

## Human Similarity theories for the semantic web

Jose Quesada

Max Planck Institute, Human development  
[quesada@gmail.com](mailto:quesada@gmail.com)

**Abstract.** The human mind has been designed to evaluate similarity fast and efficiently. When building/using a data format to make the web content more machine-friendly, can we learn something useful from how the mind represents data? We present four theories psychological theories that tried to solve the problem and how they relate to semantic web practices. Metric models (such as the vector space model and LSA) were the first-comers and still have important advantages. Advances in Bayesian methods pushed Feature models( e.g., Topic model). Structural mapping models propose that for similarity, shared structure matters more, although the formalisms that express these ideas are still developing. Transformational distance models (e.g., syntagmatic-paradigmatic -SP- model) reduce similarity to information transmission. Topic and SP models do not require preexisting classes but still have a long way to go; the need of automatically generating structure is less pressing when one of the driving forces of the semantic web is the creation of ontologies.

**Keywords:** similarity, cognition, semantics, information extraction, representation, psychology, cognitive science.

### 1. Introduction

The human mind has been “designed” to evaluate similarity fast and efficiently. When building/using a data format to make web content more machine-friendly, can we learn something useful from how the mind represents data? Are there any domain-independent findings on human representation that can inform ontology building and other semantic web activities? Can knowing humans be useful to design better for machines? I would say it might, considering that the end user of what machines using the semantic web produce is human, after all. Nature may have produced algorithms and representations that are reusable. And humans and machines dealing with lots of information may face similar problems.

There are different areas in which psychology may inform semantic web practitioners; For example, agents in the semantic web will do both inductive and deductive reasoning [1], follow causal chains [2], solve problems and make decisions [3]. All these activities depend crucially on how we represent information, and this is what similarity theories aim to explain. So in this paper we will review the major approaches to similarity in psychology and how they relate to the semantic web.

In the last 50 years, psychology has made good progress on the topic of similarity; the basic conclusion is that similarity is a hard topic, but approachable. But why is it so difficult? For a start, it is a very labile phenomenon. Murphy and Medin [4] noted that "the relative weighting of a feature (as well as the relative importance of common and distinctive features) varies with the stimulus context and task, so that there is no unique answer to the question of how similar is one object to another" (p. 296). Goodman [5] also criticized the central role of similarity as an explanatory concept. What does it mean to say that two objects  $a$  and  $b$  are similar? One intuitive answer is to say that they have many properties in common. But this intuition does not take us very far, because all objects have infinite sets of properties in common. For example, a plum and a lawnmower both share the properties of weighing less than 100 pounds (and less than 101 pounds, etc). That would imply that all objects are similar to all others (and vice versa, if we consider that they are different in an infinite set of features too). Goodman proposed that similarity is thus a meaningful concept when defined with a certain "respect". Instead of considering similarity as a binary relation  $s(a, b)$ , we should think of it as a ternary relation  $s(a, b, r)$ . But once we introduce "respects", then similarity itself has no explanatory value: the respects have. Thus, if similarity is useless when not defined "with respect to", then it is not an explanatory concept on which theories can be built: theories should be about "the respects" and similarity can leave the scenario without being missed.

Although this criticism could have been lethal for any psychological theories of similarity, it has not been. The abstract concept of similarity used by philosophers like Goodman and the psychological concept of similarity are different, the latter being more constrained: (1) There are psychological restrictions on what a respect can be. Although they can be very flexible and changeable with goals, purpose, and context, there are constraints in what form they take: they do not change arbitrarily, but systematically. These systematic variations prevent the set of common respects from being infinite, and enable their scientific study [6]. (2) Since people do not normally compare objects one "respect" at a time, but along multiple dimensions (e.g., size, color, function, etc.), the psychologically central issue is to explain the mechanism by which all these factors are combined into a single judgment of similarity. Then, respects do some, but not all of the work in explaining similarity judgments [7] (3) Goodman assumes that the set of features in which two objects can be compared is infinite (then, they have an infinite number of properties in which they are similar *and* dissimilar). However, in psychology we are interested in the similarity between two mental representations of the objects in the mind. Mental representations must be finite. Then computation of similarity can be thought to take place without the need of constraining respects. Theories of mental representation based on similarity should explain what is represented and how this is selected. The features represented cannot be arbitrary, otherwise they cannot be studied scientifically [8].

As a conclusion, what most similarity and categorization psychological theories have in common is the problem of choosing respects [8]: The feature selection and weighting process is outside of the scope of the models, that is, is set up a-priori by the researcher, not dictated by the theory. This is a very important flaw in a model of similarity, as Goodman pointed out. Semantic web practitioners face this problem too.

