

Экспрессивы

Функциональные модели в естественном языке

Экспрессивы

В формальной семантике:

1. *Independence*: Expressive content contributes a dimension of meaning that is separate from the regular descriptive content.
2. *Nondisplaceability*: Expressives predicate something of the utterance situation.
3. *Perspective dependence*: Expressive content is evaluated from a particular perspective. In general, the perspective is the speaker's, but there can be deviations if conditions are right.
4. *Descriptive ineffability*: Speakers are never fully satisfied when they paraphrase expressive content using descriptive, i.e., nonexpressive, terms.
5. *Immediacy*: Like performatives, expressives achieve their intended act simply by being uttered; they do not offer content so much as inflict it.
6. *Repeatability*: If a speaker repeatedly uses an expressive item, the effect is generally one of strengthening the emotive content, rather than one of redundancy.

Экспрессивы

В формальной семантике:

- (45) a. **damn** : $\langle e, \varepsilon \rangle$
b. $\llbracket \mathbf{damn} \rrbracket^c$ = the function f such that $f(\llbracket a \rrbracket^c)(c) = c'$, where
- $c \approx_{cJ, \llbracket a \rrbracket^c}^{\mathbf{I}} c'$;
 - the length of \mathbf{I} is not more than .5; and
 - $\mathbf{I} \subseteq [-1, 0]$
- (47) a. **fucking** : $\langle e, \varepsilon \rangle$
b. $\llbracket \mathbf{fucking} \rrbracket^c$ = the function f such that $f(\llbracket a \rrbracket^c)(c) = c'$,
- $c \approx_{cJ, \llbracket a \rrbracket^c}^{\mathbf{I}} c'$;
 - the length of \mathbf{I} is not more than .2; and
 - $\mathbf{I} \subseteq [-1, 0]$

Экспрессивы и восклицания

- Constant et al., 2009; Potts and Schwarz, 2008

Экспрессивы

Rating: 1 star

Review: I bought this book and read it half way through. It was soo damn boring to read. Just opinions on what caused what. It was absolutely dreadful and was a good way to fall asleep. Sold it back on amazon for 8 bux.

(a) A short one-star review

Rating: 3 star

Review: This is a good book to compare the various menu items and products for their calorie values. The one problem I have with it is that it only shows 3-4 good items and 3-4 bad items per restaurant/store/manufacturer. I would rather have had this broken into a series of books and gone into more detail.

(b) A short three-star review

Rating: 5 star

Review: I never had an opinion either way about Tori. I thought she was fun in Scary Movie 2, and then got totally addicted to So NoTORious. I couldn't believe how down to earth and self-deprecating she was, and it seemed so sincere. Her book confirmed that too! And I learned that in the end, little girls (rich or poor) are the same, and all we really want is that damn Barbie dream house :-)

(c) A short five-star review

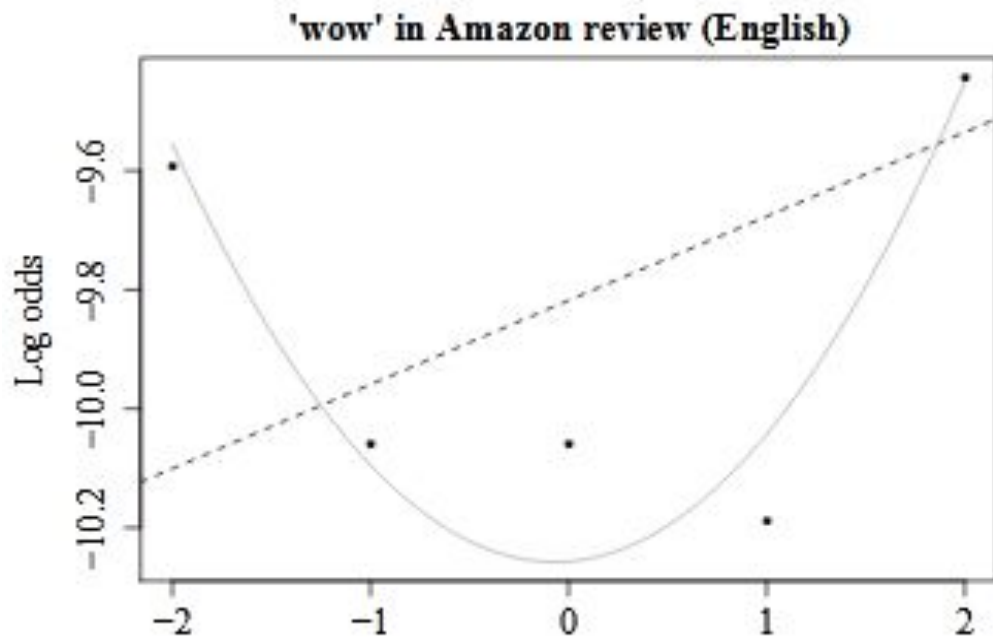
Figure 1: Example reviews from Amazon.com

