The Role of Roles in Translating Across Conceptual Systems

Robert L. Goldstone (rgoldsto@indiana.edu)
Brian J. Rogosky (brogosky@indiana.edu)
Department of Psychology
Indiana University
Bloomington, IN. 47405

Abstract

We explore one aspect of meaning, the identification of matching concepts across systems (e.g. people, theories, or cultures). We present a computational algorithm called ABSURDIST (Aligning Between Systems Using Relations Derived Inside Systems for Translation) that uses only within-system similarity relations to find between-system translations. While illustrating the sufficiency of within-system relations to account for translating between systems, simulations of ABSURDIST also indicate synergistic interactions between intrinsic, within-system information and extrinsic information.

Conceptual Meaning and Translation

There have been two major answers to the question of how our concepts have meaning. The first answer is that concepts' meanings depend on their connection to the external world (Harnad, 1990). By this account, the concept Dog means what it does because our perceptual apparatus can identify features that characterize, if not define, dogs. Dog is characterized by features that are either perceptually given, or can be reduced to features that are perceptually given. This will be called the "external grounding" account of conceptual meaning. The second answer is that concepts' meanings depend on their connections to each other (Markman & Stillwell, 2001; Saussure, 1915/1959). By this account, Dog's meaning depends on Cat, Domesticated, and Loyal, and in turn, these concepts depend on other concepts, including Dog. The dominating metaphor here is of a conceptual web in which concepts all mutually influence each other (Quine & Ullian, 1970). A concept can mean something within a network of other concepts but not by itself. This will be called the "conceptual web" account.

The goal of this article is to argue for the synergistic integration of conceptual web and externally grounded accounts of conceptual meaning. However, in pursuing this argument, we will first argue for the sufficiency of the conceptual web account for a particular task associated with conceptual meaning. Then, we will show how the conceptual web account can be ably supplemented by external grounding to establish meanings more successfully than either method could by itself.

Our point of departure for exploring conceptual meaning will be a highly idealized and purposefully simplified version of a conceptual translation task. Consider two individuals, Joan and John, who each possesses a number of concepts. Suppose further that we would like some way to tell that Joan and John both have a concept of, say, **Mushroom**. Joan and John may not have exactly the same concept of **Mushroom**. John may believe mushrooms grow from seeds whereas Joan believes they grow from spores. More generally, Joan and John will differ in the rest of their conceptual networks because of their different experiences and levels of expertise. Still, it seems desirable to say that Joan and John's **Mushroom** concepts correspond to one another. We will describe a network that translates between concepts in two systems, placing, for example, Joan and John's **Mushroom** concepts in correspondence with each other.

Translation across systems is generally desirable and specifically necessary in order to say things like "John's concept of mushrooms is less informed than Joan's." Fodor and Lepore have taken the existence of this kind of translation as a challenge to conceptual web accounts of meaning (Fodor & Lepore, 1992). By Fodor and Lepore's interpretation, if a concept's meaning depends on its role within the larger system, and if there are some differences between the systems, then the concept's meaning would be different in the two systems. A natural way to try to salvage the conceptual web account is to argue that determining corresponding concepts across systems does not require the systems to be identical, but only similar. However, Fodor (Fodor, 1998; Fodor & Lepore, 1992) insists that the notion of similarity is not adequate to establish that Joan and John both possess a **Mushroom** concept. Fodor argues that "saying what it is for concepts to have similar, but not identical contents presupposes a prior notion of beliefs with similar but not identical concepts" [Fodor, 1998, p. 32].

