Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

What is Table Understanding?

Name	Country	Born	Died	Status	Masters T.	PGA
Willie Park, Jr.	Scotland	1864	1925	Prof.	NYF	256
Harry Vardon	Wales	1871	1932	Prof.	BAT	251
Thomas Renouf	Ireland	1859	1916	Prof.	NFT	189
J.H. Taylor	England	1898	1923	Prof.	ONN	172
Harold Hilton	England	1867	1925	Prof.	CF.BU	162
David Kinnell	Scotland	1851	1932	Amat.	NBNC	161
James Kinnell	Scotland	1892	1916	Prof.	NYF	159
Freddie Tait	Wales	1843	1923	Prof.	ONN	157
Sandy Herd	Scotland	1863	1925	Prof.	NFT	156
David Herd	Scotland	1861	1932	Amat.	NYF	155

Name	Country	Born	Died	Status	Masters T.	PGA
Willie Park, Jr.	Scotland	1864	1925	Prof.	NYF	256
Harry Vardon	Wales	1871	1932	Prof.	BAT	251
Thomas Renouf	Ireland	1859	1916	Prof.	NFT	189
J.H. Taylor	England	1898	1923	Prof.	ONN	172
Harold Hilton	England	1867	1925	Prof.	CF.BU	162
David Kinnell	Scotland	1851	1932	Amat.	NBNC	161
James Kinnell	Scotland	1892	1916	Prof.	NYF	159
Freddie Tait	Wales	1843	1923	Prof.	ONN	157
Sandy Herd	Scotland	1863	1925	Prof.	NFT	156
David Herd	Scotland	1861	1932	Amat.	NYF	155

Regular Table

Club	Season	League			National Cup		Continental		Other		Total	
		Division	Apps	Goals	Apps	Goals	Apps	Goals	Apps	Goals	Apps	Goals
	2011-12		0 / 0	0	0	0	0	0			0	0
	2012-13		0 / 1	0	0 / 1	0			0 / 1	0		
	Total		0	0	1	0	0	0	0	0	1	0
	2012-13		1 / 6	0	1 / 6	0			1 / 6	w: 0		
	2013-14		13 / 15	1	13 / 15	0			13 / 15	0		
	Total		19	1	2	0	0	0	0	0	21	1
	2014-15		11 / 7	1	11 / 7	0			11 / 7	1		
	2016-17		36 / 4	3	36 / 4	0			36 / 4	l: 3		
	2017-18		24 / 31	3	24 / 31	0			24 / 31	3		
	2018-19		4 / 72	0	4 / 72	0			4 / 72	0		
	Total		64	6	5	0					69	6
Career total			83	7	8	2	0	0	0	0	91	7

Irregular Table

What is Table-to-Text Generation?

Table-to-Text Generation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

→ In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Table-to-Text Generation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

→ In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Table-to-Text Generation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

→

Hilton played for England

Renouf and Taylor scored 156

Park scored 151 points

In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Willie Park played for Scotland

There were three ties in the Championship

David Herd finished last

Table-to-Text Generation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

→

In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Table-to-Text Generation

Content Selection

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

Why do we need to improve fidelity?

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

False!

David Kinnell scored 240 points.

Why do we need to improve Table Representation?

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
T7	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
T7	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

→

Simple linearization

```
Place, Player, Country, Score, 1, Willie Park, Jr., Scotland, 151, 2, Harry vardon, Jersey, 154, T3, Thomas Renouf, Jersey, 156, T3, J.H. Taylor, England, 156, T5, Harold Hilton, England, 157, T5, David Kinnell, Scotland, 157, T7, James Kinnell, Scotland, 158, T7, Freddie Tait, Scotland, 158, 9, Sandy Herd, Scotland, 159, 10, David Herd, Scotland, 160
```

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
T7	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

→

```
Place, Player, Country, Score, 1, Willie
Park, Jr., Scotland, 151, 2, Harry vardon,
Jersey, 154, T3, Thomas Renouf, Jersey,
156, T3, J.H. Taylor, England, 156, T5,
Harold Hilton, England, 157, T5, David
Kinnell, Scotland, 157, T7, James Kinnell,
Scotland, 158, T7, Freddie Tait, Scotland,
158, 9, Sandy Herd, Scotland, 159, 10,
David Herd, Scotland, 160
```

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
T7	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

→

```
Place, Player, Country, Score, 1, Willie
Park, Jr., Scotland, 151, 2, Harry vardon,
Jersey, 154, T3, Thomas Renouf, Jersey,
156, T3, J.H. Taylor, England, 156, T5,
Harold Hilton, England, 157, T5, David
Kinnell, Scotland, 157, T7, James Kinnell,
Scotland, 158, T7, Freddie Tait, Scotland,
158, 9, Sandy Herd, Scotland, 159, 10,
David Herd, Scotland, 160
```

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
T7	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

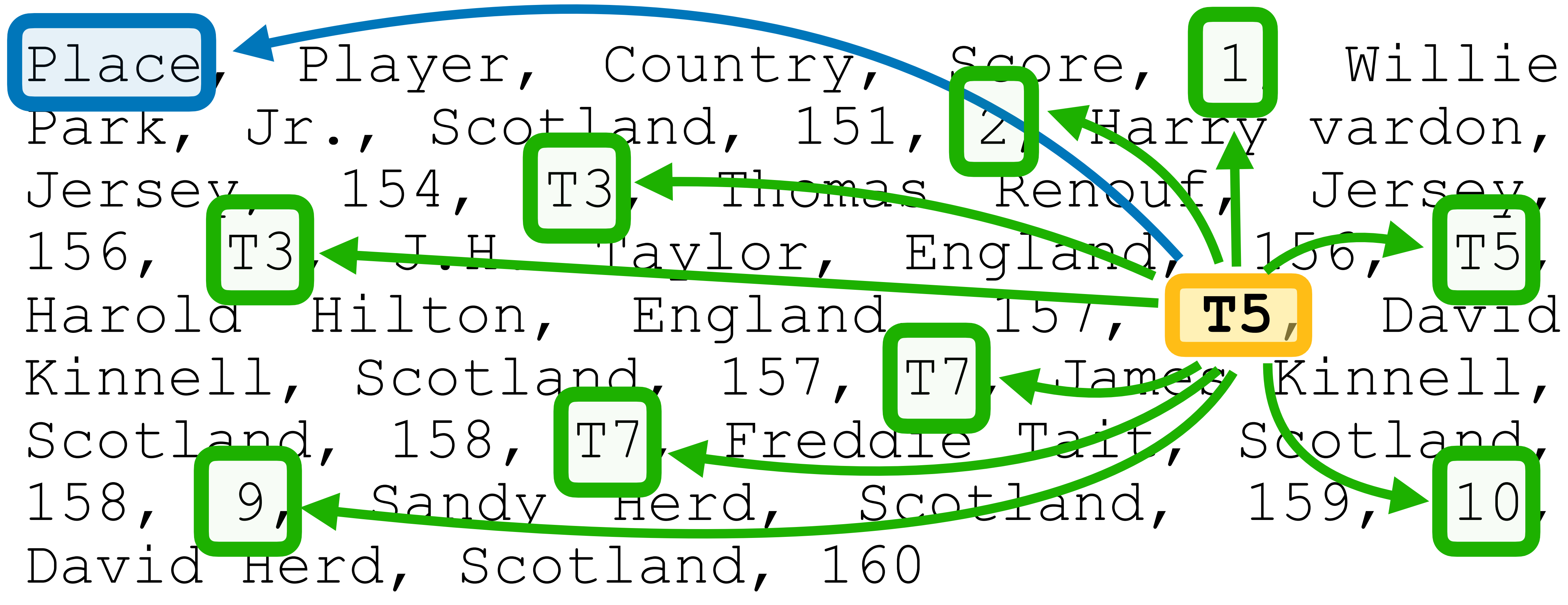


Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
T7	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

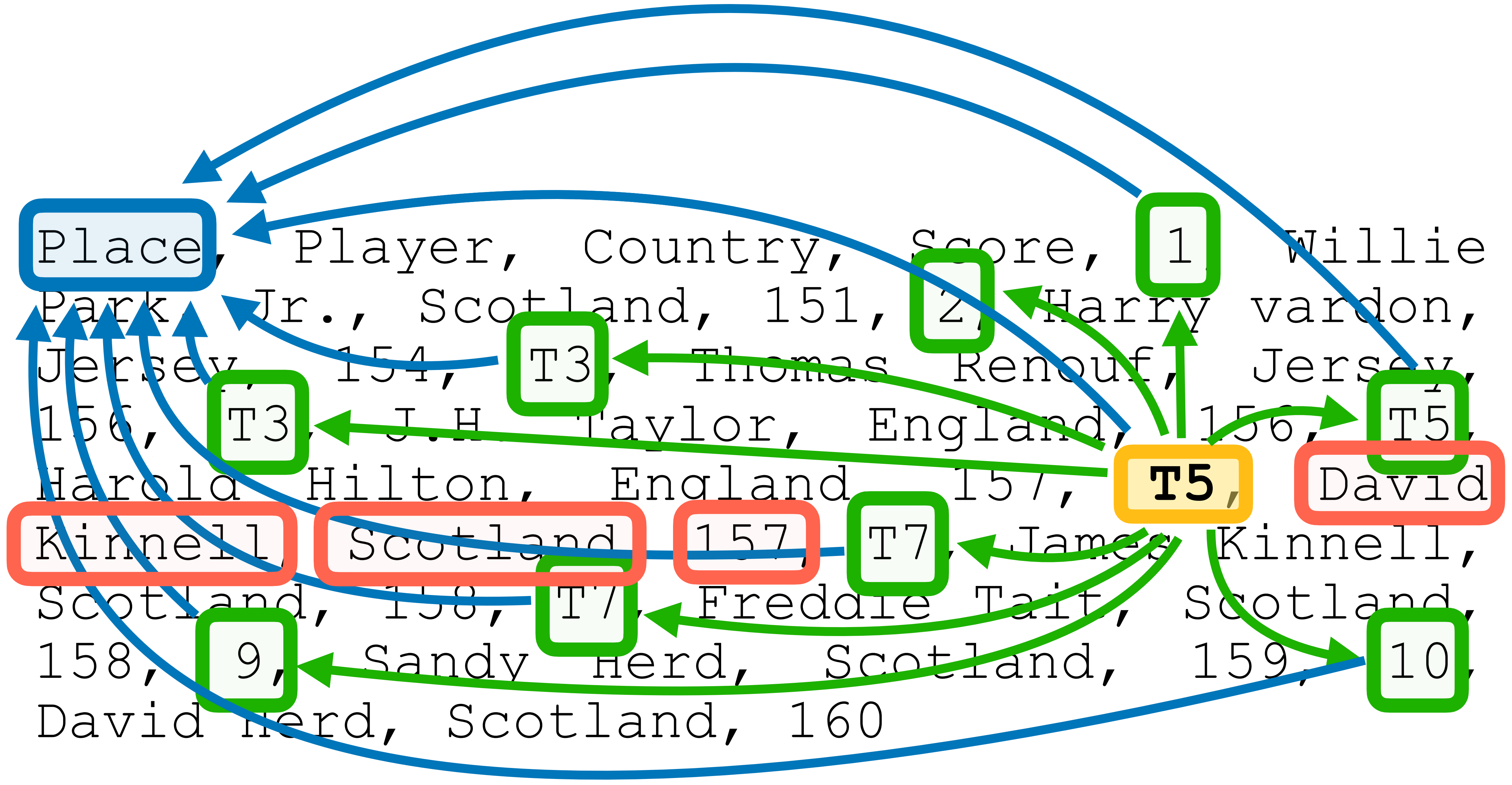


Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
T7	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

Verbose linearization

<page_title> 1898 Open Championship <page_title> <table> <row> <cell> Place </cell> <cell> Player <row_header> Place </row_header> </cell> <cell> Country <row_header> Place </row_header> <row_header> Player </row_header> </cell> <cell> Score <row_header> Place </row_header> <row_header> Player </row_header> <row_header> Country </row_header> </cell> </row> <row> <cell> 1 <col_header> Place </col_header> </cell> <cell> Willie Park, Jr. <col_header> Player </col_header> </cell> <cell> Scotland <col_header> Country </col_header> </cell> <cell> 151 <col_header> Score </col_header> </cell> <row> <cell> 2 <col_header> Place </col_header> </cell> <cell> Harry Vardon <col_header> Player </col_header> </cell> <cell> Wales <col_header> Country </col_header> </cell> <cell> 154 <col_header> Score </col_header> </cell>

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu

Universidad del País Vasco | Euskal Herriko Unibertsitatea

HiTZ

Hizkuntza Teknologiako Zentroa

Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Fidelity

Automatic Logical Forms improve fidelity in Table-to-Text generation

Fidelity

Automatic Logical Forms improve fidelity in Table-to-Text generation

Representation

Pixel-based Table-To-Text Generation

Fidelity

Automatic Logical Forms improve fidelity in Table-to-Text generation

Representation

Pixel-based Table-To-Text Generation

Beyond Table-to-Text

Lossless Table Visualisations Enhance Multimodal Table Understanding

Fidelity

Automatic Logical Forms improve fidelity in Table-to-Text generation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157

Content Selection

Willie Park, Jr.

Harold Hilton

151 157

In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157

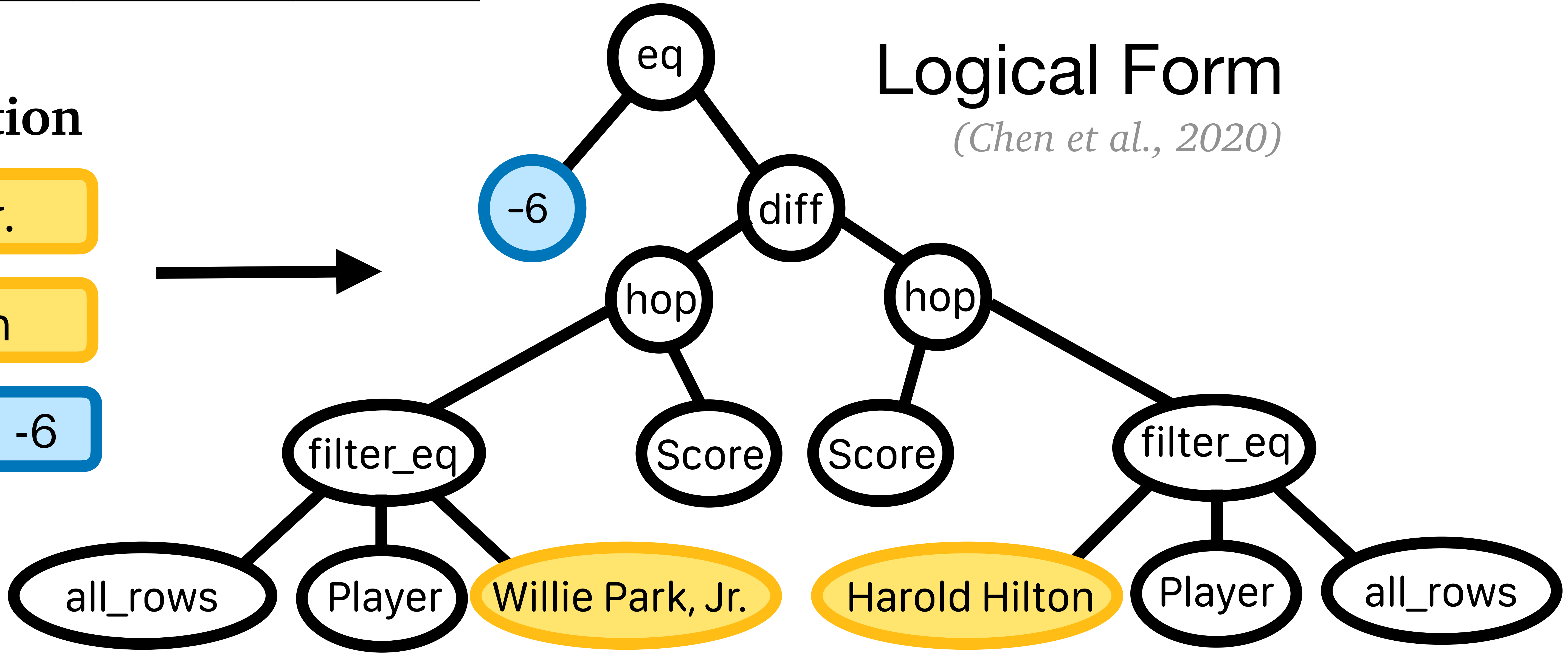
In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Content Selection

Willie Park, Jr.

Harold Hilton

151 – 157 = -6



Logical Form

(Chen et al., 2020)

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157

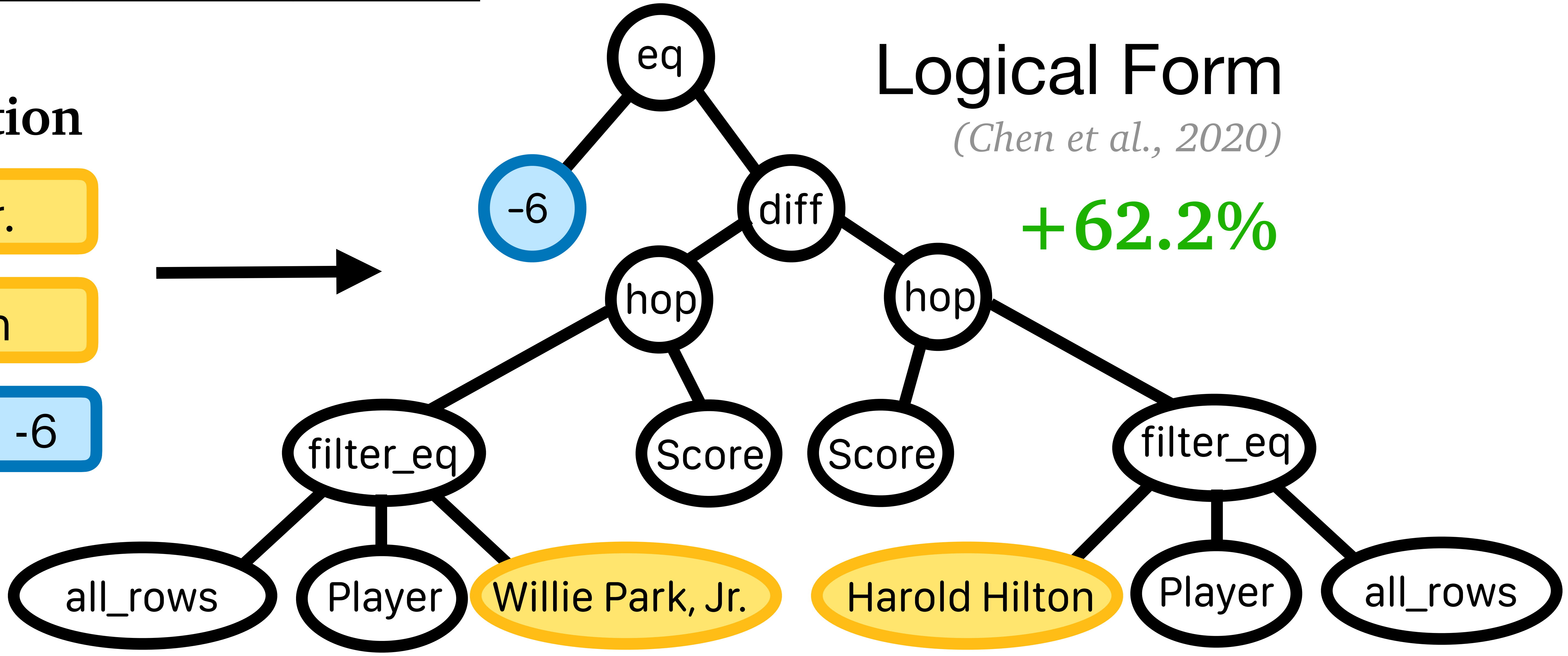
In the 1898 Open Championship, **Willie Park, Jr.** scored **six points less** than **Harold Hilton**.

Content Selection

Willie Park, Jr.

Harold Hilton

151 – 157 = -6



Logical Form

(Chen et al., 2020)

+62.2%

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157

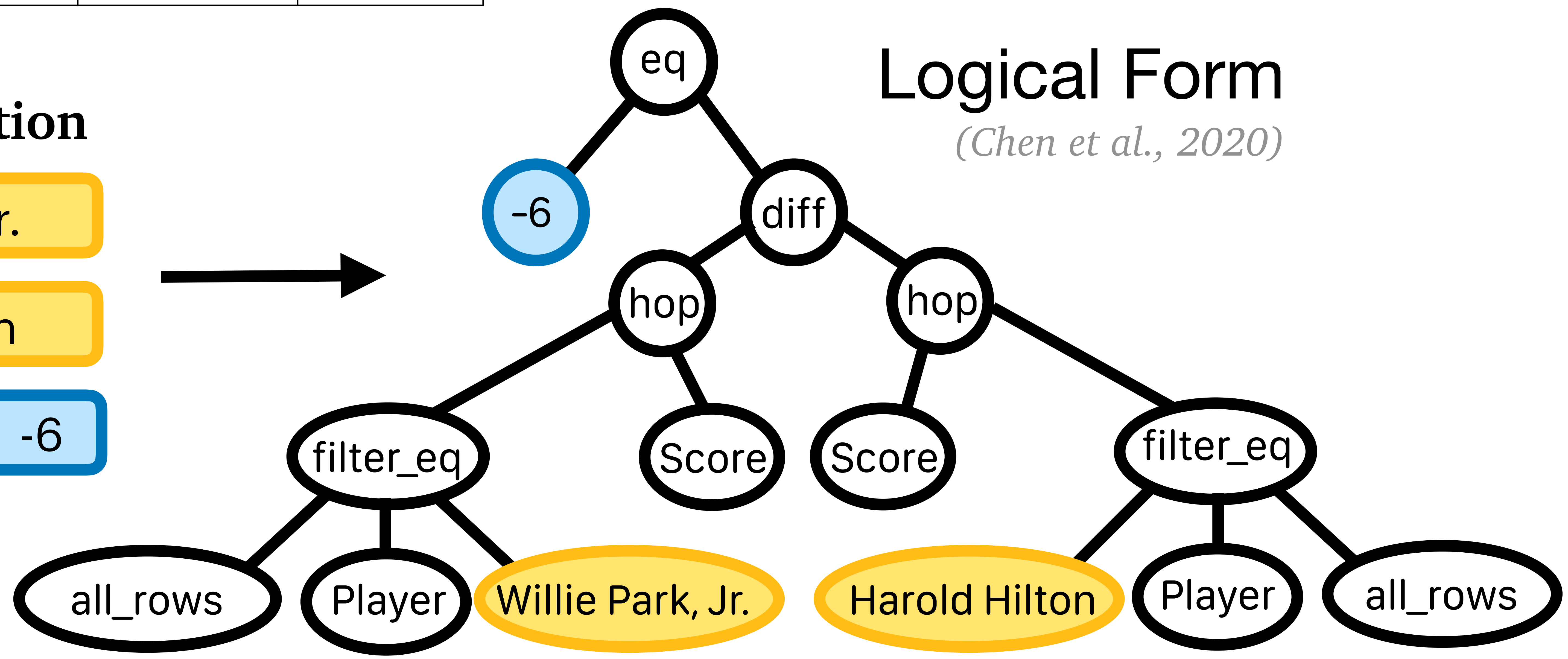
In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Content Selection

Willie Park, Jr.

Harold Hilton

151 − 157 = -6



Logical Form

(Chen et al., 2020)

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157

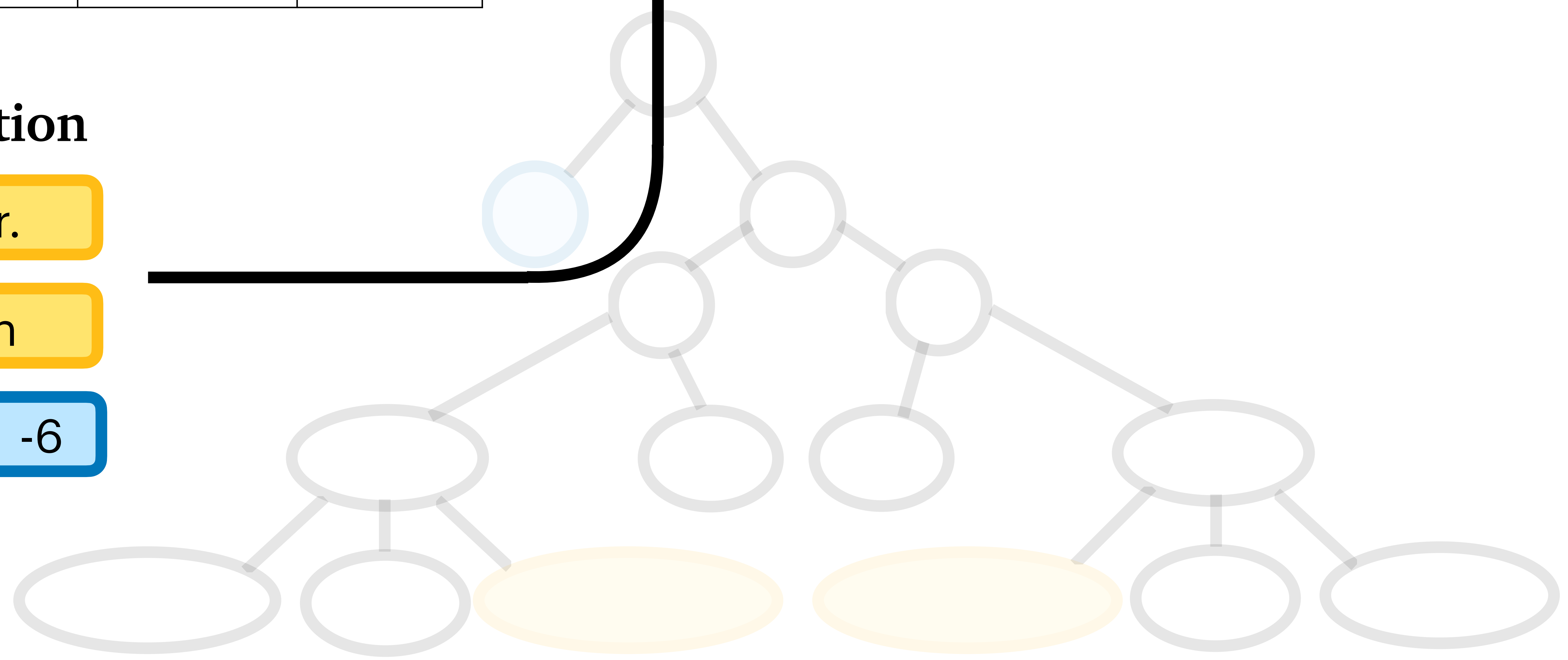
In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.

Content Selection

Willie Park, Jr.

Harold Hilton

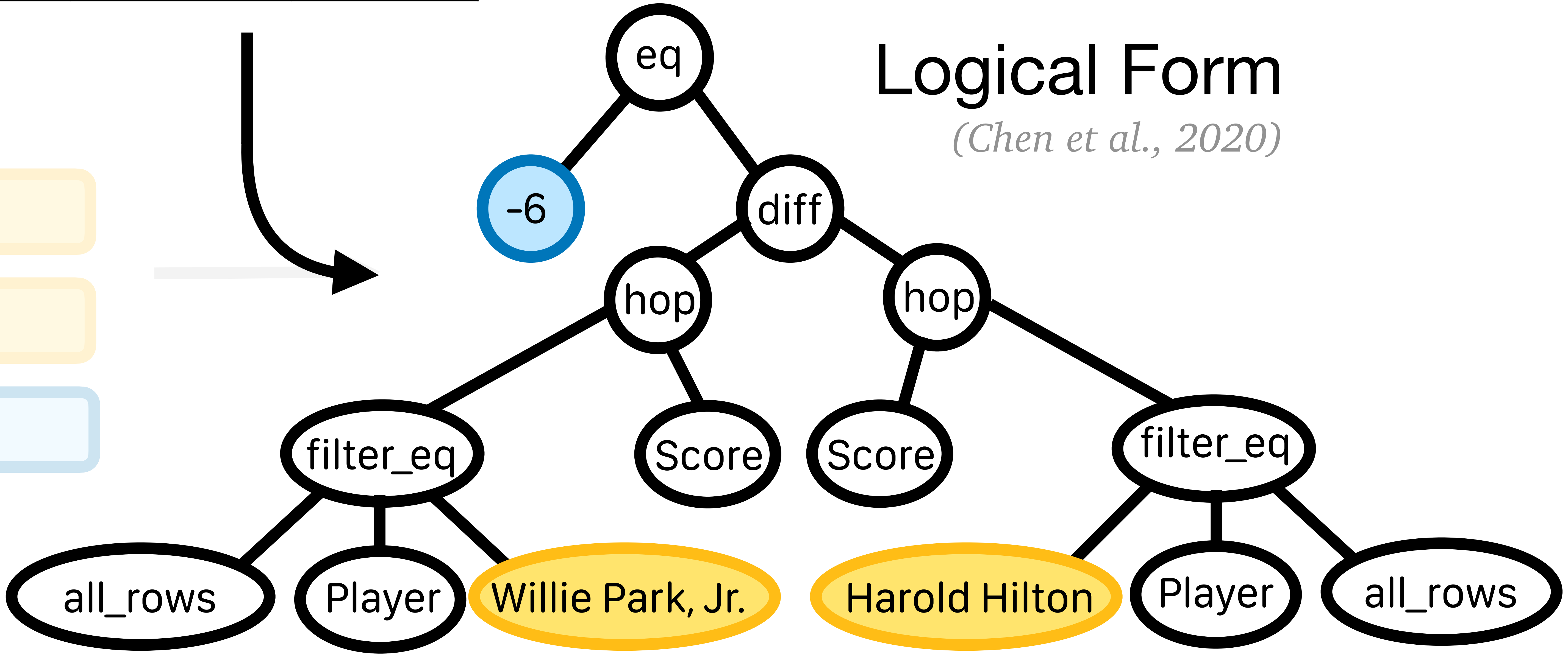
151 – 157 = -6



Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
T3	J.H. Taylor	England	156
T5	Harold Hilton	England	157
T5	David Kinnell	Scotland	157

In the 1898 Open Championship, Willie Park, Jr. scored six points less than Harold Hilton.



Logical Form
(Chen et al., 2020)

Do Logical Forms improve fidelity more than Content Selection values?

Do Logical Forms improve fidelity more than Content Selection values?

Can models automatically generate correct Logical Forms based on Content Selection?

