METAPHOR AND SYMBOL, *22*(4), 291–312

Interpretive Diversity Explains Metaphor–Simile Distinction

Akira Utsumi
Department of Systems Engineering,
University of Electro-Communications, Chofugaoka, Chofushi, Tokyo, Japan

A number of properties—aptness, topic–vehicle similarity, vehicle conventionality—have recently been used to explain a metaphorsi-mile distinction. This paper argues that interpretive diversity better explains a metaphor–simile distinction than these properties. Interpretive diversity refers to the semantic richness of the figurative interpretation of a topic–vehicle pair and is determined depending on both the number of features involved in the interpretation and the uniformity of salience distribution of those features. The interpretive diversity view predicts that interpretively more diverse pairs should be easier to comprehend via a categorization process, and thus the preference for and the relative comprehensibility of the metaphor form should be greater. Two experiments demonstrated that, as predicted, interpretive diversity was correlated positively with metaphor preference (Experiment 1) and with the relative comprehensibility of the metaphor form compared to the simile form (Experiment 2). Furthermore, interpretive diversity was found to be more important in explaining metaphor–simile distinction than aptness, similarity, and conventionality.

Although there has been a widely held consensus since Aristotle that a metaphor "An X is a Y" and a simile "An X is like a Y" express almost the same figurative meaning, some recent studies have revealed many intriguing differences between these two kinds of tropes. One critical difference is people's preference for one form over another (e.g., Chiappe & Kennedy, 1999, 2001). For example, if people are asked to make a choice between "*Life is a journey*" and "*Life is like a journey*," most of them choose the metaphor form as more appropriate (Chiappe & Kennedy, 1999). If asked to choose between "*Highways are snakes*" versus "*Highways are like snakes*," people show a strong preference for the simile form. In addition to

Correspondence should be addressed to Akira Utsumi, Department of Systems Engineering The University of Electro-Communications 1-5-1 Chofugaoka, Chofushi, Tokyo 182-8585, Japan. E-mail: utsumi@se.uec.ac.jp

such difference in form preference, metaphor and simile differ in many other ways; metaphors are comprehended faster than similes (Glucksberg, 2003; Johnson, 1996), metaphors are interpreted as being more metaphorical and having more force (Glucksberg, 2003; Zharikov & Gentner, 2002), metaphors are more likely to be used to make the discussion more interesting and to emphasize a point in conversation (Roberts & Kreuz, 1994), and metaphors are less likely to be used to compare similarities (Harris, Friel, & Mickelson, 2006). Given these observed differences between metaphor and simile, one question emerges: What properties of a topic–vehicle pair determine the preferred from of expression? This is the question addressed in this paper.

Recently a number of experimental studies have tackled this question. Chiappe and his colleagues (Chiappe & Kennedy, 1999, 2000, 2001; Chiappe, Kennedy, & Chiappe, 2003; Chiappe, Kennedy, & Smykowski, 2003) have proposed that *aptness* or *topic–vehicle similarity* affects people's preference for the metaphor form. They define aptness as the extent to which a topic–vehicle pairing captures salient features of the topic in question and then argue that the preference for the metaphor form increases as the aptness or the similarity of a topic–vehicle pair increases. They indeed demonstrated that the preference for the metaphor or simile form was predicted by aptness or topic–vehicle similarity. Another series of experimental studies (Bowdle & Gentner, 1999, 2005; Gentner & Bowdle, 2001; Zharikov & Gentner, 2002) has addressed the *conventionality of the figurative meaning of the vehicle* as a source of metaphor–simile distinction. These studies demonstrated that the simile form was preferred and more comprehensible when the vehicle of the comparison was novel, but the metaphor form became preferred and more comprehensible when the vehicle was conventionalized by repeated figurative use.

This paper proposes that the *interpretive diversity* of a topic–vehicle pair, which refers to the semantic richness of the figurative meaning of the pair, is a key property that explains metaphor–simile distinction. We argue that as the interpretive diversity of a topic–vehicle pair increases, the preference for the metaphor form increases as well. The semantic richness of the figurative meaning or interpretive diversity depends on two factors: the number of features that constitute the figurative meaning and the uniformity (or evenness) of salience distribution of those features (Utsumi, 2005). A higher value of interpretive diversity means a larger number of features and a more uniform salience distribution. When a topic–vehicle pair *A* conveys a larger number of relevant features than a topic–vehicle pair *B*, the pair *A* can be seen as more rich or interpretively more diverse than the pair *B*. On the other hand, when two topic–vehicle pairs *A* and *B* convey the equal number of meanings, the pair with a more uniform salience distribution of the conveyed features is interpretively more diverse. For example, suppose that for pair *A* one meaning is much more salient than the other meanings, but for pair *B* all the meanings are equally salient. In this case, pair *A* can be seen as interpretively more diverse than pair *B*. Interpretive diversity can be quantitatively assessed as Shannon's entropy (Utsumi, 2005).

