

# What do character-level models learn about morphology?

## The case of dependency parsing

Clara Vania, Andreas Grivas, and Adam Lopez

c.vania@ed.ac.uk, andreasgrv@gmail.com, alopez@inf.ed.ac.uk

ILCC, School of Informatics, University of Edinburgh

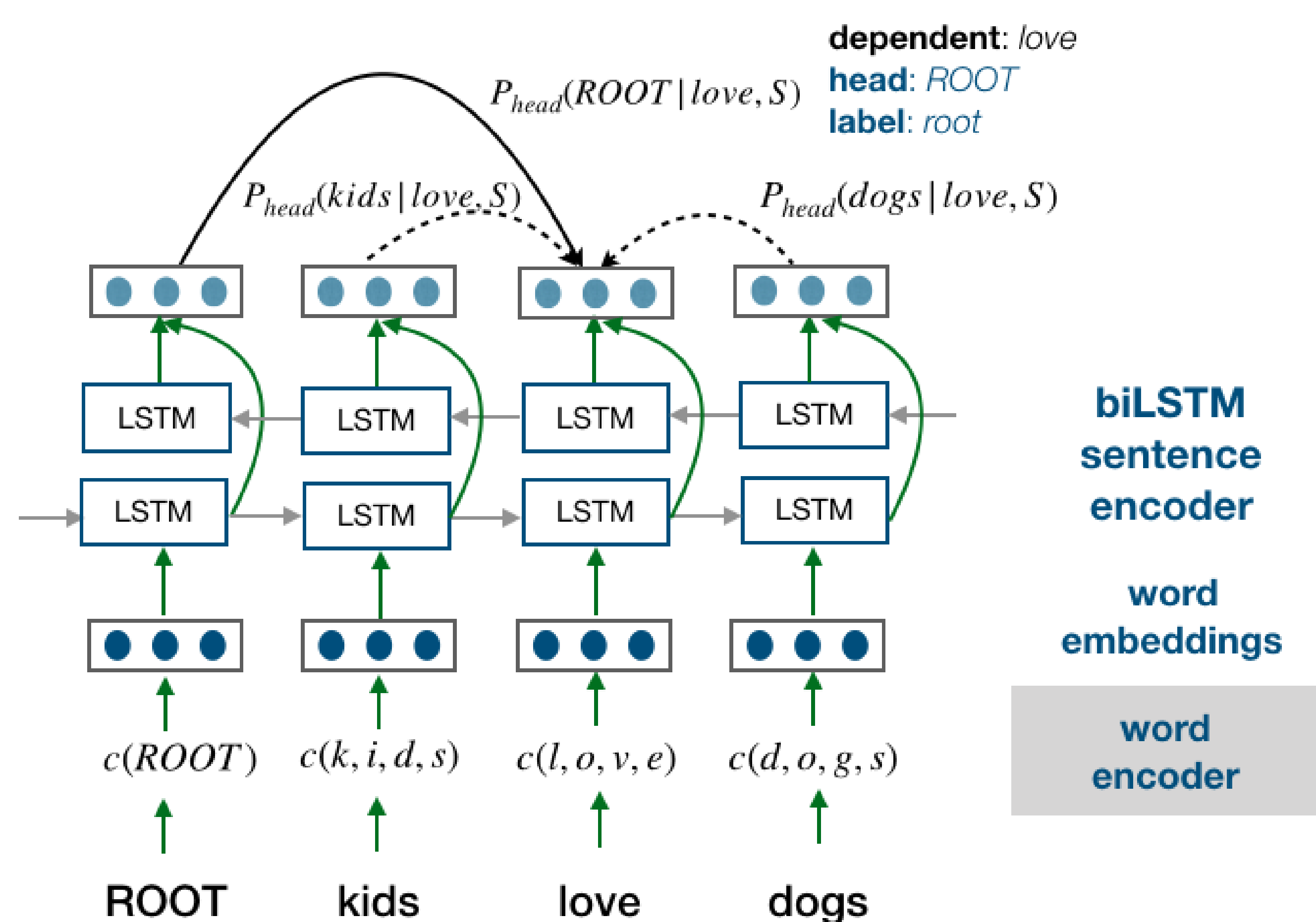


### Motivation

- Character-level models are effective for MRLs.
- Do we need to explicitly model morphology?
- Compare character-level input with morphological annotations (*oracle*):

