# Author profiling in social media

Paolo Rosso

2023-2024

DSIC

MIARFID Official Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

# Author Profiling

Language and style varies among classes of authors

**Forensics**: who is behind  an harassment

**Security**: who is behind a threat

**Marketing**: who is behind an opinion

**Socio-political analysis**: who is behind a stance

- **Gender & age**
- **Personality**
- **Native language and language variety**
- Ideological/organizational affiliation

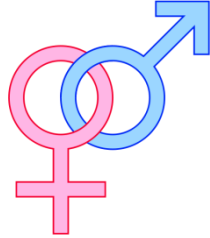
# Outline

- Profiling gender & age
- Author profiling at PAN-2013 and PAN-2014
- Profiling **bots** at PAN-2019
- Profiling **fake news spreaders** at PAN-2020
- Profiling **haters** at PAN-2021


https://pan.webis.de/

# Profiling gender & age

# Which is female/male?



My aim in this article is to show that given a relevance theoretic approach to utterance interpretation, it is possible to develop a better understanding of what some of these so-called apposition markers indicate. It will be argued that the decision to put something in other words is essentially a decision about style, a point which is, perhaps, anticipated by Burton-Roberts when he describes loose apposition as a rhetorical device. However, he does not justify this suggestion by giving the criteria for classifying a mode of expression as a rhetorical device. Nor does he specify what kind of effects might be achieved by a reformulation or explain how it achieves those effects. In this paper I follow Sperber and Wilson's (1986) suggestion that rhetorical devices like metaphor, irony and repetition are particular means of achieving relevance. As I have suggested, the corrections that are made in unplanned discourse are also made in the pursuit of optimal relevance. However, these are made because the speaker recognises that the original formulation did not achieve optimal relevance .

The main aim of this article is to propose an exercise in stylistic analysis which can be employed in the teaching of English language. It details the design and results of a workshop activity on narrative carried out with undergraduates in a university department of English. The methods proposed are intended to enable students to obtain insights into aspects of cohesion and narrative structure: insights, it is suggested, which are not as readily obtainable through more traditional techniques of stylistic analysis. The text chosen for analysis is a short story by Ernest Hemingway comprising only 11 sentences. A jumbled version of this story is presented to students who are asked to assemble a cohesive and well formed version of the story. Their re-constructions are then compared with the original Hemingway version.

[examples: Moshe Koppel]

# British National Corpus

- 920 documents labelled for
  - author gender
  - document genre
- Used 566 controlled for genre

| | Male | Fem |
|---|---|---|
| **Fiction (prose)** | **132** | **132** |
| **Non-fiction** | **151** | **151** |
| **Arts (general)** | 8 | 8 |
| **Arts (acad.)** | 12 | 12 |
| **Belief/Thought** | 12 | 12 |
| **Biography** | 27 | 27 |
| **Commerce** | 5 | 5 |
| **Leisure** | 8 | 8 |
| **Science (gen.)** | 13 | 13 |
| **Soc. Sci. (gen.)** | 26 | 26 |
| **Soc. Sci. (acad.)** | 19 | 19 |
| **World Affairs** | 21 | 21 |

M. Koppel, S. Argamon, and A. R. Shimoni. Automatically categorizing written texts by author gender. Literary and linguistic computing 17(4), 2002.

# Distinguishing features: male vs. female style

Males use more

- Determiners
- Adjectives
- *of* modifiers (e.g. *pot of gold*)

Informational features

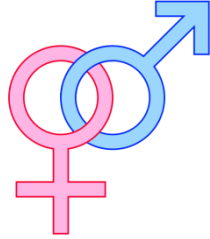
Females use more

- Pronouns *
- *for* and *with*
- Negation
- Present tense

Involvedness features

J. W. Pennebaker. The Secret Life of Pronouns: What Our Words Say about Us. Bloomsbury USA, 2013.

# Which is female/male?



My aim in this article is to show that given a relevance theoretic approach to utterance interpretation, it is possible to develop a better understanding of what some of these so-called apposition markers indicate. It will be argued that the decision to put something in other words is essentially a decision about style, a point which is, perhaps, anticipated by Burton-Roberts when he describes loose apposition as a rhetorical device. However, he does not justify this suggestion by giving the criteria for classifying a mode of expression as a rhetorical device. Nor does he specify what kind of effects might be achieved by a reformulation or explain how it achieves those effects. In this paper I follow Sperber and Wilson's (1986) suggestion that rhetorical devices like metaphor, irony and repetition are particular means of achieving relevance. As I have suggested, the corrections that are made in unplanned discourse are also made in the pursuit of optimal relevance. However, these are made because the speaker recognises that the original formulation did not achieve optimal relevance .

The main aim of this article is to propose an exercise in stylistic analysis which can be employed in the teaching of English language. It details the design and results of a workshop activity on narrative carried out with undergraduates in a university department of English. The methods proposed are intended to enable students to obtain insights into aspects of cohesion and narrative structure: insights, it is suggested, which are not as readily obtainable through more traditional techniques of stylistic analysis. The text chosen for analysis is a short story by Ernest Hemingway comprising only 11 sentences. A jumbled version of this story is presented to students who are asked to assemble a cohesive and well formed version of the story. Their re-constructions are then compared with the original Hemingway version.

# Female vs. male

**My** aim in this article is to show that given a relevance theoretic approach to utterance interpretation, it is possible to develop a better understanding of what some of these so-called apposition markers indicate. It will be argued that the decision to put something in other words is essentially a decision about style, a point which is, perhaps, anticipated by Burton-Roberts when **he** describes loose apposition as a rhetorical device. However, **he** does not justify this suggestion by giving the criteria for classifying a mode of expression as a rhetorical device. Nor does **he** specify what kind of effects might be achieved by a reformulation or explain how it achieves those effects. In this paper **I** follow Sperber and Wilson's (1986) suggestion that rhetorical devices like metaphor, irony and repetition are particular means of achieving relevance. As **I** have suggested, the corrections that are made in unplanned discourse are also made in the pursuit of optimal relevance. However, these are made because the speaker recognises that the original formulation did not achieve optimal relevance .

The main aim of this article is to propose an exercise in stylistic analysis which can be employed in the teaching of English language. It details the design and results of a workshop activity on narrative carried out with undergraduates in a university department of English. The methods proposed are intended to enable students to obtain insights into aspects of cohesion and narrative structure: insights, it is suggested, which are not as readily obtainable through more traditional techniques of stylistic analysis. The text chosen for analysis is a short story by Ernest Hemingway comprising only 11 sentences. A jumbled version of this story is presented to students who are asked to assemble a cohesive and well formed version of the story. **Their** re-constructions are then compared with the original Hemingway version.

# Female vs. male

My aim in this article is to **show** that given a relevance theoretic approach to utterance interpretation, it is possible to **develop** a better understanding of what some of these so-called apposition markers **indicate**. It will be argued that the decision to **put** something in other words is essentially a decision about style, a point which is, perhaps, anticipated by Burton-Roberts when he **describes** loose apposition as a rhetorical device. However, he does not **justify** this suggestion by giving the criteria for classifying a mode of expression as a rhetorical device. Nor does he **specify** what kind of effects might be achieved by a reformulation or explain how it **achieves** those effects. In this paper I **follow** Sperber and Wilson's (1986) suggestion that rhetorical devices like metaphor, irony and repetition are particular means of achieving relevance. As I have suggested, the corrections that are made in unplanned discourse are also made in the pursuit of optimal relevance. However, these are made because the speaker **recognises** that the original formulation did not **achieve** optimal relevance .

The main aim of this article is to **propose** an exercise in stylistic analysis which can be employed in the teaching of English language. It **details** the design and results of a workshop activity on narrative carried out with undergraduates in a university department of English. The methods proposed are intended to enable students to obtain insights into aspects of cohesion and narrative structure: insights, it is suggested, which are not as readily obtainable through more traditional techniques of stylistic analysis. The text chosen for analysis is a short story by Ernest Hemingway comprising only 11 sentences. A jumbled version of this story is presented to students who are asked to assemble a cohesive and well formed version of the story. Their re-constructions are then compared with the original Hemingway version.

# Female vs. male

My aim in this article is to show that given a relevance theoretic approach to utterance interpretation, it is possible to develop a better understanding of what some of these so-called apposition markers indicate. It will be argued that the decision to put something in other words is essentially a decision about style, a point which is, perhaps, anticipated by Burton-Roberts when he describes loose apposition as a rhetorical device. However, he does **not** justify this suggestion by giving the criteria for classifying a mode of expression as a rhetorical device. **Nor** does he specify what kind of effects might be achieved by a reformulation or explain how it achieves those effects. In this paper I follow Sperber and Wilson's (1986) suggestion that rhetorical devices like metaphor, irony and repetition are particular means of achieving relevance. As I have suggested, the corrections that are made in unplanned discourse are also made in the pursuit of optimal relevance. However, these are made because the speaker recognises that the original formulation did **not** achieve optimal relevance .

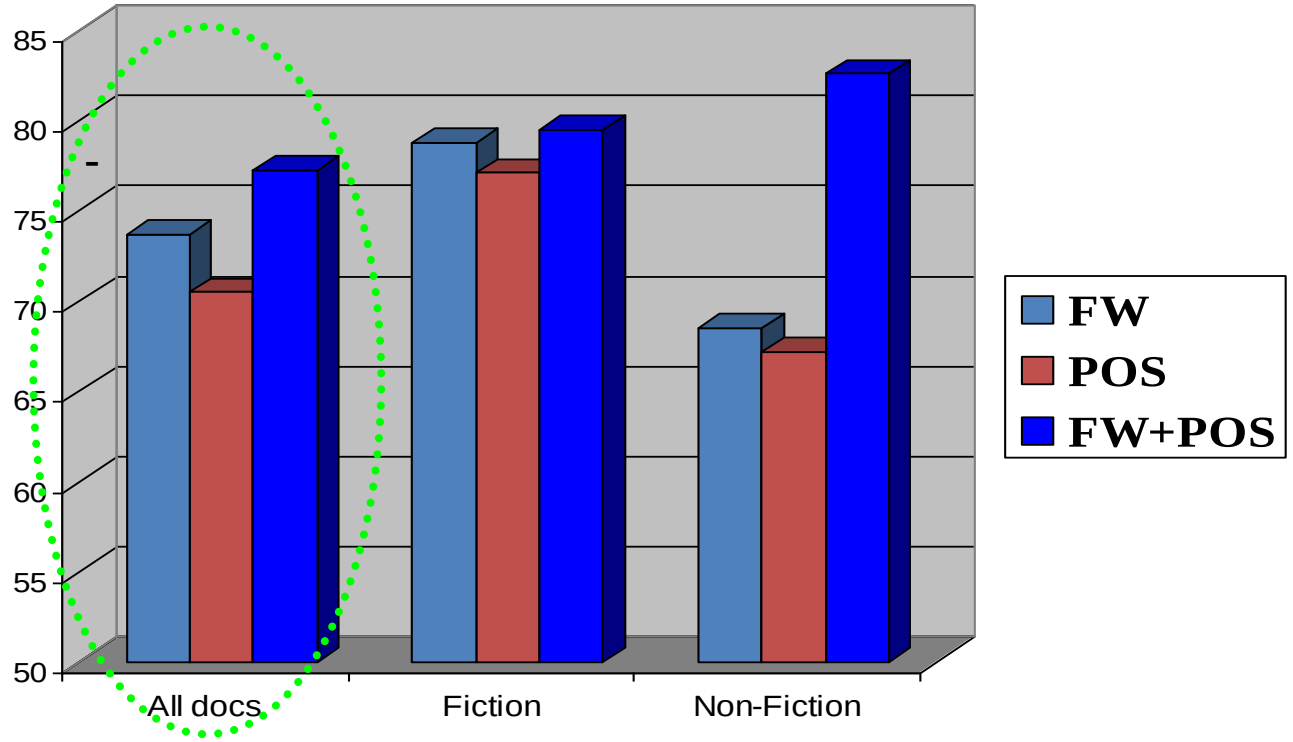
The main aim of this article is to propose an exercise in stylistic analysis which can be employed in the teaching of English language. It details the design and results of a workshop activity on narrative carried out with undergraduates in a university department of English. The methods proposed are intended to enable students to obtain insights into aspects of cohesion and narrative structure: insights, it is suggested, which are **not** as readily obtainable through more traditional techniques of stylistic analysis. The text chosen for analysis is a short story by Ernest Hemingway comprising only 11 sentences. A jumbled version of this story is presented to students who are asked to assemble a cohesive and well formed version of the story. Their re-constructions are then compared with the original Hemingway version.

# Female vs. male

My aim in this article is to show that given a relevance theoretic approach to utterance interpretation, it is possible to develop a better understanding **of** what some of these so-called apposition markers indicate. It will be argued that the decision to put something in other words is essentially a decision about style, a point which is, perhaps, anticipated by Burton-Roberts when he describes loose apposition as a rhetorical device. However, he does not justify this suggestion by giving the criteria for classifying a mode **of** expression as a rhetorical device. Nor does he specify what kind **of** effects might be achieved by a reformulation or explain how it achieves those effects. In this paper I follow Sperber and Wilson's (1986) suggestion that rhetorical devices like metaphor, irony and repetition are particular means **of** achieving relevance. As I have suggested, the corrections that are made in unplanned discourse are also made in the pursuit **of** optimal relevance. However, these are made because the speaker recognises that the original formulation did not achieve optimal relevance .

The main aim **of** this article is to propose an exercise in stylistic analysis which can be employed in the teaching **of** English language. It details the design and results **of** a workshop activity on narrative carried out with undergraduates in a university department **of** English. The methods proposed are intended to enable students to obtain insights into aspects **of** cohesion and narrative structure: insights, it is suggested, which are not as readily obtainable through more traditional techniques **of** stylistic analysis. The text chosen for analysis is a short story by Ernest Hemingway comprising only 11 sentences. A jumbled version **of** this story is presented to students who are asked to assemble a cohesive and well formed version **of** the story. Their re-constructions are then compared with the original Hemingway version.

# Results per feature set




- Handle fiction and non-fiction separately
- Use full feature set

POS: Part Of Speech  FW: Function words (*and, of, the*,..)

Teen Twenties Thirties

Male Female

# Social media: example

Yesterday we had our second jazz competition. Thank God we weren't competing. We were sooo bad. Like, I was so ashamed, I didn't even want to talk to anyone after. I felt so rotton, and I wanted to cry, but...it's ok.

Teen Twenties Thirties

Male Female

# Social media: example

Yesterday we had our second jazz competition. Thank God we weren't competing. We were sooo bad. Like, I was so ashamed, I didn't even want to talk to anyone after. I felt so rotton, and I wanted to cry, but...it's ok.

# Blog corpus

- Less-formal text
  - 85,000 blogs
  - blogger-provided profiles (gender, age, occupation, astrological sign)
  - harvested August 2004
  - all non-text ignored (formatting, quoting)

# Blog corpus

| | Gender | | |
|---|---|---|---|
| **Age** | **Female** | **Male** | **Total** |
| ~~**unknown**~~ | ~~12287~~ | ~~12259~~ | ~~24546~~ |
| **13-17** | ~~6949~~ | ***4120*** | 8240 |
| ~~**18-22**~~ | ~~7393~~ | ~~7690~~ | ~~15083~~ |
| **23-27** | ***4043*** | ~~6062~~ | 8086 |
| ~~**28-32**~~ | ~~1686~~ | ~~3057~~ | ~~4743~~ |
| **33-37** | ***860*** | ~~1827~~ | 1720 |
| **38-42** | ***374*** | ~~819~~ | 748 |
| **43-48** | ***263*** | ~~584~~ | 526 |
| ~~**>48**~~ | ~~314~~ | ~~906~~ | ~~1220~~ |
| **Total** | **9660** | **9660** | **19320** |

Final balanced corpus:

- 19,320 total blogs
  - 8240 in “10s”
  - 8086 in “20s”
  - 2994 in “30s”

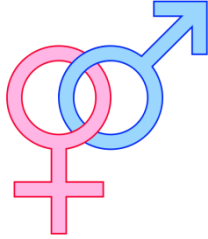
- 681,288 total posts
- 141,106,859 total words

# Gender and age classification

| Features | Gender & age (accuracy) |
|---|---|
| Style & Content | 80.0% - 77.4% |
| Style Words | 77.0% - 69.4% |
| Content Words | 73.0% - 76.2% |

J. Schler, M. Koppel, S. Argamon, and J. W. Pennebaker. Effects of age and gender on blogging. In AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, pages 199–205. AAAI, 2006.

