

IJCAI 2016

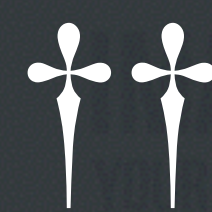
Language Sense on Computer

Radoslaw Komuda[†], Michal Ptaszynski^{††}, Rafal Rzepka[†] and Kenji Araki[†]

Recognizing and Converting Cockney Rhyming Slang for Cyberbullying and Crime Detection



Hokkaido University
Language Media Lab.



Kitami Institute of Technology
Department of Computer Science

The Outline

Introduction

What is Cockney Rhyming Slang?

Corpus and lexicon creation

Experiment results

Conclusions and future work



HOKKAIDO
UNIVERSITY

ニイハクシツ
HOKKAIDO

Why is this research relevant?

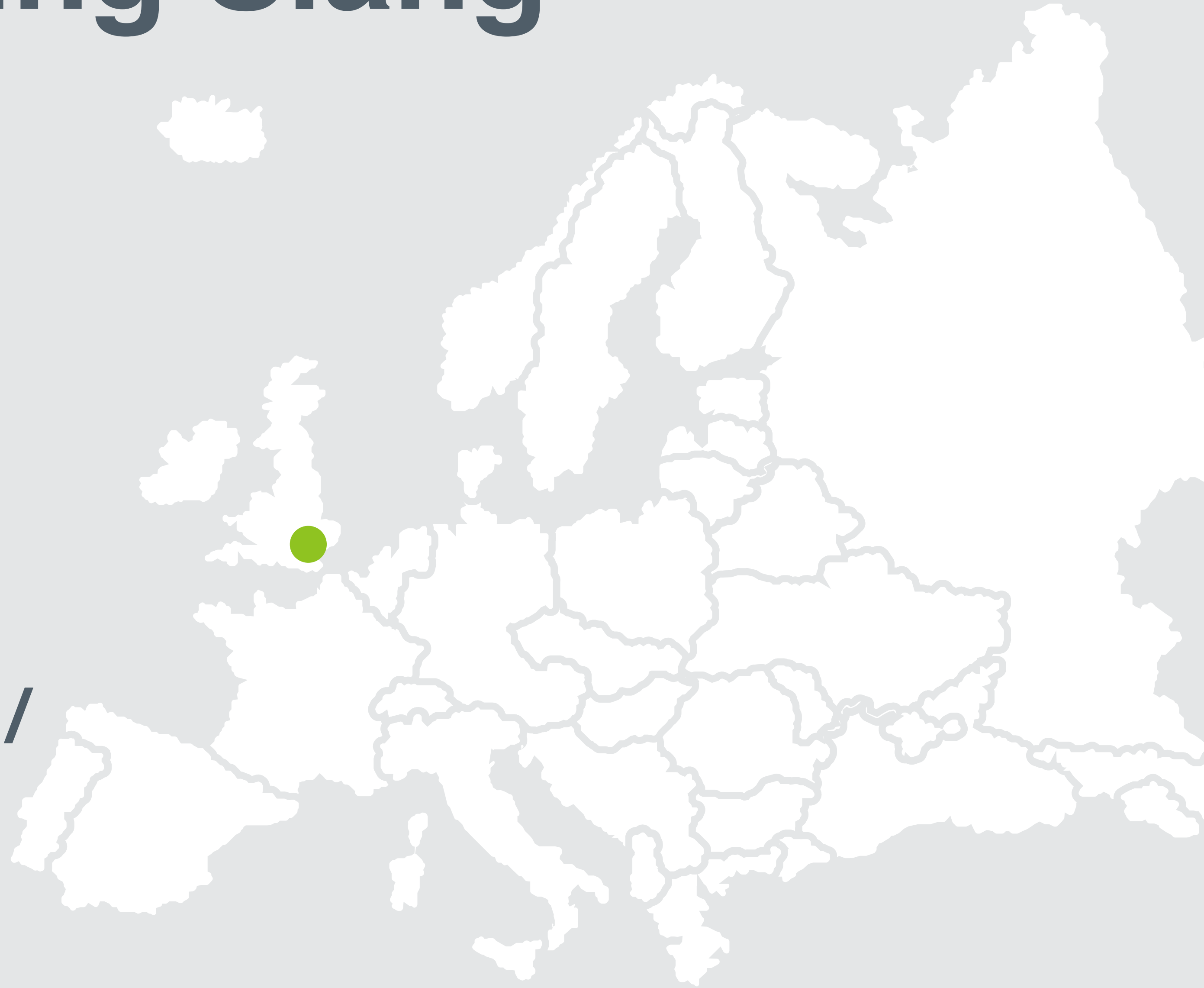


Entry:	Explanation:
dirl	“die in real life”; used to upset or encourage to commit suicide
gnoc	“get naked on camera”; used in grooming and sexting
np4np	“naked picture for naked picture”; same as above
IHML	“I hate my life.”
oreo	racist slang term for a black person who is ‘trying to be white’
mos	“mom over shoulder”
miw	“mom is watching”
pos / pob	“parents over shoulder”
taw	“teachers are watching”
kpc	“keeping parents clueless”

