

¿Capturan los modelos del lenguaje el significado?

Reflexiones desde la lingüística

Motivación: Definir el significado de las palabras



Proceso cognitivo interno



Proceso social



Interacción con referencias tangibles

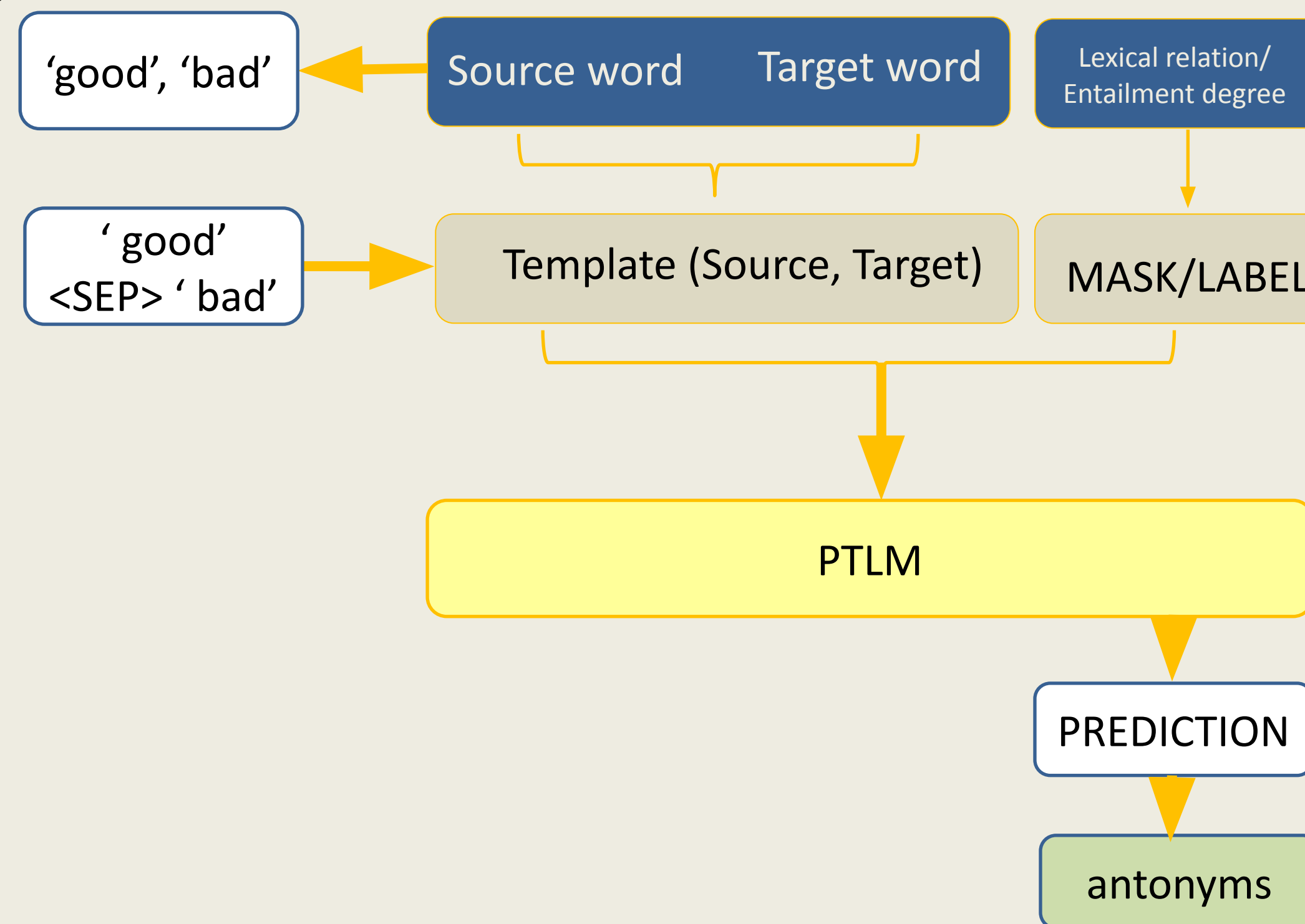


¿Basta con conocer qué terminos rodean una palabra?

