PSIMETRICA: Tools and techniques for measuring phonological similarity

Shane T. Mueller

Indiana University

January 27, 2006

in collaboration with David E. Meyer, Adam Krawitz, Tom Gruenenfelder, and members of the Brain, Cognition, & Action Lab at Univ. of Michigan

In order to test hypotheses about a variable intrinsic to a word, the variable must be confounded with the word.

- For example: word frequency, word length, meaning, spelling, etc. can not be counter-balanced.
- Noted by Clark (1973) “The language-as-fixed-effect fallacy”
- Especially true for phonological similarity in verbal working memory tasks.

The Phonological-Loop Model

- Verbal working memory (VWM) involves a phonological representation that decays with time, but is refreshed with cyclic rehearsal.
- Recall accuracy should depend on:
  - Articulatory Duration of words
  - Phonological Similarity
- Problem: Phonological similarity is an intrinsic variable that cannot be counterbalanced.
- Question: How do people control phonological similarity in VWM tasks?

Conrad & Hull, 1964

"Whenever subsets of letters or digits, or indeed any other material which can be verbalized, are compared in memory experiments ....care must be taken to equate the subsets for acoustic confusability. The results otherwise will certainly be contaminated by this factor."

How do people heed Conrad & Hull?

"The mean frequency of occurrence of the short words was 16 and of the long words 16.2." (Caplan et al., 1992)

"These sets were approximately matched in frequency and imageability." (Cowan et al., 1997)

How do people heed Conrad & Hull?

"[Subjects] were asked to judge the similarity of sound of the words in each word pair, and to rate the similarity on a 1-5 scale." (Baddeley & Andrade, 1994)

"The words were matched for phonological similarity....Fifteen subjects rated each word pair on a 5-point Likert scale..." (Lovatt et al., 2000)

Vitz & Winkler, 1973

"Previous research has amply demonstrated, at least in perceptual and short-term memory tasks, that not all phonemes are equally confusable.... our interpretation is that relatively little of the variance in the rating of complete words is due to factors existing at a lower or more molecular level than the phoneme."

Don't Ask, Don't Tell

- Researchers DO care about controlling for aspects of verbal stimuli.
- Experimenters often ignore phonological similarity.
- Subjective ratings may be insufficient.
- Unlike frequency, imageability, etc., similarity norms are not widely available (N^2 problem)
- What is needed is an objective measure of phonological similarity.....

The PSIMETRICA Technique

- PSIMETRICA stands for Phonological SImilarity METRIC Analysis.
- PSIMETRICA measures dissimilarity or distance:
  - The measure of an item to itself is 0.
  - The measure of an item to its opposite is larger (closer to 1.)
- PSIMETRICA is symmetric:
  - The measure of item A to item B is equal to the measure of item B to item A.
- PSIMETRICA is multi-dimensional:
  - The "dissimilarity profile" allows for definition of multiple dimensions that map on to psychological concepts

Why isn't similarity a single number?

- Similarity may depend on the context.
- Context could mean: the task, the instructions, experience in that task. For example, compare:
  - Maiden, weapon, station, lemon, heaven, nation.
  - Cattle, kitchen, captive, county, contents, kitten.

Round the rugged rock the ragged rascal ran

To Helen. (E. A. Poe)

Helen, thy beauty is to me
Like those Nicean barks of yore
That gently, o'er a perfum'd sea,
The weary way-worn wanderer bore
To his own native shore.
On desperate seas long wont to roam,
Thy hyacinth hair, thy classic face,
Thy Naiad airs have brought me home
To the glory that was Greece,
And the grandeur that was Rome.
Lo, in yon brilliant window-niche
How statue-like I see thee stand,
The agate lamp within thy hand,
Ah! Psyche, from the regions which
Are Holy Land!

Rhyme

Alliteration

Assonance

Consonance

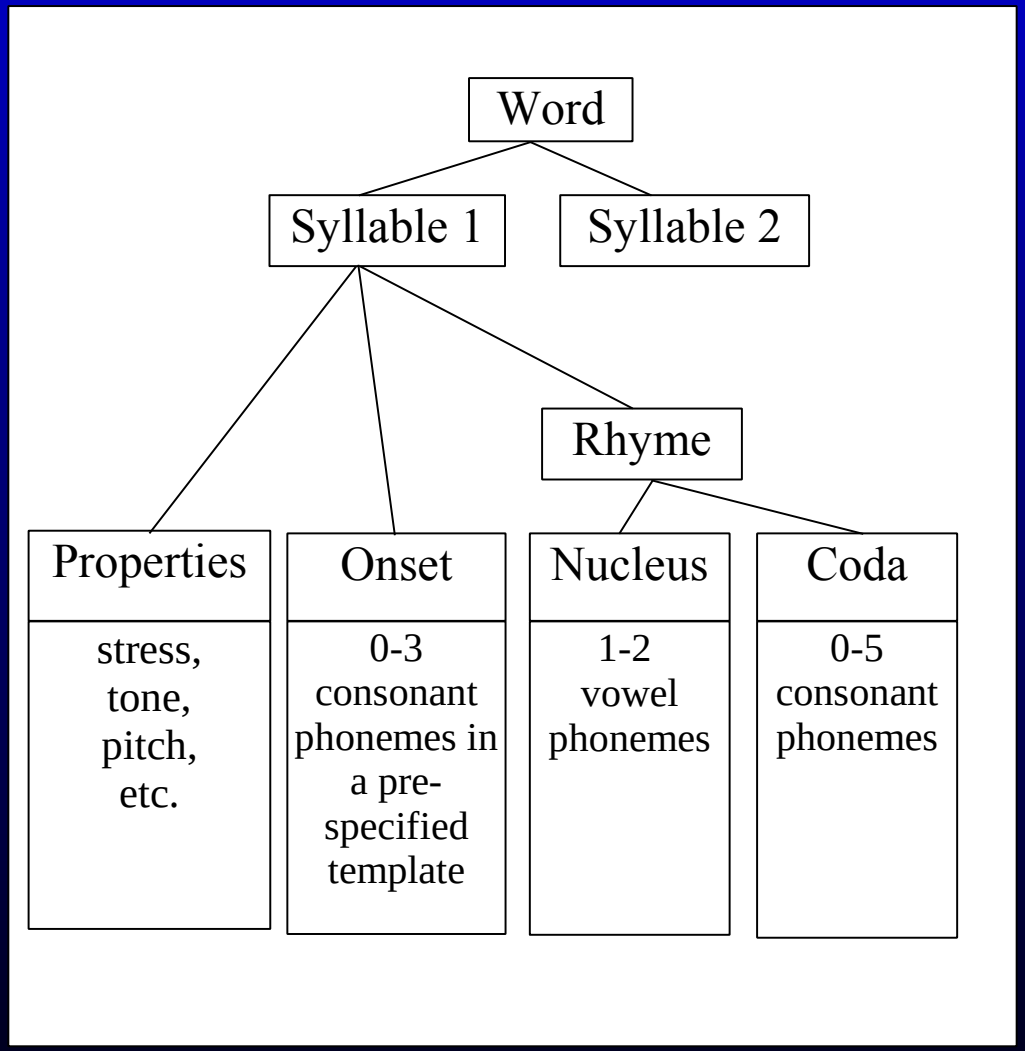
Stress Pattern

The Similarity Profile

- The similarity profile is a set of similarity measures that are psychologically interesting.
  - Syllable Onsets
  - Syllable Nuclei
  - Syllable Codas
  - Stress patterns
  - Initial Sounds (alliteration)
  - Cross-syllabic or cross-cluster (consonance)
  - Final nucleus-coda (rhyme)
  - Combinations of above

Hierarchical Representation of Words

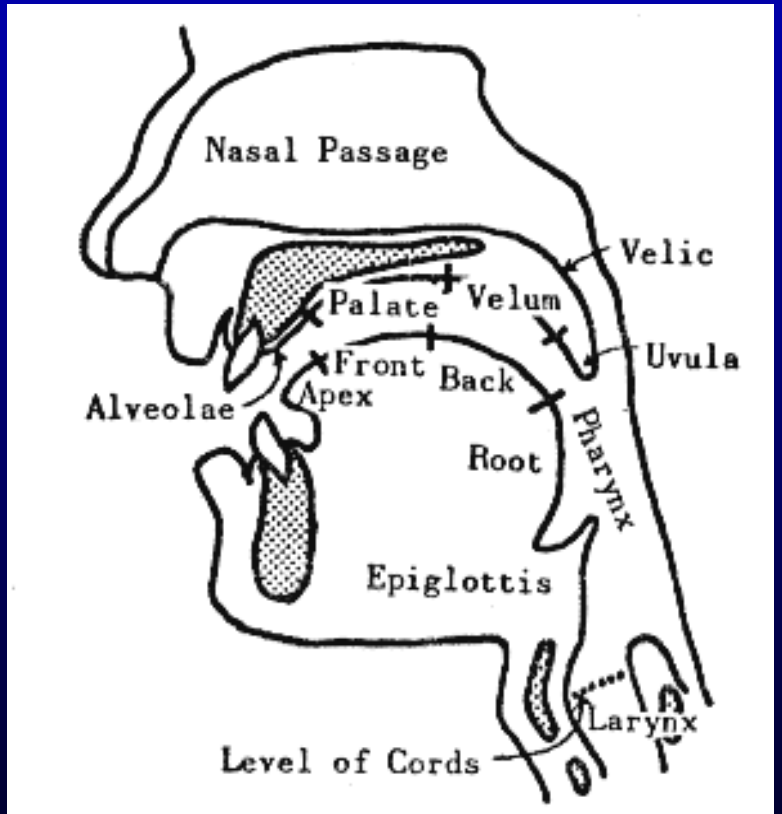
- Words are motor programs, represented in memory as hierarchical linguistic structures (as shown on the right.)
- Words contain syllables, which consist of properties and phoneme clusters, including onsets, nuclei, and codas.
- Phoneme clusters consist of phonemes.
- Phonemes consist of phonological-feature vectors.



Phonological Features

- Individual phonemes can be described by the state of the vocal articulators when a sound is being produced.

- The features used to represent phonemes in PSIMETRICA are vocalic, consonantal, high, back, low, anterior, coronal, round, tense, voice, continuant, nasal, strident.



