# Machine Translation vs. Multilingual Approaches for Entity Linking

Henry Rosales-Méndez, Barbara Poblete and Aidan Hogan

Millenium Institute for Foundational Research on Data. Department of Computer Science, University of Chile

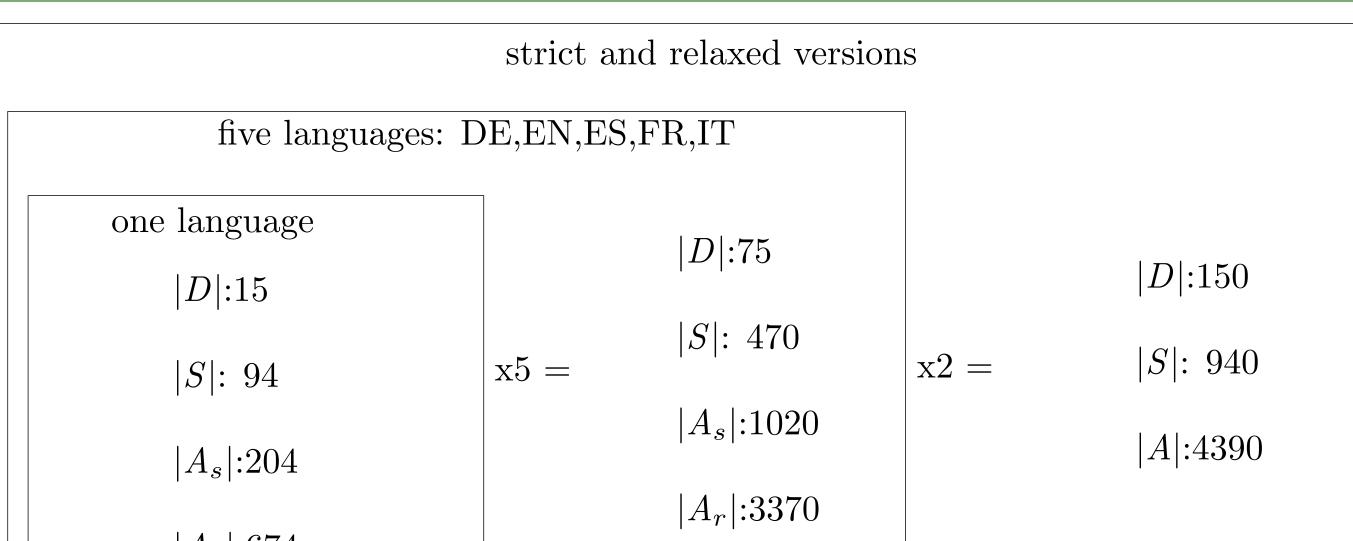


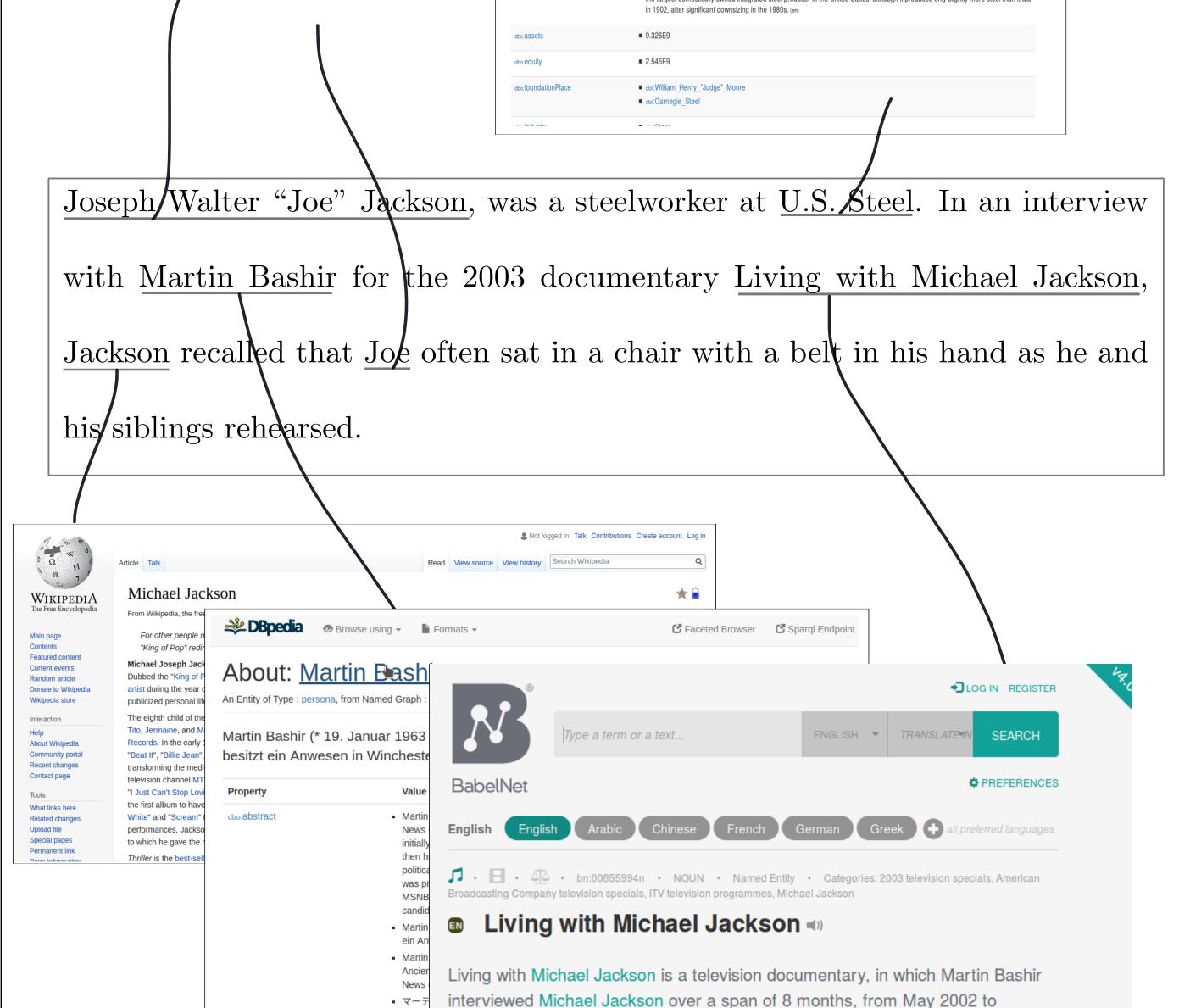
## What is *Entity Linking*?

#### Fig 1. Entity Linking example

	Item Discussio		d View history Search	n Wikidata	٩	
<b>IKIDATA</b>	Joe Jacks	son (Q361297)				
Main page Community portal Project chat Create a new item Recent changes Random item Query Service Nearby Help Donate	American music manager and father of the Jackson family Joseph Jackson   John Baptist, of Battersea Jackson   John Baptiste Jackson   Joannes Baptista Jackson   Jackson of B			Srowse us	sing 👻 🔓 Formats 👻	🖸 Faceted Browser 🛛 🖸 Sparql Endpoint
				About: U.S. Steel An Entity of Type : Public company, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org		
	Language	Label	Description	The United States Steel Corporation (NYSE: X), more commonly known as U.S. Steel, is an American integrated steel		
	English	Joe Jackson	American m father of the	producer with major production operations in the United States, Canada, and Central Europe. The company was the world's 15th largest steel producer in 2014.		
				Property	Value	
	/			dbo:abstract	major production operations in the United States, Canad producer in 2014. The company was renamed USX Corp the company spun off the oil & gas business of Marathor	ommonly known as U.S. Steel, is an American integrated steel producer with da, and Central Europe. The company was the world's 15th largest steel poration in 1986 and back to United States Steel Corporation in 2001 when n Oil and the steel business of U.S. Steel to shareholders. In 2001 it was still er in the United States, although it produced only slightly more steel than it did

