Computational Intelligence 696i

Language

Lecture 7

Sandiway Fong

Administriva

- Reminder:
  - Homework 2 due next Tuesday (*midnight*)

Last Time

- **WordNet** (Miller @ Princeton University)
    - *word-based semantic network*
    - **Logical Metonymy**
        - *gap-filling that involves computing some telic role*
        - *John enjoyed* X
            - X some non-eventive NP, e.g. *the wine*
        - *John enjoyed* [Ving] X
            - *John enjoyed* [drinking] *the wine*
        - **Generative Lexicon** (GL) Model (*wine*: telic role: *drink*)
        - **WordNet structure**
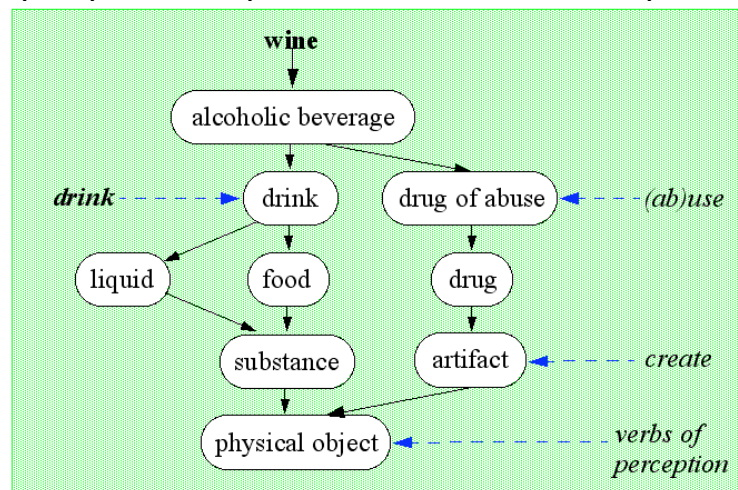
Last Time

- **WordNet** (Miller @ Princeton University)
  - some applications including:
    - concept identification/disambiguation
    - query expansion
    - (cross-linguistic) information retrieval
    - document classification
    - machine translation
    - Homework 2: GRE vocabulary
      - word matching exercise (*from a GRE prep book*)
- GRE as Turing Test
  - if a computer program can be written to do as well as humans on the GRE test, is the program **intelligent**?
  - **e-rater** from ETS (98% *accurate essay scorer*)

Today's Lecture

- Guest mini-lecture by
  - Massimo Piattelli-Palmarini on
 - *A powerful critique of semantic networks: Jerry Fodor's atomism*

Today's Lecture

but first...

- Take a look at an alternative to WordNet
  - WordNet: **handbuilt system**
  - COALS:
    - **the correlated occurrence analogue to lexical semantics**
    - (Rohde et al. 2004)
    - a instance of a vector-based statistical model for similarity
      - e.g., see also Latent Semantic Analysis (LSA)
      - Singular Valued Decomposition (SVD)
        - » sort by singular values, take top *k* and reduce the dimensionality of the co-occurrence matrix to rank *k*
    - based on weighted co-occurrence data from large corpora

COALS

- **Basic Idea:**
  - compute co-occurrence counts for (open class) words from a large corpora
  - corpora:
    - Usenet postings over 1 month
    - 9 million (distinct) articles
    - 1.2 billion word tokens
    - 2.1 million word types
      - 100,000th word occurred 98 times
  - co-occurrence counts
    - based on a ramped weighting system with window size 4
      - *excluding closed-class items*
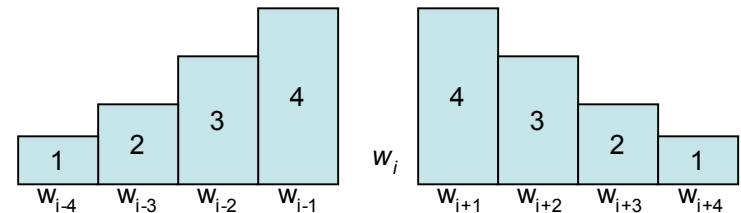
COALS

- Example:

Table 7

Step 3 of the COALS method: Negative values discarded and the positive values square rooted.

	a	as	chuck	could	how	if	much	wood	woodch.	would	,	.	?
a	0	0	0.120	0.093	0	0.291	0	0	0.310	0.262	0.291	0	0
as	0	0.175	0	0	0	0	0.364	0.320	0	0	0	0	0.365
chuck	0.120	0	0	0.306	0	0.146	0	0.177	0.220	0	0	0.297	0.175
could	0.093	0	0.306	0	0	0.182	0	0.149	0.221	0	0	0.263	0.151
how	0	0	0	0	0	0	0.438	0.265	0	0.263	0	0	0
if	0.291	0	0.146	0.182	0	0	0	0	0.291	0.076	0.372	0	0
much	0	0.364	0	0	0.438	0	0	0.358	0	0.136	0	0	0.268
wood	0	0.320	0.177	0.149	0.265	0	0.358	0	0	0.034	0	0.333	0.317
woodch.	0.310	0	0.220	0.221	0	0.291	0	0	0	0.221	0.291	0	0
would	0.262	0	0	0	0.263	0.076	0.136	0.034	0.221	0	0.246	0	0
,	0.291	0	0	0	0	0.372	0	0	0.291	0.246	0	0	0
.	0	0	0.297	0.263	0	0	0	0.333	0	0	0	0	0
?	0	0.365	0.175	0.151	0	0	0.268	0.317	0	0	0	0	0

COALS

- available online
- http://dlt4.mit.edu/~dr/COALS/similarity.php

COALS
the correlated occurrence analogue to lexical semantics

This interface allows you to compute pairwise similarity ratings of words or short phrases using the COALS vectors. You are limited to the 500,000 most-common words according to the corpus on which COALS is based.

