A Typology of Lexical Analogy in WordNet

Tony Veale,
School of Computer Science and Informatics,
University College Dublin,
Dublin D4, Ireland.
Tony.Veale@UCD.ie

Abstract

Analogy and metaphor are extremely knowledge-hungry processes, so one should question whether lightweight lexical ontologies like WordNet are sufficiently rich to support them. In this paper we argue that resources like WordNet are suited to the processing of certain kinds of lexical analogies and metaphors, for which we propose a spatially-motivated typology and a corresponding computational model. We identify two kinds of dimension that are important in lexical analogy – lexicalized (taxonomic) dimensions and ad-hoc (goal-specific) dimensions – and describe how these can be automatically identified, extracted and exploited in WordNet.

1 Introduction

Analogy is often used to highlight the inherent similarity between concepts that would otherwise appear uncomfortably alien, either because they reside in very different domains or because they are superficially incongruous. When conveyed through language, a lexical analogy is a fundamentally semiotic artifact whose interpretation hinges both on the structure of the concepts involved *and* the usage patterns of the words used to communicate these concepts. As such, lexical analogies can juxtapose words and concepts in a variety of guises, from the S.A.T-style of "Fructose is to fruit as Lactose is to what" (answer: milk) to the metaphoric style of "a milliner is a hat tailor" (e.g., see Miller, 1960; Rumelhart and Abrahamson, 1973; Turney *et al.*, 2003; Veale, 2004). This flexibility allows a model of lexical analogy to be applied not just in NLP, but in computer-aided learning, automated scholastic testing, and concept-based computer gaming.

Theories of analogy and metaphor are typically based either on graph-theoretic structure-mapping (e.g., see Falkenhainer *et al*, 1989; Veale and Keane, 1997) or on taxonomic abstraction (e.g., see Fass, 1988; Way, 1991; Veale, 2004). While the former is most associated with analogy, the latter has been a near-constant in the computational treatment of metaphor. In essence, structure-mapping assumes that the propositional meaning of a concept is expressed in an explicit, graph-theoretic form so that unifying sub-graph isomorphisms can be found between different concepts. In contrast, abstraction theories assume that analogous concepts, even if far removed in ontological terms, will nonetheless share a common subsuming category that will capture their propositional similarity. Thus, we should expect an analogous pairing like *surgeon* and *butcher* to ultimately share an abstraction that captures the notion of professionals who cut flesh (e.g., see Way, 1991; Veale, 2004).

Both perspectives have their followers and their uses. The structure-mapping perspective has been shown to be highly effective when rich, arbitrarily-complex propositions are used to describe the concepts of interest. The abstraction perspective has likewise been shown to be quite effective when the available domain knowledge is more tree-like and hierarchical (e.g., see Veale and Keane, 1997), as in an ontology with a strong taxonomic backbone. Princeton WordNet (see Miller, 1995), a lexical database of English words, is such a lightweight ontology, where the greatest concentration of discriminating structure is to be found in the abstraction hierarchy that differentiates different lexical concepts, or *syn*onym *sets*, from each other.

Veale (2003, 2004) has demonstrated that semi-structured semantic resources like WordNet are sufficiently rich to support the interpretation of S.A.T.-style lexical analogies. To more fully explore this potential, this paper describes a typology of the analogies that can be realistically supported by WordNet. By recognizing the inherent limitations imposed by WordNet, in particular its lack of explicit propositional content, we intend to identify a computationally productive sweet-spot in the space of possible lexical analogies.

2 A Spatial Perspective

A lexical ontology like WordNet provides a hierarchically-contoured space in which similarity judgments can be conceived spatially, either in terms of simple link distance, or in more nuanced information-theoretic terms as a measure of gloss-overlap. It makes sense, therefore, to also conceive a typology of lexical analogies in spatial terms. This builds on the model advocated by Rumelhart and Abrahamson (1973), who view analogy as occurring within a multi-dimensional *psychological space*. Within this space, pairs of concepts partake in a particular spatial relationship defined by a set of semantic axes, such as *size*, *ferocity* and *humanness* for animal analogies. It is the duty of a good analogy to preserve this spatial relationship when mapping pairs of concepts.

