University of Bamberg

Event-centric Emotion Analysis in Natural Language Processing

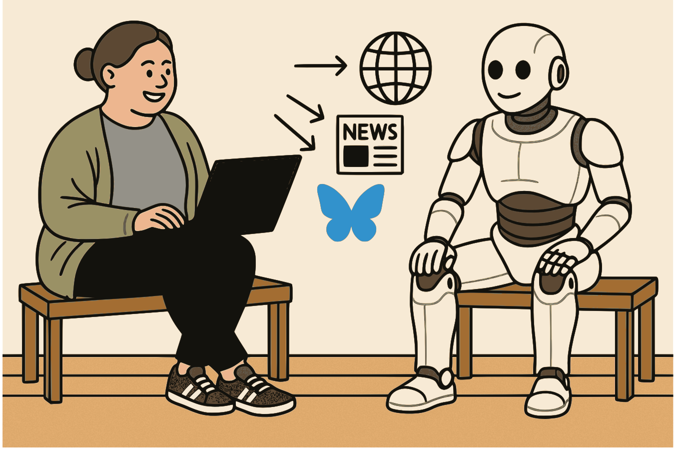
Appraisal Variables as Emotion Models

AI Meets Human Data Colloquium, Augsburg, January 12, 2026

Roman Klinger
roman.klinger@uni-bamberg.de

romanklinger.de romanklinger
https://www.bamberg.de/nlproc/
https://www.romanklinger.de/

Natural Language Processing and Understanding



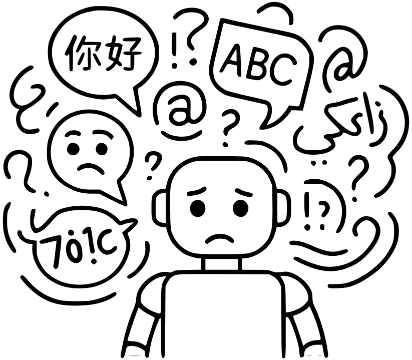
- We study how machines can understand human language
- We focus on written text

Understanding...

- What does it mean to "understand"?
- Cambridge Dictionary: "to know the meaning of something that someone says"

- How can we make computers understand?
- How can we measure if we are successful?
- How and what for can we make use of the meaning that has been understood?

Natural Language Understanding



- Desideratum: A machine that understands language as humans do?
- How to study language in its entirety? (universal language understanding ability)
- We study particular phenomena.
- We define concrete tasks to solve.

⇒ Pragmatic approach to language understanding sky news .COM.AU

World News › United States

Donald Trump mocks Bill Gates after billionaire's humiliating backflip on climate change

30 October 2025 - 04:46AM sky news .COM.AU

Donald Trump wasted no time mocking Bill Gates after the billionaire admitted that climate change will not lead to humanity's demise in a memo released yesterday.

Tasks in Natural Language Understanding

"Donald Trump mocks Bill Gates after billionaire's humiliating backflip on climate change"

Sky News Australia, Oct 30, 2025

What information is in this sentence that's worth understanding?

- Find entity names: Donald Trump ; Bill Gates
- Recognize the sentiment: negative
- Topic: climate change
- Stances: Bill Gates → opinion change on climate change.
- Relation: Donald Trump → negative opinion (Bill Gates)
- ...
- Aggregating information enables many use cases:
  diverse news recommendation, social network analysis, opinion mining, ...

Outline

1 Introduction to Natural Language Understanding

2 Emotions and Emotion Analysis

3 Appraisals

4 Generation of Explaining Context

5 How to Collect Data?

6 Appraisals to Understand Argument Convincingness

7 Other Topics

8 Wrap Up

Emotion Examples

Which emotion is associated with each example?

How did you recognize that?

- "She became angry."
- "A tear is running down his face."
- "We are going for a walk at the beach."

With this exercise, we discussed two things:

- What is an appropriate set of emotions?
- How are they expressed/recognized?

Emotion Models – Basic Emotions

How to define a categorical system of emotions?

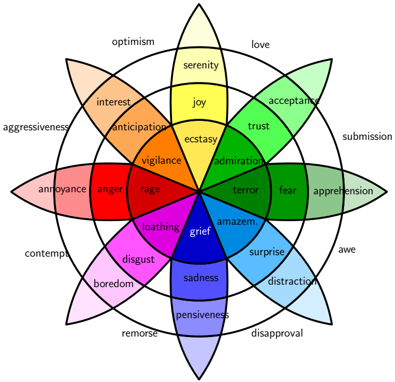
- Distinctive universal signals
- Presence in other primates
- Distinctive physiology
- Distinctive universals in antecedent events
- Coherence among emotional response
- Quick onset
- Brief duration
- Automatic appraisal
- Unbidden occurrence

Ekman (1992): An argument for basic emotions.

Ekman: What are non-basic emotions?