Vector Type:
- ◯ COALS — Use 1,000-100,000 dimensions, 14,000 recommended
- ⦿ COALS-SVD — Use 10-1,500 dimensions, 800 recommended
- ◯ COALS-SVDB — Use 10-1,500 dimensions, 500 recommended

Dimensions: 800

The word pairs should be formatted with two words per line, separated by white space. In place of a word, you can use a short phrase or set of words in curly braces. Their vectors will be averaged before the similarity is computed.

Word Pairs:
```
DOG CAT
COW {FARM ANIMAL}
```

COALS

- on homework 2

Task: Match each word in the first column with its definition in the second column

accolade	**deviation**
aberrant	keen **insight**
abate	abolish
abscond	**lessen** in intensity
acumen	**sour** or bitter
acerbic	**depart** secretly
abscission	**build**ing up
accretion	renounce
abjure	removal
abrogate	praise

Task: Match each word in the first column with its definition in the second column
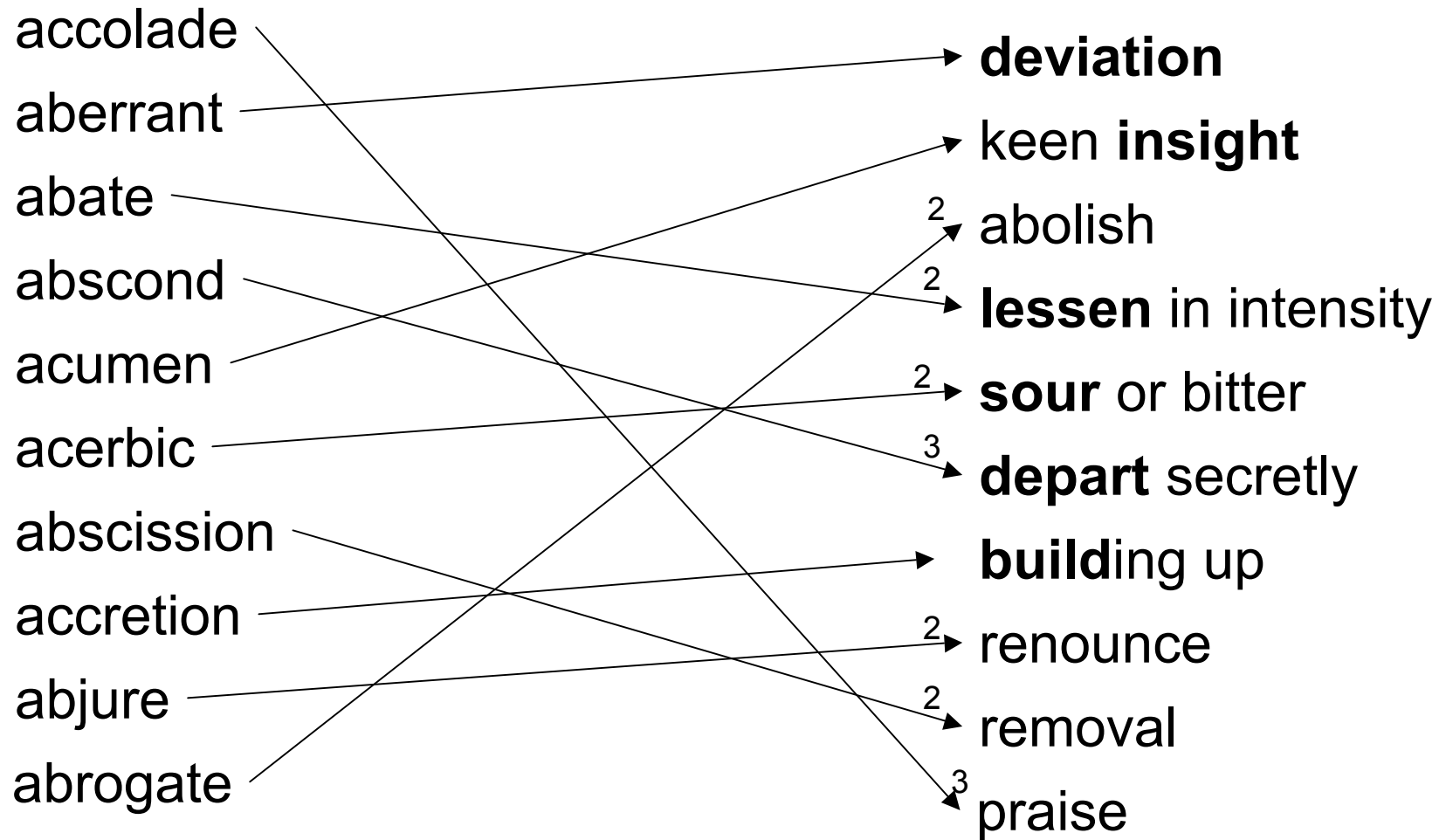
accolade

