

The Power of Natural Language Processing for Social Media Analysis

Luis Espinosa-Anke and Jose Camacho-Collados




13 October 2020

Outline

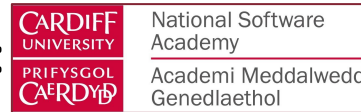
- Introduction to Natural Language Processing (NLP)
- Challenges and applications in social media:
 - Multilinguality
 - Emoji
 - Meaning shift

NLP at Cardiff University *now*

A newly created Cardiff Uni NLP group (40+ members!)

- Talks, reading groups and tutorials
- Presence in social media: @Cardiff_NLP 
- Networking, interdisciplinarity

Our research feeds into our MSc **teaching**:



- **Applied Machine Learning**, with applications in NLP and Computer Vision.
- **Manipulating and Exploiting Data**, databases & ML/NLP
- **Computational Data Science**, data science and software engineering

About us

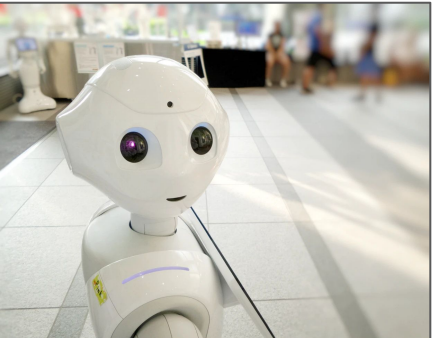
Luis Espinosa Anke

- Background in Linguistics and Translation Studies
- Erasmus Mundus MA in NLP
- PhD in Natural Language Processing, User-facing NLP experience and **Fulbright** Fellow
- Lecturer **Computer Science** (today) - Cardiff University

Jose Camacho Collados

- Background in Mathematics
- Erasmus Mundus MA in NLP
- PhD in Natural Language Processing and Google AI Doctoral Fellow
- Lecturer **Computer Science** (today) - Cardiff University

Natural Language Processing



Natural Language Processing (NLP)

- Branch of AI that deals with human language
- Multi-billion industry
- Usual suspects: spell checkers, autocomplete, machine translation, speech-to-text and text-to-speech systems, web search
- Other applications: sentiment analysis, trend prediction, digital forensics, hate speech monitoring, fake news detection, recommendation systems, chatbots.

Natural Language Processing

- How? First, we need data.

Natural Language Processing

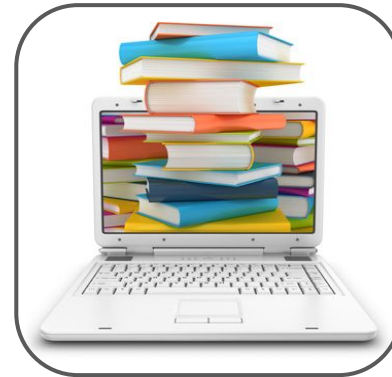
- How? First, we need data.

structured



and

organic



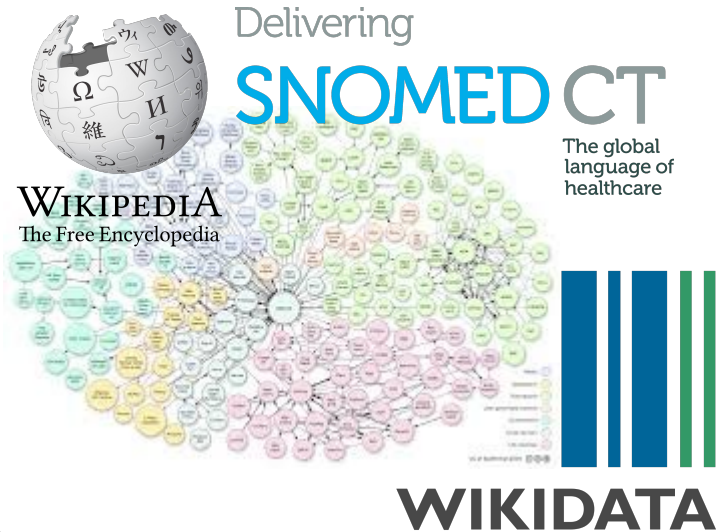
Natural Language Processing

➤ How? First, we need data.

