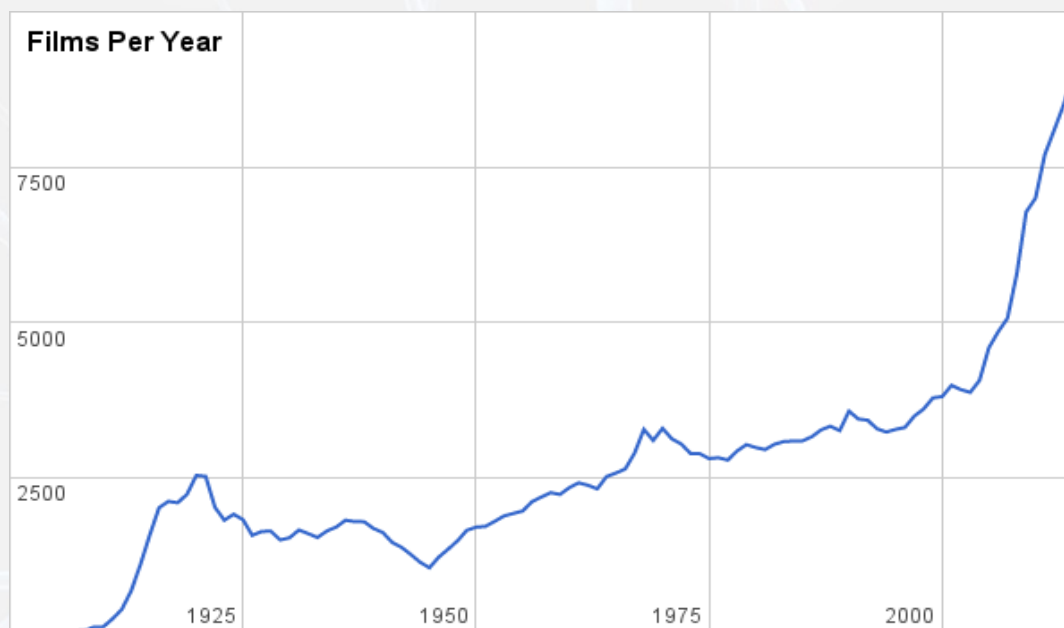# *Whose line is it anyway?*
Automatic Multilingual Emotion Annotation of Movie Dialogues

Wojciech Stokowiec

National Information Processing Institute, Natural Language Processing Laboratory, Warsaw

Motivation

- There are to many movies.
- Yet Another Approach to Movie Recommendation
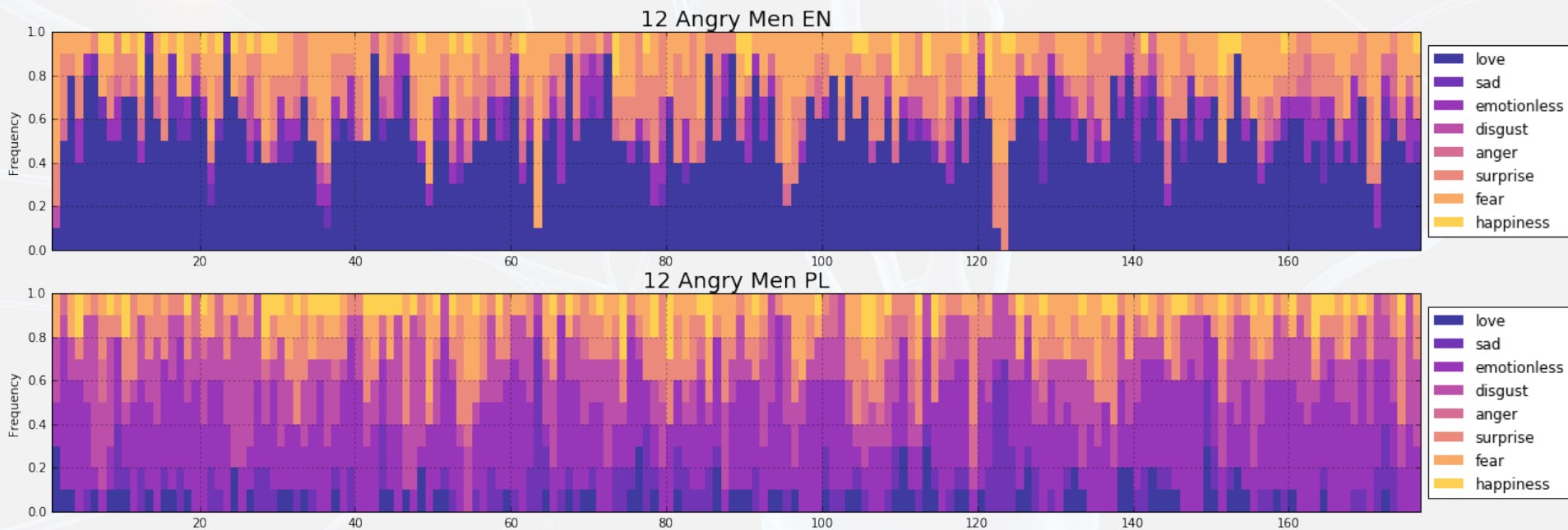- For fun mostly :-)

