

Multilingual WSD-like Constraints for Paraphrase Extraction



Wilker Aziz[†] and Lucia Specia[§]

[†] Research Group in Computational Linguistics, University of Wolverhampton
w.aziz@wlv.ac.uk

[§] Department of Computer Science, University of Sheffield
l.specia@sheffield.ac.uk



The University Of Sheffield.

Pivoting

Phrases (e_1, e_2) that typically align to the same set of foreign phrases $F = f : e_1 \rightarrow e_2$ are said to be paraphrases (Bannard and Callison-Burch, 2005).

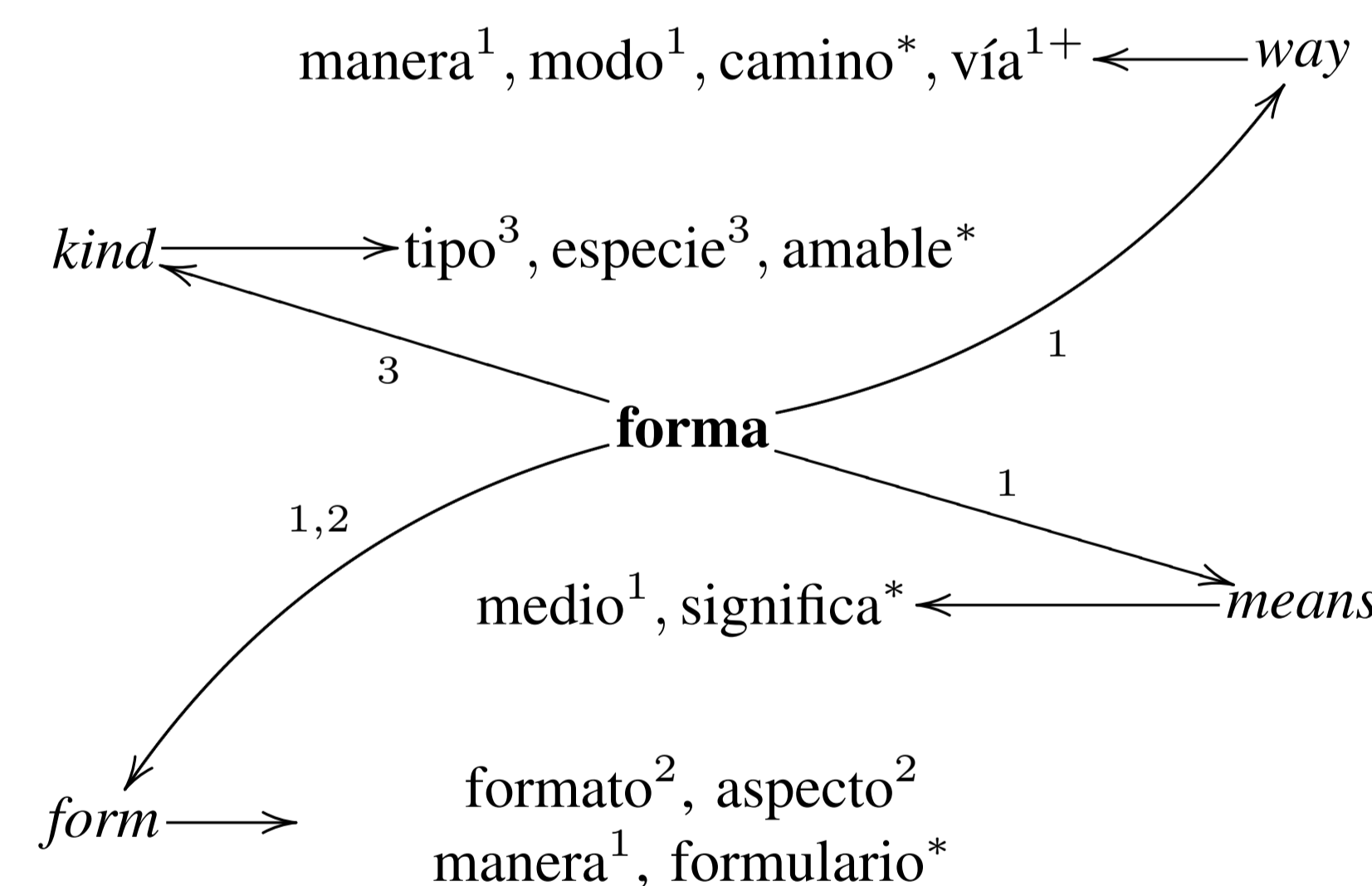
Era de vital importancia acabar con _{e1} los conflictos y evitar que el país ...	I was incredibly important immediately to terminate _f the conflicts and prevent the country from ...
considerando que sería difícil poner fin _{e2} bruscamente a unas medidas ...	whereas it would be difficult to terminate _f abruptly measures ...

