

# Constant (association for art and media)

<http://constantvzw.org>

## Scandinavian Institute for Computational Vandalism

<http://sicv.activearchives.org/>

## Algolit

<http://constantvzw.org/site/-Algolit.184-.html>

## The Botopera

<http://botopera.activearchives.org/>

## The MakeHuman bugreport

[ <http://www.makehuman.org> ]

## Mondotheque

<http://mondotheque.be>

## Relearn

<http://relearn.be>

# A differential word cloud

## Cqrrelations

Poetry to the statistician, science to the dissident and detox to the data-addict.

<http://cqrrelations.constantvzw.org>

## Computational Linguistics & Psycholinguistics (CLiPS)

... is a research center associated with the Linguistics department of the faculty of Arts of the University of Antwerp, and is the result of the fusion of the CNTS and CPL research centers.

Most of the CLiPS research is based on competitively acquired research funding. Funding agencies include the Research Foundation - Flanders, the Institute for the Promotion of Innovation by Science and Technology in Flanders, the Dutch Language Union, the European Commission and occasionally companies.

The goal of CLiPS is to produce internationally recognized top research and resources in (developmental) psycholinguistics, (corpus) linguistics, and computational linguistics, and to investigate the interdisciplinary combinations of these disciplines.

<http://www.clips.ua.ac.be>

<http://www.clips.ua.ac.be/demos>

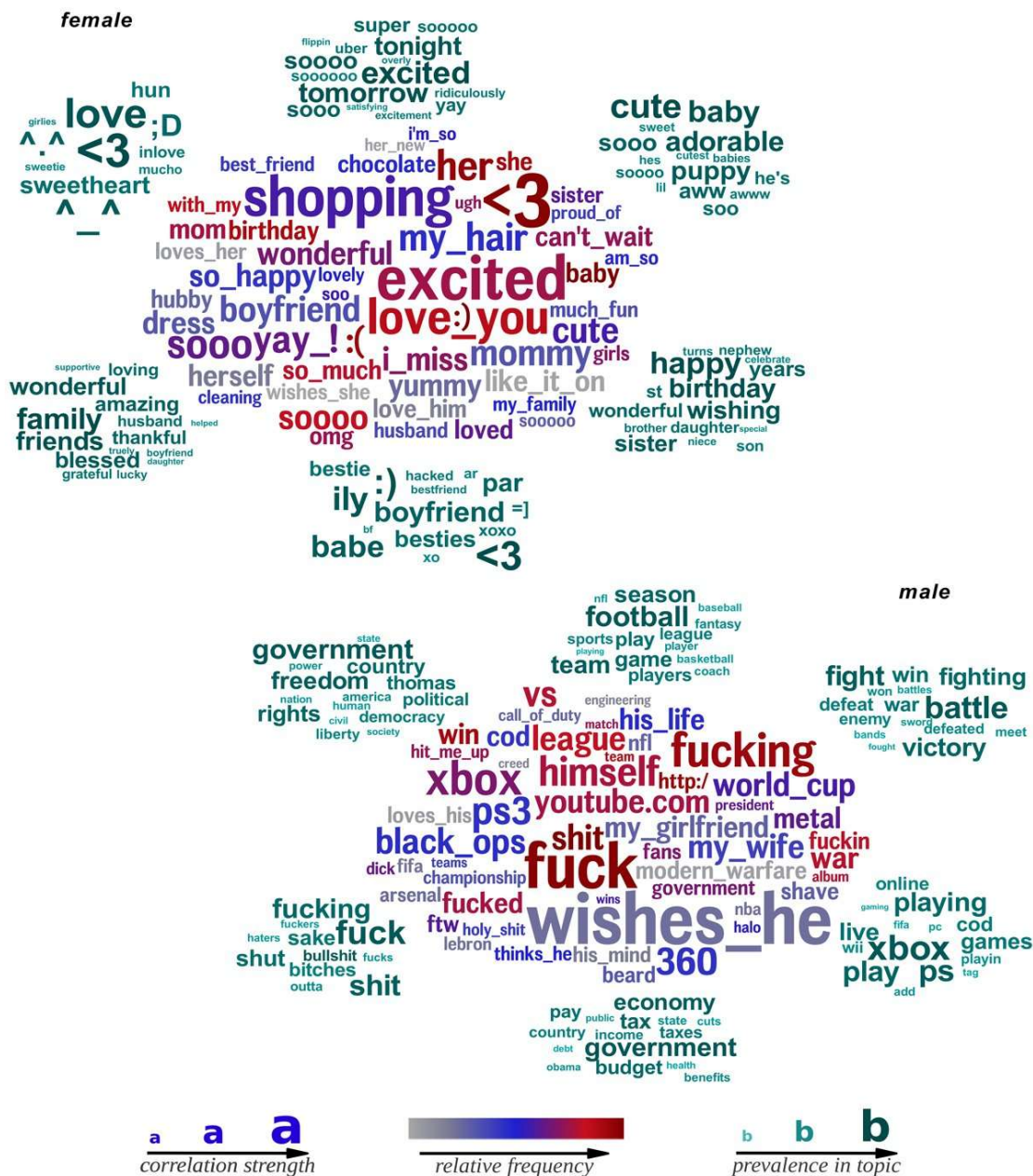
In our *open-vocabulary* technique, the data itself drives a comprehensive exploration of language that distinguishes people, finding connections that are not captured with traditional closed-vocabulary word-category analyses. Our analyses shed new light on psychosocial processes yielding results that are face valid (e.g., subjects living in high elevations talk about the mountains), tie in with other research (e.g., neurotic people disproportionately use the phrase 'sick of' and the word 'depressed'), suggest new hypotheses (e.g., an active life implies emotional stability), and give detailed insights (males use the possessive 'my' when mentioning their 'wife' or 'girlfriend' more often than females use 'my' with 'husband' or 'boyfriend'). To date, this represents the largest study, by an order of magnitude, of language and personality.