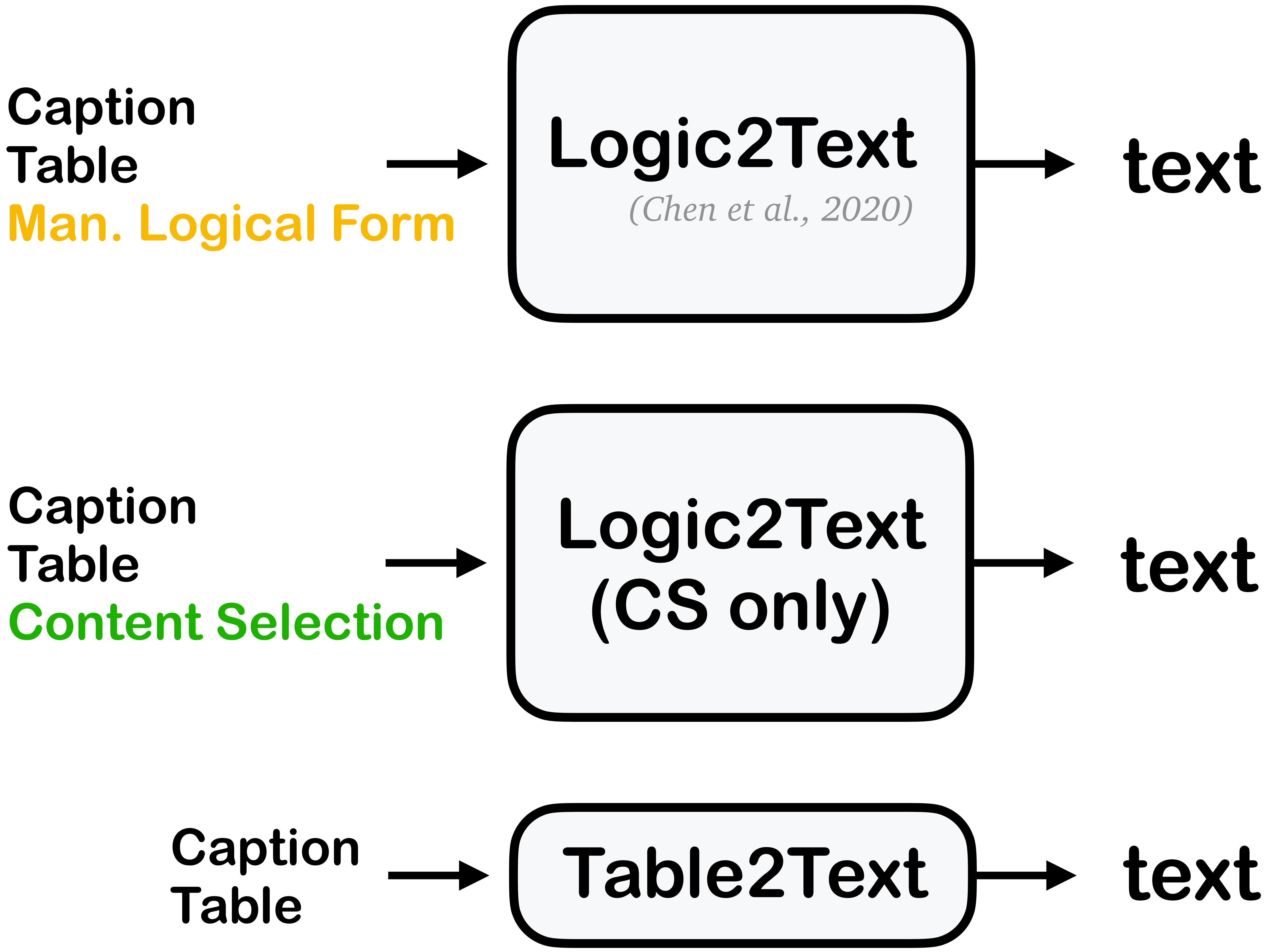
Do Logical Forms improve fidelity more than Content Selection values?

Can models automatically generate correct Logical Forms based on Content Selection?

Can automatically generated Logical Forms improve fidelity in Table-to-Text?

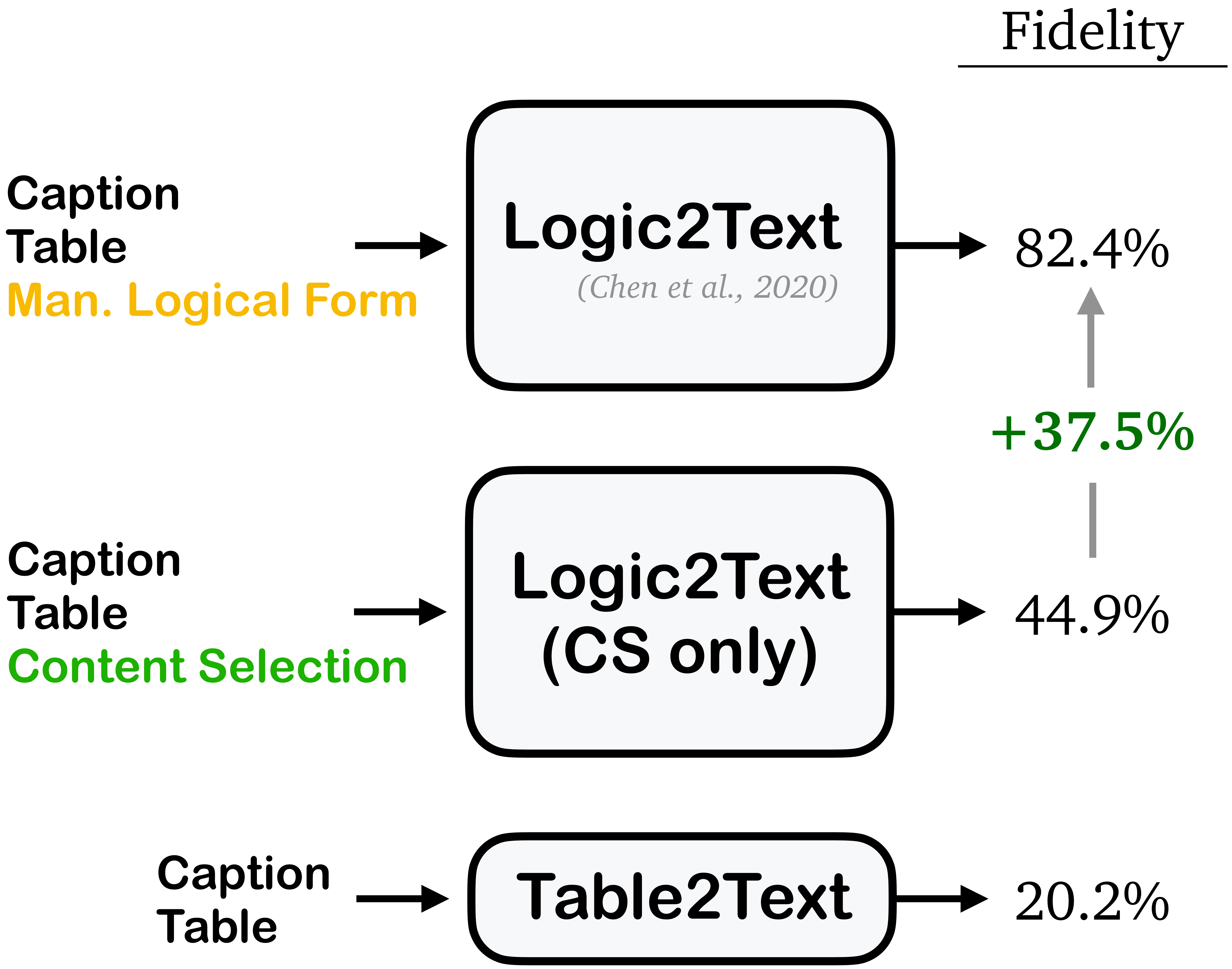
Experiments

Do Logical Forms improve fidelity more than Content Selection values?



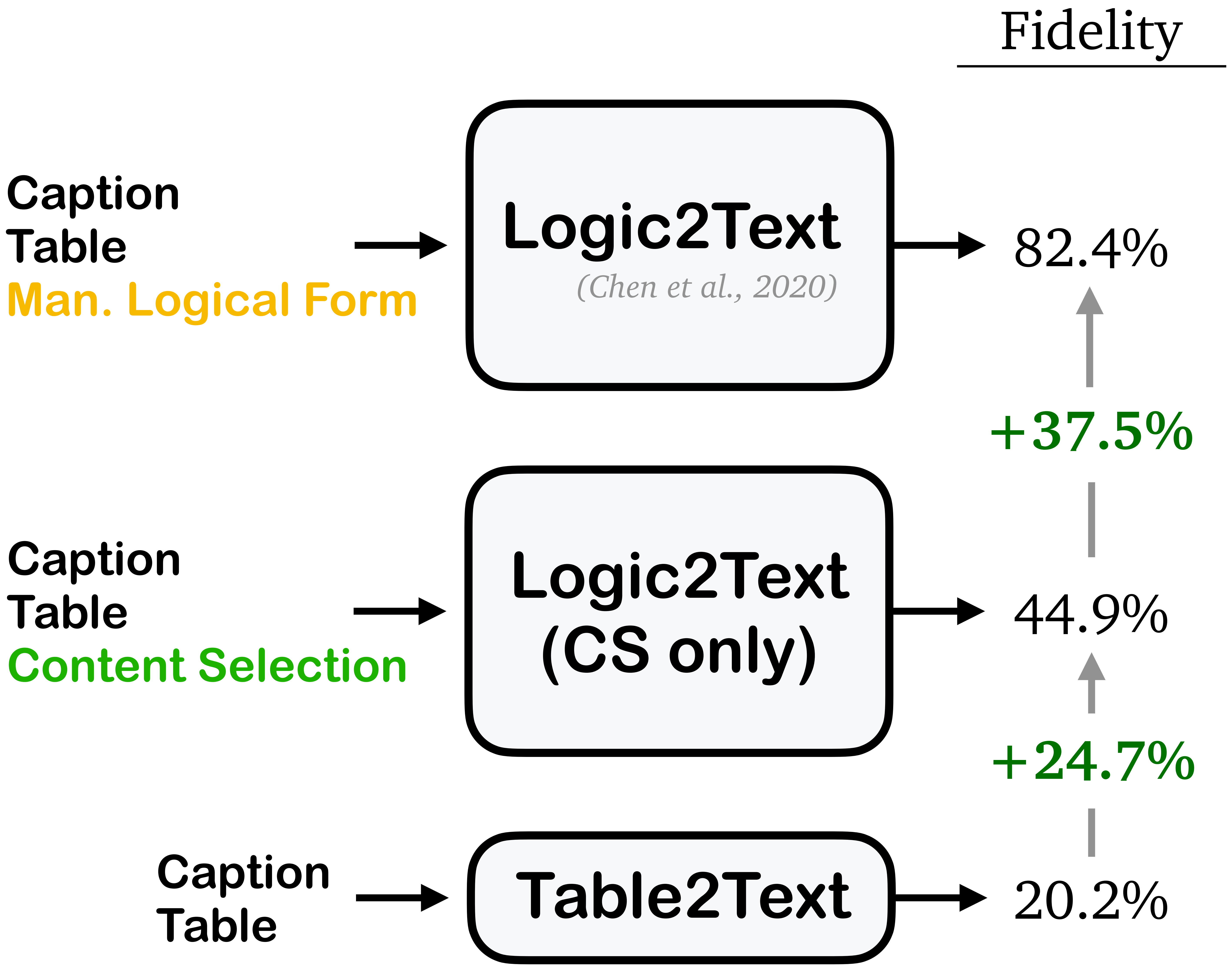
Experiments

Do Logical Forms improve fidelity more than Content Selection values?



Experiments

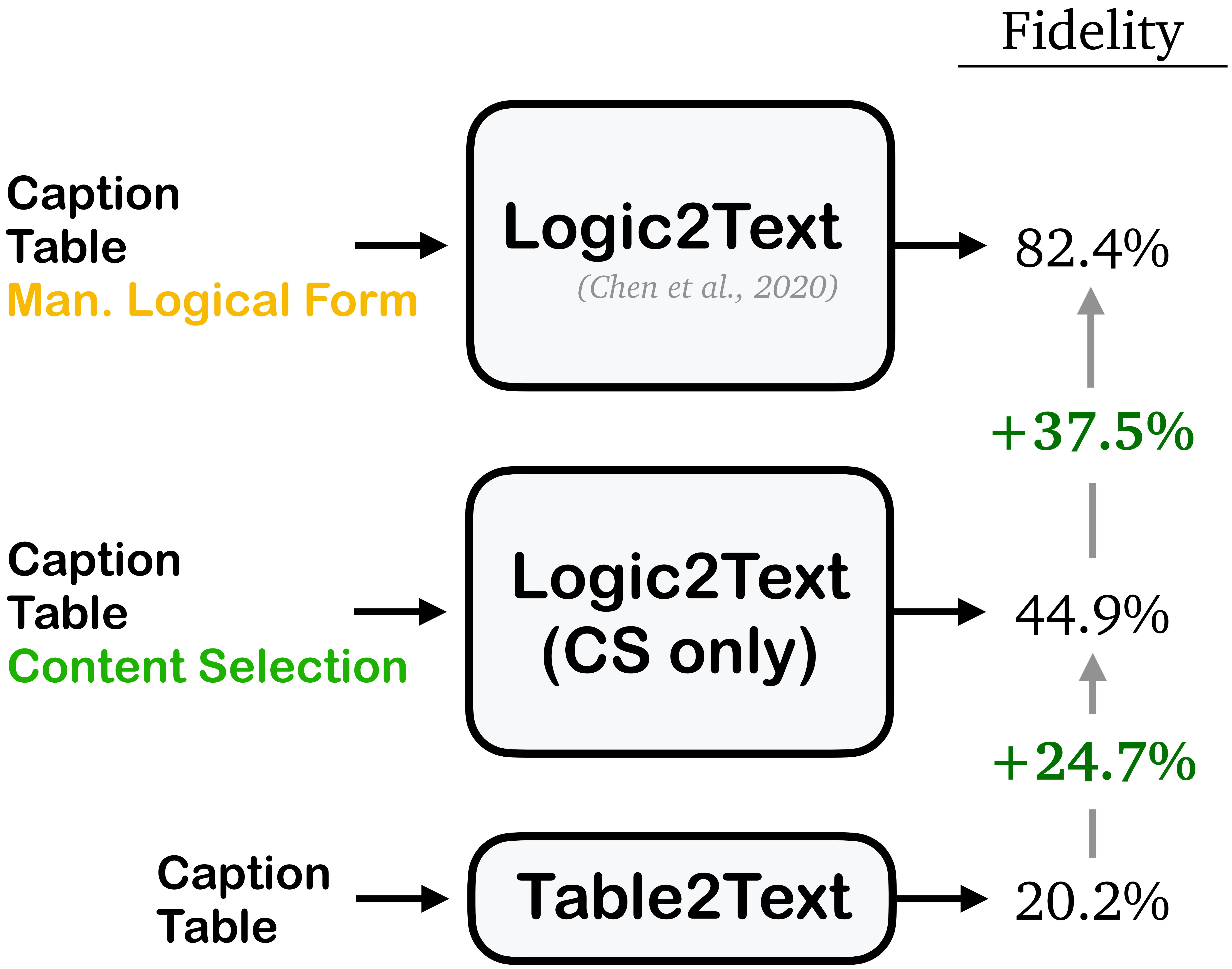
Do Logical Forms improve fidelity more than Content Selection values?



Experiments

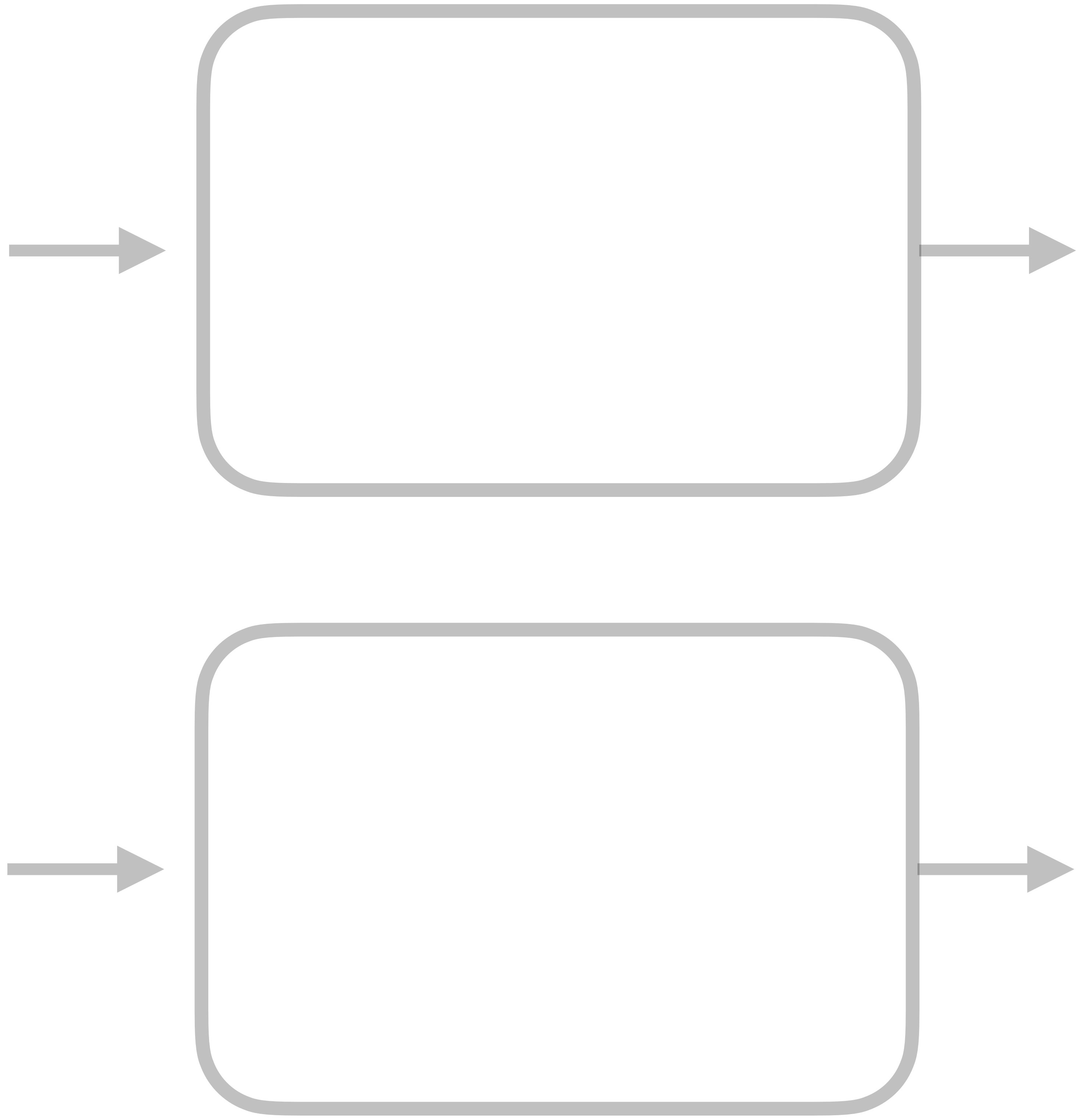
Do Logical Forms improve fidelity more than Content Selection values?

Yes



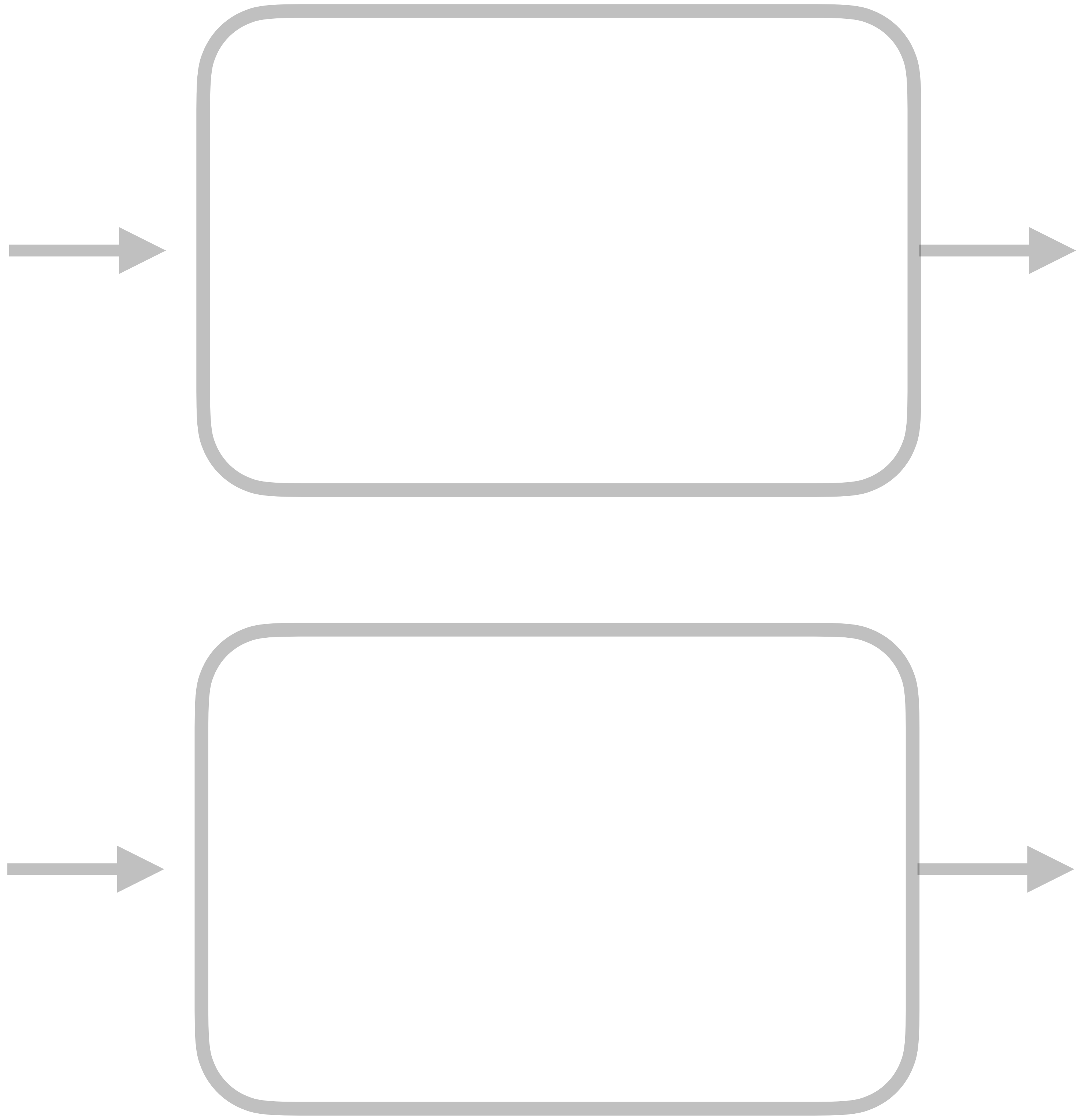
Experiments

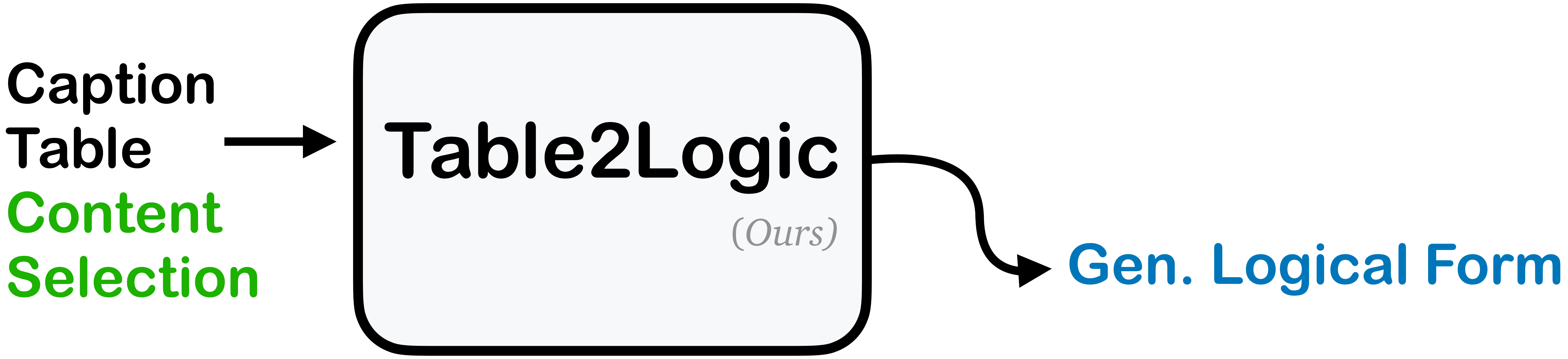
Can models automatically generate Logical Forms based on Content Selection?



Experiments

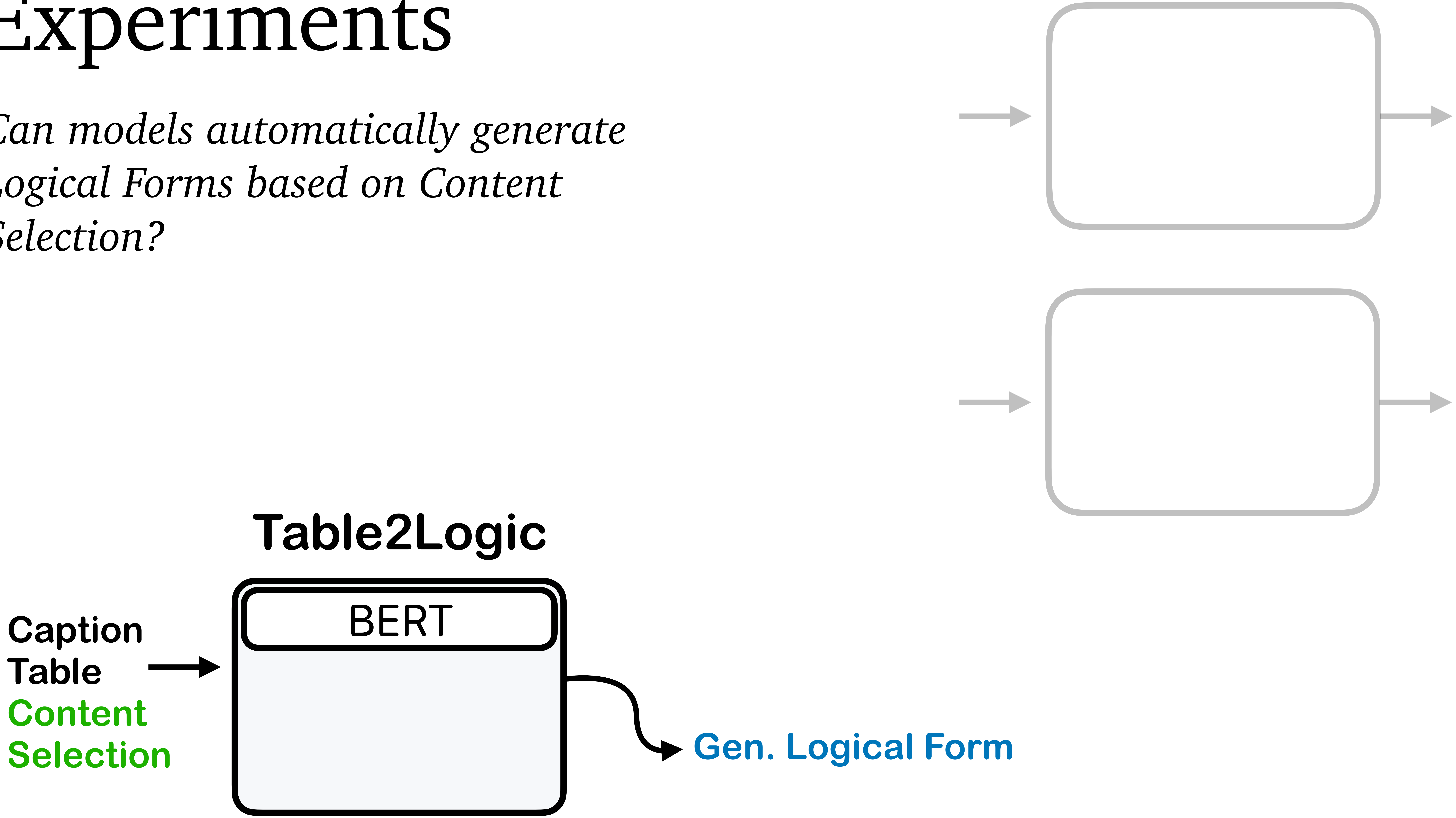
Can models automatically generate Logical Forms based on Content Selection?