#### Content of VoxEL 4

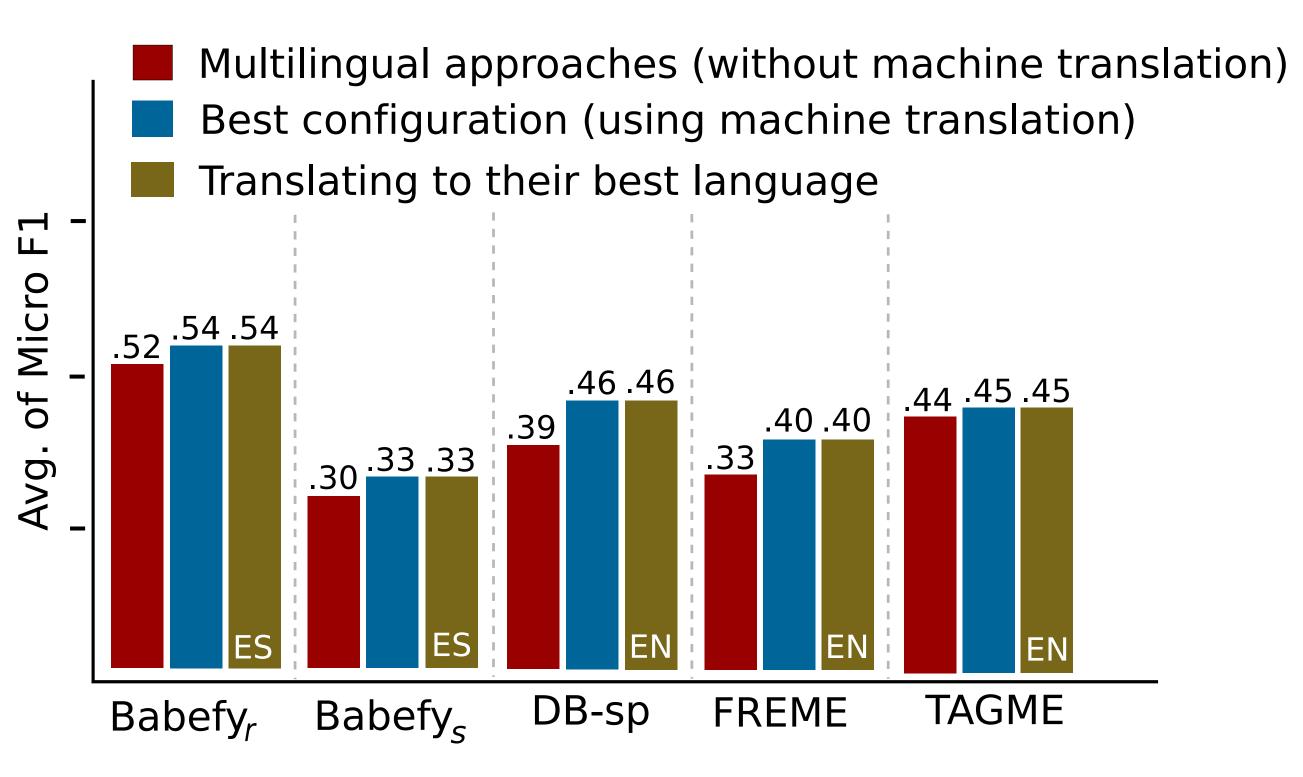




 $|A_r|$ :674

Could we not simply focus on supporting one language in 5 the EL system and translate the input text to that language?

Plotting the averaged scores of the systems with/without 6 machine translation.