Экспрессивы

Our basic perspective is the log-odds distribution, as in (1), where $\text{count}(x_n, R)$ is the number of tokens of x_n (a word-string of length n) in rating category R , and $\text{count}_n(R)$ is the number of tokens of word-strings of length n in rating category R .

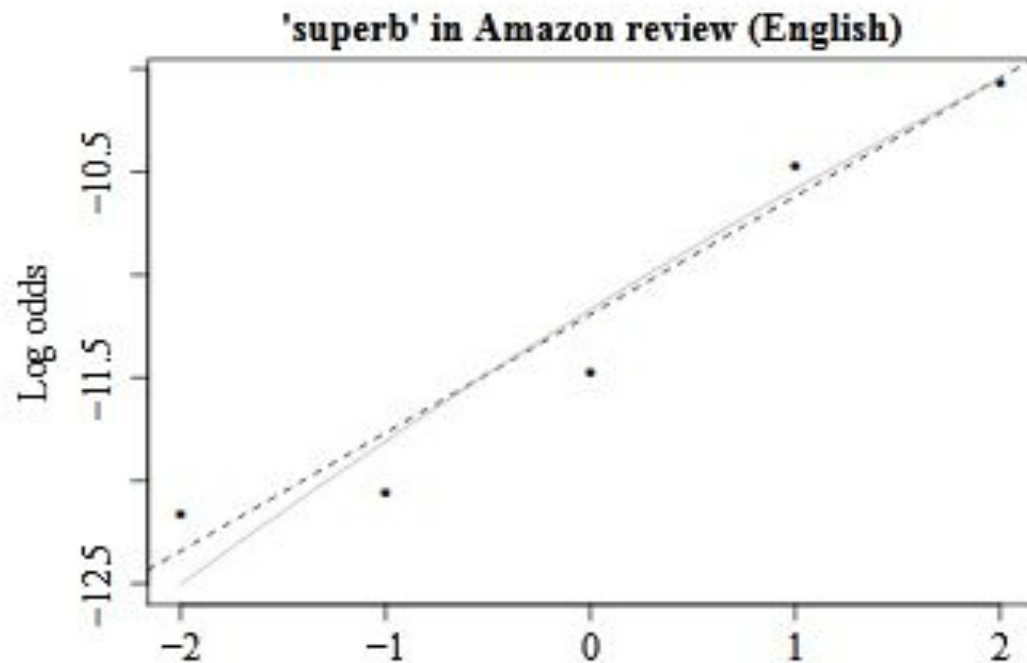
$$(1) \quad \text{log-odds}(x_n, R) \stackrel{\text{def}}{=} \ln \left(\frac{\text{count}(x_n, R)}{\text{count}_n(R) - \text{count}(x_n, R)} \right)$$

