Performance of HAL-like word space models on semantic categorization tasks.

Cyrus Shaoul
Chris Westbury

Department of Psychology, University of Alberta

Additions to Marco's List

- COALS (completed in 2005, submitted in 2007) Rohde, Plaut & Gonnerman
- BEAGLE (2006,2007) Jones & Mewhort
  - Some Fortran source code now available
  - New work on Semantic Distinctiveness of contexts

What are we working with?

Corpus → Word-Space → Clusters

Subjects → Responses → Categories

Clusters vs. Categories

- For the purposes of this talk:
  - Clusters are groupings of points in a word-space, each point representing the contextual information for that word. Grouping is based on the geometric relationship between points.
  - Categories are groupings of words gathered from studies of language behavior.
- Both are groupings of words, but is there any point of comparing categories to clusters?

Theoretical Concern

- What does a close match between a cluster and a category mean?
- **If** there is a correspondence between the human semantic space organization and a model's representation of word meaning **and** there is a correspondence between clustering algorithms and the human faculty for word categorization **then** there is reason to hope that this task is valid.

My model and how it works

- HAL-like
- No factorization or PCA
- Used USENET and GIGAWORD corpora

How HAL works:

	bank -5	bank -4	bank -3	bank -2	bank -1	bank +1	bank +2	bank +3	bank +4	bank +5
run	23	12	43	103	2	0	201	23	48	23

Apply weighting function

	bank (backwards)	bank (forwards)
run	183	295

Insert into global co-occurrence vector

	a	<---->	ban	bank	bark	bat	<---->	zoo	a	<---->	ban	bank	bark	bat	<---->	zoo
run	13210	<---->	101	183	23	1492	<---->	32	5400	<---->	21	295	9	10293	<---->	2

Insert this vector into the global co-occurrence matrix

	a	<---->	ban	bank	bark	bat	<---->	zoo	a	<---->	ban	bank	bark	bat	<---->	zoo
rub	342	<---->	2	2	34	3	<---->	3	1322	<---->	0	0	1	0	<---->	0
rump	3454	<---->	0	0	0	0	<---->	2	1233	<---->	0	2	0	0	<---->	0
run	13210	<---->	101	183	23	1492	<---->	32	5400	<---->	21	295	9	10293	<---->	432
runner	65242	<---->	34	33	0	4523	<---->	0	4321	<---->	0	2	4	22	<---->	344
runs	24556	<---->	5	546	0	5312	<---->	0	3455	<---->	0	2	24	43	<---->	23

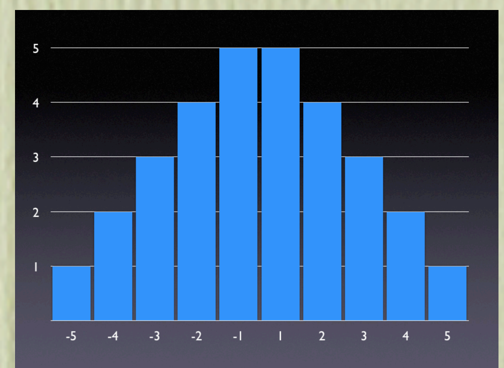
Changes we have made to HAL

- Made changes to the choice of the vectors used (Original HAL: vectors with the greatest variance. Our model: vectors with with greatest frequency.)
- Normalize vectors by using a frequency ratio.
- Added a neighborhood threshold, restricting the number of neighbors to those within a standardized distance away.

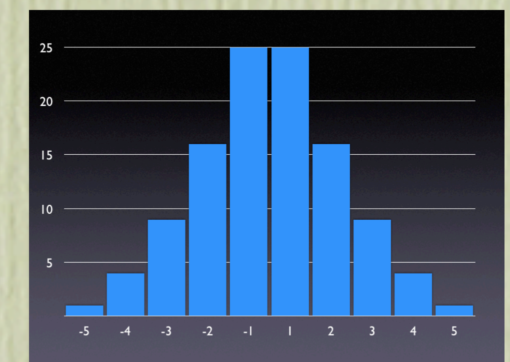
Personal Research Interest

- Can a word’s neighborhood density help predict the time it takes to access the word’s meaning?
- Does the setting of parameters of the HAL model change its ability to make this prediction?

