



IMPACT & FICTION

Measuring the impact of fiction on readers

Interpreting and Computing Valence in (Dutch) Literary Studies

Peter Boot

peter.boot@huygens.knaw.nl

DH 2023, Workshop SIG-DLS Seven Years on

July 10, 2023



Motivation: valence in literary studies

Reagan et al. 2016: Six basic shapes

Rebora 2020: Shared Emotions in Reading Pirandello

Buechel et al. 2017: Course of Emotion in Three Centuries of German Text

Grubert and Algee-Hewitt 2017: Depictions of oil and coal in American fiction and nonfiction

Nalisnick and Baird 2013: Char-to-Char Sent Analysis in Shakespeare's Plays

Contents

- Valence in psycholinguistic studies
 - Valence in consumer reviews
 - Valence in word embedding-based approaches
 - Valence beyond the word level
 - Word-based tools & dictionaries for Dutch
 - Application on corpus of narrative
 - Conclusions
-
- Limitation: only lexicon-based approaches
 - Reflects the current state of my reading and thinking

Valence in psycholinguistic studies

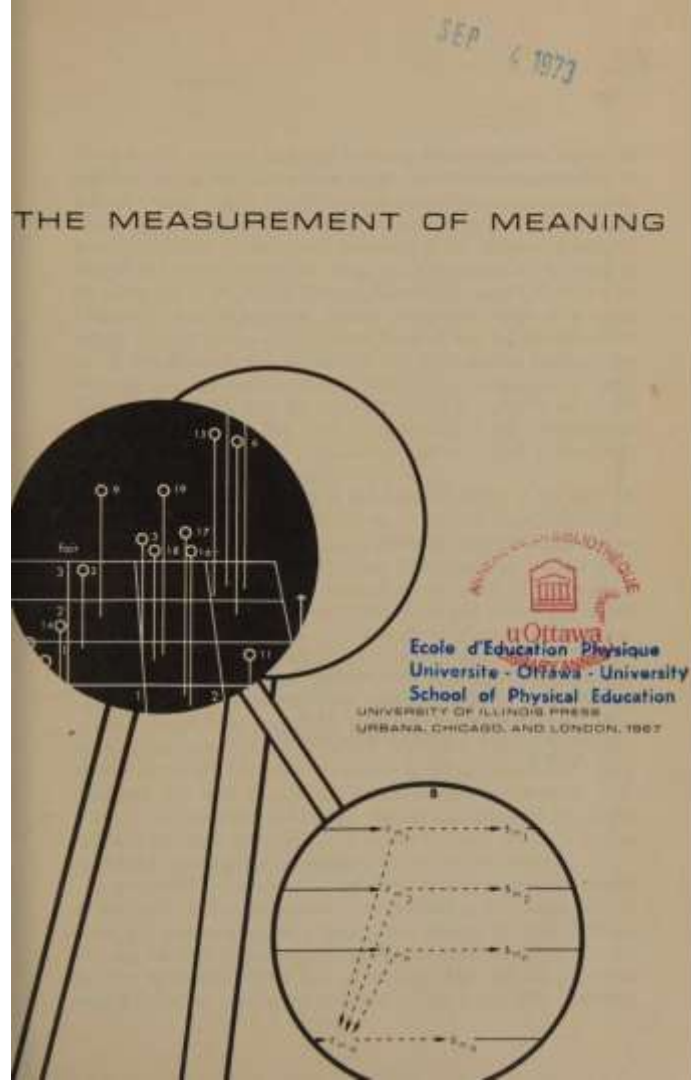
Osgood et al 1957

Words (concepts) associated to positions on scales of opposite adjectives (high - low, good - bad)

on the same scale. Each item appeared as follows:

LADY rough _____:_____:_____:_____:_____:_____:_____ smooth,
with the subject instructed to place a check-mark in that position