# Men vs. women



| LIWC category | male | female |
|---|---|---|
| job | **68.1±0.6** | 56.5±0.5 |
| money | **43.6±0.4** | 37.1±0.4 |
| sports | **31.2±0.4** | 20.4±0.2 |
| tv | **21.1±0.3** | 15.9±0.2 |
| sex | 32.4±0.4 | **43.2±0.5** |
| family | 27.5±0.3 | **40.6±0.4** |
| eating | 23.9±0.3 | **30.4±0.3** |
| friends | 20.5±0.2 | **25.9±0.3** |
| sleep | 18.4±0.2 | **23.5±0.2** |
| *pos-emotions* | *248.2±1.9* | ***265.1±2*** |
| *neg-emotions* | *159.5±1.3* | ***178±1.4*** |

J. W. Pennebaker - LIWC: Linguistic Inquiry and Word Count

# The lifecycle of the common blogger...

| Word | 10s | 20s | 30s |
|---|---|---|---|
| maths | 105 | 3 | 2 |
| homework | 137 | 18 | 15 |
| bored | 384 | 111 | 47 |
| sis | 74 | 26 | 10 |
| boring | 369 | 102 | 63 |
| awesome | 292 | 128 | 57 |
| mum | 125 | 41 | 23 |
| crappy | 46 | 28 | 11 |
| mad | 216 | 80 | 53 |
| dumb | 89 | 45 | 22 |

# The lifecycle of the common blogger...

| Word | 10s | 20s | 30s |
|---|---|---|---|
| maths | 105 | 3 | 2 |
| homework | 137 | 18 | 15 |
| bored | 384 | 111 | 47 |
| sis | 74 | 26 | 10 |
| boring | 369 | 102 | 63 |
| awesome | 292 | 128 | 57 |
| mum | 125 | 41 | 23 |
| crappy | 46 | 28 | 11 |
| mad | 216 | 80 | 53 |
| dumb | 89 | 45 | 22 |

| Word | 10s | 20s | 30s |
|---|---|---|---|
| semester | 22 | 44 | 18 |
| apartment | 18 | 123 | 55 |
| drunk | 77 | 88 | 41 |
| beer | 32 | 115 | 70 |
| student | 65 | 98 | 61 |
| album | 64 | 84 | 56 |
| college | 151 | 192 | 131 |
| someday | 35 | 40 | 28 |
| dating | 31 | 52 | 37 |
| bar | 45 | 153 | 111 |

# The lifecycle of the common blogger...

| Word | 10s | 20s | 30s |
|---|---|---|---|
| maths | 105 | 3 | 2 |
| homework | 137 | 18 | 15 |
| bored | 384 | 111 | 47 |
| sis | 74 | 26 | 10 |
| boring | 369 | 102 | 63 |
| awesome | 292 | 128 | 57 |
| mum | 125 | 41 | 23 |
| crappy | 46 | 28 | 11 |
| mad | 216 | 80 | 53 |
| dumb | 89 | 45 | 22 |

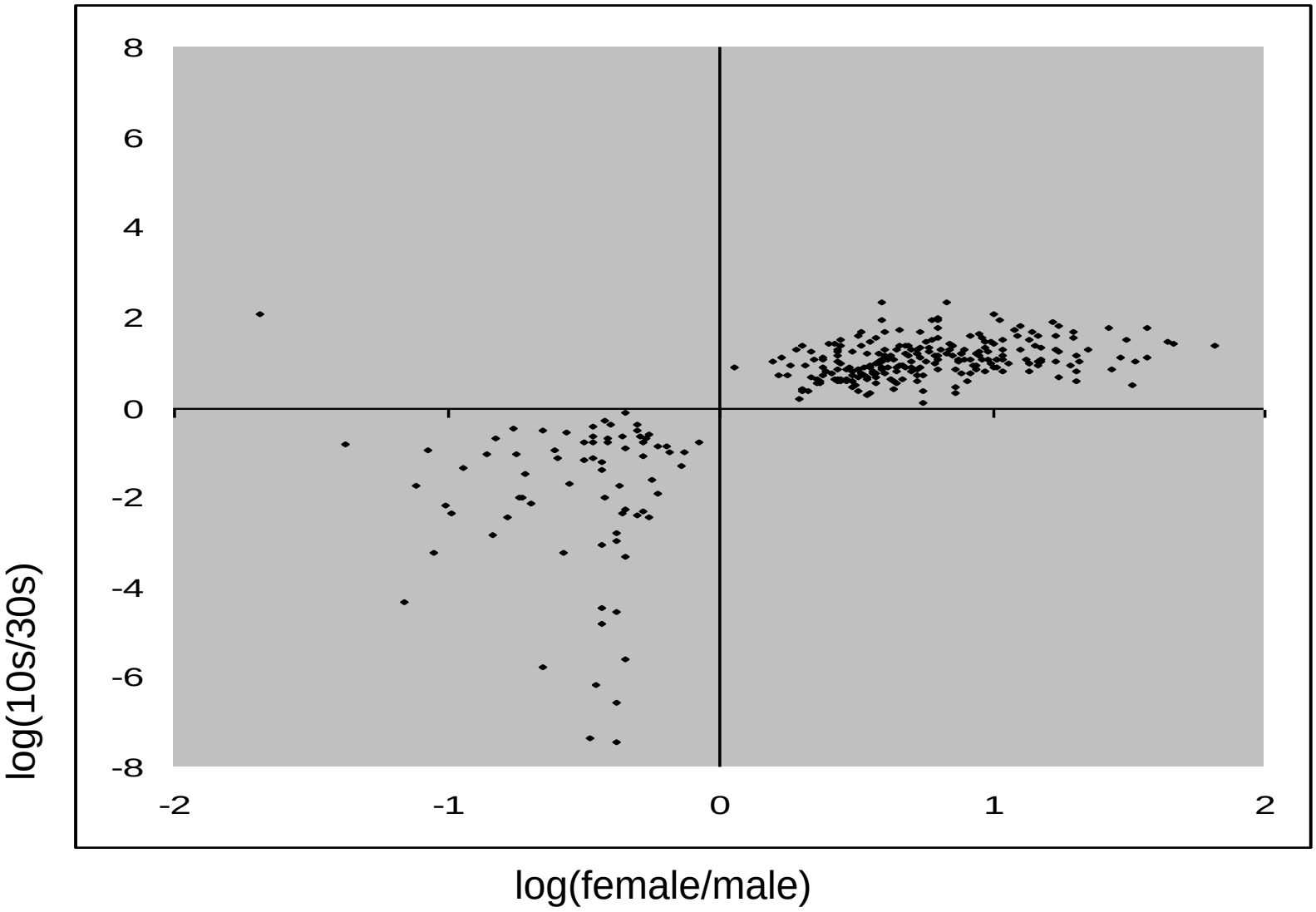
| Word | 10s | 20s | 30s |
|---|---|---|---|
| semester | 22 | 44 | 18 |
| apartment | 18 | 123 | 55 |
| drunk | 77 | 88 | 41 |
| beer | 32 | 115 | 70 |
| student | 65 | 98 | 61 |
| album | 64 | 84 | 56 |
| college | 151 | 192 | 131 |
| someday | 35 | 40 | 28 |
| dating | 31 | 52 | 37 |
| bar | 45 | 153 | 111 |

| Word | 10s | 20s | 30s |
|---|---|---|---|
| marriage | 27 | 83 | 141 |
| development | 16 | 50 | 82 |
| campaign | 14 | 38 | 70 |
| tax | 14 | 38 | 72 |
| local | 38 | 118 | 185 |
| democratic | 13 | 29 | 59 |
| son | 51 | 92 | 237 |
| systems | 12 | 36 | 55 |
| provide | 15 | 54 | 69 |
| workers | 10 | 35 | 46 |

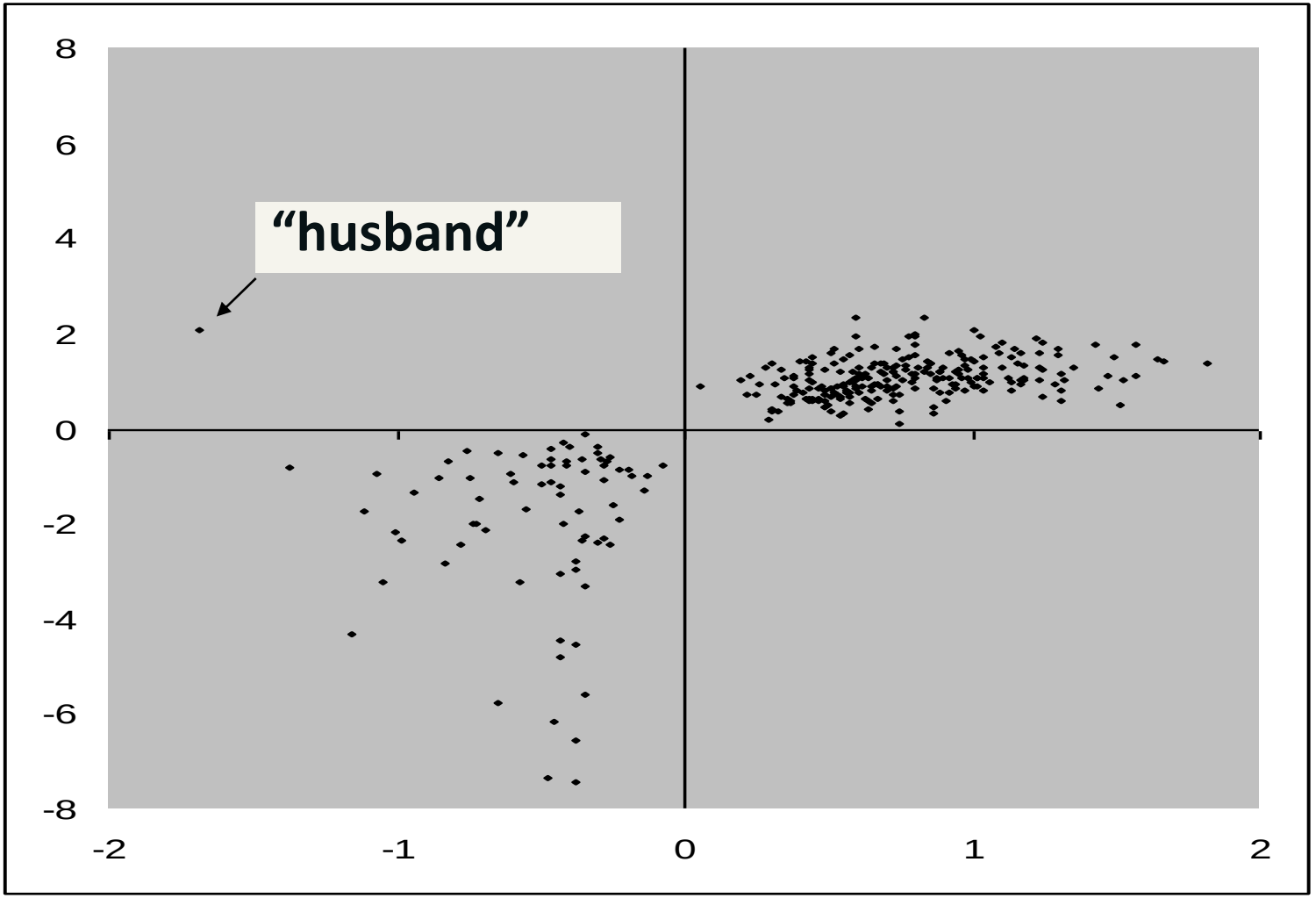
# Relating age & gender

- Now…is there a linguistic connection between age and gender?
- Consider the most distinctive words for both Age and Gender:
  - Intersect the 1000 words with **highest Age information gain** and the 1000 words with **highest Gender information gain**
  - Total of 316 words
  - Plot log(30s/10s) vs. log(male/female)

# Relating age & gender



# Relating age & gender



# Gender & age: pre PAN state of the art

| AUTHOR | COLLECTION | FEATURES | RESULTS | OTHER CHARACTERISTICS |
|---|---|---|---|---|
| **Argamon et al., 2002** | British National Corpus | Part-of-speech | Gender: 80% accuracy | |
| **Koppel et al., 2003** | Blogs | Lexical and syntactic features | Gender: 80% accuracy | Self-labeling |
| **Schler et al., 2006** | Blogs | Stylistic features + content words with the highest information gain | Gender: 80% accuracy<br>Age: 75% accuracy | |
| **Goswami et al., 2009** | Blogs | Slang + sentence length | Gender: 89.18 accuracy<br>Age: 80.32 accuracy | |
| **Zhang & Zhang, 2010** | Segments of blog | Words, punctuation, average words/sentence length, POS, word factor analysis | Gender: 72.10 accuracy | |
| **Nguyen et al., 2011, 2013** | Blogs & Twitter | Unigrams, POS, LIWC | Correlation: 0.74<br>Mean absolute error: 4.1 - 6.8 years | Manual labeling<br>Age as continuous variable |
| **Peersman et al., 2011** | Netlog | Unigrams, bigrams, trigrams and tetagrams | Gender+Age: 88.8 accuracy | Self-labeling, min 16 plus 16,18,25 |

# Author profiling @ PAN

- **CLEF 2013: Age and gender in social media**
- **CLEF 2014: Age and gender in social media, Twitter, blogs, reviews**
- CLEF 2015: Age, gender, personality in Twitter
- CLEF 2016: Cross-genre age and gender
- FIRE 2016: Personality in source code
- CLEF 2017: Gender and language variety identification in Twitter
- FIRE 2017: Native Indian language identification
- FIRE 2017: Cross-genre gender identification in Russian
- CLEF 2018: Multimodal (text + image) gender in Twitter

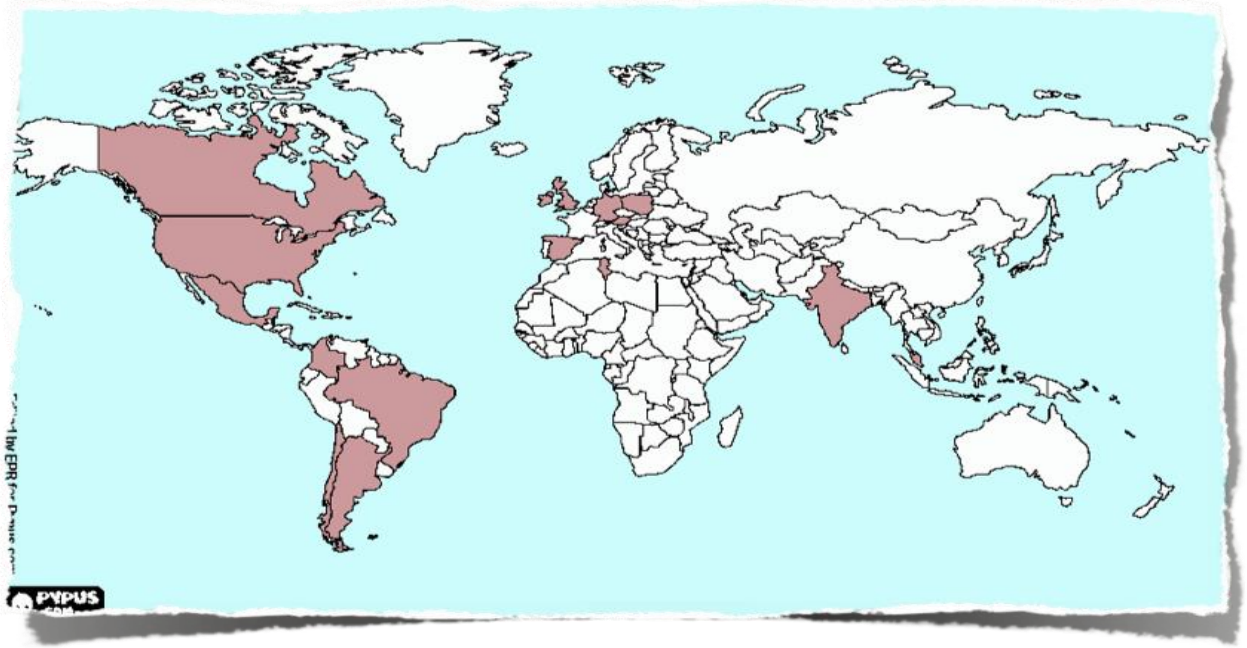
# Author profiling @ PAN

- **CLEF 2019: Bots and gender profiling on Twitter**
- **CLEF 2020: Profiling fake news spreaders on Twitter**
- **CLEF 2021: Profiling hate speech spreaders on Twitter**
- CLEF 2022: Profiling irony and stereotype spreaders on Twitter
- CLEF 2023: Profiling cryptocurrency influencers with few-shot learning

# Author profiling: PAN @CLEF 2013

- Teams submitting results: 21 (Registered teams: 64)
- (Towards) **big data**: 400,000 social media texts

including **chat lines of potential pedophiles** (task in 2012)



- **Age classes**: 10s (13-17), 20s (23-27), 30s (33-48)
- **Languages**: English and Spanish

# Results: EN vs. ES

| English | | | |
|---|---|---|---|
| Team | Total | Gender | Age |
| Meina | 0.3894 | 0.5921 | 0.6491 |
| Pastor L. | 0.3813 | 0.5690 | 0.6572 |
| Seifeddine | 0.3677 | 0.5816 | 0.5897 |
| Santosh | 0.3508 | 0.5652 | 0.6408 |
| Yong Lim | 0.3488 | 0.5671 | 0.6098 |
| Ladra | 0.3420 | 0.5608 | 0.6118 |
| Aleman | 0.3292 | 0.5522 | 0.5923 |
| Gillam | 0.3268 | 0.5410 | 0.6031 |
| Kern | 0.3115 | 0.5267 | 0.5690 |
| Cruz | 0.3114 | 0.5456 | 0.5966 |
| Pavan | 0.2843 | 0.5000 | 0.6055 |
| Caurcel Diaz | 0.2840 | 0.5000 | 0.5679 |
| H. Farias | 0.2816 | 0.5671 | 0.5061 |
| Jankowska | 0.2814 | 0.5381 | 0.4738 |
| Flekova | 0.2785 | 0.5343 | 0.5287 |
| Weren | 0.2564 | 0.5044 | 0.5099 |
| Sapkota | 0.2471 | 0.4781 | 0.5415 |
| De-Arteaga | 0.2450 | 0.4998 | 0.4885 |
| Moreau | 0.2395 | 0.4941 | 0.4824 |
| baseline | 0.1650 | 0.5000 | 0.3333 |
| Gopal Patra | 0.1574 | 0.5683 | 0.2895 |
| Cagnina | 0.0741 | 0.5040 | 0.1234 |

| Spanish | | | |
|---|---|---|---|
| Team | Total | Gender | Age |
| Santosh | 0.4208 | 0.6473 | 0.6430 |
| Pastor L. | 0.4158 | 0.6299 | 0.6558 |
| Cruz | 0.3897 | 0.6165 | 0.6219 |
| Flekova | 0.3683 | 0.6103 | 0.5966 |
| Ladra | 0.3523 | 0.6138 | 0.5727 |
| De-Arteaga | 0.3145 | 0.5627 | 0.5429 |
| Kern | 0.3134 | 0.5706 | 0.5375 |
| Yong Lim | 0.3120 | 0.5468 | 0.5705 |
| Sapkota | 0.2934 | 0.5116 | 0.5651 |
| Pavan | 0.2824 | 0.5000 | 0.5643 |
| Jankowska | 0.2592 | 0.5846 | 0.4276 |
| Meina | 0.2549 | 0.5287 | 0.4930 |
| Gillam | 0.2543 | 0.4784 | 0.5377 |
| Moreau | 0.2539 | 0.4967 | 0.5049 |
| Weren | 0.2463 | 0.5362 | 0.4615 |
| Cagnina | 0.2339 | 0.5516 | 0.4148 |
| Caurcel Diaz | 0.2000 | 0.5000 | 0.4000 |
| H. Farias | 0.1757 | 0.4982 | 0.3554 |
| baseline | 0.1650 | 0.5000 | 0.3333 |
| Aleman | 0.1638 | 0.5526 | 0.2915 |
| Seifeddine | 0.0287 | 0.5455 | 0.0512 |
| Gopal Patra | – | – | – |

