Title
Formalizing Diagnosticity in Concept Combinations

Permalink
https://escholarship.org/uc/item/4wk1c6sq

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Authors
Ferrari, Daniela
Keane, Mark

Publication Date
2005

Peer reviewed

Formalizing Diagnosticity in Concept Combinations

Daniela Ferrari (daniela.ferrari@ucd.ie)
Department of Philosophy and Social Sciences,
University of Siena, Siena, Italy

Mark Keane (mark.keane@ucd.ie)
Department of Computer Science,
University College Dublin,
Dublin 4, Ireland

Abstract

A new way to formalize diagnosticity, based on information retrieval ideas, is advanced and then used to examine the effects of diagnosticity on the understanding of novel, noun-noun compounds. Three experiments are reported in which the pattern of diagnosticities on the combination's constituent concepts is used to predict how the compound will be interpreted. Predictions about how distributions of diagnosticity values affect the number of interpretations produced in property and relational compounds are also tested. The implications of this new methodology and these results for theories of concept combination are discussed.

Introduction

Conceptual combinations are to language generativity, what fruit flies are to genetics. By examining how people understand noun-noun combinations (e.g. *night train*, *finger cup*, and so on) it has been possible to begin to understand the combinatorial richness of human language, a compositional fluency that allows people to create new meanings from previously unseen combinations of known words.

One of the key findings in this research area is that there are two main classes of interpretations produced from noun-noun compounds (see e.g., Costello & Keane, 1997; Hampton, 1997; Wisniewski & Gentner, 1991): property interpretations (e.g. when *finger cup* is interpreted as "a thin cup" by using the property predicate *thin* from the concept *finger*) and relational interpretations (e.g., when *night train* is interpreted using the relational predicate *operates-at* to connect the two concepts). Indeed, it has been shown that concept combinations tend to be interpreted predominantly in either a property or a relational way, allowing them to be termed *property compounds* or *relational compounds* (see Gagné, 2000; Tagalakis & Keane, 2004; Wisniewski, 1996; Wisniewski & Love, 1998).

There are several different accounts of how such combinations are understood, that have raised issues about whether relational compounds are more "natural" than property compounds (cf. Gagné, 2000; Gagné & Shoben, 1997). Stressing the importance of distributional knowledge about the relations associated with the compounds constituents, Gagné (2000) has proposed that relational compounds are understood by selecting a relation highly-frequently associated with the modifier (the first word in the compound) to provide an interpretation. Wisniewski and his colleagues have argued that property compounds are

understood by aligning the constituents and mapping the property from one concept to the other. Costello & Keane (2000) propose a constraint-satisfaction theory which sees property and relational compounds as being formed and understood by a unitary mechanism that shapes the meaning using the constraints of diagnosticity, plausibility and informativeness.

Costello & Keane have argued that diagnosticity plays a key role in selecting one predicate over another when interpreting all types of compounds. A diagnostic predicate is one that differentiates a concept from all other concepts (for related ideas see Hampton's, 1987, *importance*; Tversky's, 1977 *diagnosticity*; Rosch's, 1974, *cue validity*). So the interpretation of a cactus fish as "a prickly fish" appears to be a more acceptable interpretation than "a green fish" because *prickly* is more diagnostic of *cactus* than *green*. Costello & Keane (1997) confirmed this hypothesis empirically by showing that people rated interpretations using diagnostic properties as being more acceptable than those using non-diagnostic properties. Furthermore, several computational models have shown that diagnosticity operates as an important selection criterion in modelling concept combination (see Costello & Keane, 2000; Lynott, Tagalakis, & Keane, 2004).

However, the use of the diagnosticity idea is not without its problems. First, all of the empirical evidence on diagnosticity has only been shown for property compounds not relational ones. Second, and more seriously, it is not at all clear how one could operationally define diagnosticity for relational predicates. People can give ratings for the diagnosticity of the properties *prickly* and *green* for *cactus* but it does not seem sensible to ask people to rate the diagnosticity of *operates-at* for *train*. Relational predicates seem to be more external to concepts than properties, in a way that makes them inappropriate for use in direct rating tasks.