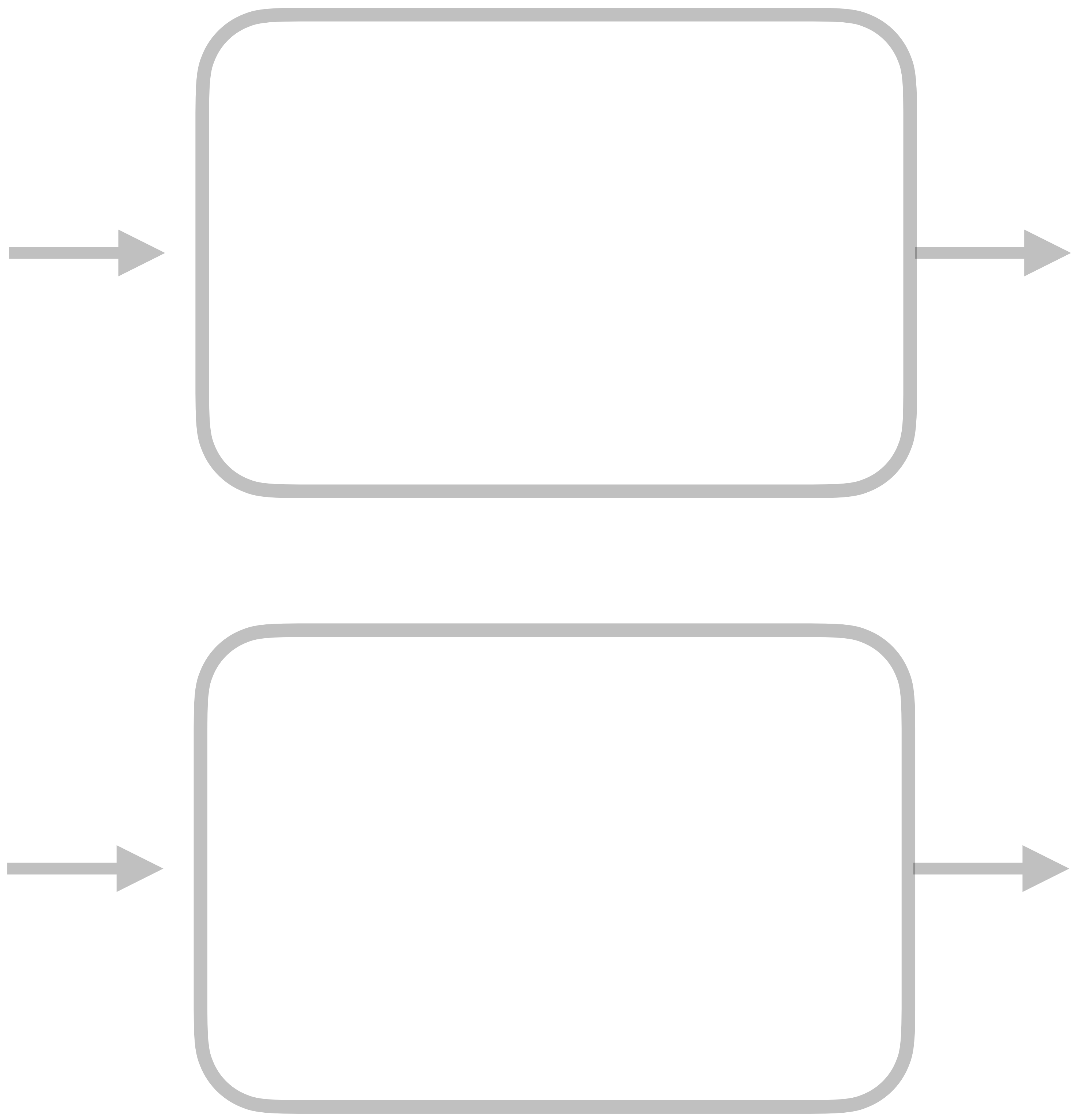
Experiments

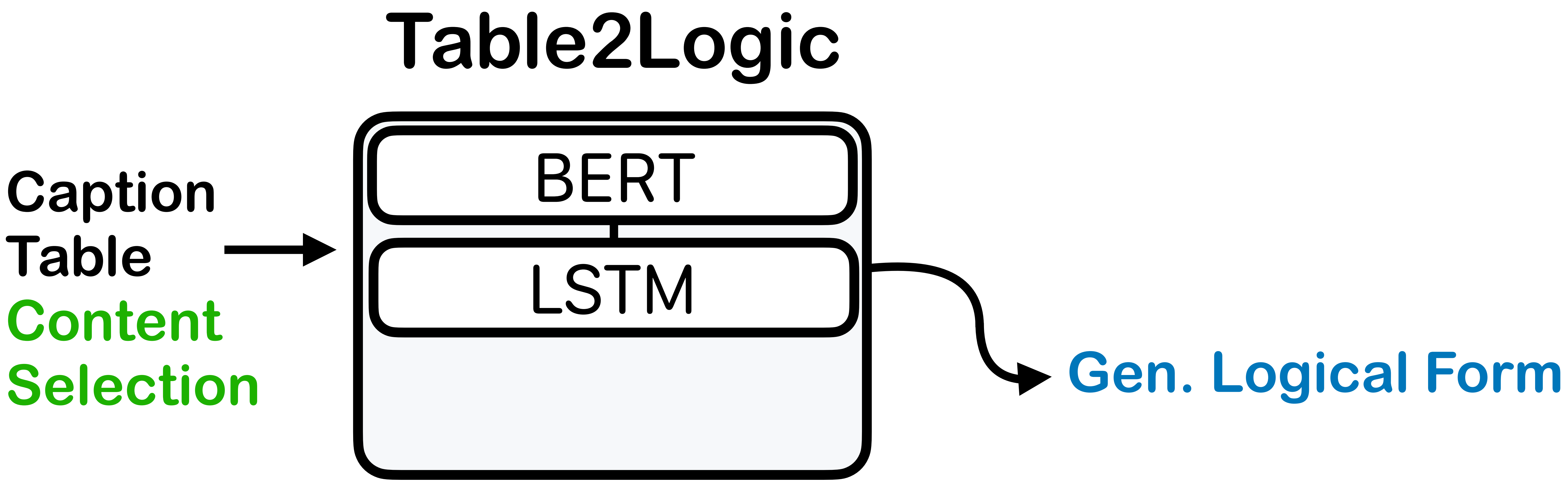
Can models automatically generate Logical Forms based on Content Selection?



Experiments

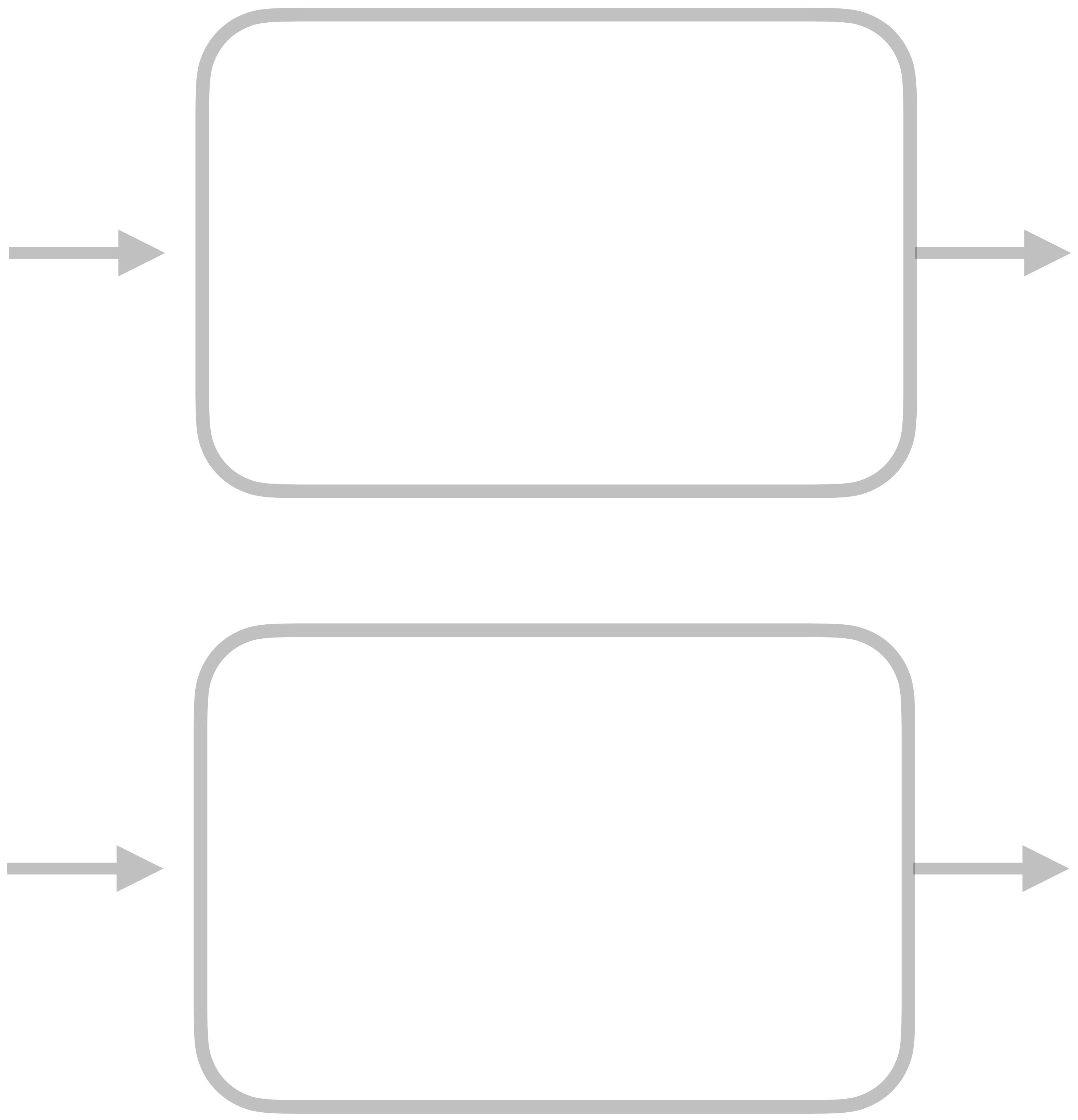
Can models automatically generate Logical Forms based on Content Selection?

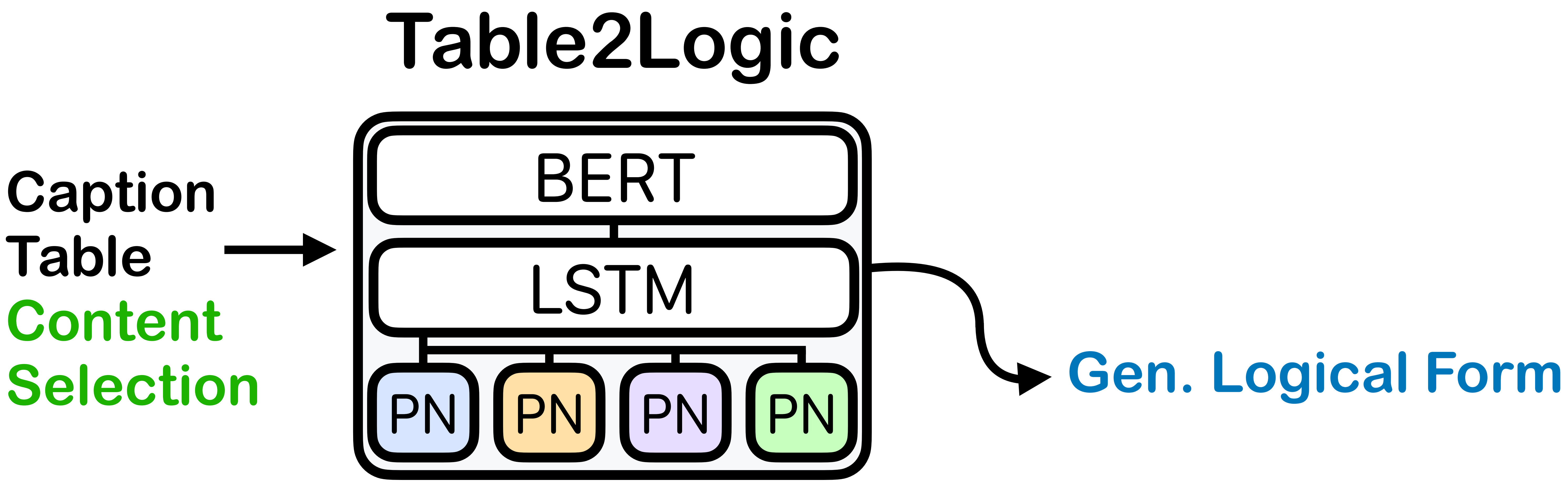




Experiments

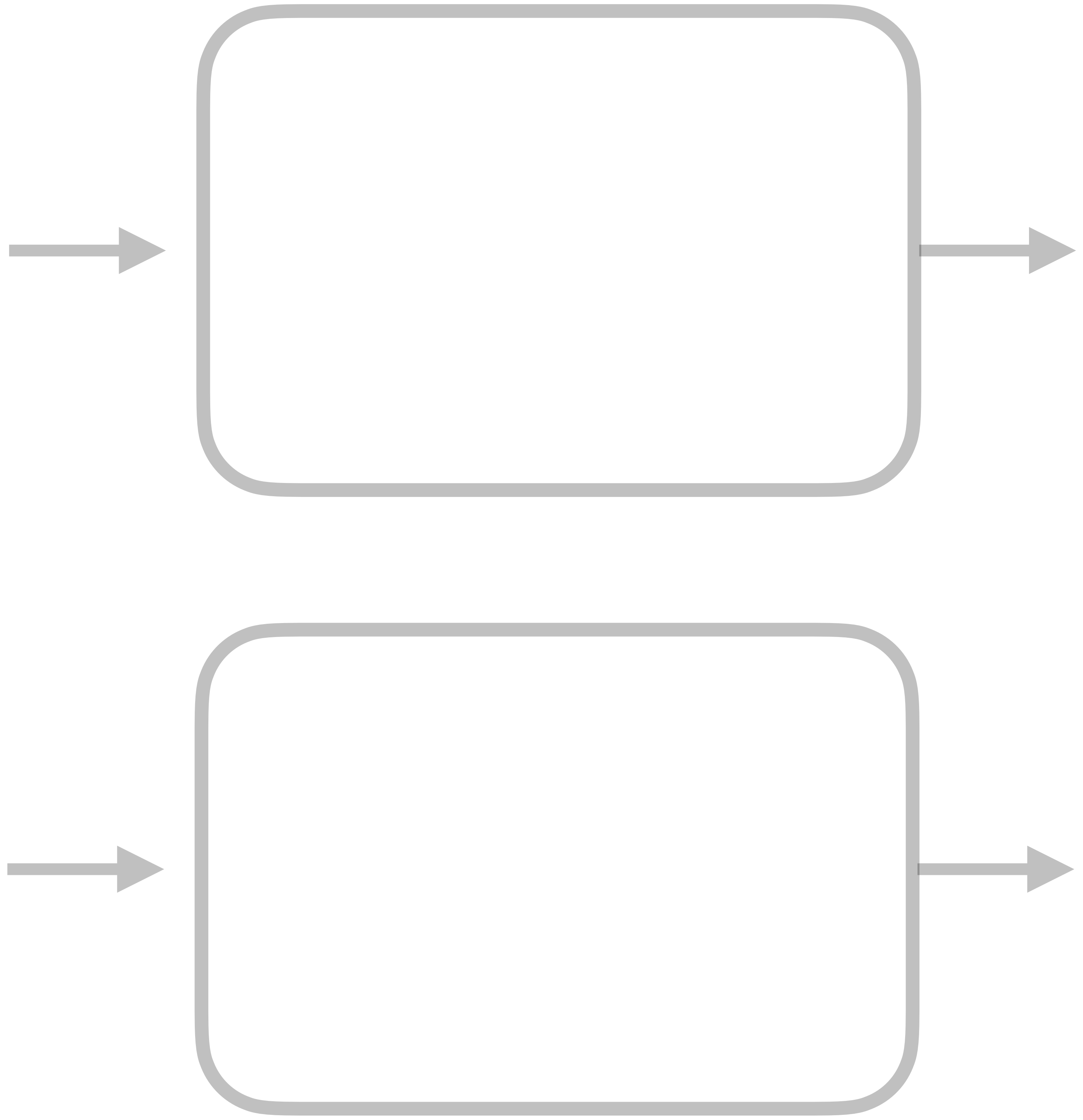
Can models automatically generate Logical Forms based on Content Selection?

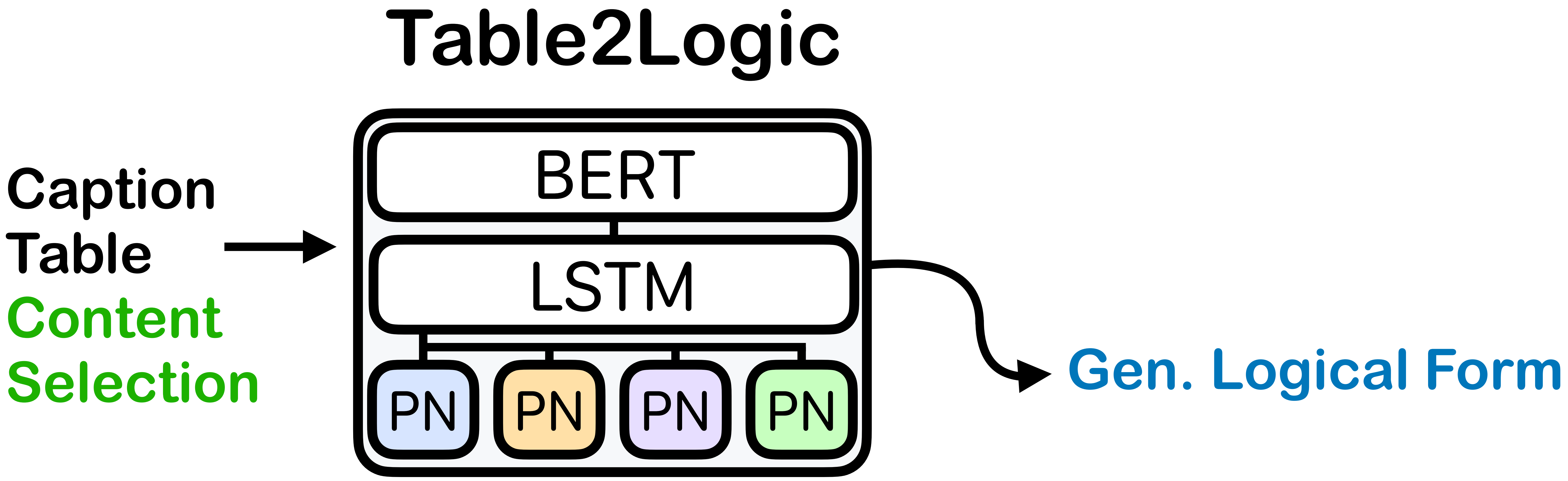


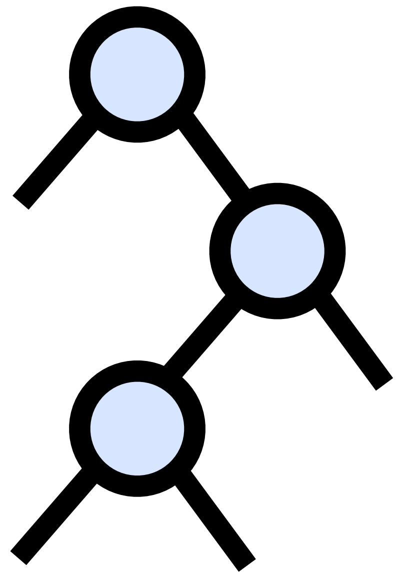


Experiments

Can models automatically generate Logical Forms based on Content Selection?

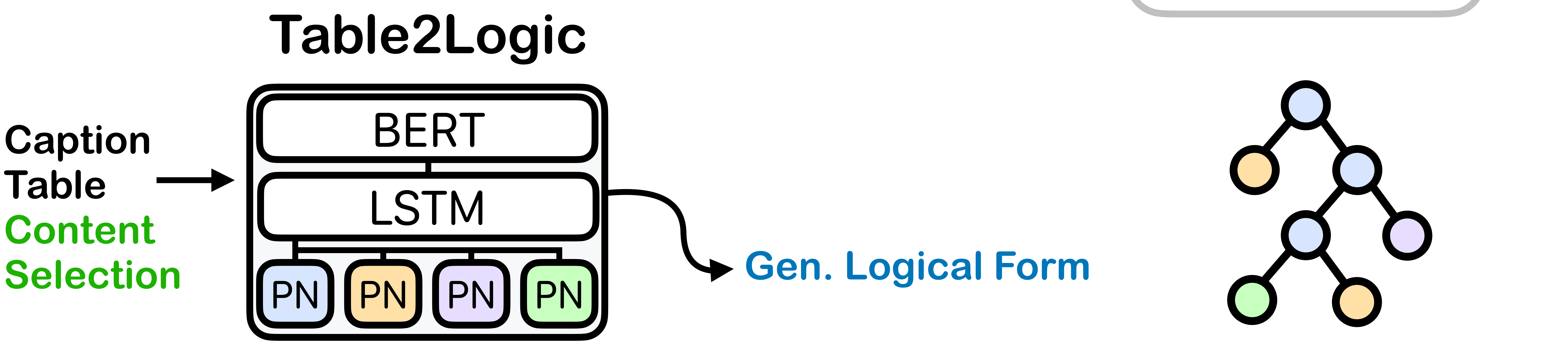






Experiments

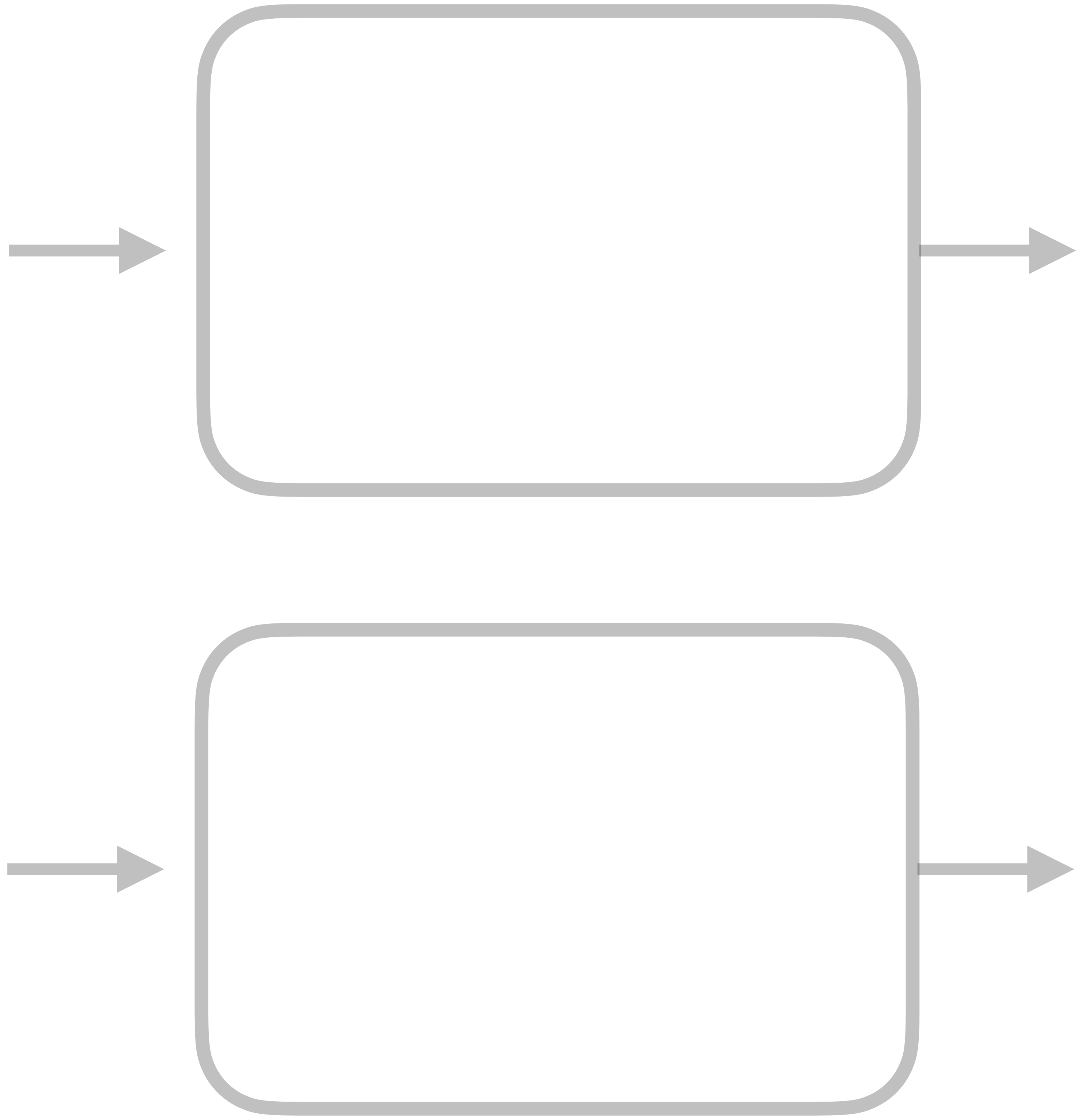
Can models automatically generate Logical Forms based on Content Selection?

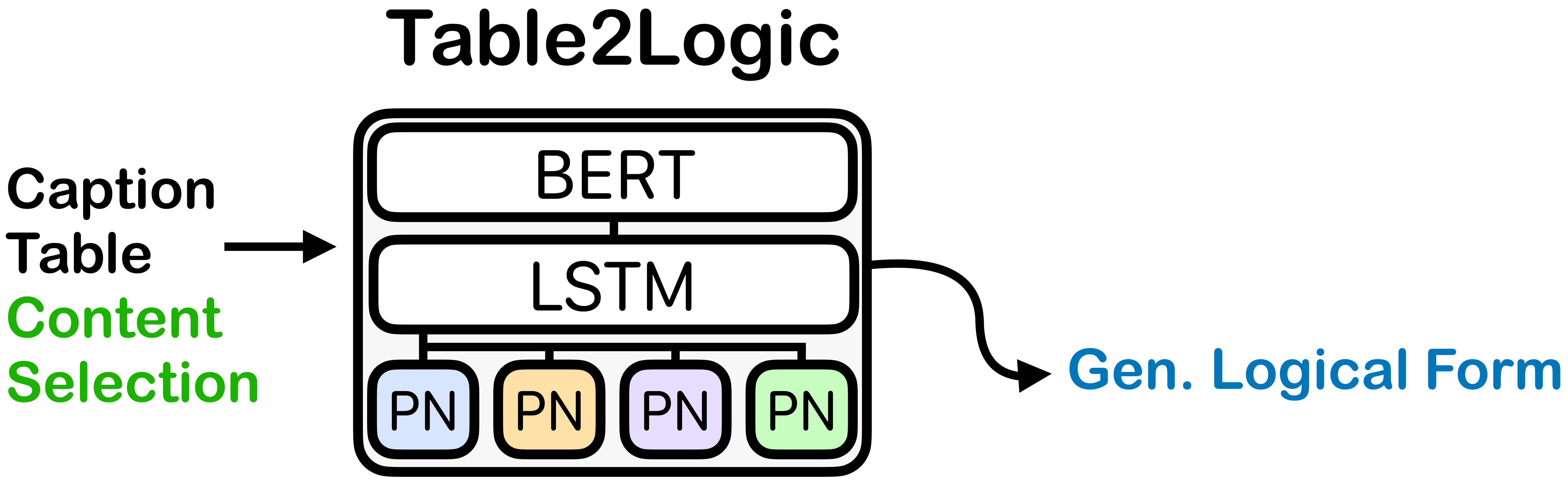


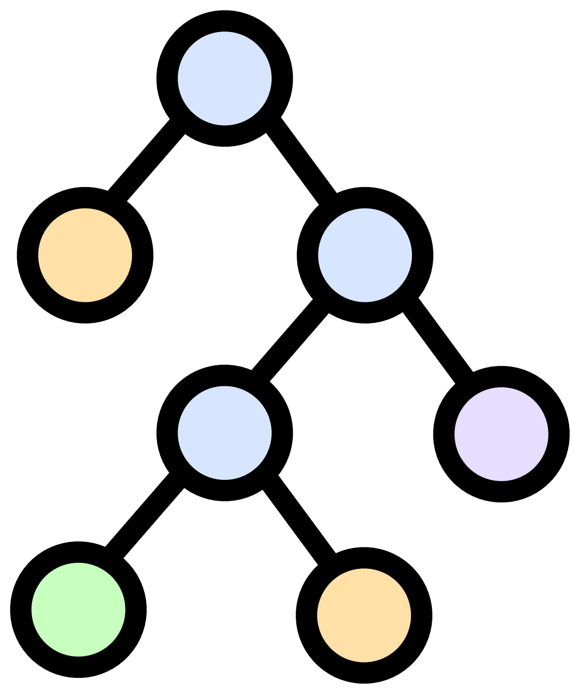
Experiments

Can models automatically generate Logical Forms based on Content Selection?

