Lab notebook

KBC

4/5/2021

2021-04-05

I extracted the terms from the current version of the RO, then got the absolute frequencies of their words:

/Users/kevincohen/Dropbox/Scripts-new/lexicalFrequency.pl experimental-outputs/ro.terms.txt

. . . which reveals imbalances in the use of words that are clear opposites, e.g.:

- during 11 before 4 after 2
- to 49 from 16 towards 1
- directly 18
- indirectly 8 indirect 2 direct 2

. . . which suggests that I could be adding a bunch of opposite terms to the ontology.

So, I ran the current version of my antonym-finding script:

code/findAntonyms.pl resources/ontologies/ro.2021-03-08.obo.txt |wc -l

. . . which finds 43 pairs of opposites.

Next step: find overlap with Mike's resources/ontologies/predicates.txt.

PATO: opposites already annotated

```
library("ggVennDiagram")
```

```
## Warning: package 'ggVennDiagram' was built under R version 4.0.2
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2
```

```
## -- Attaching packages ------------------------------------------------------------- tidyverse
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.4
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2

## Warning: package 'tidyr' was built under R version 4.0.2

## Warning: package 'readr' was built under R version 4.0.2

## Warning: package 'dplyr' was built under R version 4.0.2

## Warning: package 'forcats' was built under R version 4.0.2


## -- Conflicts ---------------------------------------------------------------------------- tidyverse_confl
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
# NB: need to add reduce/increase. Do HP, then RO.

# only found by me
# comm -23 experimental-outputs/pato.mine.terms.only.tsv resources/current-opposites.sorted.terms.only.
#       59

# only in current annotations
#comm -13 experimental-outputs/pato.mine.terms.only.tsv resources/current-opposites.sorted.terms.only.t
#       27


# in both
#comm -12 experimental-outputs/pato.mine.terms.only.tsv resources/current-opposites.sorted.terms.only.t
#       188

# https://www.datanovia.com/en/fr/blog/magnifique-diagramme-de-venn-ggplot-avec-r/

# this separates out the three entirely--not what the ggVennDiagram() function is looking for
# Here I'm just getting the terms themselves in printable fashion, separating them by who "finds" them
pato.mine.only <- as.vector(read_lines("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/e
pato.mine.only <- stringr::str_split(pato.mine.only, pattern = "\t")
#pato.mine.only
current.only <- as.vector((read_lines("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/exp
current.only <- stringr::str_split(current.only, pattern = "\t")
both.mine.and.current <- as.vector((read_lines("/Users/kevincohen/Dropbox/N-Z/translator-concept-opposi
both.mine.and.current <- stringr::str_split(both.mine.and.current, pattern = "\t")

# here I have *everything* that I found and *everything* in current, so that we can quantify overlap &
pato.mine <- as.vector(read_lines("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experi
current <- as.vector(read_lines("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/resources

x <- list(Mine = pato.mine, Current = current)
ggVennDiagram(x)
```
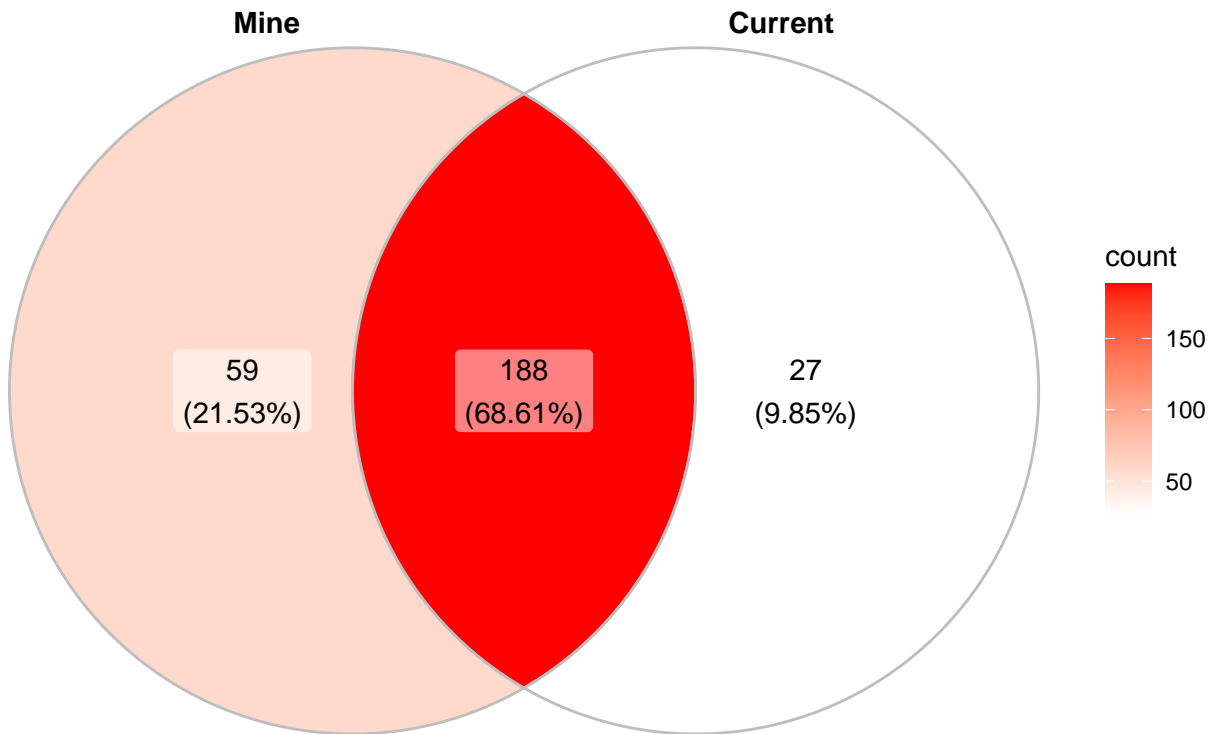
Mine   Current

