FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

Jena University Language & Information Engineering Lab

From Sentiment to Emotion:

Challenges of a More Fine-Grained Analysis of Affective Language

Sven Buechel

Jena University Language and Information Engineering (JULIE) Lab
Friedrich-Schiller-Universität Jena,
Jena, Germany

https://julielab.de

Slides: https://julielab.de/downloads/publications/slides/buechel_invited_ims_2018.pdf

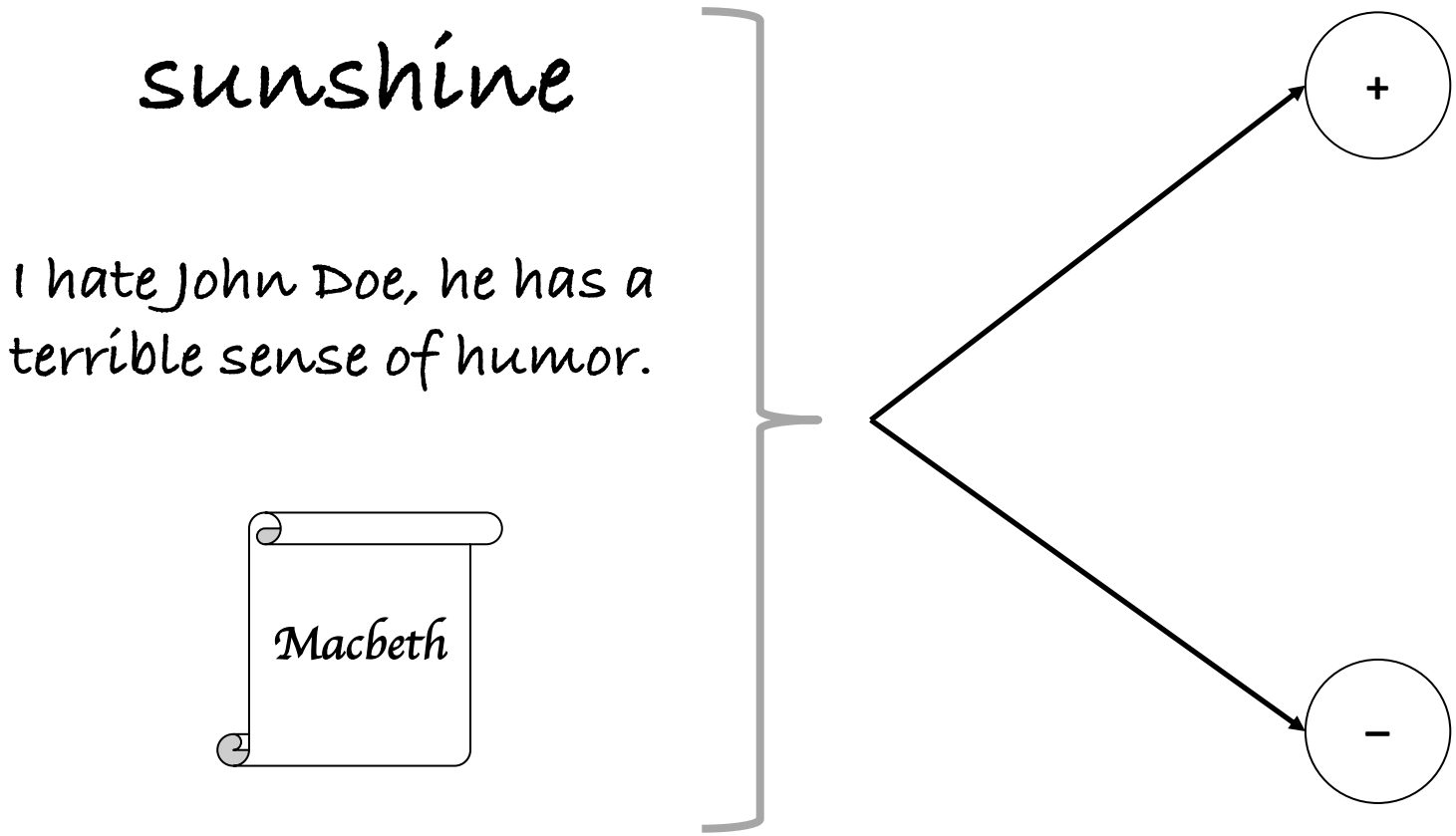
Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- Dealing with data sparsity
- Discussion and conclusion

Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- Dealing with data sparsity
- Discussion and conclusion

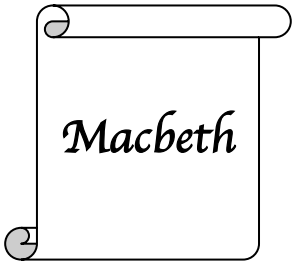
Sentiment Analysis — Two-Class Problem

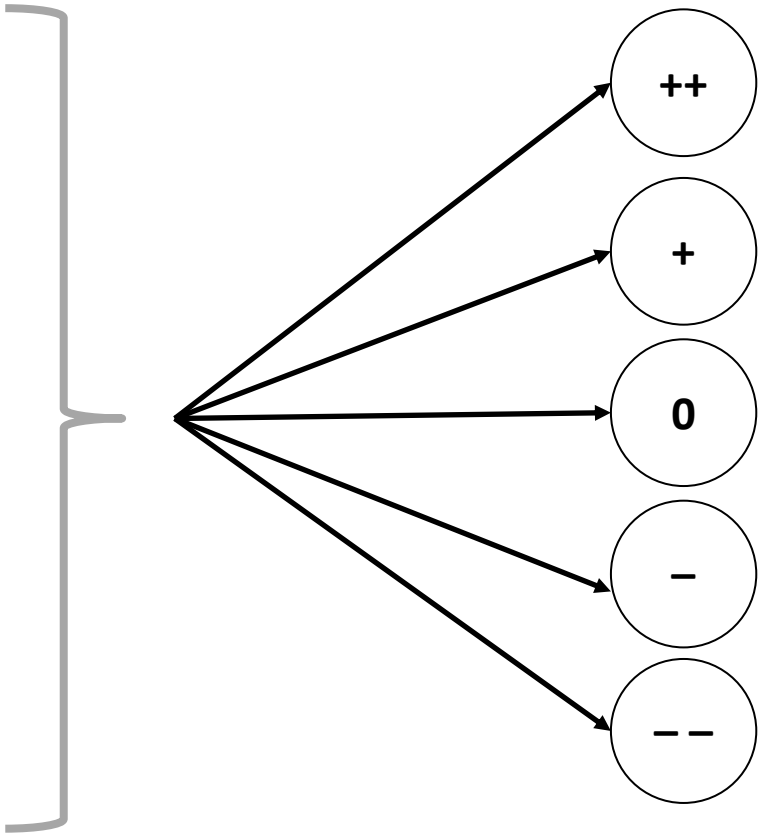


Sentiment Analysis — Multi-Class Problem

sunshine

I hate John Doe, he has a terrible sense of humor.

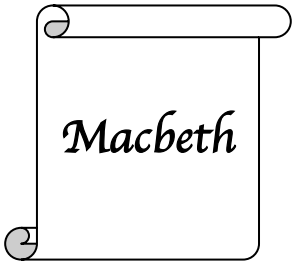


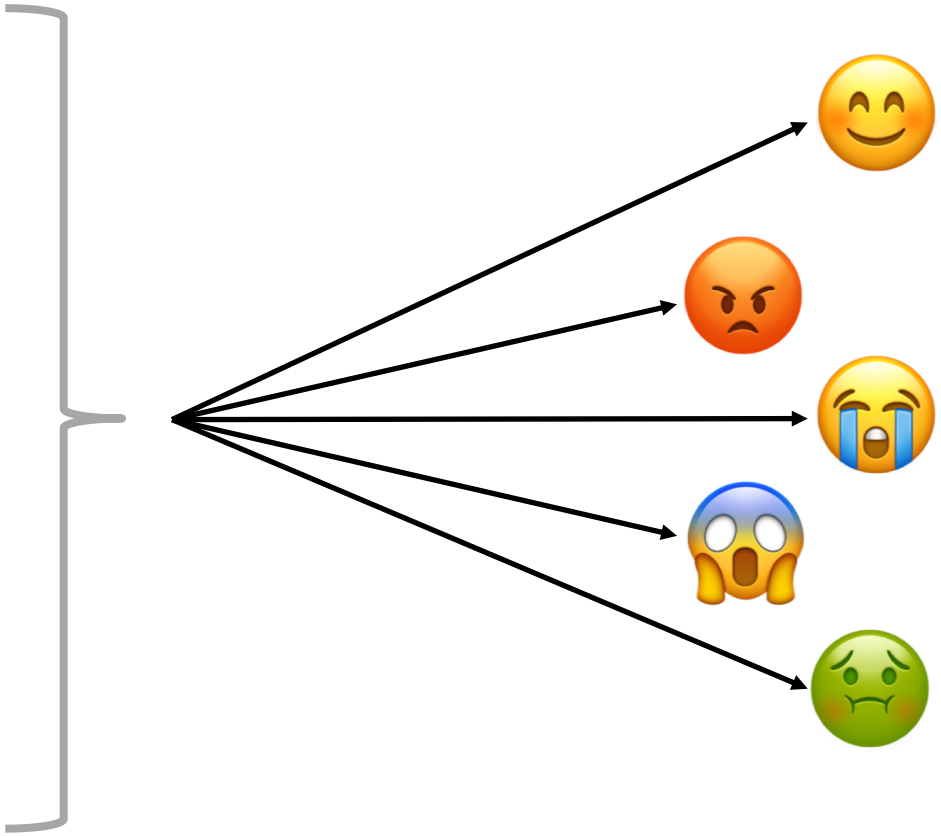


Emotion Analysis

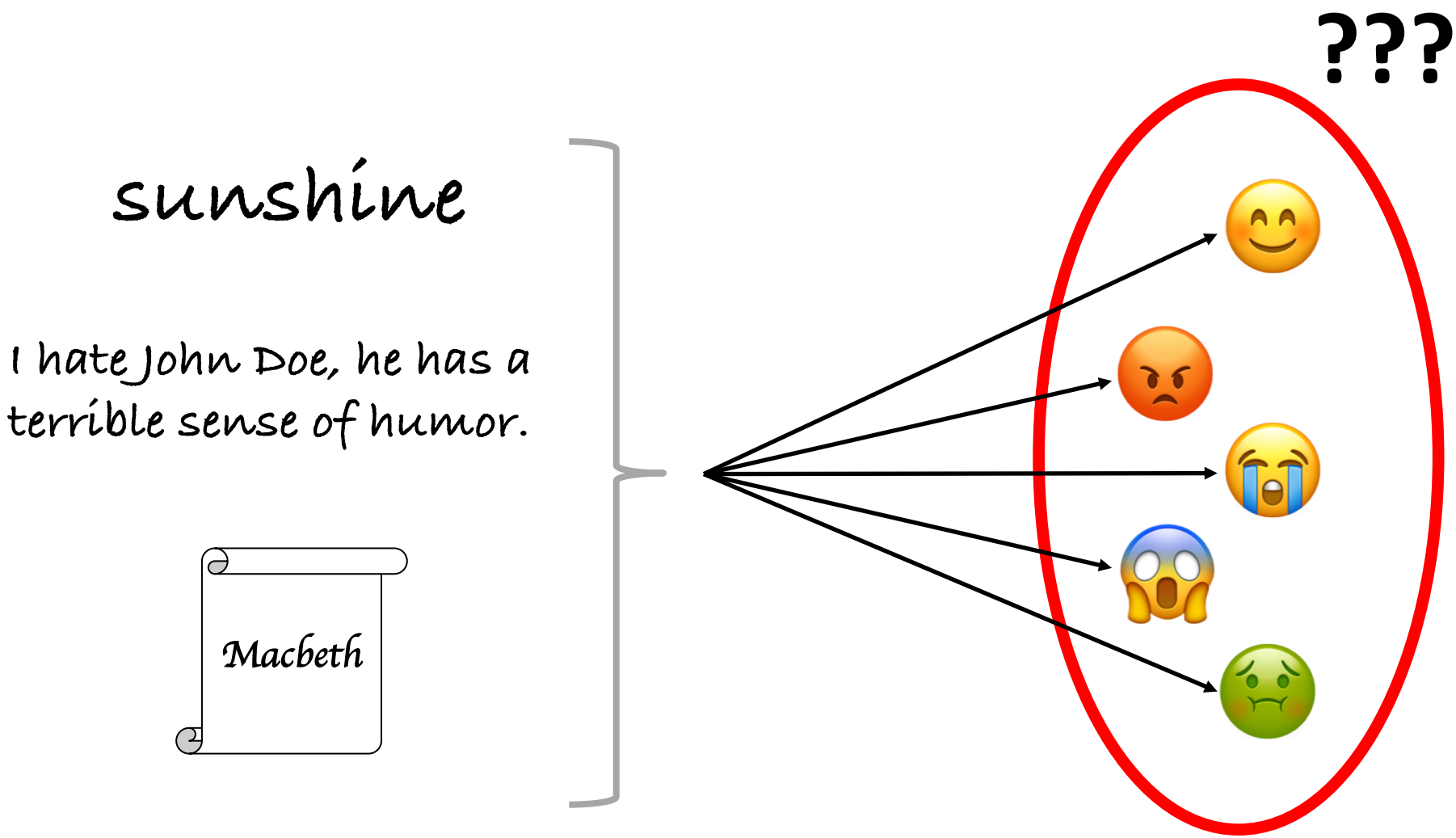
sunshine

I hate John Doe, he has a terrible sense of humor.





Emotion Analysis

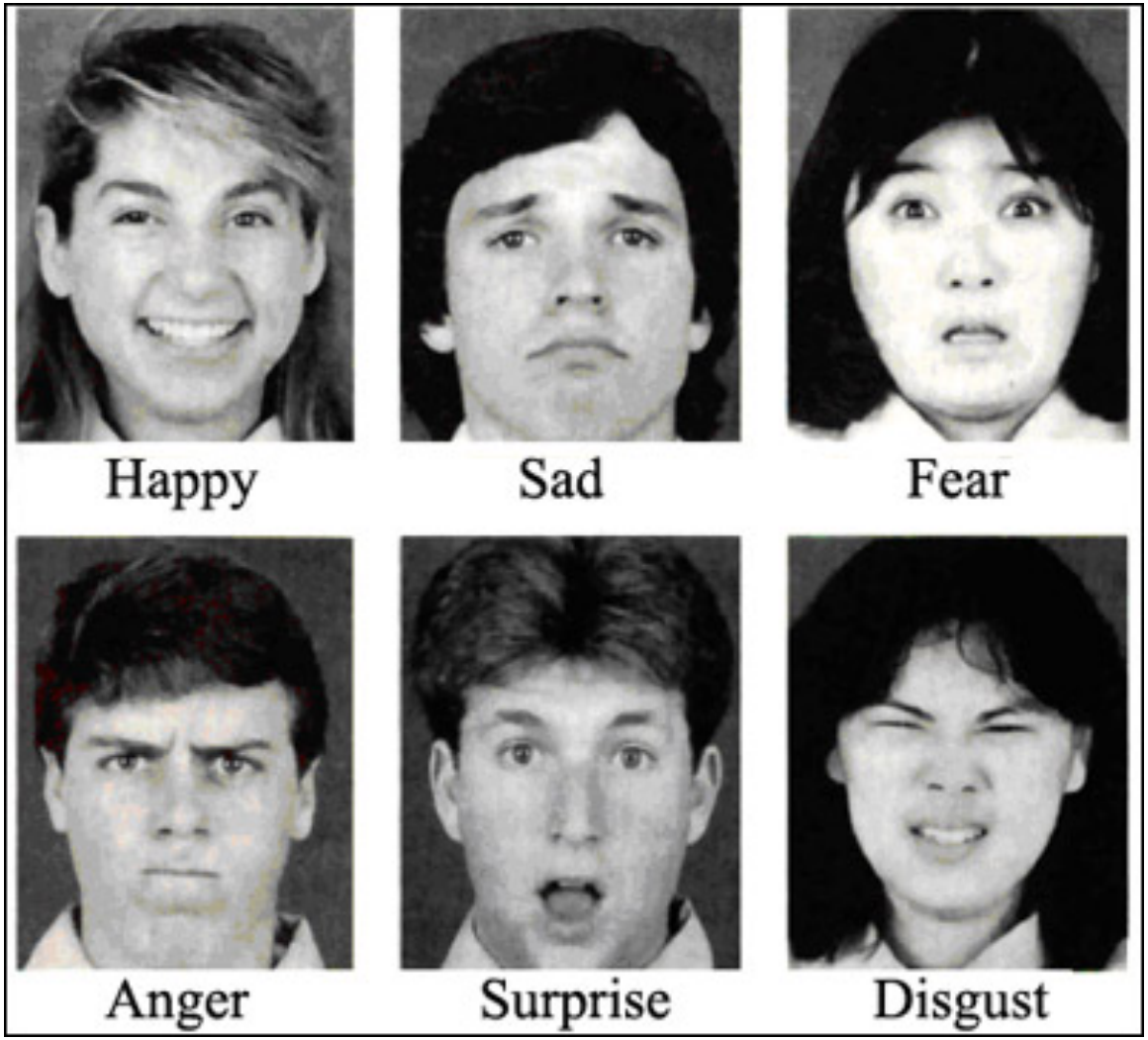


Major Approaches in Emotion Representation

Psychology Model

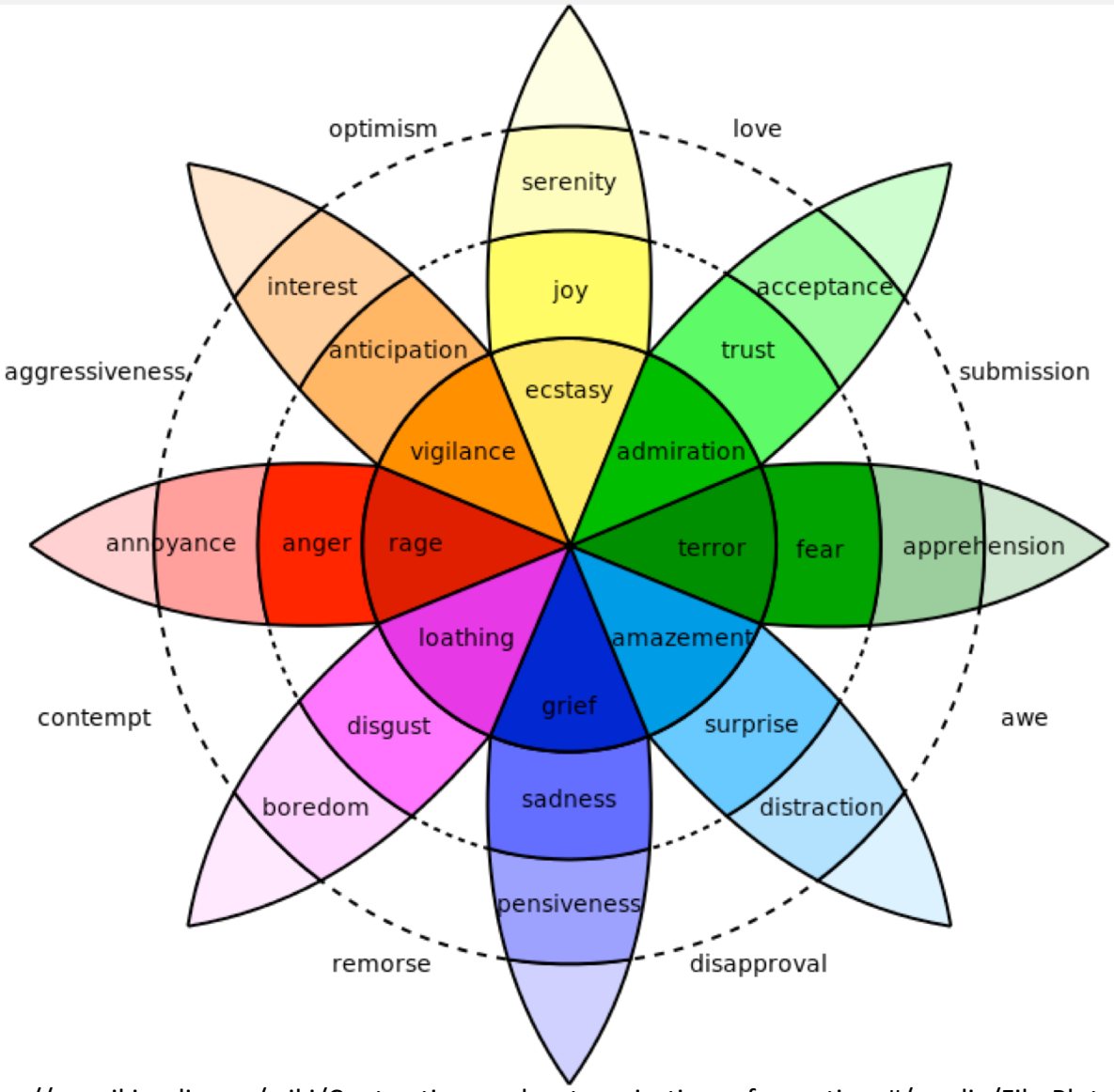
Prediction Problem	discrete emotions as instances	dimensional emotions as compositions
Regression		
Classification		

Ekman's Basic Emotions



Source: http://ocw.mit.edu/courses/brain-and-cognitive-sciences/9-00sc-introduction-to-psychology-fall-2011/emotion-motivation/discussion-emotion/

