



# Distant Reading in

## Sentiment Analysis

Simone Rebora & Giovanni Pietro Vitali

[simone.rebora@univr.it](mailto:simone.rebora@univr.it)

[giovannipetrovitali@gmail.com](mailto:giovannipetrovitali@gmail.com)



# SENTIMENT ANALYSIS

“This book is **fantastic**!  
It tells **wonderful** and **beautiful** things...”

+0.9

+1

“In those days, **sadness**  
shadowed my heart”

-0.5

-1



# SENTIMENT ANALYSIS

“[...] is the field of study that **analyzes people’s opinions, sentiments, appraisals, attitudes, and emotions** towards entities and their attributes expressed in written text. The entities can be products, services, organizations, individuals, events, issues, or topics”  
**(Liu, 2015)**



# SENTIMENT ANALYSIS

[ABOUT](#) [ARTICLES](#) [BOOKS](#) [COURSES](#) [LECTURES/EVENTS](#) [PAPERS](#) [WORKSHOPS](#) [NOTED](#)

## Matthew L. Jockers

05  
Thursday  
JUN 2014

### *A Novel Method for Detecting Plot*

POSTED BY MATTHEW JOCKERS IN COMMENTARY, TEXT-MINING

≈ COMMENTS OFF

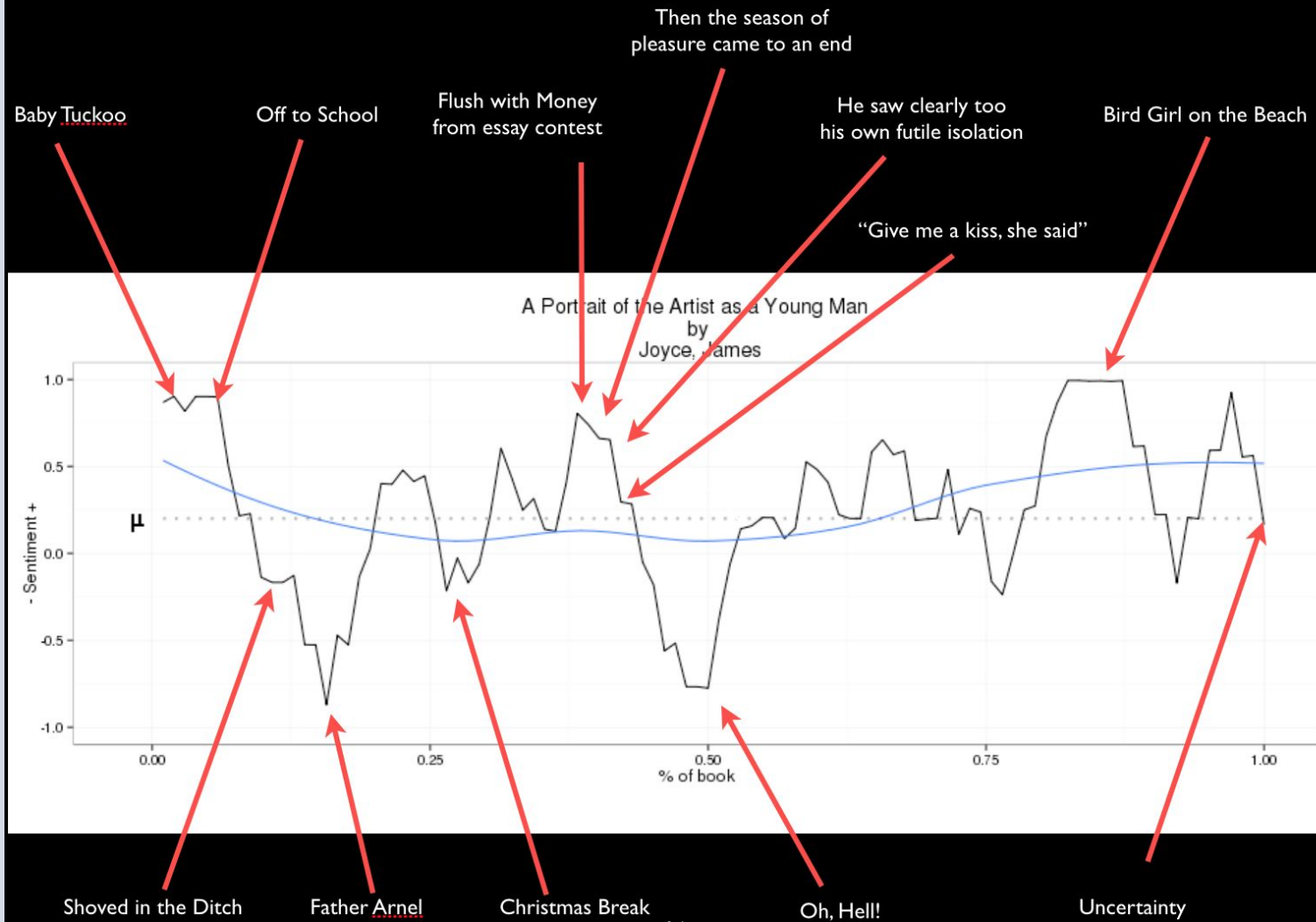
While studying anthropology at the University of Chicago, Kurt Vonnegut proposed writing a master's thesis on the shape of narratives. He argued that "the fundamental idea is that stories have shapes which can be drawn on graph paper, and that the shape of a given society's stories is at least as interesting as the shape of its pots or spearheads." The idea was rejected.

In 2011, [Open Culture](#) featured a video in which Vonnegut expanded on this idea and suggested that computers might someday be able to model the shape of stories, that is, the movement of the narratives, the plots. The video is about four minutes long; it's worth watching.

