The Analogical Thesaurus

An Emerging Application at the Juncture of Information Retrieval and Analogical Reasoning

Tony Veale,
Department of Computer,
University College Dublin, Belfield, Dublin, Ireland.

Tony.Veale@UCD.ie

Abstract

Innovative applications often occur at the juncture of radically different domains of research. This paper describes an emerging application, called the analogical thesaurus, that arises at the boundary of two very different domains, the highly applied domain of information retrieval, and the esoteric domain of lexical metaphor interpretation. This application has the potential to not just to improve the utility of conventional electronic thesauri, but to serve as an intelligent mapping component in any system that uses analogical reasoning or case-based reasoning.

1 Introduction

A conventional thesaurus, electronic or otherwise, is designed to improve recall at the cost of precision. There are two mitigating reasons for this: first, since there are few, if any, real synonyms in language (this is especially so in English, which opportunistically imbues any near synonyms with different nuances of meaning), any word substitution will incur some loss of semantic precision; secondly, the very fact that someone is inclined to use a thesaurus suggests that they are unhappy with the precise meaning or connotation of the probe word, and are thus seeking a word that is not identical, merely similar, in meaning.

The issue of semantic precision becomes a good deal more complex when one considers the analogical purposes to which a thesaurus can be put. For example, suppose one wanted to know the Hindu, Roman or Semitic equivalents of the Greek gods Zeus, Ares and Athena, or to know the Muslim version of the bible, a church or a priest? Whereas a conventional thesaurus is indexed on a single probe word, analogical queries require both a source and a target term, to permit a mapping between two domains to be constructed. Thus, instead of a simple query "church" or "bible", one can pose much more specific queries like "Muslim church" (mosque), "Hindu bible" (the Vedas), "Celtic Ares" (Morrigan) or "Jewish German" (Yiddish). Semantic precision thus takes on a very different complexion when analogy is involved: though "mosque" and "synagogue" are not even near-synonyms, one can say that each forms a perfect correspondence with the other in the analogy of a "Muslim synagogue". Thus, one should differentiate between semantic precision (the basis of synonymy), and analogical precision (the basis of analogy and metaphor).

This paper demonstrates how an analogical thesaurus can be constructed from an existing thesaurus/taxonomy like WordNet (see [Miller, 1995]), by marrying techniques from information retrieval and lexical metaphor interpretation. By viewing analog-retrieval as process of information retrieval (IR) over a semi-structured text archive such as WordNet (which imposes a hierarchical structure onto the unstructured text of dictionary word definitions, or *glosses*), techniques such as query-expansion [Jing and Croft, 1994] can be used to maximize recall. Simultaneously, analogical reasoning techniques can be used to filter the results of IR to ensure that only precise domain counterparts are ever presented to the user. We argue that analogical precision is a more reliable indicator of retrieval utility in a thesaurus than semantic similarity, for while a word may evoke many near-synonyms, a well-structured analogy will often generate a single best mapping (e.g., see [Falkenhainer *et al.*, 1989]).

2 Thesaurus Retrieval

We begin by sketching a simple information-retrieval model of how an analogical thesaurus might work, and then critique this model to reveal the extra sophistication that is needed. Remember that retrieval is not geared to find near-synonyms, but analogical correspondences, so pre-existing associative structures (like the *synsets* in WordNet) will be of little help.

Given a source word s (like "Zeus") and a target domain word t (like "Hindu"), we retrieve all word senses c whose textual glosses contain the word t (such as *{yoga}*, *{fakir}*, etc.) From each candidate c_i on this list, we evaluate the structural similarity of c_i to every sense of s using a taxonomic metric (e.g., see

[Resnik, 1999]), such as distance to the lowest common hypernym. For example, *{Zeus}* and *{Varuna}* are structurally similar to the extent that they share the grandparent *{deity, god}*. Those candidate senses c_i whose similarity to s fall below a certain threshold are rejected, while those that remain are ranked in descending order of similarity and displayed to the user. This straw-man approach works well in many cases, and can answer such queries as "Who is the French Beckett?" (Victor Hugo) and "What is Hebrew German?" (Yiddish). However, there are significant problems that make it largely unusable.