Followed by factor analysis



Labelled dimensions as

Evaluativeness

Potency

Activity

Now known as

Valence

Dominance

Arousal

Table 1

ROTATED FACTOR LOADINGS — ANALYSIS I

	I	II	III	IV	h ²
*1. good-bad	.88	.05	.09	.09	.79
2. large-small	.06	.62	.34	.04	.51
3. beautiful-ugly	.86	.09	.01	.26	.82
4. yellow-blue	-.33	-.14	.12	.17	.17
5. hard-soft	-.48	.55	.16	.21	.60
6. sweet-sour	.83	-.14	-.09	.02	.72
7. strong-weak	.19	.62	.20	-.03	.46
8. clean-dirty	.82	-.05	.03	.02	.68
9. high-low	.59	.21	.08	.04	.40
10. calm-agitated	.61	.00	-.36	-.05	.50
11. tasty-distasteful	.77	.05	-.11	.00	.61
*12. valuable-worthless	.79	.04	.13	.00	.64
13. red-green	-.33	-.08	.35	.22	.28
14. young-old	.31	-.30	.32	.01	.29
15. kind-cruel	.82	-.10	-.18	.13	.73
16. loud-soft	-.39	.44	.23	.22	.45
17. deep-shallow	.27	.46	.14	-.25	.37
18. pleasant-unpleasant	.82	-.05	.28	-.12	.77
19. black-white	-.64	.31	.01	-.03	.51
20. bitter-sweet	-.80	.11	.20	.03	.69
21. happy-sad	.76	-.11	.00	.03	.59
22. sharp-dull	.23	.07	.52	-.10	.34
23. empty-full	-.57	-.26	-.03	.18	.43

value
Bitter
activity

*

The General Inquirer

Stone et al 1966

SIGN-STRONG and SIGN-WEAK and SIGN-ACCEPT and SIGN-REJECT were devised to operationalize two of the Osgood evaluative dimensions.

Direct measurement of what was a constructed dimension

THE GENERAL INQUIRER: A Computer Approach to Content Analysis

*Philip J. Stone
Dexter C. Dunphy
Marshall S. Smith
Daniel M. Ogilvie
with associates*



The M.I.T. Press

*Massachusetts Institute of Technology
Cambridge, Massachusetts, and London, England*

Affective Norms for English Words (ANEW)

Bradley and Lang 1999

Words to be used as stimuli in psychological research

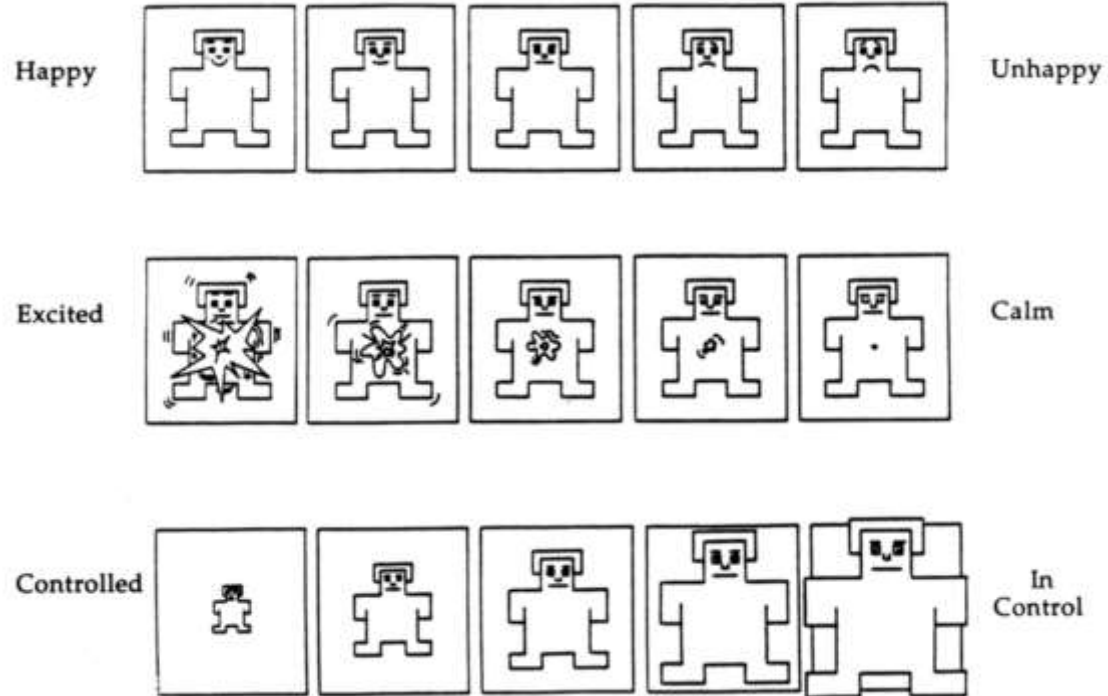
Pleasure, arousal and dominance

Scored using Self-Assessment Mannekins

Self Assessment Mannekin

"At one extreme of this scale, you are happy, pleased, satisfied, contented, hopeful. *When you feel completely happy you should indicate this by bubbling in the figure at the left.* The other end of the scale is when you feel completely unhappy, annoyed, unsatisfied, melancholic, despaired, or bored. *You can indicate feeling completely unhappy by bubbling in the figure at the right.*"

