University of Bamberg

Eliciting Textual Data from Psychological Study Participants

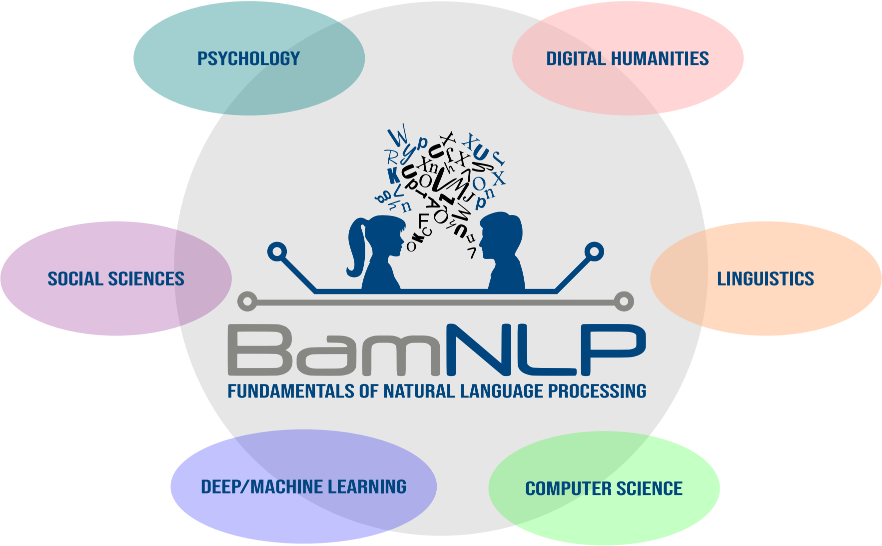
Oberseminar Computerlinguistik, January 30, 2025

Roman Klinger
roman.klinger@uni-bamberg.de

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

About Myself

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning
- 2010, 2013: Research visits at UMass Amherst:
  Probabilistic machine learning, MCMC inference
- 2011–2012: Postdoc at Fraunhofer SCAI:
  Social media mining, eGovernment
- 2013–2014: Postdoc at Bielefeld University:
  Sentiment analysis, opinion mining
- 2015: Co-Founder of Semalytix GmbH (exit 2020)
  Social Media Health Mining
- 2014–2024: (Senior) Lecturer/apl. Prof at IMS, Uni Stuttgart
  Natural Language Understanding and Generation
- 03/2024: Full Professor for Fundamentals of NLP, Bamberg


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


Bamberger Zentrum für Künstliche Intelligenz
Digitales Lehren & Lernen
KI & Ethik
KI & Gesundheit
KI & Bildung
Kognition & Interaktion
Soziotechnische Systeme
Digitale Überlieferung
Erklärbares Maschinelles Lernen
KI-Systementwicklung
Kognitive Systeme
Kulturinformatik
Medieninformatik & Information Retrieval
Sprachgenerierung & Dialogsysteme
Smart Environments
User Experience & Design
Informationsvisualisierung
Grundlagen der Sprachverarbeitung
Multimodale Intelligente Interaktion
Mensch-Computer-Interaktion
Algorithmen & Komplexitätstheorie
KI-Engineering in Unternehmen
Computational Social Science & KI
Computergrafik & ihre Grundlagen
Data Engineering
Energieeffiziente Systeme
Analyse sozialer Netzwerke

Outline

1 Motivation

2 Emotion Analysis – Annotator’s Label Reconstruction

3 Multimodal Data Acquisition – Data Representativeness

4 Deception – Intentions

5 Coping – Role Playing

6 Take Home

Annotation Tasks and Data Acquisition (1/2)

Example 1: POS Tagging ("What are the POS tags"?)

He	walks	to	the	kitchen	.
PP	VERB	PREP	DT	NOUN	.

⇒ Presumably objective task, annotation of existing texts with trained experts.

Example 2: Hate Speech Detection ("Is this, legally considered, hate speech in Europe?")

