

# A corpus-based method for estimating word meanings in semantic space

Russell Richie<sup>1</sup>, Stefan Kaufmann<sup>2</sup>, & Whitney Tabor<sup>1</sup>

<sup>1</sup> University of Connecticut, Department of Psychology, Cognitive Science Program, Language Plasticity IGERT

<sup>2</sup> University of Connecticut, Department of Linguistics, Cognitive Science Program



LANGUAGE PLASTICITY:  
GENES, BRAIN,  
COGNITION,  
COMPUTATION  
IGERT  
INTEGRATIVE GRADUATE  
EDUCATION & RESEARCH  
TRAINEESHIP



REFERENCES  
<sup>1</sup> Sagi, E., Clark, B., Kaufmann, S. (2011). In K. Allen & J. Robinson (Eds.) *Current Methods in Historical Semantics*, pgs. 161-183, De Gruyter Mouton.  
<sup>2</sup> Manin, D. Y. (2008). *Cognitive Science*, 32(7), 1075-1098.  
<sup>3</sup> Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.  
<sup>4</sup> Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66(2), 115-24.  
<sup>5</sup> Hoffman, P., Lambon Ralph, M. a, & Rogers, T. T. (2013). Semantic diversity: a measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3), 718-30.

## ACKNOWLEDGMENTS

Thanks to audiences at the 2014 UCONN Language Fest for their feedback. This work is supported by NSF IGERT 1144399 to UCONN.

## INTRODUCTION

Various accounts of word comprehension & production<sup>4</sup>, Zipf's Law<sup>2</sup>, and semantic change<sup>1</sup> rely on constructs of words' specificity and location in semantic space

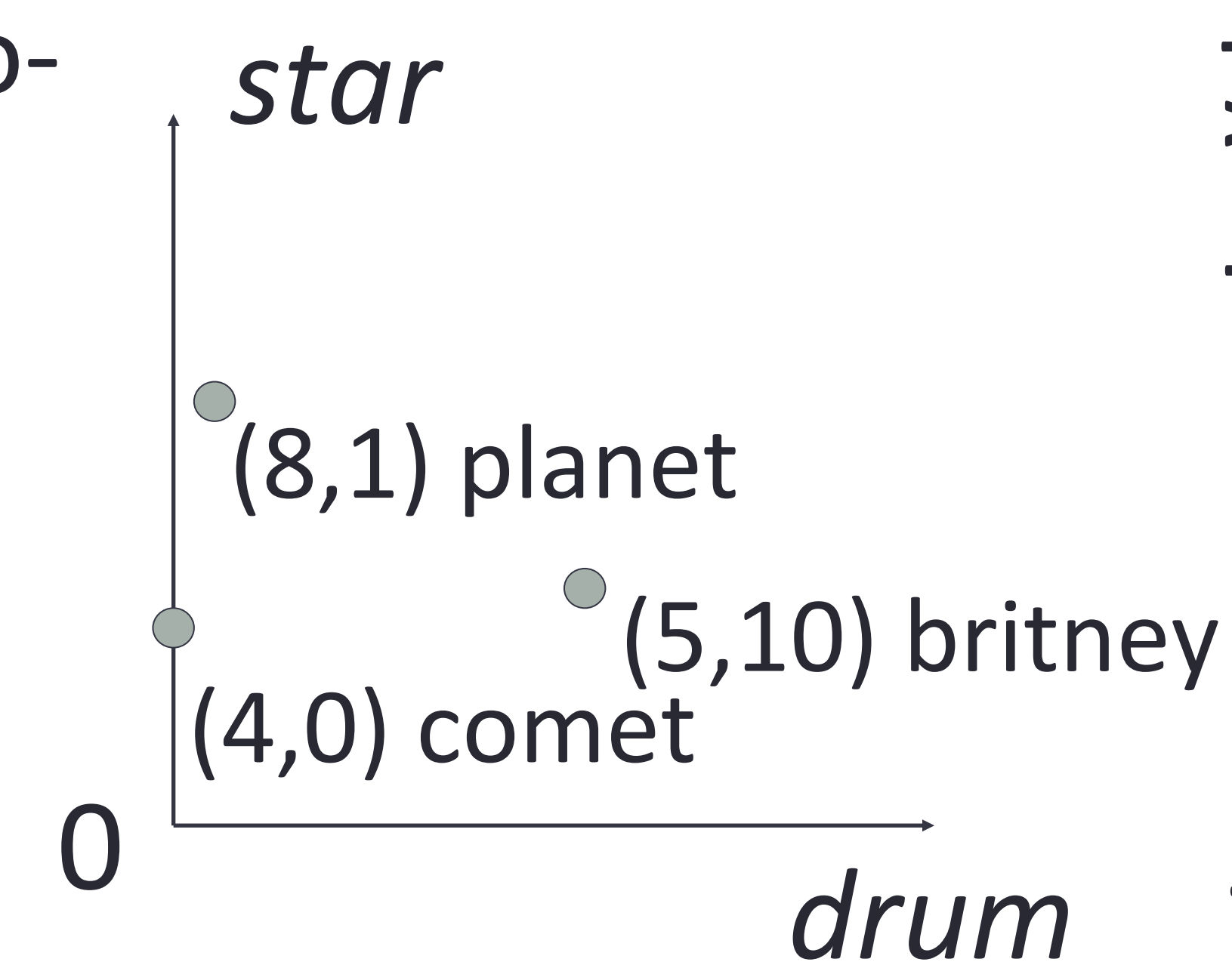
### How to measure specificity/location?