The semantic web ‘standard’ data structure language is RDF. In RDF, the fundamental concepts are resources, properties and statements. Resources are objects, like books, people or events. Resources have properties like chapters, proper names, or physical locations. Properties are a special type of resources that describe the relation between two resources. And a statement just asserts the properties of resources. In a sense, psychologists and semantic web practitioners are playing the same game: trying to model the world with a formalism. Psychologists want this formalism to be as close as possible to humans; Semantic web practitioners want it to ‘just work’. For psychologists, a better formalism is one that models even human flaws and inconsistencies. For Semantic web practitioners, a better formalism is more expressive, while being as simple as possible; if a machine using it reaches conclusions that a human won’t, so much more impressive.

The concept of similarity is very different in psychology and in machine learning too. Machine learning (and in particular, computational linguistics) use structured representations, while most of the psychologists use mainly ‘flat’ representations. But the main difference is that the machine learning group often uses representations that are not psychologically plausible. For example, some parsers use human-coded representations of syntactic dependencies from corpora like TREEBANK [9], WordNet [10] or even Google queries. Semantic similarity according to Resnik [11] refers to similarity between two concepts in a taxonomy such as WordNet [10] or CYC upper ontology. These are of course not available to the mind; even though models may perform very well on interesting tasks, they have no psychological plausibility. Still, there seems to be some level of convergence between machine-learning and psychological approaches. This paper will try to make connections particularly where they are relevant for the semantic web paradigm.

## 2. What is Similarity, anyway?

The question “What is similarity” has inspired considerable research in the past, because it affects several cognitive processes like memory retrieval, categorization, inference, analogy, and generalization, to mention a few. We have divided current efforts to answer this question into four main branches: continuous features (spatial) models, set theoretic models, hierarchical models, and transformational distance. Similar classification can be found in Goldstone [12] and in Markman [13].

## 3. Continuous features (spatial) models

Shepard can be considered the father of metric models (models that use a multidimensional metric space to represent knowledge) in psychology. Shepard’s [14] *Science* paper, ‘*Toward a universal law of generalization for psychological science*’ is his most ambitious and definitive attempt to propose multidimensional spaces as an universal law in psychology. Shepard’s [14] main proposal is that psychologists can

#### 4 Jose Quesada

utilize metric spaces to model internal representations for almost any stimulus (i.e., shapes, hues, vowel phonemes, Morse-code signals, musical intervals, concepts, etc.).

We rarely encounter the exact same situation twice. There is always some change in the environment. Usually, this new environment has some physical resemblance to an environment with which we have some history. This incremental change is the crucial element--the more similar the new environment is to something we already know, the more we will respond in a similar way.

A metric space is defined by a metric distance function  $D$ , that assigns to every pair of points a nonnegative number, called their distance, following three axioms: minimality [ $D(A,B) \geq (A,A) = 0$ ], symmetry [ $D(A,B) = D(B,A)$ ], and the triangle inequality [ $D(A,B) + D(B,C) \geq D(A,C)$ ]. The methodological tool Shepard proposed is multidimensional Scaling [MDS, 15], a now-classic approach to representing proximity data. In MDS, objects are represented as points in a multidimensional space, and proximity is assumed to be a function of the distance in the space,  $p(i,j) = g[D(i,j)]$ , where  $g$  is a decreasing function (a negative exponential). The distance in the  $n$ -dimensional metric space that the MDS generates represents similarity, and is calculated using the Minkowski power metric formula:

$$D(i, j) = \left( \sum_{k=1}^n |X_{ik} - X_{jk}|^r \right)^{(1/r)} \quad (1)$$

Where  $n$  is the number of dimensions,  $X_{ik}$  is the value of the dimension  $k$  for entity  $i$ , and  $r$  is a parameter that defines the spatial metric to be used.

The vector space model from classical information retrieval capitalizes on this finding. It maps words to a space with as many dimensions as contexts exist in a corpus. However, the basic vector space model fails when the texts to be compared share few words, for instance, when the texts use synonyms to convey similar messages. Latent Semantic Analysis (LSA) [16, 17] solves this problem by running a singular value decomposition (SVD) and then dimension reduction on the term by document matrix. LSA can model human similarity judgments for words and text, but it faces problems. Some of these problems are conceptual: negation just doesn't work on any spatial models (NOT is a ubiquitous word and it forms a vector that adds nothing to the overall meaning). LSA uses a bag of words approach where word order does not matter; the semantic web approach requires machine learning algorithms that can produce structured representations from plain text. There are also problems with the implementation (scalability): the SVD is a one-off operation that assumes a static corpus. Updating the space with new additions to the corpus is possible, but not trivial.