The ABSURDIST Algorithm for Cross-system Translation

We will now present a simple neural network called ABSURDIST (Aligning Between Systems Using Relations Derived Inside Systems for Translation) that finds conceptual correspondences across two systems (two people, two time slices of one person, two scientific theories, two developmental age groups, two language communities, etc.) using only inter-conceptual similarities, not conceptual identities, as input. Thus, ABSURDIST will take as input two systems of concepts in which every concept of a system is defined exclusively in terms of its dissimilarities to other concepts in the same system. Laakso and

Cottrell (2000) describe another neural network model that uses similarity relations within two systems to compare the similarity of the systems. ABSURDIST produces as output a set of correspondences indicating which concepts from System A correspond to which concepts from System B. These correspondences serve as the basis for understanding how the systems can communicate with each other without the assumption made by Fodor (1998) that the two systems have exactly the same concepts. The existence of ABSURDIST provides evidence against Fodor's argument that similarities between people's concepts are an insufficient basis for determining that two people share an equivalent concept. ABSURDIST is not a complete model of conceptual meaning or translation. Our point is that even if the only relation between concepts in a system were simply similarity, this would still suffice to find translations of the concepts in different systems.

Elements $A_{1..m}$ belong to System A, while elements $B_{1..n}$ belong to System B. $C_t(A_q,B_x)$ is the activation, at time t, of the unit that represents the correspondence between the qth element of A and the xth element of B. There will be m·n correspondence units, one for each possible pair of corresponding elements between A and B. In the current example, every element represents one concept in a system. The activation of a correspondence unit is bound between 0 and 1, with a value of 1 indicating a strong correspondence between the associated elements, and a value of 0 indicating strong evidence that the elements do not correspond. Correspondence units dynamically evolve over time by the equations:

$$\begin{aligned} &if\, N\left(C_t\left(A_q,B_x\right)\right) \geq 0\, then\, C_{t+1}\left(A_q,B_x\right) = C_t\left(A_q,B_x\right) + N\left(C_t\left(A_q,B_x\right)\right)\left(\max - C_t\left(A_q,B_x\right)\right)L \\ &else\, C_{t+1}\left(A_q,B_x\right) = C_t\left(A_q,B_x\right) + N\left(C_t\left(A_q,B_x\right)\right)\left(C_t\left(A_q,B_x\right) - \min\right)L \end{aligned} \quad (1).$$

If $N(C_t(A_q,B_x))$, the net input to a unit that links the qth element of A and the xth element of B, is positive, then the unit's activation will increase as a function of the net input, a squashing function that limits activation to an upper bound of max=1, and a learning rate L (set to 1). If the net input is negative, then activations are limited by a lower bound of min=0. The net input is defined as

$$N\left(C_t\left(A_q,B_x\right)\right) = \alpha E\left(A_q,B_x\right) + \beta R\left(A_q,B_x\right) - \chi I\left(A_q,B_x\right), \quad (2)$$

where the E term is the external similarity between A_q and B_x, R is their internal similarity, I is the inhibition to placing A_q and B_x into correspondence that is supplied by other developing correspondence units, and a+b+c=1. When a=0, then correspondences between A and B will be based solely on the similarities among the elements within a system, as proposed by a conceptual web account. The amount of excitation to a unit based on within-system relations is given by

$$R\left(A_q,B_x\right) = \frac{\sum_{\substack{r=1 \\ r\neq q}}^{m}\sum_{\substack{y=1 \\ y\neq x}}^{n} S\left(D\left(A_q,A_r\right),D\left(B_x,B_y\right)\right)C_t\left(A_r,A_y\right)}{Min\left(m,n\right)-1},$$

where $D(A_q,A_r)$ is the psychological distance between elements q and r in System A, and S is a negative exponential function of the absolute difference between S's two arguments. The amount of inhibition is given by

$$I\left(A_q,B_x\right) = \frac{\sum_{\substack{r=1 \\ r\neq q}}^{m} C_t\left(A_r,B_x\right) + \sum_{\substack{y=1 \\ y\neq x}}^{n} C_t\left(A_q,B_y\right)}{m+n-2}.$$

According to the equation for R, Elements q and x will tend to be placed into correspondence to the extent that they enter into similar similarity relations with other elements. For influencing alignments, the similarity between two distances is weighted by the strengths of the units that align elements that are placed in correspondence by the distances. The equation for R represents the sum of the supporting evidence (the consistent correspondences), with each piece of support weighted by its relevance (given by the S term). The inhibitory I term is based on a one-to-one mapping constraint (Falkenhainer, Forbus, & Gentner, 1989). The unit that places A_q into correspondence with B_x will tend to become deactivated if other strongly activated units place A_q into correspondence with other elements from B, or B_x into correspondence with other elements from A.