The religious group of Norse paganism is terrible and should be eliminated from our country.
Yes, negative mention of minority group and call for action.

⇒ Presumably objective task, annotation of existing texts with trained experts.

Annotation Tasks and Data Acquisition (2/2)

Example 3: Sentiment Analysis ("Do you find this text positive?")

The AFD is the only party which plans to do something good for Germany.
Good. (if you like the AFD). Bad. (otherwise)

→ Subjective task, annotation of existing texts by multiple people.

Example 4: Author-Perspective Emotion Detection ("Which emotion did the author feel in context of the described event?")

I organized the funeral service.
Pride? Sadness?

⇒ Annotators challenged to recreate author-level labels.

Subjective Tasks of Author-Related Labels

There are many such tasks:

- Author's emotional state
- Deception
- Personality
- Demographics
- Intend
- ...

Questions to discuss

- Can annotators reconstruct author level private states?
  - Use case study on event-centered emotion analysis
- When people create synthetic posts in an experiment, how unrealistic is the outcome?
  - Use case study on multimodal emotion analysis
- Can we make people to have an intention?
  - Use case study on deception detection.
- Can people successfully play to be different than they actually are?
  - Use case study on coping strategy detection

Outline

1 Motivation

2 Emotion Analysis – Annotator's Label Reconstruction

3 Multimodal Data Acquisition – Data Representativeness

4 Deception – Intentions

5 Coping – Role Playing

6 Take Home

Outline

1 Motivation

2 Emotion Analysis – Annotator’s Label Reconstruction

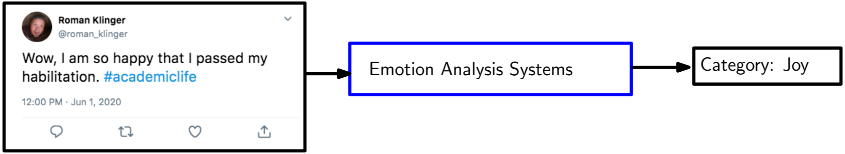
3 Multimodal Data Acquisition – Data Representativeness

4 Deception – Intentions

5 Coping – Role Playing

6 Take Home

Emotion Analysis: What we want to do.



Emotion Examples

Which emotion was felt by the author of the examples?

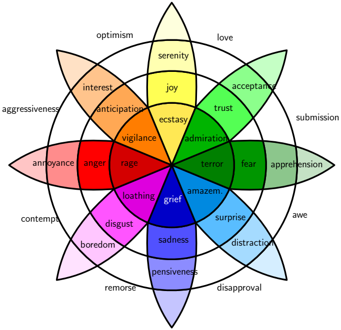
How did you recognize that?

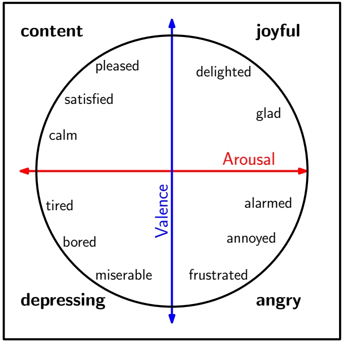
- “She became angry.”
- “A tear was running down my face.”
- “Their dog ran towards me quickly.”

With this exercise, we discussed:

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

How to define a categorical system of emotions?

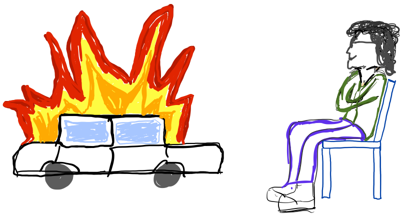




Definition of Emotions: Components

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	

Components

Fear

Name

Cognitive Appraisal in Scherer's Component Process model

Event →

Relevance	Implication	Coping	Normative Significance
Novelty	Causality: agent	Control	Internal standards
Intrinsic Pleasantness	Goal conduciveness	Adjustment	External standards
Goal Relevance	Outcome probability	Power	
	Urgency		
	Causality motive		
	Expectation discrepancy		

