The ‘Worthy of Attention’ Collostruction: Frequency, synonymy, and learnability

Gustave Hahn-Powell, Robert Nelson, & Elliot Patton
University of Alabama

Introduction

- Expansion of Hahn-Powell & Patton (2009), which asked:
  - How do people learn how to use synonymous adjectives properly?
  - How do we know which one(s) can be used at a particular time?
- Current study also examines learnability:
  - Can an Artificial Neural Network (ANN) tell us anything about how humans learn to use near-synonyms correctly?
  - Can it give us insight into the way in which meaning becomes grammatically encoded?

Previous research

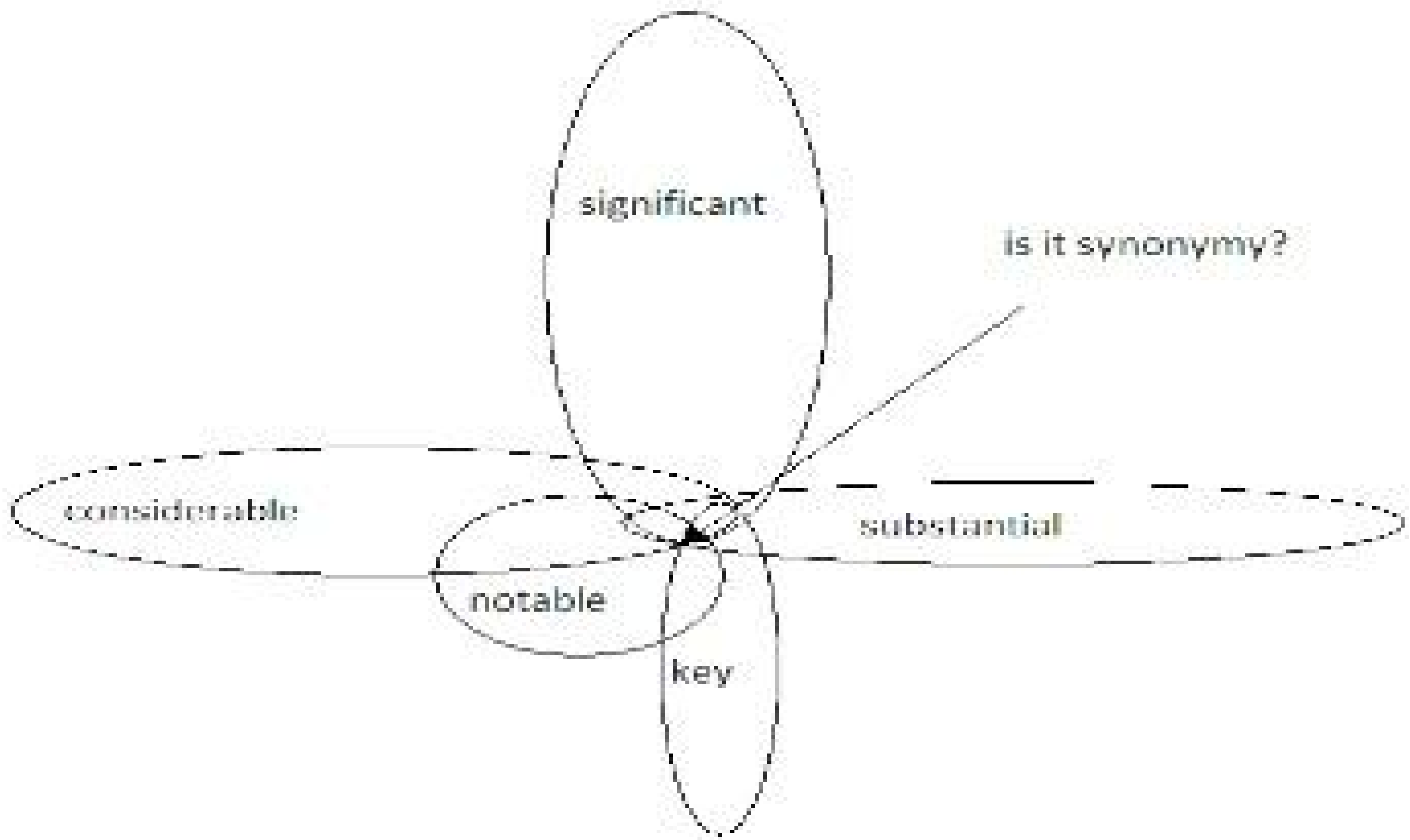
- D. Liu (2010, in press in 2008)
  - Corpus-driven
  - Five synonymous adj: *main, major, chief, primary, principal*
  - Six noun groups
- Edmonds & Hirst (2002)
  - coined the term "plesionymy" (***near*** synonymy): overlap, but with peripheral meanings
  - attempted to take a computational approach to synonymy

Hahn-Powell & Patton (2009)

- Similar in intention to Liu (2010), but opted for emphatic pre-modifiers with ostensibly closer meaning: *significant, substantial, considerable,* and *notable*
  - Included *key* as an adjectival with some overlap
  - Goals:
    - Investigate nature of noun/adjective categorization
    - Develop a methodology for investigating near-synonyms
    - Help EAP students to distinguish synonyms
    - Assess meaning in real-world context

Central meaning

Figure 1: Venn Diagram illustration of synonymy

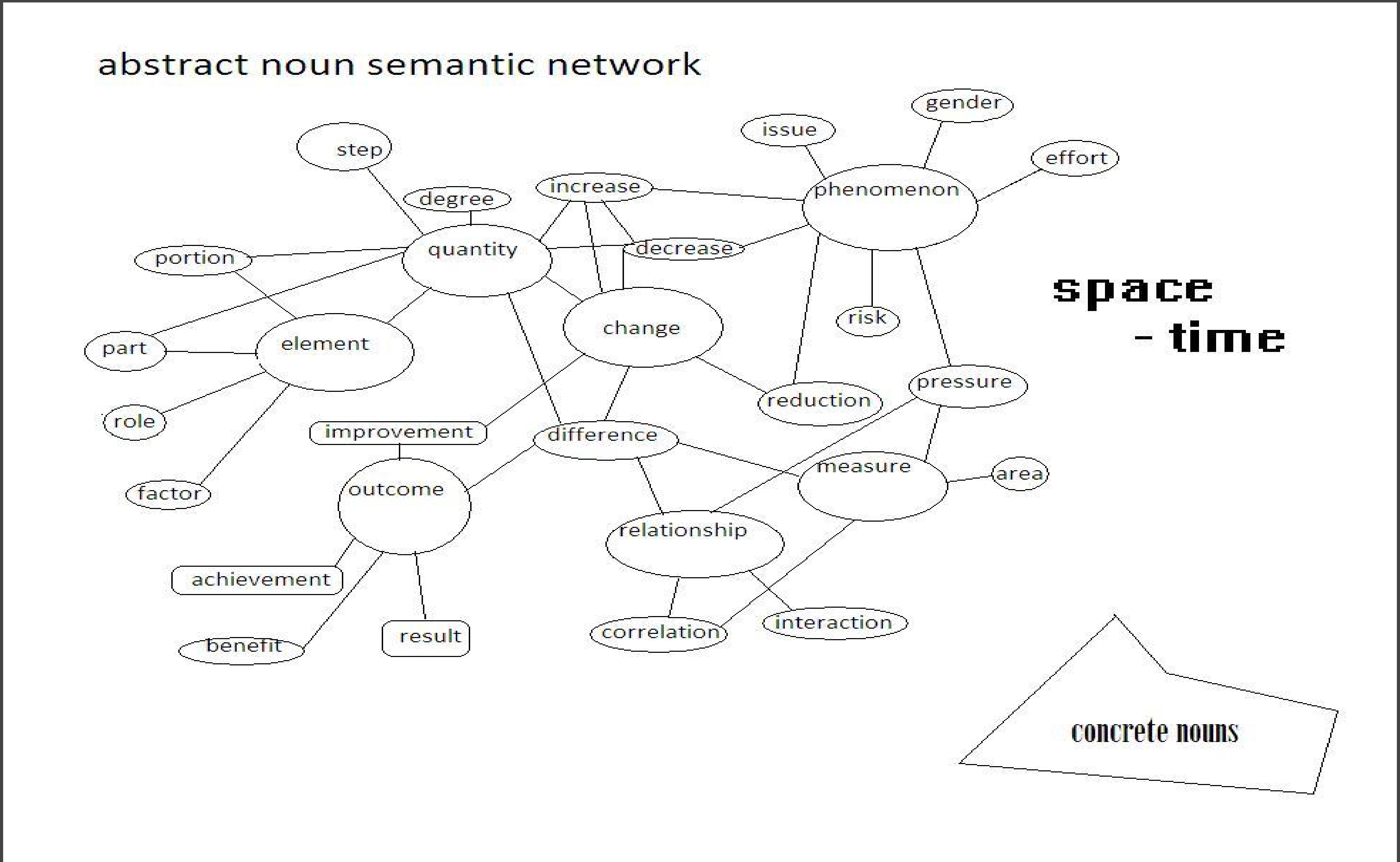


Hahn-Powell & Patton (2009)

