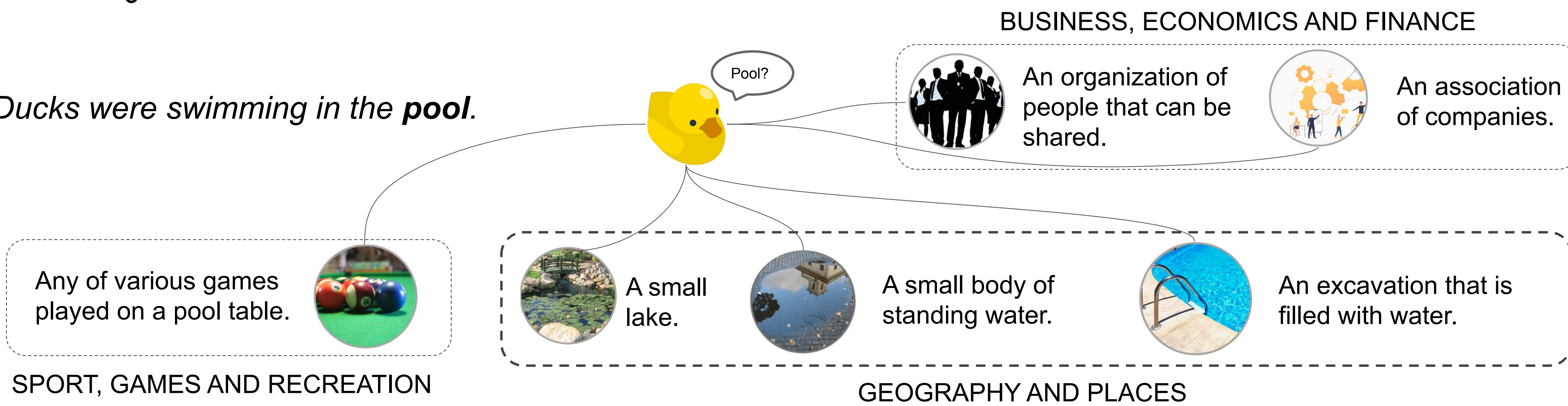




## Word Sense Disambiguation: the fine-granularity problem

Ducks were swimming in the **pool**.



## CSI

- A **Coarse Sense Inventory** (CSI) obtained by **manually** grouping Roget's categories into **45** labels and mapping them to WordNet synsets.
- **Shared** among lemmas, covering all parts of speech.
- CSI labels are **easy to use** and have a high **descriptiveness**.

## Competitors

### WordNet Domains (WND)

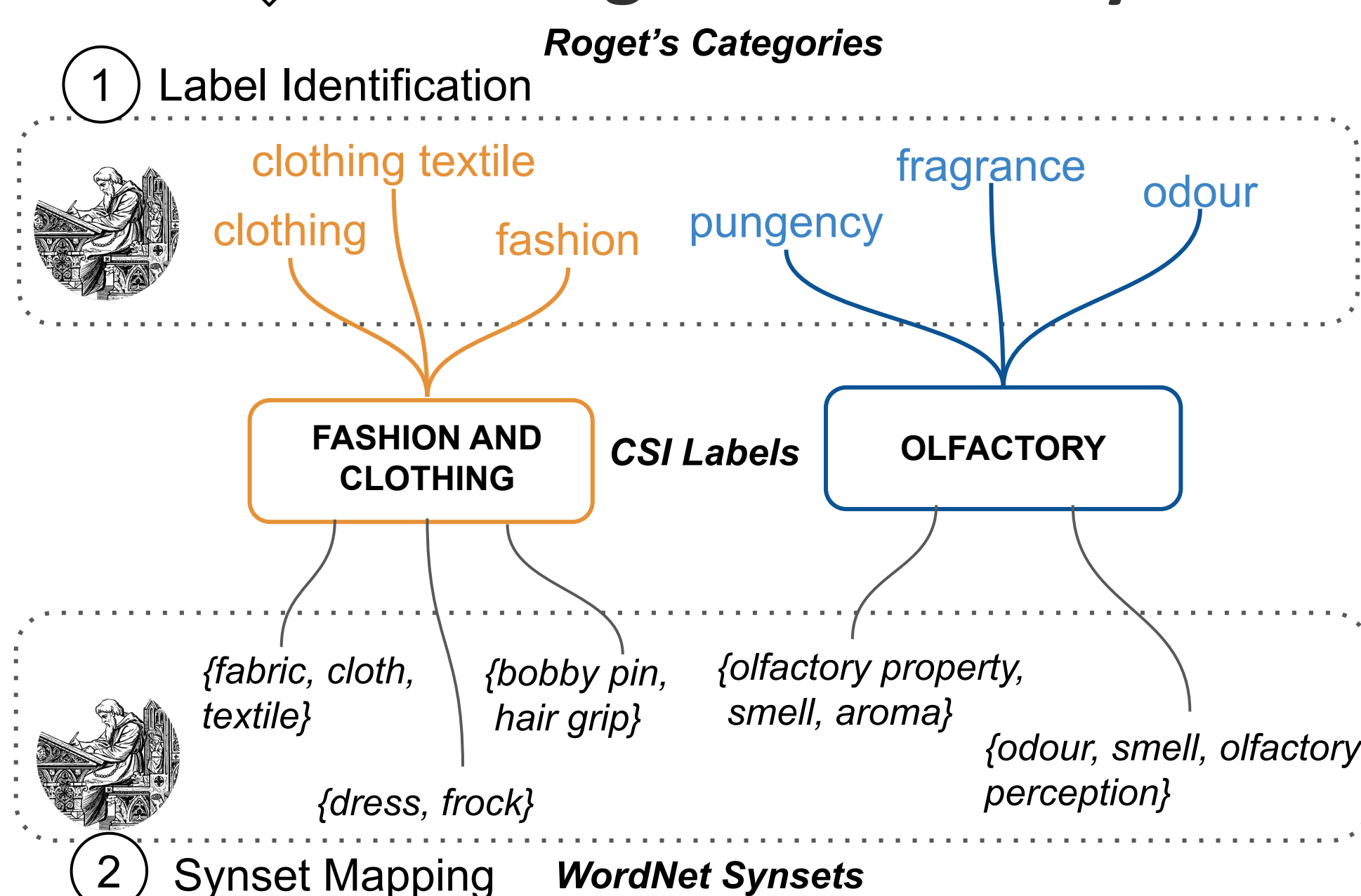
- ✓ Semi-automatic
- ✓ 200 labels
- ✓ All parts of speech included
- ✗ *Factotum* label covers more than 18% of WordNet synsets

