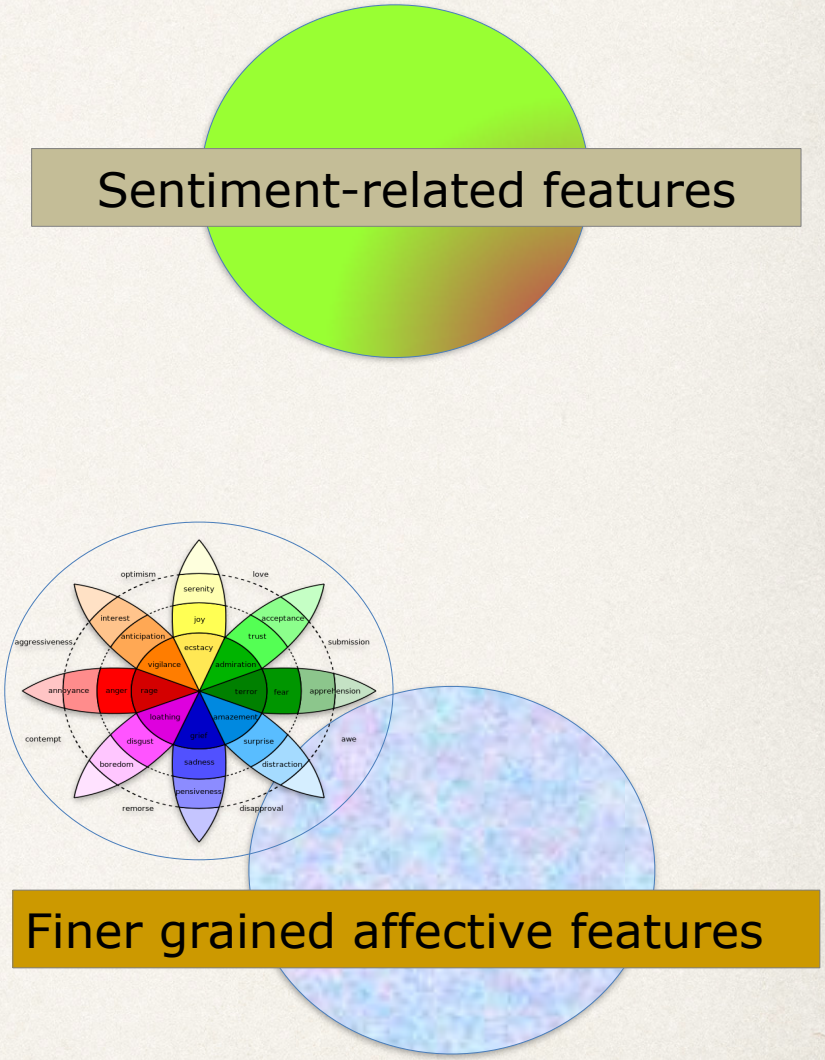
Lexicons for Sentiment, Affect, and Connotation

Viviana Patti

LR4NLP 2024



Connotation in the lexicon

- Words have connotation as well as sense
- Can we build lexical resources that represent these connotations?
- And use them in computational tasks on sentiment analysis et al.?
- Affective connotation

Lexicons for Sentiment, Affect, and Connotation Speech and Language Processing. Chapter 22 in Daniel Jurafsky & James H. Martin. Draft of August, 2024..

Scherer Typology of Affective States

Emotion: brief organically synchronized … evaluation of a major event

- *angry, sad, joyful, fearful, ashamed, proud, elated*

Mood: diffuse non-caused low-intensity long-duration change in subjective feeling

- *cheerful, gloomy, irritable, listless, depressed, buoyant*

Interpersonal stances: affective stance toward another person in a specific interaction

- *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*

Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons

- *liking, loving, hating, valuing, desiring*

Personality traits: stable personality dispositions and typical behavior tendencies

- *nervous, anxious, reckless, morose, hostile, jealous*

Morality values: …

NLP: Layers

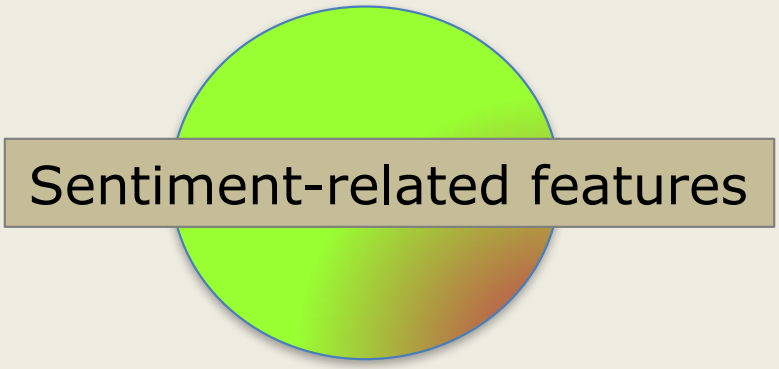
- Natural Language Processing (NLP) & Computational Linguistics (CL)
  - Morphology → how words are
  - Phonetics → how words sound
  - Syntax → how words relate
  - Semantics → what words means
  - Pragmatics → what is the intention of the speaker (context)

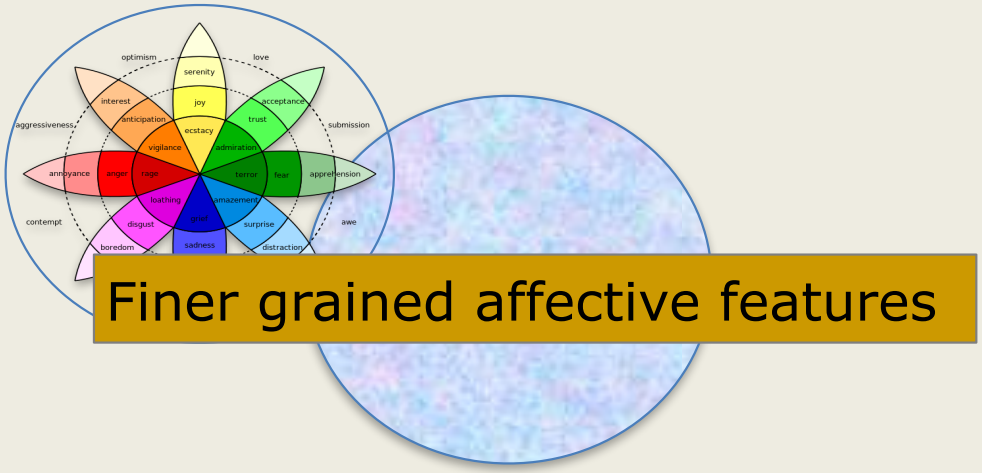
Suggested readings (overview)

- Lexicons for Sentiment, Affect, and Connotation
  Speech and Language Processing. Chapter 22 in Daniel Jurafsky & James H. Martin. Draft of February, 2024.
    - https://web.stanford.edu/~jurafsky/slp3/22.pdf
- M. Nissim, V. Patti. 2017. Semantic Aspects in Sentiment Analysis. Chapter 3 in Sentiment Analysis in Social Networks. 31-48.

Computational models and lexical resources for affect

- Different facets of the affective content
- Wide availability of lexical resources for English covering the various perspectives.
- Both sentiment and emotion lexicons, and psycholinguistic resources available for English, refer to various affective models and capture different facets of affect, including:
  - Sentiment polarity: aspects related to the polarity of words
  - Finer-grained aspects: which can be captured according to different categorical or dimensional models of emotions.





Semantic resources for SA

- Classical resources on sentiment polarity
- Beyond the polarity valence: emotion lexica, ontologies, and psycholinguistic resources
  - Emotion lexica:
    - Emotional categories: finer-grained affective lexica based on emotional categorie
      - e.g., Emolex (NRC Emotion lexicon)
    - Dimensional models: based on VAD (valence, arousal, dominance model).
      - e.g., (NRC VAD lexicon)
    - Intensity! Best-worst Scaling
  - Psycholinguistic resources and other accounts of affect:
    - e.g, LIWC, DAL
  - Concept-based resources and ontologies: toward a knowledge-based approach to sentiment analysis
    - EmoSenticNet: lexicon + ontologies
- Social media corpora annotated for sentiment and emotion categories
  - Semeval corpora

Affective lexical resources

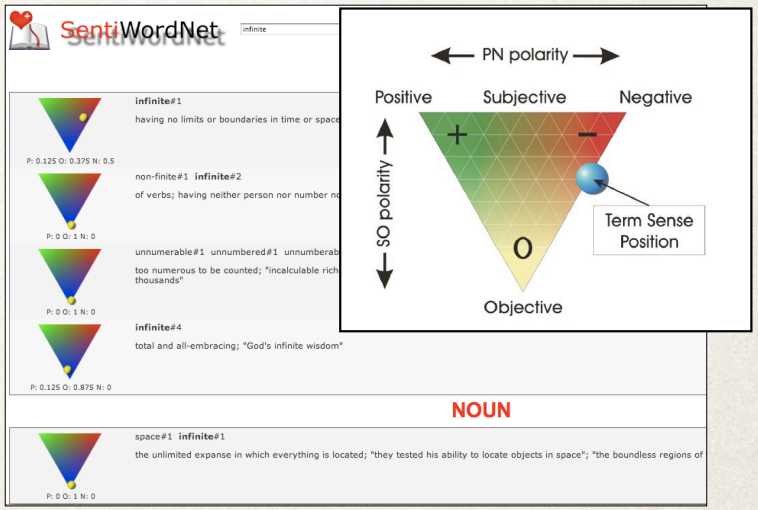
- Classic sentiment lexicons
  - SentiWordNet (Esuli and Sebastiani)
    - https://github.com/aesuli/SentiWordNet
  - WordNet Affect (Strapparava and Valitutti)
    - http://wndomains.fbk.eu/wnaffect.html
  - General Inquirer (Stone et al.)
    - https://inquirer.sites.fas.harvard.edu/
  - AFINN (Bradley and Lang)
    - https://github.com/abromberg/sentiment_analysis/tree/master/AFINN .
- Different aspects:
  - Providing a scale or an intensity of positivity/negativity;
  - Associate sentiment with words rather than with their senses
  - Some of them are based on WordNet:
    - sentiment annotation at the sense/synset level

WordNet

- https://wordnet.princeton.edu/
- WordNet is an on-line lexical reference system whose design is inspired by psycholinguistic theories of human lexical memory
- English nouns, verbs, adjectives and adverbs are organized into synsets:
  - synsets: synonym sets, each representing one underlying lexical concept

