

Multilingual Entity Linking: Comparing English and Spanish †

Henry Rosales-Méndez, Barbara Poblete and Aidan Hogan

University of Chile

{hrosales,bpoblete,ahogan}@dcc.uchile.cl

October 22nd, 2017

† LD4IE - Linked Data for Information Extraction Workshop.

Example

In November 1983 Michael Jackson and his brothers partnered with PepsiCo in a \$5 million promotional deal that broke records for a celebrity endorsement. The first Pepsi Cola campaign, which ran in the United States from 1983 to 1984 and launched its iconic "New Generation" theme, included tour sponsorship, public relations events, and in-store displays.

Example - DBpedia Spotlight

In [November](#) 1983 [Michael Jackson](#) and his [brothers](#) partnered with [PepsiCo](#) in a \$5 million [promotional deal](#) that [broke records](#) for a [celebrity endorsement](#). The first [Pepsi Cola](#) campaign, which ran in the [United States](#) from 1983 to 1984 and [launched](#) its [iconic "New Generation" theme](#), included [tour](#) sponsorship, [public relations events](#), and in-store displays.

Example - DBpedia Spotlight

<http://dbpedia.org/resource/Indium>
<http://dbpedia.org/resource/November>
http://dbpedia.org/resource/Michael_Jackson

In [November](http://dbpedia.org/resource/November) 1983 [Michael Jackson](http://dbpedia.org/resource/Michael_Jackson) and his [brothers](#) partnered with [PepsiCo](#) in a \$5 million [promotional deal](#) that [broke records](#) for a [celebrity endorsement](#). The first [Pepsi Cola](#) campaign, which ran in the [United States](#) from 1983 to 1984 and [launched](#) its [iconic "New Generation" theme](#), included [tour sponsorship](#), [public relations events](#), and in-store displays.

http://dbpedia.org/resource/The_Miami_Herald

Phases in Entity Linking

- 1 Entity Recognition.
- 2 Entity Disambiguation.

Name Variations in Entity Linking

Michael Joseph Jackson

Michael J. Jackson

King of Pop

Michael Jackson



Overview of multilingual EL approaches

Name	Year	Evaluated Languages	Demo	Src	API
KIM	2004	English, French, <i>Spanish</i>	✓	✗	✓
SDA	2011	English, French	✗	✗	✗
ualberta	2012	English, Chinese	✗	✗	✗
HITS	2012	English, <i>Spanish</i> , Chinese	✗	✗	✗
THD	2012	English, German, Dutch	✓	✓	✓
DBpedia Spotlight	2013	French, Italian, Russian, Dutch, English, German, <i>Spanish</i> , Hungarian, Danish	✓	✓	✓
TAGME	2013	English, German, Dutch	✓	✗	✓
Wang-Tang	2013	English, Chinese	✗	✗	✗
AGDISTIS	2014	English, German	✓	✓	✓
Babelify	2014	English, French German, Italian, <i>Spanish</i>	✓	✗	✓
WikiME	2016	German, <i>Spanish</i> , French, Italian, Chinese, Hebrew, Thai, Arabic, Turkish, Tamil, Tagalog, Urdu	✓	✗	✗
FEL	2017	English, <i>Spanish</i> , Chinese	✗	✓	✗

Research Questions

- How does Entity Linking performance differ between English and Spanish?

Research Questions

- Do multilingual systems configured for the language perform much better for Spanish than monolingual systems not configured for that language?

Research Questions

- What might be the possible reasons for the observed results?

Dataset of SemEval 2015 Task 13

- Composed by 4 documents, each document in English, Spanish and Italian.

	<i>Doc₁</i>	<i>Doc₂</i>	<i>Doc₃</i>	<i>Doc₄</i>
<i>Sentences</i>	38	53	22	24

- Contains 769 entity mentions with their corresponding links to DBpedia, WordNet and BabelNet.

Gold Standard

Key speakers at the social exclusion conference were (left to right) Cristina Louro, Employment, Industrial Relations and Social Affairs Directorate, European Commission Fernando Gomes, a high degree of Committee of the Regions Barbara Weiler, Member of the European Parliament José Maria Gil-Robles, Vice-President, European Parliament and John Carroll, Economic and Social Committee .

Gold Standard

Key speakers at the [social exclusion](#) conference were (left to right) Cristina Louro, Employment, [Industrial Relations](#) and [Social Affairs Directorate](#), [European Commission](#) Fernando Gomes, a high degree of [Committee of the Regions](#) Barbara Weiler, [Member of the European Parliament](#) José Maria Gil-Robles, [Vice-President](#), [European Parliament](#) and John Carroll, [Economic and Social Committee](#).