Now valence has become something that you can feel



Norms of valence, arousal, and dominance for 13,915 English lemmas

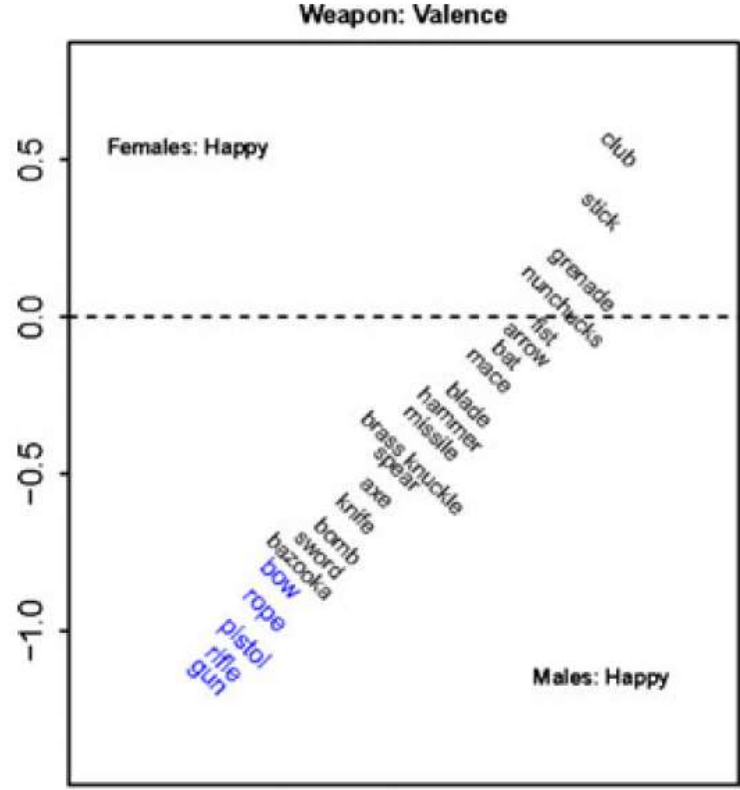
Warriner et al 2013

Amazon Mechanical Turk (US citizens)

Likert scales

"You will use a scale *to rate how you felt while reading each word*. There will be approximately 350 words. The scale ranges from 1 (happy) to 9 (unhappy). At one extreme of this scale, you are happy, pleased, satisfied, contented, hopeful. *When you feel completely happy* you should indicate this by choosing rating 1."

Gender differences (from Warriner et al.)



Norms of VAD and AoA for 4,300 Dutch words

Moors et al. 2013

"Participants in the valence/pleasantness condition were asked to judge the extent to which *the words in the study referred to something that is positive/pleasant* ("positief/aangenaam") or negative/unpleasant ("negatief/onaangenaam")"

Valence no longer about feeling, but about objects in the world.

Gender differences in the output:

- Women like babies, marriage, traditional virtues and hairdressers more than men do
- Men value beer, sex and war more than women do

Obtaining reliable human ratings of VAD for 20,000 English words

Mohammad 2018

Crowdsourced

Best-worst-scaling:

- "Q1. Which of the four words below *is associated with* the MOST happiness / pleasure / positiveness / satisfaction / contentedness / hopefulness OR LEAST unhappiness / annoyance / negativeness / dissatisfaction / melancholy / despair?"
- Q2. Which of the four words below *is associated with* the LEAST happiness / pleasure / positiveness / satisfaction / contentedness / hopefulness OR MOST unhappiness / annoyance / negativeness / dissatisfaction / melancholy / despair?"

'Is associated with': impersonal formulation, not about feeling, not about the world, but about the language system?