59 (21.53%)	188 (68.61%)	27 (9.85%)

count

```
print("In both mine and 'current':")
```

```
## [1] "In both mine and 'current':"
```

```
print(both.mine.and.current)
```

```
## [[1]]
## [1] "abnormal" "normal"
##
## [[2]]
## [1] "absent"  "present"
##
## [[3]]
## [1] "active"   "inactive"
##
## [[4]]
## [1] "acute"   "chronic"
##
## [[5]]
## [1] "adhesive"     "non-adhesive"
##
## [[6]]
## [1] "aerobic"   "anaerobic"
##
## [[7]]
```

```
## [1] "aerobic (for occurrence)"   "anaerobic (for occurrence)"
##
## [[8]]
## [1] "aligned with"   "misaligned with"
##
## [[9]]
## [1] "aromatic"      "non-aromatic"
##
## [[10]]
## [1] "arrhythmic" "rhythmic"
##
## [[11]]
## [1] "asynchronous" "synchronous"
##
## [[12]]
## [1] "back"  "front"
##
## [[13]]
## [1] "balanced"   "unbalanced"
##
## [[14]]
## [1] "branched"    "unbranched"
##
## [[15]]
## [1] "closed" "open"
##
## [[16]]
## [1] "coiled"    "uncoiled"
##
## [[17]]
## [1] "colored"    "discolored"
##
## [[18]]
## [1] "compatible"   "incompatible"
##
## [[19]]
## [1] "complete structure"   "incomplete structure"
##
## [[20]]
## [1] "condensed"   "decondensed"
##
## [[21]]
## [1] "conspicuous"   "inconspicuous"
##
## [[22]]
## [1] "continuous"    "discontinuous"
##
## [[23]]
## [1] "contractile"      "non-contractile"
##
## [[24]]
## [1] "coordinated"   "uncoordinated"
##
## [[25]]
```

```
## [1] "crowded"    "uncrowded"
##
## [[26]]
## [1] "damaged"    "undamaged"
##
## [[27]]
## [1] "decreased accumulation" "increased accumulation"
##
## [[28]]
## [1] "decreased acidity" "increased acidity"
##
## [[29]]
## [1] "decreased adhesivity" "increased adhesivity"
##
## [[30]]
## [1] "decreased affinity" "increased affinity"
##
## [[31]]
## [1] "decreased age" "increased age"
##
## [[32]]
## [1] "decreased amount" "increased amount"
##
## [[33]]
## [1] "decreased angle to" "increased angle to"
##
## [[34]]
## [1] "decreased anterior-posterior diameter"
## [2] "increased anterior-posterior diameter"
##
## [[35]]
## [1] "decreased area" "increased area"
##
## [[36]]
## [1] "decreased avidity" "increased avidity"
##
## [[37]]
## [1] "decreased behavioural activity" "increased behavioural activity"
##
## [[38]]
## [1] "decreased branchiness" "increased branchiness"
##
## [[39]]
## [1] "decreased cellular motility" "increased cellular motility"
##
## [[40]]
## [1] "decreased circumference" "increased circumference"
##
## [[41]]
## [1] "decreased coiling" "increased coiling"
##
## [[42]]
## [1] "decreased combustibility" "increased combustibility"
##
```