LSA spawned a plethora of models for extracting semantics from text corpora. Some of them partially address structured representations. For example the Topic model [18] could potentially use a generative model with several layers of topics (hierarchical models). Beagle [19] proposes methods to capture both syntax and

semantics simultaneously in a single representation using convolution. Beagle uses a moving window, so only close sequential dependencies make an impact in its understanding of syntax; it is still far from delivering a fully automatic propositional analysis of text.

Another approach is to use a large corpus of labeled articles as dimensions. For example, any text can be a weighted vector of similarities to Wikipedia articles [20]. This currently produces the highest correlation to human judgments of similarity (.72 vs .60 for LSA).

Although recent developments have addressed some implementation issues (e.g., the SVD can now be run in parallel) the direct application of LSA or any other statistical methods to semantic web problems is still not obvious. RDF operations are logical; in LSA vectors are obtained using statistical inference. Combining the logic and statistical approaches seems to be a worthwhile goal and some groups are pursuing it [21, 22].

#### 4. Discrete set theoretic models

Tversky's set-theoretic approach and Shepard's metric space approach are often considered the two classic – and classically opposed – theories of similarity and generalization (although Shepard has some research on the set-theoretic approach, e.g., [15, 23]).

Metric spaces have problems as a model for how humans represent similarities. Amos Tversky [24] pointed out that violations of the three assumptions of metric models (minimality, symmetry, and the triangle inequality) are empirically observed.

Minimality is violated because not all identical objects seem equally similar; complex objects that are identical (e.g., twins) can be more similar to each other than simpler identical objects (e.g., two squares).

Tversky [24] argued that similarity is an asymmetric relation. This is an important criticism for models that assume that similarity can be represented in a metric space, since metric distance in an Euclidean space is, of course, symmetric. He provided empirical evidence, for example, when participants were asked a direct rating, the judged similarity of North Korea to China exceeded the judged similarity China to North Korea<sup>1</sup>. A second criticism relates to the fact that similarity judgments are subjected to task and context-dependent influences, and this is not reflected in pure metric models.

---

<sup>1</sup> However, results from Aguilar and Medin 25. Aguilar, C.M., Medin, D.L.: Asymmetries of comparison. *Psychon. Bull. Rev.* **6** (1999) 328-337 suggest that similarity rating asymmetries are only observed under quite circumscribed conditions.

Another important criticism focuses on the triangle inequality axiom, which says that distances in a metric space between any two points must be smaller than the distances between each of the two points and any third point. In terms of similarities, this means that if an object is similar to each of the two other objects, the two objects must be at least fairly similar to each other [26]. However, James [27] gives an example in which this does not hold true: the moon is similar to a gas jet (with respect to luminosity) and also similar to a football (with respect to roundness), but a gas jet and a football are not at all similar.

Tversky proposed that similarity is a function of both common and distinctive features, as described in the formula:

$$S(A, B) = f(\Theta(A \cap B) - \alpha(A - B) - \beta(B - A)) \quad (2)$$

Where  $A$  and  $B$  are feature sets. The similarity of  $A$  to  $B$  is expressed as a linear combination of the measure of the common ( $A \cap B$ ) and distinctive ( $A - B, B - A$ ) features. The parameters  $\Theta$ ,  $\alpha$ , and  $\beta$  are weighing parameters given to the common and distinctive components, and the function  $f$  is often simply assumed to be additive (i.e., all features are independent and their effects combine linearly).

To respond to these criticisms, some researchers have proposed different solutions that basically extend the assumptions of metric models and enable them to explain the violation in the three assumptions. Nosofsky [28] defended the metric space approach arguing that asymmetries in judgments are not necessarily due to asymmetries in the underlying similarity relationships. For example, in word similarity judgments, if the relationship  $A \rightarrow B$  is stronger than  $B \rightarrow A$ , a simple explanation could be that word  $B$  has higher word frequency, is more salient, or its representation is more available than word  $A$ .

Krumhansl [26] has proposed that some objections to geometric models may be overcome by supplementing the metric distance with a measure of the density of the area where the objects that figure in the comparison are placed. Krumhansl argued that if  $A \rightarrow B$  is stronger than  $B \rightarrow A$ , an explanation is that  $A$  is placed in a sparser region of the space. For example, in LSA the nearest 20 neighbors of "China" range between .98 and .80. However, the 20 nearest neighbors of "Korea" range between .98 and .66, which means "China" is in a denser part of the space than "Korea". One could argue that although Krumhansl's explanation does propose a solution for the problem, the resulting modified distance function need not satisfy the metric axioms anymore.

Kintsch [29] offered yet another way of modeling asymmetric judgments using a metric model. In his predication model, Kintsch substitutes the productivity rule in LSA (addition) with more sophisticated mechanisms that related the neighborhood of the predicate and argument to create a composed vector. His model is another source of evidence of theories that, using metric underlying models, can explain phenomena

that conflict with the metric assumptions. As well, there seems to be controversy about how much the stimulus density can affect psychological similarity [30-32].