Cockney Rhyming Slang

a native of East London,

**linguistic game or a cant
(cryptolect) among traders /
thieves / immigrants
of the 19th century London**



replacing a common word with a phrase
of two or three words,
the last of which rhymes with the original word

He was the ~~win~~icken Dinner.

5 + 5 = ~~C~~ock and Hen.

What's the ~~price~~ price of leezza Rice?

My team has a very important ~~Central meeting~~ meeting this week.

I need to buy a new ~~Barack~~ Barack Obama.

The word "alone" could be replaced with:

- "Jack Jones" (singer),
- "Pat Malone" (MLB pitcher),
- "Todd Sloan" (thoroughbred horse racing jockey)
- "Toblerone" (Swiss chocolate bar brand).

All
for one
and
one for all

Conducting the Experiment

01 **Lexicon and corpus creation**

02 **CRS detection and the *hemiteleia* problem**

03 **CRS detection results**

04 **CRS conversion results**



Lexicon and the corpus

MANDY DINGLE

MANDY DINGLE IS COCKNEY SLANG FOR SINGLE.

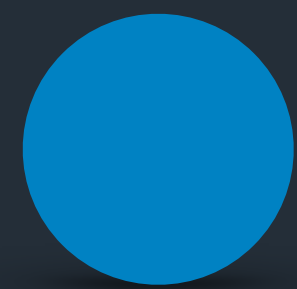
"Is that bird still mandy dingle?"

Lexicon and the corpus

"Auntie Ella": "Umbrella",
"Auntie Nell": "Bell",
"Auntie Nellie": "Belly",
"Auntie Nelly": "Telly",
"Austin Power": "Shower",
"Ave Maria": "Fire",
"Ayers Rock": "Cock",
"Aylsebury Duck": "Fuck",
"Ayrton Senna": "10 pounds (tenner)",
"Baa Lamb": "Tram",
"Babbling Brook": "Crook",
"Babe Ruth": "Truth",
"Bacardi Breezer": "Freezer",
"Bacardi Breezer": "Geezer",
"Back and Front": "Cunt"

Jus goin dahn the Bath for a couple o
The place is full of Bathroom Taps.
I've got a nasty case of the Bathroom
Let's go dahn the Battle Cruiser.
Put some Donald on me Baydon.
Hit it off the Beans on Toast.
My penalty hit the beans.
Look at that stupid Beating Heart.
I'm right Lee Marvin, think I'll nip out
Becks & Posh.
I can't tonight, I've got no Beck's.
Last night she bee'd through the plea
pain.
Can you lend us an Ayrton Senna ma
out of Bees and Honey