Correspondence unit activations are initialized to random values selected from a normal distribution with a mean of 0.5 and a standard deviation of 0.05. In our simulations, Equation (1) is iterated for a fixed number of cycles. It is assumed that ABSURDIST places two elements into correspondences if the activation of their correspondence unit is greater than or equal to 0.55 after the fixed number of iterations have been completed (4000 cylces in the simulations described below).

Assessing ABSURDIST's Performance

In assessing ABSURDIST's performance, we will assume that conceptual dissimilarities obey Euclidean distance metric assumptions, and are interpretable as distances between concepts lying in a geometric space. Our general method for evaluating ABSURDIST will be to generate a number of elements in a two dimensional space, with each element identified by its value on each of the two dimensions. These will be the elements of System A, and each is represented as a point in space. System B's elements are created by copying the points from System A and adding Gaussian noise to each of the dimension values of each of the points. Then, equation (1) is used to update correspondences across the two systems for a fixed number of iterations. The correspondences computed by ABSURDIST are then compared to the correct correspondences. Two elements correctly correspond to each other if the element in System B was originally copied from the element in System A.

Noise tolerance and system complexity

An initial set of simulations was conducted to determine how robust the ABSURDIST algorithm was to noise and how well the algorithm scaled to different sized systems. We ran a 7 X 6 factorial combination of simulations, with 7 levels of added noise and 6 different numbers of elements per system. Noise was infused into the

algorithm by varying the displacement between corresponding points across systems. The points in System A were set by randomly selecting dimension values from a uniform random distribution with a range from 0 to 1000. System B points were copied from System A, and Gaussian noise with standard deviations of 0, 0.1, 0.2, 0.3, 0.4, 0.5, or 0.6% was added to the points of B. The number of points per system was 3, 4, 5, 6, 10, or 15. α was set to 0, β was set to 0.4, and χ to 0.6. The values for β and χ were selected because they were the most balanced weights that produced fewer than 5% two-to-one correspondences. For each of the 42 combinations of noise and number of items, 1000 separate randomized starting configurations were tested. The results from this simulation are shown in Figure 1, which plots the percentage of simulations in which each of the proper correspondences between systems is recovered. For example, for 15-item systems, the figure plots the percentage of time that all 15 correspondences are recovered. The graph shows that performance gradually deteriorates with added noise, but that the algorithm is robust to at least modest amounts of noise.

More surprisingly, Figure 1 also shows that the algorithm's ability to recover true correspondences generally increases as a function of the number of elements in each system, at least for small levels of noise. One might have thought that as more elements were matched between systems that there would be greater confusion between elements, given that the size of the bounding region remains constant. As the number of elements in a system increases, the similarity relations between those elements provide increasingly strong constraints that serve to uniquely identify each element. If one generated random translations that were constrained to allow only one-to-one correspondences, then the probability of generating a completely correct translation would be 1/N!. Thus, with 0.6% noise, the 23% rate of recovering all 3 correspondences for a 3-item system is slightly above chance performance of 16.67%. However, with the same amount of noise, the 17% rate of recovering all of the correspondences for a 15-item system is remarkably higher than the chance rate of $7.6X10^{-13}$. Thus, at least in our highly simplified domain, we have support for the argument (Lenat & Feigenbaum, 1991) that establishing meanings on the basis of within-system relations becomes easier, not harder, as the size of the system increases.

Interactions between extrinsic and intrinsic determinants of alignments

The simulation above indicates that within-system relations are sufficient for discovering between-system translations, but this should not be interpreted as suggesting that the meaning of an element is not also dependent on relations extrinsic to the system. ABSURDIST offers a useful, idealized system for examining interactions between intrinsic (within-system) and extrinsic (external to

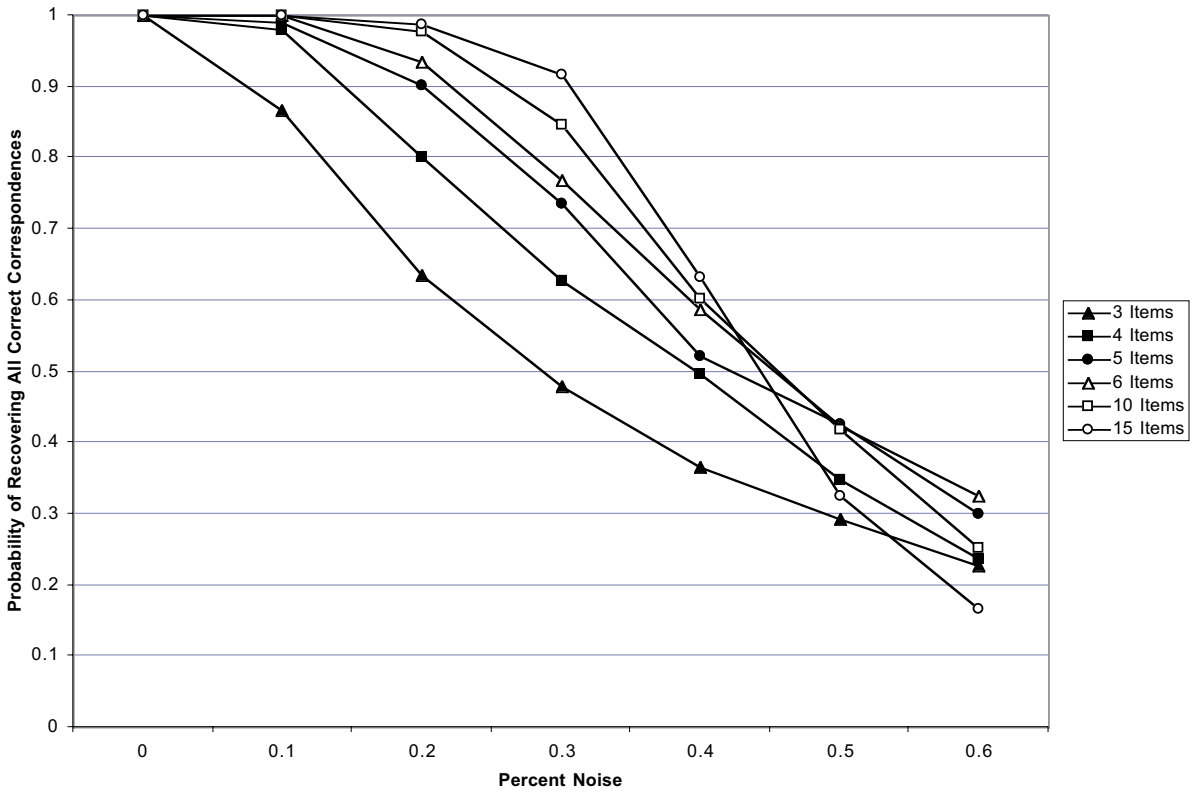

Figure 1

the system) aspects of meaning. One way to incorporate extrinsic biases into the system is by initially seeding correspondence units with values. Thus far, all correspondence units have been seeded with initial activation values tightly clustered around 0.5. However, in many situations, there may be external reason to think that two elements correspond to each other: they may receive the same label, they may have perceptual attributes in common, they may be associated with a common event, or a teacher signal may have provided a hint that the two elements correspond. In these cases, the initial seed-value may be significantly greater than 0.5.

Figure 2 shows the results of a simulation of ABSURDIST with different amounts of extrinsic support for a selected correspondence between two elements. Two systems are generated by randomly creating a set of points in two dimensions for System 1, and copying the points' coordinates to System 2 while introducing 0.6% noise to their positions. When Seed = 0.5, then no correspondence is given an extrinsically supplied bias. When Seed=0.75, then one of the true correspondences between the systems is given a larger initial activation than the other correspondences. For a system made up of 15 elements, a mapping accuracy of 31% is obtained without any extrinsic assistance (Seed=0.5). If seeding a single correct correspondence with a value of 1 rather than 0.5 allowed ABSURDIST to recover just that one correspondence with 100% probability, then accuracy would increase at most to 35.6% (((.31 * 14) + 1)/15). The reference line in Figure 2 shows these predicted increases in accuracy. For all systems tested, the observed increment in accuracy far outstretches the increase in accuracy predicted if seeding a correspondence only helped that correspondence. Moreover, the amount by which translation accuracy improves beyond the amount predicted generally increases as a function of system size. Thus, externally seeding a correspondence does more than just fix that correspondence. In a system where correspondences all mutually depend upon each other, seeding one correspondence has a ripple-effect through which other correspondences are improved.

Equation 2 provides a second way of incorporating extrinsic influences on correspondences between systems. This equation defines the net input to a correspondence unit as an additive function of the extrinsic support for the correspondence, the intrinsic support, and the competition against it. Thus far, the extrinsic support has been set to 0. The extrinsic support term can be viewed as any perceptual, linguistic, or top-down information that suggests that two objects correspond (this differs from the philosopher's use of "external meaning" to refer to the causal determinants of a concept). To study interactions between extrinsic and intrinsic support for correspondences, we conducted 1000 simulations that started with 10 randomly placed points in a two-dimensional space for System A, and then copied these points over to System B with Gaussian-distributed noise. The intrinsic, role-based support is determined by the previously described equations. The extrinsic support term of Equation 2 is given by a negative exponential function of the absolute distance between the two concepts' absolute locations. Thus, the correspondence unit connecting q and x will tend to be strengthened if q and x have similar coordinates. This is extrinsic support because the similarity of q's and x's coordinates can be determined without any reference to other elements.

In conducting this third simulation, we assigned three different sets of weights to the extrinsic and intrinsic support terms. For the "Extrinsic only" results of Figure 3, we set α=0.4, β=0, and χ=0.6. For the "Intrinsic only" results, we set α=0, β=0.4, and χ=0.6. For "Intrinsic and Extrinsic," we set α=0.2, β=0.2, and χ=0.6.

Figure 3 shows that using only information intrinsic to a system results in better correspondences than using only extrinsic information. This is because corresponding elements that have considerably different positions in their systems can often still be properly connected with intrinsic information if other proper correspondences can be

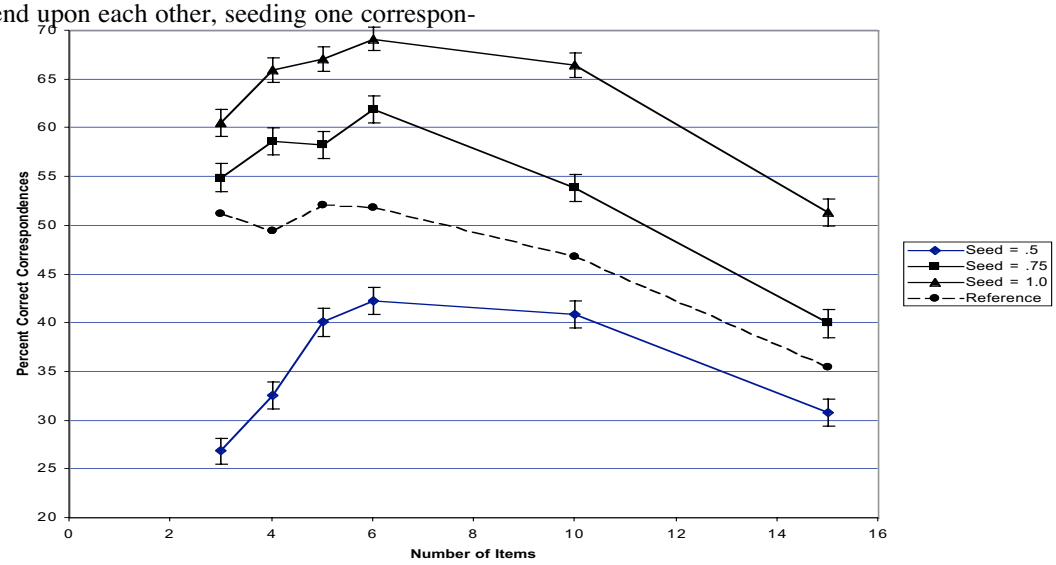

Figure 2

recovered. The intrinsic support term is more robust than the extrinsic term because it depends on the entire system of emerging correspondences. For this reason, it is surprising that the best translation performance is found when intrinsic and extrinsic information are both incorporated into Equation 2 with equal weight. The superior performance of the network that uses both intrinsic and extrinsic information derives from its robustness in the face of noise. Some distortions to points of System B adversely affect the intrinsic system more than the extrinsic system. For example, a slight distortion to a point may make its pattern of distances to other points quite similar to another point. A system that incorporates both sources of information will tend to recover well from either disruption to absolute or relative positions if the other source of information is reasonably intact.