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

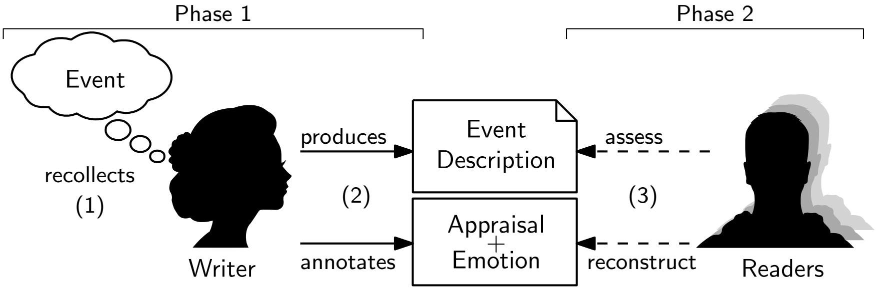
Research Questions

- Can appraisals and emotions be annotated reliably by external annotators?

E. Troiano, L. Oberländer, et al. (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	Agreement: Emotion F1 G–V	Emotion F1 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 (though 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	△
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.

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

Outline

1 Motivation

2 Emotion Analysis – Annotator's Label Reconstruction

3 Multimodal Data Acquisition – Data Representativeness

4 Deception – Intentions

5 Coping – Role Playing

6 Take Home

Introduction

- Synthetic 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

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

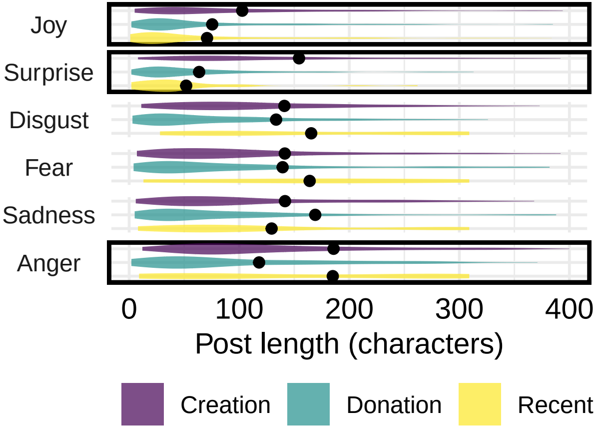
Trump supporters say ear bandages are 'sign of love'

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.

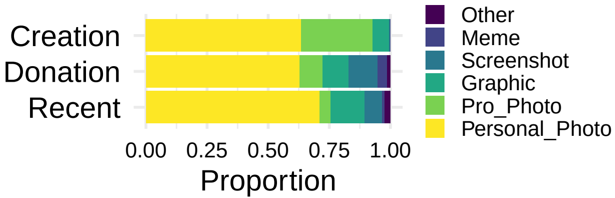
Members of the RNC's Arizona delegation said they were wearing the bandages as a sign of solidarity with the former president after he survived an assassination attempt.

Recent post labeled as anger.

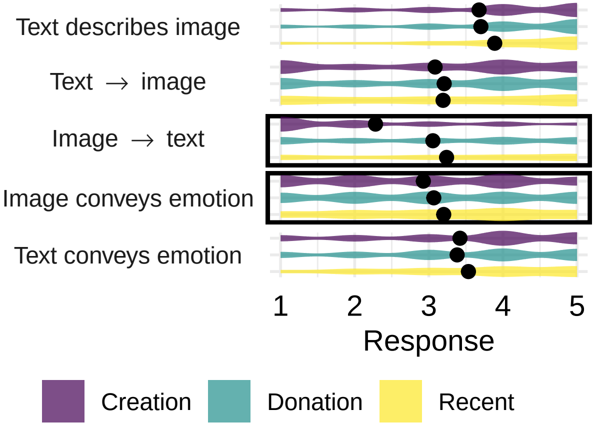
Are the subcorpora comparable? – Post Length