Escandell Vidal, M. V. (2004). Fundamentos de semántica composicional (1. ed). Ed. Ariel.

RQ 1: semántica simple

Relaciones lexico-semánticas



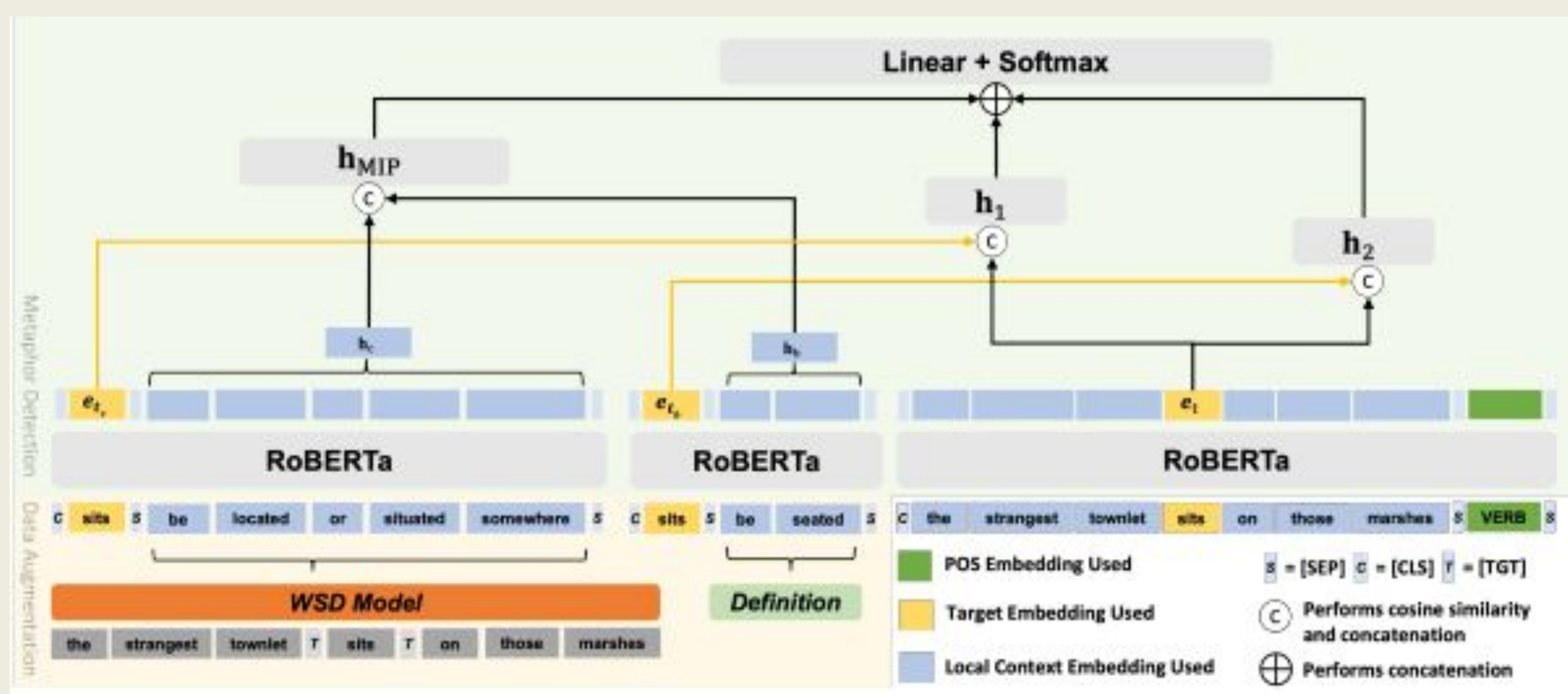
	K&H+N	BLESS	EVAL	ROOT9
RoBERTa				
T1	0.989	0.954	0.764	0.936
T2	0.989	0.955	0.757	0.936
T3	0.989	0.956	0.771	0.937
T4	0.312	0.133	0.087	0.934
TM1	0.988	0.947	0.761	0.936
TM2	0.988	0.946	0.764	0.928
TM3	0.985	0.951	0.746	0.926
SoTA				
LexNET	0.985	0.893	0.600	0.813
KEML	0.993	0.944	0.660	0.878
SphereRE	0.990	0.938	0.620	0.861
RelBERT	0.949	0.921	0.701	0.910