Method

- Analyzed randomly sampled tokens (COCA) for all adjectives
- Looked at 50 most common nouns paired with each adjective and created a master list
  - Master list: 84 nouns
- Created a "behavioral profile" for each word based on noun pairings and overlap with other adjectives (synonymy)

Abstract noun semantic network



Significant

- Paired with every noun on the list (most freq: ***difference***)
  - (Most frequently paired with *relationship, quantity* types)
- Heavy use as a mathematical (statistical) term in academic journals
- Most versatile of five adjectives, and most common (>58,000 tokens)
- Very little occurrence with concrete nouns

Substantial

- Much more limited than *significant*, much less used
- Most frequent noun: ***number***
- (Most frequent pairing: *element* types)
    - Very little occurrence with concrete nouns

Considerable

- Showed the least amount of overlap on the whole
- Most frequent pairing: ***amount***
  - (*phenomenon* types)
- Very little occurrence with concrete nouns

Notable

- Overall, least commonly occurring
- Most common pairings: ***exception***
  - (*phenomenon* and *element* types)
- Some occurrence with concrete nouns

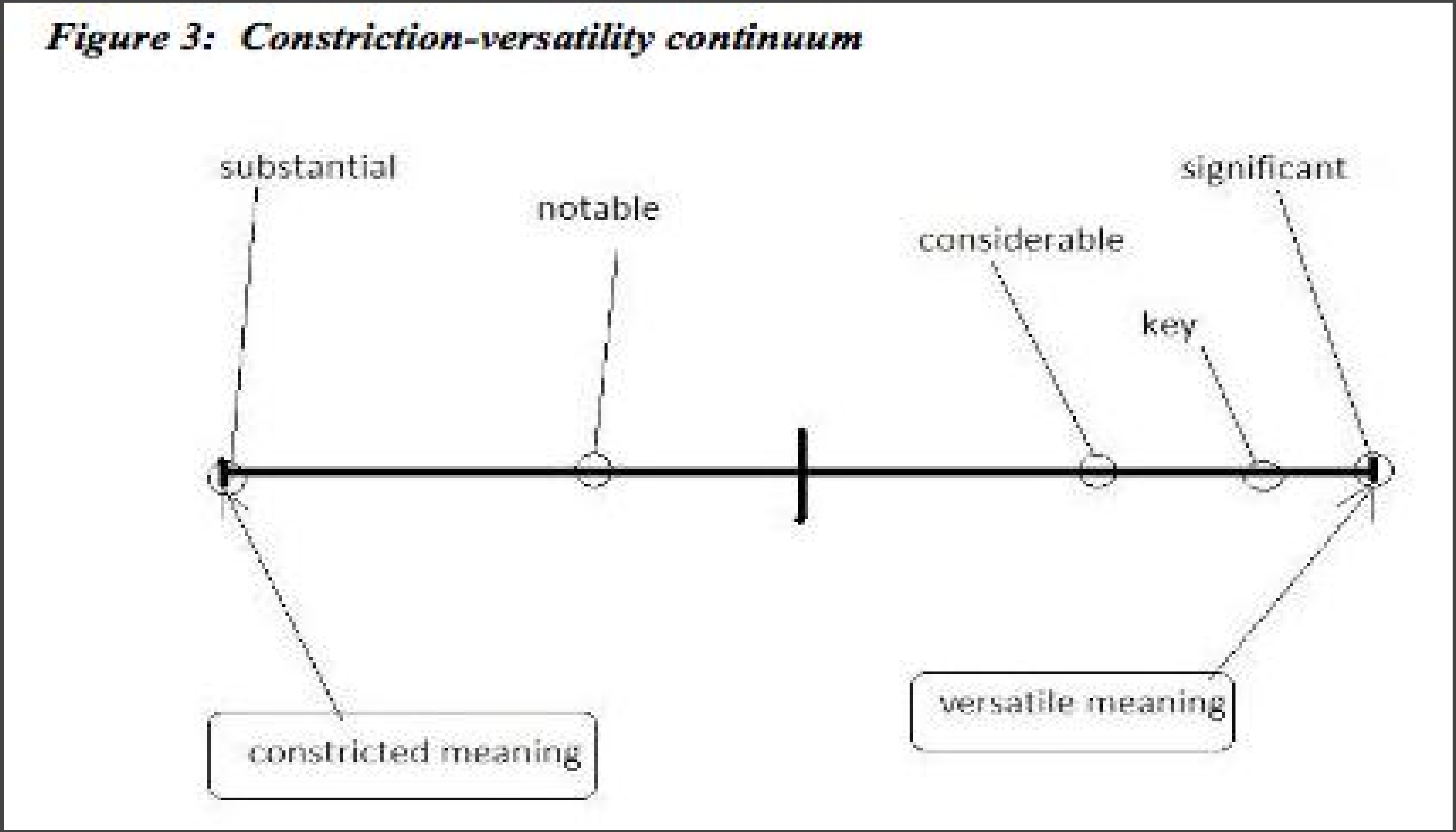
Key

- Very common as an adjective
  - OED does not acknowledge it as such
- Heavy occurrence with concrete nouns
  - Expected
- Most frequent pairing: ***role***
  - (*element* types)

Frequency Table

NOUN	Sig.	Sub.	Con.	Not.	Key	TOTAL
difference	5004	171	31	73	205	5484
role	787	46	11	8	1161	2013
number	1097	535	170	4	43	1849
change	1188	212	48	28	150	1626
issue	193	16	1	3	1063	1276
factor	446	6	2	5	737	1196
issue	421	160	60	2	464	1107
element	78	12	2	5	970	1067
effect	806	101	16	10	3	936
contribution	624	95	8	22	30	779

Meaning-Usage Continuum (MUC)



Approaching the data from a learnability perspective

Two features of this data are promising to a neural network approach to the data's learnability

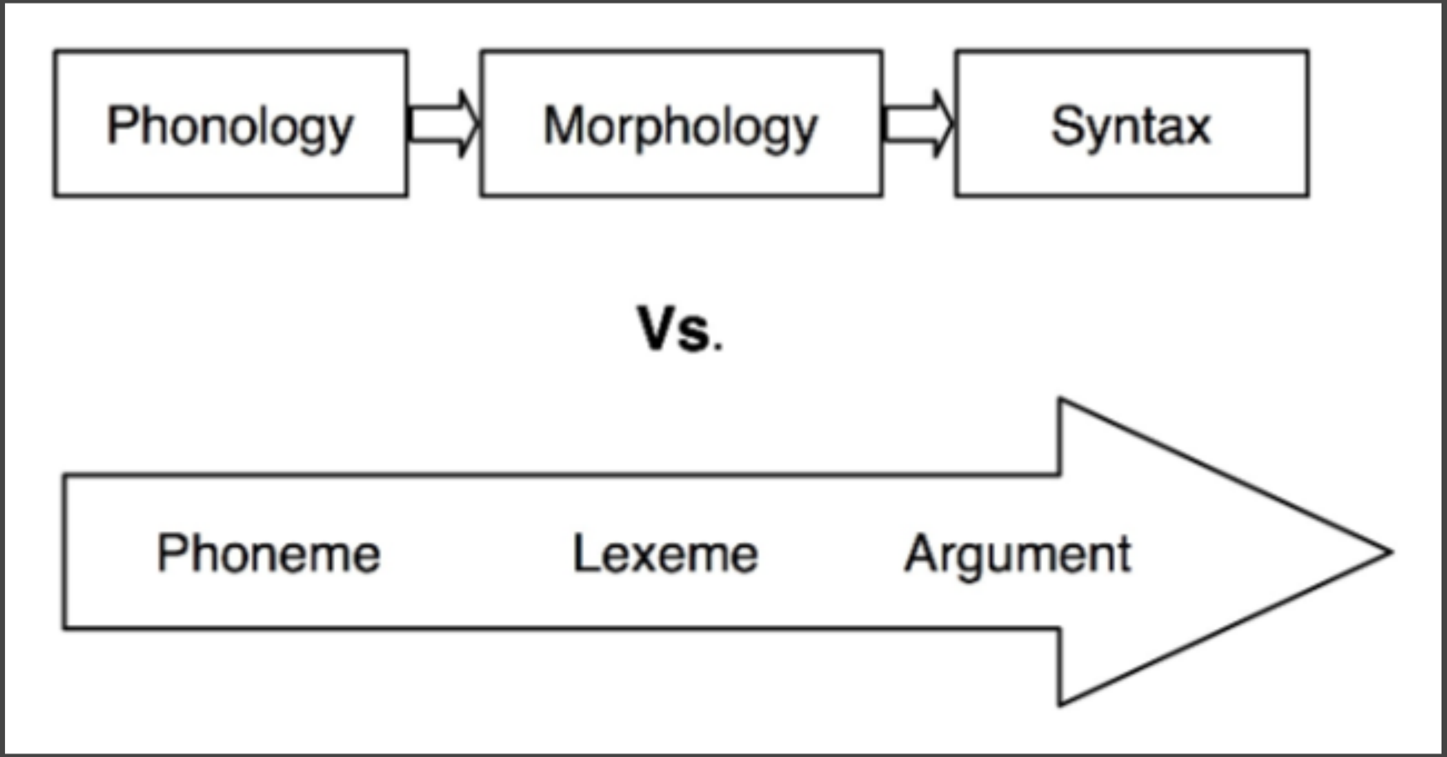
1. it involves skewed co-occurrences
2. it involves categories

This makes this data set a good place to ask: 'how might we gain insight into the processes through which meaning becomes form (i.e., grammaticalization)?'