Problem

- For each line of dialogue in a given movie assign one emotion

- Find frequencies of emotions in iconic movies

- To make it more interesting, do it for two different languages

- Check if results are stable across languages

Method – Emotion Vectors

- Skip-gram model with Negative Sampling trained on English and Polish wikipedia for word vectors.

- BabelNet used to create „Sense Trees": trees with emotion concepts (love, anger, etc.) as roots and their hyponyms as children

- Each node has been assigned a vector representation based on sense lemma

- Calculate weighted average of word vectors with weights inversely proportional to the distance from the root.

Method – Sentence Vectors

- Remove stop words

- Average word vectors for each sentence
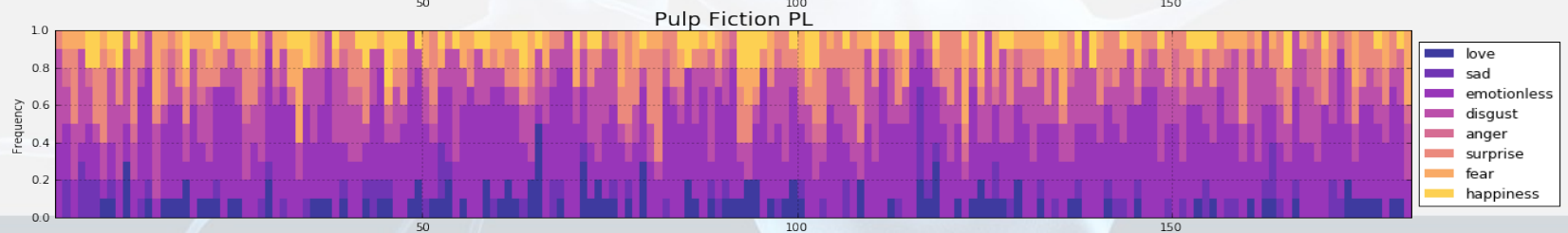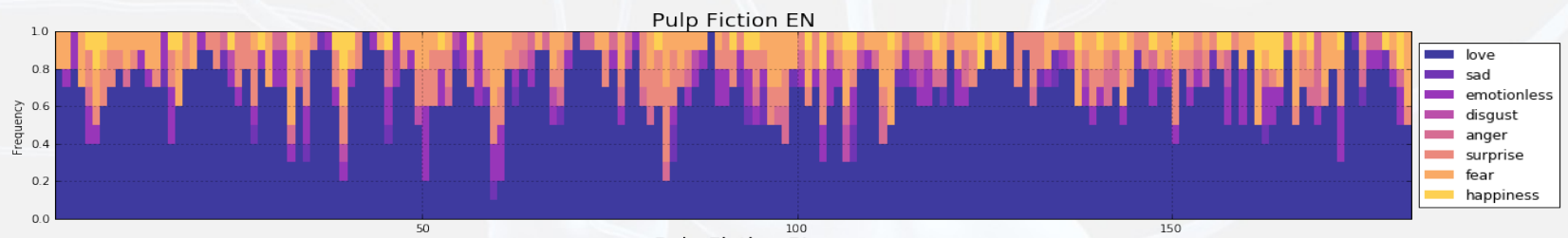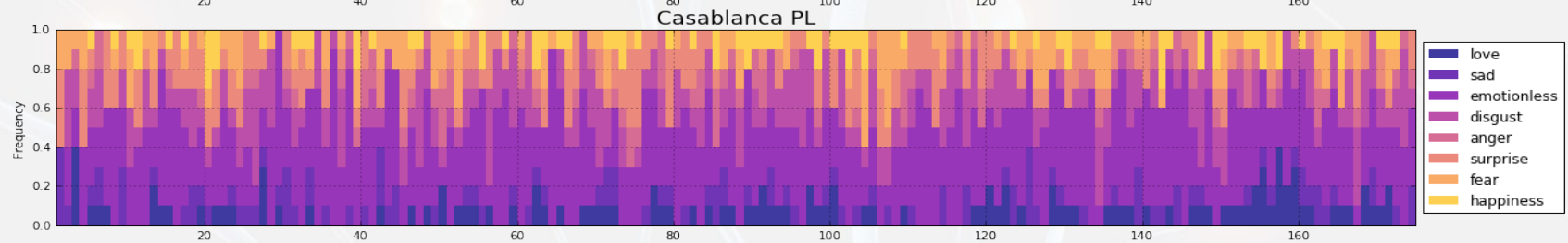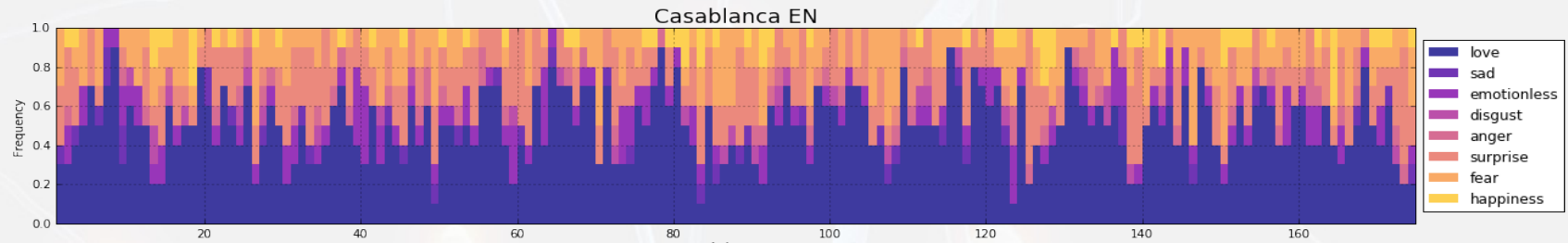
- Hope for the best :-)

Sample Results

	Casablanca		Pulp Fiction	
	EN	**PL**	**EN**	**PL**
love	851	116	1210	112
sad	42	113	33	78
emotionless	133	694	121	722
disgust	33	302	17	407
anger	56	94	61	58
surprise	320	94	157	234
fear	241	128	168	123
happiness	68	87	50	83
sum	**1744**	**1744**	**1817**	**1817**

Table 1: Comparison of emotions

Sample Results

Casablanca EN

Casablanca PL

Pulp Fiction EN

Pulp Fiction PL

Future Work

- Using Forests insted of Trees seems like a low-hanging fruit.

- Maybe word vectors and cosine distance as similarity measure are not such a great idea.

- The length of the shortest path that connects the concepts in a Semantic Tree?

Thank you!

Ośrodek Przetwarzania Informacji
Państwowy Instytut Badawczy

al. Niepodległości 188 b
00-608 Warszawa

tel.: +48 22 570 14 00
e-mail: opi@opi.org.pl