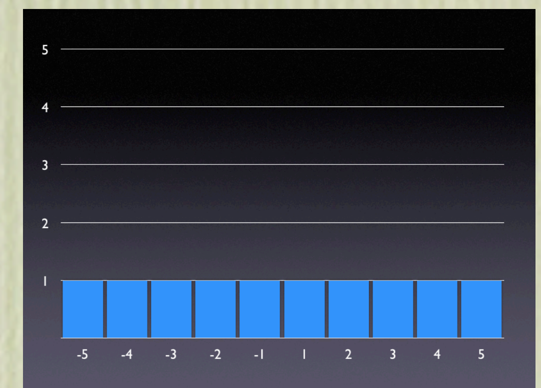
Weighing Schemes



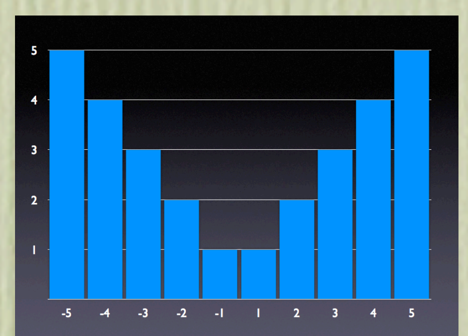
Linear Ramp



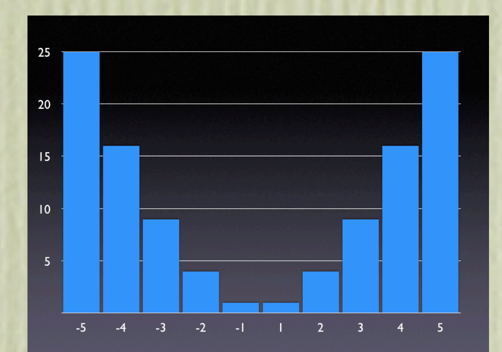
Exponential Ramp



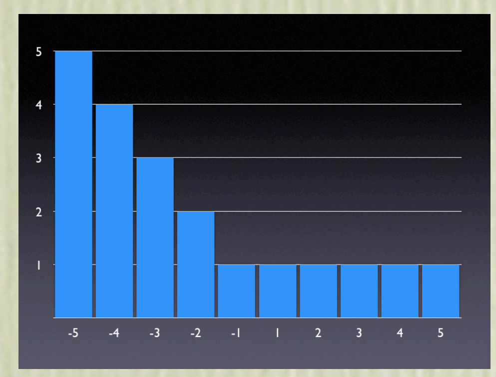
Flat



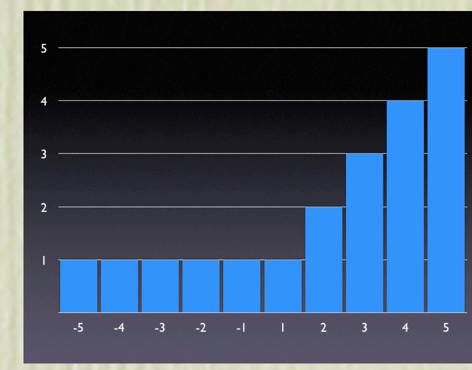
Inverse Ramp



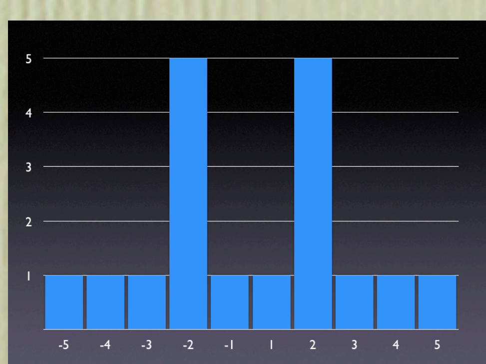
Inverse Exponential Ramp