In this paper, we tackle both of these problems in reverse order. First, we develop a new methodology for characterizing the diagnosticity of all predicates, whether they be property or relational ones, using information retrieval ideas, in particular the tf-idf schema (van Rijsbergen, 1979; see Experiment 1). Second, we report two experiments using these diagnosticity measures to test for their effects in property compounds (Experiment 2) and relational compounds (Experiment 3). Our results indicate that the role

of diagnosticity may be very different in the two compound types.

Formalizing Diagnosticity

Costello & Keane (2000) developed their diagnosticity scores by asking people to directly rate participant-derived properties. This approach will not work for relational predicates and may well be sub-optimal for property predicates. In their computational model, Costello & Keane (*ivi*) claim that a feature is diagnostic of a concept if *that feature occurs frequently in instances of that concept and rarely in instances of the other concept.* Asking people to rate such predicate features for diagnosticity is a poor proxy for this notion. Sartori and Lombardi (2000) have proposed a better method for determining, what they call, *relevance* by adapting ideas from information retrieval.

As for information retrieval, queries are based on terms. If a term occurs in a query, then its presence in a document means that the document is relevant to the query. However, different terms can be differentially important to a document. *A term is important if it has a high frequency in a document and a low frequency in the rest of the documents of the collection.* This idea is clearly analogous to Costello & Keane's definition of diagnosticity. So, a similar procedure could be used to formalize diagnosticity if we replace *terms* with *feature predicates* and *documents* with *concepts*.

In information retrieval, term frequency ($tf_{d,t}$) refers to the frequency of a term in a document and the frequency across a collection of documents is referred to as inverse document frequency (idf_t). Given that N is the number of documents in the collection and n_t the number of documents containing term t, the inverse document frequency is expressed by:

$$idf_t = \log_2 N/n_t \qquad (i)$$

The relevance of term t for document d is a composition of $tf_{d,t}$ with idf_t:

$$R_{d,t} = tf_{d,t} * idf_t \qquad (ii)$$

which, in virtue of (i), can be in turn stated as:

$$R_{d,t} = tf_{d,t} \qquad * \log_2 N/n_t \qquad (iii)$$

In reusing these ideas to compute diagnosticity we clearly need a database of predicate features and concepts. Sartori & Lombardi (2000) suggest that this database should be obtained from people's listings of features for a corpus of words.

After people have described the word concepts using a variety of descriptors, each concept description can be divided into one or more property predicates (e.g., *is red*, *is fluffy*, *is fast*) and relational predicates (e.g., *comes-from-milk*, *for-getting-attention*, *eats-wood*). These predicates and concepts can then be arranged in a Concepts Matrix. The frequency of predicate feature f in concept c ($f_{c,f}$) is computed for every feature predicate that occurs in a concept, as illustrated in Table 1. At this point transformation of the frequency values by application of formula (ii) will lead to a new matrix that contains diagnosticity values for each feature

predicate. This matrix returns varying diagnosticity values for predicates across the whole set of concepts and can give the same predicate different diagnosticity values in different concepts.

Table 1: The Concepts Matrix

	C_1	C_2	$\ldots$	C_m
F_1	$f_{1,1}$	$f_{2,1}$	$\ldots$	$f_{m,1}$
F_2	$f_{1,2}$	$f_{2,2}$	$\ldots$	$f_{m,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
F_n	$f_{1,n}$	$f_{1,n}$	$\ldots$	$f_{m,n}$

Experiment 1: Predicting Compound Types

To evaluate this methodology for computing diagnosticity we performed a general test to see if diagnosticity predicted types of compound (i.e., property or relational). If diagnosticity plays a role in comprehending compounds, then we should find that for a given compound if its property predicates have a higher diagnosticity score, then it should tend to yield property interpretations (i.e., be a property compound). Similarly, if a given compound has relational predicates with higher diagnosticity scores, then it should tend to yield relational interpretations (i.e., be a relational compound).