character → w, a, n, t, s  
oracle → want, +VB, +3rd, +SG, +Pres

### Task: Dependency Parsing



Character-level models outperform word, but not as effective as the oracle

Model → ↓ Language	word	char-lstm	oracle	imp. o/c
Finnish	80.8	88.4	<b>88.8</b>	0.4
Turkish	61.6	68.6	<b>69.5</b>	0.9 <sup>†</sup>
Czech	89.3	90.6	<b>92.0</b>	1.4 <sup>†</sup>
English	88.9	89.4	<b>89.9</b>	0.5
German	84.5	84.5	<b>86.5</b>	2.0 <sup>†</sup>
Hindi	93.1	<b>93.3</b>	<b>93.3</b>	-
Portuguese	85.5	86.0	<b>86.5</b>	0.5
Russian	90.1	92.4	<b>93.3</b>	0.9 <sup>†</sup>
Spanish	86.9	87.4	<b>87.7</b>	0.3
Urdu	87.0	<b>87.1</b>	87.0	-0.1
Arabic	70.9	72.1	<b>72.7</b>	0.6
Hebrew	69.8	69.8	<b>70.0</b>	0.2

Table 1: Label Attachment Score (LAS).

### Why do characters beat word?

- Character-level models are **effective** for **OOV handling**.

dogs vs. sloths

- Parameter sharing** between pairs of **observed** training words also **helps**.

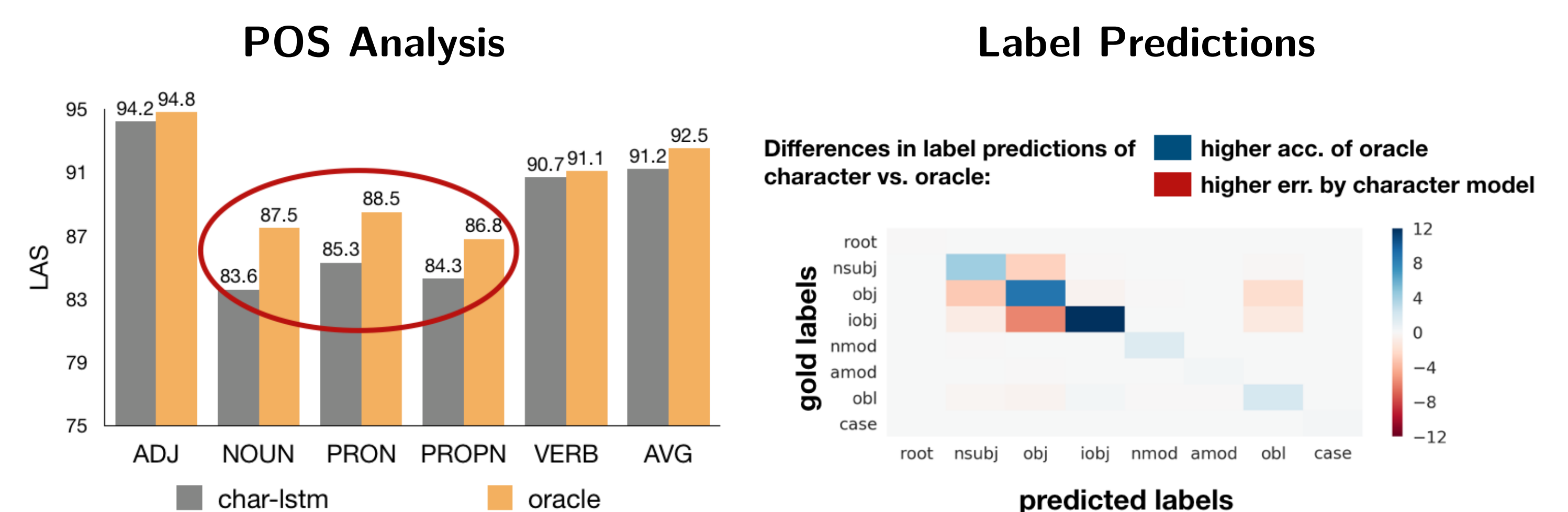
dog vs. dogs

### Why do morphemes beat characters?

- Morphological analysis **disambiguates** words.