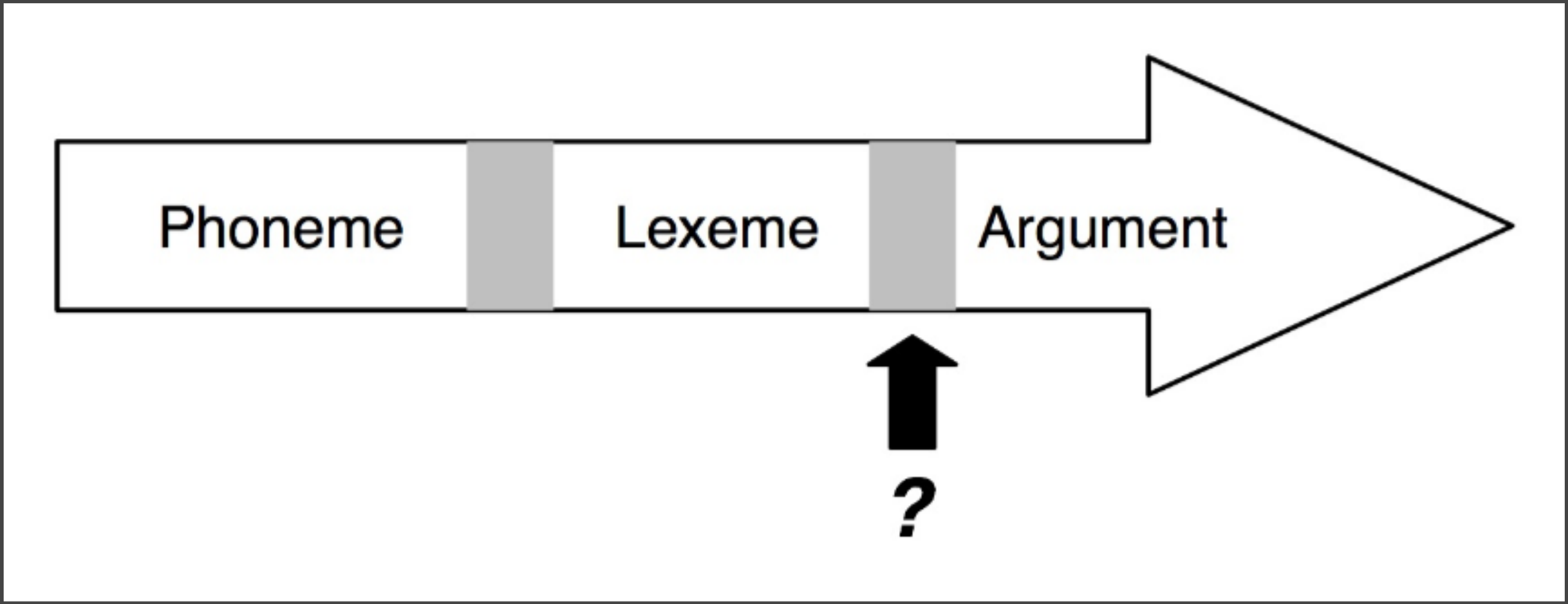
A Construction grammar approach

- Asking how a meaning becomes a form assumes that meaning can become form
- An appropriate linguistic framework is thus Construction Grammar
- A construction is a conventionalized form-meaning pair at any level of schematicity:
  - Lexical
  - Argument
    - Goldberg, 1995, 2006

Domains of knowledge vs. Continuum of Schematicity



Gray areas of linguistic knowledge



Collostruction

- A collostruction is a particular construction that *attracts* or *repels* lexemes
- Collostruct:
  - The more-or-less set phrase that attracts the word
- Collexeme(s):
  - The words that are ‘attracted’ to the collostruct

For example:

- The into-causative construction:

  (Stefanowitsch & Gries, 2003)

  - S_{agent} V $O_{patient/agent}$ into-A_{gerund} resulting action
  - She tricked him into going
  - They conned us into buying it

- Attracts words like:
  - Trick, blackmail, fool, coerce, mislead, goad, shame, etc.

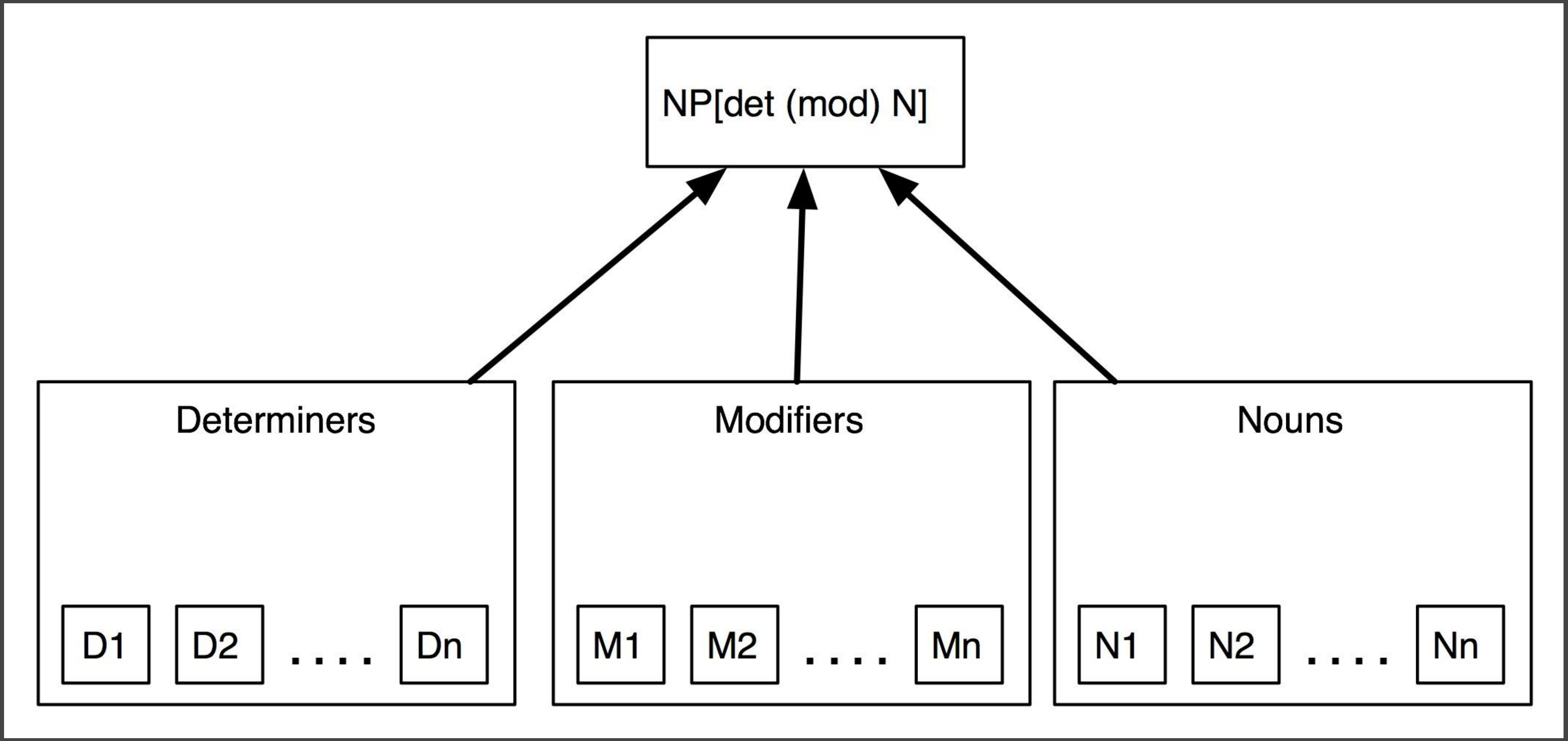
Collostructions can have lexical collostructs

- For example, the verb ‘cause’ attracts:
  - Problem, damage, harm, havoc, distress, injury, stress, change, wear, etc.
- Note that all collexemes are ‘semantically prosodic’ with each other and ‘semantically compatible’ with the collostruct.
- Lexemes that are not prosodic and compatible are repelled

Learning constructions

- Goldberg (2006) argues that argument constructions acquire their meaning from highly correlated verbs
  - So that the cause-receive construction. . .
    - S V O_i O_d
  - Has acquired the meaning ‘cause to receive’ from its frequent pairing with the verb ‘give’
    - This is a reasonable position from both statistical learning and neurophysiologic perspectives