First, the most desirable candidate words may not contain the target word t in their glosses: they may instead contain a synonym of t (like "Hindi" or "Hindustani"), or a word highly correlated with t (like "India" or "Trimurti"), or may only implicitly relate to t via a hypernym. For instance, the word sense *{Varuna}* has the following gloss in WordNet 1.6: "supreme cosmic deity"; no reference to any aspect of Hindu culture is present. However, *{Varuna}* is defined as a hyponym of *{Hindu_deity}*, and it is through this hyponym that relevance is determined.

Secondly, taxonomy-based discrimination of good candidates from mediocre and downright poor candidates is only effective on a coarse-grained basis. For instance, we can rely on the taxonomic structure of WordNet to recognize that hyponyms of *{deity, god}* (such as *{Aditi}* and *{Avatar})* make better candidates than hyponyms of *{person}* (like *{fakir}* or *{Gurkha}*). But we cannot depend on WordNet to discriminate between *{Varuna}* and *{Aditi}* on a taxonomic basis, since the relevant criterion (supremacy within one's pantheon) is not coded taxonomically in WordNet 1.6, but simply stated in the textual glosses.

The first problem can be resolved by adapting to the vagaries of word usage in WordNet sense glosses, by using query expansion to cast a wider retrieval net (e.g., see [Jing and Croft, 1994] who use a thesaurus to perform expansion). The second problem requires that we adapt WordNet itself, to make fine-grained taxonomic discrimination a reality. That is, we need to unlock the implicit structural information contained in WordNet's glosses, and reify this information to the level of taxonomic structure. Thus, *{Zeus}* and *{Varuna}* would become hyponyms of a new taxonomic node, *{Supreme_deity}*, such that the presence of this node will introduce greater discrimination into structural similarity metrics.

2.1 Query Expansion

We consider the easier of the two problems first. Suppose the query "What is the Muslim bible?" is posed to an analogical thesaurus: viewed as a problem of information retrieval, the goal is to retrieve the most similar word senses to (a sense of) "bible" that reside in the Muslim domain. Intuitively, the best answer a thesaurus can provide is *{Koran, Quran}*, which in WordNet 1.6 has the gloss: "sacred writings of Islam". However, the simple query "Muslim" will not retrieve this word sense, and clearly, neither will a query expanded with synonyms such as "Muslim or Moslem or Mohammedan".

We thus introduce the notion of 'symmetric associativity', a disciplined broadening of the notion of simple synonymy. By definition, synonyms are symmetric associates of each other since one can be substituted for the other without substantial loss of meaning (e.g., "Moslem" and "Muslim"). We broaden this notion to include terms that are so closely correlated in meaning that one can be used as a metonymic proxy for the other (e.g., "Muslim", "Islam", and "Koran"). This associativity can be determined statistically from a corpus, but a simpler and more principled method involves using the WordNet sense glosses themselves.

♦**Defn**: The *symmetric associates* of a word X comprises the set of synonyms of X, as well as the set of each word Y that appears in a definition/gloss of a sense of X such that X also appears in the definition of an individual sense of Y.

Thus "Islam" is a symmetric associate of "Muslim" since the former occurs in a definition of the latter and vice versa. Similarly by this reckoning, the symmetric associates of "Hindu" are *{Hindu, Hindoo, Hinduism, Hindustan, Hindustani, Trimurti}*, where the latter, "Trimurti", denotes a triad of divinities in Hindu mythology.

3 Taxonomic Discrimination

Taxonomies have, since antiquity (see [Hutton, 1982]), provided a systematic means of hierarchical decomposition of knowledge, whereby a domain is successively dissected via differentiation into smaller pockets of related concepts. Effective differentiation leads to effective clustering, so that similar concepts become localized to the same region of the taxonomy. This locality makes their inherent similarity easier to recognize computationally (e.g., see [Resnik, 1999]). For this reason, large-scale ontologies like Cyc [Lenat *et al.* 1990], a common-sense ontology for general reasoning, and WordNet [Miller, 1995], both organize their contents around a central taxonomic backbone.