Representing Emotion — Wheel of Emotion



Source: https://en.wikipedia.org/wiki/Contrasting_and_categorization_of_emotions#/media/File:Plutchik-wheel.svg

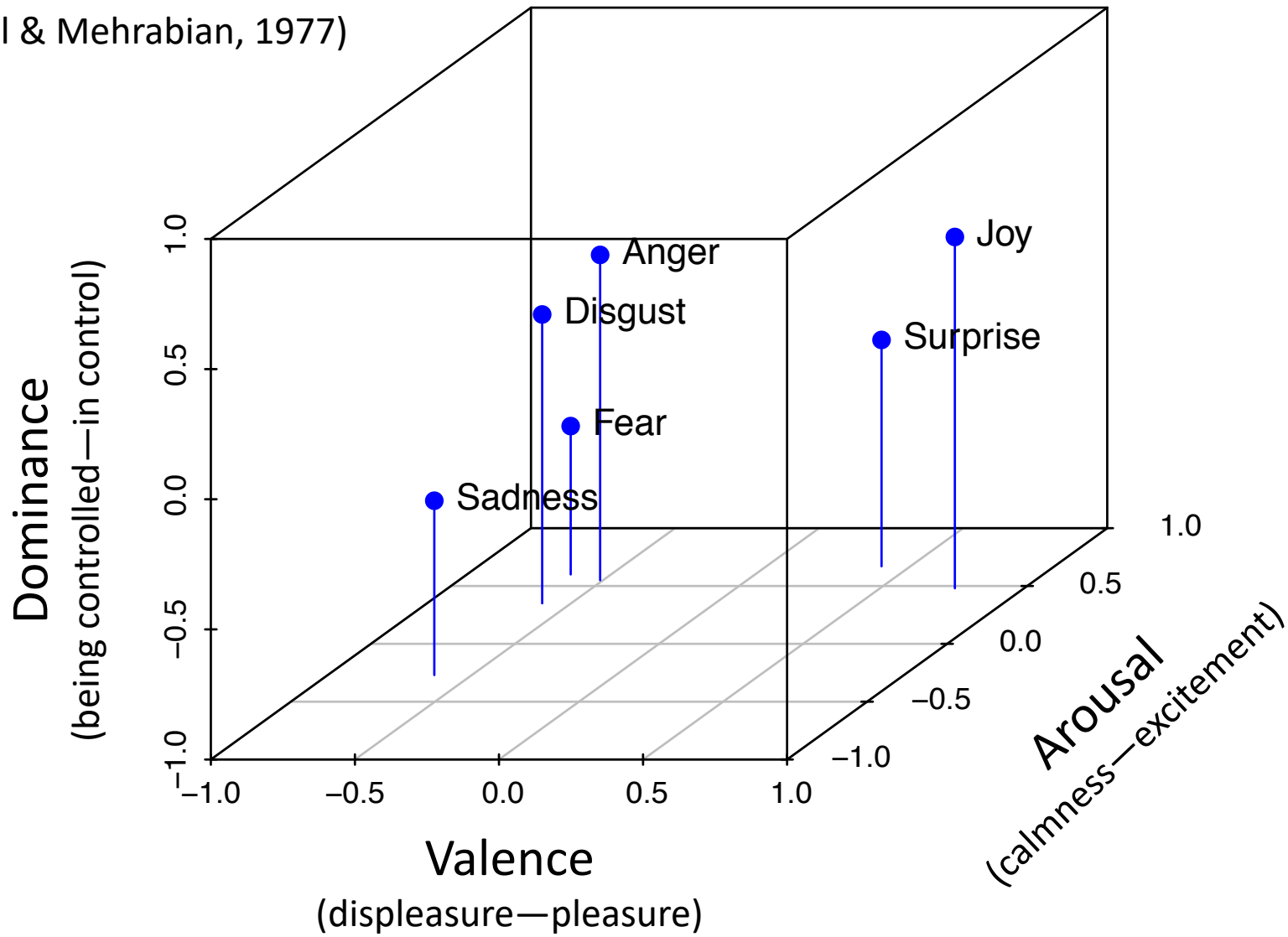
Major Approaches in Emotion Representation

Psychology Model

Prediction Problem	discrete emotions as instances	dimensional emotions as compositions
Regression	0.1 😭 0.2 😡 … 0.9 😊	
Classification	😊	

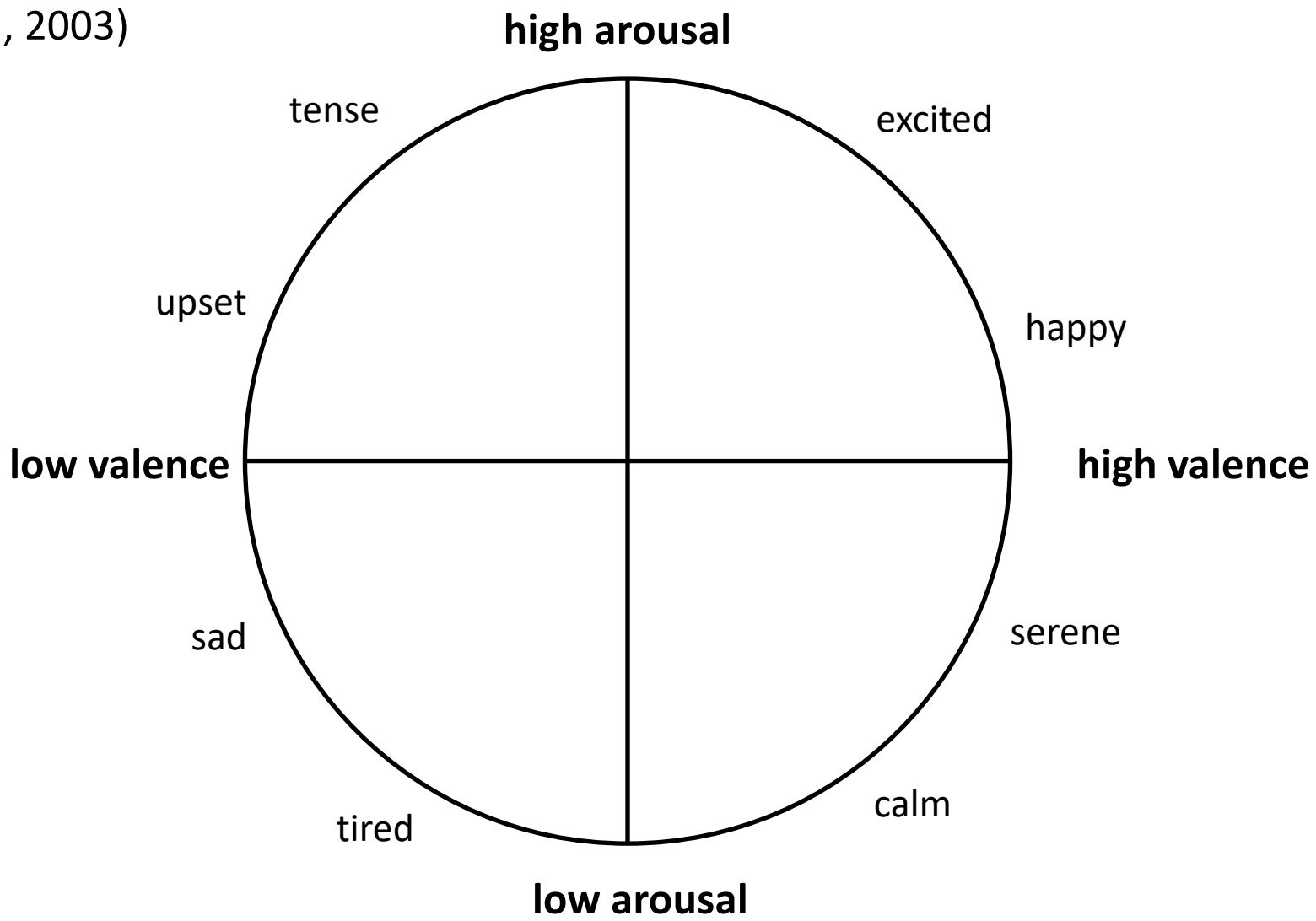
Valence-Arousal-Dominance

(Russell & Mehrabian, 1977)



Valence-Arousal

(Russell, 2003)



Major Approaches in Emotion Representation

Psychology Model

Prediction Problem	discrete emotions as instances	dimensional emotions as compositions
Regression	0.1 😭 0.2 😡 … 0.9 😊	0.8 Valence 0.6 Arousal - 0.3 Dominance
Classification	😊	+Valence +Arousal -Dominance

Major Approaches in Emotion Representation

Psychology Model

Prediction Problem	discrete emotions as instances	dimensional emotions as compositions
Regression	0.1 😭 0.2 😡 … 0.9 😊	0.8 Valence 0.6 Arousal - 0.3 Dominance
Classification	😊	+Valence +Arousal -Dominance

Current Situation in Emotion Analysis

- Huge interest
- Very messy
  - lack of agreed upon terminology
  - no consensus w.r.t. emotion representation
- Consequences
  - data sparsity
  - lack of interoperability of datasets, tools and analyses
- But getting better
  - shared tasks (SemEval 2018, 2019; WASSA 2017, 2018)
  - growing awareness of psychological work
  - work specifically aiming at enhancing interoperability e.g., Bostan & Klinger (COLING 2018); our own work

Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- Dealing with data sparsity
- Discussion and conclusion

Outline

- Introduction
- ➢ Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- Dealing with data sparsity
- Discussion and conclusion

Measuring Organizational Emotion

(Buechel et al., WASSA 2016)

- Collaboration with management and organization researchers
- Interest in anthropomorphic communication behavior of organizations (esp. targets, virtues, cognitive processes)
- Is this framework also applicable to emotions?
- Do enterprises communicate with a distinctive and persistent emotional profile?
- Analysis of annual reports and corporate social responsibility (CSR) reports

Annual Reports

The most important new model in 2002 was the Actros, which had its premiere at the International Auto Show (IAA) in Hanover and was well received by customers and automotive journalists. Its distinctive characteristics are its more powerful engines, a new axle and suspension concept, improved aerodynamics and a redesigned driver's cab.

Mercedes-Benz Vans still leads the field

The Mercedes-Benz Vans business unit sold 236,600 vehicles worldwide in 2002, nearly matching the figure for 2001. With a market share of 18% (2001: 19%) in the segment of 2 to 6 metric tons, Mercedes-Benz Vans is still the market leader in Western Europe. Whereas the Sprinter was able to maintain its strong market position in the heavy vans segment, in the segment of mid-size vans the market share of the Vito decreased due to the model changeover scheduled for 2003.

In the spring of 2002, DaimlerChrysler introduced the new Vaneo, which is positioned as a premium product in this segment.

The updated Sprinter model was introduced at the International Auto Show (IAA) in Hanover in September 2002. This new model is more attractive and, thanks to longer service intervals, more economical. Another new feature is the Electronic Stability Program (ESP). DaimlerChrysler is the first vehicle manufacturer to offer this system in this van segment. To strengthen its presence in the US van market in early 2003, DaimlerChrysler plans to offer the Sprinter, which has been sold successfully in the US under the Freightliner brand name since the middle of 2001, as a Dodge brand vehicle as well. We also plan to launch the Sprinter in Canada and Mexico.

The licensing agreement with Volkswagen AG for the production of the Sprinter van by Volkswagen was renewed to cover successor models as well.

The updated Mercedes-Benz Sprinter appeals with a new design and a world first. The Sprinter is the first van series worldwide for which all models can be supplied with the ESP electronic stability program.

Unit Sales 2002 [1]

	1,000 Units	02/01 in %
World	485	-2
of which: Vans [2]	246	-5
Trucks [3]	212	+3
Buses	25	-8
Unimogs	2	-23
Europe	287	-2
of which: Germany	103	-3
Western Europe (excluding Germany)	162	-5
of which: France	32	-10
United Kingdom	33	+14
Italy	23	+4
NAFTA	118	+11
of which: United States	100	+12
South America (excluding Mexico)	37	-14
of which: Brazil	30	-12
Asia	24	-8

1 Wholesale figures (including leased vehicles)
2 Including the Mitsubishi L200 pickup and the Mitsubishi Pajero in South Africa
3 Including schoolbuses by Thomas Built Buses and bus chassis by Freightliner

DaimlerChrysler, 2002

Annual Reports

The most important new model in 2002 was the Actros, which had its premiere at the International Auto Show (IAA) in Hanover and was well received by customers and automotive journalists. Its distinctive characteristics are its more powerful engines, a new axle and suspension concept, improved aerodynamics and a redesigned driver's cab.

Mercedes-Benz Vans still leads the field

The Mercedes-Benz Vans business unit sold 236,600 vehicles worldwide in 2002, nearly matching the figure for 2001. With a market share of 18% (2001: 19%) in the segment of 2 to 6 metric tons, Mercedes-Benz Vans is still the market leader in Western Europe. Whereas the Sprinter was able to maintain its strong market position in the heavy vans segment, in the segment of mid-size vans the market share of the Vito decreased due to the model changeover scheduled for 2003.

In the spring of 2002, DaimlerChrysler [illegible] new Vaneo, which is positioned as a prem[illegible] in this segment.

The updated Sprinter model was introd[illegible] International Auto Show (IAA) in Hanover [illegible] 2002. This new model is more attractive a[illegible] to longer service intervals, more economi[illegible] new feature is the Electronic Stability Prog[illegible] DaimlerChrysler is the first vehicle manuf[illegible] this system in this van segment. To strengthen [illegible] presence in the US van market in early 2003, DaimlerChrysler plans to offer the Sprinter, which has been sold successfully in the US under the Freightliner brand name since the middle of 2001, as a Dodge brand vehicle as well. We also plan to launch the Sprinter in Canada and Mexico.

The licensing agreement with Volkswagen AG for the production of the Sprinter van by Volkswagen was renewed to cover successor models as well.

The updated Mercedes-Benz Sprinter appeals with a new design and a world first. The Sprinter is the first van series worldwide for which all models can be supplied with the ESP electronic stability program.

The Mercedes-Benz Vans business unit sold 236,600 vehicles worldwide in 2002, nearly matching the figure for 2001. With a market share of 18% (2001: 19%) in the segment of 2 to 6 metric tons, Mercedes-Benz Vans is still the market leader in Western Europe. Whereas the

of which: Germany	103	- 3
Western Europe (excluding Germany)	162	- 5
of which: France	32	- 10
United Kingdom	33	+ 14
Italy	23	+ 4
NAFTA	118	+ 11
of which: United States	100	+ 12
South America (excluding Mexico)	37	- 14
of which: Brazil	30	- 12
Asia	24	- 8

1 Wholesale figures (including leased vehicles)
2 Including the Mitsubishi L200 pickup and the Mitsubishi Pajero in South Africa
3 Including schoolbuses by Thomas Built Buses and bus chassis by Freightliner

DaimlerChrysler, 2002

Corporate Social Responsibility (CSR) Reports

Jena University Language & Information Engineering Lab

BE BOLD. BE TRANSPARENT. BE A LEADER.

Our stakeholder engagement process has identified a number of issues that are important to society and our business, including several over-arching priorities for the future.

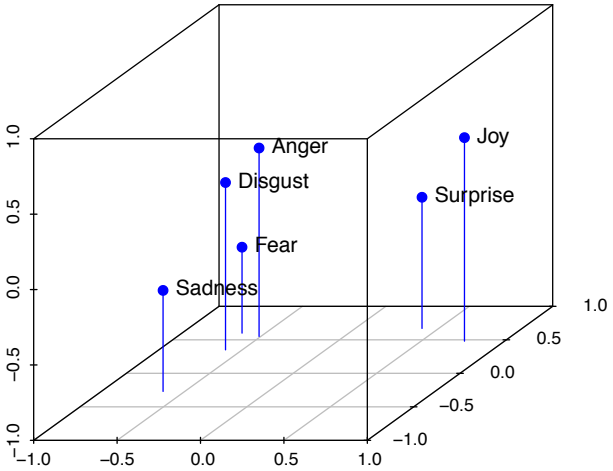
WHAT OUR STAKEHOLDER ADVISORY TEAM TOLD US	HOW IT IS INFLUENCING OUR JOURNEY
GOALS: SET LONG-TERM, MEASURABLE, AMBITIOUS GOALS.	• ESTABLISHED GOALS FOR THREE OF OUR CSR & SUSTAINABILITY PILLARS • DEVELOPING IMPROVED COLLECTION, CONSOLIDATION, AND REPORTING PROCESSES
FOOD: FOCUS ON FOOD AS THE PRIORITY.	• PARTNERED WITH THE ALLIANCE FOR A HEALTHIER GENERATION IN CONNECTION WITH THE CLINTON GLOBAL INITIATIVE TO COMMIT TO INCREASING CUSTOMERS' ACCESS TO FRUIT AND VEGETABLES AND HELPING FAMILIES MAKE INFORMED FOOD CHOICES • WORKING TO IMPROVE PRODUCT NUTRITIONALS • PROVIDING EASY ACCESS TO NUTRITION INFORMATION AND MORE CUSTOMER CHOICE
SOURCING: LEVERAGE McDONALD'S SCALE AND MARKET LEADERSHIP TO INFLUENCE CHANGE ON KEY ISSUES.	• WORKING THROUGH GLOBAL ORGANIZATIONS TO ADVANCE COLLECTIVE IMPROVEMENTS, SUCH AS SUSTAINABLE BEEF PRODUCTION • COLLABORATING WITH SUPPLIERS TO SUPPORT MORE SUSTAINABLE AGRICULTURE
PLANET: INCORPORATE CLIMATE CHANGE AND WATER RISK AVOIDANCE IN OUR STRATEGY.	• ESTIMATED SYSTEM-WIDE CARBON FOOTPRINT • DEVELOPED CLIMATE AND ENERGY CHANGE POSITION STATEMENT • INCLUDED WATER RISK IN ENVIRONMENTAL SCORECARD FOR SUPPLIERS • COMPLETED WATER STRESS MAPPING FOR ALL RESTAURANTS • DEVELOPING ENTERPRISE-WIDE WATER STRATEGY AND BEGINNING WATERSHED-LEVEL INVESTIGATIONS
PEOPLE: BE BOLDER IN OUR COMMUNICATIONS, BOTH INTERNALLY AND EXTERNALLY.	• CREATED A CSR & SUSTAINABILITY COMMUNICATIONS TEAM • ALIGNING WITH REGIONAL McDONALD'S COMMUNICATIONS PERSONNEL
COMMUNITY: LEVERAGE EMPLOYEES AS A KEY DRIVER OF COMMUNITY STRATEGY.	• PLANNING GLOBAL EMPLOYEE VOLUNTEER PROGRAM AND RECOGNITION INITIATIVE • CREATING COMMUNITY STRATEGY ALIGNED WITH BRAND AMBITION, CSR & SUSTAINABILITY FRAMEWORK

14

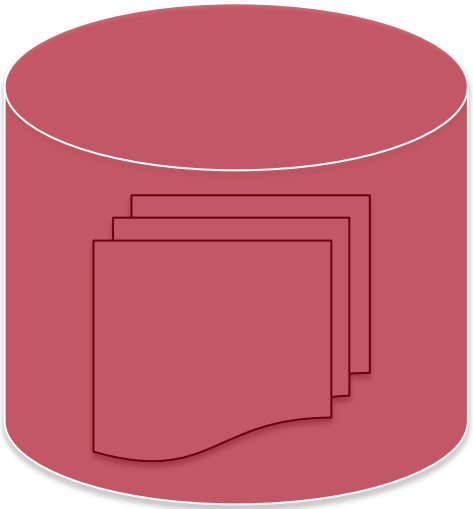
McDonald's 2012/13

Choosing an Emotion Representation

- Most of the documents are rather neutral
  - fine-grained, „high-resolution“
- Exploratory study
  - unclear what emotion categories are most relevant
- Social science application
  - interpretable outcome



Corpus Description



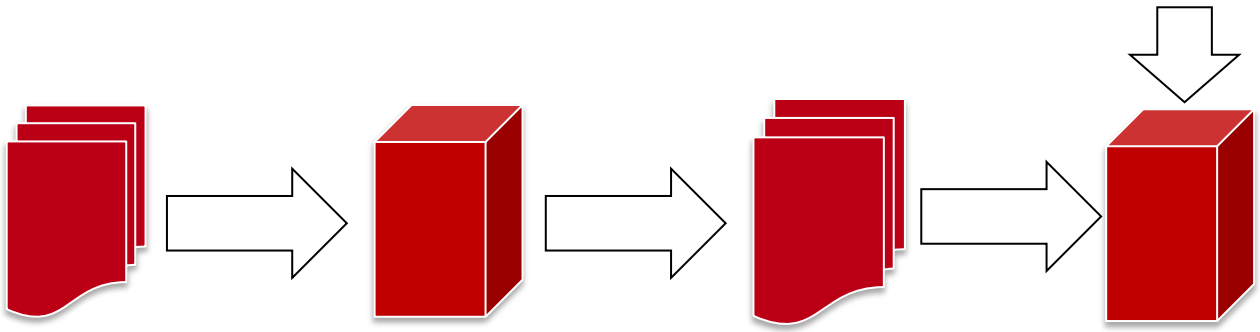
- Countries: US, UK, Germany
- 30 companies per country
  (DIJA, FTSE 100, DAX)
- 1676 documents (2/3 AR, 1/3 CSR)
- Years 1992–2015
- Successor: JOCo (Händschke et al., ECONLP @ ACL 2018)

Measuring Document Emotion: JEmAS

(Buechel & Hahn, ECAI 2016)