Given that the figurative interpretation of topic–vehicle pair A consists of n meanings or features $a_1, \ldots, a_n$ and with relative saliences $p(a_1), \ldots, p(a_n)$ where $\Sigma_{i=1}^{n} p(a_i) = 1$, the interpretive diversity H(A) of the interpretation of pair A is calculated by the following equation:

$$H(A) = -\sum_{i=1}^{n} p(a_i) \log_2 p(a_i) \quad (1)$$

Interpretive diversity takes the maximal value $\log_2 n$ if all relative saliences $p(a_i)$ are equal, while taking the minimum value 0 if one salience is 1 and all others are 0. It is worth pointing out here that the definition by Equation 1 has been most widely used to measure the biodiversity of an ecosystem (Pielou, 1975), and often referred to as Shannon's diversity index in the field of ecology. It suggests that Shannon's entropy is suitable for measuring the degree of diversity of something.

This paper reports two experiments conducted to test the interpretive diversity view. Experiment 1 examined metaphor preference, and tested the relation between metaphor preference and interpretive diversity of the features shared by the topic and the vehicle. Experiment 2 examined difference in comprehensibility between the metaphor and simile forms of a topic–vehicle pair and tested whether difference in comprehensibility between both forms was explained by the interpretive diversity of the figurative interpretation generated by the participants.

RELATION BETWEEN FORM PREFERENCE AND COMPREHENSION PROCESS

Metaphor-simile distinction in terms of preference and comprehensibility can be attributed to processing difference of metaphors, in other words, whether metaphors are comprehended via a categorization process or a comparison process. In general, metaphors of the form "An X is a Y" are encouraged to be comprehended as categorizations because they are grammatically identical to literal categorization statements (e.g., "An orange is a fruit"). On the other hand, similes of the form "An X is like a Y" are encouraged to be comprehended as comparisons because they are grammatically identical to literal comparison statements (e.g., "An orange is like a lemon").[1] Therefore, if people do comprehend metaphors via a categorization process and similes via a comparison process in accordance with this link between form and function, there should be no strong preference for metaphors or similes and no big difference in comprehensibility between metaphors and similes. However, if metaphors cannot be comprehended

[1]Bowdle and Gentner (2005) refer to this link between form and function in figurative language as grammatical concordance. Grammatical concordance is now widely accepted and supported by a number of studies (e.g., Bowdle & Gentner, 2005; Chiappe & Kennedy, 2001; Chiappe et al., 2003; Gentner & Bowdle, 2001; Glucksberg & Haught, 2006b; Gregory & Mergler, 1990), and thus it is used as a tool for examining whether a comparison process or a categorization process is used during metaphor (or simile) comprehension.

via a categorization process invited by the form and as a result both metaphors and similes are comprehended via a comparison process, similes should be preferred over and more comprehensible than metaphors. On the contrary, metaphors should be preferred over and more comprehensible than similes, if both metaphors and similes are comprehended via a categorization process because similes cannot be comprehended as comparisons.

Bowdle and Gentner's (2005) career of metaphor theory states that, although metaphors are basically processed as comparisons, conventional metaphors whose vehicle refers to a conventional metaphoric category are processed as categorizations. Therefore, the career of metaphor theory predicts that in the case of novel topic–vehicle pairs similes are preferred over and comprehensible than metaphors, while in the case of conventional pairs metaphors and similes are equally preferred and comprehensible. Bowdle and Gentner (2005) demonstrated that the experimental results were consistent with this prediction.[2]

The attributive category theory (Glucksberg, 2001; Glucksberg & Keysar, 1990) claims that metaphors are processed as categorization statements expressing that the topic is a member of an abstract superordinate category exemplified by the vehicle, but its recent development (Jones & Estes, 2005, 2006; Glucksberg, 2003; Glucksberg & Haught, 2006a, 2006b) has advocated that metaphor aptness, not vehicle conventionality, determines the choice of comprehension strategy. Apt metaphors are processed as categorizations, but less apt metaphors may be processed as comparisons after initially processed as categorizations because in the case of less apt metaphors a categorization does not make sense. Jones and Estes (2005, 2006) also demonstrated that more apt metaphors were more likely to be processed as categorizations and thus achieved higher metaphor preference than less apt metaphors.