Экспрессивы



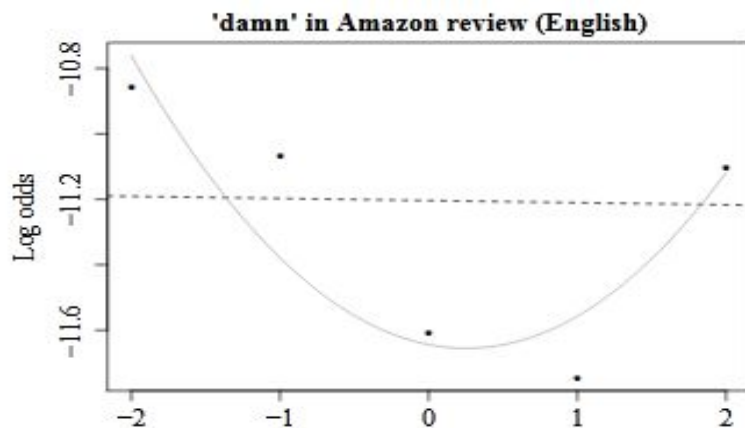
Constant et al., 2009

Экспрессивы



Constant et al., 2009

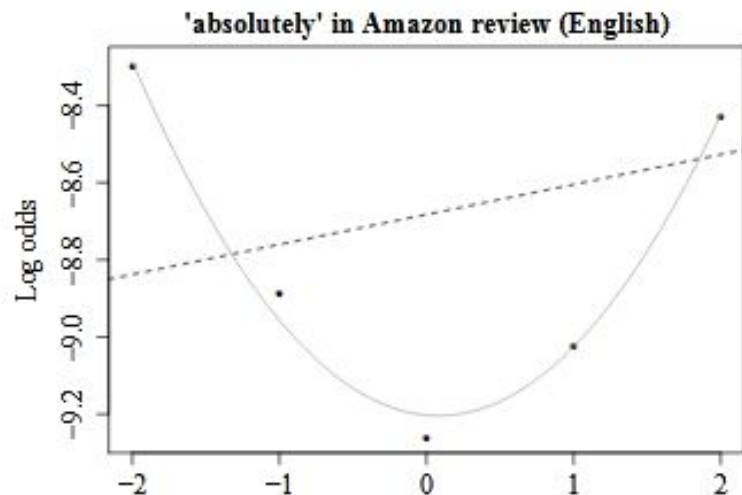
Экспрессивы



- a. Try answering the damn question.
- b. Sounds like another damn politician to me.
- c. Just a damn minute! What history books did you read?

- a. funnest damn movie i've seen all year⁶
- b. Linnea Faris, a woman from Michigan who was wearing a "Remember Alex" T-shirt, shook her head in disbelief. [...] "I've spent hours crying over that damn bird."⁷

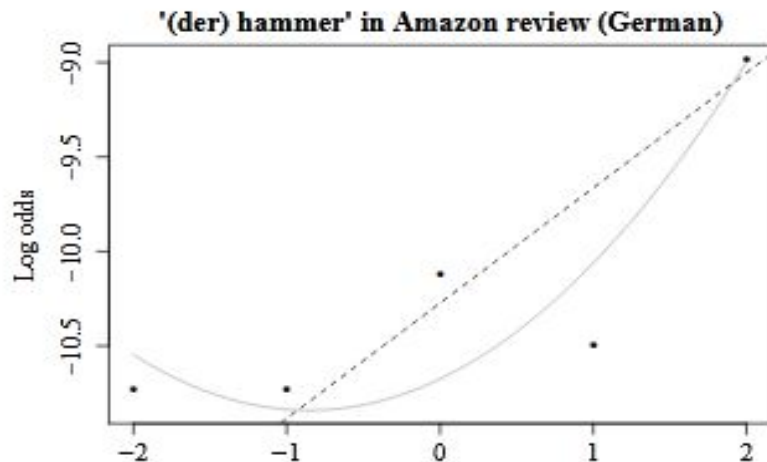
Экспрессивы



- a. I could absolutely/totally/SO jump over that fence.
- b. MTV like totally gave us TWO episodes back to back. It was like so random. The more the merrier, but it's like waay too much for one recap. (Web example from Beaver and Clark 2008: 74)

Экспрессивы: немецкий

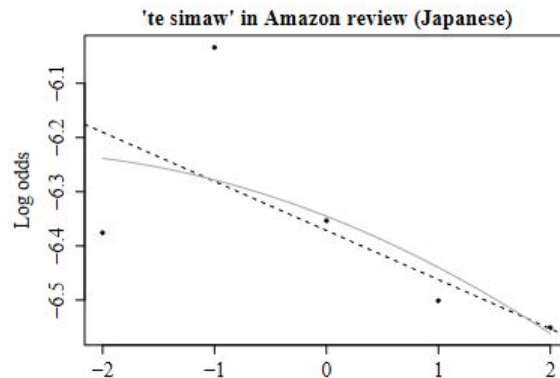
- a. Dieses Album ist **der Hammer**
This album is **the hammer**
'This album is cool.' (\approx 'the bomb', colloquially)
- b. amys stimme mischt sich mit **hammer** beats
Amy's voice mixed self with **cool** beats
'Amy's voice mixed with cool beats.'
- c. Was für ein **Hammer** Album
what for a **cool** album
'What a cool album!'