Gold Standard

<http://dbpedia.org/page/Affair>
<http://babelnet.org/synset?word=bn:00001739n> (affairs)
<https://en.wikipedia.org/wiki/Affair>
<wn:affairs%1:04:00::>

Key speakers at the [social exclusion](#) conference were ([left to right](#)) [Cristina Louro](#), [Employment, Industrial Relations and Social Affairs Directorate](#), [European Commission](#) [Fernando Gomes](#), a high degree of [Committee of the Regions](#) [Barbara Weiler](#), [Member of the European Parliament](#) [José Maria Gil-Robles](#), [Vice-President](#), [European Parliament](#) and [John Carroll](#), [Economic and Social Committee](#).

https://pt.wikipedia.org/wiki/Assembleia_Nacional
<https://en.wikipedia.org/wiki/Majlis>
<https://en.wikipedia.org/wiki/Parlamentet>
<https://fr.wikipedia.org/wiki/Parlimentaire>
<https://en.wikipedia.org/wiki/Parliament>

Overview of multilingual EL approaches

Name	Year	Evaluated Languages	Demo	Src	API
KIM	2004	English, French, <i>Spanish</i>	✓	✗	✓
SDA	2011	English, French	✗	✗	✗
ualberta	2012	English, Chinese	✗	✗	✗
HITS	2012	English, <i>Spanish</i> , Chinese	✗	✗	✗
THD	2012	English, German, Dutch	✓	✓	✓
DBpedia Spotlight	2013	French, Italian, Russian, Dutch, English, German, <i>Spanish</i> , Hungarian, Danish	✓	✓	✓
TAGME	2013	English, German, Dutch	✓	✗	✓
Wang-Tang	2013	English, Chinese	✗	✗	✗
AGDISTIS	2014	English, German	✓	✓	✓
Babelify	2014	English, French German, Italian, <i>Spanish</i>	✓	✗	✓
WikiME	2016	German, <i>Spanish</i> , French, Italian, Chinese, Hebrew, Thai, Arabic, Turkish, Tamil, Tagalog, Urdu	✓	✗	✗
FEL	2017	English, <i>Spanish</i> , Chinese	✗	✓	✗

Selection criteria

- 1 The system must support Spanish.
- 2 Details of the system must be published.
- 3 A public demo or API must be available for the system.
- 4 The system must be a complete EL system including both ER and ED phases
- 5 The system must perform linking to Wikipedia or a related resource, such as DBpedia, YAGO or BabelNet.

Overview of multilingual EL approaches

Name	Year	Evaluated Languages	Demo	Src	API
KIM	2004	English, French, <i>Spanish</i>	✓	✗	✓
SDA	2011	English, French	✗	✗	✗
ualberta	2012	English, Chinese	✗	✗	✗
HITS	2012	English, <i>Spanish</i> , Chinese	✗	✗	✗
THD	2012	English, German, Dutch	✓	✓	✓
DBpedia Spotlight	2013	French, Italian, Russian, Dutch, English, German, <i>Spanish</i> , Hungarian, Danish	✓	✓	✓
TAGME	2013	English, German, Dutch	✓	✗	✓
Wang-Tang	2013	English, Chinese	✗	✗	✗
AGDISTIS	2014	English, German	✓	✓	✓
Babelify	2014	English, French German, Italian, <i>Spanish</i>	✓	✗	✓
WikiME	2016	German, <i>Spanish</i> , French, Italian, Chinese, Hebrew, Thai, Arabic, Turkish, Tamil, Tagalog, Urdu	✓	✗	✗
FEL	2017	English, <i>Spanish</i> , Chinese	✗	✓	✗

Example annotations

Key speakers at the social exclusion conference were (left to right) Cristina Louro, Employment, Industrial Relations and Social Affairs Directorate, European Commission Fernando Gomes, a high degree of Committee of the Regions Barbara Weiler, Member of the European Parliament José Maria Gil-Robles, Vice-President, European Parliament and John Carroll, Economic and Social Committee .

Example annotations - DBpedia Spotlight

Key speakers at the [social exclusion conference](#) were (left to right) [Cristina Louro](#), [Employment, Industrial Relations and Social Affairs Directorate](#), [European Commission](#) [Fernando Gomes](#), a high degree of [Committee](#) of the [Regions](#) [Barbara Weiler](#), Member of the [European Parliament](#) [José Maria Gil-Robles](#), Vice-President, [European Parliament](#) and [John Carroll](#), [Economic and Social Committee](#).