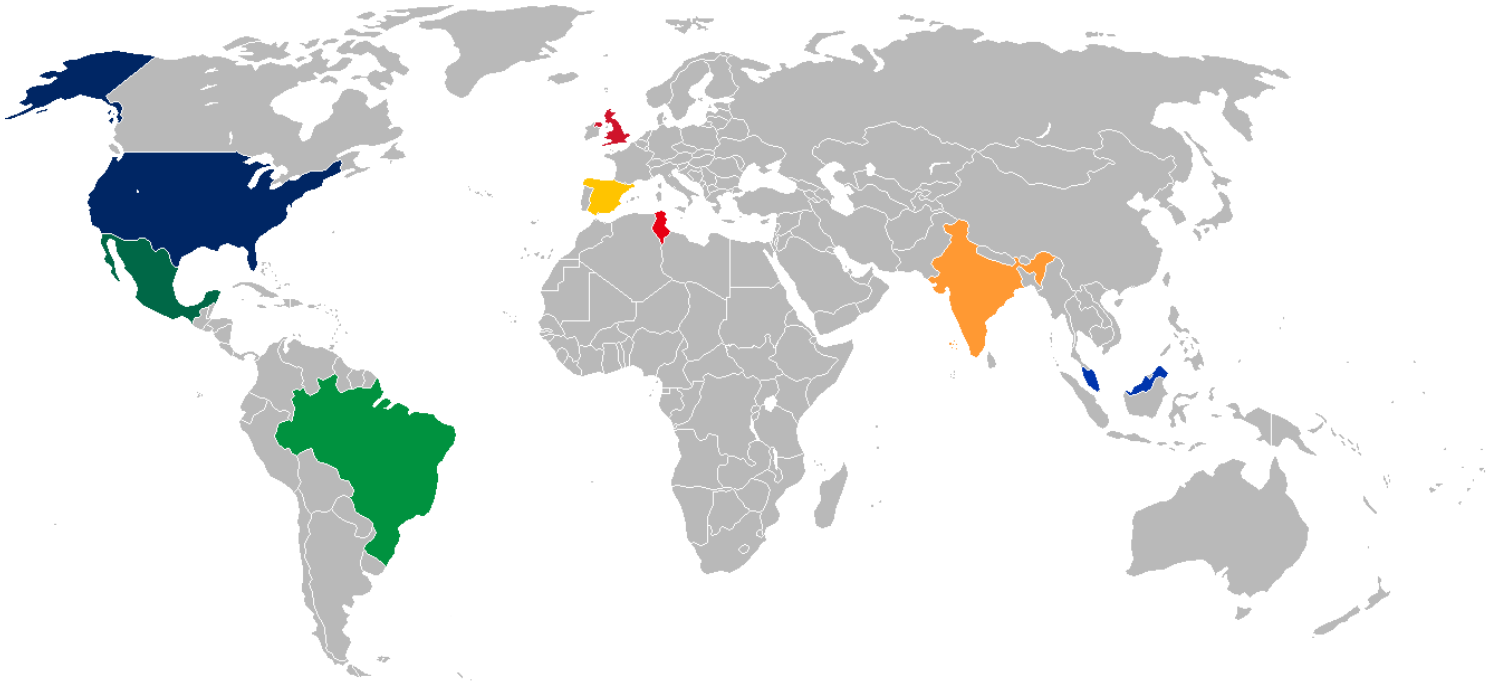
# Features

- Stylistic: frequency of punctuation marks, capital letters,…
- Part of Speech
- Readability measures
- Dictionary-based words,  topic-based words
- Collocations
- Character or word n-grams
- Slang words, character flooding
- Emoticons
- Emotion words

F. Rangel, P. Rosso, M. Koppel, E. Stamatatos, and G. Inches. Overview of the Author Profiling Task at PAN 2013 - Notebook for PAN at CLEF 2013. CEUR Workshop Proceedings Vol. 1179. 2013.

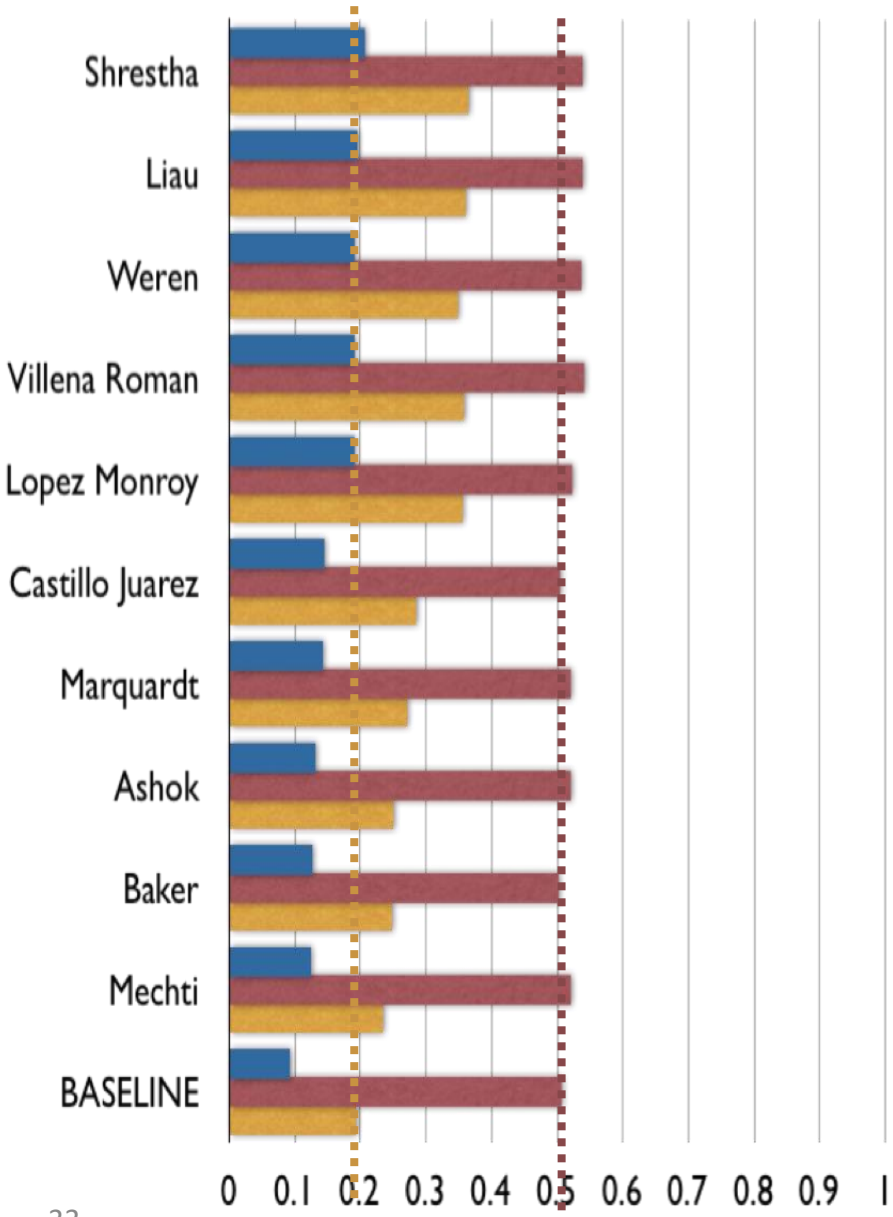
# Author profiling PAN @CLEF 2014

- Teams submitting results: 10

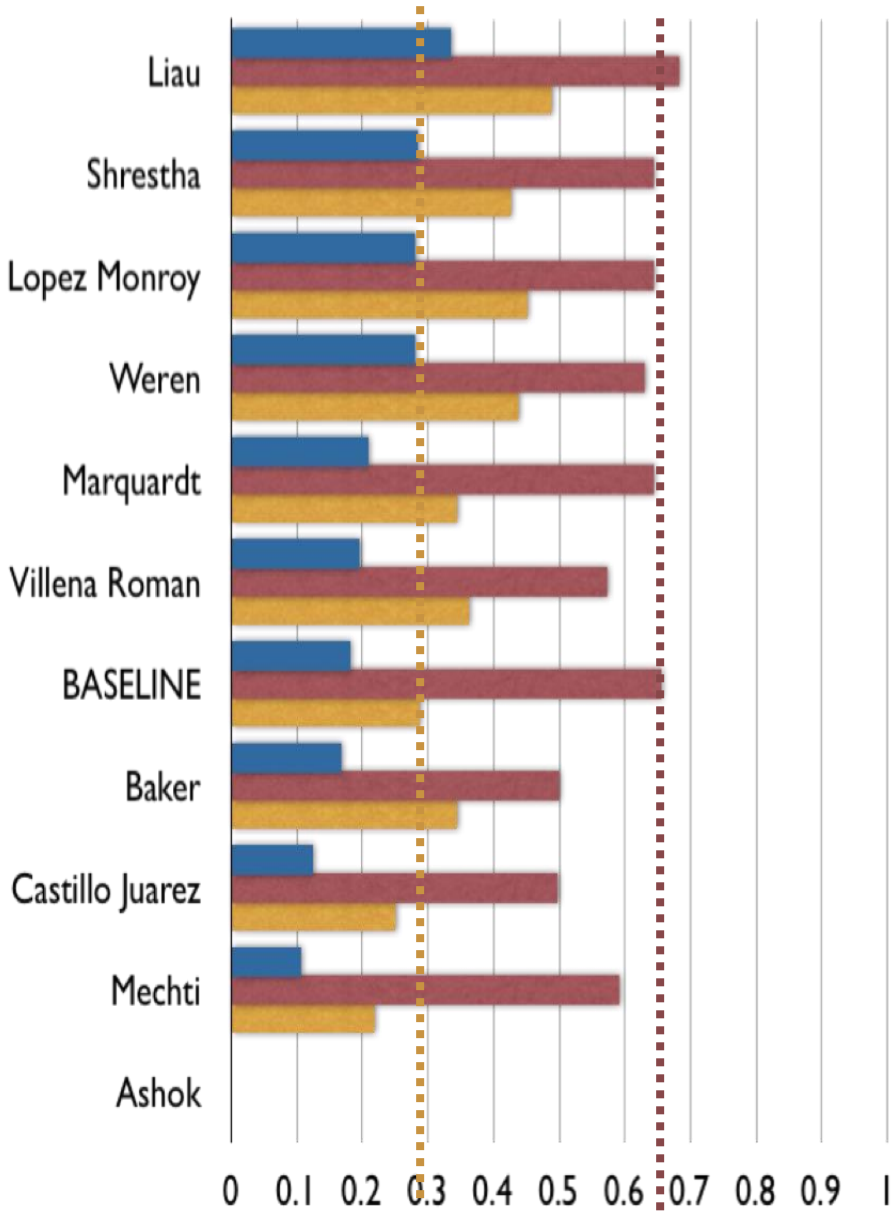


- **Social media + blogs + Twitter + reviews**
- **Age classes**: 18-24, 25-34, 35-49, 50-64, 65+
- **Languages**: English and Spanish

# Results in social media

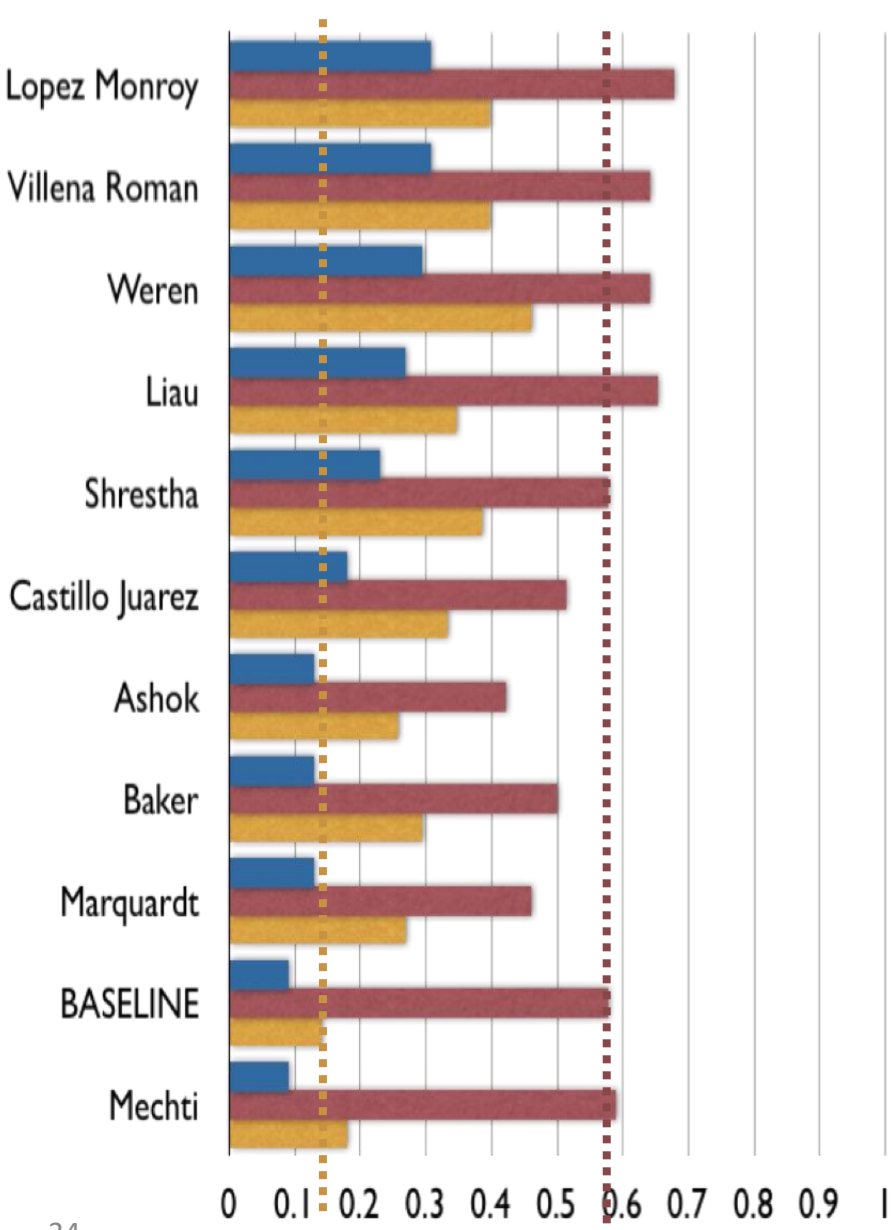



EN (joint, gender, age)

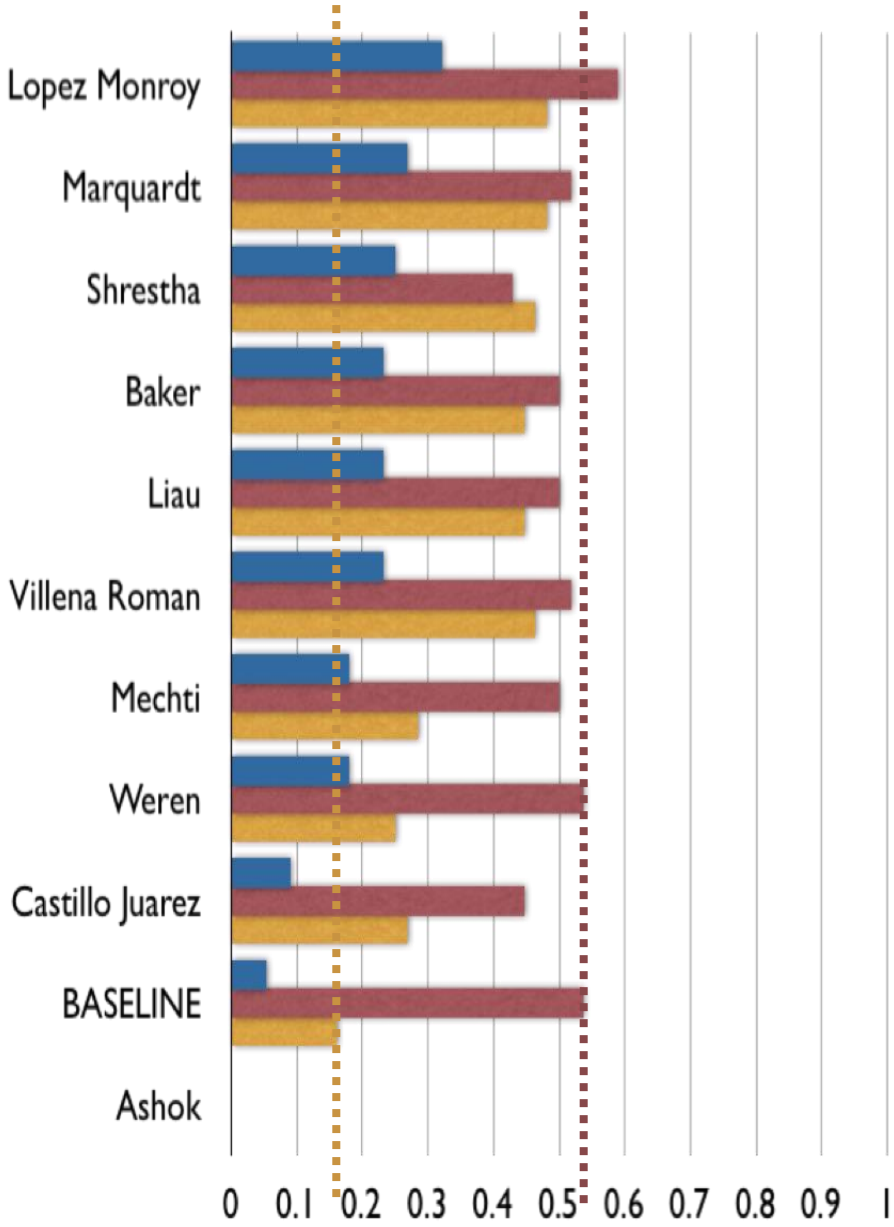



ES (joint, gender, age)