Phonological-Feature Coding

	"Placemats"	"Amount"
	/plesmæts/ /(ples) (mæts)/ /((pl)(e)(s)) ((m)(æ)(ts))/	/əmaʊnt/ /(ə) (maʊnt)/ /((Ø)(ə)(Ø)) ((m)(aʊ)(nt))/
1	((-+----+-xx----) (++---++xx++--)) (+--------+xxxx) (-+----++xx-+-+)	(+--+-----xxxx)
2	(-+----+-xx+-+-) (+---+---+xxxx) ((-+----++xx----) (-+----++xx-+-+))	(-+----+-xx+-+-) (+--++---+xxxx) (+-++---+-xxxx) ((-+----++xx+-+-) (-+----++xx----))

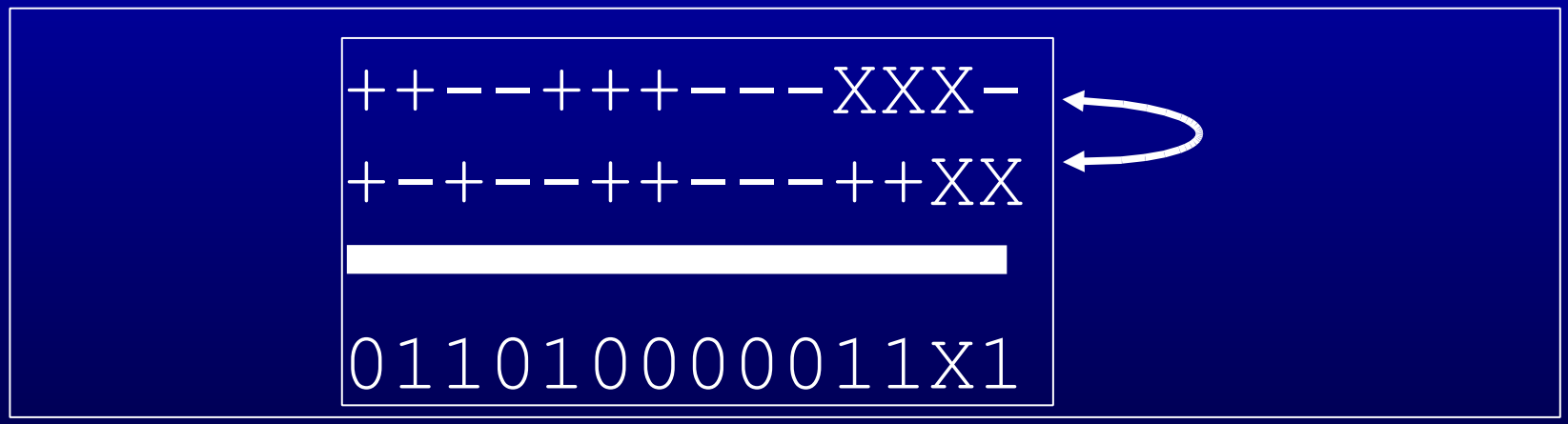
Alignment of Paired Words

- Different phoneme clusters use different alignment principles.
- Onsets are aligned according to a prespecified template.
- Nuclei are aligned by doubling shorter vowels.
- Codas are aligned to minimize dissimilarity.

Example Word	Phonemic Representation		First Syllables			Second Syllables		
			Onset	Nucleus	Coda	Onset	Nucleus	Coda
placemats	/((pl)(e)(s))	((m)(æ)(ts))/	/Øpl/	/e/	/s/	/ØmØ/	/ææ/	/Øts/
amount	/((Ø)(ə)(Ø))	((m)(aʊ)(nt))/	/ØØØ/	/ə/	/Ø/	/ØmØ/	/aʊ/	/ntØ/

Calculation of Phonological Dissimilarity between Phonemes

The dissimilarity between paired phonemes is calculated as follows:



6 features differ

12 of 13 features have values

Dissimilarity = 6/12 = 0.5 + baseline

Quantifying Phonological Dissimilarity for a Pair of Words

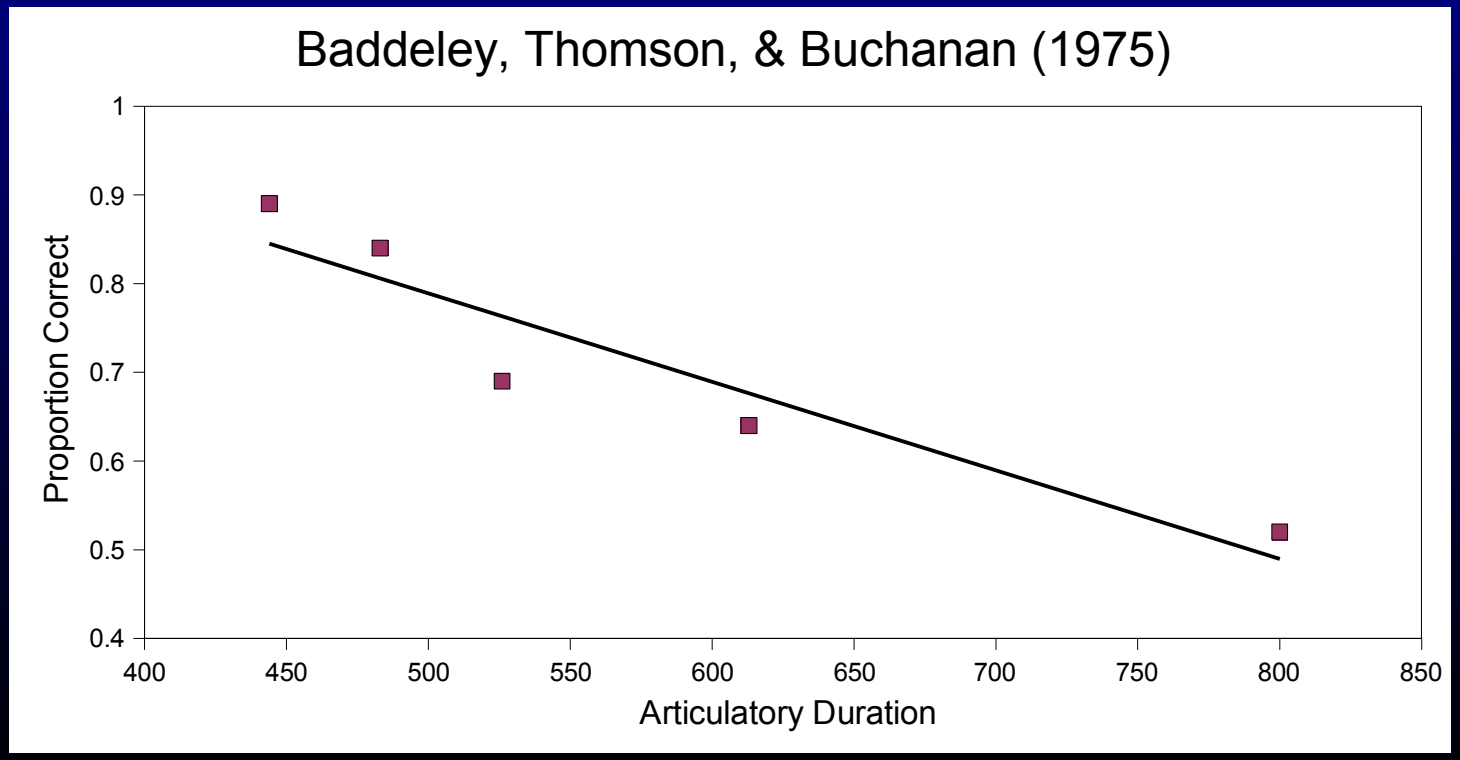
- Dimensions of the "**Dissimilarity profile**" are computed for corresponding syllables of paired words.

Phoneme Cluster	Example Words: Placemats	Example Words: Amount	Mean Phonological Dissimilarity: Phonemes	Mean Phonological Dissimilarity: Phoneme Clusters	Dissimilarity Profile
	First Syllables				
Onsets	(p l)	(Ø Ø)	0.37, 0.53	0.45	0.225
Nuclei	(e)	(ə)	0.22	0.2	
Codas	(s)	(Ø)	0.37	0.37	0.25
	Second Syllables				
Onsets	(m)	(m)	0	0	
Nuclei	(æ)	(a U)	0.11, 0.44	0.28	
Codas	(t s)	(n t)	0.37, 0, 0.37	0.25	0.31

Application: Immediate Serial Recall

Baddeley et al., 1975

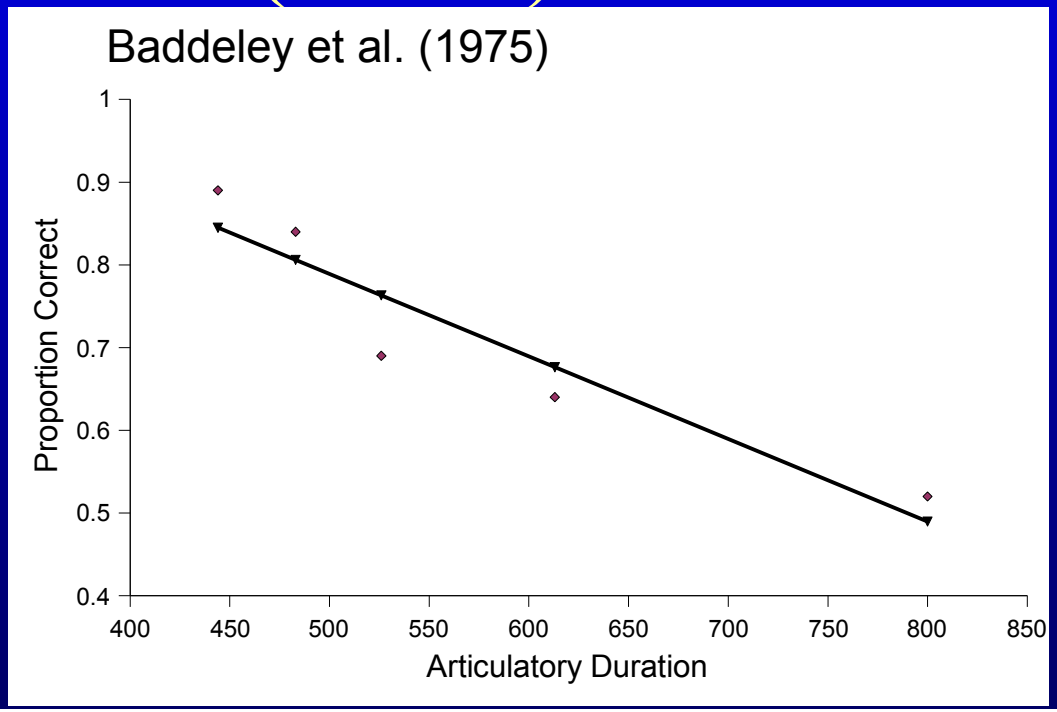
- Across 5 sets of words, probability correct recall is correlated with articulatory duration.
- The slope and intercept of these points has been used to make inferences about the phonological-loop model ("2 second decay")