- “I do not allow for non-basic emotions” (Ekman, 1999)
  - ⇒ They do no exist.
- What is love, depression, or hostility?
  - Personality traits (hostility, openness)
  - Moods (depression, anxiety, long-term disturbances are clinically relevant)
  - Emotional plots (love, grief, jealousy)

Models of Basic Emotions: Plutchik's Wheel



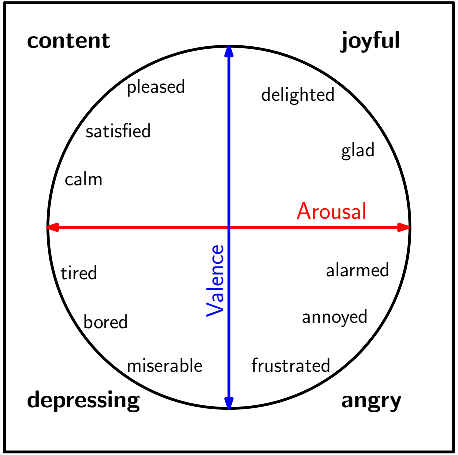
An emotion is a patterned bodily reaction that follows a function

- protection – fear
- destruction – anger
- reproduction – joy
- deprivation – sadness
- incorporation – acceptance
- rejection – disgust
- exploration – anticipation
- orientation – surprise

⇒ These are basic emotions according to Plutchik (1970)

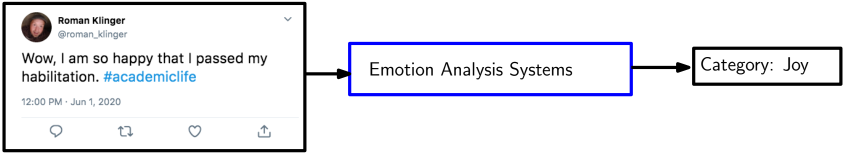
Emotion Models – Valence-Arousal Model of Affect

- Perhaps mixtures and opposites do not make sense, but there are other ways to explain the relations between emotions?

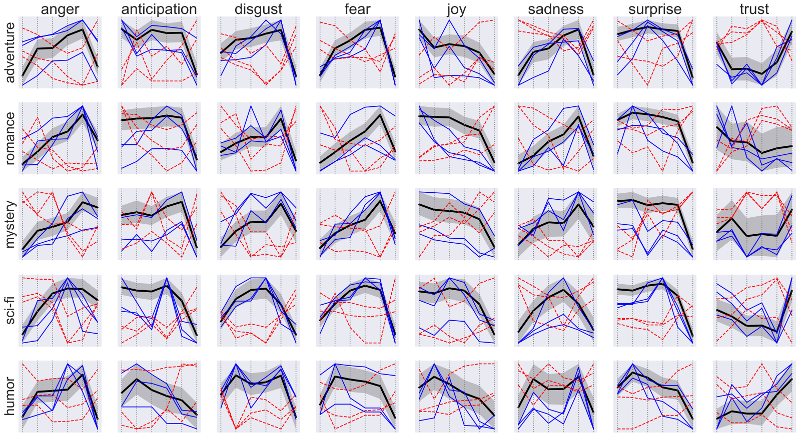


Russell, R. (1980). A Circumplex Model of Affect.

Emotion Analysis: What we want to do.



Literary Studies



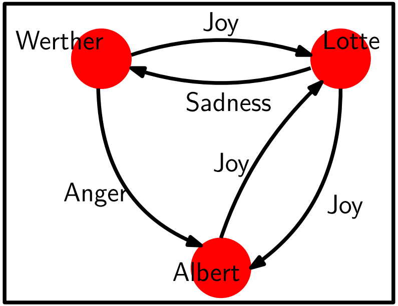
E. Kim, S. Padó, and R. Klinger (2017). "Investigating the Relationship between Literary Genres and Emotional Plot Development". In: LaTeCHCLfL

News Analysis

Emotion	Dominant Emotion	Reader Emotions
Anger	The Blaze, The Daily Wire, BuzzFeed	The Gateway Pundit, The Daily Mail, Talking Points Memo
Annoyance	Vice, NewsBusters, AlterNet	Vice, The Week, Business Insider
Disgust	BuzzFeed, The Hill, NewsBusters	Mother Jones, The Blaze, Daily Caller
Fear	The Daily Mail, Los Angeles Times, BBC	Palmer Report, CNN, InfoWars
Guilt	Fox News, The Daily Mail, Vice	The Washington Times, Reason, National Review
Joy	Time, Positive.News, BBC	Positive.News, ThinkProgress, AlterNet
Love	Positive.News, The New Yorker, BBC	Positive.News, AlterNet, Twitchy
Pessimism	MotherJones, Intercept, Financial Times	The Guardian, Truthout, The Washington Post
Neg. Surprise	The Daily Mail, MarketWatch, Vice	The Daily Mail, BBC, Breitbart
Optimism	Bussines Insider, The Week, The Fiscal Times	MarketWatch, Positive.News, The New Republic
Pos. Surprise	Positive.News, BBC, MarketWatch	Positive.News, The Washington Post, MotherJones
Pride	Positive.News, The Guardian, The New Yorker	Daily Kos, NBC, The Guardian
Sadness	The Daily Mail, CNN, Daily Caller	The Daily Mail, CNN, The Washington Post
Shame	The Daily Mail, The Guardian, The Daily Wire	Mother Jones, National Review, Fox News
Trust	The Daily Signal, Fox News, Mother Jones	Economist, The Los Angeles Times, The Hill

L. A. M. Bostan, E. Kim, and R. Klinger (2020). “GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception”. In: LREC

Social Networks



E. Kim and R. Klinger (2019). "Frowning Frodo, Wincing Leia, and a Seriously Great Friendship: Learning to Classify Emotional Relationships of Fictional Characters". In: NAACL

Emotions and Events

Emotions and Events are linked in (at least) two ways:

Emotions are events

- "Donald is happy about his birthday present."
- FrameNet Emotion Directed Frame:
  - Event: "happy"
  - Experiencer: "Donald"
  - Stimulus: "his birthday present"
  - ...

