py
python .\code\clec_dependency_plots.py --write-tables
```

输出与环境

- 核心输出：`output/ST3_dependency_collocations.json`
 - `overall_top_30` / `overall_top_30_normalized`
- `group_comparison`（低分组 / 高分组对比）
- `differential_top_30_normalized`（含 `log_likelihood`、`significance`）
- 常用结果表：`output/differential_top_30_normalized.csv`（可直接用于论文结果表）
- Python：3.11
- 需要安装的依赖：`pip install stanza,matplotlib`