Lexicon categorization statistics



NEUTRAL



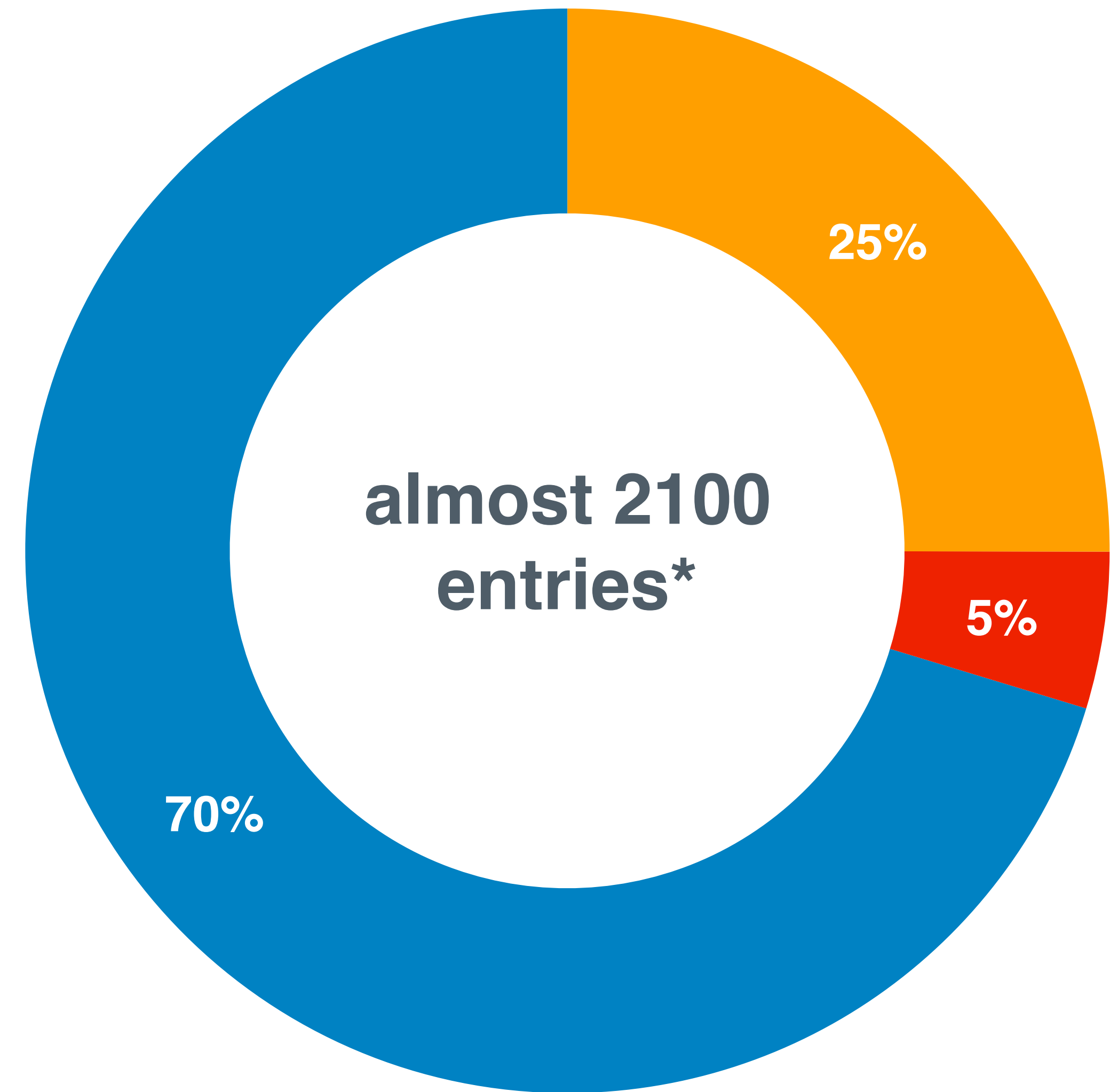
CYBERBULLYING

cursing, bad words, insults and words that can be use for sexual harassment and hate speech



CRIME

guns and drugs, pedophilia, prostitution, stealing



***) SOURCE:** Cockney Rhyming Slang: London's Famous Secret Language website: <http://cockneyrhymingslang.co.uk/>

corpus in numbers:

BASELINE:
519 cases (55.43%)

HEMITELEIA:
382 cases (40.64%)

FINAL CRS RECOGN.*:
95.85%

940 actual
CRS samples


949 sentences




***) 39 cases not recognized:** misspelling and omitting “h” (74%), irregular plural forms (13%), wrong entries (13%)

	OVERALL	NEUTRAL	CYBER BULLYING	CRIME
detection recall	0.851 (92 / 108)	0.873 (61 / 68)	0.750 (24 / 32)	0.875 (7 / 8)
F1-measure	0.920	0.932	0.857	0.933

	OVERALL	NEUTRAL	CYBER BULLYING	CRIME
conversion recall	0.398 (42 / 108)	0.426 (29 / 68)	0.313 (10 / 32)	0.375 (3 / 8)
F1-measure	0.555	0.597	0.476	0.500

 **CRS convert:**
F1 = 0.555

 **CRS recog:**
F1 = 0.920

 **CRS convert:**
P@n = 0.977

 **CRS recog:**
P@n = 1.000



System if usefull...

... but working with slang is a very input-sensitive process.

Main FW tasks:

- extend with other British slangs,
- develop the idea,



Thank You

■ For Your Attention ■

	OVERALL	NEUTRAL	CYBER BULLYING	CRIME
PRECISION	1.000 (92 / 92)	1.000 (61 / 61)	1.000 (24 / 24)	1.000 (7 / 7)
RECALL	0.851 (92 / 108)	0.873 (61 / 68)	0.750 (24 / 32)	0.875 (7 / 8)
F1-measure	0.920	0.932	0.857	0.933

	OVERALL	NEUTRAL	CYBER BULLYING	CRIME
PRECISION	0.977 (42 / 43)	1.000 (29 / 29)	1.000 (10 / 10)	0.750 (3 / 4)
RECALL	0.398 (42 / 108)	0.426 (29 / 68)	0.313 (10 / 32)	0.500 (3 / 8)
F1-measure	0.555	0.597	0.476	0.500