MWE as WSD: Solving Multiword Expression Identification with Word Sense Disambiguation

*Joshua Tanner
Mantra Inc.

*Jacob Hoffman
Project Ronin

*Both authors contributed equally to this work.

Motivation

- Determining whether a group of words constitutes a multiword expression (MWE) requres an understanding of that MWE's meaning and whether it matches the context

kick_the_bucket

Grandpa kicked the bucket at the age of 102.

I kicked the bucket down the hill.

No MWE

- Approaches using gloss information (definition text) to represent meaning have recently been very successful in word sense disambiguation (WSD)
- Extending these gloss-based WSD methods to MWE identification allows us to efficiently solve both tasks with a single Bi-Encoder model

Data

- We train on a modified SemCor dataset with added synthetic negative MWE examples and a small number of manual MWE annotations

SemCor after each addition of data

	Positive MWEs	Negative MWEs
SemCor	12409	0
+Annotation	12907	658
+SyntheticNegatives	12907	14688

Examples of each annotation type

Context	MWE	Type
What effort **do** you **make** to assess...	**make do**	Synthetic Negative
..your in **plant** feeding **operation?**	**in operation**	Annotated Negative
...**works** full-time **on** some other assignment?	**work on**	Annotated Positive

- We also experiment with fine-tuning on the PARSEME and DiMSUM training datasets

Results

Test set results on PARSEME 1.1 English and DiMSUM for MWE identification

Training data is listed in parenthesis: S=SemCor, P=PARSEME, D=DiMSUM

PARSEME1.1							DiMSUM		
MWE-based			Token-based			System	MWEs		
P	R	F	P	R	F		P	R	F
-	-	36.0	-	-	40.2	Taslimipoor+(2019) Kirilin+(2016)	73.5	48.4	58.4
-	-	**41.9**	-	-	-	Rohanian+(2019) Williams(2017)	65.4	56.0	60.4
36.1	**45.5**	40.3	40.2	**52.0**	**45.4**	Liu+(2021)	47.9	52.2	50.0
16.3	39.9	23.1	19.2	43.9	26.7	Rule-basedPipeline	57.7	55.5	56.6
28.2	38.5	32.5±0.4	30.7	39.0	34.3±0.4	Rules+DCA(*S*)	70.9	53.0	60.6±0.1
35.7	39.3	37.4±0.6	37.7	38.6	38.1±0.4	Rules+DCA(*S* / *D*)	78.2	51.8	**62.3**±0.1
47.1	33.8	39.4±0.3	**48.3**	32.1	38.6±0.2	Rules+DCA(*S* / *P*)	75.7	49.4	59.8±0.1
45.4	33.2	38.3±0.1	46.9	31.9	38.0±0.2	Rules+DCA(*S* / *P* / *D*)	**80.4**	49.5	61.3±0.4

- Our approach improves precision substantially, achieving SOTA on DiMSUM
- DCA Poly-encoder is our most perfomant model
- Recall is limited by our dependence on a MWE lexicon; work in lexicon expansion could further improve scores

English WSD all-words task F1

System	F1	System	F1
Blevins+	79.0	PolyEnc(*S*)	73.8±0.2
DCA(*S*)	77.2±0.1	BiEnc(*S*)	77.4±0.6
DCA(*S* / *P* / *D*)	74.4±0.6	BiEnc(*S* / *P* / *D*)	74.2±1.0

- Our models retain most of their WSD performance, but lose around 6% when fine-tuned for MWE identification

