Guided Beam Search to Improve Generalization in Low-Resource Data-to-Text Generation

INLG 2023
Nicolas Garneau & Luc Lamontagne
Université Laval

On the menu 

1. Data-to-text Generation
2. Dataset
3. Human Evaluation of 3 Neural Architectures
4. Guided Beam Search

1. Data-to-text generation

Generating Court Dockets Descriptions using Neural Networks

UNIVERSITÉ LAVAL

Court Dockets

```
ACC.  DOE JOHN
      1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
      NAIS 01/01/1979
      AVO. DOUGH JANE
                                INFRACTION DATE 01/12/2019
                                OPENING DATE    01/01/2020
PLA.  TREMBLAY SARAH
      1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
      AVO. BOULAY JEAN

ORG.  CITY POLICE DEPARTMENT
      NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

UNIVERSITÉ LAVAL

Court Dockets

```
ACC.  DOE JOHN
      1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
      NAIS 01/01/1979
      AVO. DOUGH JANE
                                INFRACTION DATE 01/12/2019
                                OPENING DATE    01/01/2020
PLA.  TREMBLAY SARAH
      1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
      AVO. BOULAY JEAN

ORG.  CITY POLICE DEPARTMENT
      NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

*"John Doe **pleaded guilty** to **failure to comply with an order** and **mischief to property** on **October 1st, 2015**."*

Court Dockets

```
ACC.  DOE JOHN
      1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
      NAIS 01/01/1979
      AVO. DOUGH JANE
                                INFRACTION DATE 01/12/2019
                                OPENING DATE    01/01/2020
PLA.  TREMBLAY SARAH
      1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
      AVO. BOULAY JEAN

ORG.  CITY POLICE DEPARTMENT
      NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

*"John Doe **pleaded guilty** to **failure to comply with an order** and **mischief to property** on **October 1st, 2015**."*

Court Dockets

```
ACC.  DOE JOHN
      1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
      NAIS 01/01/1979
      AVO. DOUGH JANE
                                INFRACTION DATE 01/12/2019
                                OPENING DATE    01/01/2020
PLA.  TREMBLAY SARAH
      1130, ROUTE PRINCIPALE QUEBEC (QUEBEC) G2G - 2G2
      AVO. BOULAY JEAN

ORG.  CITY POLICE DEPARTMENT
      NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE                  FED
 01 *733.1(01)A)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
    2 YEARS PROBATION UNSUPERVISED PROBATION NO FEES
 02 *430(01)A) *430(04)B)
    01/10/2015 09:38 PLEAS GUILTY
    01/10/2015 09:38 SENTENCE
    FEES
    SURCHARGE WITH DELAYS  45 DAYS
    PENALTY INFLICTED WITHOUT CUSTODY:  39 DAYS
    DENTENTION CUSTODY GRANTED:   9 DAYS
    PENALTY INFLICTED OF  30 DAYS
```

*"John Doe **pleaded guilty** to **failure to comply with an order** and **mischief to property** on **October 1st, 2015**."*

Court Dockets

```
ACC.  DOE JOHN
      1 DE L'ÉTANG QUEBEC, QUEBEC G1G - 1G1
      NAIS 01/01/1979
      AVO. DOUGH JANE
                              INFRACTION DATE 01/12/2019
                              OPENING DATE    01/01/2020
PLA.  TREMBLAY SARAH
      1130, ROUTE PRINCIPA
      AVO. BOULAY JEAN