aberrant

abate

abscond

acumen

acerbic

abscission

accretion

abjure

abrogate

deviation

keen **insight**

[2] abolish

[2] **lessen** in intensity

[2] **sour** or bitter

[3] **depart** secretly

building up
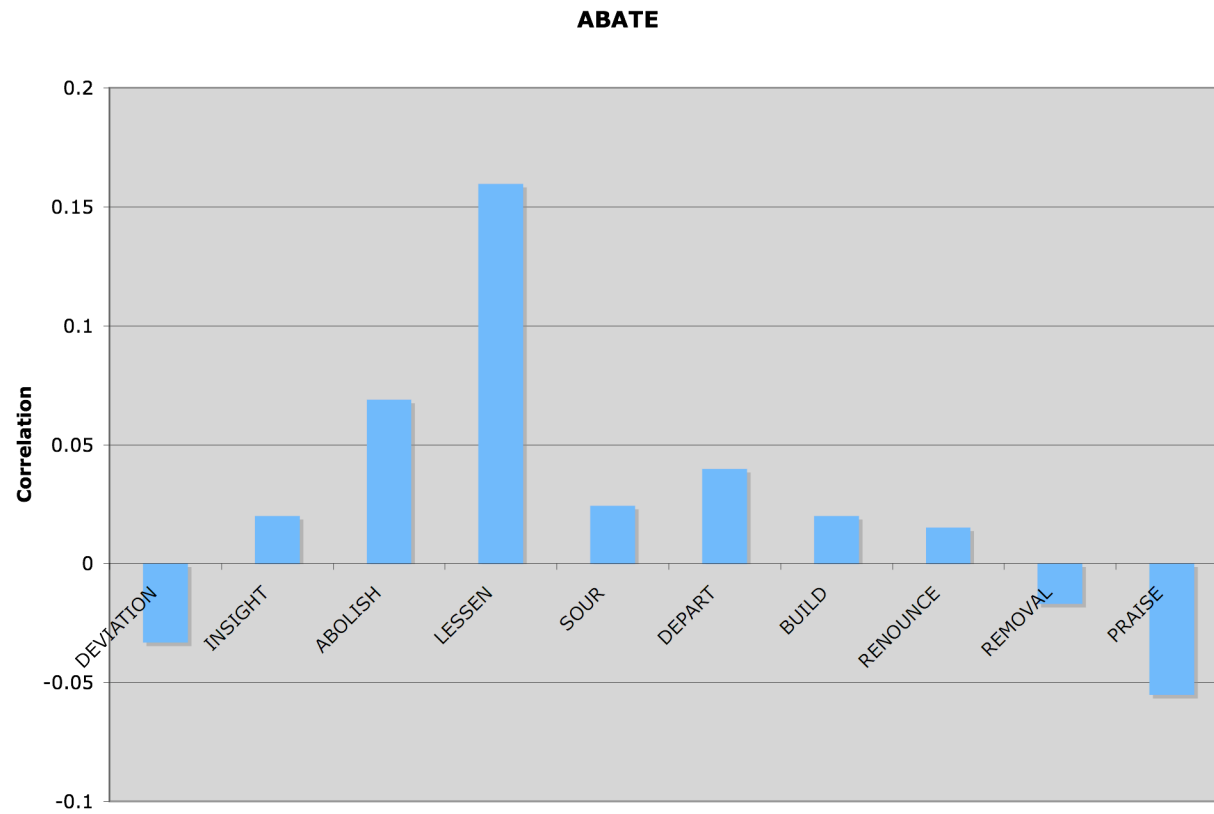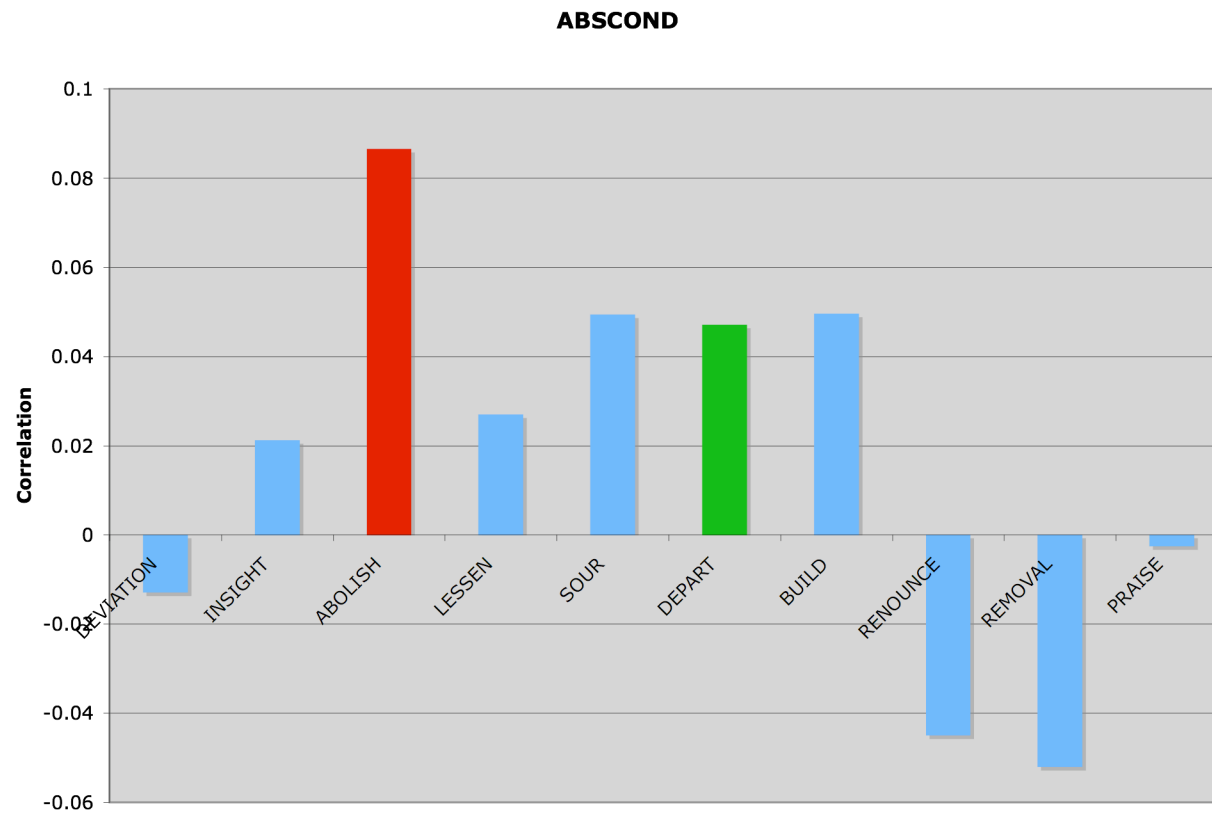