### SuperSenses (SuS)

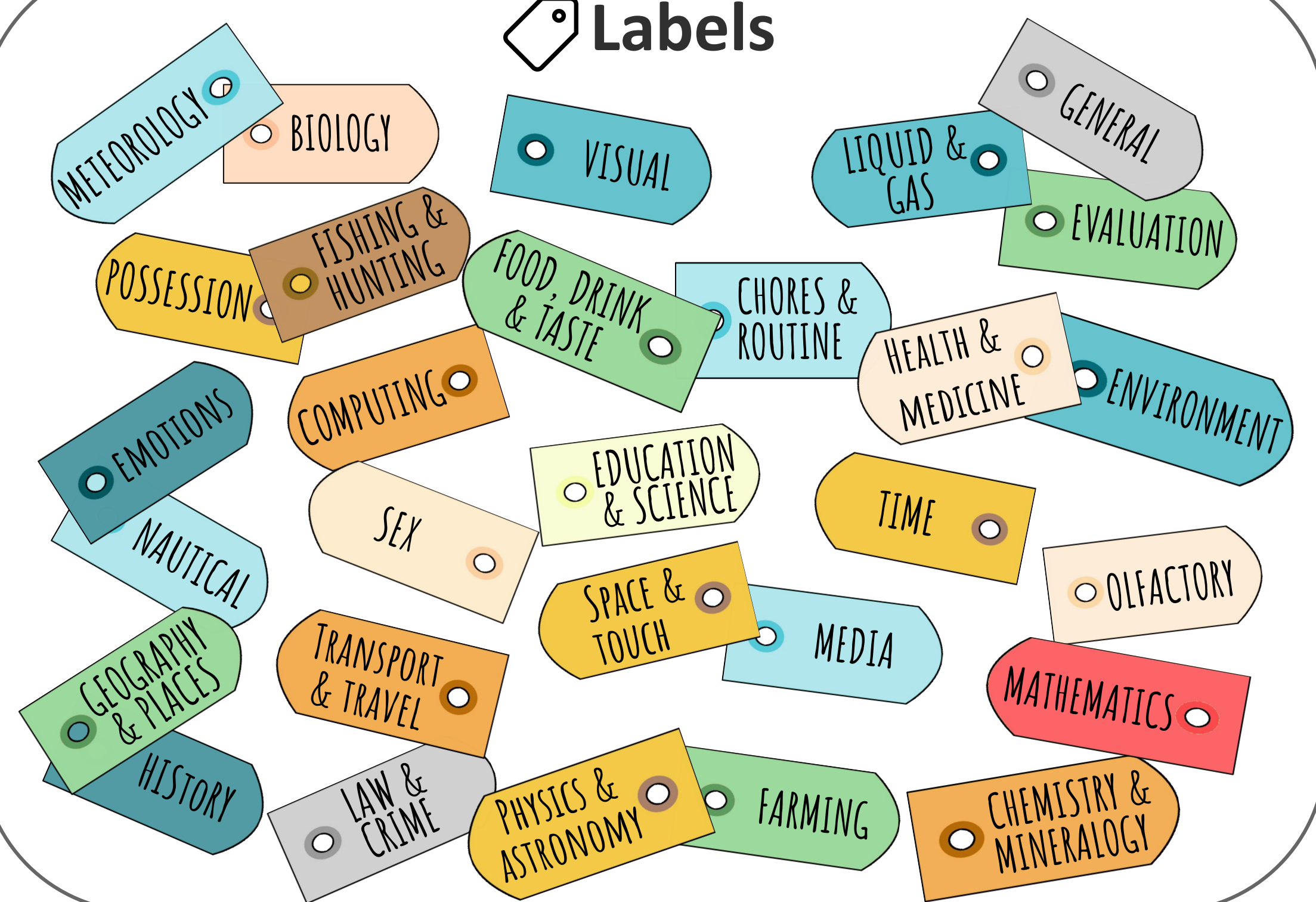
- ✓ Manual
- ✓ 45 labels
- ✗ Only nouns and verbs are meaningfully clustered



## Building the Inventory



## Labels



## Descriptiveness Task

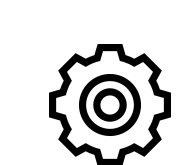
200 target words from SemCor annotated by 3 annotators.

The street drowsed on an August afternoon in the **shade** of the curbside tree, and **silence** was a weight.

- ★★★★ Physics & astronomy
- ★★ Noun.state
- ★ Factotum
- ★★★★ Acoustics
- ★★★ Music, sound & dancing
- ★ Noun.attribute

CSI WND SuS

	K-IAA [0-1]	Descriptiveness [1-3]
CSI	<b>0.81</b>	<b>2.23</b> ★★★★★
WND	0.74	1.80 ★★★★★
SuS	0.69	2.04 ★★★★★
WordNet	0.51	-



