Whissell's Dictionary of Affect in Language
Technical Manual and User's Guide
Cynthia Whissell
Laurentian University

The new Dictionary of Affect in Language (called DAL or Dictionary for short) is an instrument designed to measure the emotional meaning of words and texts. It does this by comparing individual words to a word list of 8742 words which have been rated by people for their activation, evaluation, and imagery.

Dictionary Construction

WORD SELECTION

Words included in the DAL were selected in an ecologically valid manner. There were three steps involved in the selection.

STEP 1: The Kucera and Francis 1969 corpus of 1,000,000 words was sampled from print media in the early 1960's. Words from this corpus with frequencies greater than 10 which also appeared in more than one subsample were included in the DAL list. This insured that the starting words in the set would not be rare ones, or ones specific to one type of print source. Proper names were removed from the sample.

STEP 2: The word set was then compared to four text samples generated by individuals rather than media. It was also compared to a large sample from juvenile literature. Unique words found in these sources were added to the list.

All of the samples employed at this step had been gathered by researchers at Laurentian University:

1.   Students' retelling of a story, 16309 words (source: Terri-Lynn Dittburner, Dr. M. Persinger)
2.   Interviews on the topic of abuse, 6085 words (source: Carolyn Djaferis)
3.   Adolescents' descriptions of their emotions, 15929 words (source: Louise Wood)
4.   University students' essays, 14807 words (source: Katie Lemega)
5.   Juvenile fiction of the 50's, 60's, 70's, 80's, and 90's, 82865 words (source: Michael Dewson and Laurie Steven)

STEP 3: The DAL list which contained approximately 8700 words at the end of step 2 was tested on 16 new, blindly selected, samples. It was also tested on a corpus of 350,000 words of English text collected by Whissell from many sources. The DAL demonstrated a hit rate or matching rate of approximately 90%.
The hit rate of 90% meant that one would expect NINE OUT OF TEN WORDS IN MOST ENGLISH TEXTS to be matched by the DAL.
The 90% hit rate compares very favourably with that of the original DAL (15%-25%).

WORD RATING

Data Collection Period
Data were collected in the latter half of the 1990's.

Volunteers
Volunteers for the rating task were mostly university students. Both men and women participated. A very small number of volunteers (less than 5%) were paid. Many volunteers received experimental participation credits. A total of over 200 volunteers was involved.

Rating Dimensions
The words of the DAL list were rated along the dimensions of PLEASANTNESS, ACTIVATION, and IMAGERY. In each case the scale used was a three-point scale.

(1) Unpleasant	(2) In between	(3) Pleasant
(1) Passive	(2) In between	(3) Active
(1) Hard to imaging	(2) In between	(3) Easy to imagine

Method
Roughly 50% of the ratings for Pleasantness and Activation were gathered using a computer-administered task. The remaining 50% and all ratings for Imagery were gathered in a paper and pencil task.
Different volunteers rated different numbers of words, and some rated words along more than one dimension. Occasionally volunteers returned to be retested on a second set of words.
Most volunteers were able to make about 200 rating judgments before showing signs of boredom, inattention, or fatigue (the task was self-paced and could be terminated).
The DATA used to create the DAL involved MORE THAN 186,000 DIFFERENT RATING JUDGMENTS ABOUT WORDS. Each word was rated for Activation and Pleasantness an average of 8 times and for Imagery 5 times.

THE 8742 WORDS IN THE DAL LIST

Descriptive Statistics

	Mean	SD	Skewness	Kurtosis
Pleasantness	1.84	.44	.27	-.37
Activation	1.85	.39	.39	-.29
Imagery	1.94	.63	.18	-1.18

Correlations

Correlations among rating dimensions were extremely weak, and shared variance among scales was minimal (<1%).

	Pleasantness	Activation	Imagery
Pleasantness	1.0	.10	.06
Activation	.10	1.0	.05
Imagery	.06	.05	1.0

POPULATION VALUES FOR TEXT SAMPLES

Not all DAL words are used with equal frequency. As a result, population parameters for text samples are somewhat different than DAL parameters for individual words. The broad corpus of 350,000 words of English text collected by Whissell was used as the normative text corpus. Descriptive statistics were calculated for this corpus.

	Mean	**SD**
Pleasantness	1.85	.36
Activation	1.67	.36
Imagery	1.54	.63

In addition, the text corpus was evaluated for the appearance of certain classes of extreme words.