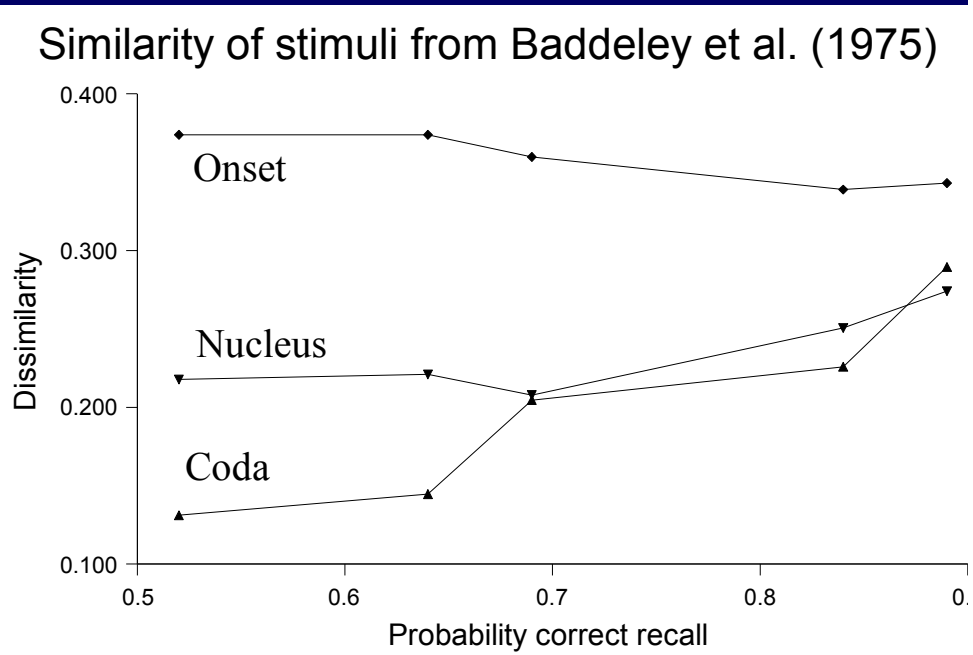


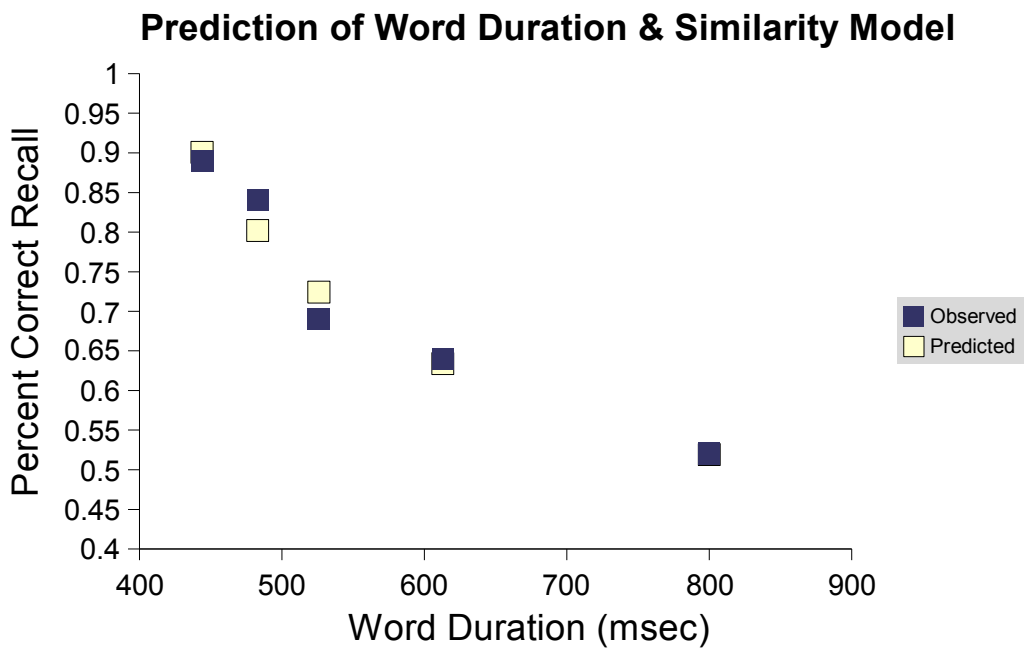
Baddeley et al. (1975)

Articulatory duration r = -.94
Onset similarity: r = -.94
Nucleus similarity: r = .84
Coda similarity: r = .94
Partial r's are high (-.67, .92, .49).

Systematic confounding of similarity to artic. duration.

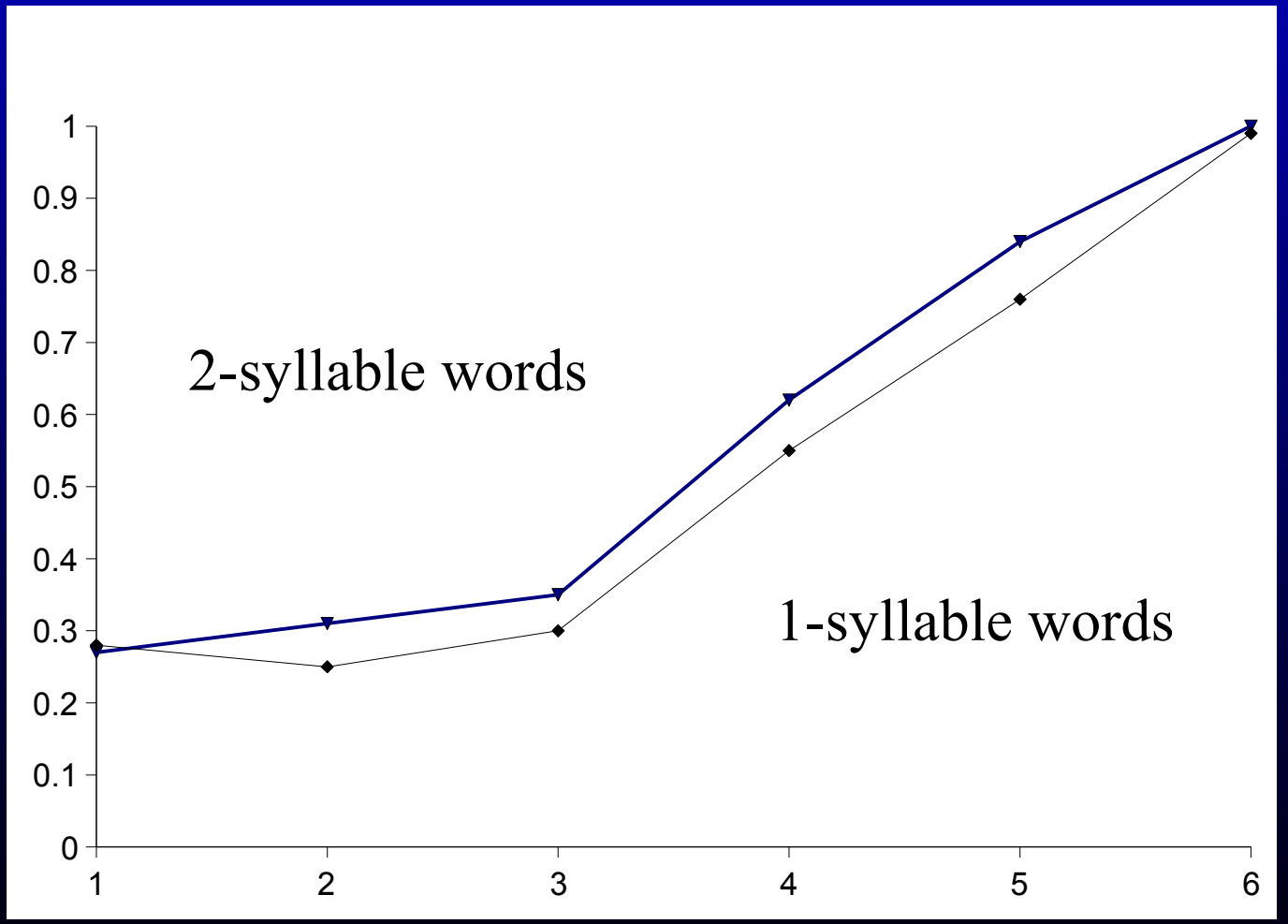






Cowan et al., 1997

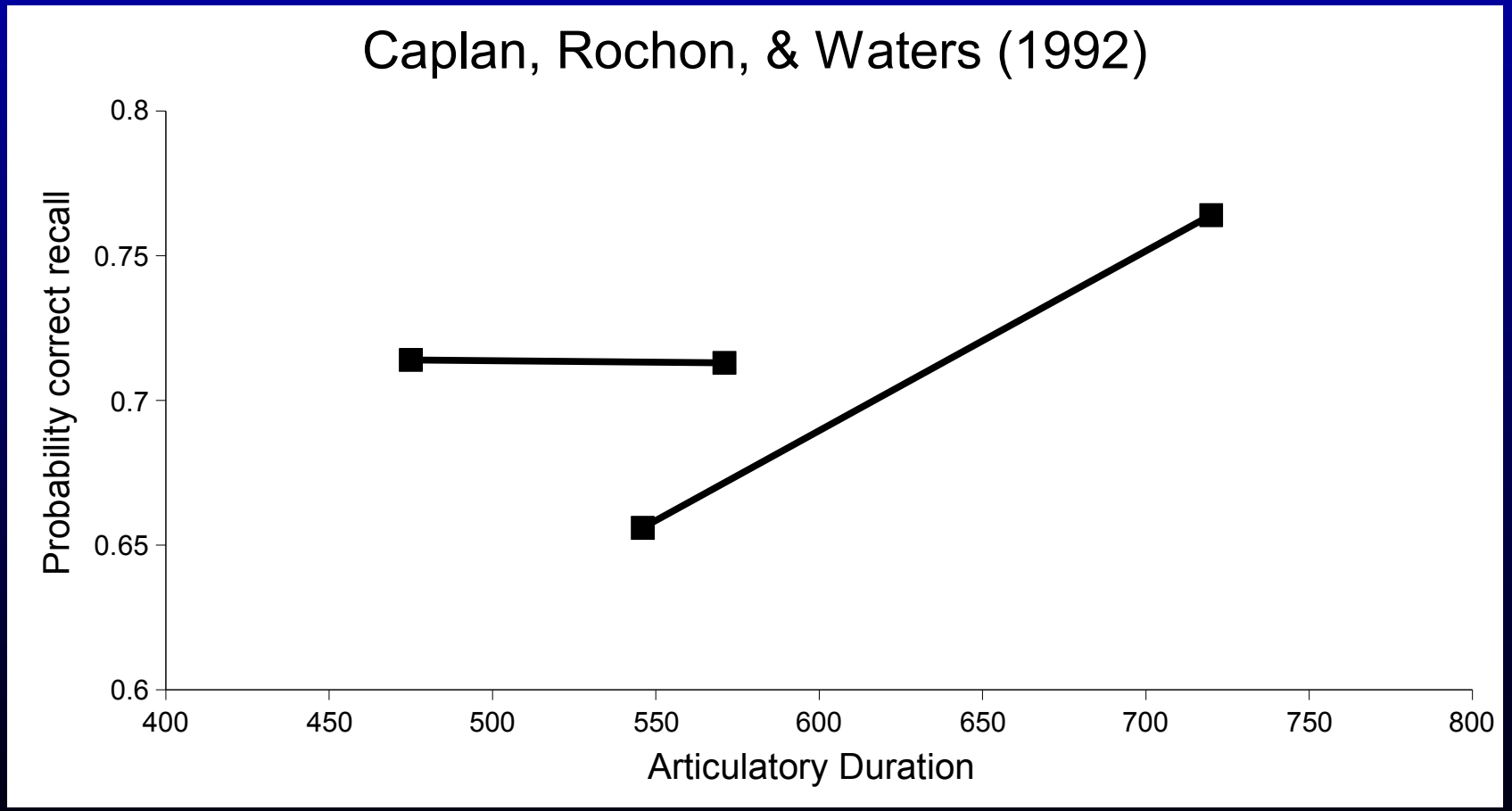
- Cowan et al. (1997) created 'long' words out of 'short' words.



Cowan et al., 1997

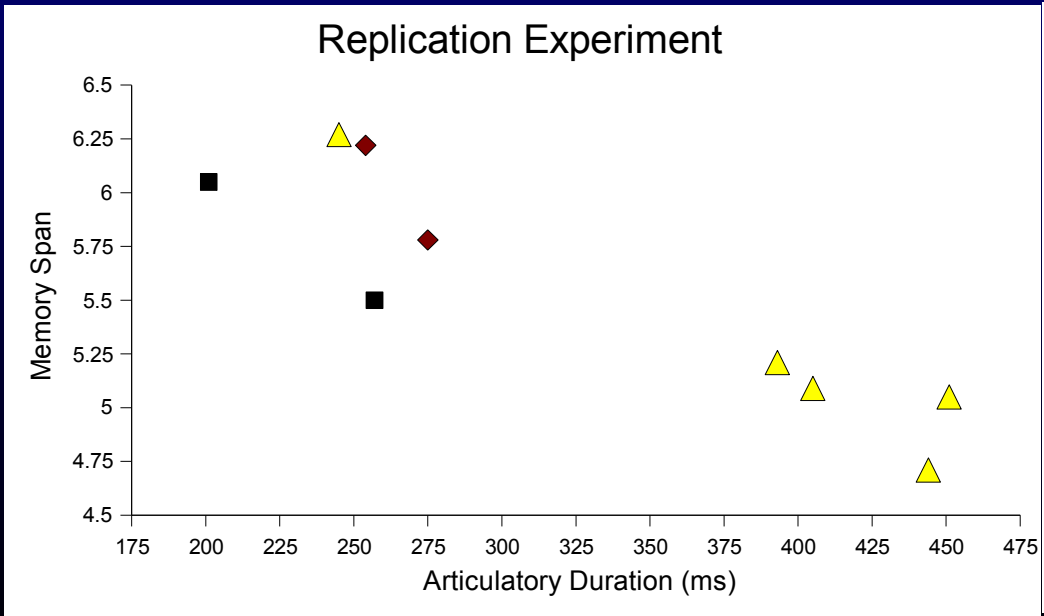
- PSIMETRICA can be used to evaluate their similarity.
- Evaluation shows: (Long versus Short)
  - Onsets: .42 vs. .37
  - Nuclei: .26 vs. .20
  - Codas: .25 vs. .275
- Long words were less similar than short!