Differences based on

- Gender
- Age
- Big 5 personality characteristics (self-assessed)

As literary scholars have known all along: there are only perspectives on the text

Valence in consumer reviews

Also called polarity or sentiment

No longer: how does reader respond

But: what did author of text want to express

ANEW: typically nouns

In consumer reviews typically adjectives



Pattern

De Smedt and Daelemans 2012

Toolset for NLP, web scraping and sentiment analysis

"Seven human annotators were presented with the list in random order and asked to classify each adjective in terms of positive-negative polarity and subjectivity."

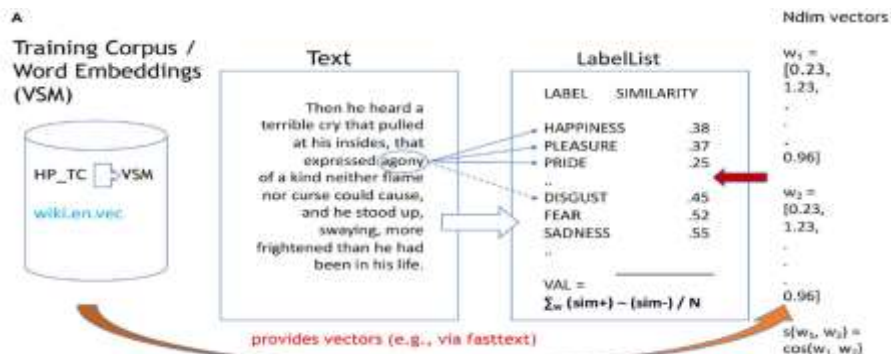
Distinguishes sense of words (without using them)

No longer test subjects but expert annotators.

	form	wordnet_id	pos	sense	polarity	subjectivity	intensity	confidence
2983	scheidsrechterlijk	a-02637730	JJ	van scheidsrechter	0.0	0.3	1.0	0.7
2984	schelden	v-00824767	VB	NaN	-0.2	0.1	1.0	0.7
2985	schemerachtig	a-00274551	JJ	schemerig	0.0	0.3	1.0	0.7
2986	scherp	a-01810189	JJ	niet bot	0.1	0.4	1.0	1.0
2987	scherp	a-01214255	JJ	puntig	0.1	0.4	1.0	1.0
2988	scherp	a-01214430	JJ	van geluid	-0.4	0.6	1.0	1.0
2989	scherp	a-00033077	JJ	van smaak en geur	-0.1	0.8	1.0	1.0
2990	scherp	a-00648614	JJ	bits	-0.6	1.0	1.0	1.0
2991	scherp	a-00438909	JJ	scherpzinnig	0.6	0.9	1.0	1.0
2992	scherp	NaN	JJ	(van munitie)	0.1	0.2	1.0	1.0
2993	scherp	a-00834198	JJ	goed, effectief	0.6	0.9	1.1	1.0

Valence in word embedding-based approaches

Example: Jacobs 2019



But Jacobs 2017

'Embedding': Germanet

Labels: satisfaction, relief, joy, happiness, contentment, pride, surprise, satisfied, relieved, joyful, happy, satisfied, proud, satisfy, relieve, rejoice, gladden, delight

Seems objective, because we don't need annotators anymore.

In a sense: meaning now comes from the language system

Except of course it isn't: seed words, word embedding

When is it appropriate to use a specific embedding?

What are the consequences of using one embedding rather than another?

Valence beyond the word

1 to few sentences - ANET, Bradley and Lang 2007

1 tweet – SemEval task, Rosenthal et al 2017

1 paragraph - Reading Pirandello, Rebora 2020

100 words – Section *Sandman*, Jacobs and Kinder 2020

180 words – IDEST short story database, Kaakinen et al 2022

→ Requires quite a leap of faith to apply this at book level

Not clear what it would mean to assess valence of a book or a book chapter:

Valence is always seen as an intuitive response, and reading a book takes too long for an intuitive response.