In spatial terms, one can view analogy as either a process of *projection* or of *translation*, occurring within a coordinate space defined by conceptual dimensions. Now, the projection

Petr Sojka, Key-Sun Choi, Christiane Fellbaum, Piek Vossen (Eds.): GWC 2006, Proceedings, pp. 105–110. © Masaryk University, 2005

view is more compatible with a structure-mapping perspective on analogy, since this entails that concepts will be represented as graphs whose topology should be preserved by analogical projection. In contrast, the translation view better suits WordNet's lack of explicit propositional structure, as lexical concepts are thus more usefully conceived of as points in a space. Analogy therefore becomes a process of spatial translation or relocation, whereby one or more conceptual dimensions are modified simultaneously.

Consider a culturally-situated word-concept like "Cognac". By translating this French concept along the dimension of Nationality, we can generate the cross-cultural equivalents such as *Rum Jamaica*, *Whiskey* (Ireland), *Port* (Portugal), *Vodka* (Russia) and *Sherry* (Spain). The more dimensions that can be identified for a given concept, the greater the potential for generating analogies, since any subset of these dimensions can be modified to translate the concept to a different part of the ontological space. A successful lexical analogy ensues if the source concept is translated to a point in space occupied by another lexical concept.

Now consider the possible analogues for the concept *Athena*, the Greek deity of wisdom and prudent warfare: by varying Nationality, one can generate *Minerva* (◇ *Roman*) or *Ganesh* (◇ *Hindu*); by varying Gender, one can also generate *Ares* (*female* ◇ *male*); and by varying Symbolism, one can generate other deities of the same culture and gender that personify a different theme, such as *Aphrodite* (*wisdom* ◇ *love*) or *Diane* (*wisdom* ◇ *moon*). By varying several dimensions simultaneously, an even wider range of candidate analogies is generated.

How many dimensions are needed to support lexical analogy? We might sidestep this issue using a statistical technique like Latent Semantic Indexing (see Landauer and Dumais, 1997) with which it is possible to assign each lexical concept in WordNet to a unique position in a coordinate space of, typically, 300 numeric dimensions, simply by treating the gloss of each concept as a representative bag of words. However, a symbolic approach can more transparently and economically identify just those conceptual dimensions that form a local analogy space around a given source concept. For example, given the source concept Athena, its local analogy space will be characterized by just the dimensions Nationality, Gender and Symbolism. Translation can occur along one or more dimensions simultaneously, to generate a range of meaningful candidate analogies.

3 A Dimensional Typology

We pragmatically characterize an analogy by the number of identifiable dimensions that are shared by the local spaces around its source and target concepts:

1-D: the local spaces share a single identifiable dimension, which is modified by the analogy. E.g., the *Beckett/Aeschylus/Hugo* space of writer analogies is 1-D, defined by the shared dimension of Nationality. Likewise, *Astronaut/Airman/Gondolier* space is 1-D, defined by the shared dimension Conveyance.

2-D: the local spaces share two identifiable dimensions, one or both of which can be modified. E.g., analogies like "Retsina is Greek Cognac" are 2-D, defined by the shared dimensions of Nationality (*French/Greek*, etc.) and Flavoring (*resin, sugar*, etc.).

3-D: the local spaces share three identifiable dimensions, any subset of which can be modified. E.g., the space of *Athena/Minerva/ Ares/Mars* analogies is 3-D, involving the three shared dimensions of Nationality, Gender and Symbolism.

n-D: the local spaces share n dimensions, though examples of analogies with $n > 3$ are rare in WordNet.

We can further enrich the typology by considering two additional factors: a) the number of dimensions that are actually modified in a given analogy; and b) the semantic distance between the source and target concepts. We can thus characterize lexical analogies as being of type $x/y/z$, $0 < x \leq y, z > 0$, where y is the number of identifiable dimensions shared by source and target, x is the number that must be modified to reach the target from the source, and z is the number of ISA links that must be traversed to find a common hypernym between source and target. By this reckoning, the analogy *Astronaut:Airman* is $1/1/1$ (the only shared dimension, Conveyance, is modified, while each share the same immediate hypernym), *Athena:Minerva* is $1/3/2$ (Nationality is altered while Gender and Symbolism are preserved, and both share a taxonomic grandparent), *Athena:Mars* is a $2/3/2$ analogy (Nationality + Gender altered), and so on.