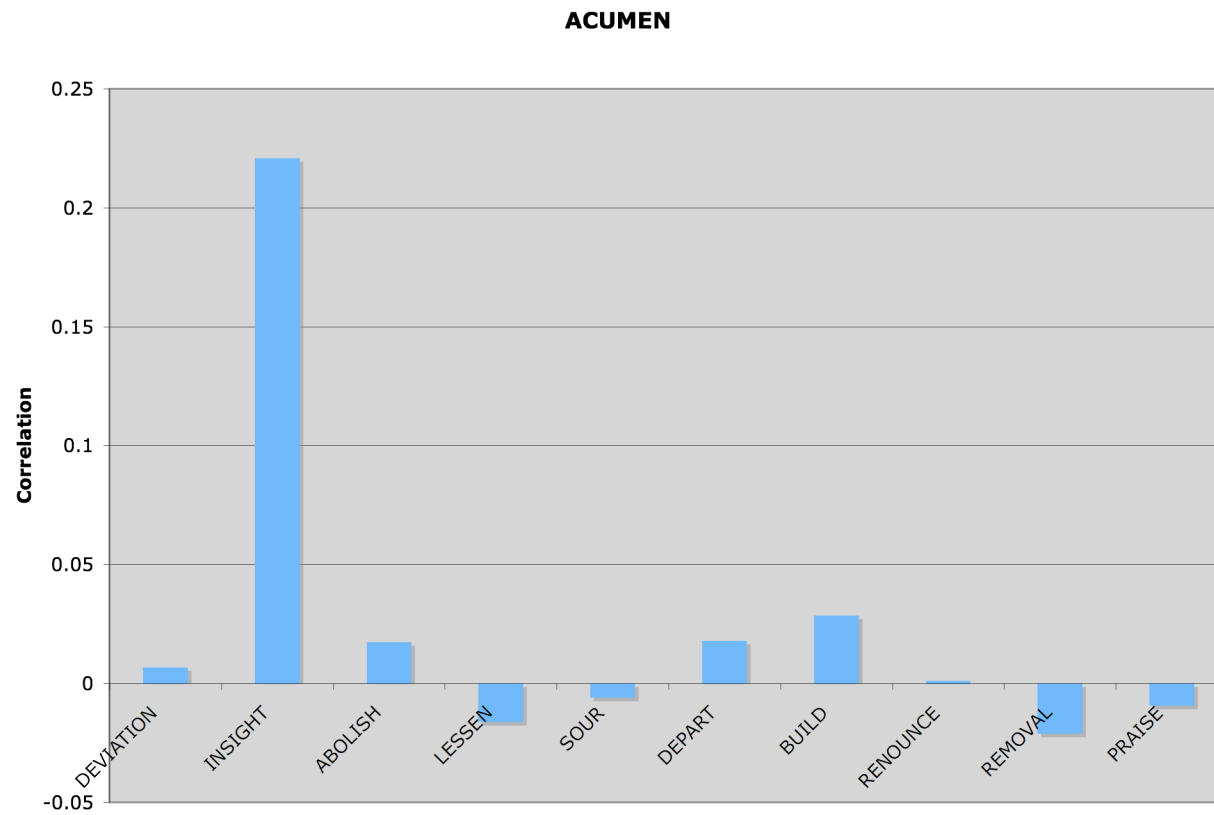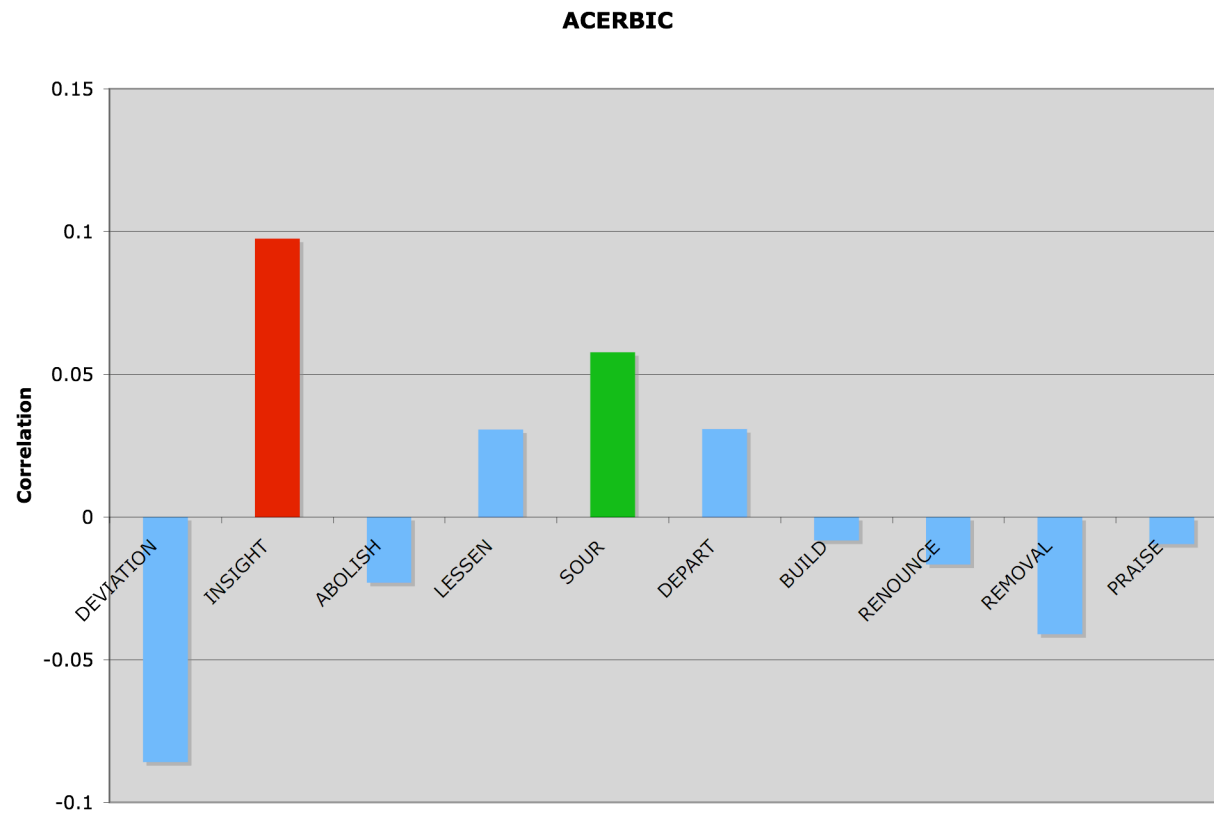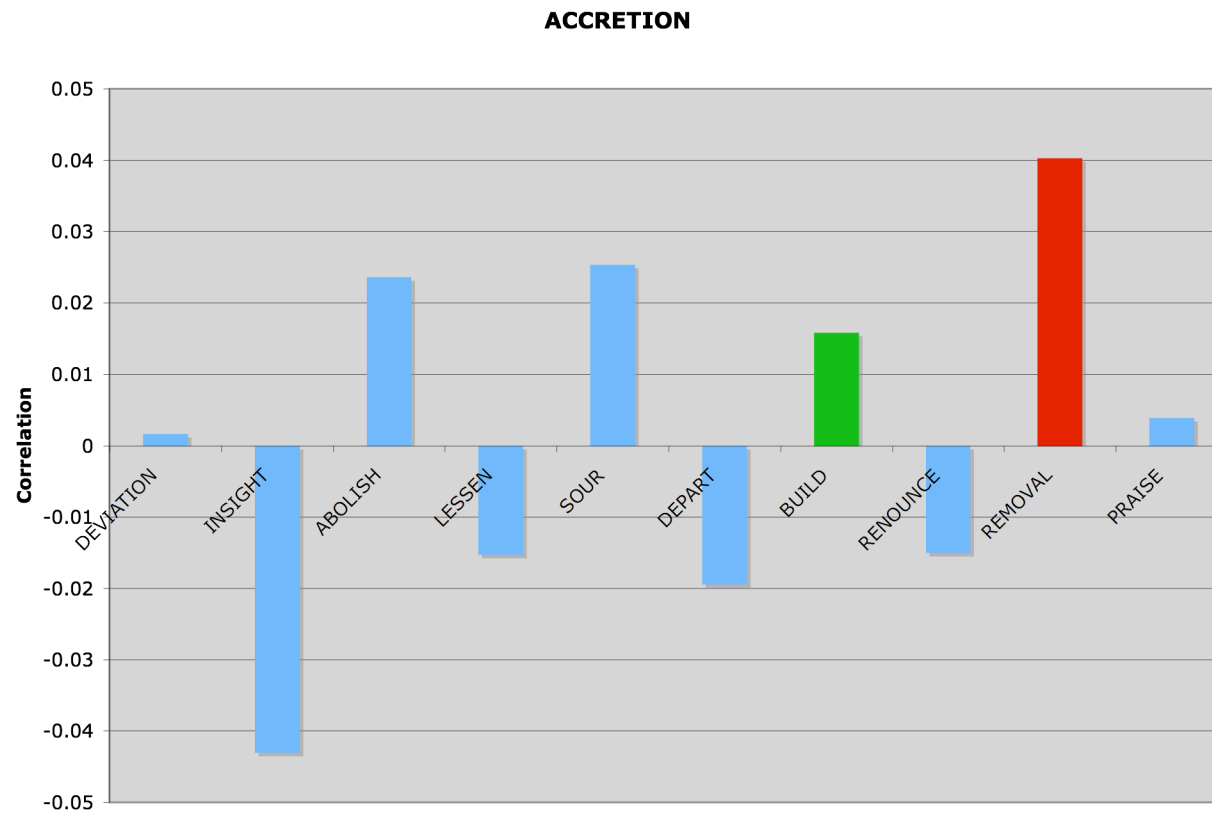
[2] renounce