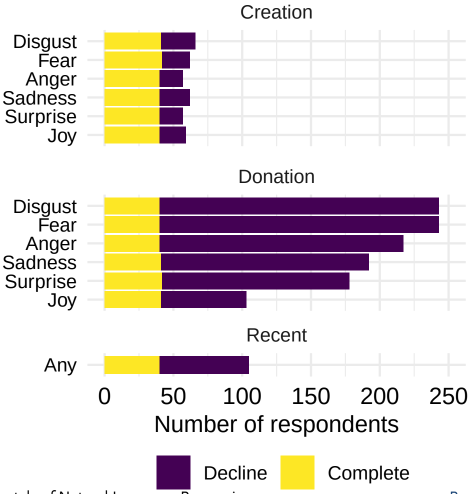
Are the subcorpora comparable? – Image Type



Are the subcorpora comparable? – Text–Image Relation



Are the subcorpora comparable? – Participant acceptance



Are the differences a problem? (ongoing work)

Experiment (Text only, work in progress)

- Fine-tune RoBERTa on Creation/Donation subsets
- Test on Creation/Donation, zero-shot predictions (minicpm-v)

Results

- Training on Creation: performance on Creation is higher (F score .55 vs. .40)
- Training on Donation: performance on Donation is lower (F score .50 vs. .39)
- ⇒ Creation cannot generalize well to real data.
- Zero-shot: Creation shows higher performance than Donation test data (.57 vs. .51)
- ⇒ Creation data result too optimistic!

Summary

- Training on synthetic data is not the best approach
- Testing zero-shot predictors not realistic
- Participating in Creation data is more accepted

Outline

1 Motivation

2 Emotion Analysis – Annotator's Label Reconstruction

3 Multimodal Data Acquisition – Data Representativeness

4 Deception – Intentions

5 Coping – Role Playing

6 Take Home

Deception

Deception

The term "deception" refers to the intentional act of causing someone to hold a false belief, which the deceiver knows to be false or believes to be untrue.

Examples: Lies, exaggerations, omissions

A. Velutharambath, A. Wührl, et al. (2024). "Can Factual Statements be Deceptive? The DeFaBel Corpus of Belief-based Deception". In: LREC-COLING

Linguistic Cues of Deception

- Deceptive statements have fewer self-references
- More ambiguous statements
- Longer sentences, more details
- Readability is lower

Cross-Corpus Deception Detection

Dataset	Domain	Truthful	Deceptive	Total	TC	SC
Bluff the listener (BLUFF)	game	251 (33.3%)	502 (66.7%)	753	241.66	11.5
Diplomacy dataset (DIPLOMACY)	game	16402 (94.9%)	887 (5.1%)	17289	24.53	1.7
Mafiascum dataset (MAFIASCUM)	game	7439 (76.9%)	2237 (23.1%)	9676	4690.69	362.8
Multimodal Decep. in Dialogues (BOXOFLIES)	game	101 (20.2%)	400 (79.8%)	501	12.2	1.6
Miami University Decep. Detection Db. (MU3D)	interview	160 (50.0%)	160 (50.0%)	320	131.7	5.7
Real-life trial data (TRIAL)	interview	60 (49.6%)	61 (50.4%)	121	79.85	3.9
Cross-cultural deception (CROSSCULTDE)	opinion	600 (50.0%)	600 (50.0%)	1200	80.0	4.5
Deceptive Opinion (DECOP)	opinion	1250 (50.0%)	1250 (50.0%)	2500	65.56	4.0
Boulder Lies and Truth Corpus (BLTC)	review	1041 (69.8%)	451 (30.2%)	1492	116.92	6.5
Deception in reviews (DEREV2014)	review	118 (50.0%)	118 (50.0%)	236	145.22	6.7
Deception in reviews (DEREV2018)	review	1552 (50.0%)	1552 (50.0%)	3104	176.6	8.1
Deceptive opinion spam (OPSPAM)	review	800 (50.0%)	800 (50.0%)	1600	170.5	9.5
Online deceptive reviews (ONLINEDE)	review	101431 (85.9%)	16694 (14.1%)	118125	171.5	7.2
Open Domain Deception (OPENDOMAIN)	statement	3584 (50.0%)	3584 (50.0%)	7168	9.33	1.0
		134789 (82.1%)	29296 (17.9%)	164085	436.88	31.05

