Emotion Annotations
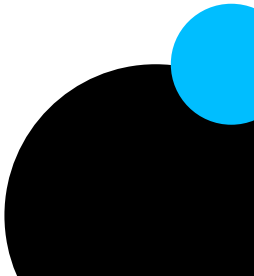## Understanding Annotators' Disagreements

Enrica Troiano
Institute for Natural Language Processing

enrica.troiano@ims.uni-stuttgart.de
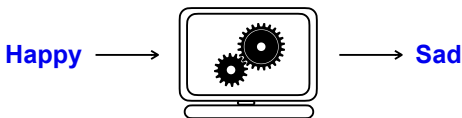
March 11, 2021

About Myself

PhD (ongoing)

- IMS, University of Stuttgart, Germany
- Computational Emotion Analysis (EA)
  Sup: Roman Klinger, Sebastian Padó
- Emotion Style Transfer
  Goal: Paraphrase a text and change its emotion

Happy ⟶ [laptop with gears] ⟶ **Sad**

About Myself

PhD (ongoing)

- IMS, University of Stuttgart, Germany
- Computational Emotion Analysis (EA)
  Sup: Roman Klinger, Sebastian Padó
- Emotion Style Transfer
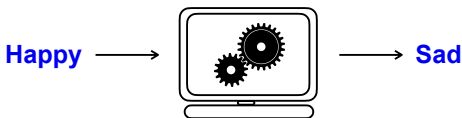 Goal: Paraphrase a text and change its emotion

Happy ⟶ ⟶ **Sad**

...how about data?

An Emotion is...

✗ **a feeling** – conscious perception

✗ **a mood** – vague, weak

An Emotion is...

✗ **a feeling** – conscious perception

✗ **a mood** – vague, weak

✔ **based on goals and beliefs**

✔ **happy *because*, angry at, sad *about***

An Emotion is...

an episode of interrelated, synchronized **changes** [...] in response to the **evaluation** of an external or internal **stimulus event** as relevant to major concerns of the organism.

Scherer, 2005

Changes in Response to Evaluations of **Events**

Emotion Ingredients

Event – seeing a crocodile

Changes in response to Evaluations of Events

Emotion Ingredients

Event – seeing a crocodile

Bodily symptoms – trembling

Action tendencies – run

Expression – eyes widening

Feeling – very bad

Changes in response to Evaluations of Events

Emotion Ingredients

Event – seeing a crocodile

Bodily symptoms – trembling

Action tendencies – run

Expression – eyes widening

Feeling – very bad

Cognitive Appraisal –
is this unexpected? can control it?

Changes in response to Evaluations of Events

Emotion Ingredients

...realized in text

Event – seeing a crocodile

A crocodile appeared.

Bodily symptoms – trembling

I couldn't stop shaking.

Action tendencies – run

RUN!!

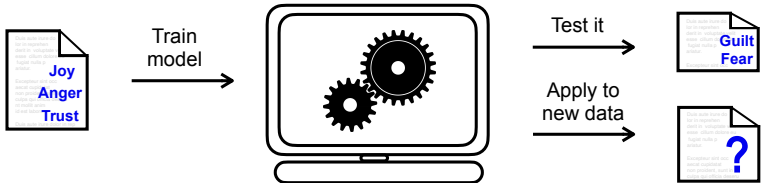**Expression** – eyes widening

My eyes went wide open.

Feeling – very bad

That didn't feel good.

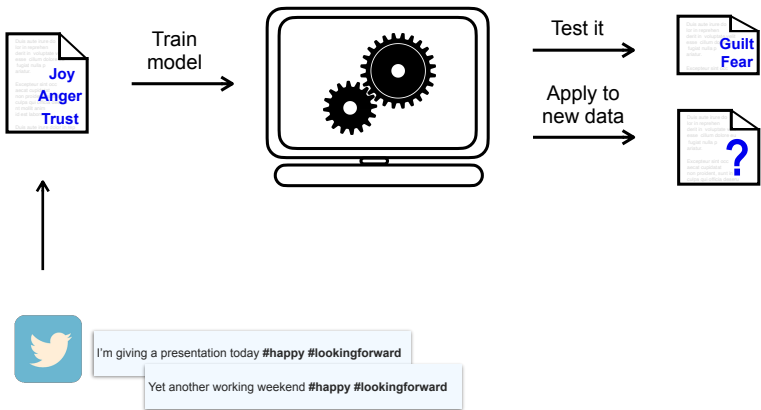**Cognitive Appraisal** –
is this unexpected? can control it?