structured

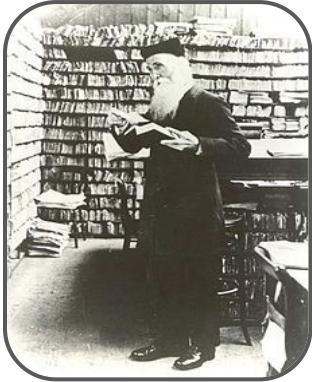
and

organic



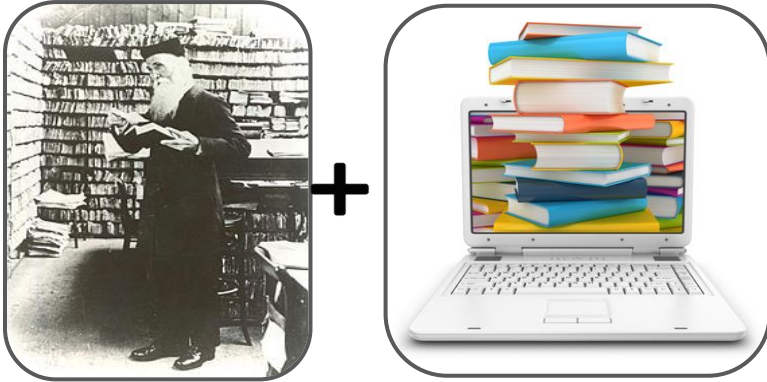
Natural Language Processing

➤ Our goal



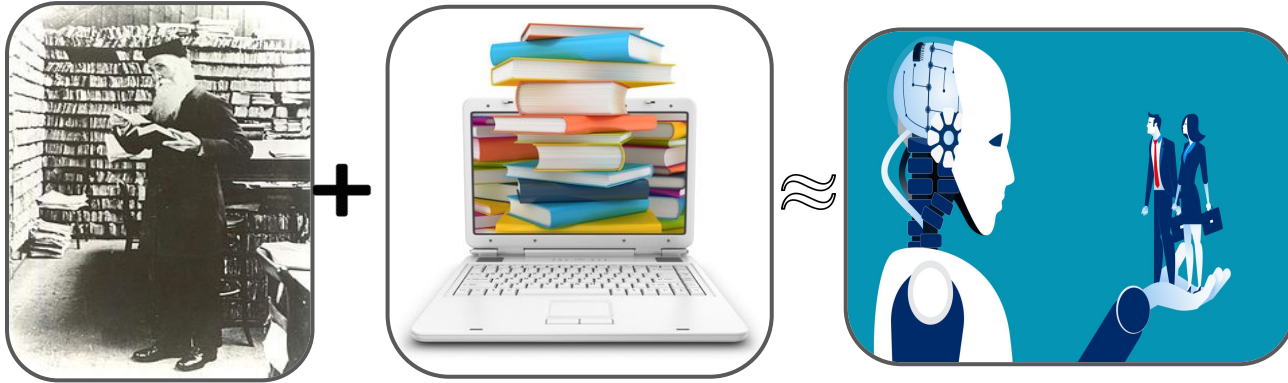
Natural Language Processing

➤ Our goal



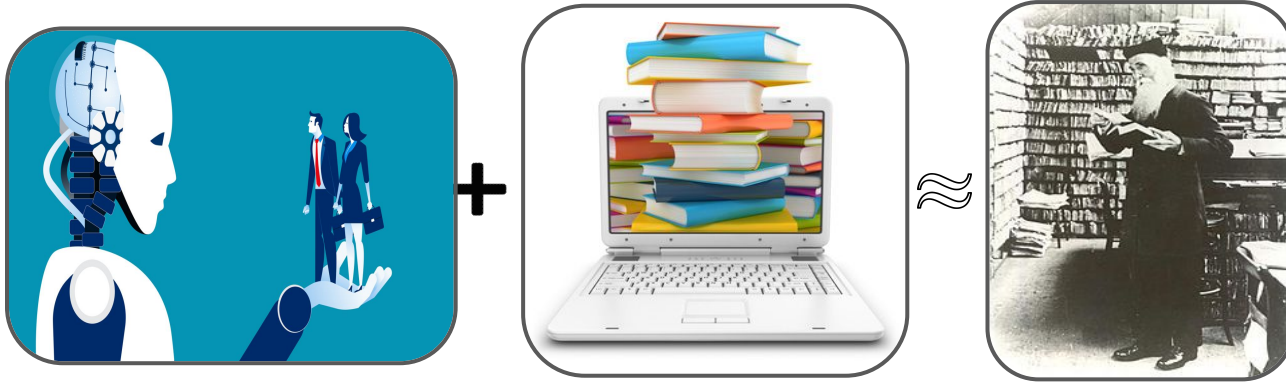
Natural Language Processing

➤ Our goal



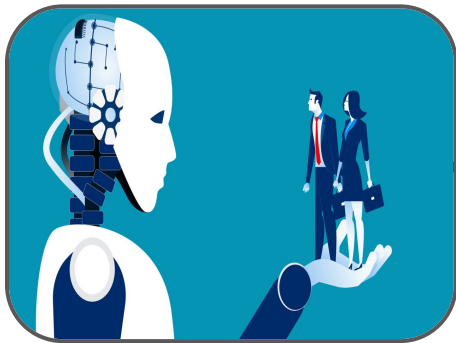
Natural Language Processing

➤ Our goal

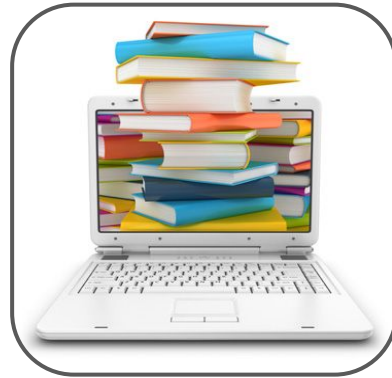


Natural Language Processing

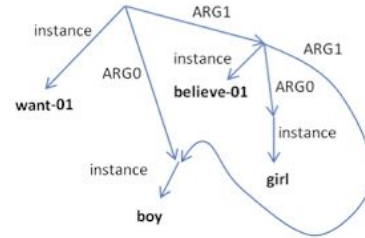
➤ Our goal



+

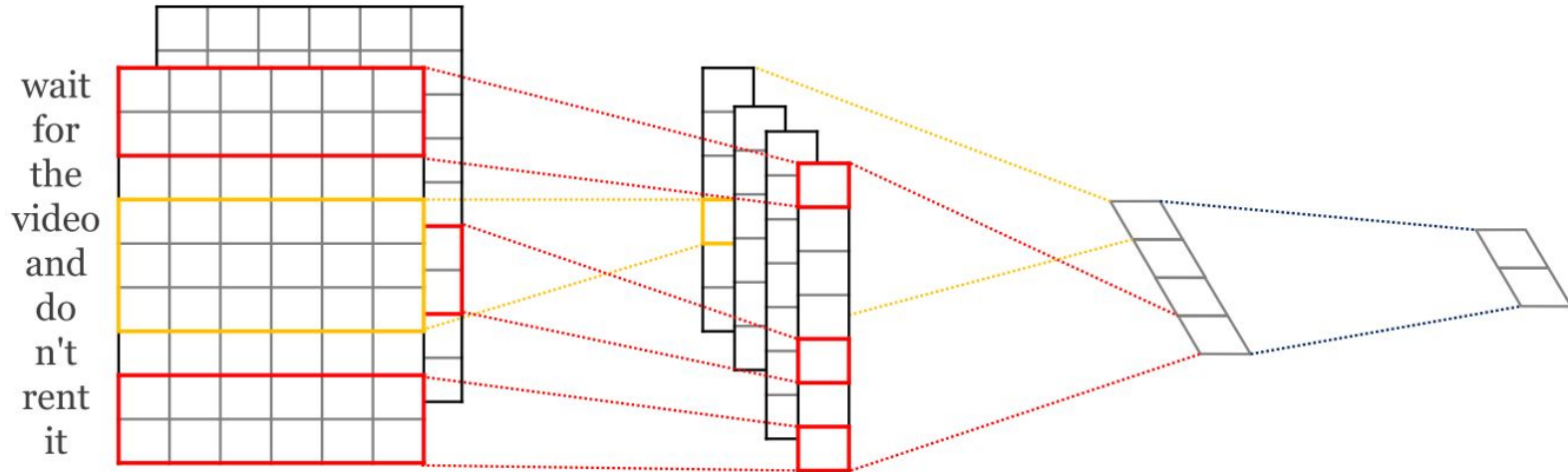


AI goal: language understanding



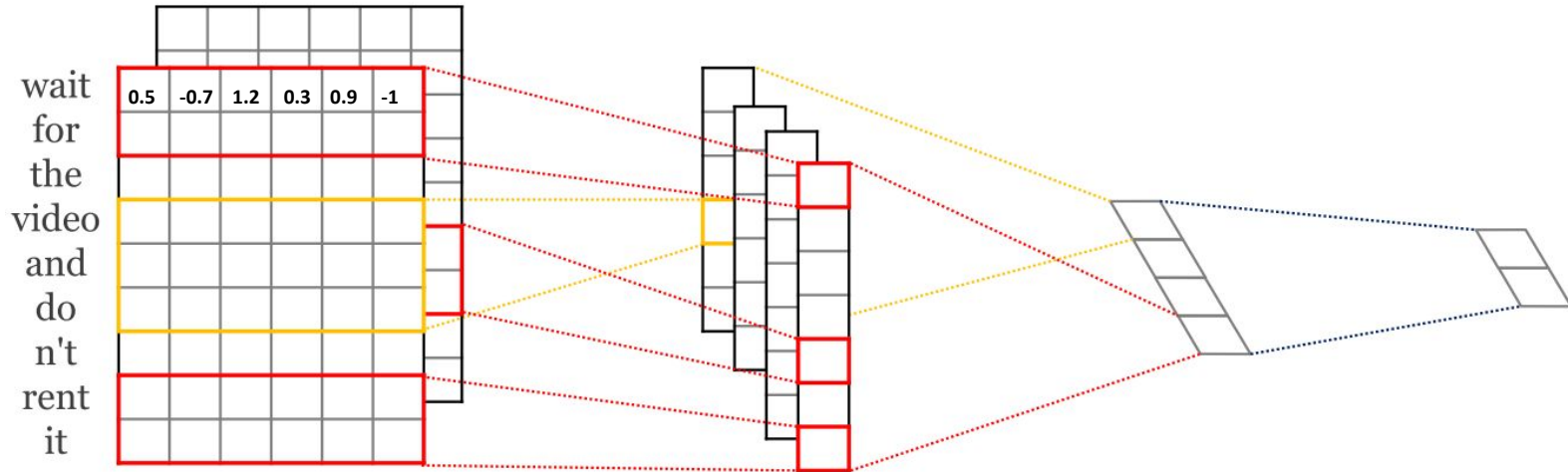
Natural Language Processing

- Then, algorithms (machine learning).



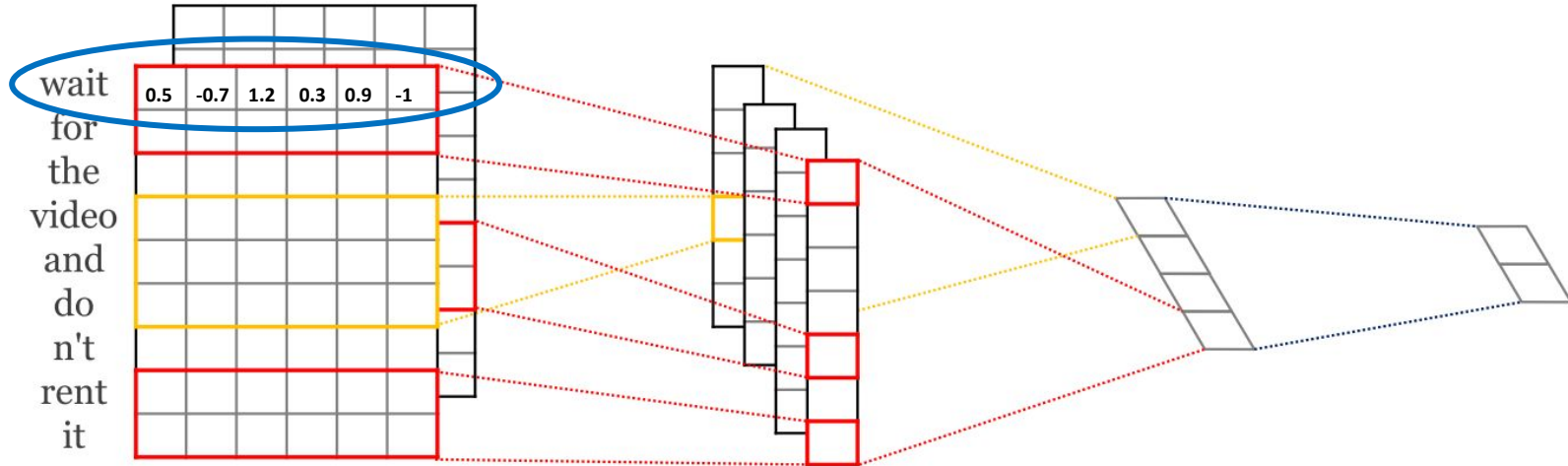
Natural Language Processing

- Then, algorithms (machine learning).