{r,e,a,d} vs. {read, SG, 3rd, Pres}  
{r,e,a,d} vs. {read, SG, 1st, Past}

- Morphology helps for **nouns** → **subjects** and **objects**.



### Characters and Case Syncretism

Maša čitaet **pis mo**

Masha read.3SG **letter.ACC**

'Masha reads a letter.'

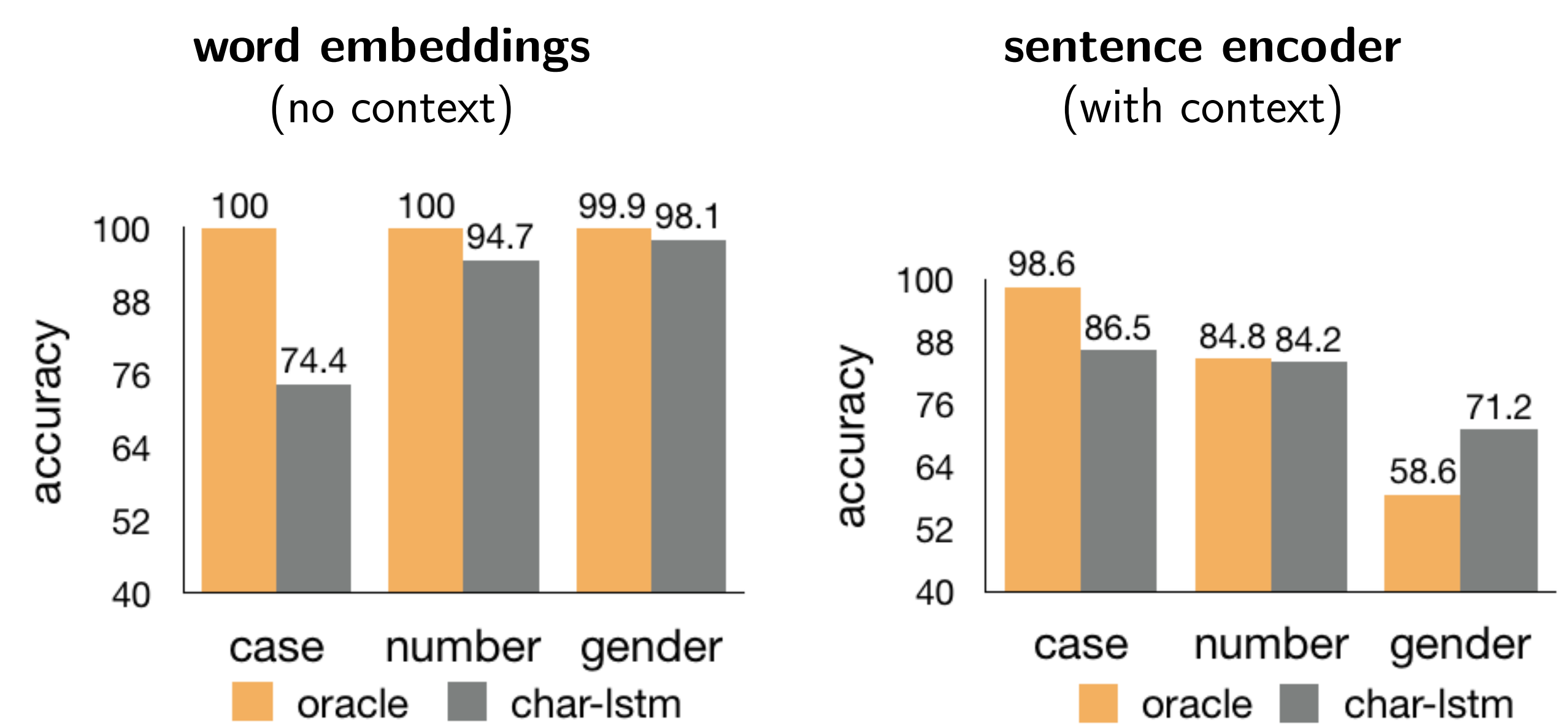
Na stole ležit **pis mo**

on table lie.3SG **letter.NOM**

'There's a letter on the table.'