No clues good clues: out of context Lexical Relation Classification (Pitarch et al., ACL 2023)

Incluso con prompts muy sencillos (con mínima información) los ML demuestran capturar información semántica simple (e.g., sinonimia y antonimia)

RQ 2: semántica compleja

Lenguaje figurado



ContrastWSD: Enhancing Metaphor Detection with Word Sense Disambiguation Following the Metaphor Identification Procedure (Elzohbi & Zhao, LREC-COLING 2024)

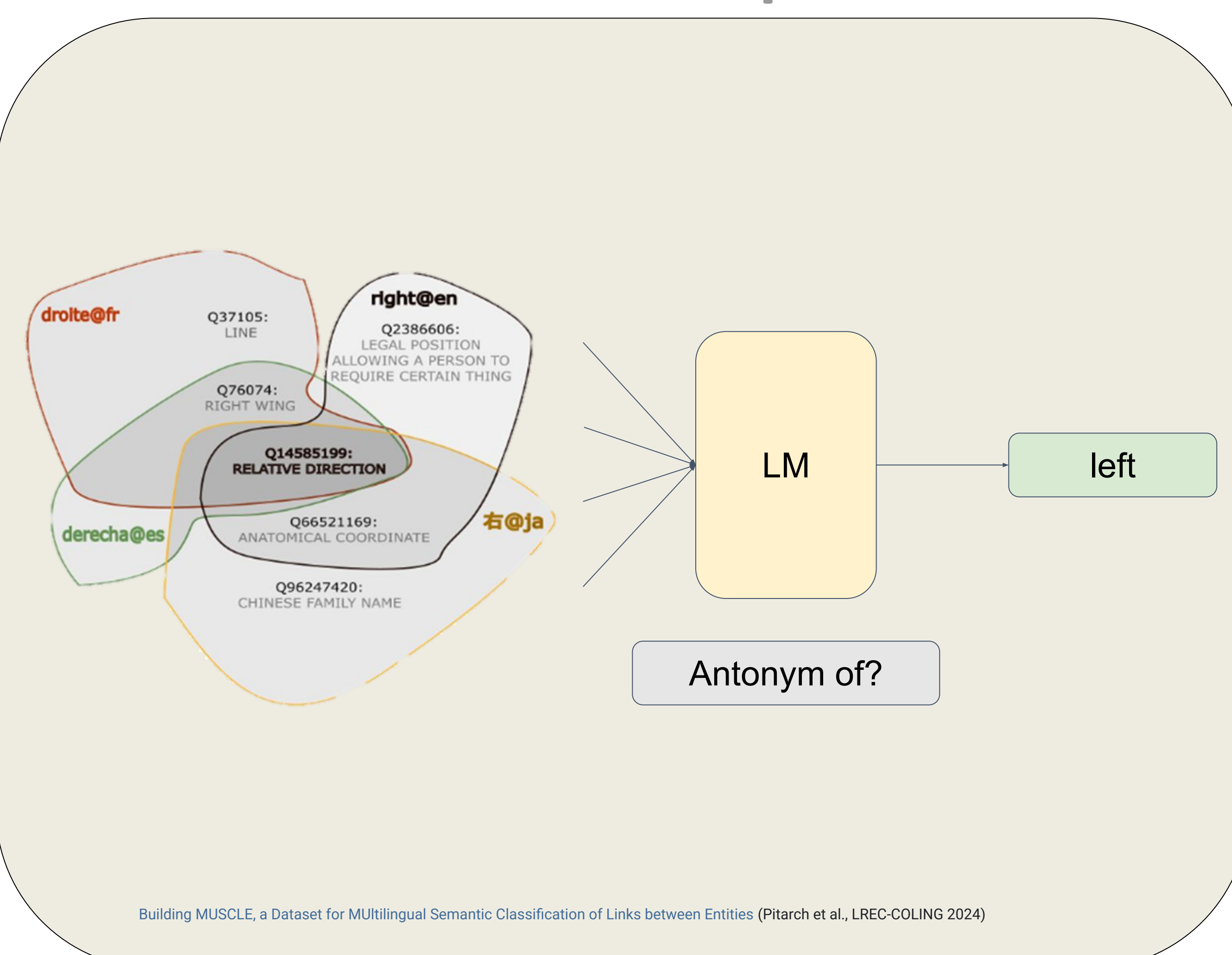
Dataset	Model	Rec	Prec	F1
VUA-18	MeiBERT	77.5	79.87	78.66
	MsW_cos	77.88	80.31	79.07
	FrameBERT	76.78	79.33	78.03
	ContrastWSD	78.85	80.16	79.50
VUA-18 (-)	MeiBERT	73.34	71.44	72.35
	MsW_cos	70.52	74.77	72.55
	FrameBERT	70.33	73.81	72.02
	ContrastWSD	75.23	72.38	73.77
VUA-20	MeiBERT	69.51	75.58	72.40
	MsW_cos	69.98	75.64	72.70
	FrameBERT	69.30	75.62	72.31
	ContrastWSD	69.89	76.60	73.09
VUA-20 (-)	MeiBERT	74.61	71.26	72.88
	MsW_cos	77.22	68.39	72.51
	FrameBERT	75.07	71.59	73.28
	ContrastWSD	75.57	72.95	74.22