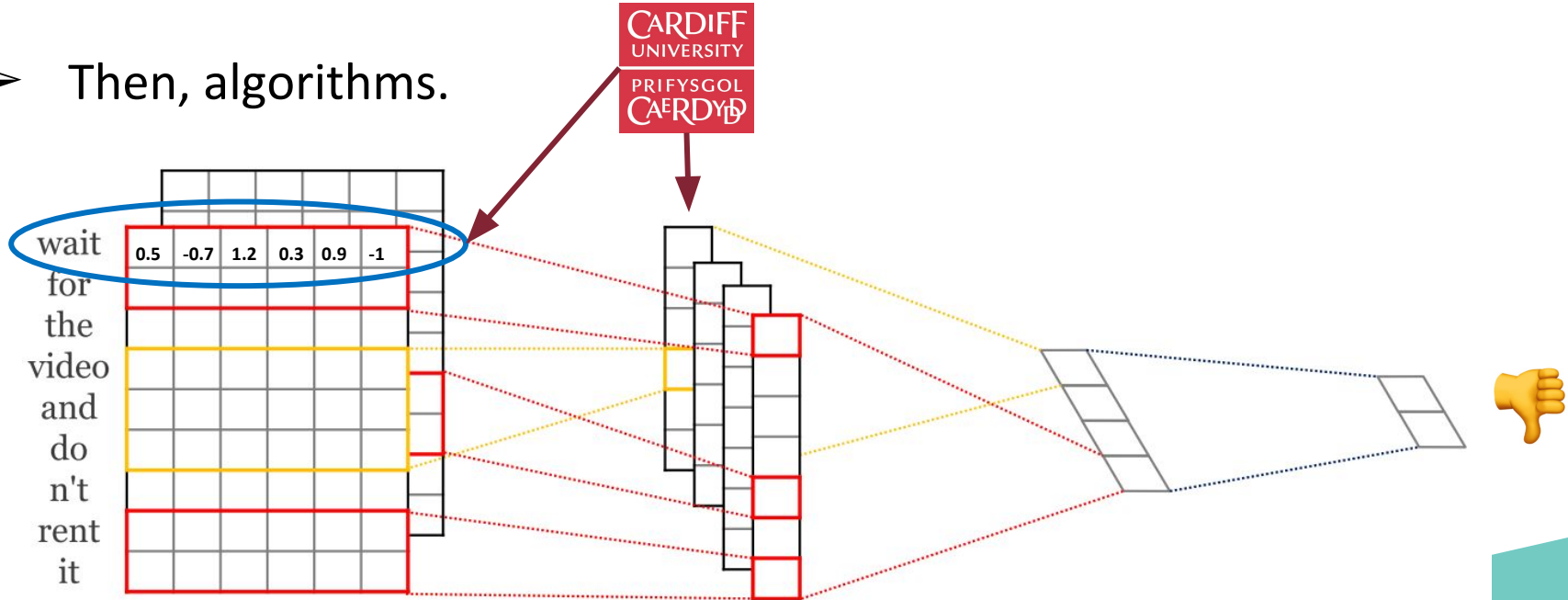
Natural Language Processing

- Then, algorithms (machine learning).



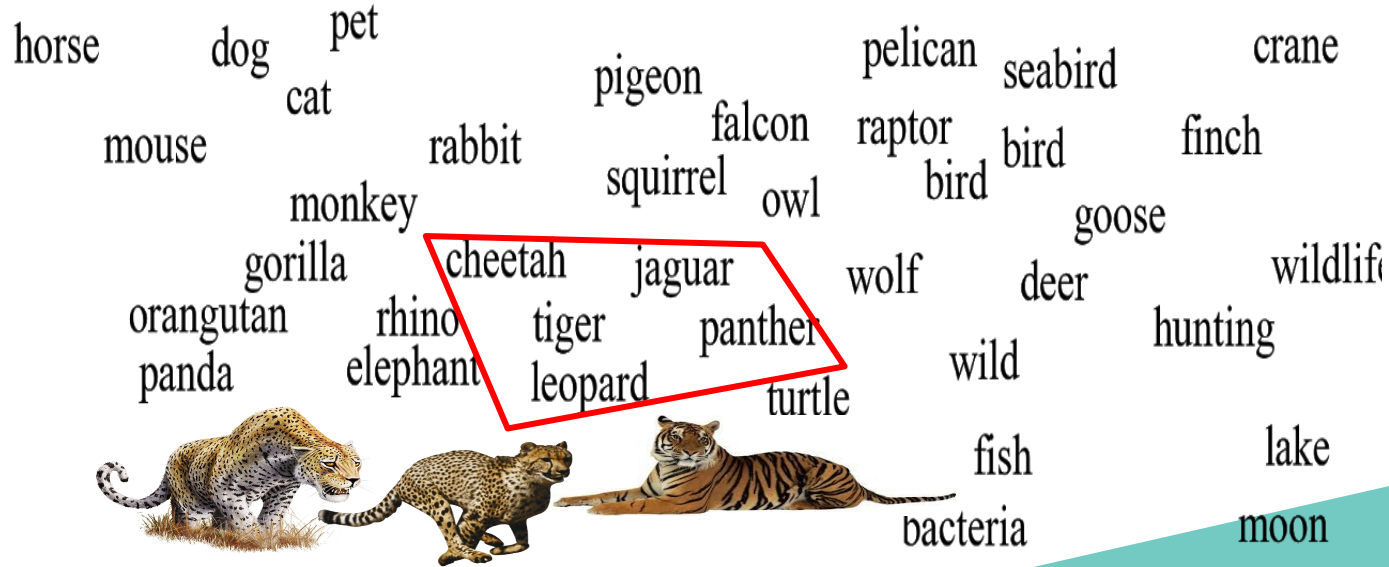
Natural Language Processing

➤ Then, algorithms.

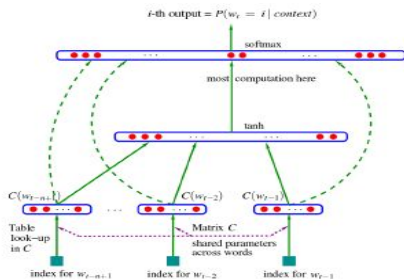


Word vector space models

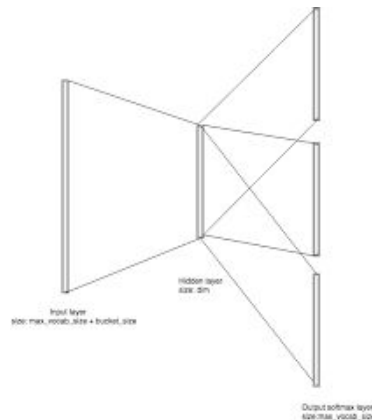
Words are represented as vectors: semantically similar words are close in the vector space



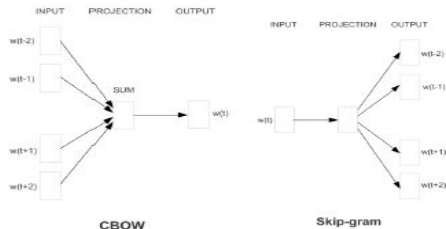
Neural networks for learning word vector representations from text corpora -> word embeddings



Bengio et al. (2003)



Bojanowski et al. (2017)



Mikolov et al. (2013)