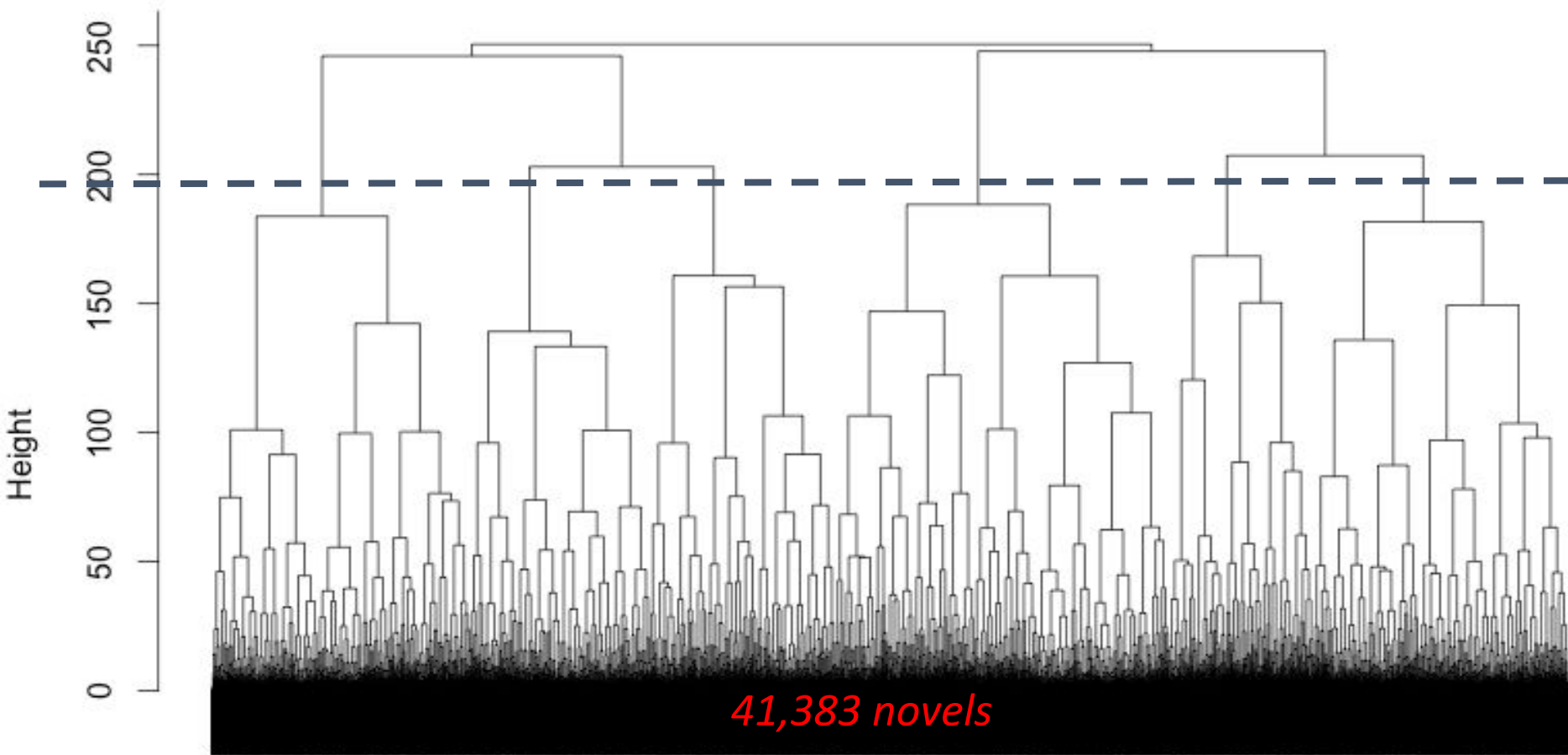
#### ✦ CONTACT

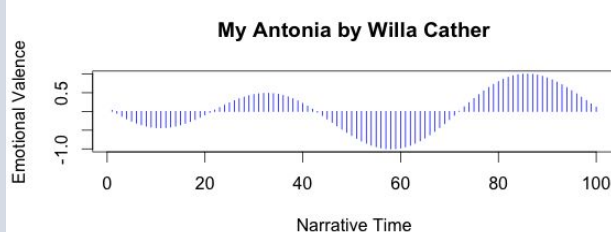
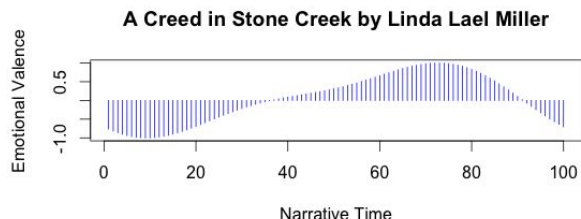
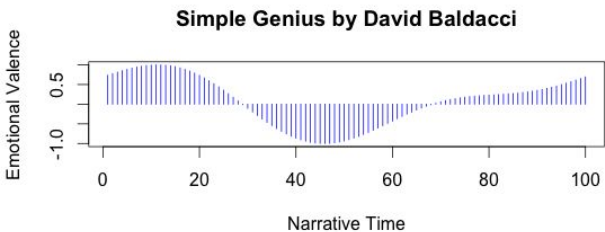
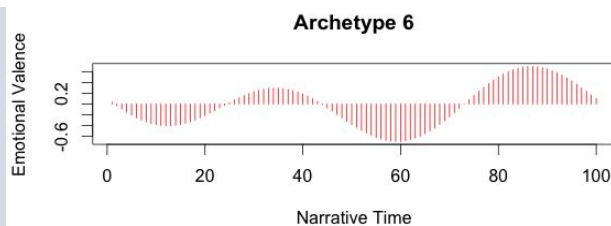
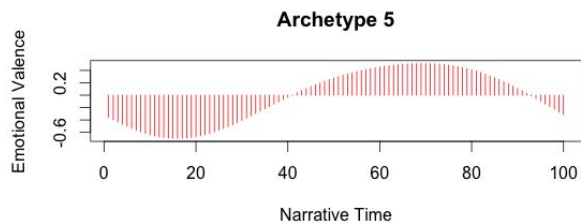
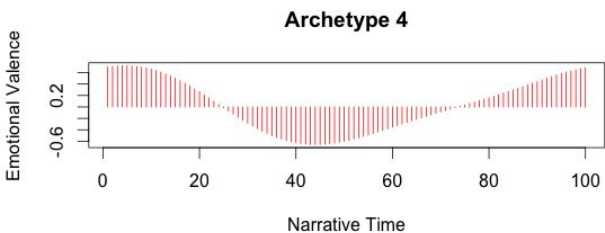
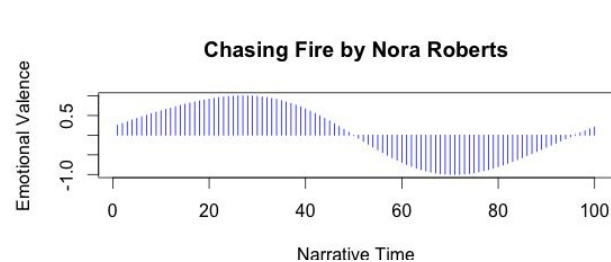
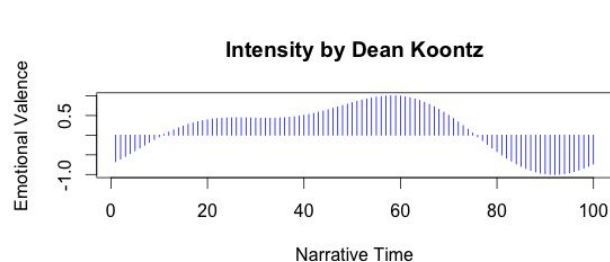
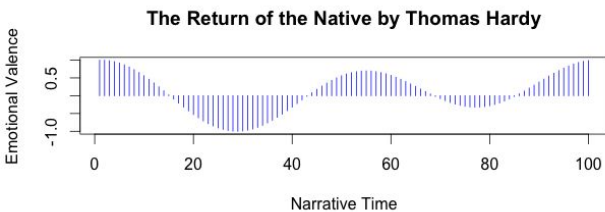
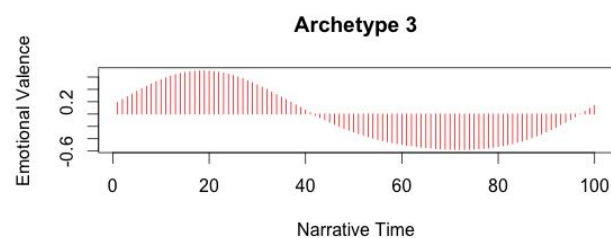
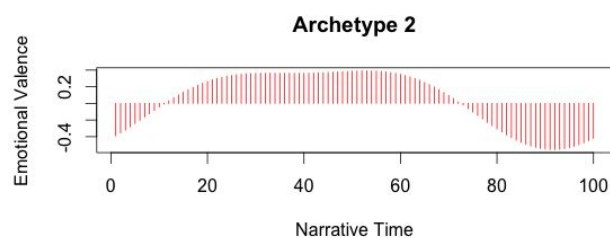
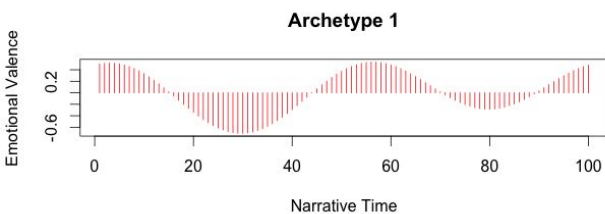