In: *Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach* <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3783449/>

## Open Vocabulary Ethics Statement

In seeking insights from language use about personality, gender, and age, we explore two approaches. The first approach, serving as a replication of the past analyses, counts word usage over manually created *a priori* word-category lexica. The second approach, termed *DLA*, serves as our main method and is *open-vocabulary* – the words and clusters of words analyzed are determined by the data itself.

In: *Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach* <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3783449/>



### Words, phrases, and topics most highly distinguishing females and males

Female language features are shown on top while males below. Size of the word indicates the strength of the correlation; color indicates relative frequency of usage. Underscores ( ) connect words of multiword phrases. Words and phrases are in the center; topics, represented as the 15 most prevalent words, surround. (: females and males; correlations adjusted for age; Bonferroni-corrected ).

In: *Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach* <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3785449/>

### Antoinette Rouvroy @ Discrimination & Big Data

Well, about the social normativities. I completely agree that algorithmic normativity, despite the fact that it appears completely a-normative in fact, is a reflection of social or unreflected upon social normativities. An increase. An encouragement of such normativities. But also a naturalisation of these normativities. Which become invisible. Unspeakable. Because they have been translated into ones and zeroes.

*Discrimination and Big Data*. With Geoffrey Bowker, Solon Barocas, Antoinette Rouvroy and Seda Guerses. January 2015, Constant in collaboration with Vlaams-Nederlands Huis de Buren and CDP. <http://video.constantvzw.org/cqrelations/bigdatadiscrimination.webm> + <http://sound.constantvzw.org/cqrelations/big-data-discrimination.mp3>



## Aligning humans and algorithms



### Pattern for Python

Pattern is a web mining module for the Python programming language. It has tools for data mining (Google, Twitter and Wikipedia API, a web crawler, a HTML DOM parser), natural language processing (part-of-speech taggers, n-gram search, sentiment analysis, WordNet), machine learning (vector space model, clustering, SVM), network analysis and <canvas> visualization. The module is free, well-document and bundled with 50+ examples and 350+ unit tests.