I had no way out.

Typical EA Pipeline

Typical EA Pipeline – Text Collection

Introduction
o

Emotion Analysis
000●000

Confidence and Intensity
00000000

Appraisals
000000000

Conclusions
00

Typical EA Pipeline – Text Collection

Annotating Emotion Phenomena

A crocodile appeared. → **?**

Annotating Emotion Phenomena

A crocodile appeared. → ? → *fear*

Plutchik, 1990

Discrete Emotions
fear, surprise, joy, sadness, shame, anger

Ekman, 1992

Joy   Anger   Disgust

Fear   Sadness   Surprise

University of Stuttgart   Enrica Troiano

Annotating Emotion Phenomena

A crocodile appeared. → → **?**   *high*

Discrete Emotions
fear, surprise, joy, sadness, shame, anger

Intensity
emotion magnitude   *[0-1] low, medium, high*

Annotating Discrete Emotions

Plutchik, 1990

Reaching high inter-annotator agreement is difficult.

Explicit	*I'm scared of crocodiles*	– fear
Implicit	*A crocodile appeared*	– fear and surprise
		– neutral (text)
		– fear (writer)
		– joy (reader)

Past experience, cultural background, personal goals are involved.

Today's Talk

Challenge: Obtaining consistent emotion annotations

We will: Understand disagreements with...

1. confidence and intensity ratings (Troiano, Padó and Klinger, 2021);

2. annotation of emotion components, i.e. appraisals (Hofmann, Troiano and Klinger, 2021).

Emotion Ratings:
How Intensity, Annotation Confidence and Agreements are Entangled

Enrica Troiano, Sebastian Padó and **Roman Klinger**

Institut für Maschinelle Sprachverarbeitung

University of Stuttgart, Germany

{firstname.lastname}@ims.uni-stuttgart.de

Starting Point

Goal: Understand disagreements wrt measurable factors

Starting Point

Goal: Understand disagreements wrt measurable factors
Building Blocks:

- Task Subjectivity

Starting Point

Goal: Understand disagreements wrt measurable factors
Building Blocks:

- Task Subjectivity
- Raters' Confidence

Starting Point

Goal: Understand disagreements wrt measurable factors
Building Blocks:

- Task Subjectivity
- Raters' Confidence
- Emotion Intensity
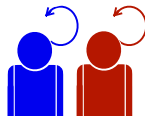
Starting Point

Goal: Understand disagreements wrt measurable factors
Building Blocks:

- Task Subjectivity
- Raters' Confidence
- Emotion Intensity
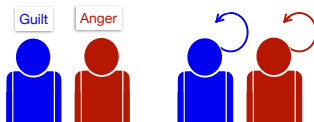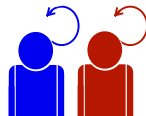

↑disag. ↓conf.

↑int. ↑conf.

Starting Point

Goal: Understand disagreements wrt measurable factors
Building Blocks:

- Task Subjectivity
- Raters' Confidence
- Emotion Intensity
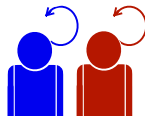
| **RQ1: Are disagreements related to confidence?** | ↑disag. ↓conf. |

| **RQ2: Are confidence and intensity related?** | ↑int. ↑conf. |

Annotation Setup

Data 700 sentences from COCA

Participants 3 female Master students, previous experience

Guidelines   (EMO) Is the text emotional?   *emotion*/*neutral*

　　　　　　　 (CONF) How confident are you about *emotion*?   *1 to 3*

　　　　　　　　 (INT) And how intense is such *emotion*?   *1 to 3*

Annotation Setup – Try it yourself!

a. The new version of our product will be released soon.
b. She reached her boiling point.
c. I passed the exam.

Guidelines (EMO) Is the text emotional? *emotion*/*neutral*

(CONF) How confident are you about *emotion*? *1 to 3*

(INT) And how intense is such *emotion*? *1 to 3*

Is the text emotional?   **emotional/neutral**

	Cohen's κ	Fleiss' κ
Ann1–Ann2	.38	.34
Ann2–Ann3	.43	
Ann3–Ann1	.30	

- Inter-annotator agreement [.3, .43] : Substantial disagreements
- Reader's perspective

RQ1: Are disagreements related to confidence?

