

- Word Sense Disambiguation!

Kiamehr Rezaee

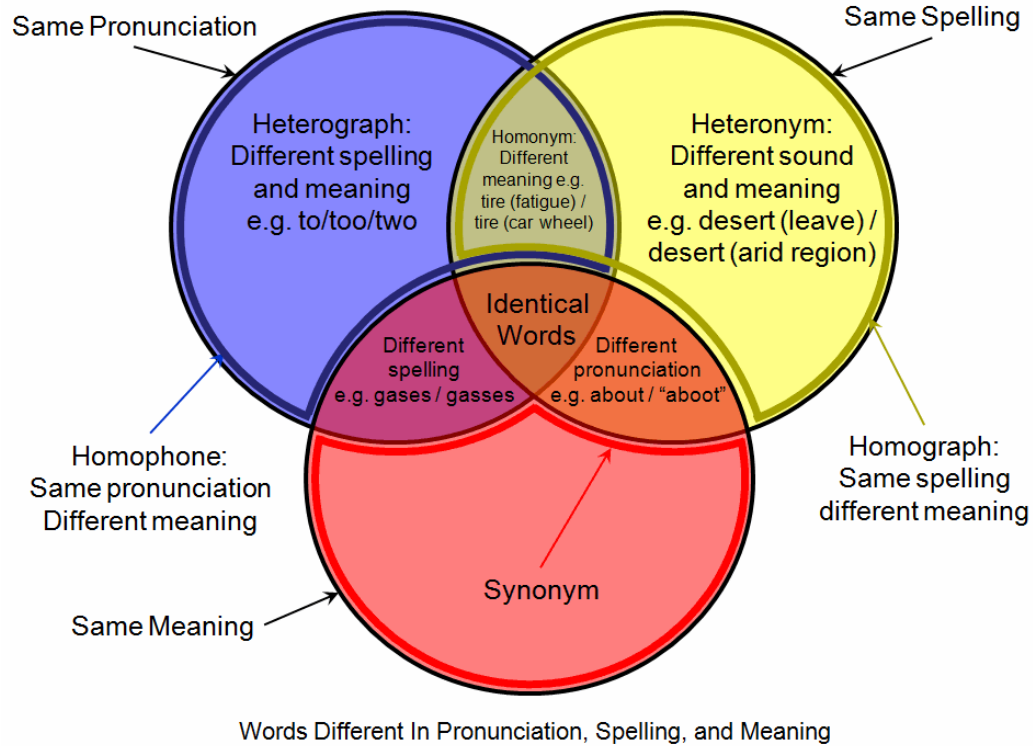


1

Ambiguous Words

What is the main problem?

Ambiguous Words



- Ambiguous Words
- I've got a gaming mouse and a pet mouse!