<http://www.clips.ua.ac.be/pattern>

### Common applications

- Sentiment mining
- Age prediction
- Gender prediction
- Personality prediction
- Level of education prediction
- Deception detection
- Authorship attribution

### The AMiCA project

The AMiCA (“Automatic Monitoring for Cyberspace Applications”) project aims to mine relevant social media (blogs, chat rooms, and social networking sites) and collect, analyse, and integrate large amounts of information using text and image analysis. The ultimate goal is to trace harmful content, contact, or conduct in an automatic way. Essentially, we take a cross-media mining approach that allows us to detect risks “on-the-fly”. When critical situations are detected (e.g. a very violent communication), alerts can be issued to moderators of the social networking sites. When used on aggregated data, the same technology can be used for incident collection and monitoring at the scale of individual social networking sites. In addition, the technology can provide accurate quantitative data to support providers, science, and government in decision-making processes with respect to child safety online.

Sponsor: IWT - Agentschap voor Innovatie door Wetenschap en Technologie (Agency for Innovation by Science and Technology)

<http://amicaproject.be/>

## Steps

- The availability of abundant machine-readable data or sources
- A corpus of pre-analysed and pre-parsed data, for testing and training purposes
- A 'Gold Standard' derived from manually (validated by humans) annotated corpi
- Standard parsing algorithms that can pre-process texts for more efficient analysis
- Pattern recognition algorithms [TF-IDF, K-nn]
- Training software (Machine learning) that allows the algorithms to be optimized against the Gold Standard

## Remembering annotation

Generally I start with the sky and then I continue adding everything else. For enclosed spaces, the first thing that I label is the ceiling. The order of the annotations does not really matter, but one has to find what is more enjoyable or easier. Once I annotate the ceiling I label the walls and then all the other elements in the room, finishing with the floor.

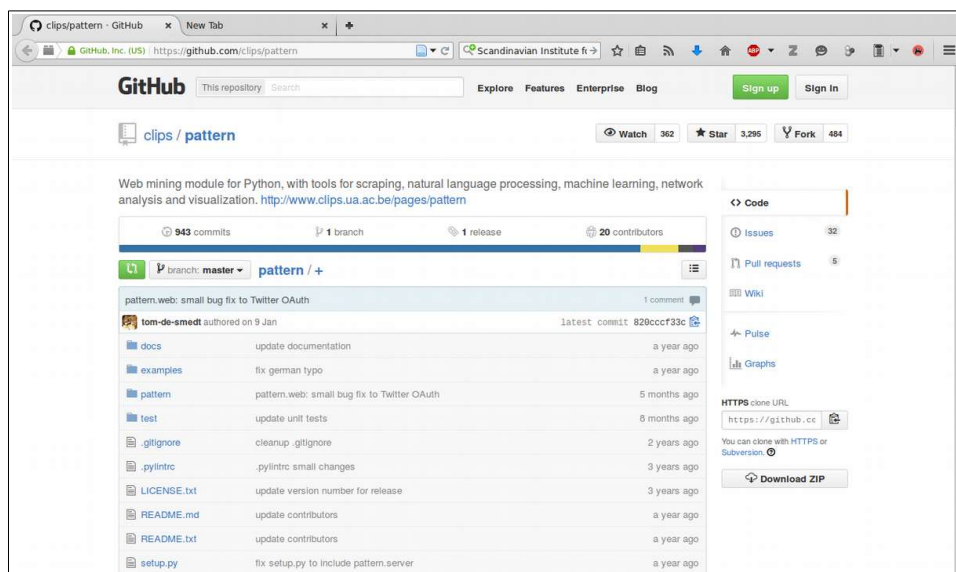
Adela Barriuso, Antonio Torralba: Notes on Image Annotation. Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology <http://people.csail.mit.edu/torralba/publications/memories.pdf>

LabelMe <http://labelme.csail.mit.edu/Release3.0>

Machine-learning algorithms that partially automate data processing still need to be trained for every new form, or every new kind of topic the algorithm might deal with. (...) Such work of alignment is not a bug — it is the condition of possibility for keeping humans and automation working in the same world.