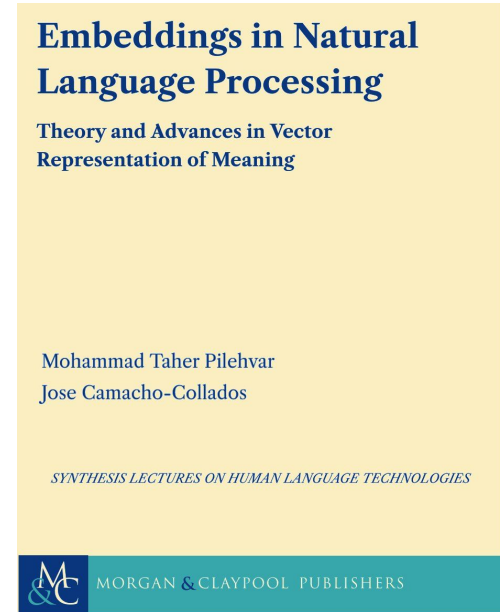
Probability and Ratio	$k = solid$	$k = gas$	$k = water$
$P(k ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}
$P(k steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}
$P(k ice)/P(k steam)$	8.9	8.5×10^{-2}	1.36

Pennington et al. (2014)

Book on “Embeddings in Natural Language Processing”

First draft freely available online:

josecamachocollados.com/book_embNLP_draft.pdf



Social media

Social media

- Social media is a goldmine for modeling human behaviour
- But it presents major communication challenges:
 - Multilinguality
 - Emoji
 - Meaning shift



Multilinguality

Multilinguality



Maribel.R.L
@maribelr1

Replying to [@IkerCasillas](#)

Muchas felicidades campeón 🎂🎂🎂🎂 😘😘

[Translate Tweet](#)

8:54 PM · May 20, 2019 from [Getafe, Spain](#) · [Twitter for Android](#)



Bimba Dop
@BimbaDop

[@IkerCasillas](#) Buon compleanno Iker 🎂 [#HalaMadrid](#)

[Translate Tweet](#)

1:58 PM · May 20, 2020 · [Twitter for Android](#)



Mimi Vardista #ThankYouMaria
@veronamimi42

Replying to [@IkerCasillas](#)

Happy birthday 🌟🎉🎁🎉🌟🎂🎉🎁 Legend.



3:37 PM · May 20, 2019 · [Twitter for Android](#)



Rami
@RamiArs11

[#halamadrid](#) iker تولدت مبارك 🎂




11:08 AM · Feb 22, 2019 · [Twitter Web Client](#)

Emoji

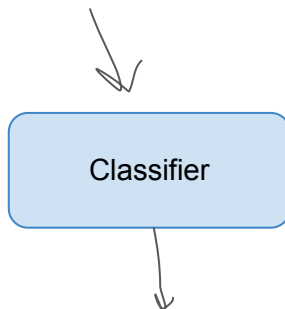
Are Emojis Predictable?



Roger Federer  @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 

Can't wait to visit #Chicago for the first time next Week. What shall I do while I am in town? #LaverCup



Are Emojis Predictable?

F Barbieri, M Ballesteros, H Saggion

EACL 2017

Emoji prediction

- Autocomplete systems
- Enabling studies on network dynamics
- Better NLP for social media understanding (because you can't just treat an emoji just like another word)
- etc.

Are Emojis Predictable?



Roger Federer
Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup



Roger Federer  @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 🤔



Are Emojis Predictable?

F Barbieri, M Ballesteros, H Saggion


EACL 2017



Are Emojis Predictable?

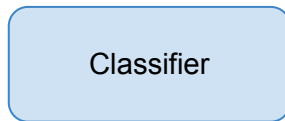


Roger Federer  @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 



Can't wait to visit #Chicago for the first time next Week. What shall I do while I am in town? #LaverCup



Are Emojis Predictable?

F Barbieri, M Ballesteros, H Saggion

EACL 2017

Emoji prediction

Emoji prediction

- Autocomplete systems
- Enabling studies on network dynamics
- Better NLP for social media (because you can't just treat an emoji just like another word)
- etc.

Emoji prediction

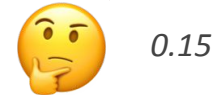


Roger Federer  @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 🤔

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup

Classifier



Emoji prediction



Roger Federer  @rogerfederer · Mar 15

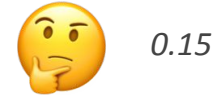
Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 🤔

*Can't wait to visit #Chicago for the first time next week. **What** shall I do while I am in town? **#LaverCup***

Classifier + attention



0.80



0.15



0.02



0.01



0.008

...

0.00..

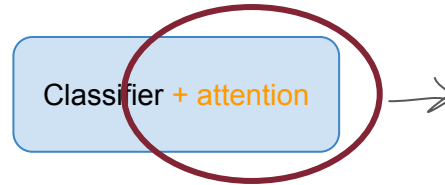
Emoji prediction



Roger Federer @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 🤔

Can't wait to visit #Chicago for the first time next week. **What** shall I do while I am in town? **#LaverCup**



0.80



0.15



0.02



0.01



0.008



0.00..

Emoji prediction



Roger Federer ✓ @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 🤔

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup

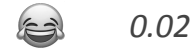
Classifier + attention



0.80



0.15



0.02



0.01



0.008



0.00..

Emoji prediction



Roger Federer ✓ @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup 🤔

Can't wait to visit #Chicago for the first time next week. **What** shall I do while I am in town? #LaverCup

Classifier + attention



0.80



0.15



0.02



0.01



0.008



0.00..

Emoji prediction

praying we have a snow day tomorrow 🙏

Emoji prediction

praying we have a snow day tomorrow 🙏

Standard Att: 🙏 0.709, ❄️ 0.126, 😊 0.017
praying we have a snow day tomorrow

Label-Wise Att: 🙏 0.510, ❄️ 0.153, 🙌 0.027
praying we have a snow day tomorrow (🙏)
praying we have a snow day tomorrow (❄️)
praying we have a snow day tomorrow (🙌)

EMOJI as anchor between languages



Maribel.R.L
@maribelr1

Replying to @IkerCasillas

Muchas felicidades campeón 🍰🍰🍰🍰💋💋

[Translate Tweet](#)

8:54 PM · May 20, 2019 from Getafe, Spain · [Twitter for Android](#)



Bimba Dop
@BimbaDop

[@IkerCasillas](#) Buon compleanno Iker 🍰 #HalaMadrid

[Translate Tweet](#)

1:58 PM · May 20, 2020 · [Twitter for Android](#)



Mimi Vardista #ThankYouMaria
@veronamimi42

Replying to @IkerCasillas

Happy birthday 🌟🎉🎁🎉🌟🍰🎉🎁 Legend.



3:37 PM · May 20, 2019 · [Twitter for Android](#)



Rami
@RamiArs11

[#halamadrid](#) iker تولدت مبارك 🍰



11:08 AM · Feb 22, 2019 · [Twitter Web Client](#)

EMOJI as anchor between languages



Maribel.R.L.
@maribelr1

Replying to @IkerCasillas

Muchas felicidades campeón 🍰🍰🍰🍰💋💋

[Translate Tweet](#)

8:54 PM · May 20, 2019 from Getafe, Spain · [Twitter for Android](#)



Mimi Vardista #ThankYouMaria
@veronamimi42

Replying to @IkerCasillas

Happy birthday 🌟🎉🎁🎉🌟🍰🎉🎁 Legend.



3:37 PM · May 20, 2019 · [Twitter for Android](#)



Bimba Dop
@BimbaDop

@IkerCasillas Buon compleanno Iker 🍰 #HalaMadrid

[Translate Tweet](#)

1:58 PM · May 20, 2020 · [Twitter for Android](#)



Rami
@RamiArs11

#halamadrid iker تولدت مبارك 🍰



11:08 AM · Feb 22, 2019 · [Twitter Web Client](#)

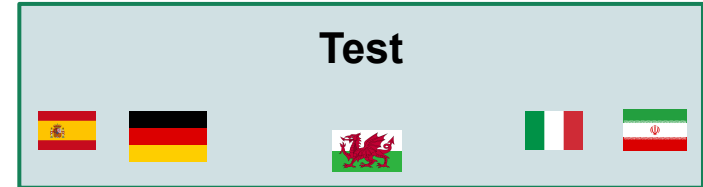
Cross-lingual transfer

Why cross-lingual transfer?