A. Velutharambath and R. Klinger (2023). "UNIDECOR: A Unified Deception Corpus for Cross-Corpus Deception Detection". In: WASSA

Cross-Corpus Deception Detection

Features	BLTC	Bluff	BoxOfLies	CrossCultDe	DecOp	DeRev2014	DeRev2018	Diplomacy	MafiaScum	MU3D	OnlineDe	OpenDomain	OpSpam	Trial
	Datasets													
Analytic	**.13**	−.04	**.12**	.01	.02	**−.25**	**.23**	**.02**	**−.02**	**.14**	**.10**	**.05**	**.15**	**.25**
Authentic	.03	−.05	.00	**.28**	**.22**	**.28**	**−.05**	**−.03**	−.02	.07	.00	**−.04**	**−.09**	−.09
BigWords	.02	.00	**.18**	.04	**.05**	**−.21**	**.24**	.01	−.01	**.18**	**−.01**	**.03**	**−.08**	.09
Clout	.00	.00	.02	**−.11**	**−.28**	**−.45**	.00	**.02**	**.02**	.03	**−.05**	.01	**.10**	**.26**
Cognition	**−.08**	**.17**	−.05	.02	**.07**	−.06	**−.13**	−.01	−.01	**−.17**	.00	**−.09**	**−.06**	**−.28**
GunningFog	**.18**	**−.21**	**.12**	**.21**	**.25**	.01	**.13**	**−.09**	**−.03**	−.04	**.13**	.02	.02	.06
Kincaid	**.18**	**−.21**	**.14**	**.2**	**.24**	.01	**.13**	**−.08**	**−.03**	−.04	**.13**	**.03**	.02	.06
Linguistic	**−.07**	**.10**	**−.15**	.04	**.10**	**.29**	**−.14**	**−.02**	**−.03**	**−.16**	**−.05**	**−.05**	**−.18**	−.08
Period	.01	**−.07**	.02	**−.11**	**−.18**	**.26**	**−.07**	.00	.00	.03	**.01**	**.03**	**.24**	−.06
Physical	.02	.03	**.15**	−.04	**−.16**	**−.25**	**.06**	.00	**.03**	.04	**−.15**	−.01	−.01	.06
WC	**.18**	**−.21**	.04	**.22**	**.25**	.02	**.13**	**−.10**	.01	−.04	**.13**	−.02	.02	.06
auxverb	**−.08**	**.12**	−.06	**−.08**	**−.09**	**.22**	**−.12**	−.01	**.02**	**−.15**	.00	**.03**	**−.08**	**−.21**
focusfuture	**−.09**	**.09**	−.02	−.04	**−.08**	**−.17**	**−.2**	−.01	.02	−.04	**.01**	**−.04**	**−.16**	.08
function	−.05	**.13**	−.03	.00	**.10**	**.25**	**−.06**	**−.04**	**−.03**	**−.15**	**−.03**	**−.05**	**−.23**	**−.23**
i	**−.06**	**−.15**	−.07	**.13**	**−.3**	**.39**	**−.16**	**−.05**	.02	−.01	**−.12**	**−.04**	**−.33**	−.13
shehe	.01	−.11	−.03	**−.15**	.00	**−.17**	**−.07**	.00	**−.04**	**−.14**	**.04**	**−.04**	−.01	**−.18**
verb	**−.11**	.07	**−.09**	**−.06**	**−.07**	**.16**	**−.26**	**−.02**	.00	**−.14**	**−.07**	−.01	**−.16**	−.14
you	**−.10**	**.17**	−.03	−.05	**−.07**	**−.19**	**−.23**	.01	**.03**	−.08	**−.05**	**−.05**	.01	−.05

- We cannot find a consistent property of deception across corpora.