Sentiment lexicons: SentiWordNet

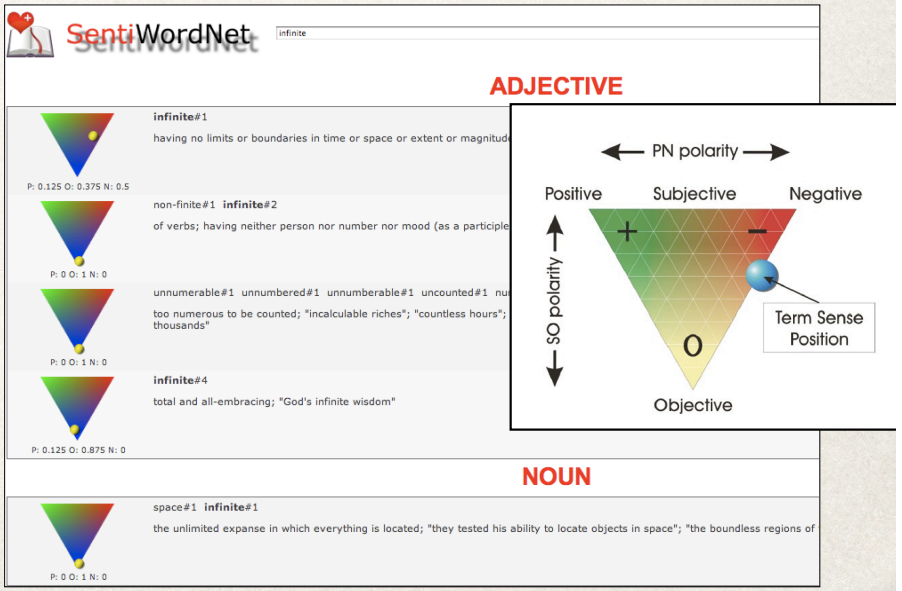
- SentiWordNet: synset of WordNet 3.0 are annotated according to their degree of neutrality, positiveness and negativity.
- https://github.com/aesuli/SentiWordNet
- Each synset ***s*** is associated the scores Pos(s), Neg(s) and Obj(s) indicating how neutral (Obj) or affective (Pos and Neg) the terms contained in the synset are.
- Each score ranges in [0.0,1.0] and and their sum is 1.0 for each synset.
- For a given sense, the values PosScore and NegScore are the positivity and negativity score assigned by SentiWordNet to the synset.



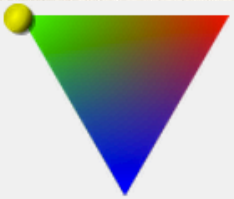
- The objectivity score can be calculated as: 1 - (Pos(s) + Neg(s)).

Sentiment lexicons: SentiWordNet

- SentiWordNet: synset of WordNet 3.0 are annotated according to their degree of neutrality, positiveness and negativity.
- https://github.com/aesuli/SentiWordNet
- Word senses belong to a syntactic category (Name, Verb, Adjective, and a word can be ambiguous w.r.t different categories.
- Different senses of the same term may have different opinion-related properties.
- Intuition: 1 -> the term (sense) is objective;
  0 -> the term conveys some strong sentiment (positive or negative) meaning



SentiWordNet: example

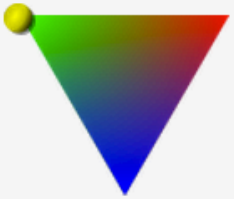


P: 1 O: 0 N: 0

estimable#1 00904163

deserving of respect or high regard

Feedback on SentiWordNet values: They are OK. Suggest your values.

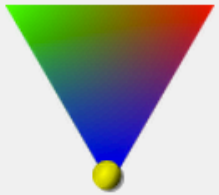


P: 1 O: 0 N: 0

respectable#2 honorable#4 good#4 **estimable**#2 01983162

deserving of esteem and respect; "all respectable companies give guarantees"; "ruined the family's good name"

Feedback on SentiWordNet values: They are OK. Suggest your values.



P: 0 O: 1 N: 0

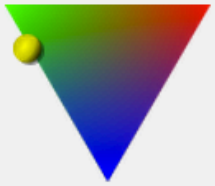
estimable#3 computable#1 00301432

may be computed or estimated; "a calculable risk"; "computable odds"; "estimable assets"

Feedback on SentiWordNet values: They are OK. Suggest your values.

SentiWordNet: example

ADJECTIVE

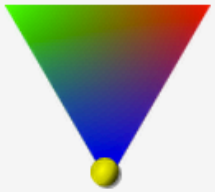


P: 0.75 O: 0.25 N: 0

good#1 01123148

having desirable or positive qualities especially those suitable for a thing specified; "good news from the hospital"; "a good report card"; "when she was good she was very very good"; "a good knife is one good for cutting"; "this stump will make a good picnic table"; "a good check"; "a good joke"; "a good exterior paint"; "a good secretary"; "a good dress for the office"

Feedback on SentiWordNet values: They are OK. Suggest your values.

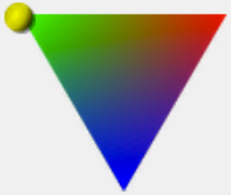


P: 0 O: 1 N: 0

good#2 full#6 00106020

having the normally expected amount; "gives full measure"; "gives good measure"; "a good mile from here"

Feedback on SentiWordNet values: They are OK. Suggest your values.



P: 1 O: 0 N: 0

good#3 01129977

morally admirable

Feedback on SentiWordNet values: They are OK. Suggest your values.

SentiWordNet_3.0.0_20100908.txt

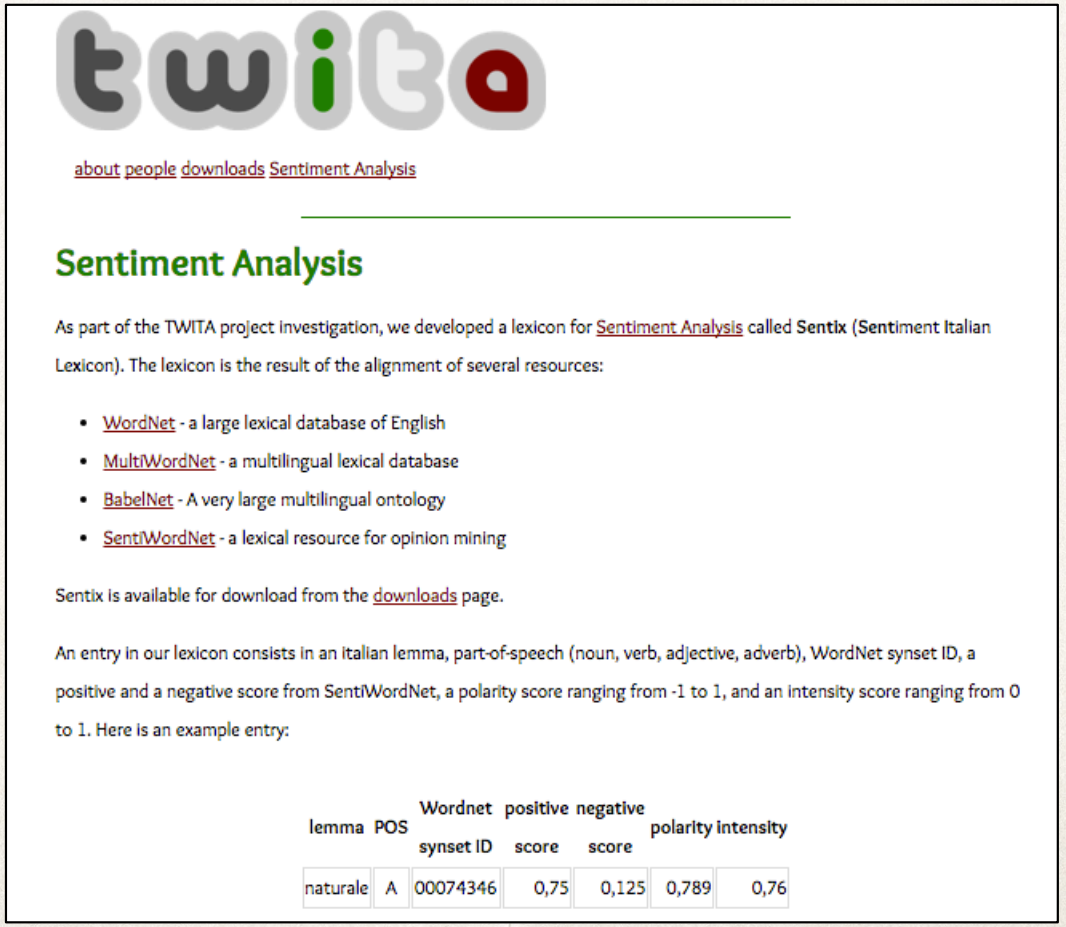
SentiWordNet_3.0.0_20100908.txt

```
# SentiWordNet v3.0.0 (1 June 2010)
# Copyright 2010 ISTI-CNR.
# All right reserved.
#
# For any information about SentiWordNet:
# Web: http://sentiwordnet.isti.cnr.it
# -------
#
# Data format.
#
# SentiWordNet v3.0 is based on WordNet version 3.0.
# WordNet website: http://wordnet.princeton.edu/
#
# The pair (POS,ID) uniquely identifies a WordNet (3.0) synset.
# The values PosScore and NegScore are the positivity and negativity
# score assigned by SentiWordNet to the synset.
# The objectivity score can be calculated as:
# ObjScore = 1 - (PosScore + NegScore)
# SynsetTerms column reports the terms, with sense number, belonging
# to the synset (separated by spaces).
#
# -------
#
# POS   ID  PosScore    NegScore    SynsetTerms Gloss
a   00001740    0.125   0   able#1  (usually followed by `to') having the necessary means or skill or know-how or authority to do something;
"able to swim"; "she was able to program her computer"; "we were at last able to buy a car"; "able to get a grant for the project"
```

```
retina
a   00105746    0   0   ample#1 more than enough in size or scope or capacity; "had ample food for the party"; "an ample supply"
a   00106020    0   0   good#2 full#6   having the normally expected amount; "gives full measure"; "gives good measure"; "a good mile from
here"
a   00106182    0.25    0   generous#3  more than adequate; "a generous portion"
a   00106277    0   0   wide-cut#1 wide#6 full#8    having ample fabric; "the current taste for wide trousers"; "a full skirt"
a   00106456    0   0.625   stingy#2 scrimpy#1 meagre#1 meagerly#1 meager#1 deficient in amount or quality or extent; "meager resources";
"meager fare"
a   00106821    0   0.5 spare#5 scanty#1 bare#2 lacking in amplitude or quantity; "a bare livelihood"; "a scanty harvest"; "a spare diet"
a   00107017    0.125   0.625   exiguous#1  extremely scanty; "an exiguous budget"
a   00107128    0   0.5 hand-to-mouth#1 providing only bare essentials; "a hand-to-mouth existence"
a   00107250    0   0.375   hardscrabble#1  barely satisfying a lower standard; "the sharecropper's hardscrabble life"
a   00107384    0   0   paltry#2 miserable#6 measly#1   contemptibly small in amount; "a measly tip"; "the company donated a miserable $100
```

Sentix/TWITA

- http://valeriobasile.github.io/twita/sentix.html



twita

about people downloads Sentiment Analysis

Sentiment Analysis

As part of the TWITA project investigation, we developed a lexicon for Sentiment Analysis called **Sentix** (**Sent**iment Italian Lexicon). The lexicon is the result of the alignment of several resources:

- WordNet - a large lexical database of English
- MultiWordNet - a multilingual lexical database
- BabelNet - A very large multilingual ontology
- SentiWordNet - a lexical resource for opinion mining

Sentix is available for download from the downloads page.