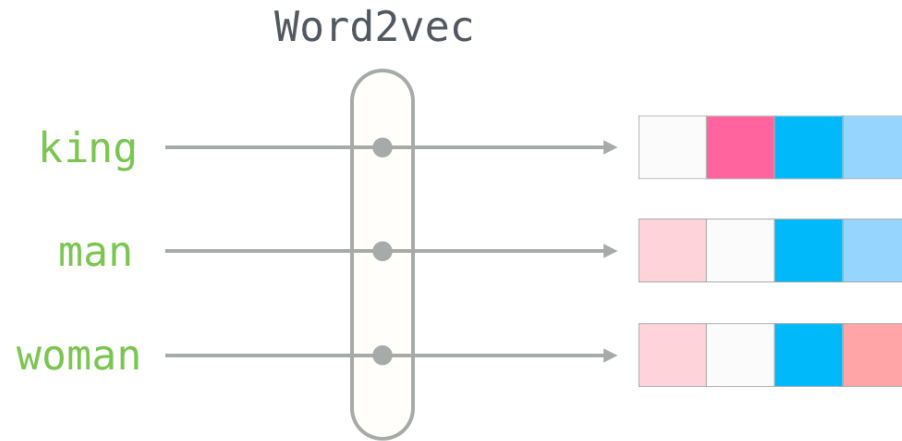


● Examples

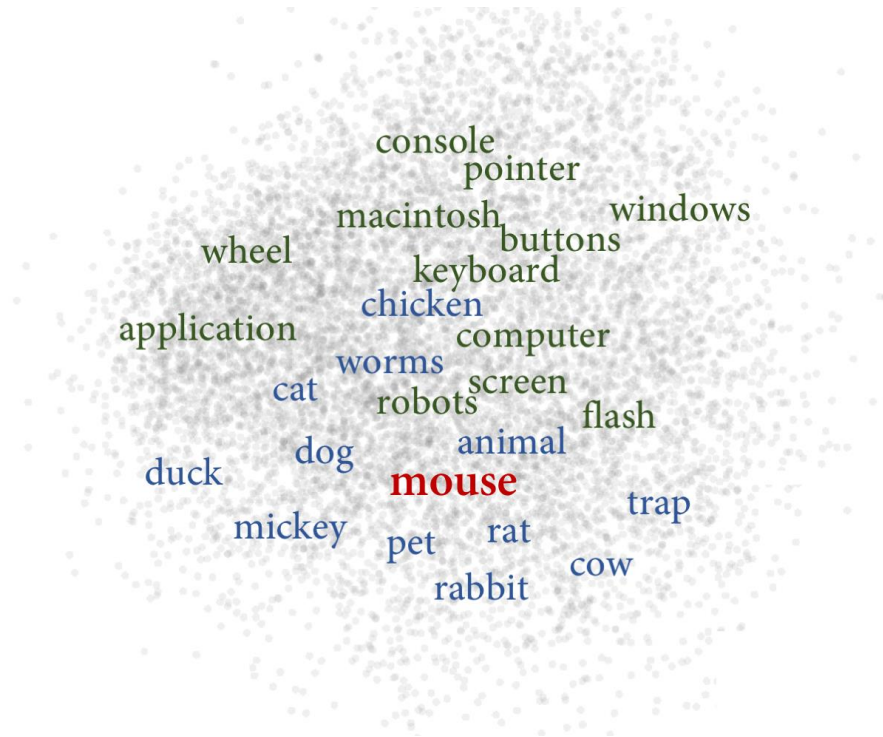
- - Machine Translation (play in Spanish)
- Document Retrieval (Jaguar)
- Text to Speech (lead)
- Spell Correction (aid/aide)

- Meaning Conflation Deficiency

- Static word vectors



- Meaning Conflation Deficiency



Sense Inventories

WordNet

- ▶ S: (n) **mouse**
 - ▷ S: (n) rodent, gnawer
- ▶ S: (n) shiner, black eye, **mouse**
 - ▷ S: (n) bruise, contusion
- ▶ S: (n) **mouse**
 - ▷ S: (n) person, individual, someone, somebody, mortal, soul
- ▶ S: (n) **mouse**, computer mouse
 - ▷ S: (n) electronic device

Word Sense Disambiguation

WordNet 

- ▶ S: (n) **mouse**
 - ▷ S: (n) rodent, gnawer
- ▶ S: (n) shiner, black eye, **mouse**
 - ▷ S: (n) bruise, contusion
- ▶ S: (n) **mouse**
 - ▷ S: (n) person, individual, someone, somebody, mortal, soul
- ▶ S: (n) **mouse**, computer mouse
 - ▷ S: (n) electronic device

*"a **mouse** takes much more room than a trackball"*

- Resources

- Unified WSD Benchmark

	#Docs	#Sents	#Tokens	#Annotations	#Sense types	#Word types	Ambiguity
Senseval-2	3	242	5,766	2,282	1,335	1,093	5.4
Senseval-3	3	352	5,541	1,850	1,167	977	6.8
SemEval-07	3	135	3,201	455	375	330	8.5
SemEval-13	13	306	8,391	1,644	827	751	4.9
SemEval-15	4	138	2,604	1,022	659	512	5.5
SemCor	352	37,176	802,443	226,036	33,362	22,436	6.8
OMSTI	-	813,798	30,441,386	911,134	3,730	1,149	8.9

- Resources

- Word in Context

F There's a lot of trash on the *bed* of the river — I keep a glass of water next to my *bed* when I sleep

F *Justify* the margins — The end *justifies* the means

T *Air* pollution — Open a window and let in some *air*

T The expanded *window* will give us time to catch the thieves — You have a two-hour *window* of clear weather to finish working on the lawn

2

Systems

How to solve the problem?