5

```
## [[43]]
## [1] "decreased concentration" "increased concentration"
##
## [[44]]
## [1] "decreased contractility" "increased contractility"
##
## [[45]]
## [1] "decreased coordination" "increased coordination"
##
## [[46]]
## [1] "decreased curvature" "increased curvature"
##
## [[47]]
## [1] "decreased degree of illumination" "increased degree of illumination"
##
## [[48]]
## [1] "decreased depth" "increased depth"
##
## [[49]]
## [1] "decreased diameter" "increased diameter"
##
## [[50]]
## [1] "decreased distance" "increased distance"
##
## [[51]]
## [1] "decreased distribution" "increased distribution"
##
## [[52]]
## [1] "decreased duration" "increased duration"
##
## [[53]]
## [1] "decreased duration of temperature" "increased duration of temperature"
##
## [[54]]
## [1] "decreased efficacy" "increased efficacy"
##
## [[55]]
## [1] "decreased efficiency" "increased efficiency"
##
## [[56]]
## [1] "decreased elasticity" "increased elasticity"
##
## [[57]]
## [1] "decreased elevation" "increased elevation"
##
## [[58]]
## [1] "decreased fatigability" "increased fatigability"
##
## [[59]]
## [1] "decreased fecundity" "increased fecundity"
##
## [[60]]
## [1] "decreased female fertility" "increased female fertility"
##
```

```
## [[61]]
## [1] "decreased female receptivity" "increased female receptivity"
##
## [[62]]
## [1] "decreased fertility" "increased fertility"
##
## [[63]]
## [1] "decreased flexibility" "increased flexibility"
##
## [[64]]
## [1] "decreased fluid flow" "increased fluid flow"
##
## [[65]]
## [1] "decreased fluorescence" "increased fluorescence"
##
## [[66]]
## [1] "decreased force" "increased force"
##
## [[67]]
## [1] "decreased fragility" "increased fragility"
##
## [[68]]
## [1] "decreased frequency" "increased frequency"
##
## [[69]]
## [1] "decreased functionality" "increased functionality"
##
## [[70]]
## [1] "decreased height" "increased height"
##
## [[71]]
## [1] "decreased humidity" "increased humidity"
##
## [[72]]
## [1] "decreased intensity" "increased intensity"
##
## [[73]]
## [1] "decreased length" "increased length"
##
## [[74]]
## [1] "decreased life span" "increased life span"
##
## [[75]]
## [1] "decreased magnetism" "increased magnetism"
##
## [[76]]
## [1] "decreased magnitude" "increased magnitude"
##
## [[77]]
## [1] "decreased male fertility" "increased male fertility"
##
## [[78]]
## [1] "decreased male receptivity" "increased male receptivity"
##
```

```
## [[79]]
## [1] "decreased mass" "increased mass"
##
## [[80]]
## [1] "decreased mass density" "increased mass density"
##
## [[81]]
## [1] "decreased mobility" "increased mobility"
##
## [[82]]
## [1] "decreased object quality" "increased object quality"
##
## [[83]]
## [1] "decreased occurrence" "increased occurrence"
##
## [[84]]
## [1] "decreased odor" "increased odor"
##
## [[85]]
## [1] "decreased osmolality" "increased osmolality"
##
## [[86]]
## [1] "decreased osmolarity" "increased osmolarity"
##
## [[87]]
## [1] "decreased perimeter" "increased