An entry in our lexicon consists in an italian lemma, part-of-speech (noun, verb, adjective, adverb), WordNet synset ID, a positive and a negative score from SentiWordNet, a polarity score ranging from -1 to 1, and an intensity score ranging from 0 to 1. Here is an example entry:

lemma	POS	Wordnet synset ID	positive score	negative score	polarity	intensity
naturale	A	00074346	0,75	0,125	0,789	0,76

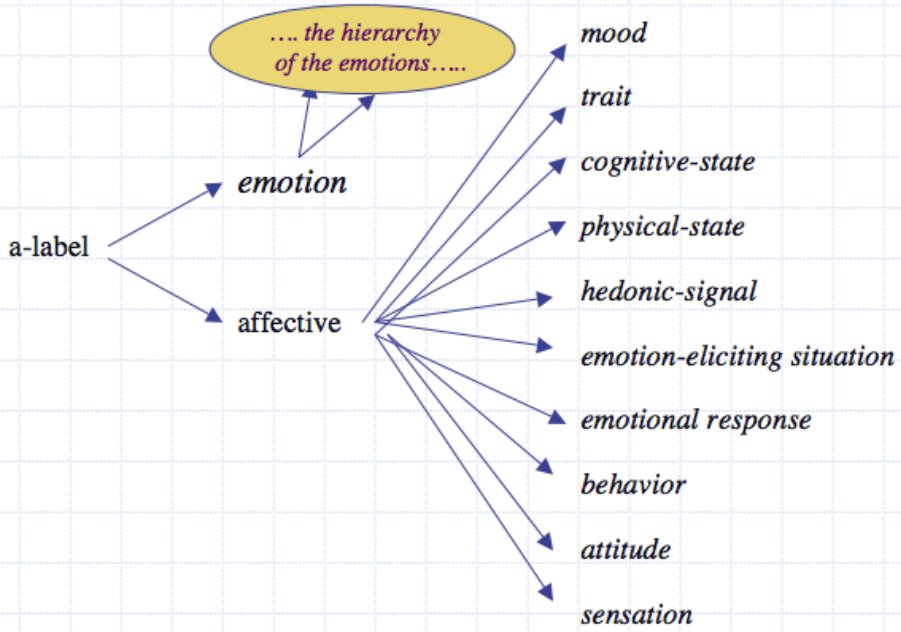
WN-Affect

- http://wndomains.fbk.eu/wnaffect.html
- Affective lexical resource, built for affective computing, computational humor, text analysis, etc.
- It is a lexical repository of the direct affective words: different perspective!
- What about words conveying an indirect affective meaning:
  - All words can potentially convey affective meaning
  - Even those not directly related to emotions can evoke pleasant or painful experiences
  - Some of them are related to the individual story, but for many others the affective power is part of the collective imagination (e.g. mum, ghost, war, …)
- The resource, named WordNet-Affect, started from WordNet, through selection and labeling of synsets representing affective concepts.

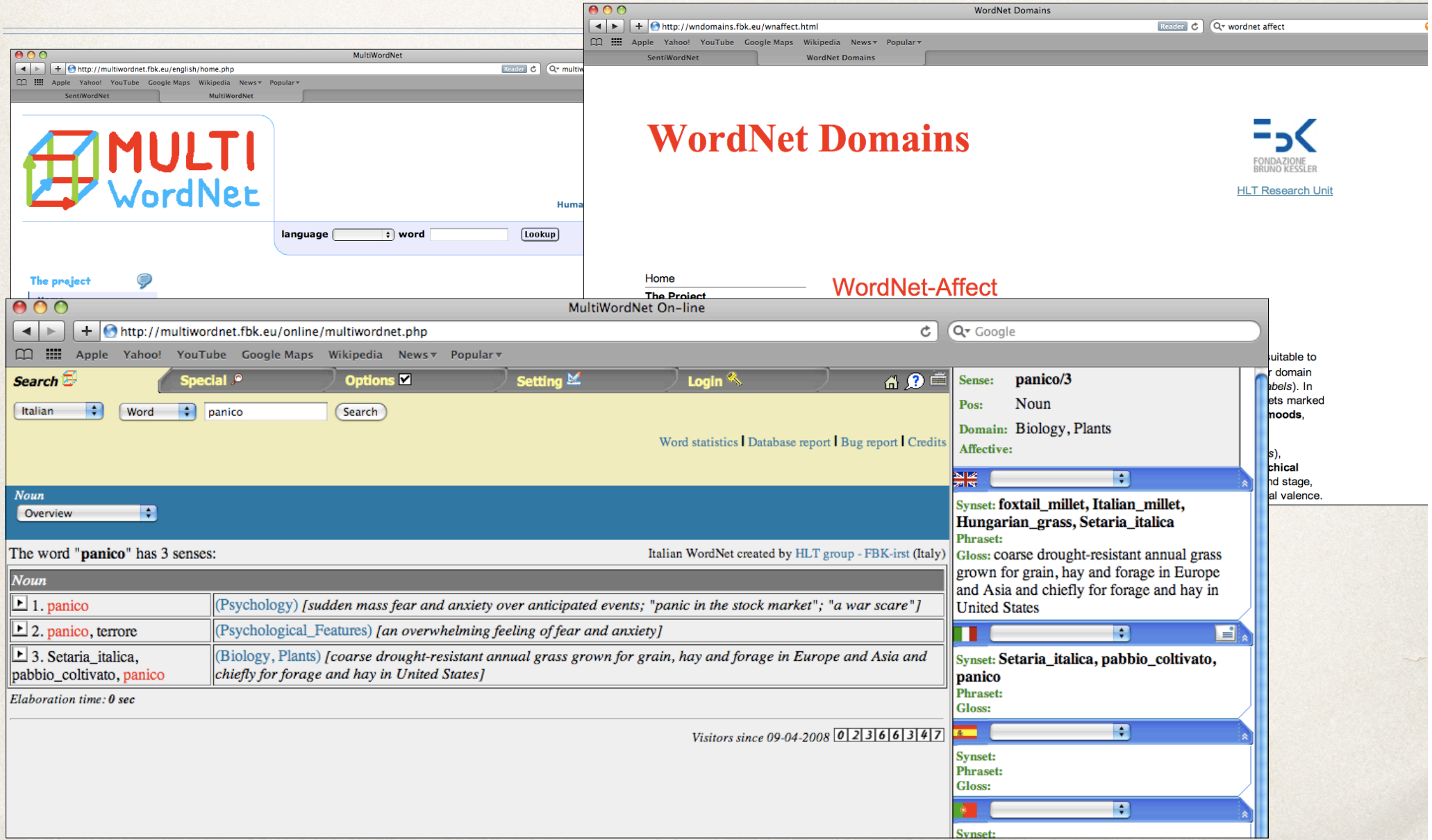
WN-Affect

Affective domain labels

- Ortony distinguishes between *emotional* and *affective-not-emotional* words
- Among affective-not-emotional words many categories from the literature (e.g. Johnson-Laird and Oatley, 1988 Ortony et al., 1988)



MultiWN + WN-Affect Domain



MultiWN + WN-Affect Domain

MultiWordNet On-line

http://multiwordnet.fbk.eu/online/multiwordnet.php

Apple Yahoo! YouTube Google Maps Wikipedia News Popular

Search Special Options Setting Login

Italian Word panico Search

Word statistics | Database report | Bug report | Credits

Noun

Overview

The word "**panico**" has 3 senses: Italian WordNet created by HLT group - FBK-irst (Italy)

Noun	
1. panico	(Psychology) *[sudden mass fear and anxiety over anticipated events; "panic in the stock market"; "a war scare"]*
2. panico, terrore	(Psychological_Features) *[an overwhelming feeling of fear and anxiety]*
3. Setaria_italica, pabbio_coltivato, panico	(Biology, Plants) *[coarse drought-resistant annual grass grown for grain, hay and forage in Europe and Asia and chiefly for forage and hay in United States]*

Elaboration time: 0 sec

Visitors since 09-04-2008 0 2 3 6 6 3 4 7

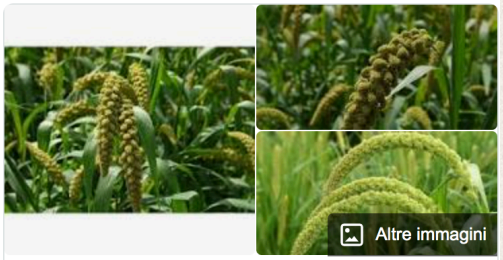
Sense: **panico/3**
Pos: Noun
Domain: Biology, Plants
Affective:

Synset: **foxtail_millet, Italian_millet, Hungarian_grass, Setaria_italica**
Phraset:
Gloss: coarse drought-resistant annual grass grown for grain, hay and forage in Europe and Asia and chiefly for forage and hay in United States

Synset: **Setaria_i**
panico
Phraset:
Gloss:

Synset:
Phraset:
Gloss:

Synset:



Panico

Cereali

La Setaria italica è un cereale appartenente alla famiglia delle Poaceae, a grano nudo, comunemente chiamato panìco per le sue pannocchie, molto simile al miglio. Wikipedia

