



The E2E Dataset

- Restaurant information domain
- Collected via **CrowdFlower**
 - using **pictorial instructions** to get more diverse texts



Flat MR

```
name[Loch Fyne],
eatType[restaurant],
food[French],
priceRange[less than £20],
familyFriendly[yes]
```

NL reference

Loch Fyne is a family-friendly restaurant providing wine and cheese at a low cost.

Loch Fyne is a French family friendly restaurant catering to a budget of below £20.

Loch Fyne is a French restaurant with a family setting and perfect on the wallet.

- More challenging than previous NLG datasets

E2E NLG Challenge

- take part!
- download the data at: <http://bit.ly/e2e-nlg>
- submit results by **October 15**

Lexical Richness

- Higher **diversity** (Mean Segmental Type-Token Ratio)
- Higher proportion of **rare words** (Lexical Sophistication)
- A challenge to word-based generators that don't handle **out-of-vocabulary** words

Dataset	Tokens	Types	LS	TTR	MSTTR
E2E	65,710	945	0.57	0.01	0.75
SFRest	45,791	1,187	0.43	0.03	0.62
Bagel	1,071	70	0.42	0.04	0.41

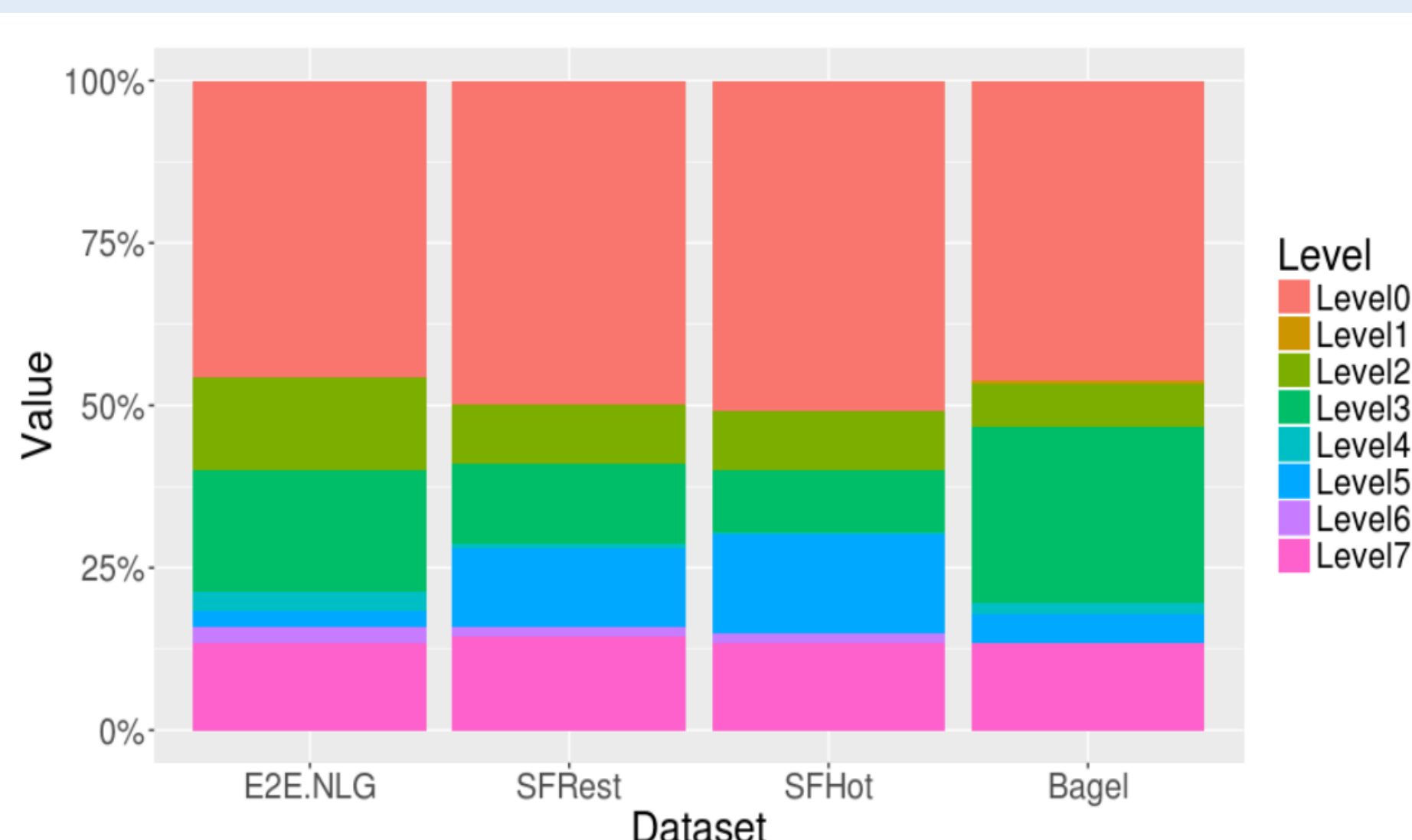
Dataset Size

- Significantly **more data**, more references per MR → more attractive
- **Larger number of sentences** per MR → more challenging
- **Longer** sentences → more challenging

	No. of instances	No. of unique MRs	Refs/MR	Slots/MR	W/Ref	W/Sent	Sents/Ref
E2E	50,602	5,751	8.1 (2–16)	5.43	20.1	14.3	1.5 (1–6)
SFRest	5,192	1,950	1.82 (1–101)	2.86	8.53	8.53	1.05 (1–4)
Bagel	404	202	2 (2–2)	5.41	11.54	11.54	1.02 (1–2)

Syntactic Variation

- Slightly fewer syntactically simple sentences
- Slightly more sentences with complex syntactic structure



The Vaults is an Indian restaurant.
 Cocum is a very expensive restaurant but the quality is great.
 The coffee shop Wildwood has fairly priced food, while being in the same vicinity as the Ranch.
 Serving cheap English food, as well as having a coffee shop, the Golden Palace has an average customer rating and is located along the riverside.