perimeter"
##
## [[88]]
## [1] "decreased permeability" "increased permeability"
##
## [[89]]
## [1] "decreased phosphorylation" "increased phosphorylation"
##
## [[90]]
## [1] "decreased photosensitivity" "increased photosensitivity"
##
## [[91]]
## [1] "decreased pigmentation" "increased pigmentation"
##
## [[92]]
## [1] "decreased porosity" "increased porosity"
##
## [[93]]
## [1] "decreased position" "increased position"
##
## [[94]]
## [1] "decreased pressure" "increased pressure"
##
## [[95]]
## [1] "decreased process quality" "increased process quality"
##
## [[96]]
## [1] "decreased quality" "increased quality"
##
```

```
## [[97]]
## [1] "decreased radioactivity" "increased radioactivity"
##
## [[98]]
## [1] "decreased radiopacity" "increased radiopacity"
##
## [[99]]
## [1] "decreased rate" "increased rate"
##
## [[100]]
## [1] "decreased resistance to" "increased resistance to"
##
## [[101]]
## [1] "decreased sensitivity of a process" "increased sensitivity of a process"
##
## [[102]]
## [1] "decreased sensitivity of a process to oxygen"
## [2] "increased sensitivity of a process to oxygen"
##
## [[103]]
## [1] "decreased sensitivity to irradiation"
## [2] "increased sensitivity to irradiation"
##
## [[104]]
## [1] "decreased sensitivity toward" "increased sensitivity toward"
##
## [[105]]
## [1] "decreased size" "increased size"
##
## [[106]]
## [1] "decreased solubility" "increased solubility"
##
## [[107]]
## [1] "decreased speed" "increased speed"
##
## [[108]]
## [1] "decreased stability" "increased stability"
##
## [[109]]
## [1] "decreased strength" "increased strength"
##
## [[110]]
## [1] "decreased susceptibility toward" "increased susceptibility toward"
##
## [[111]]
## [1] "decreased temperature" "increased temperature"
##
## [[112]]
## [1] "decreased tendency" "increased tendency"
##
## [[113]]
## [1] "decreased thickness" "increased thickness"
##
## [[114]]
```

```
## [1] "decreased threshold" "increased threshold"
##
## [[115]]
## [1] "decreased tolerance to" "increased tolerance to"
##
## [[116]]
## [1] "decreased tonicity" "increased tonicity"
##
## [[117]]
## [1] "decreased turgor" "increased turgor"
##
## [[118]]
## [1] "decreased variability" "increased variability"
##
## [[119]]
## [1] "decreased variability of color" "increased variability of color"
##
## [[120]]
## [1] "decreased variability of rate" "increased variability of rate"
##
## [[121]]
## [1] "decreased variability of size" "increased variability of size"
##
## [[122]]
## [1] "decreased variability of temperature"
## [2] "increased variability of temperature"
##
## [[123]]
## [1] "decreased velocity" "increased velocity"
##
## [[124]]
## [1] "decreased viscosity" "increased viscosity"
##
## [[125]]
## [1] "decreased volume" "increased volume"
##
## [[126]]
## [1] "decreased waxiness" "increased waxiness"
##
## [[127]]
## [1] "decreased weight" "increased weight"
##
## [[128]]
## [1] "decreased wetness" "increased wetness"
##
## [[129]]
## [1] "decreased width" "increased width"
##
## [[130]]
## [1] "decreased width and length" "increased width and length"
##
## [[131]]
## [1] "defasciculated" "fasciculated"
##
```

