

本研究での取り組み

非流暢性タグを付与したデータを作成し利用することで、除去すべき非流暢性が明示的に分かるようになり、流暢なテキストへのEnd-to-End音声翻訳を容易にする

背景

タスク 非流暢性を含む原言語音声 (Disfluent Speech) → 音声翻訳 ST → 流暢な目的言語テキスト (Fluent Text)



課題 明示的にどれか非流暢性がわからないため、流暢なテキストへの直接翻訳は困難なタスク

まとめと今後の方針

まとめ 目的言語テキストに対し非流暢性タグを付与したデータを用いた音声翻訳→流暢なテキストへの翻訳性能向上

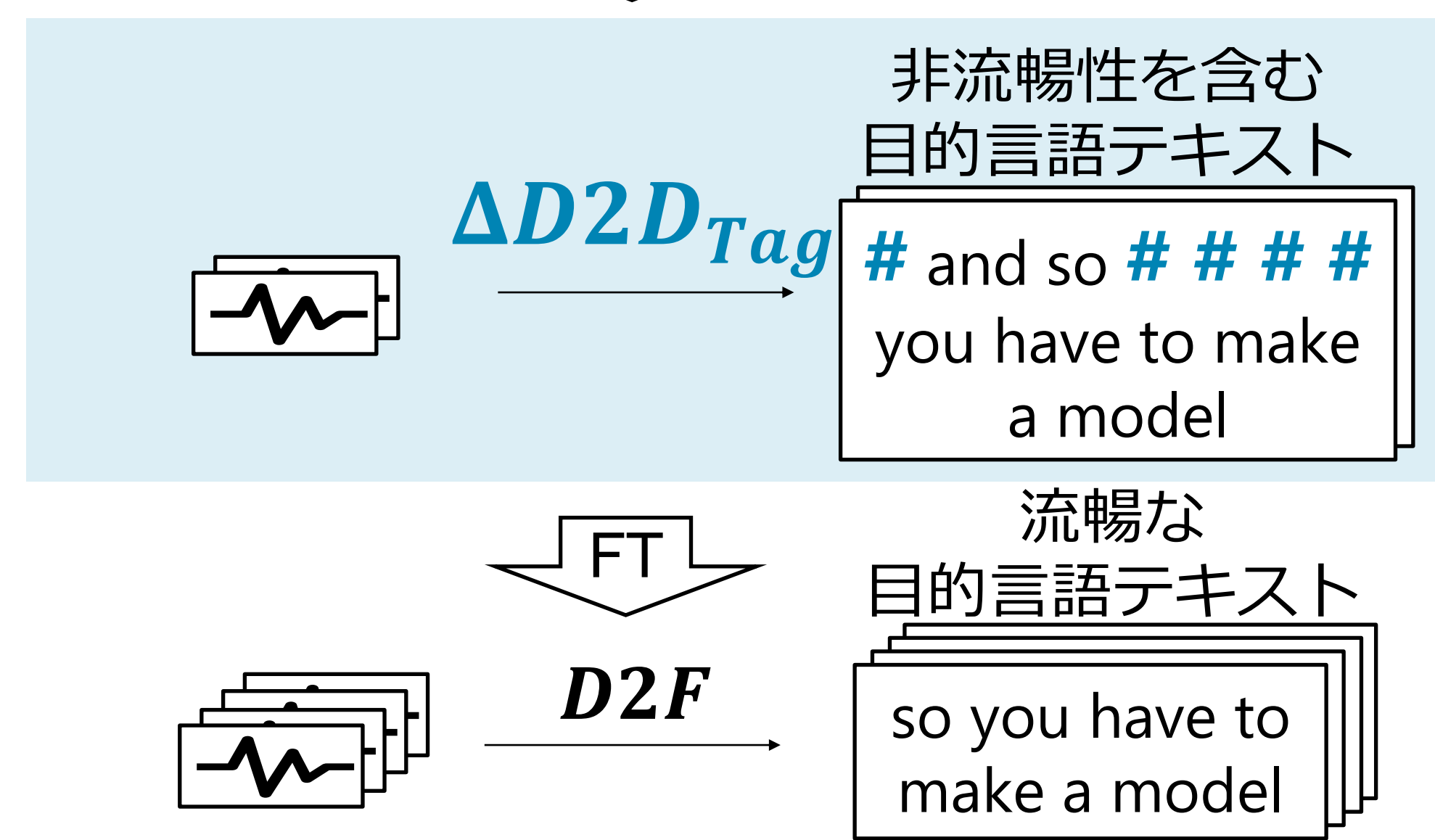
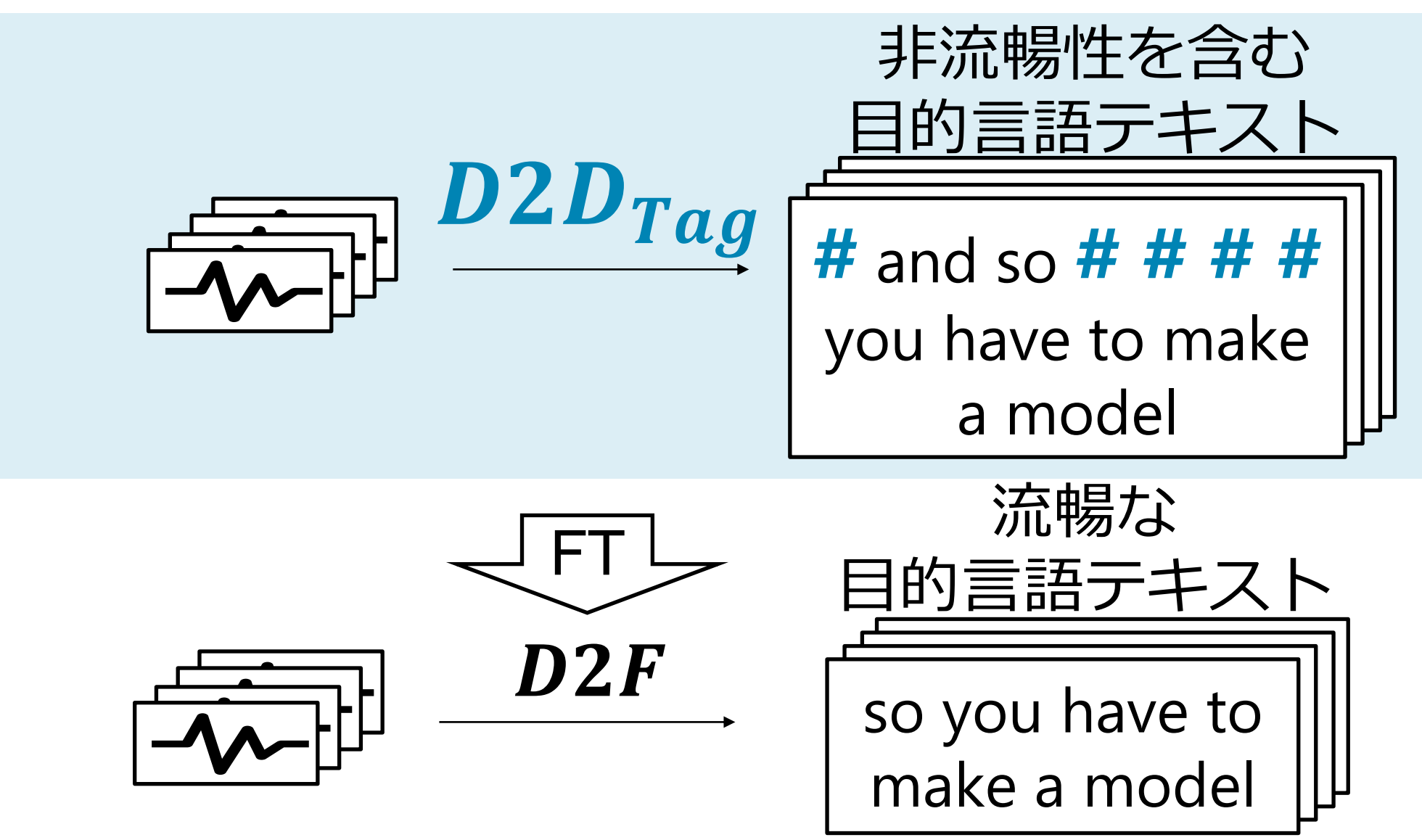
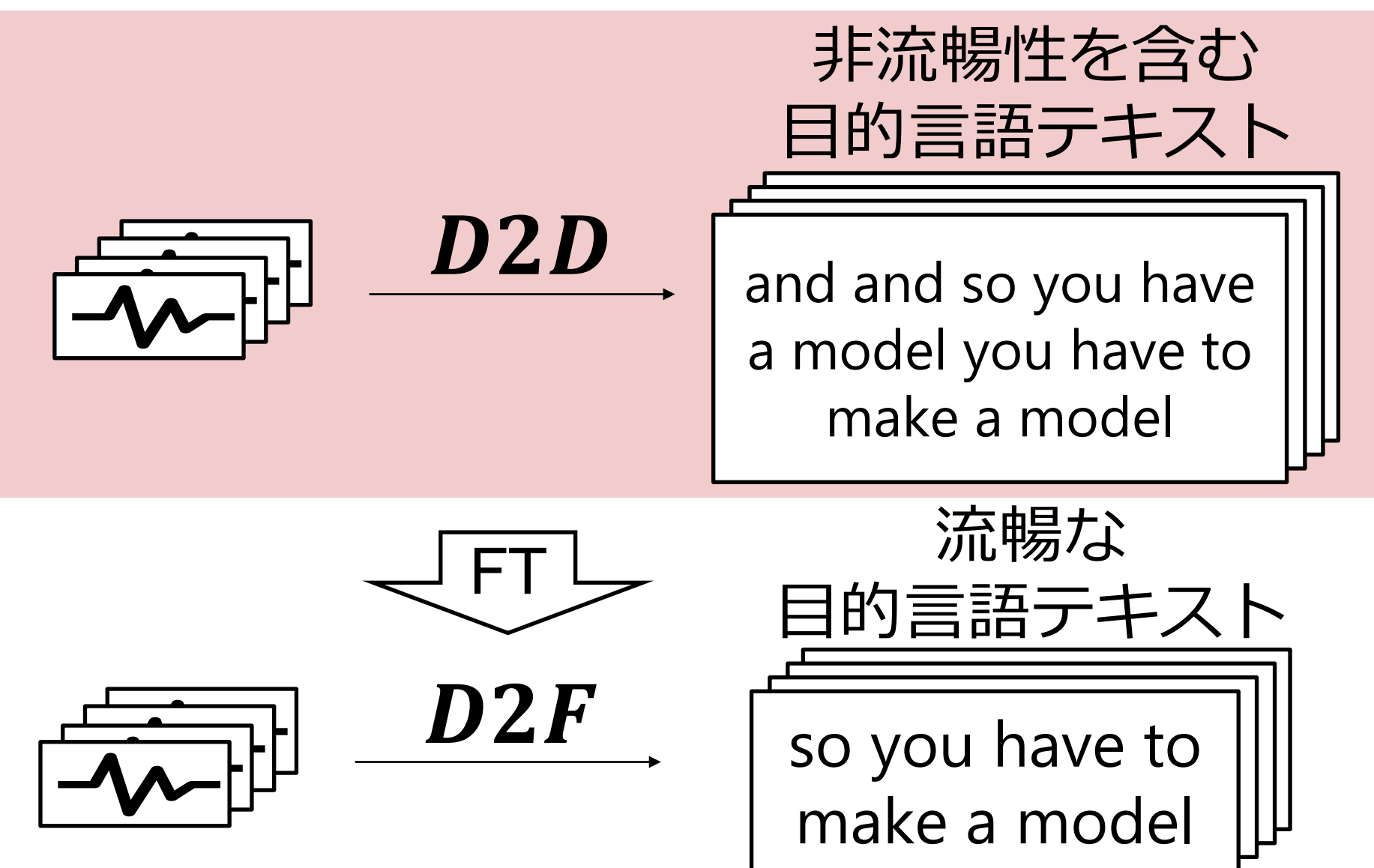
今後の方針 出力文に含まれる非流暢性タグの数を制御できる学習, 出力位置の予測精度の向上