⇒ Emotion role labeling (not the topic of today's talk)

Events cause emotions

- "There is a car on fire."
  - Relevant event for the speaker, might cause fear.
  - Requires interpretation of events to infer possible emotions.
- (main part of today's talk)

How are emotions expressed?
Do we need to deal with event descriptions (Twitter/Literature)?

Component	Example	Fraction (T/L)	
Physiology	Loves when a song makes your heart race	5	8
Action	sometimes when i think bout you i want to beat the shit out of your face	18	19
Expression	when I walk in the room and my nephew recognises me his face lights up with the biggest smile	13	44
Feeling	Feelin a bit sad today	32	17
Appraisal	Thinks that mel had a great 50th birthday party	75	61

F. Casel, A. Heindl, and R. Klinger (2021). "Emotion Recognition under Consideration of the Emotion Component Process Model". In: KONVENS

Outline

1 Introduction to Natural Language Understanding

2 Emotions and Emotion Analysis

3 Appraisals

4 Generation of Explaining Context

5 How to Collect Data?

6 Appraisals to Understand Argument Convincingness

7 Other Topics

8 Wrap Up

Definition of Emotions: Components

Emotion (Scherer, 2005)

Emotions are "an episode of interrelated, synchronized changes in the states of [...] five organismic subsystems in response to the evaluation of a [...] stimulus-event ..."

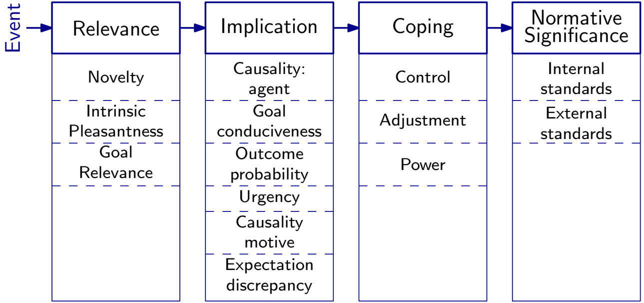
Event

Feeling	Expression	Bodily Symptom
Action Tendency	Cognitive Appraisal	
Fear		

Components

Name

Cognitive Appraisal in Scherer's Component Process model



K.R. Scherer (2001). Appraisal Considered as a Process of Multilevel Sequential Checking.

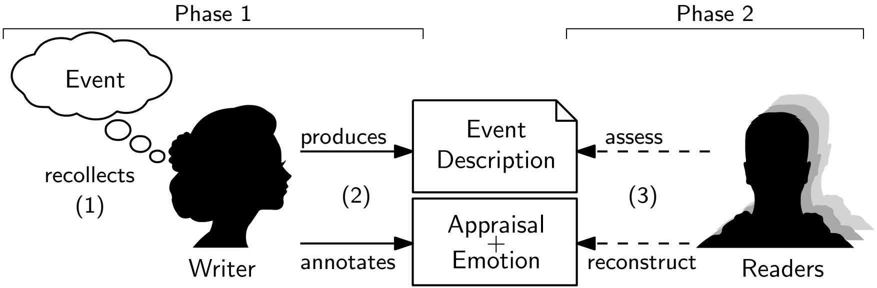
Research Questions

- Can appraisals and emotions be annotated reliably by external annotators?
- Can we computationally model appraisals and does it help emotion categorization?

E. Troiano, L. Oberländer, and R. Klinger (2023). "Dimensional Modeling of Emotions in Text with Appraisal Theories: Corpus Creation, Annotation Reliability, and Prediction". In: Computational Linguistics 49.1

J. Hofmann et al. (2020). "Appraisal Theories for Emotion Classification in Text". In: COLING

Approach



- Production: 550 event descriptions for anger, boredom, disgust, fear, guilt/shame, joy, pride, relief, sadness, surprise, trust, no emotion
- Five readers for subset of produced texts

Examples

pride I baked a delicious strawberry cobbler.

fear I felt ... when there was a power outage in my home. That day, my wife and I were cuddling in the sitting room when a thunderstorm started. Then ... filled me when thunder hit our roof and all the lights went off.

joy I found the perfect man for me, and the more time goes on, the more I realized he was the best person for me. Every day is a ....