[2] removal

[3] praise

COALS and the GRE

COALS and the GRE

COALS and the GRE

ABATE

COALS and the GRE

ABSCOND

COALS and the GRE
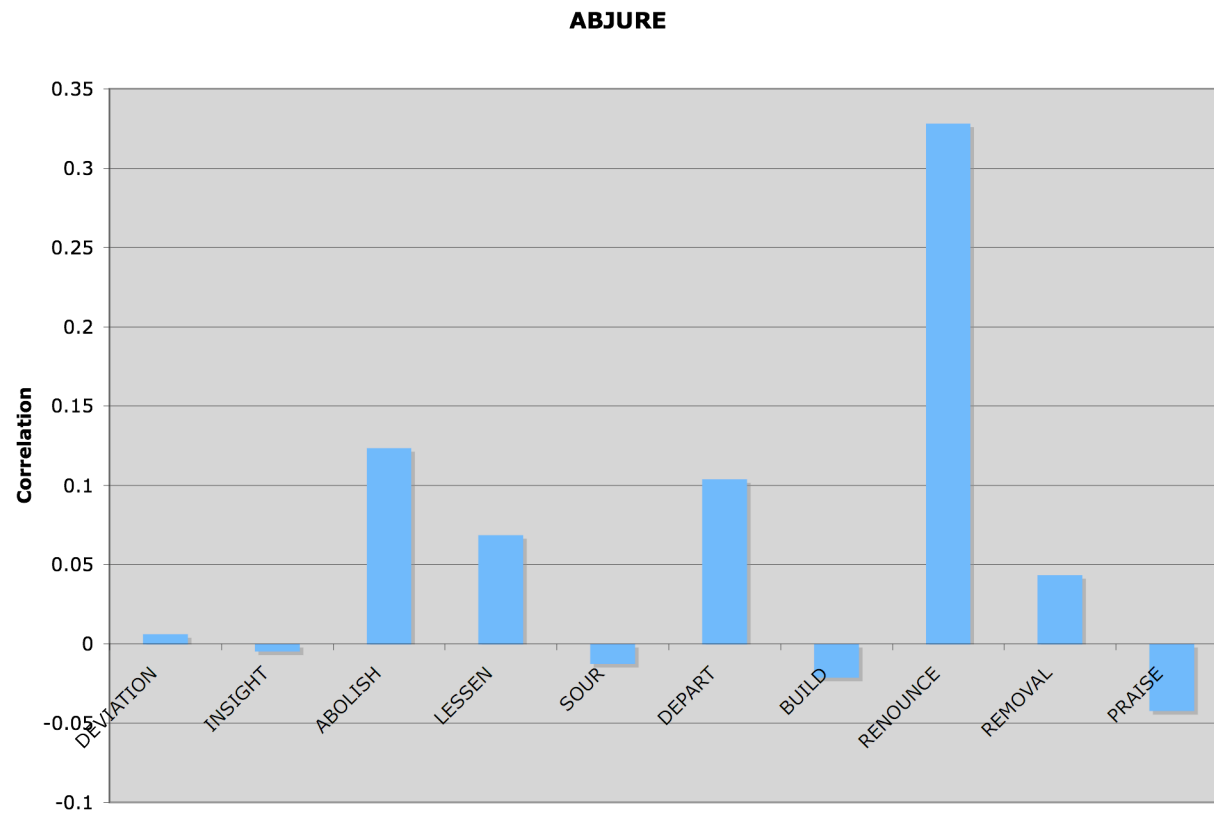
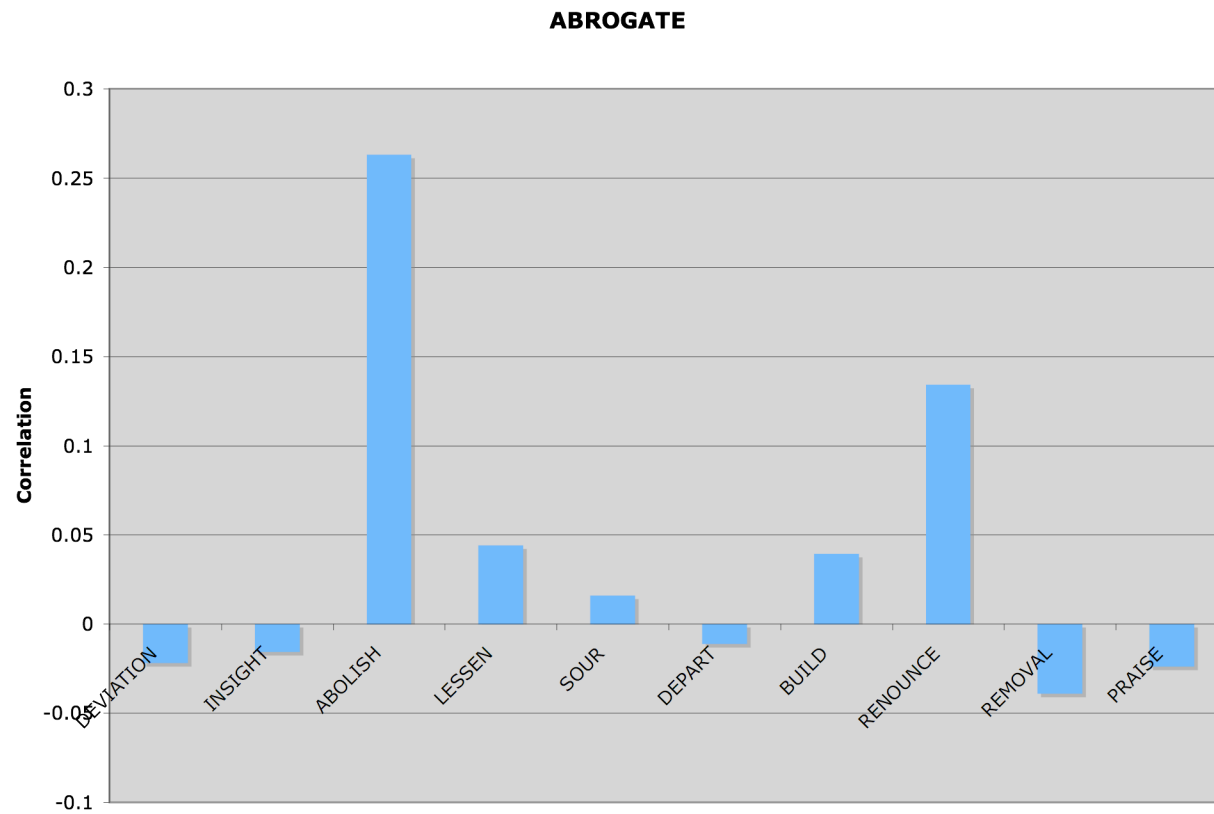# COALS and the GRE