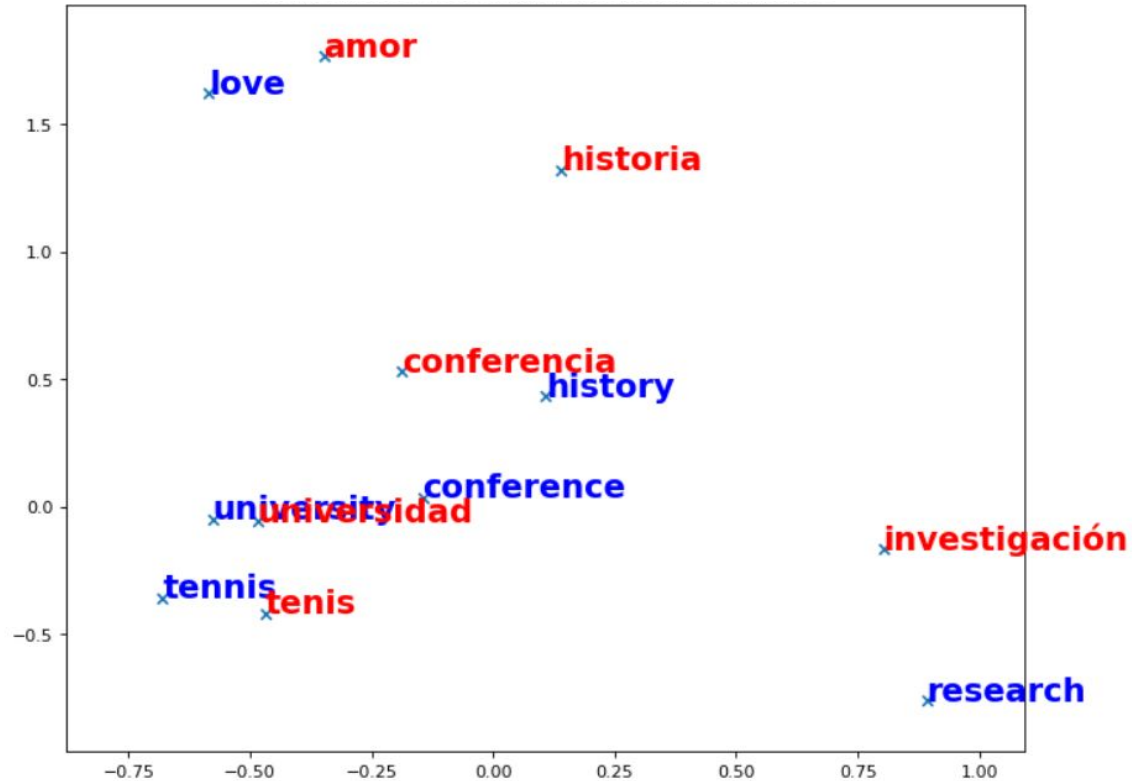
Classic
Machine Learning pipeline



Cross-lingual pipeline

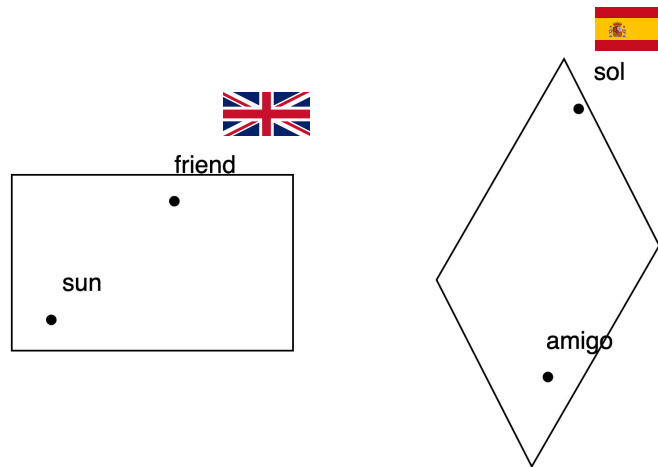


Cross-lingual embeddings



Slide credit: MUSE

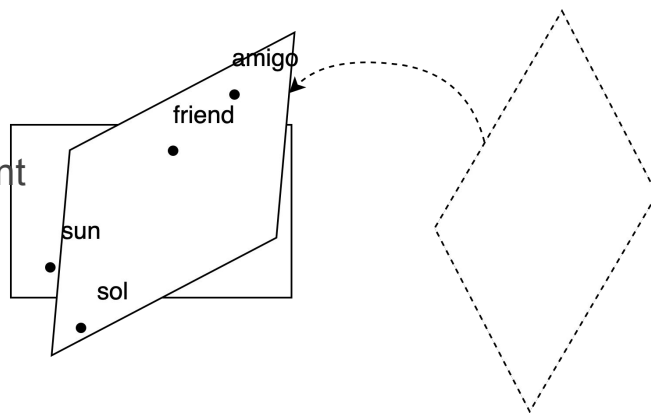
Learning cross-lingual embeddings



- **Word2Vec** (Mikolov et al. 2013)
- **FastText** (Bojanowski et al. TACL 2017)

Learning cross-lingual embeddings

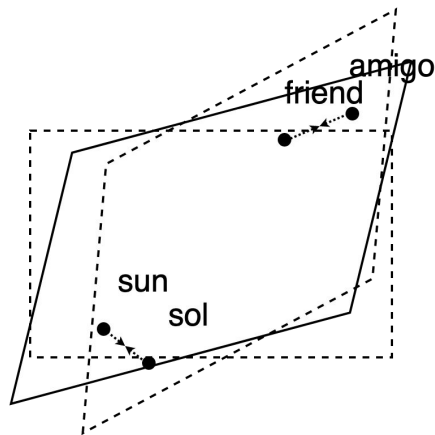
- Step 2: Orthogonal alignment



- **MUSE** (Conneau et al. ICLR 2018)
- **VecMap** (Artetxe et al. AAAI 2018)

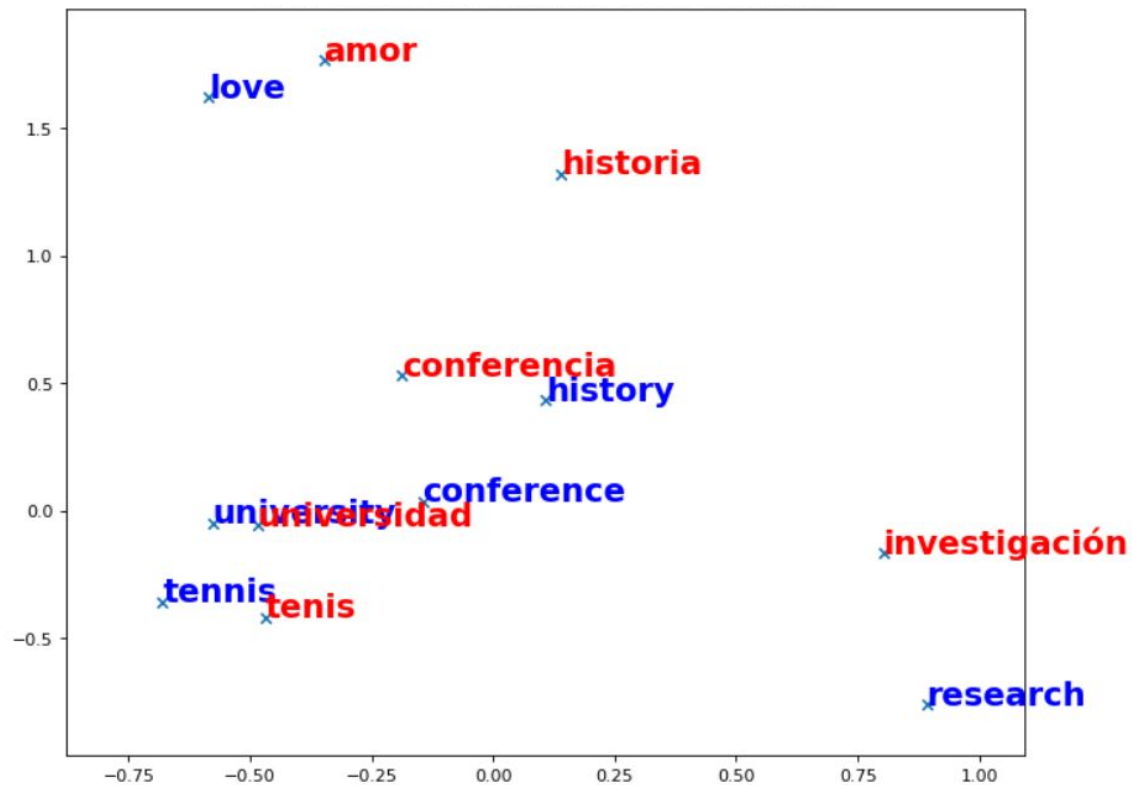
Learning cross-lingual embeddings

- Step 3: Linear mean mapping



➤ **Meemi** (Doval et al. EMNLP 2018)