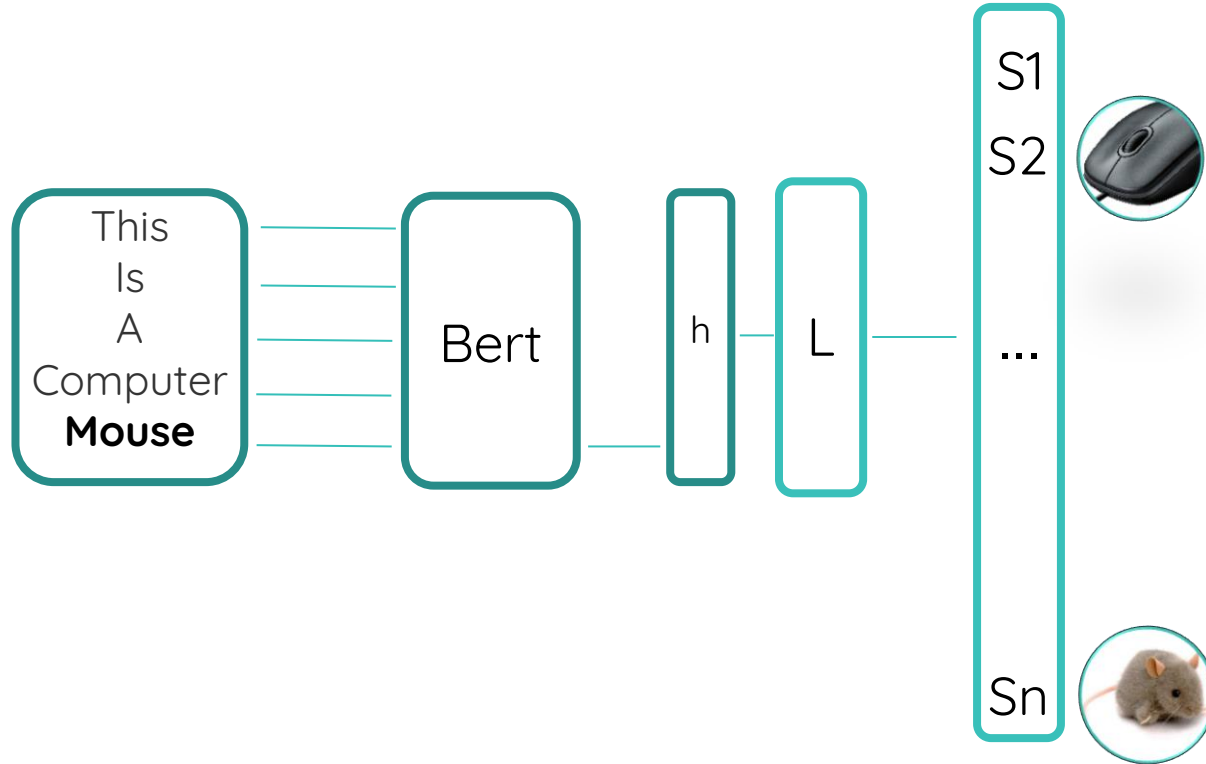
● Paradigms

- Knowledge-Based
- Supervised
- Hybrid

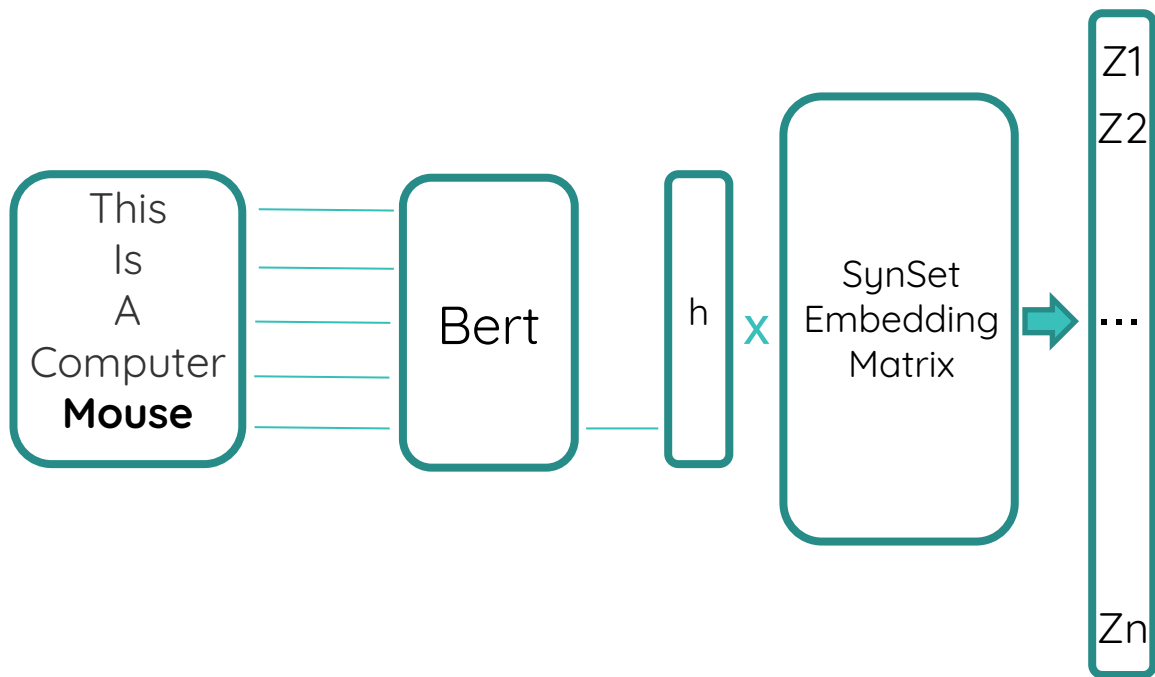
Paradigms

Type	System	SE2		SE3		SE07		SE13		SE15		ALL		
		FN	CS	FN	CS	FN	CS	FN	CS	FN	CS	FN	CS	
KB	Lesk _{ext} +emb	63.0	74.9	63.7	75.5	56.7	71.6	66.2	77.4	64.6	73.9	63.7	75.3	
	Babel _{fy} †	67.0	78.4	63.5	77.5	51.6	68.8	66.4	77.0	70.3	79.1	65.5	77.3	
	TM	69.0	-	66.9	-	55.6	-	65.3	-	69.6	-	66.9	-	
	UKB	68.8	81.2	66.1	78.1	53.0*	70.8	68.8	79.1	70.3	77.4	67.3*	78.7*	
	SyntagRank	71.6	-	72.0	-	59.3	-	72.2	-	75.8	-	71.7	-	
Supervised	SVM	IMS	70.9	81.5	69.3	80.8	61.3	74.3	65.3	77.4	69.5	75.7	68.4	79.1
		IMS+emb	72.2	82.8	70.4	81.5	62.6	75.8	65.9	76.9	71.5	76.7	69.6	79.8
	1NN	Context2vec	71.8	82.6	69.1	80.5	61.3	74.5	65.6	78.0	71.9	76.6	69.0	79.7
		ELMo	71.6	82.8	69.6	80.9	62.2	74.7	66.2	77.7	71.3	77.0	69.0	79.6