Handcrafted ontologies, e.g. dictionaries or WordNet? Too rigid.

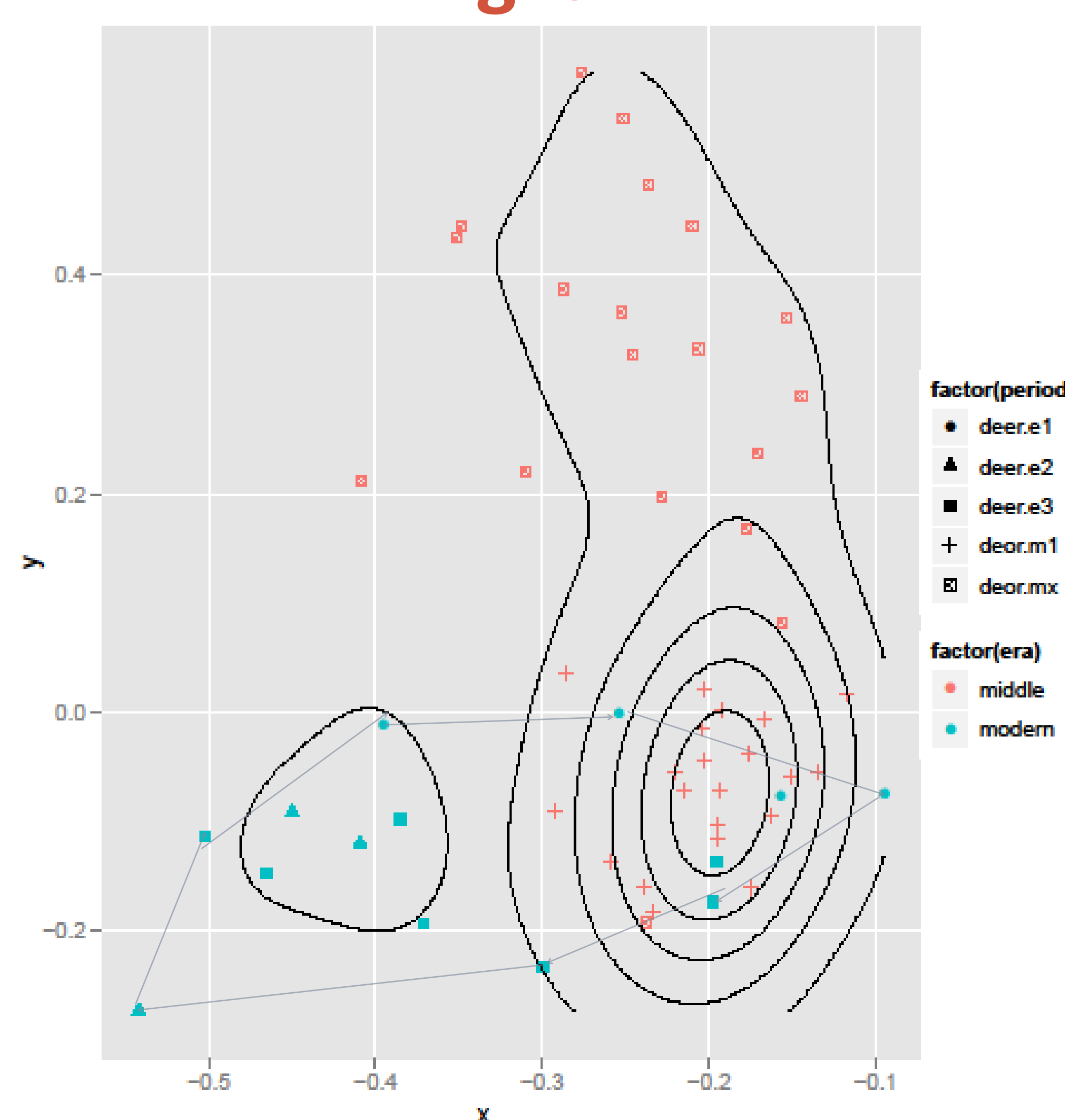
**BETTER:** word/word and word/document co-occurrence in corpora:

Contextual Diversity<sup>4</sup>, Semantic Distinctiveness<sup>4</sup>, & Semantic Diversity<sup>5</sup> → estimates of word meaning size/diversity, **not location**

	star	drum
planet	8	1
comet	4	0
britney	5	10



Low-dim view of 'deer' in early & modern English<sup>1</sup>

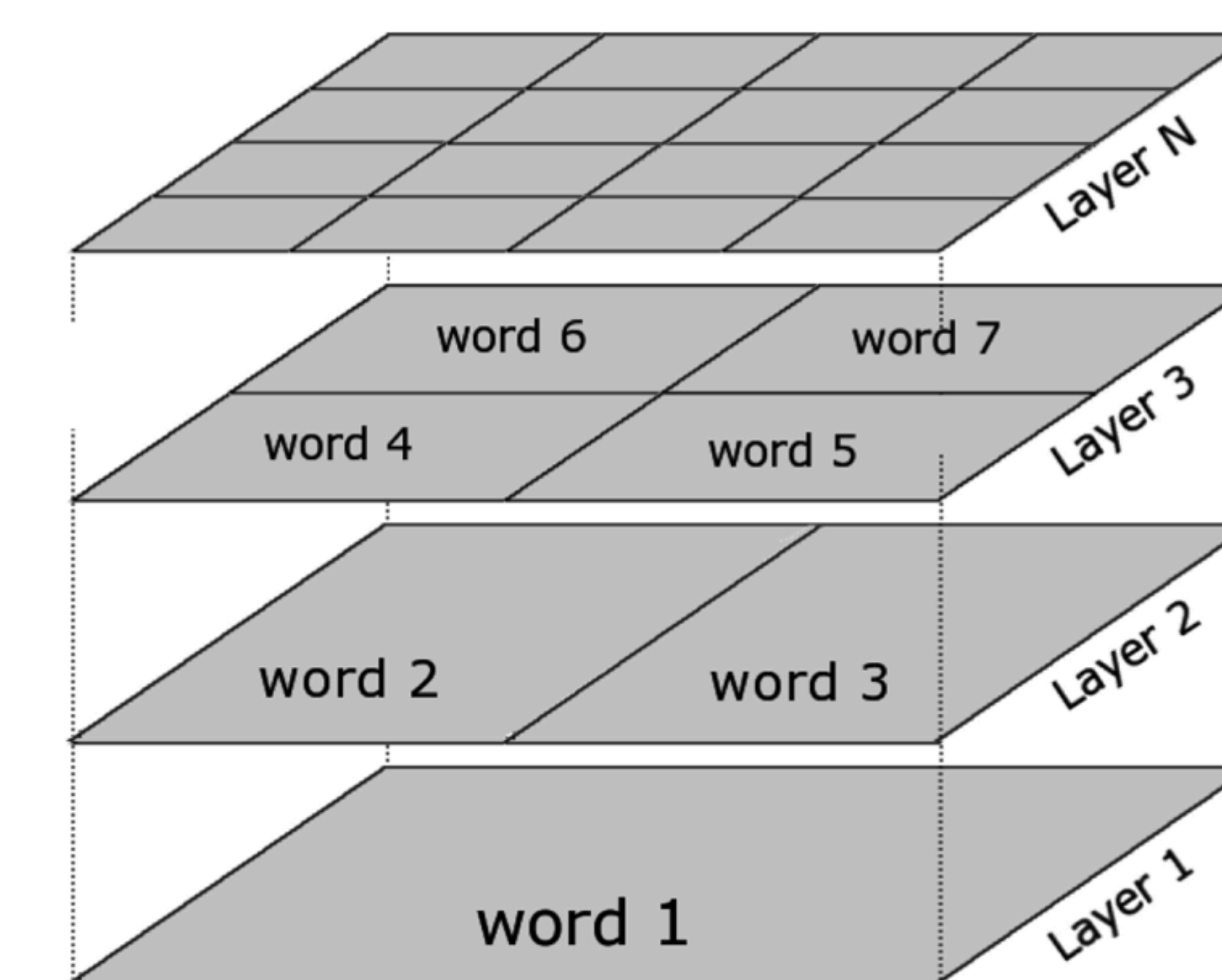
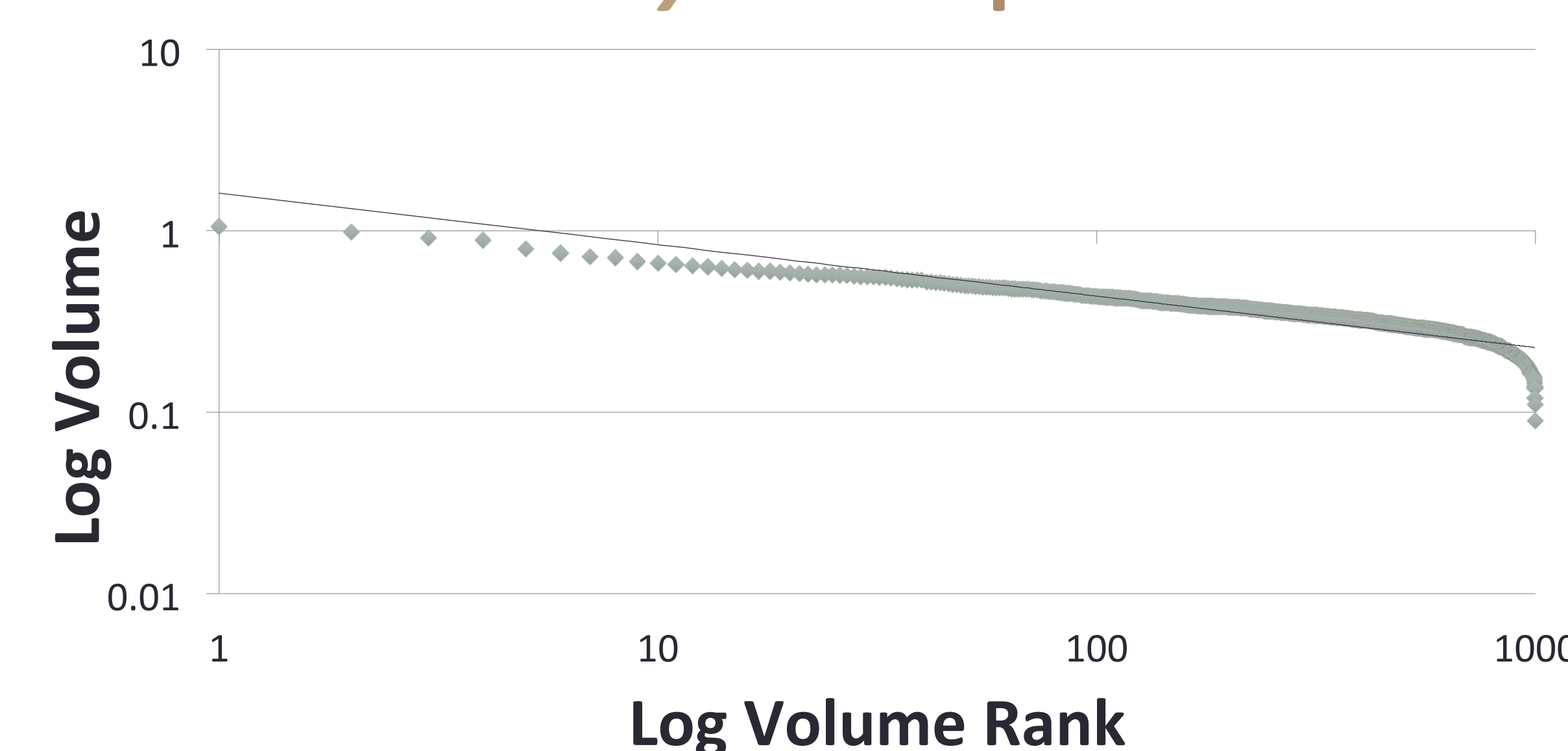