Cross-lingual embeddings

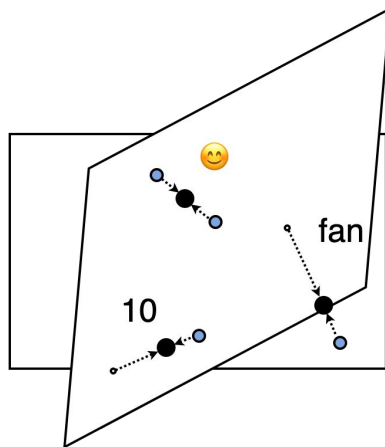


Applications of cross-lingual embeddings

- NLP tasks for which training data exists in one language (e.g. English) but not in other (e.g. Welsh).
- Case study on Twitter

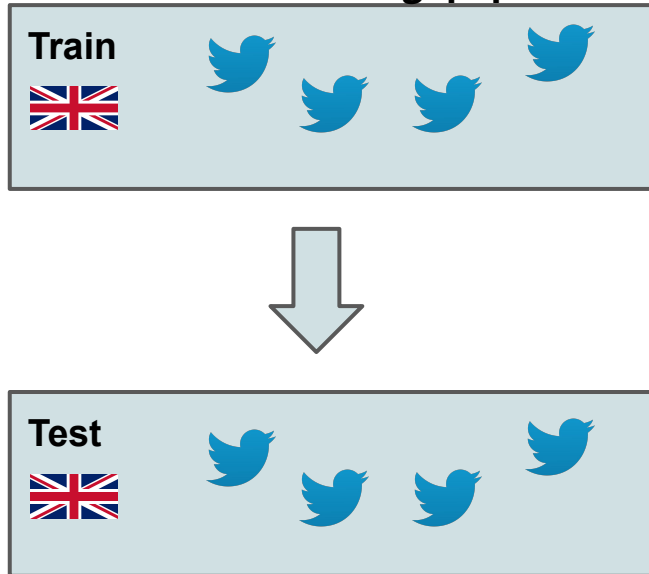
Cross-lingual NLP for Twitter

Cross-lingual NLP for Twitter

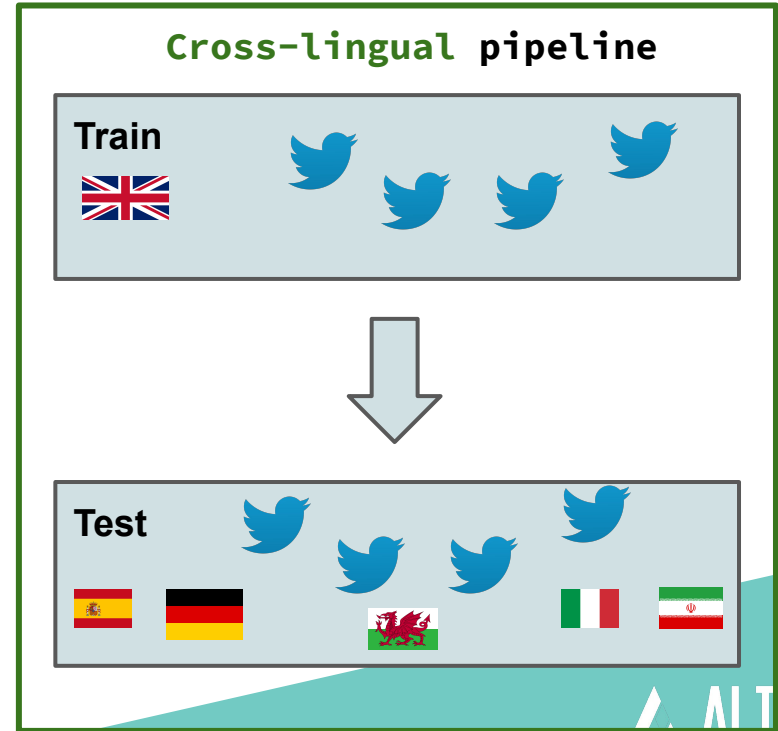


Cross-lingual NLP for Twitter

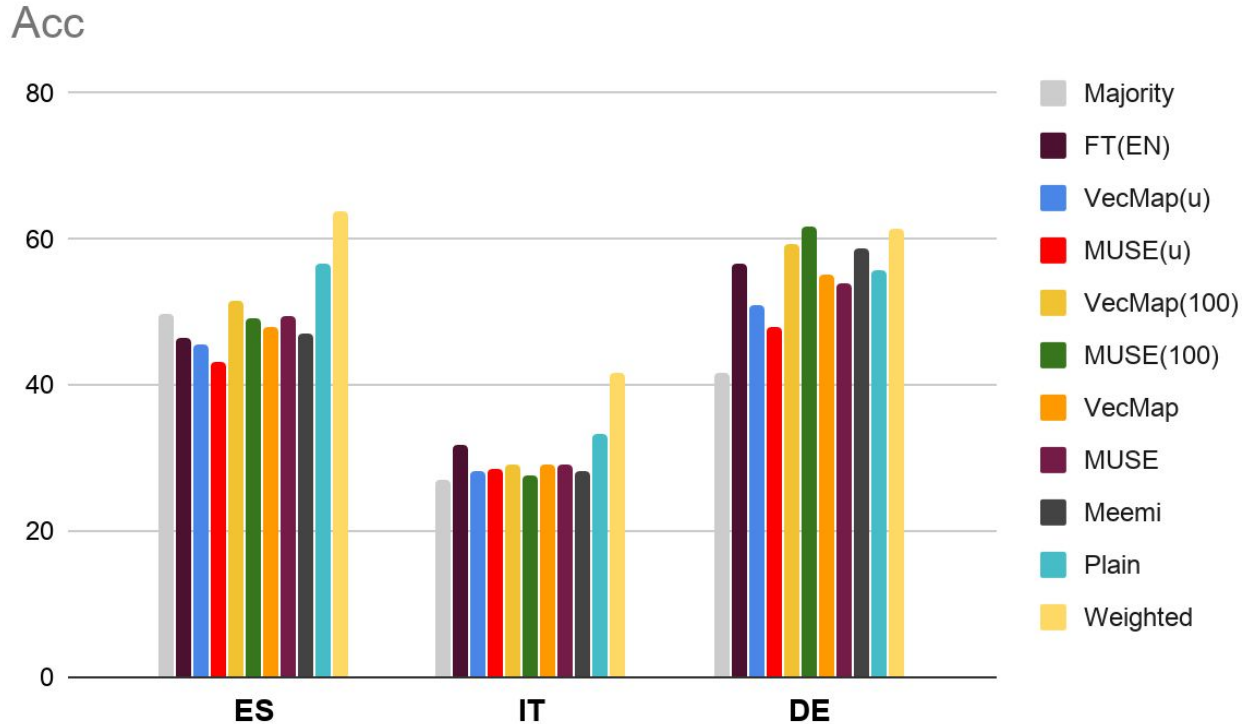
Classic Machine Learning pipeline



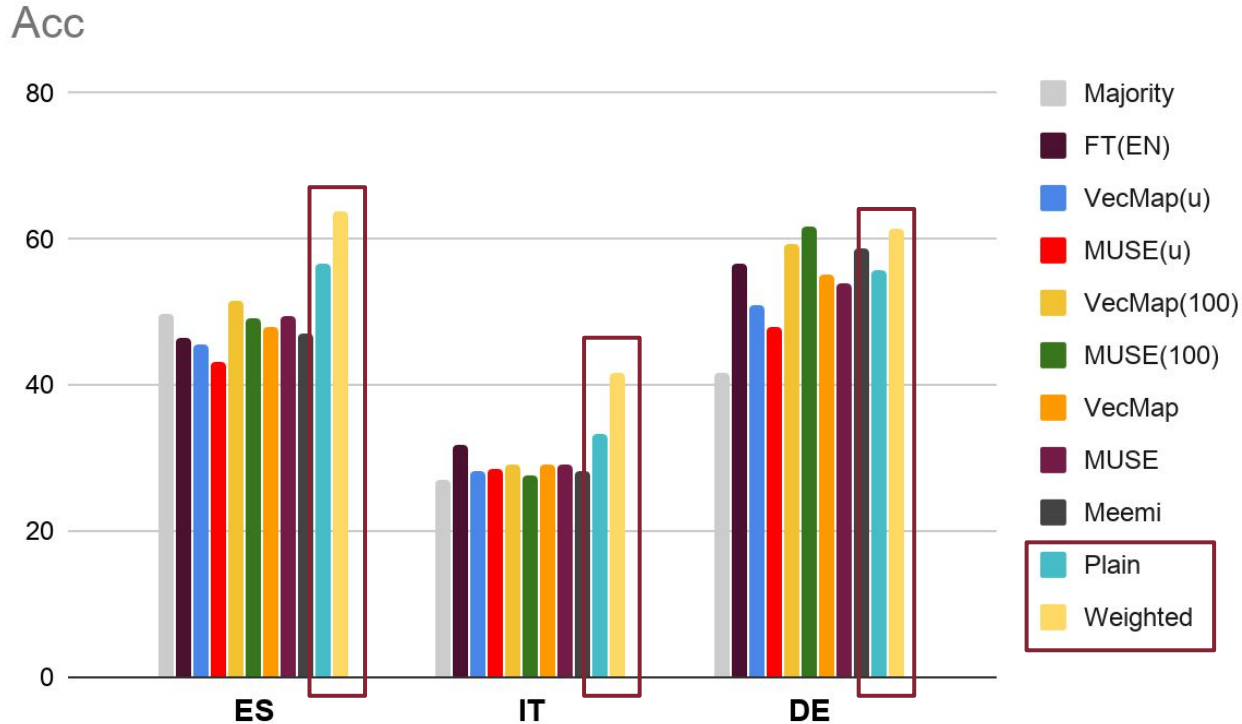
Cross-lingual pipeline



Cross-lingual sentiment analysis



Cross-lingual sentiment analysis



Cross-lingual NLP for Twitter