```
## [[132]]
## [1] "degenerate"     "non-degenerate"
##
## [[133]]
## [1] "dephosphorylated" "phosphorylated"
##
## [[134]]
## [1] "differentiated"   "undifferentiated"
##
## [[135]]
## [1] "discriminate"   "indiscriminate"
##
## [[136]]
## [1] "disorganized" "organized"
##
## [[137]]
## [1] "disoriented" "oriented"
##
## [[138]]
## [1] "distributed"   "undistributed"
##
## [[139]]
## [1] "down" "up"
##
## [[140]]
## [1] "dry" "wet"
##
## [[141]]
## [1] "edible"   "inedible"
##
## [[142]]
## [1] "efficient"   "inefficient"
##
## [[143]]
## [1] "elastic"   "inelastic"
##
## [[144]]
## [1] "elongated" "shortened"
##
## [[145]]
## [1] "female fertile" "female sterile"
##
## [[146]]
## [1] "female semi-fertile" "female semi-sterile"
##
## [[147]]
## [1] "fertile" "sterile"
##
## [[148]]
## [1] "flexible"   "inflexible"
##
## [[149]]
## [1] "fragile"     "non-fragile"
##
```

```
## [[150]]
## [1] "functional"     "non-functional"
##
## [[151]]
## [1] "hard" "soft"
##
## [[152]]
## [1] "high brightness" "low brightness"
##
## [[153]]
## [1] "high saturation" "low saturation"
##
## [[154]]
## [1] "immature" "mature"
##
## [[155]]
## [1] "immobile" "mobile"
##
## [[156]]
## [1] "imperforate" "perforate"
##
## [[157]]
## [1] "impermeable" "permeable"
##
## [[158]]
## [1] "inflated"   "uninflated"
##
## [[159]]
## [1] "insensitive toward" "sensitive toward"
##
## [[160]]
## [1] "insoluble in" "soluble in"
##
## [[161]]
## [1] "insufficient" "sufficient"
##
## [[162]]
## [1] "insusceptible toward" "susceptible toward"
##
## [[163]]
## [1] "invariant" "variant"
##
## [[164]]
## [1] "invariant color" "variant color"
##
## [[165]]
## [1] "invariant shape" "variant shape"
##
## [[166]]
## [1] "invariant temperature" "variant temperature"
##
## [[167]]
## [1] "irregular duration" "regular duration"
##
```

```
## [[168]]
## [1] "irregular sleep pattern" "regular sleep pattern"
##
## [[169]]
## [1] "irregular spatial pattern" "regular spatial pattern"
##
## [[170]]
## [1] "lumenized"    "unlumenized"
##
## [[171]]
## [1] "magnetic"      "non-magnetic"
##
## [[172]]
## [1] "male fertile" "male sterile"
##
## [[173]]
## [1] "male semi-fertile" "male semi-sterile"
##
## [[174]]
## [1] "necessary (continuant)"    "unnecessary (continuant)"
##
## [[175]]
## [1] "necessary (occurrent)"    "unnecessary (occurrent)"
##
## [[176]]
## [1] "negative charge" "positive charge"
##
## [[177]]
## [1] "non-porous" "porous"
##
## [[178]]
## [1] "non-potable" "potable"
##
## [[179]]
## [1] "non-progressive" "progressive"
##
## [[180]]
## [1] "old"    "young"
##
## [[181]]
## [1] "opaque"       "transparent"
##
## [[182]]
## [1] "ornamented"    "unornamented"
##
## [[183]]
## [1] "pigmented"    "unpigmented"
##
## [[184]]
## [1] "responsive to"    "unresponsive to"
##
## [[185]]
## [1] "rough"   "smooth"
##
```

```