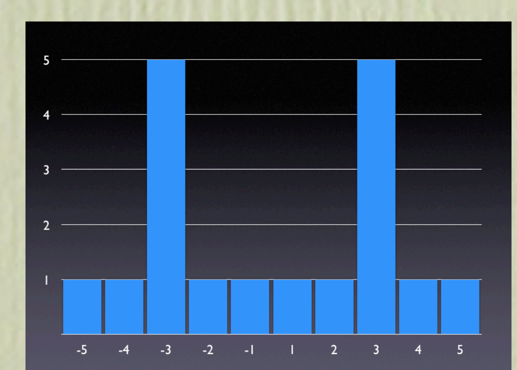
Linear Ramp Behind



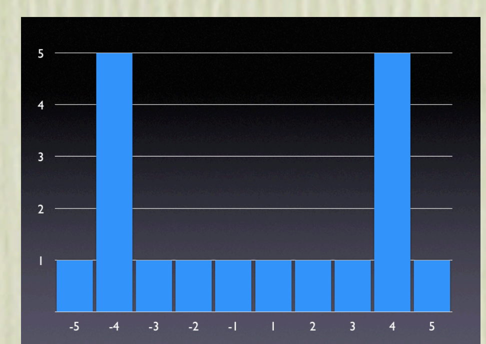
Linear Ramp Ahead



Second Word



Third Word

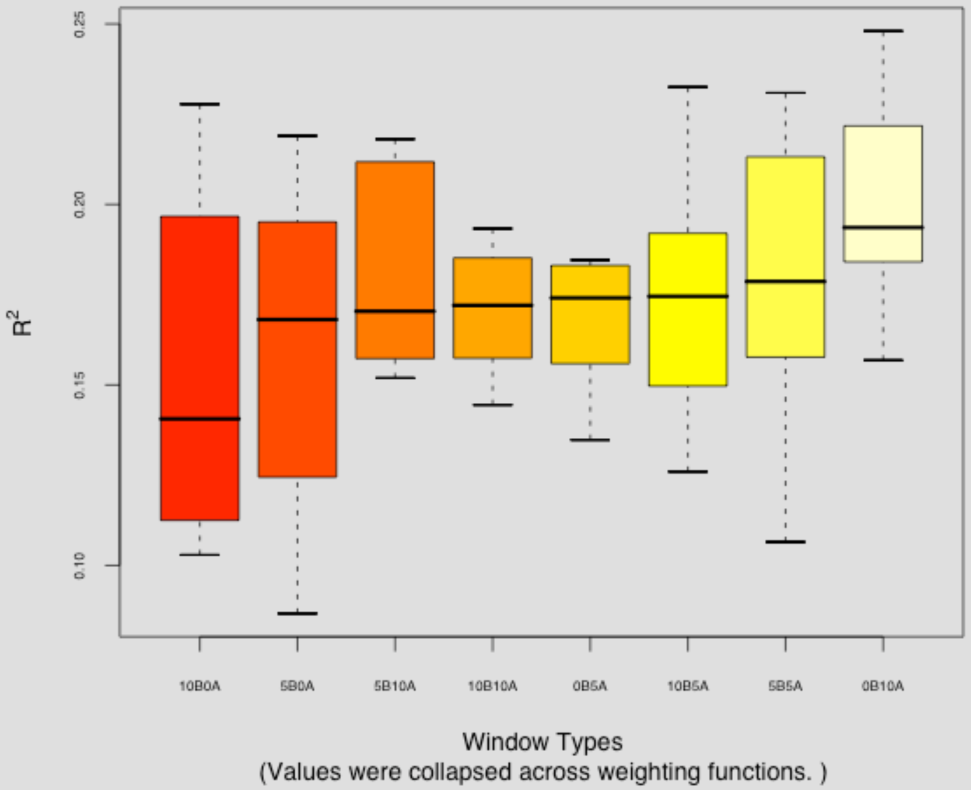


Fourth Word

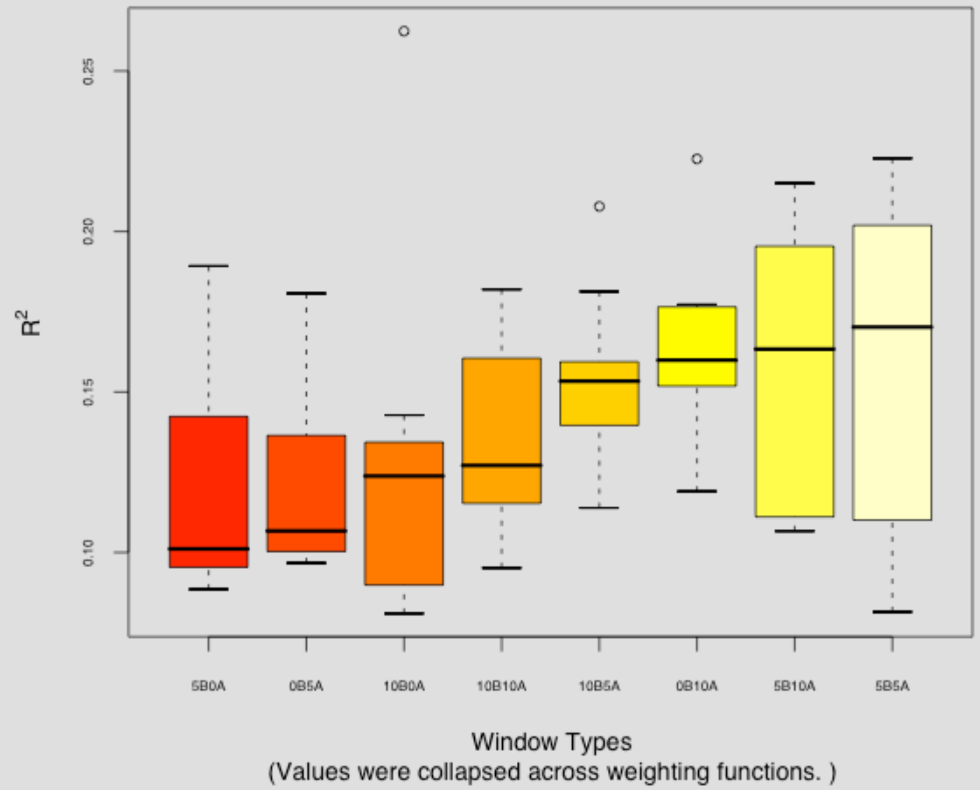
Lexical and Semantic Decision Results

- Each combination of parameter settings produces differing measures of neighborhood density