Экспрессивы: японский

- (11) harii pottaa to wa nagai o-tukiai de, owat-**tesimat**-te totemo kanasii
Harry Potter with TOPIC long HON-acquaintance be end-ANTI-HON-INF very sad
omoi-ga si-masu
thought-NOM do-HON
'I had a long acquaintance with the Harry Potter series, and was sad to see it end.'

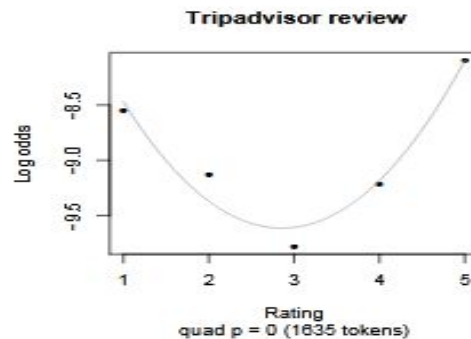
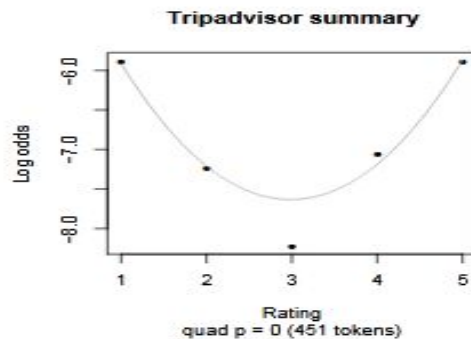
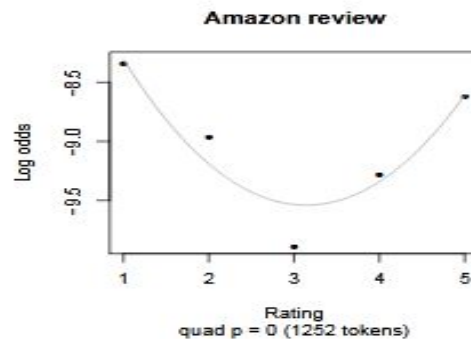
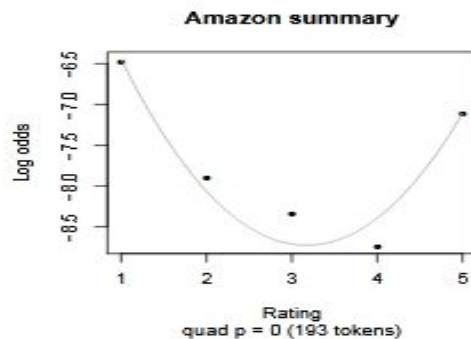
- (12) nesugoshi-**chimat**-ta.
overslept-ANTI-HON-PAST
'I overslept, which sucks.'