Matthew L. Jockers  
Susan J. Rosowski Associate  
Professor of English  
Department of English  
University of Nebraska-Lincoln  
Lincoln, NE 68588



# Cluster Dendrogram







# The Shapes of Stories

by Kurt Vonnegut



## Man in Hole

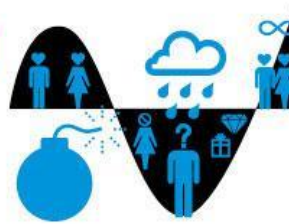


The main character gets into trouble then gets out of it again and ends up better off for the experience.



Arsenic and Old Lace  
Harold & Kumar Go To White Castle

## Boy Meets Girl

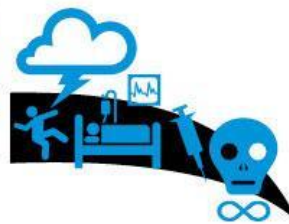


The main character comes across something wonderful, gets it, loses it, then gets it back forever.



Jane Eyre  
Eternal Sunshine of the Spotless Mind

## From Bad to Worse



The main character starts off poorly then gets continually worse with no hope for improvement.



The Metamorphosis  
The Twilight Zone

## Which Way Is Up?



The story has a lifelike ambiguity that keeps us from knowing if new developments are good or bad.



Hamlet  
The Sopranos

## Creation Story



In many cultures' creation stories, humankind receives incremental gifts from a deity. First major staples like the earth and sky, then smaller things like sparrows and cell phones. Not a common shape for Western stories, however.

## Old Testament



Humankind receives incremental gifts from a deity, but is suddenly ousted from good standing in a fall of enormous proportions.



Great Expectations with original ending

## New Testament



Humankind receives incremental gifts from a deity, is suddenly ousted from good standing, but then receives off-the-charts bliss.



Great Expectations with revised ending

## Cinderella



It was the similarity between the shapes of Cinderella and the New Testament that thrilled Vonnegut for the first time in 1947 and then over the course of his life as he continued to write essays and give lectures on the shapes of stories.





# SENTIMENT ANALYSIS - CRITICAL ASPECTS

## From a **theoretical point of view**:

- Jockers called his software «syuzhet», referring to Russian formalism (see Vladimir Propp) and narratology (the “science of narration”)
- However, traditional narratological studies (see Gérard Genette, Mieke Bal, et al.) do not consider emotions at all
- Only Patrick Colm Hogan wrote a book on *Affective Narratology* (2012), but looking at much more complex phenomena than “plot arcs”

## From a **practical point of view**:

- Emotions are subjective (of course!), so their measurement can be unreliable
- SA software are generally unstable: you modify a few parameters, and you get completely different results



# SENTIMENT ANALYSIS

## **SA and cognitive literary studies**

Jacobs, A. M., Schuster, S., Xue, S., and Lüdtke, J. (2017). What's in the brain that ink may character... A quantitative narrative analysis of Shakespeare's 154 sonnets for use in (Neuro-)cognitive poetics. *Scientific Study of Literature*, 7(1): 4-51.

## **SA for the study of secondary literature**

Mellmann, K. and Du, K. (2018). "Sentimentanalyse in Unstrukturierten Texten (Am Bsp. Literaturgeschichtlicher Rezeptionsanalyse)." In DHd 2018 Konferenzabstracts, 305–8. Cologne: Universität zu Köln.

## **SA for the study of social reading**

Rebora, S. and Pianzola F. (2018). A New Research Programme for Reading Research: Analysing Comments in the Margins on Wattpad. *DigitCult - Scientific Journal on Digital Cultures*, 3(2): 19–36

## **SA for Italian language**

Sprugnoli, R., Tonelli, S., Marchetti, A., and Moretti, G. (2016). Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities*, 31(4): 762-772.

## **SA for German language**

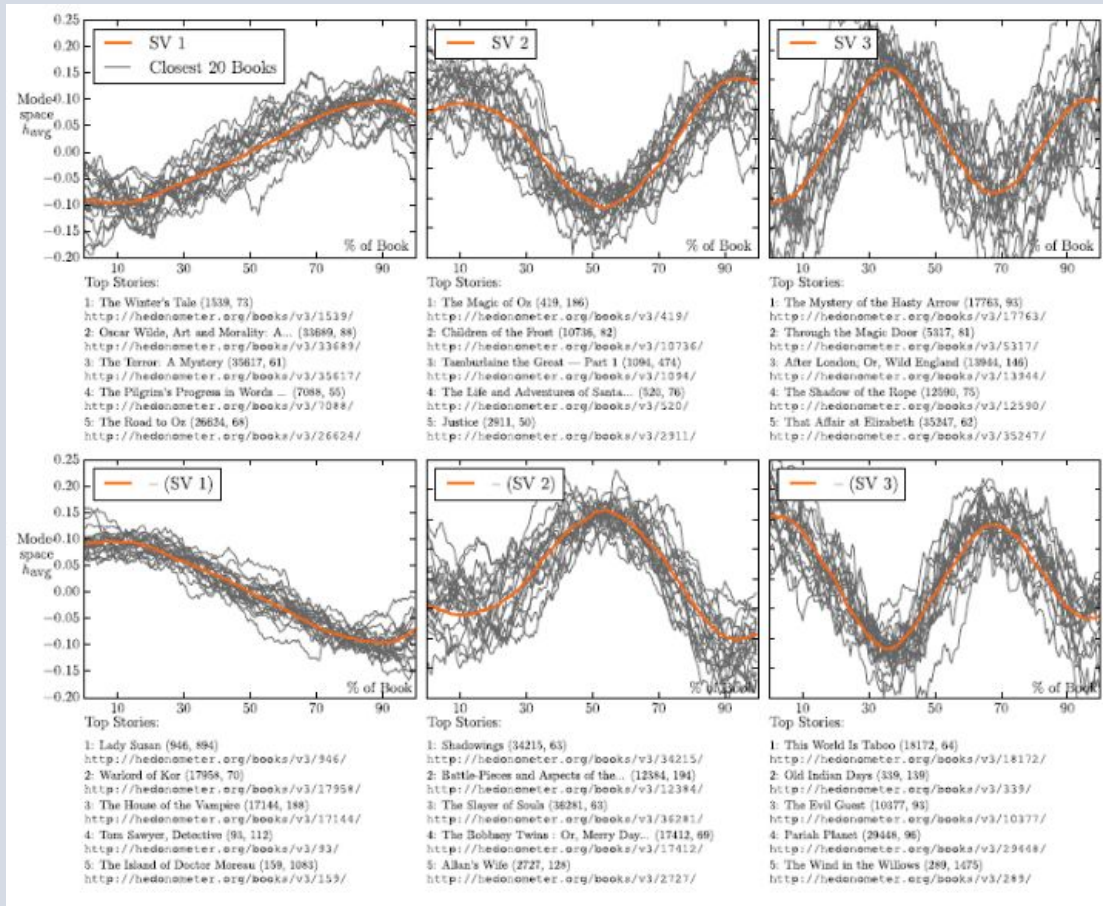
Zehe, A., Becker, M., Jannidis, F., and Hotho, A. (2017). Towards Sentiment Analysis on German Literature. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*. Cham: Springer, pp. 387-394.