Taxonomic systematicity implies that related or analogous domains should be differentiated in the same ways, so that similarity judgments in each domain can be comparable. But in very large taxonomies, this systematicity is often lacking. For example, in WordNet 1.6, the concept *{alphabet}* is differentiated culturally into *{Greek_alphabet}* and *{Hebrew_alphabet}*, but the concept *{letter, alphabetic_character}* is not similarly differentiated into *{Greek_letter}* and *{Hebrew_letter}*. Rather, every letter of each alphabet, such as *{alpha}* and *{aleph}*, is located under exactly the same hypernym, *{letter, al-*

phabetic_character}. This means that on structural grounds alone, each letter is equally similar to every other letter, no matter what alphabet they belong to (e.g., *alpha* is as similar to *aleph* as it is to *beta*).

An analogical thesaurus would thus be unable to separate good analogues from bad using structural similarity, and in examples such as "Jewish alpha", would return the enter Hebrew alphabet as candidates. To achieve competent analogical mapping then, it is vital that these deficiencies are automatically recognized and repaired. We thus identify an important class of taxonomic support structure for analogies that we dub an "analogical pivot", and show how taxonomies like WordNet, which contain relatively few natural pivots, can be automatically enriched with thousands of new pivots that significantly expand its potential for analogical reasoning. Though we limit our discussion to WordNet, we predict that these techniques are also applicable to ontologies like Cyc.

3.1 Analogical Composition

Consider again the analogical query "Hindu Zeus" and how one might resolve it using WordNet. The goal is to find a counterpart for the source concept Zeus (the supreme deity of the Greek pantheon) in the target domain of Hinduism. In WordNet 1.6, *{Zeus}* is a daughter of *{Greek_deity}*, which is turn is a daughter of *{deity, god}*. Now, because WordNet also defines an entry for *{Hindu_deity}*, it requires just a simple composition of ideas to determine that the "Hindu Zeus" will be daughter of the node *{Hindu_deity}*. More generally, one finds the lowest parent of the head term ("Zeus") that, when concatenated with the modifier term ("Hindu") or some synonym or symmetric associate thereof, yields an existing WordNet concept. In effect, the mapping process uses the pivot to construct a target counterpart of the source concept that significantly narrows the space of possible correspondences.

So the Hindu counterpart of Zeus is not *{Hindu_deity}*, but one of the relatively few daughter nodes of this target-domain differentiation of the pivot *{deity, god}*.

Compare this approach with the conventional one of taxonomic reconciliation, due to Aristotle [Hutton, 1982], in which two nodes can be considered analogous if they share a common superordinate. This approach still finds considerable traction in computational models today (e.g., see [Fass, 1998], [Way, 1991]), but it is easily trivialized: in a well designed taxonomy, any two nodes will always share at least one superordinate (even if it is the root node), and so any two concepts will always be potential analogues in such a system.

The current approach uses a much stricter notion of taxonomic analogy: two nodes are potentially analogous if they each possess superordinates that are themselves analogous differentiations of the same direct parent (the pivot of the analogy). Zeus and Varuna are analogous because *{Greek_deity}* and *{Hindu_deity}* are analogous, by virtue of being different domain specializations of the same pivot. This constraint is the taxonomic equivalent of the 'squaring rule' described in [Veale and Keane, 1997] to ensure that there is strong structural support for every analogical mapping. The approach is also constructive: it indicates how the target counterpart of the pivot, *{Hindu_deity}* is to be constructed from the source analogue (Zeus). The key is to first locate the pivot of the analogy and then follow its differentiation path into the target domain.