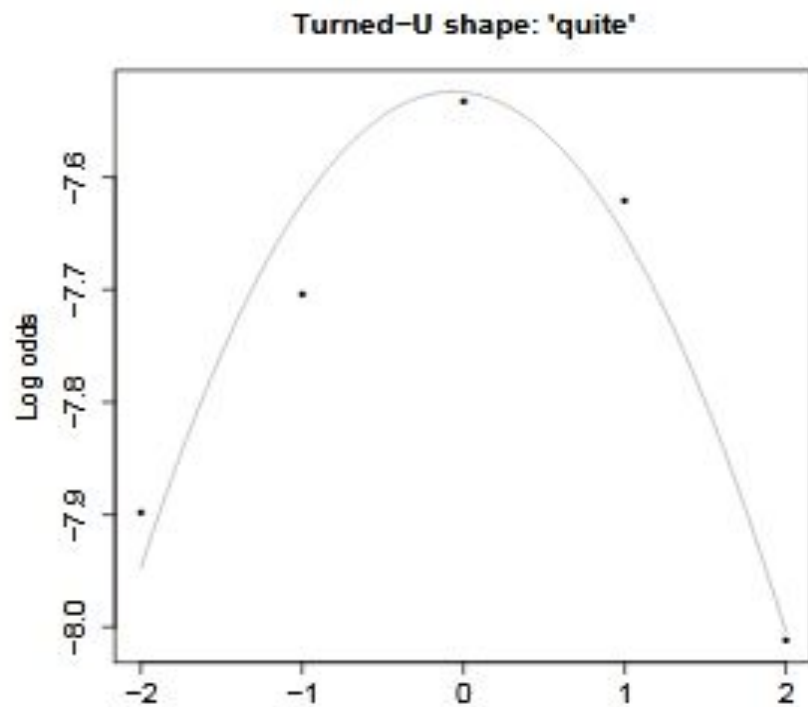
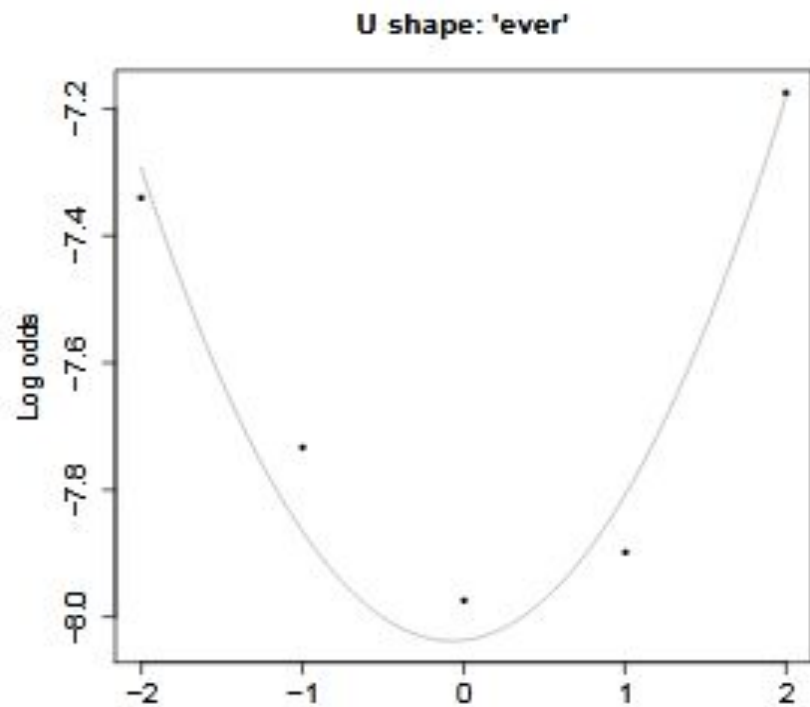


Восклицания

- What a Find!
- What an awesome place!!
- What a mess!
- What an Overpriced Dump

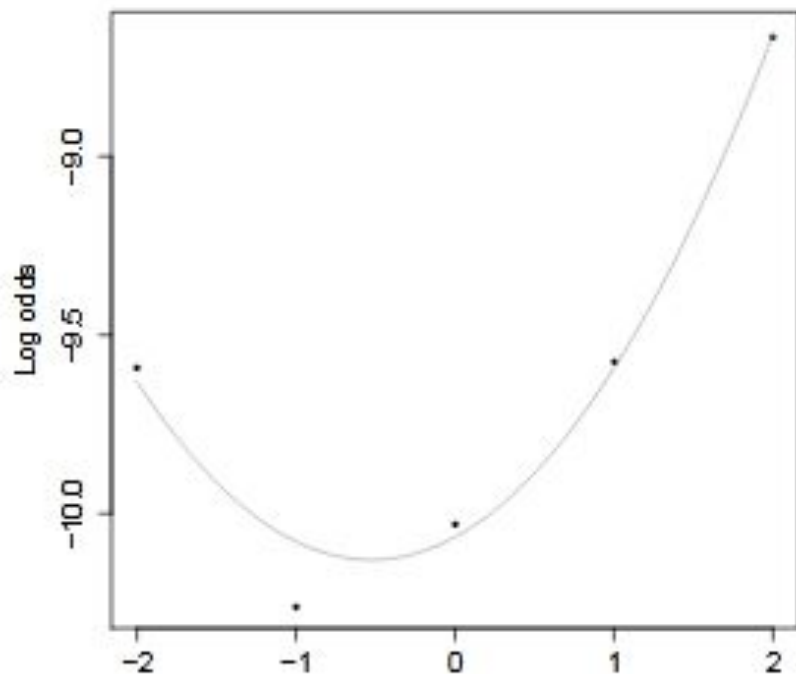
(from a five-star Tripadvisor review)
(from a five-star Tripadvisor review)
(from a one-star Tripadvisor review)
(from a one-star Tripadvisor review)