Lilly Irani: Justice for Data Janitors (2015) <http://www.publicbooks.org/nonfiction/justice-for-data-janitors>

## pattern.en.paternalism



## Setting paternalism detection as a task

### Guidelines for the Fine-Grained Analysis of Polar Expressions

Polar facts and some polar resultative causatives do not explicitly express sentiment towards a target entity. They contain factual information from which a positive or negative evaluation of a certain entity can be inferred using common sense or world knowledge. In other words, in order to determine the polar expression's polarity, interpretation is needed. This poses a problem in certain text types because

- sometimes interpretation requires domain-specific knowledge
- sometimes a polar fact or polar resultative causative can be interpreted from different perspectives

Annotators are encouraged to annotate any expression which they think to be polar, even if they are not entirely sure

whether the expressed sentiment is positive or negative. Using the 'unknown' label for polar expressions with unknown or ambiguous polarity enables us to determine the polarity of these expressions with the help of domain experts in a later annotation stage. Of course, the polarity labels 'positive', 'negative' and 'other' should be used as often as possible.

Marjan Van de Kauter and Bart Desmet : Guidelines for the Fine-Grained Analysis of Polar Expressions . LT3 Technical Report – LT3 13-01 . University College Ghent (2013)

**PATERNALISM (OR PARENTALISM)** is behavior, by a person, organization or state, which limits some person or group's liberty or autonomy for that person's or group's own good. Paternalism can also imply that the behavior is against or regardless of the will of a person, or also that the behavior expresses an attitude of superiority.

<https://en.wikipedia.org/wiki/Paternalism>

**LE PATERNALISME** est une doctrine politique qui définit comme moralement souhaitable qu'un agent privé ou public puisse décider à la place d'un autre pour son bien propre. Cette doctrine s'oppose au libéralisme.

From: <https://fr.wikipedia.org/wiki/Paternalisme>

**PATERNALISME** verwijst naar een houding of beleid vergelijkbaar met het hiërarchische familiepatroon waarbij de vader (pater in het Latijn) aan het hoofd van de familie staat en de vader beslissingen neemt voor de andere familieleden (vrouw en kinderen), ook als die beslissing niet in overeenstemming is met wat zij wensen.

<https://nl.wikipedia.org/wiki/Paternalisme>

**PATERNALÍSM** (Ec. Pol.) Concepție care desemnează interesul pe care îl manifestă patronii pentru bunăstarea muncitorilor sau pentru atmosfera familială din întreprindere, raporturile dintre patroni și muncitori caracterizate prin afecțiune reciprocă, autoritate și respect. (*Political economy*) *Notion referring to the interest patrons have in the welfare of the workers or in the family-like atmosphere at the work place, relations between employers and workers characterized by mutual affection, authority, and respect.*

Dicționarul explicativ al limbii române, 2009 <http://dexonline.ro/definitie/paternalism> (there is no definition for paternalism in Romanian Wikipedia)

## Data-set

### Gutenberg project

J. B. Bury, The Idea Of Progress, 1920, <http://www.gutenberg.org/cache/epub/4557/pg4557.txt>

Maud Churton Braby, Modern Marriage and How To Bear It, 1908, <https://www.gutenberg.org/files/31529/31529-0.txt>

Harriet Martineau, How to Observe Morals and Manners, 1838,

<http://www.gutenberg.org/cache/epub/33944/pg33944.txt>

Irwin Edman, Human Traits and their Social Significance, 1920,

<http://www.gutenberg.org/cache/epub/22306/pg22306.txt>

James Hayden Tufts, The Ethics of Cooperation, 1918, <http://www.gutenberg.org/cache/epub/29508/pg29508.txt>

James Harvey Robinson, The Mind in the Making: The Relation of Intelligence to Social Reform, 1921,

<http://www.gutenberg.org/cache/epub/8077/pg8077.txt>

Helen Kendrick Johnson, Woman And The Republic, 1897, <https://www.gutenberg.org/cache/epub/7300/pg7300.txt>