	Fleiss' κ			Ann1$_{neut}$ vs. Ann2$_{emo}$
	CONF			CONF
1			*1*	
2			*2*	
3			*3*	
	IAA			Disagreement Counts

- Agreements on EMO increase by increasing CONF
- Disagreements on EMO are rare with high CONF

RQ1: Are disagreements related to confidence?

	Fleiss' κ		Ann1$_{neut}$ vs. Ann2$_{emo}$
	CONF		CONF
1	$-.001$	*1*	
2		*2*	
3		*3*	
	IAA		Disagreement Counts

- Agreements on EMO increase by increasing CONF
- Disagreements on EMO are rare with high CONF

RQ1: Are disagreements related to confidence?

	Fleiss' κ		Ann1$_{neut}$ vs. Ann2$_{emo}$
	CONF		CONF
1	−.001	*1*	
2	.03	*2*	
3		*3*	
	IAA		Disagreement Counts

- Agreements on EMO increase by increasing CONF
- Disagreements on EMO are rare with high CONF

RQ1: Are disagreements related to confidence?

	Fleiss' κ		Ann1$_{neut}$ vs. Ann2$_{emo}$
	CONF		CONF
1	$-.001$	*1*	
2	.03	*2*	
3	**.39**	*3*	
	IAA		Disagreement Counts

- Agreements on EMO increase by increasing CONF
- Disagreements on EMO are rare with high CONF

RQ1: Are disagreements related to confidence?

	Fleiss' κ
	CONF
1	−.001
2	.03
3	**.39**
	IAA

	Ann1$_{neut}$ vs. Ann2$_{emo}$
	CONF
1	10
2	1
3	**0**
	Disagreement Counts

- Agreements on EMO increase by increasing CONF
- Disagreements on EMO are rare with high CONF

RQ1: Are disagreements related to confidence?

	Fleiss' κ			Ann1$_{neut}$ vs. Ann2$_{emo}$	
	CONF	INT		CONF	INT
1	−.001	.04	*1*	10	5
2	.03	.20	*2*	1	6
3	**.39**	**.30**	*3*	**0**	**0**
	IAA			Disagreement Counts	

- Agreements on EMO increase by increasing CONF
- Disagreements on EMO are rare with high CONF
- Same holds for INT

RQ2: Are Confidence and Intensity related?

	Spear.'s ρ
Ann1	.5
Ann2	.51
Ann3	.64

- Positive intra-annotator correlation between CONF and INT

RQ2: Are Confidence and Intensity related?

	Spear.'s ρ
Ann1	.5
Ann2	.51
Ann3	.64

Ann3

	Int 1	Int 2	Int 3
Conf3	3	93	38
Conf2	144	152	0
Conf1	120	15	0

• Positive intra-annotator correlation between CONF and INT

RQ2: Are Confidence and Intensity related?

	Spear.'s ρ
Ann1	.5
Ann2	.51
Ann3	.64

Ann3

|	Int 1	Int 2	Int 3
Conf 3	3	93	38
Conf 2	144	152	0
Conf 1	120	15	0

- Positive intra-annotator correlation between CONF and INT
- Items with low INT are rated with low CONF
- With strong INT, annotators have high CONF

Confidence, Intensity, Disagreements are Entangled

RQ1: Disagreements are related to confidence (and intensity).

- Discrete labels alone not sufficient to understand what is going on
- Confidence reveals when people expect to disagree

RQ2: Confidence and intensity are correlated.
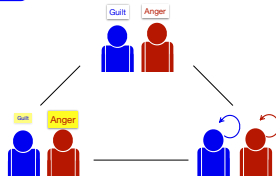
Emotion Components: Appraisals

A crocodile appeared. ⟶ ⟶ **?**

novelty
effort
urgency

Discrete Emotions
fear, surprise, joy, sadness, shame, anger

Intensity
emotion magnitude

⤳**Cognitive Appraisal**
no control, unexpected, ...