Available:
`https://github.com/JULIELab/JEmAS`

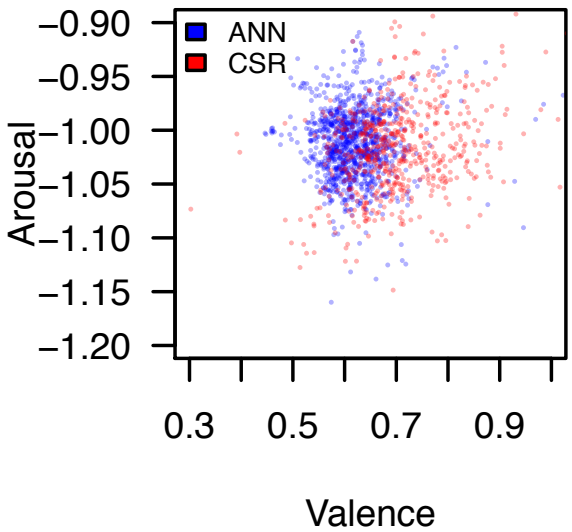
emotion lexicon

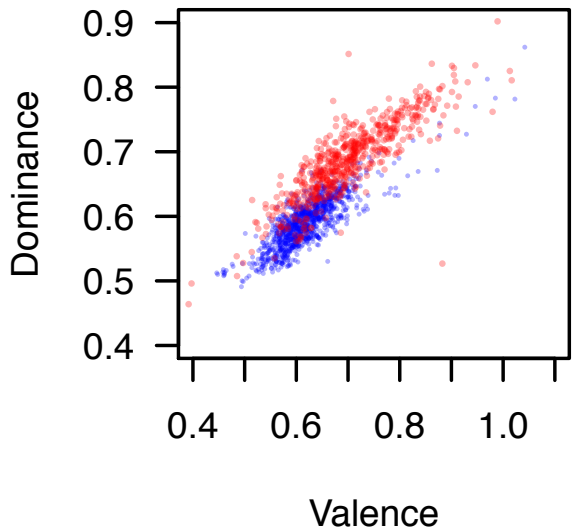


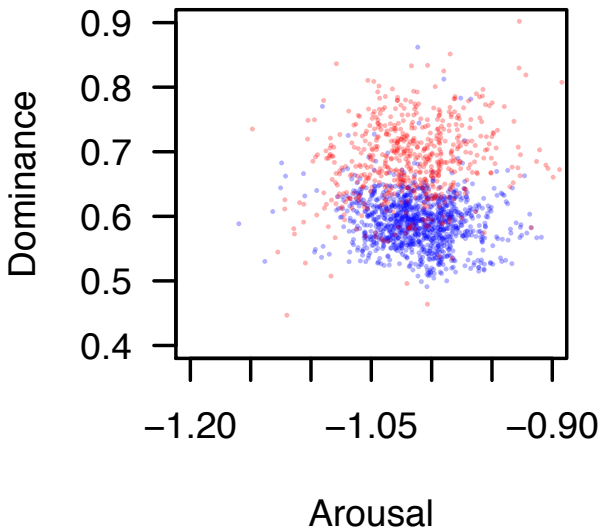
full text documents → linguistic normalization → BOW representation → VAD score calculation

Results — Annual vs. CSR Reports

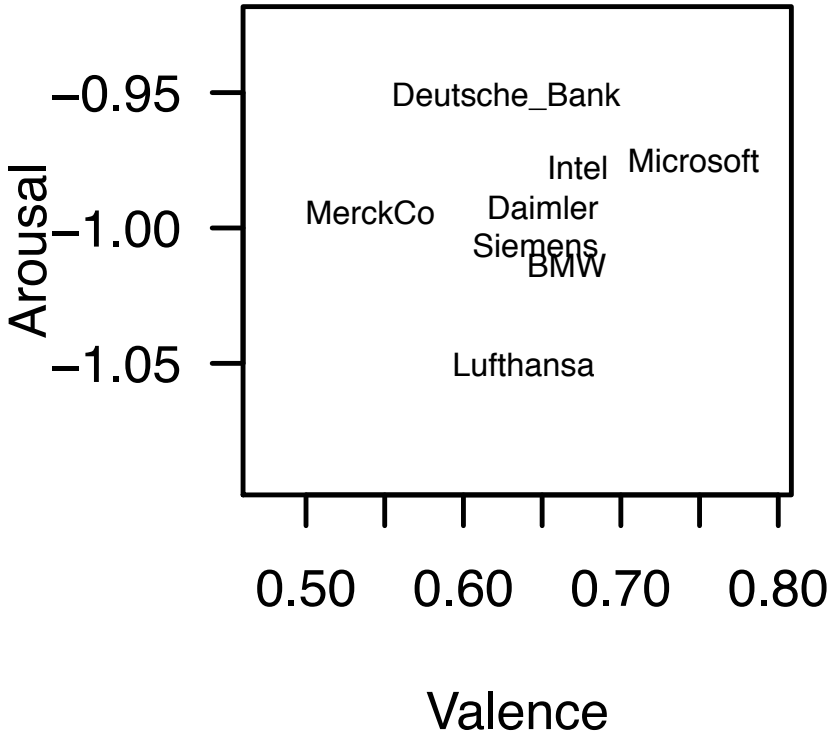
(Buechel et al., WASSA 2016)

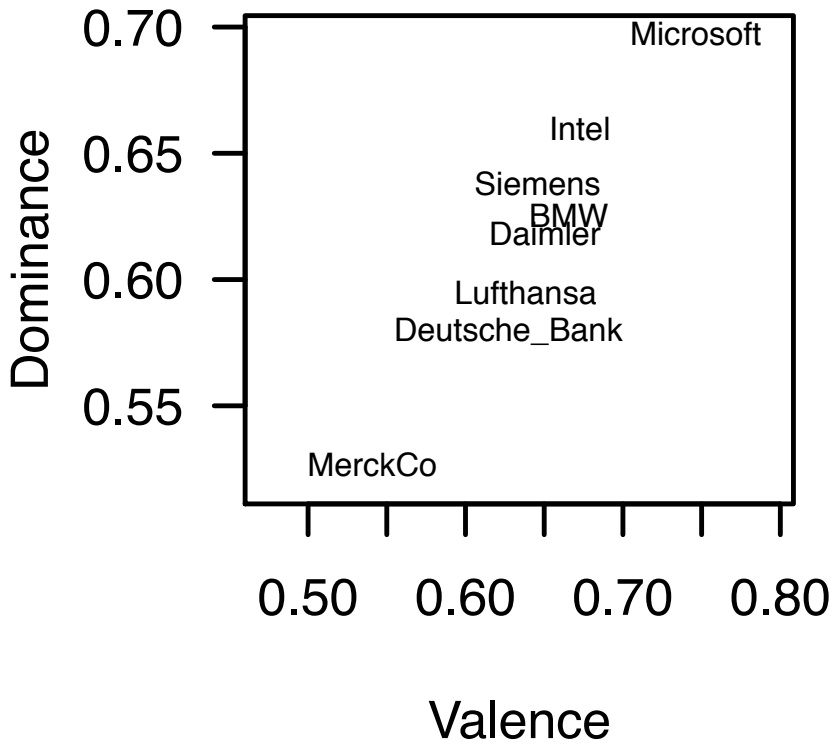






Results — Emotional Profiling of Organizations





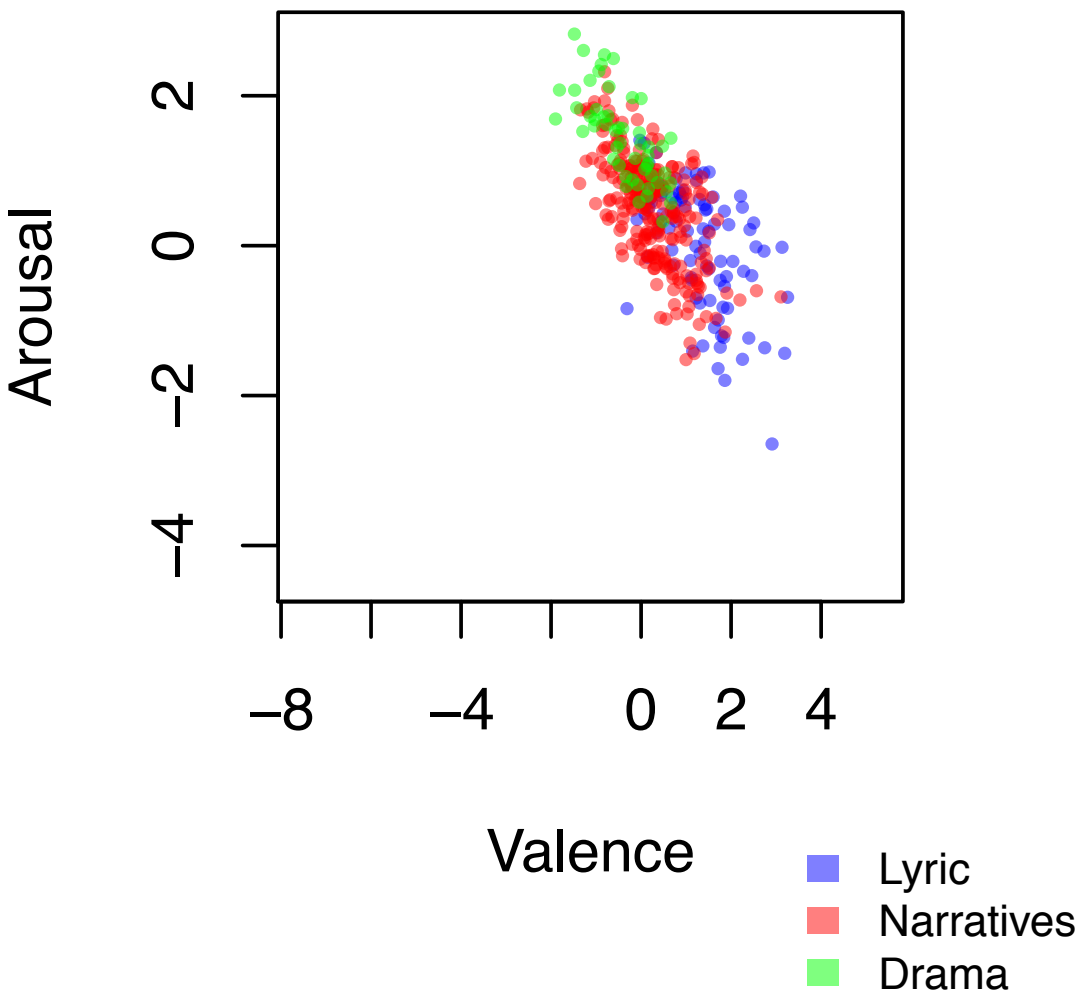
- Statistical analysis revealed that…
  - authoring company explains most of variability in VAD score
  - VAD scores are rather time invariant
- Companies have distinct and persistent emotional profile

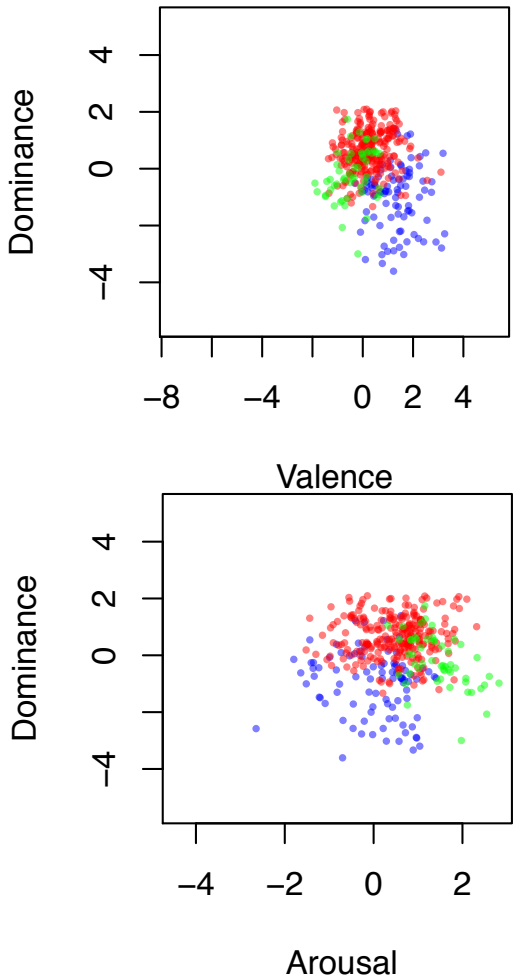
DH Application: Emotional Profiling in the DTA

Source and License: Charles Hackley via https://flic.kr/p/qSsjHA (CC-BY 2.0)

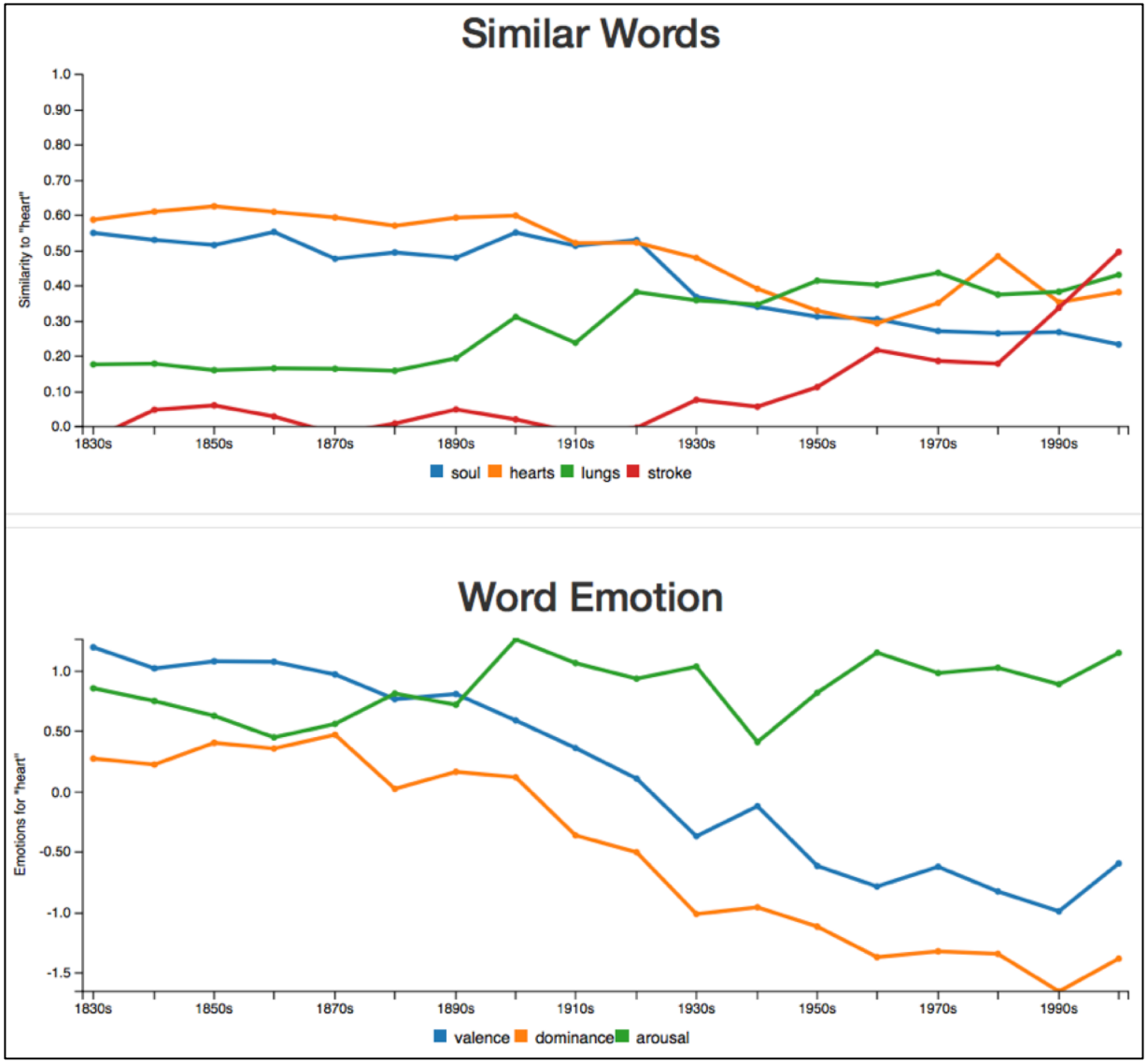
Emotional Profiles of Literary Forms in the DTA

(Buechel et al., LT4DH 2016, DH 2017)





Exploring Historical Word Emotions: *heart*



JeSemE.org

(Hellrich et al., COLING 2018)

Interim Conclusion

- Great potential of emotion analysis for DH and CSS
- Fine-grained representations more informative than polarity
- Quite simple methodologies

Outline

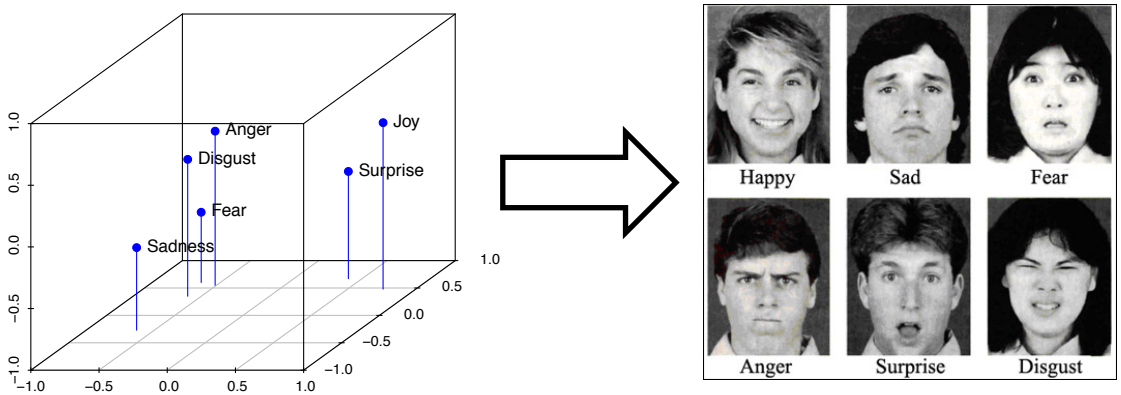
- Introduction
- ➢ Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- Dealing with data sparsity
- Discussion and conclusion

Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- ➢ Dealing with lack of interoperability
- Dealing with data sparsity
- Discussion and conclusion

Emotion Representation Mapping

- How to compare JEmAS against previous work?
- Basic idea: find a mapping that converts VAD to BE scores
- Also interesting for psych. theory: what is the relationship between discrete and dimensional emotion representations?
- Psychologist already created double annotated lexicons for this reason!