## Experimental Setup



BERT contextualized embeddings + dense layer



+ BERT contextualized embeddings + bi-directional LSTM



+ ELMo contextualized embeddings + bi-directional LSTM

**Training:** SemCor

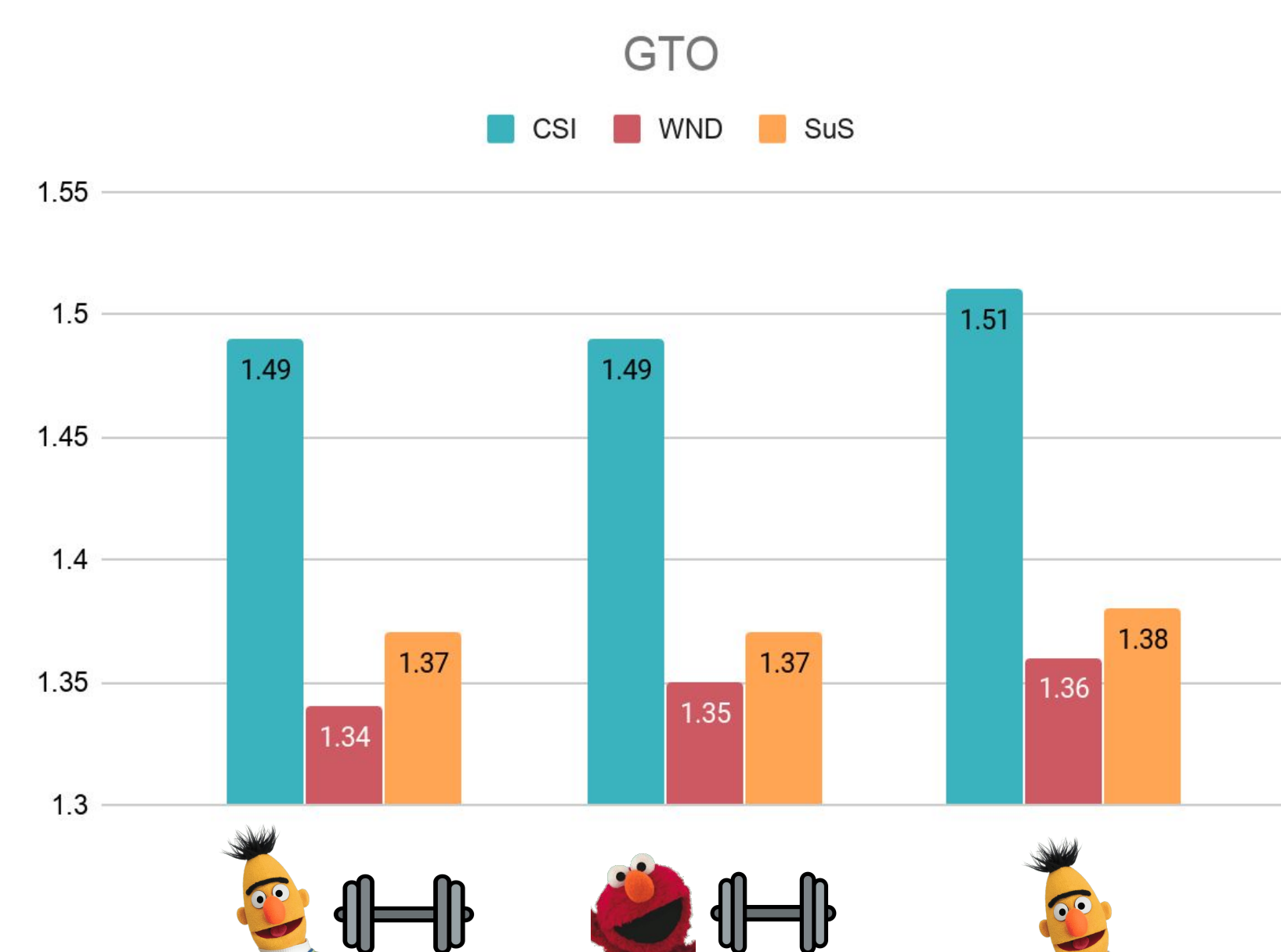
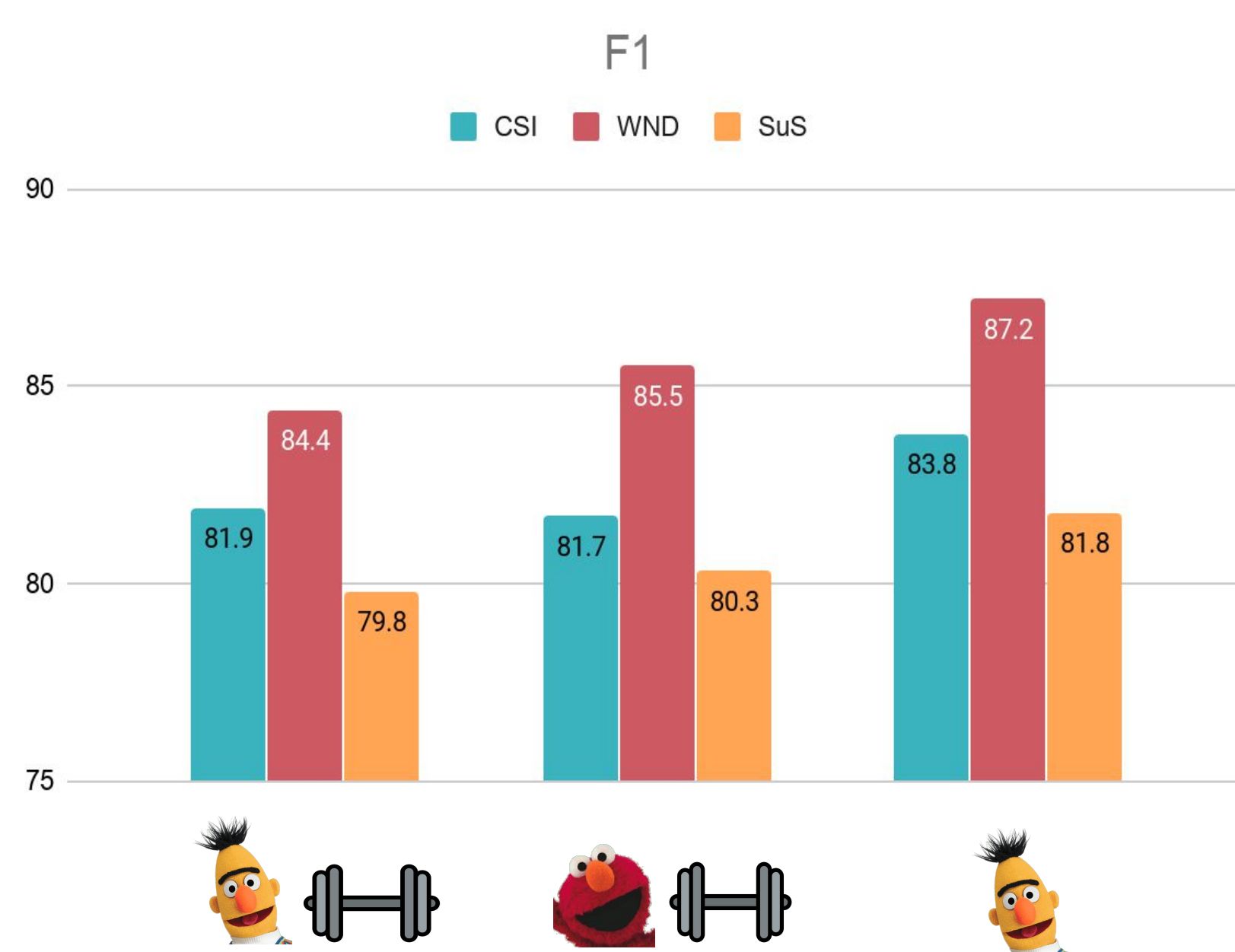
**Dev:** SemEval-07