Method

Material Sixty novel, noun-noun compounds were drawn from the larger corpus of Tagalakis and Keane (2003); none of these compounds were lexicalized and none used nouns that could be understood as adjectives. Half of these 60 were property compounds and half relational compounds, as established by previous studies. Using the British National Corpus (BNC: Burnard, 1995), half of these 30 property compounds were classified as infrequent (no hit in the corpus) and half as frequent (one or more hits in the corpus). The relational compounds were similarly split into frequent and infrequent. Independent ratings gathered by Tagalakis & Keane (2003) show that almost all (90%) of the infrequent items were consistently rated as "unfamiliar" by people, while all the frequent items were rated as "familiar". All compounds were broken down into their constituents, giving 120 individual concepts to which another 120 were added as fillers randomly chosen from categories of common living things and artifacts.

Participants and Procedure Eight native English-speaking students from University College Dublin were paid to take part in the experiment. They received 5 booklets, one at the time, with instructions to describe "how the thing looks", "what it is used for", and "any other information that comes to mind" (following Sartori & Lombardi, 2000). This procedure was designed to encourage participants to describe

both the perceptual and functional features of the concepts, and additional associative or encyclopedic relevant information. The task took between three and four hours, during which time participants were given rest breaks.

The descriptions given were broken down into distinct predicate representations. Where possible, complex meanings were broken down into simpler ones that had already been used by other subjects (e.g. *sweetening* into *sweet* and *flavouring*, where *flavouring* could be used for the concept *salt* as well). Synonyms and other closely related concepts were grouped (e.g. *ornament*, *aesthetics*, *decoration*). The distinction between active and passive forms was kept (e.g., *burns fuel* and *burned as fuel* were considered two different meanings). Only positive attributes were considered (e.g., *hairless*, *don't fly*, *don't have teeth* were excluded). Disjunctions were treated as inclusive (e.g. if the concept *apple* was described as 'green or red' each colour was counted once).

Analysis Using the Concept Matrix Using the above procedure a total of 722 distinct predicate features were identified, each classified as a property predicate or relational predicate. Property predicates typically involved perceptual or physical features of objects (e.g. *is hollow*, *has big ears*, *smells*). Relational predicates were ones that put the concept in connection with other objects (e.g. *eats bananas*, *is in desert*) or were related to functions (e.g., *used for sitting on*). A 240 (concepts) x 722 (predicates) Concept Matrix was thus built in which the diagnosticity of each predicate was computed using (ii) as described above.

Table 2: Diagnosticity Values for *Devil woman*

Modifier		Head		Combination		
TPD	TRD	TPD	TRD	TPD	TRD	TPD -TRD
161.15	23.72	82.65	67.99	243.81	91.71	152.09

By summing the values of all the property predicates and relational predicates mentioned in descriptions of that concept, each concept was described by two values: the total property diagnosticity (TPD) and total relational diagnosticity (TRD). Therefore for any given combination, both concepts, the first concept (or modifier) and the second concept (the head), have TPD and TRD scores. Table 2 shows these values for the modifier and head of *devil woman*, along with the combined TRD and TPD scores for the combination as a whole. It also shows the difference between the TPD and TRD (i.e., TPD-TRD). This Diagnosticity Difference score is important because if it is positive it means that the diagnosticity of the property predicates dominates that of the relational predicates and vice versa if the Diagnosticity Difference is negative.

Results & Discussion

The key prediction of interest here is whether Diagnosticity Difference scores predict the class of compounds. Specifically, whether compounds with a positive Diagnosticity Difference value tend to be property compounds and compounds with a negative Diagnosticity Difference value tend to be relational compounds. This key prediction was confirmed by results.

Overall, the diagnosticity values of predicates in the Concept Matrix ranged from 0 (when the feature did not appear in the description of a given concept) to 63.25 (*has-eight-legs* for concept *octopus*). The mean diagnosticity of all predicates was 12.22. After computing the TPD and TRD scores for all concepts we combined these scores to determine the Diagnosticity Difference value for each compound.

A t-test analysis of the Diagnosticity Difference values showed that property compounds (M = 107.79) were reliably different to relational compounds (M = 17.52; t(58) = 5.49, p<.001). A chi-square test also showed a reliable relationship between the sign of Diagnosticity Difference value (negative or positive) and the class of the compounds (property or relational), $\chi^2(1) = 16.7$, $p < .001$.