## [[186]]
## [1] "serrated"    "unserrated"
##
## [[187]]
## [1] "stratified"    "unstratified"
##
## [[188]]
## [1] "structured"    "unstructured"
```

```r
print("In mine only:")
```

```
## [1] "In mine only:"
```

```r
print(pato.mine.only)
```

```
## [[1]]
## [1] "abaxial to" "axial to"
##
## [[2]]
## [1] "acyclic cyclicity" "cyclic cyclicity"
##
## [[3]]
## [1] "alobate" "lobate"
##
## [[4]]
## [1] "anaplastic" "aplastic"
##
## [[5]]
## [1] "aneuploid" "euploid"
##
## [[6]]
## [1] "asymmetrical" "symmetrical"
##
## [[7]]
## [1] "atonicity" "tonicity"
##
## [[8]]
## [1] "complex" "simple"
##
## [[9]]
## [1] "curved dorsal"  "curved ventral"
##
## [[10]]
## [1] "dark blue"  "light blue"
##
## [[11]]
## [1] "dark brown"  "light brown"
##
## [[12]]
## [1] "dark cyan"  "light cyan"
##
## [[13]]
## [1] "dark green"  "light green"
```

```
##
## [[14]]
## [1] "dark grey"  "light grey"
##
## [[15]]
## [1] "dark magenta"  "light magenta"
##
## [[16]]
## [1] "dark orange"  "light orange"
##
## [[17]]
## [1] "dark phase"  "light phase"
##
## [[18]]
## [1] "dark purple"  "light purple"
##
## [[19]]
## [1] "dark red"  "light red"
##
## [[20]]
## [1] "dark red brown"  "light red brown"
##
## [[21]]
## [1] "dark violet"  "light violet"
##
## [[22]]
## [1] "dark yellow"  "light yellow"
##
## [[23]]
## [1] "dark yellow brown"  "light yellow brown"
##
## [[24]]
## [1] "decreased linear velocity" "increased linear velocity"
##
## [[25]]
## [1] "decreased propagation velocity" "increased propagation velocity"
##
## [[26]]
## [1] "decreased proportionality to" "increased proportionality to"
##
## [[27]]
## [1] "decreased spatial extent of a process"
## [2] "increased spatial extent of a process"
##
## [[28]]
## [1] "desaturated blue" "saturated blue"
##
## [[29]]
## [1] "desaturated brown" "saturated brown"
##
## [[30]]
## [1] "desaturated cyan" "saturated cyan"
##
## [[31]]
```

```
## [1] "desaturated green" "saturated green"
##
## [[32]]
## [1] "desaturated magenta" "saturated magenta"
##
## [[33]]
## [1] "desaturated orange" "saturated orange"
##
## [[34]]
## [1] "desaturated purple" "saturated purple"
##
## [[35]]
## [1] "desaturated red" "saturated red"
##
## [[36]]
## [1] "desaturated violet" "saturated violet"
##
## [[37]]
## [1] "desaturated yellow" "saturated yellow"
##
## [[38]]
## [1] "disposition" "position"
##
## [[39]]
## [1] "dorsal orientation"  "ventral orientation"
##
## [[40]]
## [1] "dorsal to"  "ventral to"
##
## [[41]]
## [1] "dystonicity" "tonicity"
##
## [[42]]
## [1] "glutinous"     "non-glutinous"
##
## [[43]]
## [1] "ictal"     "non-ictal"
##
## [[44]]
## [1] "immobile relative to" "mobile relative to"
##
## [[45]]
## [1] "nonpolar polarity" "polar polarity"
##
## [[46]]
## [1] "obsolete backcross fertile" "obsolete backcross sterile"
##
## [[47]]
## [1] "obsolete decreased action potential" "obsolete increased action potential"
##
## [[48]]
## [1] "obsolete decreased consumption" "obsolete increased consumption"
##
## [[49]]
```

```