Cross-Corpus Deception Detection

Within-corpus and cross-corpus results for RoBERTa

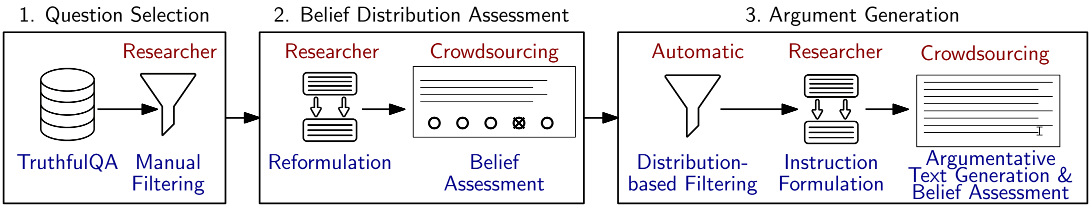
Tested on \ Trained on	Bluff	BoxOfLies	Diplomacy	Mafiascum	MU3D	Trial	CrossCultDe	DecOp	BLTC	DeRev2014	DeRev2018	OnlineDe	OpSpam	OpenDomain
Bluff	0.81	0.8	0	0	0.55	0.69	0.37	0.39	0.66	0.74	0.56	0.0077	0.29	0.69
BoxOfLies	0.41	0.89	0	0	0.62	0.82	0.82	0.81	0.89	0	0.4	0.019	0.6	0.51
Diplomacy	0.064	0.097	0	0	0.11	0.098	0.092	0.085	0.097	0.002	0.11	0.01	0.083	0.1
Mafiascum	0.37	0.38	0	0	0.3	0.36	0.12	0.09	0.37	0	0.37	0	0	0.28
MU3D	0.52	0.67	0	0	0.64	0.53	0.46	0.23	0.57	0.22	0.59	0.011	0.57	0.34
Trial	0.41	0.67	0	0	0.57	0.64	0.54	0.38	0.67	0.029	0.36	0	0.27	0.53
CrossCultDe	0.37	0.67	0	0	0.39	0.52	0.71	0.54	0.67	0.22	0.47	0.28	0.53	0.45
DecOp	0.36	0.67	0	0	0.16	0.52	0.74	0.8	0.67	0.34	0.31	0.34	0.55	0.56
BLTC	0.57	0.82	0	0	0.65	0.72	0.64	0.36	0.82	0.34	0.45	0.15	0.76	0.51
DeRev2014	0.54	0.67	0	0	0.4	0.61	0.23	0.068	0.67	0.89	0.77	0.075	0.68	0.12
DeRev2018	0.5	0.67	0	0	0.54	0.61	0.34	0.18	0.7	0.59	0.94	0.13	0.56	0.31
OnlineDe	0.17	0.24	0	0	0.21	0.22	0.24	0.15	0.26	0.25	0.31	0.75	0.38	0.16
OpSpam	0.53	0.67	0	0	0.49	0.55	0.42	0.14	0.66	0.15	0.48	0.16	0.91	0.44
OpenDomain	0.38	0.67	0	0	0.37	0.63	0.66	0.64	0.66	0.014	0.23	0.046	0.57	0.64

Model does not generalize across corpora.

Research Hypotheses

- Something is wrong here...
- We assume that model's mostly learn topic/domain specific properties of lies.

Belief-based Deception Framework and Corpus (DeFaBel)



A. Velutharambath, A. Wührl, et al. (2024). "Can Factual Statements be Deceptive? The DeFaBel Corpus of Belief-based Deception". In: LREC-COLING

"Wenn man einen Regenwurm durchschneidet, entstehen zwei Regenwürmer" – Who believed it?