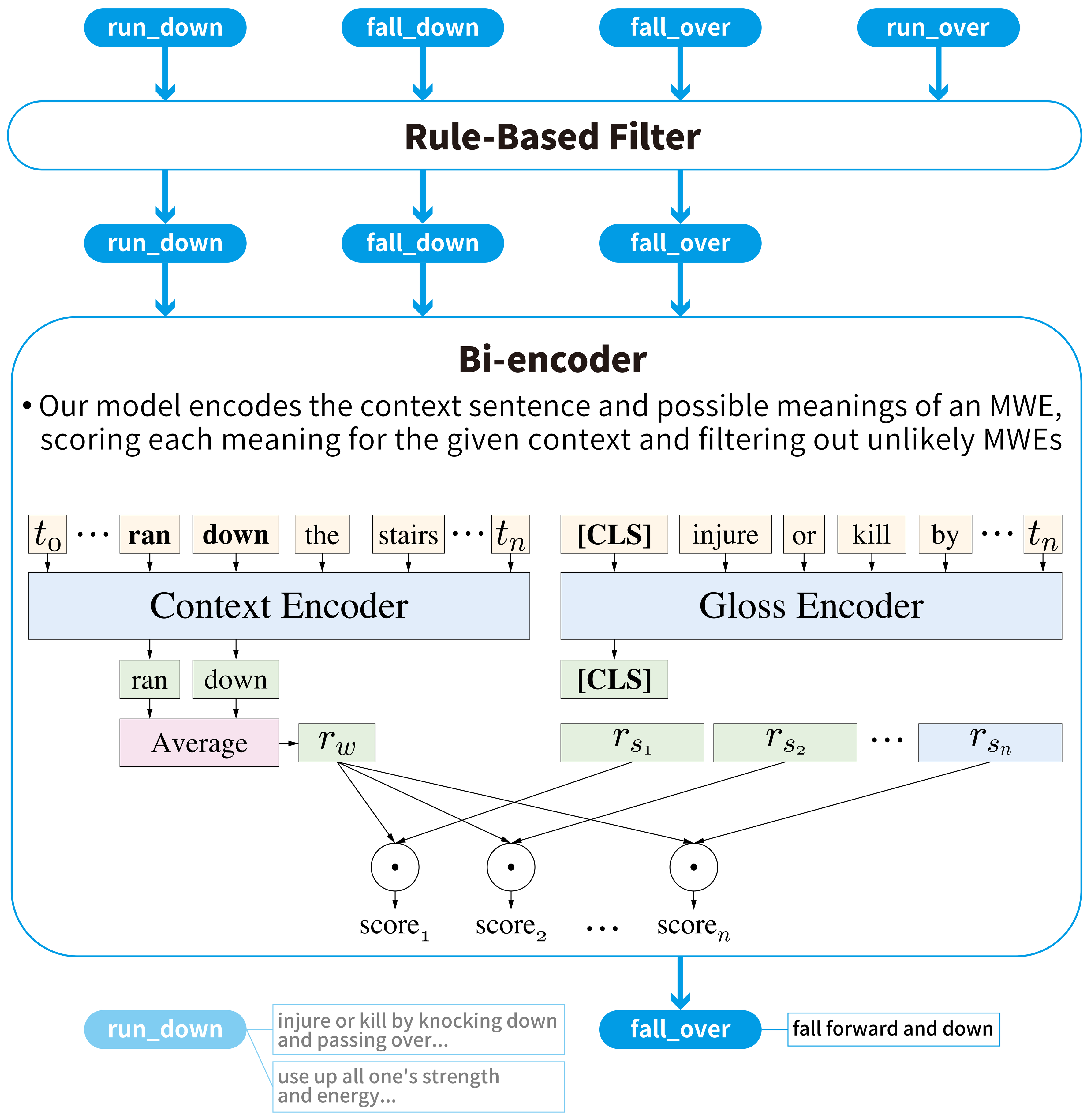
Methodology

run_down fall_down fall_over

I ran down the stairs and fell over.