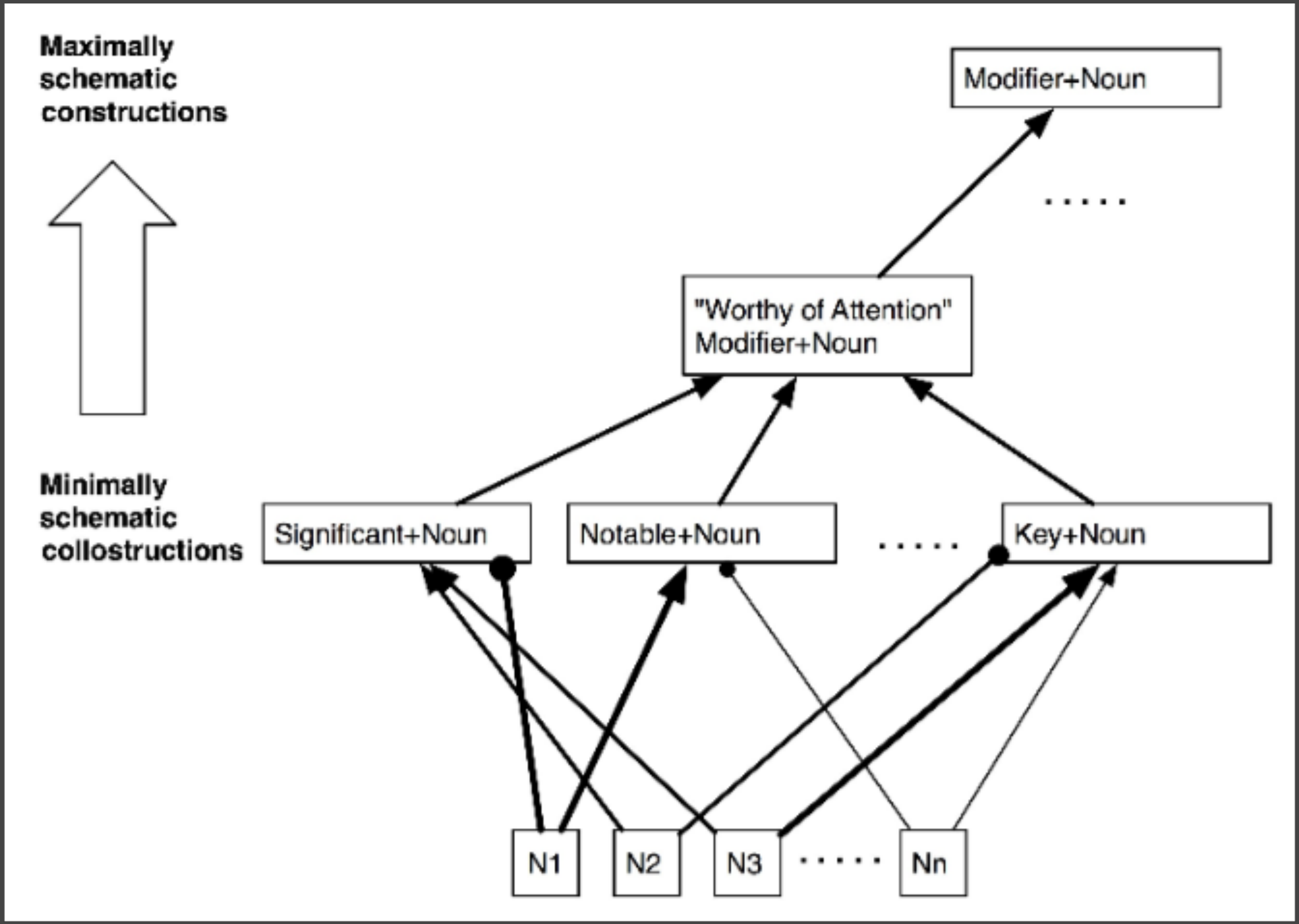
A strict hierarchy of ModN constructions



Evidence against a strict hierarchy

- Adjective ordering in English
  - Unlike many languages, English has preferential adjective positioning
  - so:
    - big red dog
  - not:
    - red big dog
- ESL errors
  - Adjective-noun order errors are not all or none
    - (Lightbown & Spada, 1990)

An open hierarchy



A more open hierarchy implies. . .

1. Statistical properties of the input are important

2. the structure of the learning system (i.e., the cortex) is important

- network
- hierarchy

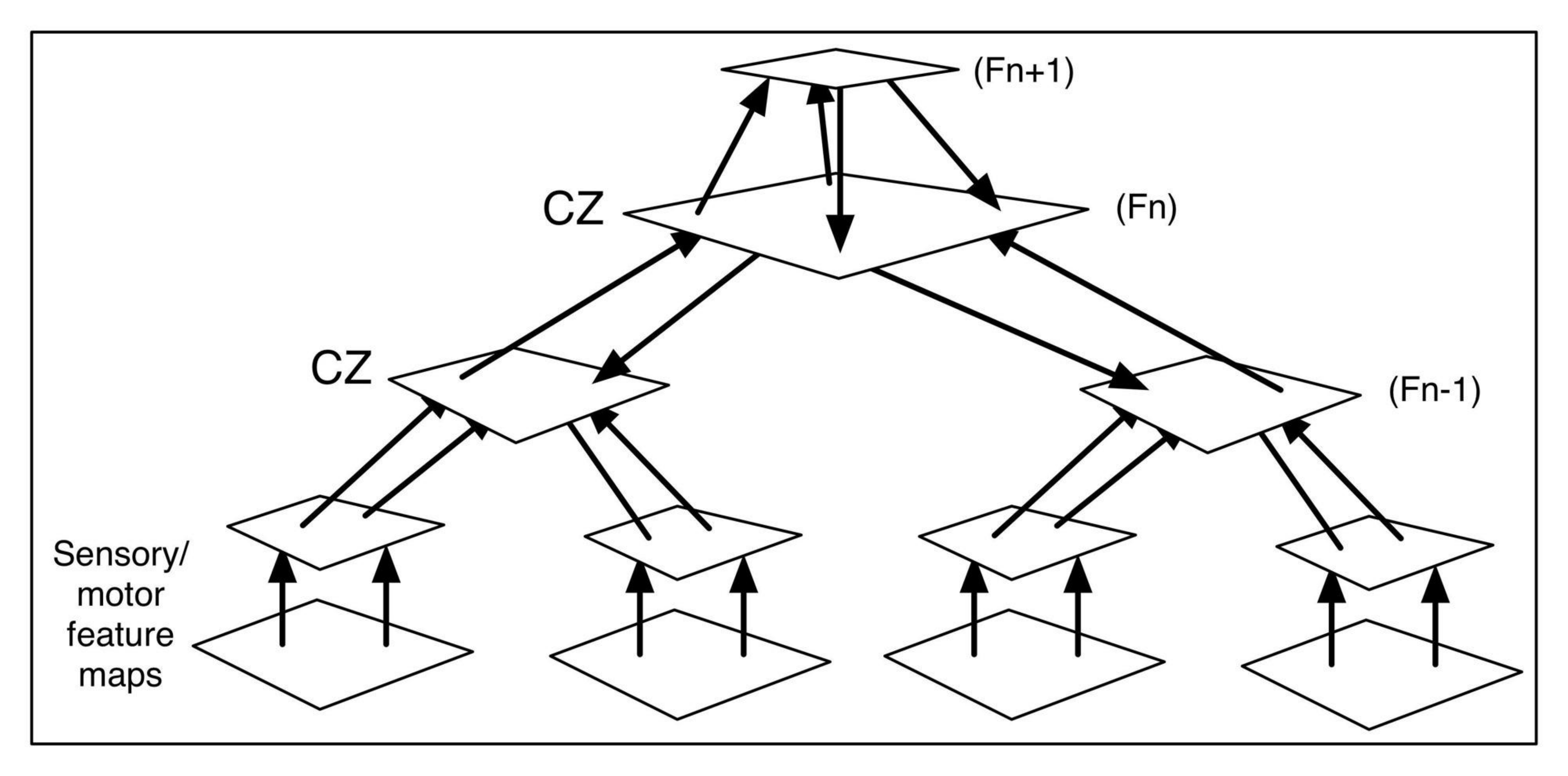
Statistics and learning

- Children use the statistics of sound stream to locate both word and phrase boundaries (Saffran, 2003)
- We integrate multiple probabilistic cues in the online comprehension of language and the learning of complex argument structures (Chritiansen & Chater, 2001)

Cortical memory networks

- Memory networks of the posterior association cortex (where lexical memories are mostly located) show rich connectivity with other memory networks.
  - Related memories tend to be partially superposed over the same connections
  - Unrelated memory networks are not superposed, and may have no connections, or inhibitory connections
    - (Fuster, 1997, 2003)

Cortical lamina



Two grammars

- Flat frequency
- All words equiprobable

- Biased frequency
- Word frequencies match those found in COCA by Patton & Hahn-Powell, 2009

Creating a Finite State Grammar (FSG)

- Built the two Finite State Grammars in Mathematica
    - FSGs were used to generate two large corpora of five types of argument structures, some of which contained the Mod+N combination we are interested in ('Crit NP')
    - Example structures:
        1. s[np[det N] vp[V np[det N]]]
        2. s[np[det N] vp[V np[Crit NP]]]
        3. s[np[det N comp['that' np[det N]]] vp[V np[det N]]]
        4. s[np[det N] vp[Vdt np1[det N] pp['to' np2[det N]]]]
        5. s[np[det N] vp[Vlink Mod]]