## **SA for emotional arcs (again)**

Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., and Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1): 31.



Andrew J Reagan et al. 2016. “The emotional arcs of stories are dominated by six basic shapes.”  
*EPJ Data Science*

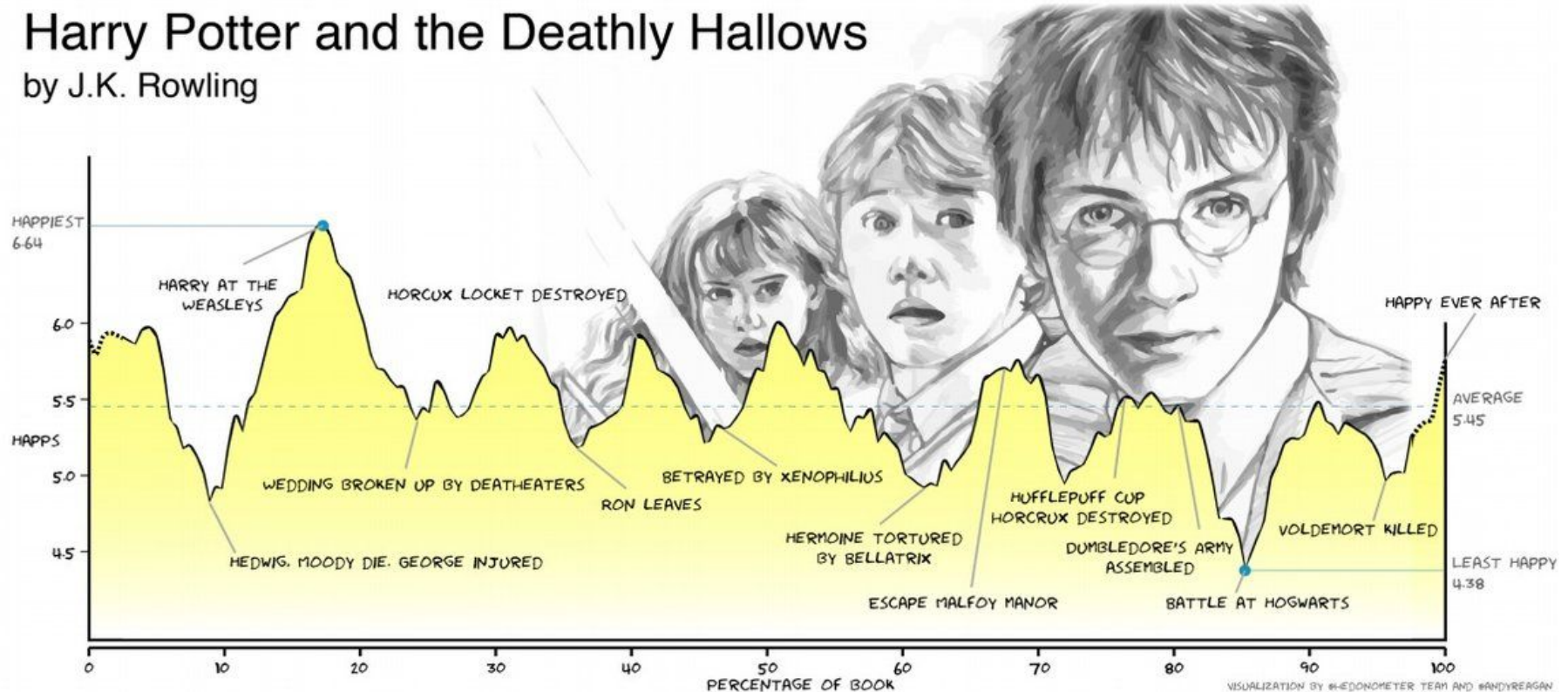




<http://hedonometer.org/>

# Harry Potter and the Deathly Hallows

by J.K. Rowling





# SENTIMENT ANALYSIS A (SIMPLE) TAXONOMY

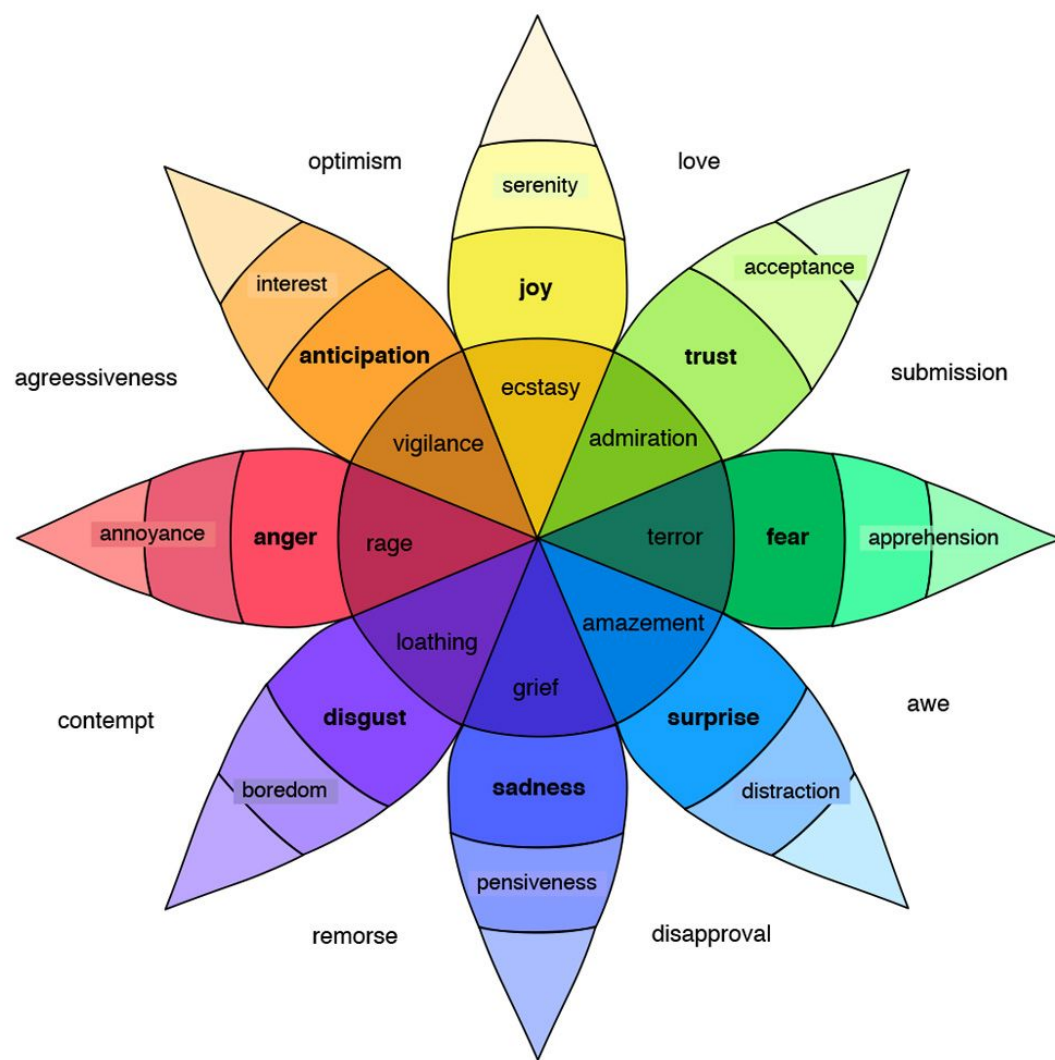
## Three **defining** elements:

- the emotion theory adopted by the tool
- the technique to accomplish the analysis
- the emotion resources



# EMOTION THEORIES

## 1. *Plutchik's eight basic emotions*

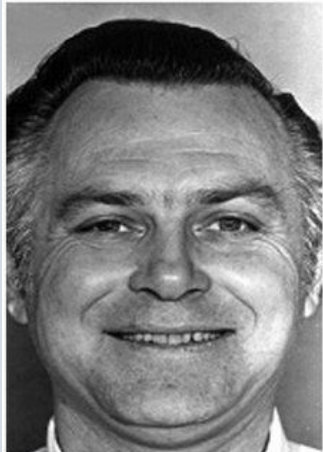
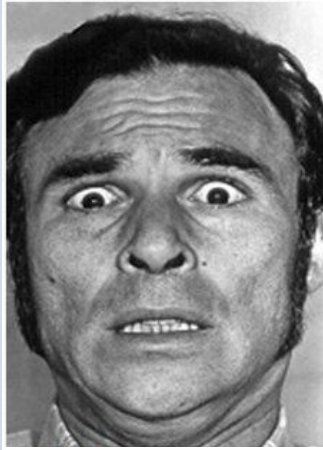






# EMOTION THEORIES

## *2. Ekman's Theory of (Six) Basic Emotions*

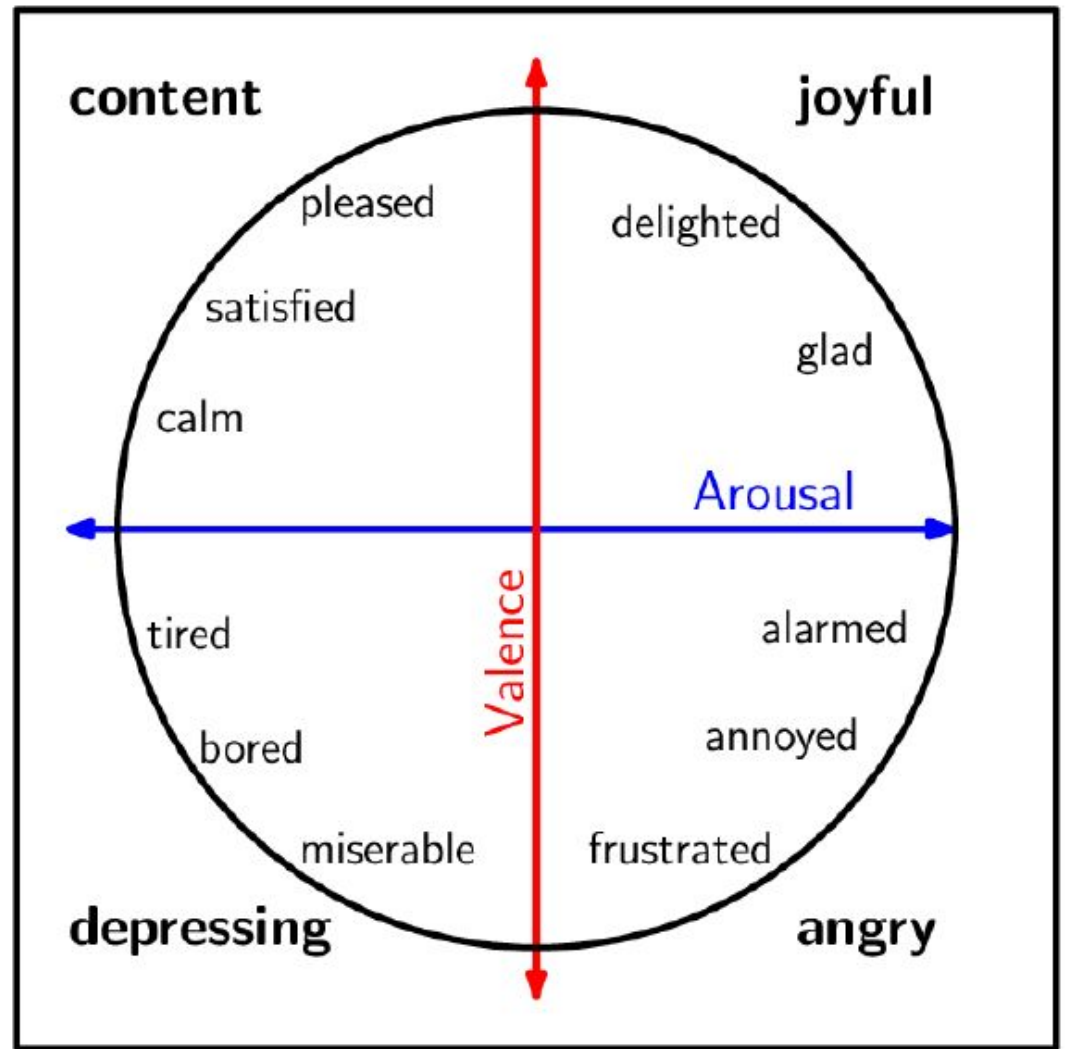






# EMOTION THEORIES

## 3. *Russel's Circumplex Model*





# SA - HOW IT WORKS

Approach\_1: **Syuzhet** (Jockers, 2015)

**Sentiment  
Dictionary**

English Word	Arabic Translation (Google Translate)	Basque Translation (Google Translate)	Positive (Valence)	Negative (Valence)
aback	الى الوراء	aback	0	0
abacus	طبلية تاج	abako	0	0
abandon	تخلي	alde batera utzi	0	1
abandoned	مهجور	abandonatu	0	1
abandonment	هجر	abandono	0	1
abate	انحسر	abate	0	0
abatement	انحسار	murritzeko	0	0
abba	أبا	abba	1	0
abbot	رئيس الدير	abade	0	0
abbreviate	اختصر	laburtu	0	0
abbreviation	الاختصار	laburdura	0	0
abdomen	بطن	abdominalak	NRC-Emotion-Lexicon-v0.92 (Mohammad and Turney, 2013)	0
abdominal	البطن	sabeleko		0
abduction	اختطاف	urrunketa		1
aberrant	شاذ	aberranteari	0	1
aberration	انحراف	aberrazio	0	1
abeyance	abeyance	etena	0	0
abhor	مقت	gaitzesten	0	1
abhorrent	مشمئز	nazkagarria	0	1