ORG.  CITY POLICE DEPARTM
      NO. QUE150807017(1


 2 CHARGES

 CRIMINAL CODE
 01 *733.1(01)A)
    01/10/2015 09:38 PLEA
    01/10/2015 09:38 SENT
    FEES
    SURCHARGE WITH DELAYS
    PENALTY INFLICTED WIT
    DENTENTION CUSTODY GR
    PENALTY INFLICTED OF
    2 YEARS PROBATION UNS
 02 *430(01)A)  430(04)B)
    01/10/2015 09:38 PLEA
    01/10/2015 09:38 SENT
    FEES
    SURCHARGE WITH DELAYS
    PENALTY INFLICTED WIT
    DENTENTION CUSTODY GR
    PENALTY INFLICTED OF
```

Search

Funding | Canada's System of Justice | Laws

solidated Acts ➜ R.S.C., 1985, c. C-46 - Table of Contents ➜ R.S.C., 1985, c. C-46

Criminal Code (R.S.C., 1985, c. C-46)

Full Document: HTML (Accessibility Buttons available) | XML [4844 KB] | PDF [7392 KB]

Act current to 2022-11-16 and last amended on 2022-10-26. Previous Versions

Previous Page | Table of Contents | Next Page

Mischief

430 (1) Every one commits mischief who wilfully

(a) destroys or damages property;

(b) renders property dangerous, useless, inoperative or ineffective;

(c) obstructs, interrupts or interferes with the lawful use, enjoyment or operation of property; or

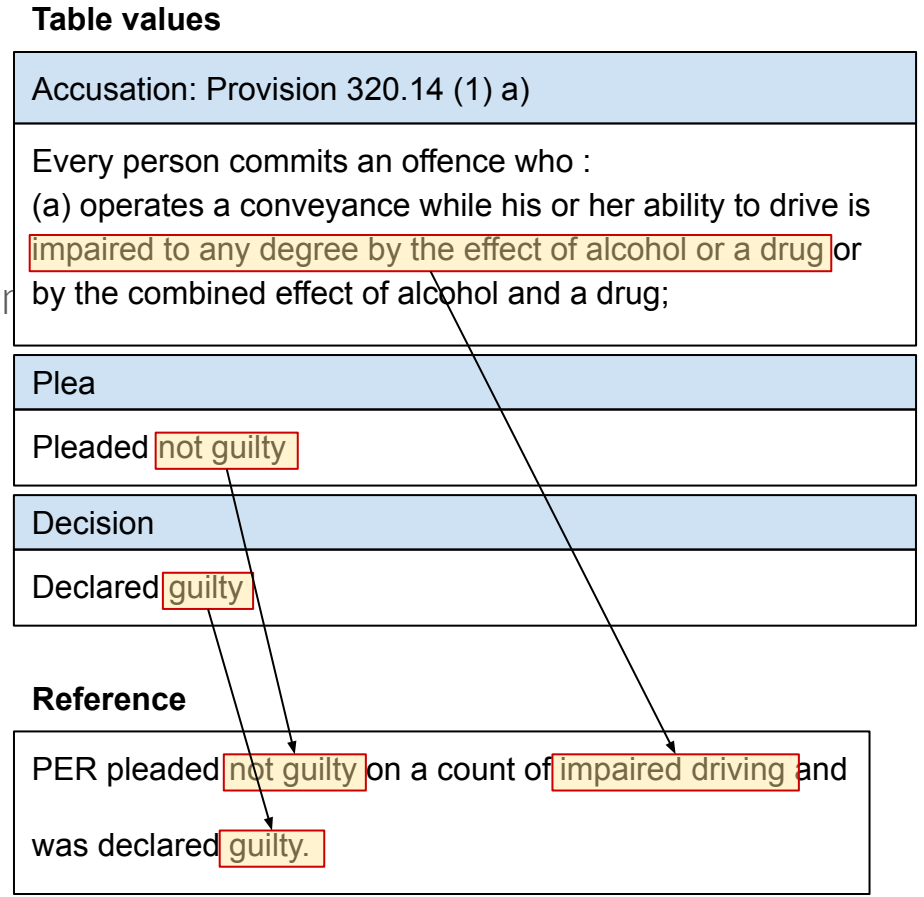
(d) obstructs, interrupts or interferes with any person in the lawful use, enjoyment or operation of property.

2. Dataset

931 training examples, 424 test examples

2. Dataset

931 training examples, 424 test exar



3. Human evaluation

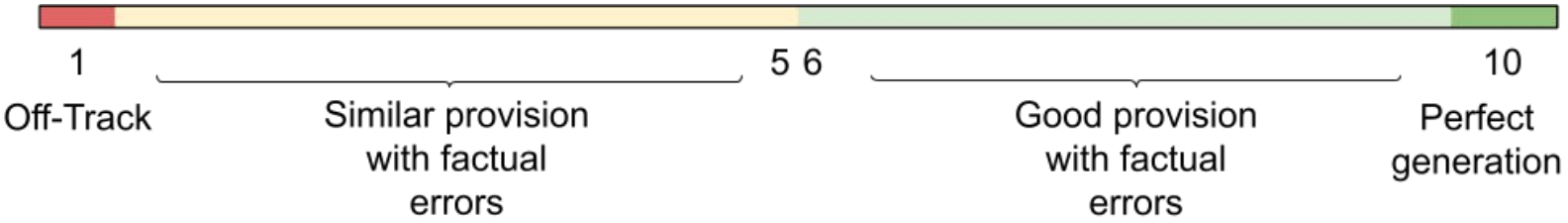
Neural Networks are known to omit and/or hallucinate facts

Evaluating Legal Accuracy of Neural Generators on the Generation of Criminal Court Dockets Description

Nicolas Garneau[†], Eve Gaumond[‡], Luc Lamontagne[†], and Pierre-Luc Déziel[‡]
Université Laval, Québec, Canada
Computer Science Department[†] and Faculty of Law[‡]
{nicolas.garneau,luc.lamontagne}@ift.ulaval.ca
eve.gaumond@observatoire-ia.ulaval.ca
pierre-luc.deziel@fd.ulaval.ca

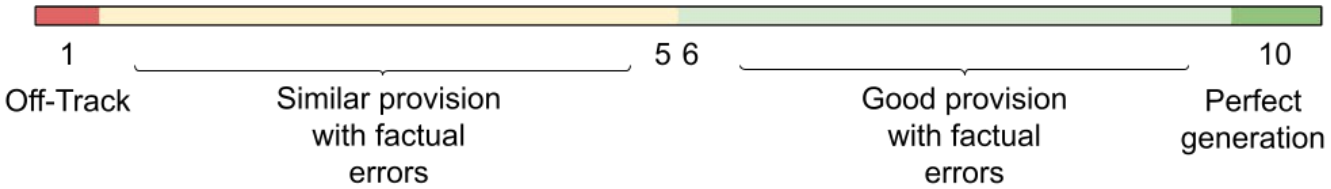