.
.
Novelty
Urgency to Act
Coping Potential
Expected Outcome
Congruency with Goals

Underlie Discrete Emotions

Pleasantness the situation was agreeable

Effort physical or mental, expected to be caused by the event

Certainty understanding of what is happening

Attention need or will to attend the emotion event

Responsibility held by the experiencer for the situation

Control of the experiencer over the situation

(Smith & Ellsworth, 1985)

Underlie Discrete Emotions

joy ≠ *fear*

Pleasantness the situation was agreeable   ↑   ↓

Effort physical or mental, expected to be caused by the event   ↓   ↑

Certainty understanding of what is happening   ↑   ↓

Attention need or will to attend the emotion event

Responsibility held by the experiencer for the situation

Control of the experiencer over the situation   ↑   ↓

(Smith & Ellsworth, 1985)

Why Appraisals?

1 Fundamental component of emotions
   - account for how emotions are elicited
     (not stimulus, but stimulus as appraised)

2 Evaluation is needed when annotating
   - e.g., *A crocodile appeared*

Annotating fine-grained Emotion Information

Emotion-Aware, Emotion-Agnostic, or Automatic:
Corpus Creation Strategies to Obtain
Cognitive Event Appraisal Annotations

Jan Hofmann, Enrica Troiano, and **Roman Klinger**
Institut für Maschinelle Sprachverarbeitung, University of Stuttgart, Germany
{jan.hofmann,enrica.troiano,roman.klinger}@ims.uni-stuttgart.de

Annotation Setting

Data: Event descriptions

> *I felt ... when my mom offered me curry*
> *I felt ... because I can't stand when people lie*

- enISEAR corpus (Troiano et al., 2019)
- labels: *anger, disgust, fear, guilt, joy, sadness, shame*
- ...but no emotion words
- 210 of 1001 (English) texts

Annotators: 2 judges, background in EA and annotation experience

Annotation Setting

Data: Event descriptions

I felt ... when my mom offered me curry – joy
I felt ... because I can't stand when people lie – disgust

- enISEAR corpus (Troiano et al., 2019)
- labels: *anger, disgust, fear, guilt, joy, sadness, shame*
- ...but no emotion words
- 210 of 1001 (English) texts

Annotators: 2 judges, background in EA and annotation experience

Annotation Guidelines

Most probably, at the time when the event happened, the writer...

Pleasantness	*found that the event was pleasant*
Effort	*had to expend mental or physical effort to deal with the situation*
Certain	*was certain about what was happening*
Attention	*wanted to devote further attention to the event*
Responsibility	*was responsible for the situation*
self-Control	*found that he/she was in control of the situation*
situational-Control	*found that the event could not have been changed or influenced by anyone*

➙ *1 or 0*

Annotation Guidelines – Try it yourself!

```
I felt ...  when I was abseiling down a cliff-face
```

Most probably, at the time when the event happened, the writer...

Pleasantness	*found that the event was pleasant*
Effort	*had to expend mental or physical effort to deal with the situation*
Certain	*was certain about what was happening*
Attention	*wanted to devote further attention to the event*
Responsibility	*was responsible for the situation*
self-Control	*found that he/she was in control of the situation*
situational-Control	*found that the event could not have been changed or influenced by anyone*

➙ *1 or 0*

Annotation Strategies

Challenge: impersonating in the writer, possible disagreements on both discrete emotions and appraisals

❶ Show only text (EMOHIDE)

$p(appraisal|text)$

I felt ... when my mom offered me curry

Annotation Strategies

Challenge: impersonating in the writer, possible disagreements on both discrete emotions and appraisals

① Show only text (EMOHIDE)

p(appraisal|text)

I felt ... when my mom offered me curry

② Show emotion (EMOVIS)

p(appraisal|emotion, text)

I felt joy when my mom offered me curry

Inter-annotator Agreement

	EMOHIDE
Appraisal	Cohen's κ
Pleasantness	.87
Effort	.38
Certainty	.43
Attention	.30
Responsibility	.64
self-Control	.71
sit.-Control	.54
Macro $\varnothing$	**.55**

Inter-annotator Agreement

Appraisal	EMOHIDE Cohen's κ	EMOVIS κ	Δ
Pleasantness	.87	.93	+.06
Effort	.38	.44	+.06
Attention	.30	.55	+.25
Certainty	.43	.71	+.28
Responsibility	.64	.80	+.16
self-Control	.71	.66	−.05
sit.-Control	.54	.65	+.11
Macro $\varnothing$	**.55**	**.68**	+ **.13**

More information, less Disagreements

Knowing the emotion allows for more consistent appraisal reconstructions.