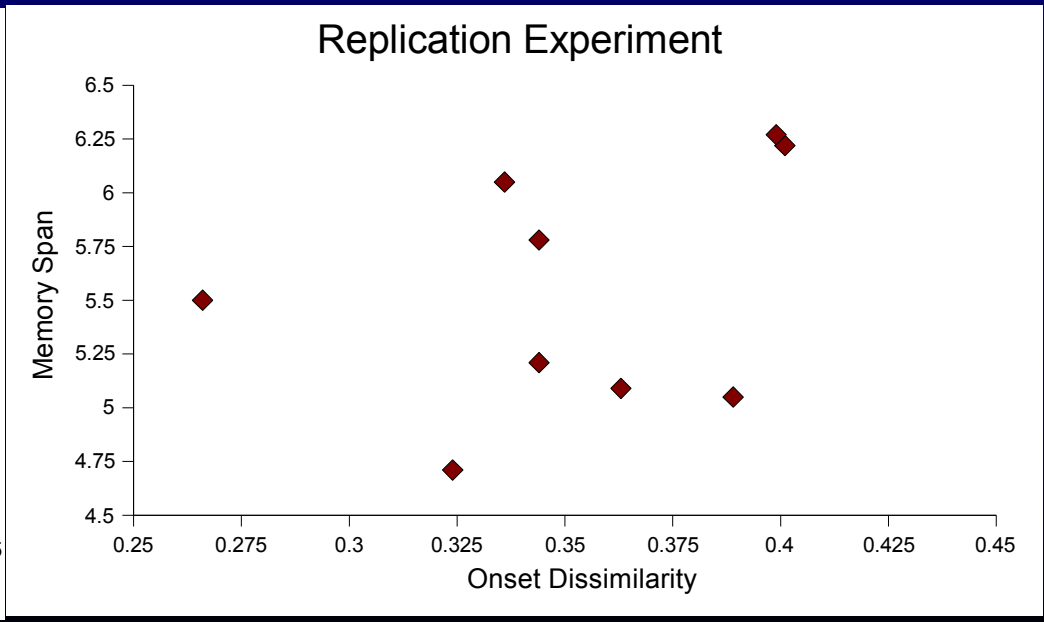
Caplan, Rochon, and Waters (1992) showed that articulatory duration has no effect or an effect opposite of that predicted by the phonological-loop model.



New Experimentation

- Mueller et al. (2004) replicated design of Caplan et al. (1992), adding new word sets to aid in calibration
- Phonological similarity and articulatory duration alone are each poor predictors of memory span

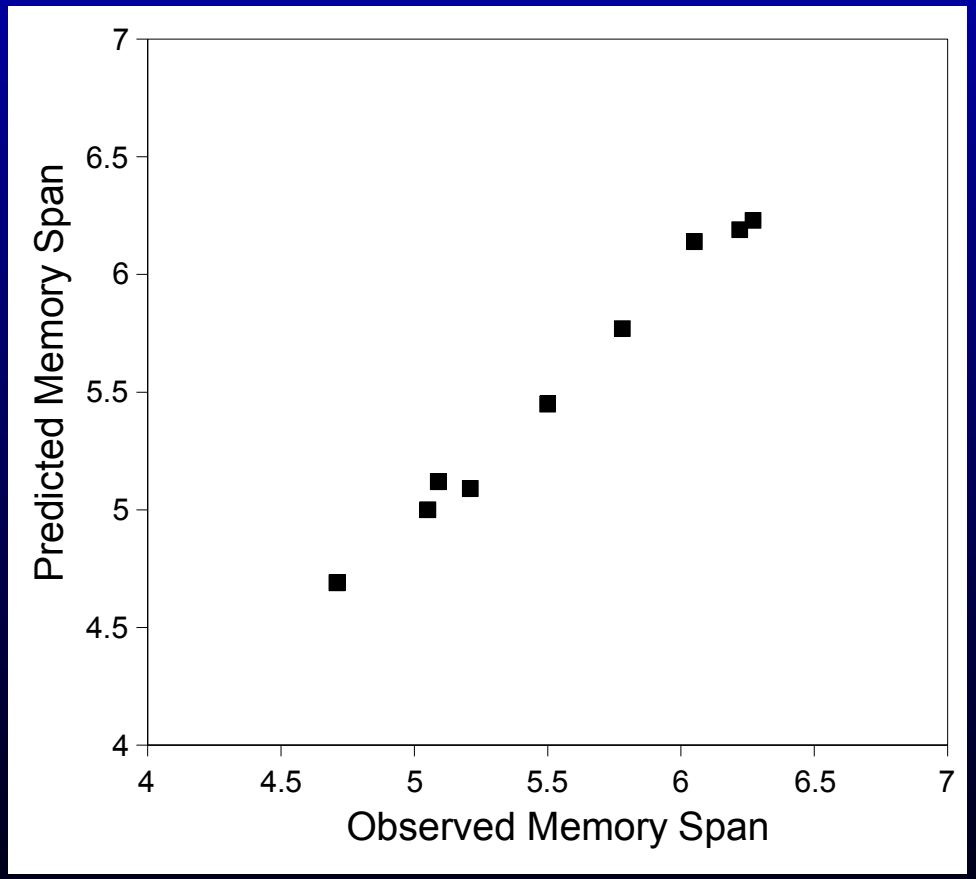




Duration and Similarity

- Together, articulatory duration and phonological dissimilarity can explain and predict performance in the memory-span task (see right panel.)
- The similarity of the onset was able to account for deviations from the performance expected by articulatory duration.
- These conclusions were confirmed by using 0-parameter prediction to a new set of data.

Results from Experiments 1 and 2



Interim Discussion

- In verbal working memory tasks, duration and onset similarity together appear to account for many obtained violations of the articulatory loop model.
- Onset similarity seems to play an important role.
- PSIMETRICA can be used for other types of psychological data....

Other Applications

- Useful for selecting stimuli to manipulate/control for similarity
- Searching databases based on similarity of sound
- Evaluating other memory and linguistic tasks
  - Subjective Ratings Tasks
  - Understanding similarity/ neighborhood effects in spoken word recognition

Database Search

- Goal: Search a database of words based on phonological similarity
- Allow different aspects of similarity to be selected and weighted.
- Demo: Database of 3000 most common written words in English.

Syllable Onset matches to SERVICE

Onsets:

"CIVIL"	0.0
"SERVICE"	0.0
"SERVING"	0.0
"SEVEN"	0.0
"SEVERE"	0.0
"SOVIET"	0.0
"SURVEY"	0.0
"SERVICES"	0.0093
"C"	0.021
"CELL"	0.021
"CELLS"	0.021

Stress-weighted onsets:

"CIVIL"	0.0
"SERVICE"	0.0
"SERVING"	0.0
"SEVEN"	0.0
"SOVIET"	0.011
"SEVERAL"	0.027
"SERVICES"	0.028
"SUFFERED"	0.035
"SURFACE"	0.035
"SOUTHERN"	0.046
"SYMBOL"	0.046

Overall Matches to SERVICE

Onset+Nucleus+Coda:

"SERVICE"	0.0
"SURFACE"	0.026
"CERTAINLY"	0.097
"CERTAIN"	0.11
"SOURCES"	0.12
"SIR"	0.13
"SERVICES"	0.13
"CIRCLE"	0.13
"SURVEY"	0.13
"SURPRISE"	0.16
"SERVING"	0.16

(O+N+C)*Stress:

"SERVICE"	0.0
"SURFACE"	0.026
"CERTAIN"	0.11
"SOURCES"	0.12
"CIRCLE"	0.13
"SERVING"	0.17
"CERTAINLY"	0.20
"PURPOSE"	0.20
"STARTED"	0.21
"SHERMAN"	0.21
"SCIENCE"	0.23

O*N*C*S:

"SERVICE"	0.0
"SURFACE"	0.026
"CERTAIN"	0.15
"CIRCLE"	0.17
"SOURCES"	0.20
"PURPOSE"	0.20
"SERVING"	0.22
"PURCHASE"	0.25
"CERTAINLY"	0.26
"SHERMAN"	0.28
"SERVICES"	0.28

Especially Useful for Names

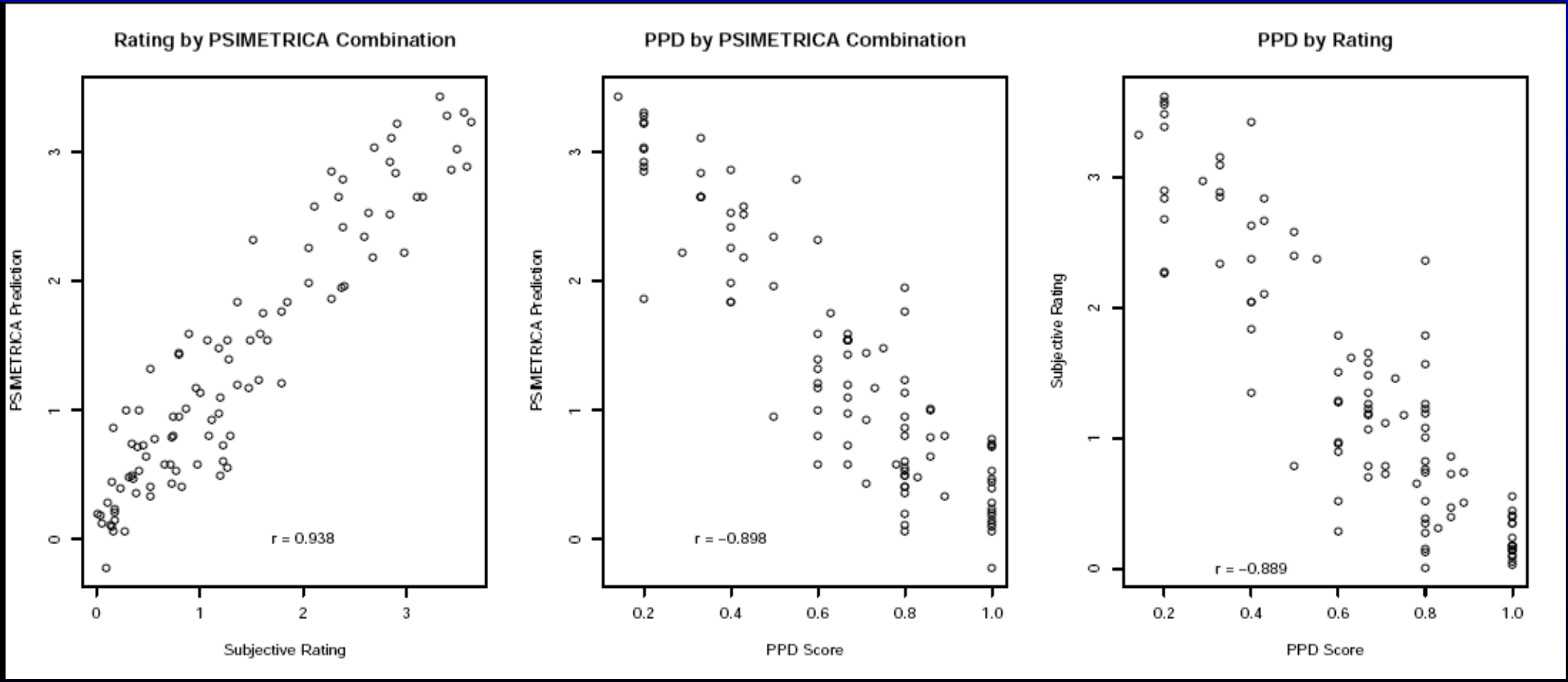
Matches to SHIRLEY:

"Shirley"	0.0
"Sheri"	0.072
"Shelly"	0.072
"Shelby"	0.20
"Marina"	0.50
"Kimberly"	0.52
"Stephanie"	0.52
"Rodney"	0.52
"Gregory"	0.64

Similarity Ratings Data

- Vitz & Winkler (1972) compared 4 sets of 25 words to 4 different comparison words.
- Obtained subjective ratings of similarity from subjects
- Compared ratings to an edit distance where differences between individual phonemes were all equal.
- Achieved high correlations (-.8 to -.95) between ratings and PPD score.
- Did not use sub-phonemic similarity levels

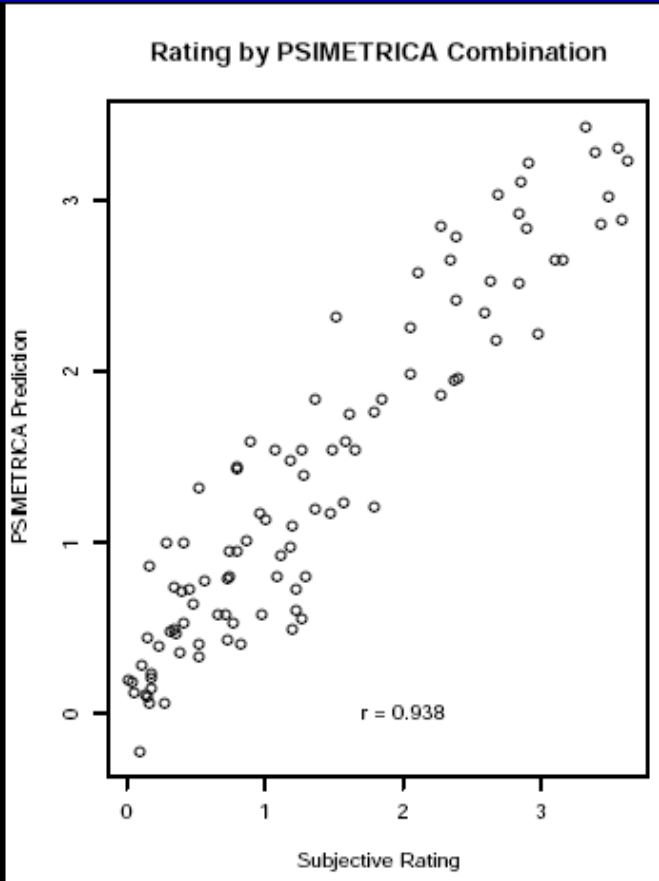
Vitz/Winkler and PSIMETRICA Comparison Results

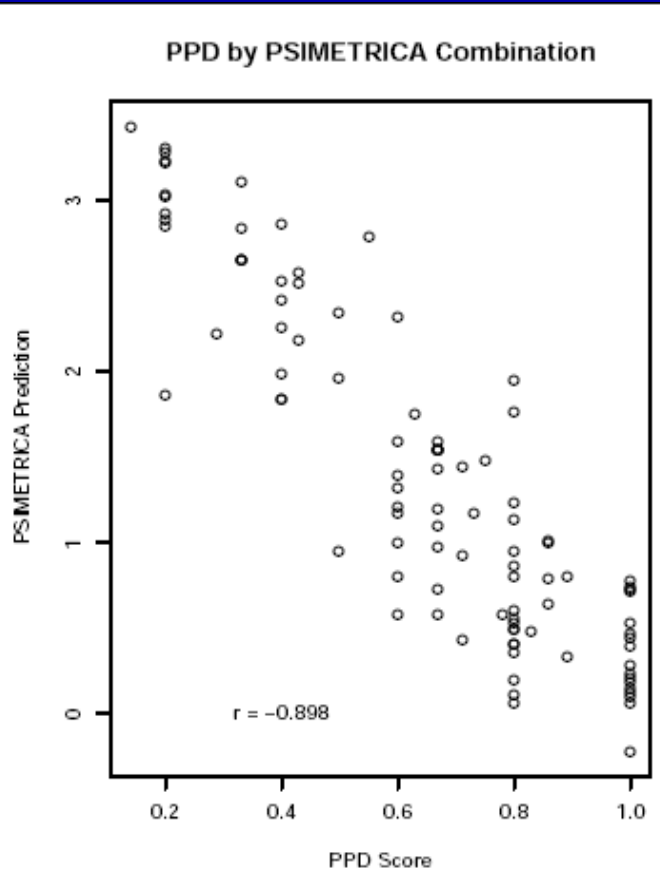


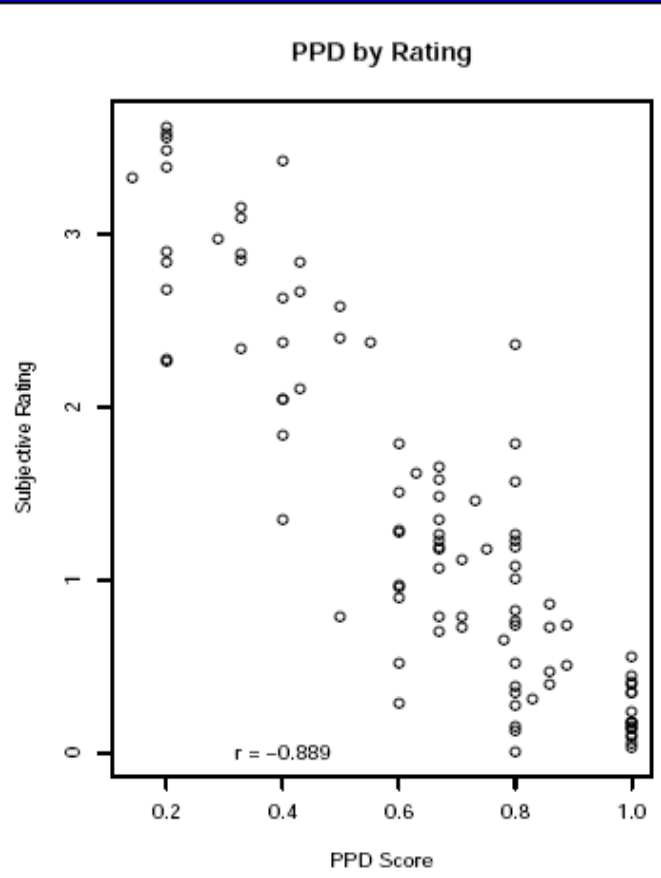
PSIMETRICA Results

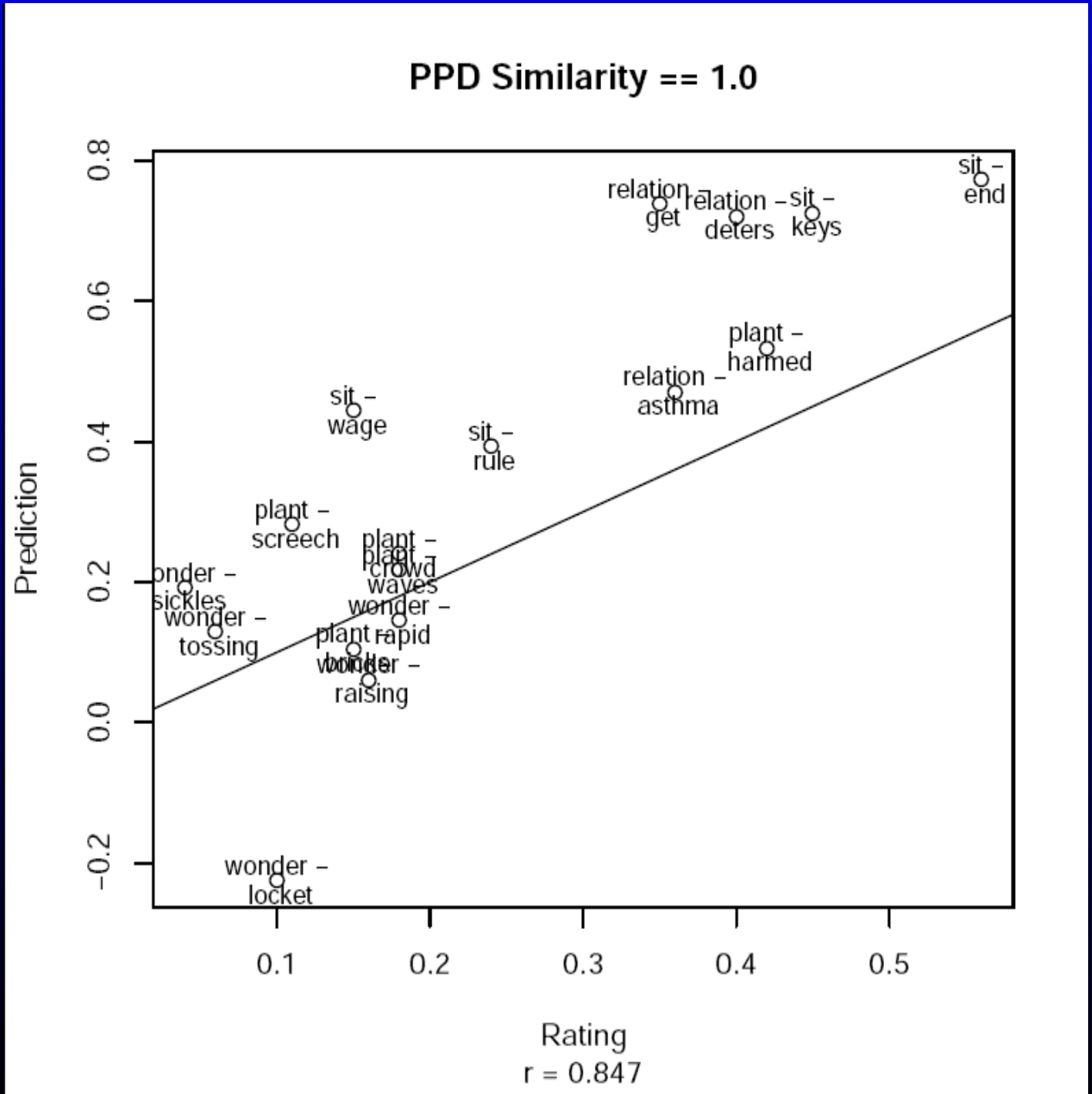
- When applied to similarity ratings data sets, PSIMETRICA produces predicted values with similar predictive power (r = .8 to .95)
- PSIMETRICA uses linear model with multiple parameters (and so has advantage)
  - onset nucleus coda o:n:c
    -1.680 -2.155 -1.538 1.512
- PSIMETRICA has some additional parameters we can adjust to improve similarity.
- We can assess ratings versus predictions for pairs PPD model assigns identical scores.