Discussion

The ABSURDIST model makes two theoretically important points. First, translations between two systems can be found using only information about the relations between elements within a system. The claim is that the concept in Person A that matches a concept in Person B can be found considering only the relations between concepts in Person A, and the relations between concepts in Person B. ABSURDIST demonstrates how a holistic conception of meaning is compatible with the goal of determining correspondences between concepts across individuals. Two people need not have exactly the same systems, or even the same number of concepts, to create proper conceptual correspondences. Contra Fodor (Fodor, 1998; Fodor & Lepore, 1992) information in the form of inter-conceptual similarities suffices to find inter-system translations between concepts. It is often easier to find translations for large systems than small systems.

The second important theoretical contribution of ABSURDIST is to formalize some of the ways that intrinsic, within-system relations and extrinsic, perceptual information synergistically interact in determining conceptual alignments. Intrinsic relations suffice to determine cross-concept translations, but if extrinsic information is available, more robust, noise-resistant translations can be found. The synergistic benefit of combining intrinsic and extrinsic information sheds new light on the debate on accounts of conceptual meaning. It is common to think of intrinsic and extrinsic accounts of meaning as being mutually exclusive, or at least zero-sum. Seemingly, either a concept's meaning depends on information within its conceptual system or outside of its conceptual system, and to the extent that one dependency is strengthened, the other dependency is weakened. In opposition to this zero-sum perspective on intrinsic and extrinsic meaning, ABSURDIST offers a framework in which a concepts' meaning is both intrinsically and extrinsically determined (see also two-factor theories in philosophy such as Block, 1986), and the external grounding makes intrinsic information more, not less, powerful. To claim that all concepts in a system depend on all of the other concepts in a system is perfectly compatible with claiming that all of these concepts have a perceptual basis.

We have focused on the application of ABSURDIST to the problem of translating between different people's conceptual systems. However, the algorithm is applicable to a variety of situations in which elements from two systems must be placed in correspondence in an efficient and reasonable (though not necessarily optimal) manner. A