3. Human evaluation

Legal accuracy scale



Legal accuracy scale

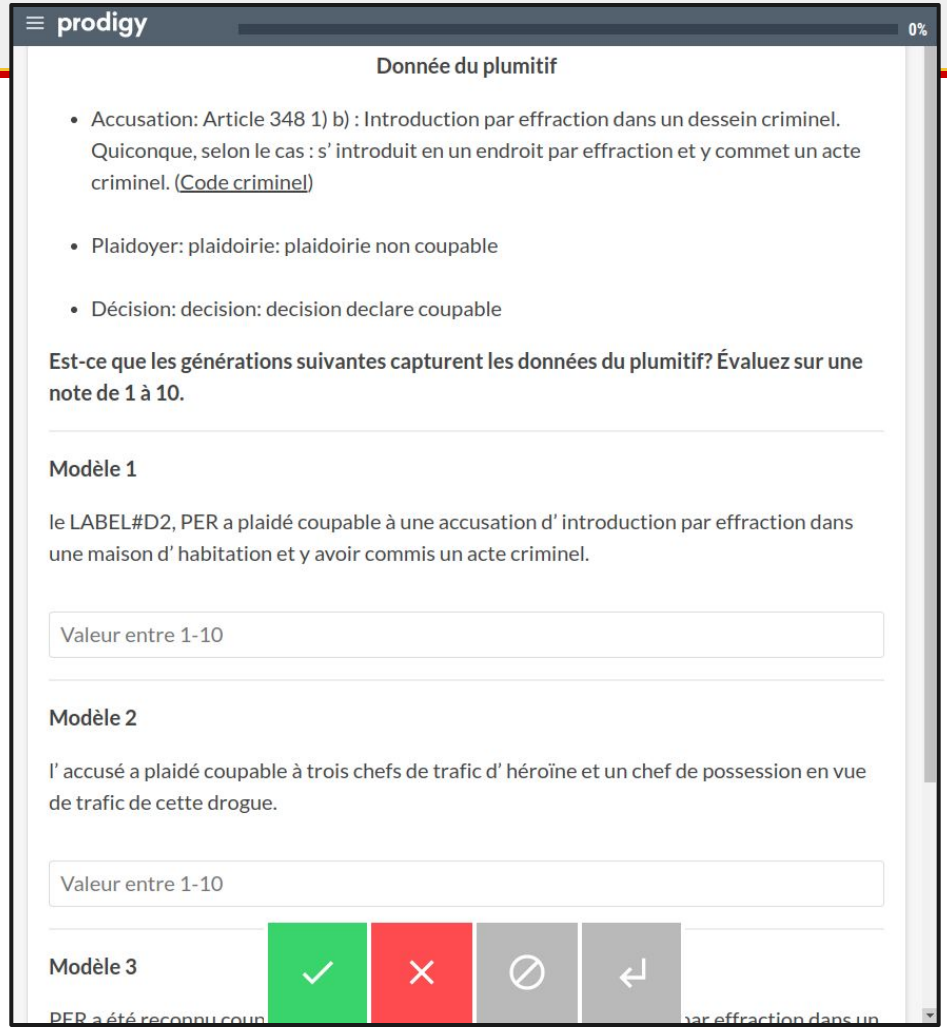
1. **Theme**: some provisions are similar to others (e.g. *Trafficking in substance* ≅ *Possession for purpose of trafficking* ≇ *Driving under the influence*)
    - Position on the scale
2. **Precision (factual errors)**:
    - Hallucination: Anything not supported by the table
    - Omission: Table value not verbalized
        - Points on the scale



3. Human evaluation

We trained 3 models with different priors;

1. LSTM from scratch (*no prior*)
2. BARThez (*language prior*)
3. *CriminelBART* (*language and domain prior*)


prodigy
0%
Donnée du plumitif
• Accusation: Article 348 1) b) : Introduction par effraction dans un dessein criminel. Quiconque, selon le cas : s' introduit en un endroit par effraction et y commet un acte criminel. (Code criminel)
• Plaidoyer: plaidoirie: plaidoirie non coupable
• Décision: decision: decision declare coupable
Est-ce que les générations suivantes capturent les données du plumitif? Évaluez sur une note de 1 à 10.
Modèle 1
le LABEL#D2, PER a plaidé coupable à une accusation d' introduction par effraction dans une maison d' habitation et y avoir commis un acte criminel.
Valeur entre 1-10
Modèle 2
l' accusé a plaidé coupable à trois chefs de trafic d' héroïne et un chef de possession en vue de trafic de cette drogue.
Valeur entre 1-10
Modèle 3

3. Human evaluation

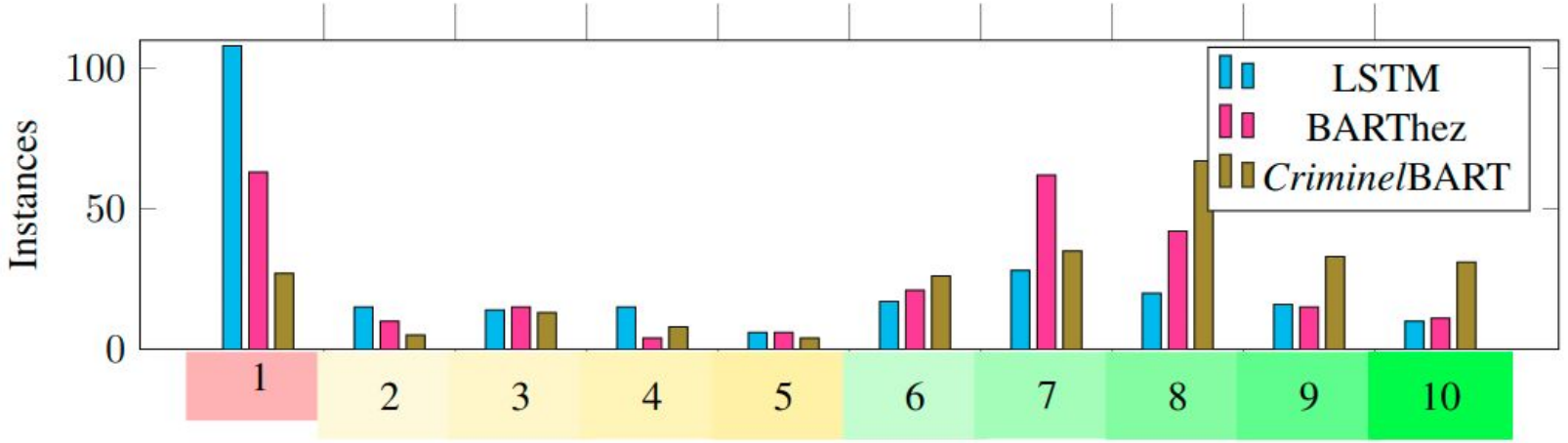


Figure 3: Results of the human evaluation according to the legal accuracy scale. We present the results of the vanilla LSTM (no prior), BARThez (language prior), and *Criminel*BART (language and domain prior).

3. Human evaluation

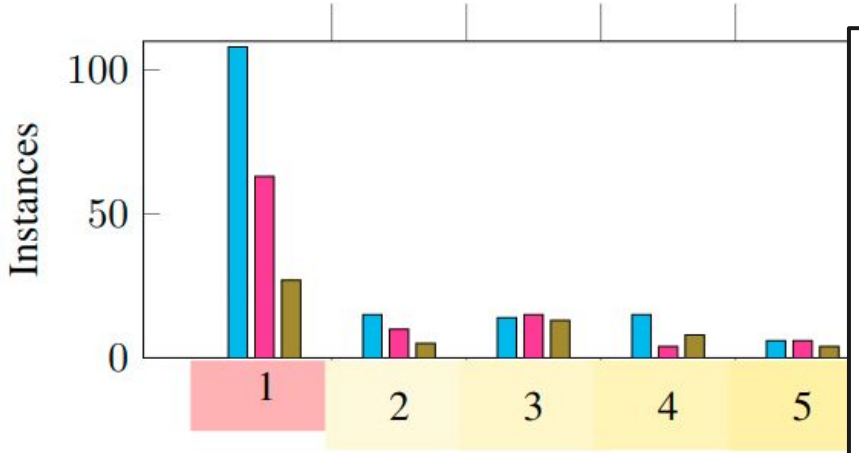


Figure 3: Results of the human evaluation according to the vanilla LSTM (no prior), BARThez (language prior), and *Cr*

	LSTM	BARThez	*Criminel*BART
Ann. 1	4.4±2.8	5.2±2.9	6.3±2.6
Ann. 2	3.7±3.2	5.2±3.0	6.8±2.8
