Actionable Entities Recognition Benchmark for Interactive Fiction

Alexey Tikhonov and Ivan P. Yamshchikov

"One must never place a loaded rifle on the stage if it is not going to go off. It is wrong to make promises you do not mean to keep."

A. Chekhov

[Actionable Entities Recognition] (AER) - recognition of entities that protagonists could interact with for further **[plot development]**.

Interactive fiction


```
Junction                                    Score: 0        Moves: 4

Endless Stair
You are at the bottom of a seemingly endless stair, winding its way upward
beyond your vision. An eerie light, coming from all around you, casts strange
shadows on the walls. To the south is a dark and winding trail.
Your old friend, the brass lantern, is at your feet.

>take lantern
Taken.

>turn on lantern
The lamp is now on.

>s
Junction
You are at the junction of a north-south passage and an east-west passage. To
the north, you can make out the bottom of a stairway. The ways to the east and
south are relatively cramped, but a wider trail leads to the west.
Standing before you is a great rock. Imbedded within it is an Elvish sword.

>take sword
The sword is deeply imbedded within the rock. You can't even begin to budge it.

>_
```

[5550 locations from 995 different games] with one or more AE in each of them.

Detected Actionable Entities (AEs) and Actual Action Targets (ATs)

Table 1: Relationship between AEs detected by the XLMR model traind on BAER and ATs extracted from the Club Floyd dataset.

	All payer action targets (AT)				Unique ATs			
AER model threshold	p >0.5	p >0.65	p>0.8	p>0.95	p >0.5	p >0.65	p>0.8	p>0.95
Share of AEs that occur in AT list	0.38	0.43	0.48	**0.57**	0.22	0.25	0.29	0.36
Share of ATs labelled as AEs	**0.84**	0.65	0.30	0.05	**0.84**	0.65	0.30	0.04

AER is not NER

Table 2: Pre-trained XLMR t-NER hardly detects AEs out of the box. Yet after fine-tuning on BAER for entity span prediction the quality significantly improves.

	Accuracy	F1-score
Pre-trained T-NER	0.05	0.00
T-NER fine-tuned on BAER	0.50	0.51

AER is not NER

Table 3: Top t-NER categories with the highest frequencies of AEs across four different datasets.

BAER 1 117 texts		ClubFloyd 43 795 texts		TV-MAZE 299 197 texts		WikiPlots 2 070 449 texts	
T-NER category	$\frac{v_{AER}}{v_{ner}}$	T-NER category	$\frac{v_{AER}}{v_{ner}}$	T-NER category	$\frac{v_{AER}}{v_{ner}}$	T-NER category	$\frac{v_{AER}}{v_{ner}}$
product	18	chemical	7.5	person	36.5	product	31.2
person	2.6	product	7.3	product	12.2	corporation	22.7
work of art	2.0	person	5.5	corporation	3.9	person	20.1
organization	0.7	other	2.1	organization	3.3	group	5.4
location	0.5	work of art	1.3	chemical	3.1	chemical	4.9

AEs and Narrative Structure
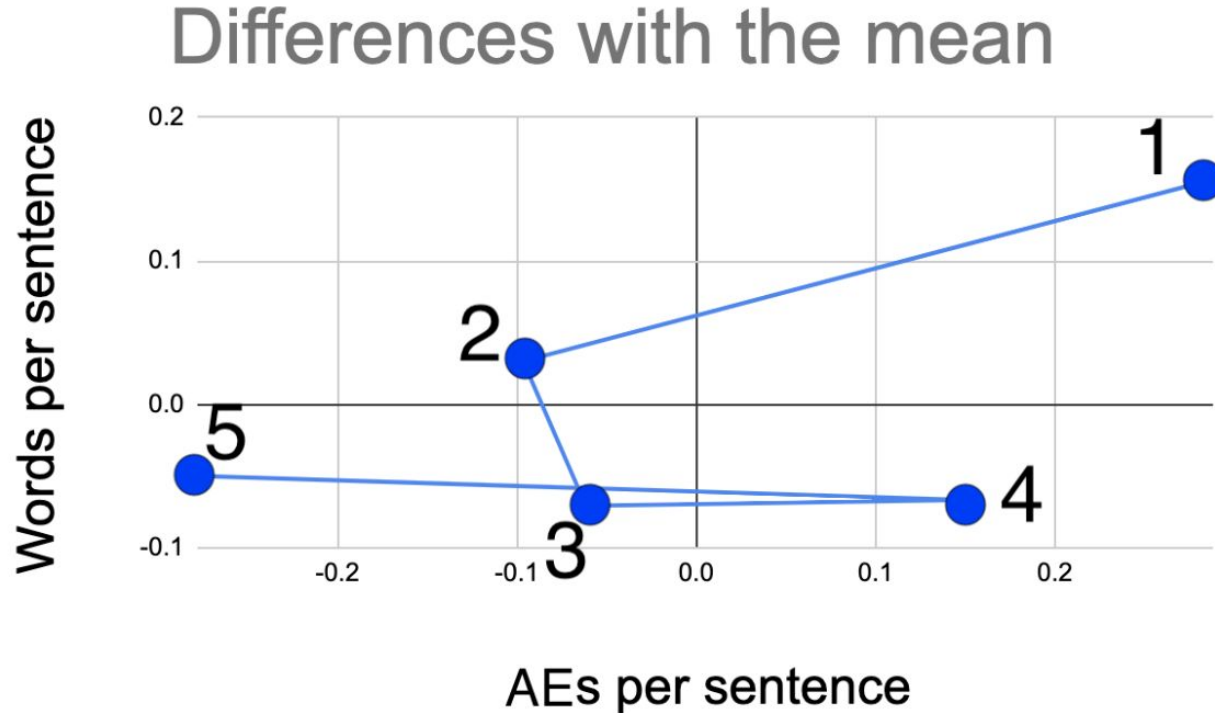
Conclusion

- presents a **[new task]** — **[AER]**;

- presents a **[new BAER dataset]** to introduce the **[benchmark for AER]**;

- shows **[correlation]** between AEs and human attempts to interact with entities;

- illustrates the **[differences]** between **[AER]** and **[NER]** tasks;

- analyzes the **[connection]** between **[AEs]** in the text and the **[structure of the narrative]**.