Charles Darwin, On the Origin of species, 1859, <http://www.gutenberg.org/cache/epub/1228/pg1228.txt>

Emma Goldman, Anarchism and other essays, 1910, <http://www.gutenberg.org/cache/epub/2162/pg2162.txt>

John F. Hume, The Abolitionists (Together With Personal Memories Of The Struggle For Human Rights), 1830-1864,

<http://www.gutenberg.org/cache/epub/13176/pg13176.txt>

### Wikipedia

Mining: <https://en.wikipedia.org/wiki/Mining>

Textile Industry: [https://en.wikipedia.org/wiki/Textile\\_industry](https://en.wikipedia.org/wiki/Textile_industry)

History of computing hardware: [https://en.wikipedia.org/wiki/History\\_of\\_computing\\_hardware](https://en.wikipedia.org/wiki/History_of_computing_hardware)

Marissa Mayer: [https://en.wikipedia.org/wiki/Marissa\\_Mayer](https://en.wikipedia.org/wiki/Marissa_Mayer)

Larry Page: [https://en.wikipedia.org/wiki/Larry\\_Page](https://en.wikipedia.org/wiki/Larry_Page)

Liberty: <https://en.wikipedia.org/wiki/Liberty>

Choice: <https://en.wikipedia.org/wiki/Choice>

Sabotage: <http://en.wikipedia.org/wiki/Sabotage>

Social Darwinism : [http://en.wikipedia.org/wiki/Social\\_Darwinism](http://en.wikipedia.org/wiki/Social_Darwinism)

Anarchism: <https://en.wikipedia.org/wiki/Anarchism>

# The Annotators

- Annotator 001 (f, 1982) is a French author living in Belgium. She is currently involved in a research to the life of Anna Kavan, and is interested in digital writing.
- Annotator 002 (f, 1969) is a Dutch designer/artist living in Belgium. She is a feminist and interested in tools, practice and Free Software.
- Annotator 003 (m, 1990) is a Dutch artist living in The Netherlands. He is interested in infrastructures and networks.
- Annotator 004 (f, 1988) is a French artist living in The Netherlands. She is interested in the physical location of the web, and enjoys the act of making web pages.
- Annotator 005 (f, 1989) is a Dutch designer living in The Netherlands. She is interested in language philosophy and computational linguistics.
- Annotator 006 (f, 1991) is a Romanian curator living in The Netherlands. She is interested in the conditional aspect of interfaces.
- Annotator 007 (m, 1982) is a Hungarian researcher living in Spain. He is interested in collaborative production practices and cybernetics as an ideological formation.
- Annotator 008 (m, 1984) wishes that he was 007. Why? Because 7, 8, 9... No 008 has an English mother-tongue and has been annotating considerably in professional and non professional contexts (including with building ontological frameworks since 2008 (fuck me, that was a long time ago). Besides, the reading material themes in 008s remit is close to 'natural reading environment' He has been permitted to retire at 250, unless somebody is able to catch up with him!
- Annotator 009 (f, 1975) is a French researcher/teacher living in France. She is interested in bots.

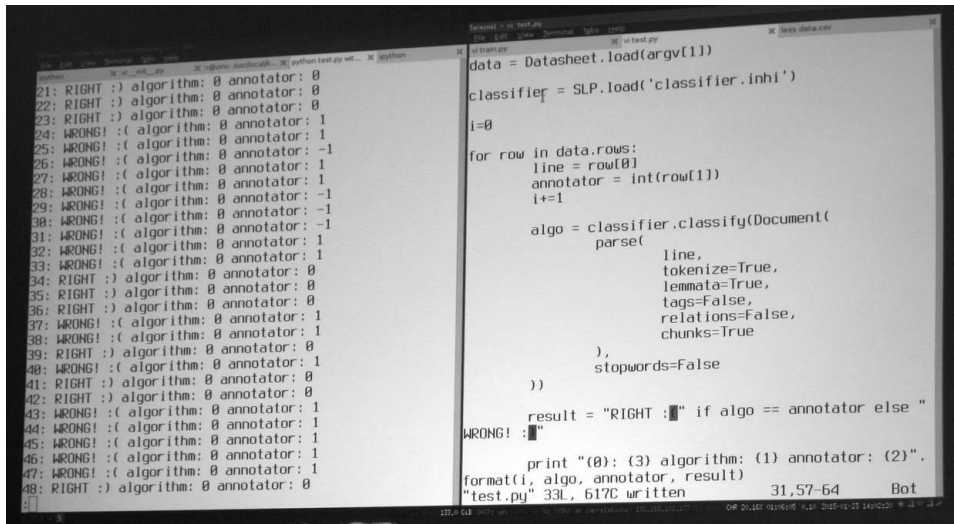
## Notes (sample)