Yes

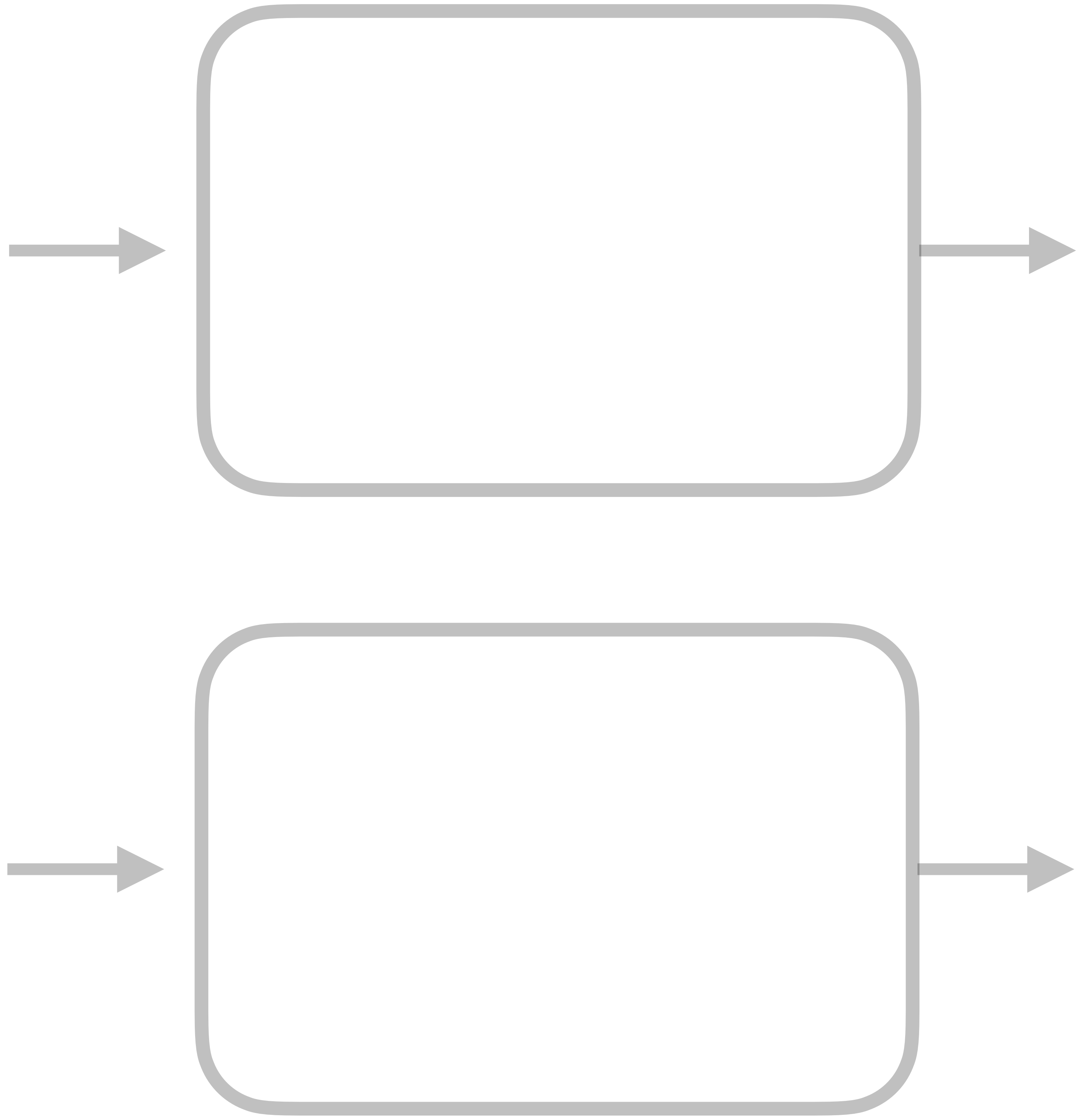






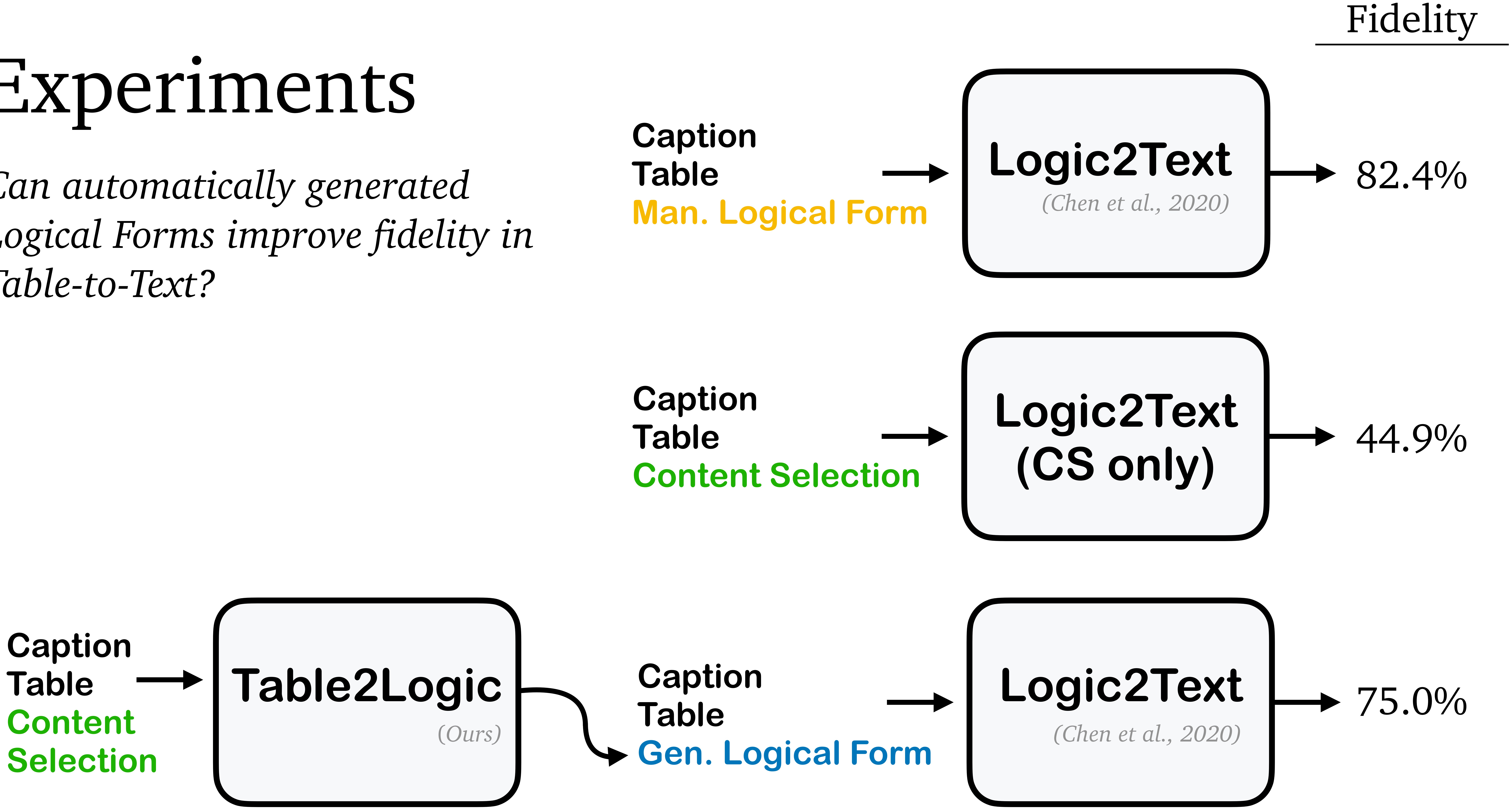
Experiments

Can automatically generated Logical Forms improve fidelity in Table-to-Text?



Experiments

Can automatically generated Logical Forms improve fidelity in Table-to-Text?



Experiments

Can automatically generated Logical Forms improve fidelity in Table-to-Text?

Yes

Fidelity

Caption
Table
Man. Logical Form
→ Logic2Text (Chen et al., 2020) → 82.4%

Caption
Table
Content Selection
→ Logic2Text (CS only) → 44.9%

+30.1%

Caption
Table
Content
Selection
→ Table2Logic (Ours) →

Caption
Table
Gen. Logical Form
→ Logic2Text (Chen et al., 2020) → 75.0%

Experiments

Can automatically generated Logical Forms improve fidelity in Table-to-Text?

Yes

Fidelity

Caption
Table
Man. Logical Form

Logic2Text
(Chen et al., 2020)

82.4%

Caption
Table
Content Selection

Logic2Text (CS only)

-7.4%

Caption
Table
Content
Selection

Table2Logic
(Ours)

Caption
Table
Gen. Logical Form

Logic2Text
(Chen et al., 2020)

75.0%

Conclusions

- *Logical Forms improve fidelity compared to using only Content Selection values.*

- *Logical Forms can be generated automatically based on the Content Selection values.*

- *Automatic Logical Forms improve fidelity in Table-to-Text generation.*

Expert Systems With Applications 238 (2024) 121869

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Automatic Logical Forms improve fidelity in Table-to-Text generation

Iñigo Alonso *, Eneko Agirre

HiTZ Basque Center for Language Technologies - Ixa NLP Group, University of the Basque Country (UPV/EHU), M. Lardizabal 1, Donostia, 20018, Basque Country, Spain

ARTICLE INFO

Keywords:
Natural Language Generation
Table-to-Text
Deep learning
Logical forms
Faithfulness
Hallucinations

ABSTRACT

Table-to-text systems generate natural language statements from structured data like tables. While end-to-end techniques suffer from low factual correctness (fidelity), a previous study reported fidelity gains when using manually produced graphs that represent the content and semantics of the target text called Logical Forms (LF). Given the use of manual LFs, it was not clear whether automatic LFs would be as effective, and whether the improvement came from the implicit content selection in the LFs. We present **T𝓁T**, a system which, given a table and a set of pre-selected table values, first produces LFs and then the textual statement. We show for the first time that automatic LFs improve the quality of generated texts, with a 67% relative increase in fidelity over a comparable system not using LFs. Our experiments allow to quantify the remaining challenges for high factual correctness, with automatic selection of content coming first, followed by better Logic-to-Text generation and, to a lesser extent, improved Table-to-Logic parsing.

1. Introduction

Data-to-text generation is the task of taking non-linguistic structured input such as tables, knowledge bases, tuples, or graphs, and automatically producing factually correct[1] textual descriptions of the contents of the input (Covington, 2001; Gatt & Krahmer, 2018; Reiter & Dale, 1997). Real-world applications include, among others, generating weather forecasts from meteorological data (Goldberg, Driedger, & Kittredge, 1994), producing descriptions from biographical information (Lebret, Grangier, & Auli, 2016), or generating sport summaries using game statistics (Wiseman, Shieber, & Rush, 2017). In these applications, the goal is to represent relevant information in the input data using natural language descriptions. Therefore, generating text that faithfully and accurately represents the underlying information in the source becomes critical. It should be noted that the task is underspecified, in the sense that the same table may be described by multiple textual descriptions, all of them correct, as each one can focus on different, relevant subsets of the input data. This makes the use of manual evaluation of fidelity key to measure the quality of the generated text. Our work focuses on how to improve faithfulness automatically.

Various Data-to-Text approaches have emerged to address this challenge. Methods include leveraging the structural information of the input data (Chen, Su, Yan, & Wang, 2020; Puduppully, Dong, & Lapata, 2019b; Wiseman et al., 2017), using neural templates (Wiseman, Shieber, & Rush, 2018), or focusing on content ordering (Puduppully, Dong, & Lapata, 2019a). Recent techniques (Aghajanyan et al., 2022; Chen, Chen, Su, Chen, & Wang, 2020; Chen, Chen, Zha et al., 2020; Kasner & Dusek, 2022) leverage large-scale pre-trained models (Devlin, Chang, Lee, & Toutanova, 2019), and report significant performance gains in terms of fluency and generalization with respect to previous work that did not use such models.

However, these end-to-end systems struggle with fidelity as they are still susceptible to produce hallucinations, i.e. they generate text that, despite its fluency, does not describe in a faithful way the input data (Koehn & Knowles, 2017; Maynez, Narayan, Bohnet, & McDonald, 2020).

In this context Chen, Chen, Zha et al. (2020) propose to reformulate Data-to-Text as a Logic-to-Text problem. Alongside the usual table information, the input to the language realization module in this approach also includes a tree-structured graph representation of the semantics of the target text called logical form (LF). Logical forms follow compositional semantics (Carnap, 1947) to formalize the underlying meanings represented in the target text. When provided alongside tables in this case, the meaning conveyed by LFs is related to a semantic context as defined in Wang, Liu, Ip, Zhang, and Deters (2014), Zhang (1994). In this case, the semantic context is given by the table. An example of how LFs represent this meaning can be seen in Fig. 2. Although the LFs were applied to tables in this paper, the proposal could be easily extended to other Data-to-Text problems.

* Corresponding author.
E-mail addresses: inigoborja.alonso@ehu.eus (I. Alonso), e.agirre@ehu.eus (E. Agirre).

[1] We use the terms factual correctness, faithfulness, and fidelity indistinctly.

https://doi.org/10.1016/j.eswa.2023.121869
Received 24 April 2023; Received in revised form 18 September 2023; Accepted 26 September 2023
Available online 12 October 2023

Expert Systems With Applications

Automatic Logical Forms improve fidelity in Table-to-Text generation

Iñigo Alonso, and Eneko Agirre

Representation

Pixel-based Table-To-Text Generation

Representation

Pixel-based Table-To-Text Generation

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

<page_title> 1898 Open Championship
<page_title> <table> <row> <cell> Place </
cell> <cell> Player <row_header> Place </
row_header> </cell> <cell> Country
<row_header> Place </row_header>
<row_header> Player </row_header> </cell>
<cell> Score <row_header> Place </
row_header> <row_header> Player </
row_header> <row_header> Country </
row_header> </cell> </row> <row> <cell> 1
<col_header> Place </col_header> </cell>
<cell> Willie Park, Jr. <col_header>
Player </col_header> </cell> <cell>
Scotland <col_header> Country </
col_header> </cell> <cell> 151
<col_header> Score </col_header> </cell>
<row> <cell> 2 <col_header> Place </
col_header> </cell> <cell> Harry Vardon
<col_header> Player </col_header> </cell>
<cell> Wales <col_header> Country </
col_header> </cell> <cell> 154
<col_header> Score </col_header> </cell>

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

ToTTo *(Parikh et al., 2020)*

Table Representation

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

<page_title> 1898 Open Championship </page_title> <table> <cell> Willie Park, Jr. <col_header> Player </col_header> </cell> <cell> 76-75=151 <col_header> Score </col_header> </cell> <cell> Harold Hilton (a) <col_header> Player </col_header> </cell> <cell> 76-81=157 <col_header> Score </col_header> </cell> </table>

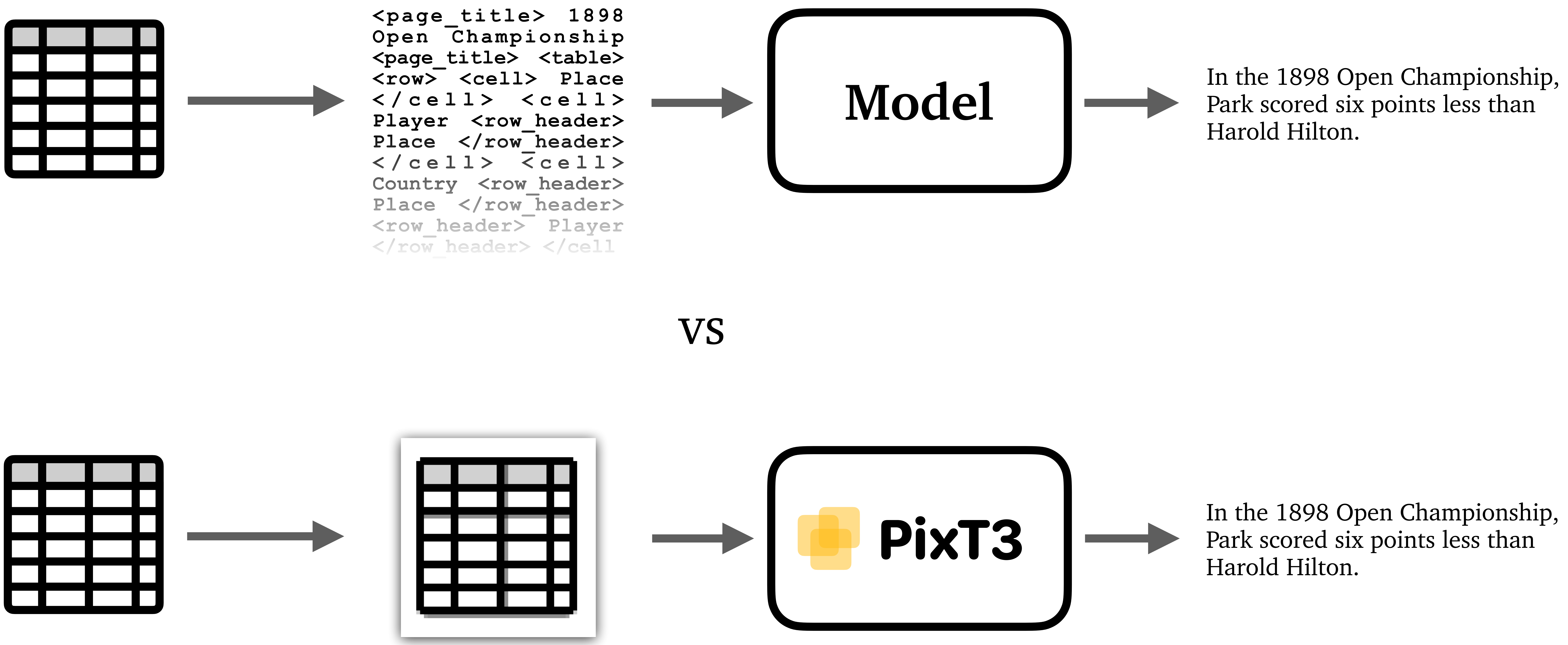
In the 1898 Open Championship, Park scored six points less than Harold Hilton.

ToTTo *(Parikh et al., 2020)*

Club	Season	League			National Cup		Continental		Other		Total	
		Division	Apps	Goals	Apps	Goals	Apps	Goals	Apps	Goals	Apps	Goals
	2011-12		0 / 0	0	0	0	0	0			0	0
	2012-13		0 / 1	0	0 / 1	0			0 / 1	0		
	Total		**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**
	2012-13		1 / 6	0	1 / 6	0			1 / 6	w: 0		
	2013-14		13 / 15	1	13 / 15	0			13 / 15	0		
	Total		**19**	**1**	**2**	**0**	**0**	**0**	**0**	**0**	**21**	**1**
	2014-15		11 / 7	1	11 / 7	0			11 / 7	1		
	2016-17		36 / 4	3	36 / 4	0			36 / 4	l: 3		
	2017-18		24 / 31	3	24 / 31	0			24 / 31	3		
	2018-19		4 / 72	0	4 / 72	0			4 / 72	0		
	Total		**64**	**6**	**5**	**0**					**69**	**6**
Career total			**83**	**7**	**8**	**2**	**0**	**0**	**0**	**0**	**91**	**7**

Irregular Table

Table-to-Text as Image-to-Text



Can Vision-Language Models perform Table-to-Text Generation?

Can Vision-Language Models perform Table-to-Text Generation?

Can this approach maintain the same level of fidelity as its unimodal counterparts?

Can Vision-Language Models perform Table-to-Text Generation?

Can this approach maintain the same level of fidelity as its unimodal counterparts?

Are images a space-efficient modality for representing tables for Table-to-Text Generation?

Advancements in

Visual Language Understanding

- Dessurt *(Davis et al., 2022)*
- Donut *(Kim et al., 2022)*
- Pix2Struct *(Lee et al., 2022)*

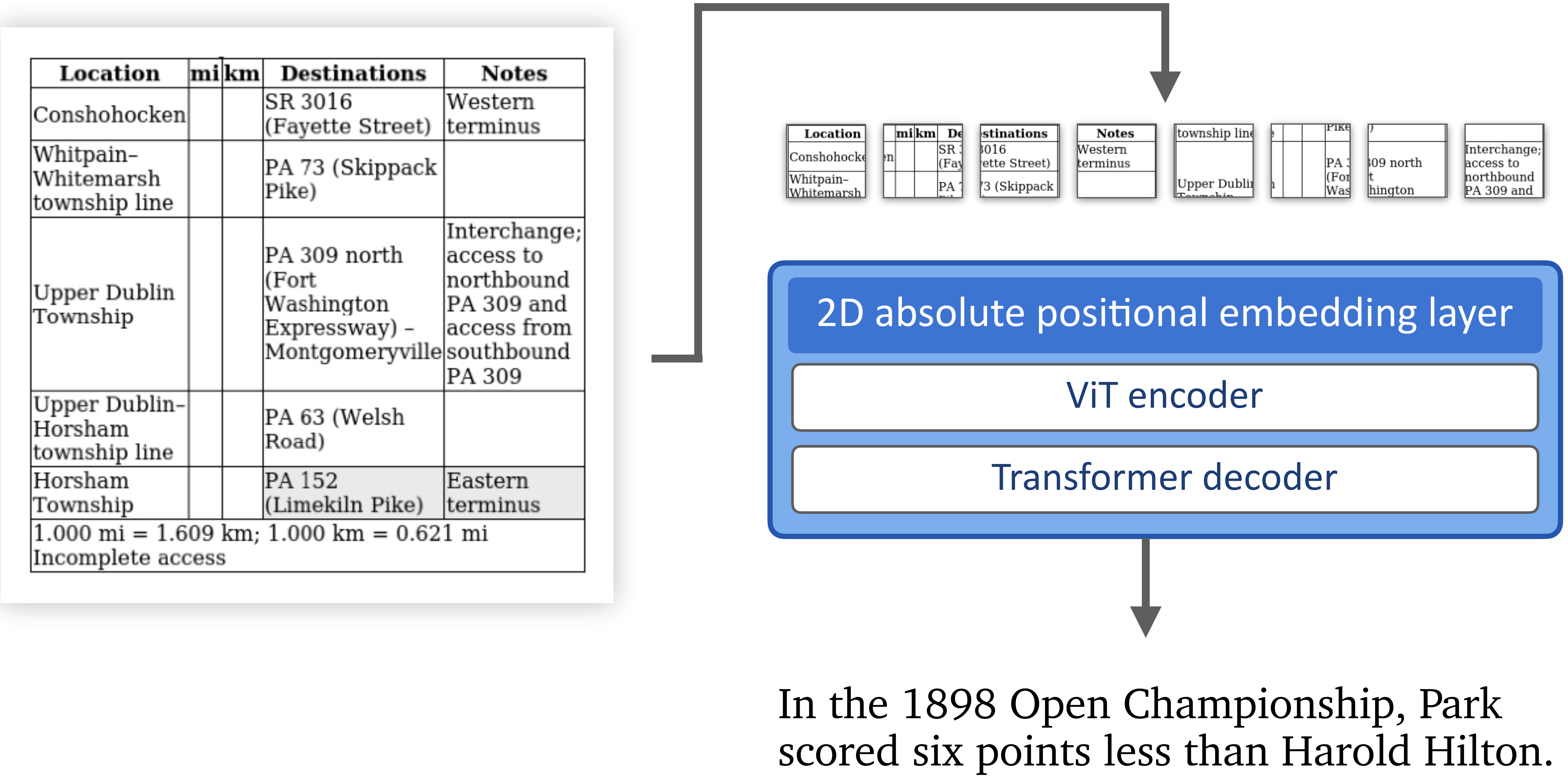
Advancements in

Visual Language Understanding

- Dessurt *(Davis et al., 2022)*
- Donut *(Kim et al., 2022)*
- Pix2Struct *(Lee et al., 2022)*

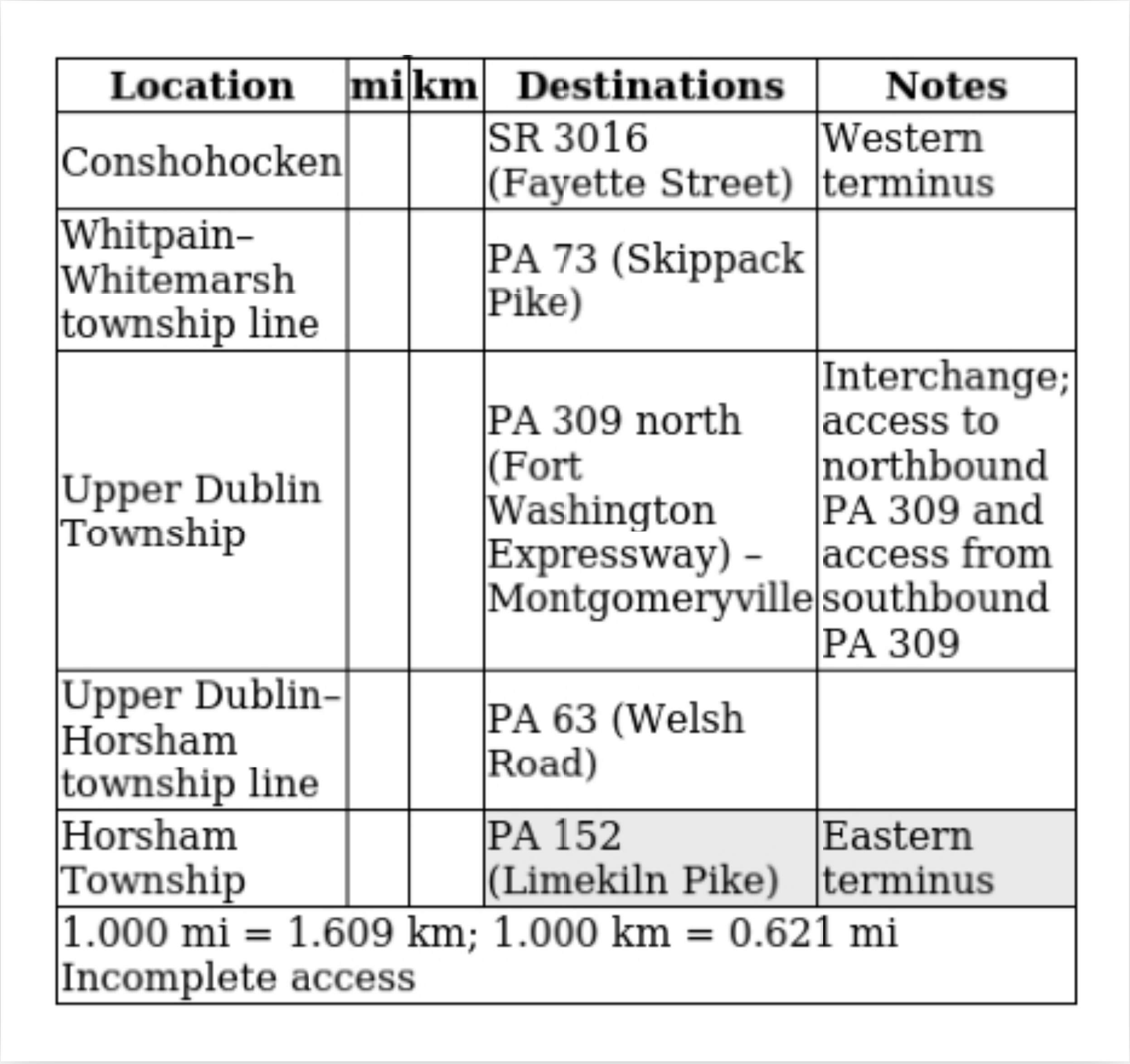
PixT3

Architecture of **PixT3**

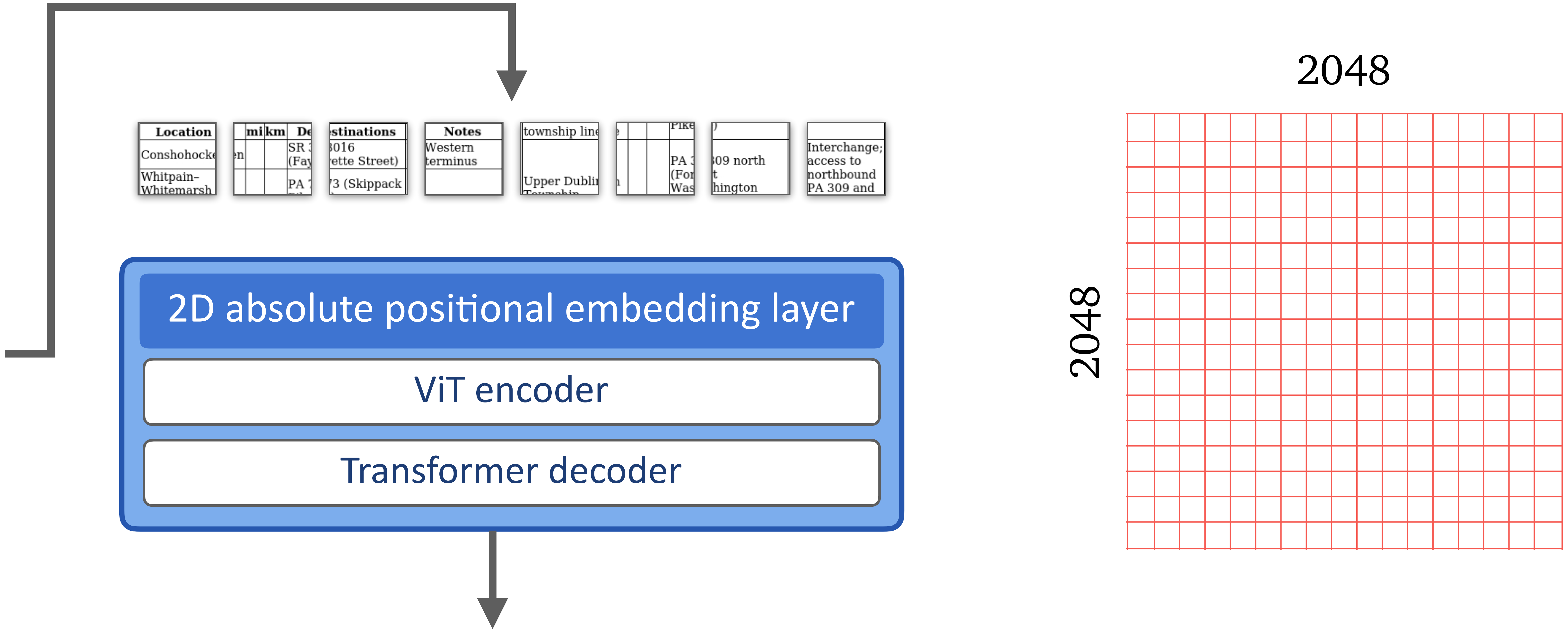


Location	mi	km	Destinations	Notes
Conshohocken			SR 3016 (Fayette Street)	Western terminus
Whitpain–Whitemarsh township line			PA 73 (Skippack Pike)	
Upper Dublin Township			PA 309 north (Fort Washington Expressway) – Montgomeryville	Interchange; access to northbound PA 309 and access from southbound PA 309
Upper Dublin–Horsham township line			PA 63 (Welsh Road)	
Horsham Township			PA 152 (Limekiln Pike)	Eastern terminus
1.000 mi = 1.609 km; 1.000 km = 0.621 mi Incomplete access				

Architecture of **PixT3**

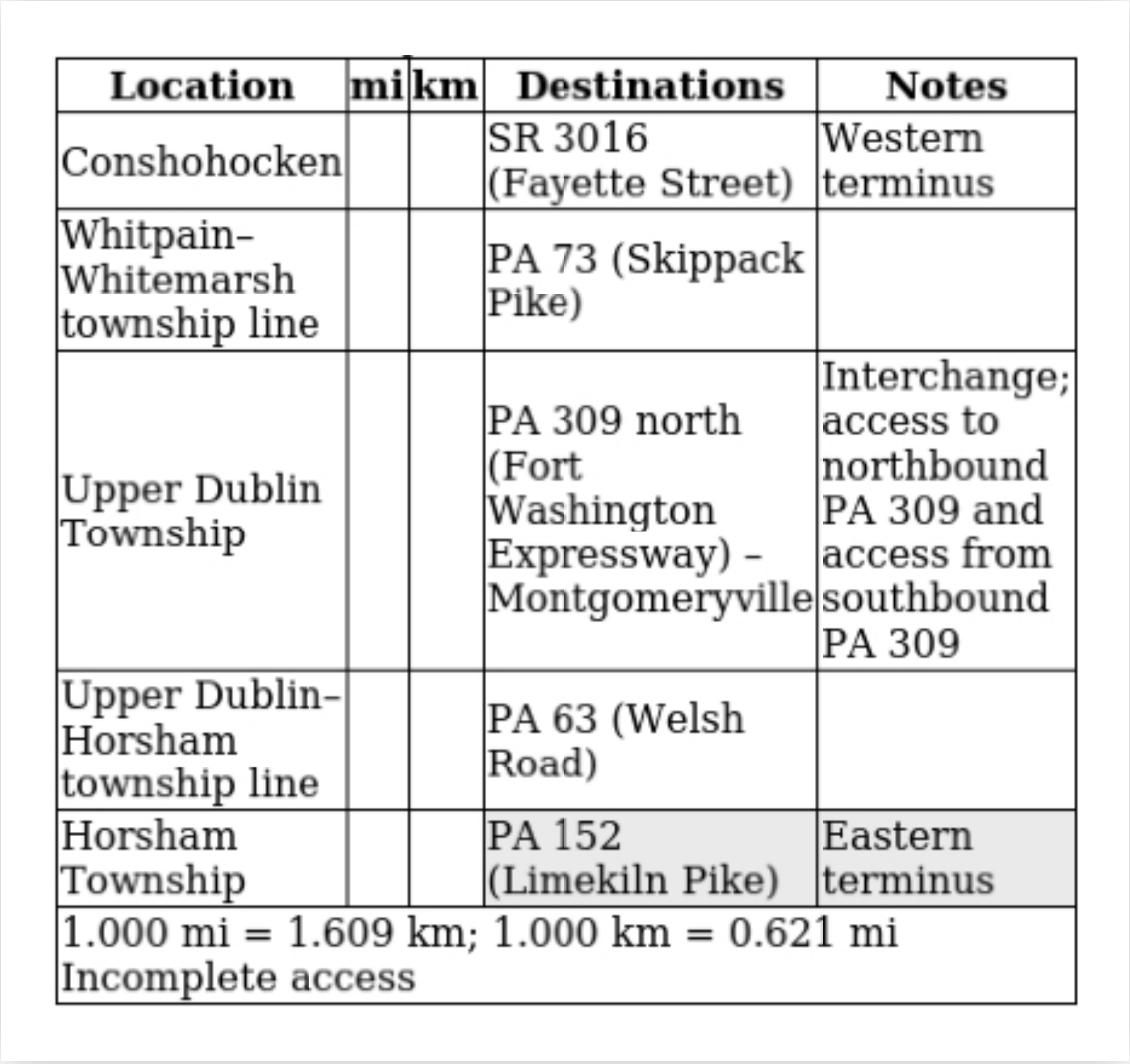


Location	mi	km	Destinations	Notes
Conshohocken			SR 3016 (Fayette Street)	Western terminus
Whitpain–Whitemarsh township line			PA 73 (Skippack Pike)	
Upper Dublin Township			PA 309 north (Fort Washington Expressway) – Montgomeryville	Interchange; access to northbound PA 309 and access from southbound PA 309
Upper Dublin–Horsham township line			PA 63 (Welsh Road)	
Horsham Township			PA 152 (Limekiln Pike)	Eastern terminus
1.000 mi = 1.609 km; 1.000 km = 0.621 mi				
Incomplete access				