Ein Regenwurm hat im Gegensatz zu ändern Tieren oder Säugetieren kein gehirn sondern ein dezentrales Nervensystem, welches seine Funktionen steuert. Ebenso hat er kein Herz oder andere singuläre Organe, die für ihr lebenswichtig sind. Verdauung, Atmung sind nicht an einen Ort gebunden. Das führt dazu, dass ein durchgeschnittener Regenwurm zwei Teile bildet, die unabhängig voneinander lebensfähig sind. Nach einer gewissen Zeit, wachsen an den Enden jeweils Schwanz/Kopf, die mit den ursprünglichen Enden des Wurm vergleichbar sind - es sind zwei neue, lebensfähige Regenwürmer entstanden.

Deceptive

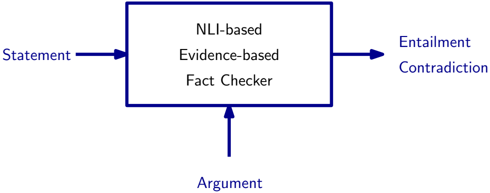
Schneidet man einen Regenwurm durch, so verdoppelt sich das Tier sozusagen, weil sich die beiden Hälften des durchgeschnittenen Wurmes zu eigenständigen Wesen entwickelt. Das liegt daran, dass der Regenwurm ein verblüffend komplexes Wesen ist. Er hat die Fähigkeit, seine inneren Organe, sein Herzkreislaufsystem und sein Gehirn bei Bedarf zu duplizieren. Das liegt in der Entwicklungsgeschichte des Regenwurms begründet. So nützlich er im Garten ist, so leicht wird er auch vom Menschen aus Versehen geteilt. Das weiß jeder Gärtner, der im Übereifer beim Jäten schon einmal einen Regenwurm geteilt hat. Der Regenwurm hat sich in seiner Evolution diesen tragischen Unfällen angepasst, indem er die Fähigkeit entwickelt hat, sich bei Bedarf aus zwei Hälften neu entstehen zu lassen. Praktisch, oder?

Non-Deceptive

Deception modeling in DeFaBel

Work in Progress:

- Current models do not recognize deception in this corpus
- We do not find the linguistic markers known to indicate deception in English
- But:
  Deceptive arguments are less suitable to
  fact-check the original statement than real arguments!



Deception Results

- Existing corpora lead to non-random classifiers
- Questionable if they actually model deception
- We propose a corpus in which authors do lie
  (but may change their opinion as part of the experiment)
- Models don't work...

Perhaps deception features do not hold in German?
Perhaps "established" deception detection methods don't actually do detect deception?
Perhaps something is wrong with our corpus.

Outline

1 Motivation

2 Emotion Analysis – Annotator's Label Reconstruction

3 Multimodal Data Acquisition – Data Representativeness

4 Deception – Intentions

5 Coping – Role Playing

6 Take Home

Classification of Coping Approaches

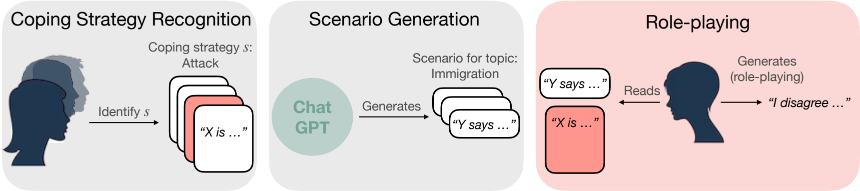
- Goal:
  - Develop a corpus of descriptions how people cope with challenging situations
- ...conditioned on
  - Personality types of coping strategies
  - Susceptibility with particular challenging triggers
- ⇒ Prefiltering participants for typically used coping strategy would be too costly.
- ⇒ We prefilter for susceptability and ask for role-playing the coping strategy.