OMFG 

			
diooo	mioddio	maaah	نمیفهمیدی
wtfff	ommiodio	njaahah	نمیبخشم
dioooo	oddiooo	hahaha	ک***

Cross-lingual NLP for Twitter

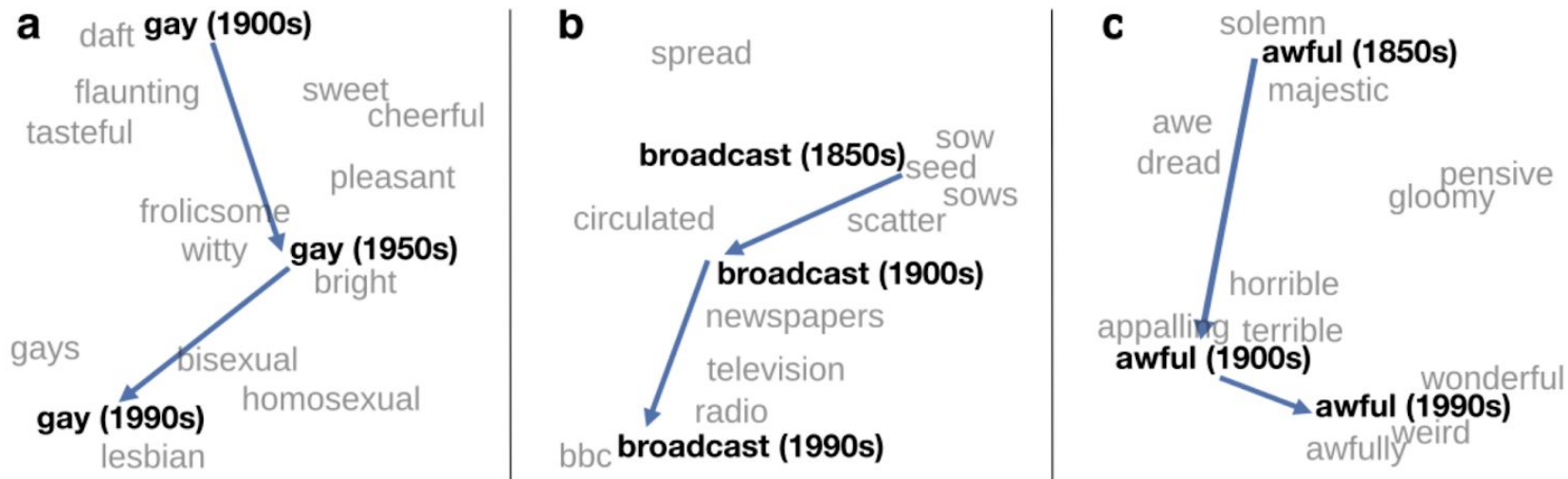
CHILLAX 

			
relajadito	rilassando	entspan	خواب
relaxx	rilasso	gemütlich	بيدارشو
relajaito	rilassa	relaxte	بيدارشب

Meaning Shift in Social Media

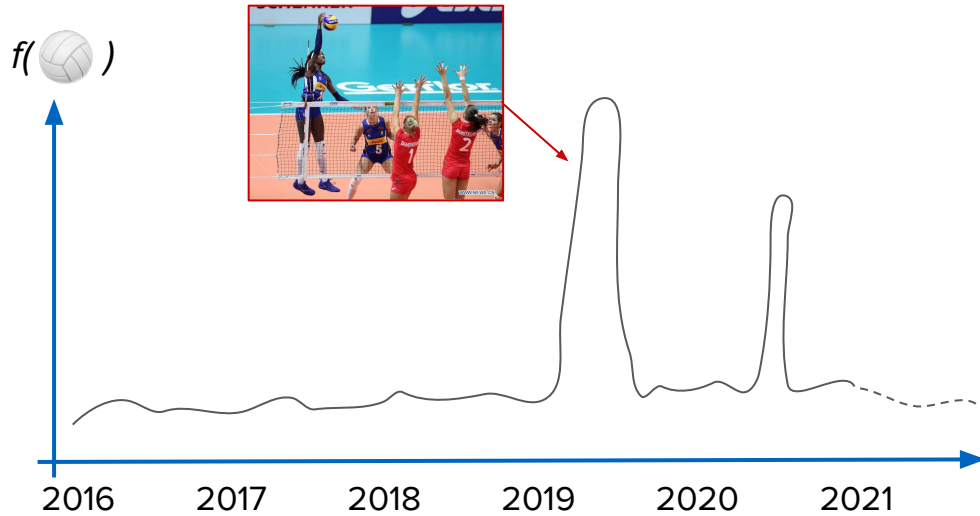
Word Meaning Shift

What is the expiration date of an NLP classifier?



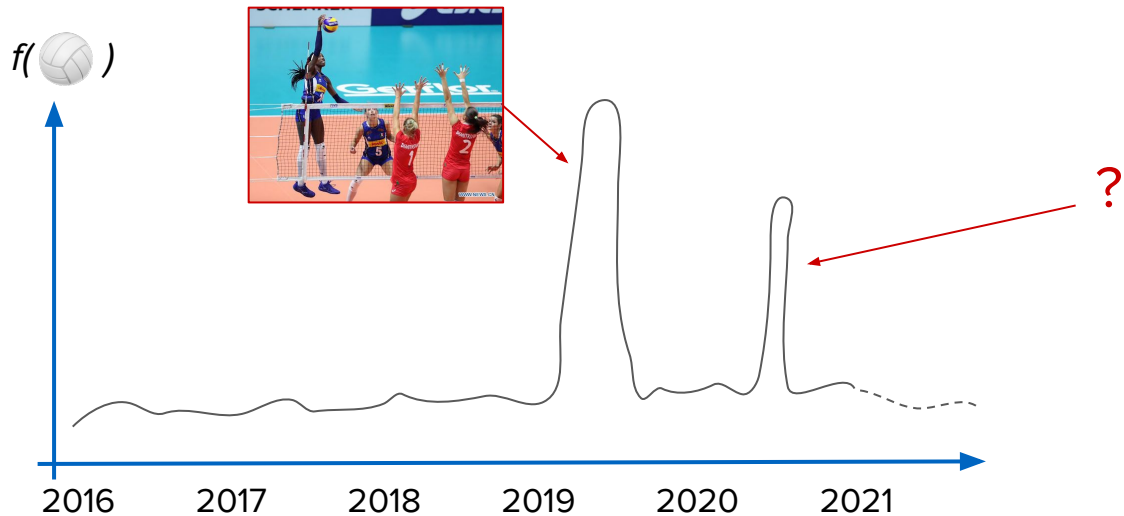
Word Meaning Shift in Social Media

Example of quick meaning shift that does not persist over time
(here, meaning shift for the *volleyball* emoji was only temporary)