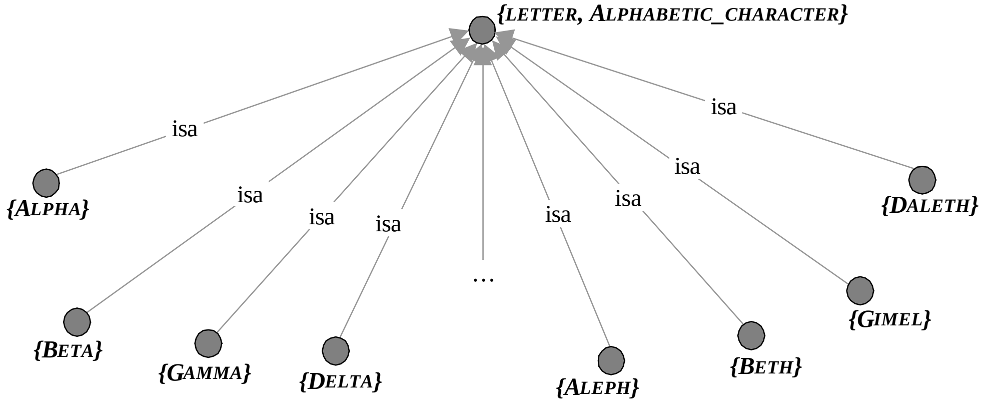

Figure 1: The impoverished sub-taxonomy of {letter, alphabetic_character} *as found in WordNet 1.6*

The above approach is still very fragile on a number of accounts. Firstly, natural pivots like *{deity, god}* are extremely rare in WordNet, which has not been explicitly constructed for analogical purposes. For instance, as noted earlier, the WordNet concept *{letter, alphabetic_character}* is not culturally differentiated, so a mapping cannot be constructed for "Jewish delta" → *{Hebrew_letter}*. Figure 1 illustrates the structure (and lack thereof) of the letter domain in WordNet 1.6.

Secondly, even when pivots do exist to facilitate a mapping, what is produced is a target hypernym rather than a specific domain counterpart. One still needs to go from *{Hindu_deity}* to *{Varuna}* (like Zeus a supreme cosmic deity, but of Hinduism), or from *{Hebrew_letter}* to *{daleth}* (like “delta” the fourth letter, but of the Hebrew alphabet).

4 Adding Fine-Grained Distinctions

Both problems can be solved by creating additional differentiating nodes that will dissect the taxonomy in new ways. In turn, this will convert the existing hypernyms of these nodes into analogical pivots whose hyponyms are explicitly labeled by domain.

For example, the creation of differentiator nodes like *{Greek_letter}* and *{Hebrew_letter}* will transform *{letter, alphabetic_character}* into a pivot that extends into the Greek and Hebrew domains (see Figure 2 below). Nodes such as these act as signposts from the pivot into specialized areas of the taxonomy and thus allow the first cut of the analogical mapping to occur. In contrast, other differentiator nodes may be less ambitious: a new node like *{1st_letter}* will unite just two hyponyms, *{alpha}* and *{aleph}*. However, these lower-level differentiators allow for finer-grained mapping within the established target domain, once the appropriate area of the taxonomy has been identified using the pivot.

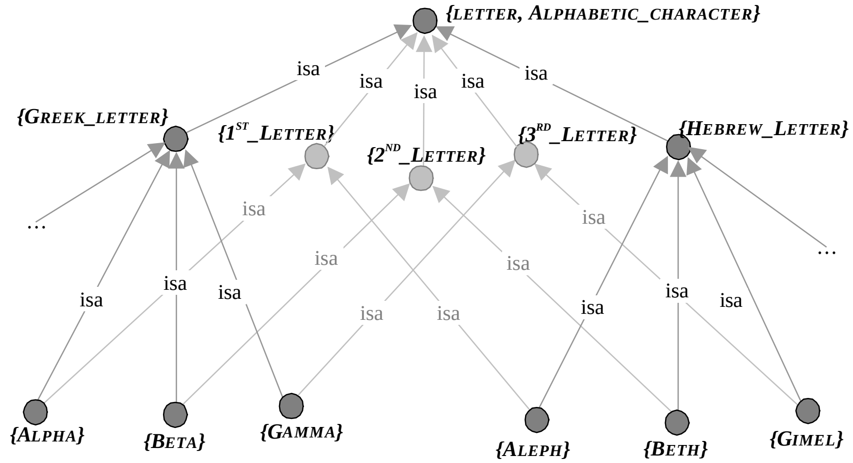

Figure 2: The taxonomic structure of {letter, alphabetic_character} *becomes a richly structured lattice when enriched with a variety of new types like* {Greek_letter} *and* {1st_letter}