In summary, it seems that *supplemented* metric models can explain most of the criticisms attributed to them, and that some of the traditional effects such as context effects and asymmetry of similarities can be due to additional factors not considered in the classical explanations.

There used to be no feature models able to work with plain text corpora and generate, but recently the Bayesian camp has proposed a few. The most successful of these is the Topic model. Griffiths, Steyvers, and Tenenbaum [18] propose that representation might be a language of discrete features and generative Bayesian models instead of continuous spaces. This bottom-up approach has the advantage of generating ‘topics’ instead of unlabelled dimensions, so the resulting representations are ‘explainable’. The Topic model can also explain asymmetries in similarities, because conditional probabilities are indeed asymmetrical ( $P(A|B) \neq P(B|A)$  necessarily).

The Topic model is indeed a feature model because ‘the association between two words is increased by each topic that assigns high probability to both and is decreased by topics that assign high probability to one but not the other, in the same way that Tverksy claimed common and distinctive features should affect similarity’ [18 p. 223].

At the implementation level, the Topic model is not memory-intensive; since it is a Markov chain Montecarlo model, it simply allocates words to topics in an iterative way.

The combination of explainable dimensions and possibility to handle structured representations makes the Topic model an interesting choice for the representation problems the semantic web encounters. Still, the level of structural complexity that current topic models can derive from text is very basic. Future implementations may be able to accommodate more realistic structures because the overall probabilistic framework is more flexible than previous vector space models. For promising new ways of combining ontologies with bottom-up topics, see [33, 34].

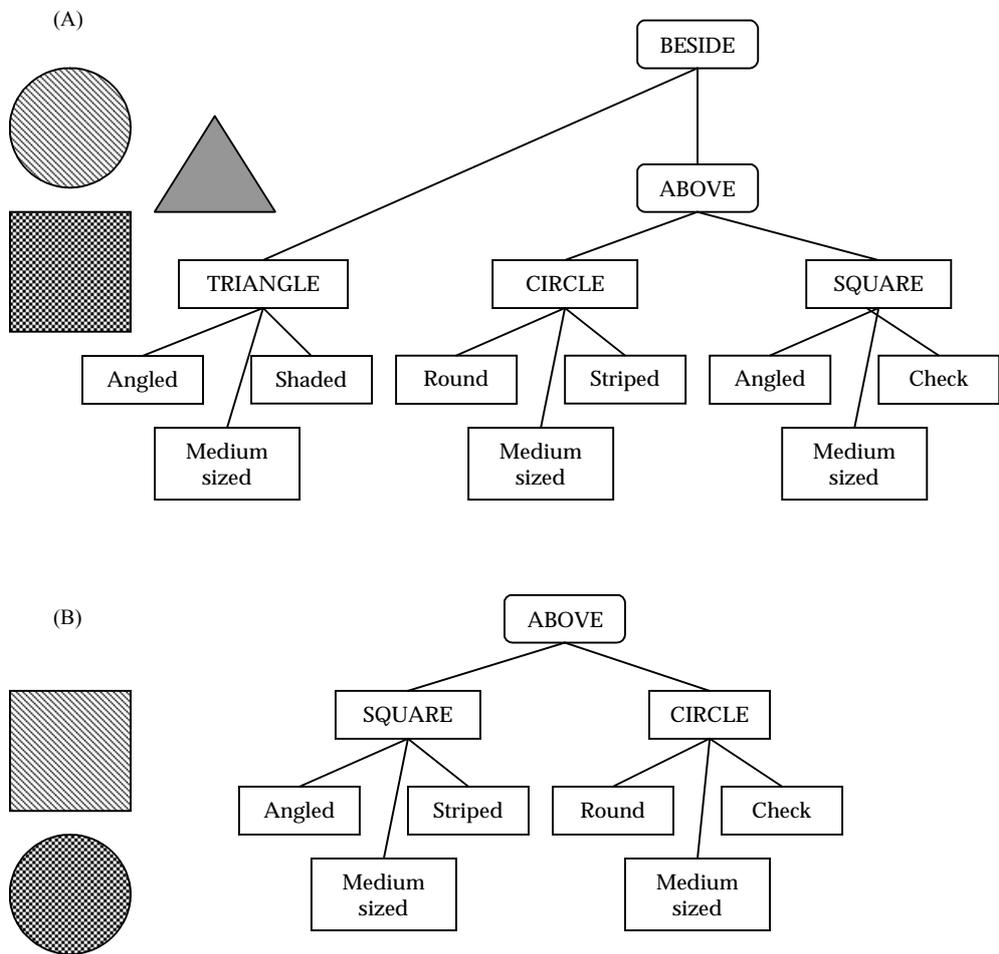
## 5. Hierarchical models and alignment-based models

Some researchers [e.g., 7, 12, 35] argued that neither spatial models nor discrete set theoretic models are well suited to model human representation. In several experiments humans show evidence of using structured representations rather than a collection of coordinates or features.