Ann. 3	3.6±3.3	5.4±3.2	7.0±2.8
Avg.	3.9±2.9	5.3±2.9	6.7±2.6
ρ	0.76	0.85	0.84

Table 1: Average score and standard deviation per annotator and the overall score for each model. We also provide the annotator agreement ρ per model. The overall agreement is 0.84.

3. Human evaluation

Poor generalization to unseen provisions

Provision	LSTM	BARThez	*Criminel*BART
445.1 (1) a)	1.0	1.0	1.0
150	2.3	5.0	4.6
83.181	1.0	1.0	1.0
241	1.0	2.7	2.0
467.12	1.0	1.0	8.7
810.2	1.0	1.0	1.0
172	1.0	1.0	1.33
320.14	1.0	6.3	7.3

Table 3: Analysis of the generalization capabilities of the models on unseen provisions. We provide details on the provisions in Appendix D.

4. Guided Beam Search

4. Guided Beam Search

Input Data	**Provision 341**: Attack on premises, residence or transport of internationally protected person; Every one who commits a violent attack on the official premises, private accommodation or means of transport of an internationally protected person that is likely to endanger the life or liberty of such a person is guilty of an indictable offence and liable to imprisonment for a term of not more than fourteen years.
***Criminel*BART** 1.0	The or on the LABELD, at LOC, LOC, exercised assault and battery against PER, thereby committing the criminal act under Section 266(a) of the Criminal Code.

4. Guided Beam Search

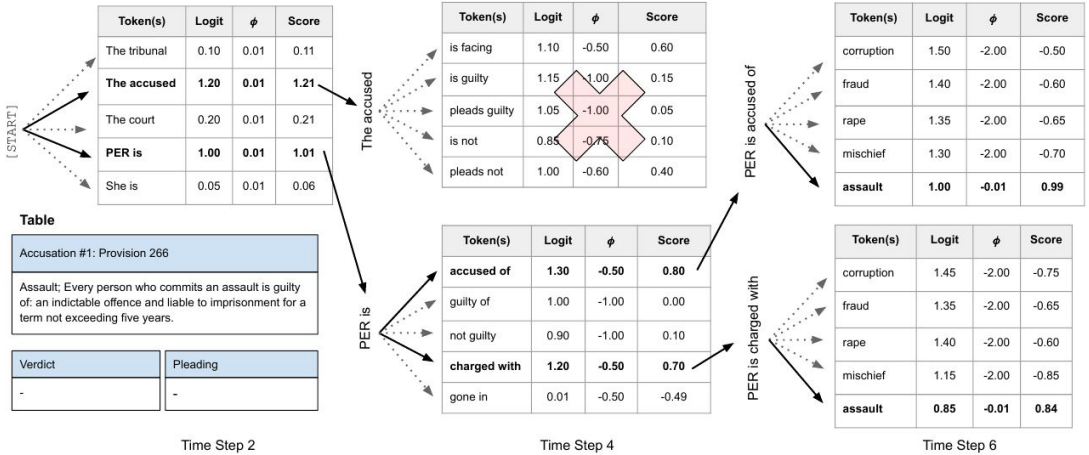


Figure 7.6 – A weighted beam search example with a beam size of 2 where, for example, "corruption" would be a more common offense according to the training set distribution, but the ϕ term puts more weight to the "assault" offense since the generated sequence does not contain omissions nor hallucinations. Note that the *phi* term will grow overtime because of the tangent hyperbolic function.

4. Guided Beam Search

Based on the constrained beam search of *Hafez* (Ghazvininejad et al., 2017)

$$score(b_i, w) = score(b_{i-1}) + \log Gen(w)$$

4. Guided Beam Search