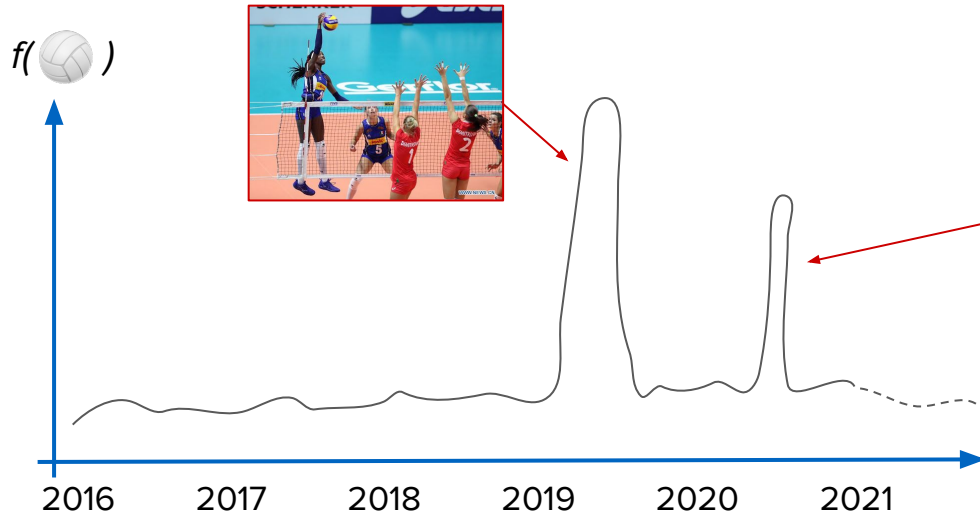
Word Meaning Shift in Social Media

Example of quick meaning shift that does not persist over time
(here, meaning shift for the *volleyball* emoji was only temporary)



Word Meaning Shift in Social Media

Example of quick meaning shift that does not persist over time
(here, meaning shift for the *volleyball* emoji was only temporary)



Tom Hanks, Rita Wilson Test Positive For COVID-19 Coronavirus



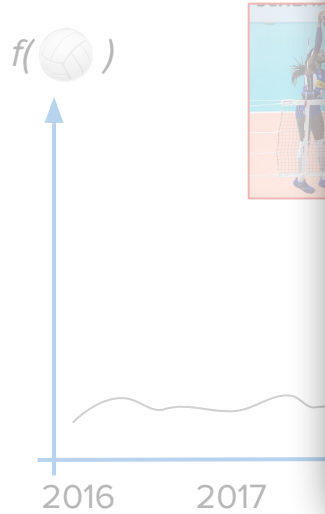
Bruce Y. Lee Senior Contributor ©
Healthcare

I am a writer, journalist, professor, systems modeler, computational and digital health expert, avocado-eater, and entrepreneur, not always in that order.



Word Meaning Shift in Social Media

Example of quick
(the meaning shif



Wilson Test
-19

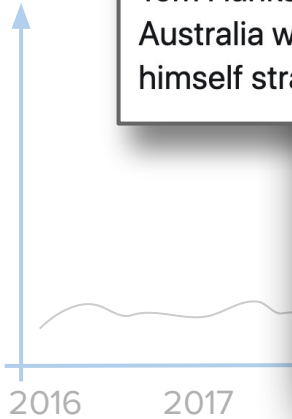
later, computational and digital health
ways in that order.



Word Meaning Shift in Social Media

Example of quick
(the meaning shift)

$f(\text{Wilson})$



Rob Not Bob
@RBecker262

Tom Hanks and his wife Rita Wilson are quarantined in Australia with the [#Coronavirus](#). Once again, Tom finds himself stranded on an island with Wilson. 🏐



Tracy AMH Trump 🇺🇸 🦅 @TracyTrump6 · Mar 11

Stop freaking out about [@tomhanks](#) being quarantined in Australia for the [coronavirus](#). He survived on an island for 5 year's with a darn volleyball. [#Wilson](#) 🏐

Wilson Test
-19

later, computational and digital health
ways in that order.

Word Meaning Shift in Social Media

 **Dr. Bibi Neumann** ❤️🌱🍷
@DrBibiNeumann1

Manhattan Beach 'Karen' Calls Cops on Black Woman in Racist Rant

[Traducir Tweet](#)



Manhattan Beach 'Karen' Calls Cops on Black Woman in Racist Rant
A new "Karen" has emerged in supposedly liberal and all-accepting Manhattan Beach, CA.
[tmz.com](#)

2:19 p. m. · 10 oct. 2020 · Twitter for iPhone

Word Meaning Shift in Social Media

 **Dr. Bibi Neumann** ❤️🌱🍷
@DrBibiNeumann1

Manhattan Beach 'Karen' Calls Cops on Black Woman
in Racist Rant

[Traducir Tweet](#)

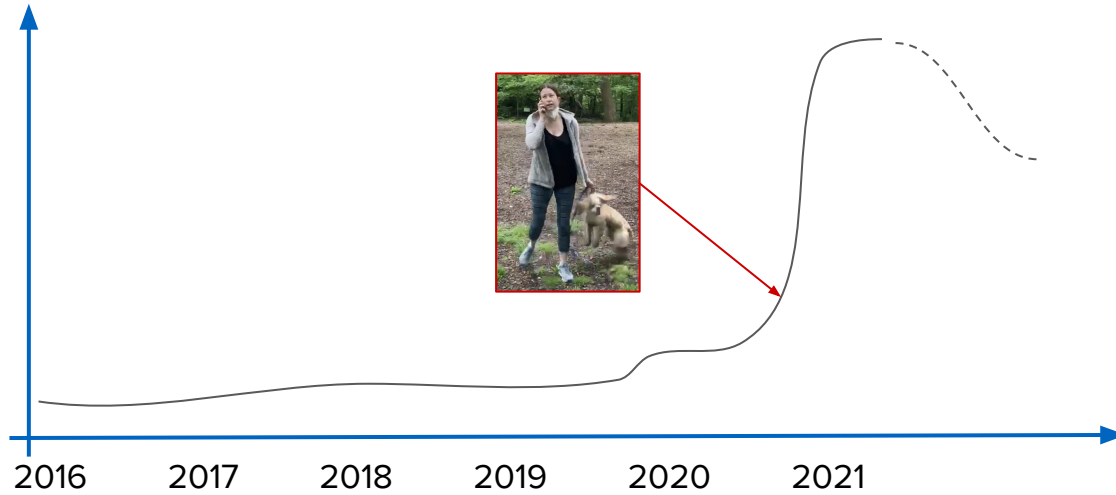


Manhattan Beach 'Karen' Calls Cops on Black Woman in Racist Rant
A new "Karen" has emerged in supposedly liberal and all-accepting Manhattan Beach, CA.
[tmz.com](#)

2:19 p. m. · 10 oct. 2020 · Twitter for iPhone

Word Meaning Shift in Social Media

$f(\text{"karen"})$



Hope you enjoyed the ride



Ongoing and past projects

Ongoing:

- Snapchat (NLP in social media)
- Welsh Government (cross-lingual NLP with Welsh language)
- Crime and Security Research Institute (misinformation in social media)
- Welsh SMEs (e.g., MarketMate)

Past:

- Google
- Spanish police and Spanish Royal Academy
- Recent publications with: Amazon, IBM, Snapchat, Pandora, MIT and others.


Conclusions

- Internet (text) data is crucial for understanding online communities, language evolution, commonalities and differences between languages. We have seen challenges involved in:
 - Social media understanding
 - Enabling cross-lingual and cross-domain knowledge transfer
 - Understanding real-world NLP challenges



Thank you!

Questions please!

 [@camachocollados](#)
[@espinosaanke](#)