The structural matching theory assumes that mental representations consist of hierarchical systems that encode objects, attributes of objects, relations between

objects, and relations between relations [13]. Structure mapping models are then the closest to the data structures that the semantic web uses (RDF).

The two sets of objects (A) and (B) in Figure 1 would be represented by the hierarchical structures (a) and (b). What are represented as a hierarchical system are the features of one objects, and the comparison between two mental representations consists on aligning the two structures so the matching is maximal. The best structural matching possible determines the similarity between the two objects. In Figure 1, page 8, the best interpretation involves matching the "above" relations, since they are a higher-level connected relational structure than, e.g., "circle".



**Fig. 1:** Example of structured representations, and structural alignment [adapted from 13, p. 122]. The trees represent the features, keeping the structure. Rounded boxes are relationships,

uppercase square boxes are objects, and lowercase boxes are features. The “above” relation is directional; “Above” (square, circle) is different than “above” (circle, square).

The details on how the matching is done vary with the different models; The structure mapping engine SME [36] was the original; it works by forcing one-to-one mappings. That is, it limits any element in one representation to corresponding to at most one element in the other representation. SIAM [37] is an spreading activation model; it consists of a network of nodes that represent all possible feature-to-feature, object-to-object, and role-to-role correspondences between compared stimuli. The activation of a particular node indicates the strength of the correspondence it represents. SIAM treats one-to-one mapping as a soft constraint.

Structured representations gain some of their power from the ability to create increasingly complex representations of a situation by embedding relations in other relations and creating higher-order relational structures. These higher-order structures can encode important psychological elements like causal relations and implications [13]. In fact, RDF as a data structure has this property (reification, also called compositionality [38]). Currently compositionality is hard to implement for metric models and feature models.

So how are current structure-matching models in psychology different from the similarity models used in semantic web applications? The psychological models use very simple and artificial materials, like those in Figure 1. Most published papers contain a few examples where the model works (i.e., the solar system mapped to Rutherford’s model of the atom) but not about where it fails. There is no published study on how general a model is (i.e., using a large selection of objects) nor what the boundary conditions are. More thorough testing and model comparison is needed. The overall impression is that fine-tuning the model to the examples in the paper took a good amount of time for the experimenter, so doing this for a large representative sample of structures may be time consuming. Second, psychological similarity models stress the importance of working memory capacity limitations, which have no relevance for machine learning and general usage in applications. Working memory limitations may help the model explain human patterns such as common errors, but do not contribute to better applications. Third, scaling may be an issue. The Rutherford example requires 42 and 33 nodes to represent the solar system and atom, respectively, and it is one of the largest mappings published. Semantic web applications can easily deal with knowledge bases several orders of magnitude larger (Although see [39, 40] for some examples of SME applications with larger knowledge bases). Last, all these theories use hand-built representations. Information extraction is a type of information retrieval whose goal is to automatically extract structured information, i.e. categorized and contextually and semantically well-defined data from a certain domain, from unstructured machine-readable documents. To date, no psychological theories of the structured kind do information extraction or propose an alternative solution to avoid hand-built representations.

So, is there no way to derive structured representation automatically from text to avoid all the above problems? The next section includes the latest, and most promising line of work: transformational distance.

## 6. Models based on Transformational distance

For transformational distance theories similarity of two entities is inversely proportional to the number of operations required to transform an entity so as to be identical to another [e.g., 41, 42-45]. The idea of similarity as transformation is promising in that it is very general and seems able to solve some of the previous theories problems.

We will review the representational distortion theory [8, 46], and the SP model [45, 47]. The representational distortion theory of Hahn and Chater [8, 46] uses a measure of transformation called Kolmogorov complexity,  $K(x|y)$  of one object,  $x$ , given another object,  $y$ . This is the length of the shortest program which produces  $x$  as output using  $y$  as input. The main assertion of the theory is that representations that can be generated by a short program are simple, and the ones that require longer programs are more complex. For example, a representation consisting in a million zeroes, although long, is very simple, whereas the sentence "Mary loves roses" is shorter but more complex. With this Kolmogorov measure of complexity, a similarity measure can be defined as the length of the shortest program that takes representation  $x$  and produces  $y$ . That is, the degree to which two representations are similar is determined by how many instructions must be followed to transform one into another. This approach to similarity implements the minimality and triangle assumptions (like metric theories), but enables the relationships between items to be asymmetrical, escaping one of the most pervasive criticisms of metric theories, namely the asymmetry in human similarity judgments. Note that the representational distortion theory needs to propose a vocabulary of basic representational units and basic possible transformations; but this vocabulary is currently not specified. However feature theories do not explain where features come from, so the transformational view is not at a disadvantage.