Emotion Representation Mapping

Word	Val	Aro	Dom	Joy	Ang	Sadn	Fear	Disg
sunshine	7.3	3.2	6.8	4.1	1.3	1.2	1.1	1.3
terrorism	1.8	8.1	4.2	1.4	4.1	3.2	3.8	3.6
earthquake	1.9	8.3	1.9	1.2	3.2	3.8	4.3	2.7

Emotion Representation Mapping

(Buechel & Hahn, ECAI 2016)

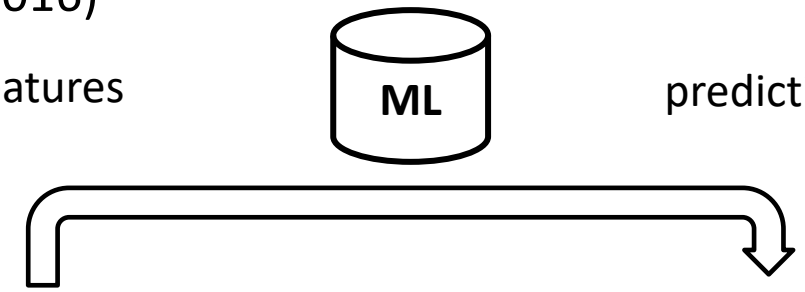
features ML prediction

Word	Val	Aro	Dom	Joy	Ang	Sadn	Fear	Disg
sunshine	7.3	3.2	6.8	4.1	1.3	1.2	1.1	1.3
terrorism	1.8	8.1	4.2	1.4	4.1	3.2	3.8	3.6
earthquake	1.9	8.3	1.9	1.2	3.2	3.8	4.3	2.7

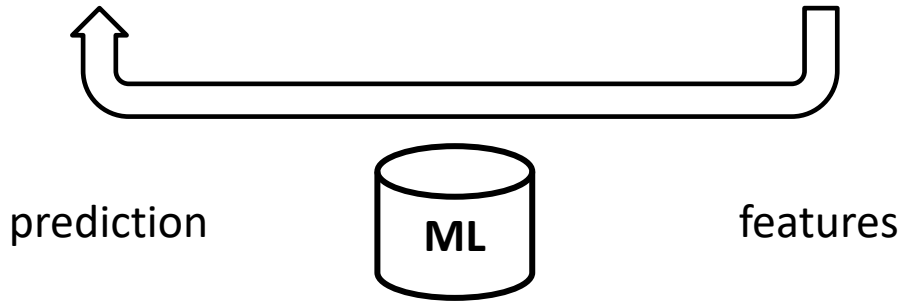
prediction ML features

Emotion Representation Mapping

(Buechel & Hahn, ECAI 2016)

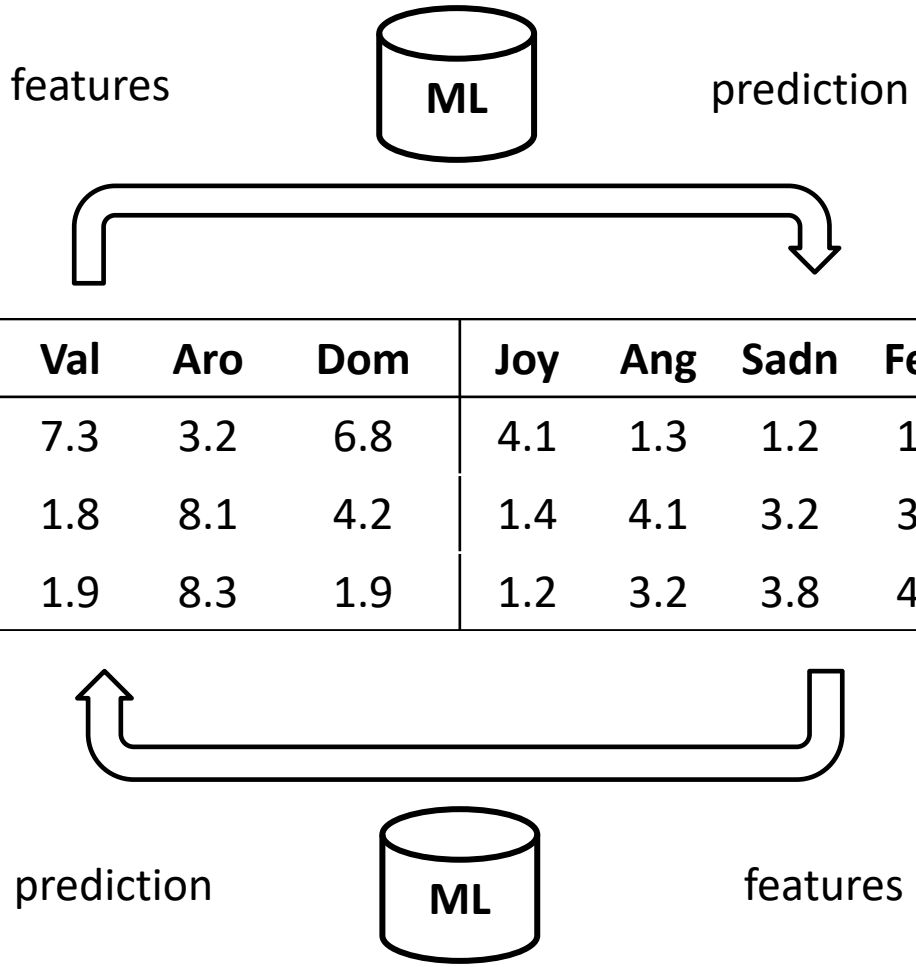


Word	Val	Aro	Dom	Joy	Ang	Sadn	Fear	Disg
sunshine	7.3	3.2	6.8	4.1	1.3	1.2	1.1	1.3
terrorism	1.8	8.1	4.2	1.4	4.1	3.2	3.8	3.6
earthquake	1.9	8.3	1.9	1.2	3.2	3.8	4.3	2.7



Map JEmAS output to BE — SOTA in three emotion categories!

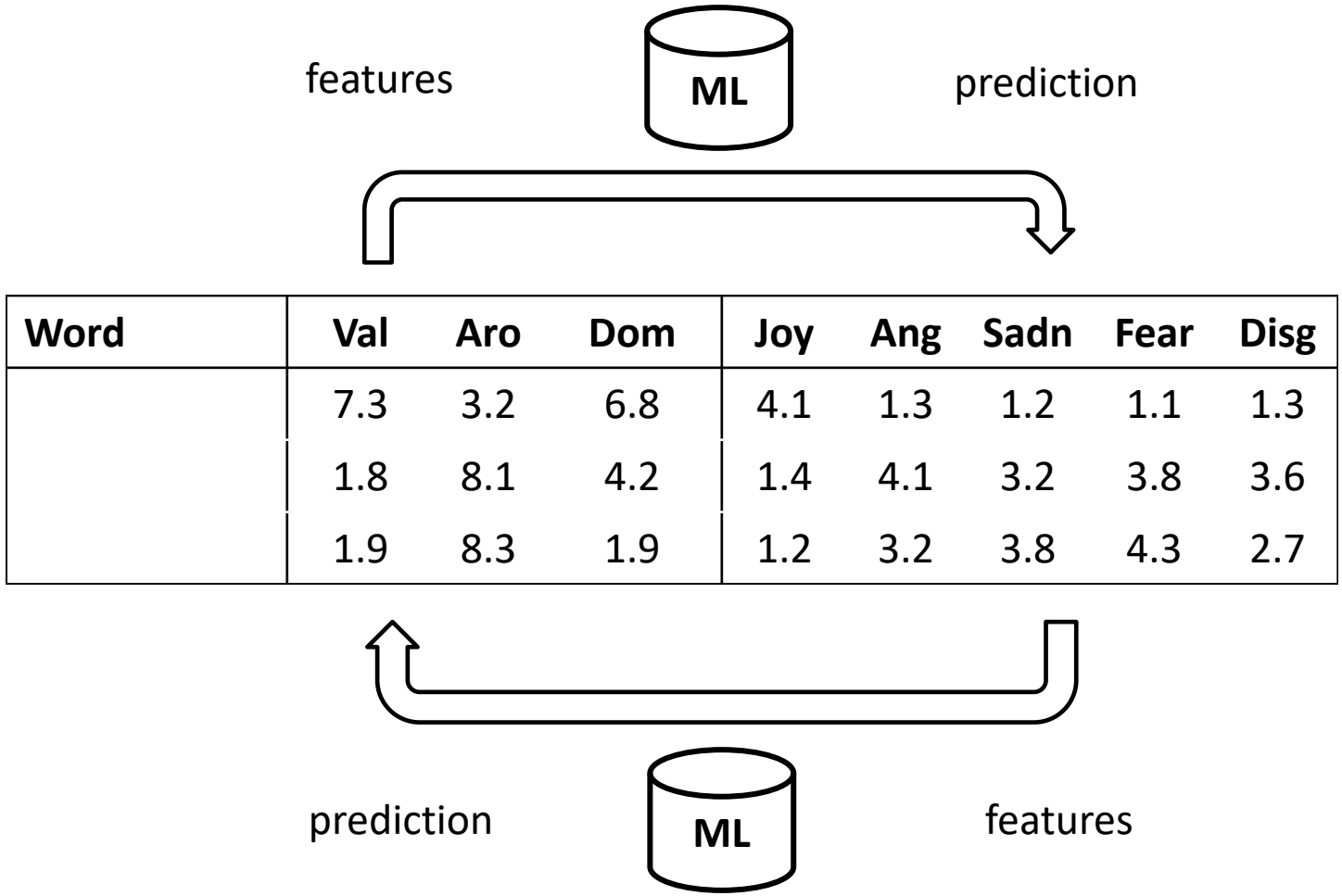
Crosslingual Application



Word	Val	Aro	Dom	Joy	Ang	Sadn	Fear	Disg
sunshine	7.3	3.2	6.8	4.1	1.3	1.2	1.1	1.3
terrorism	1.8	8.1	4.2	1.4	4.1	3.2	3.8	3.6
earthquake	1.9	8.3	1.9	1.2	3.2	3.8	4.3	2.7

(Buechel & Hahn, EACL 2017, CogSci 2017, LREC 2018)

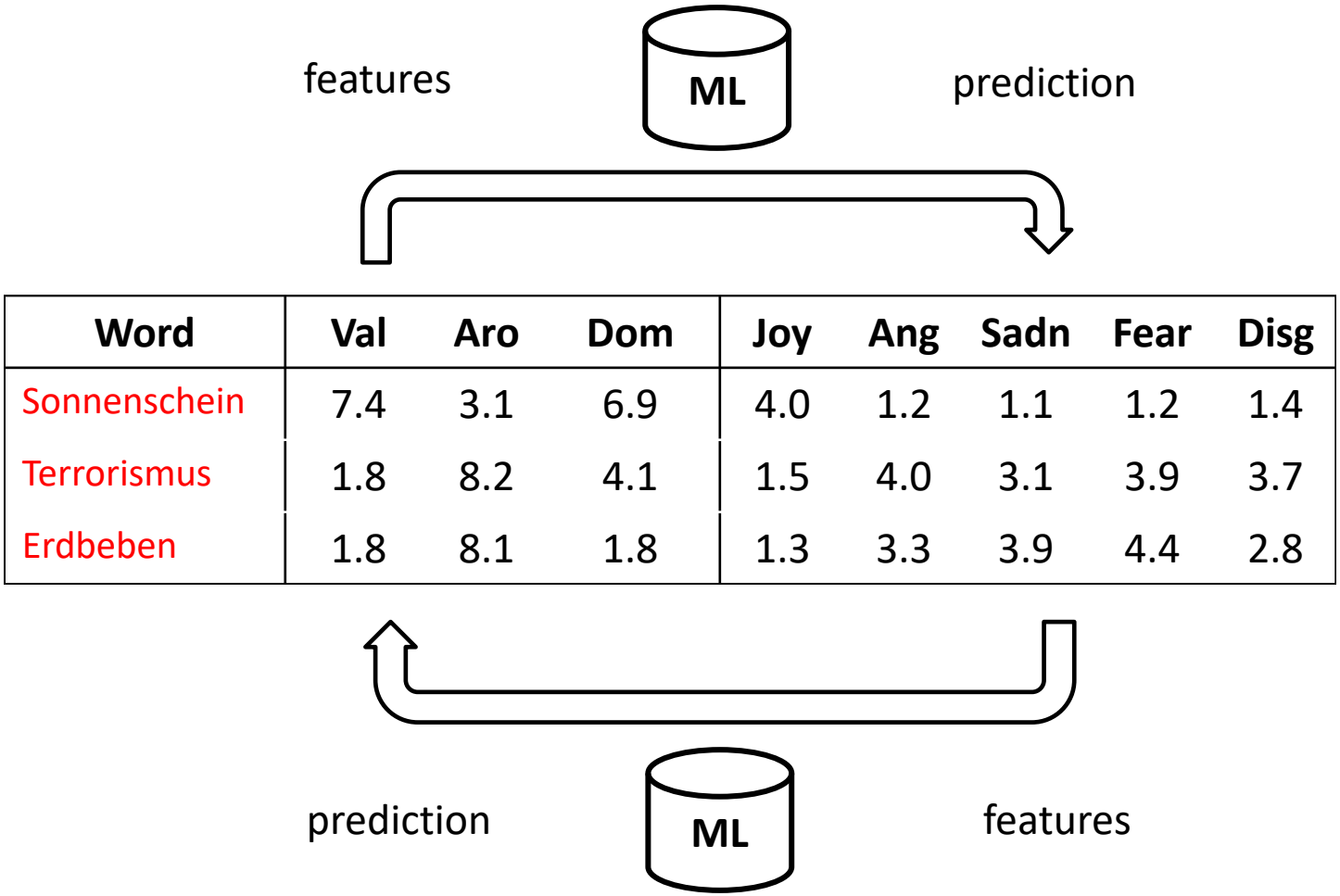
Crosslingual Application



Word	Val	Aro	Dom	Joy	Ang	Sadn	Fear	Disg
	7.3	3.2	6.8	4.1	1.3	1.2	1.1	1.3
	1.8	8.1	4.2	1.4	4.1	3.2	3.8	3.6
	1.9	8.3	1.9	1.2	3.2	3.8	4.3	2.7

(Buechel & Hahn, EACL 2017, CogSci 2017, LREC 2018)

Crosslingual Application



Word	Val	Aro	Dom	Joy	Ang	Sadn	Fear	Disg
Sonnenschein	7.4	3.1	6.9	4.0	1.2	1.1	1.2	1.4
Terrorismus	1.8	8.2	4.1	1.5	4.0	3.1	3.9	3.7
Erdbeben	1.8	8.1	1.8	1.3	3.3	3.9	4.4	2.8

(Buechel & Hahn, EACL 2017, CogSci 2017, LREC 2018)

Comparison against Human Reliability

- Collected 8 double-annotated pairs of datasets (en, es, de, pl)
- New technique to allow for standardized comparison against **split-half reliability**

(Buechel & Hahn, COLING 2018)

Comparison against Human Reliability

- Collected 8 double-annotated pairs of datasets (en, es, de, pl)
- New technique to allow for standardized comparison against **split-half reliability**

	r1	r2	r3	r4	r5	r6
i1						
i2						
i3						
i4						
i5						
i6						

(Buechel & Hahn, COLING 2018)

Comparison against Human Reliability

- Collected 8 double-annotated pairs of datasets (en, es, de, pl)
- New technique to allow for standardized comparison against **split-half reliability**

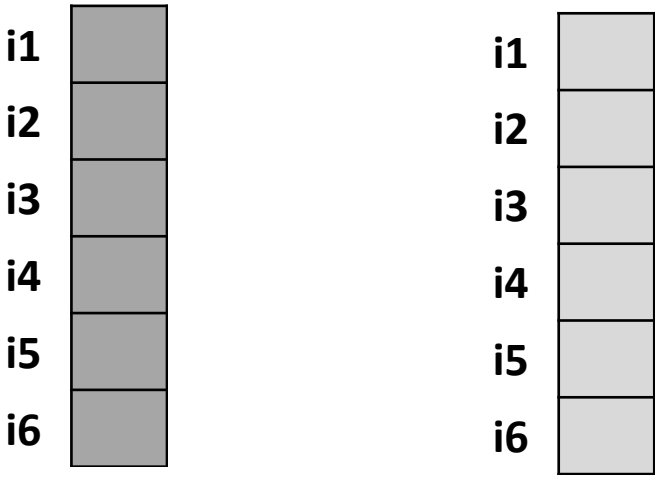
	r1	r4	r5
i1			
i2			
i3			
i4			
i5			
i6			

	r2	r3	r6
i1			
i2			
i3			
i4			
i5			
i6			

(Buechel & Hahn, COLING 2018)

Comparison against Human Reliability

- Collected 8 double-annotated pairs of datasets (en, es, de, pl)
- New technique to allow for standardized comparison against **split-half reliability**



(Buechel & Hahn, COLING 2018)

Comparison against Human Reliability

- Collected 8 double-annotated pairs of datasets (en, es, de, pl)
- New technique to allow for standardized comparison against **split-half reliability**
- *Does the model agree more with gold data than two random groups of ten people would agree with each other?*

➢ In over 50% of the cases (also in crosslingual setup): ***Yes!***

(Buechel & Hahn, COLING 2018)

Generating New Emotion Lexicons

- Identify VA(D) or BE lexicons which do not have complementary ratings for that language
- Apply models for prediction
- Gold quality

➢ New ratings for 13 languages, up to 13k entries each (en, es, de, pl, it, nl, pt, zh, id, fr, gr, fn, sv)

(Buechel & Hahn, COLING 2018)

Interim Conclusion II

- Multitude of competing emotion representation formats endangers interoperability
- Proposed *emotion representation mapping*
- Automatically converted ratings are as reliable as gold data

Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- ➢ Dealing with lack of interoperability
- Dealing with data sparsity
- Discussion and conclusion

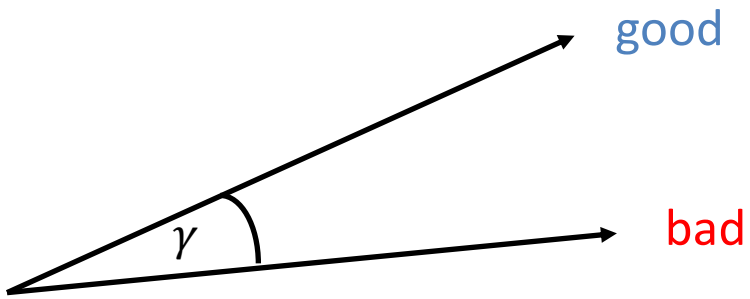
Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- ➢ Dealing with data sparsity
- Discussion and conclusion

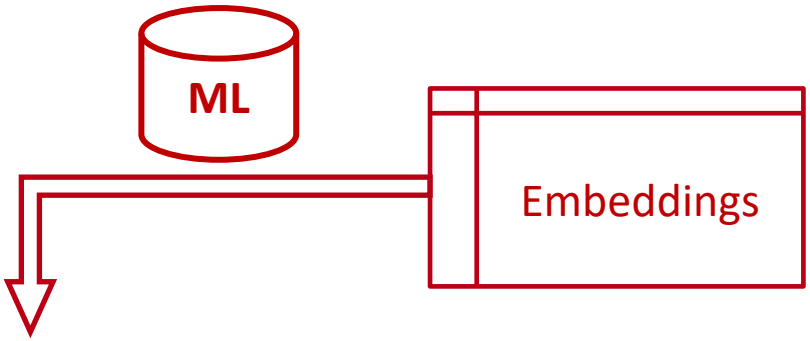
Two Popular Misconceptions about DL?

- Enormous data requirements
  - cf. WASSA 2017 shared task
- Insufficient affective information in pre-trained embeddings

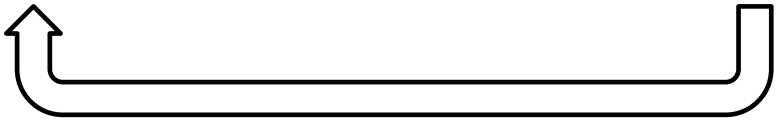
(Tang et al., 2014)



Word Emotion Induction



Word	Val	Aro	Dom	Joy	Ang	Sadn	Fear	Disg
sunshine	??	??	??	4.1	1.3	1.2	1.1	1.3
terrorism	??	??	??	1.4	4.1	3.2	3.8	3.6
earthquake	??	??	??	1.2	3.2	3.8	4.3	2.7



prediction

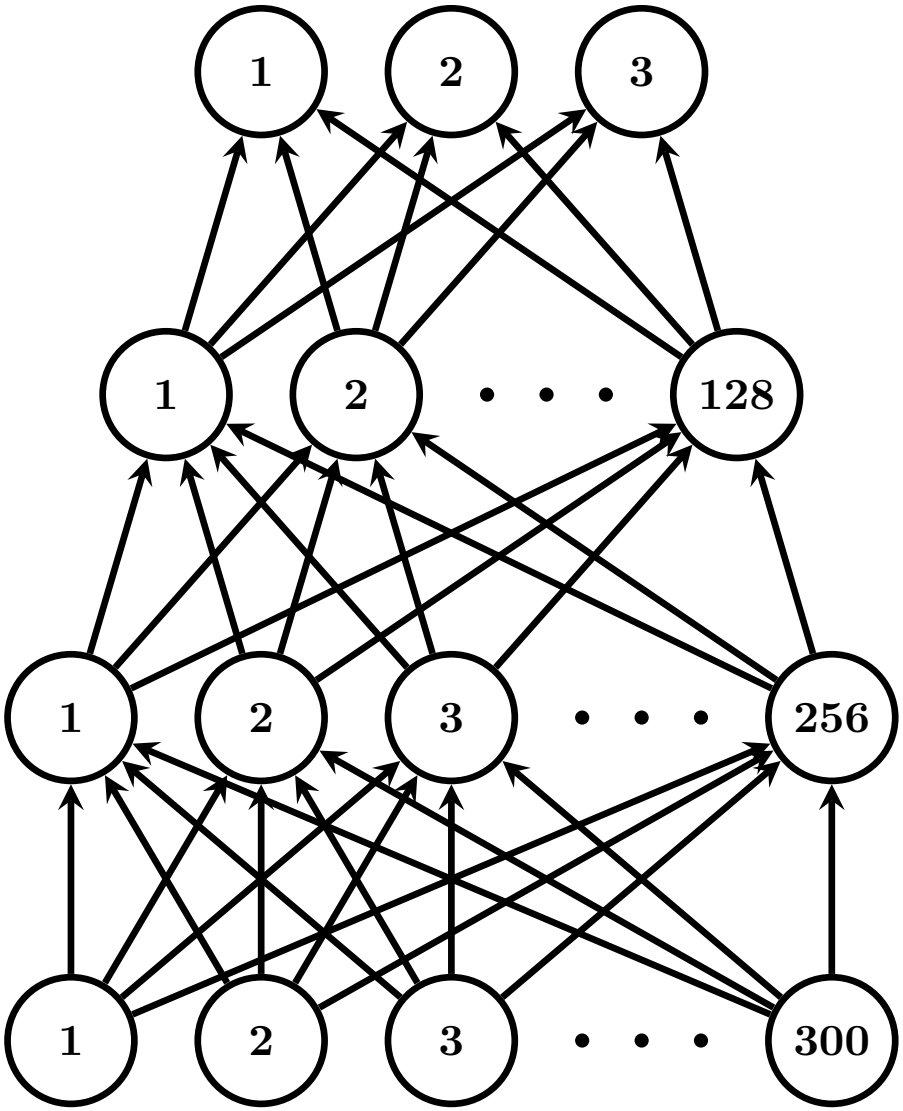
features

Emotion Lexicons

Source	ID	Language	Format	# Entries
Bradley and Lang (1999)	EN	English	VAD	1,034
Warriner et al. (2013)	EN+	English	VAD	13,915
Redondo et al. (2007)	ES	Spanish	VAD	1,034
Stadthagen-Gonzalez et al. (2017)	ES+	Spanish	VA	14,031
Schmidtke et al. (2014)	DE	German	VAD	1,003
Yu et al. (2016a)	ZH	Chinese	VA	2,802
Imbir (2016)	PL	Polish	VAD	4,905
Montefinese et al. (2014)	IT	Italian	VAD	1,121
Soares et al. (2012)	PT	Portuguese	VAD	1,034
Moors et al. (2013)	NL	Dutch	VAD	4,299
Sianipar et al. (2016)	ID	Indonesian	VAD	1,490