**Test:** Senseval-2 + Senseval-3 + SemEval-13 + SemEval-15 (**ALL**)



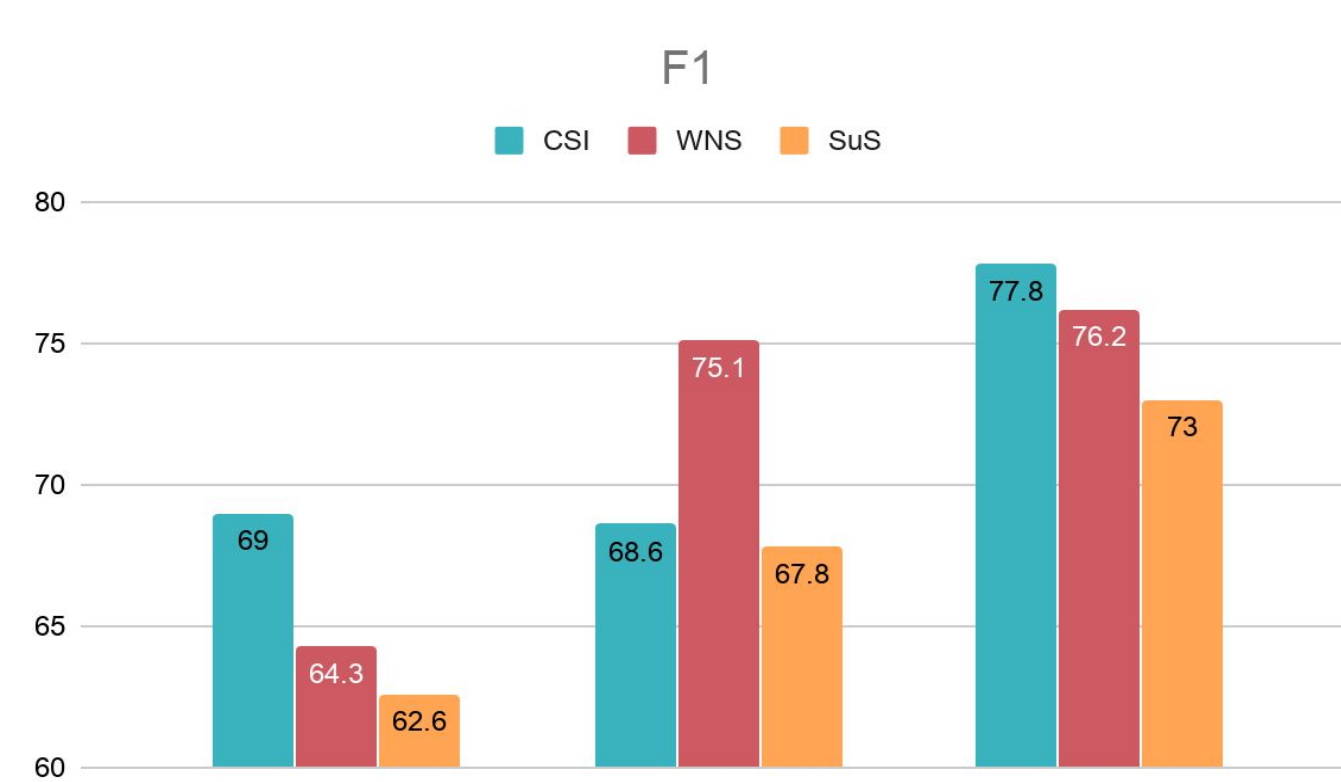
## English All-Words WSD

Results on the ALL dataset in terms of F1 and perplexity/F1 trade-off (**GTO**).

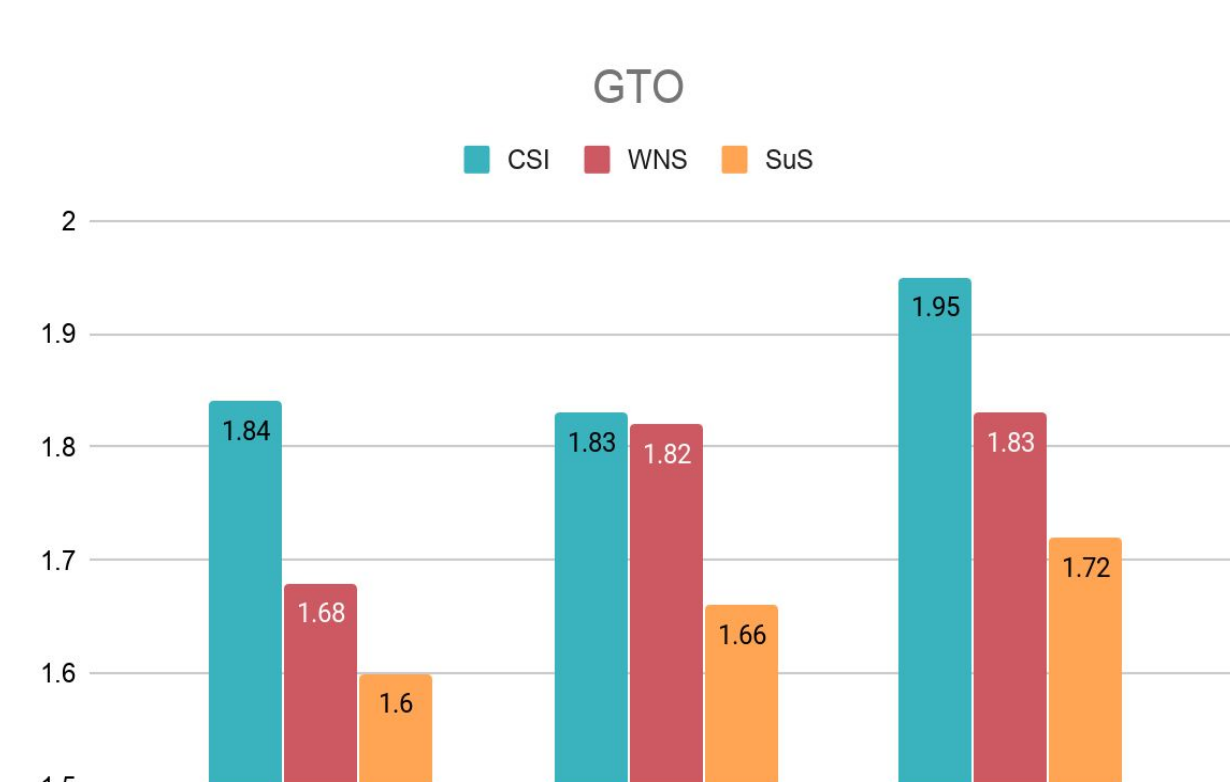


## Zero- and Few-Shot Learning

CSI increases the **sample efficiency** of a supervised WSD model.



CSI enables BERT to attain the best performance on **out-of-vocabulary** words.