Reliability Results

Condition	Val.	#Pairs	Emotion F$_1$ G–V	Emotion F$_1$ V–V	Emotion Acc. G–V	Emotion Acc. V–V	Appraisal RMSE G–V	Appraisal RMSE V–V
All Data		6600 12000	.49	.50	*.49	*.52	*1.57	*1.48
Gender match	M–M	631 1113	.50	*.45	.51	*.49	1.55	1.50
	F–F	2405 1377	.49	*.52	.51	*.55	1.57	*.1.50
	≠	2962 3920	.49	*.48	.50	*.52	1.57	*.1.48
Age diff.	> 7	3089 7991	.49	*.48	.51	*.51	*1.58	1.48
	≤ 7	2076 3939	.49	*.51	.50	*.54	*1.56	1.48
Validators' Event Fam.	> 3	1386 540	.49	.44	.51	.47	*1.60	*1.42
	≤ 3	2099 676	.48	.45	.49	.48	*1.58	*1.47
Validators' Openness	+	2685 1472	.49	.49	.50	.52	1.57	1.47
	–	3000 1568	.49	.48	.50	.51	1.57	1.48
Validators' Conscien.	+	3151 1638	*.48	.51	*.49	.53	*1.57	*1.49
	–	2589 1426	*.50	.51	*.51	.54	*1.56	*1.46
Validators' Extraversion	+	2878 1685	.49	*.48	.50	*.51	*1.58	*1.51
	–	2812 1535	.50	*.52	.51	*.55	*1.56	*1.46
Validators' Agreeabl.	+	2675 1451	.49	*.51	.51	*.54	*1.58	1.47
	–	2930 1553	.48	*.45	.49	*.49	*1.56	1.47
Validators' Emot. Stab.	+	2838 3009	*.48	*.48	*.49	*.51	*1.57	*1.50
	–	2792 2897	*.50	*.51	*.51	*.54	*1.56	*1.46

- Validators agree more with each other than with the generator
- V–G agreements:
  - Higher agreement for Female pairs
  - Low age difference leads to higher agreement
- V properties only:
  - Event familiarity hurts agreement for appraisal
  - We expected Open annotators to perform better.
  - Emotional stability "hurts" emotion annotation.
  - Extraversion, Conscient., Agreeableness help.
- Most differences are quite small (but significant)

Examples (writer/reader/avg. writer–reader agreement as error)

- All writers/readers agree on emotion, high average appraisal agreement

  pride, .65 — I baked a delicious strawberry cobbler

  fear, .84 — A housemate came at me with a knife

- All writers/readers agree on emotion, low average appraisal agreement

  disgust, 2.0 — His toenails where massive

  fear, 2.1 — I felt ... going in to hospital

- All readers agree on the emotion, but not with the writer, high appraisal agreement

  trust, joy, .87 — I am with my friends

  anger, fear, 1.1 — My waters broke early during pregnancy

- All readers agree on the emotion, but not with the writer, low appraisal agreement

  pride, sadness, 1.7 — That I put together a funeral service for my Aunt

Appraisals add additional information to emotion analysis

That I put together a funeral service for my Aunt

Dimension	Writer	Readers	$\triangle$
Emotion	Pride	Sadness	
Suddenness	4	3.6	0.4
Familiarity	1	2.0	−1.0
Predictability	1	1.8	−0.8
Pleasantness	4	1.0	3.0
Unpleasantness	2	4.8	−2.8
Goal-Relevance	4	2.6	1.4
Chance-Resp.	4	4.4	−0.4
Self-Resp.	1	1.2	−0.2
Other-Resp.	1	1.4	−0.4
Conseq.-Predict.	2	1.8	0.2
Goal Support	1	1.2	−0.2
Urgency	2	3.8	−1.8
Self-Control	5	3.2	1.8
Other-Control	3	2.0	1.0
Chance-Control	1	4.6	−3.6
Accept-Conseq.	4	2.4	1.6
Standards	1	2.4	−1.4
Social Norms	1	1.2	−0.2
Attention	4	4.4	−0.4
Not-Consider	1	3.8	−2.8
Effort	4	4.6	−0.6

Emotion Annotation Result

Conclusion

Annotators can quite well reconstruct authors emotion,
but there is a small and significant agreement drop.

Challenge

Authors recall "important" events. We do (presumably) not get a realistic subsample of event descriptions as they appear in the wild.

- Appraisals explain subjectivity
- Not shown: appraisals help to disambiguate emotion categories in automatic models

Potential Reason for V–G Discrepancy

- Isolated events are not sufficient
- Subjectivity is not only personality and demographics

Outline

1 Introduction to Natural Language Understanding

2 Emotions and Emotion Analysis

3 Appraisals

4 Generation of Explaining Context

5 How to Collect Data?

6 Appraisals to Understand Argument Convincingness

7 Other Topics

8 Wrap Up

Idea: Generate backstories to explain emotions/appraisals

Event

"The loudspeaker suddenly malfunctioned and went silent."

- Many emotion interpretations possible.

⇒ We autogenerate stories that explain such event for a given emotion.

J. Schäfer and R. Klinger (2025). Shaping Event Backstories to Estimate Potential Emotion Contexts. arXiv: 2508.09954 [cs.CL]. URL: https://arxiv.org/abs/2508.09954

Backstories

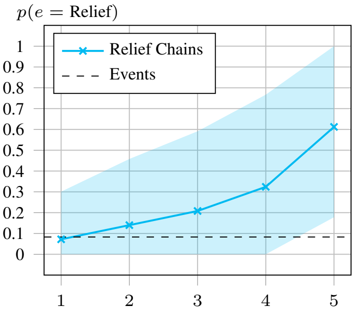
Relief: I was tasked with giving a presentation to a large crowd. The sound system malfunctioned, amplifying my voice to an ear-piercing level. The sound technician ignored the problem and chatted with someone. The audience covered their ears and looked at me with discomfort. The loudspeaker suddenly malfunctioned and went silent.