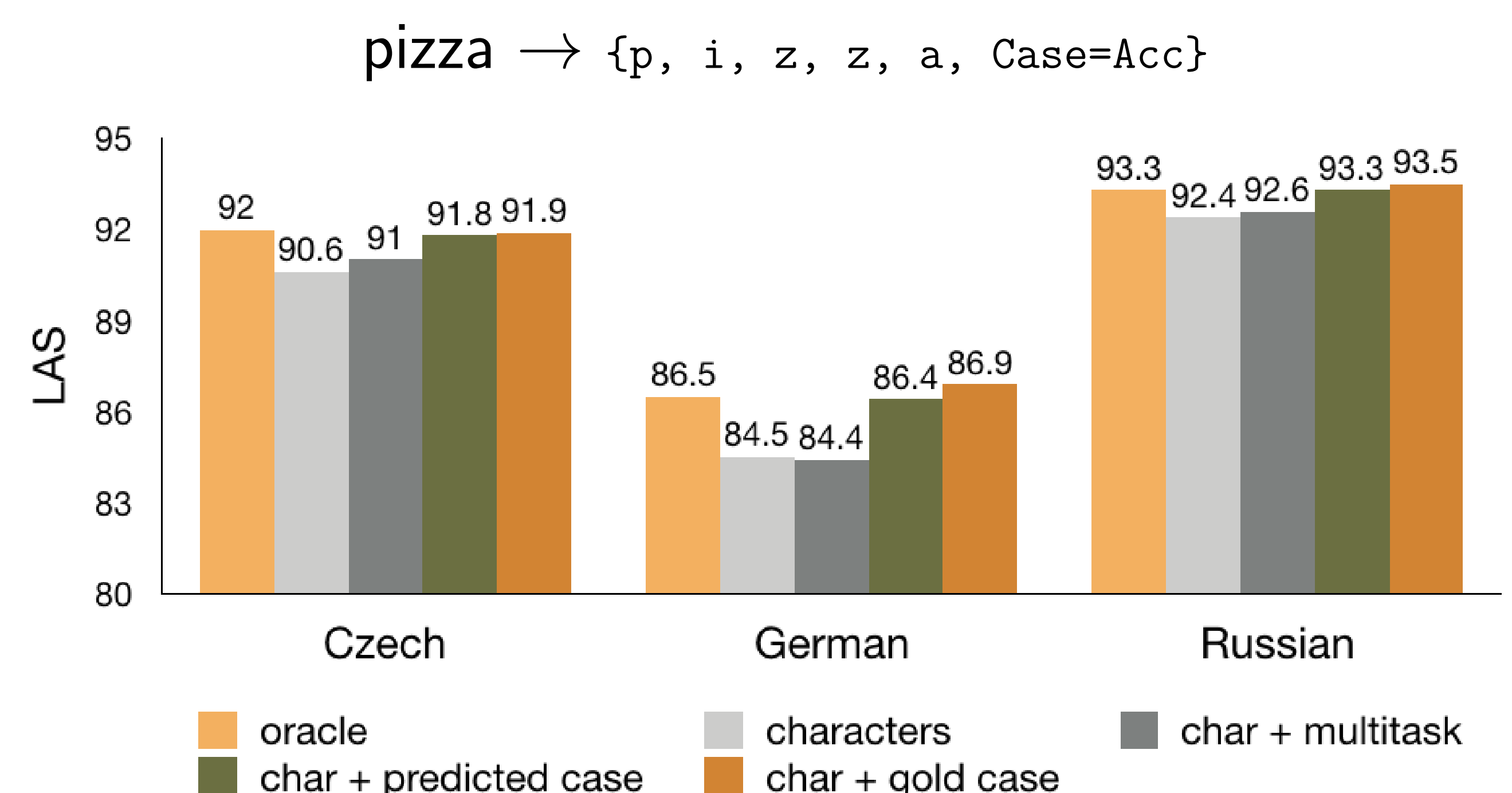
### The oracle relies on case

- Train a classifier to predict case given a learned representation from specific layers, i.e., output of:



### Explicitly modeling case improves parsing accuracy

- Multi-task learning**: use case prediction as an auxiliary task.
- Augment character input with **predicted case\***:



(\* We use the same dataset for case prediction.

### Conclusion

- Character-level models learn **some aspects of morphology**.
- For dependency parsing, **case syncretism** shown to be **difficult** for character-level models.
- Explicitly modeling case improves** neural dependency parsing.