Based on the constrained beam search of *Hafez* (Ghazvininejad et al., 2017)

$$score(b_i, w) = score(b_{i-1}) + \log Gen(w) + \sum_j \alpha_j * f_j(w); \forall w \in V_{suc}$$

4. Guided Beam Search

Based on the constrained beam search of *Hafez* (Ghazvininejad et al., 2017)

$$score(b_i, w) = score(b_{i-1}) + \log Gen(w) + (\omega \cdot (v_i - o_i) - \gamma \cdot h_i)$$

4. Guided Beam Search

Based on the constrained beam search of *Hafez* (Ghazvininejad et al., 2017)

$$score(b_i, w) = score(b_{i-1}) + \log Gen(w) + (\omega \cdot \underbrace{(v_i - o_i)}_{\text{omission reward}} - \gamma \cdot h_i)$$

4. Guided Beam Search

Based on the constrained beam search of *Hafez* (Ghazvininejad et al., 2017)

$$score(b_i, w) = score(b_{i-1}) + \log Gen(w) + (\omega \cdot \underbrace{(v_i - o_i)}_{\text{omission reward}} - \underbrace{\gamma \cdot h_i}_{\text{hallucination penalty}})$$

4. Guided Beam Search

We need two models, one that predicts the number of omissions, another that predicts the number of hallucinations

$$o_i = m_o(V_i, s_i)$$
$$h_i = m_h(V_i, s_i)$$

4. Guided Beam Search

To this end, we create a new training dataset from the original training data

4. Guided Beam Search

To this end, we create a new training dataset from the original training data

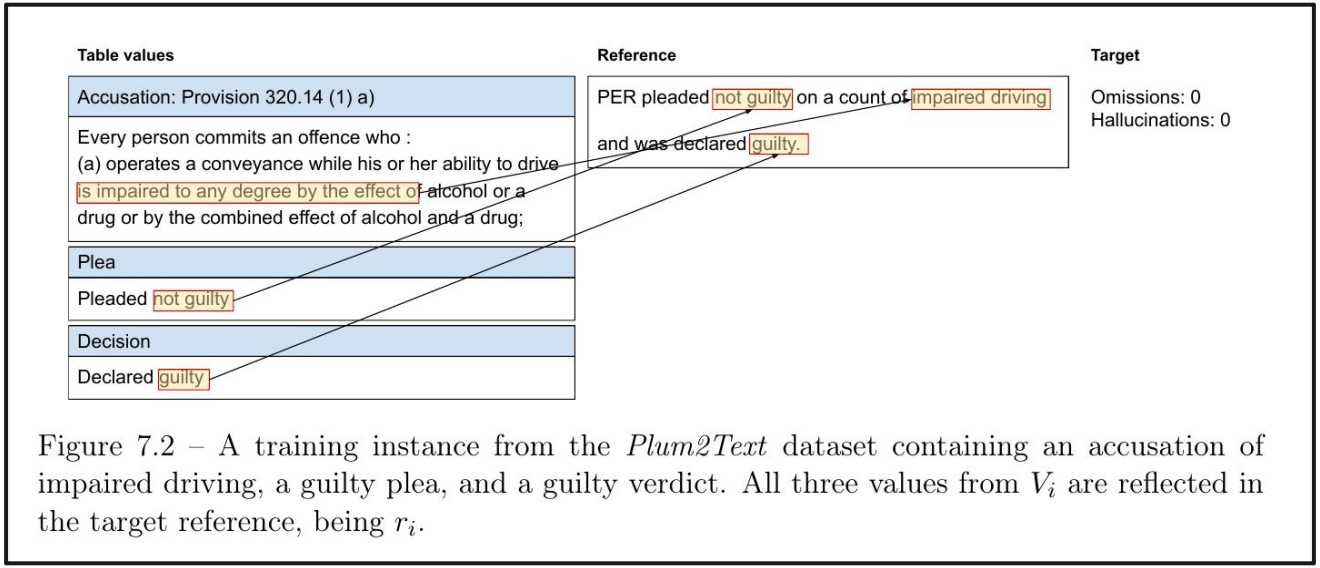
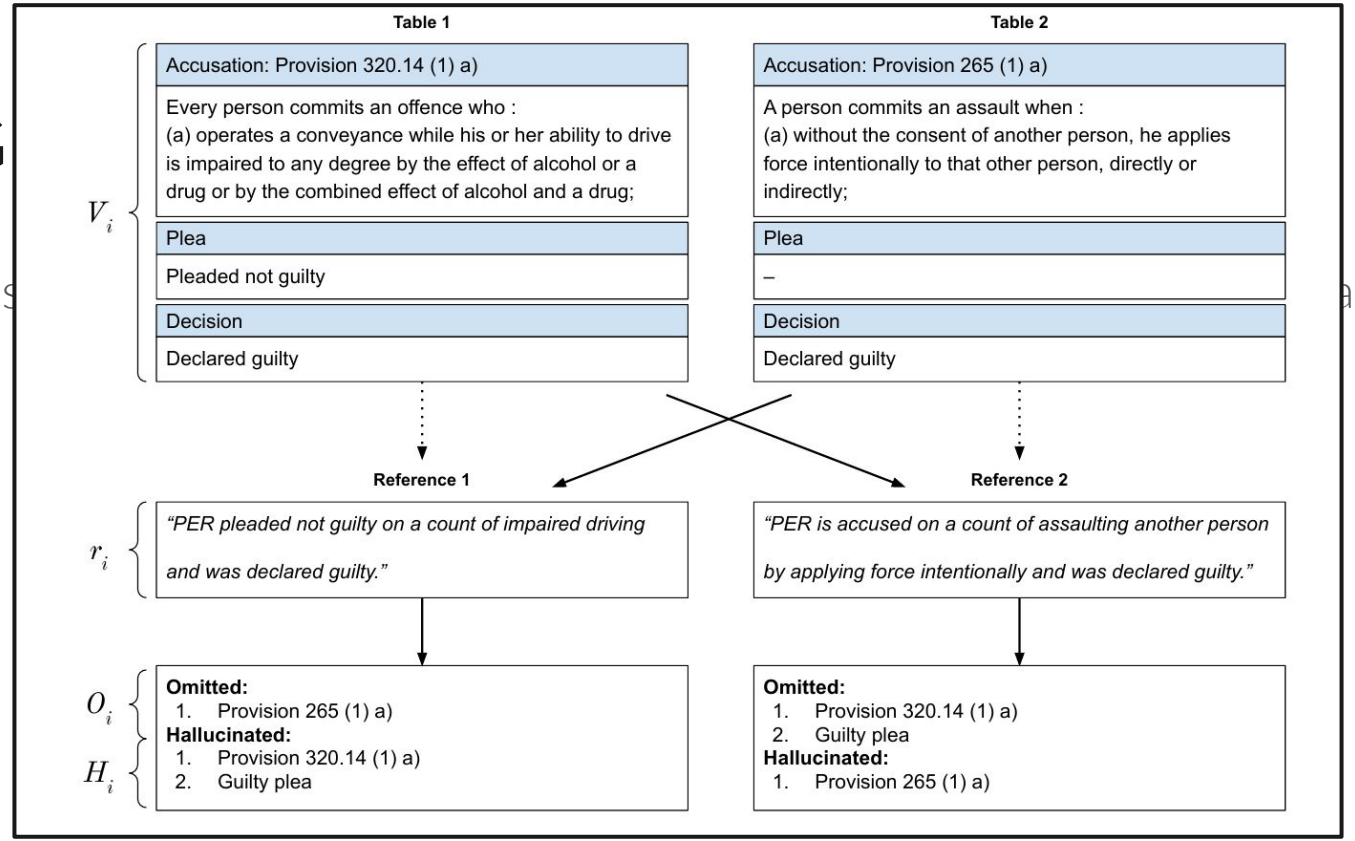


Figure 7.2 – A training instance from the *Plum2Text* dataset containing an accusation of impaired driving, a guilty plea, and a guilty verdict. All three values from V_i are reflected in the target reference, being r_i.

4. G

To this ata



4. Guided Beam Search

To this end, we create a new training dataset from the original training data

$$o_i = m_o(V_i, s_i)$$
$$h_i = m_h(V_i, s_i)$$

4. Guided Beam Search

To this end, we create a new training dataset from the original training data

Model	MSE	RMSE	MAE	$\mathcal{R}^2$	Accuracy
Omission	0.05	0.23	0.10	0.99	0.96
Hallucination	0.05	0.22	0.08	0.99	0.97