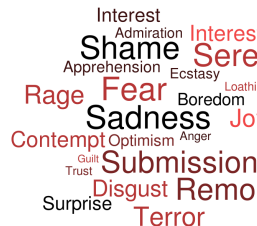**1** **Show only text**   **2** **Show emotion**

```
I felt fear when I was abseiling down a cliff-face
```

$+$**Effort**
$+$ **Pleasantness**
$+$ **Responsibility**
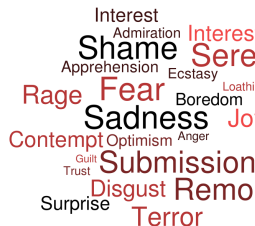$+$ **situational-Control**

What have we learned?

1 Disagreements are systematic

What have we learned?

1 Disagreements are systematic
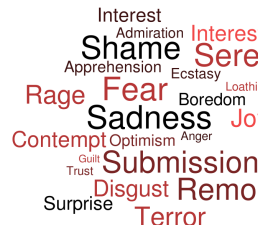- Disagreements ≠ Unreliability

What have we learned?

1 Disagreements are systematic

- Disagreements ≠ Unreliability

- Typical measures of IAA underestimate quality of data
  (do not capture patterns in disagreement)

  ..but we can give different weight to different disagreements:
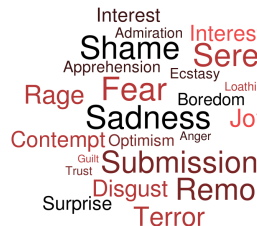 *neut-emo* CONF3 > *neut-emo* CONF1
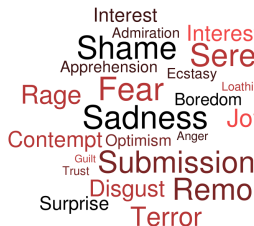
What have we learned?

❷ Annotation is a multi-dimensional problem
- Observing interaction is beneficial
  (complex phenomenon ≈ richer annotation)

Interest

Admiration   Interes

Shame   Sere

Apprehension   Ecstasy

Loathi

Rage   Fear   Boredom

Sadness   Jo

Contempt Optimism Anger

Guilt

Trust Submission

Disgust Remo

Surprise

Terror

What have we learned?

❷ Annotation is a multi-dimensional problem

- Observing interaction is beneficial
  (complex phenomenon ≈ richer annotation)

- Do intensity corpora contain right annotation?
  ...to be explored!

Interest
Admiration Interes
Shame Sere
Apprehension Ecstasy
Loathi
Rage Fear Boredom
Sadness Jo
Contempt Optimism Anger
Guilt Submission
Trust
Disgust Remo
Surprise
Terror

What have we learned?

❷ Annotation is a multi-dimensional problem

- Observing interaction is beneficial
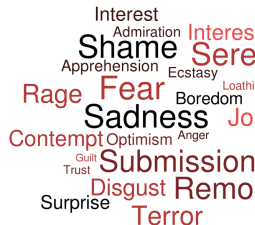 (complex phenomenon ≈ richer annotation)

- Do intensity corpora contain right annotation?
  ...to be explored!

- Give information to annotators
  ...but data leakage.

Interest

Admiration Interes

Shame Sere

Apprehension Ecstasy

Loathi

Rage Fear Boredom

Sadness Jo

Contempt Optimism Anger

Guilt Submission

Trust

Disgust Remo

Surprise Terror

Future Work – Emotion Style Transfer

Given **joy**

Yes, sure I've loved it! The concert hall was vibrant.

$\approx$

I want **annoyance**

I loved it, yeah, sure... The concert hall was bursting with chaos.

Future Work – Emotion Style Transfer

Given **joy**

≈

I want **annoyance**

Yes, sure I've loved it! The concert hall was vibrant.

I loved it, yeah, sure... The concert hall was bursting with chaos.

Future Work – Emotion Style Transfer

Given **joy**   ≈   I want **annoyance**

Yes, sure I've loved it! The concert hall was vibrant.

I loved it, yeah, sure... The concert hall was bursting with chaos.