The model

$$p(e_2|e_1) = \sum_{f \in F} p(f|e_1)p(e_2|f)$$

Pitfalls

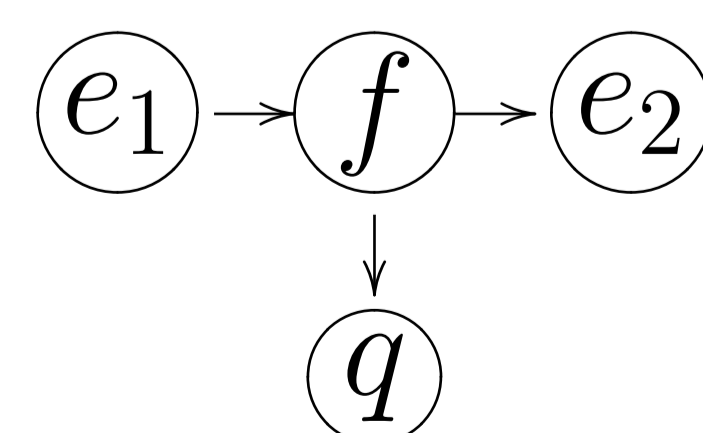
- different senses of the input are mixed together in one distribution
- very low probability to less frequent senses
- certain senses become unreachable