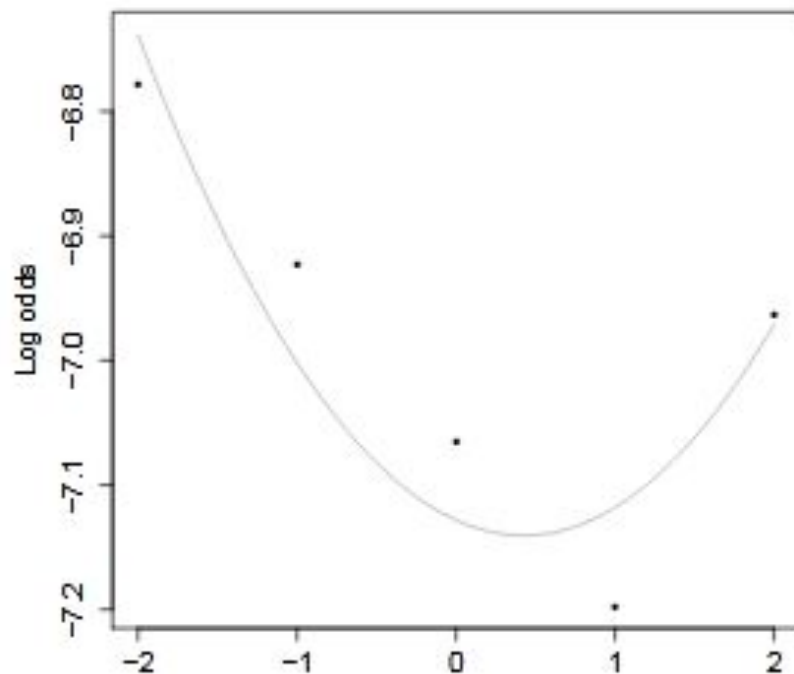


Восклицания

J shape: 'thanks'



Reverse-J shape: 'never'



Восклицания

!!	!	absolutely	all
best	ever	every	i've ever
i've	i	it !	my
the best	this	what a	wow

(a) The items whose shapes are limited to U, J, and Reverse J for all four corpora.

Восклицания

!!	!	absolute	absolutely
again !	all	am	any
anyone	best	book	couldn't
even	ever !	ever	ever had
every	have ever	i	i am
i could	i have	i've	i've ever
is the	it !	life	must
my	never	new	one of
simply	the best	this	this is
time	what	what a	will
will never	wow !	wow	

(b) The items whose shapes are limited to U, J, and Reverse J in at least three of the four corpora.

Восклицания

- but	a few	a good	average
basic	but nothing	but some	but still
cons	decent	few	fine
little too	mostly	near the	not bad
not quite	otherwise	part	points
pretty	pros	quite	short
some	somewhat	though	two
with some			

(c) The items whose shapes are limited to Turned U in at least three of the four corpora.