3.1 Lexicalized and Ad-hoc Dimensions

Notice how membership in some dimensions seems to be determined by simple taxonomic subsumption, while membership in others is determined by an altogether more arbitrary kind of relationship. For instance, a dimension like Nationality contains all those values that one would expect to find organized under Nationality in a taxonomy like WordNet, such as *Greek, Irish, French*, etc. In contrast, a dimension like Symbolism, as exhibited by deity concepts, is defined by the tacit knowledge of what a deity can meaningfully represent to a given culture. Such dimensions seem to correspond to what Barsalou (1983) terms "ad-hoc" or "goal-directed" categories.

We refer to dimensions like Nationality as *lexicalized dimensions*, since membership is concisely defined by a single node in a lexical ontology. Following Barsalou, we refer to dimensions like Symbolism as *ad-hoc dimensions*, since it is difficult to see how mythological themes like *War, Wisdom, Love* and *Agriculture* can be taxonomically united under any single lexical concept. Indeed, because they are goal-directed, ad-hoc dimensions tend to cut across taxonomic categories, as is the case with the diverse value set *{fruit, milk, malt, wood}* that forms the ad-hoc dimension "things from which sugar can be extracted". The arbitrary goal-specific relationship that connects the members of an ad-hoc dimension must, for the most part, be identified on

a case-by-case basis with no explicit guidance from the taxonomy.

4 Generating Lexical Analogies

Given an n-dimensional source concept C_S, it is relatively straightforward to generate analogical targets by a process of translation along one or more dimensions. Simply select all subsets of these n dimensions and enumerate the possible values that each can take. When a given arrangement of values describes an existing lexical concept C_T, then an analogy C_S:C_T of type $m/n/z$, is identified, $m \leq n$ and $z \geq 1$.

For instance, given the 1-D source concept *Airman*, with the single dimension Conveyance=*Aircraft*, consider targets with alternative values for this dimension, such as Conveyance=*Spacecraft*. Since this latter setting is compatible with *Astronaut*, the lexical analogy *Airman*:*Astronaut (1/1/1)* is identified. Given the 2-D concept *Isaac-Newton*, any or both of the dimensions Nationality=*English* or Scientist-Type= *Mathematician* can be modified; the combination of Nationality=*English* (unchanged) and Scientist-Type=*Chemist* yields Newton's fellow countryman, *John-Dalton*; the combination Nationality=*Greek* (changed) and Scientist-Type=*Mathematician* (unchanged) yields Eratosthenes, among others; and the combination Nationality=*French* and Scientist-Type=*Chemist* (both changed) yields *Louis-Pasteur*.

In general, we can expect analogies of type $x/y/z$ to become more meaningful and more precise as y tends towards ∞ and both x and z tends toward 0. The $y - x$dimensions that remain unchanged serve to ground the analogy in literal similarity. For instance, in a lexical analogy between *Pizza* and *French-Toast*, the dimensions Nationality and Dairy-Product-Type are both modified (*Italian* $\diamond$ *French* and *Cheese* $\diamond$ *Milk*) while the lexicalized dimension of Baked-Good-Type remains unchanged (as *Bread*). But a literal grounding can also be provided by taxonomic similarity between C_S and C_T, allowing $y/y/1$ analogies to be quite satisfying also; e.g., consider the possible pairings of *Koran*, *Bible* and *Torah* (all $1/1/1$).

However, as x tends toward y for high values of y, the space of candidate targets can increase rapidly, as the number of possible translations grows. This can cause the best analogies to become swamped in a sea of implausible candidates. What is needed is a way of ranking analogies so that the most satisfying correspondences bubble to the top. However, if the list of candidate analogies is sorted in descending order of the value $y^{-1}xz$, the best candidates should achieve the highest ranking.

5 Identifying Dimensions in WordNet

To implement a model of analogy around a third-party knowledge source like WordNet, one must adopt a pragmatic perspective on conceptual representation. Since the notion of dimension is not explicit in WordNet, we take the dimensions of a lexical concept C to be just those that can be automatically extracted. The richest vein to mine comprises those elements of WordNet that carry meaning but which are denied explicit semantic representation, such as definition glosses.