- 11 data sets
- 1 to 14k entries
- 9 languages

Model Details



output layer
affine transformation

two hidden layers
shared across VAD
.5 dropout
LReLU activation

embedding layer
.2 dropout

Word Embeddings

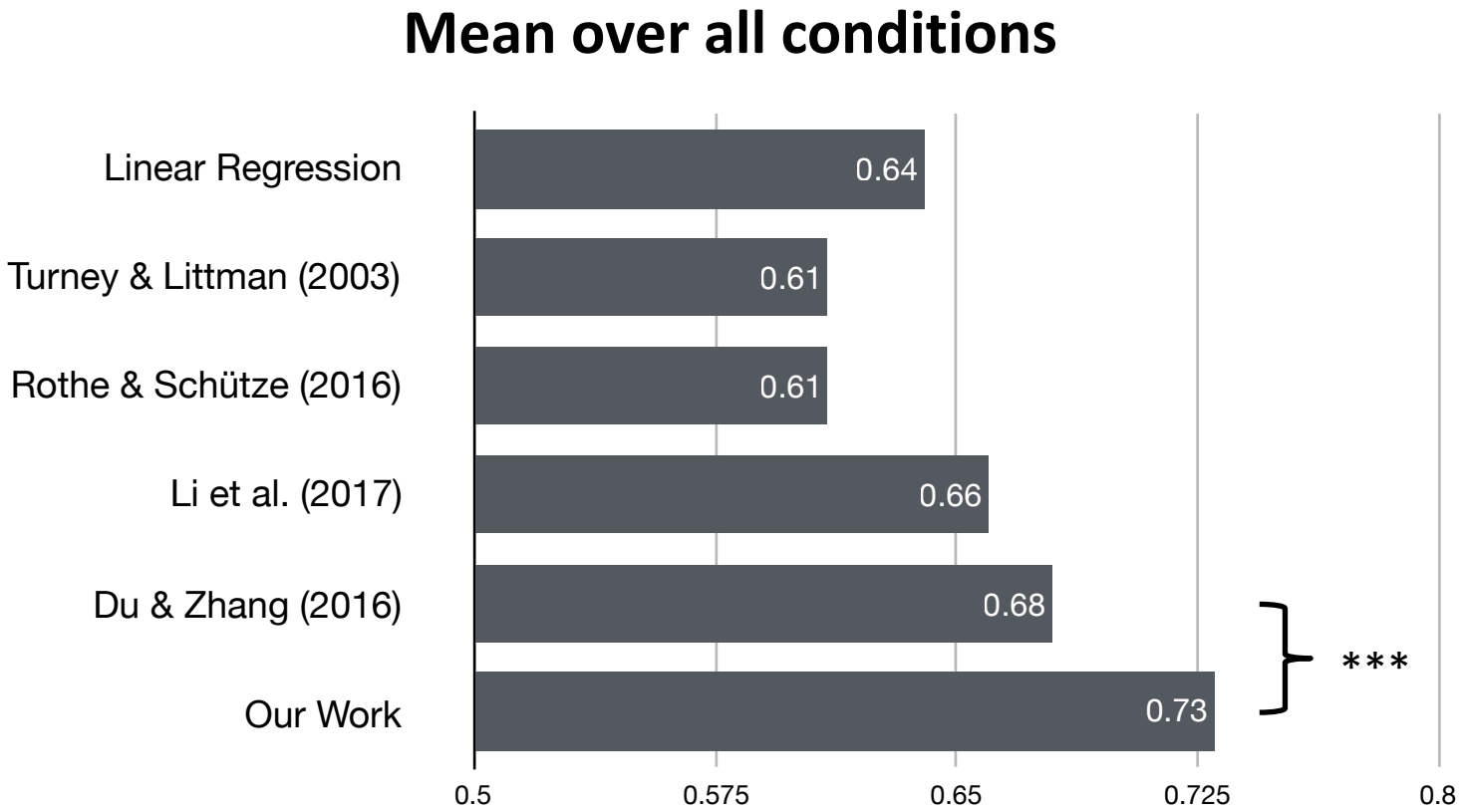
- All languages: FastText vectors trained on Wikipedias (Graves et al., LREC'18)
- English
  - Google News (SGNS, 100B)
  - Common Crawl (FastText, 600B)
- Not updated during training

Experimental Setup

- Compare our model against 5 reference methods
  - Linear regression baseline
  - Similarity to seed words (Turney & Littman, 2003)
  - Densifier (Rothe & Schütze, 2016)
  - Ridge regression (Li et al., 2017)
  - Boosted MLP (Du & Zhang, 2016)
- Evaluate on 11 data sets
- 3 distinct embedding models for English

New State-of-the-Art Results

(Buechel & Hahn, NAACL 2018)



- Very close to human performance (SHR and ISR)
- Word embeddings do not contain affective information???

Sentence-Level EA in Small Datasets

(Buechel et al., arXiv 2018)

- How much gold data is needed for sentence-level prediction?
- Chose four datasets
  - between 192 and 1000 instances
  - English, Polish, Portuguese
  - VAD and BE
- Same embeddings models as last study

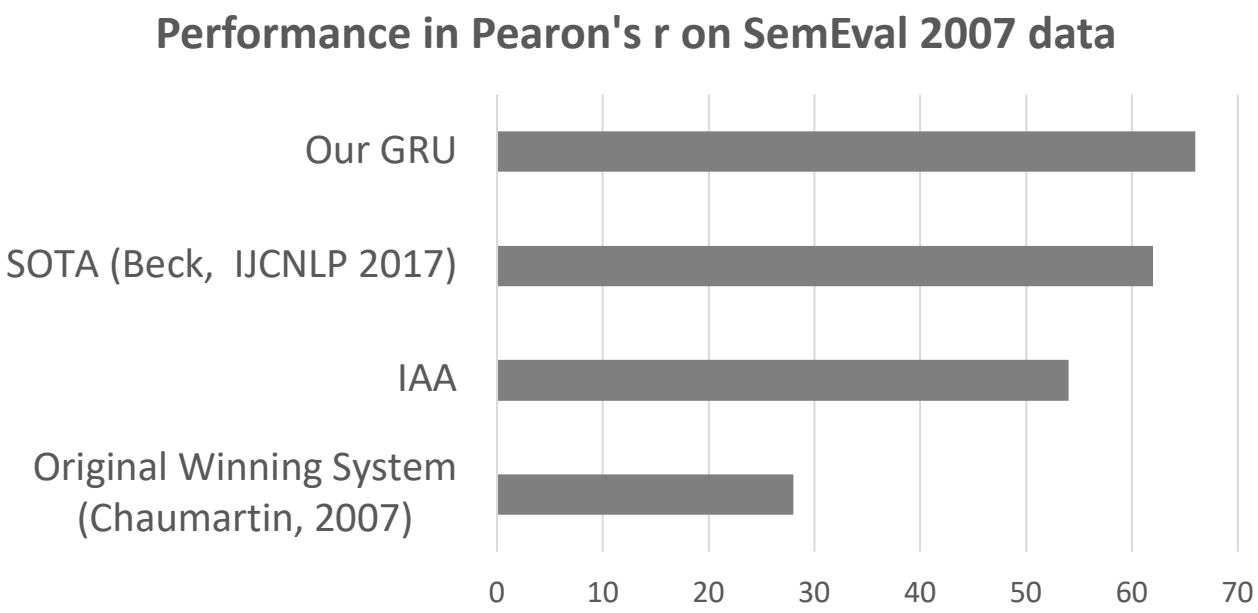
Small Sized Models of Different Architectures

- Baseline
  - BoW Ridge Regression
  - Bag-of-Vectors Ridge Regression
- DL models:

Model	Filters	Recurrent	1st Dense	2nd Dense
FFN	-	-	256	128
CNN	128	-	128	-
GRU	-	128	128	-
LSTM	-	128	128	-
CNN-LSTM	128	128	128	-

Results

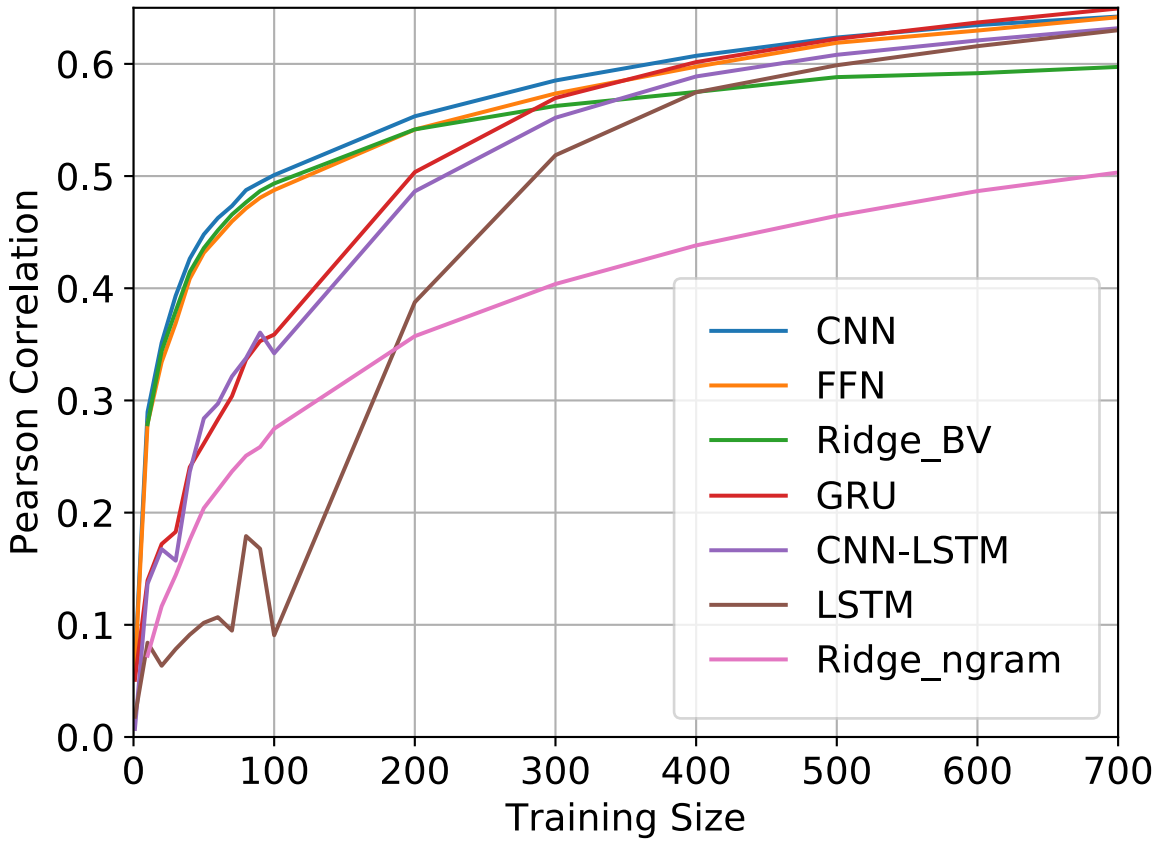
- All DL systems did surprisingly well on all datasets
- GRU performed best by 1%-pt over all datasets
- Beats (weak) IAA and previous SOTA on SemEval 2007 data



(Buechel et al., EMNLP 2018, arXiv 2018)

Influence of Training Size on Performance

(Buechel et al., arXiv 2018)



- GRU feasible down to 300 samples
- CNN and FFN feasible down to 100 samples

Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- ➢ Dealing with data sparsity
- Discussion and conclusion

Outline

- Introduction
- Applications of emotion analysis in DH and CSS
- Dealing with lack of interoperability
- Dealing with data sparsity
- ➢ Discussion and conclusion

Applications of Emotion Analysis

- Emotion more expressive than sentiment
- Advantageous in interdisciplinary applications
- VA(D) seems quite feasible
  - general purpose
  - easy to visualize
  - good value for money

Dealing with Lack of Interoperability

- Many different emotion representation formats
- Endanger interoperability of tools, datasets, and analyses
- Emotion representation mapping tackles this problem by allowing to convert between formats
- Mapped gold data is as reliable as actual gold data, probably even in cross-lingual applications

Dealing with Data Sparsity

- Turns out to be surprisingly unproblematic
- Multi-task learning helps a bit
- Small models and strong, pre-trained embeddings
- Word embeddings contain plenty of affective information (as opposed to popular claims in the literature)

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

Jena University Language & Information Engineering Lab

From Sentiment to Emotion:

Challenges of a More Fine-Grained Analysis of Affective Language

Sven Buechel

Jena University Language and Information Engineering (JULIE) Lab
Friedrich-Schiller-Universität Jena,
Jena, Germany

https://julielab.de

Slides: https://julielab.de/downloads/publications/slides/buechel_invited_ims_2018.pdf

References

Sven Buechel, João Sedoc, H. Andrew Schwartz, and Lyle Ungar. 2018. Learning Neural Emotion Analysis from 100 Observations: The Surprising Effectiveness of Pre-Trained Word Representations. In **arXiv**:1810.10949.

Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, João Sedoc. 2018. Modeling Empathy and Distress in Reaction to News Stories. In ***EMNLP 2018***.

Johannes Hellrich, **Sven Buechel** and Udo Hahn. 2018. JeSemE: A Website for Exploring Diachronic Changes in Word Meaning and Emotion. In ***COLING 2018: System Demonstrations***.

Sven Buechel and Udo Hahn. 2018. Emotion Representation Mapping for Automatic Lexicon Construction (Mostly) Performs on Human Level. In ***COLING 2018.***

Sebastian G.M. Händschke, **Sven Buechel**, Jan Goldenstein, Philipp Poschmann, Tinghui Duan, Peter Walgenbach and Udo Hahn. 2018. A Corpus of Corporate Annual and Social Responsibility Reports: 280 Million Tokens of balanced Organizational Writing. In ***ECONLP* @ ACL 2018**.

Sven Buechel and Udo Hahn. 2018. Word Emotion Induction for Multiple Languages as a Deep Multi-Task Learning Problem. In ***NAACL 2018*** .

Sven Buechel and Udo Hahn. 2018. Representation Mapping: A Novel Approach to Generate High-Quality Multi-Lingual Emotion Lexicons. In ***LREC 2018****.*

Sven Buechel, Johannes Hellrich and Udo Hahn: The Course of Emotion in Three Centuries of German Text: A Methodological Framework. In ***DH 2017****.*

Sven Buechel and Udo Hahn. 2017. A Flexible Mapping Scheme for Discrete and Dimensional Emotion Representations: Evidence from Textual Stimuli. In ***CogSci 2017.***

Sven Buechel and Udo Hahn. 2017. EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis. In ***EACL 2017.***

Sven Buechel and Udo Hahn. 2016. Emotion analysis as a regression problem - Dimensional models and their implications on emotion representation and metrical evaluation. In ***ECAI 2016***.

Sven Buechel, Udo Hahn, Jan Goldenstein, Sebastian G. M. Händschke, and Peter Walgenbach. 2016. Do enterprises have emotions? In ***WASSA @ NAACL 2016****.*

Backup Slides

Introduction: Sentiment and Emotion

NLP before Sentiment Analysis

- High-level NLP tasks used to be centered around **facts**
  - information/relation extraction
  - document classication
  - semantic parsing
  - natural language inference
- Then, around 2000, something happend...

Growing Interest in *Subjective* Language

Jena University Language & Information Engineering Lab

semantic polarity of words
(Hatzivassiloglou & McKeown, 1997)

good *fantastic*
great *mediocre*
boring *poor*

evaluative statements
(Pang et al., 2002)

The pizza was great!

The service was aweful...

expression of feelings

I just love the peace and quietness after a summer rain.

I hate John Doe, he has a terrible sense of humor.

Different „flavors“ of sentiment analysis

- Polarity prediction (SA as „document classification“)
- Aspect-based
- Opinion holder and target identification
- Related task: detecting subjectivity, irony, empathy, hate speech, offensive language

Application Domains

- Product reviews / analytics
  - Restaurant (Yelp)
  - Online retailers (Amazon)
  - Movies (RottenTomatoes, IMDB)
- Social media (esp. Twitter)
  - Political science
  - Public relations
  - Stock market prediction

All in all, I prefer a Marvel movie that doesn't take itself seriously, but the nonstop unseriousness of Deadpool 2 can wear you down, too.

May 18, 2018 | Rating: B- | Full Review…

Peter Rainer
Christian Science Monitor
★ Top Critic

rottentomatoes.com

Xavidub @jimdoherty09 · 2 Min.
I see **#Erdogan** is on course to get 101% of the vote 🙁

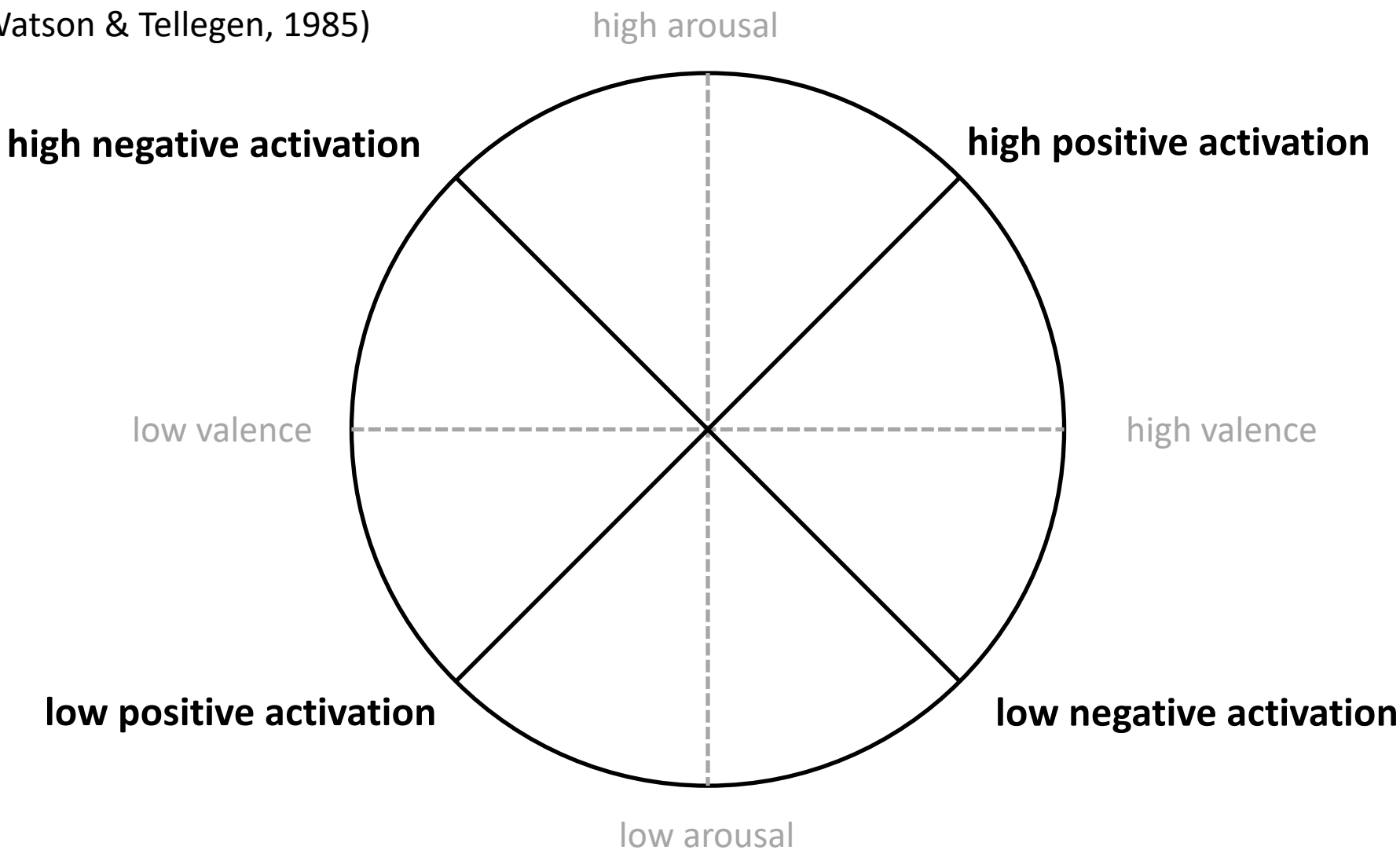
twitter.com

Tiffany Couch @TheTiffanyCouch · 22 Std.
Audi CEO arrested over **diesel** cheating scandal. #fraud **#audi** @FinancialTimes @FCPA

twitter.com

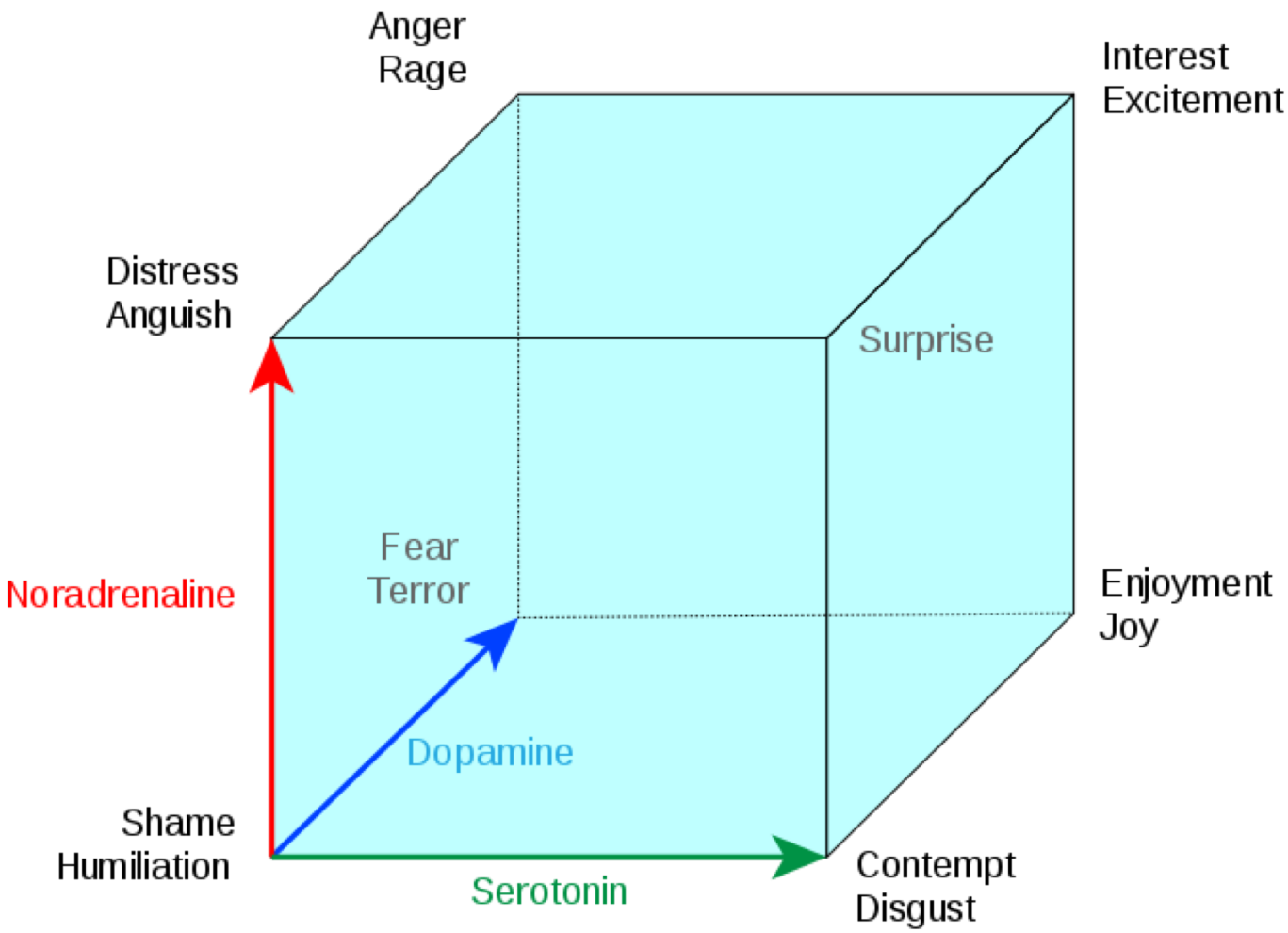
Positive Activation – Negative Activation (PANA)