Another approach to measure transformational distance is string edit theory. The string edit theory centers on the idea that a string (composed by words, actions, states, amino acids, or any other element) can be transformed into a second string using a series of "edit" operations. String edit theory uses basic transformations like (insert, delete, match, and substitute), although this basic set varies in different implementations. Each "edit" operation for each particular item has a probability of occurrence associated. For example, in a perceptual word recognition task, the probability of substituting M for N could be higher than the probability of substituting M for B. These probabilities are defined a-priori and reflect the "cost" of the operation, but can also be learned for each problem. There is always more than one sequence of operations that can transform a string into a second string. Each sequence of operations has a probability too, which is the average of the probabilities of the transformations that form part of it.

The most well-developed model of cognition based on string edit is the syntagmatic paradigmatic (SP) model [45]. SP proposes that people use large amounts of verbal knowledge in the form of constraints derived from the occurrences of words in different slots. The constraints are categorized in two types: (1) syntagmatic

associations that are thought to exist between words that often occur together, as in "run" and "fast" and (2) paradigmatic associations that exist between words that may not appear together but can appear in the same sentence context, such as "run" and "walk". The SP model proposed that verbal cognition is the retrieval of sets of syntagmatic and paradigmatic constraints from sequential and relational long-term memory and the resolution of these constraints in working memory. When trying to interpret a new sentence, people retrieve similar sentences from memory and align these with the new sentence. The set of alignments is an interpretation of the sentence. For instance, to build an interpretation of the sentence "Mary is loved by John" they might retrieve from memory "Ellen is adored by George", "Sue who wears army fatigues is loved by Michael", and "Pat was cherished by Big Joe", leading to the following interpretation:

Mary					is	loved	by	John
Ellen					is	adored	by	George
Sue	who	wears	army	fatigues	is	loved	by	Michael
Pat					was	cherished	by	Big Joe

The set of words that aligns with each word from the target sentence represents the role that the word plays in the sentence. So, in the example [Ellen, Sue, Pat] represents the lovee role and [George, Michael, Joe] the lover role. The model assumes that any two sentences convey similar factual content to the extent that they contain similar words aligned with similar sets of words. Note that SP does not assume any previous knowledge (i.e., syntax). The model can solve basic question-answering tasks such as which tennis player won a match when trained on a specific plain text corpus of such news [47].

Both XML and RDF are data languages of labeled trees, and of course tree edit distance is a subclass of string edit theory [48]. There are several algorithms proposed to match such structures efficiently. For example Bertino et al [49] propose a way to match an XML tree to a set of trees (DTDs) in polynomial time. Thus, once the starting knowledge base is in a structured form, there are algorithms to do similarity operations either efficiently or in a cognitively plausible way, but not both. The remaining step is to get from a flat form to a structure that satisfies the requirements of the algorithms, which has proven not to be easy. This step is not necessary for models such as SP, since they work from plain text. In this sense this is a promising venue. Contrary to the semantic web idea to create domain-specific data languages by agreement and force that structure onto existing text in the wild, SP proposes no structure a priori. In fact, SP captures meaning as sentence exemplars. The difficult task of either defining or inducing semantic categories is avoided.

Both theories (string edit theory and on Kolmogorov complexity) deal with structured representations, feature representations and continuous representations if needed. Of course, feature theories can argue that each of the transformations proposed can be added as a feature without leaving the feature approach. However, adding higher order relationships as features makes evident one of the weak points of feature theories: anything can be a feature. Which transformations are allowed? What do people actually use? Is there a general transformation vocabulary that works for

any domain? Such vocabulary, if it exists, should be independent of the transformations' characteristics (for example, their salience); otherwise, the description in feature terms becomes redundant, and could be eliminated without losing explanatory power. Because of this, the representational distortion theory proposes transformations as explanatorily prior. Feature models constitute a subset of the family of representational distortion theories, where similarity between objects is defined using a very limited set of transformations: feature insertion, feature deletion, or feature substitution. These are exactly the same transformation sets that the SP model proposes for sentence processing. However, the SP model escapes the former criticism because the "features" (in this case, words) are not generated ad-hoc, but learned empirically by experience with real-world text corpora. But the question of whether there is a viable universal transformation language still stands.

Transformational distance models could be more general than Tversky's contrast model. This view is shared by Hahn and Chater [8, pp. 71-72]: "indeed, the [Kolmogorov complexity] model can be viewed as a generalization of the feature and spatial models of similarity, to the extent that similar sets of features (nearby points in space) correspond to short programs". Chater and Vitanyi [50, 51] have mathematical proof that any similarity measure reduces to information distance.