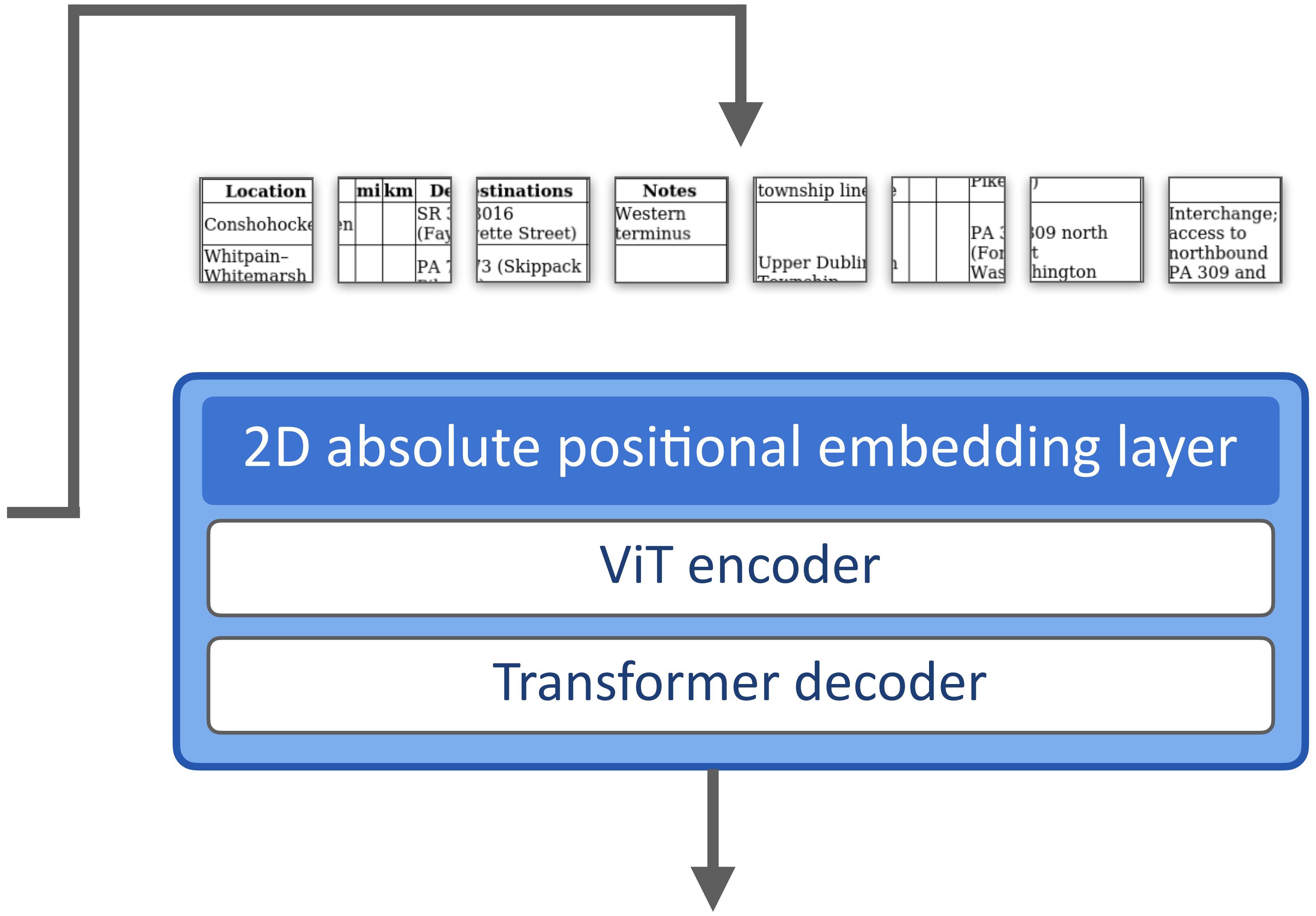
In the 1898 Open Championship, Park scored six points less than Harold Hilton.

Architecture of **PixT3**

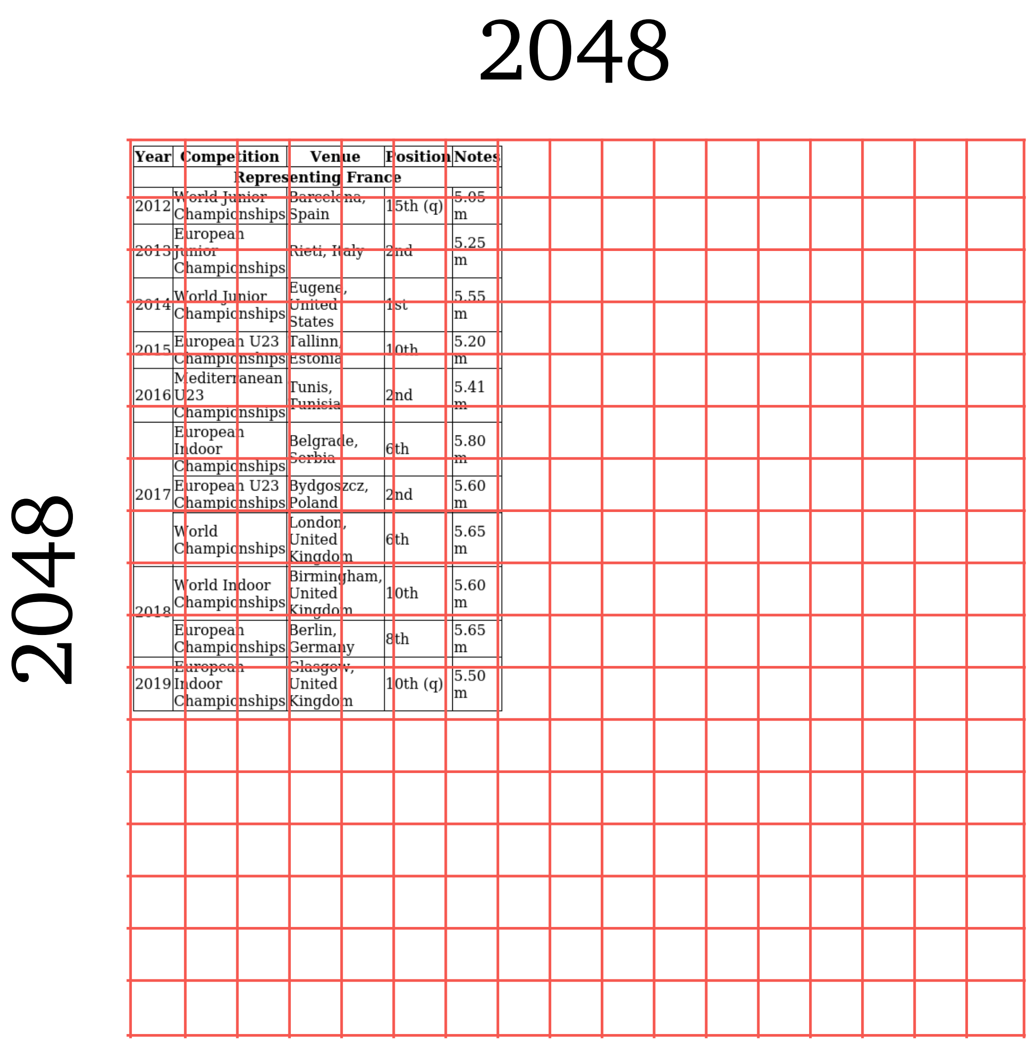


Location	mi	km	Destinations	Notes
Conshohocken			SR 3016 (Fayette Street)	Western terminus
Whitpain–Whitemarsh township line			PA 73 (Skippack Pike)	
Upper Dublin Township			PA 309 north (Fort Washington Expressway) – Montgomeryville	Interchange; access to northbound PA 309 and access from southbound PA 309
Upper Dublin–Horsham township line			PA 63 (Welsh Road)	
Horsham Township			PA 152 (Limekiln Pike)	Eastern terminus

1.000 mi = 1.609 km; 1.000 km = 0.621 mi
Incomplete access

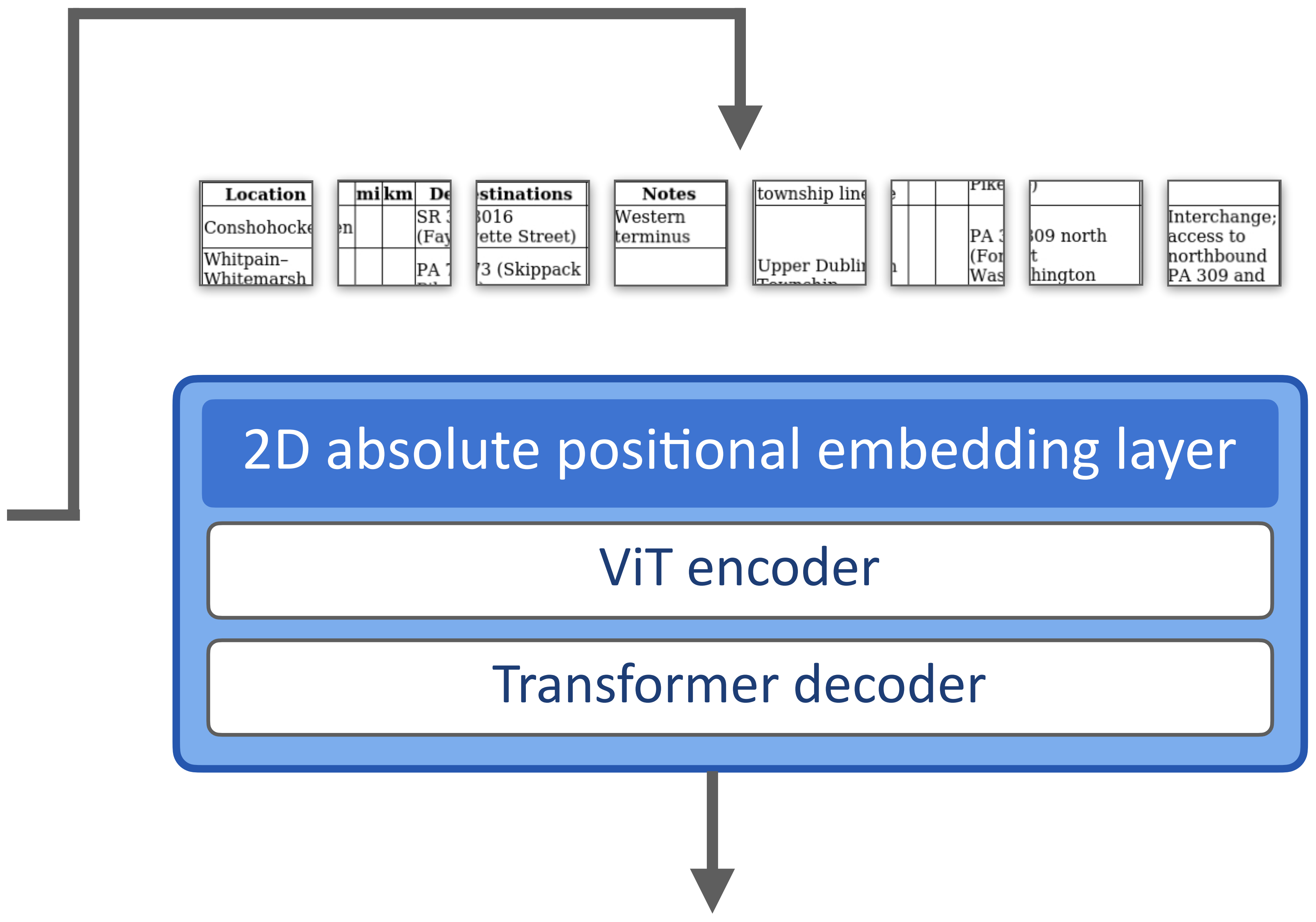


In the 1898 Open Championship, Park scored six points less than Harold Hilton.

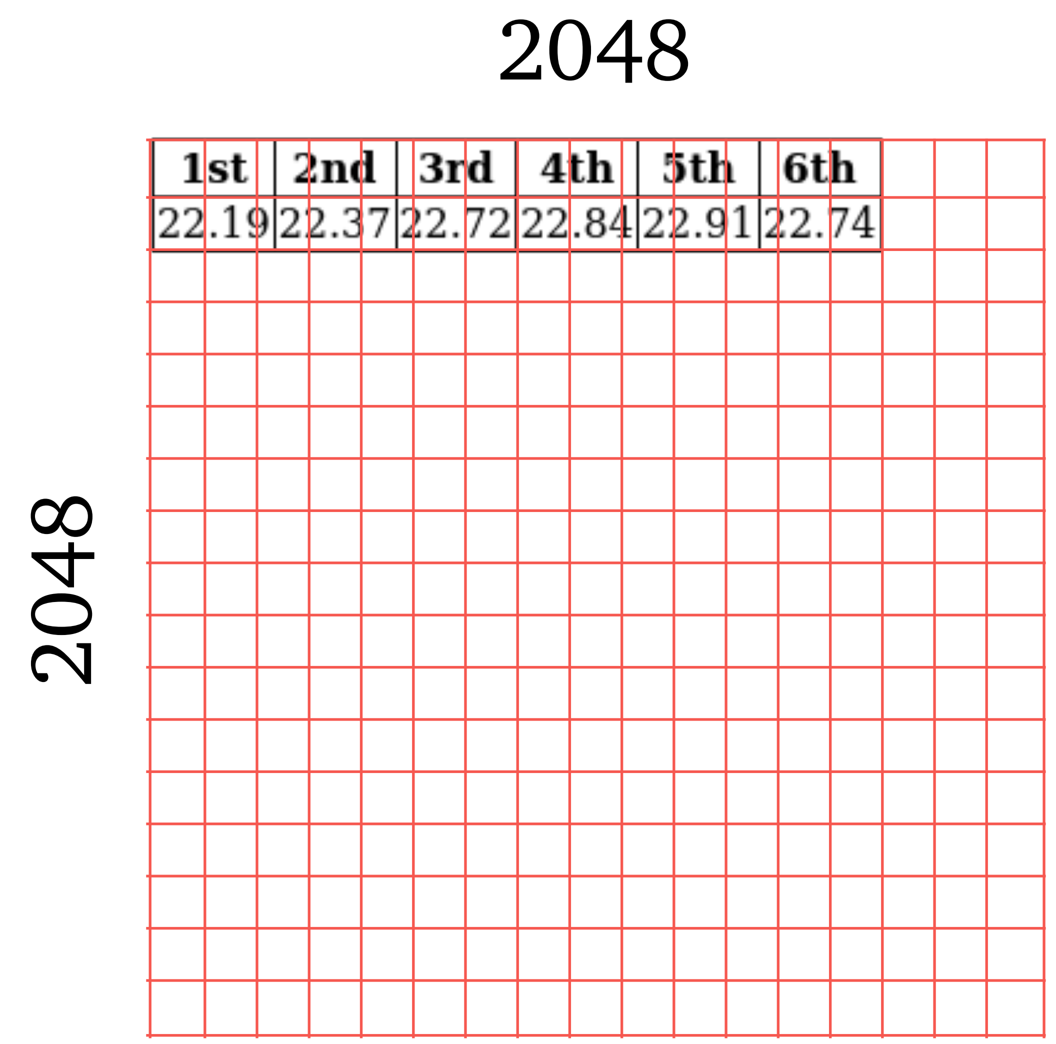


Architecture of **PixT3**

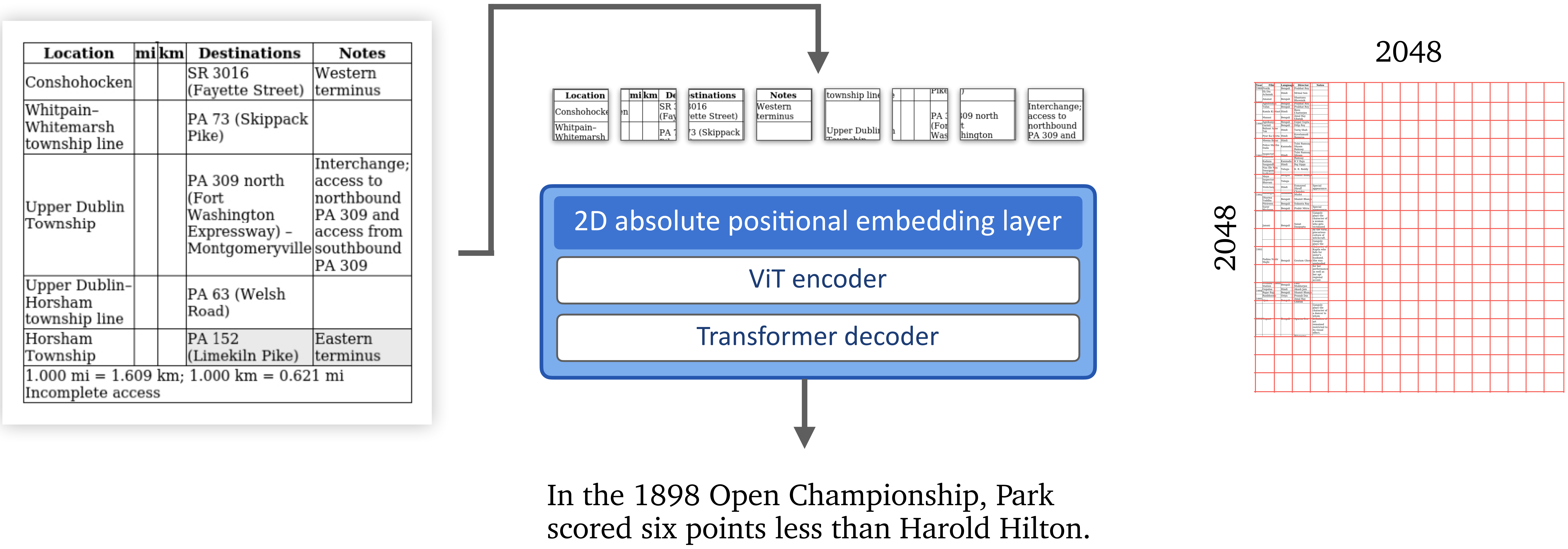
Location	mi	km	Destinations	Notes
Conshohocken			SR 3016 (Fayette Street)	Western terminus
Whitpain–Whitemarsh township line			PA 73 (Skippack Pike)	
Upper Dublin Township			PA 309 north (Fort Washington Expressway) – Montgomeryville	Interchange; access to northbound PA 309 and access from southbound PA 309
Upper Dublin–Horsham township line			PA 63 (Welsh Road)	
Horsham Township			PA 152 (Limekiln Pike)	Eastern terminus
1.000 mi = 1.609 km; 1.000 km = 0.621 mi Incomplete access				



In the 1898 Open Championship, Park scored six points less than Harold Hilton.



Architecture of **PixT3**



Location	mi	km	Destinations	Notes
Conshohocken			SR 3016 (Fayette Street)	Western terminus
Whitpain–Whitemarsh township line			PA 73 (Skippack Pike)	
Upper Dublin Township			PA 309 north (Fort Washington Expressway) – Montgomeryville	Interchange; access to northbound PA 309 and access from southbound PA 309
Upper Dublin–Horsham township line			PA 63 (Welsh Road)	
Horsham Township			PA 152 (Limekiln Pike)	Eastern terminus
1.000 mi = 1.609 km; 1.000 km = 0.621 mi				
Incomplete access				

Table Structure Awareness

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

False!

David Kinnell scored 154.

Table Structure Awareness

Title: 1898 Open Championship

Place	Player	Country	Score
1	Willie Park, Jr.	Scotland	151
2	Harry Vardon	Jersey	154
T3	Thomas Renouf	Jersey	156
	J.H. Taylor	England	156
T5	Harold Hilton	England	157
	David Kinnell	Scotland	157
T7	James Kinnell	Scotland	158
	Freddie Tait	Scotland	158
9	Sandy Herd	Scotland	159
10	David Herd	Scotland	160

False!

David Kinnell scored 154.

23%

unfaithful sentences due to structural errors

Structure Learning Curriculum

Table:

oY	io	HG	eG2S
Z4ikU	O1	aRU	mubk6
URa	dAF		I
I86	GAe	Ob	sUr5
L1	3	Vf1	Svaq2

Target:

<<<dAF><<<URa><I>>><<<io><O1><GAe>
<3>><<HG><aRU><Ob><Vf1>>>>>

Structure Learning Curriculum

Table:

oY	io	HG	eG2S
Z4ikU	01	aRU	mubk6
URa	dAF		I
I86	GAe	Ob	sUr5
L1	3	Vf1	Svaq2

Target:

<<<dAF>

Structure Learning Curriculum

Table:

oY	io	HG	eG2S
Z4ikU	O1	aRU	mubk6
URa	dAF		I
I86	GAe	Ob	sUr5
L1	3	Vf1	Svaq2

Target:

<<<dAF><<<URa><I>>>

Structure Learning Curriculum

Table:

oY	io	HG	eG2S
Z4ikU	01	aRU	mubk6
URa	dAF		I
I86	GAe	Ob	sUr5
L1	3	Vf1	Svaq2

Target:

<<<dAF><<<URa><I>>><<<io><01><GAe>
<3>><<HG><aRU><Ob><Vf1>>>>>

Structure Learning Curriculum

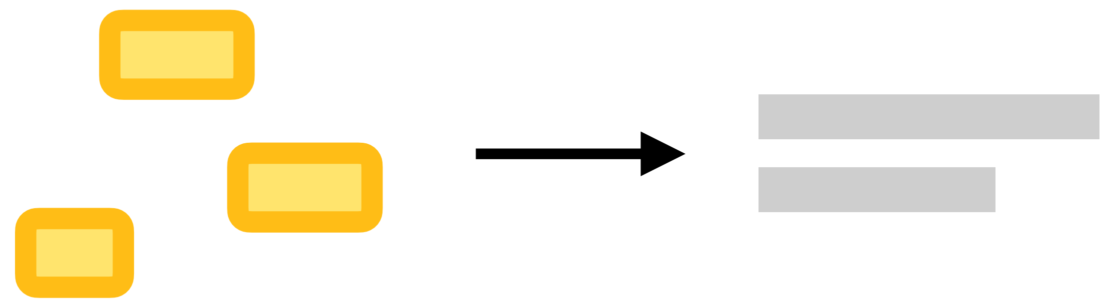
23% → 7%

69.6% reduction in structural faithfulness errors

Three evaluation settings

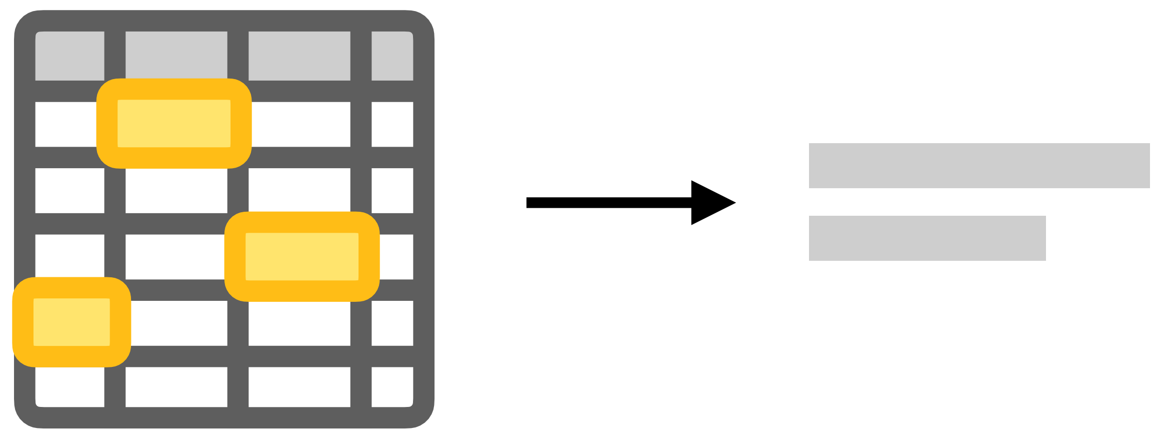
Tightly Control

Highlighted cells only



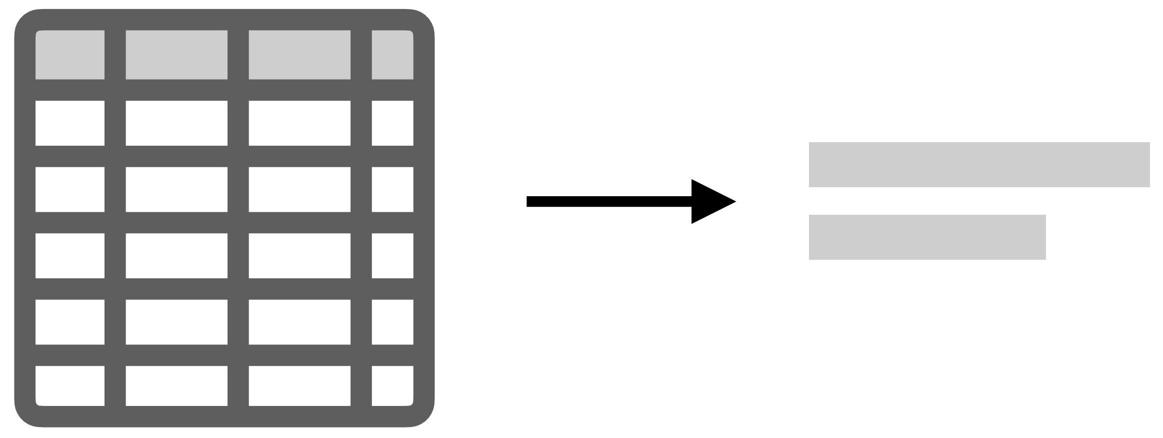
Loosely Control

Table + highlighted cells



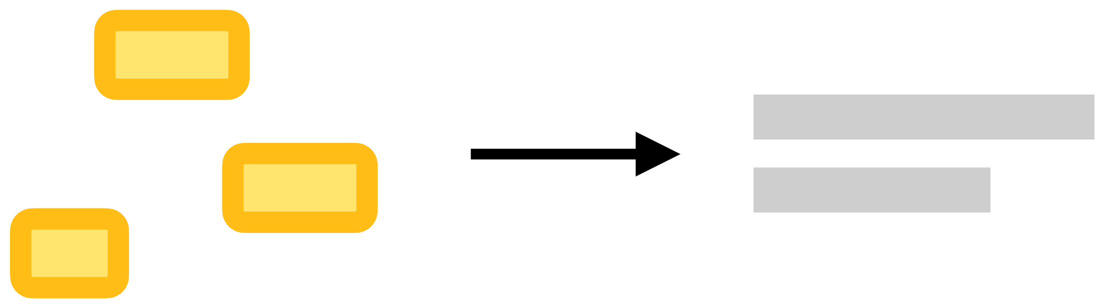
Open Ended

Table only

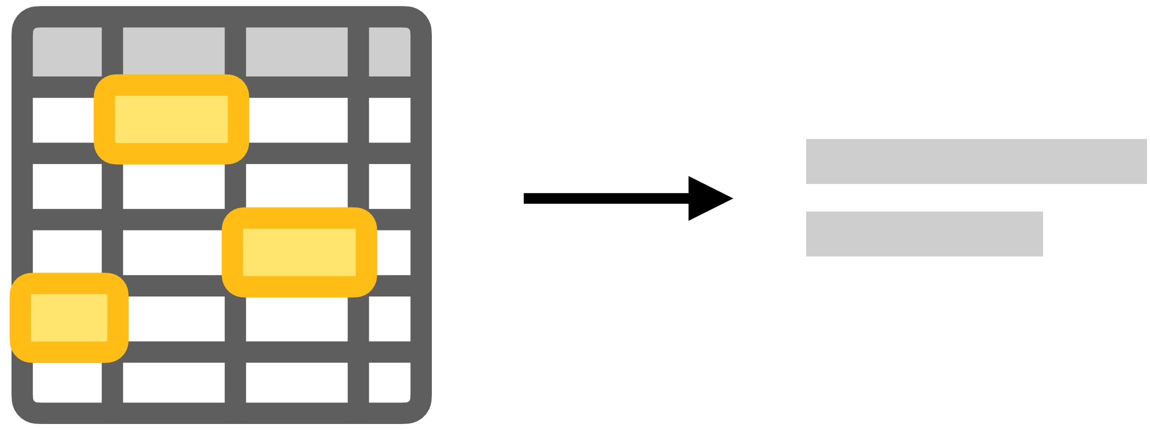


Three evaluation settings

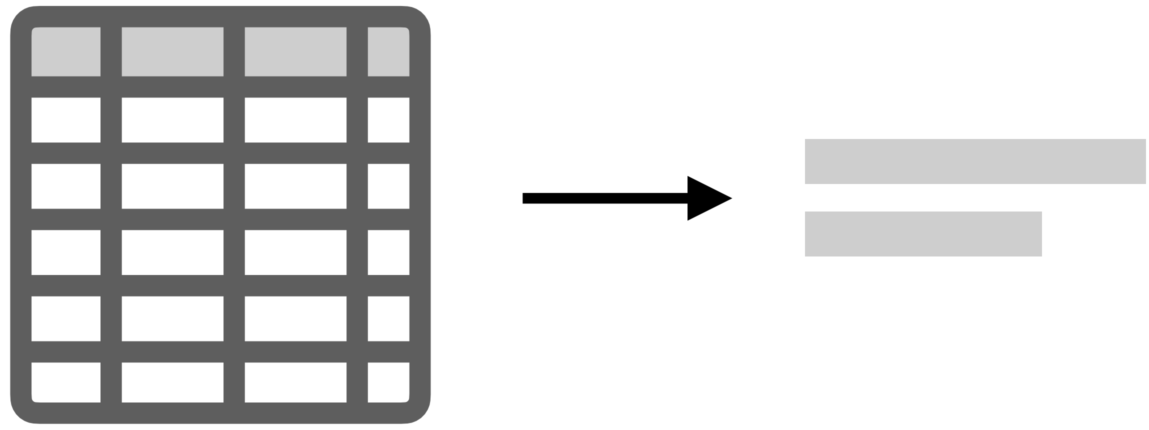
TControl
Highlighted cells only



LControl
Table + highlighted cells



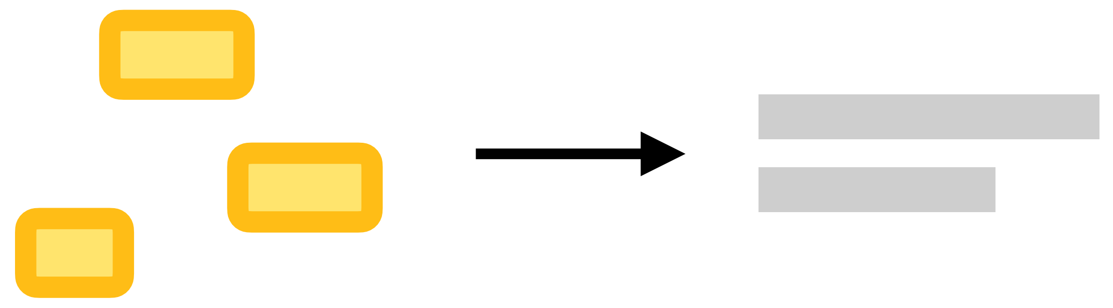
OpenE
Table only



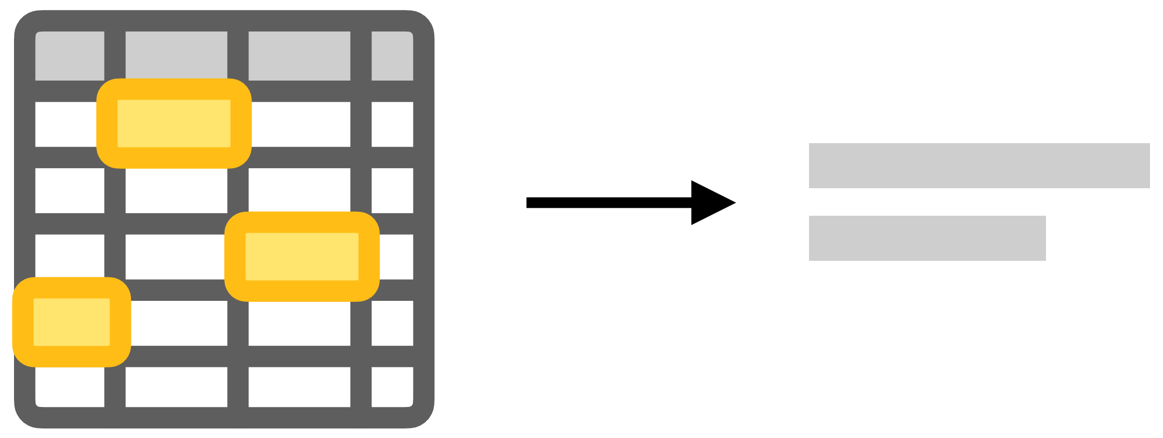
BLEU **PARENT** **Fidelity**

Three evaluation settings

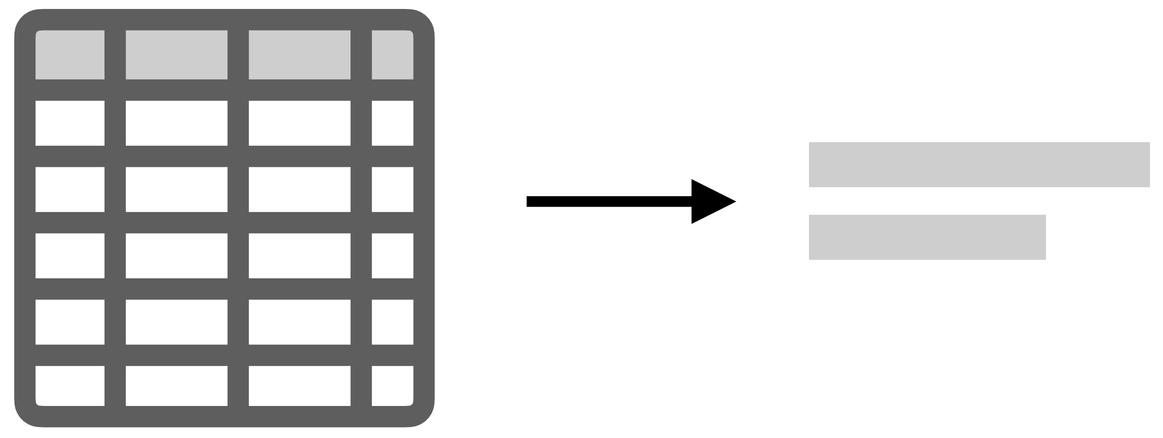
TControl
Highlighted cells only



LControl
Table + highlighted cells



OpenE
Table only

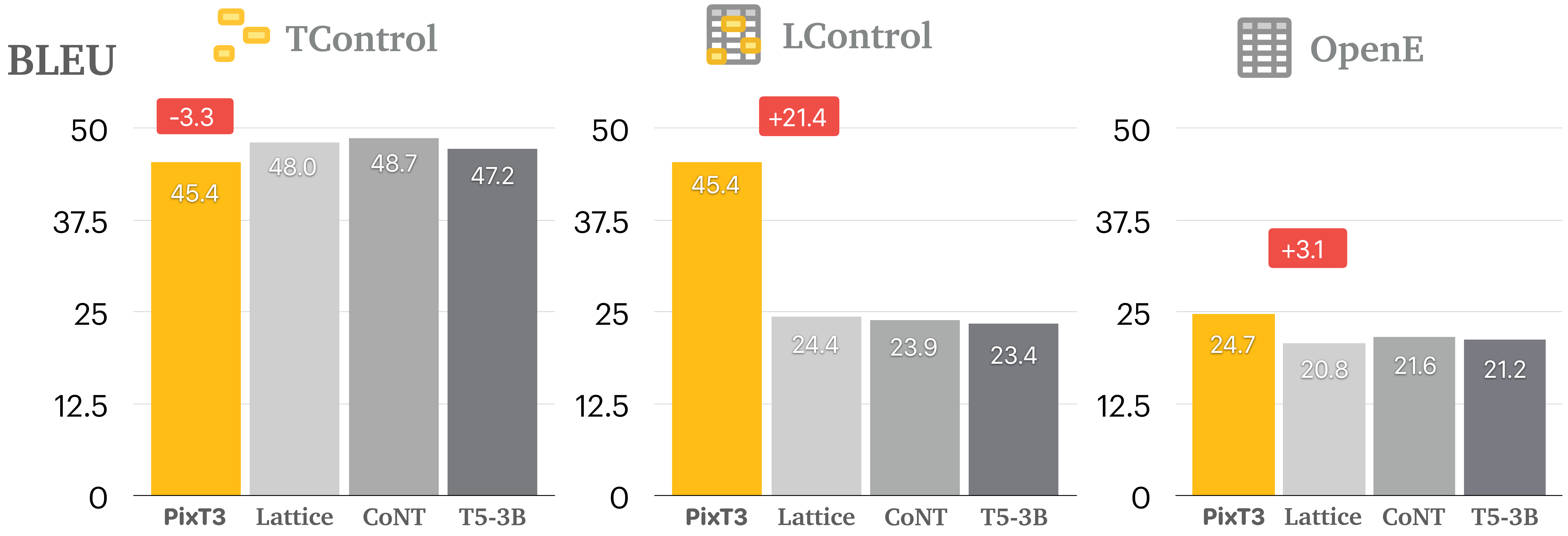


BLEU

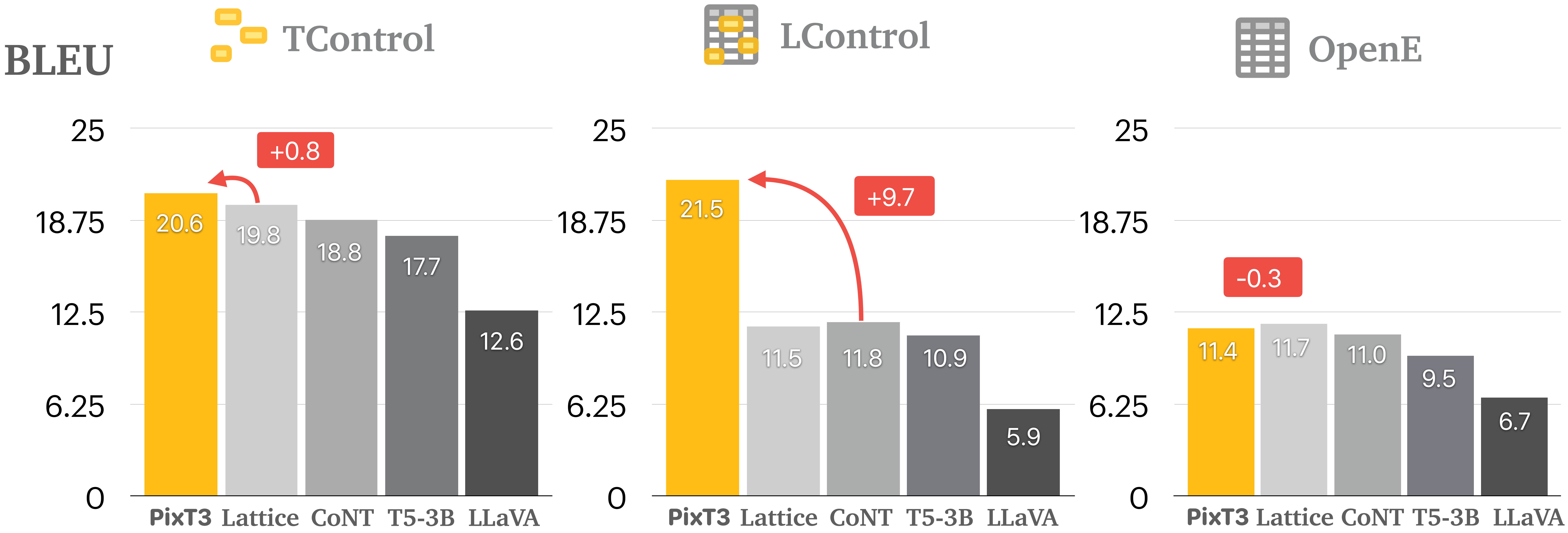
PARENT
(Dhingra et al., 2019)

Fidelity

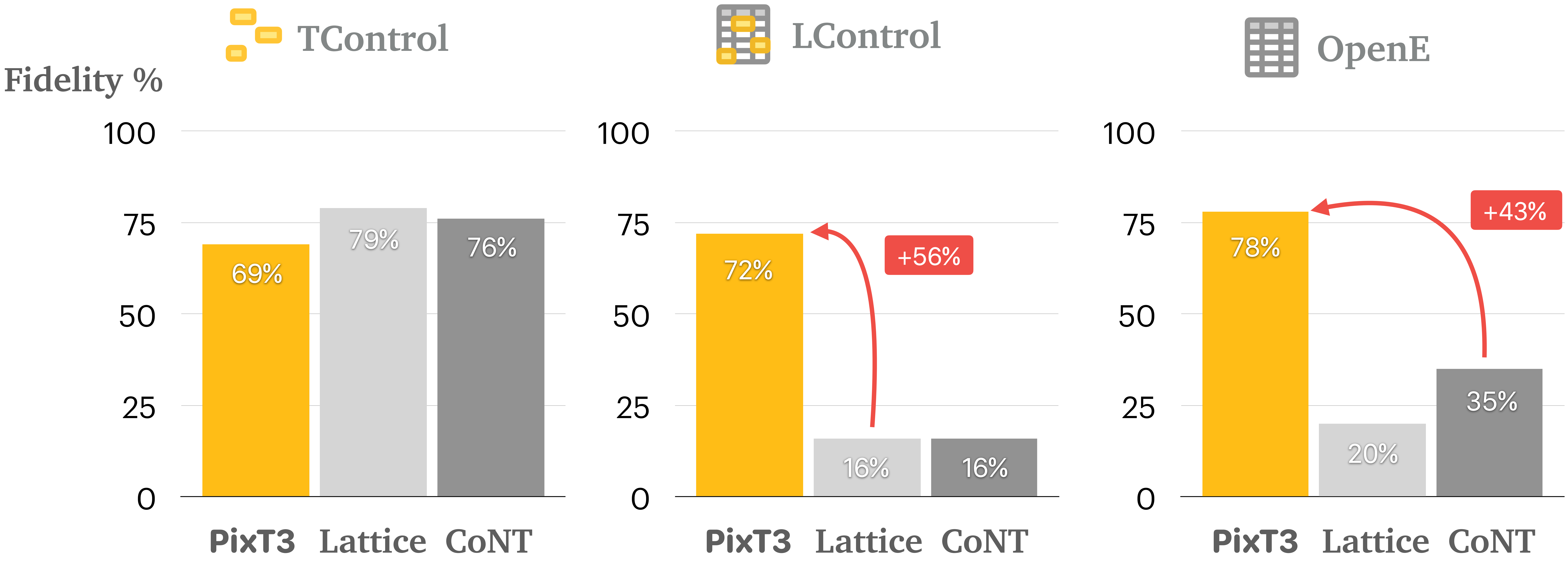
Automatic: In-domain *(ToTTo)*



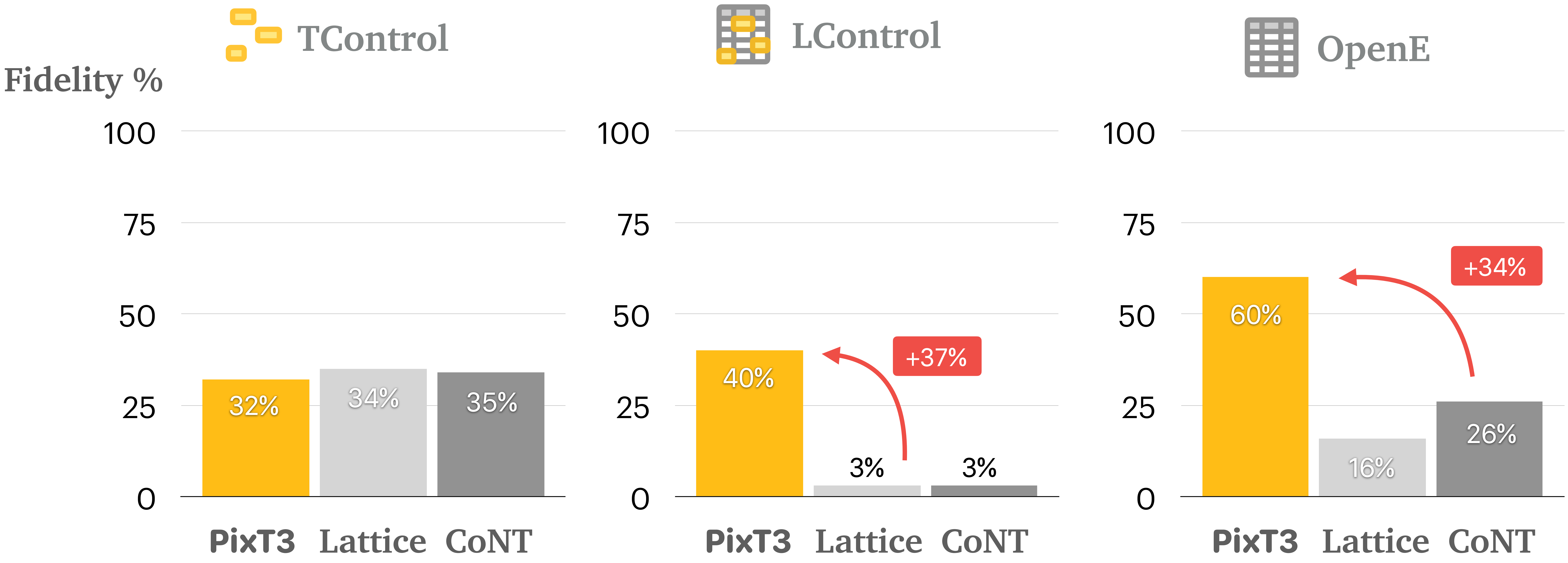
Automatic: Out-of-domain *(Logic2Text)*



Fidelity: In-domain *(ToTTo)*



Fidelity: Out-of-domain *(Logic2Text)*



Can Vision-Language Models perform Table-to-Text Generation?

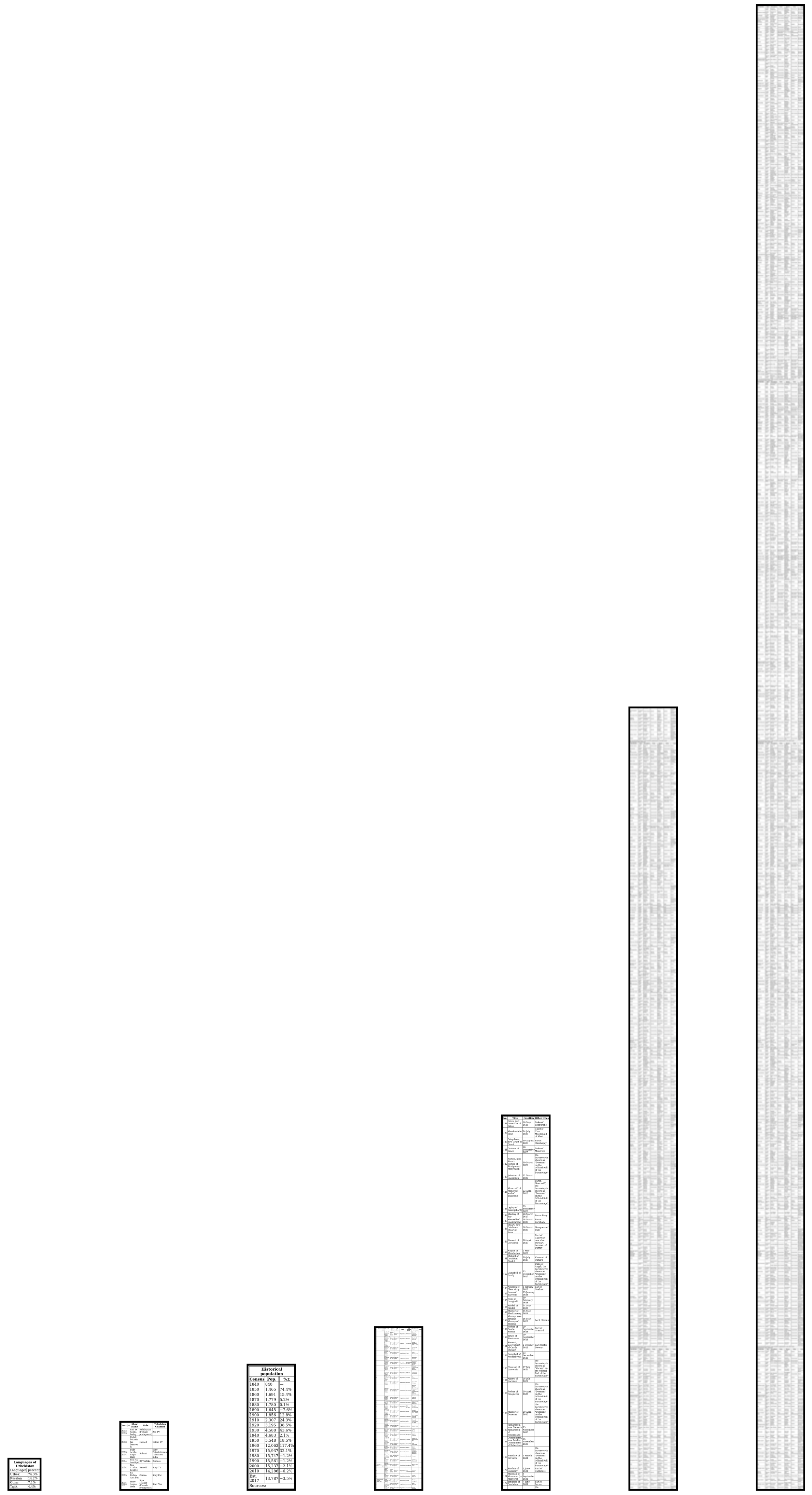
Yes

Can this approach maintain the same level of fidelity as its unimodal counterparts?

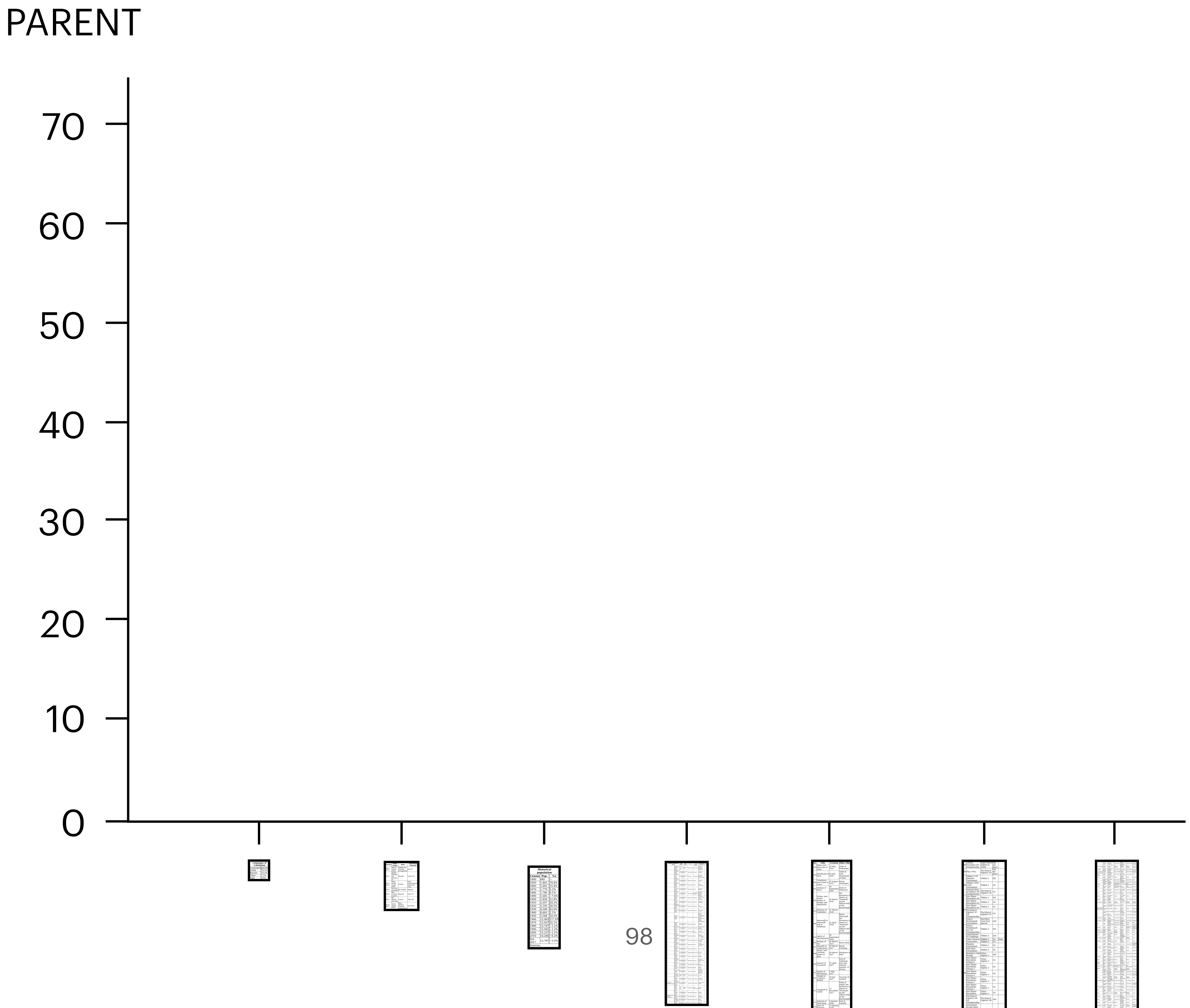
Yes

Are images a space-efficient modality for representing tables for Table-to-Text Generation?