Nome scientifico: Setaria italica

Classificazione superiore: Setaria

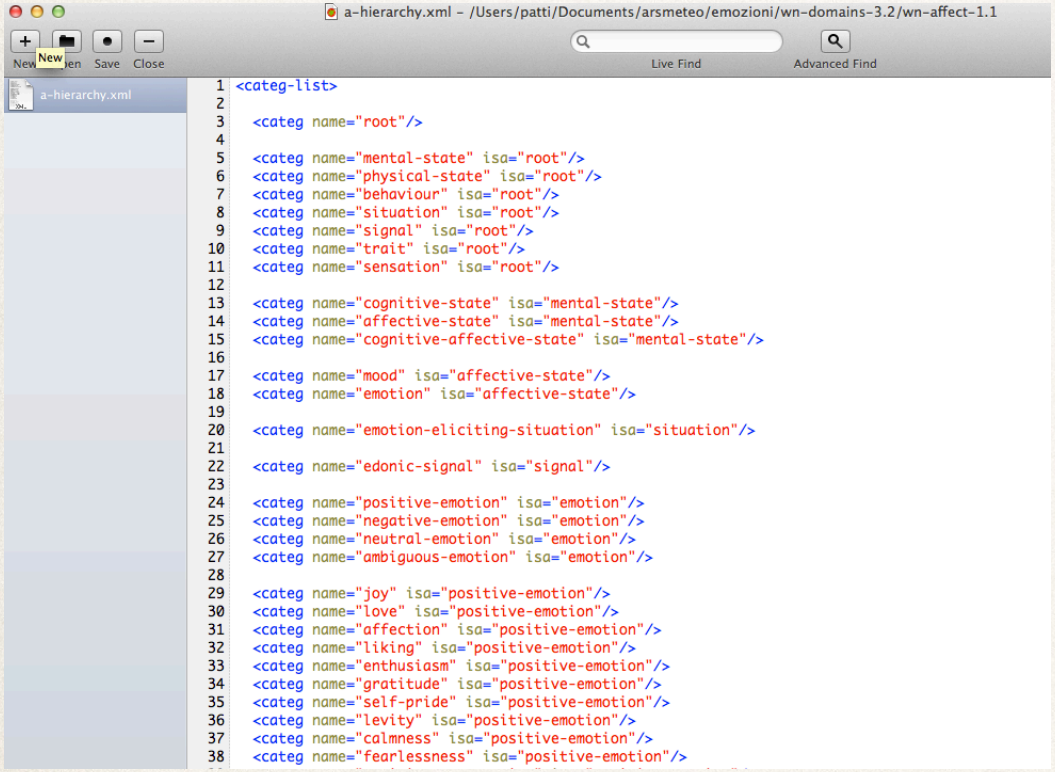
Categoria tassonomica: Specie

Famiglia: Poaceae

Valence tagging in WN-Affect

- Distinguishing synsets according to emotional valence
  - Positive emotions (e.g. joy#1, enthusiasm#1),
  - Negative emotions (e.g. fear#1, horror#1),
  - Ambiguous, when the valence depends on the context (e.g. surprise#1)
  - Neutral, when the synset is considered affective but not characterized by valence (e.g. indifference#1)

WN-Affect Hierarchy

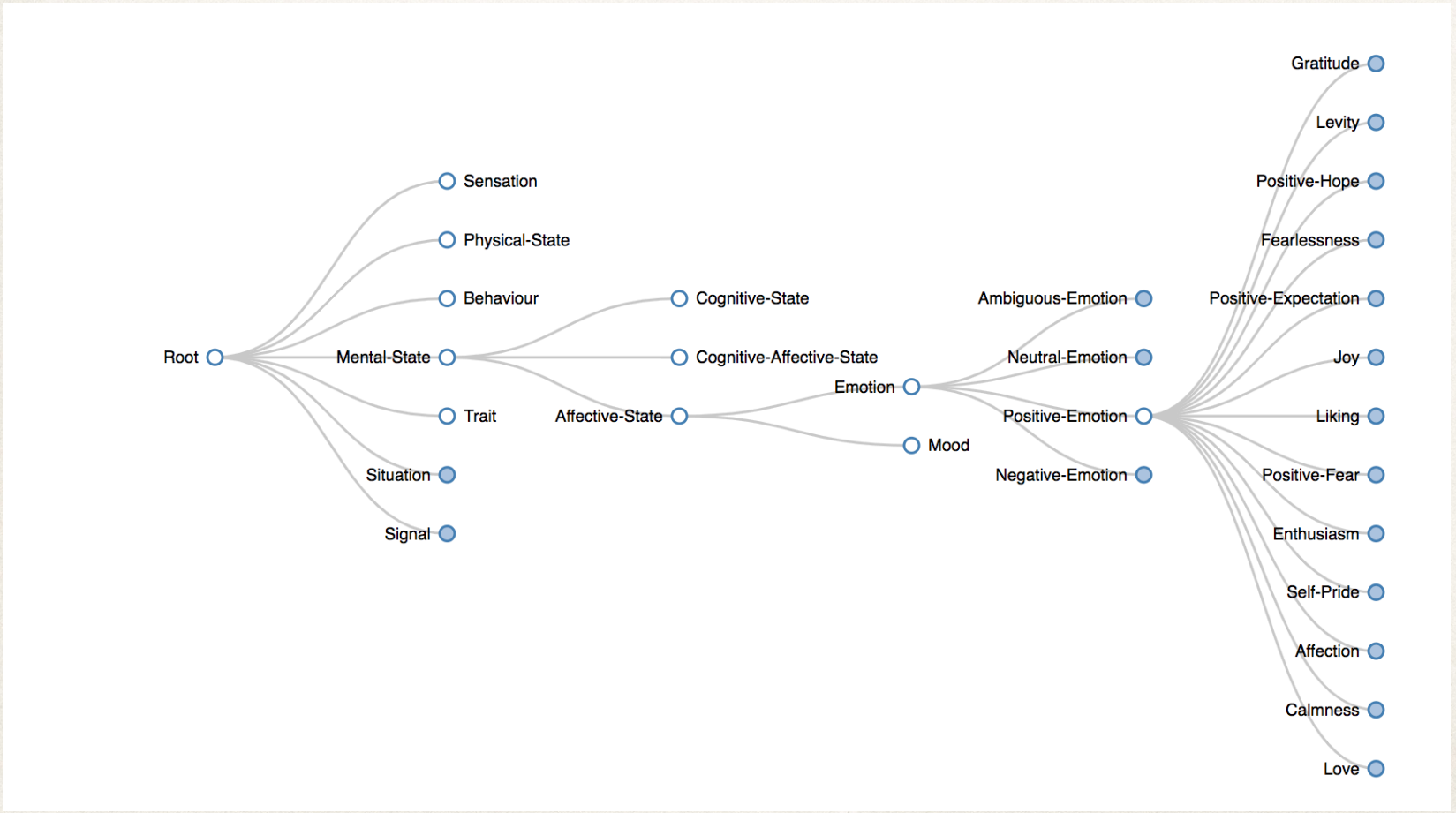



```
<categ-list>

  <categ name="root"/>

  <categ name="mental-state" isa="root"/>
  <categ name="physical-state" isa="root"/>
  <categ name="behaviour" isa="root"/>
  <categ name="situation" isa="root"/>
  <categ name="signal" isa="root"/>
  <categ name="trait" isa="root"/>
  <categ name="sensation" isa="root"/>

  <categ name="cognitive-state" isa="mental-state"/>
  <categ name="affective-state" isa="mental-state"/>
  <categ name="cognitive-affective-state" isa="mental-state"/>

  <categ name="mood" isa="affective-state"/>
  <categ name="emotion" isa="affective-state"/>

  <categ name="emotion-eliciting-situation" isa="situation"/>

  <categ name="edonic-signal" isa="signal"/>

  <categ name="positive-emotion" isa="emotion"/>
  <categ name="negative-emotion" isa="emotion"/>
  <categ name="neutral-emotion" isa="emotion"/>
  <categ name="ambiguous-emotion" isa="emotion"/>

  <categ name="joy" isa="positive-emotion"/>
  <categ name="love" isa="positive-emotion"/>
  <categ name="affection" isa="positive-emotion"/>
  <categ name="liking" isa="positive-emotion"/>
  <categ name="enthusiasm" isa="positive-emotion"/>
  <categ name="gratitude" isa="positive-emotion"/>
  <categ name="self-pride" isa="positive-emotion"/>
  <categ name="levity" isa="positive-emotion"/>
  <categ name="calmness" isa="positive-emotion"/>
  <categ name="fearlessness" isa="positive-emotion"/>
```

http://www.gsi.upm.es:9080/ontologies/wnaffect/

WN-Affect Hierarchy

http://www.gsi.upm.es:9080/ontologies/wnaffect/fulltree.html
(EUROSENTIMENT EU project)



WN-Affect Synset list: nouns, adjectives, verbs, adv

a-synsets.xml – /Users/patti/Documents/arsmeteo/emozioni/wn-domains-3.2/wn-a

Live Find Advanced Find

```
92  <noun-syn id="n#05610210" categ="misogamy"/>
93  <noun-syn id="n#05610544" categ="misocainea"/>
94  <noun-syn id="n#05601051" categ="misery"/>
95  <noun-syn id="n#05610135" categ="misanthropy"/>
96  <noun-syn id="n#05615937" categ="mercifulness"/>
97  <noun-syn id="n#05600345" categ="melancholy"/>
98  <noun-syn id="n#05582161" categ="meekness"/>
99  <noun-syn id="n#05613388" categ="malice"/>
100 <noun-syn id="n#05613127" categ="malevolence"/>
101 <noun-syn id="n#05613291" categ="maleficence"/>
102 <noun-syn id="n#05609709" categ="loyalty"/>
103 <noun-syn id="n#05609517" categ="lovingness"/>
104 <noun-syn id="n#05607724" categ="love"/>
105 <noun-syn id="n#05588618" categ="lividity"/>
106 <noun-syn id="n#05573285" categ="liking"/>
107 <noun-syn id="n#05584043" categ="levity"/>
108 <noun-syn id="n#05563906" categ="neutral-languor"/>
109 <noun-syn id="n#05587782" categ="positive-languor"/>
110 <noun-syn id="n#05615380" categ="kindheartedness"/>
111 <noun-syn id="n#05603196" categ="joylessness"/>
112 <noun-syn id="n#05596218" categ="joy"/>
113 <noun-syn id="n#05614225" categ="jollity"/>
114 <noun-syn id="n#05597701" categ="jocundity"/>
115 <noun-syn id="n#05594897" categ="jitteriness"/>
116 <noun-syn id="n#05612865" categ="jealousy"/>
117 <noun-syn id="n#05577212" categ="isolation"/>
118 <noun-syn id="n#05614716" categ="irascibility"/>
119 <noun-syn id="n#05593389" categ="fear-intimidation"/>
120 <noun-syn id="n#05607353" categ="despair-intimidation"/>
121 <noun-syn id="n#05594325" categ="insecurity"/>
122 <noun-syn id="n#05578483" categ="ingratitude"/>
123 <noun-syn id="n#05588725" categ="infuriation"/>
124 <noun-syn id="n#05588960" categ="indignation"/>
125 <noun-syn id="n#05579322" categ="indifference"/>
126 <noun-syn id="n#05586117" categ="impatience"/>
```