5.1 Lexicalized and Ad-hoc Dimensions

The most populous lexicalized dimension in WordNet is that of Nationality, which affords an analogical bridge between foods, people and artifacts. WordNet contains over 6000 glosses that mention a proper-named nationality, and over 8000 that mention a proper-named country. By generalizing these references, we can situate the culture-specific concepts that mention them along a national dimension.

In general then, the relevant lexicalized dimensions for a lexical concept C can be identified by generalizing the other concepts mentioned in the gloss of C. It is important to generalize enough to allow any analogy to transcend domain boundaries, but not so much as to trivialize the process. A good balance is achieved by considering the most specific non-compound hypernyms of any taxonomic terms that are mentioned. As such, any set of elements that can be united by a common hypernym can constitute a lexicalized dimension. For instance, the concepts *Bible, Torah, Koran, Avesta* and *Upanishad* are recognizable as analogical counterparts by virtue of having glosses that mention the religions *Christianity, Judaism, Islam, Zoroastrianism*, and *Hinduism* respectively. These terms are relatively unambiguous, but since WordNet glosses are primarily for human consumption, it helps if glosses have already been disambiguated and sense-tagged, as in *eXtended WordNet* (see Mihalcea and Moldovan, 2001).

Since lexicalized dimensions correspond directly to nodes in a taxonomy, it is straightforward to enumerate their allowable values. For instance, *Cake/Bread/Pastry* are all substitutable values for the dimension Baked-Good-Type, facilitating analogies between *Pizza, Sandwich, Layer-Cake* and *Turnover*. Likewise, *Philosopher, Theologian, Scientist* and *Alchemist* define the lexical dimension *Intellectual*, supporting analogies among *Kant, Aquinas, Ohm* and *Faust*.

5.2 Identifying Ad-Hoc Dimensions

In ontologies like WordNet that permit, but tend to eschew, multiple-inheritance, ad-hoc categories are often secondary taxonomic indexes that fail to be represented hierarchically. Consider all compounds of the form *X-god* in WordNet. Each such compound is defined as a specialization of the concept *God*, yet in each case the modifier X has only a lexical, rather than a semantic, representation. Nonetheless, the diverse modifier set *{sea, sun, forest, war, earth, goat, snake, father}* forms a meaningful ad-hoc dimension that can be glossed as *"things that can be personified as gods"*. Of course, deriving such a name is beyond a system powered by WordNet alone, but a useful schematic name can be derived automatically for this ad-hoc dimension: God-Modifier-Type. Once this dimension has been identified, it can be used to generate analogies between deities, e.g., *Poseidon:sea::Apollo:sun*. In many cases, these ad-hoc dimensions remain unchanged in an analogy, and provide instead the literal pivot on which the analogy turns. For instance,

"Poseidon is the Greek Neptune" is more satisfying than "Apollo is the sun Poseidon".

In general, whenever WordNet contains a synset pair {C_S, M_S-H, ...} and {C_T, M_T-H, ...}, the lexical analogy C_S:M_S::C_T:M_T can be generated via the ad-hoc dimension H-Modifier-Type = {M_S, M_T, ...}. For instance, the synset pair *{selenolatry, moon-worship}* and *{heliolatry, sun-worship}* suggests the analogy *selenolatry:moon::heliolatry:sun*. Here the relevant ad-hoc dimension is named Worship-Modifier-Type, and unites the WordNet modifiers *{sun, moon, devil, animal, fish, Bible, word, idol, miracle, woman, fire, tree, symbol, snake, place, self}* into the goal-specific category "things that are often worshipped".

Ad-hoc dimensions can also be identified via the textual glosses of WordNet entries. However, because the relationship underlying these dimensions is arbitrary – unlike the purely subsumptive basis of lexicalized dimensions – more complex pattern-matching is required. We employ a template-based approach to extract dimensions, matching variations of the syntactic pattern *X-who|which|that-Verb-Y* against WordNet glosses to extract relationships of the form *Verb(X, Y)*. For instance, the set {newspaper, milk, coal, panegyric, eulogy, mail, speech, oration} represents the ad-hoc dimension "things that are delivered by specific people", to which the schematic name Deliver-Object-Type is automatically assigned.