Table 7.1 – Performance of both the omission and hallucination models on *Plum2Text* w.r.t the mean squared error (MSE), the root mean squared error (RMSE), the mean average error (MAE), $\mathcal{R}^2$, and the accuracy.

4. Guided Beam Search

To this end, we create a new training dataset from the original training data

$$o_i = m_o(V_i, s_i)$$
$$h_i = m_h(V_i, s_i)$$

$$score(b_i, w) = score(b_{i-1}) + \log Gen(w) + (\omega \cdot (v_i - o_i) - \gamma \cdot h_i)$$

4. Guided Beam Search

Automatic Evaluation

				BLEU									Rates	
ω	γ	β	τ	1	2	3	4	ROUGE	METEOR	BScore	NLI	RANK	Hal.	Om.
0.0	0.0	5	-	0.73	0.58	0.47	0.41	0.42	0.38	0.78	0.34	0.72	0.28	0.24
0.2	0.2	15	-	0.73	**0.59**	**0.48**	**0.43**	**0.44**	0.38	**0.79**	0.34	**0.76**	0.13	0.11
Post processing				0.73	**0.58**	**0.48**	**0.42**	**0.43**	0.37	**0.79**	0.34	**0.78**	0.11	0.11

Table 7.3 – Automatic evaluation results of the best performing original *CriminelBART* ($\omega = 0.0$, $\gamma = 0.0$, $\beta = 5$, without temporal weighting), the best performing model using the weighted beam search algorithm ($\omega = 0.2$, $\gamma = 0.2$, $\beta = 15$, without temporal weighting), and that same model using the post processing finalization step.

4. Guided Beam Search

Human Evaluation

1. We gathered 45 unseen provisions from the Criminal Code
2. Asked 3 students from the faculty of law to evaluate the generation w.r.t. to the evaluation guidelines using the legal accuracy scale

4. Guided Beam Search

Human Evaluation

1. We gathered 45 unseen provisions from
2. Asked 3 students from the faculty of law to the evaluation guidelines using the leg

Provision	*CriminelBART*	Guided *CriminelBART*
46	1.00	8.00
57	3.00	8.00
58	2.33	7.00
83.04	2.67	8.00
83.08	3.00	8.00
83.21	5.33	8.00
83.181	1.00	8.00
123	1.00	8.00
148	7.67	8.67
150	3.67	8.33
170	2.33	5.00
173	2.33	8.33
202	1.00	4.67
218	1.00	5.67
243	4.33	6.67
245	2.00	7.33
253	6.00	8.00
267	6.33	8.00