- Some parameter sets give higher correlations with RT than others.
- Performed a search through parameter space.
- Best LDRT predictor: Inverse Ramp, 10 words behind, 5 words ahead.

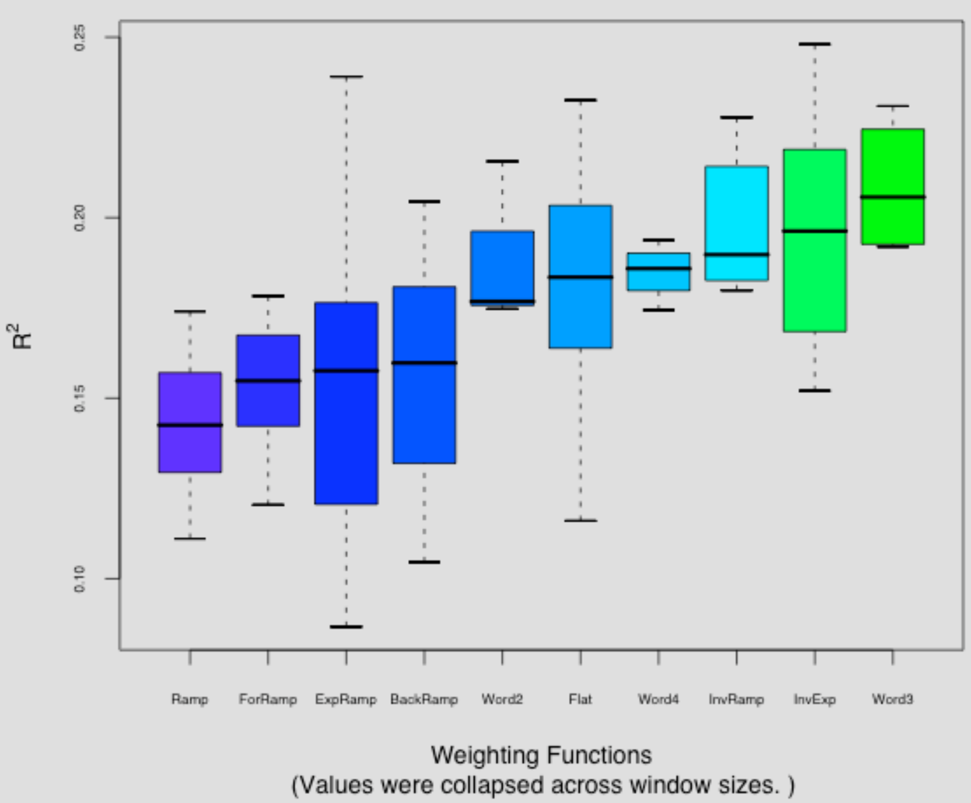
R^2 of $\frac{1}{\text{NCOUNT}+1}$ with LDRT for different window sizes.