Word-based valence lexicons/tools for Dutch

Jacobs: based on 11,000 novel embedding, word freq > 50 (Impact and Fiction)

LiLaH: correction of NRC translation (Ljubešić et al 2020)

LIWC 2007: manual translation of 2007 dictionary (Boot et al 2017)

LIWC 2015: automatic translation of 2015 dictionary (Van Wissen and Boot 2017)

Moors (Moors et al. 2013)

Pattern (De Smedt and Daelemans 2012)

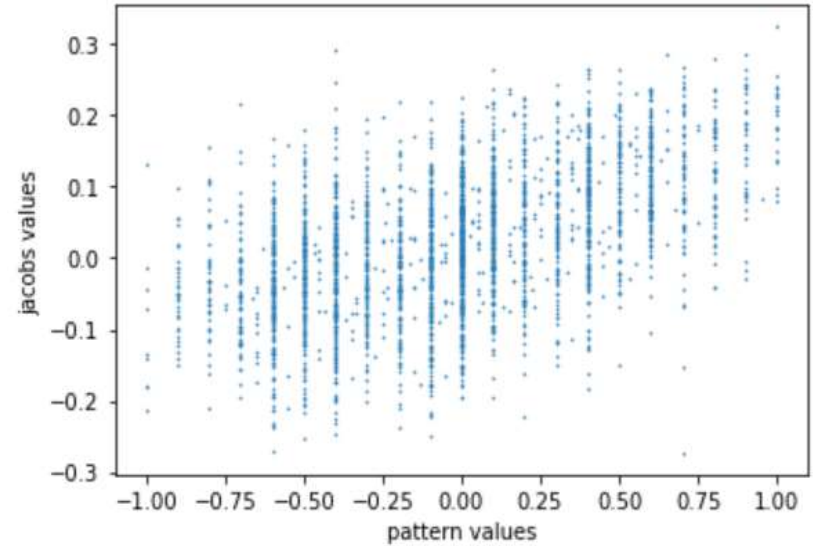
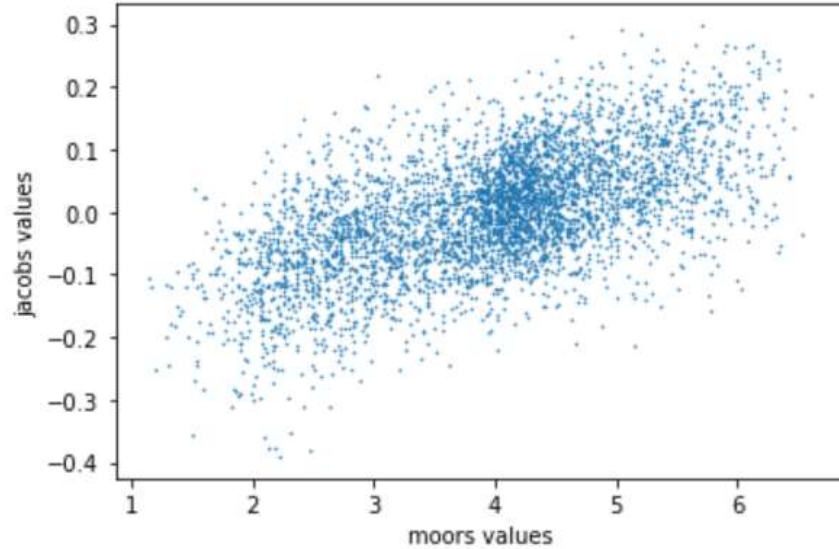
Overlap

		in l	in l not in r	in l and r	in r not in l	in r
	l					
	r					
jacobs	lilah	129853	124838	5015	731	5746
	liwc15	129853	127842	2011	674	2685
	moors	129853	125595	4258	41	4299
	patt	129853	126935	2918	386	3304
lilah	liwc15	5746	5043	703	1982	2685
	moors	5746	4265	1481	2818	4299
	patt	5746	4842	904	2400	3304
liwc15	moors	2685	2224	461	3838	4299
	patt	2685	2361	324	2980	3304
moors	patt	4299	3573	726	2578	3304

Except with Jacobs, overlap between lexicons is small.

→ plenty of room for the lexicons to disagree,
even if on the shared words they have similar ratings.

Agreement



At word level, tools broadly agree.

But is this enough if the tools should measure the same concept?