T<sub>n</sub> : Training corpus with *n* annotated sentences per target word



## Achievements

**Descriptive labels**, that are easy to use.

**Trade-off** between granularity and performance.

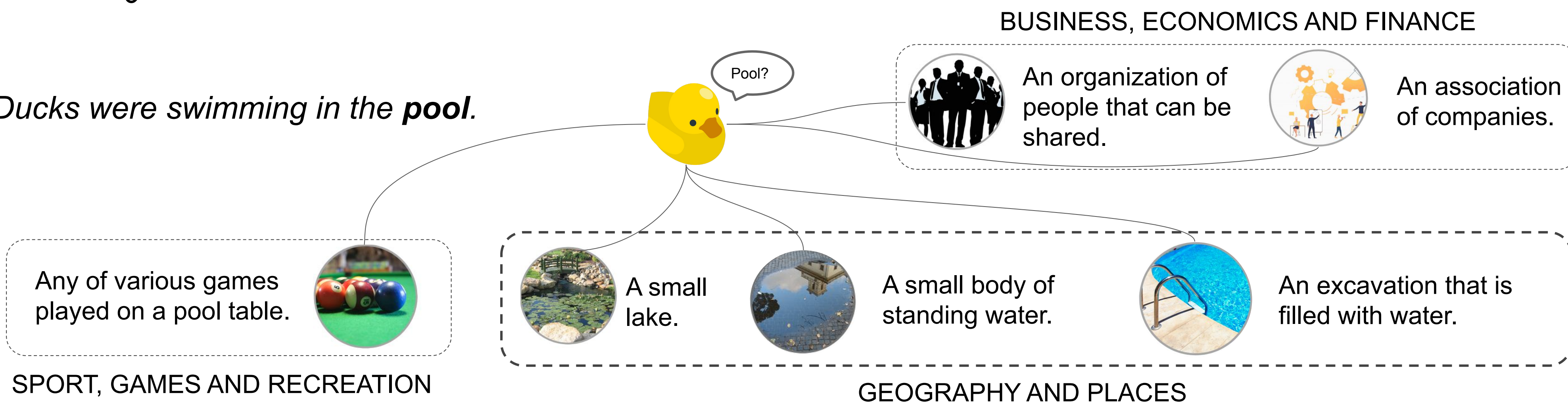
**Reduce** the need for manually annotated data.





## Word Sense Disambiguation: the fine-granularity problem

Ducks were swimming in the **pool**.



## CSI

- A **Coarse Sense Inventory** (CSI) obtained by **manually** grouping Roget's categories into **45** labels and mapping them to WordNet synsets.
- **Shared** among lemmas, covering all parts of speech.
- CSI labels are **easy to use** and have a high **descriptiveness**.



## Competitors

### WordNet Domains (WND)

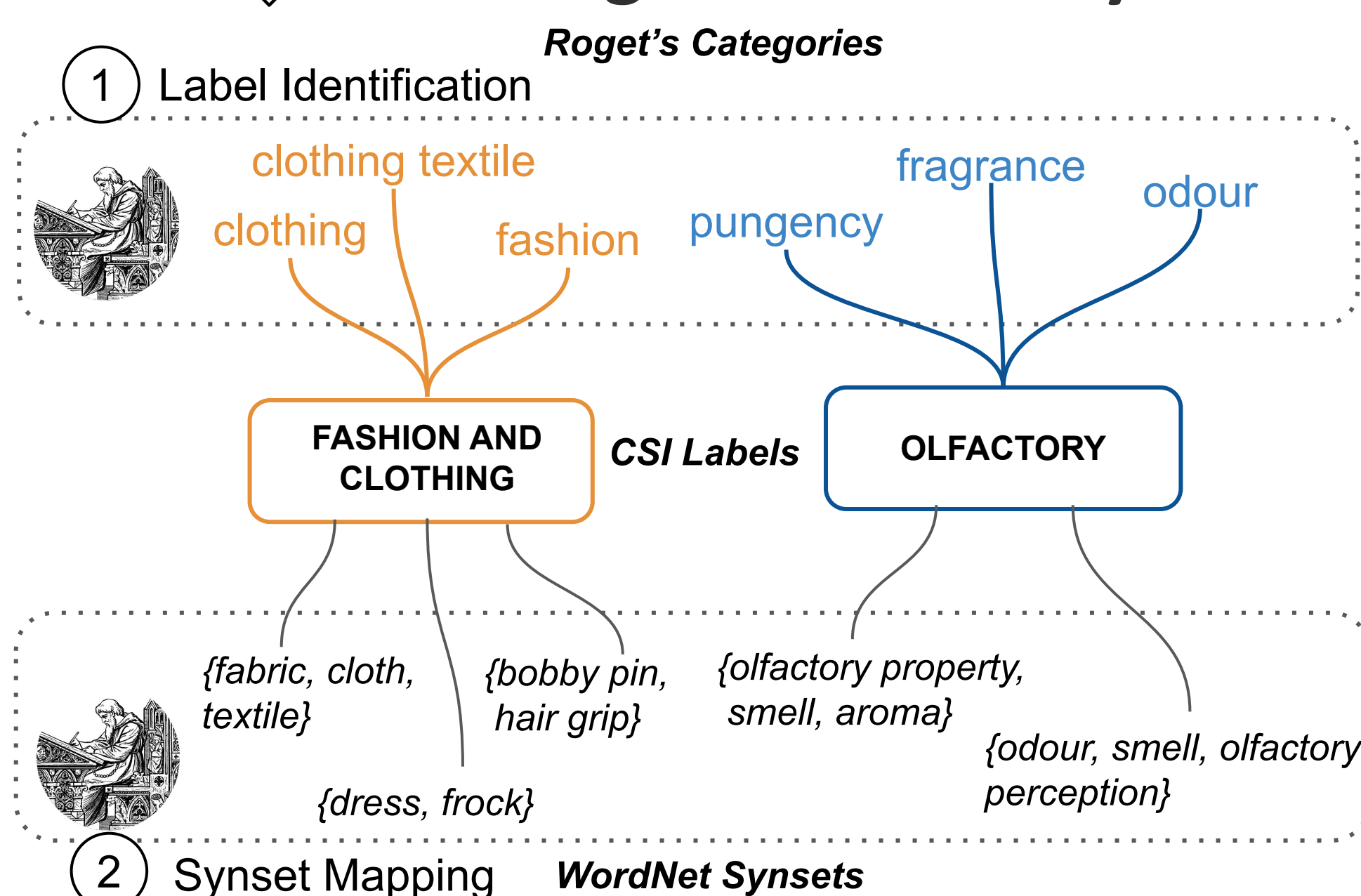
- ✓ Semi-automatic
- ✓ 200 labels
- ✓ All parts of speech included
- ✗ *Factotum* label covers more than 18% of WordNet synsets