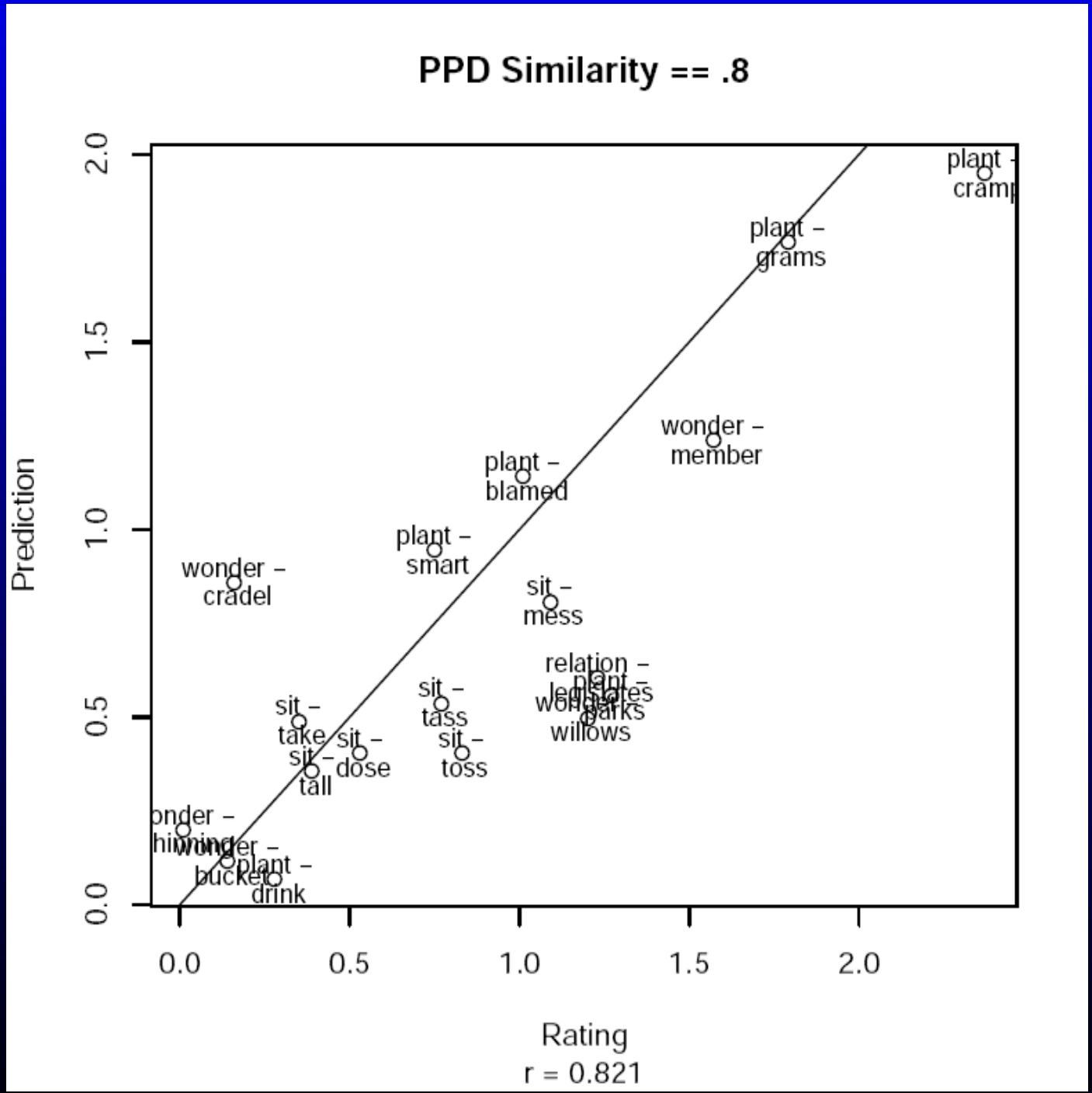
Vitz/Winkler and PSIMETRICA Comparison Results