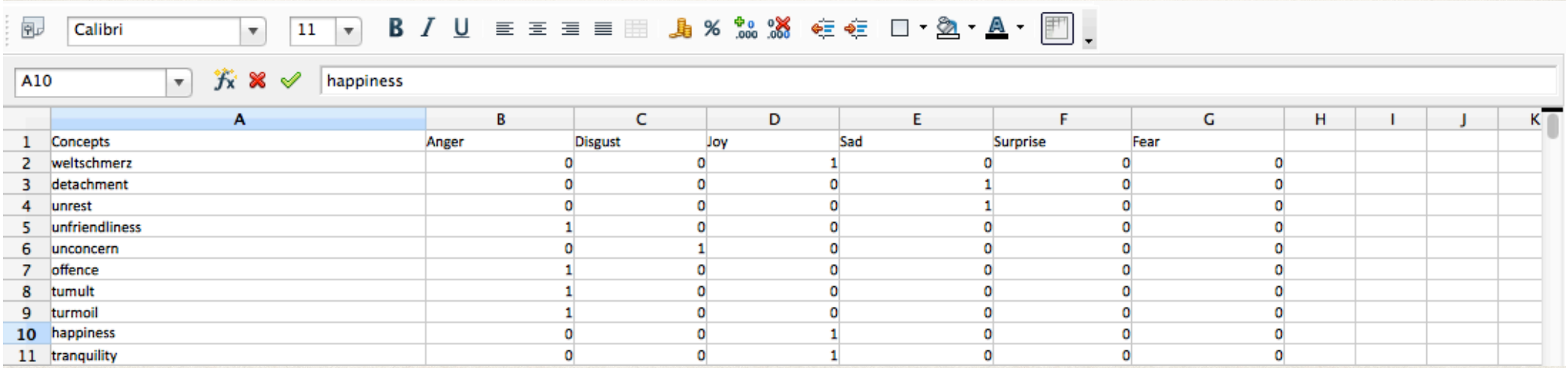
WN-Affect – Ekman Big 6

- SemEval-2007 Task 14: Affective Text
- https://www.aclweb.org/anthology/S07-1013/
- Starting from WordNet-Affect, for Task 14 at Semeval 2007, a new version of the resource has been provided.
- This resource only includes a portion of WordNet-Affect, as it was re-annotated at a finer-grained level using the six emotional category labels from Ekman:

joy, fear, anger, sadness, disgust, surprise

EmoSenticNet

- SenticNet + WN-Affect (Ekman Big 6)
- http://www.gelbukh.com/emosenticnet/

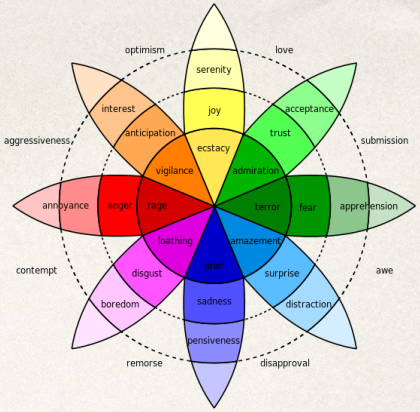


Calibri 11

A10 happiness

	A	B	C	D	E	F	G	H	I	J	K
1	Concepts	Anger	Disgust	Joy	Sad	Surprise	Fear				
2	weltschmerz	0	0	1	0	0	0				
3	detachment	0	0	0	1	0	0				
4	unrest	0	0	0	1	0	0				
5	unfriendliness	1	0	0	0	0	0				
6	unconcern	0	1	0	0	0	0				
7	offence	1	0	0	0	0	0				
8	tumult	1	0	0	0	0	0				
9	turmoil	1	0	0	0	0	0				
10	happiness	0	0	1	0	0	0				
11	tranquility	0	0	1	0	0	0				

Affective lexical resources



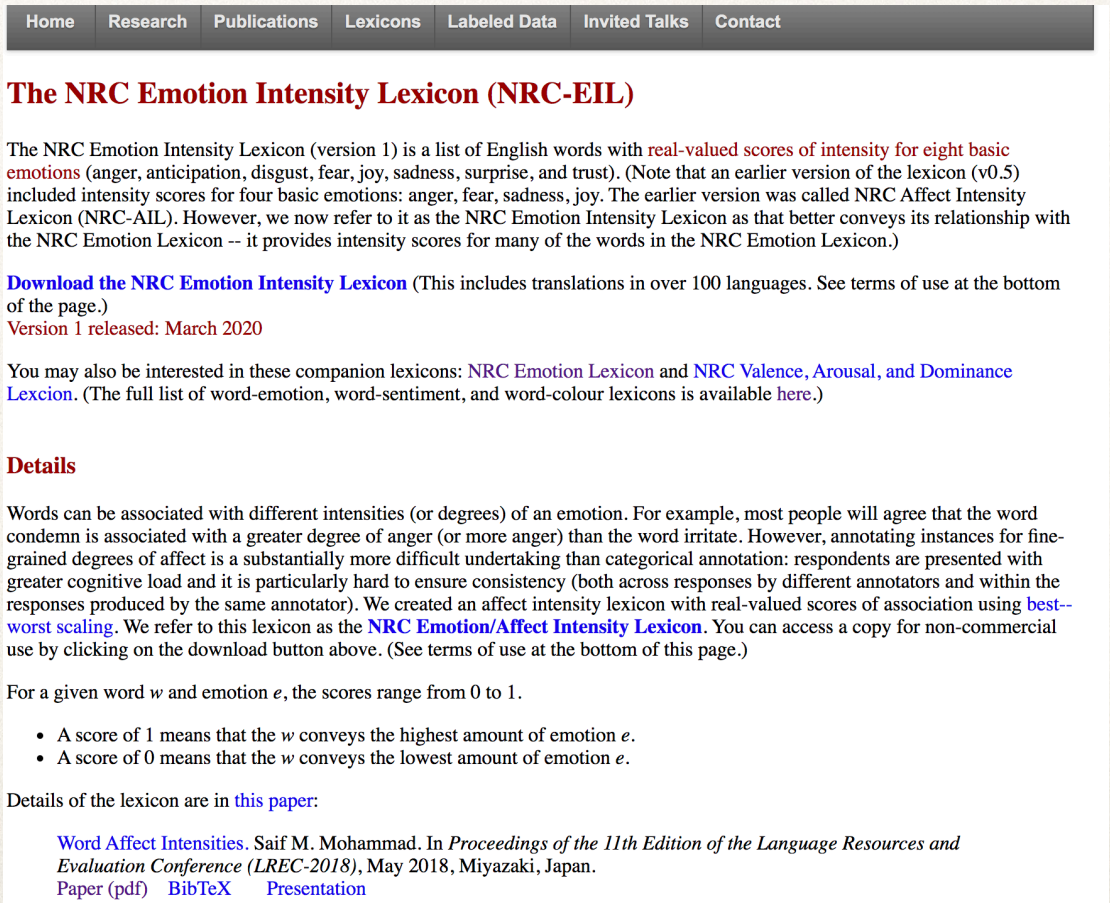
- Emo-lex: NRC Canada
  - http://saifmohammad.com/WebPages/lexicons.html
  - NRC EMOTION LEXICON
  - The NRC emotion lexicon is a list of words and their associations with eight emotions from the Plutchik's model (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).
  - The annotations were manually done through Amazon's Mechanical Turk. Refer to publications listed in the README for more details.

Semantic resources for SA

- Emotion lexica:
  - Emotional categories: finer-grained affective lexica based on emotional categorie
    - e.g., Emolex (NRC Emotion lexicon)
  - Dimensional models: based on VAD (valence, arousal, dominance Russell's model).
    - *Valence* o Valenza edonica (da dispiacere a piacere), *Arousal* o Grado di attivazione (da disattivo ad attivo), *dominance* (da *submisiveness* a *dominance*)
    - e.g., (NRC VAD lexicon)
  - Intensity!
  - Rating vs ranking
    - G. N. Yannakakis, R. Cowie and C. Busso, "The Ordinal Nature of Emotions: An Emerging Approach," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2018.2879512
      https://ieeexplore.ieee.org/document/8521685
  - Best-worst scaling annotations

NRC Emotion Intensity Lexicon

- http://saifmohammad.com/WebPages/AffectIntensity.htm



Home | Research | Publications | Lexicons | Labeled Data | Invited Talks | Contact

The NRC Emotion Intensity Lexicon (NRC-EIL)

The NRC Emotion Intensity Lexicon (version 1) is a list of English words with real-valued scores of intensity for eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). (Note that an earlier version of the lexicon (v0.5) included intensity scores for four basic emotions: anger, fear, sadness, joy. The earlier version was called NRC Affect Intensity Lexicon (NRC-AIL). However, we now refer to it as the NRC Emotion Intensity Lexicon as that better conveys its relationship with the NRC Emotion Lexicon -- it provides intensity scores for many of the words in the NRC Emotion Lexicon.)

Download the NRC Emotion Intensity Lexicon (This includes translations in over 100 languages. See terms of use at the bottom of the page.)
Version 1 released: March 2020

You may also be interested in these companion lexicons: NRC Emotion Lexicon and NRC Valence, Arousal, and Dominance Lexcion. (The full list of word-emotion, word-sentiment, and word-colour lexicons is available here.)

Details

Words can be associated with different intensities (or degrees) of an emotion. For example, most people will agree that the word condemn is associated with a greater degree of anger (or more anger) than the word irritate. However, annotating instances for fine-grained degrees of affect is a substantially more difficult undertaking than categorical annotation: respondents are presented with greater cognitive load and it is particularly hard to ensure consistency (both across responses by different annotators and within the responses produced by the same annotator). We created an affect intensity lexicon with real-valued scores of association using best--worst scaling. We refer to this lexicon as the **NRC Emotion/Affect Intensity Lexicon**. You can access a copy for non-commercial use by clicking on the download button above. (See terms of use at the bottom of this page.)

For a given word *w* and emotion *e*, the scores range from 0 to 1.

- A score of 1 means that the *w* conveys the highest amount of emotion *e*.
- A score of 0 means that the *w* conveys the lowest amount of emotion *e*.

Details of the lexicon are in this paper:

> Word Affect Intensities. Saif M. Mohammad. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, May 2018, Miyazaki, Japan.
> Paper (pdf) BibTeX Presentation

Terms of use

Terms of use:

1. **Research Use:** The lexicon mentioned in this page can be used freely for non-commercial research and educational purposes.

2. **Citation:** Cite the papers associated with the lexicon in your research papers and articles that make use of them.

3. **Media Mentions:** In news articles and online posts on work using the lexicon, cite the lexicon. For example: "We make use of the <resource name>, created by <author(s)> at the National Research Council Canada." We would appreciate a hyperlink to the lexicon home page and an email to the contact author (saif.mohammad@nrc-cnrc.gc.ca). (Authors and homepage information provided at the top of the README.)