### SuperSenses (SuS)

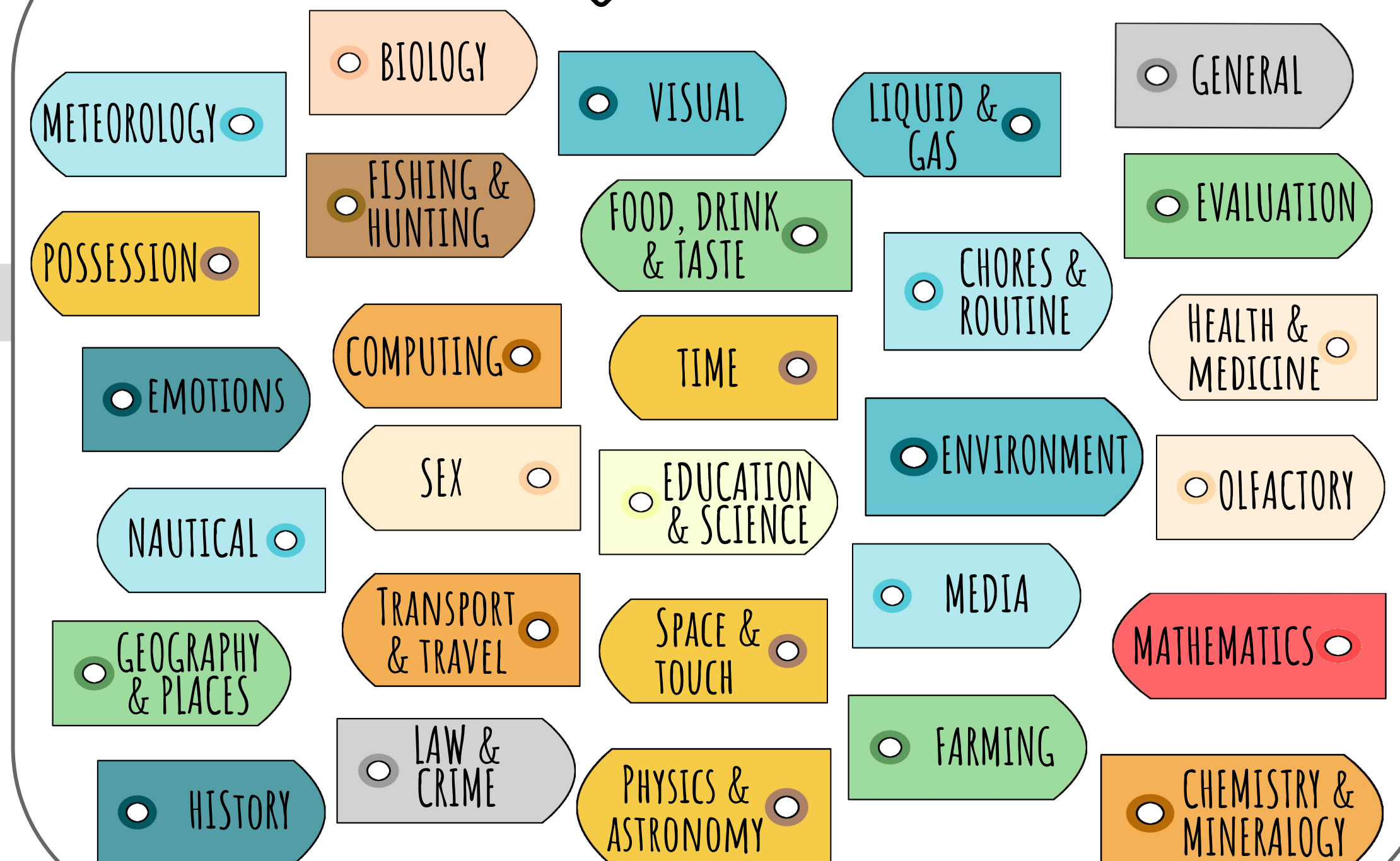
- ✓ Manual
- ✓ 45 labels
- ✗ Only nouns and verbs are meaningfully clustered



## Building the Inventory



## Labels



## Descriptiveness Task

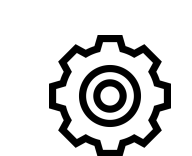
200 target words from SemCor annotated by 3 annotators.

The street drowsed on an August afternoon in the **shade** of the curbside tree, and **silence** was a weight.

- ★★★★ Physics & astronomy
- ★★★ Noun.state
- ★ Factotum
- ★★★★ Acoustics
- ★★★ Music, sound & dancing
- ★ Noun.attribute

CSI WND SuS

	K-IAA [0-1]	Descriptiveness [1-3]
CSI	<b>0.81</b>	<b>2.23</b> ★★★★★
WND	0.74	1.80 ★★
SuS	0.69	2.04 ★★★★★
WordNet	0.51	-