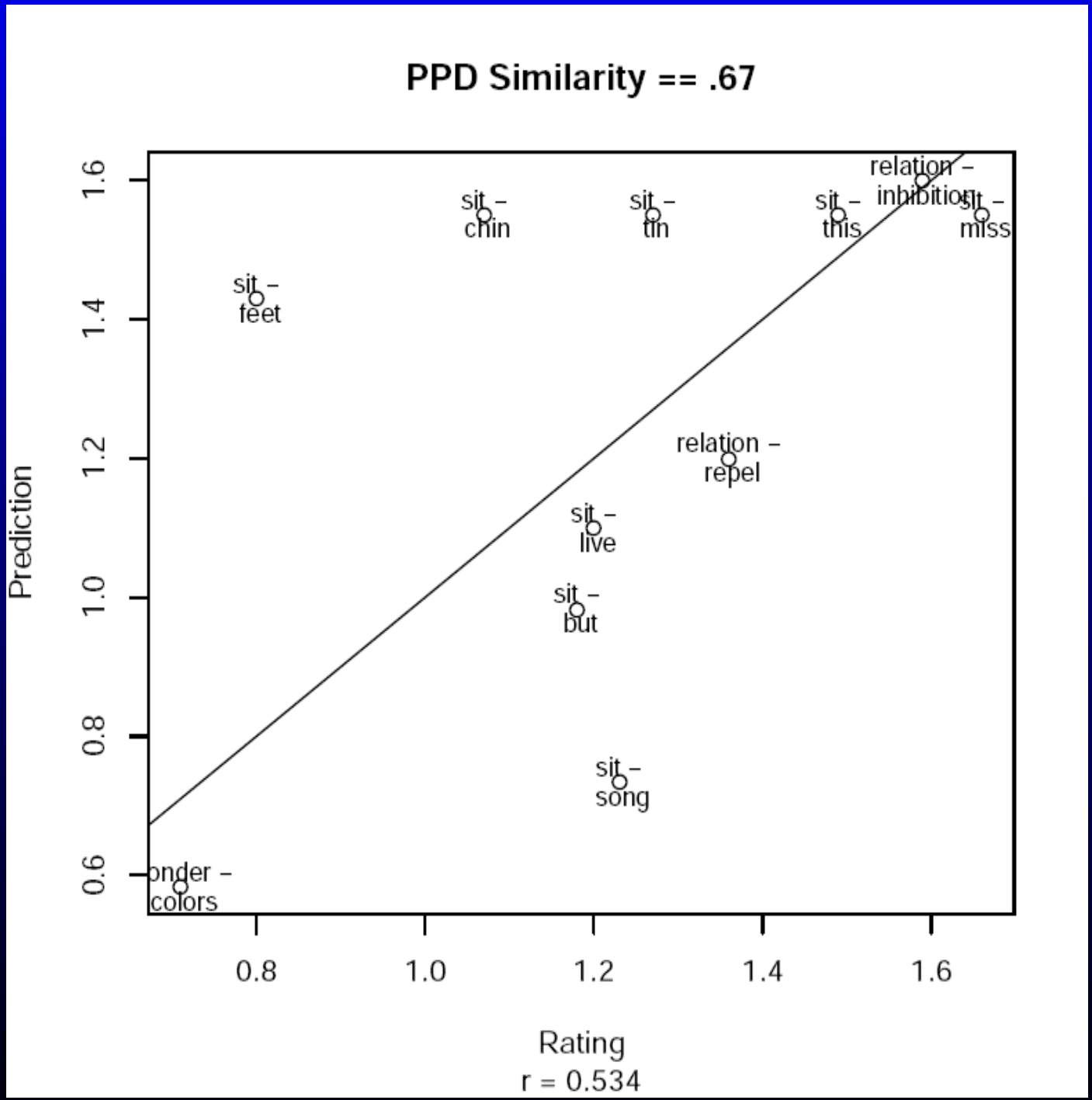


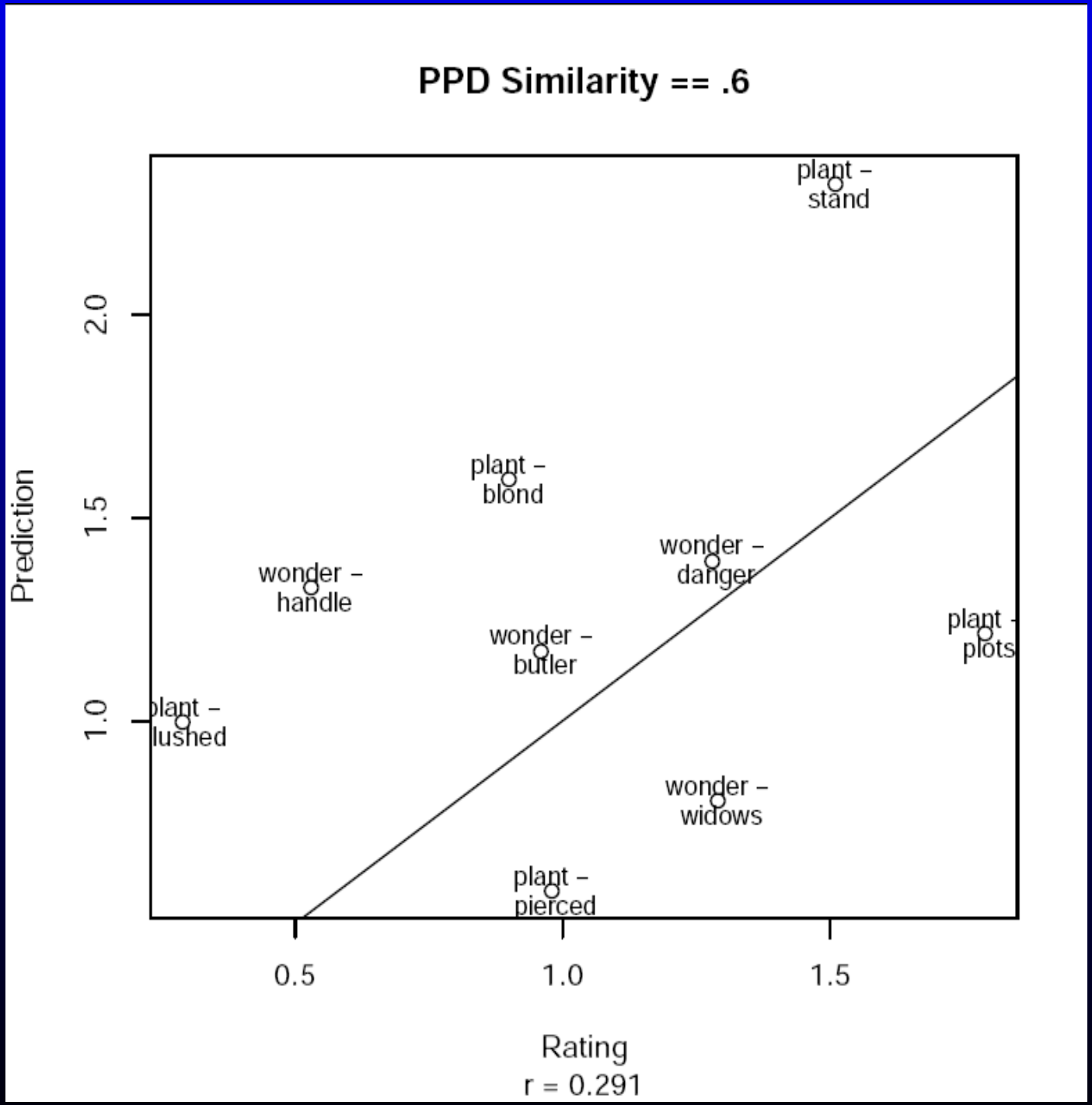


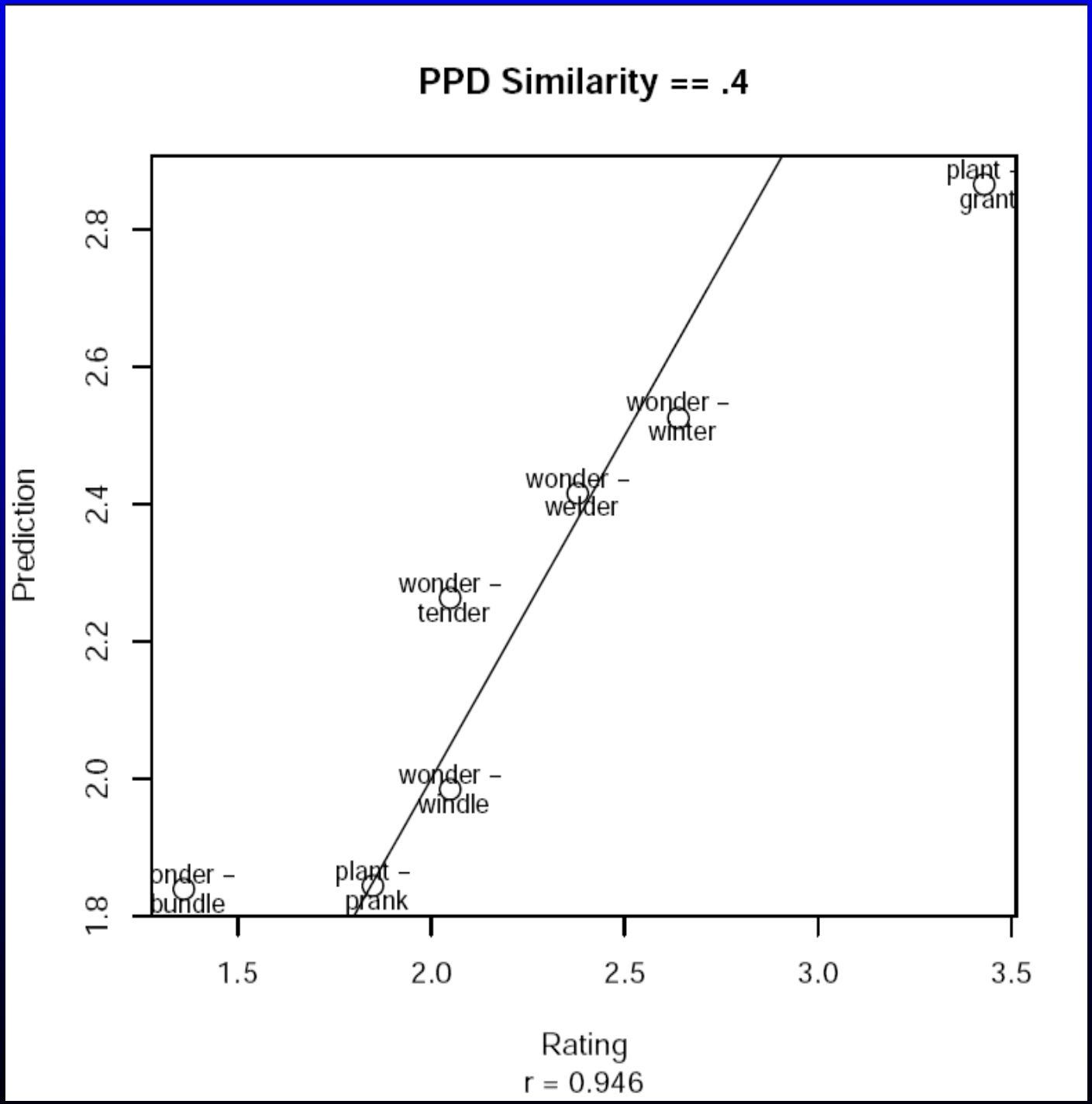


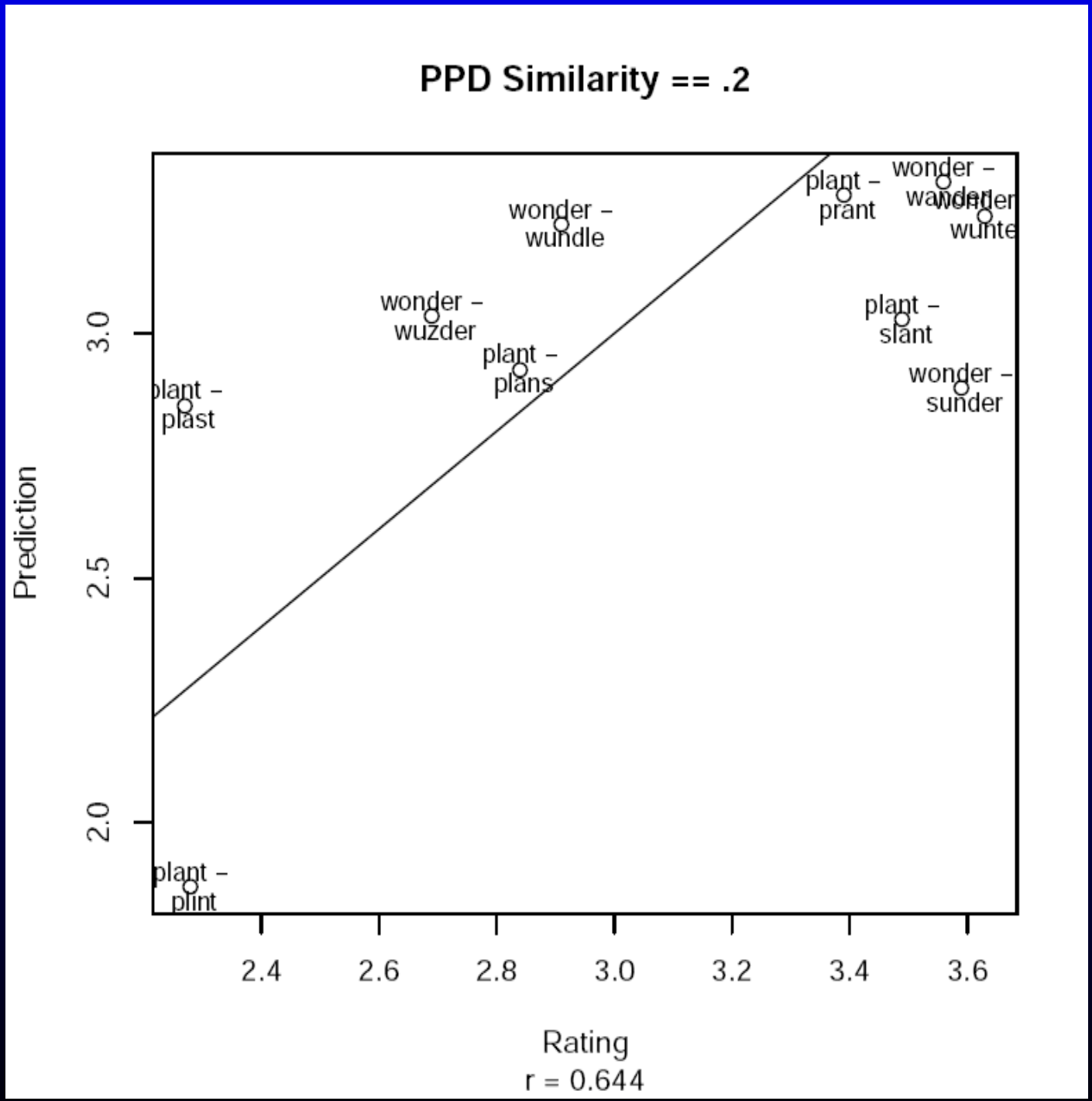

PPD Similarity == 1.0
Prediction
Rating
r = 0.847
0.8
0.6
0.4
0.2
0.0
−0.2
0.1
0.2
0.3
0.4
0.5
sit – end
relation – get
relation – deters
sit – keys
plant – harmed
relation – asthma
sit – wage
sit – rule
plant – screech
plant – crowd
plant – waves
wonder – sickles
wonder – tossing
wonder – rapid
plant – raising
wonder – locket


PPD Similarity == .8
Prediction
Rating
r = 0.821
0.0
0.5
1.0
1.5
2.0
plant – cramp
plant – grams
wonder – member
plant – blamed
plant – smart
wonder – cradel
sit – mess
relation –
sit – tass
sit – take
sit – dose
sit – toss
sit – tall
willows
drink


PPD Similarity == .67
Prediction
Rating
r = 0.534
0.6
0.8
1.0
1.2
1.4
1.6
relation – inhibition
sit – chin
sit – tin
sit – this
sit – miss
sit – feet
relation – repel
sit – live
sit – but
sit – song
onder – colors


PPD Similarity == .6
plant – stand
plant – blond
wonder – danger
wonder – handle
plant – plots
wonder – butler
plant – lushed
wonder – widows
plant – pierced
Prediction
2.0
1.5
1.0
0.5
1.0
1.5
Rating
r = 0.291


PPD Similarity == .4
Prediction
1.8
2.0
2.2
2.4
2.6
2.8
1.5
2.0
2.5
3.0
3.5
Rating
r = 0.946
plant – grant
wonder – winter
wonder – welder
wonder – tender
wonder – windle
plant – prank
onder – bundle

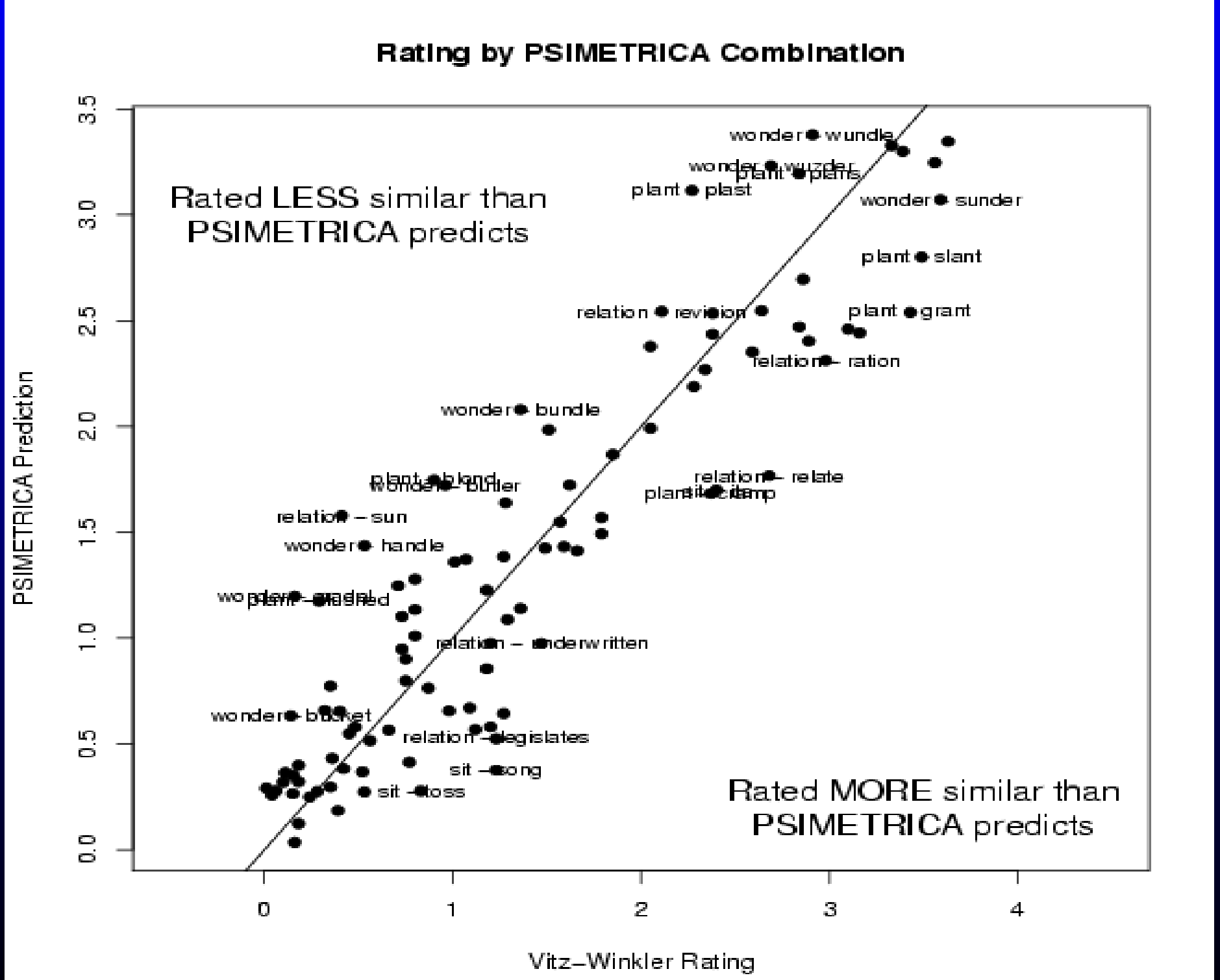

PPD Similarity == .2
Prediction
Rating
r = 0.644
2.0
2.5
3.0
2.4
2.6
2.8
3.0
3.2
3.4
3.6
plant – plint
plant – plast
wonder – wuzder
plant – plans
wonder – wundle
plant – prant
wonder – wunder
plant – slant
wonder – sunder

Similarity Ratings

- Within a single PPD level, PSIMETRICA can predict observed differences in ratings
  - Many of the improvements stem from sub-phoneme matches.