IN TEXTS,

Very Pleasant Words (in the top 10% of all rated words) occurred .060 of the time(sd=.24).
Fun or Cheerful Words (in the top 25% of all rated words for Pleasantness and in the top 25% for Activation) occurred .049 of the time (sd=.215).
Very Active Words (top 10% of all rated words) occurred .042 of the time (sd=.201).
Nasty Words (top 25% for Activation, bottom 25% for Pleasantness) occurred .032 of the time (sd=.175).
Very Unpleasant words (bottom 10%) occurred .038 of the time (sd=.191).
Very Sad words (bottom 10% for both Activation and Pleasantness) occurred .052 of the time (sd=.221).
Very Passive words (bottom 10% for Activation) occurred .195 of the time (sd=.396).
Nice or Soft words (top 25% for Pleasantness, bottom 25% for Activation) occurred .046 of the time (sd=.21).
Highly Imaged words (top 10%) occurred .045 of the time (sd=.209) and
Poorly Imaged words (bottom 10%) .399 of the time (sd=.490).

Differences in word usage patterns explain the fact that DAL list means for Activation and Imagery are higher than text means. The more frequent use of more passive and more poorly imaged words brings the text mean down lower than the list mean (1.84 to 1.67 for Activation, 1.93 to 1.53 for Imagery). There were weak negative correlations between the frequency of use for individual words and Activation (-.06) and Imagery (-.08).

RELIABILITY AND VALIDITY

Reliability Information

Full DAL ratings for Pleasantness and Activation were correlated with subsamples of ratings.

Subsample	N	Pleasantness r	Activation r
1	2746	.68	.55
2	2118	.72	.61
3	8740	.87	.86
4	8466	.81	.69

Validity Information

The new and original DAL share 2165 words. The correlation for Activation of these words was .45, that for Pleasantness .71.
The new DAL shares 1556 words with the Children's DAL. The correlation for Activation was .51, that for Pleasantness .62.
The new DAL shares 1703 words with a list of nouns which had been rated for imagery. The correlation was .47.

POSSIBLE USES OF THE DICTIONARY

The DAL could be used to:

1.   Select words for learning, memory, and cognitive experiments.
2.   Score texts of many different kinds including descriptions of subjective feelings, essays, newspaper reports, poetry, literature.
3.   Score responses in a free-association task, or other projective task.
4.   Score memory for stories for emotional memory.
5.   Score song lyrics for emotional content.

LANGUAGE is not the most frequent human behaviour (heartbeat and breathing exceed it in number of occurrences per lifetime!) but it is VERY COMMON. **ANY sample of language gathered in ANY manner can be scored for its emotionality and imagery using the Dictionary of Affect in Language**.

Oscar Wilde's poem ballad of Reading Goal was written in 1896, and was dedicated (in memoriam) to a member of the Royal Horse Guards who died there (Reading Goal was a prison).

The poem ends with the lines:
And all men kill the thing they love,
By all let this be heard,

Some do it with a bitter look,
Some with a flattering word,
The coward does it with a kiss,
The brave man with a sword!

HDC Program Output for The Ballad of Reading Goal

Mean Pleasantness 1.80
Mean Activation 1.65
Mean Imagery 1.60
-Known Adult Words 3605
Child Pleasantness 4.22
Child Activation 3.99
-Known Child Words 1019
Total words 4188
Mean Frequency 2479.2
Sentence Length 32.98
Sentences (periods) 26
Exclamation Marks 26
Question Marks 7
% Nice 2.75
% Pleasant 5.05
% Fun 3.74
% Active 3.08
% Nasty 3.52
% Unpleasant 6.05
% Sad 4.94
% Passive 19.39
% High Imagery 6.24
% Low Imagery 45.8

From the data here it is obvious that the Ballad is somewhat less pleasant (1.80 vs 1.85), better imaged (1.60 vs 1.53), and marginally less active (1.65 vs 1.67) than the normative text corpus. 3605 of the 4188 words in the poem were matched, giving the DAL a hit rate of 86%. Values for the emotional dimensions and categories could all be compared to the text corpus using z tests and the means and standard deviations reported earlier for the corpus.

For example, the standardized score for pleasantness would be calculated as

z = (Sample mean - Corpus mean) / Corpus sd
or 1.80-1.85/.36
or .14

The z test testing the null of equality of sample and corpus means would involve additionally dividing the standard error by the square root of the number of scored words, namely

(1.80 - 1.85) / .36 / sqr(3605)
or 8.4 (p<.01)

USING THE HDC SCORING PROGRAM

1.   Download the Whissell Dictionary of Affect from the HDC pages.
2.   Run setup.
3.   Run the .exe file or double click the icon.
4.   Clear the screen
5.   Type in any desired text OR
6.   Drag in any highlighted material from a WORD or WORD PERFECT program OR
7.   Cut and Paste material from such programs OR
8.   Name a file which is an ASCII or DOS file for analysis.
9.   To analyze more than 1 samples at a time, separate each sample by entering aaa between samples

You will receive two output files with a name chosen by yourself, one is a summary file and will look like the Reading Goal example, the other is a data file and will list each and every scored word with all information for that particular word.