4. **Credit:** If you use the lexicon in a product or application, then acknowledge this in the 'About' page and other relevant documentation of the application by stating the name of the resource, the authors, and NRC. For example: "This application/product/tool makes use of the <resource name>, created by <author(s)> at the National Research Council Canada." We would appreciate a hyperlink to the lexicon home page and an email to the contact author (saif.mohammad@nrc-cnrc.gc.ca).

5. **No Redistribution:** Do not redistribute the data. Direct interested parties to the lexicon home page. You may not rent or license the use of the lexicon nor otherwise permit third parties to use it.

6. **Proprietary Notice:** You will ensure that any copyright notices, trademarks or other proprietary right notices placed by NRC on the lexicon remains in evidence.

7. **Title:** All intellectual property rights in and to the lexicon shall remain the property of NRC. All proprietary interests, rights, unencumbered titles, copyrights, or other Intellectual Property Rights in the lexicon and all copies thereof remain at all times with NRC.

8. **Commercial License:** If interested in commercial use of the lexicon, contact the author: saif.mohammad@nrc-cnrc.gc.ca

9. **Disclaimer:** National Research Council Canada (NRC) disclaims any responsibility for the use of the lexicon and does not provide technical support. NRC makes no representation and gives no warranty of any kind with respect to the accuracy, usefulness, novelty, validity, scope, or completeness of the lexicon and expressly disclaims any implied warranty of merchantability or fitness for a particular purpose of the lexicon. That said, the contact listed above welcomes queries and clarifications.

10. **Limitation of Liability:** You will not make claims of any kind whatsoever upon or against NRC or the creators of the lexicon, either on your own account or on behalf of any third party, arising directly or indirectly out of your use of the lexicon. In no event will NRC or the creators be liable on any theory of liability, whether in an action of contract or strict liability (including negligence or otherwise), for any losses or damages incurred by you, whether direct, indirect, incidental, special, exemplary or
consequential, including lost or anticipated profits, savings, interruption to business, loss of business opportunities, loss of business information, the cost of recovering such lost information, the cost of substitute intellectual property or any other pecuniary loss arising from the use of, or the inability to use, the lexicon regardless of whether you have advised NRC or NRC has advised you of the possibility of such damages.

NRC lexicon in other languages

Home | About Saif | Research | Publications | Sentiment and Emotion Lexicons | Sentiment and Emotion Labeled Data | Invited Talks | Contact

NRC Word-Emotion Association Lexicon (aka EmoLex)
The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing.

Email: saif.mohammad@nrc-cnrc.gc.ca
Follow @Saifm

Association Lexicon	Version	# of Terms	Categories	Association Scores	Method of Creation	Papers
Word-Emotion and Word-Sentiment Association Lexicon						
NRC Word-Emotion Association Lexicon (also called EmoLex) README	0.92 (2010)	14,182 unigrams (words) ~25,000 senses*	**sentiments:** negative, positive **emotions:** anger, anticipation, disgust, fear, joy, sadness, surprise, trust	0 (not associated) or 1 (associated) not associated, weakly, moderately, or strongly associated	**Manual:** By crowdsourcing on Mechanical Turk. **Domain:** General	**Crowdsourcing a Word-Emotion Association Lexicon**, Saif Mohammad and Peter Turney, *Computational Intelligence*, 29 (3), 436-465, 2013. Paper (pdf) BibTeX **Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon**, Saif Mohammad and Peter Turney, In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, June 2010, LA, California. Abstract Paper (pdf) Presentation

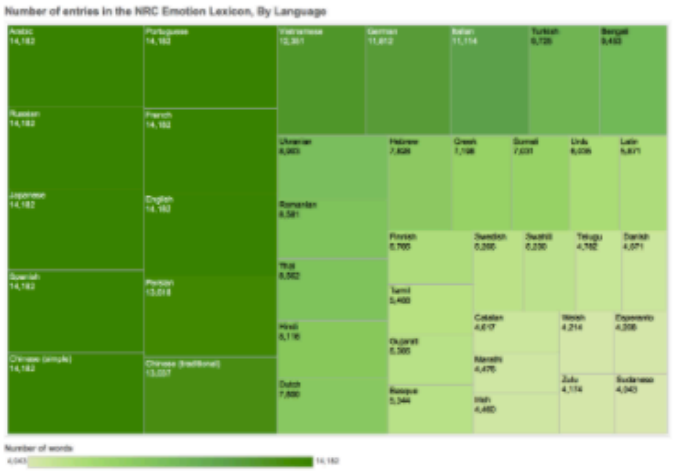
* The sense-level annotations provided by individual annotators for the eight emotions can be downloaded by clicking here.
Access various other word-emotion, word-sentiment, and word-colour lexicons here.)

NRC Emotion Lexicon in Other Languages

The NRC Emotion Lexicon has affect annotations for **English** words. Despite some cultural differences, it has been shown that a majority of affective norms are stable across languages. Thus we provide versions of the lexicon in over twenty languages by translating the English terms using Google Translate (July 2015). Go here to obtain it. We currently have translations into these languages: Arabic, Basque, Bengali, Catalan, Chinese, Chinese, Danish, Dutch, Esperanto, Finnish, French, German, Greek, Gujarati, Hebrew, Hindi, Italian, Irish, Japanese, Latin, Marathi, Persian, Portuguese, Romanian, Russian, Somali, Spanish, Sudanese, Swahili, Swedish, Tamil, Telugu, Thai, Turkish, Ukranian, Urdu, Vietnamese, Welsh, and Zulu. Email me if interested in a language not listed here.

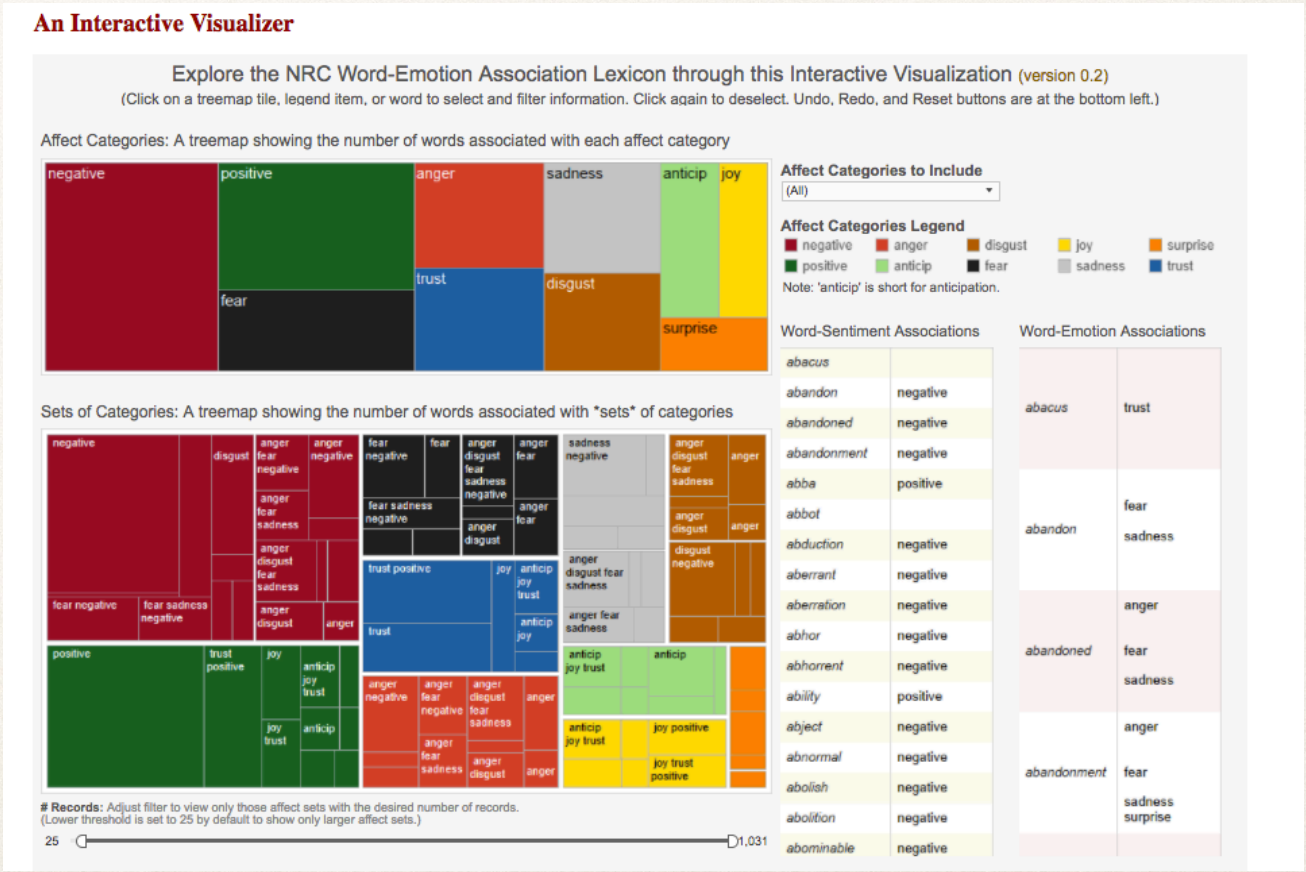
See README of the NRC Emotion Lexicon for more details about the lexicon and the terms of use.

Note that some translations by Google Translate may be incorrect or they may simply be transliterations of the original English terms. If you have a list of corrections for any language, we will be happy to hear from you.