Correlations on book fragments

	polarity_patt	polarity_jacobs	polarity_lilah	polarity_liwc07	polarity_liwc15	polarity_moors
polarity_patt	1.000000	0.328739	0.468759	0.527935	0.466739	-0.021068
polarity_jacobs	0.328739	1.000000	0.398416	0.344903	0.318249	-0.358199
polarity_lilah	0.468759	0.398416	1.000000	0.548289	0.502073	-0.038768
polarity_liwc07	0.527935	0.344903	0.548289	1.000000	0.756389	0.129475
polarity_liwc15	0.466739	0.318249	0.502073	0.756389	1.000000	0.099935
polarity_moors	-0.021068	-0.358199	-0.038768	0.129475	0.099935	1.000000

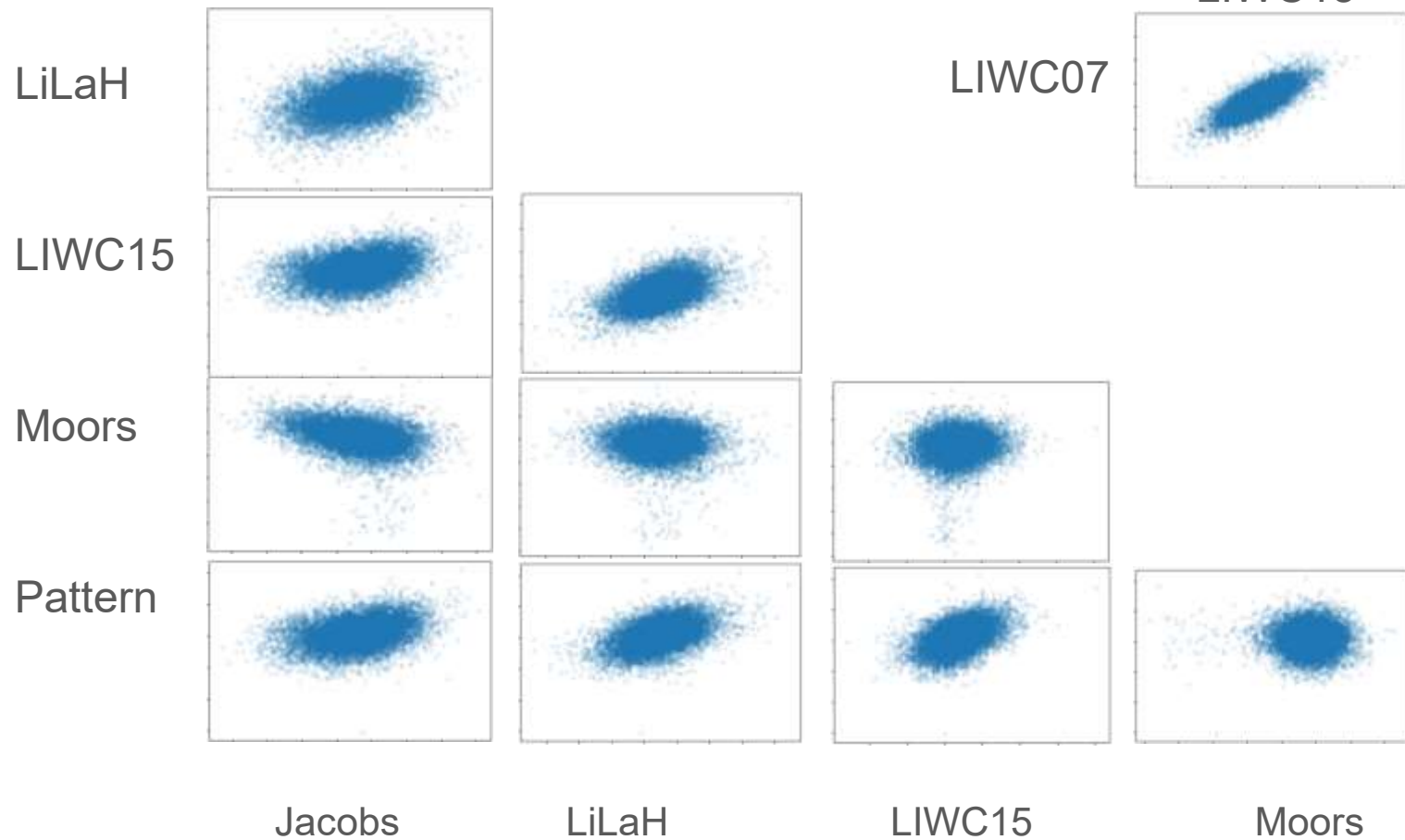
11,000 novels, random fragment of 500 words, average valence.

LIWC 2007 does best in agreeing with the others.

Result for Moors is strange.

In all very disappointing.