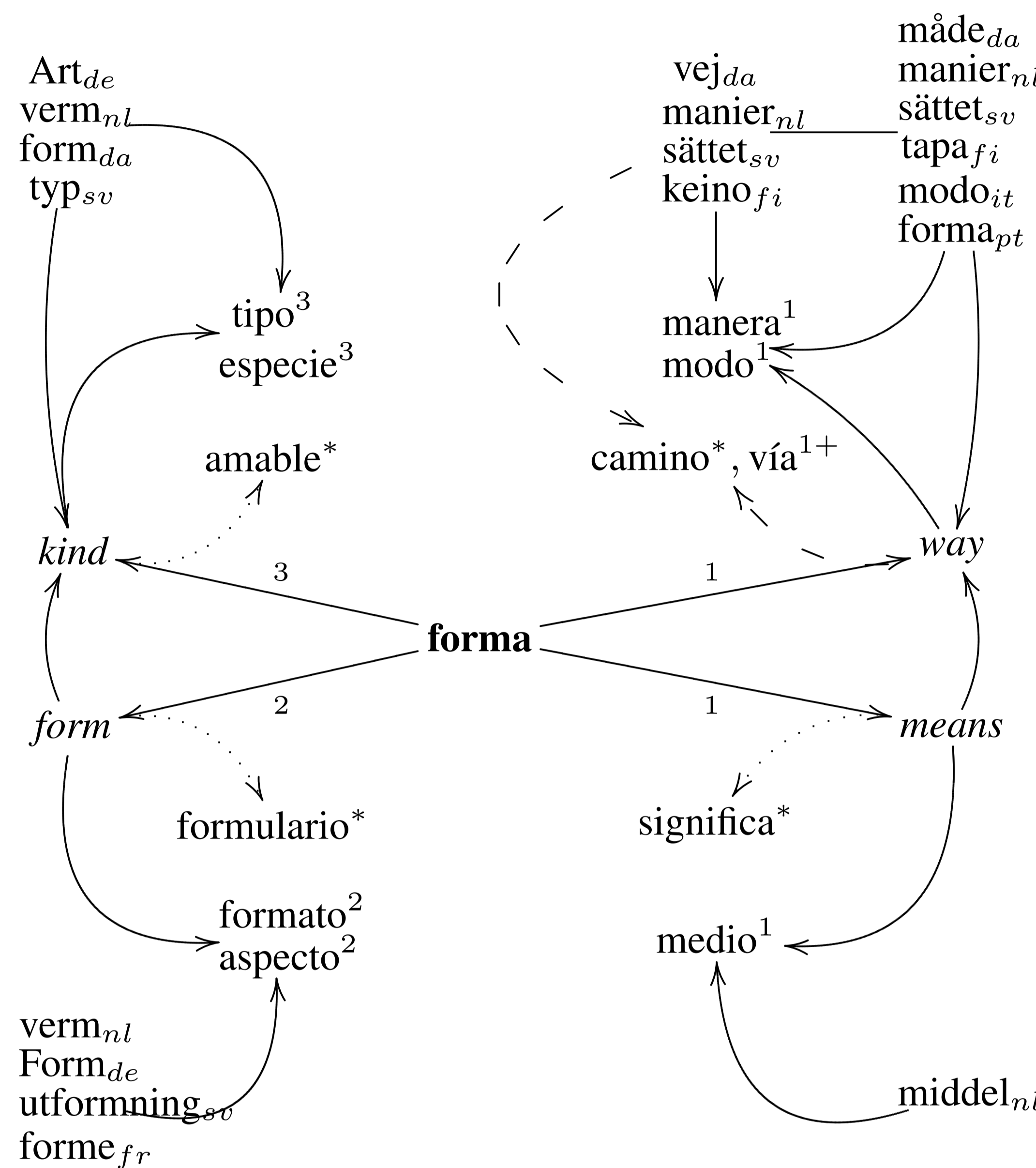
Pivoting with WSD-like constraints

The *WSD assumption* is that different translations of a phrase discriminate different senses of the input.

The model



$$p(e_2|e_1, q) = \frac{1}{z} \sum_{f \in F} p(e_2|f)p(q|f)p(f|e_1)$$



Phrases to paraphrase

Unambiguous	Ambiguous
concreto, fondos, haber, amor propio, sangre fría, dar a luz, dar con, tomar el pelo	derecho, comercial, real, particular, legal, justo, común, cerca, esencial, especial, fuerte, puesto, oficial, figura, informe, parte, cuenta, forma, claro, clave, tiempo, seguro, respuesta, trabajar, responder, garantizar, volver, aumentar, incluir, tratar, ofrecer, establecer, pasar, dejar, realizar, punto de vista, llevar a cabo, dar vueltas, tener que, acabar con

Paraphrasing in context

La idea de conceder a la Unión Europea su propia competencia fiscal - la **palabra clave** es el "impuesto por Europa" - está siendo debatida.
The idea of granting the EU its own tax competence - the **keyword** is the "Europe tax" - is being discussed.

Experiments

Models

CCB

standard pivoting (9 languages)

CCB-wsd

pivoting constrained to one pivot phrase (Callison-Burch, 2007)

multi

pivoting with WSD-like constraints (9 languages)

LM re-ranking (5-gram)