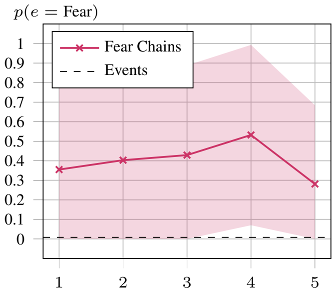
Fear: I arrived at a remote wilderness survival training camp, where the instructors emphasized the importance of following loudspeaker instructions for safety. The instructors warned us about the toxic waste site nearby and explained that the loudspeaker would alert us to any changes in air quality. During the first exercise, I struggled to navigate the challenging terrain, but the loudspeaker provided crucial guidance, helping me stay on track. I completed a difficult obstacle course, relying heavily on the loudspeaker's instructions to avoid hazards and find the safest route. The loudspeaker suddenly malfunctioned and went silent.

Pride: I spent the entire morning upgrading the sound system with a new backup system to prevent technical issues. The event organizer informed me that the conference was running 30 minutes behind schedule, giving me extra time to test the new backup system. I used the extra time to run a series of tests on the sound system, trying to simulate potential failures. The keynote speaker began to talk, and the sound system was working flawlessly, but I was still waiting for a real test of the new backup system. The loudspeaker suddenly malfunctioned and went silent.

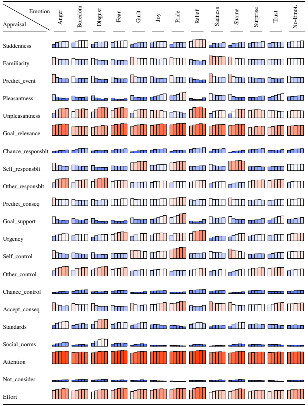
Results in a Nutshell

- Backstories make interpretation more clear for models and annotators (details not shown for time reasons).
- Effect more pronounced for some emotions than others



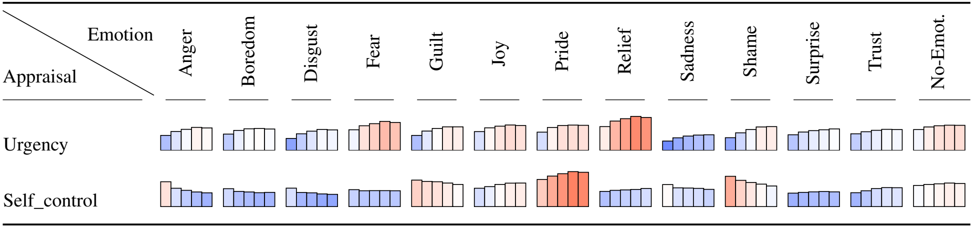


Shape of Appraisal Trajectories also Matters



J. Schäfer, J. Wagner, and R. Klinger (2026). "Appraisal Trajectories in Narratives Reveal Distinct Patterns of Emotion Evocation". In: Submitted to WASSA

Shape of Appraisal Trajectories also Matters (subset)



J. Schäfer, J. Wagner, and R. Klinger (2026). "Appraisal Trajectories in Narratives Reveal Distinct Patterns of Emotion Evocation". In: Submitted to WASSA

Outline

1 Introduction to Natural Language Understanding

2 Emotions and Emotion Analysis

3 Appraisals

4 Generation of Explaining Context

5 How to Collect Data?

6 Appraisals to Understand Argument Convincingness

7 Other Topics

8 Wrap Up

Introduction

- Prompting humans for data creation has advantages:
  - Direct access to the author's assessment
  - Privacy: authors are aware what they share and can filter
- Potential issues:
  - Data is not realistic
  - People recall particularly "prototypical" events
  - Type of data might differ due to missing post creation triggers

C. Bagdon et al. (2025). "Donate or Create? Comparing Data Collection Strategies for Emotion-labeled Multimodal Social Media Posts". In: ACL

Approach: Data elicitation strategies

- Creation:
  - “Think of an event that caused an emotion X in you.”
  - “Write a social media post text about that.”
  - “Select an image you want to share from a CC image data base.”
- Donation:
  - “Pick a multimodal post from your social media timeline that you made because the associated event caused emotion X.”
  - “Copy paste the text and the image.”
- Recent:
  - “Pick the 10 most recent posts from your social media timeline.”
  - “Annotate them for the following emotion set.”

Data Example

Name
@Username

Absolutely insane, what is going on?!

Creation post labeled as surprise.

Name2
@Username2

Exhibit 2

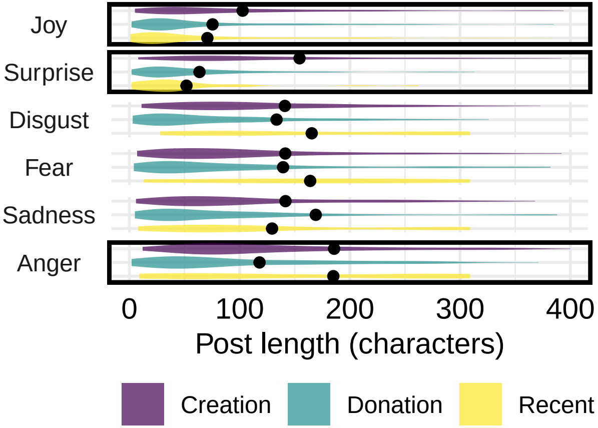
Trump supporters say ear bandages are 'sign of love'

Several supporters of former President Donald Trump wore bandages on their ears to the third night of the Republican National Convention (RNC) in Milwaukee, Wisconsin.