Enhancing the differentiating power of WordNet is essentially a task of feature reification. WordNet (like other taxonomies, such as Cyc [Lenat *et al.* 1990]) expresses some of its structure explicitly, via *isa*-links, and some of it implicitly, in textual glosses intended for human rather than machine consumption. Fortunately, these glosses are consistent enough to permit automatic extraction of structural features (e.g., see [Harabagiu *et al.* 1999], who extract lateral connections between concepts from these glosses). What is needed then is a means to recognize the word features in these glosses with the most analogical potential, so that they may be ‘lifted’ to create new taxonomic nodes. Now, the noun sense glosses of WordNet 1.6 collectively contain over 40,000 unique content words (excluding articles, prepositions, etc.), but clearly only a small fraction of these words can be profitably reified. We thus employ two broad criteria to identify the word forms worth reifying, ‘differentiation potential’ and ‘alignment potential’:

♦**Defn**: A lemmatized word-form has *differentiation potential* if it occurs in more than one gloss, but not in too many (e.g., more than 1000). Additionally, there must be a precedent for using the word as an explicit differentiator in at least one existing taxonomic entry.

♦**Defn**: A word-form has *alignment potential* if it can be found in multiple locations of the taxonomy at the same relative depth from a potential pivot.

Consider the word “supreme”, which occurs in 42 different WordNet 1.6 glosses, enough to demonstrate cross-domain potential but not too many to suggest vagueness. Additionally, there are three WordNet precedents – *{supreme_court}, {supreme_authority} and {supreme_being}* – for its explicit use as a differentiator. And of the concepts that “supreme” is used to gloss, six – *{Zeus}, {Jove}, {Jupiter}, {Cronos}, {Wotan}* and *{Varuna}* – are grand-daughters of the concept *{deity, god}*. The symmetry one expects in analogy is thus present, since all that occupy the same depth in the taxonomy relative to the potential pivot *{deity, god}*. The word "supreme" thus has alignment potential relative to the concept *{deity, god}*, suggesting that "supreme" can be reified to create a new taxonomic concept *{supreme_deity}*. This situation is illustrated in Figure 3.

How does one identify the potential pivot nodes against which alignability is measured? In general, any interior non-leaf node of the taxonomy can be a potential pivot node, but from a practical perspective, it makes sense to only consider the atomic concepts that have not already been differentiated. Thus, *{deity, god}* is a potential pivot but *{greek_deity}* is not, since the latter is already specific to the *{Greek}* domain. For efficiency reasons, we currently employ the following heuristic for pivot identification:

♦**Defn**: A hypernym X is a potential pivot relative to a hyponym Y if X is the lowest, undifferentiated (atomic) hypernym of Y.

Thus, when we consider the word forms in the gloss of *{Zeus}*, alignability is determined relative to the concept *{deity, god}* rather than *{greek_deity}*, so that any reification that is performed will create a new differentiation of the former. We assume that an analogical thesaurus will exploit a reverse-index of gloss words to the concepts that are defined by them, making the identification of alignable features very efficient. The thesaurus simply needs to examine each concept reachable via the index entry for a probe word and consider only those at the same relative depth from the potential pivot.