(Watson & Tellegen, 1985)



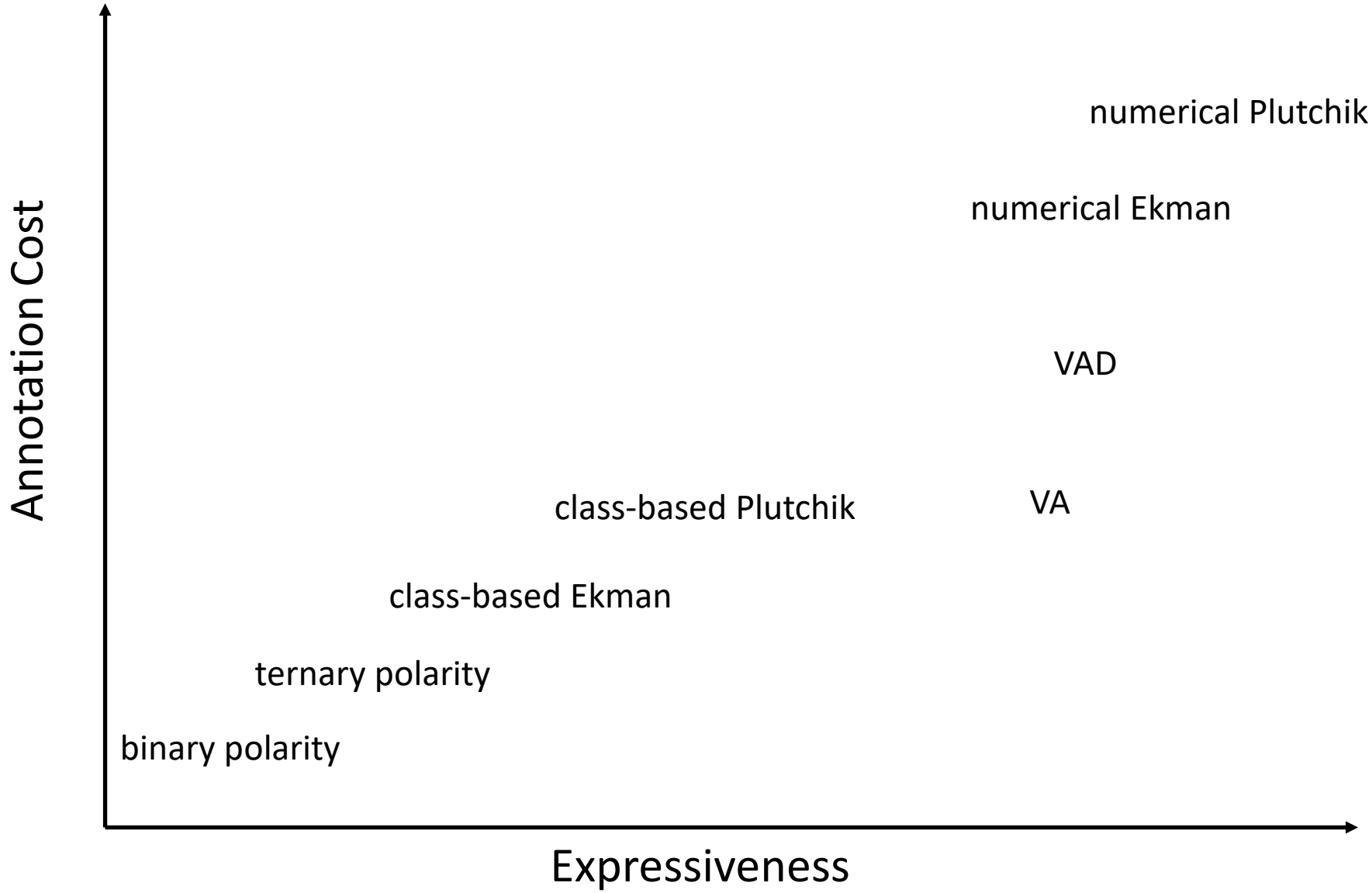
Lövheim Cube of Emotion

(Lövheim, 2012)



source: https://en.wikipedia.org/wiki/L%C3%B6vheim_cube_of_emotion

Annotation Cost vs. Expressiveness



Arguments in Favor of Dimensional Models

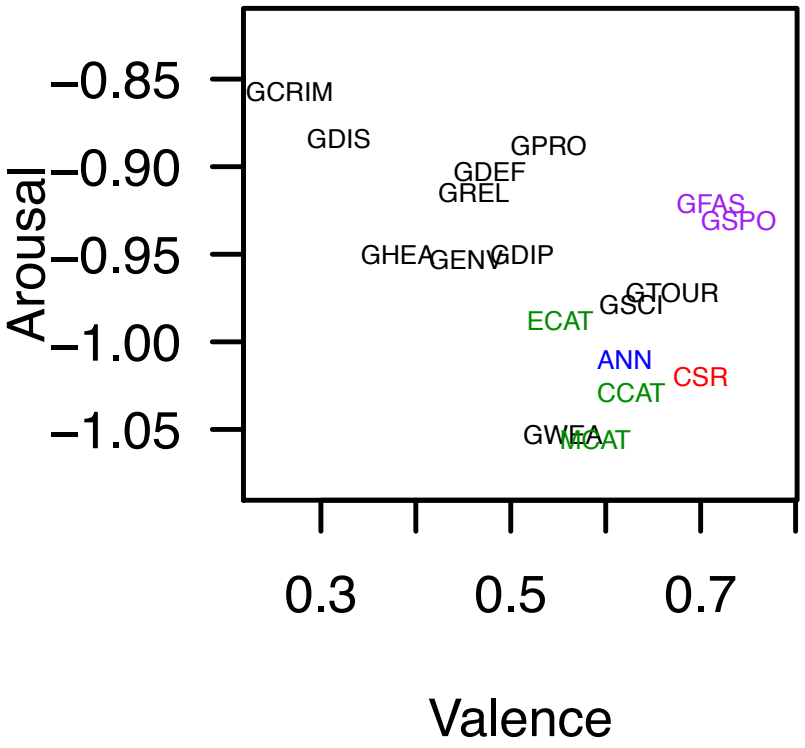
- Good value for money
- General purpose (one set of variables fits all use cases)
- Large overlap with psychology
- Interpretability
  - Intuitive to understand (in contrast to PANA, Lövheim)
  - Nice visualizations

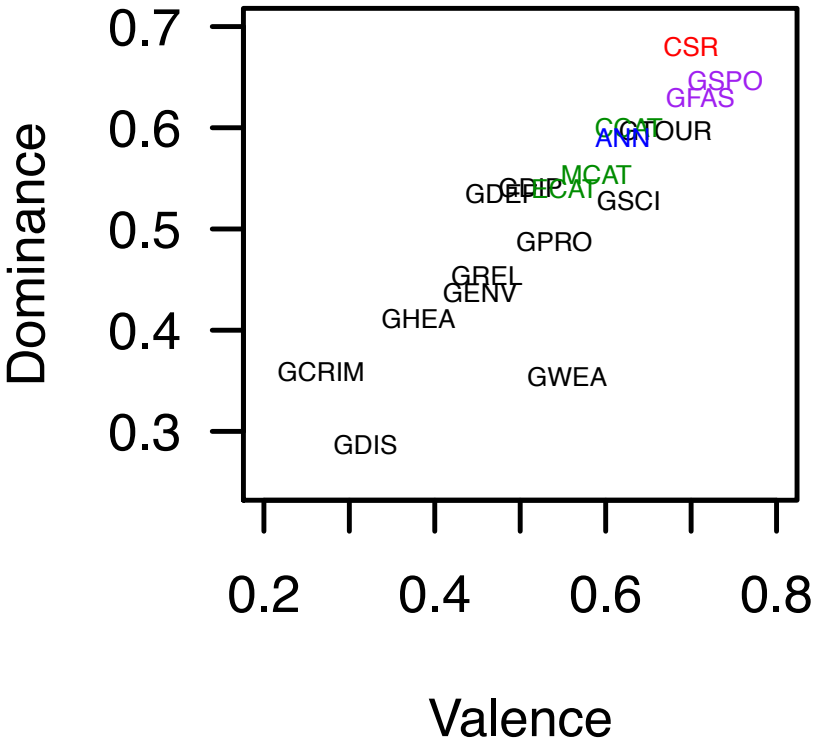
Organizational Emotion (WASSA, ECONLP)

JOCO Corpus Statistics

- 280M Tokens (for comparison: BNC has 100M),
- 5K reports
- Equal distribution by country
- 250K tokens of annual vs. 35K tokens of CSR reports

Results — Organizational Writing vs. News Topics

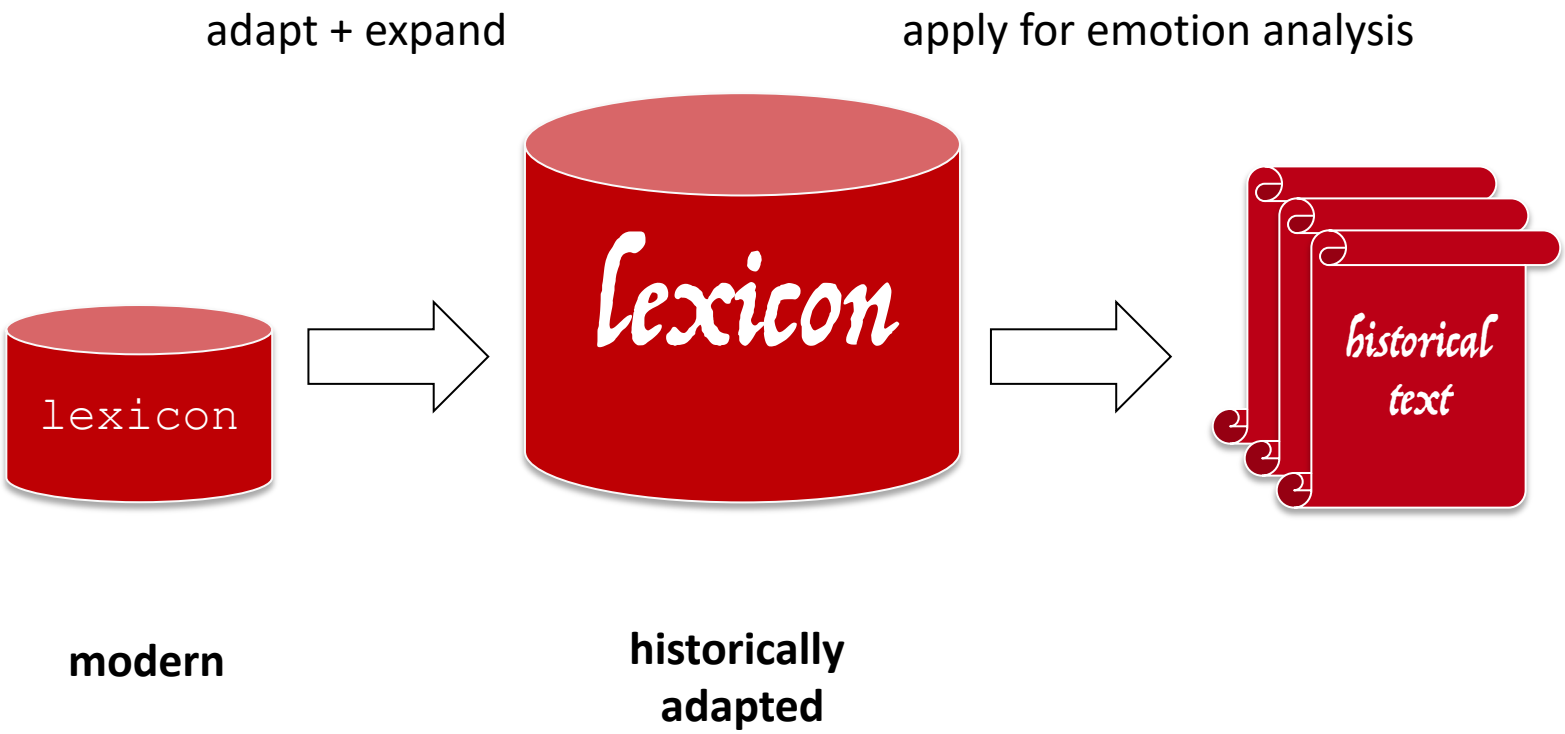




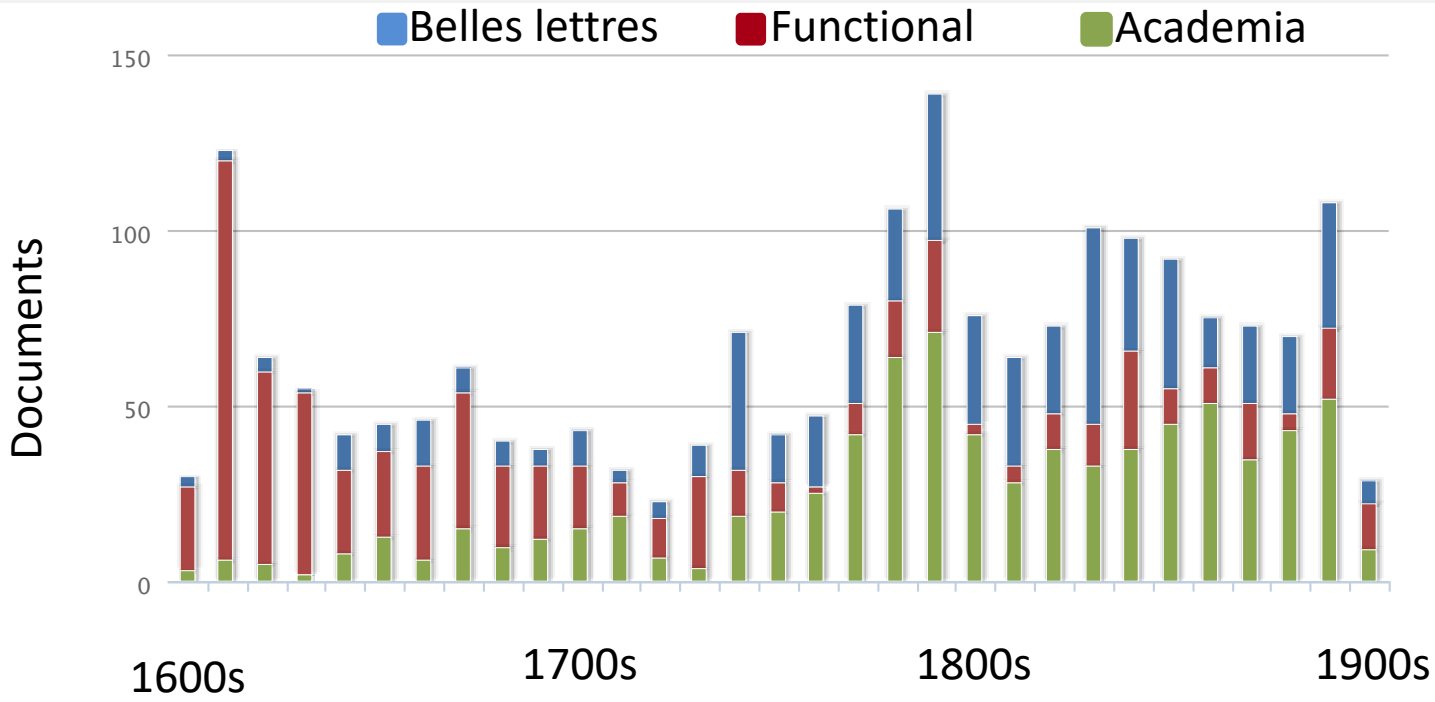
- Based on Reuters Corpus Volume 1 (RCV1)
- 800k newswire documents
- Hierarchy of 103 topic codes

Historical Emotions (LT4DH, DH, COLING)

Methodological Framework



Target Corpus: DTA



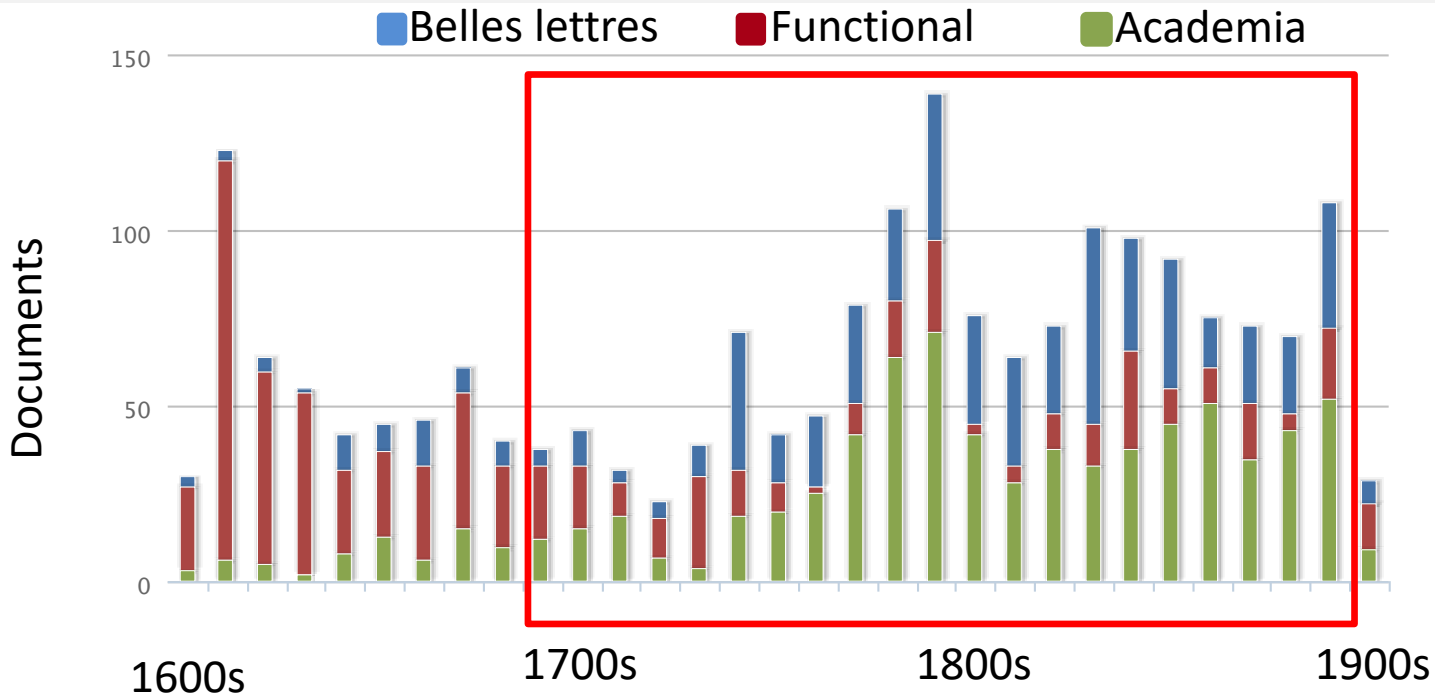
http://www.deutsches-textarchiv.de/doku/textauswahl