In confirming such an intuitive hypothesis, that the most diagnostic predicates will be reflected in the class of the compound, we can have some confidence that this measure of diagnosticity holds some validity. In the next two experiments, we go further in examining the distribution of predicate diagnosticities in property (Experiment 2) and relational compounds (Experiment 3) with a view to developing more precise predictions.

Experiment 2: Property Compounds

We have seen that overall property compounds have positive Diagnosticity Difference values (whereas relational compounds have negative values); that is, the TPD for the two concepts in the compounds is higher than the TRD. However, this is quite a rough characterisation of the diagnosticities of a compound. Each concept in the compound could have several property predicates, each of which could have a different diagnosticity score. The TPD merely sums up all of these diagnosticity values for each concept but it says nothing about the distribution of those scores. Yet, it is clear that these property predicates could be distributed in different ways. A given concept could have a *peaked distribution* where there is a single, highly-diagnostic predicate (e.g., *prickly* for *cactus*) among many low-diagnostic predicates or it could have a *flat distribution*, where all the predicates have roughly equal diagnosticity values. In this experiment, we examine predictions about compounds with *peaked* and *flat* distributions.

Our first prediction is that if the constituent concepts of a compound have a peaked distribution there should be a tendency to produce fewer different interpretations from it across a group of participants, as interpretations will focus on using the single, highly-diagnostic predicate. Conversely, when the distribution is flat, there should be a tendency to produce many different interpretations based on the different equally-diagnostic predicates.

There is one further twist to the distribution prediction. In Experiment 1, we did not distinguish between the head and modifier of the compound, but clearly we must consider that both head and modifier will have different predicate distributions that may be peaked or flat. In the interpretation of property compounds a property predicate of the modifier is asserted of the head (e.g., a *bullet train* is a *fast* train). So, we

would expect that the greatest effect of the peaked/flat distribution should be found for modifiers rather than heads.

The second prediction we make, following on from Experiment 1, is that compounds with a higher TPD than TRD (i.e., a positive Diagnosticity Difference) should be interpreted as property compounds. That is, they should predominantly be interpreted using property rather than relational interpretations.

Method

Materials Sixty novel, noun-noun compounds were created from the concepts analysed in Experiment 1. All of these compounds had positive Diagnosticity Difference values. The set of compounds had 15 compounds in which the modifier had a peaked distribution (PEAK MODIFIER), 15 compounds with a modifier with a flat distribution (FLAT MODIFIER), 15 compounds with a head with a peaked distribution (PEAK HEAD), 15 compounds with a head with a flat distribution (FLAT HEAD). In each of these conditions the corresponding head/modifier paired in the compound was counterbalanced (roughly half had peaked distributions, while the other half had flat distributions). A peaked distribution for a concept was defined as one in which there was one predicate with a diagnosticity score one standard deviation above the mean score (M = 12.22) of the whole feature set; this means it had to have a value ≥ 23.10. Thirty filler relational compounds were used to obviate the development of specialised interpretation strategies.

Subjects, Procedure & Scoring Thirty native English-speaking undergraduates from the University College Dublin participated in the experiment for partial course credits. Combinations were presented on a PC, one combination per screen shown, in random order. Subjects were given instructions to type in their first interpretation next to the combination. Subjects completed the task in approximately twenty minutes. Interpretations were classified into five categories: Property, Relation, Mixed, Like and Other. Interpretations in which a property feature was involved were classified as Property. Those in which a relation was established were classified as Relation. Interpretations that used both a property and a relation were classed as Mixed. Interpretations using 'like' without specification of a dimension were considered a class on their own, for there is disagreement in the literature as to their status. Unclear interpretations and interpretations that just renamed the combinations with synonyms constituted the class Other. Ratings by the first author were validated by an independent judge who was unaware of the hypotheses of the experiment (with 90% agreement).

Results and Discussion

Overall, the two predictions made were confirmed. First, as predicted, there was a main effect of condition on the number of different interpretations produced as revealed by a one-way ANOVA (F(58) = 4.714, p<.02). The mean number of property interpretations produced in the FLAT MODIFIER condition (M = 6.6) was reliably greater than that found in the

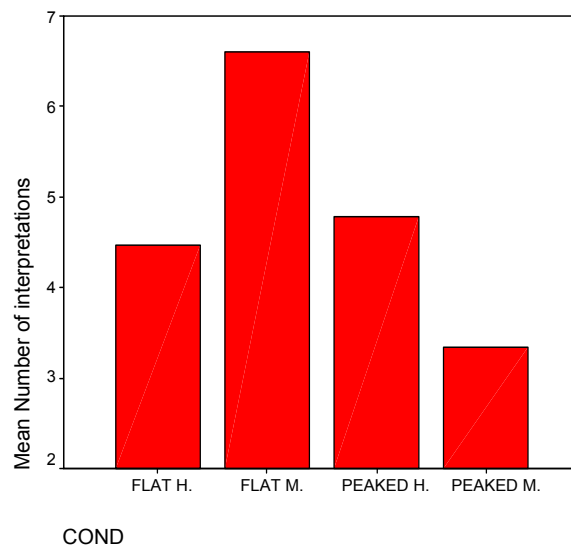PEAK MODIFIER one (M = 3.33; t(28) = 3.408, p <.05; see Figure 1).

Figure 1: Experiment 2. Mean property interpretations generated across conditions.

In contrast, no such difference was found between the FLAT HEAD (M = 4.47) and PEAK HEAD (M = 4.79) conditions, $t(27)^1$ = -394, p>.05. This result essentially shows that when a modifier concept has a set of predicates with a flat distribution then many interpretations are produced, whereas fewer are produced from modifiers with peaked distributions. Head concepts do not have the same impact at all on frequency of production of different interpretations.

Second, as expected we found that most of the compounds were interpreted as property compounds (see Figure 2). Participants produced many more property interpretations (M = 17.40) than relational interpretations (M = 6.50; t(59) = 6.375, p <.001) Furthermore, a one-way ANOVA on the differences between the mean frequencies of property interpretations versus relational interpretations, reveals a reliable main effect of condition, F (59) = 3.875, p <.02. Pair-wise comparisons show that this effect is most pronounced between the PEAK MODIFIER (M = 19.47) and FLAT MODIFIER (M = 6.60) conditions, t(28)= -3.84, p<.001. There is no reliable difference between the PEAK HEAD (M = 11.73) and FLAT HEAD conditions (M = 5.80; t = -1.091 p>0.5).

Taken together these results indicate that diagnosticity plays a key role in the generation of interpretations to conceptual combinations. Further confirming the result of Experiment 1, we have found that compounds with a positive Diagnosticity Difference are interpreted as property compounds. Moreover, the locus of this effect mainly lies in modifier concept rather than the head one. If the compound's modifier has many property predicates with a flat distribution of diagnosticity values then many property interpretations are

---

[1] Degrees of freedom are different because one item was responded to completely in a relational way.

produced. If the compound's modifier has a peaked distribution then fewer interpretations are produced.
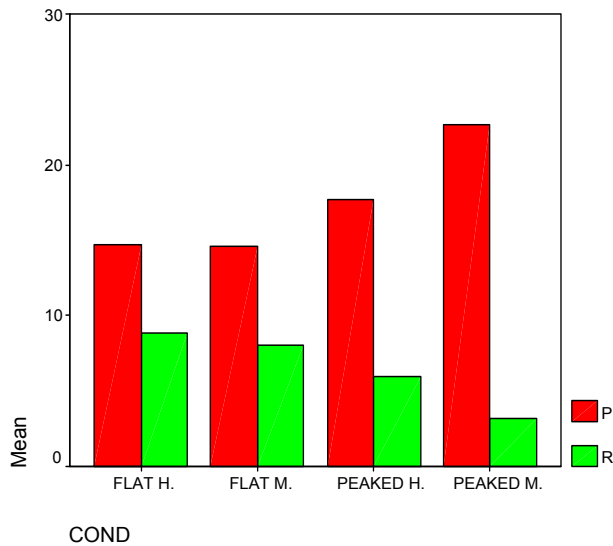

Figure 2: Experiment 2. Mean frequencies of property and relation interpretations across conditions.