The same information as scatterplots



Provisional conclusions

There is no single well-defined concept of valence

Definitions vary between perspective of author, world and reader

Various methods of establishing word valence

Most assume that context is irrelevant

No methods allow for ambivalence

Demographic groups differ in their perception of (word) valence

Can valence be validated in texts > few hundred words?

On narrative texts, tools / lexicons (for Dutch) give widely different results

If a concept is theoretically doubtful and hard to measure consistently – what is the use of this concept?

Literature

- Boot, Peter, Hanna Zijlstra, and Rinie Geenen. 2017. "The Dutch translation of the Linguistic Inquiry and Word Count (LIWC) 2007 dictionary." *Dutch Journal of Applied Linguistics* 6 (1): 65-76.
- Bradley, Margaret M, and Peter J Lang. 1999. *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical report C-1. Center for research in psychophysiology. U. of Florida.
- Bradley, Margaret M, and Peter J Lang. 2007. "Affective Norms for English Text (ANET): Affective ratings of text and instruction manual." Technical Report. D-1, University of Florida, Gainesville, FL.
- Buechel, Sven, Johannes Hellrich, and Udo Hahn. 2017. "The Course of Emotion in Three Centuries of German Text-A Methodological Framework." DH 2017.
- De Smedt, Tom, and Walter Daelemans. 2012. "'Vreselijk mooi!' (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives." Proceedings of the 8th Language Resources and Evaluation Conference (LREC'12).
- Hu, Mingqing, and Bing Liu. 2004. "Mining and summarizing customer reviews." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.
- Jacobs, Arthur M. 2017. "Quantifying the beauty of words: a neurocognitive poetics perspective." *Frontiers in Human Neuroscience* 11: 622.
- Jacobs, Arthur M. 2019. "Sentiment analysis for words and fiction characters from the perspective of computational (Neuro-) poetics." *Frontiers in Robotics and AI* 6: 53.
- Jacobs, Arthur M, and Annette Kinder. 2020. "Computing the affective-aesthetic potential of literary texts." *AI* 1 (1): 11-27.
- Kaakinen, Johanna K, Egon Werlen, Yvonne Kammerer, Cengiz Acartürk, Xavier Aparicio, Thierry Baccino, Ugo Ballenghein, Per Bergamin, Núria Castells, and Armanda Costa. 2022. "IDEST: International Database of Emotional Short Texts." *PLoS ONE* 17 (10).
- Ljubešić, Nikola, Ilija Markov, Darja Fišer, and Walter Daelemans. 2020. "The LiLaH Emotion Lexicon of Croatian, Dutch and Slovene." Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media.
- Mohammad, Saif. 2018. "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words." *Proceedings of the 56th annual meeting of the association for computational linguistics* (volume 1: Long papers).
- Moors, Agnes, Jan De Houwer, Dirk Hermans, Sabine Wanmaker, Kevin Van Schie, Anne-Laura Van Harmelen, Maarten De Schryver, Jeffrey De Winne, and Marc Brysbaert. 2013. "Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words." *Behavior research methods* 45 (1): 169-177.
- Nalisnick, Eric T, and Henry S Baird. 2013. "Character-to-character sentiment analysis in Shakespeare's plays." Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).
- Osgood, Charles Egerton, George J Suci, and Percy H Tannenbaum. 1957. *The measurement of meaning*
- Reagan, Andrew J, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. "The emotional arcs of stories are dominated by six basic shapes." *EPJ Data Science* 5 (art. no. 31).
- Rebora, Simone. 2020. "Shared Emotions in Reading Pirandello. An Experiment with Sentiment Analysis." AIUCD 2020.
- Rosenthal, Sara, Noura Farra, and Preslav Nakov. 2019. "SemEval-2017 task 4: Sentiment analysis in Twitter." Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017).
- Stone, Philip J, Dexter C Dunphy, and Marshall S Smith. 1966. *The general inquirer: A computer approach to content analysis*. the MIT Press.
- Teodorescu, Daniela, and Saif M Mohammad. 2022. "Frustratingly Easy Sentiment Analysis of Text Streams: Generating High-Quality Emotion Arcs Using Emotion Lexicons." arXiv:2210.07381.
- Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert. 2013. "Norms of valence, arousal, and dominance for 13,915 English lemmas." *Behavior research methods* 45: 1191-1207.
- Van Wissen, Leon, and Peter Boot. 2017. "An Electronic Translation of the LIWC Dictionary into Dutch." eLex Conference 2017, Leiden.