# Results in blogs

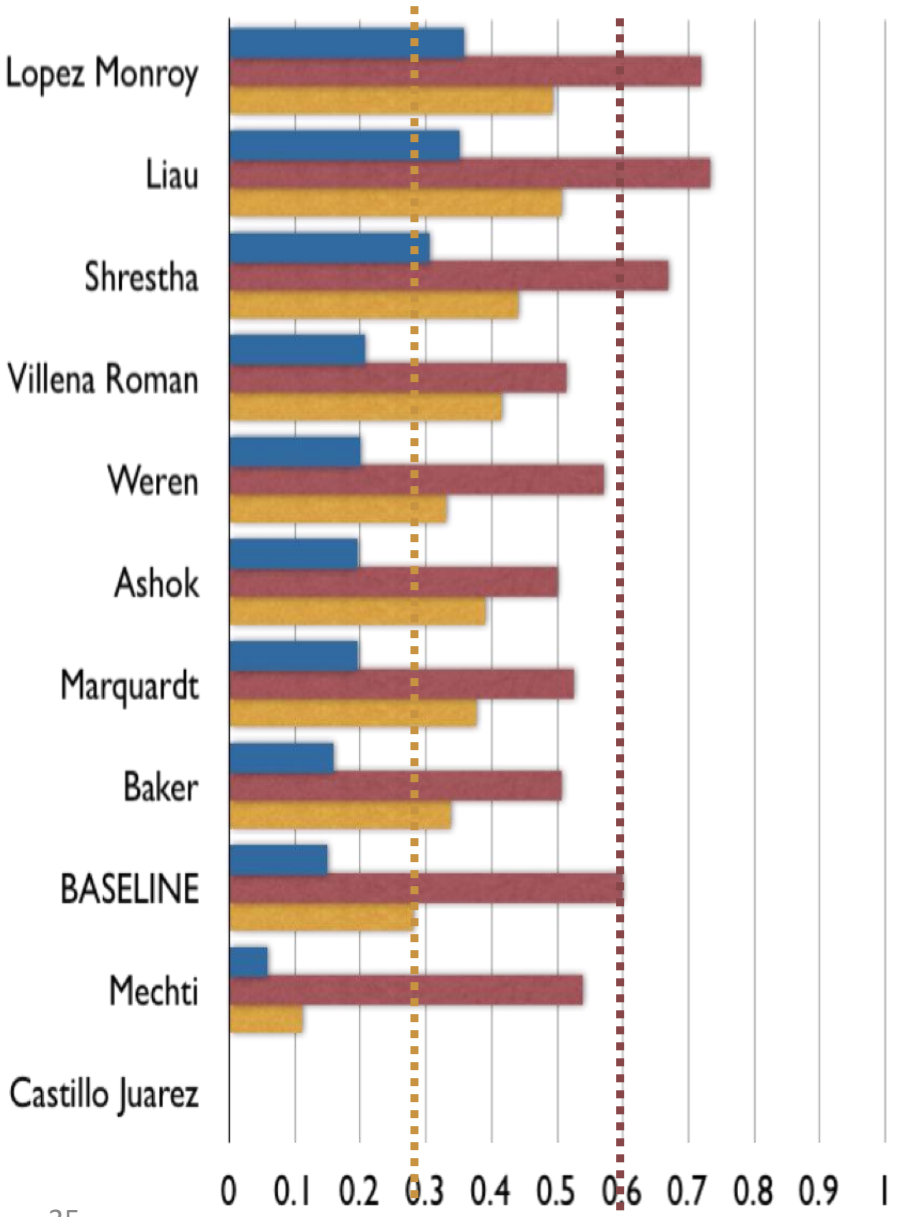



EN (joint, gender, age)

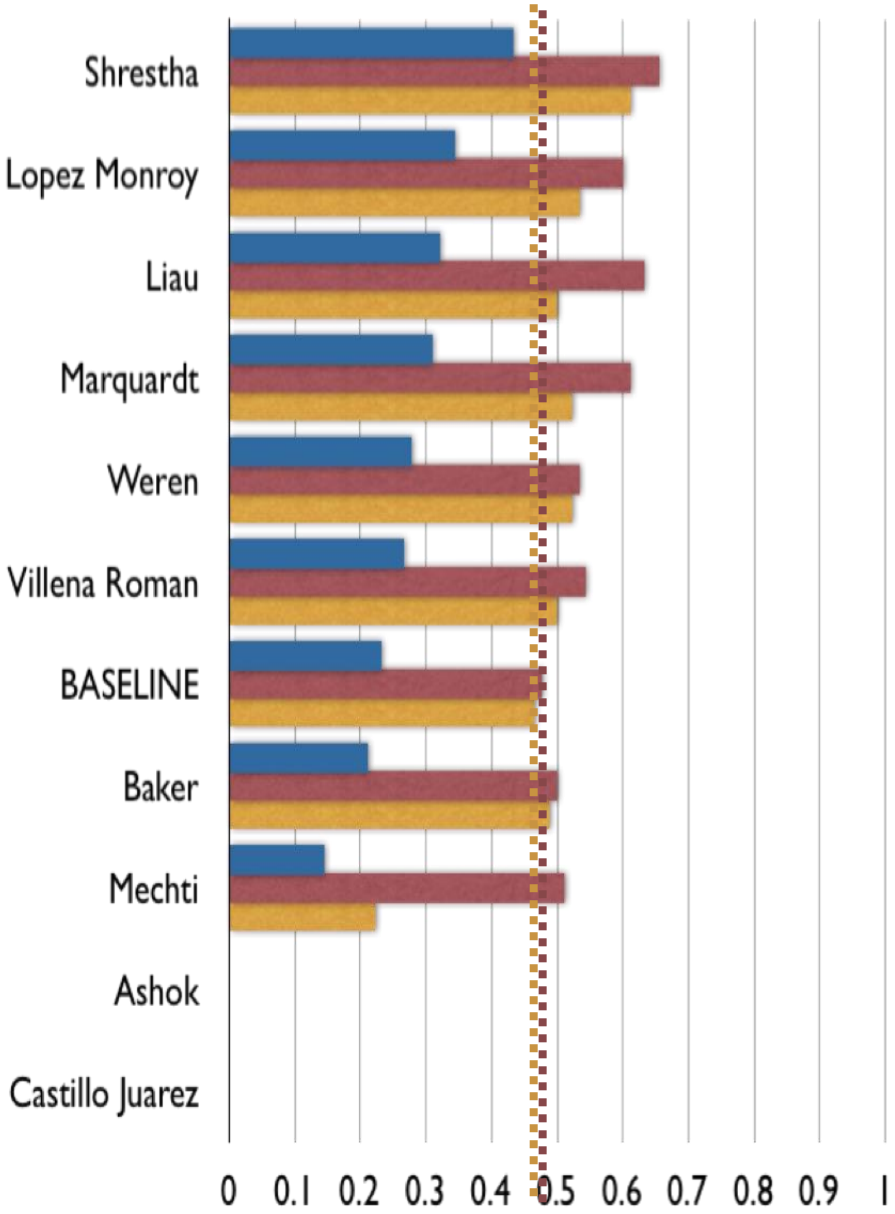



ES (joint, gender, age)

# Results in Twitter

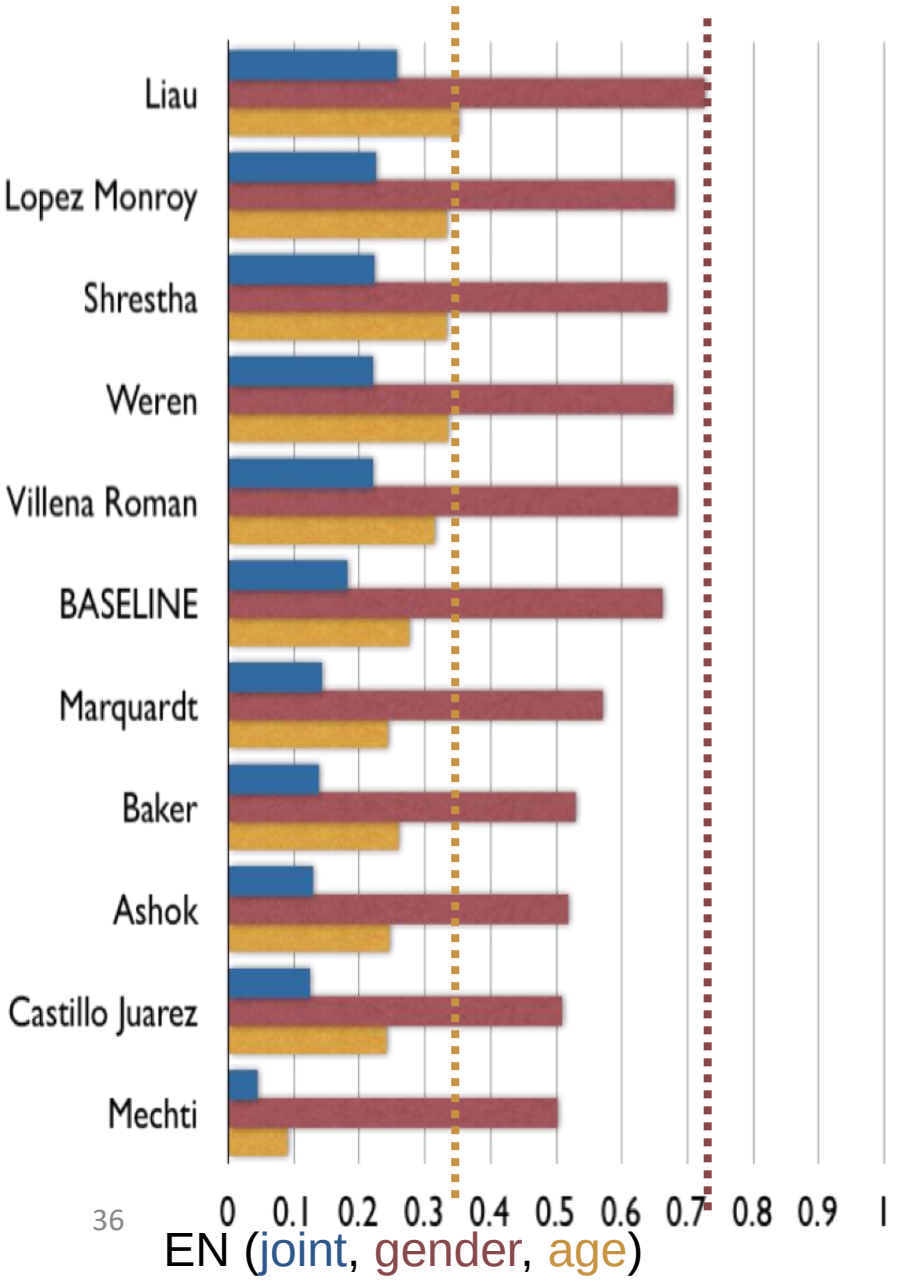



EN (joint, gender, age)



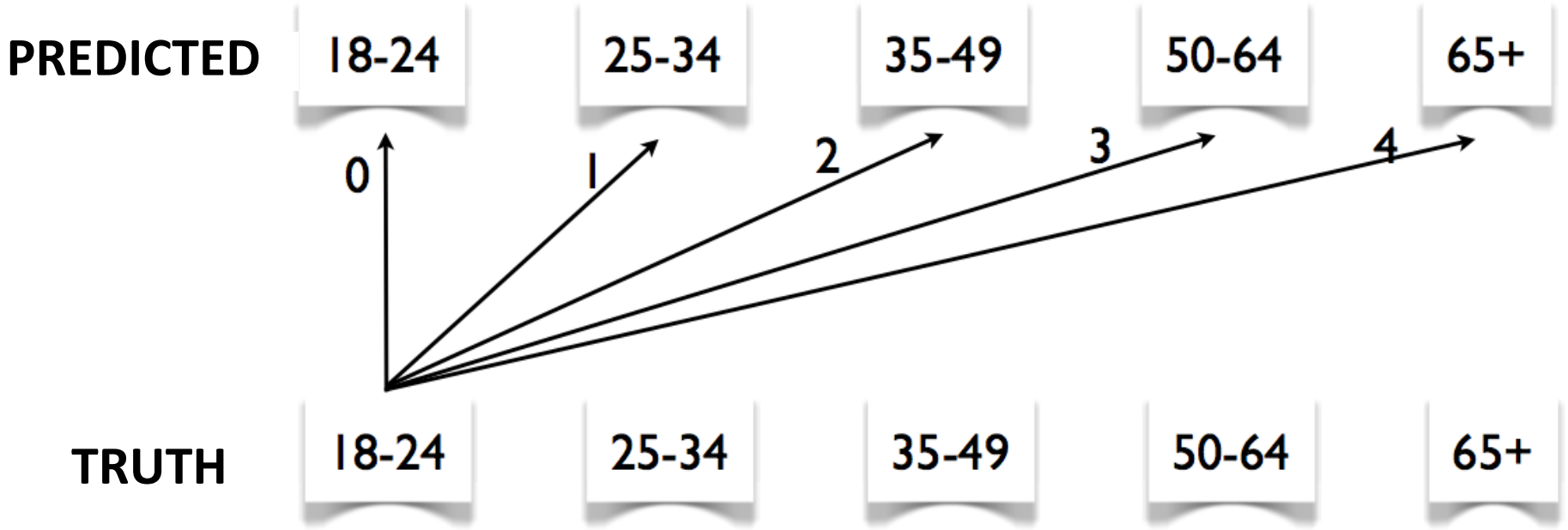

ES (joint, gender, age)

# Results in reviews

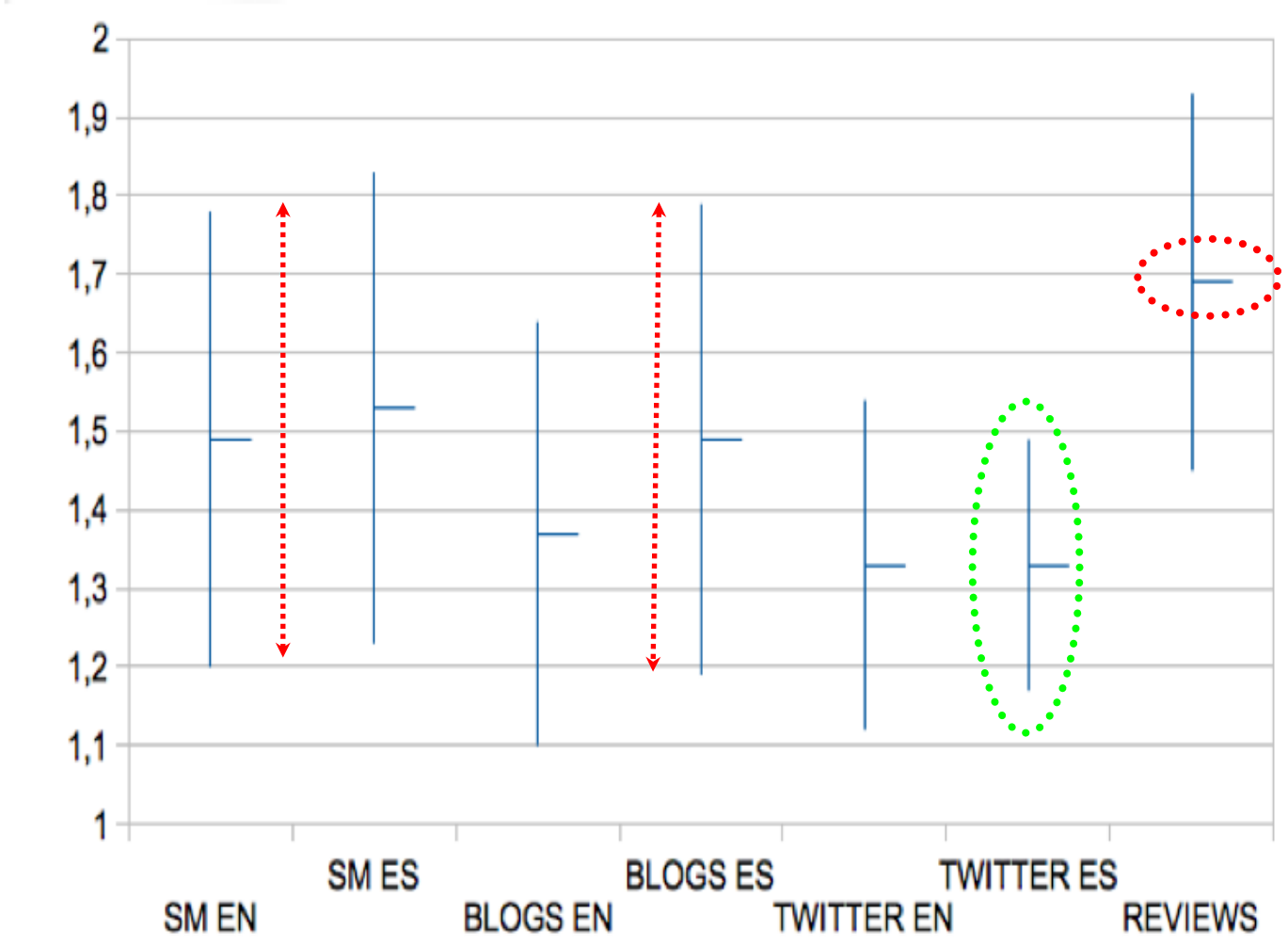



The problem of deceptive opinions

# Distances in misclassified age



# Distances in misclassified age




Twitter: more spontaneous way to communicate

# Approaches: features

- Similar features than in 2013:

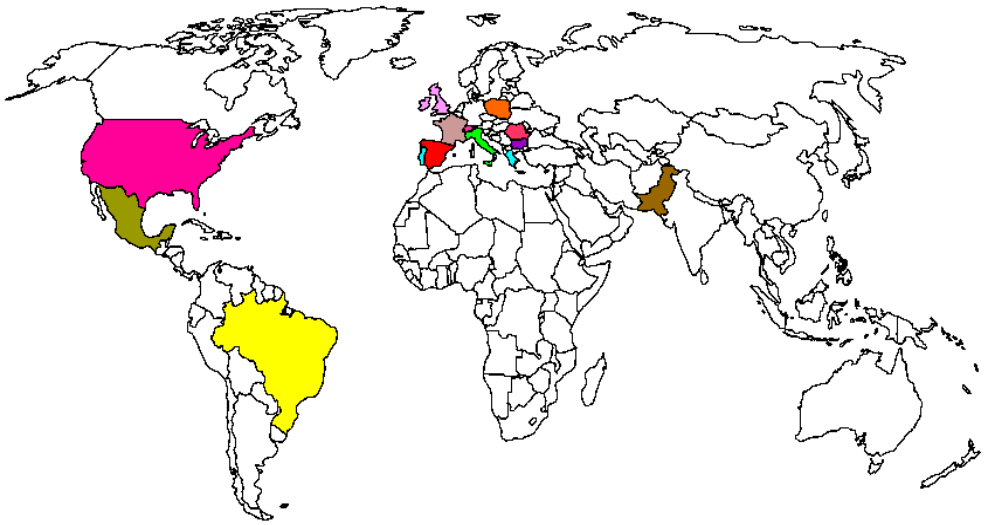
  content (bag of words, word n-grams) and stylistic

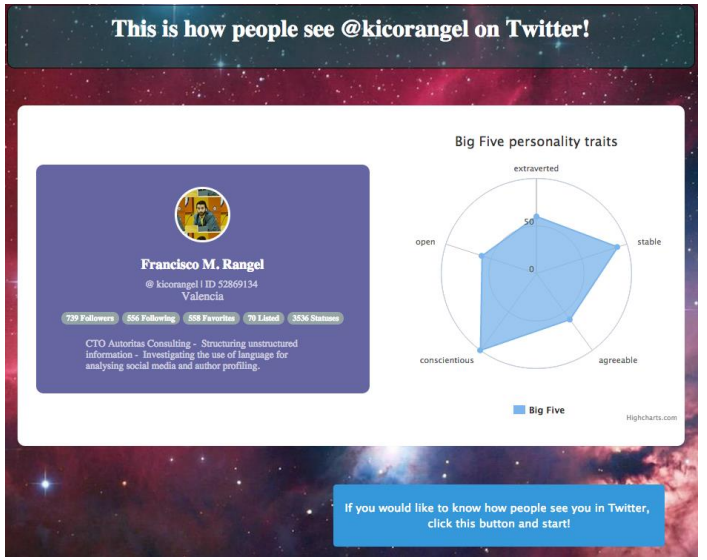
- frequency of words related to different psycholinguistic concepts, extracted from: LIWC and MRC psycholinguistic database

F. Rangel, P. Rosso, I. Chugur, M. Potthast, M. Trenkman, B. Stein, B. Verhoeven, and W. Daelemans. Overview of the 2nd Author Profiling Task at PAN 2014—Notebook for PAN at CLEF 2014. CEUR Workshop Proceedings Vol. 1180, pp. 898-927, 2014.

# 2015: Age, gender and personality

- Teams submitting results: 22






- **Personality:** http://your-personality-test.com/
- **Age classes**: 18-24, 25-34, 35-49, 50+
- **Languages**: English, Spanish, Dutch, Italian
- **Results**: **Gender** EN: 85.9% **Age** EN: 83.8% **Personality** openness (root-mean-square error 0.097)

# Personality questionnaire

1. I am a reserved person
2. I trust other people
3. I tend to be lazy
4. I am generally relaxed, not stressed
5. I have few artistic interests
6. I am sociable
7. I tend to find fault with others
8. I do my job well
9. I get nervous easily
10. I have an active imagination

(answers from 1 to 5: http://mypersonality.autoritas.net/ )

# Big Five personality traits

**Big Five** personality traits (given a text, determine if the author is):

**O**pen to new experiences
**C**onscientious: tends to be careful and scrupuolus
**E**xtroverted: gets energy from being around people
**A**greeable: prefers to agree with others
**N**eurotic: tends to worry about things

# PR-SOCO : PAN @FIRE 2016

## Personality Recognition in SOurce Code

- Big five personality traits from Java source codes
- 11 teams submitted 49 runs
- http://www.autoritas.es/prsoco/
- Best results for the openness trait

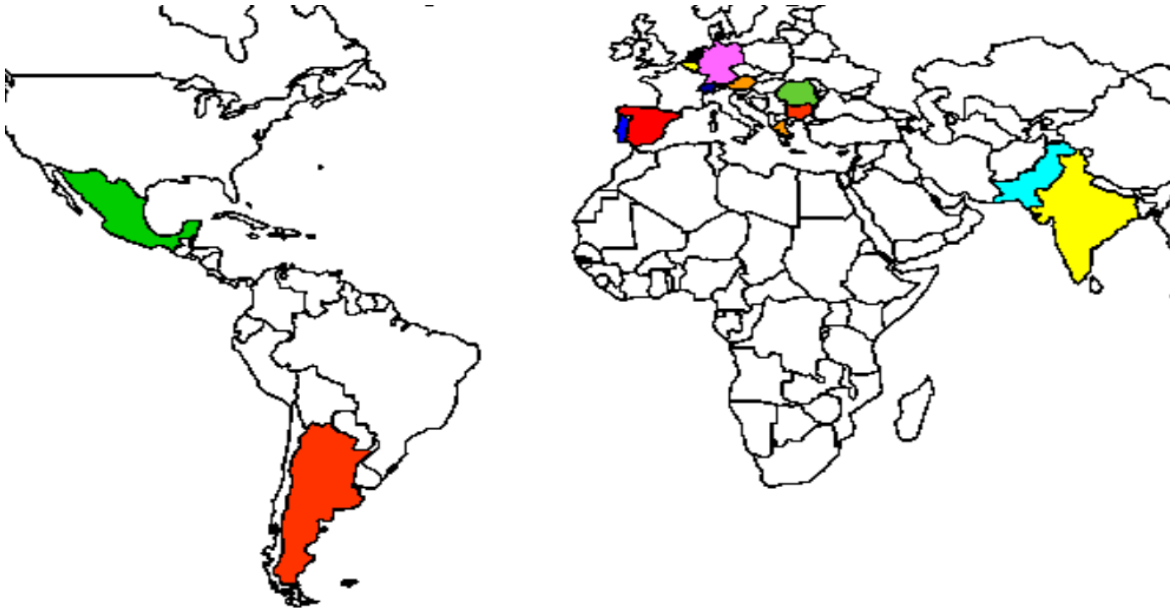
  in line with the results obtained on Twitter data

  at PAN @ CLEF 2015

Rangel F., González F., Restrepo-Calle F., Montes M., Rosso P. PAN at FIRE: Overview of the PR-SOCO Track on Personality Recognition in SOurce COde. Notebook for PAN at FIRE 2016. CEUR Workshop Proceedings. CEUR-WS.org, vol. 1737, pp. 1-5. 2016

# 2016: Cross-genre age and gender

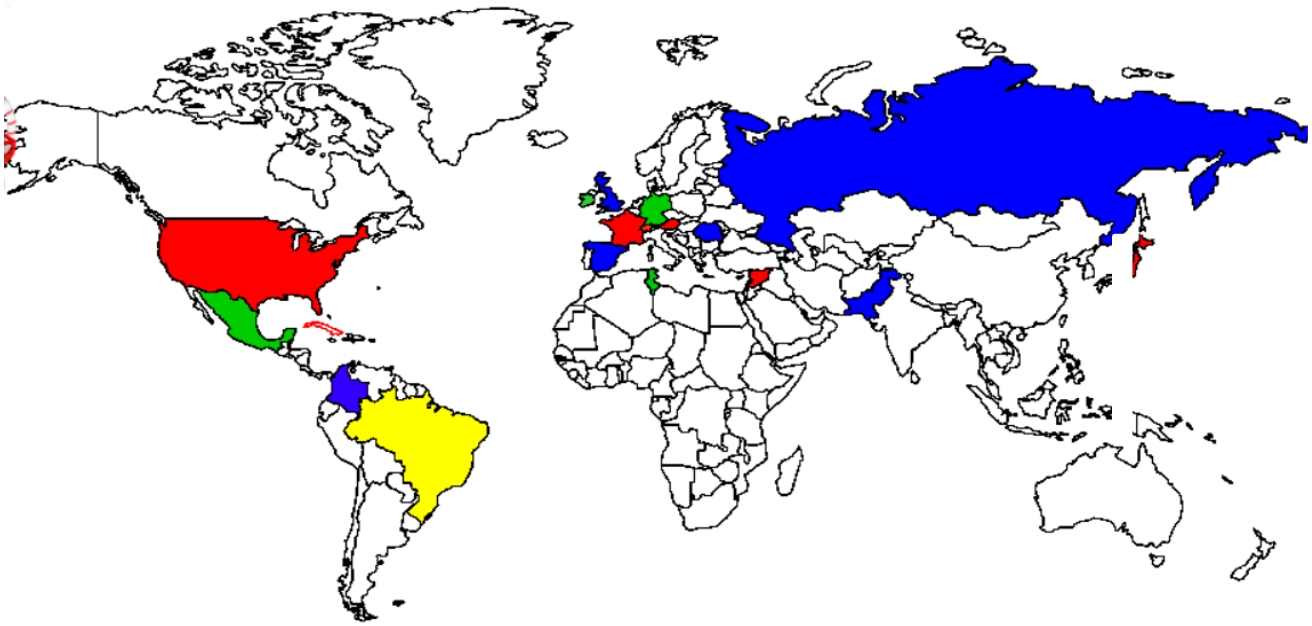
- Teams submitting results: 22
- **Training** Test: social media and blogs



- **Age classes**: 18-24, 25-34, 35-49, 35-49, 50-64, 65+
- **Languages**: English, Spanish and Dutch
- **Results (blogs)**: **Gender** EN: 75.6% **Age** EN: 58.9%

# 2017: Gender and language variety

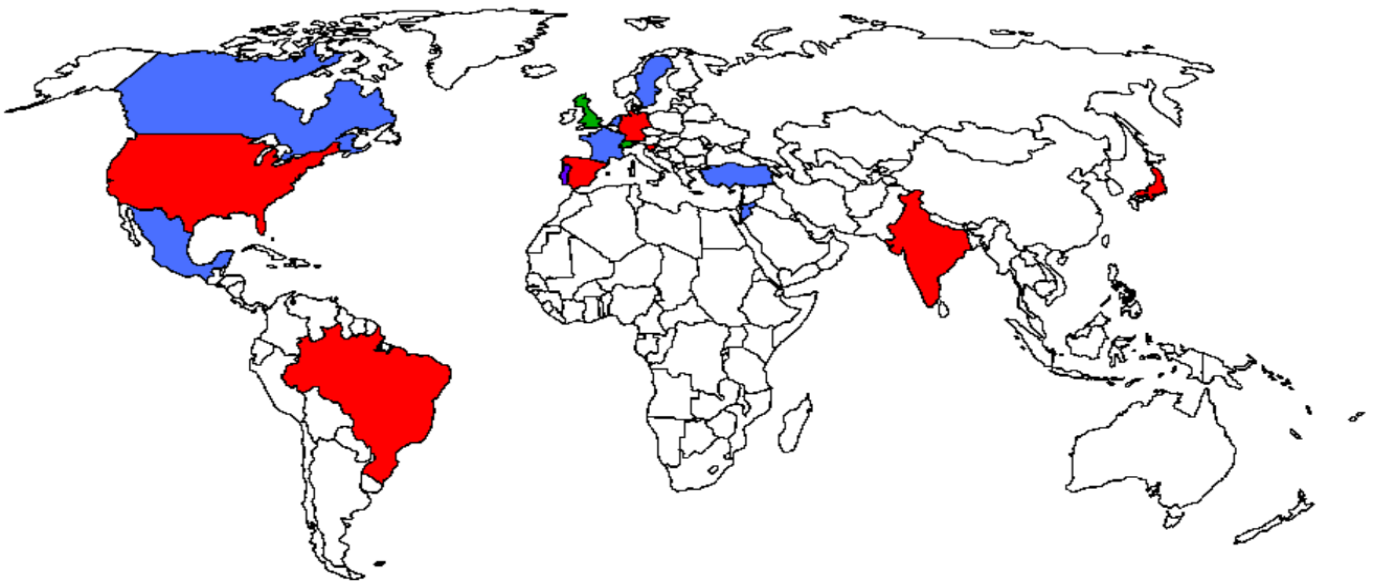
- Teams submitting results: 22



- **Languages**: English, Spanish, Portuguese, Arabic
- **Results**: **Gender** PT: 87% EN: 82% ES: 83% AR: 80%
  **Language variety** (lowest) EN: 90% AR: 83%

# 2018: Multimodal gender identification

- Teams submitting results: 23



- **Languages**: English, Spanish, Arabic
- **Results**: EN, ES: 82% AR: 81.7%

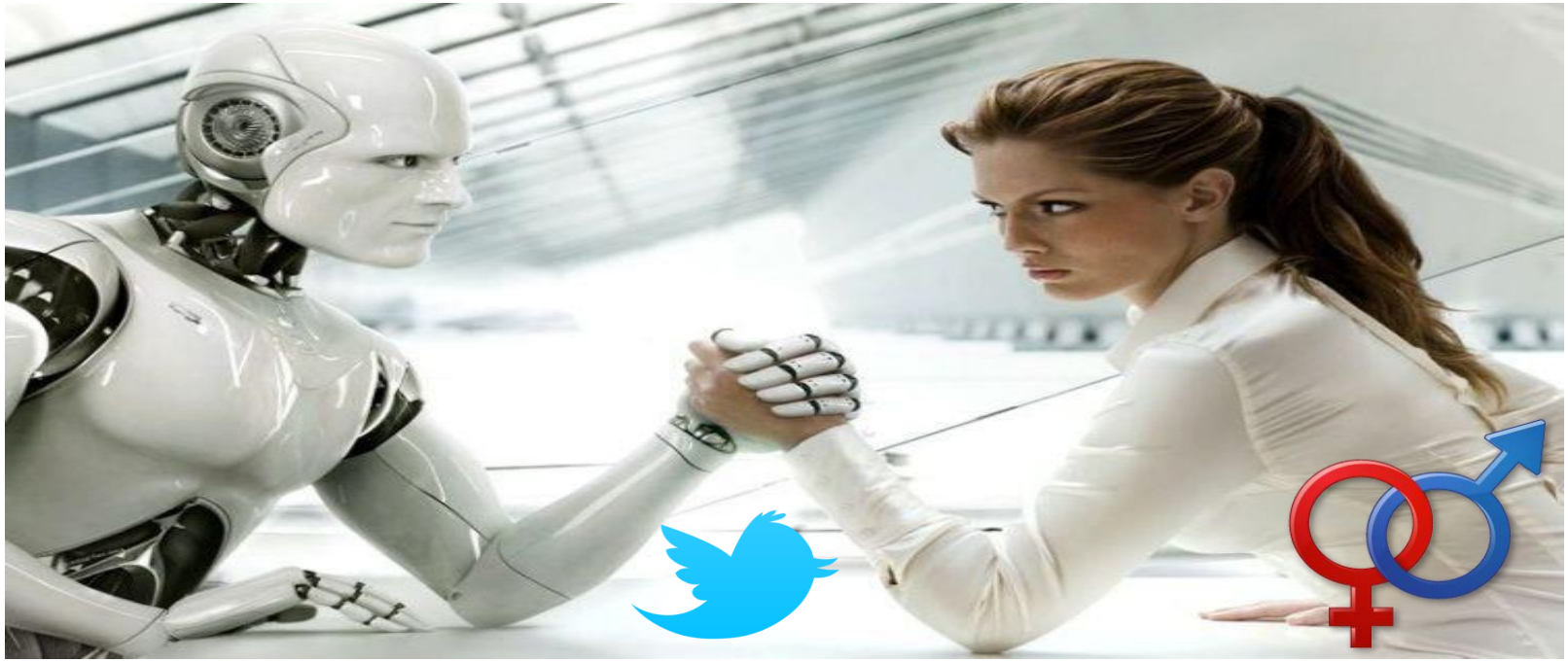
# Outline

- Profiling gender & age
- Author profiling at PAN-2013 and PAN-2014
- Profiling **bots** at PAN-2019
- Profiling **fake news spreaders** at PAN-2020
- Profiling **haters** at PAN-2021


https://pan.webis.de/

# Author profiling: PAN@CLEF 2019

## Bots and gender profiling



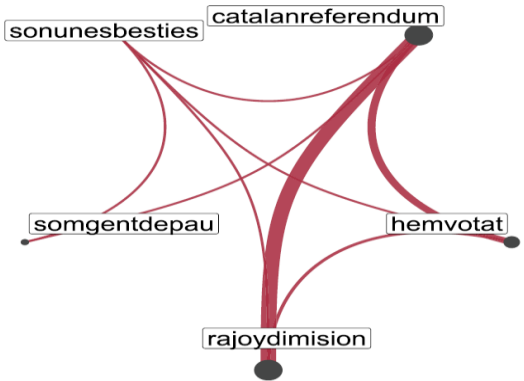
Rangel F., Rosso P. Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling in Twitter. In: L. Cappellato, N. Ferro, D. E. Losada and H. Müller (eds.) CLEF 2019 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings.CEUR-WS.org, vol. 2380, 2019