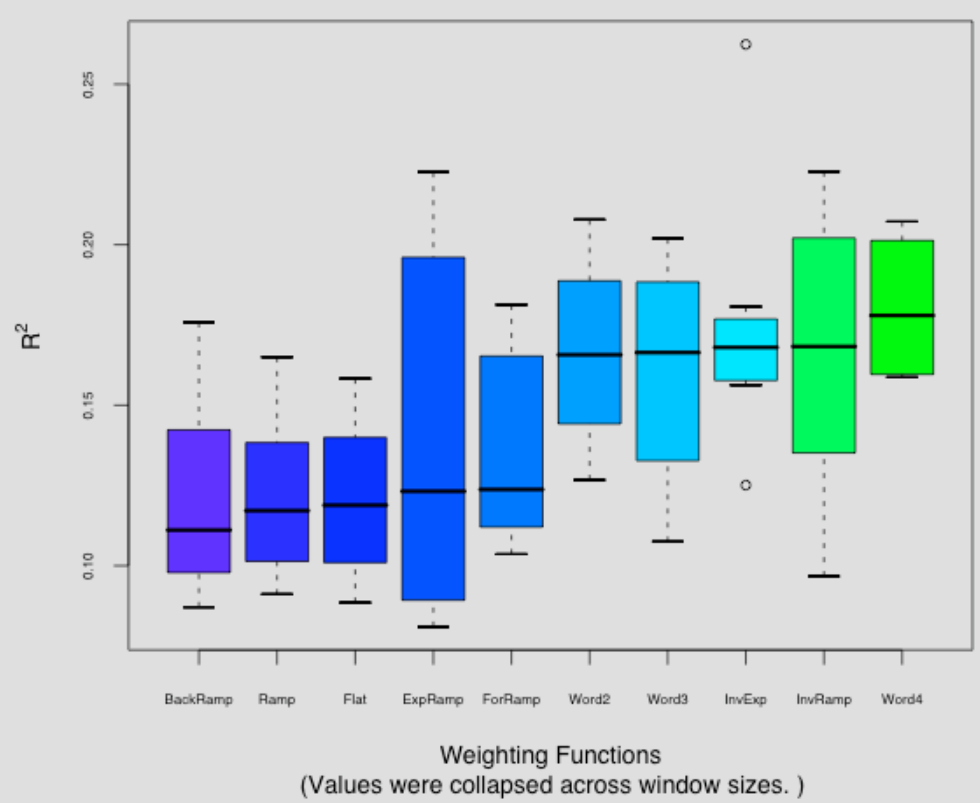
R^2 of ARC with LDRT for different window sizes.