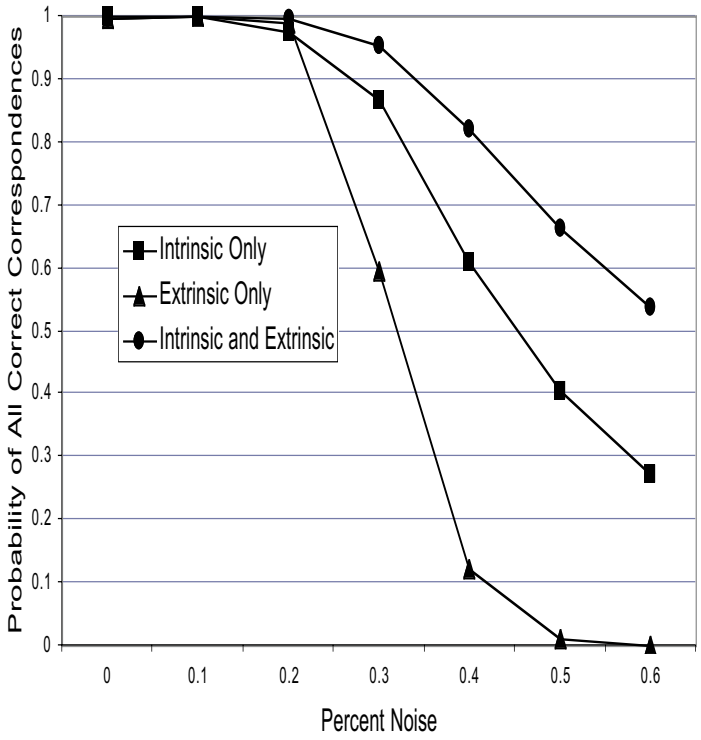

Figure 3

combination of properties makes ABSURDIST particularly useful for applications in cognitive science: 1) the algorithm can operate solely on relations within a system, 2) the within-system relations can be as simple as generic similarity relations, 3) the algorithm can combine within-system and between-systems information when each is available, 4) the algorithm has a strong bias to establish one-to-one correspondences, and 5) the algorithm does not require larger numbers of iterations for convergence as the number of elements per system increases. Some of the domains of application for ABSURDIST include object recognition, analogy, and automatic translation.

Object recognition

The ABSURDIST algorithm can be applied to the problem of object recognition that is invariant to translation, rotation, and reflection. For this application, a pictorial object is the system, and points on the object are elements of the system. A standard solution to recognizing rotated objects is to find matching landmark points that are identifiable on a known object and an input object to be recognized (Ullman, 1996). Once identified, these landmarks can reveal how the input would need to be rotated to match the known object. Even if no extrinsically aligned landmarks can be identified, ABSURDIST can still match the objects by taking advantage of the wealth of information contained in within-object proximity relations (Edelman, 1999).

Analogy

ABSURDIST offers a complementary approach to analogical reasoning between domains. Most existing models of analogical comparison represent the domains to be compared in terms of richly structured propositions (Hummel & Holyoak, 1997; Eliasmith & Thagard, 2001). In many cases, such as single words or pictures, it is difficult to come up with propositional encodings that capture an item's meaning. In such cases, ABSURDIST's unstructured similarity relations are a useful addition to existing models of analogical reasoning.

Automatic dictionary translation

The small-scale simulations conducted here leave open the promise of applying ABSURDIST to much larger translation tasks, such as dictionaries, thesauri, encyclopedias, and organizational structures. ABSURDIST could provide automatic translations between dictionaries of two different languages, using only co-occurrence relations between words within each dictionary (Burgess & Lund, 2000; Landauer & Dumais, 1997), perhaps supplemented by a small number of external hints (e.g. that French "chat" and English "cat" might correspond to each other because of their phonological similarity).

Acknowledgments

We would like to thank Gary Cottrell, Eric Dietrich, Shimon Edelman, Stevan Harnad, John Hummel, Michael Lynch, Art Markman, and Mark Steyvers for comments on an earlier version of this research. This research was funded by NIH grant MH56871 and NSF grant 0125287.

References

Block, N. (1986). Advertisement for a semantics for psychology. Midwest Studies in Philosophy, 10, 615-78.

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Diettrich & A. B. Markman (Eds.) Cognitive dynamics: Conceptual change in humans and machines. (pp. 117-156). Mahwah, NJ: Lawrence Erlbaum Associates.

Edelman, S. (1999). Representation and recognition in vision. Cambridge, MA: MIT Press.

Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. Cognitive Science, 25, 245-286

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. Artificial Intelligence, 41, 1-63.

Fodor, J. (1998). Concepts: Where cognitive science went wrong. Oxford: Clarendon Press.

Fodor, J., & Lepore, E. (1992). Holism. Oxford, UK: Blackwell.

Harnad, S. (1990). The symbol grounding problem. Physica D, 42, 335-346.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. Psychological Review, 104, 427-466.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.

Laakso, A., & Cottrell, G. (2000). Content and cluster analysis: Assessing representational similarity in neural systems. Philosophical Psychology, 13, 47-76.

Lenat, D. B., & Feigenbaum, E. A. (1991). On the thresholds of knowledge, Artificial Intelligence, 47, 185-250.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. Journal of Experimental and Theoretical Artificial Intelligence, 13, 329-358.

Quine, W. V., & Ullian, J. S. (1970). The Web of Belief. New York: McGraw-Hill.

Saussure, F. (1915/1959). Course in general linguistics. New York: McGraw-Hill.

Ullman, S. (1996). High-level vision. Cambridge, MA: MIT Press.