- Annotator 008: When I reached 50, I went back to the beginning and decided to make some of the earlier annotations more neutral. Some paragraphs can not be definitive without wider context – therefore I took a more conservative approach.
- Annotator 003: Made most of my decisions based on 'keywords' that I thought the algorithm should learn to 'flag' as paternalist. Found it difficult to gauge the level of more factual texts.
- Annotator 005: wasn't that familiar with the term 'paternalism', which evolved along the process of labeling. An other strategy was applied later in the process: to focus more on writing style, rather than the content. Although, this was not always possible to apply. In the later classifications, there are comments valued with a capital "W" for a focus on the writing style, and a capital "C" for a focus on the content.
- Annotator 006: Took context (year of publication) into account.
- Annotator 002: Felt it was most difficult to decide whether style or content should be taken into account.



## Comments (sample)

- Annotator 007 on paragraph #549: Reduces questions of political agency to physiological problems.
- Annotator 008 on paragraph #334: Analysis of individual's history and philosophical outlook.
- Annotator 003 on paragraph #416: strenuously controlling sex
- Annotator 006 on paragraph #416: For 1908 it raises feminist issues
- Annotator 009 on paragraph #416: I write 0 not because it's neutral but as a kind of balance as I couldn't choose between -1 and 1. there are elements that can be considered emancipatory, against paternalism (the text is from 1908), but there are also elements which are paternalist as well.
- Annotator 004 on paragraph #442: What gives Bernard Shaw the aptitude to reveal the deep nature of men and woman?



```
data = Datasheet.load(argv[1])
classifier = SLP.load('classifier.inhi')
i=0
for row in data.rows:
    line = row[0]
    annotator = int(row[1])
    i+=1
    algo = classifier.classify(Document(
        parse(
            line,
            tokenize=True,
            lemmata=True,
            tags=False,
            relations=False,
            chunks=True
        )
    ),
    stopwords=False
))
    result = "RIGHT :█" if algo == annotator else "
WRONG! :█"
    print "{0}: {3} algorithm: {1} annotator: {2}"
format(i, algo, annotator, result)
"test.py" 33L, 617C written          31,57-64 Bot
```

## The Removal of Pascal

### Features before removal

[u'w', u'g', u'ineffa\xe7able', u'hp', u'magazine', u't', u'autre', u's', u'r', u'il', u'laugh', u'48', u'100', u'suitable', u'une']

Une différente coutume donnera d'autres principes naturels. Cela se voit par expérience; et s'il y en a d'ineffaçables à la coutume, il y en a aussi de la coutume ineffaçables à la nature. *A different custom will cause different natural principles. This is seen in experience; and if there are some natural principles ineradicable by custom, there are also some customs opposed to nature, ineradicable by nature, or by a second custom.*

Blaise Pascal, quoted in Harriet Martineau: *How to observe morals and manners*

### Features after removal

[u'laugh', u'g', u'hp', u'magazine', u's', u'r', u'une', u'w', u'48', u'100', u'suitable', u't']

