

# **How does the pre-training objective affect what large language models learn about linguistic properties**

Ahmed Alajrami, Nikolaos Aletras

Presented by Giuseppe Attanasio

MilaNLP

**“Models trained on representations using  
non-linguistically motivated objectives  
[...] achieve competitive results”**

## A linguistically motivated objective: MLM

“Models trained on representations using  
**non-linguistically motivated objectives**  
[...] achieve competitive results”

## A linguistically motivated objective: MLM

“Models trained on representations using  
**non-linguistically motivated objectives**  
[...] achieve **competitive results**”

Probing representations  
for linguistic properties

## Linguistically Motivated

- Masked Language Modeling
- Select and Replace (S+R): 10% shuffle, 10% random

## Non-Linguistically Motivated (NoMo)

- First Char Prediction
  - [c]at and [c]omputer have the same class
- ASCII
  - Sum ASCII && mod5
- Random (5-way)

Target of "cat":

$$(99 + 95 + 116) \bmod 5 = 0$$

# Probing tasks



# Probing tasks

SentLen

TreeDepth

TopConst

BShift

Tense

SubjNum

ObjNum

Semantic Odd Man Out

CoordInv

# words in the sentence

Depth of its parse tree

Top constituents of the tree

Adjacent words swap

Present/past tense

Singular/plural subject

Singular/plural direct object

Noun or verb replaced

Inverted coordinate clauses

What you can cram into a single \$&!\* vector:  
Probing sentence embeddings for linguistic properties

Alexis Conneau  
Facebook AI Research  
Université Le Mans  
aconneau@fb.com

German Kruszewski  
Facebook AI Research  
germank@fb.com

Guillaume Lample  
Facebook AI Research  
Sorbonne Universités  
glample@fb.com

Loïc Barrault  
Université Le Mans  
loic.barrault@univ-lemans.fr

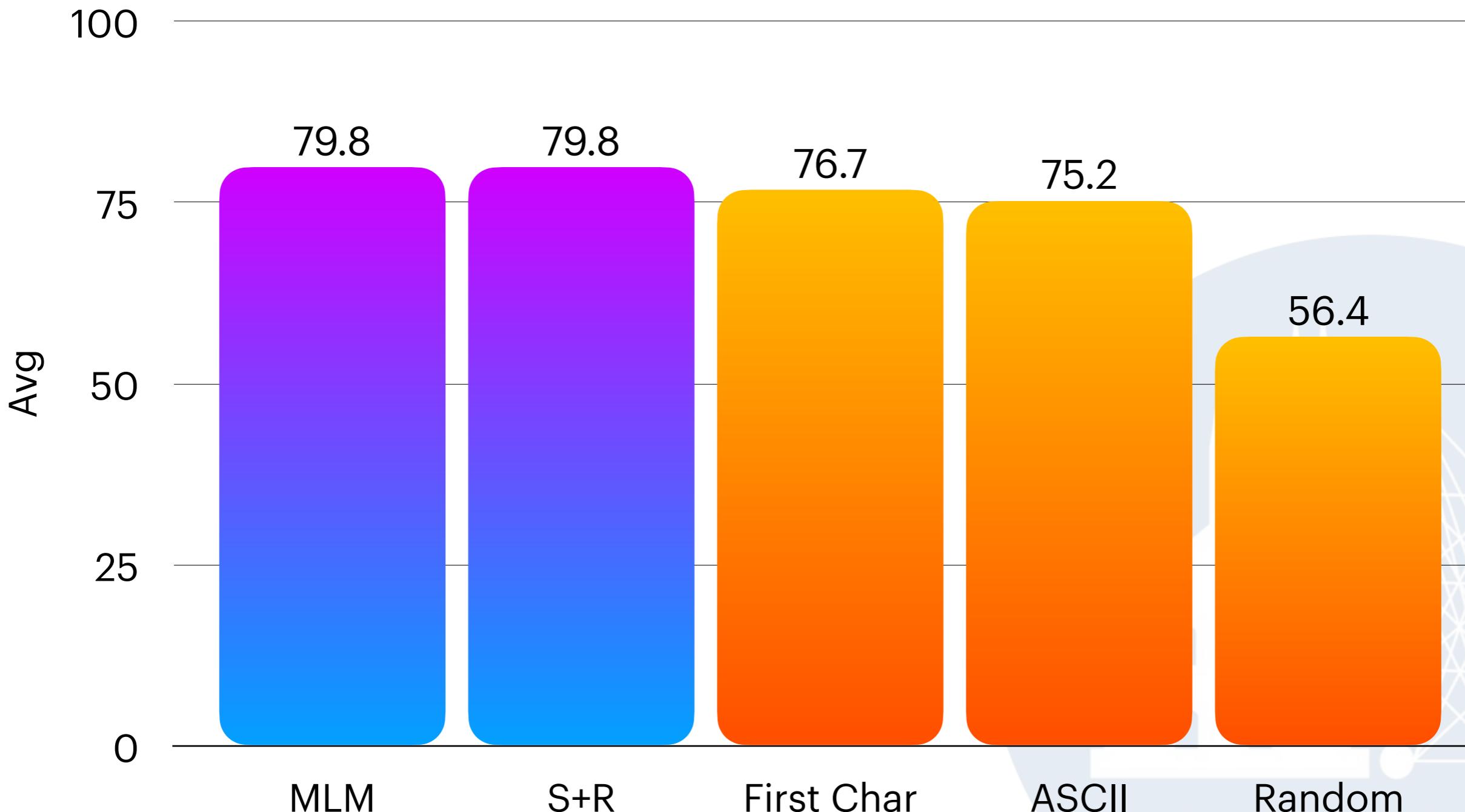
Marco Baroni  
Facebook AI Research  
mbaroni@fb.com