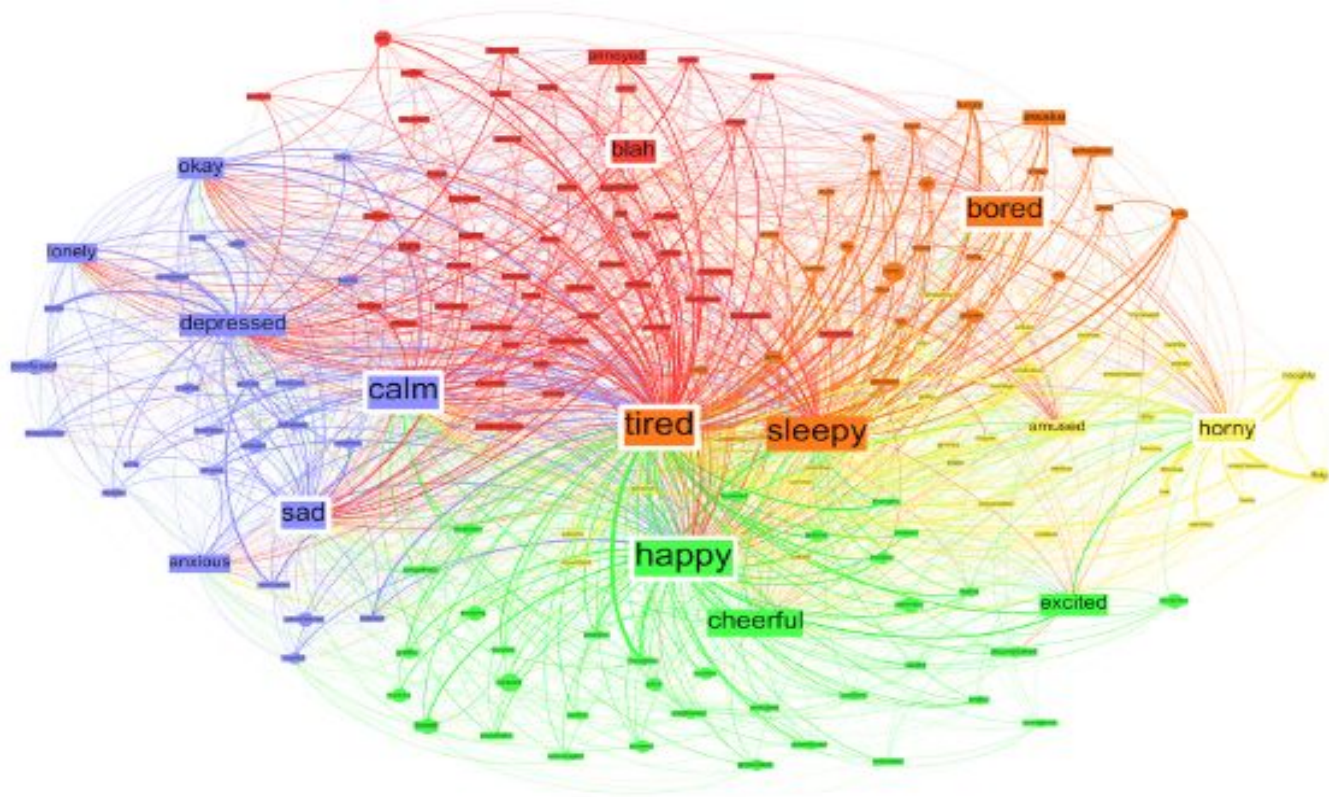
Sentiment analysis

Концептуальные проблемы

Какие из следующих высказываний выражают отношение (sentiment)? Какова полярность отношения (sentiment polarity: pos/neg), если оно выражено? Достаточно ли нам градации pos/neg?

1. There was an earthquake in California.
2. The team failed to complete the physical challenge. (We win/lose!)
3. They said it would be great.
4. They said it would be great, and they were right.
5. They said it would be great, and they were wrong.
6. The party fat-cats are sipping their expensive imported wines.
7. Oh, you're terrible!
8. Её муж – какой-то сантехник.
9. Отличное кино!
10. Опять надо выгуливать эту чертову собаку!
11. Of 2001, “Many consider the masterpiece bewildering, boring, slow-moving or annoying, . . .”
12. long-suffering fans, bittersweet memories, hilariously embarrassing moments, . . .

Измерения аффекта (valence=polarity; arousal = intensity), отношения между элементами, и переходы



Связанные задачи

- Subjectivity (Pang and Lee 2008)
- Bias (Recasens et al. 2013; Pryzant et al. 2020)
- Stance (Anand et al. 2011)
- Hate-speech (Nobata et al. 2016)
- Microaggressions (Breitfeller et al. 2019)
- Condescension (Wang and Potts 2019)
- Sarcasm (Khodak et al. 2017)
- Deception and betrayal (Niculae et al. 2015)
- Online trolls (Cheng et al. 2017)
- Polarization (Gentzkow et al. 2019)
- Politeness (Danescu-Niculescu-Mizil et al. 2013)
- Linguistic alignment (Doyle et al. 2016)

Некоторые релевантные дейтасеты:

- IMDb movie reviews (50K) (Maas et al. 2011):

<http://ai.stanford.edu/~amaas/data/sentiment/index.html>

- Datasets from Lillian Lee's group:

<http://www.cs.cornell.edu/home/llee/data/>

- Datasets from Bing Liu's group:

<https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

- RateBeer (McAuley et al. 2012; McAuley and Leskovec 2013):

