⑥ 二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、

「二人称代名詞（du/dich/dir/ihr/euch、敬称 Sie/Ihnen/Ihr …）」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で処理）。Aliasリストは参照不要だが、

同一発話内の\*\*呼格（Vocative）検出\*\*には④の名詞エッジ検出と同等のロジック（語境界・内部正規化・一意性判定）を内部利用してよい。

# 出力ヘッダ（④/⑤と完全一致・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>

（作品タイトル）

</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でエスケープ）。

- \*\*対象は二人称の人称代名詞のみ\*\*：`du/dich/dir`（2人称単数）、`ihr/euch`（2人称複数）、

\*\*敬称\*\* `Sie/Ihnen/Ihr`（大文字、2人称扱い）。

- 文頭の「Sie」は曖昧。\*\*2人称敬称と確信できない場合は採用せず、エラー行を生成\*\*。

- Wikisource等のブロック番号は使用しない。位置は \*\*Speech\_Id + Span\*\*（発話内オフセット）で特定。

- 1出現=1行（同一発話に複数あれば複数行）。

# 決定ルール（いずれか満たした場合のみ Target を一意決定）

\*\*優先度：① Vocative ＞ ② 隣接応答 ＞ ③ 単一在席\*\*

1) Vocative（最優先）

- 同一発話内で\*\*呼格（名前/役職＋区切り記号等）\*\*が\*\*一意\*\*にノードへマッチ → そのノードを Target とする。

- 例：`Herr Stauffacher, … du …` → Target=Stauffacher（Node\_Id）。

- 呼格判定は④相当の名詞一致規則（`Unique\_in\_Scope=y` の語形に限定）を内部利用して良い。

2) 隣接応答（他のVocativeが同一発話に存在しないことが条件）

- \*\*A) 質問への応答\*\*：直前の発話が `?` で終わる質問で、現在の発話に二人称代名詞が出現 → \*\*前発話の話者\*\*を Target。

- \*\*B) ターンテイキング\*\*：直前の発話者が1名で、直前発話本文に\*\*現在の発話者\*\*を指す名詞呼称が\*\*一意\*\*にマッチ → \*\*前発話の話者\*\*を Target。

（Bの名指し検出は④の名詞一致規則に準拠）

3) 単一在席

- \*\*在席者の定義\*\*：Scene内で (A) その場で一度でも発話した、または (B) ト書きで登場が指示されてから退場指示まで在席。

- 現在の発話時点で\*\*話者以外の在席ノードがちょうど1名\*\*のみ → その1名を Target。

# 曖昧回避と不採用規則

- 同一発話に\*\*複数のVocative\*\*がある／候補者が複数／在席者が2名以上 → \*\*通常出力は行わず、エラー行\*\*（Excluded\_Flag=y）。

- `ihr/euch` の\*\*複数宛\*\*は、\*\*一意の集合または個体1名に確定できる場合のみ\*\*採用。確定できなければ\*\*エラー行\*\*。

- 敬称 `Sie/Ihnen/Ihr` は\*\*大文字のみ\*\*を対象とし、2人称敬称と断定できない場合は\*\*エラー行\*\*。

# 手順（要点）

1) 発話抽出と Speech\_Id 付与

- <SCOPE> に従い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 を採番

（形式：`I-3-U0007`。Sceneなしは `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（分解不能なら通常出力は行わず、エラー行へ）。

3) 代名詞トークンの検出

- 対象：`du/dich/dir`、`ihr/euch`、`Sie/Ihnen/Ihr`（大文字）— \*\*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：`dich`、`Ihnen`）を入れる。

4) アドレス先の決定（上記ルール①→②→③の順で判定）

- 成功した場合：

- `Target\_Node\_Id` を確定。

- `Matched\_By=pronoun-coref-2p`

- `Excluded\_Flag=n`

- `Speech\_Quote`：該当箇所を含む最大10語抜粋（CSV安全）

- `Alias\_Id`：`PRON2P-S{Unit}{Value}-N{Target\_Node\_Id}-{Speech\_Id}-{k}`（同発話内連番=k）

5) ⚠ 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（監査可能性のため）。

- \*\*話者ノードが特定できない\*\*：`Source\_Node\_Id` 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ID、`Remarks="集合話者・個体分解不能"`。

- \*\*複数候補／複数Vocative／単一在席不成立\*\*：`Remarks` に具体的要因（例："候補複数"、"vocative複数"、"在席者>1"）。

- \*\*敬称Sieの曖昧\*\*：`Remarks="Sie曖昧（2人称敬称不確定）"`。

- \*\*詩的表現等で指示不成立\*\*（歌詞・合唱の反復など）：`Remarks="詩的表現（指示不成立）"`。

共通規約：

- `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-2p`。

- `Alias\_Id`：`PRON2P-ERR-S{Unit}{Value}-{Speech\_Id}-{k}`。

- \*\*Edge\_Id\*\*（プレースホルダ規約）：

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

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

- `TNA`：ターゲット未確定トークン。

6) Edge\_Id（主キー）の生成（決定的・通常/エラー共通）

- 通常出力：

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

- エラー出力：5) 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