Incluso con arquitecturas mucho más complejas, relaciones complejas como la metáfora que necesita de conocimiento contextual y social, se mantienen como retos a resolver en los modelos del lenguaje.

Solución: Exploración métodos híbridos

+ info socio-cultural: Mapas semánticos

+ datos cognitivos



Building MUSCLE, a Dataset for Multilingual Semantic Classification of Links between Entities (Pitarch et al., LREC-COLING 2024)

word	Original_an notation	definitions	Prec	ICPrec	Img	Conc	Phy
chicken	1	the flesh of a chicken used for food ()	5.0	3.41	4.12	4.21	3.14
chicken	1	a domestic fowl bred for flesh or eggs; believed to have been developed from the red jungle fowl ()	4.57	3.42	4.45	4.21	3.00
chicken	0	a person who lacks confidence, is irresolute and wishy-washy ()	5.62	0.16	4.79	2.98	3.13
chicken	0	a foolhardy competition; a dangerous activity that is continued until one competitor becomes afraid	4.07	1.63	4.72	2.78	2.68
chicken	0	easily frightened ()	0.0	Na	4.42	2.04	3.86

	Estimate	Standard Error	p
(Intercept)	1.205	0.208	< .001
nb17_cosines_definitions_mean	0.445	0.344	0.195
w2v_cosines_definitions_mean	-2.634	0.373	< .001
familiarity_definitions_max	-0.001	0.000	< .001
imageability_definitions_max	-0.043	0.020	0.033
concreteness_definitions_max	-0.048	0.020	0.013
physicality_definitions_max	-0.028	0.026	0.276
precision_definitions_max	-0.046	0.007	< .001
ic_precision_definitions_max	-0.000	0.000	< .001