# Bots: propaganda, fake news, inflammatory content

- Bots may **influence** users with **commercial, political or ideological** purposes…
- **Polarization** and spread **disinformation and fake news**
- US 2016 Presidencial election, Brexit, 1 Oct 2017 referendum for the Catalan independence:

# Bots: propaganda, fake news, inflammatory content

- Bots may **influence** users with **comercial, political or ideological** purposes…
- **Polarization** and spread **disinformation and fake news**
- US 2016 Presidencial election, Brexit, 1 Oct 2017 referendum for the Catalan independence: **23.5%** of 3.6 million tweets generated **by bots 19%** of the interactions were **from bots to humans**

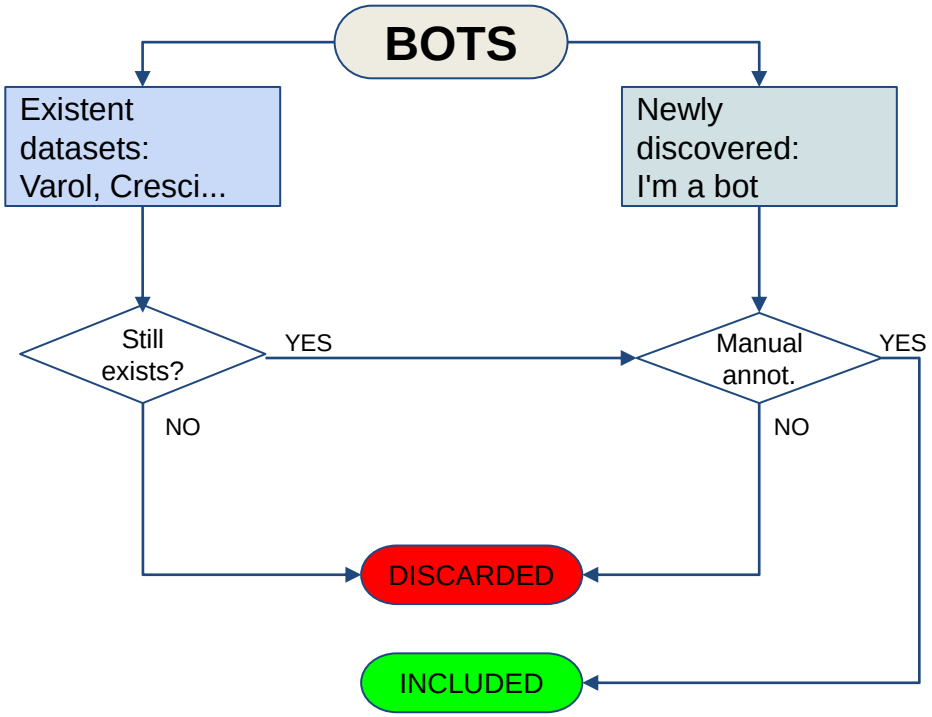



Massimo Stella, Emilio Ferrara, and Manlio De Domenico. Bots increase exposure to negative and inflammatory content in online social systems. Proc. of the National Academy of Sciences of the United States of America, 115(49):12435–12440, 2018.

# Bots and gender profiling

- How difficult / easy is to discriminate **bots from humans** on the basis only on **textual features**?
- What are the **most difficult types of bots**?

# Bots and humans accounts




**Humans** selected from PAN-AP'17 author profiling + manual annotation

# Dataset

- Twitter accounts identified as bots in existing datasets + new ones

- Each **author (bot or human) feed** is composed by exactly **100 tweets**

| | | (EN) English | | | | (ES) Spanish | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Bots | Humans | | Total | Bots | Humans | | Total |
| | | | F | M | | | F | M | |
| **Training** | **Training** | 1,440 | 720 | 720 | 2,880 | 1,040 | 520 | 520 | 2,080 |
| | **Development** | 620 | 310 | 310 | 1,240 | 460 | 230 | 230 | 920 |
| | **Total** | 2,060 | 1,030 | 1,030 | 4,120 | 1,500 | 750 | 750 | 3,000 |
| **Test** | | 1,320 | 660 | 660 | 2,640 | 900 | 450 | 450 | 1,800 |
| **Total** | | 3,380 | 1,690 | 1,690 | 6,760 | 2,400 | 1,200 | 1,200 | 4,800 |

# Types of bots

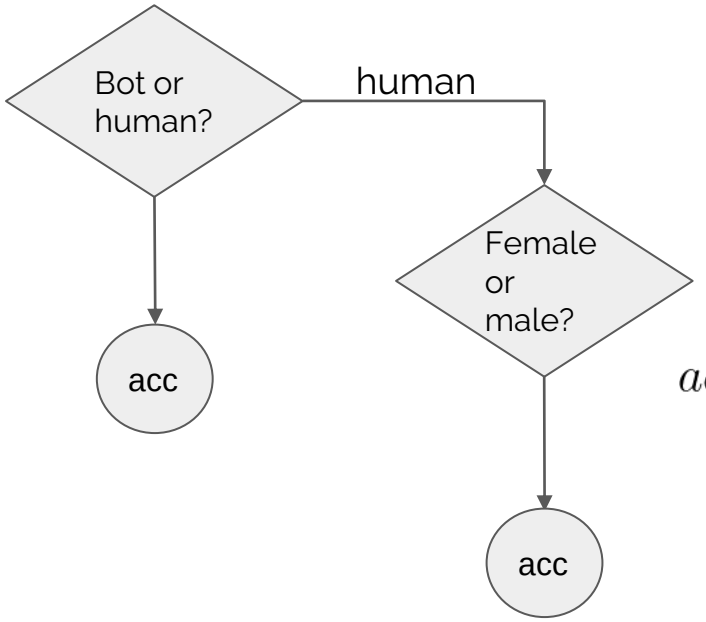
| | |
|---|---|
| **TEMPLATE** | The Twitter feed responds to a **predefined structure or template**, such as for example a Twitter account giving the state of the earthquakes in a region or **job offers** in a sector |
| **FEED** | The Twitter feed retweets or **shares news about a predefined topic**, such as for example regarding Trump's policies |
| **QUOTE** | The Twitter feed reproduces **quotes from famous books or songs, from celebrities** or people, or jokes |
| **ADVANCED** | Twitter feeds whose **language is generated** on the basis of more elaborated technologies such as Markov chains, **metaphors**, or in some cases, randomly choosing and merging texts from big corpora |

# Metaphormagnet

For example, the bot
**@metaphormagnet**
was developed by
**Tony Veale** and **Goufu Li**
to automatically generate metaphorical language


# Evaluation measures

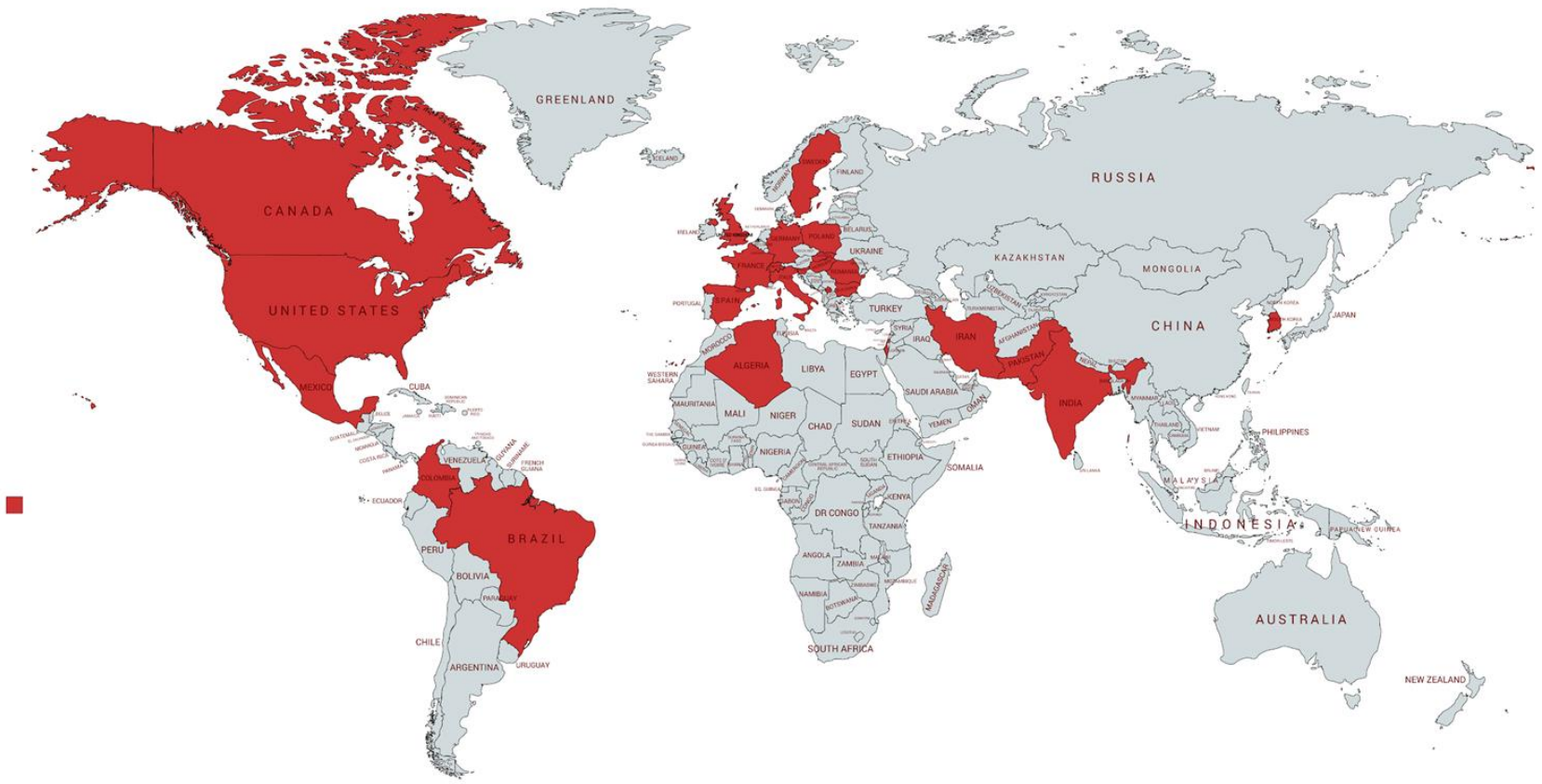
**Accuracy** is calculated per language and task:




$$acc_{[en|es]} = \frac{acc_{bots} + acc_{gender}}{2}$$

$$ranking = \frac{acc_{en} + acc_{es}}{2}$$

# Statistics




55+1 participants
26 countries

# Approaches

What kind of ...

Preprocessing

Features

Methods

... did the teams perform?

# Approaches: Preprocessing

| | |
|---|---|
| Twitter elements (URLs, users, hashtags, ...) | Van Halteren; Vogel; Polignano; Giachanou; Gishamer; Puertas; Saeed; Petritk; Valencia; Onose; Babaei; Yacob; Zhechev; Mahmood |
| Word segmentation | Gishamer; Joo |
| Tokenisation | Van Halteren; Polignano; Gishamer; Joo; Bacciu; Petritk; Goubin; Zhechev; Mahmood |
| Stemming / lemmatisation | Ikae; Joo; Saeed; Bacciu; Basile; Petritk; Babaei; Goubin; Zhechev; |
| Punctuation marks | Vogel; Saeed; Onose; Ribeiro; Goubin; Yacob; Zhechev; |
| Lowercase | Van Halteren; Vogel; Giachanou; Saeed; Ribeiro |
| Stopwords | Joo; Saeed; Babaei; Zhechev; |
| Character flooding | Vogel; Gishamer; Goubin |
| Latent Semantic Analysis | Rakesh |
| Short words | Vogel |
| Infrequent words | Ikae; Gishamer |
| Contractions and acronyms | Joo; Saeed |

# Approaches: Features

| | |
|---|---|
| Stylistic features:<br>- Number of occurrences<br>- Verbs, adjs, pronouns<br>- Number of hashtags, mentions, URLs...<br>- Upper vs. lower case<br>- Punctuation marks<br>- ... | Joo; Goubin; Ashraf; Cimino; Oliveira; Ikae; De la Peña; Johansson; Giachanou; Martinc; Przybyla; Van Halteren; Fernquist |
| N-gram models | Ispas; Bounaama; Rakesh; Valencia; Mahmood; Fahim; Espinosa; Pizarro; Martinc; Martinc; Dias; Vogel; Giachanou; De la Peña; Babaei; Saeed; Joo; Bacciu; Johansson; Fernquist; HaCohen; Gishamer |
| Emotional features | Cimino; Giachanou; Oliveira |
| Lexicon-based features | Gamallo |
| Compression algorithms | Fernquist |
| DNA-based approach | Kosmajac |
| Embeddings | Polignano; Fagni; Halvani; Onose; López-Santillán; Staykovsky; Joo |

# Approaches: Methods

| | | | |
|---|---|---|---|
| SVM | Vogel; Cimino; Fagni; Pizarro; Jimenez; HaCohen; Bacciu; Goubin; Srinivasarao; Mahmood; Yacob; Ribeiro; Babaei; Rakesh; Gishamer; Moryossef; Giachanou | | |
| Logistic regression | Gishamer; Moryossef; Valencia; Bolonyai; Przybyła | CatBoost | Fernquist |
| SpaCy | Moryossef | kNN | Ikae |
| Random Forest | Johansson; Moryossef | Multilayer Perceptron | Staykovski |
| Stochastic Gradient Descent | Giachanou; Bounaama | RNN | Dias; Petrik; Bolonyai; Onose |
| Decision Trees | Saeed | CNN | Dias; Petrik; Polignano; Farber |
| Multinomial BayesNet | Saeed | BERT | Joo |
| Naive Bayes | Gamallo | Feedforward NN | Halvani; De la Peña |
| Adaboost | Bacciu | LSTM | Zhechev |

# Baselines

| | |
|---|---|
| MAJORITY | A statistical baseline that always predicts the majority class in the training set. In case of balanced classes, it predicts one of them |
| RANDOM | A baseline that randomly generates the predictions among the different classes |
| CHAR N-GRAMS | With values for n from 1 to 10, and selecting the 100, 200, 500, 1,000, 2,000, 5,000 and 10,000 most frequent ones |
| WORD N-GRAMS | With values for n from 1 to 10, and selecting the 100, 200, 500, 1,000, 2,000, 5,000 and 10,000 most frequent ones |
| W2V | Texts are represented with two word embedding models: Continuous Bag of Words (CBOW); and Skip-Grams |
| LDSE | This method represents documents on the basis of the probability distribution of occurrence of their words in the different classes. The key concept of LDSE is a weight, representing the probability of a term to belong to one of the different categories: human / bot, male / female. The distribution of weights for a given document should be closer to the weights of its corresponding category. LDSE takes advantage of the whole vocabulary |

# Global ranking

| Ranking | Team | Bots vs. Human | | Gender | | Average |
|---|---|---|---|---|---|---|
| | | EN | ES | EN | ES | |
| 1 | Pizarro | 0.9360 | **0.9333** | 0.8356 | **0.8172** | **0.8805** |
| 2 | Srinivasarao & Manu | 0.9371 | 0.9061 | 0.8398 | 0.7967 | 0.8699 |
| 3 | Bacciu et al. | 0.9432 | 0.9078 | 0.8417 | 0.7761 | 0.8672 |
| 4 | Jimenez-Villar et al. | 0.9114 | 0.9211 | 0.8212 | 0.8100 | 0.8659 |
| 5 | Fernquist | 0.9496 | 0.9061 | 0.8273 | 0.7667 | 0.8624 |
| 6 | Mahmood | 0.9121 | 0.9167 | 0.8163 | 0.7950 | 0.8600 |
| 7 | Ipsas & Popescu | 0.9345 | 0.8950 | 0.8265 | 0.7822 | 0.8596 |
| 8 | Vogel & Jiang | 0.9201 | 0.9056 | 0.8167 | 0.7756 | 0.8545 |
| 9 | Johansson & Isbister | **0.9595** | 0.8817 | 0.8379 | 0.7278 | 0.8517 |
| 10 | Goubin et al. | 0.9034 | 0.8678 | 0.8333 | 0.7917 | 0.8491 |
| 11 | Polignano & de Pinto | 0.9182 | 0.9156 | 0.7973 | 0.7417 | 0.8432 |
| 12 | Valencia et al. | 0.9061 | 0.8606 | **0.8432** | 0.7539 | 0.8410 |
| 13 | Kosmajac & Keselj | 0.9216 | 0.8956 | 0.7928 | 0.7494 | 0.8399 |
| 14 | Fagni & Tesconi | 0.9148 | 0.9144 | 0.7670 | 0.7589 | 0.8388 |
| | char nGrams | 0.9360 | 0.8972 | 0.7920 | 0.7289 | 0.8385 |
| 15 | Glocker | 0.9091 | 0.8767 | 0.8114 | 0.7467 | 0.8360 |
| | word nGrams | 0.9356 | 0.8833 | 0.7989 | 0.7244 | 0.8356 |
| 16 | Martinc et al. | 0.8939 | 0.8744 | 0.7989 | 0.7572 | 0.8311 |
| 17 | Sanchis & Velez | 0.9129 | 0.8756 | 0.8061 | 0.7233 | 0.8295 |
| 18 | Halvani & Marquardt | 0.9159 | 0.8239 | 0.8273 | 0.7378 | 0.8262 |
| 19 | Ashraf et al. | 0.9227 | 0.8839 | 0.7583 | 0.7261 | 0.8228 |
| 20 | Gishamer | 0.9352 | 0.7922 | 0.8402 | 0.7122 | 0.8200 |
| 21 | Petrik & Chuda | 0.9008 | 0.8689 | 0.7758 | 0.7250 | 0.8176 |
| 22 | Oliveira et al. | 0.9057 | 0.8767 | 0.7686 | 0.7150 | 0.8165 |
| | W2V | 0.9030 | 0.8444 | 0.7879 | 0.7156 | 0.8127 |
| 23 | De La Peña & Prieto | 0.9045 | 0.8578 | 0.7898 | 0.6967 | 0.8122 |
| 24 | López Santillán et al. | 0.8867 | 0.8544 | 0.7773 | 0.7100 | 0.8071 |
| | LDSE | 0.9054 | 0.8372 | 0.7800 | 0.6900 | 0.8032 |
| 25 | Bolonyai et al. | 0.9136 | 0.8389 | 0.7572 | 0.6956 | 0.8013 |