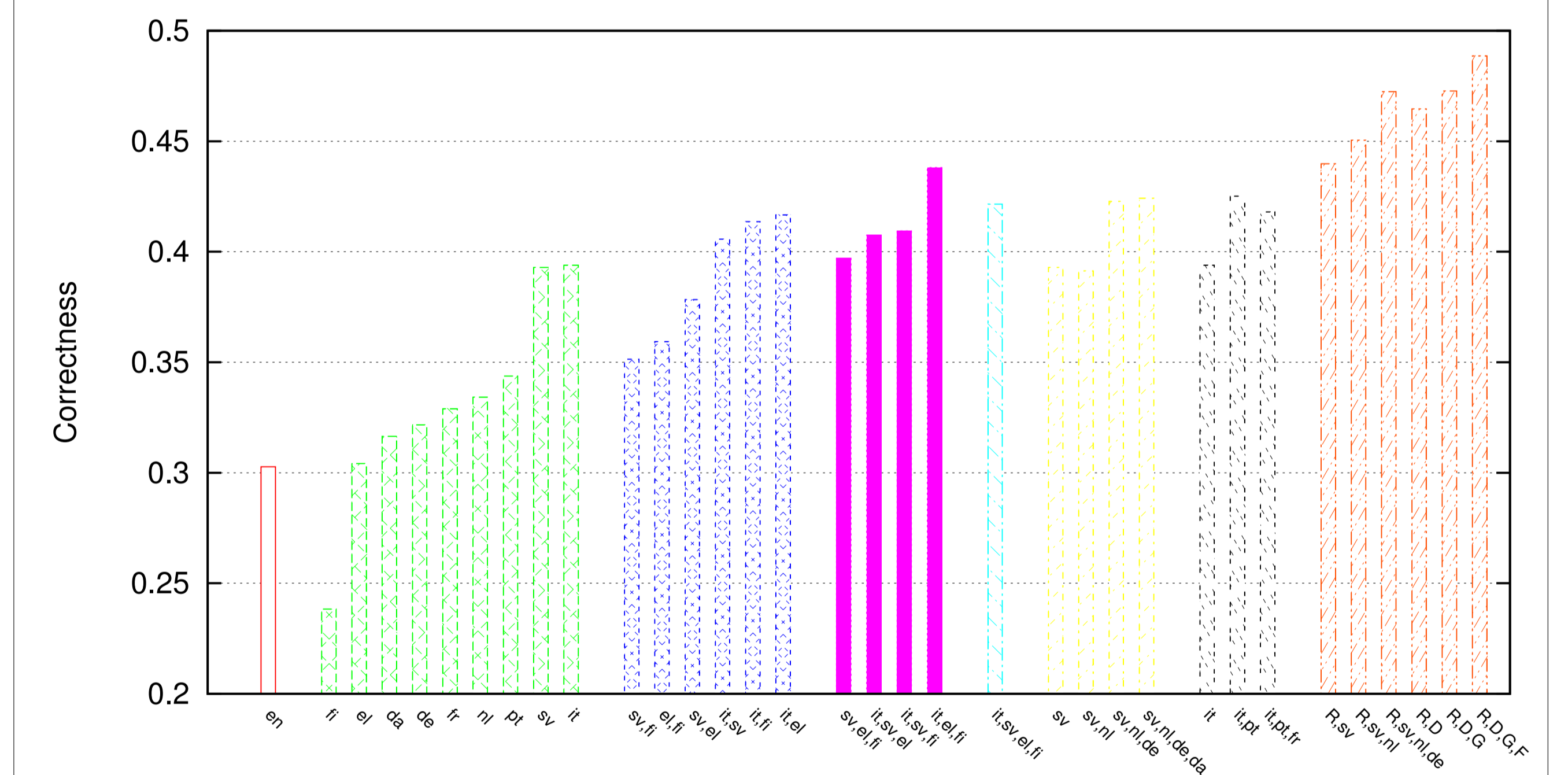
Paraphrasing in context

Method	Top	Human Translation				Machine Translation					
		M	G	Correct	F1	M	G	Correct	F1		
CCB	1	32	28	25	25	25					
CCB-wsd	1	61	38	34	28	30	71	39	34	32	33
multi	1	62	55	59	42	49	69	55	50	45	48
CCB	2	41	37	33	33	33					
CCB-wsd	2	68	44	40	33	36	79	46	40	38	39
multi	2	71	64	66	47	55	82	69	63	57	60
CCB	3	46	42	37	37	37					
CCB-wsd	3	71	47	45	36	40	83	50	44	41	42
multi	3	74	67	71	50	59	85	74	69	62	65

Without LM re-ranking

Method	Human Translation		Correct	Machine Translation		Correct
	M	G		M	G	
CCB	33	23	22			
CCB-wsd	19	9	8	16	8	7
multi	64	43	37	63	40	35

Contribution of different families



Correctness

Pivot Languages

- CCB-wsd
- 1 pivot
- 2 families
- 3 families
- 4 families
- Germanic (1-4)
- Romance (1-3)
- Romance-All (4-9)

Example of paraphrases

forma → kind		forma → means	
multi	tipo	multi	*amables
ccb-wsd	*amables	medio	*significa contar
multi	forma	através	medios que tene
ccb-wsd	*kind	instrumento	*significa
multi	especie		
ccb-wsd	especie		