従来手法

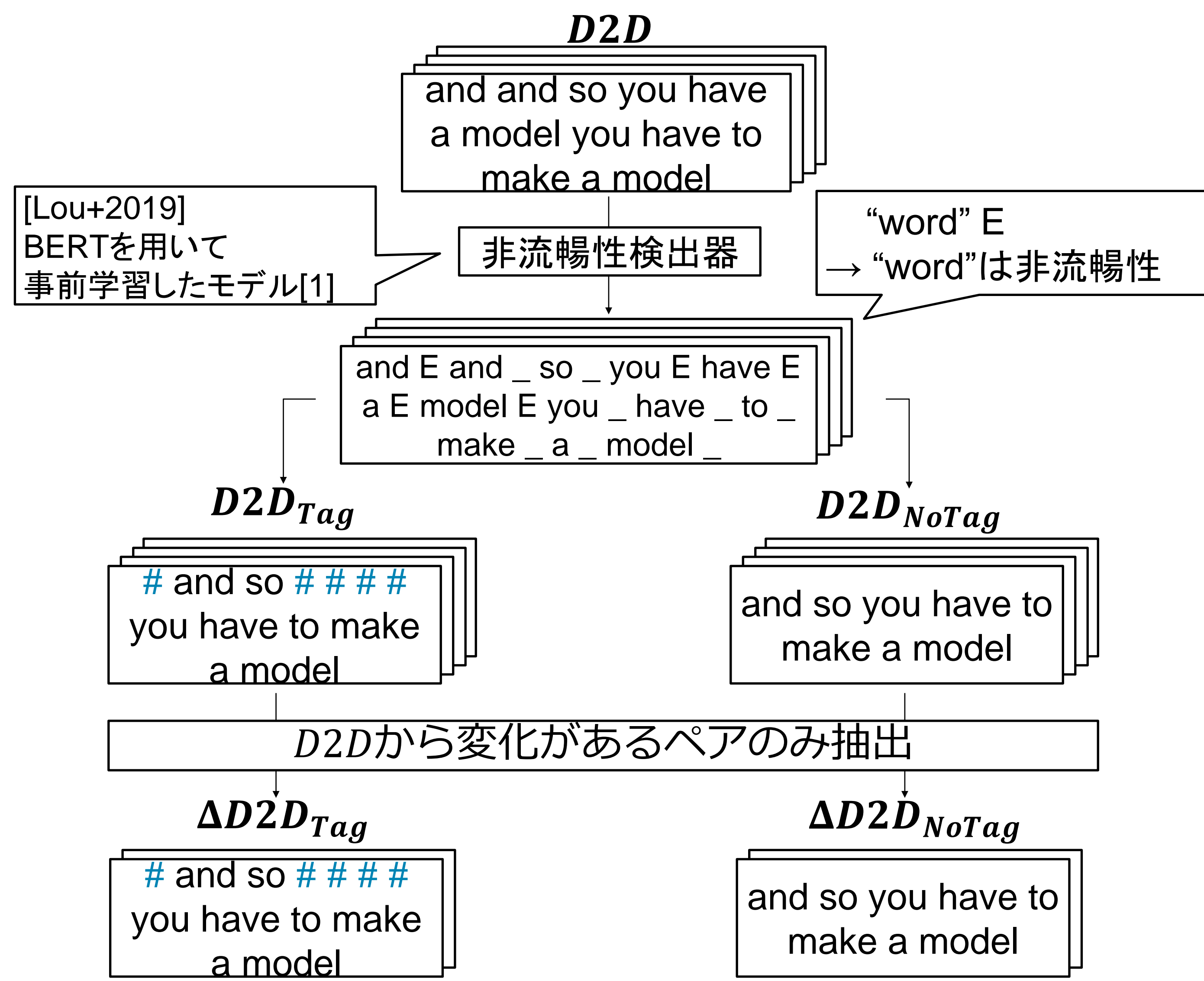
D2D: Disfluent to Disfluent → 非流暢
D2F: Disfluent to Fluent → 流暢
 (FT: Fine-tuning)