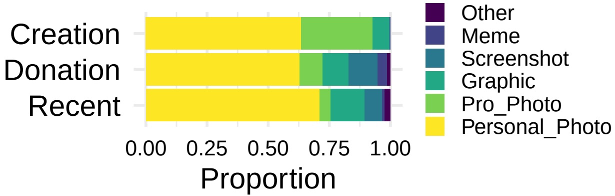
Members of the RNC's Arizona delegation said they were wearing the bandages as a sign of solidarity with the former president after he survived an assassination attempt.

Recent post labeled as anger.

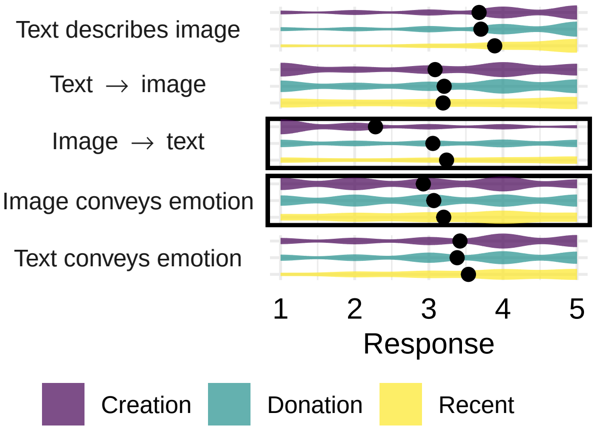
Are the subcorpora comparable? – Post Length



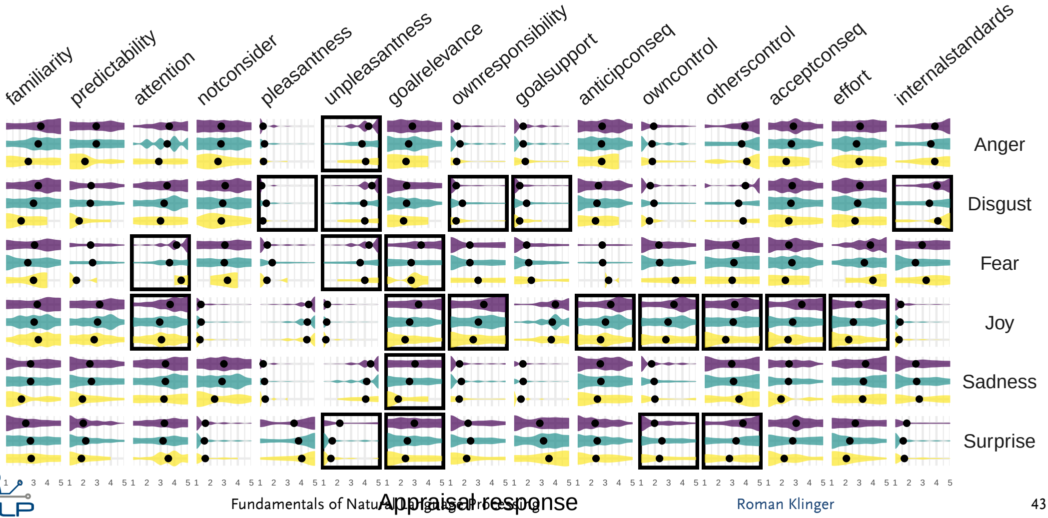
Are the subcorpora comparable? – Image Type



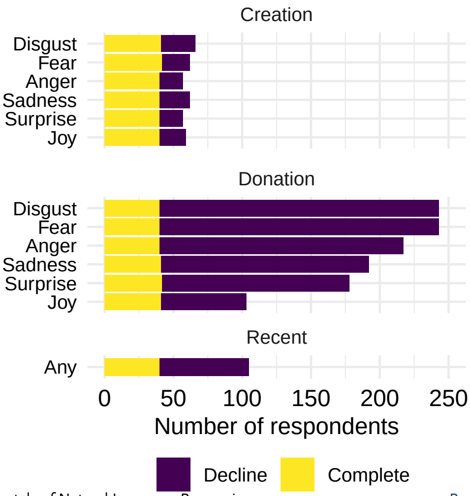
Are the subcorpora comparable? – Text–Image Relation



Are the subcorpora comparable? – Appraisal–Emotion



Are the subcorpora comparable? – Participant acceptance



Are the differences a problem?

Experiment

- Fine-tune RoBERTa with CLIP/early fusion to predict emotions
  - Train on Donation vs. train on Creation

Results

- No big performance differences: F score .38 vs. .40
  - ⇒ The experimentally elicited data is fine to optimize a model.
- But: The estimate on donated data is overall optimistic!
  F score of .60 and .62.
  - ⇒ Real data is required to estimate model performance.
- Zero-Shot prompting (Llama3.2-vision) leads to slightly better results for donated data.