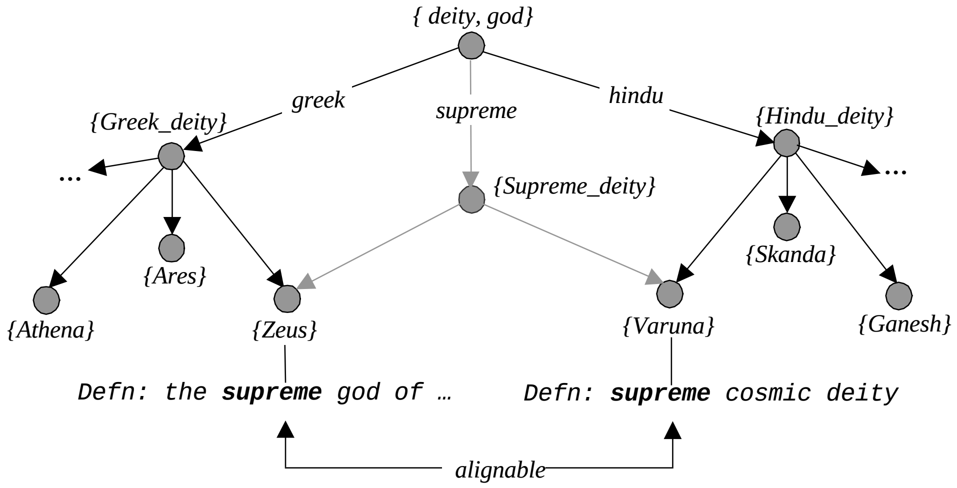

Figure 3: Analysis of the gloss for {Zeus} suggests that the word-form "supreme" has analogical potential, since it is alignable with another use in {Varuna} at the same relative depth from {deity, god}

Note that this process will sometimes create differentiation nodes that correspond to existing WordNet concepts. For example, the concepts *{bingo}* and *{billiards}* are each hyponyms of the concept *{game, contest}*, and the gloss of each contains the word-form "ball". This suggests that a new differentiator, *{ball_game}*, should be created to unite these and other concepts. However, while the meaning of differentiator nodes is entirely compositional, this cannot be assumed of existing WordNet entries. In WordNet 1.6, "ball game" is actually a synonym of "baseball", so to employ *{ball_game}* as a differentiator would erroneously make *{bingo}* and *{billiards}* a type of baseball. For reasons of safe inference then, new differentiator nodes are rejected if they correspond to existing WordNet entries.

5 Experimental Evaluation

To evaluate the effectiveness of an analogical thesaurus based on a pivot-enriched version of WordNet, 69,780 unique noun senses in WordNet 1.6 were analysed, to determine which would benefit structurally from the reification of one or more textual features. The glosses of these unique senses collectively contain 35,397 unique (unlemmatized) content words, but because of the strict reification criteria for feature-lifting from glosses, only 2806 of these content words are reified, to add 9822 new differentiator nodes, like *{cheese_dish}*, to WordNet. These nodes serve to differentiate 2737 existing nodes in WordNet, like *{dish}*, transforming these nodes into analogically-useful pivots.

In total, 18922 noun concepts (27% of the sample) are connected to the new differentiator nodes, via the addition of 28,998 new isa-links to WordNet. Each differentiator thus serves to unite an average of 3 daughters apiece. But in a subsequent pass over the new differentiator nodes, 1258 additions (or 12.8%) were culled because they did not sufficiently differentiate their parents to be worthwhile. For example, the node *{Greek_gorgon}* is worthless since all gorgons in WordNet are Greek.

A review of the other 87.2% of differentiators reveals that WordNet is being dissected in new and useful ways, both from the perspective of simple similarity judgments (e.g., the new types achieve a fine-grained clustering of similar ideas) and from the perspective of analogical potential. Overall, the most differentiating feature is "Mexico", which serves to differentiate 34 different pivots (such as *{dish}*, to group together *{taco}*, *{burrito}* and *{refried_beans}*), while the most differentiated pivot is *{herb, herbaceous_plant}*, which is differentiated into 134 subcategories (like *{prickly_herb}*). To consider just a few other domains: sports are differentiated into team sports, ball sports, court sports, racket sports and net sports; constellations are divided according to northern and southern hemispheric locales; food dishes are differentiated according to their ingredients, into cheese dishes, meat dishes, chicken dishes, rice dishes, etc.; letters are differentiated both by culture, giving Greek letters and Hebrew letters, and by relative position, so that "alpha" is both a *{1st_letter}* and a *{Greek_letter}*, while "Aleph" becomes both a *{1st_letter}* and a *{Hebrew_letter}*; and deities are further differentiated to yield *{war_deity}*, *{love_deity}*, *{wine_deity}*, *{sea_deity}*, *{thunder_deity}*, *{fertility_deity}*, and so on.