Sometimes the relational verb is too vague to carry an analogy, as with "is", "have" or "does", while ofttimes the verb is too specific to cross domain boundaries, as with "brews", "tans" and "foments" (each is found in only one context). Between these easily filtered extremes is a productive middle ground which yields over 100 verb-based dimensions with two or more WordNet instantiations, such as "repair" (8 instances), "play" (18), "practice" (25), "sell" (25), "worship" (19), "study" (15) and "make" (59). The choicest verbs are polysemous, for since their multiple related senses are already linked via figurative extension, they can readily be employed for analogical purposes, e.g., "deliver" (messages and products), "play" (instruments and games), and "specialize in" (disciplines, activities, foods, etc.).

6 WordNet's Analogical Potential

The utility of this dimension-based perspective hinges on our ability to identify enough dimensions in WordNet to support a rich body of analogies. Given the taxonomic emphasis of WordNet, lexicalized dimensions are certainly the most numerous and the easiest to identify. However, a lexicalized dimension is only worth identifying only if: a) it serves to connect two or more different concepts at the same degree of specificity (e.g., if they reside at the same depth in the WordNet taxonomy); and b) it exhibits a range of values (at least two or more) over which analogical translation can occur.

Table 1 reveals how many *noun:noun* analogies of different dimensionalities can be generated from the lexicalized dimensions of eXtended WordNet (Mihalcea and Moldovan, 2001). The source and target concepts of each analogy are drawn from the same depth of the noun taxonomy, while their distance is estimated as the number of ISA links to their *Most Specific Common Ancestor* or MSCA.

Table 1: Number of analogies (in 1000s) of each dimensionality that can be generated from lexicalized dimensions in WordNet.

Distance	**1-D**	**2-D**	**3-D**	**4-D**	**5-D**
1	840k	130k	28k	14k	5k
2	1887k	287k	26k	12k	1.5k
3	2264k	331k	40k	17k	3.4k
4	2454k	338k	36k	18k	3.4k
5	2412k	276k	27k	12k	1.6k
6	2189k	233k	22k	8k	0.73k

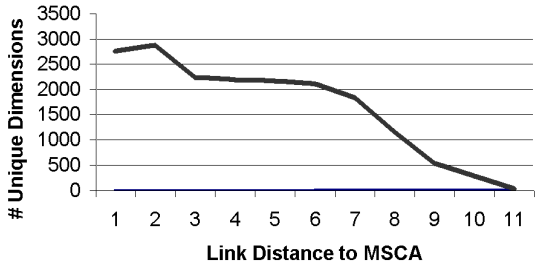

Figure 1: No. of unique lexicalized dimensions by distance.

Table 1 shows that the number of analogies grows with the allowable distance between source and target concepts, as one might expect, supporting our intuition that analogies of type $x/y/z + 1$ are more numerous than those of type x/y/z. In contrast, analogies of type or type $x/y/z$ appear more numerous than those of type or type $x/y + 1/z$, a finding also in keeping with intuition. Overall then, the space of potential lexical analogies that can be generated with WordNet appears both sizable and rich.

Why do higher distances allow for more analogies? It may be the case that the number of available dimensions grows in tandem with distance to the MSCA. However, Figure 1 reveals, somewhat surprisingly, that the opposite is the case. The total number of lexicalized dimensions shared by all pairs of concepts at the same depth actually shrinks as the distance between concepts increases, almost certainly because most dimensions are specific to particular areas of the taxonomy. Clearly then, to generate more analogies, these fewer dimensions must be made to do more work by expressing a greater range of potential values.

Overall then, some analogies juxtapose concepts that are already quite similar, while others relate very dissimilar ones. To determine whether lexicalized and ad-hoc dimensions qualitatively differ in this respect, we also considered the source:target similarity distribution of analogies generated from different dimension types. In particular, we

wanted to consider whether analogies based on lexicalized dimensions are different from analogies based on ad-hoc dimensions derived from compound terms (such as Worship-Modifier-Type) or indeed, from ad-hoc dimensions derived from glosses (such as Deliver-Object-Type).

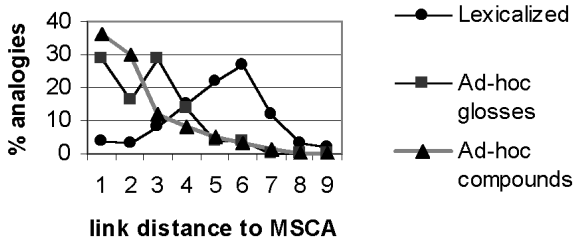

Figure 2: Distribution of analogies by Source:Target similarity.

Figure 2 suggests that lexicalized dimensions produce a body of analogies whose inter-concept similarity is normally distributed, ranging from the literal to the creatively distant, while ad-hoc dimensions tend to produce analogies (like *Lactose:Fructose*) that have a higher degree of literal similarity. This is especially true of ad-hoc dimensions extracted from compounds more so than glosses, since these dimensions are often based on a shared semantic head. Nonetheless, these more literal-minded analogies are also more *lateral*-minded, requiring a listener/student to dynamically construct an apropos ad-hoc category on the fly.

7 Palimpsest: A Relational Overlay

Consider the lexical concepts Work-Crew and Jury: both are social groups (and represented in WordNet as such), but their analogical similarity extends further than this. In English, the leader of a work crew is called a "foreman", as is the leader of a jury, which suggests that the concept Jury is commonly conceived of as a group of workers (a "work detail") that is assembled for a specific purpose and then disbanded. To capture this analogical similarity, a system would need to induce a new type of the form *Groups-whose-leaders-are-called-Foreman*, which seems to take us perilously close to the baroque typing of Way's (1991) dynamic type hierarchy. Yet this is the precisely the kind of analogically-useful adhoc category we strive for when using WordNet. However, such a grouping requires a level of textual analysis and semantic insight that is beyond the simple techniques outlined here. Ultimately, we need to consider annotating WordNet with such relational information by hand, or at the very least, semi-automatically.

In many cases the pivotal dimensional value for an analogy will a multi-place relation that is lexicalized as a verb. For instance, the WordNet lexical concepts Deliveryman, Roundsman, Deliverer and Bailor (none of which are synonymous) are similar by virtue of sharing the relational property *delivers:merchandise*. Such relational types can in turn be generalized further, to support an analogy between Paperboy and Milkman (each deliver something, suggesting the abstracted type *delivers:?*), between Cobbler and Plumber (each repair something, suggesting the abstracted type *repairs:?*), and so on. These relations might be extracted from WordNet by parsing the corresponding glosses, but WordNet glosses are neither regular enough in form, nor consistent enough in content, to make this approach a practical reality. That is, while we section 5.2 shows that we can parse these glosses to recognize dimensions that are analogically useful, we cannot reliably associate these dimensions with all of the word-concepts that should be tagged with them.

These issues lead us to a conclusion that is at once both dispiriting and liberating: semantic tagging with these dimensions cannot reliably be performed via automatic means, but must be done manually, or at most, semi-automatically. Certainly, such an effort is time-consuming, but ultimately, we believe it is less so than a fully automated approach, which would require constant tweaking to overcome both under-generation and over-generalization. To ensure a comprehensive coverage, and to allay fears that this effort is a representational *toy* designed to work only for specially selected examples, we conceive of the effort as a new lexical ontology in its own right, but one that is to be overlaid on the existing taxonomic structure of WordNet.

The new ontology, named *Palimpsest*, will attempt to offer an explicit property-theoretic description of each lexical concept in WordNet, in a form that will allow for the dynamic generation of new categories as they are needed.

Each Palimpsest property is associated with a specific WordNet synset, and relates that synset/concept to another WordNet lexical entry. Properties comprise both a predicate/relation and an object, and may be marked in a variety of ways to indicate salience. The general form of a property is given as follows:

[*]relation[/abstraction]:object[*]

Each property can be seen as specifying a category of concepts that all share that property. Thus, Jury and Crew both share the property *leader/manager:Foreman*, which becomes the de facto name for the ontological category that embraces both concepts. Each relation may be specified relative to a common abstraction that is shared by multiple relations. Thus, the relations *cleans/restores* and *repairs/restores* can be seen as specializations of the abstract relation *restores*, while the relations *sells/provides* and *pays/provides* are specializations of the abstract relation *provides*. This two-level specification of relations allows a system to perform functional abstraction on a concept, to e.g., recognize the similarity between cleaners (who restore by cleaning), repairmen (who restore by repairing) and surgeons (who restore by curing).