Outline

1 Introduction to Natural Language Understanding

2 Emotions and Emotion Analysis

3 Appraisals

4 Generation of Explaining Context

5 How to Collect Data?

6 Appraisals to Understand Argument Convincingness

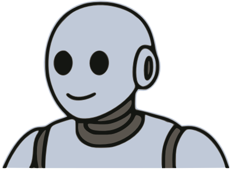
7 Other Topics

8 Wrap Up

Emotion Detection in Arguments

"You are an expert in emotion annotation in arguments. Which emotion does the reader of the following argument likely feel?"

We should ban plastic bottles, because they harm the environment.



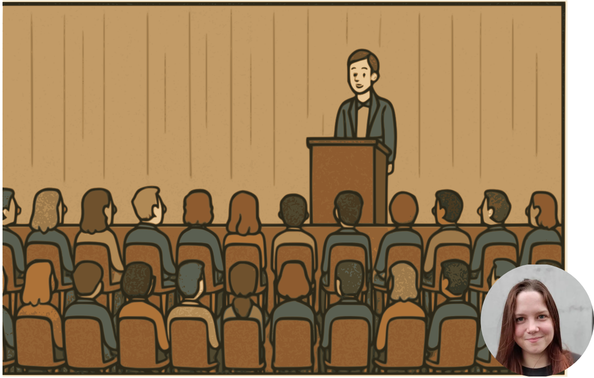
??

- Models lack access to context.
- They tend to predict fear or anger.

L. Greschner and R. Klinger (2025). "Fearful Falcons and Angry Llamas: Emotion Category Annotations of Arguments by Humans and LLMs". In: NLP4DH

How can we conduct contextualized emotion detection (and convincingness assessment) in arguments?

- We need to know for whom we make predictions!
- That is a challenge, we need annotated arguments with information about the annotator!
- How to get such data?
- We asked people to role play a debate and annotate arguments they read.



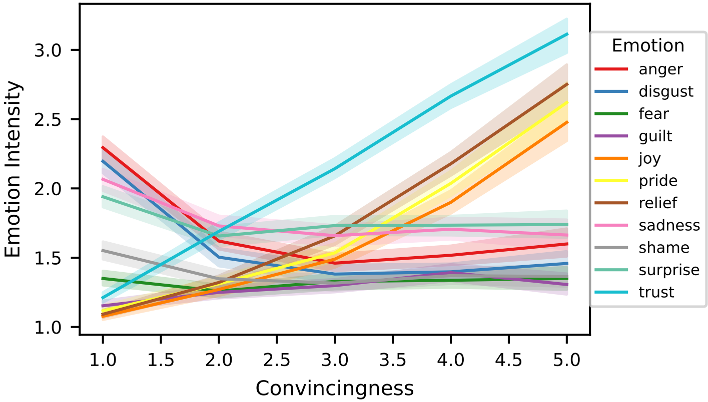
L. Greschner, S. Weber, and R. Klinger (2025). Trust Me, I Can Convince You: The Contextualized Argument Appraisal Framework. under review for LREC 2026. arXiv: 2509.17844 [cs.CL]

How can we conduct contextualized emotion detection (and convincingness assessment) in arguments?

- We need to develop and evaluate methods to collect data in context.
- Who's the speaker? Who's the listener?
- We need to develop methods to integrate contextual information in computational models.

Correlations of Emotions with Convincingness

Emotion	r
Trust	0.570
Relief	0.511
Pride	0.458
Joy	0.435
Guilt	0.105
Fear	0.006
Surprise	−0.072
Shame	−0.073
Sadness	−0.153
Anger	−0.265
Disgust	−0.264



Correlations of Appraisals with Convincingness

Appraisal	r
Pleasantness	0.566
Positive Consequentiality	0.392
Familiarity	0.327
Negative Consequentiality	0.203
Consequential Importance	0.141
Consequence Manageability	−0.034
Cognitive Effort	−0.061

Appraisal	r
Internal Check	−0.103
Argument Internal Check	−0.109
Response Urgency	−0.242
Suppression	−0.326
Suddenness	−0.342
External Check	−0.355
Unpleasantness	−0.385
Argument External Check	−0.497

- Pleasant arguments whose outcomes are good for the self and which are familiar are more convincing.
- Surprising arguments and those which go against laws or social standards are less convincing (and cause anger and disgust).

Outline

1 Introduction to Natural Language Understanding

2 Emotions and Emotion Analysis

3 Appraisals

4 Generation of Explaining Context

5 How to Collect Data?

6 Appraisals to Understand Argument Convincingness

7 Other Topics

8 Wrap Up

For whom do models make predictions?

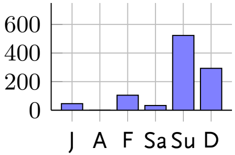
- If we don’t tell the model for whom it should make a prediction, with whose annotations is it best aligned?
- Models best reconstruct a person’s annotation when they are white, comparably young, and male.
- We need to understand biases and make models work well for everybody.