E. Troiano, S. Labat, et al. (2024). "Dealing with Controversy: An Emotion and Coping Strategy Corpus Based on Role Playing". In: EMNLP Findings

Coping Strategies (Roseman, 2013)

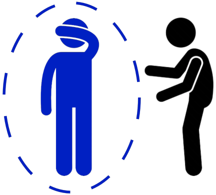
Coping Strategy	Emotions	Behavioral Function
Attack	Anger, frustration, guilt	Move against stimuli
Contact	Joy, hope, love, pride, relief	Increase contact and interaction with stimuli
Distance	Dislike, distress, fear, regret, sadness	Decrease contact and interaction with stimuli
Reject	Contempt, disgust, shame	Move stimuli away

Setup


Coping Strategy Recognition
Coping strategy s:
Attack
Identify s
"X is …"
Scenario Generation
Scenario for topic:
Immigration
Chat GPT
Generates
"Y says …"
Role-playing
"Y says …"
Reads
"X is …"
Generates
(role-playing)
"I disagree …"

Explaining Coping Strategies

Attack



Distance

Contact

Reject

Scenario example: cope by contact with racism

Definition	This character comes across as a calm, understanding, and very approachable person. For X, communication serves to unite people. It is an opportunity to exchange opinions, acknowledging the diversity of perspectives among individuals. When problems or unpleasant situations arise, this character responds with a constructive attitude. X expresses ideas with confidence, trying to solve problems in a respectful manner. X can effectively engage in discussions also with people having contrasting opinions.
Scenario (topic: racism)	During a university class discussion on historical racial events, Y confidently states, “People keep talking about systemic racism, but I believe that’s just an excuse for those who don’t want to work hard. If you look around, everyone has the same opportunities today.”
Generated reply	I understand your point Y, but it is not the case for everyone. Our group is a select handful of people who have been brought up this way.
Additional annotations	Description of X’s non-verbal behavior; rating of X’s emotional responses; comparison with own reaction.

Does the role playing work? (F_1, classification w/ DestilBERT/RoBERTa)

		Answer	Behavior
Topics	Abortion	.623	.630
	Drugs	.513	.488
	Immigration	.365	.483
	LGBTQ+	.508	.619
	Racism	.570	.457
Labels	Attack	.500	.545
	Contact	.647	.617
	Distance	.539	.560
	Reject	.408	.428

Yes.

Outline

1 Motivation

2 Emotion Analysis – Annotator's Label Reconstruction

3 Multimodal Data Acquisition – Data Representativeness

4 Deception – Intentions

5 Coping – Role Playing

6 Take Home

Summary

- We discussed cases of NLP tasks in which we need access to author labels.
- Data acquisition methods are similar to psychological experiments.
- Such creation methods are standard in psychology.
- Issues with experimental setups are known in that field, there is a tendency to move to the analysis of passively generated data.
- Therefore, NLP moves, to some degree, in the opposite direction than psychology.
- It is important to keep both fields in mind to join advantages of both approaches.

Take Home

We showed:

- Annotators can not always recreate author labels (emotion use case)
- Experimentally elicited synthetic instances differ from real posts (multimodal emotion).
- We (probably) can motivate participants to have an interest to act realistically (deception).
- In cases in which we do not have access to study participants with particular properties, we can ask them to mimic those via role playing (coping).

Important next research step:

- Systematic study of all these variables across multiple concepts.

Thank you for your attention.
Questions? Remarks?

?

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Thanks to

- Study 1: Enrica Troiano, Laura Oberländer
- Study 2: Christopher Bagdon, Aidan Combs, Carina Silberer
- Study 3: Aswathy Velutharambath, Kai Sassenberg
- Study 4: Enrica Troiano, Sofie Labat, Marco Antonio Stranisci, Rossana Damiano, Viviana Patti
- Generally: All of BamNLP (Bamberg) and IMS (Stuttgart)

University of Bamberg

Eliciting Textual Data from Psychological Study Participants

Oberseminar Computerlinguistik, January 30, 2025

Roman Klinger
roman.klinger@uni-bamberg.de

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