270.1	3.33	8.67
318	7.00	8.33
342	8.67	9.00
342.1	2.33	9.67
344	4.00	8.67
345	7.67	1.00
347	1.00	6.00
351	7.00	9.00
354	3.00	8.00
355	5.00	7.67
356	1.00	7.67
364	1.00	8.67
368	7.33	9.00
374	4.67	5.00
382.1	8.33	4.00
398	8.00	6.00
402.2	8.00	8.33
406	3.33	8.00
431	1.00	8.33
432	5.00	4.33
437	1.00	4.33
438	5.67	8.33
439	2.33	8.33
445.1	3.00	9.00
446	2.33	8.67
467.111	8.33	8.67
810.2	2.33	5.67
Average	**3.9**	**7.4**

Table 7.4 – Human evaluation of the original version of CriminelBART and the one using guided beam search on the 45 unseen provisions.

4. Guided Beam Search

Human

1. W
2. As w.r.t. to

Input Data	**Provision 341**: Attack on premises, residence or transport of internationally protected person; Every one who commits a violent attack on the official premises, private accommodation or means of transport of an internationally protected person that is likely to endanger the life or liberty of such a person is guilty of an indictable offence and liable to imprisonment for a term of not more than fourteen years.
***Criminel*BART** 1.0	The or on the LABELD, at LOC, LOC, exercised assault and battery against PER, thereby committing the criminal act under Section 266(a) of the Criminal Code.

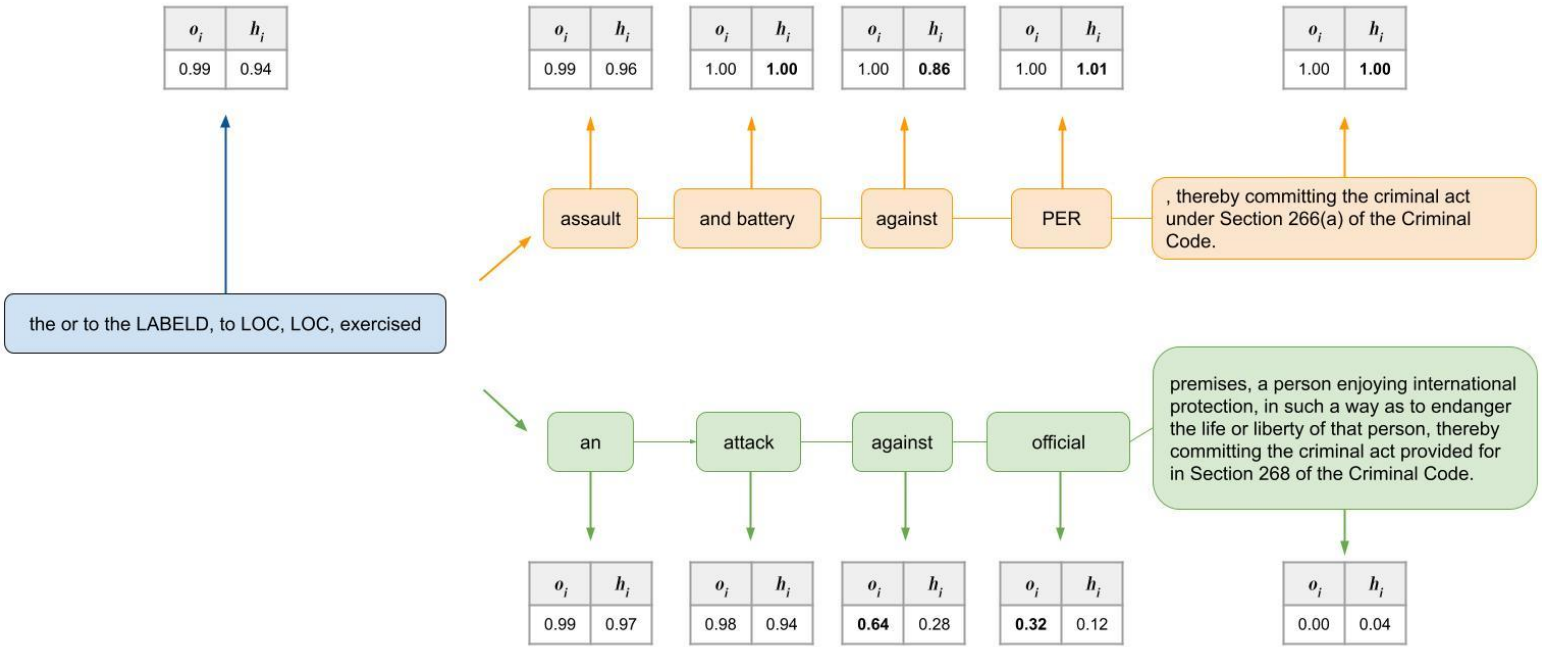
4. Guided Beam Search

Human

1. W
2. As w.r.t. to

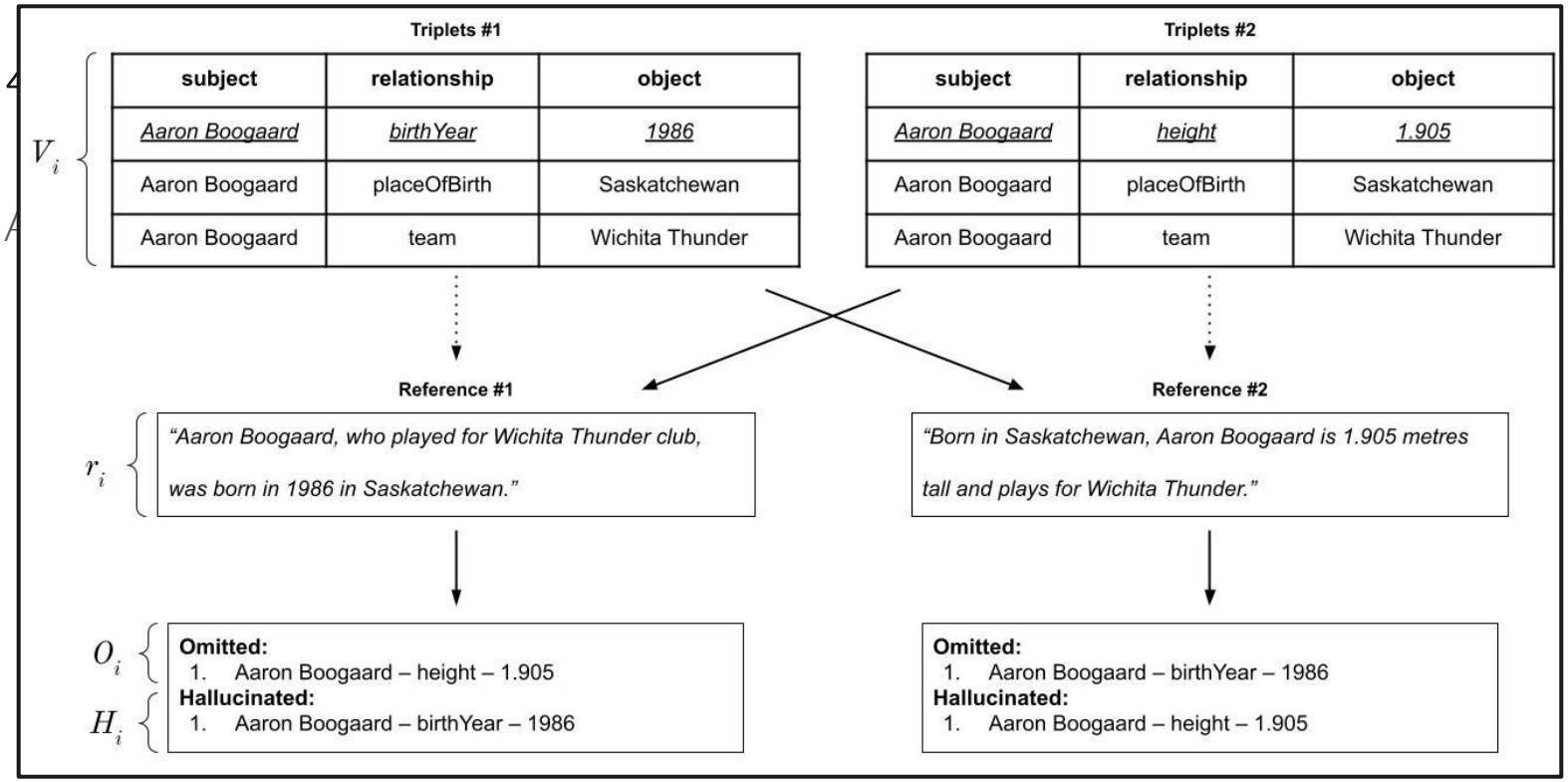
Input Data	**Provision 341**: Attack on premises, residence or transport of internationally protected person; Every one who commits a violent attack on the official premises, private accommodation or means of transport of an internationally protected person that is likely to endanger the life or liberty of such a person is guilty of an indictable offence and liable to imprisonment for a term of not more than fourteen years.