Regression Analysis

- Best regressions included Onset, Nucleus, Coda as main effects and a O-N-C interaction.
- Only one experiment had words that differed in syllable/stress patterns, and for only this experiment did stress predict ratings.
- Outliers included cross-syllable similarity (tip versus pit)
  - cross-sim measure was not reliable predictor overall
    - It was reliable within the “SIT” experiment.
    - Too few examples to yield reliability
    - Adaptive measures may be required


Rating by PSIMETRICA Combination
Rated LESS similar than PSIMETRICA predicts
Rated MORE similar than PSIMETRICA predicts
PSIMETRICA Prediction
Vitz–Winkler Rating
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
1
2
3
4
wonder – wundle
wonder – wurder
plant – plans
plant – plast
wonder – sunder
plant – slant
relation – revision
plant – grant
relation – ration
wonder – bundle
relation – relate
relation – sun
wonder – handle
relation – underwritten
wonder – basket
relation – legislates
sit – song
sit – toss

Reading tea leaves in outlier analysis

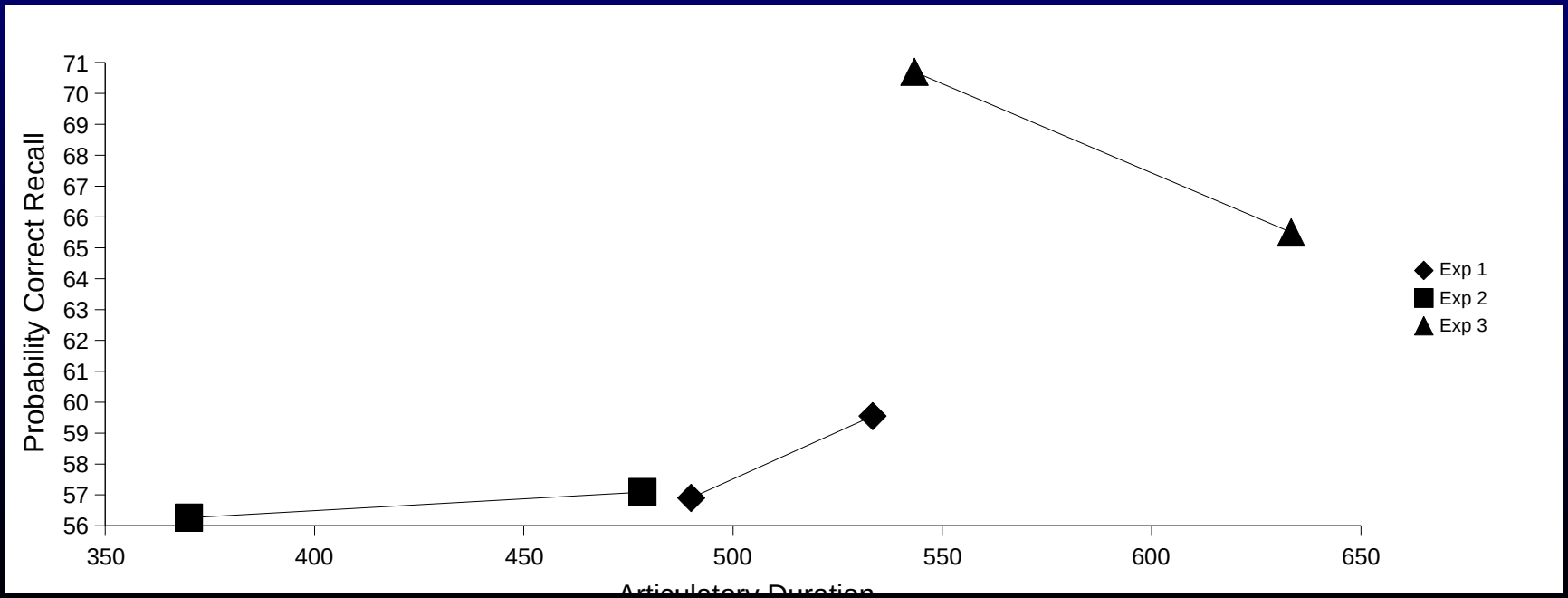
- Overprediction on words that differ by l-r
  - wundle-wonder, wonder-bundle, wonder-cradle, wonder-handle
  - Suggests differential weighting of features
- Underprediction on cross-consonant matches
  - sit-toss, sit-tass, its-sit
- Underpredict S-X matches: sit-hit, sit-pit, slant-plant
  - re-evaluate onset templating/null comparisons
- Underprediction on relation-long words
  - relation-ration, relation-legislates, relation-relate, relation-underwritten

Conclusions

- PSIMETRICA is a useful technique for measuring phonological dissimilarity.
- Multi-dimensional “similarity profile”
- Born out of work on verbal working memory
- Applied to subjective ratings
- With Tom, I am working on looking at speech identification/perception data.
  - Compute similarity space for 20,000 words.
- Other data from members of group?

Lovatt, Avons, and Masterson (2000)

- To test Baddeley's (1986) phonological-loop model, Lovatt et al. (2000) conducted three serial recall experiments with six different sets of 2-syl. words.
- Across word sets, phonological dissimilarity was "equated" on the basis of introspective judgments.

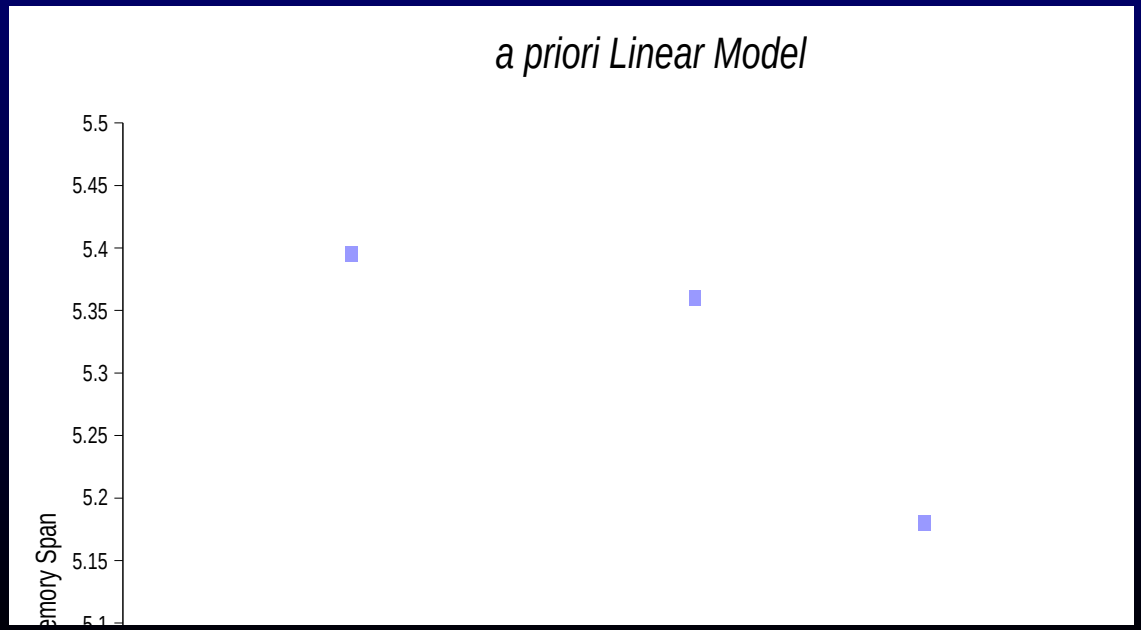


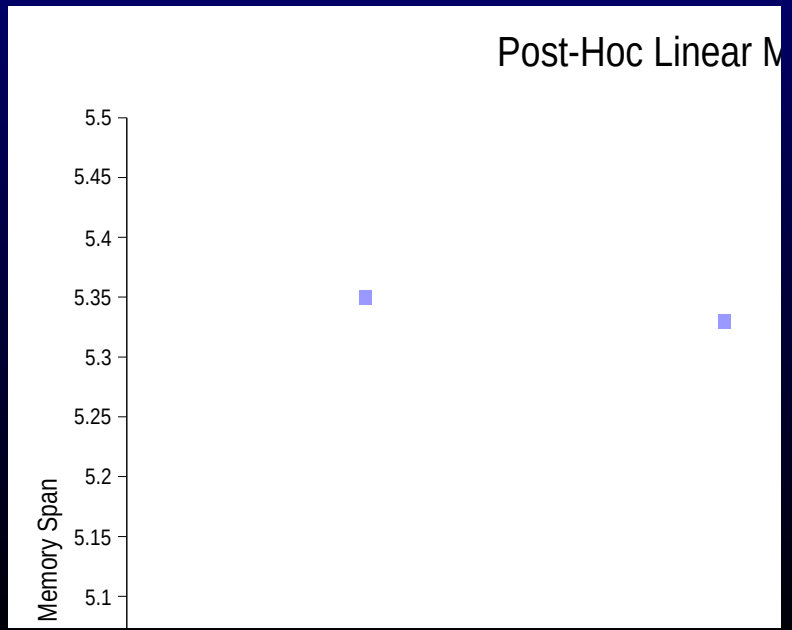
Replication of Lovatt et al. (2000)

- Four sets of words (from Lovatt et al., 2000), Exps. 1 and 2 were included.
- Articulatory duration was measured with a list recall procedure.
- Memory span was measured with an adaptive staircase method.
- Phonological dissimilarity was measured with PSIMETRICA.

Results of Replicating Lovatt et al.

- Memory spans were predicted in two ways: (1) on an a priori basis with the linear model for results from our earlier experiments, and (2) with a new linear model that re-estimated the effects of articulatory duration and phonological dissimilarity.
- Observed (solid diamonds and predicted (open diamonds) memory spans are shown below.





Detailed Results

More results and model coefficents from our replication of Lovatt et al. (2000) are shown below:

Measured Variable	Word Set				
	1-"Short"	1-"Long"	2-"Short"	2-"Long"	
Phonological Dissimilarity	0.297	0.318	0.337	0.348	
Articulatory Duration (ms)	381	315	370	349	
Observed Memory Span	5.16	5.35	5.29	5.32	
Predicted Memory Span (a priori model)	4.9	5.39	5.18	5.36	
Predicted Memory Span (new model)	5.35		5.27	5.33	5.17

Linear Coefficients

Parameter	a priori Model	New Model
Intercept	5.5	5.33
Articulatory Duration	-.00573	-.00199
Onset Similarity	5.35	2.02