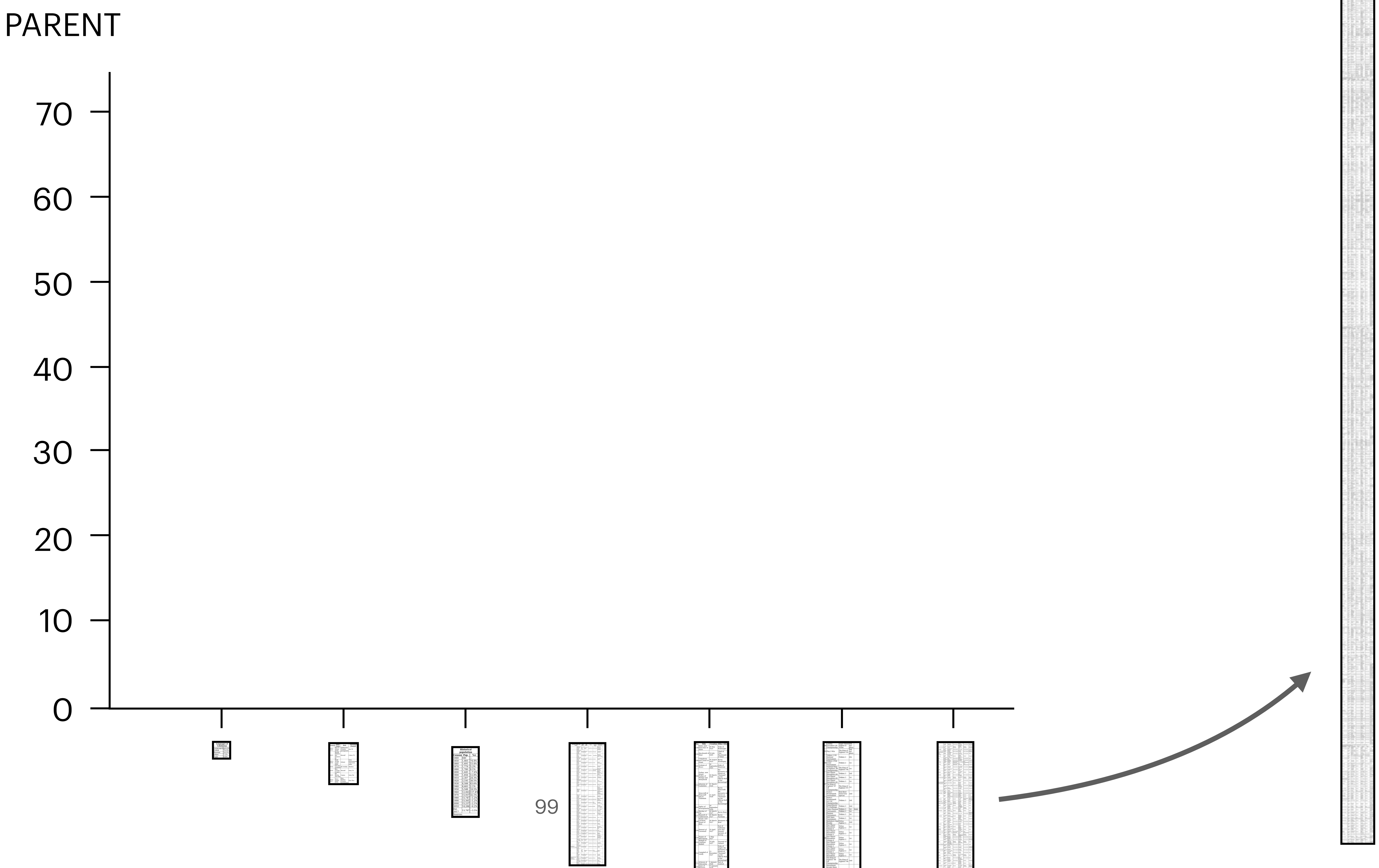
Performance degradation over size



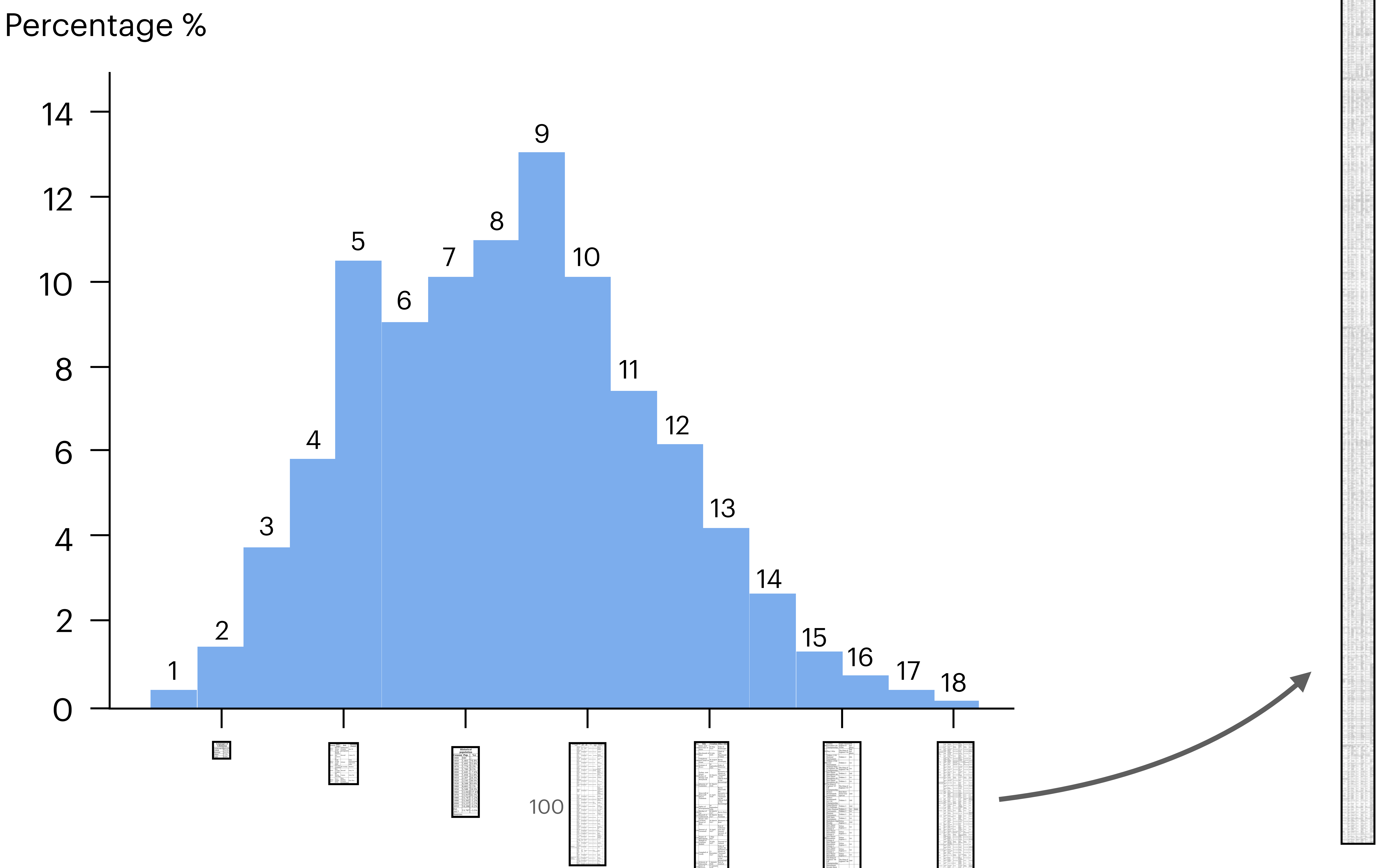
Performance degradation over size



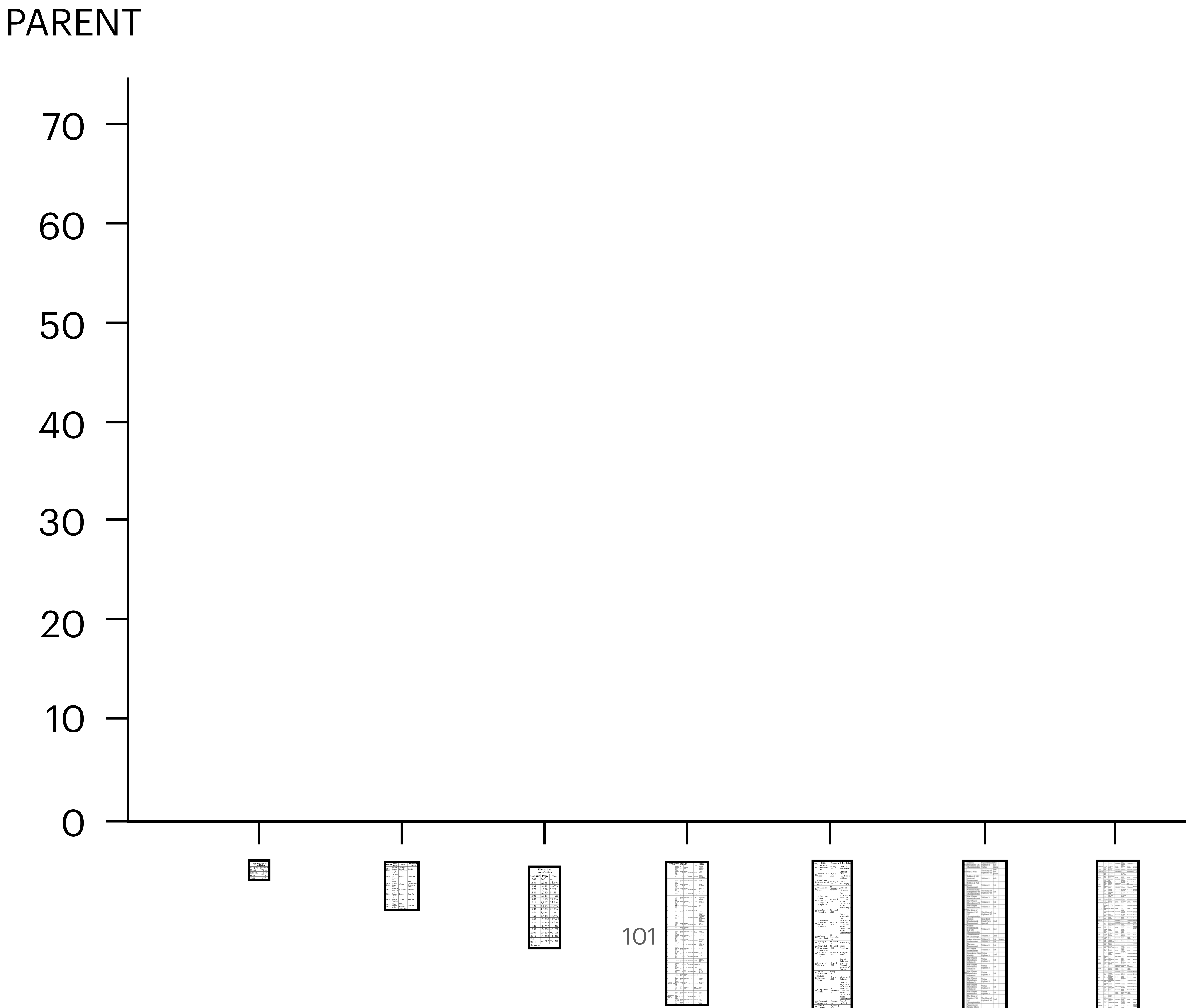
Performance degradation over size



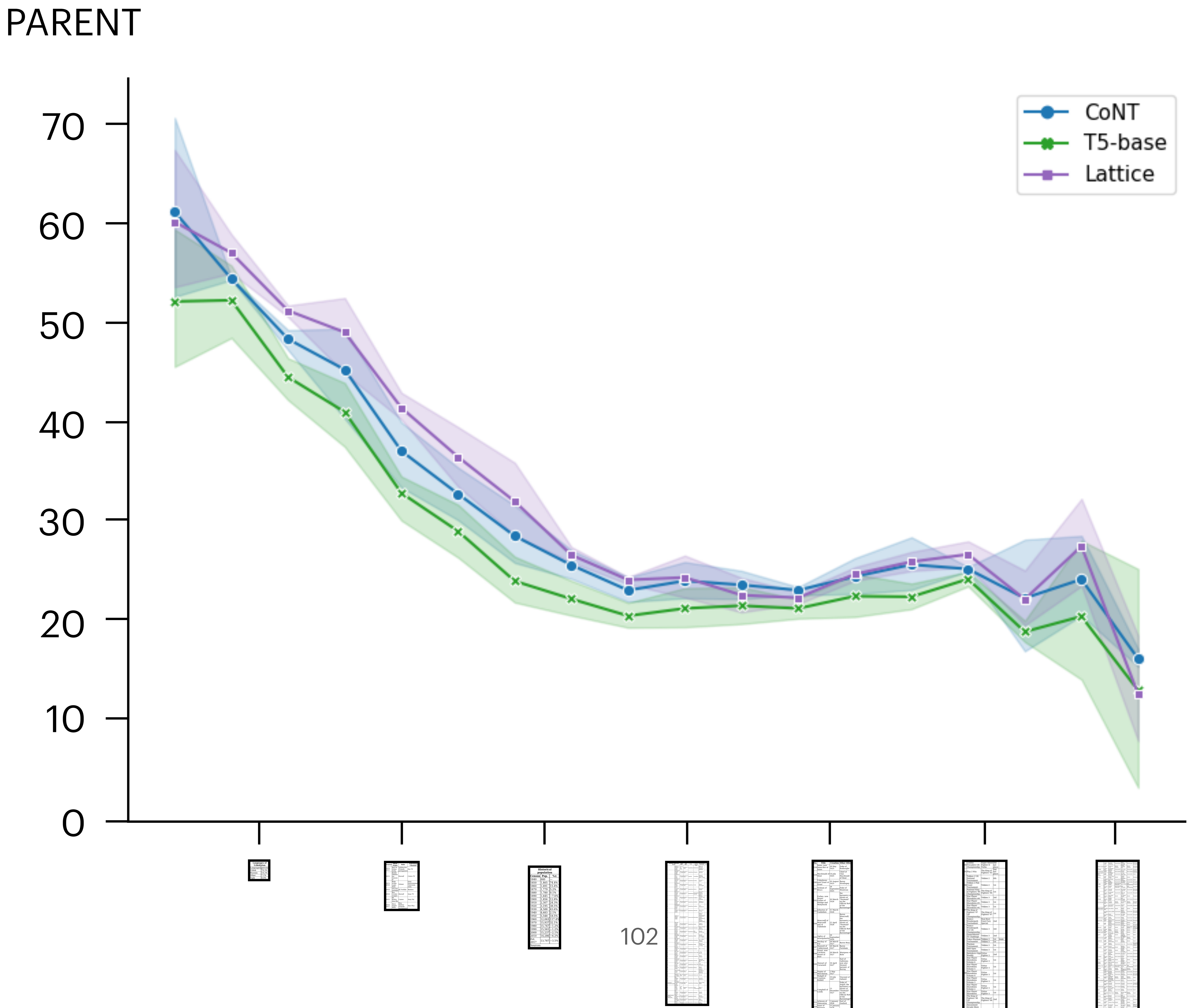
Performance degradation over size



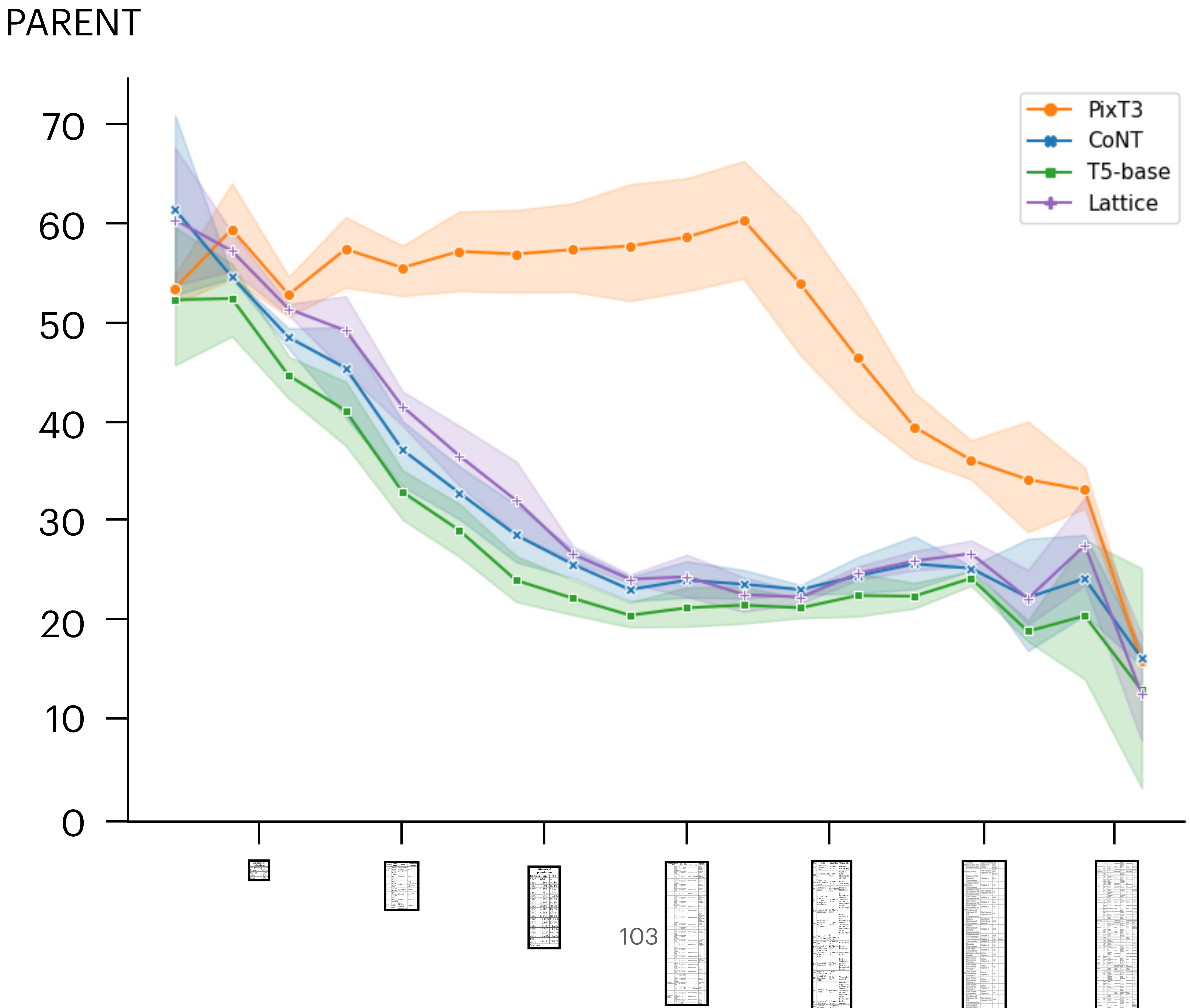
Performance degradation over size



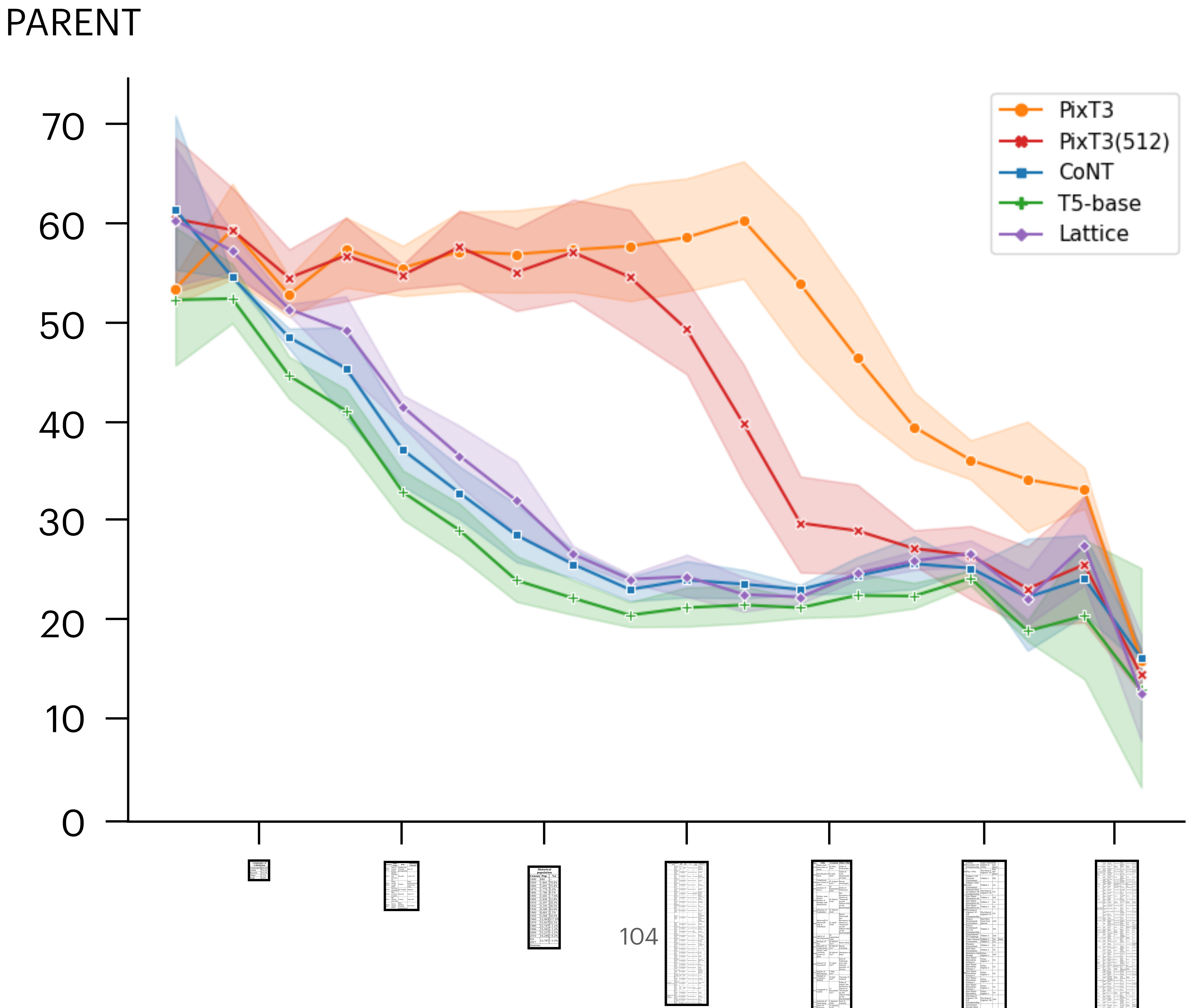
Performance degradation over size



Performance degradation over size



Performance degradation over size



Are images a space-efficient modality for representing tables for Table-to-Text Generation?

Yes

Conclusions

- PixT3 transforms table-to-text generation into a visual recognition task, eliminating the need to render input tables as strings.

- Our Structure Learning Curriculum improves the structural awareness of tables in our multimodal table-to-text models.

- PixT3 performs competitively and often surpasses state-of-the-art models across various table sizes and domains, showcasing less degradation on large tables.

PixT3: Pixel-based Table-To-Text Generation

Iñigo Alonso,
HiTZ Center - Ixa,
University of the
Basque Country UPV/EHU
inigoborja.alonso@ehu.eus

Eneko Agirre
HiTZ Center - Ixa,
University of the
Basque Country UPV/EHU
e.agirre@ehu.eus

Mirella Lapata
Institute for Language,
Cognition and Computation,
University of Edinburgh
mlap@inf.ed.ac.uk

Abstract

Table-to-text generation involves generating appropriate textual descriptions given structured tabular data. It has attracted increasing attention in recent years thanks to the popularity of neural network models and the availability of large-scale datasets. A common feature across existing methods is their treatment of the input as a string, i.e., by employing linearization techniques that do not always preserve information in the table, are verbose, and lack space efficiency. We propose to rethink data-to-text generation as a visual recognition task, removing the need for rendering the input in a string format. We present PixT3, a multimodal table-to-text model that overcomes the challenges of linearization and input size limitations encountered by existing models. PixT3 is trained with a new self-supervised learning objective to reinforce table structure awareness and is applicable to open-ended *and* controlled generation settings. Experiments on the ToTTo (Parikh et al., 2020a) and Logic2Text (Chen et al., 2020c) benchmarks show that PixT3 is competitive and, in some settings, superior to generators that operate solely on text.[1]

1 Introduction

Generating text from structured inputs such as tables, tuples, or graphs, is commonly referred to as data-to-text generation (Reiter and Dale, 1997; Covington, 2001; Gatt and Krahmer, 2018). This umbrella term includes several tasks ranging from generating sport summaries based on boxscore statistics (Wiseman et al., 2017), to producing fun facts from superlative Wikipedia tables (Korn et al., 2019), and creating textual descriptions given biographical data (Lebret et al., 2016). From a modeling perspective, data-to-text generation is challenging as it is not immediately obvious how to best describe the given input. For instance, the table in Figure 1 can be verbalized in different ways, depending on the specific content we choose to focus on. In *controlled* data-to-text generation (Parikh et al., 2020a), models are expected to generate descriptions for pre-selected parts of the input (see the *highlighted* cells in Figure 1).

Regardless of the generation setting, numerous approaches have emerged in recent years with different characteristics. A few exploit the structural information of the input (Puduppully et al., 2019; Chen et al., 2020b; Wang et al., 2022), use neural templates (Wiseman et al., 2018), or resort to content planning (Su et al., 2021; Puduppully et al., 2022). While others (Chen et al., 2020a,c; Aghajanyan et al., 2022; Kasner and Dusek, 2022) improve on fluency and generalization by leveraging large-scale pre-trained language models (Devlin et al., 2019; Raffel et al., 2020). A common feature across these methods is their treatment of tabular input as a string, following various linearization methods. As an example, Figure 1 shows the representation of tabular data (top) as a sequence of (Column, Row, Value) tuples (bottom).

Problematically, representing tabular information as a linear sequence results in a verbose representation that often exceeds the context window limit of popular Transformer models (Vaswani et al., 2017). The challenge of processing such long sequences has fostered the development of even more controlled methods which refrain from encoding the table as a whole, concentrating exclusively on highlighted content (e.g., *only* the yellow cells in Figure 1). Unfortunately, models trained on abridged input have difficulty generalizing to new domains while being practically ineffective in scenarios where content selection is not provided.

In this paper we propose to rethink data-to-text generation as a visual recognition task, allowing us to represent and preserve tabular information compactly. Vision Transformers (ViTs; Dosovitskiy et al. 2021) have significantly advanced

[1] Our code, models, and data are available at https://github.com/alonsoapp/PixT3.

Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 6721–6736
August 11-16, 2024 ©2024 Association for Computational Linguistics

Proceedings of the 62nd Annual Meeting of the ACL

PixT3: Pixel-based Table-To-Text Generation

Iñigo Alonso, Eneko Agirre, and Mirella Lapata

eman ta zabal zazu

Universidad del País Vasco | Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Edinburgh NLP
University of Edinburgh
Natural Language Processing

Beyond Table-to-Text

Lossless Table Visualisations Enhance Multimodal Table Understanding

Beyond Table-to-Text Generation

Lossless Table Visualisations Enhance Multimodal Table Understanding

Table Understanding Tasks

Table-to-Text Generation

Table Understanding Tasks

Table Question Answering

Table-to-Text Generation

Table Understanding Tasks

Table Question Answering

Table Fact Verification

Table-to-Text Generation

Table Understanding Tasks

Table Question Answering

Table Fact Verification

Table-to-Text Generation

Table Numerical Reasoning

Table Understanding Tasks

Table Question Answering

Table Fact Verification

Table-to-Text Generation

Table Numerical Reasoning

Column Type Annotation

Table Understanding Tasks

Entity Linking

Table Question Answering

Table Fact Verification

Key-Value Pair Natural Language Inference

Hybrid Question Answering

Relation Extraction

Table-to-Text Generation

Free-form Table Question Answering

Hierarchical Table Question Answering

Loosely Controlled Table-to-Text

Open Ended Table-to-Text

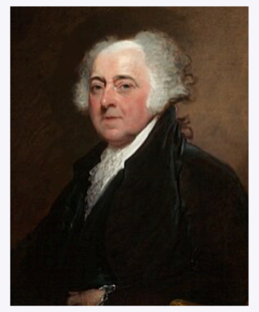
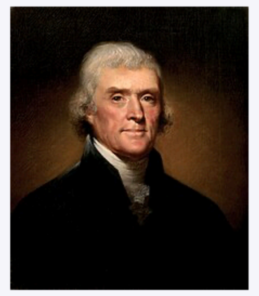
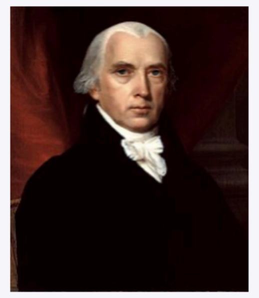
Table Numerical Reasoning

Column Type Annotation

Structure Aware Parsing

Multimodal Table Understanding Dataset

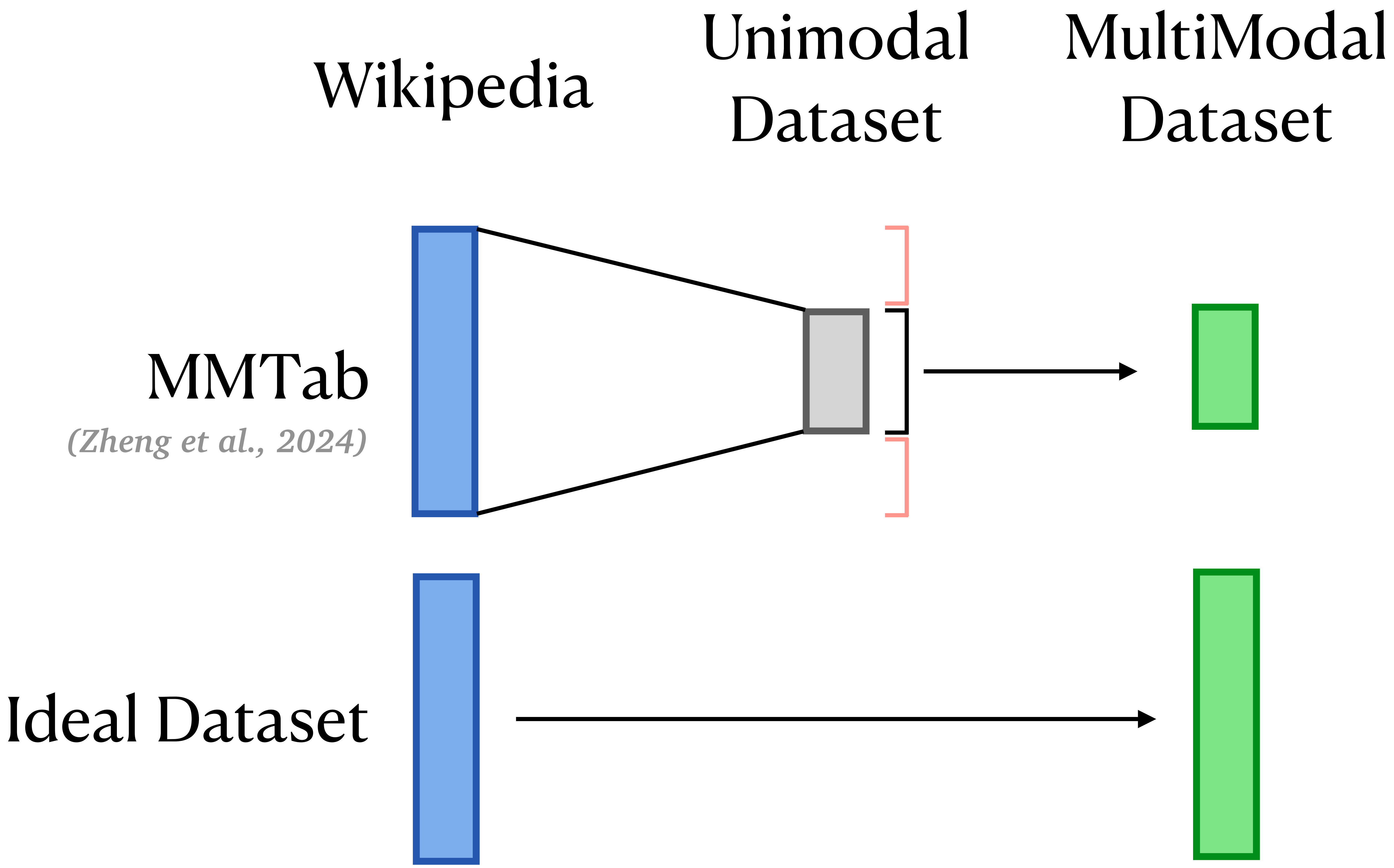
Current multimodal table datasets are lossy

No. [a]	Portrait	Name (birth–death)	Term[16]	Party[b][17]	Election
1		George Washington (1732–1799) [19]	April 30, 1789 – March 4, 1797	*Unaffiliated*	1788–89 1792
2		John Adams (1735–1826) [21]	March 4, 1797 – March 4, 1801	Federalist	1796
3		Thomas Jefferson (1743–1826) [23]	March 4, 1801 – March 4, 1809	Democratic-Republican	1800 1804
4		James Madison (1751–1836) [24]	March 4, 1809 – March 4, 1817	Democratic-Republican	1808 1812
5		James Monroe (1758–1831) [26]	March 4, 1817 – March 4, 1825	Democratic-Republican	1816 1820

<page_title> 1898 Open Championship <page_title> <table> <row> <cell> Place </cell> <cell> Player <row_header> Place </row_header> </cell> <cell> Country <row_header> Place </row_header> <row_header> Player </row_header> </cell

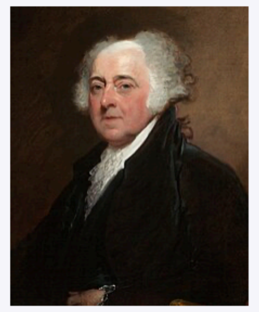
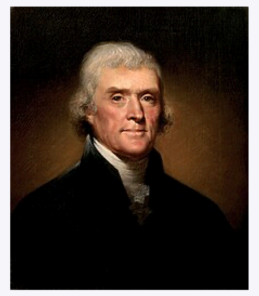
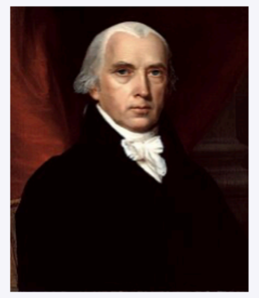
	Name	Took Office	Left Office	Party
1	Vasso Papandreou	21 February 1996	19 February 1999	Panhellenic Socialist Movement
2	Evangelos Venizelos	19 February 1999	13 April 2000	Panhellenic Socialist Movement
3	Nikos Christodoulakis	13 April 2000	24 October 2001	Panhellenic Socialist Movement
4	Akis Tsochatzopoulos	24 October 2001	10 March 2004	Panhellenic Socialist Movement
5	Dimitris Sioufas	10 March 2004	19 September 2007	New Democracy
6	Christos Folias	19 September 2007	8 January 2009	New Democracy
7	Kostis Hatzidakis	8 January 2009	7 October 2009	New Democracy

MMTab (Zheng et al., 2024)

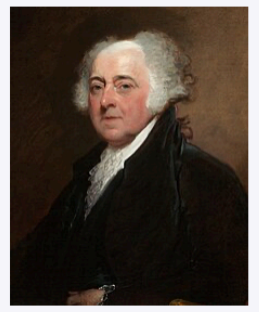
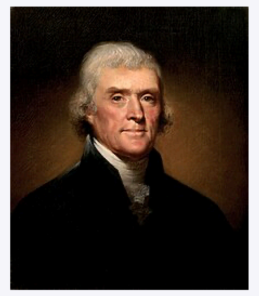
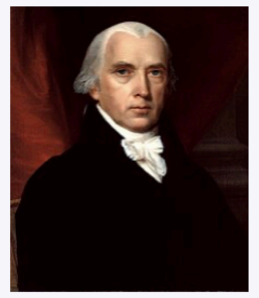


Information in raw source tables
Information in serialised tables
Information in image tables in multimodal datasets

Our Multimodal Table Understanding Dataset

No. [a]	Portrait	Name (birth–death)	Term[16]	Party[b][17]	Election
1		George Washington (1732–1799) [19]	April 30, 1789 – March 4, 1797	*Unaffiliated*	1788–89 1792
2		John Adams (1735–1826) [21]	March 4, 1797 – March 4, 1801	Federalist	1796
3		Thomas Jefferson (1743–1826) [23]	March 4, 1801 – March 4, 1809	Democratic-Republican	1800 1804
4		James Madison (1751–1836) [24]	March 4, 1809 – March 4, 1817	Democratic-Republican	1808 1812
5		James Monroe (1758–1831) [26]	March 4, 1817 – March 4, 1825	Democratic-Republican	1816 1820