## APPLICATION – ZIPF'S LAW

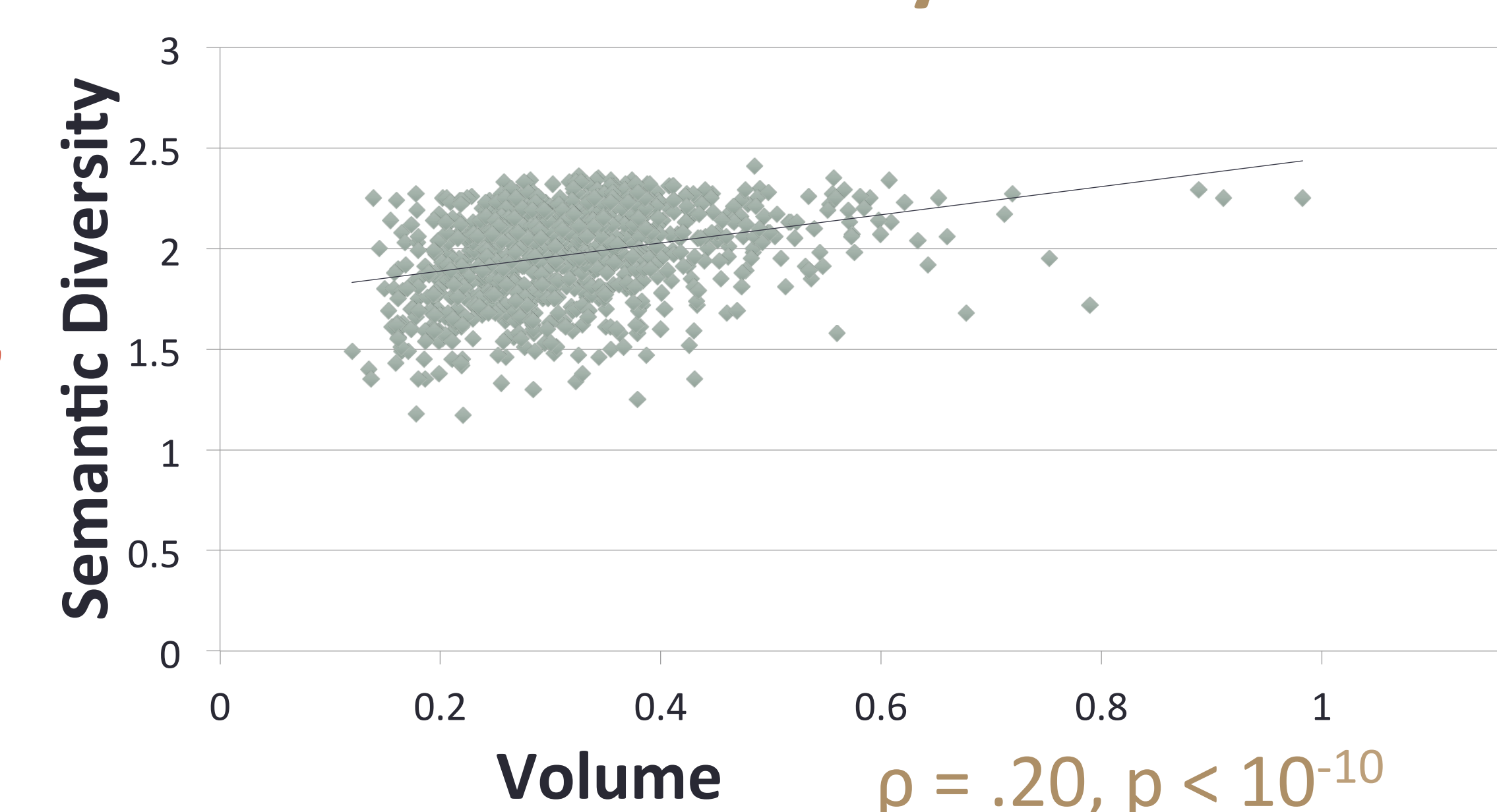
Zipf's Law: word frequency  $\propto^{-1}$  frequency rank; One explanation<sup>2</sup>:

1) Word meanings broaden/narrow to avoid excessive synonymy, yielding 'Zipfian covering'. 2) Word meaning  $\propto$  word frequency