Example annotations - DBpedia Spotlight

Key speakers at the [social exclusion conference](#) were (left to right) [Cristina Louro](#), [Employment, Industrial Relations and Social Affairs Directorate](#), [European Commission](#) [Fernando Gomes](#), a [high](#) degree of [Committee](#) of the [Regions](#) [Barbara Weiler](#), Member of the [European Parliament](#) [José María Gil-Robles](#), Vice-President, [European Parliament](#) and [John Carroll](#), [Economic and Social Committee](#).

Spanish

Los principales oradores en la conferencia sobre [exclusión social](#) fueron (de izquierda a derecha) [Cristina Louro](#), [Dirección General de Empleo, Relaciones Laborales y Asuntos Sociales de la Comisión Europea](#) [Fernando Gomes](#), [Comité](#) de las [Regiones](#) [Barbara Weiler](#), [diputada al Parlamento Europeo](#) [José María Gil-Robles](#) Gil-Delgado, Vicepresidente, [Parlamento Europeo](#), [John Carroll](#) [Comité Económico y Social](#).

Entity Linking approaches for English

- AIDA
- THD
- TAGME

Example annotations - TAGME

Key speakers at the social exclusion conference were (left to right) Cristina Louro, Employment, Industrial Relations and Social Affairs Directorate, European Commission Fernando Gomes, a high degree of Committee of the Regions Barbara Weiler, Member of the European Parliament José Maria Gil-Robles, Vice-President, European Parliament and John Carroll, Economic and Social Committee .

Example annotations - TAGME

Key speakers at the social exclusion conference were (left to right) Cristina Louro, Employment, Industrial Relations and Social Affairs Directorate, European Commission Fernando Gomes, a high degree of Committee of the Regions Barbara Weiler, Member of the European Parliament José Maria Gil-Robles, Vice-President, European Parliament and John Carroll, Economic and Social Committee.

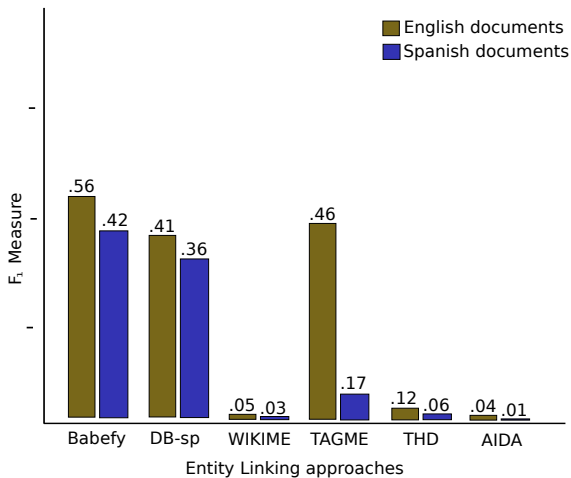
Example annotations - TAGME

Key speakers at the [social exclusion](#) conference were (left to right) [Cristina Louro](#), [Employment, Industrial Relations and Social Affairs Directorate, European Commission](#) [Fernando Gomes](#), a high degree of [Committee of the Regions](#) [Barbara Weiler](#), [Member of the European Parliament](#) [José María Gil-Robles](#), [Vice-President, European Parliament](#) and [John Carroll](#), [Economic and Social Committee](#).

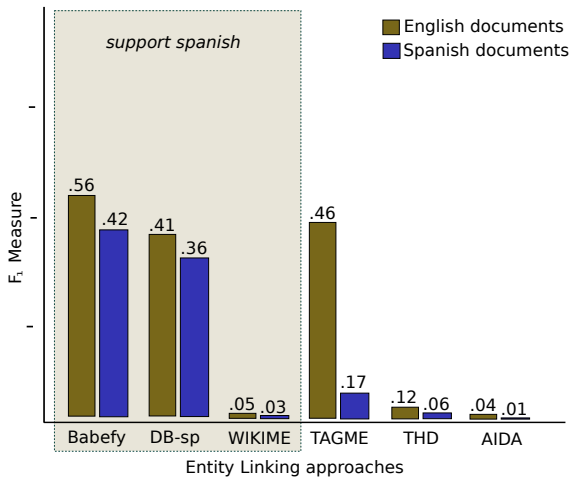
Spanish

Los ~~principales~~ oradores en la conferencia sobre ~~exclusión social~~ fueron (de izquierda a derecha) ~~Cristina Louro~~, Dirección ~~General~~ de Empleo, Relaciones Laborales y Asuntos Sociales de la Comisión ~~Europea~~ ~~Fernando Gomes~~, Comité de las ~~Regiones~~ [Barbara Weiler](#), diputada al Parlamento Europeo [José María Gil-Robles](#) Gil-Delgado, Vicepresidente, Parlamento Europeo, ~~John Carroll~~ ~~Comité~~ Económico y Social.