Collected from the source

No. [a]	Portrait	Name (birth–death)	Term[16]	Party[b][17]	Election
1		George Washington (1732–1799) [19]	April 30, 1789 – March 4, 1797	*Unaffiliated*	1788–89 1792
2		John Adams (1735–1826) [21]	March 4, 1797 – March 4, 1801	Federalist	1796
3		Thomas Jefferson (1743–1826) [23]	March 4, 1801 – March 4, 1809	Democratic-Republican	1800 1804
4		James Madison (1751–1836) [24]	March 4, 1809 – March 4, 1817	Democratic-Republican	1808 1812
5		James Monroe (1758–1831) [26]	March 4, 1817 – March 4, 1825	Democratic-Republican	1816 1820

Instruction datasets

TableInstruct (Zhang et al., 2024)
DocStruct4M (Hu et al., 2024)
MMTab (Zheng et al., 2024)

Seed datasets

TURL (Deng et al., 2020)
ToTTo (Parikh et al., 2020)
TabFact (Chen et al., 2020b)
WikiTab-QA (Pasupat and Liang, 2015)
HybridQA (Chen et al., 2020c)
NSF (National Science Foundation, 2019)
StatCan (Statistics Canada, 2024)
PubTabNet (Zhong et al., 2020)
TABMWP (Lu et al., 2023)
TAT-QA (Zhu et al., 2021)
InfoTabs (Gupta et al., 2020)

Instruction datasets

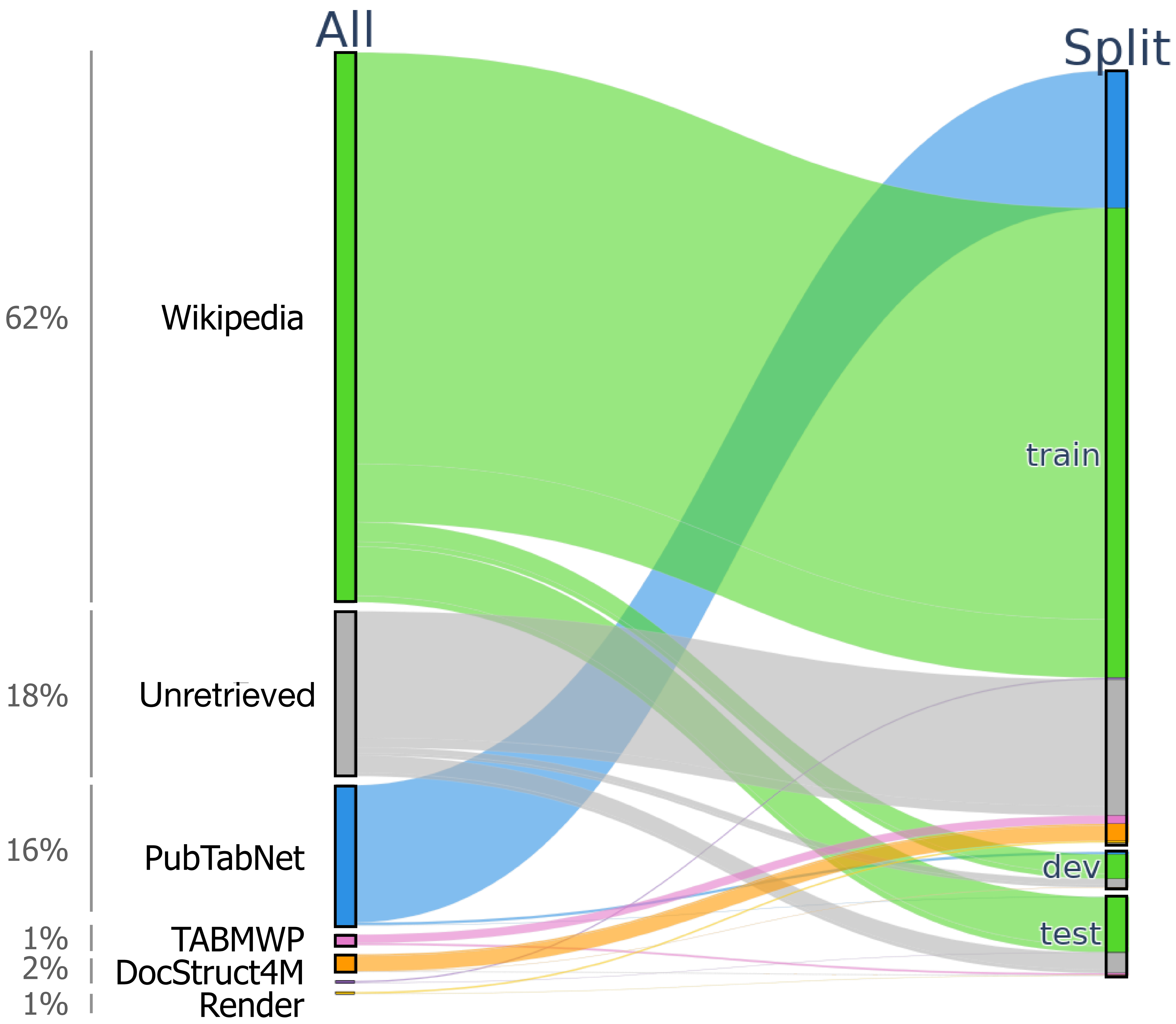
TableInstruct (Zhang et al., 2024)
DocStruct4M (Hu et al., 2024)
MMTab (Zheng et al., 2024)

2.55M

Instruction examples

1.15M

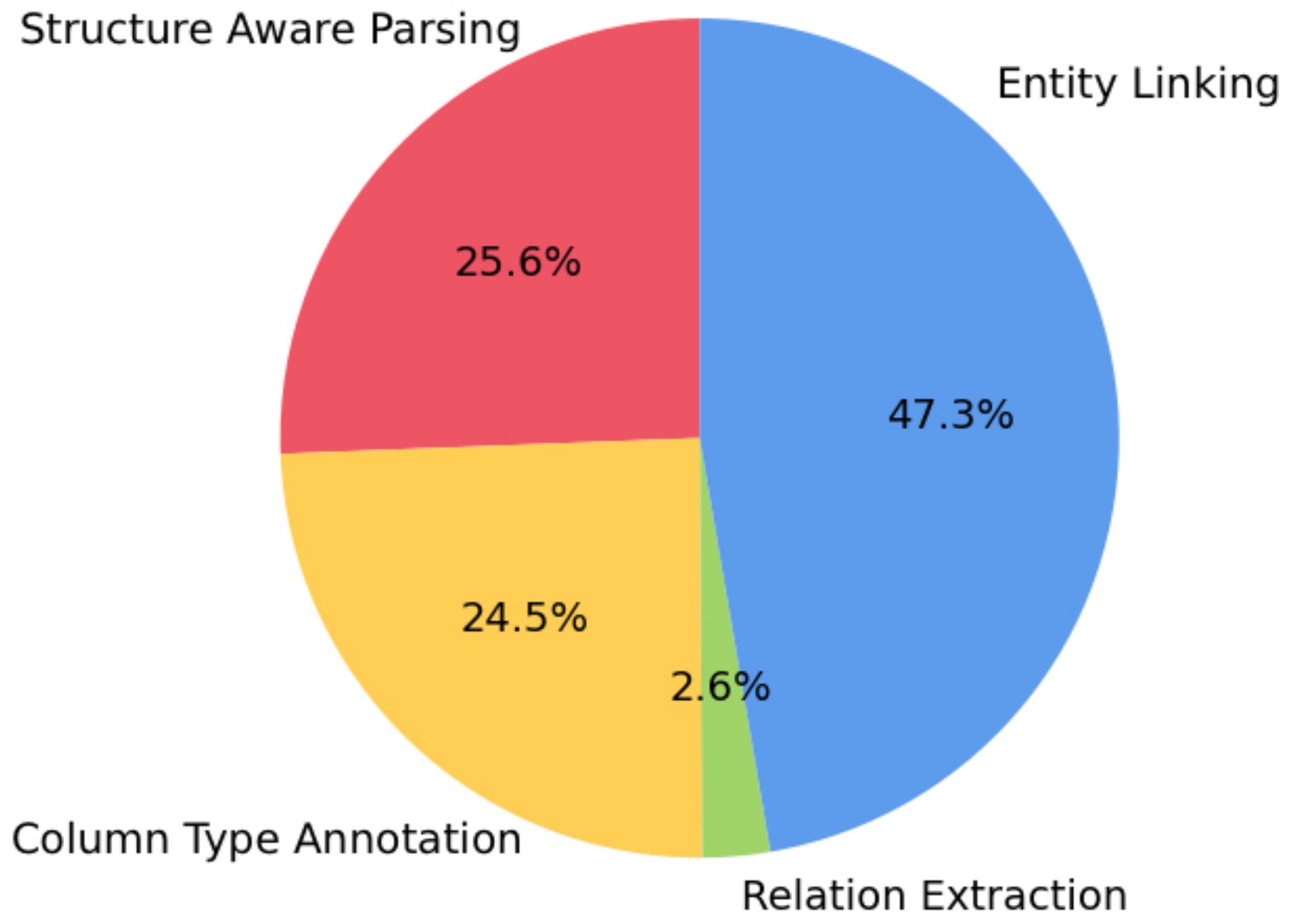
Table images



Tasks

Stage 1

Column Type Annotation
Entity Linking
Structure Aware Parsing
Relation Extraction



Stage 2

FeTaQA *(Free-form TabQA)*
HiTab *(Hierarchical TableQA)*
Table Numerical Reasoning *(Table-Reasoning)*
TabFact *(Table Fact Verification)*
Infotabs *(Table Fact Verification)*
ToTTo *(Table-to-Text)*
HybridQA *(Hybrid TableQA)*
WikiTableQuestions *(TableQA)*

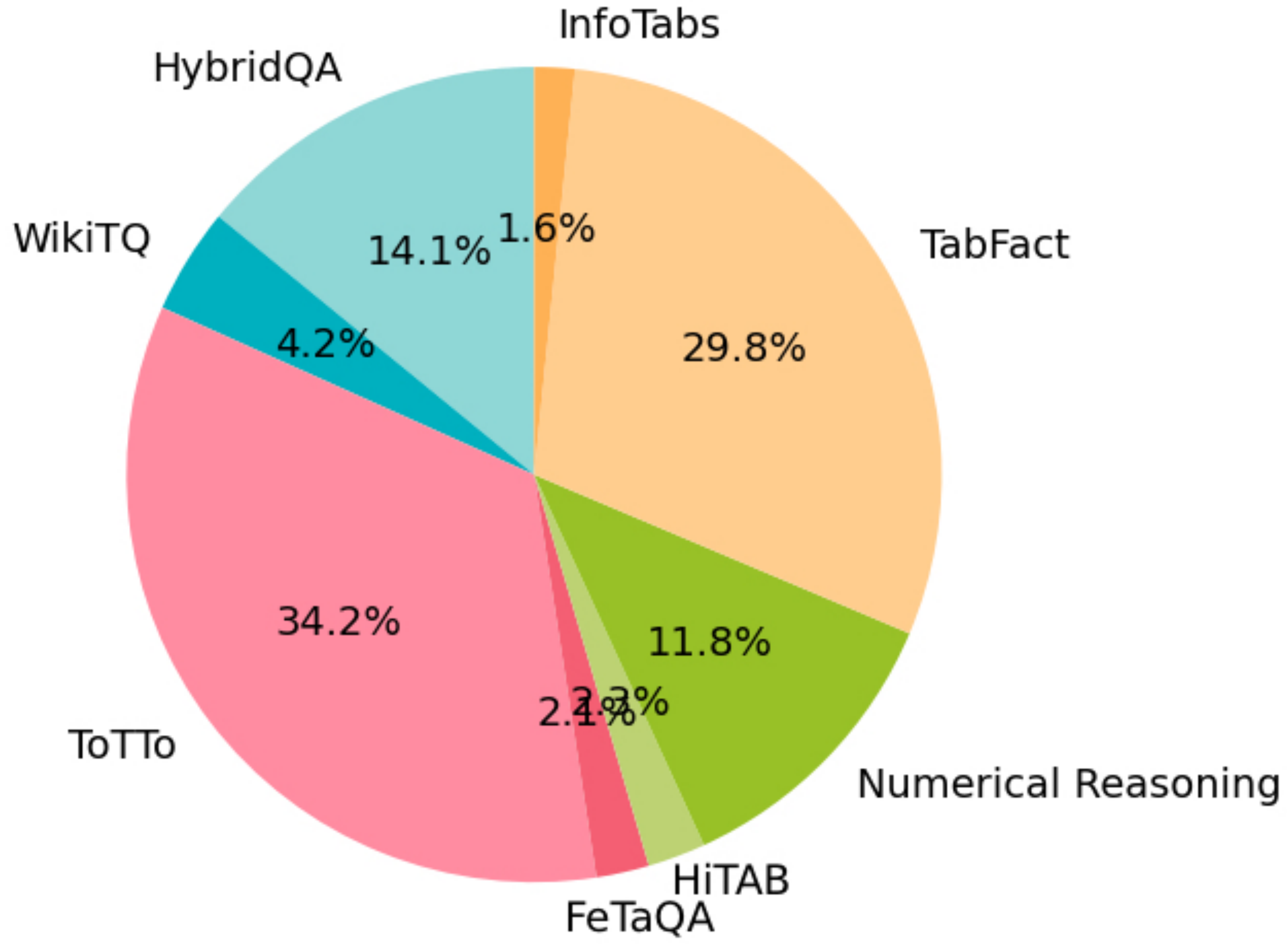
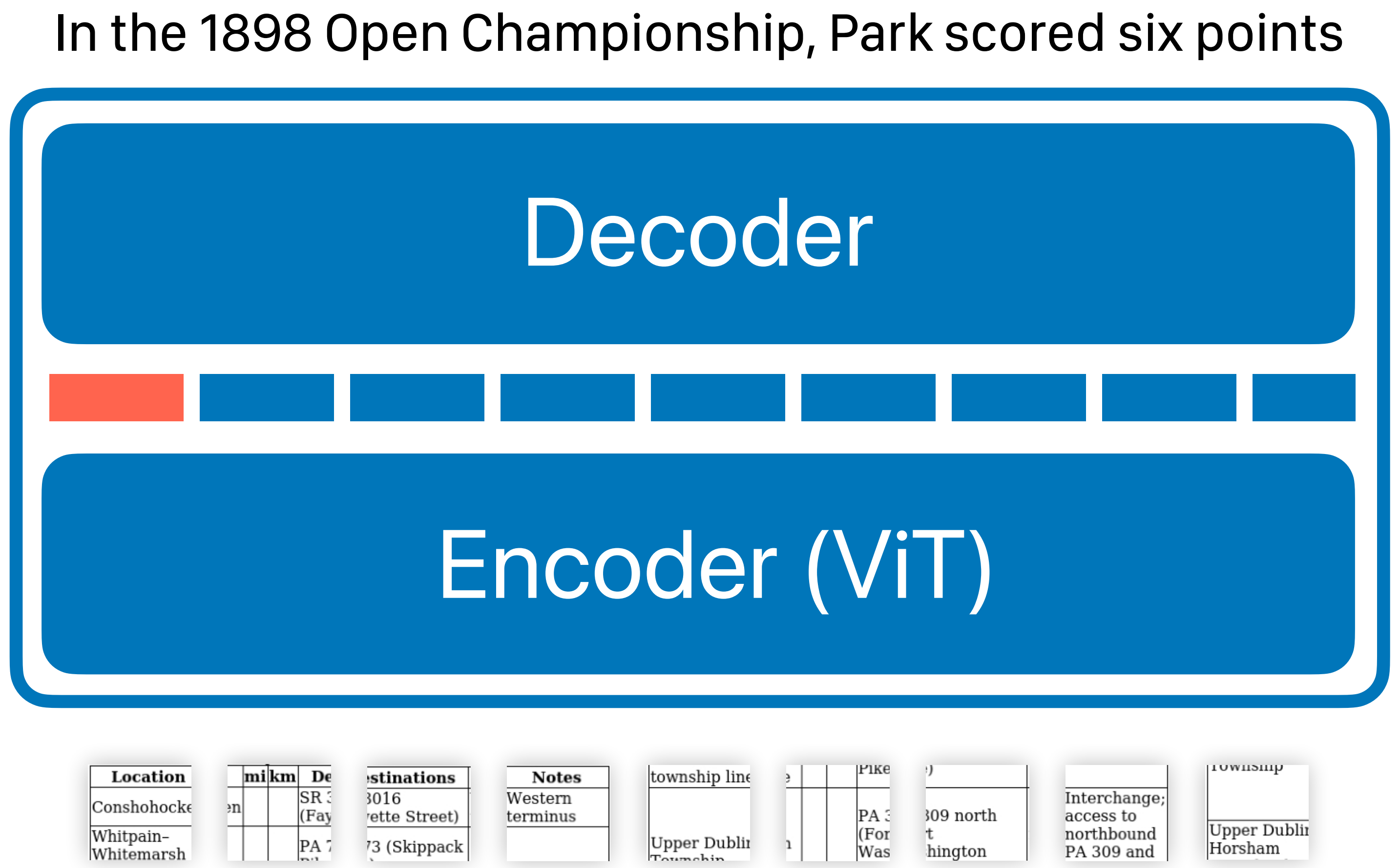


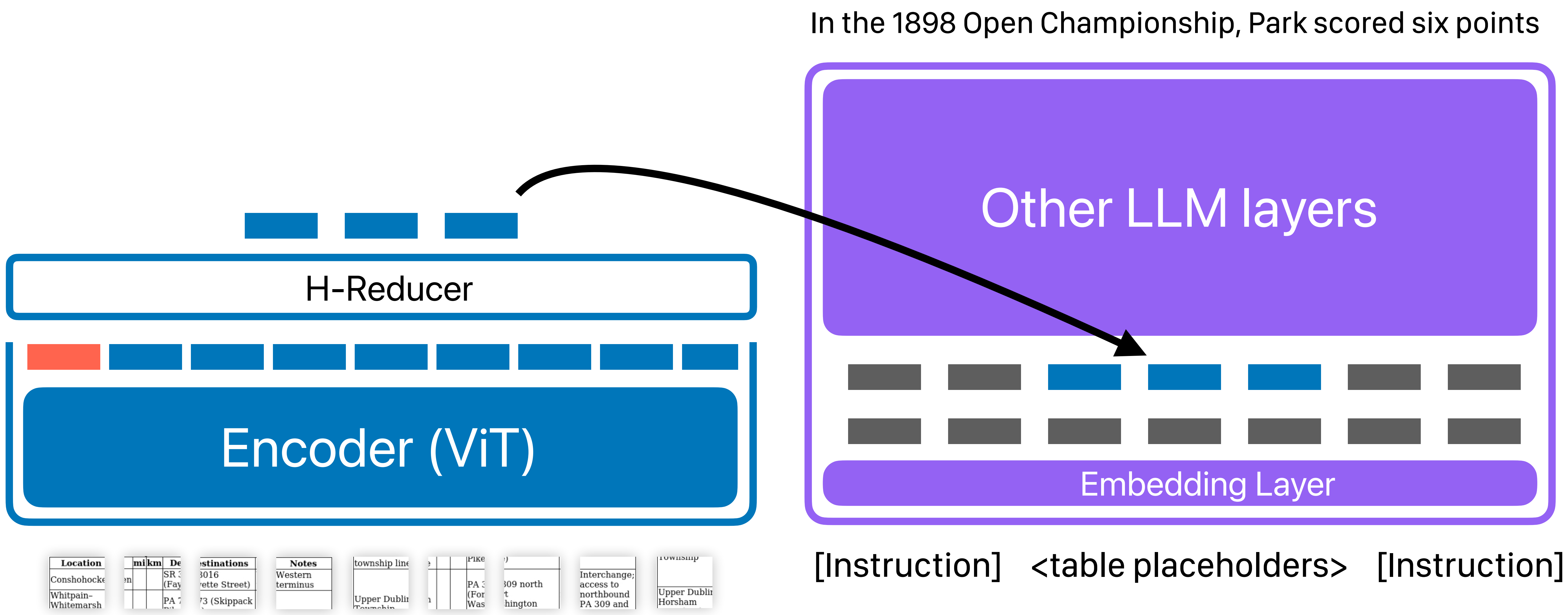
Table Understanding Vision-Language Model

PixT3



mPLUG-DocOwl 1.5

(Hu et al., 2024)



mPLUG-DocOwl 1.5

Baseline	Stage 1 (Hu et al., 2024)	+	Stage 2 (Hu et al., 2024)
Ours	Stage 1 (Hu et al., 2024)	+	Stage 2 **(Our dataset)**

mPLUG-DocOwl 1.5

Baseline: Stage 1 (Hu et al., 2024) + Stage 2 (Hu et al., 2024)

Ours: Stage 1 (Hu et al., 2024) + Stage 2 **(Our dataset)**

Results

Model	FeTaQA	HiTab	HybridQA	InfoTabs	TabFact	TaBMWP	TAT-QA	ToTTo	WikiTQ
Baseline	2.5*	17.6*	35.5*	29.9*	68.3	10.9*	12.7*	10.1*	**33.7**
Ours	**66.0**	**41.9**	**50.7**	**60.2**	**72.9**	**86.2**	**43.7**	**41.6**	32.2

* held-out dataset

mPLUG-DocOwl 1.5

Ours

Stage 1
(Hu et al., 2024)

+

Stage 2
(Our dataset)

Results

Model	FeTaQA	HiTab	HyQA	InfoTabs	TabFact	TaBMWP	TATQA	ToTTo	WikiTQ
DocOwl1.5 (Ours)	**66.0**	**41.9**	**50.7**	60.2	**72.9**	**86.2**	**43.7**	**41.6**	**32.2**
Table-LLaVA (7B)	25.8	10.4	35.6*	**63.0**	53.7	57.9	16.7	26.1	11.1

* held-out dataset

mPLUG-DocOwl 1.5

Ours

Stage 1
(Hu et al., 2024)

+

Stage 2
(Our dataset)

Results

Model	FeTaQA	HiTab	HyQA	InfoTabs	TabFact	TaBMWP	TATQA	ToTTo	WikiTQ
DocOwl1.5 (Ours)	**66.0**	**41.9**	**50.7**	60.2	**72.9**	**86.2**	**43.7**	**41.6**	**32.2**
Table-LLaVA (7B)	25.8	10.4	35.6*	**63.0**	53.7	57.9	16.7	26.1	11.1

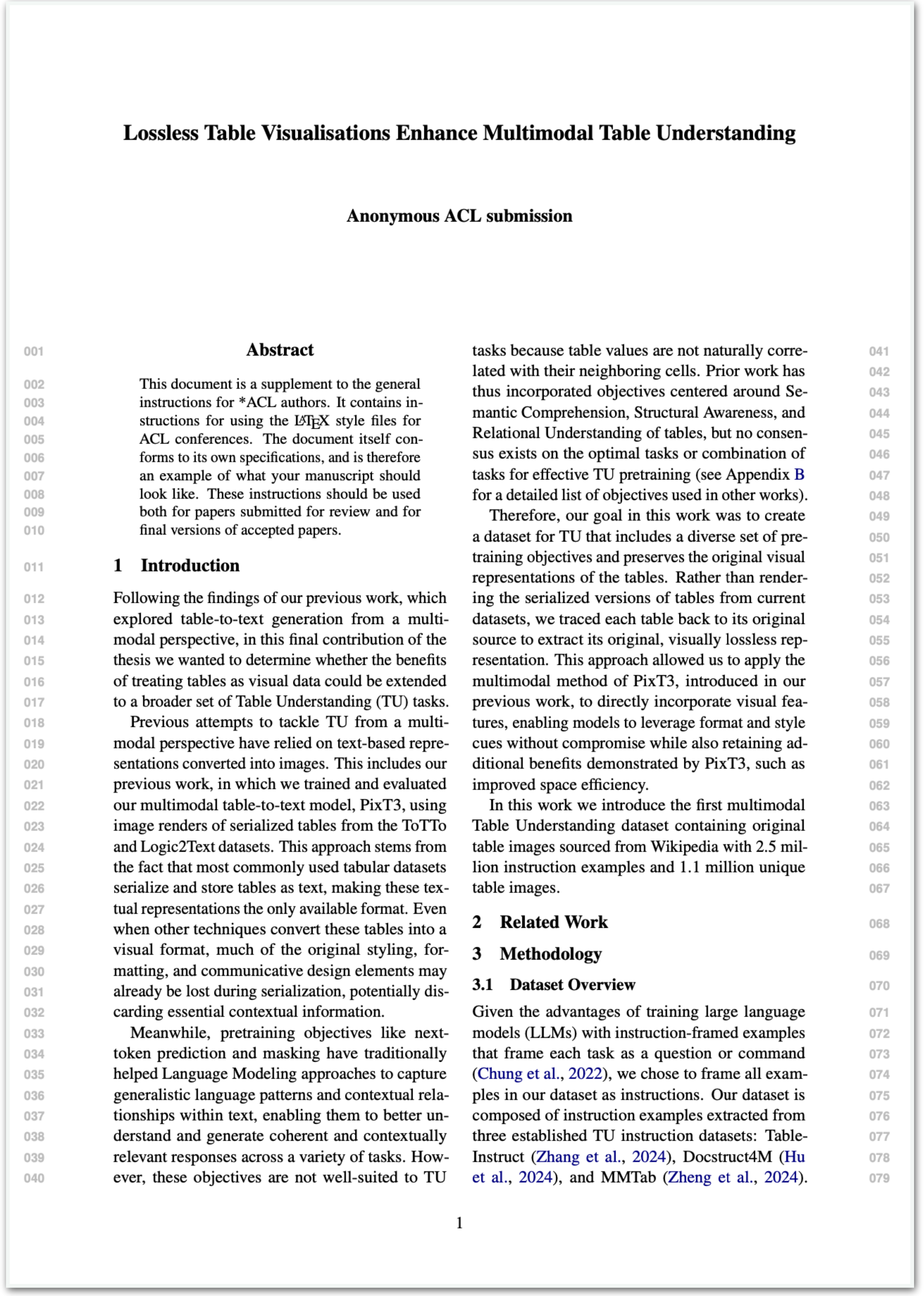
* held-out dataset

Model	FeTaQA	HiTab	HyQA	InfoTabs	TabFact	TaBMWP	TATQA	ToTTo	WikiTQ
DocOwl1.5 (Ours)	**66.0**	41.9	**50.7**	**60.2**	72.9	**86.2**	**43.7**	**41.6**	**32.2**
TableLlama	39.1	**59.8**	36.5*	10.2*	**82.9**	11.2*	6.3*	21.5*	17.1*

* held-out dataset

Conclusions

- Largest multimodal Table Understanding dataset at the time of writing with 2.5M examples.
- First multimodal Table Understanding dataset focused on original table visualisations including 1.1 original table images.
- High quality Stage 2 subset enables baseline VLM to outperform current state-of-the-art VLMs across a diverse set of tasks.



Lossless Table Visualisations Enhance Multimodal Table Understanding

Anonymous ACL submission

Abstract

This document is a supplement to the general instructions for *ACL authors. It contains instructions for using the LaTeX style files for ACL conferences. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used both for papers submitted for review and for final versions of accepted papers.

1 Introduction

Following the findings of our previous work, which explored table-to-text generation from a multimodal perspective, in this final contribution of the thesis we wanted to determine whether the benefits of treating tables as visual data could be extended to a broader set of Table Understanding (TU) tasks.

Previous attempts to tackle TU from a multimodal perspective have relied on text-based representations converted into images. This includes our previous work, in which we trained and evaluated our multimodal table-to-text model, PixT3, using image renders of serialized tables from the ToTTo and Logic2Text datasets. This approach stems from the fact that most commonly used tabular datasets serialize and store tables as text, making these textual representations the only available format. Even when other techniques convert these tables into a visual format, much of the original styling, formatting, and communicative design elements may already be lost during serialization, potentially discarding essential contextual information.

Meanwhile, pretraining objectives like next-token prediction and masking have traditionally helped Language Modeling approaches to capture generalistic language patterns and contextual relationships within text, enabling them to better understand and generate coherent and contextually relevant responses across a variety of tasks. However, these objectives are not well-suited to TU tasks because table values are not naturally correlated with their neighboring cells. Prior work has thus incorporated objectives centered around Semantic Comprehension, Structural Awareness, and Relational Understanding of tables, but no consensus exists on the optimal tasks or combination of tasks for effective TU pretraining (see Appendix B for a detailed list of objectives used in other works).

Therefore, our goal in this work was to create a dataset for TU that includes a diverse set of pretraining objectives and preserves the original visual representations of the tables. Rather than rendering the serialized versions of tables from current datasets, we traced each table back to its original source to extract its original, visually lossless representation. This approach allowed us to apply the multimodal method of PixT3, introduced in our previous work, to directly incorporate visual features, enabling models to leverage format and style cues without compromise while also retaining additional benefits demonstrated by PixT3, such as improved space efficiency.

In this work we introduce the first multimodal Table Understanding dataset containing original table images sourced from Wikipedia with 2.5 million instruction examples and 1.1 million unique table images.

2 Related Work

3 Methodology

3.1 Dataset Overview

Given the advantages of training large language models (LLMs) with instruction-framed examples that frame each task as a question or command (Chung et al., 2022), we chose to frame all examples in our dataset as instructions. Our dataset is composed of instruction examples extracted from three established TU instruction datasets: Table-Instruct (Zhang et al., 2024), Docstruct4M (Hu et al., 2024), and MMTab (Zheng et al., 2024).

1

In preparation...

Lossless Table Visualisations Enhance Multimodal Table Understanding

Iñigo Alonso, Imanol Miranda, Eneko Agirre, and Mirella Lapata

eman ta zabal zazu

Universidad del País Vasco | Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Edinburgh NLP
University of Edinburgh
Natural Language Processing

Fidelity

Automatic Logical Forms improve fidelity in Table-to-Text generation

Representation

Pixel-based Table-To-Text Generation

Beyond Table-to-Text

Lossless Table Visualisations Enhance Multimodal Table Understanding

Future Work

Fidelity

Extend Logical Forms to irregular tables

Representation

Explore how Vision Language Models process tabular data

Table Understanding

Explore the full potential of our dataset with custom VLM architectures

Publications beyond this thesis

Artificial Intelligence in Medicine

MedExpQA: Multilingual Benchmarking of Large Language Models for Medical Question Answering.

Iñigo Alonso, Maite Oronoz, Rodrigo Agerri

Submitted to ACL 2025

Vision-Language Models Struggle to Align Entities across Modalities.

Iñigo Alonso, Ander Salaberria, Gorka Azkune, Jeremy Barnes, Oier Lopez de Lacalle

Improving Fidelity and Table Representation in Table Understanding and Table-to-Text Generation

a PhD thesis by

Iñigo Alonso

Supervised by

Eneko Agirre

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology