HÖGSKOLAN I BORÅS
INSTITUTIONEN BIBLIOTEKS- OCH INFORMATIONSVETENSKAP

Connecting the dots: mass, energy, word meaning, and particle-wave duality

Sándor Darányi and Peter Wittek

Swedish School of Library and Information Science
University of Borås

QI-12, Paris, June 27-29, 2012

Background considerations

- Paper at: http://hdl.handle.net/2320/11004
- Our immediate frame of thought is an evolving digital library (DL) setting: documents (D), index terms (T), and user queries (Q) as vectors or functions for text categorization (TC) and information retrieval (IR)
  - Here and now we focus on word meaning only
  - But sentence semantics can be included
- In a broader setting we consider language as a quantum-like (QL) system, with fitting research tasks for language technology (LT)
  - E.g. investigate compute-intensive methods borrowed from QM for data-intensive large-scale LT applications in the cloud

Using physics to model TC and IR

- Physics: an emergent modeling paradigm
- Media: any, hence relevant for DL, but in focus here is text
- Model of what:
  - ”Binding forces”:
    - Lexical attraction (LA), aka “syntactic word affinity”/sentence cohesion (cf. LA = mutual information for dependence grammars (Yuret 1998)) , also called “infomagnetism” (pointwise mutual information, Hutchens 2001)
    - Term dependence based on their co-occurrence
    - Cross-textual cohesion and coherence (White 2002)

BINDING FORCES

	Physics		*Linguistics*
1	Strong nuclear force	*Strongest*	Word uninterruptability (binds morphemes into words)
2	Electromagnetism	*Less strong*	Grammar (binds words into sentences)
3	Weak nuclear force	*Less strong*	Texture/cohesion; coherence (binds sentences into texts)
4	Gravity	*Weakest*	Intercohesion; intercoherence (binds texts into literatures)

Mainstream linguistics traditionally deals with Forces 1 and 2, while discourse analysis and text linguistics are particularly concerned with Force 3. The field most identified with the study of Force 4 is information science.

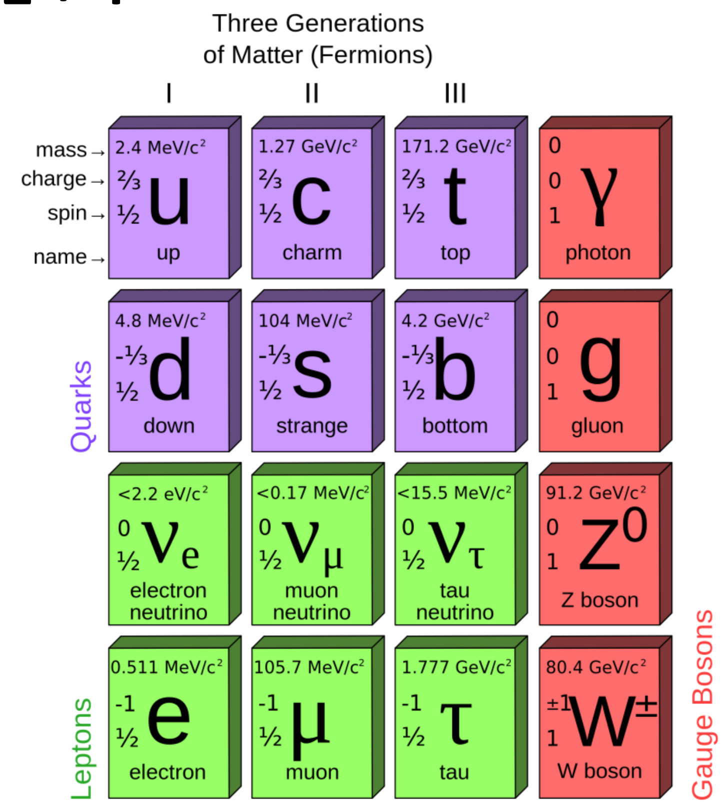
  - Classification process: decision making, supervised vs. unsupervised; by the concept of “energy”
- The metaphor is limited

Word of warning: White's four forces do not completely overlap with those in the Standard Model (SM)

Differences

- The SM describes Nature as a huge interaction between 12 types of particles and 4 types of forces
- Gravity is included in White but not in the SM
- The *graviton* is the theoretically predicted quantum of the gravitational field. If a quantum theory of gravity exists, this would be the particle which mediates the gravitational force much like the photon mediates electromagnetism for quantum electrodynamics (QED).

12 + 4



- Source: Wikipedia, Particle physics

Concepts of energy

- Three takes: on energy, exactly vs. inexactly located semantic content, and the "term mass" conjecture
- Since forces and energy are connected, how about the latter?
- Greek *energeia* (Aristotle *Met*. ix.8 1050a22), here: "work capacity, work content" [in a structure]
- As always, it is application dependent. Senses used in ML:
  - Mathematical energy
    - Signal energy in calculations, devoid of physical content (e.g. Park 2003). "Signals that arise from strictly mathematical processes and have no apparent physical equivalent are commonly considered to represent some form of mathematical energy" (Bruce 2007)
    - Loss functions (in machine learning)
    - Local density of values in mathematical object: "Energy of a (part of a) vector is calculated by summing up the squares of the values in the (part of the) vector" (Wang and Wang 2001)
  - Physical energy
    - By a force: gravitation, electromagnetic force, nuclear forces (implies field nature of subject matter)
    - The latter leading to the eigen conjecture that terms have "mass", i.e. word meaning is similar to energy
    - Methods also utilize potentials (Khrennikov 2010, Blekas & Lagaris 2007, Horn & Gottlieb 2001, Weinstein & Horn 2009)

Energy as cost of [wrong] decision (cf. LeCun *et al*. 2006)

Match right image with right class label

Right naming decisions due to minima in energy landscape

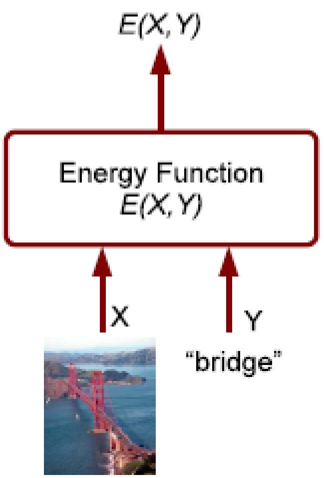


Figure 1. Energy Based Inference: Inference in *EBL* is done by finding Y that minimizes $E(X,Y)$ for a given input X. Input can be a set of image pixels and the output to be predicted can be a class label as shown.

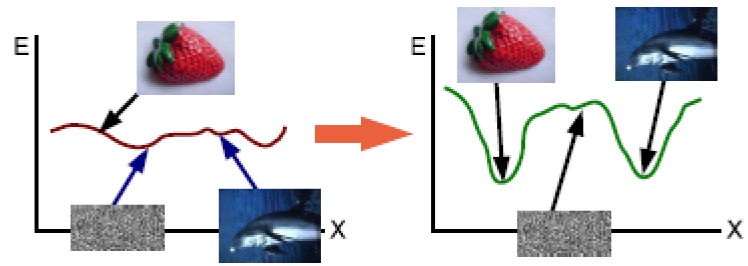
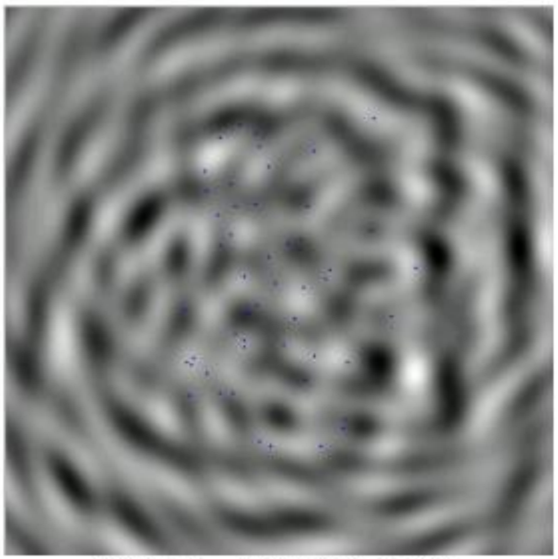
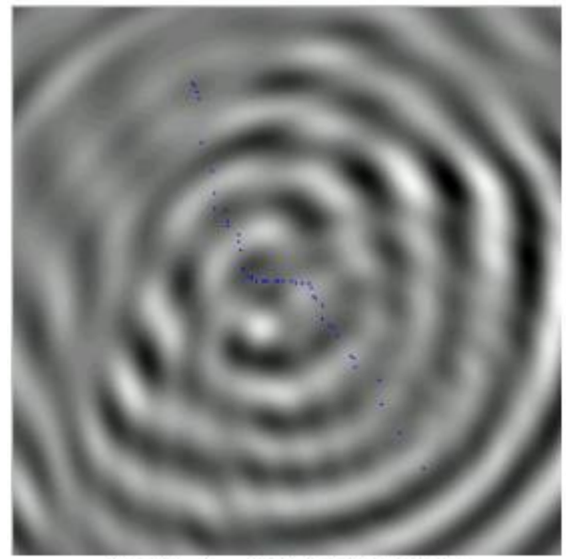


Figure 2. Energy Based Training: Before training, the energy surface produced by an *EBL* model is not distinctive around training data. After training, the energy surface is shaped lower around training data.

Back to documents as wave interference patterns (Azzopardi 2008); see also (Walmsley 2001)



Plot 1.4 – 50 C data points



Plot 1.6 – 50 D data points

Is semantic content exactly or inexactly (regionally) located in vector space?

- Depends on the tools – what do we want to model by mathematical objects?
  - Vectors capture word meaning as something exact
  - Functions capture word meaning as something inexact, fuzzy
- However, in her *vector space* model, Erk (2009) argues for regionality:
  - Many models of categorization in psychology represent concept as regions, characterized by feature vectors with dimension weights
  - Offers two computational models (for monosemous vs. polysemous words)
  - Both models can host soft region boundaries
  - In one, regionality implies gradually decreasing similarity between concepts as word type vectors
- Gradually decreasing similarity: see also lexical attraction (Beeferman *et al* 1997) :
  - Likelihood of a syntactic relation, i.e. pertains to sentence meaning
  - Decays over distance like a force
- Hence both terms and sentence components can be considered as regions as well
- Since (1) term vectors can represent both exactly and inexactly located word meaning, further since (2) at the same time word meaning as something regional refers to its wavelike nature, we start to sense a dual nature here, just like in particle-wave duality (more on this soon)

Indications of regionally located semantic content from linguistics

Lexical attraction and repulsion (Beeferman *et al.* 1997)

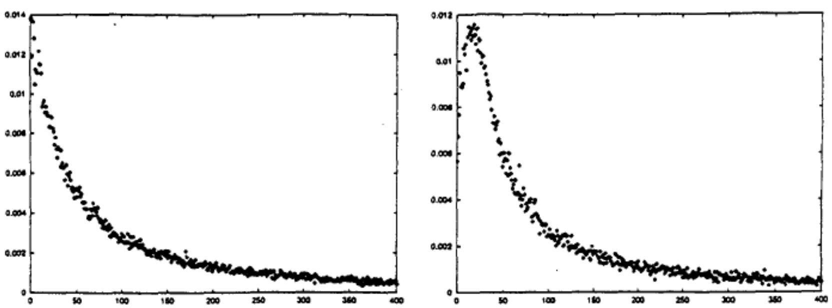
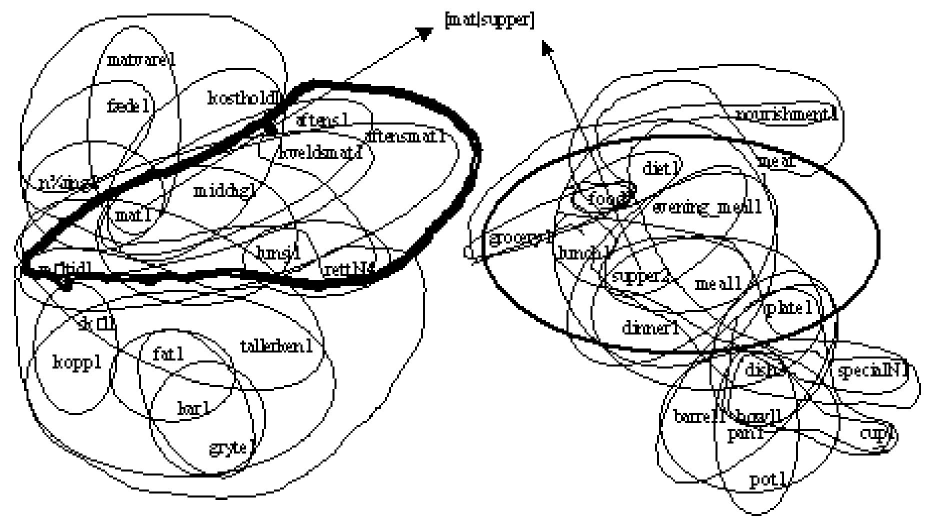


Figure 3: The observed distance distributions—collected from five million words of the Wall Street Journal corpus—for one of the non-self trigger groups (left) and one of the self trigger groups (right). For a given distance $0 \leq k \leq 400$ on the x-axis, the value on the y-axis is the empirical probability that two trigger words within the group are separated by exactly $k + 2$ words, conditional on the event that they co-occur within a 400 word window. (We exclude separation of one or two words because of our use of distance models to improve upon trigrams.)

Paired semantic fields from Norwegian and English (Dyvik 2005)



“Term mass” (eigen) conjecture

- Both several latent structure methods, and quantum mechanics, use eigen decomposition to identify the whereabouts of their objects:
  - $Ax = \lambda x$ vs. $H\psi = E\psi$ (i.e. the time-independent Schrödinger equation)
  - Let $A = XX^T$ be a symmetric term co-occurrence matrix, where X = the vocabulary; hence co-occurrence is an operator
  - Eigen decomposition is closely related to singular value decomposition
  - T, D, Q exactly located in latent semantic analysis (LSA)
  - Particles are in superposition, but the eigenstates can be exactly identified
- Our current assumption is that concepts (perhaps due to rates of word use?) have “mass”, i.e. word meaning behaves *as if* it had an energetic nature
  - Work capacity stored in electron shells vs. in term dependencies
- This “mass” via energy seems to connect to (intellectual) work

Connecting the dots

- We have seen that:
  - Exactly located word meaning (”terms as particles”) can be modelled by position vectors in Euclidean space
  - Inexactly located word meaning (”terms as regions”) implies wave nature, therefore terms as regions can be modelled by waves (functions) in Hilbert (L2) space
  - Word meaning as an observable seems to show a dual nature, depending on the observation apparatus
- Next we consider that:
  - Terms as particles with ”mass” can be demonstrated on an evolving DL as a classical mechanics (CM) system
  - Wave functions behaving like particles can be shown by Dynamic Quantum Clustering (DQC, Weinstein & Horn 2009)

”Term mass” and CM: a toy example

- Salton’s dynamic library (1976) with document cluster centroids displaced due to update
- Example: expanding TD matrices with fixed vocabulary, D and T have temporal indices as well
- Similarity by the cosine of two T (D, Q) vectors with the same temporal index
- Dislocation of the same term: cosine of two T vectors with consecutive indices (-> distance *d*)
- With TD updates at time units, term velocity *v* = *d*
- Term acceleration (*a*) = difference between term velocities over time units
- With similarity as a force, *F=m/a*

$t=0$	Doping	Football	Performance	Skiing	Training
d_1	5	2	0	0	0
d_2	4	0	0	3	1
d_3	0	0	4	0	5
d_4	6	0	2	0	0
d_5	0	3	0	0	4
$t=1$					
d_1	5	2	0	0	0
d_2	4	0	0	3	1
d_3	0	0	4	0	5
d_4	6	0	2	0	0
d_5	0	3	0	0	4
d_6	2	3	0	1	1
d_7	1	0	0	4	5
$t=2$					
d_1	5	2	0	0	0
d_2	4	0	0	3	1
d_3	0	0	4	0	5
d_4	6	0	2	0	0
d_5	0	3	0	0	4
d_6	2	3	0	1	1
d_7	1	0	0	4	5
d_8	5	6	1	1	0
d_9	2	1	1	3	0

Table 1. Evolution of an indexing vocabulary over time

$$|F|/|a| = m$$

	Doping	Football	Performance	Skiing	Training
v_1	9	9	0	25	36
v_2	49	49	4	16	0
a	40	40	4	-9	-36
F	1.56	1.28	1.24	1.35	1.37
m	0.039	0.032	0.31	0.15	0.038

Table 2. Calculation of term mass over t_0-t_1

”Term mass” and QM: DQC

Ehrenfest’s teorem:

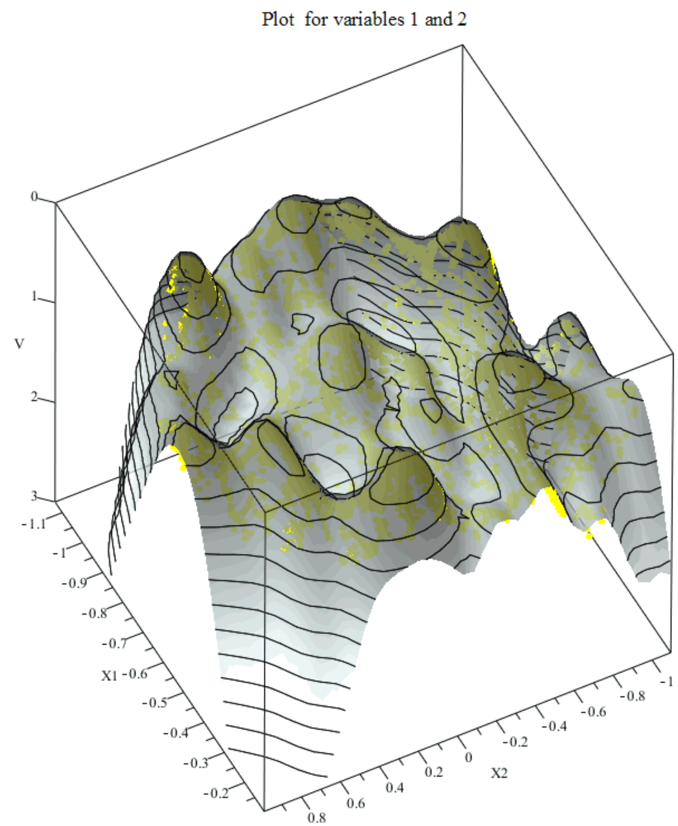
the time-dependent expectation value of the position operator

$$\langle\psi(t)|x|\psi(t)\rangle = \int \psi(x,t)^* x \psi(x,t) dx$$

satisfies the equation

$$\frac{d^2\langle x(t)\rangle}{dt^2} = \langle\psi(t)|\nabla V(x)|\psi(t)\rangle.$$

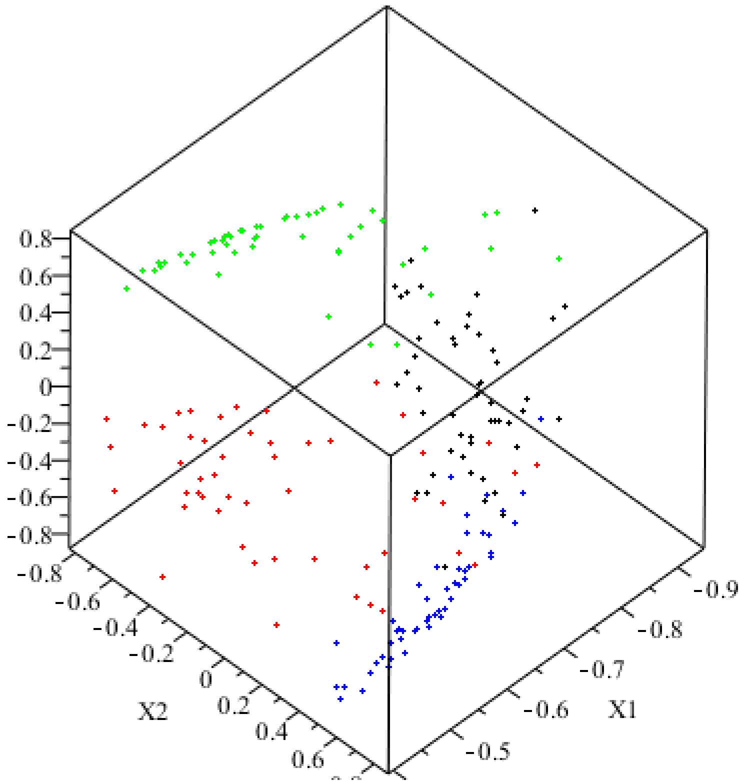
- The yellow dots in the static image are the original data points, but since they are at the centre of their corresponding (Gaussian) wave functions, they are also expectation values, i.e. the average values of their probability densities



Source: http://www.slac.stanford.edu/~niv/index_files/images/DQCOverview_193.gif

DQC – Wave functions as quasi-particles

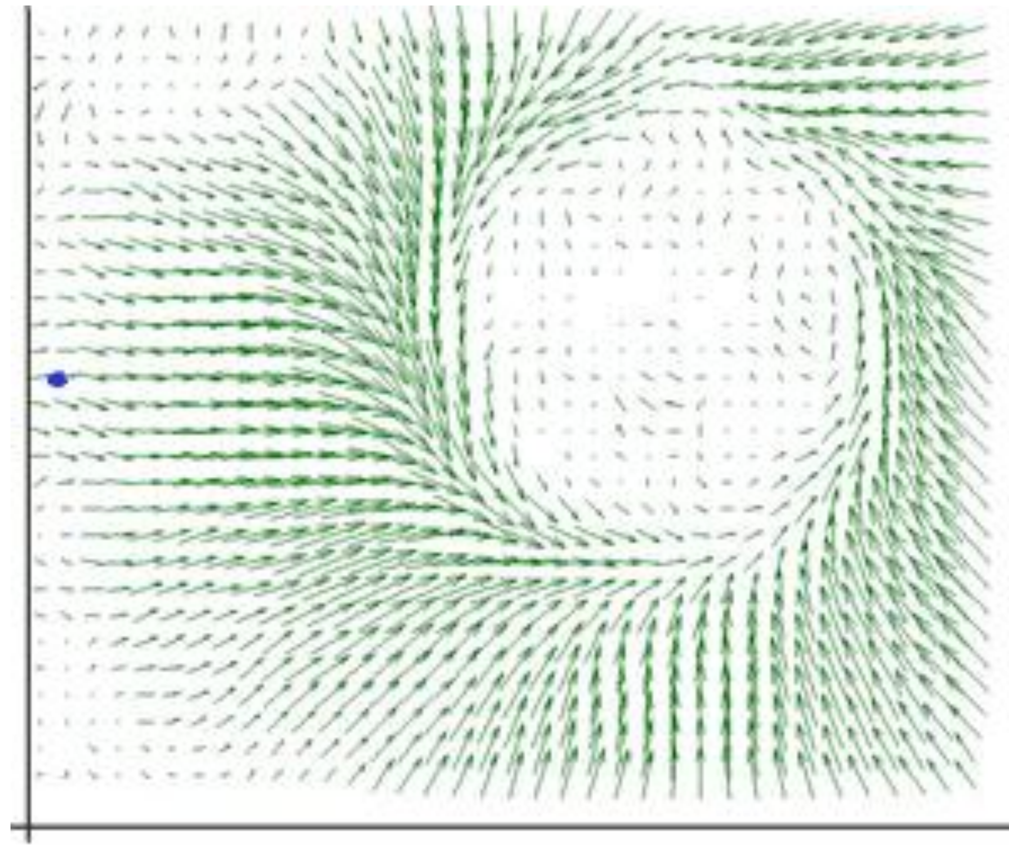
- The animation shows how expectation values of the data points roll into their respective lowest potentials
- E.g. the green dots are converging toward a "centroid", some lowest potential, and they resemble locations ("particles") in this sense



Source: http://www.slac.stanford.edu/~niv/index_files/images/DQCOverview_150.gif

Roadmap: Sentence meaning and work – a thought experiment

- The trajectory of a particle along a curve inside a vector field. At the bottom are the vectors of the field seen by the particle as it travels along the curve. The sum of the dot products of these vectors with the tangent vector of the curve at each point of the trajectory results in the line integral.
- Sentence = concatenated term vectors in a field, a vector-valued function (VVF)
- Energy (work content) of a sentence is its line integral



- Source: Wikipedia, Line integral

Future research

Lexical attraction subject to inverse square law suggests a field-like explanation

Sentence meaning as a convoluted VVF results from the field view

Need to find theories of word and sentence meaning fitting eigen decomposition (e.g. componential semantics [Katz, Pottier, Wierzbicka]), or field theory (lexical/semantic fields [Trier, Haas])