COALS and the GRE

ACCRETION

COALS and the GRE

ABJURE

COALS and the GRE

ABROGATE

Task: Match each word in the first column with its definition in the second column

accolade

aberrant

abate

abscond

acumen

acerbic

abscission

accretion

abjure

abrogate

deviation

keen **insight**

abolish

lessen in intensity

sour or bitter

depart secretly

building up

renounce

removal

praise
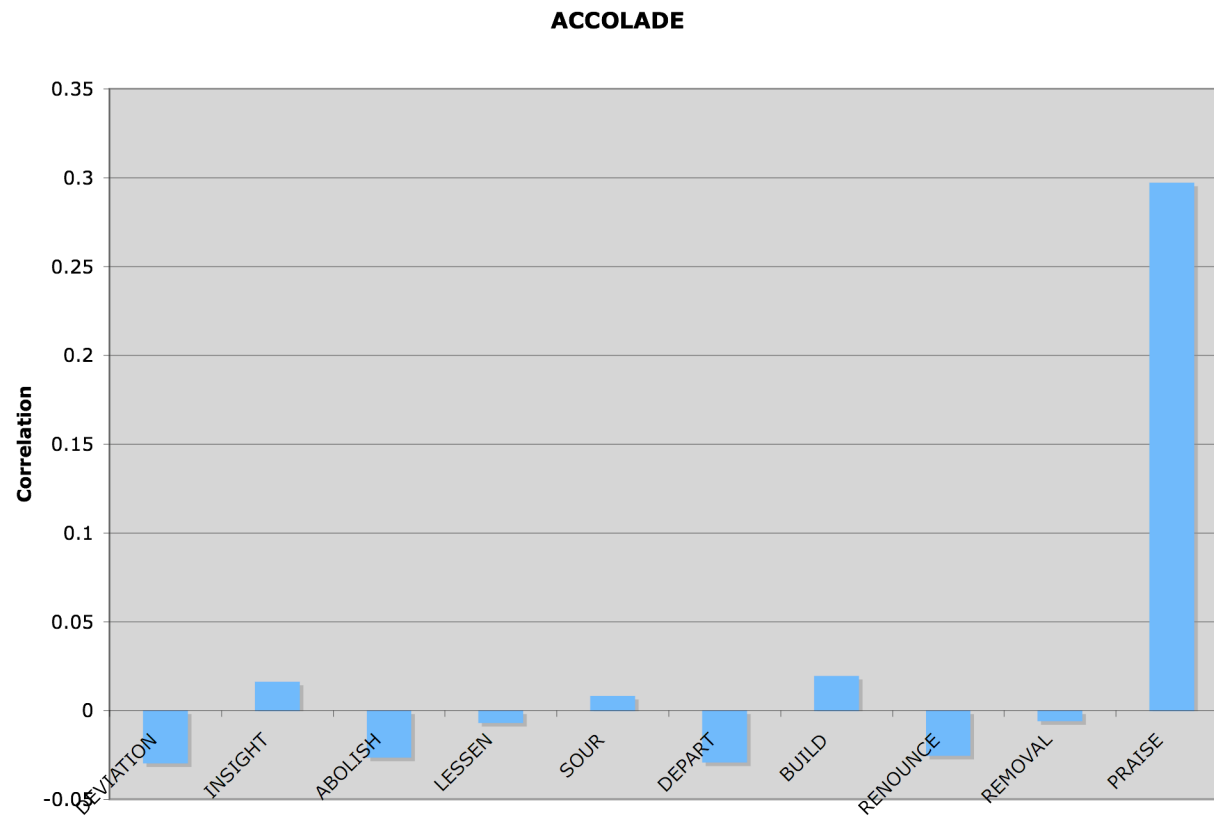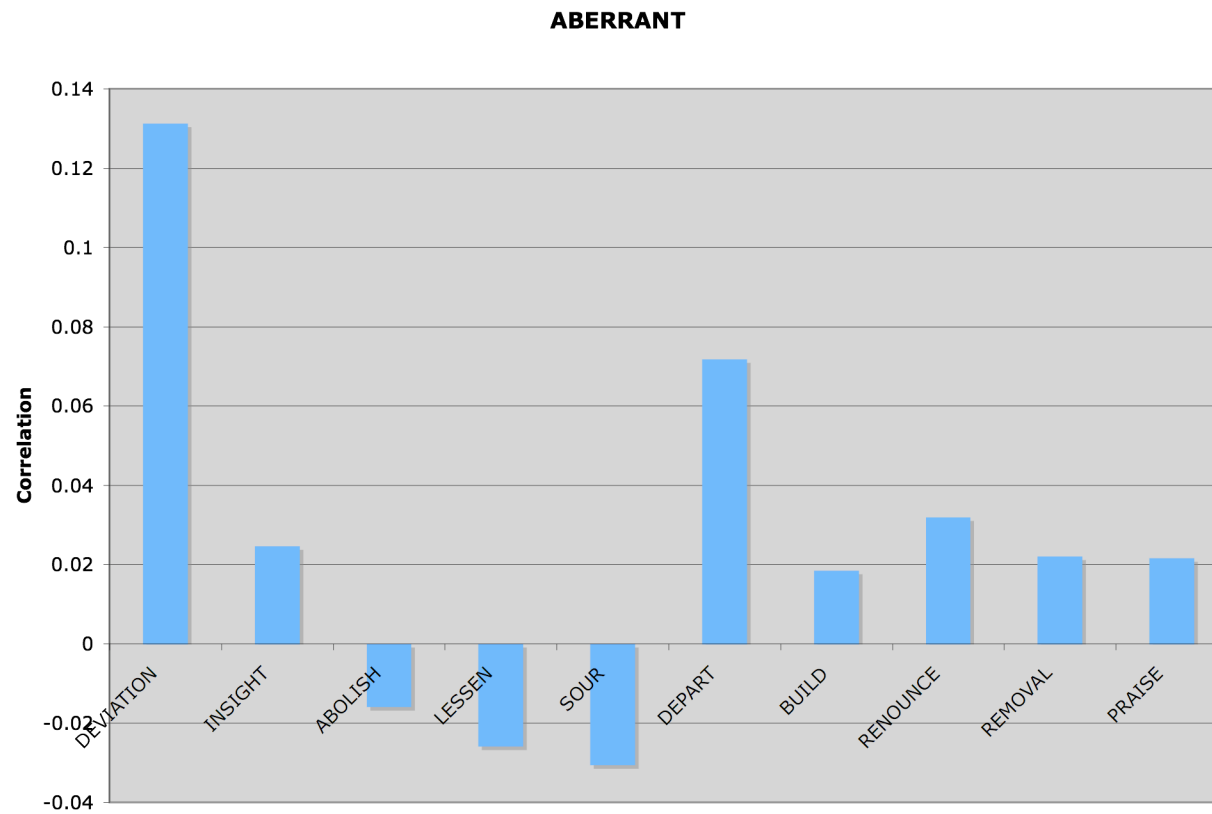
Heuristic: competing words, pick the strongest

accolade

aberrant

abate

abscond

acumen

acerbic

abscission

accretion

abjure

abrogate

7 out of 10

deviation

keen **insight**

abolish

lessen in intensity

sour or bitter

depart secretly

building up

renounce

removal

praise