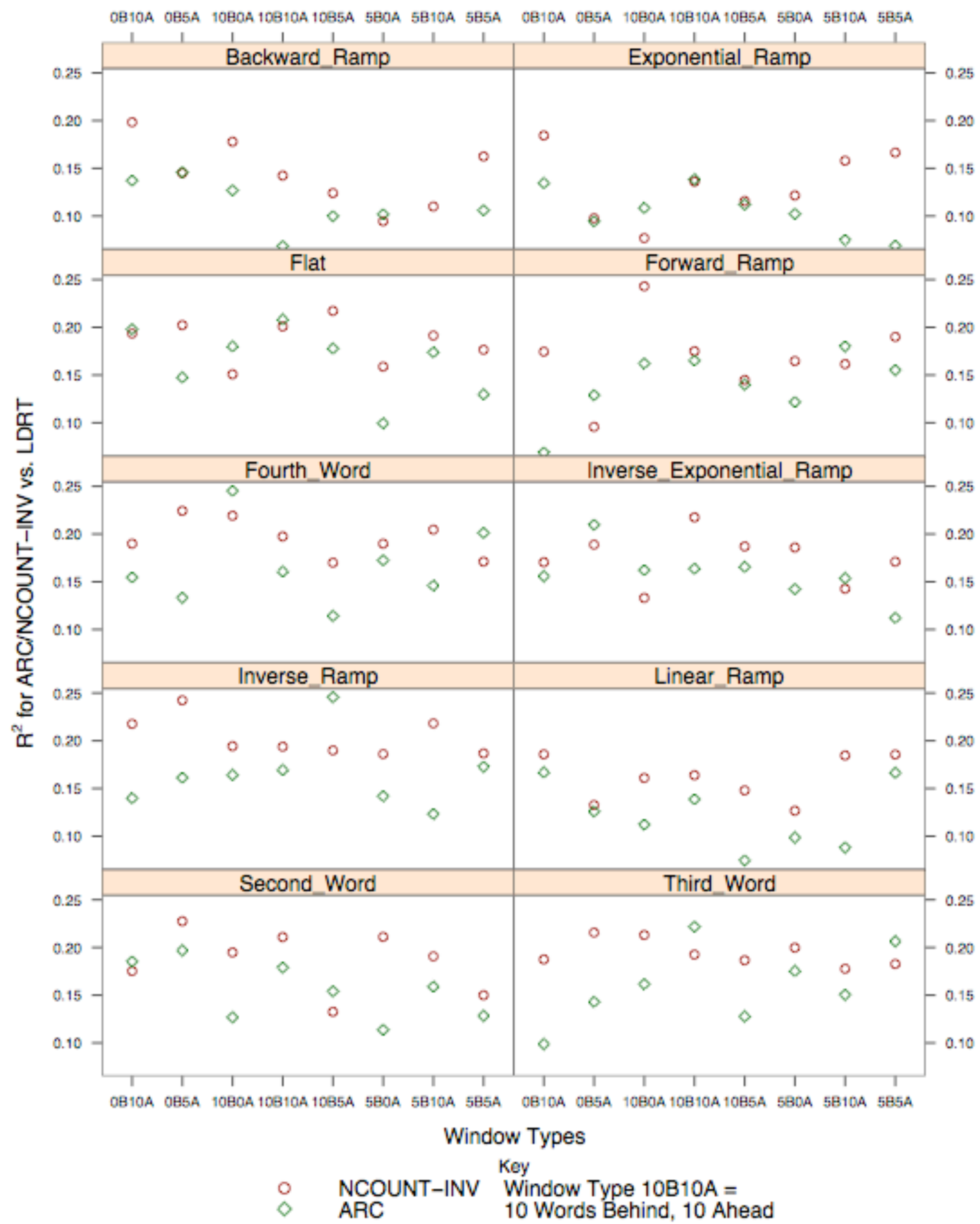
R^2 of $\frac{1}{\text{NCOUNT}+1}$ with LDRT for different weighting functions.



R^2 of ARC with LDRT for different weighting functions.



R^2 of ARC/NCOUNT−INV with LDRT for different weighting functions and window types.



ESSLLI 2008: Primary Goals

- Investigate the clusters produced by my HAL-like model to see how they compared to the lexical categories in the tasks.
- Compare the quality of the clusters when using different parameter settings to the model.