# Global ranking

| | | | | | | |
|---|---|---|---|---|---|---|
| 26 | Moryossef | 0.8909 | 0.8378 | 0.7871 | 0.6894 | 0.8013 |
| 27 | Zhechev | 0.8652 | 0.8706 | 0.7360 | 0.7178 | 0.7974 |
| 28 | Giachanou & Ghanem | 0.9057 | 0.8556 | 0.7731 | 0.6478 | 0.7956 |
| 29 | Espinosa et al. | 0.8413 | 0.7683 | 0.8413 | 0.7178 | 0.7922 |
| 30 | Rahgouy et al. | 0.8621 | 0.8378 | 0.7636 | 0.7022 | 0.7914 |
| 31 | Onose et al. | 0.8943 | 0.8483 | 0.7485 | 0.6711 | 0.7906 |
| 32 | Przybyla | 0.9155 | 0.8844 | 0.6898 | 0.6533 | 0.7858 |
| 33 | Puertas et al. | 0.8807 | 0.8061 | 0.7610 | 0.6944 | 0.7856 |
| 34 | Van Halteren | 0.8962 | 0.8283 | 0.7420 | 0.6728 | 0.7848 |
| 35 | Gamallo & Almatarneh | 0.8148 | 0.8767 | 0.7220 | 0.7056 | 0.7798 |
| 36 | Bryan & Philipp | 0.8689 | 0.7883 | 0.6455 | 0.6056 | 0.7271 |
| 37 | Dias & Paraboni | 0.8409 | 0.8211 | 0.5807 | 0.6467 | 0.7224 |
| 38 | Oliva & Masanet | 0.9114 | 0.9111 | 0.4462 | 0.4589 | 0.6819 |
| 39 | Hacohen-Kerner et al. | 0.4163 | 0.4744 | 0.7489 | 0.7378 | 0.5944 |
| 40 | Kloppenburg | 0.5830 | 0.5389 | 0.4678 | 0.4483 | 0.5095 |
| | MAJORITY | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |
| | RANDOM | 0.4905 | 0.4861 | 0.3716 | 0.3700 | 0.4296 |
| 41 | Bounaama & Amine | 0.5008 | 0.5050 | 0.2511 | 0.2567 | 0.3784 |
| 42 | Joo & Hwang | 0.9333 | - | 0.8360 | - | 0.4423 |
| 43 | Staykovski | 0.9186 | - | 0.8174 | - | 0.4340 |
| 44 | Cimino & Dell'Orletta | 0.9083 | - | 0.7898 | - | 0.4245 |
| 45 | Ikae et al. | 0.9125 | - | 0.7371 | - | 0.4124 |
| 46 | Jeanneau | 0.8924 | - | 0.7451 | - | 0.4094 |
| 47 | Zhang | 0.8977 | - | 0.7197 | - | 0.4044 |
| 48 | Fahim et al. | 0.8629 | - | 0.6837 | - | 0.3867 |
| 49 | Saborit | - | 0.8100 | - | 0.6567 | 0.3667 |
| 50 | Saeed & Shirazi | 0.7951 | - | 0.5655 | - | 0.3402 |
| 51 | Radarapu | 0.7242 | - | 0.4951 | - | 0.3048 |
| 52 | Bennani-Smires | 0.9159 | - | - | - | 0.2290 |
| 53 | Gupta | 0.5007 | - | 0.4044 | - | 0.2263 |
| 54 | Qurdina | 0.9034 | - | - | - | 0.2259 |
| 55 | Aroyehun | 0.5000 | - | - | - | 0.1250 |

# Best results

**Johansson**
- Stylistic features
- Random Forest

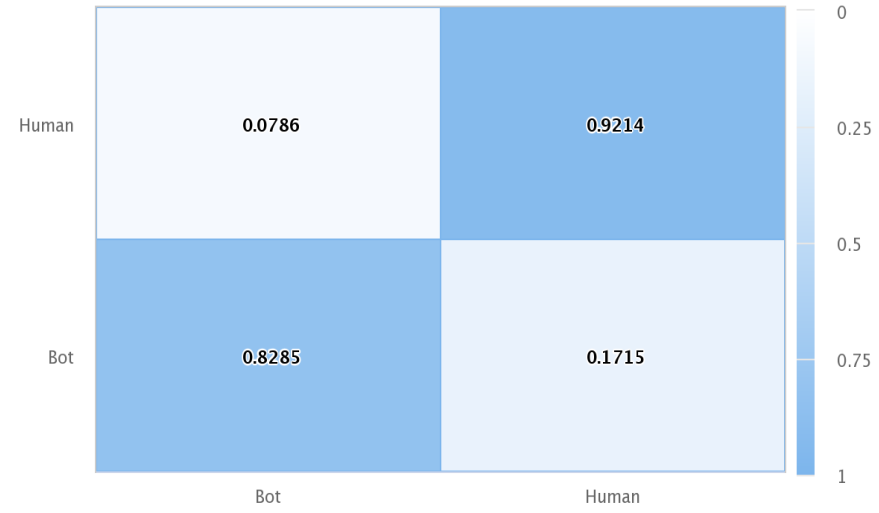
**Valencia**
- n-grams
- Logistic Regression

| Language | Bots vs. Human | Gender |
|---|---|---|
| English | 0.9595 | 0.8417 |
| Spanish | 0.9333 | 0.8172 |

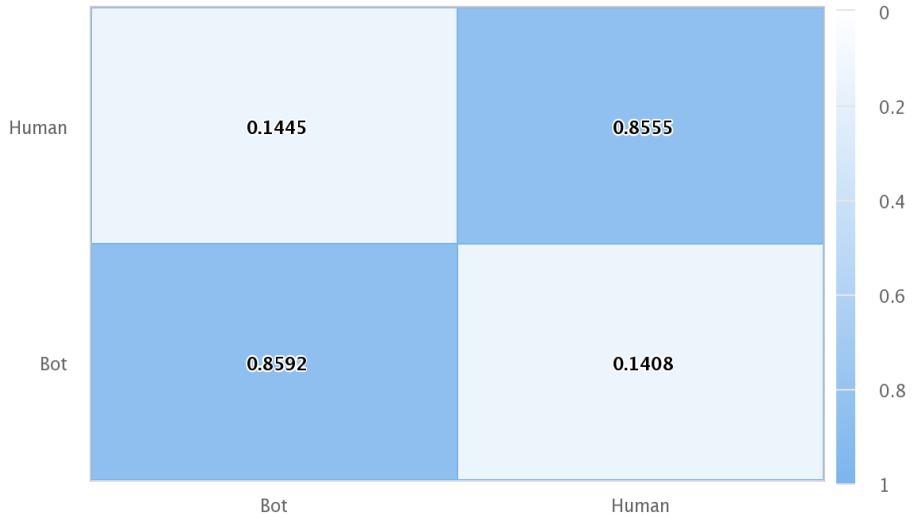
**Pizarro**
- n-grams
- SVM

# Confusion matrices: bots vs. humans

English

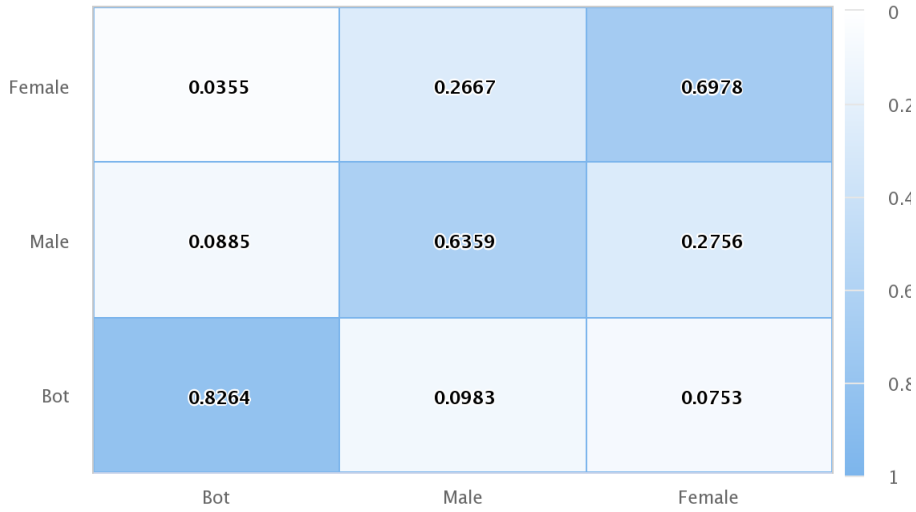



Spanish

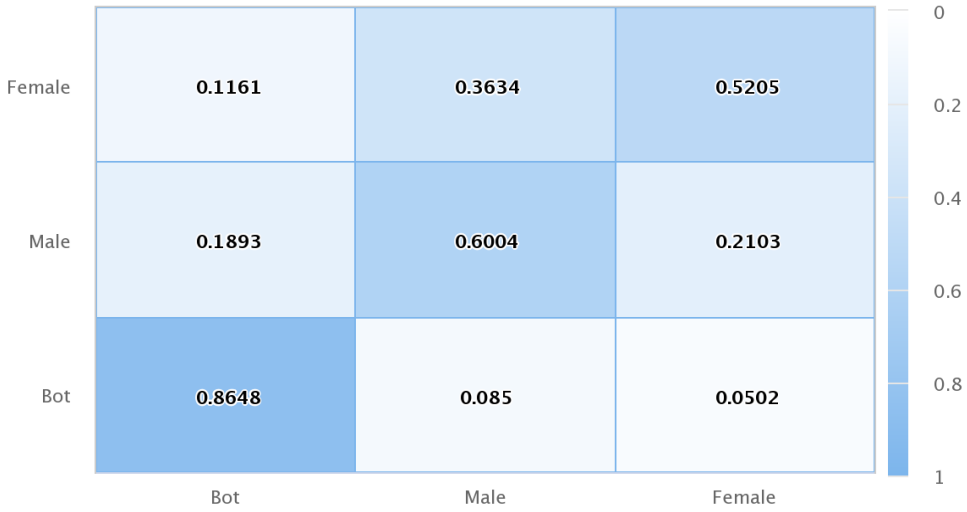


# Confusion matrices: gender

English



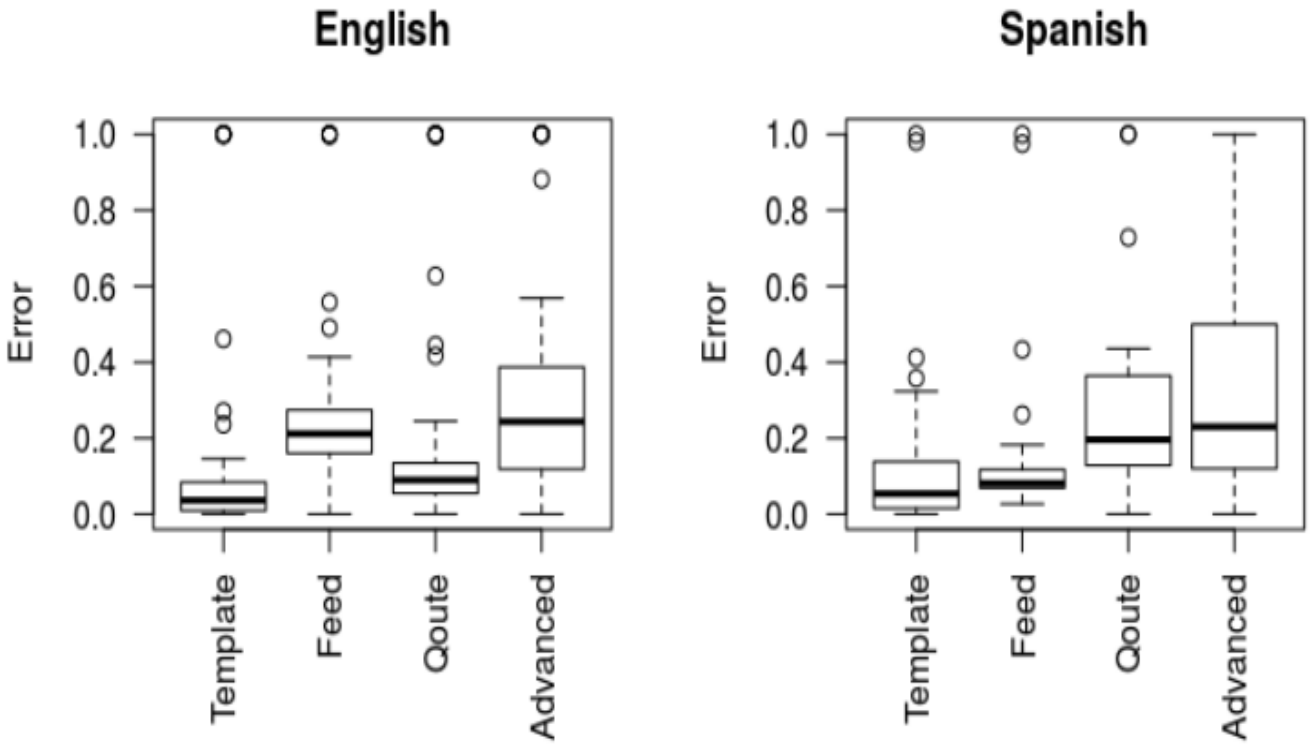

Spanish



# Test dataset: % types of bots

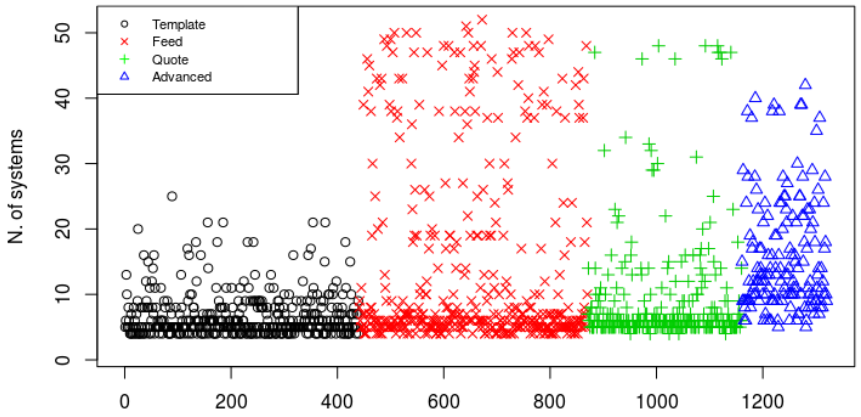
| TYPE | ENGLISH | SPANISH |
|---|---|---|
| **TEMPLATE** | 33.30% | 25.56% |
| **FEED** | 32.58% | 50% |
| **QUOTE** | 21.97% | 15.56% |
| **ADVANCED** | **12.12%** | **8.88%** |