# SA - HOW IT WORKS

## Approach\_1: Syuzhet (Jockers, 2015)

**Sentiment  
Dictionary**



**Text**

He was a happy man



**Sentiment**

**+1**



# SA - HOW IT WORKS

Approach\_1: **Syuzhet** (Jockers, 2015)

**Sentiment  
Dictionary**



**Text**

He was a **sad** and  
**boring** man



**Sentiment**

**-2**



# SA - HOW IT WORKS

Approach\_1: **Syuzhet** (Jockers, 2015)

**Sentiment  
Dictionary**



**Text**

He was neither **sad**  
nor **boring**



**Sentiment**

**-2**



# SA - HOW IT WORKS

## Approach\_1: Syuzhet (Jockers, 2015)

**Sentiment  
Dictionary**



**Text**

Well, he was like a  
potato



**Sentiment**

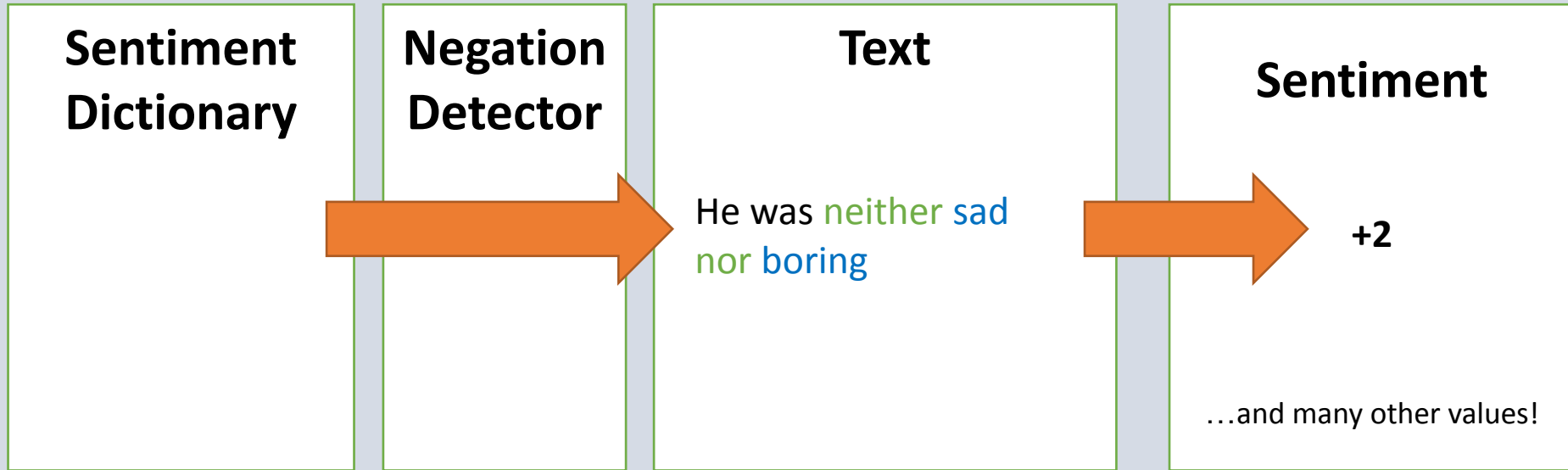
+2





# SA - HOW IT WORKS

Approach\_2: **SEANCE** (Crossley et al., 2017)



Number	Index	Variable description	POS	Negation Filtered	examples (not POS specific)
	1filename	name of source file	NA	NA	NA
	2nwords	number of words in source file	NA	NA	NA
	3Admiration/Awe_GALC	Admiration/Awe	All	No	admir*, ador*, awe*, dazed, dazzl*
	4Amusement_GALC	Amusement	All	No	amus*, fun*, humor*, laugh*, play*
	5Anger_GALC	Anger	All	No	anger, angr*, cross*, enrag*, furious
	6Anxiety_GALC	Anxiety	All	No	anguish*, anxi*, apprehens*, diffiden*, jitter*
	7Beingtouched_GALC	Beingtouched	All	No	affect*, mov*, touch*
	8Boredom_GALC	Boredom	All	No	bor*, ennui, indifferen*, languor*, tedi*
	9Compassion_GALC	Compassion	All	No	commiser*, compass*, empath*, pit*,
	10Contempt_GALC	Contempt	All	No	contempt*, denigr*, deprec*, deris*, despi*
	11Contentment_GALC	Contentment	All	No	comfortabl*, content*, satisf*
	12Desperation_GALC	Desperation	All	No	deject*, desolat*, despair*, desperat*, despond*
					comedown, disappoint*, discontent*, disenchant*, disgruntl*
	13Disappointment_GALC	Disappointment	All	No	
	14Disgust_GALC	Disgust	All	No	abhor*, avers*, detest*, disgust*, dislik*
	15Dissatisfaction_GALC	Dissatisfaction	All	No	dissatisf*, unhapp*
	16Envy_GALC	Envy	All	No	envious*, envy*
	17Fear_GALC	Fear	All	No	afraid*, aghast*, alarm*, dread*, fear*
	18Feelinglove_GALC	Feelinglove	All	No	affection*, fond*, love*, friend*, tender*
	19Gratitude_GALC	Gratitude	All	No	grat*, thank*
	20Guilt_GALC	Guilt	All	No	blame*, contriti*, guilt*, remorse*, repent*
	21Happiness_GALC	Happiness	All	No	cheer*, bliss*, delect*, delight*, enchant*
	22Hatred_GALC	Hatred	All	No	acrimon*, hat*, rancor*