## Experimental Setup



BERT contextualized embeddings + dense layer



BERT contextualized embeddings + bi-directional LSTM



ELMo contextualized embeddings + bi-directional LSTM

Training: SemCor

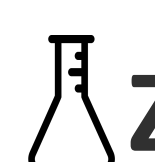
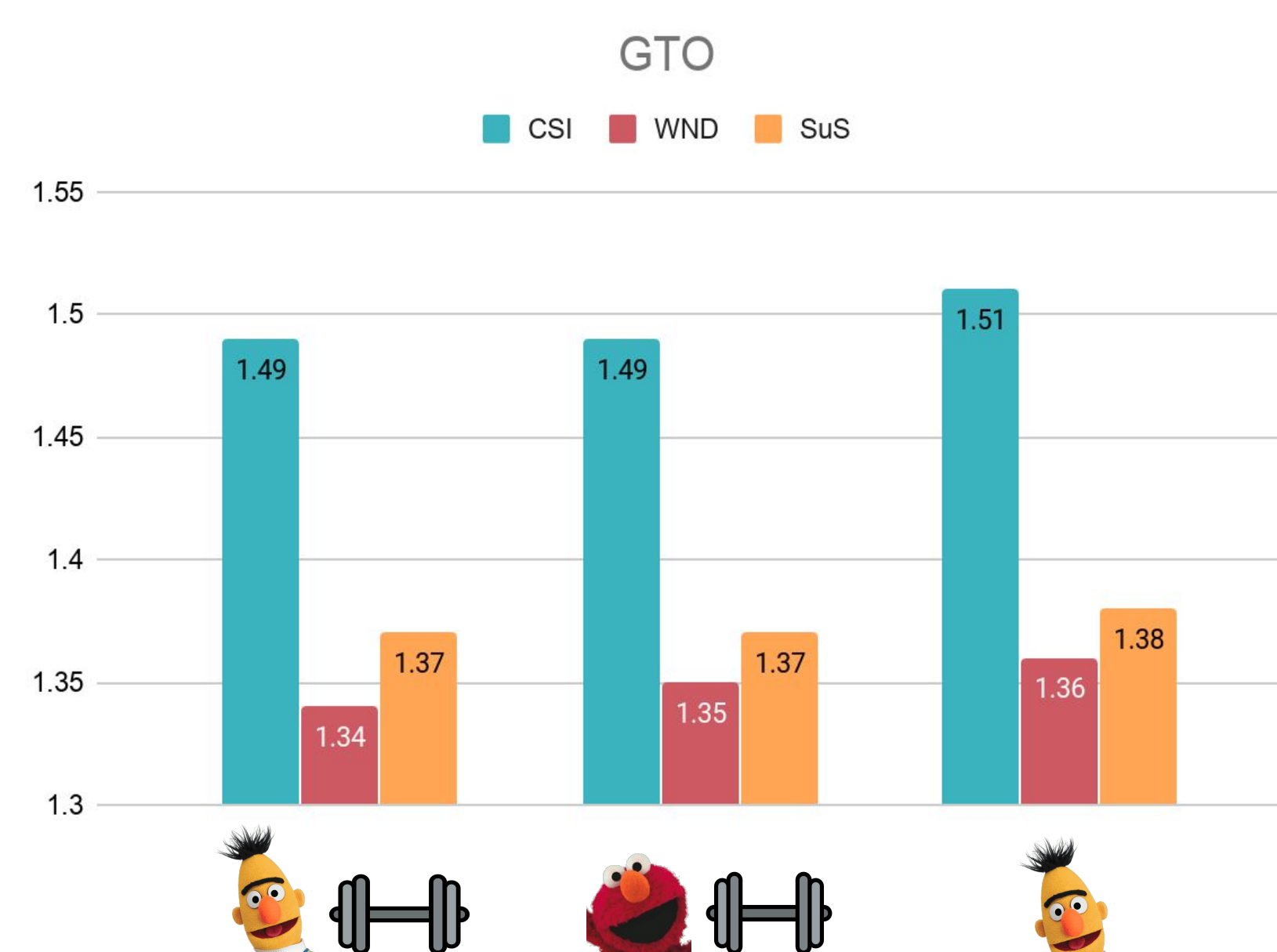
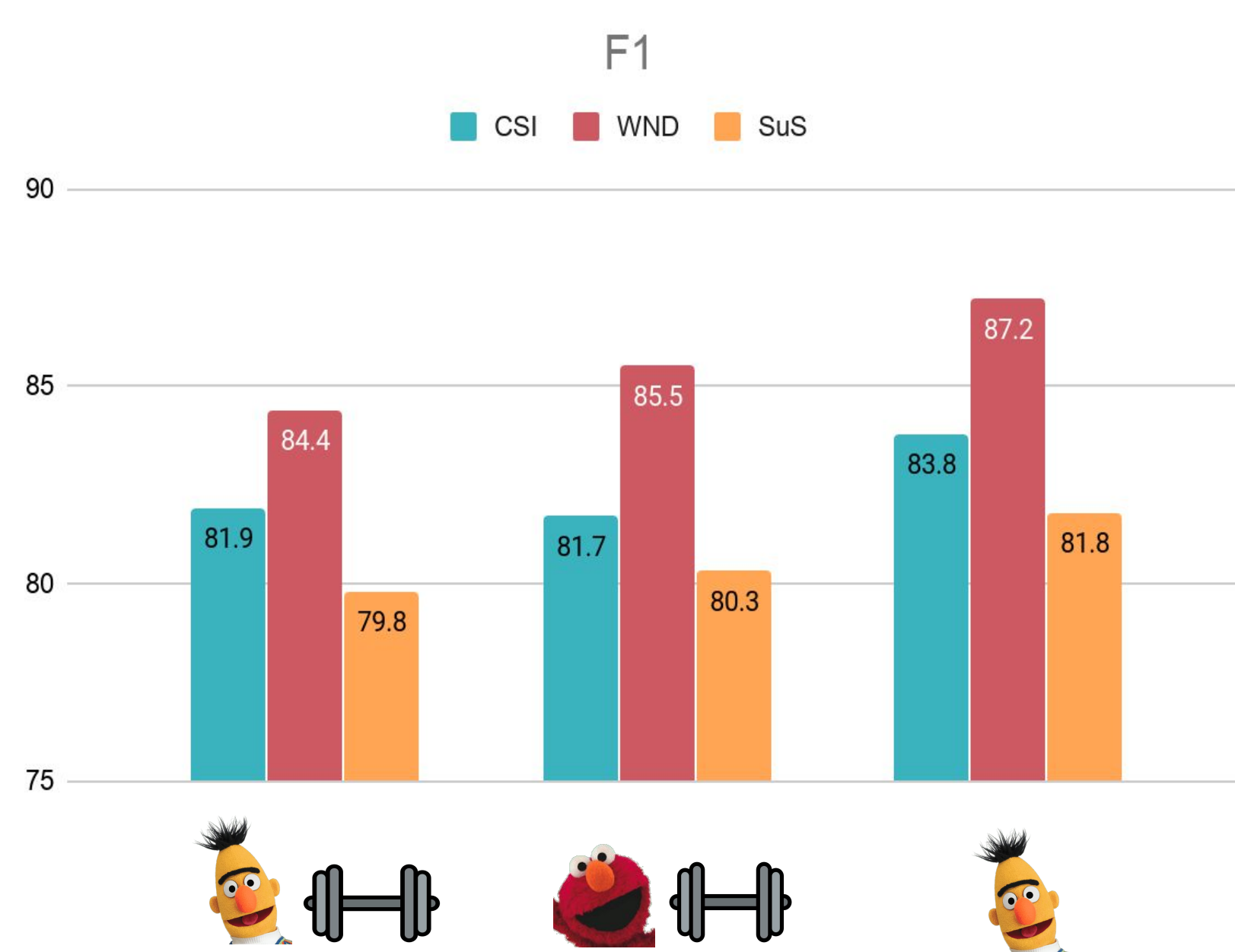
Dev: SemEval-07

Test: Senseval-2 + Senseval-3 + SemEval-13 + SemEval-15 (ALL)