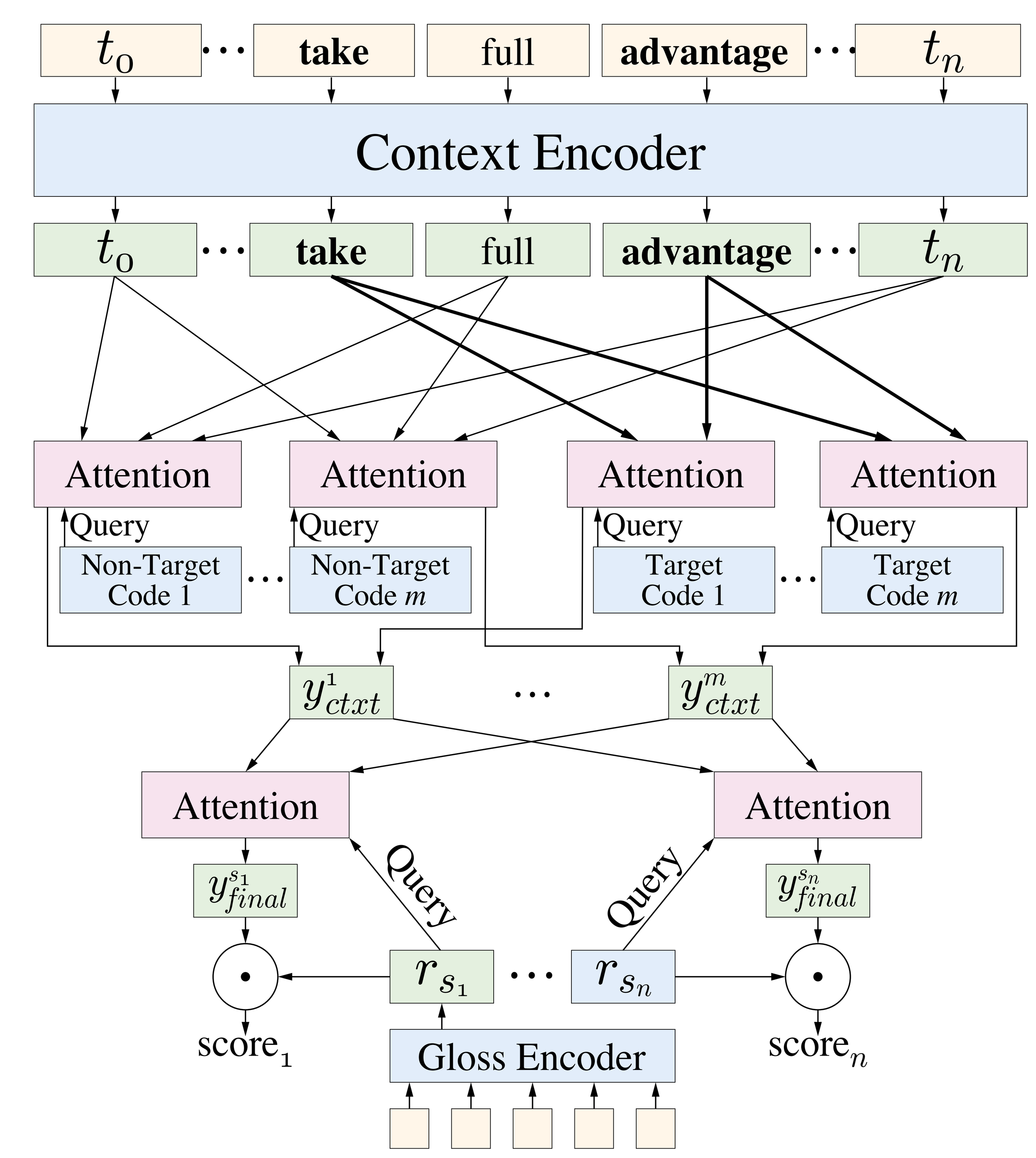
run_over

- We exhaustively generate MWE candidates with a rule-based approach and then filter them using simple rule-based filters and our Bi-Encoder model, which excludes MWEs with meanings inappropriate for the context sentence



DCA Poly-Encoder

- We also introduce a modified Poly-encoder (distinct codes attention=DCA) architecture for solving token-level tasks which outperforms the Bi-Encoder



EMNLP 2023

Data, Code, Paper: https://github.com/Mindful/MWEasWSD

Contact: josh@mantra.co.jp