## **7. Summary and Conclusion**

We have presented why similarity is a hard problem and four major psychological theories that tried to solve it. We started the discussion presenting metric models and their flaws; which were partially addressed by feature theories. Then we presented structural alignment models, explaining how they relate to current work on structured data such as RDF. We concluded with transformational distance models as the closest to an ideal solution.

One recurring theme is that once the starting knowledge base is in a structured form, there are algorithms to do similarity operations either efficiently [49] or in a cognitively plausible way [52] (but not both). The remaining step is to get from a flat form to a structure that satisfies the requirements of the algorithms, which has proven not to be easy. Currently the SP model and the Topic model show promise as bottom-up models that start with plain text and generate structured representations. The immediate advantage when compared with traditional machine learning information extraction tools is that they do not require preexisting classes (as they are inferred). Admittedly, both SP and Topic models still have a long way to go, and up to now they have focused in extraction of syntactic categories (and in an imperfect way). The semantic web of course needs an entire universe of different categories (not only syntactic).

The semantic web practitioners however are perfectly happy manually creating domain-specific languages to describe their domains (i.e., RDF-schema). This is good news because it increases the number of similarity models one can choose from. SP

and the Topic model have the head start of making no *a priori* commitment to particular grammars, heuristics, or ontologies. But this may not be a tremendous advantage in a world that seems to be eager to produce ontologies and fit all existing knowledge into those structures. Time will tell if bottom-up approaches will proliferate or fade away.

## References

1. Heit, E., Rotello, C.: Are There Two Kinds of Reasoning? Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society (2005)
2. Glymour, C.: The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology. MIT Press, Boston (2001)
3. Newell, A., Simon, H.A.: Human Problem Solving. Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1972)
4. Murphy, G.L., Medin, D.L.: The Role of Theories in Conceptual Coherence. *Psychol Rev* **92** (1985) 289-316
5. Goodman, N.: Seven strictures on similarity. In: Goodman, N. (ed.): *problems and projects*. Bobbs Merrill, Indianapolis (1972) 437-450
6. Medin, D.L., Goldstone, R.L., Gentner, D.: Respects for Similarity. *Psychol Rev* **100** (1993) 254-278
7. Goldstone, R.L.: The Role of Similarity in Categorization - Providing a Groundwork. *Cognition* **52** (1994) 125-157
8. Hahn, U., Chater, N.: Concepts and similarity. In: Lamberts, K., Shanks, D. (eds.): *Knowledge, concepts, and categories*. MIT Press, Cambridge, MA (1997)
9. Marcus, M., Marcinkiewicz, M., Santorini, B.: Building a large annotated corpus of English: the penn treebank. *Computational Linguistics* **19** (1993) 313-330
10. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Introduction to WordNet: An On-line Lexical Database\*. *International Journal of Lexicography* **3** (1990) 235-244
11. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. *Proceedings of the 14th International Joint Conference on Artificial Intelligence* **1** (1995) 448-453
12. Goldstone, R.L.: Similarity. In: Wilson, R.A., Keil, F.C. (eds.): *MIT encyclopedia of the cognitive sciences*. MIT Press, Cambridge, MA (1999) 763-765
13. Markman, A.B.: *Knowledge representation*. Lawrence Erlbaum Associates, Mahwah, NJ (1999)
14. Shepard, R.N.: Toward a universal law of generalization for psychological science. *Science* **237** (1987) 1317-1323
15. Shepard, R.N.: Multidimensional scaling, three-fitting, and clustering. *Science* **214** (1980) 390-398
16. Landauer, T., McNamara, D., Dennis, S., Kintsch, W.: *LSA: A road to meaning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc (2007)

