Measuring Text Readability with **Machine Comprehension**

Marc Benzahra[1,2], *François Yvon*[1]
[1]*LIMSI*, CNRS, Université Paris-Saclay, Orsay, France
[2]*Glose*, 53 rue du rocher, Paris, France
{marc.benzahra,francois.yvon}@limsi.fr

Download paper

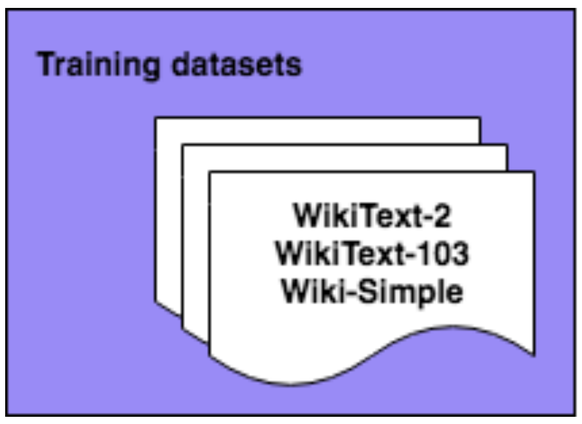
The **simpler** a **text** is, the **better** it should be **understood**.

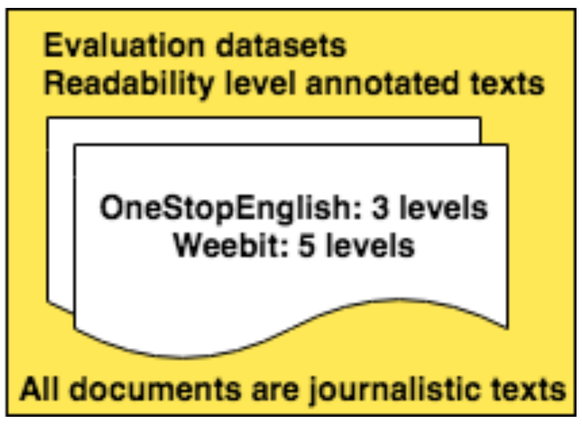
Estimate **machine reading comprehension** with the **performance of Language Models at infilling text**.

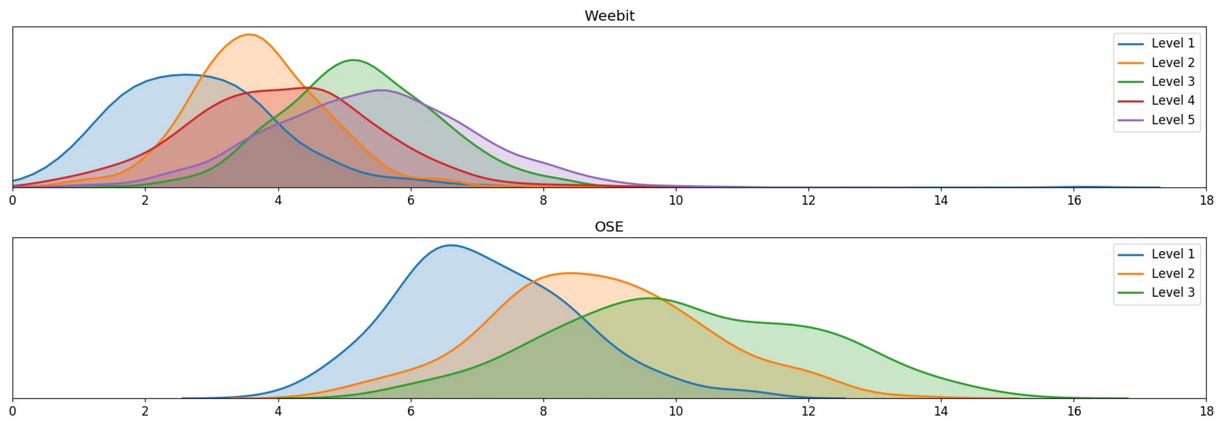
Correlation between **performance** and **readability levels** is **barely visible**.

Datasets

Datasets are used both to **train language models** and **correlate** their **text infilling performance** with **text readability categories**.