Coding Procedure for FSG

- For Simple Recurrent Network (SRN), opted to code everything in a binary pattern (after Elman, 1990)
- Main issue: Provide semantic categories for the network that allowed discrimination
  - Goal: Training, allowing network to ***discover*** answers
- Majority of semantic categories derived from Lakoff's (1987) definitions of image schema

Examples of Image Schema

- Source
- Path
- End-of-path
- Cycle
- Part-whole
- Fictive motion
- Link
- Implied verticality

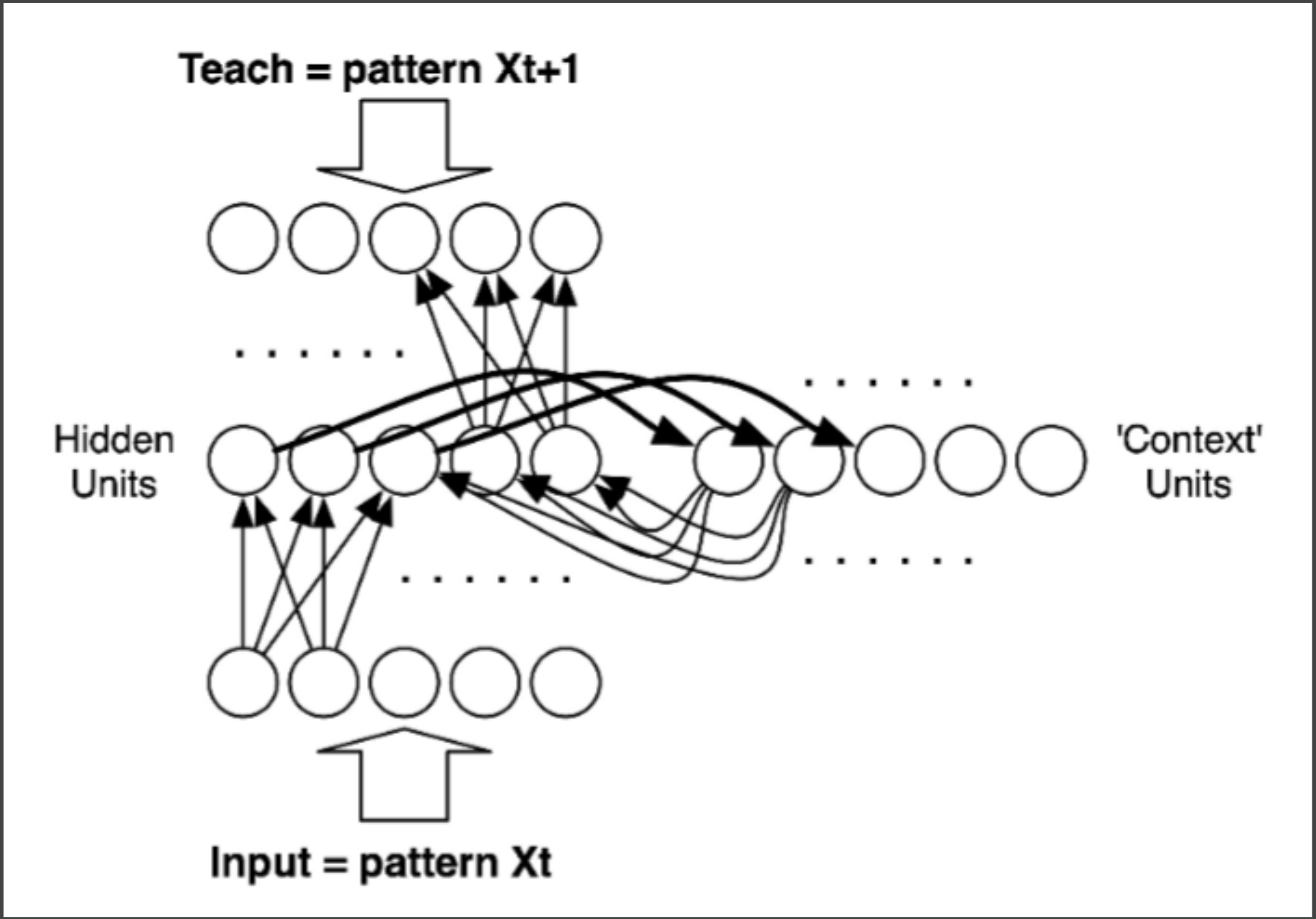
Example: **INCREASE:** fictive motion = 1; source = 0; path = 1; end-of-path = 0; force = 0; etc

Procedure

- Also coded for:
  - nouns that could be both concrete and abstract
  - nominalizations (Halliday & Martin, 1993)
  - thematic agents (Jackendoff, 1983)
- Verbs:
  - Selected 13 verbs that paired with all five adjectives using COCA
- Used COCA to generate normalized frequencies for all words
- Neural Network Coding