# Errors per bot type

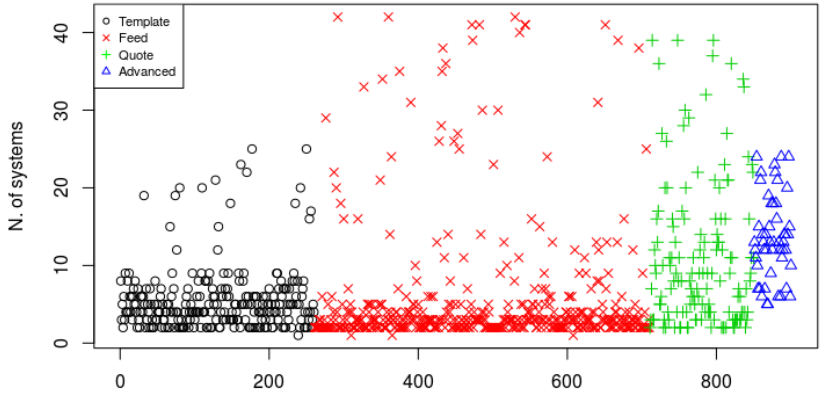


# Errors per bot type

ENGLISH



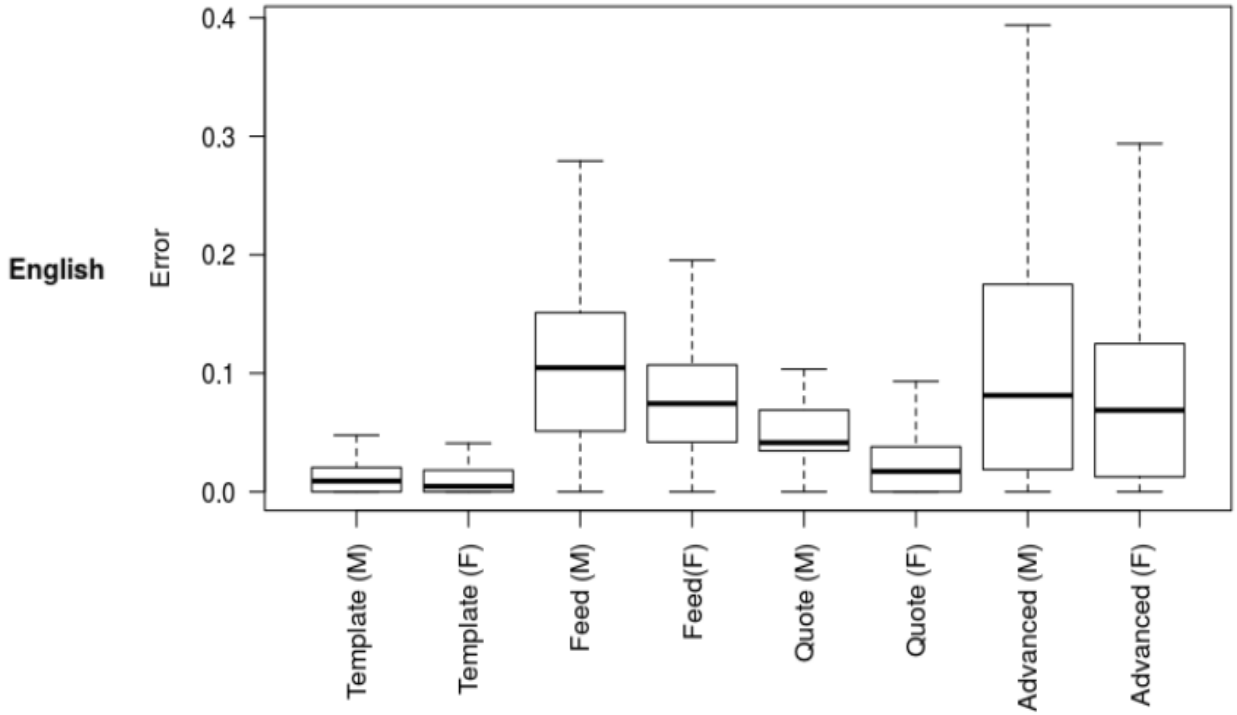

SPANISH



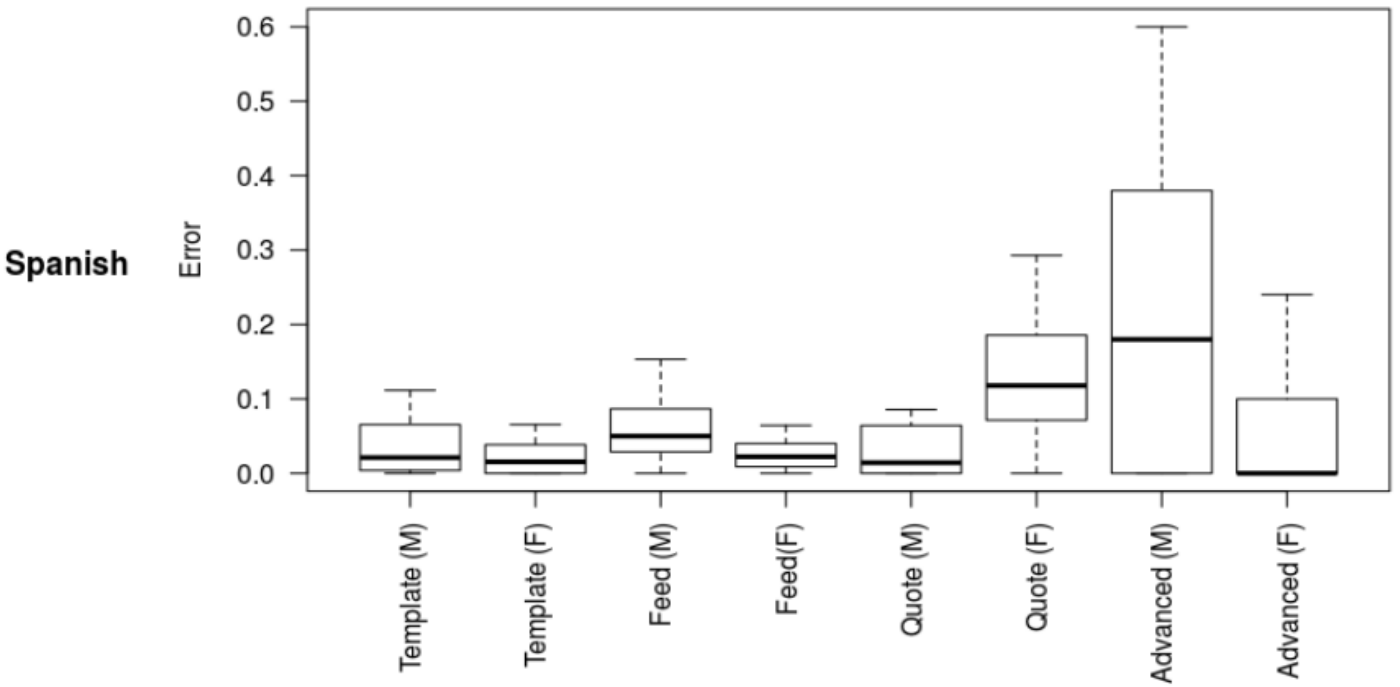
# Errors per bot type

| Author Id. | Twitter Account | Type | N. Systems |
|---|---|---|---|
| caf6d82d5dca1598beb5bfac0aea4161 | @NasaTimeMachine | template | 21 / 53 |
| *@wylejw You must be cool, I'll follow you!* | | | |
| 4c27d3c7a10964f574849b6be1df872d | @rarehero | feed | 52 / 53 |
| *Get a doll, drape fabric and spray the hell out of it with Fabric Quick Stiffening Spray ... https://t.co/C9Ub6xXZWI via @duckduckgo* | | | |
| 8d08e3a0e1fea2f965fd7eb36f3b0b07 | @MessiQuote | quote | 48 / 53 |
| *.@PedroPintoUEFA: "Messi is unstoppable and we should feel privileged to be watching a player who may be the best of all time." https://t.co/TmCR6qCzO2* | | | |
| 6a6766790e1f5f67813afd7c0aa1e60d | @markov_chains | advanced | 42 / 53 |
| *I have transferred to the local library go you! Just be Crazy John's prepaid sim card.* | | | |

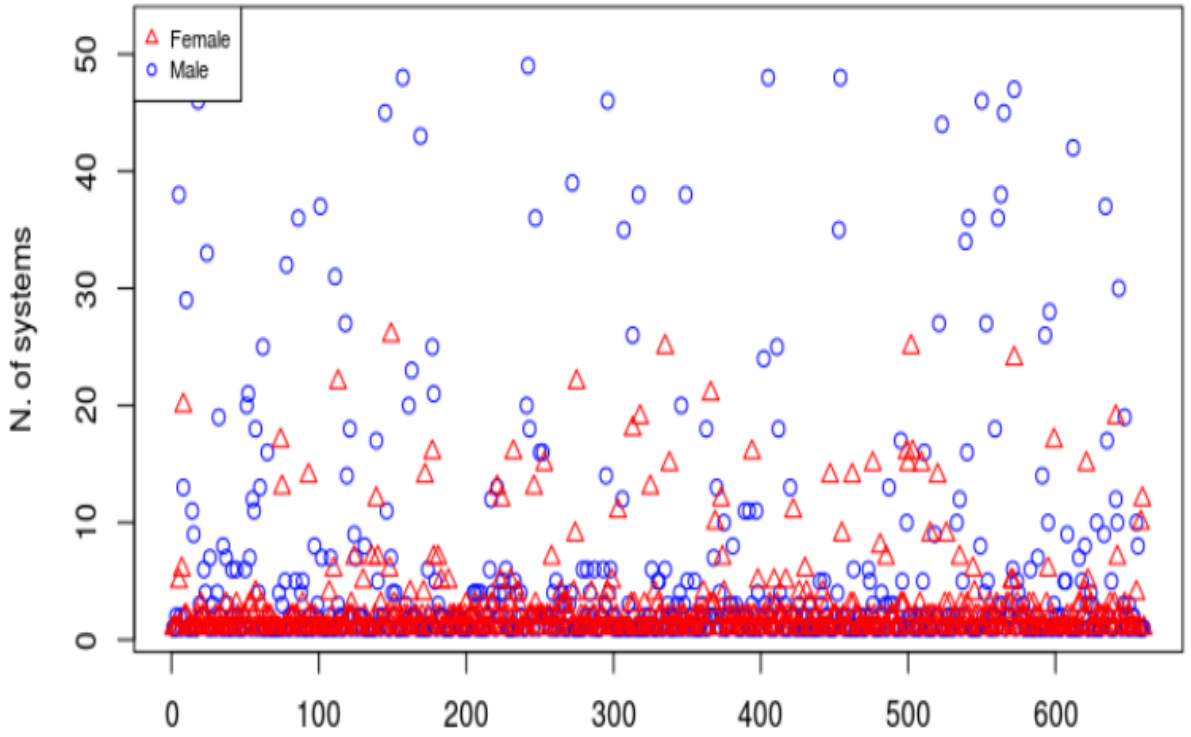

# Bot to human per gender errors



# Bot to human per gender errors



# Human to bot errors



# Human to bot errors

| Author Id. | Twitter Account | Gender | N. Systems |
|---|---|---|---|
| 63e4206bde634213b3a37343cf76e900 | @Ask_KFitz | male | 49 / 53 |
| *#Electric Imp Smart Refrigerator https://t.co/qigh5Womd7 https://t.co/JNVsRKvRQ8* | | | |
| b11ffeeed0b38eb85e4e288f5c74f704 | @iqbalmustansar | male | 45 / 53 |
| *Trend - What's Dominating Digital Marketing Right Now? - https://t.co/dWp7ovqzCM* | | | |
| ba0850ae38408f1db832707f1e0258fd | @CharBar_tweets | female | 26 / 53 |
| *Hollywood boll #bowling #legs #Sundayfunday https://t.co/cLq9ZlNM38* | | | |
| d64be10ecfbbb81d0c6e5b3115c335a5 | @RheaRoryJames | female | 25 / 53 |
| *RT @realDonaldTrump: Employment is up, Taxes are DOWN. Enjoy!* | | | |


https://botometer.iuni.iu.edu

# Human to bot errors


| Author Id. | Twitter Account | Gender | N. Systems |
|---|---|---|---|
| a22edd53bb04de0c06a52df897b13dd0 | @carlosguadian | male | 39 / 42 |
| *Tres días para analizar el presente y futuro de la Administración pública: lo que trae el Congreso NovaGob 2018 - NovaGob 2018 https://t.co/Ofc4cDTeym #novagob2018* | | | |
| cf520c8e810a6a9bae9171d6f23c29be | @kicorangel | male | 35 / 42 |
| *Google prepara una versión de pago para Youtube http://t.co/UvZdao68wc* | | | |
| 8e4340e95667c8add31f427a09dd3840 | @EmaMArredondoM | female | 30 / 42 |
| *@andrespino007 ¿Se ha preguntado cómo alguien llega a ser científico? Pequeña muestra chilena: https://t.co/fLjJsV0I0J* | | | |
| 6730bdf9686769c4a8a79d2f766a7f67 | @Annie$_{H}go$ | female | 24 / 42 |
| *Wow!! Nuevamente rebasamos expectivas... https://t.co/PJ8bHA1SrG* | | | |

# Conclusions

- Several approaches to tackle the task:
  - Best approach: n-grams + SVM
- Best results in **bots vs. human**:
  - **Over 84% on average** (EN 86.15%; ES 84.08%)
  - English (95.95%): Johansson - Stylistic features + Random Forest
  - Spanish (93.33%): Pizarro - n-grams + SVM
- Error analysis:
  - **Highest confusion from bots to humans (17.15% vs. 7.86% EN; 14.45% vs. 14.08% ES)**
    - ...**mainly towards males** (9.83% vs. 7.53% EN; 8.50% vs. 5.02% ES)
    - ...**males more confused with bots** (8.85% vs. 3.55% EN; 18.93% vs. 11.61% ES)
  - **Error per bot type**:
    - **Advanced bots: 30.11% EN; 32.38% ES**
    - EN: quote (12.64%); template (17.94%); feed (27.89%)
    - ES: quote (26.51%); template (13.20%); feed (14.28%)
    - **Mainly towards males**, except quote bots in ES (6.75% vs. 15.29% towards males)

# Conclusions

Looking at the results, we can conclude:

- It is feasible to automatically identify bots in Twitter with high precision
  - ...even when **only textual features** are used.

- There are specific cases where the task is difficult due to:
  - ...**the language used by the bots e.g., advanced bots ***
  - **...the way the humans use the platform (e.g., to share news)**

In both cases, although the precision is high, a major effort needs to be made to take into account **false positives**

***previous to GPT…**

# Industry

Organisation

autoritas
nuevas ideas, nuevas soluciones

Sponsors

symanto GROUP

THE LOGIC VALUE

IBM PartnerWorld

meaning cloud

Participants

DATYS
SOLUCIONES TECNOLÓGICAS

FUJI Xerox

ANCHORMEN

KNOW Center

OTS
OPTICAL TECH & SUPPORT

ISG

Bitdefender

# Not only industry

Organisation

autoritas
nuevas ideas, nuevas soluciones

Sponsors

symanto GROUP

THE LOGIC VALUE

IBM PartnerWorld

meaning cloud

Participants

DATYS SOLUCIONES TECNOLÓGICAS

FUJI Xerox

FOI

ANCHORMEN

Swedish Defence Research Agency

KNOW Center

OTS OPTICAL TECH & SUPPORT

iSG

Fraunhofer SIT

Bitdefender

# 2020: Profiling FAKE NEWS spreaders on Twitter


Rangel F., Giachanou A., Ghanem B., Rosso P. (2020) Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter. In: CLEF 2020 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings.CEUR-WS.org, vol. 2696

# Dataset

- **500 authors** in English and in Spanish
- 250 authors fake news spreaders + 250 not
- Each **author feed** is composed by **100 tweets** (the tweets are concatenated)

| Language | Training | Test | Total |
|---|---|---|---|
| English | 300 | 200 | 500 |
| Spanish | 300 | 200 | 500 |

| Preprocessing | Features | Methods |
| --- | --- | --- |
| Twitter elements (RT, VIA, FAV) | Stylistic features:<br>- Number of occurrences<br>- Verbs, adjs, pronouns<br>- Number of hashtags, mentions, URLs…<br>- Capital vs. lower letters<br>- Punctuation marks<br>- … | SVM |
| Emojis and other non-alphanumeric chars | | Logistic regression |
| Lemmatisation | | Random Forest |
| Tokenisation | N-gram models | Ensembles |
| Punctuation signs | | Multilayer Perceptron |
| Numbers | | NN with Dense Layer |
| Lowercase | Emotional and personality features | Fully-Connected NN |
| Stopwords | | CNN |
| Character flooding | Embeddings | LSTM |
| Infrequent terms | | bi-LSTM |
| Short texts | …BERT | |

# Results

**Buda and Bolonyai**
- n-grams
- stylistic features
- Logistic Regression ensemble

**Pizarro**
- word and char n-grams
- SVM

| English | Spanish |
|---|---|
| Buda and Bolonyai [9] (0.750) | Pizarro [45] (0.820) |

- **66** teams
- Best performance (accuracy)

# 2021:Profiling HATE speech speadeRS on Twitter


Rangel F., De La Peña G., Chulvi B., Fersini E., Rosso P. (2021) Profiling Hate Speech Spreaders on Twitter Task at PAN 2021. In: CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, vol. 2936, pp. 1772-1789

# Dataset

- **300 authors** in English and in Spanish
- 150 authors haters + 150 not
- Each **author feed** is composed by **200 tweets** (the tweets are concatenated)

| | (EN) English | | | (ES) Spanish | | |
|---|---|---|---|---|---|---|
| | **Keen to spread hate speech** | **Not keen to spread hate speech** | **Tota** | **Keen to spread hate speech** | **Not keen to spread hate speech** | **Total** |
| **Training** | 100 | 100 | 200 | 100 | 100 | 200 |
| **Test** | 50 | 50 | 100 | 50 | 50 | 100 |
| **Total** | 150 | 150 | 300 | 150 | 150 | 300 |

# Features

Stylistic features:
- Number of occurrences
- Verbs, adjs, pronouns
- Number of hashtags, mentions, URLs...
- Capital vs. lower letters
- Punctuation marks
- ...

N-gram models

Emotional and personality features

Specialised lexicons (HS)

Embeddings

...Lexical+statistical+syntactical+phonetical

...Semantic-emotion-based

Transformers

...BERT

...SBERT

...ALBERT

...RoBERTa

...BERTTweet

...BETO

LDSE at char level

Fine-tuned transformer to modify the Impostor Method

Topics aggregation combined with ELMo to represent the user

CNN to extract features from external data

# Methods

SVM

| |
|---|
| Logistic regression |
| Random Forest |
| Ensembles |
| Adaboost, Ridge, Naive Bayes, KNN, XGBoost, AutoML, ... |
| Custom architecture |
| RNN |
| CNN |
| LSTM |
| bi-LSTM |

# Results

- **66** teams
- Best performance (accuracy)

**Dukic and Sovic**
- BERT
- Logistic Regression

**Siino et al.**
- 100-dim word-embedding
- CNN

| English | | Spanish | |
|---|---|---|---|
| Dukić and Sović | (0.75) | Siino et al. | (0.85) |

# 2022: Profiling IRONY and STEREOTYPE Spreaders on Twitter

- Author profiling perspective: Profiling **irony and stereotype** spreaders
- Usage of irony and stereotypes for (often implicitly) conveying hurtfulness, although... not always
- Subtask on Profiling **stereotype stance** of ironic authors

How wonderful a **Jew** actually said something bad about Israel. I'm sooo impressed.

If Australia doesn't "DEPORT" 100K **Muslims** a year, what do you propose? Concentration camps?

Didn't you know if they rub against you that you can become **gay**?! Talk about sharing a foxhole!!!

Ortega-Bueno R., Chulvi B., Rangel F., Rosso P., Fersini E. (2022) Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO). Overview for PAN at CLEF 2022. In: CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, vol. 3180, pp. 2314-2343