Flesch-Kincaid grade level distribution overlaps between readability levels

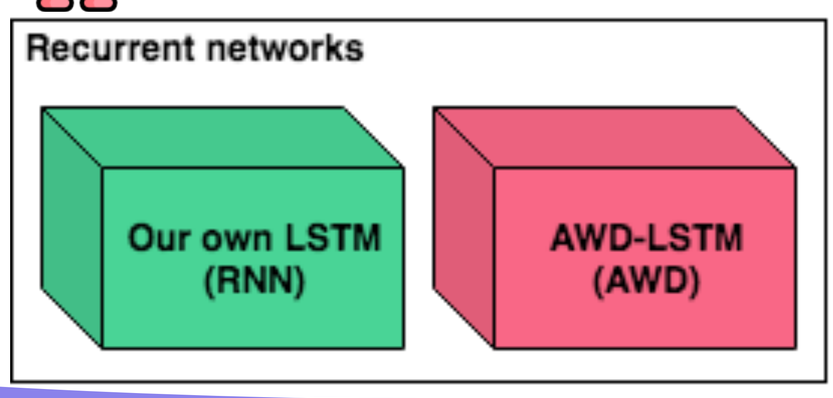
Cloze tests

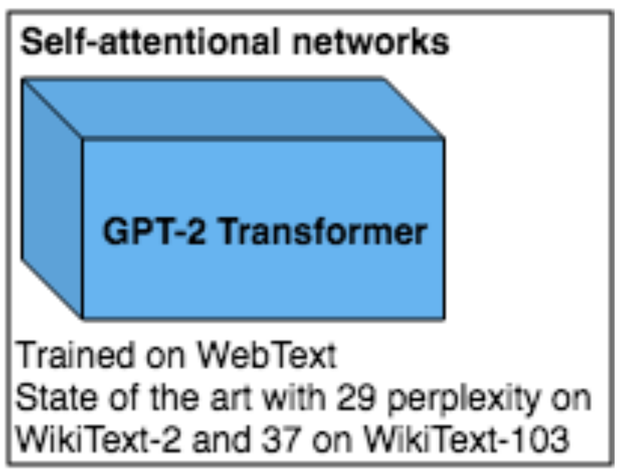
- Evaluation documents split into 5 chunks
- For each chunk, 3 blank positions chosen at random
- Prediction deemed correct when reference word appears in the first N (p@N) candidates ranked by a LM

They have twin daughters named Barbara and Jenna ___(A)
The family has a dog ___(B)
Barney and a ___(C) named India.

(A) **.** (B) **named** (C) **cat**

Language models





Model	WT2	WT103	WTSimple
RNN	90	87	51
AWD	78	137	65

Language models perplexity values

Results

- Completion rates are overall distributed as expected
- Kendall's tau-b correlations between difficulty level and completion rates are much smaller than with Flesch-Kincaid and grade level scores

Infilling performance and its correlation with difficulty level on OneStopEnglish corpus

	WikiText-2				WikiText-103				Wiki-Simple			
p@...	1	5	25	50	1	5	25	50	1	5	25	50
RNN (1)	0.10	0.21	0.33	0.39	0.12	0.24	0.37	0.44	0.12	0.27	0.40	0.47
RNN (2)	0.09	0.20	0.29	0.35	0.11	0.23	0.34	0.41	0.11	0.23	0.36	0.43
RNN (3)	0.08	0.19	0.28	0.34	0.10	0.22	0.33	0.39	0.10	0.22	0.33	0.39
RNN (τ)	0.05	0.06	0.11	0.12	0.06	0.06	0.09	0.10	0.11	0.13	0.17	0.19
AWD (1)	0.11	0.22	0.34	0.40	0.12	0.23	0.35	0.42	0.12	0.25	0.36	0.43
AWD (2)	0.10	0.20	0.31	0.37	0.11	0.21	0.32	0.37	0.11	0.22	0.32	0.38
AWD (3)	0.09	0.19	0.30	0.35	0.10	0.21	0.31	0.37	0.10	0.21	0.30	0.36
AWD (τ)	0.05	0.08	0.09	0.11	0.06	0.07	0.09	0.11	0.11	0.11	0.15	0.17