The optional * at either end of the property marks either the relation (if before) or the object (if after) as especially salient and foregrounded within the definition of the concept. For instance, in the definition of Merchant, the relation *sells/provides*, and the object, merchandise, are both foregrounded, as follows:

*{merchant, merchandiser}: *sells/provides:merchandise**

When an object like *Merchandise* is foregrounded in this way, it means it is central to the definition of the host concept. When a relation is foregrounded, it means that the host concept (e.g., *Merchant*) can be used to suggest relation in an analogical or metaphoric manner. For instance, another concept in which the relation *sells/provides* is foregrounded is *Peddler*; we can thus reconceptualize the concept Merchant as a "merchandise peddler", a Stockbroker as a "stock merchant" or "stock peddler", a prostitute as a "sex merchant" or "sex peddler", and so on.

Adhoc categories arise in Palimpsest when properties are generalized or composed. For instance, the property *sells/provides: merchandise* can be generalized either as *provides:merchandise*, *sells/provides:?* or *sells/provides:?*. Whenever a property is generalized, its corresponding category is widened to admit new members, permitting those new members to be seen as similar in ways that were not previously perceived. In contrast, composition is a category-narrowing operation, allowing multiple properties to be conjoined into one, as when two properties in the same host concept share the same object, allowing the corresponding relations to be merged. For instance, a *Dealer* buys and sells merchandise, while a *Stockbroker* buys and sells shares. In each case, a new composite relation can be created and then generalized, to yield *buy_and_sell:?*. Properties can be composed in a variety of other ways (e.g., hierarchically), allowing new categories with a more exclusive membership to be constructed on the fly.

8 Conclusions

As a lexical ontology, WordNet is both lightweight and flawed. Nonetheless, when viewed from the right dimensional perspective, it possesses enough tacit structure to yield a rich vein of lexical analogies. We have identified two different kinds of conceptual dimension – *lexicalized* and *adhoc* – through which analogically useful structure can be imposed on WordNet's lexical content. By mining this vein, we expect to use WordNet as a generator of educational analogies for scholastic tests, as well as a creator of challenging puzzles for computer games. The space of potential analogies that WordNet can generate (and subsequently reason about and manipulate) is encouragingly large, allowing both on-line tests and computer games to dynamically provide challenging new content each time they are used. This generativity will be improved greatly by the addition of a new relational overlay for WordNet, called Palimpsest, that will make explicit the complex dimensions on which analogies can pivot.

Palimpsest is still at an experimental and immature stage of development. Only with maturity will it settle on an appropriate set of relations and relational abstractions, and maturity will only come via scale. However, development is progressing apace, and Palimpsest, now nearing the 10,000 concept mark, will soon be ready for general release to the research community (where it will be free for research purposes).

References

Barsalou, L. (1983). Ad-Hoc Categories. *Memory and Cognition*, 11(3), 211–227.

Fass, D. (1988). 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 AI*. Morgan Kauffman: CA.

Landauer, T. K. and 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.

Mihalcea, R. and Moldovan, D. I. (2001). eXtended Wordnet: progress report. *Proc. of NAACL 2001—Workshop on WordNet and Other Lexical Resources*, Pittsburgh, PA.

Miller, G. (1995). WordNet: A Lexical Database for English. *Comms. of the ACM*, 38(11).

Miller, W. S. (1960). *Manual of the Miller Analogies Test*. New York: The Psychological Corporation.

Falkenhainer, B., Forbus, K. and Gentner, D. (1989). Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41.

Rumelhart, D. E. and Abrahamson, A. A. (1973). A model for analogical reasoning. *Cognitive Psychology* 5, pp. 1–28.

Turney, P. D., Littman, M. L., Bigham, J. and Shnayder, V. (2003). Combining independent modules to solve multiple-choice synonym and analogy problems. *Proc. of International Conf. on Recent Advances in Natural Language Processing.*

Veale, T. and Keane, M. (1997). The Competence of Sub-Optimal Structure Mapping on Hard Analogies. *Proc of IJCAI '97, the Int. Joint Conf. on Artificial Intelligence.* Morgan Kauffman: CA.

Veale, T. (2004). WordNet sits the S.A.T.: A Knowledge-Based Approach to Lexical Analogy. *The proceedings of ECAI 2004, the 16th European Conf. on Artificial Intelligence*. John Wiley: London.

Way, E. C. (1991). Knowledge Representation and Metaphor. *Studies in Cognitive systems*. Kluwer Academic Publishers: Holland.