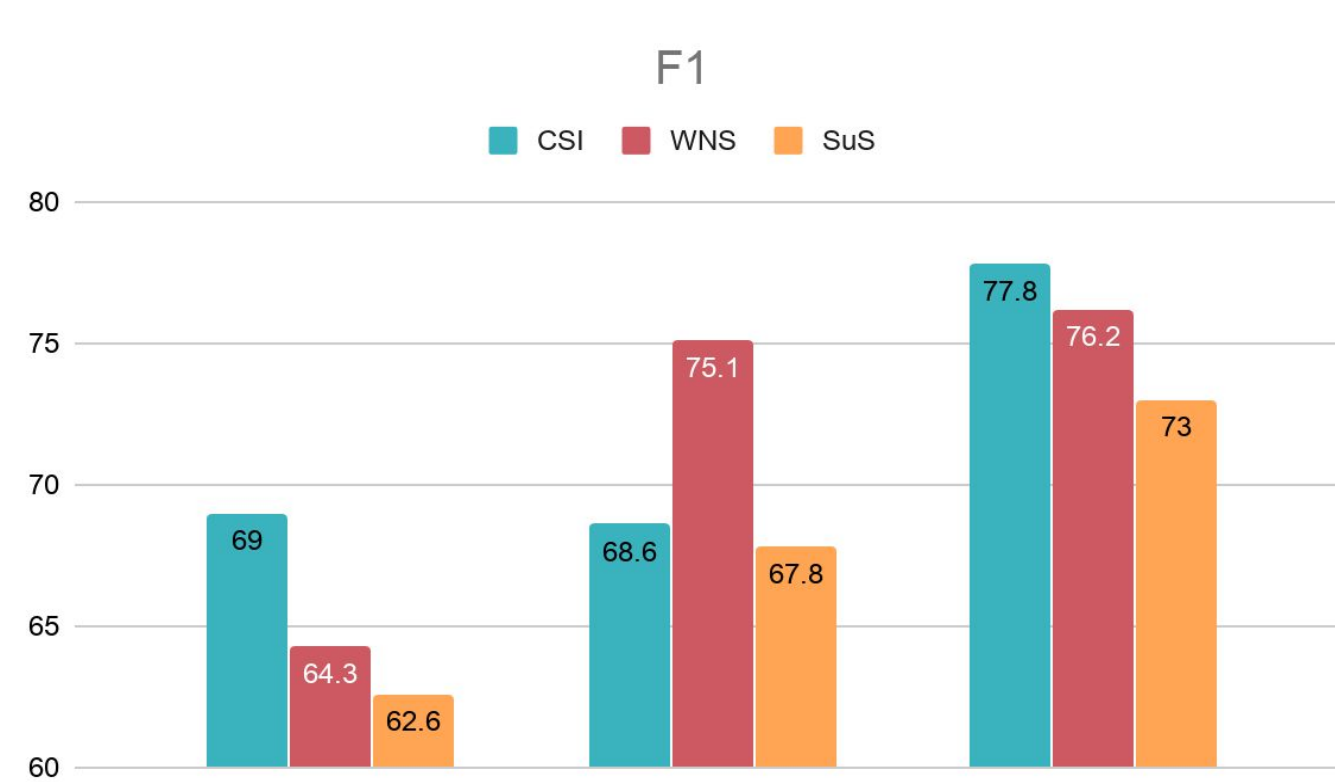
## English All-Words WSD

Results on the ALL dataset in terms of F1 and perplexity/F1 trade-off (GTO).

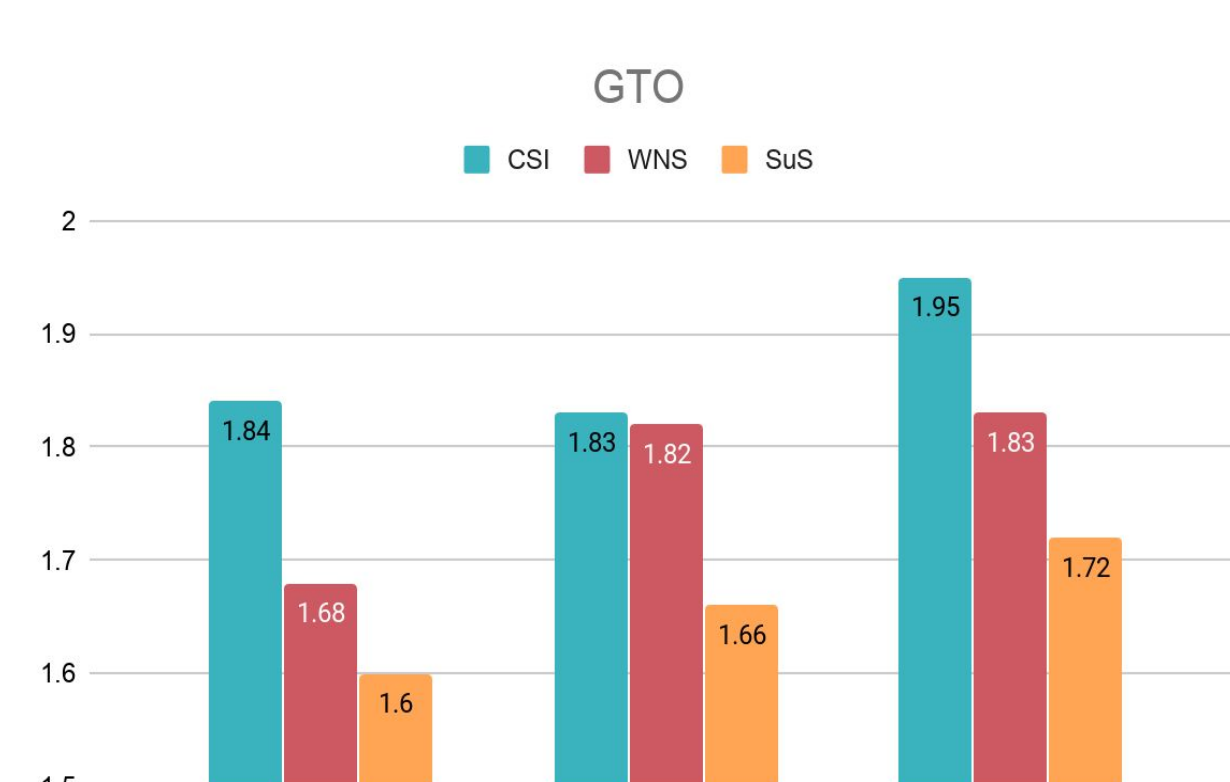


## Zero- and Few-Shot Learning

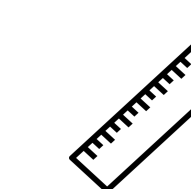
CSI increases the **sample efficiency** of a supervised WSD model.



CSI enables BERT to attain the best performance on **out-of-vocabulary** words.



T<sub>n</sub> : Training corpus with *n* annotated sentences per target word



## Metrics

- We need to consider at the same time the **difficulty** of the task and the **performance** of the model.
- Geometric Trade-Off (GTO): geometric mean of F1 and the perplexity (PPL) of a random guessing model:

$$GTO = \sqrt{F_1 \cdot PPL}$$



## Achievements

**Descriptive labels**, that are easy to use.

**Trade-off** between granularity and performance.

**Reduce** the need for manually annotated data.