- 1st third shows different genre distribution
- Individual decades comprise too little text

➢ Aggregate 30-years slices

➢ Select 1690-1899 (~ 1k documents, 7 slices)

Target Corpus: DTA



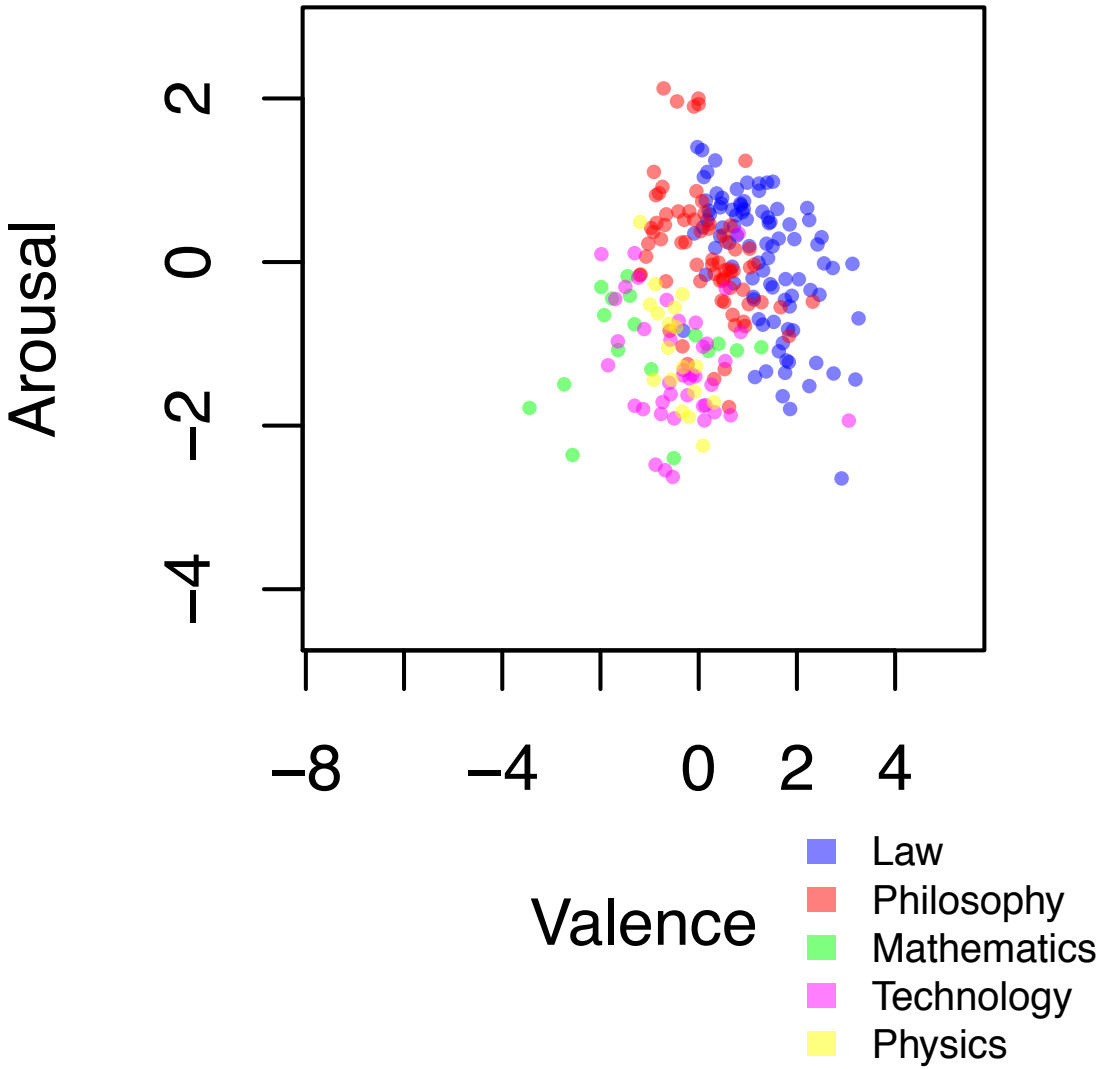
http://www.deutsches-textarchiv.de/doku/textauswahl

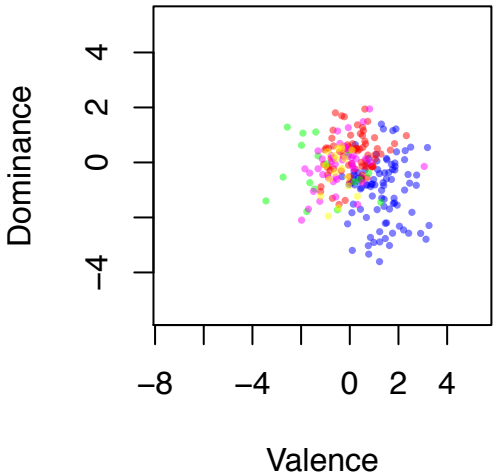
- 1st third shows different genre distribution
- Individual decades comprise too little text

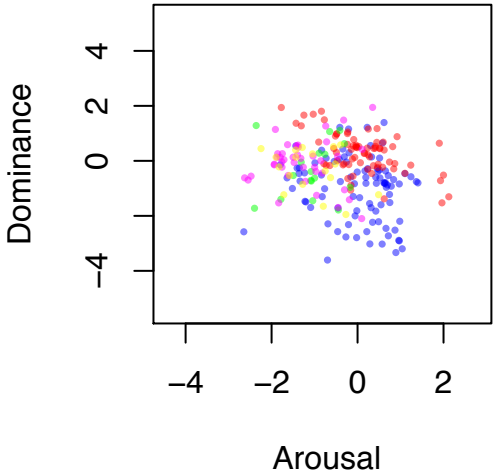
➢ Aggregate 30-years slices

➢ Select 1690-1899 (~ 1k documents, 7 slices)

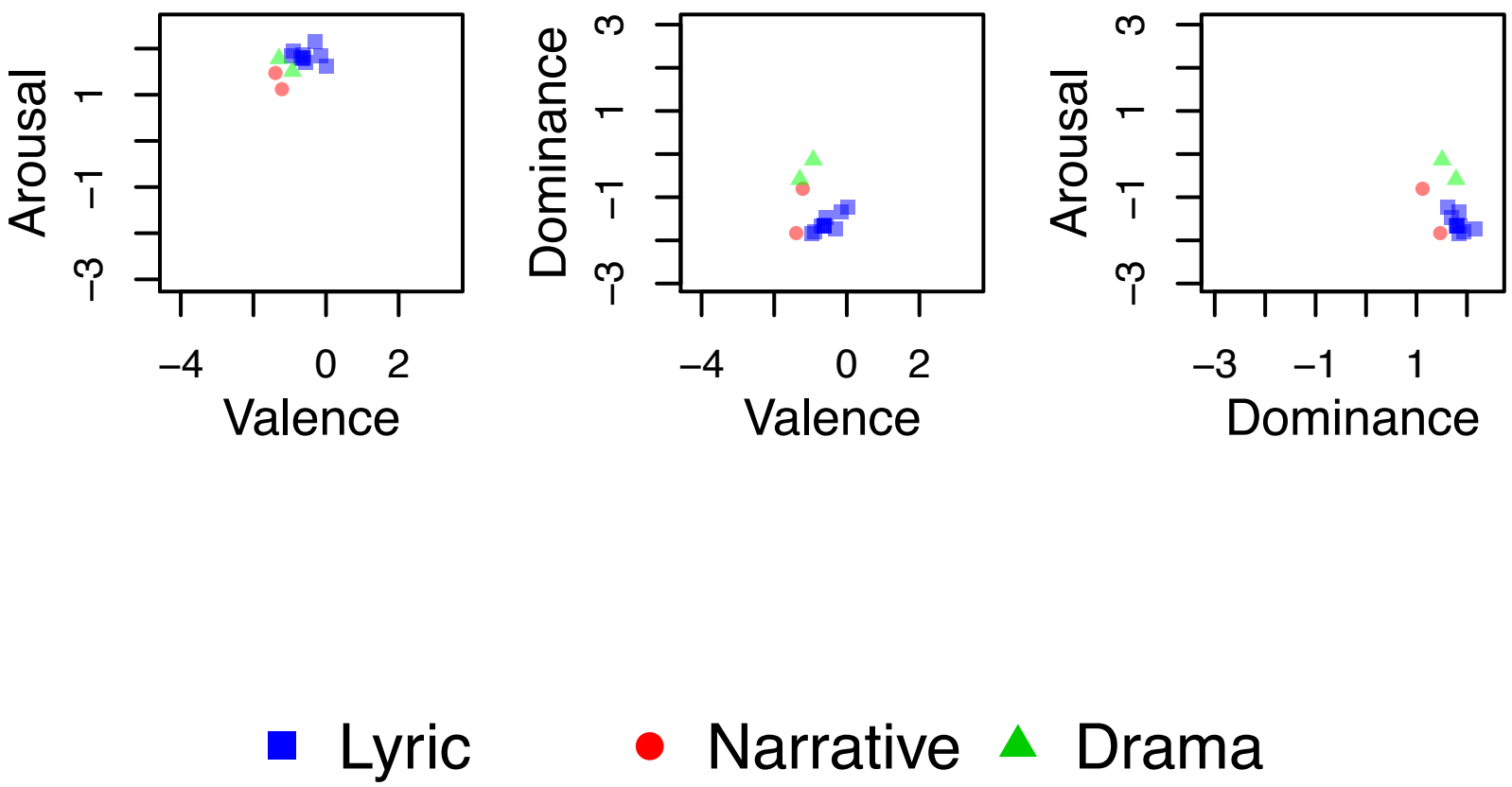
Distinction of Academic Subclasses



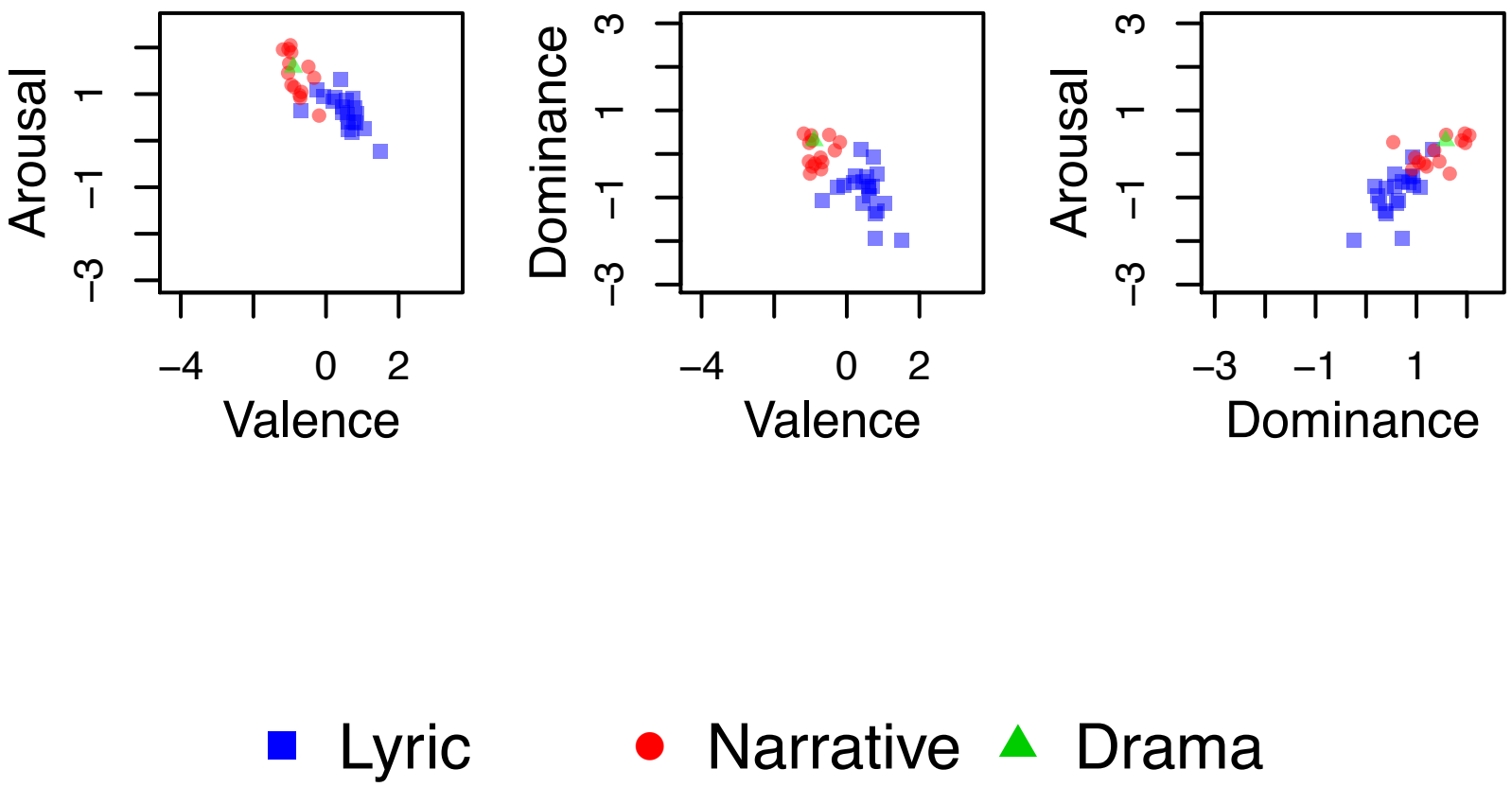




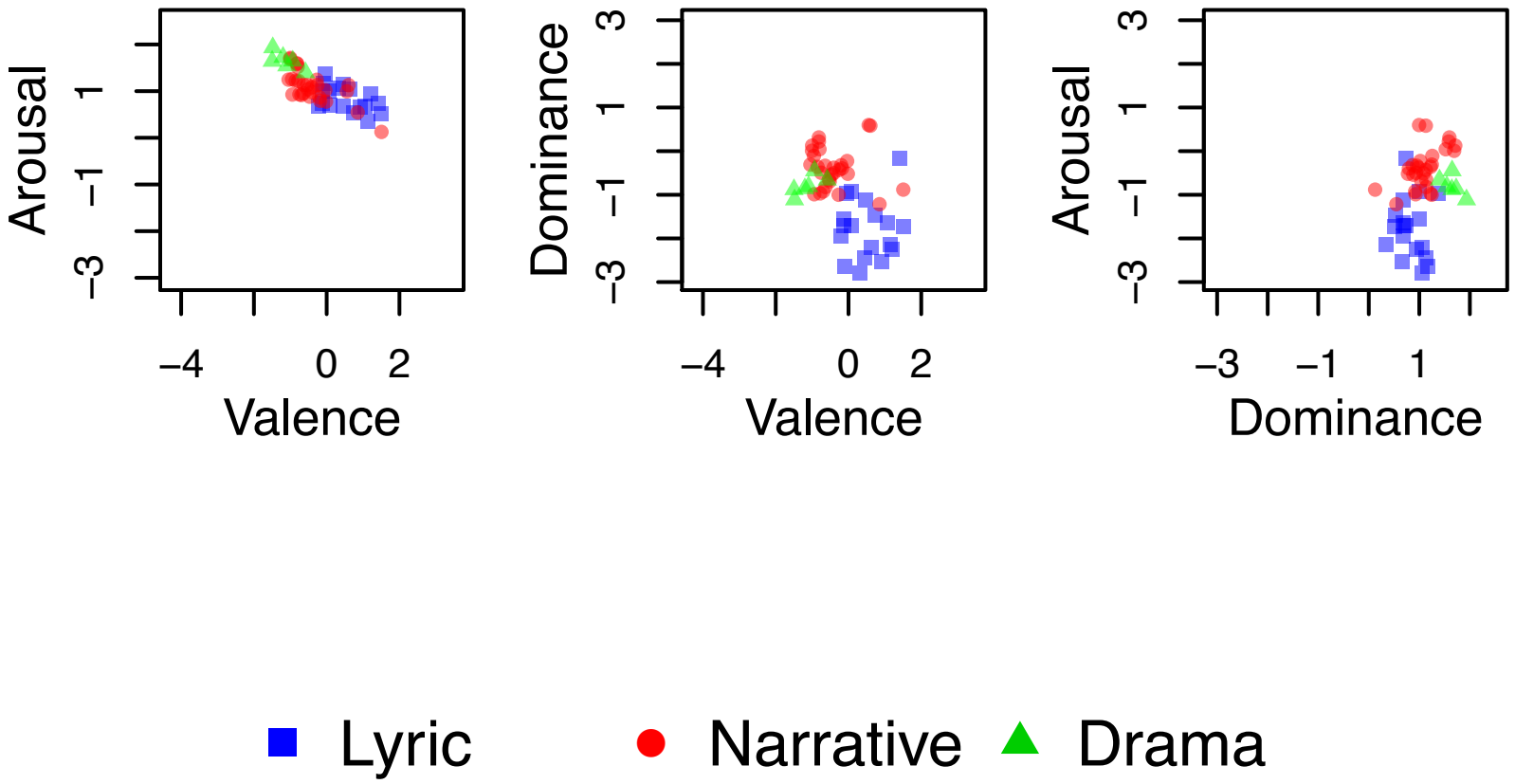
Development of Literary Forms (1690-1719)



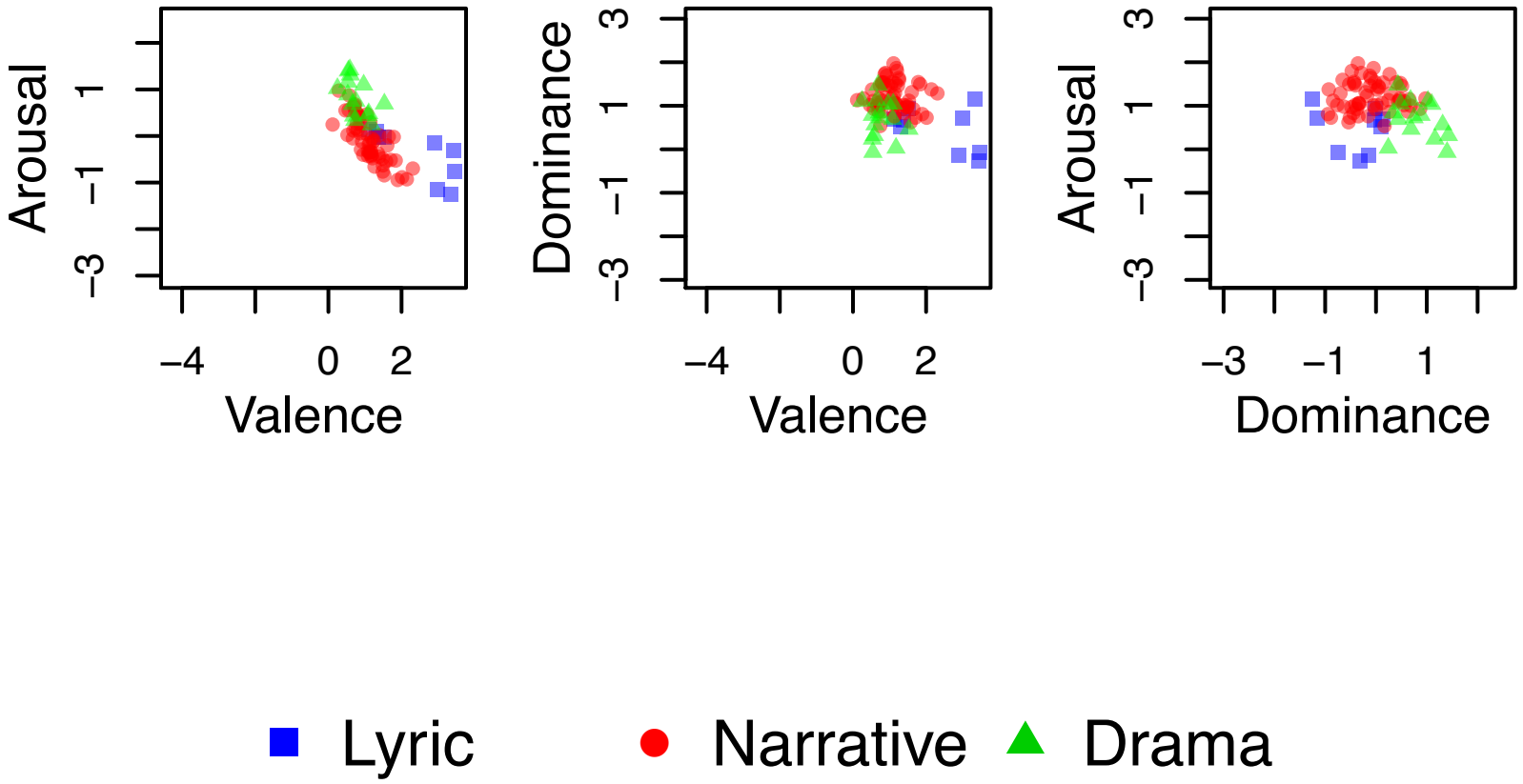
Development of Literary Forms (1720-1749)



