⑤ 三人称代名詞エッジ抽出プロンプト（幕/場スコープ対応・感情なし・名詞由来は対象外・エラー行出力あり）

あなたは、指定スコープ（幕/場）の脚本テキストから、

「三人称代名詞（er/ihn/ihm/sie/ihr/ihnen…）に基づく有向辺」を抽出し、CSVのみを出力します。

名詞に基づく辺は対象外（名詞は別プロンプト）。Aliasリストは参照しません。

# 出力ヘッダ（順序固定・厳密一致）

Edge\_Id,Work\_Name,Speech\_Id,Source\_Node\_Id,Target\_Node\_Id,Alias\_Id,Alias\_Label,Span\_Start,Span\_End,Matched\_By,Excluded\_Flag,Speech\_Quote,Remarks

# 入力

<WORK\_NAME> … </WORK\_NAME>

<SCOPE>

Unit: act | scene

Value: 例) I または I-3

</SCOPE>

<NODE\_LIST\_CSV>

Node\_Id,Name,Work Name,First Line,Grammatical Number,Existed on initial list?,Remarks

…（このスコープを含むノードリストCSV）…

</NODE\_LIST\_CSV>

<SCOPE\_SCRIPT>

（指定スコープ＝該当の幕/場のみの原文。発話ラベルと本文を含む）

</SCOPE\_SCRIPT>

# 制約

- CSVのみ出力（解説・空行・コードフェンス禁止／RFC 4180エスケープ）。

- 対象は三人称の\*\*人称代名詞のみ\*\*：er/ihn/ihm、sie/ihr/ihnen 等。

- \*\*所有代名詞（sein/seine/ihr/ihre…）は対象外\*\*。

- \*\*Sie（敬称2人称）や文頭の「Sie」は対象外\*\*（二人称プロンプトで扱う）。

- 位置は \*\*Speech\_Id + Span\*\* のみで特定（外部ブロック番号は使わない）。

# 手順（要点）

1) 発話抽出と Speech\_Id 付与

- <SCOPE> に従って当該スコープのみ処理。

- <SCOPE\_SCRIPT> から「話者ラベル＋本文」を列挙し、Scene単位で `Speech\_Id` を付与。

形式：Act-Scene-U連番（例：I-3-U0007）。Sceneが無い場合は Act 内連番（I-U0007）。

2) 話者（Source\_Node\_Id）の同定

- 話者ラベルを <NODE\_LIST\_CSV> に突合して Source\_Node\_Id を確定。

- \*\*グループ話者\*\*は個体に一意分解できる場合のみ各個体に複製（分解不能なら通常は辺を生成しないが、下記エラー行規則を適用）。

3) 代名詞トークンの検出

- 対象：er/ihn/ihm、sie/ihr/ihnen 等の\*\*人称代名詞\*\*のみ（語境界必須）。

- 各出現ごとに `Span\_Start / Span\_End` を取得し、`Alias\_Label` に実文字列（例：ihn）を入れる。

4) 候補ターゲット集合の構成（同一Scene内）

- 当該発話より\*\*前方へ最大3発話\*\*を後方探索し、\*\*名指し（固有名・役称）で明示された固有対象\*\*を候補に集める。

- 候補収集は\*\*話者本人を除外\*\*。

- 名指し判定は名詞由来プロンプトの一致規則（語境界・内部正規化）に準拠。

5) 一致条件（通常出力）

- 形態整合：`er/ihn/ihm`→ masc.sg、`ihnen`→ dat.pl、

`sie/ihr` は文脈から fem.sg / pl / dat.fem.sg を\*\*明確に判定できる場合のみ\*\*採用。

- 候補が\*\*ちょうど1名\*\*に絞れるときのみ \*\*Target\_Node\_Id=その候補\*\* で出力。

- 出力値：

- `Matched\_By=pronoun-coref-3p`

- `Excluded\_Flag=n`

- `Speech\_Quote`：該当箇所を含む短い抜粋（最大10語）

- `Alias\_Id`：`PRON3P-S{Unit}{Value}-N{Target\_Node\_Id}-{Speech\_Id}-{k}`

6) エラー行の生成（Excluded\_Flag=y）

下記の場合は\*\*行を生成\*\*し、`Excluded\_Flag=y`、`Remarks` に理由を明記する。

- \*\*話者ノードが特定できない\*\*：`Source\_Node\_Id` を空欄、`Remarks="話者特定不能"`。

- \*\*話者が集合ノードで個体に分解できない\*\*：`Source\_Node\_Id` に集合ノードID、`Remarks="集合話者・個体分解不能"`。

- \*\*詩的表現等で通常の指示と見なせない\*\*（歌詞・合唱の反復など）：`Remarks="詩的表現（指示不成立）"`。

付与値の規約：

- `Target\_Node\_Id` は空欄（未確定）。

- `Alias\_Label` は検出した代名詞そのもの。

- `Matched\_By=pronoun-coref-3p`（検出経路自体は同じ）。

- `Alias\_Id` はターゲット未確定用の形式：`PRON3P-ERR-S{Unit}{Value}-{Speech\_Id}-{k}`。

- \*\*Edge\_Id\*\* 生成時のプレースホルダ：

`{WorkSlug}|{Speech\_Id}|S{SourceOrSNA}|TNA|{Alias\_Id}|X{Span\_Start}-{Span\_End}`

- `SourceOrSNA`：`Source\_Node\_Id` が無ければ `SNA` を用いる。

- `TNA`：Target 未確定を示す固定トークン。

7) Edge\_Id（主キー）の生成（決定的・通常/エラー共通）

- `{WorkSlug}|{Speech\_Id}|S{Source\_Node\_Id}|T{Target\_Node\_Id}|{Alias\_Id}|X{Span\_Start}-{Span\_End}`

- ただしエラー行は 6) のプレースホルダ規約を適用。

- `WorkSlug`：Work\_Name を小文字ASCIIに正規化（ä→ae, ö→oe, ü→ue, ß→ss、記号除去、先頭12文字程度）。

# 出力

- 次のヘッダで CSV のみ出力（順序・綴り厳守）：

Edge\_Id,Work\_Name,Speech\_Id,Source\_Node\_Id,Target\_Node\_Id,Alias\_Id,Alias\_Label,Span\_Start,Span\_End,Matched\_By,Excluded\_Flag,Speech\_Quote,Remarks