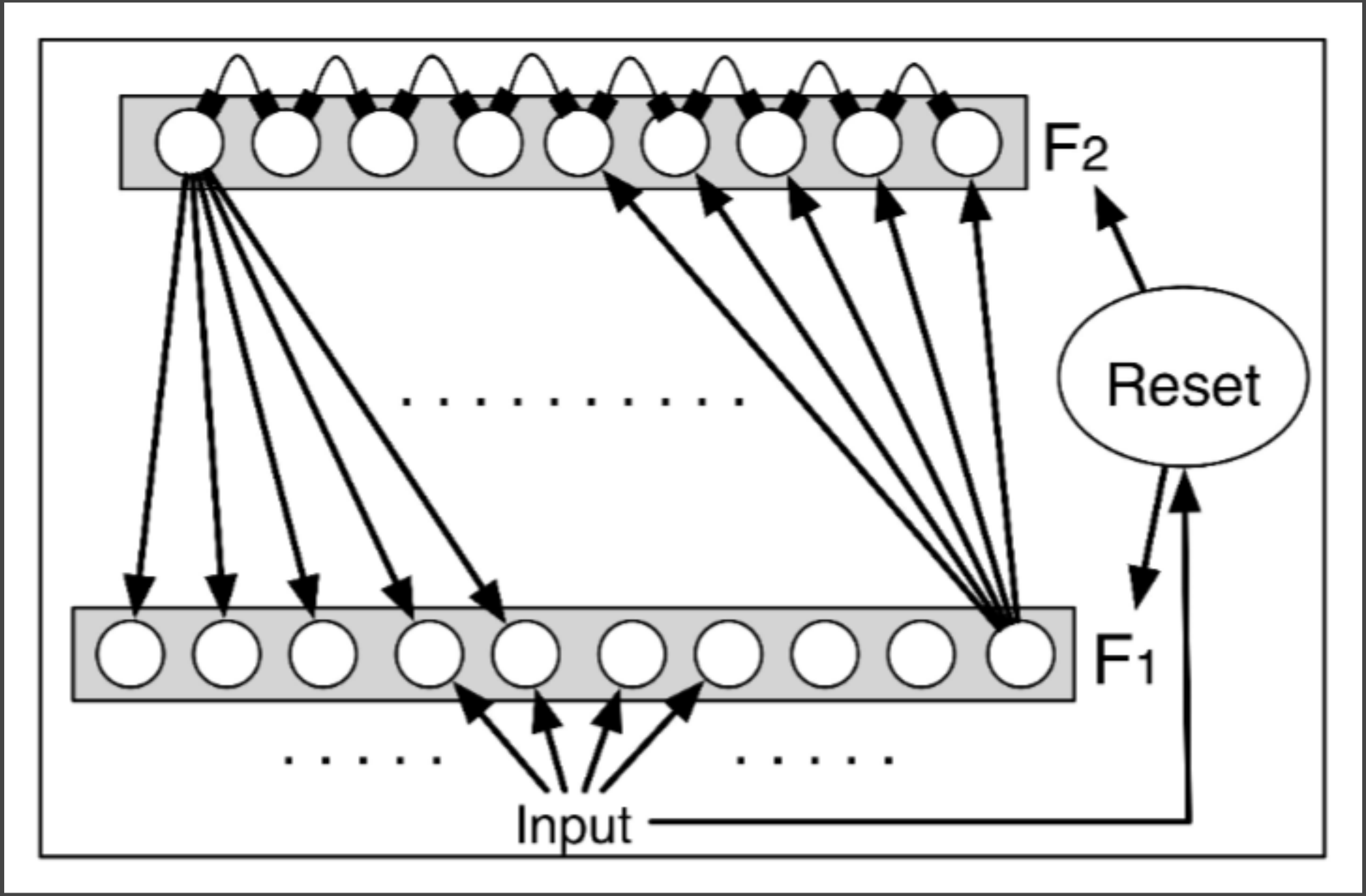
Simple Recurrent Network

Elman, 1990



Adaptive Resonance Theory

Carpenter and Grossberg, 1987

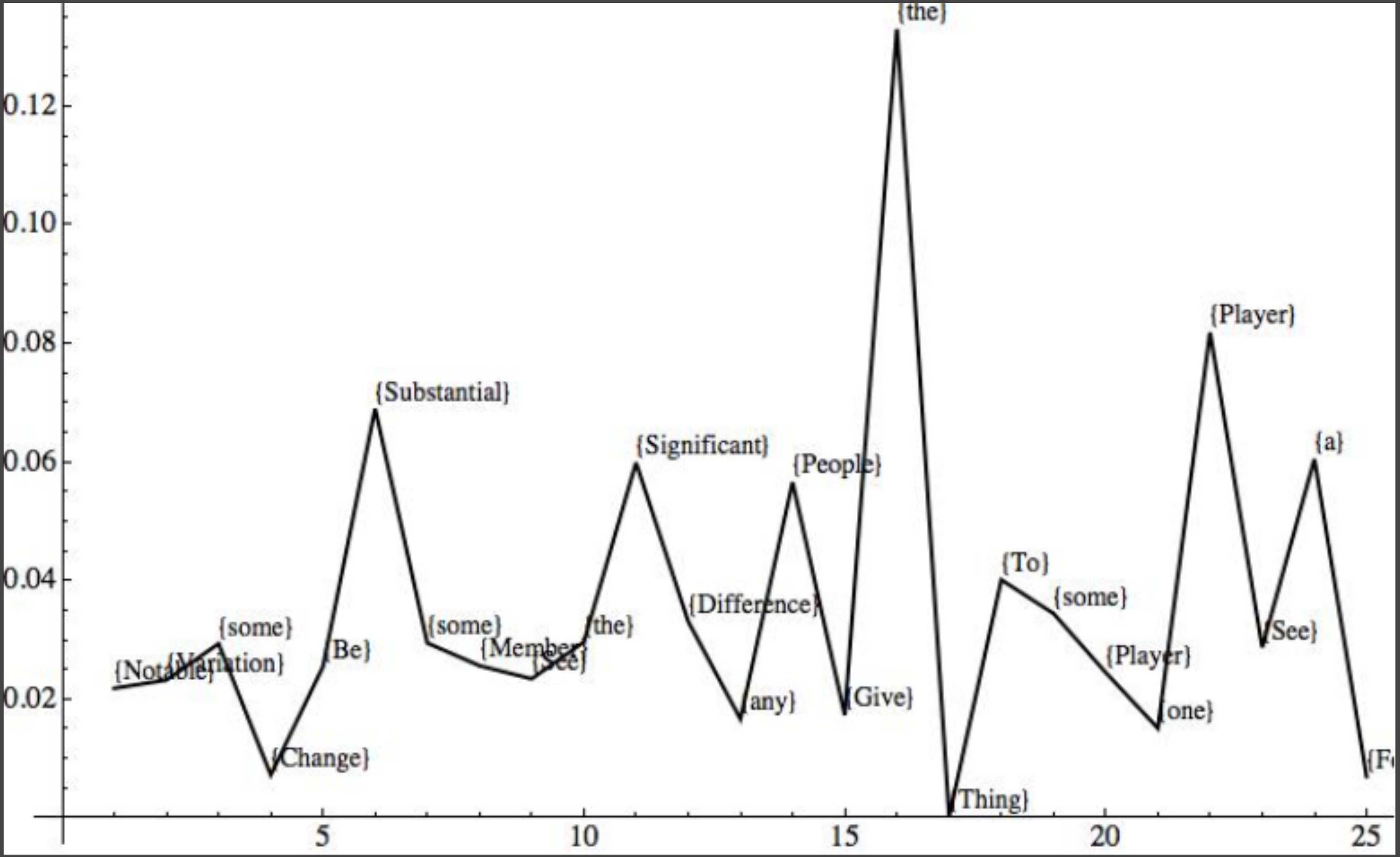


SRN results from FSG input

- SRN was able to use the generated corpus to output frequencies
  - Frequencies were then used as input, along with the binary data, to train the network
  - MSE for SRN: 0.03

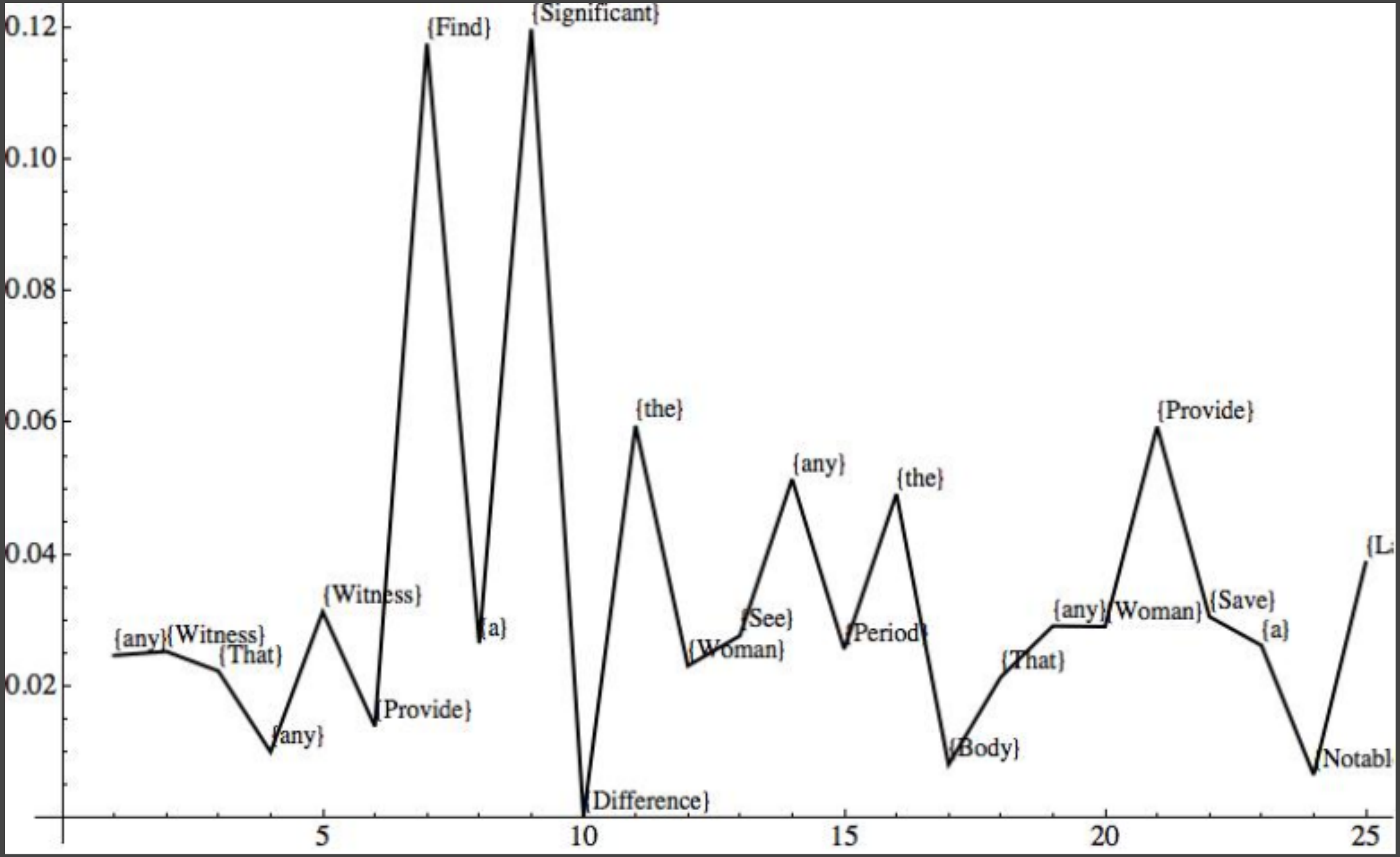
Error rates SRN:

Flat Frequency



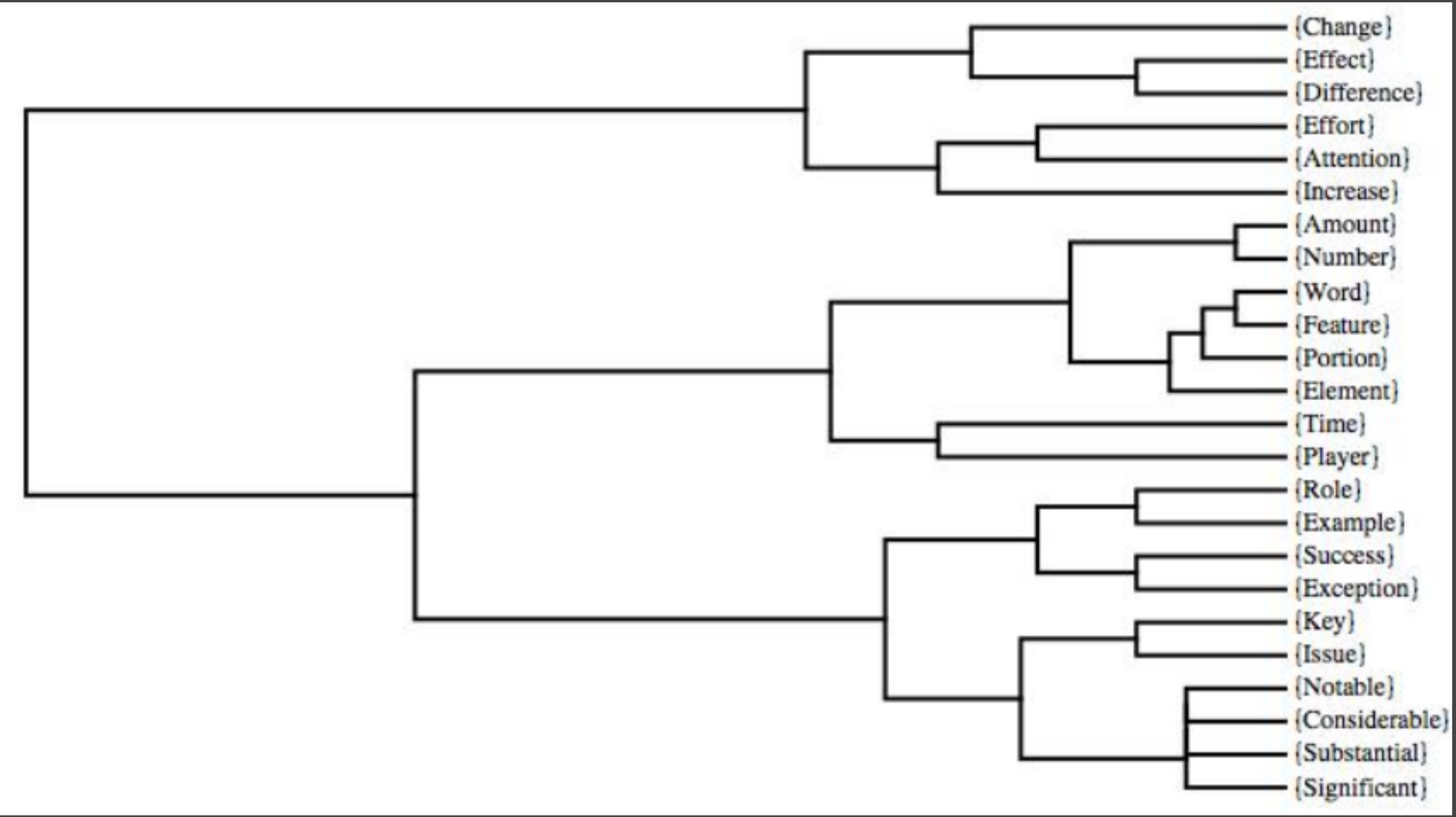
Error rates SRN

Biased Frequency



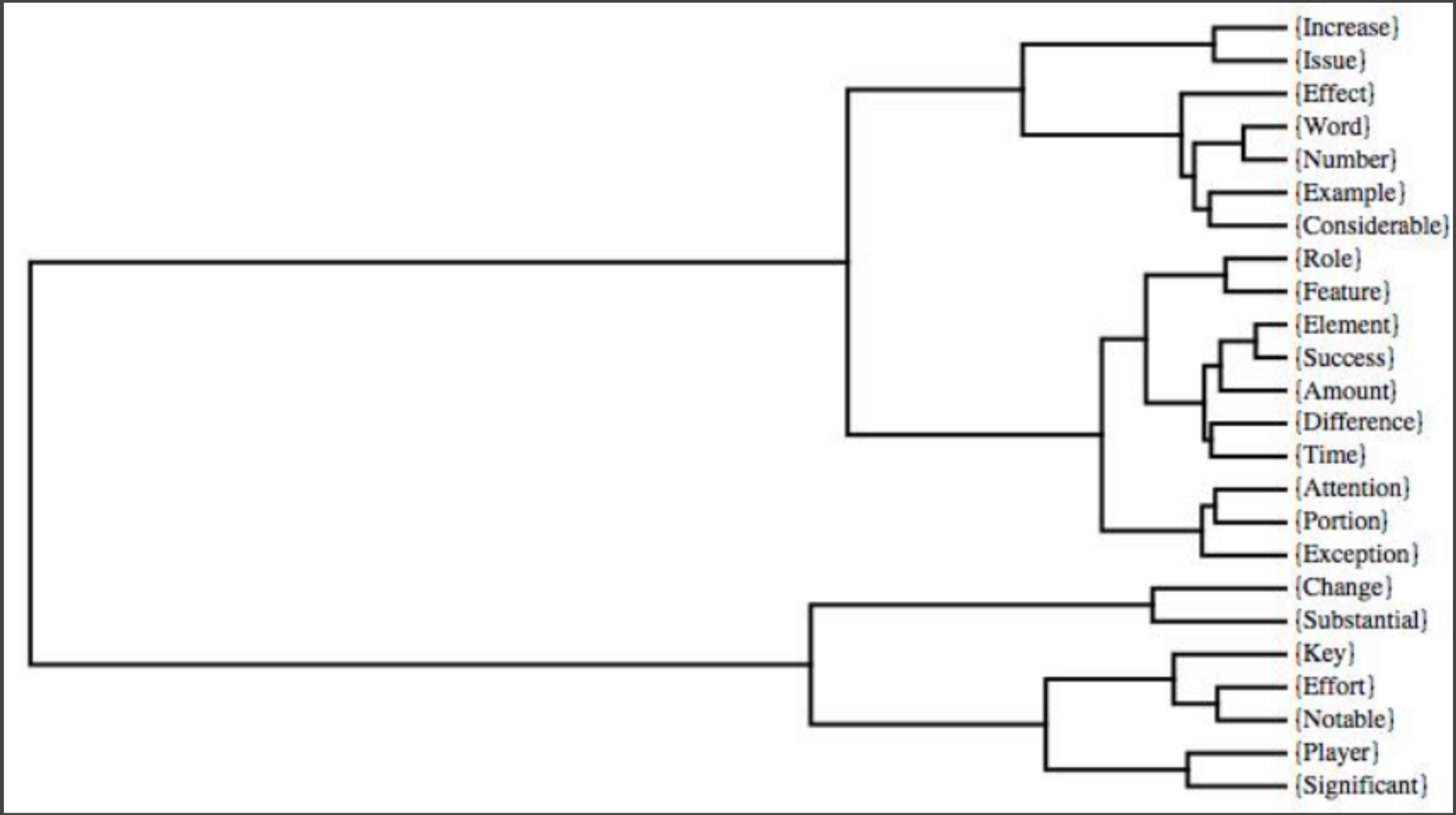
Cluster Analysis:

raw data



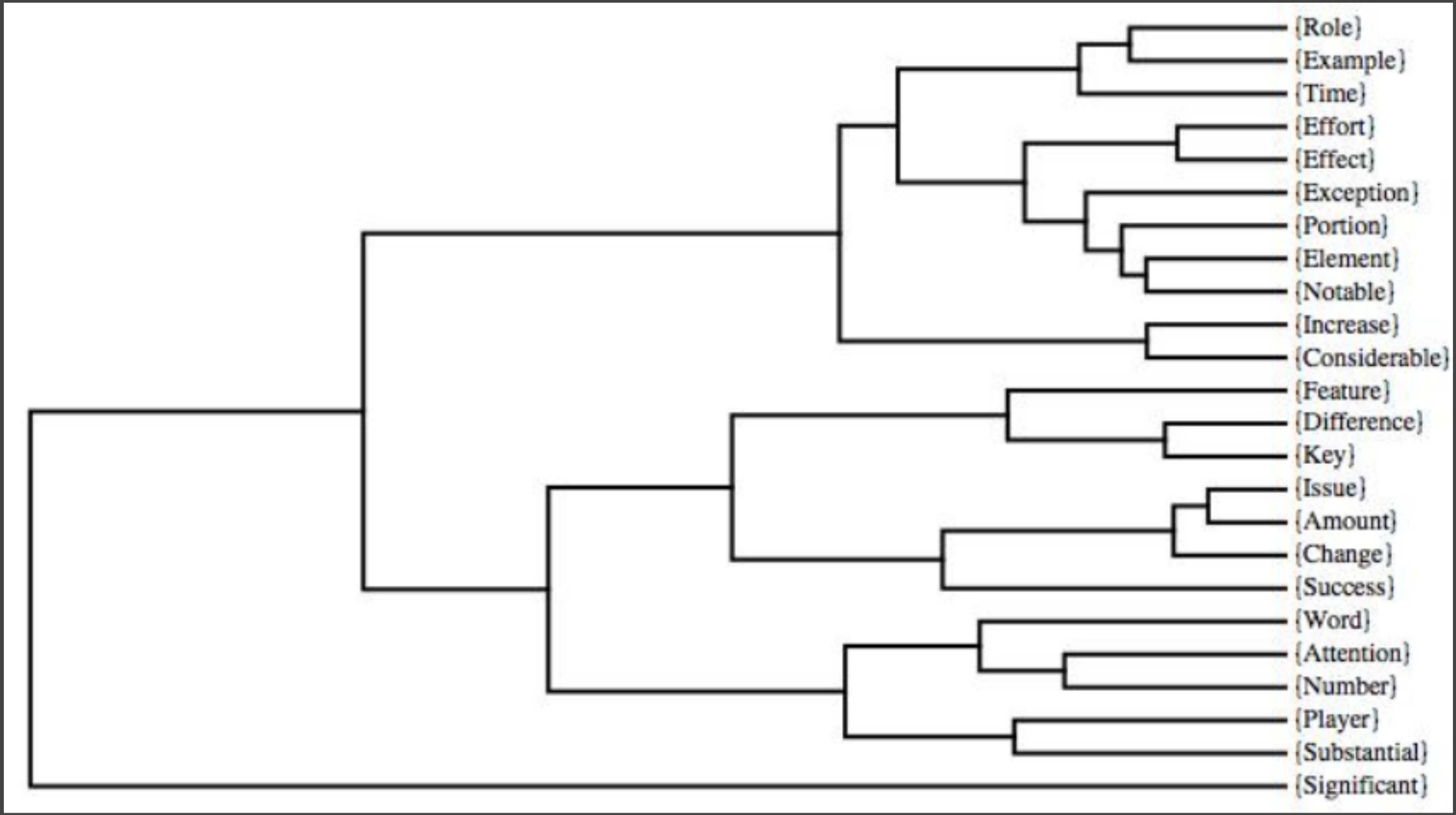
Cluster Analysis:

Flat frequency SRN HU



Cluster Analysis:

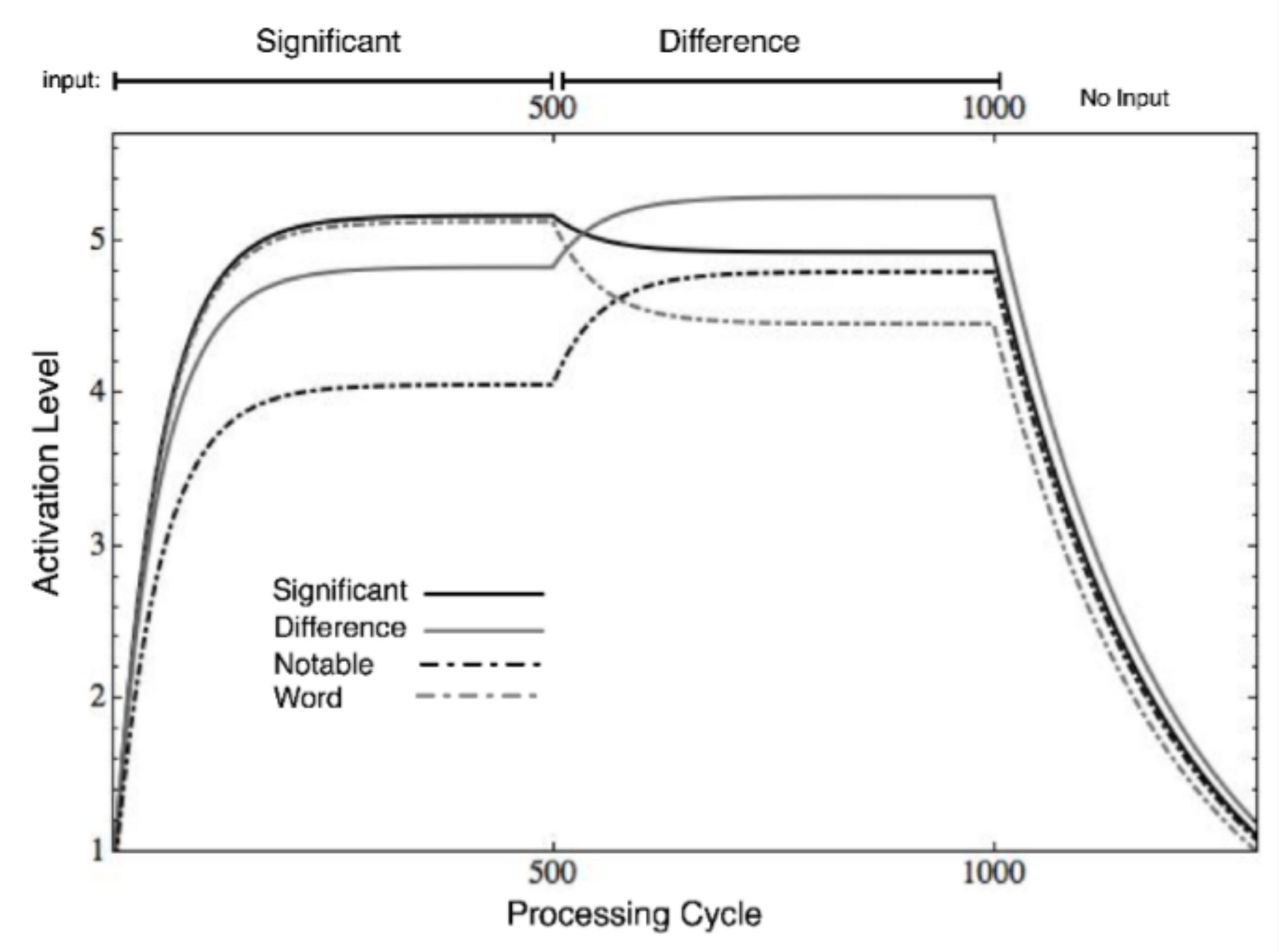
Biased frequency SRN HU



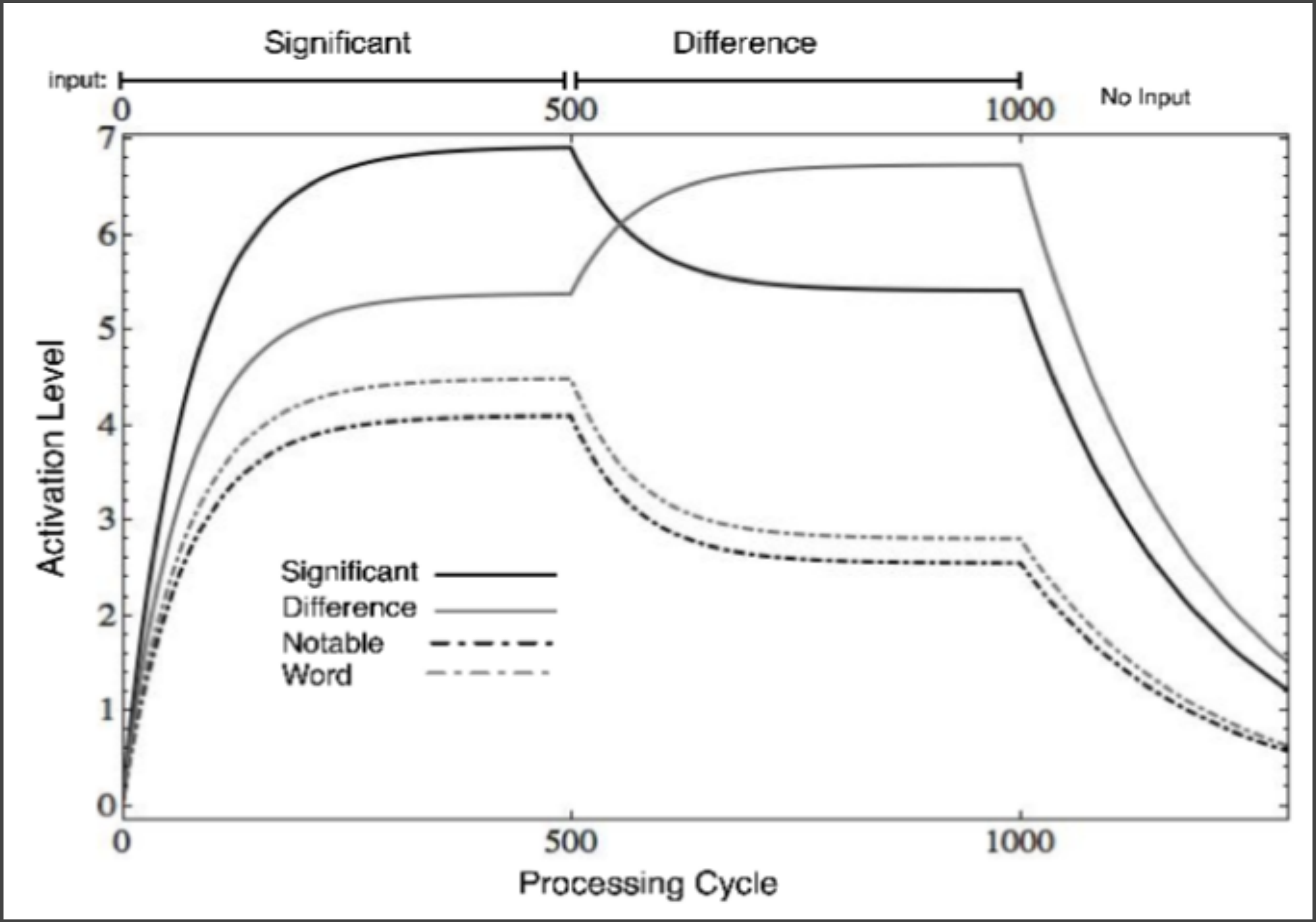
ART results

- HU activations from SRN were fed into an Adaptive Resonance Theory (ART) network
- ART networks use a competitive learning algorithm
  - We can see how lexical memory patterns repel and attract each other

ART behavior after learning from the unbiased grammar



ART behavior after learning from the biased grammar



Conclusion

These data suggest that . . .

Semantic, corpus-driven approaches can be unified with computational approaches. Why does this matter?

- This allows us to study how meaning becomes grammaticalized
- We can buttress a cognitive-semantic coding scheme with a model that emulates human memory
- We can better understand which other structures are required by knowledge of language to support those that are phenomenologically available
- All this is available in precise, measurable results