17. Landauer, T.K., Dumais, S.T.: A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychol Rev* **104** (1997) 211-240
18. Griffiths, T.L., Steyvers, M., Tenenbaum, J.: Topics in semantic representation. *Psychol Rev* **in press** (2007)
19. Jones, M.N., Mewhort, D.J.K.: Representing Word Meaning and Order Information in a Composite Holographic Lexicon. *Psychol Rev* **114** (2007) 1-37
20. Gabrilovich, E., Markovitch, S.: Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence* (2007) 1606–1611
21. Bernstein, A., Kiefer, C.: Imprecise RDQL: towards generic retrieval in ontologies using similarity joins. *Proceedings of the 2006 ACM symposium on Applied computing* (2006) 1684-1689
22. Kiefer, C., Bernstein, A., Stocker, M.: The Fundamentals of iSPARQL: A Virtual Triple Approach for Similarity-Based Semantic Web Tasks. *LECTURE NOTES IN COMPUTER SCIENCE* **4825** (2007) 295
23. Shepard, R.N., Arabie, P.: Additive Clustering - Representation of Similarities as Combinations of Discrete Overlapping Properties. *Psychol Rev* **86** (1979) 87-123
24. Tversky, A.: Features of similarity. *Psychol Rev* **84** (1977) 327-352
25. Aguilar, C.M., Medin, D.L.: Asymmetries of comparison. *Psychon. Bull. Rev.* **6** (1999) 328-337
26. Krumhansl, C.: Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychol Rev* **85** (1978) 445-463
27. James, W.: *principles of psychology*. Holt, New York (1890)
28. Nosofsky, R.: Stimulus Bias, Asymmetric similarity, and classification. *Cognitive Psychol* **23** (1991) 94-140
29. Kintsch, W.: Predication. *Cognitive Science* **25** (2001) 173-202
30. Krumhansl, C.L.: Testing the Density Hypothesis - Comment. *J Exp Psychol Gen* **117** (1988) 101-104
31. Corter, J.E.: Testing the Density Hypothesis - Reply. *J Exp Psychol Gen* **117** (1988) 105-106
32. Corter, J.E.: Similarity, Confusability, and the Density Hypothesis. *J Exp Psychol Gen* **116** (1987) 238-249
33. Chemudugunta, C., Holloway, A., Smyth, P., Steyvers, M.: Modeling Documents by Combining Semantic Concepts with Unsupervised Statistical Learning. *7th International Semantic Web Conference, Karlsruhe* (2008)
34. Chemudugunta, C., Smyth, P., Steyvers, M.: Combining Concept Hierarchies and Statistical Topic Models. *ACM 17th conference on Information and Knowledge Management* (2008)
35. Markman, A.B., Gentner, D.: Structural Alignment During Similarity Comparisons. *Cognitive Psychol* **25** (1993) 431-467
36. Falkenhainer, B., Forbus, K., Gentner, D.: The Structure-Mapping Engine: Algorithm and Examples. *Artif. Intell.* **41** (1989) 1-63
37. Goldstone, R.L.: Similarity, Interactive Activation, and Mapping. *J. Exp. Psychol.-Learn. Mem. Cogn.* **20** (1994) 3-27

38. Fodor, J.A., Pylyshyn, Z.W.: Connectionism and Cognitive Architecture - a Critical Analysis. *Cognition* **28** (1988) 3-71
39. Klenk, M., Forbus, K., IL, N.U.E.: Cognitive Modeling of Analogy Events in Physics Problem Solving From Examples. Proceedings of the 29th Annual Meeting of the Cognitive Science Society meeting. NORTHWESTERN UNIV EVANSTON IL (2007)
40. Hinrichs, T., Forbus, K.: Analogical Learning in a Turn-Based Strategy Game. IJCAI - International Joint Conference on Artificial Intelligence, Hyderabad (2007)
41. Chater, N.: Cognitive science - The logic of human learning. *Nature* **407** (2000) 572-573
42. Chater, N.: The search for simplicity: A fundamental cognitive principle? *Q. J. Exp. Psychol. Sect A-Hum. Exp. Psychol.* **52** (1999) 273-302
43. Pothos, E.M., Chater, N.: A simplicity principle in unsupervised human categorization. *Cognitive Science* **26** (2002) 303-343
44. Pothos, E., Chater, N.: Categorization by simplicity: a minimum description length approach to unsupervised clustering. In: Hahn, U., Ramscar, M. (eds.): *Similarity and categorization*. Oxford University Press, Oxford (2001)
45. Dennis, S.: A memory-based theory of verbal cognition. *Cognitive Science* **29** (2005) 145-193
46. Hahn, U., Chater, N., Richardson, L.B.: Similarity as transformation. *Cognition* **87** (2003) 1-32
47. Dennis, S.: An unsupervised method for the extraction of propositional information from text. *Proceedings of the National Academy of Sciences* **101** (2004) 5206-5213
48. Rice, S., Bunke, H., Nartker, T.: Classes of Cost Functions for String Edit Distance. *Algorithmica* **18** (1997) 271-280
49. Bertino, E., Guerrini, G., Mesiti, M.: Measuring the structural similarity among XML documents and DTDs. *Journal of Intelligent Information Systems* (2008) 1-38
50. Chater, N., Vitanyi, P.: Simplicity: a unifying principle in cognitive science? *Trends Cogn Sci* **7** (2003) 19-22
51. Chater, N., Vitanyi, P.: The generalized universal law of generalization. *J Math Psychol* **47** (2003) 346-369
52. Larkey, L.B., Love, B.C.: CAB: Connectionist analogy builder. *cognitive Science* **27** (2003) 781-794