		BERT-Base	75.5	84.9	71.5	81.4	65.1	78.9	69.8	82.1	73.4	78.1	72.2	82.0
		BERT-Large	76.3	84.8	73.2	82.9	66.2	80.0	71.7	83.1	74.1	79.1	73.5	82.8
		ALBERT-XXL	76.6	85.6	73.1	82.6	67.3	80.1	71.8	83.5	74.3	78.3	73.7	83.0
Hybrid	Seq2Seq Att+Lex+PoS	70.1	-	68.5	-	63.1*	-	66.5	-	69.2	-	68.6*	-	
	Sense Compr. Ens.	79.7	-	77.8	-	73.4	-	78.7	-	82.6	-	79.0	-	
	LMMS ₁₀₂₄	75.4	-	74.0	-	66.4	-	72.7	-	75.3	-	73.8	-	
	LMMS ₂₀₄₈	76.3	84.5	75.6	85.1	68.1	81.3	75.1	86.4	77.0	80.8	75.4	84.4	
	EWISER	73.8	-	71.1	-	67.3*	-	69.4	-	74.5	-	71.8*	-	
	KnowBert† _{WN+WK}	76.4	85.6	76.0	85.1	71.4	82.6	73.1	83.8	75.4	80.2	75.1	84.1	
	GlossBERT	77.7	-	75.2	-	72.5*	-	76.1	-	80.4	-	77.0*	-	
	BEM	79.4	-	77.4	-	74.5*	-	79.7	-	81.7	-	79.0*	-	
	EWISER†	80.8	-	79.0	-	75.2	-	80.7	-	81.8*	-	80.1*	-	
-	MFS Baseline	65.6	77.4	66.0	77.8	54.5	70.6	63.8	74.8	67.1	75.3	64.8	76.2	