Semantic Adequacy

More challenging – requires **content selection**:

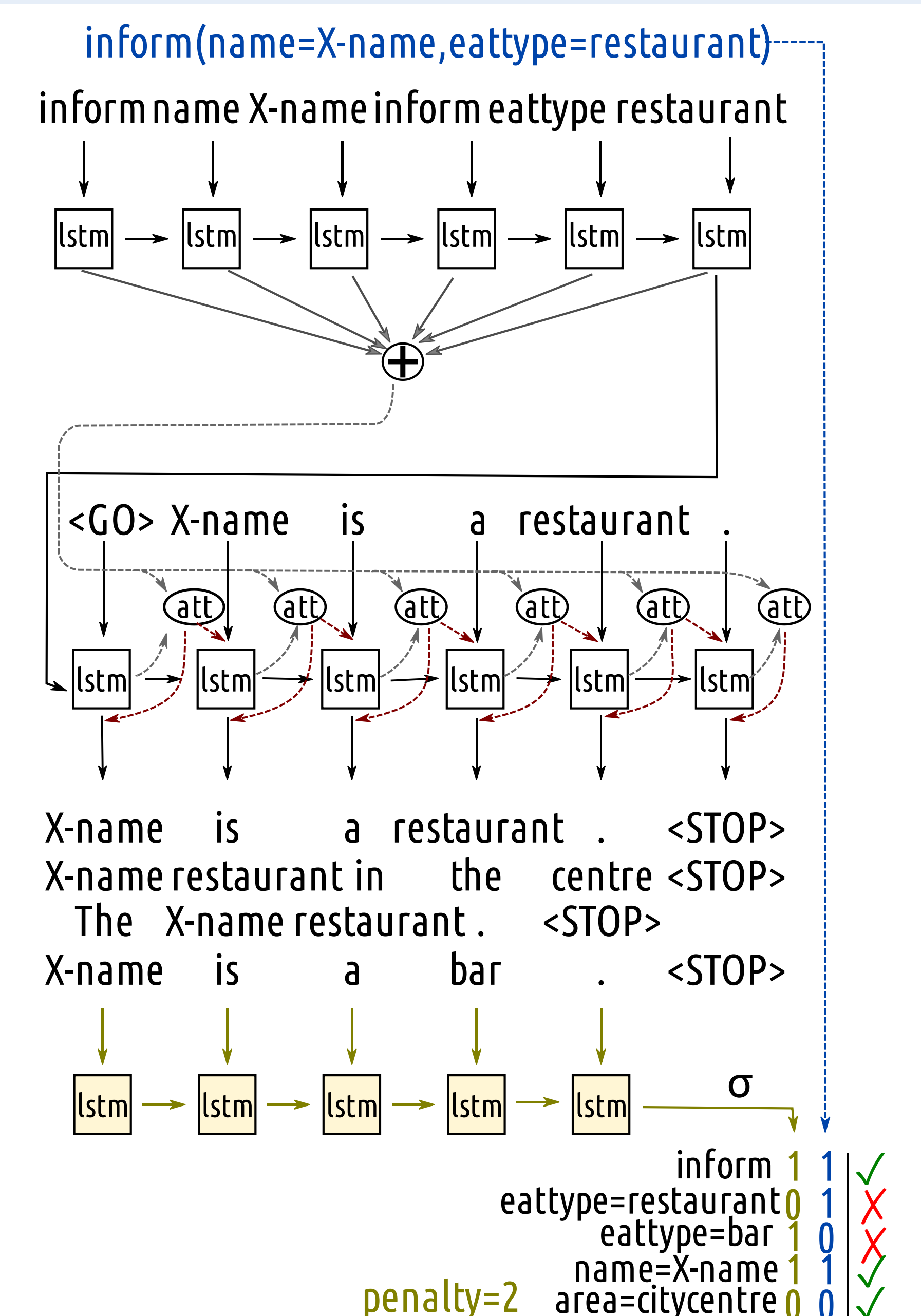
- Crowd workers asked to realize *relevant* information
 - Some (redundant) information is omitted
 - Some information is added

Dataset	Omit	Add	Exact
E2E NLG	22%	18%	60%
SFRest	0%	6%	94%
Bagel	0%	0%	100%

Baseline System

TGen: <http://bit.ly/tgen-nlg>

- LSTM **sequence-to-sequence** model with attention
- Beam search
- Reranking by MR classification
 - penalize differences w.r.t. the input MR
- Requires **delexicalization**



Development set performance:

Metric	Value
BLEU	0.6925
NIST	8.4781
METEOR	0.4703
ROUGE-L	0.7257
CIDEr	2.3987