

編集操作予測に基づく語彙制約による制御可能なテキスト平易化

大阪大学大学院 情報科学研究科 舌 達也

Accepted by EMNLP 2022 Workshop (TSAR)

制御可能なテキスト平易化

- 特定の難易度に向けて文章を平易に変換するタスク
- 子供や非母国語話者の言語学習を支援する

Original

So Yan, a widow since her husband's death nearly a decade ago, spends every weekday at a modest community center near her home.

Grade 8

She spends every weekday at a community center near her home.

Grade 5

Yan's husband died almost 10 years ago.

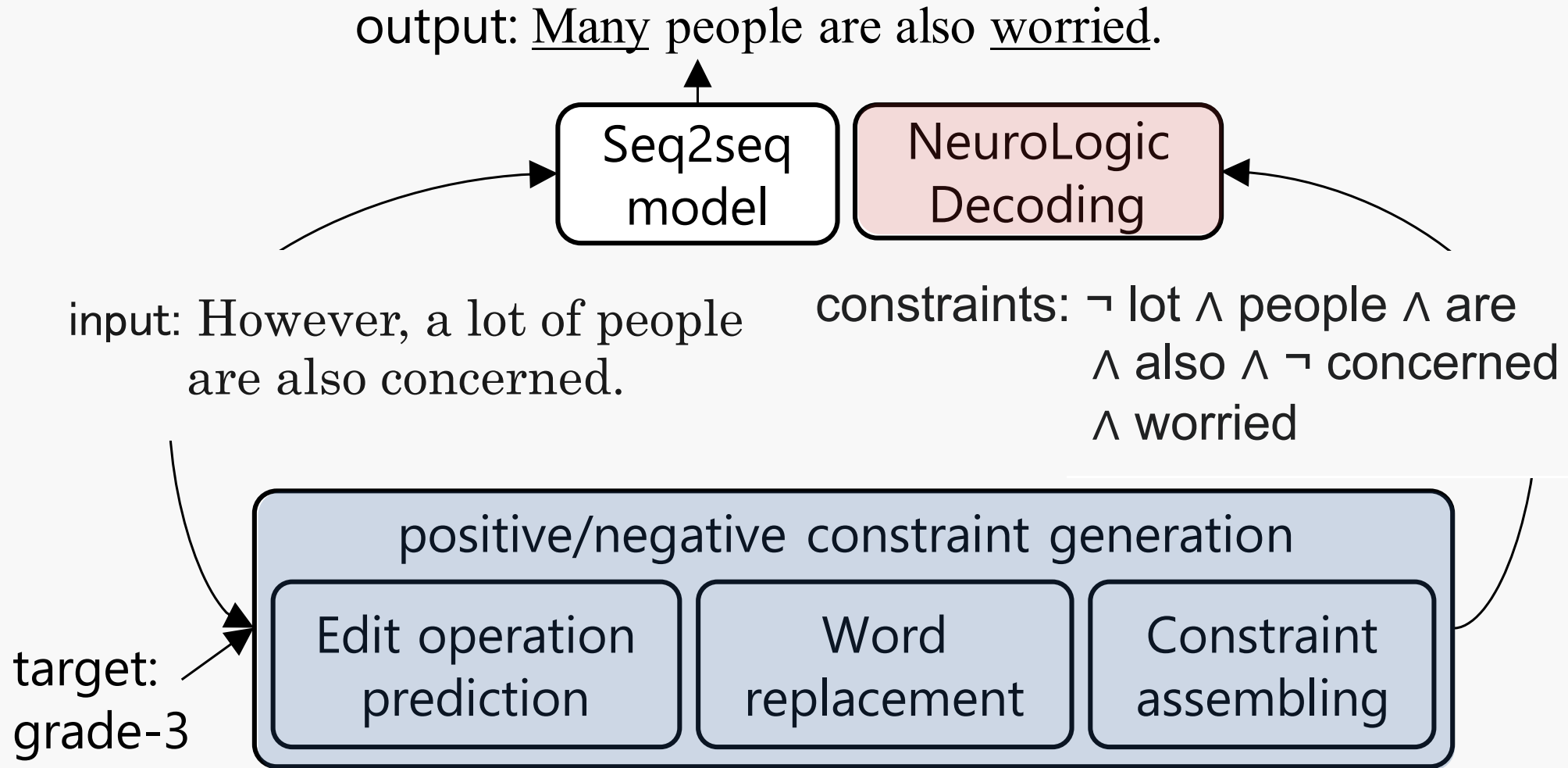
Grade 2

Yan is a widow.

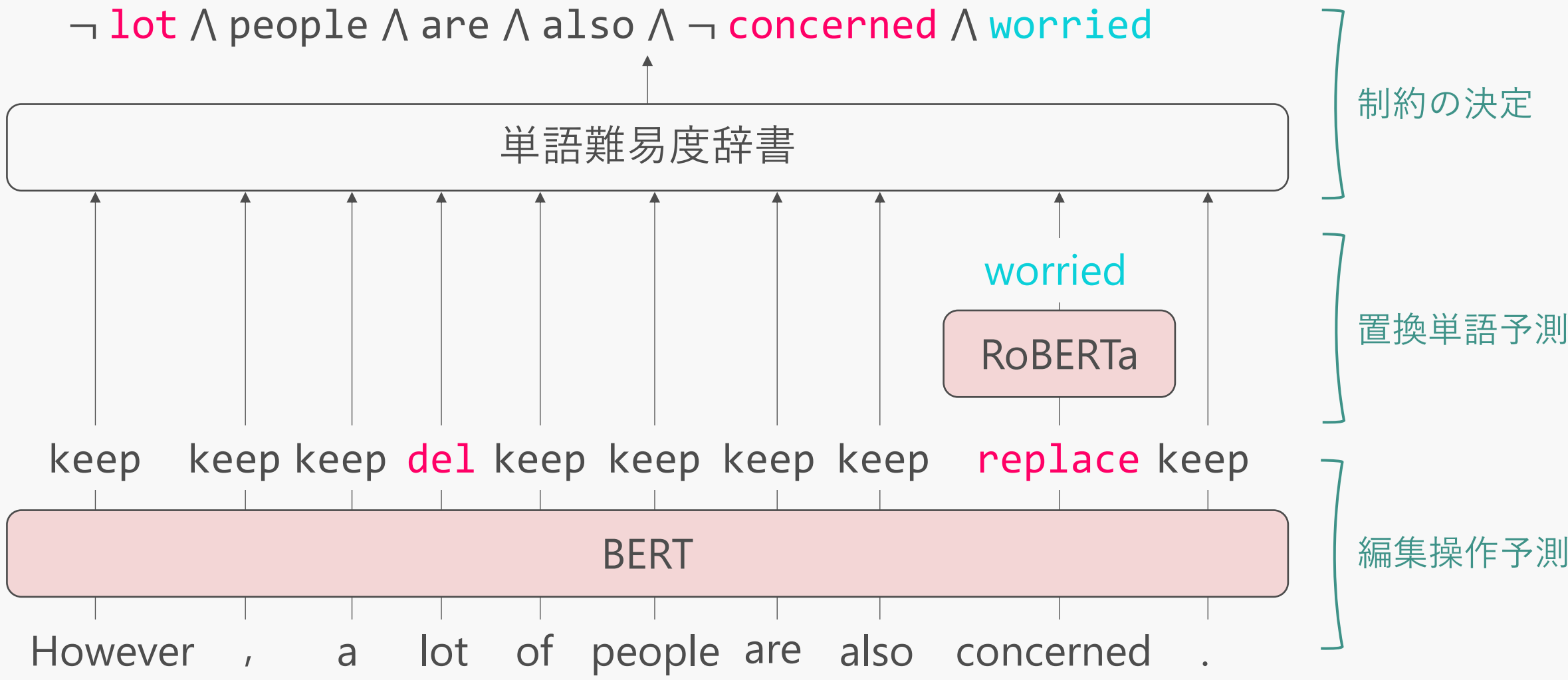
- NeuroLogic decoding [7] を用いて平易化モデルに語彙制約を加える
 - 正の制約: 出力文に出現すべき単語
 - 負の制約: 出力文に出現すべきでない単語
- 生成確率が高く (≡自然な文) かつできるだけ制約を満たした出力文を探索

$$\mathbf{y} = \arg \max_{\mathbf{y} \in \mathcal{Y}} P(\mathbf{y}|\mathbf{x}) - \lambda \sum_{i=1}^{\dots\dots\dots} (1 - C_i)$$

Number of unsatisfied constraints



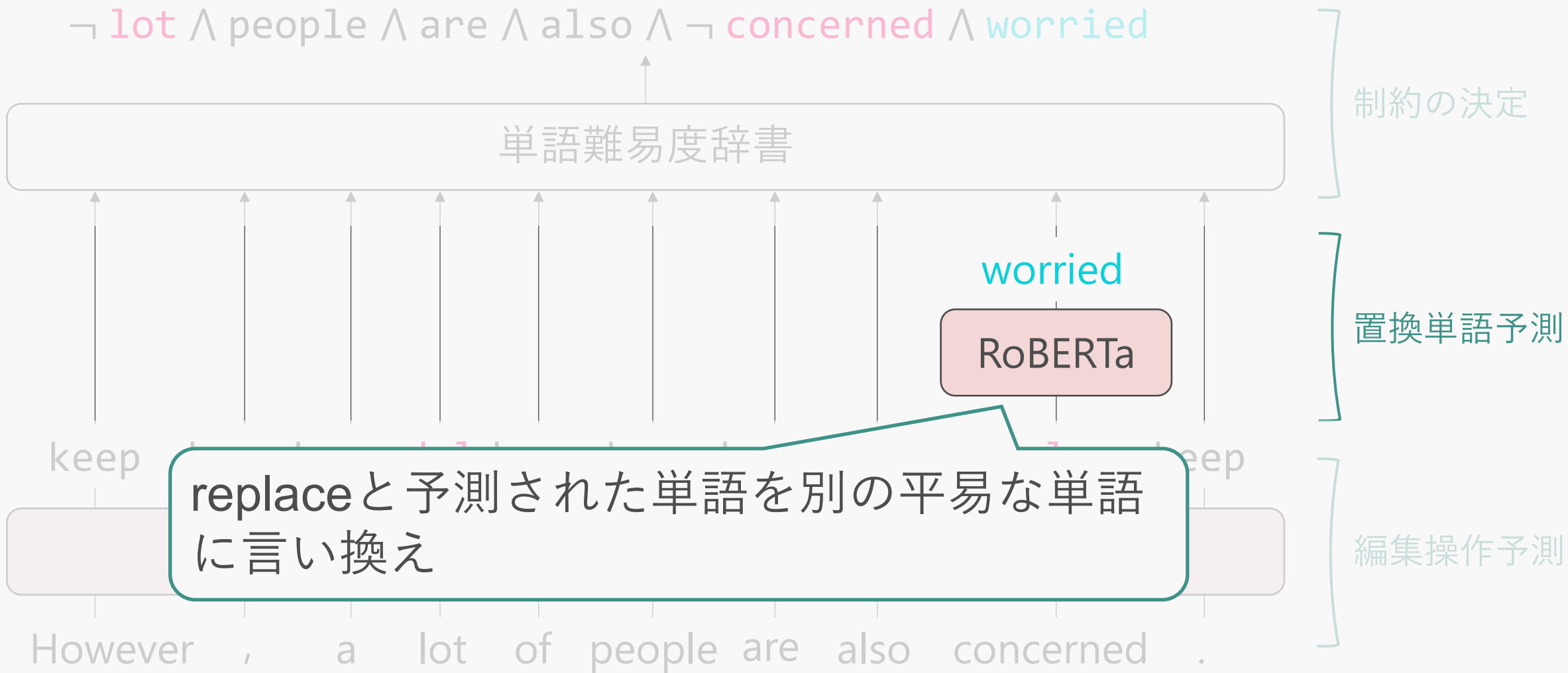
正/負の制約生成



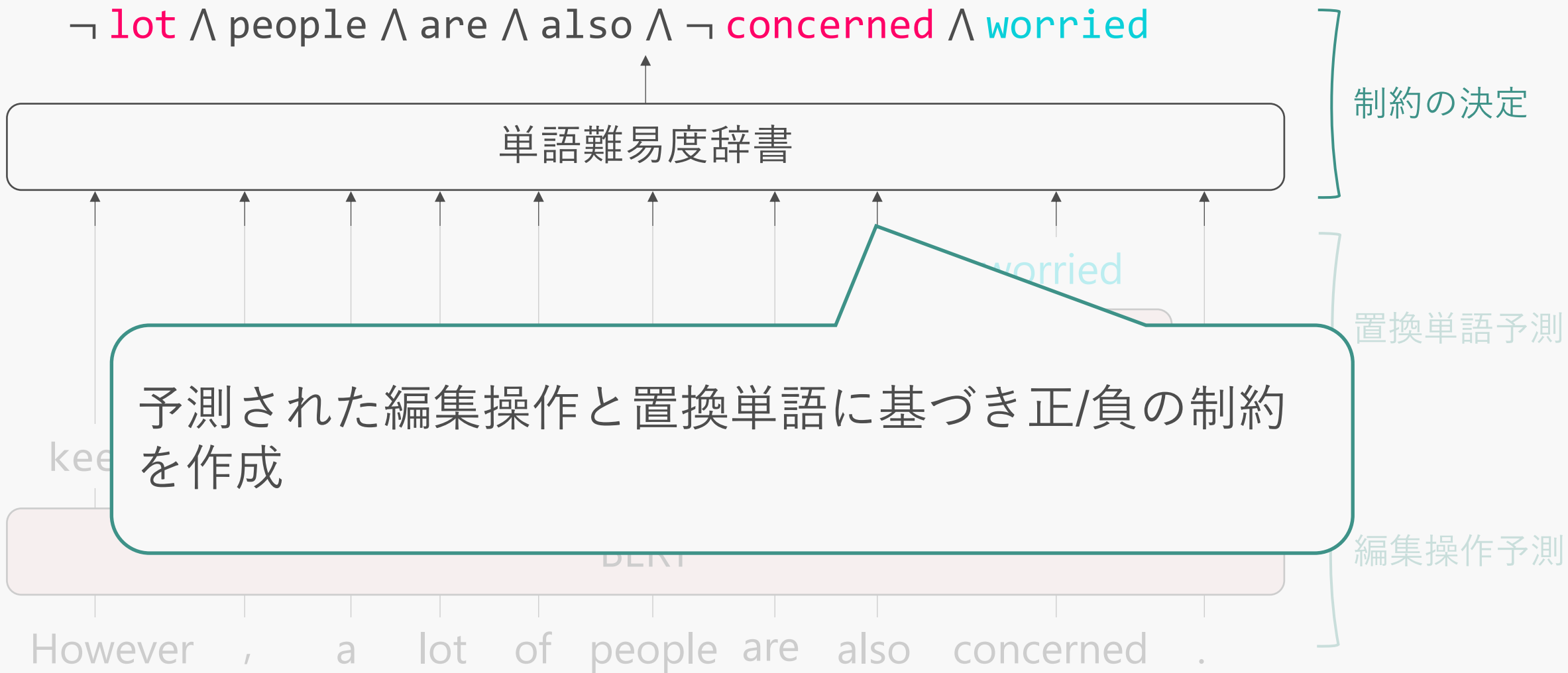
正/負の制約生成



正/負の制約生成



正/負の制約生成



制約の決定

- 単語の難易度から出力レベルの制御に関連する編集操作を選択
- 全ての単語が正/負の制約になるわけではない
- 制約にしなかった単語の出力の要否はモデルに委ねる

対象学年: l	replace	keep	delete
単語難易度 $\leq l$	-	正	-
単語難易度 $> l$	負, 正	-	負

制約の決定

- 単語の難易度から出力レベルの制御に關連する操作操作を選択
- 全
- 制

予測された操作がdelかつ対象学年より難解な単語
→負の制約

対象学年: l	replace	keep	delete
単語難易度 $\leq l$	-	正	-
単語難易度 $> l$	負, 正	-	負

制約の決定

- 単語の難易度から出力レベルの制御に關連する編集操作を選択
- 全
- 制

予測された操作がkeepかつ対象学年より平易な単語
→ 正の制約

対象学年: l	replace	keep	delete
単語難易度 $\leq l$	-	正	-
単語難易度 $> l$	負, 正	-	負

制約の決定

- 単語の難易度から出力レベルの制御に関する操作を選択
- 全
- 制

予測された操作がreplかつ対象学年より難解な単語→**負の制約**
その平易な言い換え→**正の制約**

対象学年: l	replace	keep	delete
単語難易度 $\leq l$	-	正	-
単語難易度 $> l$	負, 正	-	負

制約の決定

- 単語の難易度から出力レベルの制御に関連する編集操作を選択
- 全ての単語が正/負の制約になるわけではない
- 制約にしなかった単語の出力の要否はモデルに委ねる

対象学年: l	replace	keep	delete
単語難易度 $\leq l$	-	正	-
単語難易度 $> l$	負, 正	-	負

- データセット

- Newsela
- Newsela-Auto

- 評価指標

- SARI
- FKGL
- Pearson correlation coefficient (対象学年とFKGL)
- MSE (対象学年とFKGL)

Model	SARI	Add	Keep	Delete	FKGL	PCC	MSE
Source	12.24	0.00	36.72	0.00	9.18	0.338	47.2
Reference	100.0	100.0	100.0	100.0	3.96	100.0	0.0
DRESS (Zhang and Lapata, 2017)	38.03	2.43	42.20	69.47	4.97	0.388	13.0
MUSS (Martin et al., 2022)	41.20	6.02	35.88	81.70	2.43	0.362	13.3
BART	38.54	3.64	40.59	71.40	4.63	0.350	13.6
EditNTS (Dong et al., 2019)	37.05	1.23	36.55	73.37	3.82	0.266	16.1
(Kumar et al., 2020)	38.37	1.01	36.51	77.58	2.95	0.334	12.6
(Kajiwara, 2019)	38.48	4.55	43.41	67.47	5.01	0.417	12.2
(Dehghan et al., 2022)	40.01	3.06	36.53	80.43	3.20	-	-
Proposed	42.65	4.55	42.49	80.90	3.74	0.420	11.1
Proposed (Oracle)	54.73	10.98	66.07	87.14	4.07	0.591	8.0

Model	SARI	Add	Keep	Delete	FKGL	PCC	MSE
Source	12.04	0.00	36.12	0.00	10.11	0.393	57.5
Reference	100.0	100.0	100.0	100.0	4.34	1.000	0.0
BART	39.66	4.16	39.17	75.65	4.38	0.342	16.4
EditNTS (Dong et al., 2019)	37.43	0.97	34.78	76.53	3.12	0.215	20.4
(Kajiwara, 2019)	38.30	4.42	40.51	69.96	5.03	0.371	16.0
Proposed	43.09	4.41	42.74	82.13	3.89	0.391	15.1
Proposed (Oracle)	51.75	7.45	61.14	86.66	4.64	0.611	9.9

- より良い制約を作成できれば提案手法はさらに改善
- しかし正解ラベルにおける出現数の少なさからreplaceの予測は難しい

Edit Operation	Newsela			Newsela-Auto		
	precision	recall	F1	precision	recall	F1
replace	0.28	0.21	0.24	0.28	0.15	0.19
keep	0.58	0.57	0.57	0.58	0.57	0.58
delete	0.70	0.73	0.72	0.73	0.77	0.75

- モデルの保守的な問題に対処するための正/負の制約の作成手法を提案
- **Newsela**と**Newsela-Auto**において高い平易性と制御性を達成
- 単語レベルの制約のみでなく、フレーズレベルの制約の検討

- [1] Zhang and Lapata 2017. Sentence simplification with deep reinforcement learning. In Proc. of EMNLP 2017
- [2] Martin et al. 2022. MUSS: Multilingual unsupervised sentence simplification by mining paraphrases. In Proc. of LREC 2022
- [3] Dong et al. 2019. EditNTS: An neural programmer-interpreter model for sentence simplification through explicit editing. In Proc. of ACL 2019
- [4] Kumar et al. 2020. Iterative edit-based unsupervised sentence simplification. In Proc. of ACL 2020
- [5] Kajiwara 2019. Negative lexically constrained decoding for paraphrase generation. In Proc. of ACL 2019
- [6] Dehghan et al. 2022. GRS: Combining generation and revision in unsupervised sentence simplification. In Proc. of ACL 2022
- [7] Lu et al. 2021. NeuroLogic decoding: (un)supervised neural text generation with predicate logic constraints. In Proc. of NAACL-HLT 2021