# SA - HOW IT WORKS

## Approach\_2: SEANCE (Crossley et al., 2017)

**Sentiment  
Dictionary**

**Negation  
Detector**

**Text**

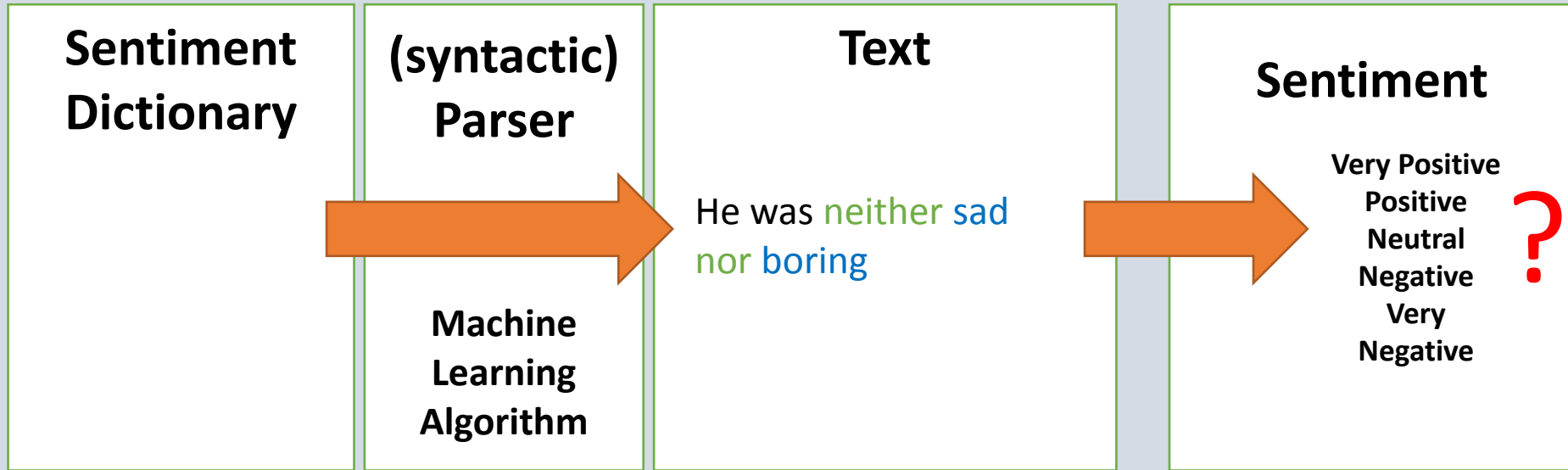
He was neither sad  
nor boring

Admiration/Awe_GALC	1
Amusement_GALC	0
Anger_GALC	1
Anxiety_GALC	2
Beingtouched_GALC	-1
Boredom_GALC	0
Compassion_GALC	1
Contempt_GALC	1
Contentment_GALC	1
Desperation_GALC	0
Disappointment_GALC	0
Disgust_GALC	0
Dissatisfaction_GALC	1
Envy_GALC	0
Fear_GALC	1
Feelinglove_GALC	2
Gratitude_GALC	-1
Guilt_GALC	0
Happiness_GALC	1
Hatred_GALC	1
Hope_GALC	1
Humility_GALC	0
Interest/Enthusiasm_GALC	0
Irritation_GALC	0
Jealousy_GALC	1
Joy_GALC	0
Longing_GALC	1
Lust_GALC	2
Pleasure/Enjoyment_GALC	-1
Pride_GALC	0
Relaxation/Serenity_GALC	1
Relief_GALC	1
Sadness_GALC	1
Shame_GALC	0
Surprise_GALC	0
Tension/Stress_GALC	0

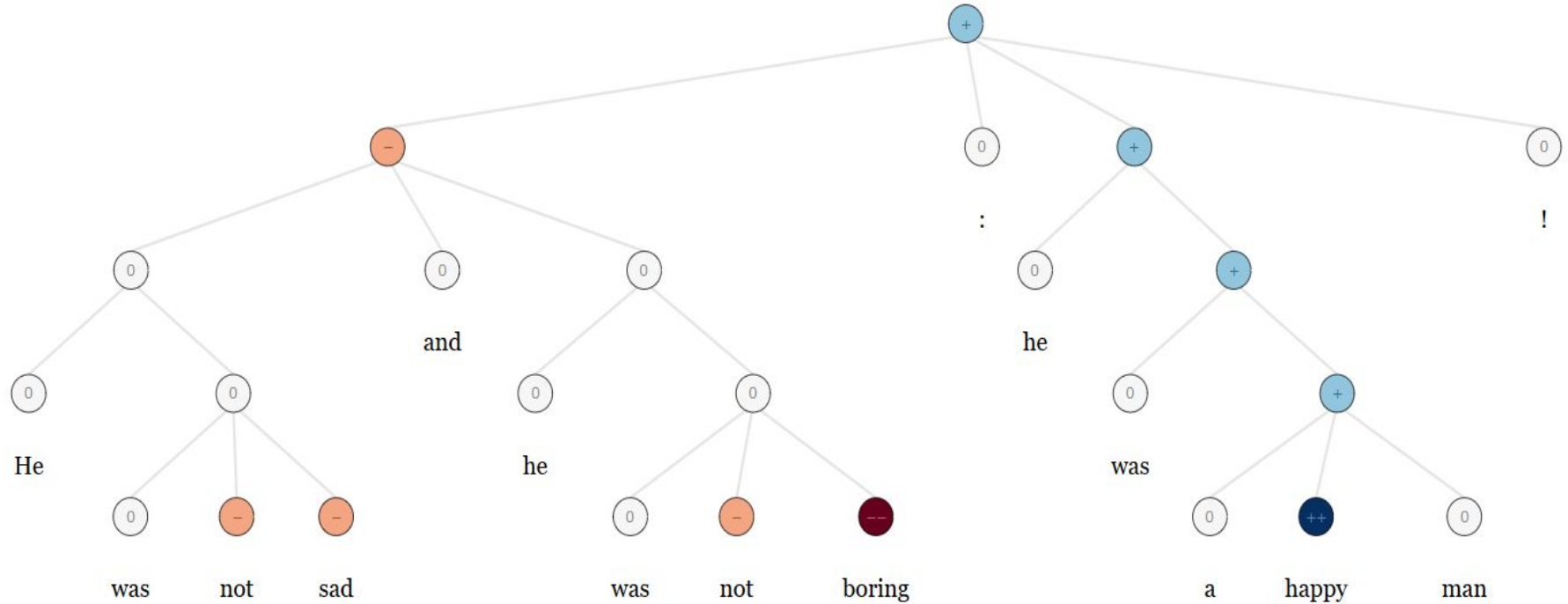


# SA - HOW IT WORKS

Approach\_3: **Stanford SA** (Socher et al., 2013)



He was not sad and he was not boring: he was a happy man!



All labels are now correct



# SA RESOURCES

## **Sentiment dictionaries**

- manually encoded (by experts, or via crowdsourcing)
- automatically created

## **Annotated datasets**

- for aspect-based SA
- for machine learning



# CROWDSOURCING



See NRC Lexicon (Mohammad and Turney, 2013)

## Issue 1.

“The task and compensation may attract cheaters (who may input random information) and even malicious annotators (who may deliberately enter incorrect information).”

Solution: control questions

Prompt word: *startle*

Q1. Which word is closest in meaning (most related) to *startle*?

- *automobile*
- *shake*
- *honesty*
- *entertain*

Q2. How positive (good, praising) is the word *startle*?

CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON

11

- *startle* is not positive
- *startle* is weakly positive
- *startle* is moderately positive
- *startle* is strongly positive

Q3. How negative (bad, criticizing) is the word *startle*?





# CROWDSOURCING



See NRC Lexicon (Mohammad and Turney, 2013)

Issue 2.