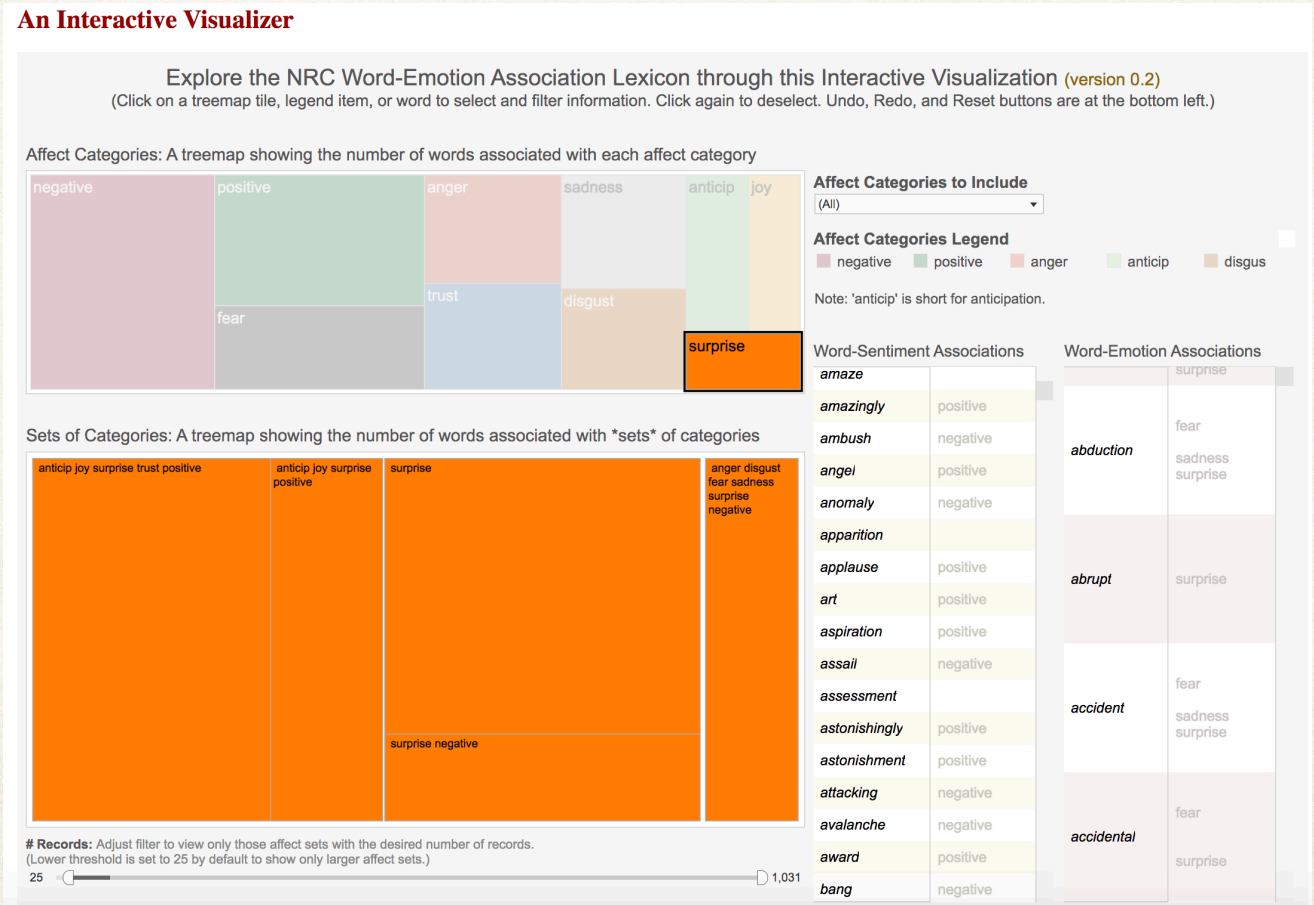
Interactive visualizer

https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm



Interactive visualizer

https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm



Affective lexical resources

- AFINN: This dictionary is a list of 2,477 English manually labeled words collected by (Nielsen, 2011). Polarity varies from – 5 up to +5.
  - Intensity of the polarity
- LIWC: Linguistic Inquiry and Word Counts Dictionary: The 2007 version of LIWC comprises almost 4,500 words and word stems distributed in categories for analyzing psycho-linguistic features in texts.
- Categories for positive and negative emotions:
  - http://www.liwc.net

  (Pennebaker et al., 2001).

Affective lexical resources

- DAL: Dictionary of Affective Language developed by Whissell (Whissell, 2009)
  - includes 8,742 English words rated in a three-point scale along three dimensions:
  - Activation (degree of response that humans have under an emotional state);
  - Imagery (how difficult is to form a mental picture of a given word);
  - Pleasantness (degree of pleasure produced by words).

NRC Valence, Arousal, Dominance (VAD) lexicon

https://saifmohammad.com/WebPages/nrc-vad.html

20,000 words, 3 emotional dimensions:

- **valence** (the pleasantness of the stimulus)
- **arousal** (the intensity of emotion provoked by the stimulus)
- **dominance** (the degree of control exerted by the stimulus)

Best-worst scaling: valence

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?

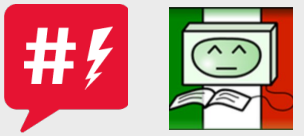
vacation, consolation, whistle, torture

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?

Best-worst scaling: tools

- https://saifmohammad.com/WebPages/BestWorst.html
- Litescale: A Lightweight Tool for Best-worst Scaling Annotation by Valerio Basile and Christian Cagnazzo
  - https://github.com/valeriobasile/litescale
  - http://gingerbeard.alwaysdata.net/litescale-web/home

Hate Words Lexicons



Hate Words Lexicon: 17 HurtLex Categories

1. Ethnic Slurs
2. Location and Demonyms
3. Profession and Occupation
4. Physical Disabilities and Diversity
5. **Cognitive Disabilities and Diversity** ✓
6. Moral Behavior and Defect
7. Words Related to Social and Economic Disadvantage
8. Words Related to Plants
9. Words Related to Animals
10. **Words Related to Male Genitalia** ✓
11. Words Related to Female Genitalia
12. **Words Related Prostitution** ✓
13. Words Related Homosexuality
14. Descriptive Words with Potential Negative Connotations
15. Derogatory Words
16. Felonies and Words Related to Crime and Immoral Behavior
17. Words Related to the Seven Deadly Sins of the Christian Tradition

Lexical Clues

- Presence of Swear Words:
  → www.noswearing.com
- Presence of Sexist Slurs,
- Presence of "Woman" Words,
  → wife, girl, her, mom, mother...
- Presence of Words from HurtLex
  → Tullio De Mauro's "Parole per Ferire"

#makemeasandwich
#YesAllMen

Structural Clues

- Presence of Hashtags,
- Presence of Emojis,
- Presence of links.

- De Mauro e il lavoro della Commissione Jo Cox: http://www.camera.it/leg17/1264
- Hurtlex: https://github.com/valeriobasile/hurtlex
- Parole per ferire: https://www.internazionale.it/opinione/tullio-de-mauro/2016/09/27/razzismo-parole-ferire

Elisa Bassignana, Valerio Basile, Viviana Patti: Hurtlex: A Multilingual Lexicon of Words to Hurt. Proc. of CLiC-it 2018.

Lexicons for detecting document affect: Simplest unsupervised method

- Lexicons can be used in a rule-based algorithm for classification.
- The simplest version is just to use the ratio of positive to negative words:

Sentiment:

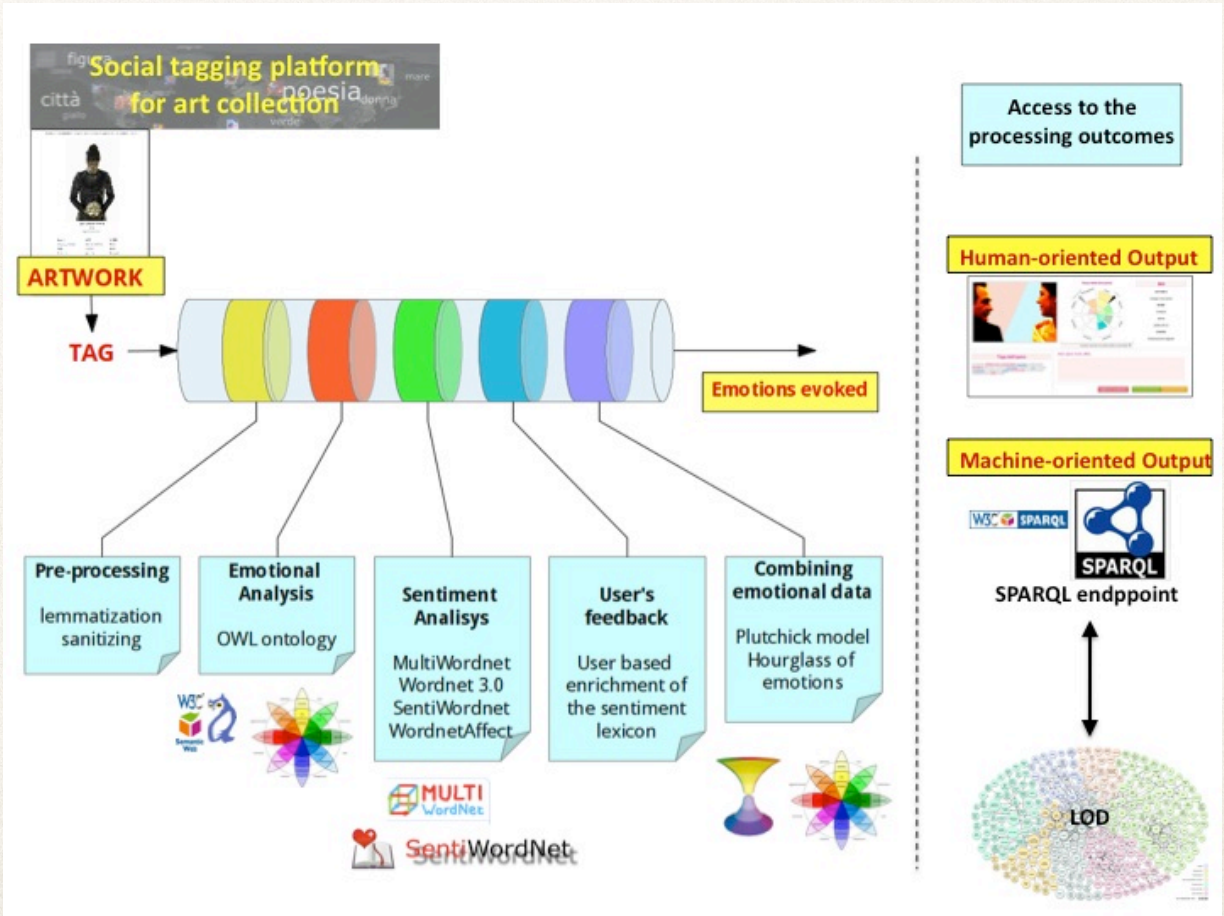
- If a document has more positive than negative words (using the lexicon to decide the polarity of each word in the document), it is classified as positive.
- Often a threshold λ is used, in which a document is classified as positive only if the ratio is greater than λ.
- If the sentiment lexicon includes positive and negative weights for each word, these can be used as well.

Emotion:

- Do the same for each emotion lexicon

Exploiting WN-Affect

- ArsEmotica



Sentiment analysis et al.: how it works

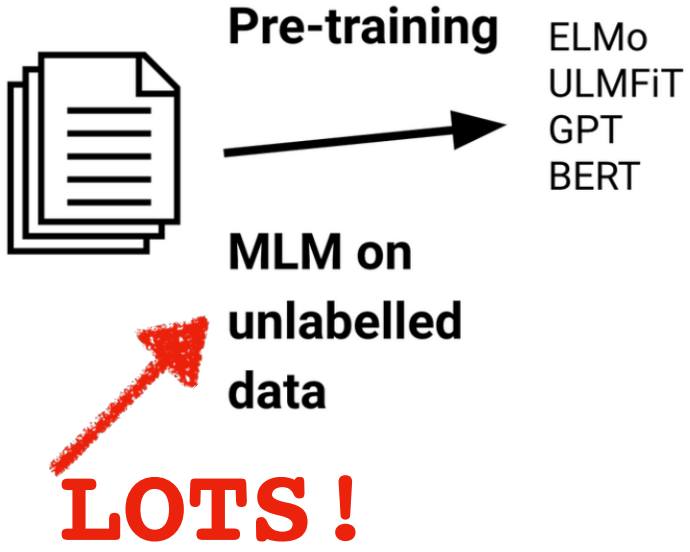
- Feature sets: capture the basic information about each input that should be used to classify / discriminate, e.g between positive and negative sentiment or between ironic and not ironic…
  - Bag-of-words (limits)
  - Punctuation, elongated words, url, expressive signals (emoji, emoticons)
  - Lexical resources for affect. These resources dictionaries can be used to derive features to be used in machine learning algorithms applied to SA.