Experiment 3: Relational Compounds

In this experiment, we essentially repeated Experiment 2 but this time used a set of compounds with negative Diagnosticity Difference values (i.e., compounds that should turn out to be relational compounds); that is, compounds in which the balance of diagnosticity scores were higher for the relational predicates than the property predicates (i.e., TRD>TPD).

If the way diagnosticity operates in these compounds is the same as in property compounds then we should see the same pattern of results for these compounds: (i) compounds with a flat distribution on their modifier should produce more interpretations than those with a peaked distribution on the modifier; (ii) more relational interpretations should be produced from these negative Diagnosticity Difference compounds as the balance of their diagnosticities is in favour of relations.

Method

Material Sixty combinations were made using the same method as in Experiment 2, according to the constraint that TRD scores were higher than TPD scores. Here, the distribution of relational predicates was modified instead of the property predicates for heads and modifiers. The resulting four sets of materials were: 15 compounds in with a modifier with a peaked distribution (PEAK MODIFIER), 15 with a modifier with a flat distribution (FLAT MODIFIER), 15 with a head with a peaked distribution (PEAK HEAD), 15 with a head with a flat distribution (FLAT HEAD). In each set the other constituent was counter-balanced in its distribution.

Subjects and Procedure Thirty native English-speaking undergraduate and postgraduate students from the University College Dublin voluntarily participated in the experiment. The procedure and scoring were as in
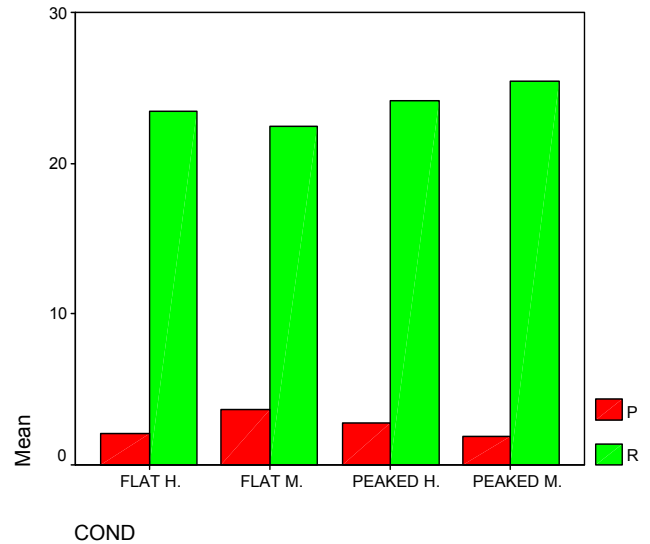

Figure 3: Experiment 3. Mean frequencies of property and relation interpretations across conditions.

Experiment 2. Ratings were carried out by the first author and an independent judge who was unaware of the hypotheses of the experiment (agreement between raters was 92.3%).

Results and Discussion

Overall, the first prediction was not supported but the second was confirmed. Taking the bad news first, all of the conditions produced relatively uniform numbers of interpretations. T-tests revealed no reliable differences between the four conditions, which had the following means: PEAK MODIFIER (M = 3.40), FLAT MODIFIER (M = 3.40), PEAK HEAD (M = 3.07), and FLAT HEAD (M = 3.80). These results show in a fairly unambiguous fashion that diagnostic, relational predicates do not operate in anything like the same way as diagnostic, property predicates. The reasons for these effects are not immediately apparent, though one possibility is that relations cannot be used as flexibly as properties to construct interpretations. Within constraint theory, this would be seen as a result of another constraint, plausibility, overwhelming the influence of diagnosticity. The plausibility constraint prevents even highly diagnostic relations from being used in full-fledged interpretations. Costello & Keane (2000) envisaged such a possibility in the interpretation of compounds but they did not realise that it could occur in such a widespread fashion in relational compounds.