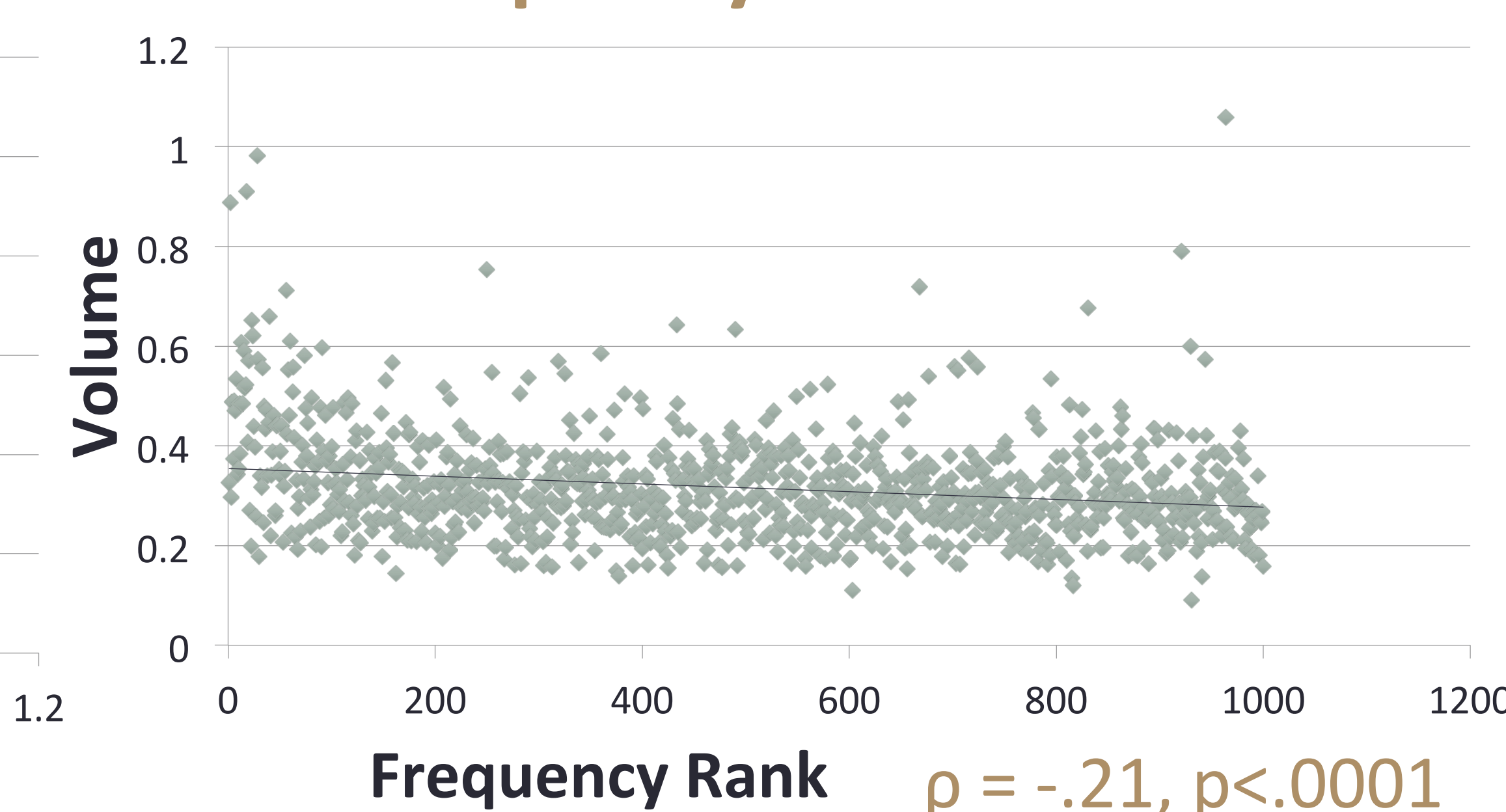
Volumes may follow power law



Semantic Diversity  $\propto$  Volume



Frequency  $\propto$  Volume



## OUR METHOD

- 1) Word-word cooccurrences → vector representing word **type** general meaning (*type vector*)
- 2) Sum of **type** vectors of words around a word **token** → **context vector** → meaning of that token
- 3) **Context vectors** of a given **type** → cloud in semantic space
- 4) Volume of convex hull<sup>3</sup> of **context vectors** → meaning specificity *and* location

## CONCLUSION

Our method replicates known findings about meaning specificity, has promise for novel applications. Next Steps:

1. Test 2nd premise of Manin (2008), based on **locations**
2. Better measure than convex hulls of meaning specificity (e.g., kernel density estimation)
3. Further comparison with existing measures
4. Psycholinguistic applications of our method<sup>5</sup>?