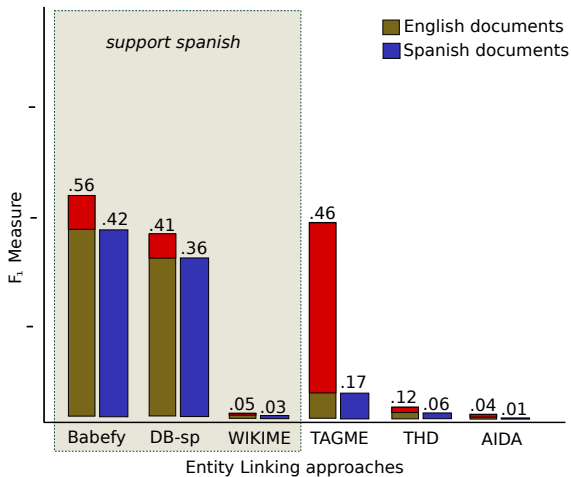
Overall Evaluation of Entity Linking



Overall Evaluation of Entity Linking



Overall Evaluation of Entity Linking



Results and Discussion

- All approaches obtain the best results for English.
- Entity Linking for Spanish performs much worse for monolingual approaches than multilingual approaches that support Spanish.
- The best score is obtained for Babelfy.

Results and Discussion

We propose that this result may be due to one (or more) of the following issues faced by multilingual systems

- The knowledge base contains different information for both languages (DBpedia-Spotlight, WIKIME).
- Models/techniques change according to the target language (Babelfy).
- Variations in the languages themselves.

Results and Discussion

Issues faced by monolingual systems

- TAGME is based on the analysis of anchor text of the English Wikipedia pages.
- THD selects candidates using the Search API of English Wikipedia
- AIDA is based on an English part-of-speech tagger.

Conclusion

- We survey the main multilingual Entity Linking approaches.
- We performed experiments to compare selected approaches in Spanish and English.
- We propose some potential explanations for the observed results.

Future Work

- Incorporate more languages.
- Incorporate more quality measures: Accuracy, BCubed, etc.
- Incorporate approaches that do not have Demo and APIs available, but do have the source code.
- Run experiments for validating explanations for differences.

Multilingual Entity Linking: Comparing English and Spanish †

Henry Rosales-Méndez, Barbara Poblete and Aidan Hogan

University of Chile

{hrosales,bpoblete,ahogan}@dcc.uchile.cl

October 22nd, 2017

† LD4IE - Linked Data for Information Extraction Workshop.

Evaluation of the Entity Recognition Phase

System	Spanish			English			%
	P	R	F1	P	R	F1	
<i>Babelfy</i>	0.727	0.540	0.620	0.820	0.644	0.721	85.99
<i>DBpedia-Spotlight</i>	0.298	0.607	0.400	0.556	0.554	0.555	72.07
<i>WikiMe</i>	0.737	0.018	0.036	0.656	0.028	0.053	67.92
TAGME	0.240	0.319	0.274	0.583	0.687	0.631	43.42
THD	0.281	0.061	0.100	0.587	0.080	0.142	70.42
AIDA	0.750	0.008	0.015	0.688	0.029	0.057	26.32

Multilingual Entity Linking: Comparing English and Spanish †

Henry Rosales-Méndez, Barbara Poblete and Aidan Hogan

University of Chile

{hrosales,bpoblete,ahogan}@dcc.uchile.cl

October 22nd, 2017

† LD4IE - Linked Data for Information Extraction Workshop.

Evaluation of the Entity Recognition Phase

System	Spanish			English			%
	P	R	F1	P	R	F1	
<i>Babelfy</i>	0.599	0.324	0.420	0.725	0.467	0.568	73.94
<i>DBpedia-Spotlight</i>	0.482	0.293	0.364	0.581	0.322	0.414	87.92
<i>WikiMe</i>	0.929	0.017	0.033	0.952	0.026	0.051	64.71
TAGME	0.371	0.118	0.179	0.568	0.391	0.463	38.66
THD	0.596	0.036	0.069	0.738	0.059	0.120	57.50
AIDA	0.667	0.005	0.010	0.773	0.022	0.044	22.73