January 2003, about different aspects of his life. IN Wikipedia One definitions

### 2 Are the Entity Linking systems targetting the same entities?

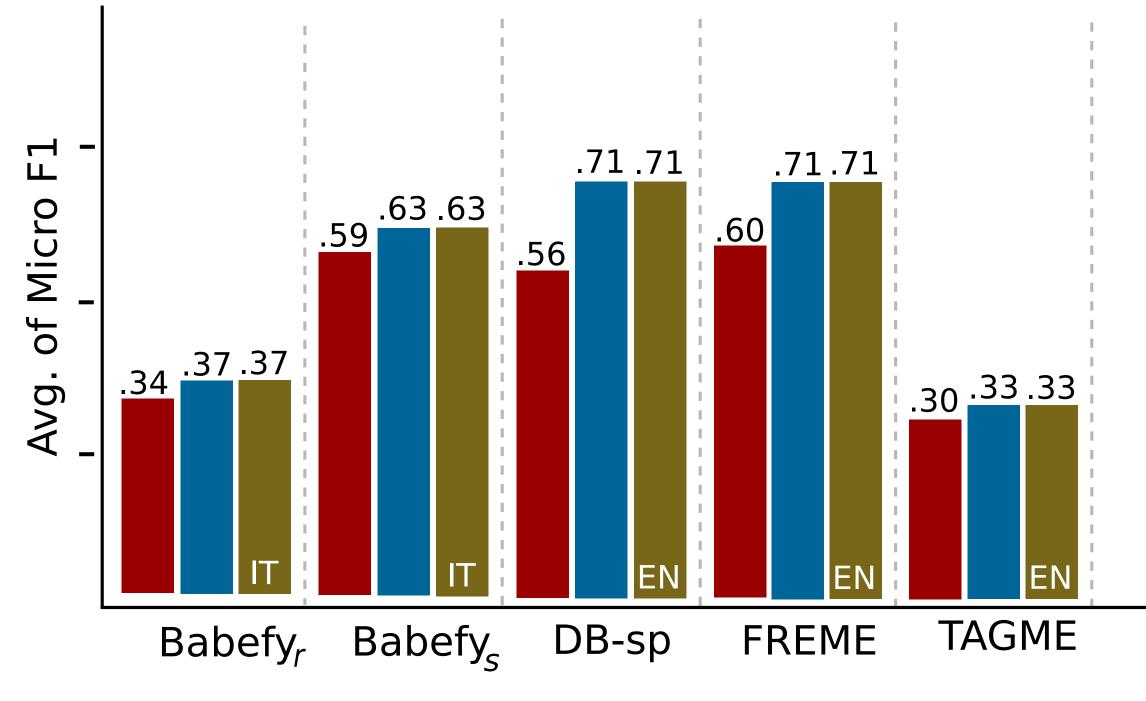
Fig 2. Example annotations produced by four EL systems: AIDA (A), Babelfy (B), DBpedia Spotlight (D) and TAGME (T).

$\boxed{ \begin{array}{c} \text{Joseph Walter "Joe" Jackson, was} \\ D \end{array} }$	a steelworker $\frac{D T}{D T}$	at <u>U.S.</u> Steel. $\frac{B D}{A B T}$	. In an interview $\frac{D T}{D T}$
with Martin Bashir for the 2003 $A D T$	$\frac{\text{documentary}}{D}$	$\frac{\text{Living with I}}{B T}$	Michael Jackson, A B D
$\frac{\text{Jackson}}{A B} \text{ recalled that } \frac{\text{Joe}}{A} \text{ often sat}$	in a chair with $\frac{\text{chair}}{T}$	th a $\frac{\text{belt}}{D T}$ in his	s $\frac{\text{hand}}{T}$ as he and
his <u>siblings</u> rehearsed. $\frac{D T}{D T}$			

#### VoxEL: Benchmark Dataset 3

- Curated multilinugal source: VoxEurop.eu.
- Same content in five languages.
- Sentences and annotations aligned cross-languages.

#### (a) Results of the Relaxed Version of VoxEL



(b) Results of the Strict Version of VoxEL

#### Conclusion

- In the context of the VoxEL dataset, all evaluated systems get better results using machine translation instead of using their own multilingual models.
- Surprisingly, in this experiment Babelfy obtains its best score when it is

#### • Two versions: relaxed and strict.

# Relaxed: only contains annotations of PER, ORG and PLACES. Strict: allows annotations to any Wikipedia page.

configured in Spanish/Italian for the relaxed/strict versions of VoxEL and translating the texts into these languages. On the other hand, DBpedia Spotlight, FREME and TAGME obtain their best value when they are configured in English.



Acknowledgements The work of Henry Rosales-Méndez was supported by CONICYT-PCHA/Doctorado Nacional/2016-21160017. The work was also supported by the Millennium Institute for Foundational Research on Data (IMFD) and by Fondecyt Grant No. 1181896.