Need to identify components of the Hamiltonian in the DL model

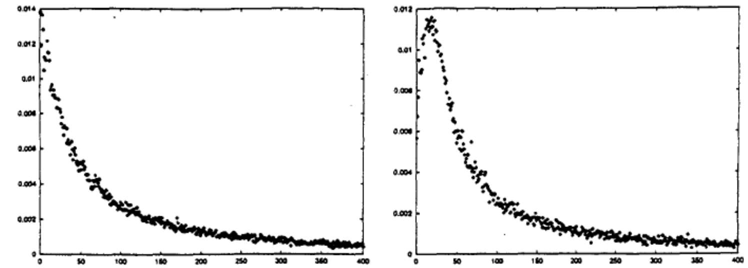
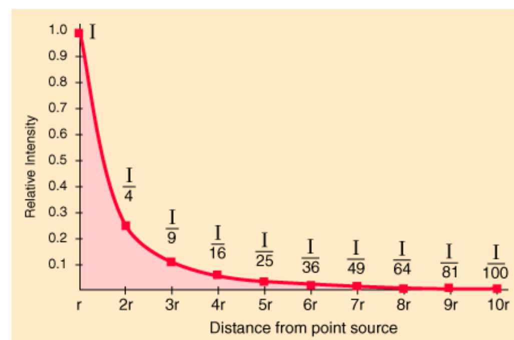
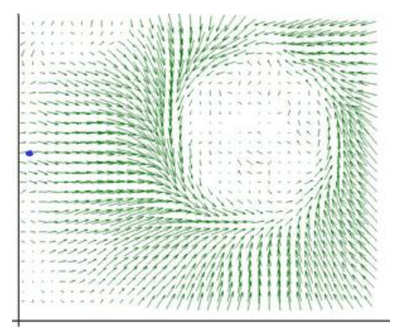


Figure 3: The observed distance distributions—collected from five million words of the Wall Street Journal corpus—for one of the non-self trigger groups (left) and one of the self trigger groups (right). For a given distance $0 \leq k \leq 400$ on the x-axis, the value on the y-axis is the empirical probability that two trigger words within the group are separated by exactly $k + 2$ words, conditional on the event that they co-occur within a 400 word window. (We exclude separation of one or two words because of our use of distance models to improve upon trigrams.)





Selected references

- Azzopardi . L. (2008). Wave Motion: A New Metaphor for 2D Information Visualization – The Exploration of a Metaphor. *CiteSeerX 2008-07-01* (Document is wave pattern)
- Beeferman, D., Berger, A., and Lafferty, J. (1997). A model of lexical attraction and repulsion. In *ACL/EACL '97*. (Lexical attraction degrades over neighbourhood, i.e. is force-like)
- Bruce , E.N. (2007). *Biomedical Signal Processing And Signal Modeling*. Wiley India Pvt. Ltd..
- Dyvik, H. (2005). Translations as a semantic knowledge source (Manuscript, illustrates overlapping semantic fields)
- Erk , K. (2009). Representing words as regions in vector space. *ACL Proceedings CoNLL '09* (Semantic content is region, inexactly located)
- Khrennikov, A. (2010). *Ubiquitous quantum structure: from psychology to finance*. Springer: Heidelberg. (QL rules)
- LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., and Huang, F.J. (2006). *A Tutorial on Energy-Based Learning*. In G. Bakir, T. Hofman, B. Schölkopf, A. Smola, B. Taskar (Eds), *Predicting Structured Data*. MIT Press. (Energy is loss function)
- Park, L.A.F. (2003*). Spectral Based Information Retrieva*l. PhD thesis. The University of Melbourne.
- Sermanet, P., Kavukcuoglu, K., and LeCun, Y. (2009). EBLearn: Open-Source Energy-Based Learning in C++. *Proceedings of ICTAI'09.*
- Yuret, D. (1998). *Discovery of Linguistic Relations Using Lexical Attraction*. PhD thesis, MIT (MI and IC yield dependence grammar with bonds and "gradient").
- Dorrer, C., Londero, P., Anderson,M., Wallentowitz , S. and Walmsley, I.A. (2001). Computing with interference: All-Optical SIngle-Query 50-Element Database Search. *QELS. 149-150.*(Document is color interference)
- Wang, C. and Wang; X.S. (2001). Indexing very high-dimensional sparse and quasi-sparse vectors for similarity searches. *The VLDB Journal* (9), 344–361. (Energy is local value density)
- White, H. (2002). *Cross-textual cohesion and coherence*. The CHI 2002 Discourse Architectures Workshop.