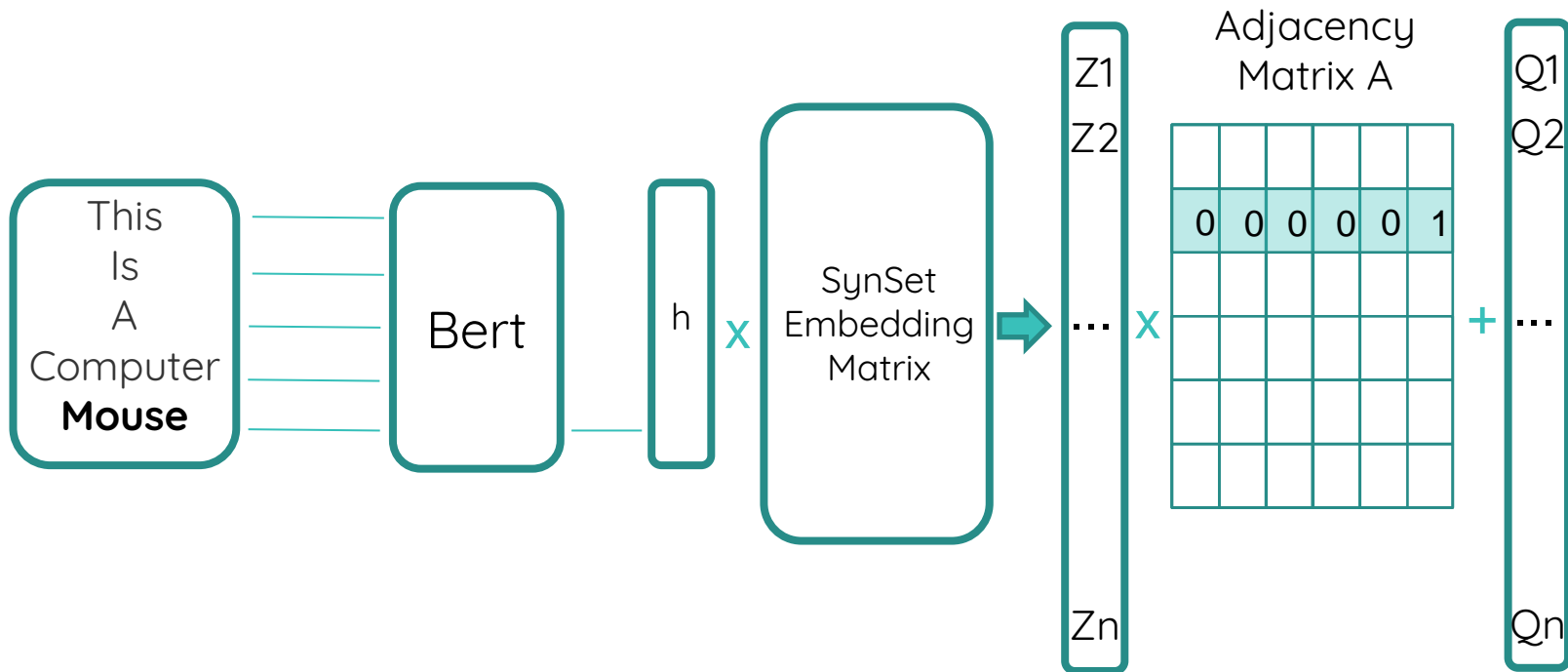
EWISER



EWISER



EWISER



3

Contributions

What I did!



CoarseWSD-20

Language Models and Word Sense Disambiguation: An
Overview and Analysis



WiC-TSV

WiC-TSV: An Evaluation Benchmark for Target Sense
Verification of Words in Context



ANY QUESTIONS?

Thanks!