J. Schäfer, A. Combs, et al. (2025). “Which Demographics do LLMs Default to During Annotation?” In: ACL

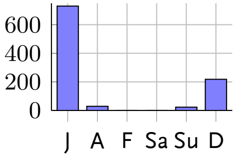
Human annotation varies – Should LLM's annotation also vary?

- You are an expert in emotion annotation. The label set is {LS}.
  The instance to classify is "The dog ran towards me."

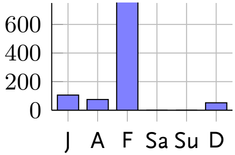
- anger, fear, joy, disgust, sadness, surprise.



- surprise, sadness, disgust, joy, fear, anger.



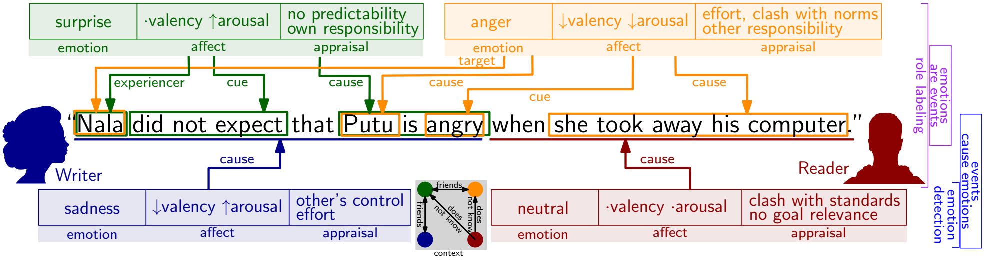
- angr, feer, joy, disgst, sadnes, suprise.



- Human susceptibility to prompt changes differs from LLM's brittleness.
- Do we want model's outputs to vary as human's output does?
- If yes, how to achieve that? If no, what should they do?

J. Li, S. Papay, and R. Klinger (2025). "Are Humans as Brittle as Large Language Models?" In: IJCNLP–AACL

Integration of Appraisal Analysis with Role Labeling



R. Klinger (2023). "Where are We in Event-centric Emotion Analysis? Bridging Emotion Role Labeling and Appraisal-based Approaches". In: Big Picture Workshop

Outline

1 Introduction to Natural Language Understanding

2 Emotions and Emotion Analysis

3 Appraisals

4 Generation of Explaining Context

5 How to Collect Data?

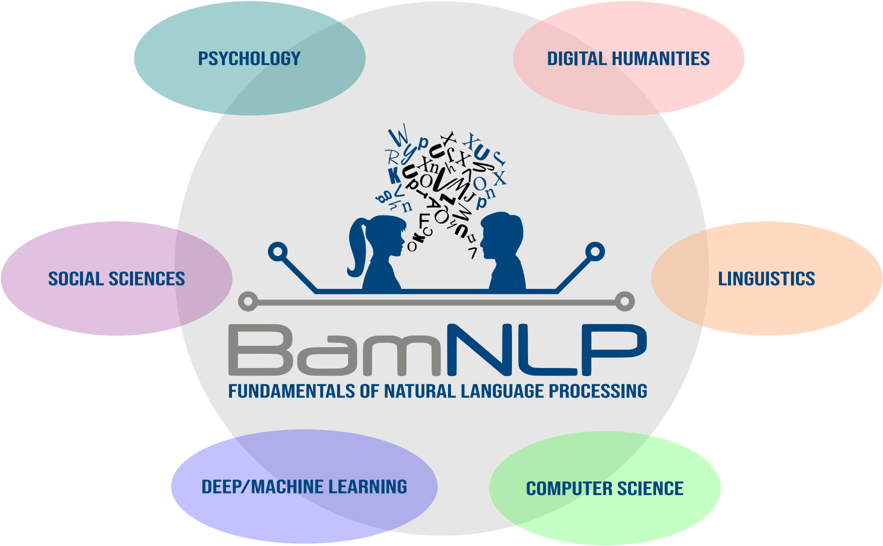
6 Appraisals to Understand Argument Convincingness

7 Other Topics

8 Wrap Up

Take Home

- Emotion analysis is a subjective natural language understanding task
- Event-centric appraisals and context explain subjectivity
- Collecting data is challenging, and data donations are better than prompting humans for data creation
- Context matters, and accessing individual knowledge is hard
- We need to better understand how variance of predictions of models should be aligned with humans
- Many open research tasks in emotion analysis


PSYCHOLOGY
DIGITAL HUMANITIES
SOCIAL SCIENCES
LINGUISTICS
BamNLP
FUNDAMENTALS OF NATURAL LANGUAGE PROCESSING
DEEP/MACHINE LEARNING
COMPUTER SCIENCE

Thank you for
your attention.
Questions? Remarks?

?

Thanks to:

- All research groups I was part of so far and all collaborators.
- All of you for your interest!
- Please reach out if you want to talk, chat, discuss, meet us, drink coffee, work with us, collaborate, ...

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

University of Bamberg

Event-centric Emotion Analysis in Natural Language Processing

Appraisal Variables as Emotion Models

AI Meets Human Data Colloquium, Augsburg, January 12, 2026

Roman Klinger
roman.klinger@uni-bamberg.de

romanklinger.de romanklinger
https://www.bamberg.de/nlproc/
https://www.romanklinger.de/

BamNLP