Disagreement between different annotators

Solution: majority rule

Emotion	Fleiss's $\kappa$	Interpretation
anger	0.39	fair agreement
anticipation	0.14	slight agreement
disgust	0.31	fair agreement
fear	0.32	fair agreement
joy	0.36	fair agreement
sadness	0.39	fair agreement
surprise	0.18	slight agreement
trust	0.24	fair agreement
<b>micro-average</b>	<b>0.29</b>	<b>fair agreement</b>



# EXPERT WORK

See LIWC Dictionary (Tausczik and Pennebaker, 2010)

[HOW IT WORKS](#)[COMPARE VERSIONS](#)[COMPARE DICTIONARIES](#)[INTERPRETING LIWC](#)[LIWC API](#)[CONTACT US](#)[BUY NOW](#)

## HOW IT WORKS

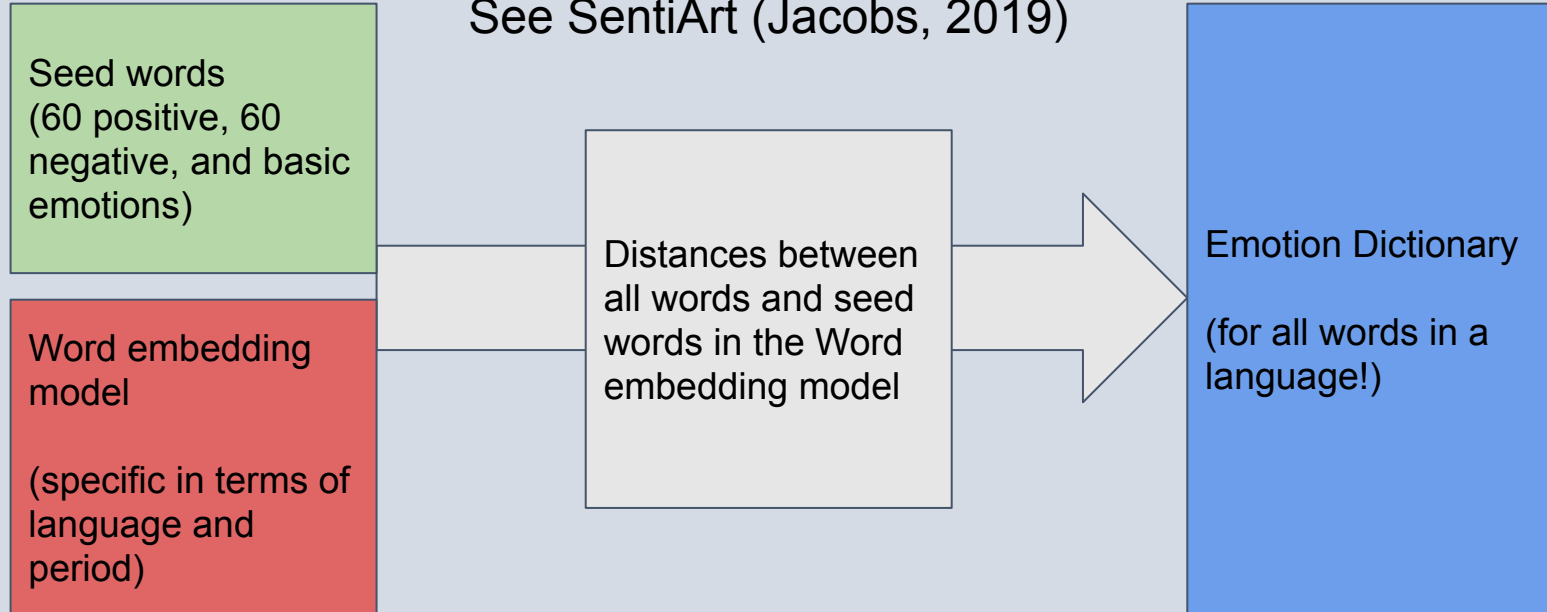
The way that the Linguistic Inquiry and Word Count (LIWC) program works is fairly simple. Basically, it reads a given text and counts the percentage of words that reflect different emotions, thinking styles, social concerns, and even parts of speech. Because LIWC was developed by

researchers with interests in social, clinical, health, and cognitive psychology, the language categories were created to capture people's social and psychological states.



# AUTOMATIC

See SentiArt (Jacobs, 2019)





# ANNOTATED DATASETS

## Question:

Are emotions expressed by single words, or by parts of a text?  
(sentences, clauses, expressions...)



# ASPECT-BASED SA

See REMAN Dataset (Kim & Klinger, 2018), 1720 sentences from novels in English (downloaded from Project Gutenberg)

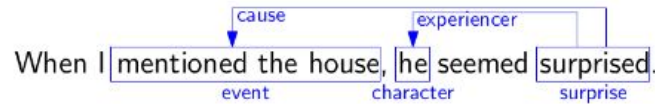


Figure 1: Example annotation from Hugo (1885), with one character, an emotion word, and event and cause and experiencer annotations.

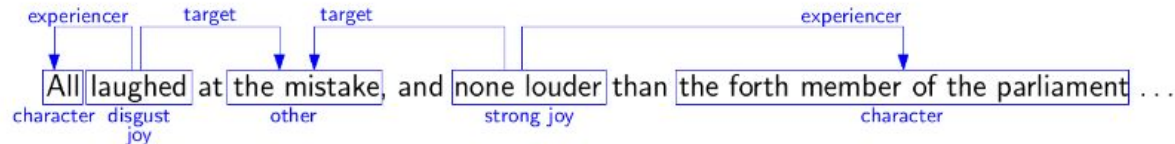


Figure 2: Example annotation from Stimson (1943), with two characters who are experiencers of different emotions. Disgust and joy are annotated as a mixture of emotions. Both emotions have the same target.



# ASPECT-BASED SA

## Distant Reading Swiss Literature



Ernst Ludwig Kirchner (1880-1938):  
Rückkehr der Tiere, 1919

[View My GitHub Profile](#)

[members](#)

[publications](#)

[conferences &  
talks](#)

[coding tutorials and  
scripts](#)

### Welcome to our webpage!

We are a team of researchers interested in the computational investigation of German-Swiss literature around 1900. Our field is digital humanities, with a focus on the application of sentiment analysis and named entity recognition.

### *High Mountains Low Arousal? Distant Reading Topographies of Sentiment in German Swiss Novels in the early 20th Century*

- Swiss National Science Foundation (SNSF)-COST-Project, see [SNSF-database](#)
- Working Group Prof. Berenike Herrmann at [Bielefeld University](#)

### Summary

The Bielefeld-based project “High Mountains Low Arousal?” works in close collaboration with the International COST Action “[Distant Reading the European Novel](#).” By means of a distant reading focusing on sentiment and emotion in the fictional spaces represented in German-Swiss novels, it aims at pioneering comparative historiographical and systematic research on the German-language novel of the early 20th Century, using digital



# MACHINE LEARNING

## Examples

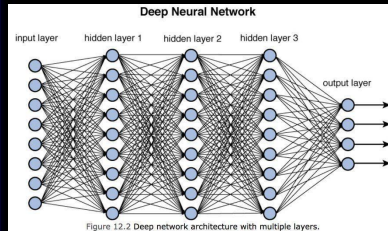
"He was  
terrified by  
machine  
learning"

**FEAR**

"She was glad"

**JOY**

## The ML "black box"



## New texts

"I am scared by  
monsters"

"He's happy to  
be a monster"





# MACHINE LEARNING

## Examples

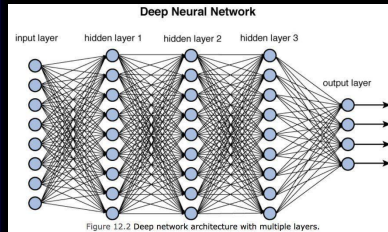
"He was  
terrified by  
machine  
learning"

**FEAR**

"She was glad"

**JOY**

## The ML "black box"



## New texts

"I am scared by  
monsters"

**FEAR**

"He's happy to  
be a monster"

**JOY**



# SA EVALUATION

- Even the most advanced SA tools hardly reach 90% accuracy in distinguishing positive from negative emotions
- Accuracy drops to below 50%, when distinguishing more fine-grained emotions (Ekman, Plutchick, etc.)

**(Rojas-Barahona, 2016)**