Dynamic types are primarily intended to increase the precision, rather than the recall rate, of analogical mapping. Consider, for example, the alphabet mapping task, in which the 24 letters of the Greek alphabet are mapped onto the 23 letters of the Hebrew alphabet (as represented in WordNet), and vice versa. The recall rate for the Hebrew to Greek letter task, for both dynamic and static WordNet hierarchies, is 100%, while for the reverse task, Greek to Hebrew, it is 96% (since Greek contains an extra letter). However, the precision of the static hierarchy is only 4%, since every letter of the target alphabet appears equally similar as a candidate mapping (Fig. 2), while for the dynamic hierarchy it is 96% (Greek to Hebrew alphabets) and 100% (Hebrew to Greek alphabets).

Table 1 presents a cross-section of the various subdomains of *{deity, god}* in WordNet as they are organized by dynamic types such as *{supreme_deity}*. Where a mapping is unavailable for cultural reasons, N/A is used to fill the corresponding cell. In two cases, marked by (*), an adequate mapping could not be generated when one was culturally available; in the case of *{Odin}*, this is due to the gloss provided by WordNet 1.6, which defines Odin as a "ruler of the Aesir" rather than the supreme deity of his pantheon; as for *{Apollo}*, a Greco-Roman deity, the failure is due to this entity being solely defined as a Greek deity in WordNet 1.6.

Common Basis	Greek	Roman	Hindu	Norse	Celtic
supreme	Zeus	Jove	Varuna	Odin *	N/A
wisdom	Athena	Minerva	Ganesh	N/A	Brigit
beauty, love	Aphrodite	Venus	Kama	Freyja	Arianrhod
sea	Poseidon	Neptune	N/A	N/A	Ler
fertility	Dionysus	Ops	N/A	Freyr	Brigit
queen	Hera	Juno	Aditi	Hela	Ana
war	Ares	Mars	Skanda	Tyr	Morrigan
hearth	Hestia	Vesta	Agni	N/A	Brigit
moon	Artemis	Diana	Aditi	N/A	N/A
sun	Apollo	Apollo *	Rahu	N/A	Lug

Table 1: *Mappings between sub-domains of the type {deity, god} in WordNet 1.6*

The data of Table 1 allows for 20 different mapping tasks in the deities domain (Greek to Roman, Roman to Hindu, etc.).

For the dynamic hierarchy approach (WordNet 1.6 augmented with pivot identification and creation techniques of section 3), the average recall rate is 61%. This moderate performance is as good as one can expect given the nature of the deity systems in these different cultures, since some pantheons are less fleshed out than others (e.g., the Norse to Hindu mapping has a precision of just 30% for this reason).

For the static hierarchy approach (WordNet 1.6 without additional pivot creation techniques), average recall is significantly lower at 34%, since many concepts (such as Varuna and Aphrodite) are not indexed on the appropriate reference terms due to poorly defined glosses (e.g., Varuna is defined as "supreme cosmic deity" in WordNet 1.6, with no explicit reference of Hinduism).

Average precision for the dynamic hierarchy approach is 93.5%, with the loss of 6.5% precision due to the items marked (*) in Table 1. In contrast, average precision for the static hierarchy approach is just 11.5%, and would be lower still if not for the indexing issues arising out of incomplete glosses, which help to reduce the number of incorrect answers the static hierarchy approach can actually retrieve.

6 Conclusions

Manually constructed representations on the ambitious scale of WordNet and Cyc are naturally prone to problems of incompleteness and imbalance. The 'one-size-fits-all' nature of the task results in a taxonomy that is often to undifferentiated for precise similarity judgments and too lopsided to support metaphor and analogical mapping. A symptom of this incompleteness is the fact that English glosses or commentaries provide the ultimate level of differentiation, so that one cannot truly differentiate two concepts without first understanding what the glosses mean. This understanding is vital to operation of an analogical thesaurus, so our goal is to lift implicit discriminators out of the flat text of the glosses and insert them into the taxonomy proper, to facilitate finer similarity judgments and richer analogical mappings.

We view the analogical thesaurus as a useful end-application in its own right, serving to enhance the creative reach of existing thesauri and writer's tools, while simultaneously increasing the precision of these tools. This is perhaps the most counter-intuitive aspect of the analogical thesaurus, as it allows a user to retrieve semantically distant word senses with greater precision than conventional thesauri allow the retrieval of near-synonyms. This paradox of sorts arises out of the distinction between semantic precision and analogical precision: though a more ambitious form of retrieval, analogical retrieval is guided by both a source and target marker to guide its deliberation, and one can talk of a perfect correspondence between highly dissimilar words (e.g., Bible and Koran, Zeus and Varuna, Gamma and Gimel, etc.).