***Criminel*BART** 1.0	The or on the LABELD, at LOC, LOC, exercised assault and battery against PER, thereby committing the criminal act under Section 266(a) of the Criminal Code.
Guided *Criminel*BART 8.33	The or on the LABELD, at LOC, LOC, exercised an attack against official premises, a person enjoying international protection, in such a way as to endanger the life or liberty of that person, thereby committing the criminal act provided for in Section 268 of the Criminal Code.

4. Guided Beam Search



4. Guided Beam Search

Applied on WebNLG


Triplets #1
subject
relationship
object
Aaron Boogaard
birthYear
1986
Aaron Boogaard
placeOfBirth
Saskatchewan
Aaron Boogaard
team
Wichita Thunder
Triplets #2
subject
relationship
object
Aaron Boogaard
height
1.905
Aaron Boogaard
placeOfBirth
Saskatchewan
Aaron Boogaard
team
Wichita Thunder
V_i
Reference #1
"Aaron Boogaard, who played for Wichita Thunder club, was born in 1986 in Saskatchewan."
Reference #2
"Born in Saskatchewan, Aaron Boogaard is 1.905 metres tall and plays for Wichita Thunder."
r_i
O_i
Omitted:
1. Aaron Boogaard – height – 1.905
H_i
Hallucinated:
1. Aaron Boogaard – birthYear – 1986
Omitted:
1. Aaron Boogaard – birthYear – 1986
Hallucinated:
1. Aaron Boogaard – height – 1.905

4. Guided Beam Search

Applied on WebNLG

				BLEU									Rates	
ω	γ	β	τ	1	2	3	4	ROUGE	METEOR	BScore	NLI	RANK	Hal.	Om.
0.0	0.0	5	-	0.81	0.71	0.64	0.58	0.55	0.54	0.94	0.63	0.64	0.11	0.00
0.2	0.2	15	-	**0.83**	**0.73**	**0.65**	**0.59**	0.53	0.54	0.94	**0.68**	**0.65**	0.10	0.00
Post processing				**0.84**	**0.74**	**0.66**	**0.60**	0.54	0.54	0.94	**0.68**	**0.66**	0.10	0.00

Table 7.7 – Automatic evaluation results of the best performing model on WebNLG ($\omega = 0.0$, $\gamma = 0.0$, $\beta = 5$, without temporal weighting) and the best performing model using the weighted beam search algorithm ($\omega = 0.2$, $\gamma = 0.5$, $\beta = 10$, without temporal weighting).

Conclusion

1. Guided beam search algorithm enables a better exploration of the generation tree
2. By predicting the number of omissions/hallucination, offers a level of interpretability