The good news is that the second hypothesis was confirmed (see Figure 3). Overall, participants consistently produced more relational (M = 23.86) than property interpretations (M = 2.60; t(59) = 20.803, p= <.001), as predicted. This effect was marked in all conditions: in condition MODIFIER PEAK relational interpretation (M = 25.40) > property interpretation (M = 1.87, t(14) = 15.668, p<.001); MODIFIER FLAT relational interpretation (M = 22.47) > property interpretation (M = 3.67, t(14) = 6.785, p<.001); HEAD PEAK relational interpretation (M = 24.13) > property interpretation (M =

2.73, t(14) = 11.107, p<.001); HEAD FLAT relational interpretation (M = 23.47) > property interpretation (M = 2.13, t(14) = 11.985, p<.001).

General Discussion

In this paper, we have advanced a new method for formalizing diagnosticity for property and relational predicates. We have then shown that this metric is accurately reflected in the known property and relational compounds (Experiment 1). In Experiment 2, we have shown that when one constructs compounds with higher diagnosticities on their property predicates, they turn out to be interpreted, as predicted, as property compounds. In Experiment 3, we saw that that when one constructs compounds with higher diagnosticities on their relational predicates, they turn out to be interpreted, as predicted, as relational compounds. Furthermore, we have shown that for property compounds, when the modifier has a flat distribution, many more interpretations are produced than in any other condition (Experiment 2). However, this effect is unique to property compounds, it does not occur in relational compounds (Experiment 3).

These results open up a whole new vista of possible empirical tests of theories of conceptual combination. Armed with this new instrument for characterizing diagnosticity it should be possible to poke the cognitive fruit fly much more accurately than before.

Acknowledgments

References

Burnard, L. (1995). *Users guide for the British National Corpus*. British National Corpus Consortium, Oxford University Computing Service.

Costello, F. & Keane, M., T. (1977). Polysemy in conceptual combination: Testing the constraint theory of combination. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 147-142). Hillsdale, NJ: Erlbaum.

Costello, F. & Keane, M., T. (2000). Efficient creativity: constraint-guided conceptual combination. *Cognitive Science* 24(2), 299-349.

Costello, F. & Keane, M., T. (2001). Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology:LMC, 27, 255-271.*

Cree, G. S. & McRee, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, cheese, and cello (and many other such concrete nouns). *Journal of experimental psychology: general*, 132 (2) 163,201.

Gagné, C. L. (2000). Relation-based combinations versus property-based combinations: A test of the CARIN theory and dual-process theory of conceptual combination. *Journal of Memory and Language*, 42, 365-389.

Gagné, C. L. & Shoben, E. J., (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, memory, and Cognition*, 23, 71-78.

Hampton, J., A., (1987) Inheritance of attributes in natural concept conjunctions. *Memory and cognition* 15 (1), 55-71.

Hampton, J. (1997). Conceptual combination. In Lamberts K., Shanks D. (Eds.) *Knowledge, concepts and categories*. Cambridge, MA: The MIT Press.

Lynott, D., Tagalakis, G., & Keane, M.T. (2004). Conceptual Combination with PUNC. *Artificial Intelligence Review*, 21(3-4), 353-374.

Rosch, E., (1974). Linguistic relativity. In A. Silverstein (Ed.), *Human communication: theoretical explorations*. Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C.B. (1975). Family resemblance: studies in the internal structure o categories. *Cognitive Psychology*, 3, 382-407.

Sartori, G. & Lombardi, L., (2004). Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience*, 16, 439-452.

Tagalakis, G. & Keane, M.T. (2003). Modelling the understanding of noun-noun compounds: The role of familiarity. *Proceedings the European Cognitive Science Conference (EuroCogSci03)* (pp 319-324). Mahwah, NJ: Erlbaum.

Tversky, A. (1977). Features of similarity. *Psychological review*, 84 (4), 327-352.

van Rijsbergen, C. (1979). *Information Retrieval*. London: Butterworth.

Wisniewski, E., J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language*, 35, 434-453.

Wisniewski, E. J. & Gentner, D. (1991). On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning. In. Simpson G.B (Ed.) *Understanding word and sentence*. Amsterdam: North Holland.

Wisniewski, E. J. & Love, B.C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language*, 38, 177-202.