## Dissent

- Group A (001, 004, 007): Paragraphs labeled: 174. 20 of those paragraphs were annotated by 1 person (and not taken into account). Disagreements: 18
- Group B (002, 005, 008): Paragraphs labeled: 55. 5 of those paragraphs were annotated by 1 person (and not taken into account). Disagreements: 21
- Group C (003, 006, 009): Paragraphs labeled: 61. 10 of those paragraphs were annotated by 1 person (and not taken into account). Disagreements: 10

After one day, 244 paragraphs were classified and ready for training. Annotators disagreed on whether a paragraph was paternalist on 49 occasions, bringing the annotator disagreement rate to 20.08967213114754%

- #214 Mining of stone and metal has been done since pre-historic times. Modern mining processes involve prospecting for ore bodies, analysis of the profit potential of a proposed mine, extraction of the desired materials, and final reclamation of the land after the mine is closed.
- #215 Mining techniques can be divided into two common excavation types: surface mining and sub-surface (underground) mining. Today, surface mining is much more common, and produces, for example, 85% of minerals (excluding petroleum and natural gas) in the United States, including 98% of metallic ores.[26]
- #216 Modern anarchism sprang from the secular or religious thought of the Enlightenment, particularly Jean-Jacques Rousseau's arguments for the moral centrality of freedom.
- #217 Modern conceptions of democracy are founded on the idea of popular sovereignty.
- #218 Most girls are aware from a very early age of the social advantages and importance of marriage, and grow up with a keen desire to accomplish itin due course, although secretly dreading it, because of their absurdperverted ideas of its physical side. Why cannot girls--and boys too,for that matter--be taught the plain truth (in suitable language of course) that sex is the pivot on which the world turns, that the instincts and emotions of sex are common to humanity, and in themselvesnot base or degrading, nor is there any cause for shame in possessing them, although it is necessary that they should be strenuously controlled. Why cannot girls be taught that \_all love\_, even theromantic love which occupiess so large a portion of their dreams, \_springs from the instinct of sex\_?[4] This may be thought a dangerous lesson, but the present policy of silence on this subject is far more dangerous, inducing as it does a tendency to brood over the forbidden theme.
- #219 Most of us do not stop to think of the conditions of an animal existence. When we read the descriptions of our nature as given by William James, McDougall, or even Thorndike, with all his reservations,we get a rather impressive idea of our possibilities, not a picture of uncivilized life. When we go camping we think that we are deserting civilization, forgetting the sophisticated guides, and the pack horsesladen with the most artificial luxuries, many of which would not havebeen available even a hundred years ago. We lead the simple life with Swedish matches, Brazilian coffee, Canadian bacon, California cannedpeaches, magazine rifles, jointed fishing rods, and electric flashlights. We are elaborately clothed and can discuss Bergson'sviews or D. H. Lawrence's last story. We naïvely imagine we are returning to "primitive" conditions because we are living out of doors or sheltered in a less solid abode than usual, and have to go to the brook for water.(...)

- #214           The Algorithm: It is 'neutral'  
              The Annotator: It is paternalist  
              (they disagree)
- #215           The Algorithm: It is 'neutral'  
              The Annotator: It is 'neutral'  
              (they agree)
- #216           The Algorithm: It is 'neutral'  
              The Annotator: It is paternalist  
              (they disagree)
- #217           The Algorithm: It is 'neutral'  
              The Annotator: It is paternalist  
              (they disagree)
- #218           The Algorithm: It is 'neutral'  
              The Annotator: It is 'neutral'  
              (they agree)
- #219           The Algorithm: It is 'neutral'  
              The Annotator: It is paternalist  
              (they disagree)

The Annotators and The Algorithm discussed 228 cases. They agreed on 111 case(s) and disagreed on 117 case(s). The Algorithm missed paternalism on 86 cases(s) but invented paternalism on 1 case(s). The Annotators had an Algorithm-Annotator-disagreement-rate of 51.3157894737%

pattern.en.paternalism: Catherine Lenoble, Anne Laforet, Femke Snelting, Roel Roscam Abbing, Manetta Berends, Julie Boschat Thorez, Cristina Cochior, Maxigas and Johnny.