A large, glowing blue and purple jellyfish against a dark background. The jellyfish has a translucent, glowing body with a bright blue center and purple edges. Its long, flowing tentacles are also illuminated with a similar color palette, creating a sense of depth and movement.

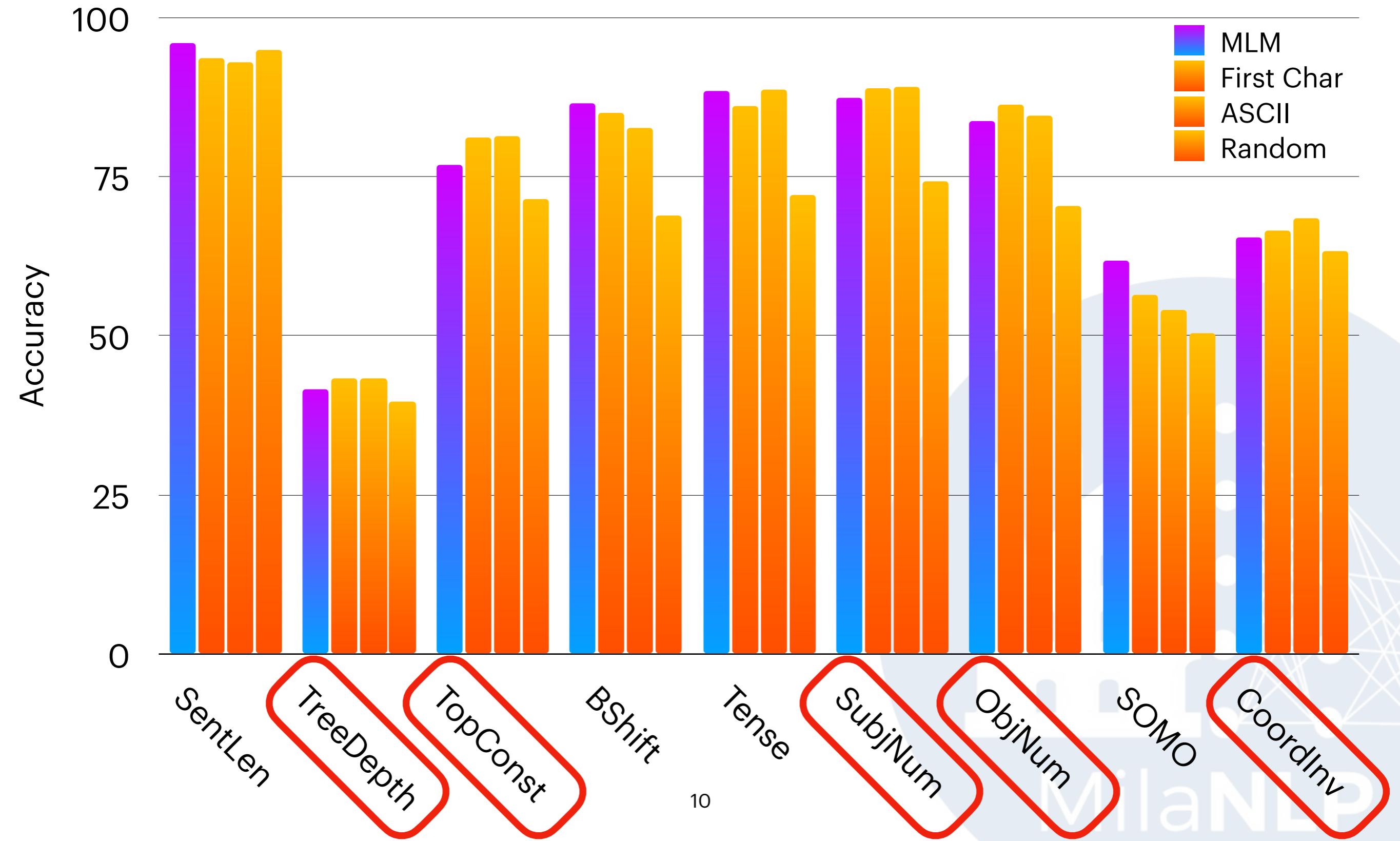
# Results

# GLUE

Makes sense!



# Linguistic Probing



# Caveats...

- Only 500k steps pretraining
- Best performing layer for linguistic probing
  - Potentially, layer 1 vs 12

... but

We should question  
our training procedures

No need for linguistic capabilities to solve GLUE

... or, linguistic probing tests are flawed

... or, GLUE tells only part of the story

What are your thoughts?