提案手法

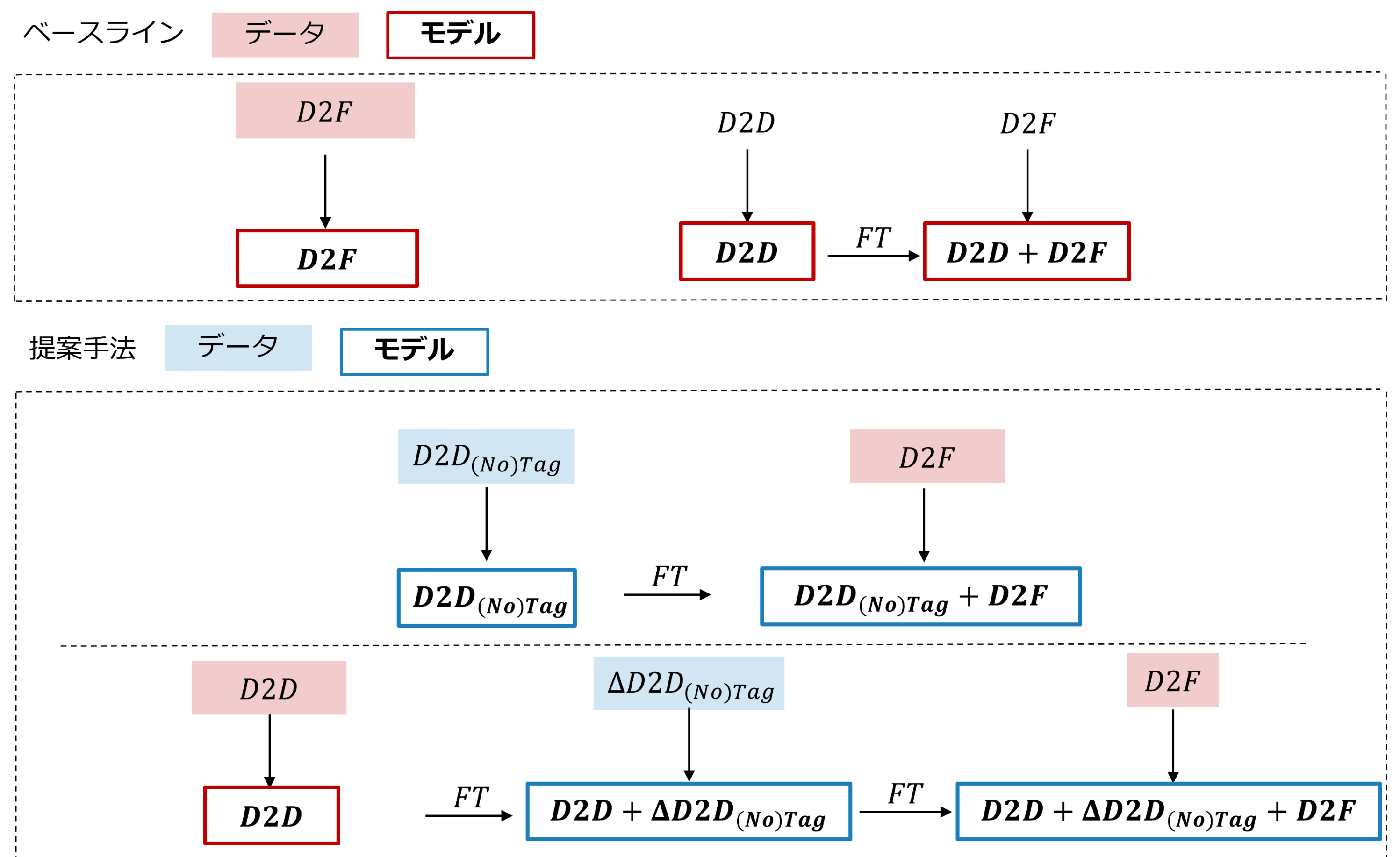
D2D_{Tag}: → タグ付き非流暢
ΔD2D_{Tag}: → タグ付き非流暢 (D2Dから変化があるペアのみ抽出)



非流暢性タグを付与したデータの作成



提案モデル学習の流れ



実験設定

データ: Fisher Spanish (Es-En) **実装**: Transformer
辞書: SentencePiece Es-En 8k **評価指標**: BLEU

結果

タグの情報を付与したモデルでBLEU値向上

原音声 → 目的言語	D2F	D2D _{NoTag}
ベースライン		
D2D	9.1	13.7
D2F	10.7	22.8
D2D + D2F	11.8	22.0
提案手法		
D2D _{Tag}	9.3	21.1
D2D_{Tag} + D2F	12.2	22.4
D2D _{NoTag}	9.8	23.4
D2D_{NoTag} + D2F	12.1	22.1
D2D + ΔD2D _{Tag}	5.1	16.6
D2D + ΔD2D_{Tag} + D2F	11.5	20.7
D2D + ΔD2D _{NoTag}	10.4	25.1
D2D + ΔD2D_{NoTag} + D2F	12.1	21.8

	D2D _{Tag} test	タグの割合
D2D _{Tag} test	-	1760 / 39538 (4.5%)
D2D _{Tag}	16.9	8926 / 41543 (21.5%)
D2D + ΔD2D _{Tag}	12.4	16877 / 44366 (38.0%)

➤ D2D_{Tag} testと比較しタグが多く出力されているが D2D_{Tag}を用いた D2D_{Tag} + D2Fで性能向上

	ベースライン
D2D	yes sure <u>that's</u> <u>that's</u> <u>that's</u> <u>that's</u> it's it that's it's an advantage that's true that there's less cold
D2D + D2F	yes <u>that's</u> <u>that's</u> definitely it's an advantage that there's less cold
提案手法	
D2D _{Tag}	yes sure <#><#><#> this is definitely it's an advantage that there's less cold
D2D _{Tag} + D2F	yes sure <u>that's</u> definitely it's an advantage right that there's less rules
D2F test	that is definitely an advantage

➤ D2D: that'sの繰り返し
 → D2D + D2F: that's の繰り返しの傾向が残っている
 ➤ D2D: that's の繰り返し部分がタグで置き換えられ出力
 → D2D_{Tag} + D2F: that's の繰り返しがなく流暢なテキストらしい出力

[1] https://github.com/pariajm/joint-disfluency-detector-and-parser/releases/download/naacl2019/swbd_fisher_bert_Edev.0.9078.pt