<http://snap.stanford.edu/data/web-RateBeer.html>

- Amazon Customer Review data:

<https://s3.amazonaws.com/amazon-reviews-pds/readme.html>

Amazon Product Data (McAuley et al. 2015; He and McAuley 2016):

<http://jmcauley.ucsd.edu/data/amazon/>

- Sentiment and social networks together (West et al. 2014)

<http://infolab.stanford.edu/~west1/TACL2014/>

- Stanford Sentiment Treebank (SST; Socher et al. 2013)

<https://nlp.stanford.edu/sentiment/>

https://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf

Лексика

- Bing Liu's Opinion Lexicon:
nltk.corpus.opinion_lexicon
- SentiWordNet: nltk.corpus.sentiwordnet
- MPQA subjectivity lexicon:
<http://mpqa.cs.pitt.edu>
- Harvard General Inquirer
- Linguistic Inquiry and Word Counts (LIWC): <https://liwc.wpengine.com>
- Hamilton et al. (2016): SocialSent
<https://nlp.stanford.edu/projects/socialsent/>
- Brysbaert et al. (2014): Norms of valence, arousal, and dominance for 13,915 English lemmas

Download:

http://www.wjh.harvard.edu/~inquirer/spreadsheet_guide.htm

Documentation:

<http://www.wjh.harvard.edu/~inquirer/homecat.htm>

Литература 1

- Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 1–9, Portland, Oregon. Association for Computational Linguistics.
- Luke Breitfeller, Emily Ahn, Aldrian Obaja Muis, David Jurgens, and Yulia Tsvetkov. 2019. Finding microaggressions in the wild: A case for locating elusive phenomena in social media posts. In *Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics.
- Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3):904–911.
- Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17*, pages 1217–1230, New York, NY, USA. ACM.
- Noah Constant; Christopher Davis; Christopher Potts; and Florian Schwarz. 2009. [The pragmatics of expressive content: Evidence from large corpora](#). *Sprache und Datenverarbeitung* 33(1-2):5-21.
- Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 2013 Annual Conference of the Association for Computational Linguistics*, pages 250–259, Stroudsburg, PA. Association for Computational Linguistics.
- Gabriel Doyle, Dan Yurovsky, and Michael C. Frank. 2016. A robust framework for estimating linguistic alignment in twitter conversations. In *Proceedings of the 25th International World Wide Web Conference, WWW '16*, pages 637–648, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 595–605, Austin, Texas. Association for Computational Linguistics.

Литература 2

- Matthew Gentzkow, Jesse M. Shapiro, and Matt Taddy. 2019. Measuring group differences in high-dimensional choices: Method and application to congressional speech. Ms, Stanford University, Brown University and Amazon.
- Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. 2017. A large self-annotated corpus for sarcasm. *arXiv preprint arXiv:1704.05579*.
- Vlad Niculae, Srikanth Kumar, Jordan Boyd-Graber, and Cristian Danescu-Niculescu-Mizil. 2015. Linguistic harbingers of betrayal: A case study on an online strategy game. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1650–1659, Beijing, China. Association for Computational Linguistics.
- Chikashi Nobata, Joel Tetreault, Achint Thomas, Yashar Mehdad, and Yi Chang. 2016. Abusive language detection in online user content. In *Proceedings of the 25th International Conference on World Wide Web, WWW '16*, pages 145–153, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- Christopher Potts. 2007. [The expressive dimension](#). *Theoretical Linguistics* 33(2):165-197.
- Christopher Potts and Florian Schwarz. 2008. [Exclamatives and heightened emotion: Extracting pragmatic generalizations from large corpora](#). Ms., UMass Amherst.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1):1–135.
- Reid Pryzant, Richard Diehl Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, and Diyi Yang. 2020. Automatically neutralizing subjective bias in text. In *Proceedings of AAAI*.
- Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659, Sofia, Bulgaria. Association for Computational Linguistics.
- Moritz Sudhof, Andrés Gómez Emilsson, Andrew L. Maas, and Christopher Potts. 2014. Sentiment expression conditioned by affective transitions and social forces. In *Proceedings of 20th Conference on Knowledge Discovery and Data Mining*, pages 1136–1145, New York. ACM.
- Zijian Wang and Christopher Potts. 2019. TalkDown: A corpus for condescension detection in context. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3702–3710, Stroudsburg, PA. Association for Computational Linguistics.