Analogical precision can alleviate the 'potpourri' effect in thesaurus construction – namely, the tendency of thesauri to include long unordered lists of representative examples for a concept (like "orator", "general" or "church"), without providing a means of identifying those examples that are must pertinent to the user (see [Landau, 1989]). These lists serve to remind rather than truly inform, since a user can only directly exploit those elements that are already known and understood. In contrast, analogical queries allow a user to precisely and concisely state (yet in a non-committal way, that avoids the use of specific features) the nature of the desired term, and to be actually informed and surprised by the results. Thus, one can seek the "French Kant" (Blaise Pascal, perhaps) or the "Roman Hannibal" (Scipio Africanus) without being overloaded with long lists of other, irrelevant figures from history.

We also view the analogical thesaurus as a useful component in other intelligent applications, especially those that involve the comprehension of analogies and metaphors (see [Falkenhainer *et al.*, 1989], [Way, 1991], [Veale and Keane, 1997]). In particular, the analogical thesaurus should provide a substantial degree of additional semantic grounding to models of analogy that are primarily based on structure-mapping theory (e.g., [Falkenhainer *et al.*, 1989], [Veale and Keane, 1997]). Structure-mapping theory suggests that good analogies arise from the systematic mapping of two richly structured domain descriptions, such that entities within these descriptions are put in correspondence by virtue of occupying structurally similar positions. By relying on structure, distant analogies can be understood without requiring a common semantic basis. Nonetheless, when such a basis is available, it would be remiss to ignore it when evaluating the appropriateness of competing interpretations. For example, given two closely competing structural interpretations of a historical analogy, one would surely prefer an interpretation that maps Hannibal to Scipio than one with greater structural support that maps Hannibal to a tank. In this respect, an analogical thesaurus should prove invaluable in discriminating the most believable from the least tenable interpretations.

References

[Fass, 1988] Dan Fass. An Account of Coherence, Semantic Relations, Metonymy, and Lexical Ambiguity Resolution. In: S. Small, G. W. Cottrell and M. K. Tanenhaus (eds.): Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology and Artificial Intelligence. Morgan Kauffman: San Mateo, CA, 1988.

[Falkenhainer *et al*., 1989] Brian Falkenhainer, Kenneth Forbus, Dedre Gentner**.** (1989). Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41, pages 1-63, 1989.

[Harabagiu *et al.*, 1999] Sanda Harabagiu, George Miller and Dan Moldovan. WordNet 2 - A Morphologically and Semantically Enhanced Resource. *The Proceedings of the ACL SIGLEX Workshop: Standardizing Lexical Resources,* College Park, MD, USA, 1999.

[Hutton, 1982] James Hutton. *Aristotle's Poetics*. Norton: New York, 1982.

[Jing and Croft, 1994] Y. Jing and W. Bruce Croft. An association thesaurus for information retrieval. *The Proceedings of {RIAO}-94, 4th International Conference ``Recherche d'Information Assistee par Ordinateur''*. New York, NY. 1994.

[Landau, 1989] Sidney Landau. *Dictionaries: the Art and Craft of Lexicography*. Cambridge: Cambridge University Press, 1989.

[Lenat *et al*, 1990] Douglas B. Lenat, R. V. Guha. *Building Large Knowledge-Based Systems*. Addison Wesley: Reading, MA, 1990.

[Miller, 1995] George Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11).

[Resnik, 1999] Philip Resnik. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research* 11 pages 95-130, 1999.

[Veale and Keane, 1997] Tony Veale, Mark. T. Keane**.** The Competence of Sub-Optimal Structure Mapping on Hard Analogies. *The proceedings of IJCAI'97, the International Joint Conference on Artificial Intelligence,* Nagoya, Japan. 1997.

[Way, 1991] Eileen Cornell Way. Knowledge Representation and Metaphor. *Studies in Cognitive systems*. Kluwer Academic: Holland, 1991.