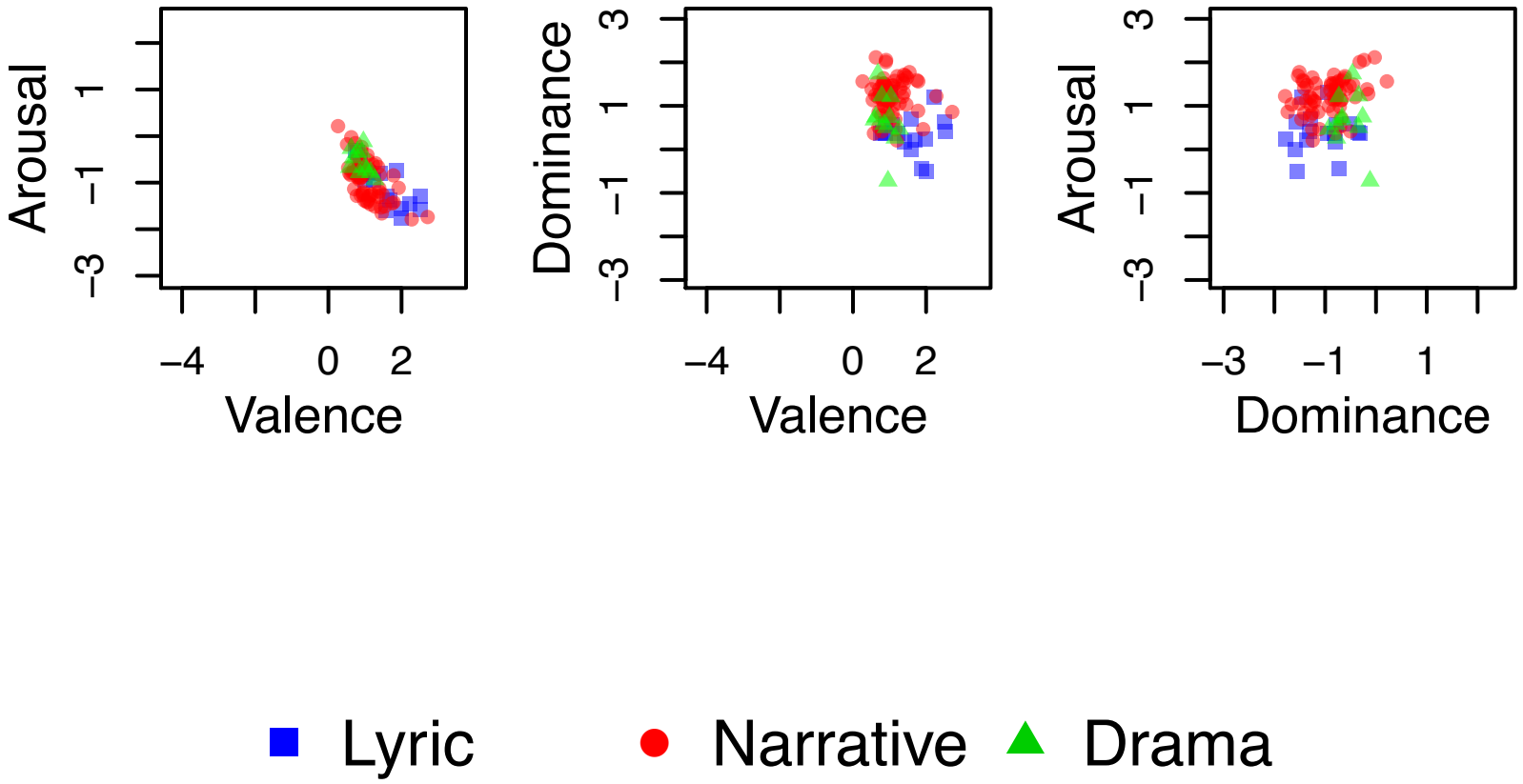
Development of Literary Forms (1750-1779)



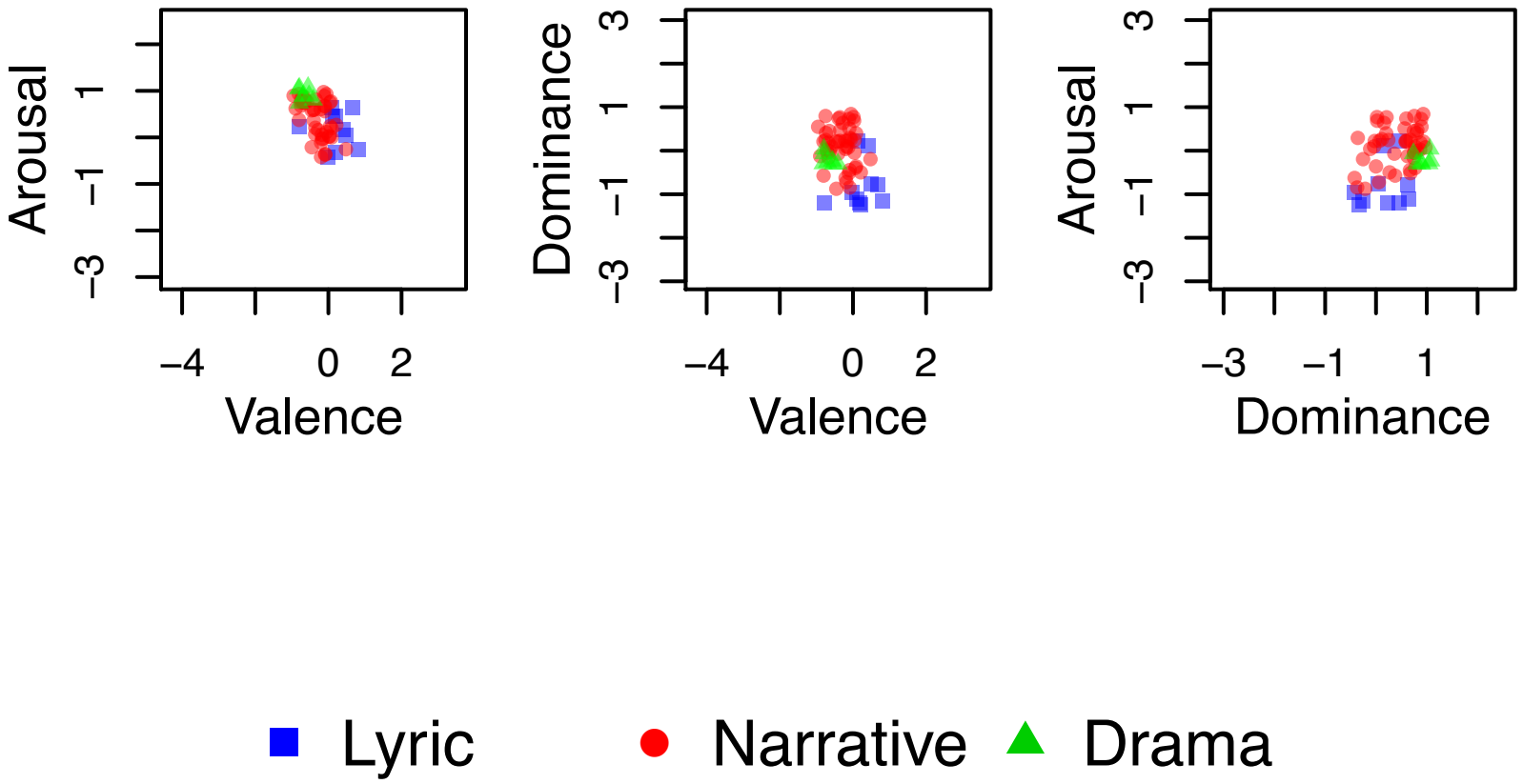
Development of Literary Forms (1780-1809)



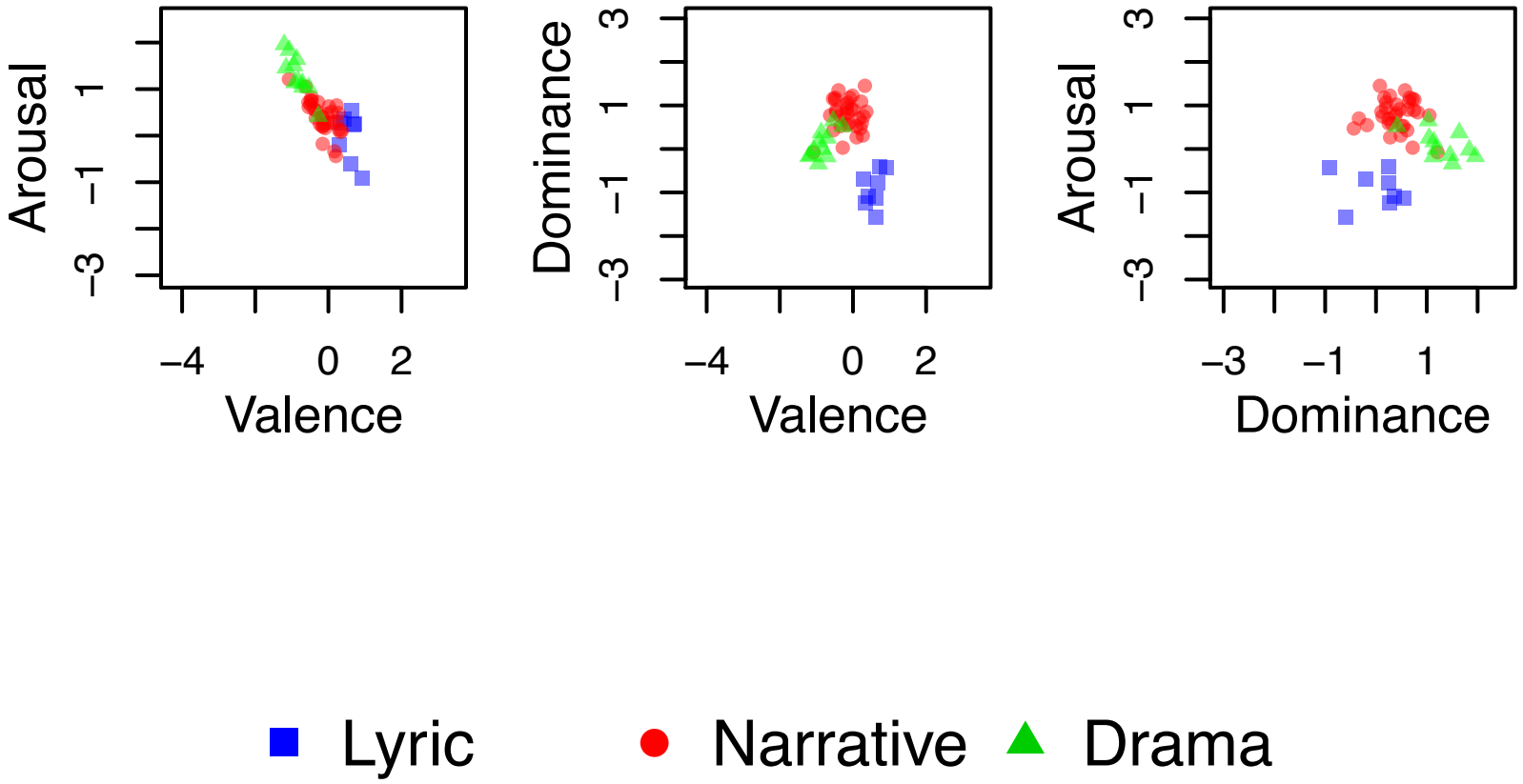
Development of Literary Forms (1810-1839)



Development of Literary Forms (1840-1869)



Development of Literary Genres (1870-1899)



Emotion Representation Mapping (ECAI, EACL, CogSci, LREC, COLING)

Results of JEmAS

(Buechel & Hahn, ECAI 2016)

- Outperforms all systems but one

  (10 reference systems in total)

  - 1st r ≈ .448 Staiano & Guerini (2014)
  - 2nd r ≈ .419 Our System
  - 3rd r ≈ .356 Neviarouskaya et al. (2011)

- State-of-the-art in 3 out of 6 emotional categories

Crowdsourcing a Large-Scale VAD Corpus

- EmoBank (Buechel & Hahn, EACL 2017)
- 10k sentences with VAD annotation from [1, 5]
- Comes with two kinds of double-annotation
  - Each sentence is annotated according to reader and writer perspective (pilot study was not fully conclusive (Buechel & Hahn, LAW 2017))
  - A subset (around 1.2k) has previously been annotated for BE5 (Strapparava & Mihalcea, SemEval 2007)
- Compare performance of EmoMap against IAA

IAA in the SemEval Dataset

	Rater 1	Rater 2	Rater 3
Item 1			
Item 2			
Item 3			
Item 4			

- For each rater
  - compute average annotation of remaining raters
  - compute correlation between this rater and average annotation
- Average over all raters
- Weak point of comparison because based on single human

Split-Half Reliability

- Correlation-based (numerical values)
- Increasingly popular within CL (Mohammad et al.)

	r1	r2	r3	r4	r5	r6
i1						
i2						
i3						
i4						
i5						
i6						

Split-Half Reliability

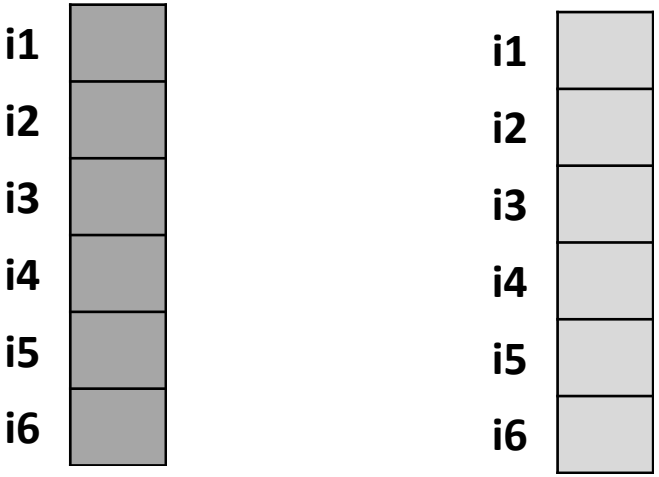
- Correlation-based (numerical values)
- Increasingly popular within CL (Mohammad et al.)

	r1	r4	r5
i1			
i2			
i3			
i4			
i5			
i6			

	r2	r3	r6
i1			
i2			
i3			
i4			
i5			
i6			

Split-Half Reliability

- Correlation-based (numerical values)
- Increasingly popular within CL (Mohammad et al.)



Spearman-Brown Adjustment

- SHR heavily influenced by number of raters thus not comparable between studies
- Solution: Spearman-Brown Adjustment, estimates reliability *r** if number of raters was increased by factor *k*

$$r^* := \frac{k\ r}{1 + (k-1)\ r}$$

Comparison against Human Reliability

- Compare model performance against adjusted SHR for 20 raters (arbitrarily chosen but tough comparison)
- Outperforming adjusted SHR:
  Model agrees more with gold data than two random groups of ten people would agree with each other.

Comparison against Human Performance

Abbrev.	VA(D)	BE5	Dom?	Overlap
en_1	Bradley and Lang (1999)	Stevenson et al. (2007)	✓	1,028
en_2	Warriner et al. (2013)	Stevenson et al. (2007)	✓	1,027
es_1	Redondo et al. (2007)	Ferré et al. (2017)	✓	1,012
es_2	Hinojosa et al. (2016b)	Hinojosa et al. (2016a)	✓	875
es_3	Stadthagen-Gonzalez et al. (2017b)	Stadthagen-González et al. (2017a)	✗	10,491
de_1	Võ et al. (2009)	Briesemeister et al. (2011)	✗	1,958
pl_1	Riegel et al. (2015)	Wierzba et al. (2015)	✗	2,902
pl_2	Imbir (2016)	Wierzba et al. (2015)	✓	1,272

- „Monolingual“ Evaluation: 10-CV on one pair of datasets
- „Crosslingual“ Evaluation: fixed test set, train on all other languages

Results: Monolingual

	Val	Aro	Dom	Joy	Ang	Sad	Fea	Dsg
en_1	.969	.741	.848	.962	.876	.871	.873	.805
en_2	.964	.704	.861	.942	.868	.821	.860	.799
es_1	.974	.771	.863	.957	.854	.833	.869	.752
es_2	.986	.828	.720	.977	.913	.867	.878	.807
es_3	.915	.692	—	.846	.839	.857	.842	.744
de_1	.929	.745	—	.894	.778	.644	.785	.461
pl_1	.963	.787	—	.946	.872	.826	.805	.826
pl_2	.947	.768	.760	.935	.844	.805	.790	.819
Avg.	.956	.754	.810	.932	.855	.816	.838	.752

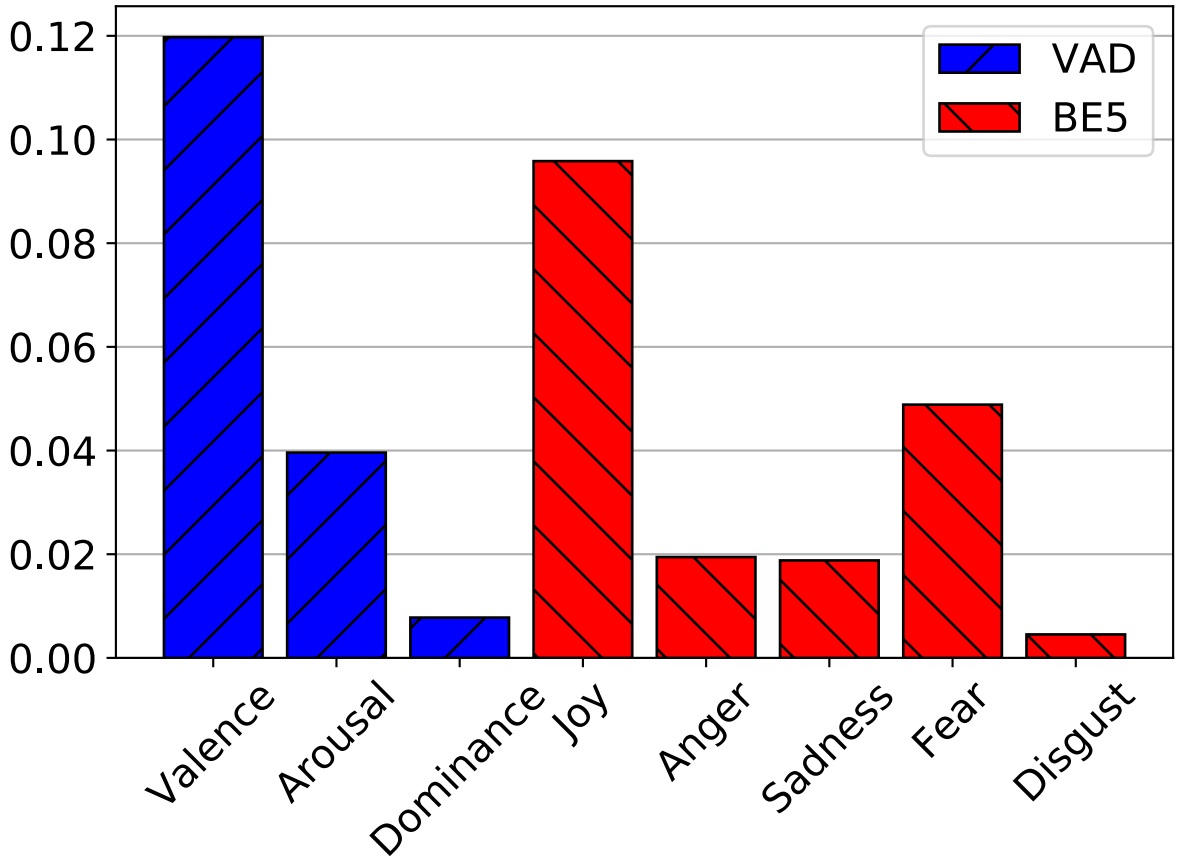
- Outperforming human reliability in 66% of cases

Results: Cross-Lingual

	Val	Aro	Joy	Ang	Sad	Fea	Dsg
en_1	.966	.683	.955	.858	.838	.817	.781
en_2	.956	.642	.934	.855	.810	.791	.800
es_1	.973	.692	.951	.786	.802	.782	.682
es_2	.985	.735	.974	.881	.860	.835	.787
es_3	.908	.548	.839	.821	.850	.807	.728
de_1	.927	.708	.889	.767	.618	.760	.458
pl_1	.957	.666	.937	.848	.784	.745	.801
pl_2	.938	.720	.932	.816	.785	.751	.809
Avg.	.951	.674	.926	.829	.793	.786	.731

- Outperforming human reliability in 54% of cases

How Important is Dominance anyway?



Not very!

Newly Generated Emotion Ratings

Mth	Lng	Format	Source	#Words
m	en	BE5	Warriner et al. (2013)	12,884
m	es	VAD	Stadthagen-González et al. (2017a)	10,489
m	de	BE5	Võ et al. (2009)	944
m	pl	BE5	Imbir (2016)	3,633
c	it	BE5	Montefinese et al. (2014)	1,121
c	pt	BE5	Soares et al. (2012)	1,034
c	nl	BE5	Moors et al. (2013)	4,299
c	id	BE5	Sianipar et al. (2016)	1,487
c	zh	BE5	Yu et al. (2016a); Yao et al. (2017)	3,797
c	fr	BE5	Monnier and Syssau (2014)	1,031
c	gr	BE5	Palogiannidi et al. (2016)	1,034
c	fn	BE5	Eilola and Havelka (2010)	210
c	sv	BE5	Davidson and Innes-Ker (2014)	99