Tasks

- Attempted:
  - Categorization
  - Salient Property Generation
- Did not attempt:
  - Modeling free association

Concrete Noun Clustering

- Results:

Original HAL Params	Cluster Entropy	Cluster Purity
2-way	0.931	0.545
3-way	0.844	0.523
6-way	0.77	0.409

LDRT Optm Params	Cluster Entropy	Cluster Purity
2-way	0.981	0.545
3-way	0.869	0.523
6-way	0.719	0.386

Concrete Noun Clustering

Error Analysis: HAL Confusion Matrix

Cluster	Bird	Tree	Green Veg	Ground Animal	Tool	Vehicle
#0	0	0	0	0	1	0
#1	0	0	0	1	0	0
#2	0	1	0	0	0	0
#3	3	0	0	0	0	0
#4	1	1	0	1	6	3
#5	3	2	5	6	6	4

Concrete Noun Clustering

Error Analysis: Optimized Params Confusion Matrix

Cluster	Bird	Tree	Green Veg	Ground Animal	Tool	Vehicle
#0	0	0	0	0	1 Chisel	0
#1	0	1 Cherry	0	0	0	0
#2	3 Owl Eagle Penguin	0	0	0	0	0
#3	0	0	0	0	1	1
#4	2	3	5	3	6	6
#5	2	0	0	5	5	0

Abstract/Concrete Noun Discrimination Task

- Best purity and entropy results for this task compared to all other tasks.

Abstract/Concrete Noun Discrimination Task

- Results

Original HAL Params	Cluster Entropy	Cluster Purity
2-way	0.84	0.6
Optimized LDRT Params	**Cluster Entropy**	**Cluster Purity**
2-way	0.647	0.725

Abstract/Concrete Noun Discrimination Task

Error Analysis: HAL Params Confusion Matrix

Cluster	High Imageability	Low Imageability	Intermediate Imageability
#0	15	6	5
#1	1	9	4

Abstract/Concrete Noun Discrimination Task

Error Analysis: Optimized Params Confusion Matrix

Cluster	High Imageability	Low Imageability	Intermediate Imageability
#0	15	1 Mystery	6
#1	1 Ache	14	3

Verb Clustering

- Results

Original HAL Params	Cluster Entropy	Cluster Purity
5-way	0.755	0.467
9-way	0.572	0.422

LDRT Optm Params	Cluster Entropy	Cluster Purity
5-way	0.715	0.511
9-way	0.709	0.333

Verb Clustering

Error Analysis: HAL Params Confusion Matrix

Cluster	Exchange	Motion	Change State	Body	Cognition
#0	1	0	0	0	0
#1	0	0	0	2	5
#2	1	1	0	1	1
#3	3	6	3	2	2
#4	0	8	2	5	2

Verb Clustering

Error Analysis: Optimized Params Confusion Matrix

Cluster	Exchange	Motion	Change State	Body	Cognition
#0	1 Pay	0	0	0	0
#1	0	0	0	0	1 Evaluate
#2	1 Breathe	0	0	1 Lend	0
#3	2	4	2	9	8
#4	1	11	3	0	1

Property Generation

- Can the model find properties of words? (DUCK and FLIES)
- Used HiDEx to generate 200 closest neighbors for all words in the list.
- All precision averages were under 0.02
- Neighborhoods from our HAL-like models did not contain many property terms.

Summary

- Clustering was most human-like for the Abstract-Concrete Noun Discrimination task, using our optimized parameter set.
- Our models did not accurately predict categorization on the other tasks.

Future Directions

- There may be a different parameter set for our model that will produce better clusters.
- New project: search through parameter space again, and look for the best parameter settings for these tasks.
- Question: What does this say about model’s generalizability?

Dank u wel.

Danke schön.

- This work was supported by the National Science and Engineering Research Council of Canada.
- Thanks to TAPoR, AICT and Westgrid for their support.
- Geoff Hollis and Emilio Gagliardi also contributed time and effort to this project.

Contact Info

- cyrus.shaoul@ualberta.ca
- chrisw@ualberta.ca
- http://www.psych.ualberta.ca/~westburylab/

- All data for clustering analysis is available.
- Please contact me for the open source release of HiDEx.