## [1] "obsolete decreased percentage" "obsolete increased percentage"
##
## [[50]]
## [1] "obsolete decreased relaxation" "obsolete increased relaxation"
##
## [[51]]
## [1] "obsolete f1 fertile" "obsolete f1 sterile"
##
## [[52]]
## [1] "obsolete f2 fertile" "obsolete f2 sterile"
##
## [[53]]
## [1] "obsolete high" "obsolete low"
##
## [[54]]
## [1] "obsolete high enzyme function value" "obsolete low enzyme function value"
##
## [[55]]
## [1] "obsolete high percentage" "obsolete low percentage"
##
## [[56]]
## [1] "obsolete high yield" "obsolete low yield"
##
## [[57]]
## [1] "obsolete intercross fertile" "obsolete intercross sterile"
##
## [[58]]
## [1] "obsolete negative regulation" "obsolete positive regulation"
##
## [[59]]
## [1] "splayed dorsal"  "splayed ventral"
```

```
print("In 'current' only:")
```

```
## [1] "In 'current' only:"
```

```
print(current.only)
```

```
## [[1]]
## [1] "anterodistal orientation" "anterodorsal orientation"
##
## [[2]]
## [1] "auxotrophic"  "prototrophic"
##
## [[3]]
## [1] "behavioural active"   "behavioural inactive"
##
## [[4]]
## [1] "biconcave" "biconvex"
##
## [[5]]
## [1] "concave" "convex"
##
```

```
## [[6]]
## [1] "concave 3-D shape" "convex 3-D shape"
##
## [[7]]
## [1] "distal orientation" "dorsal orientation"
##
## [[8]]
## [1] "distal to" "dorsal to"
##
## [[9]]
## [1] "east" "west"
##
## [[10]]
## [1] "fused with"   "unfused from"
##
## [[11]]
## [1] "has extra parts of type" "has fewer parts of type"
##
## [[12]]
## [1] "hydrophilic" "hydrophobic"
##
## [[13]]
## [1] "hydrophilicity" "hydrophobicity"
##
## [[14]]
## [1] "hyperelliptic" "hypoelliptic"
##
## [[15]]
## [1] "hyperplastic" "hypoplastic"
##
## [[16]]
## [1] "hyperresponsive to" "hyporesponsive to"
##
## [[17]]
## [1] "hypertrophic" "hypotrophic"
##
## [[18]]
## [1] "hypertrophic growth" "hypotrophic growth"
##
## [[19]]
## [1] "increased virulence" "reduced virulence"
##
## [[20]]
## [1] "large (value)" "small (value)"
##
## [[21]]
## [1] "north" "south"
##
## [[22]]
## [1] "paedomorphic growth" "peramorphic growth"
##
## [[23]]
## [1] "photoinsensitive" "photosensitive"
##
```

```
## [[24]]
## [1] "posterodistal orientation" "posterodorsal orientation"
##
## [[25]]
## [1] "recurrent" "sporadic"
##
## [[26]]
## [1] "relative large (value)" "relative small (value)"
##
## [[27]]
## [1] "resistance to"          "susceptibility toward"
```

Comparison to Chris et al.'s stuff

```
chris <- read.table("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-outputs,
colnames(chris) <- c("term01", "term02")
chris <- as_tibble(chris)
hp.mine <- read.table("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-outpu
colnames(hp.mine) <- c("term01", "term02")
hp.mine <- as_tibble(hp.mine)

# PICK IT UP FROM HERE--I NEED TO RERUN MY STUFF ON HP, CLEARLY
chris.as.vector <- as.vector(paste(chris$term01, chris$term02))
hp.mine.as.vector <- as.vector(paste(hp.mine$term01, hp.mine$term02))
x <- list(Mine = hp.mine.as.vector, Chris = chris.as.vector)
ggVennDiagram(x)
```