Exploiting affective resources

- Irony detection/Rumor detection

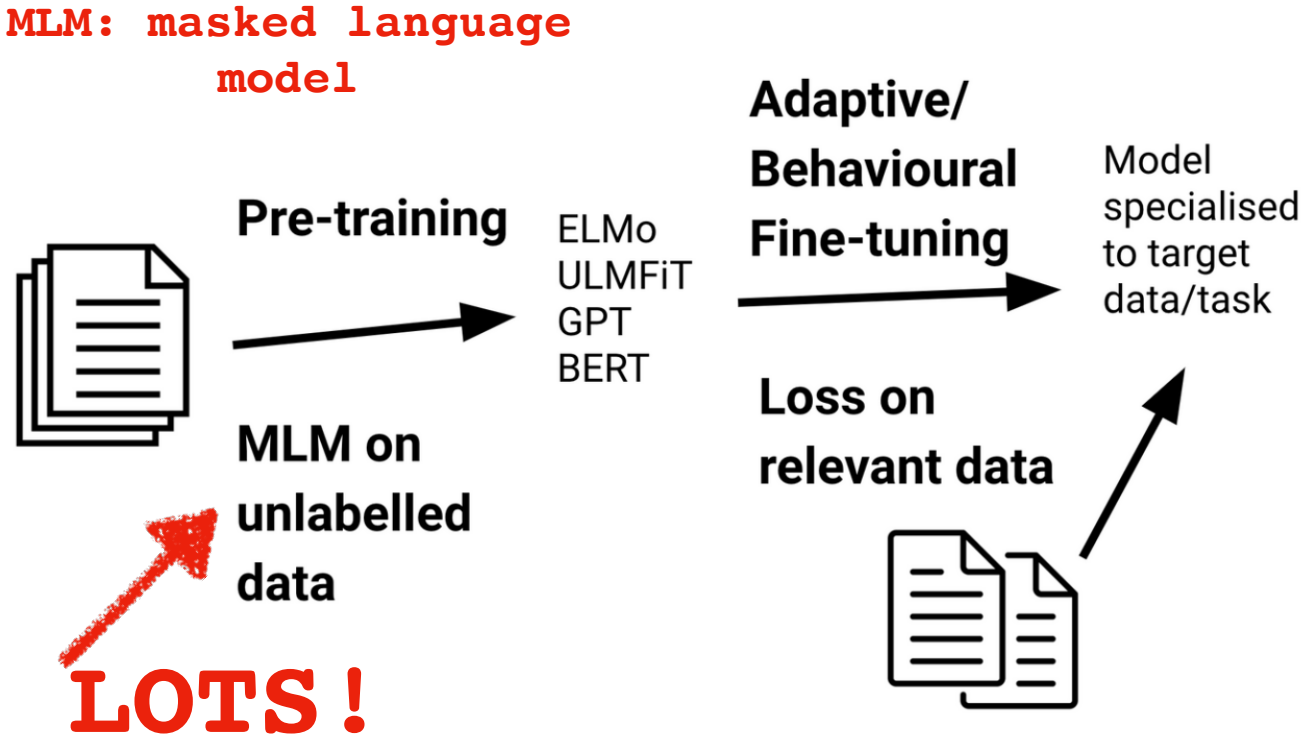
Base Models + Further Training + Fine-Tuning

MLM: masked language model



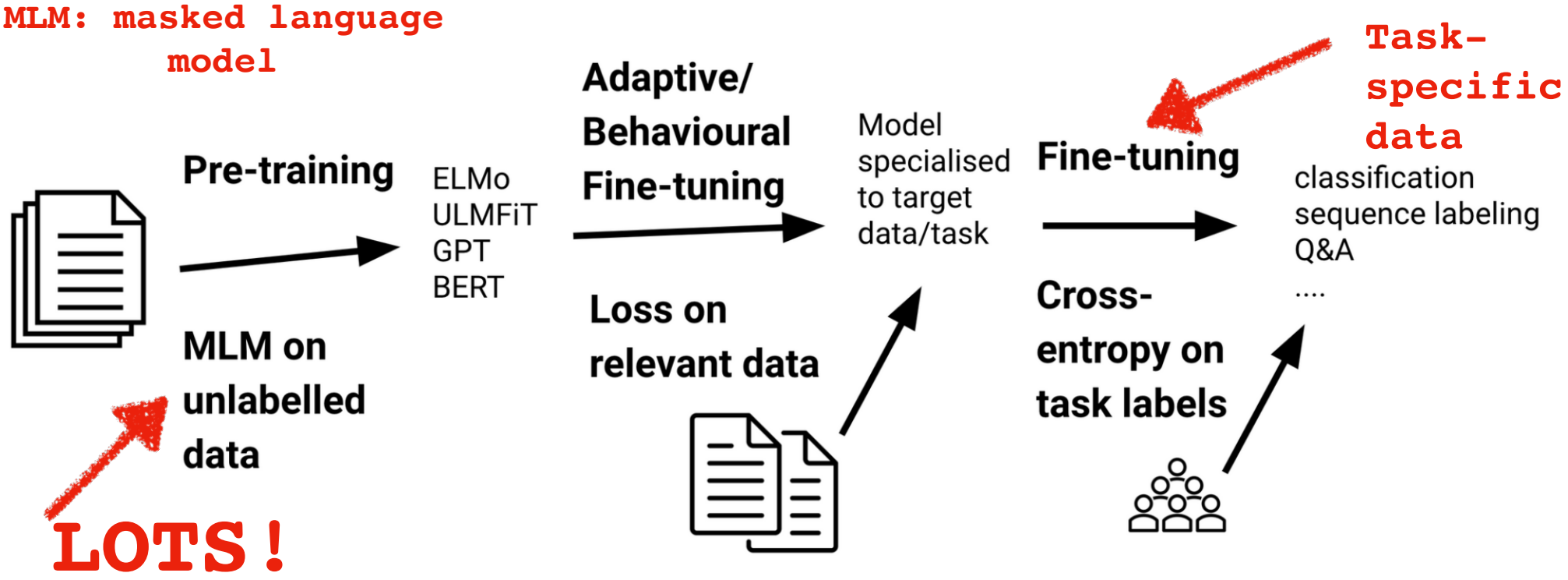
https://ruder.io/recent-advances-lm-fine-tuning/

Base Models + Further Training + Fine-Tuning



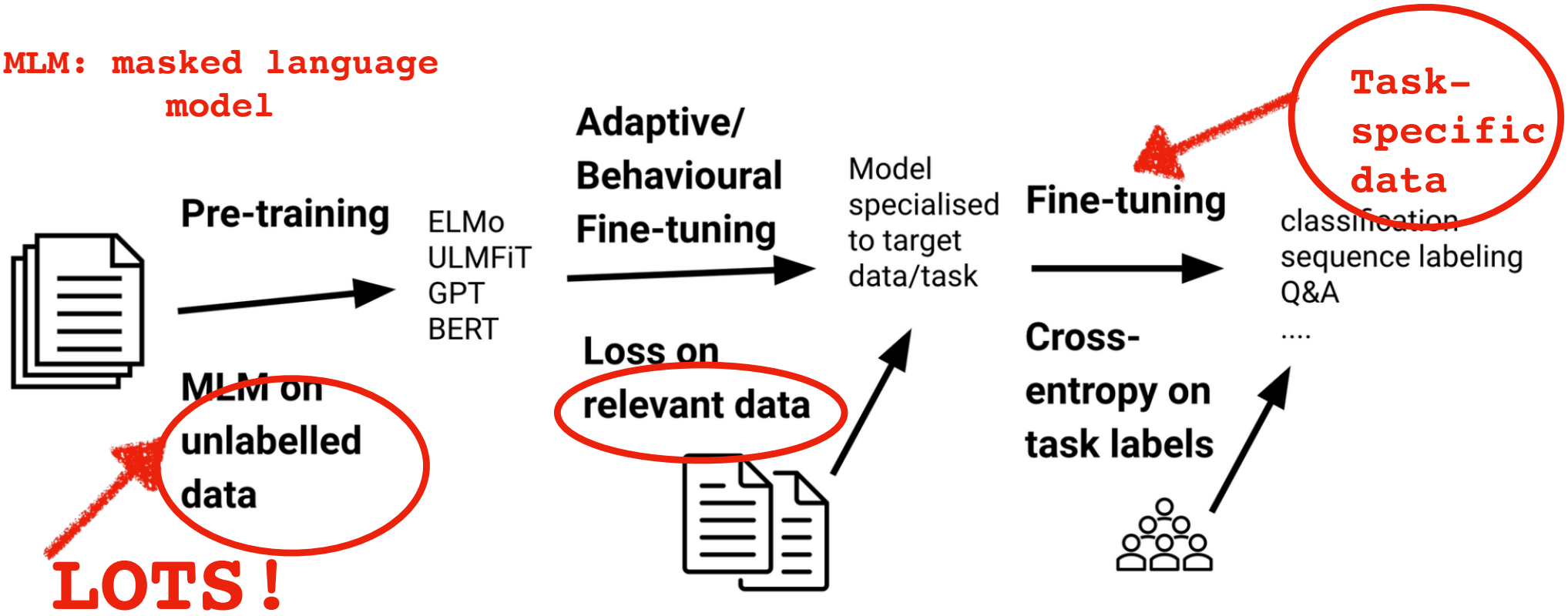
https://ruder.io/recent-advances-lm-fine-tuning/

Base Models + Further Training + Fine-Tuning



https://ruder.io/recent-advances-lm-fine-tuning/

Base Models + Further Training + Fine-Tuning



https://ruder.io/recent-advances-lm-fine-tuning/

The Search for Emotions, Creativity, and Fairness in Language: Saif M. Mohammad

https://www.youtube.com/watch?v=q_ehexD_drA

The Search for Emotions, Creativity, and Fairness in Language. The Alan Turing Institute. March 11, 2019. London, UK.

Slides

Invited talk by Saif Mohammad, NRC Canada