Against these theories, the interpretive diversity view argues that interpretively more diverse pairs should be easier to process as categorizations, and thereby easier to interpret as metaphors. Hence, although metaphors are initially processed as categorizations, less diverse metaphors fail to be processed as categorizations and thus they must be reinterpreted as comparisons. The rationale for why interpretive diversity determines the extent to which metaphors are comprehended via a categorization process may lie in the nature of categorization. When an object X is a member of a category Y, which is often expressed by a categorization statement "An X is a Y," X is expected to have many equally salient features to be included in the category Y because members of a category inherit many features of the category by default. Hence the categorization process proceeds more easily when more

[2]One finding that appears to be inconsistent with the prediction was that conventional pairs were comprehended faster as metaphors than similes. But Bowdle and Gentner (2005) pointed out that such shorter comprehension time of conventional metaphors, which they argued were understood as categorizations, was explained by the assumption that the process of comparison required more complex alignment between the topic and the vehicle than the process of categorization.

diverse features of the category Y can be attributed to X, in other words, a topic-vehicle pair X–Y is more diverse. As a result, diverse metaphors are comprehended via a categorization process but less diverse metaphors are comprehended via a comparison process.

EXPERIMENT 1

The interpretive diversity view predicts a positive correlation between the interpretive diversity of topic–vehicle pairs and their preference for the metaphor form. Specifically, when topic–vehicle pairs are highly diverse, there should be no strong preference for either form, because they may be comprehended directly as either comparisons or categorizations according to the link between form and function. On the other hand, in the case of less diverse pairs, the simile form is predicted to be preferred over the metaphor form, because they are difficult to comprehend as categorizations.

In Experiment 1, I tested these predictions by asking participants to indicate which grammatical form they preferred for each pair. I also assessed interpretive diversity of each pair by asking participants to list features shared by the topic and the vehicle and to rate the salience of each listed feature. Furthermore, in order to test whether the interpretive diversity view is superior to the other views (i.e., the aptness view, the similarity view, and the conventionality view), I collected aptness, similarity, and conventionality ratings for each pair. I also addressed an additional property of topic–vehicle pairs, *familiarity* of a topic–vehicle pairing, and collected familiarity ratings. The reason for addressing familiarity is that familiarity may possibly affect metaphor preference (Chiappe & Kennedy, 1999, 2001); for example, since Shakespeare's words "*Juliet is the sun*" is very famous and familiar, people may strongly prefer to express this comparison as metaphors independently of other properties of the comparison.

Method

Participants. One hundred and sixty-four undergraduate and graduate students at the University of Electro-Communications participated as volunteers. All participants were native speakers of Japanese and they did not get any credit for their participation in the experiment. None of them was familiar with metaphor research prior to the experiment.

Materials. Thirty pairs of Japanese topic and vehicle words were used for the experiment. (See Appendix A for the pairs.) These pairs were derived from a pilot study.

Pilot study. For the pilot study, I used 40 Japanese topic–vehicle pairs. They were composed of 22 pairs chosen as appropriate for Japanese figurative comparisons from among a list of 30 English pairs used in Chiappe and Kennedy's (1999) experiments, and 18 pairs chosen from a list of figurative expressions frequently used for Japanese metaphors or similes. Forty undergraduate and graduate students at the University of Electro-Communications participated in the pilot study. None of them participated in the main study. They were assigned to all the 40 topic–vehicle pairs and asked to rate each pair with respect to topic–vehicle similarity on a 5-point scale ranging from 1 (*not at all similar*) to 5 (*extremely similar*). The ratings for each pair were then averaged across participants. Based on the mean similarity ratings, 30 pairs were chosen for the main study such that they were distributed as uniformly as possible with respect to similarity. These pairs consisted of 14 pairs chosen from among Chiappe and Kennedy's (1999) English pairs and 16 pairs chosen from Japanese metaphors.

Procedure. The main study of Experiment 1 consisted of six tasks: metaphor preference rating task, shared feature listing task, feature salience rating task, aptness rating task, vehicle conventionality rating task and familiarity rating task. Participants were assigned to 30 topic–vehicle pairs and carried out only one of the six tasks. The presentation order of 30 pairs was randomized for each participant.

In the metaphor preference rating task, 30 participants were presented with the 30 topic–vehicle pairs both in the metaphor form (e.g., "*Deserts are ovens*" ["Sabaku ha o-bun da" in Japanese]) and in the simile form (e.g., "*Deserts are like ovens*" ["Sabaku ha o-bun no you da" in Japanese]). They were asked to rate which form (i.e., metaphor or simile) was preferable or more appropriate on a 5-point scale ranging from 1 (*simile is preferable*) through 3 (*not sure which is preferable*) to 5 (metaphor is preferable). In the shared feature listing task, other 30 participants were asked to read each topic–vehicle pair and to list two or more features that they thought were shared by the topic and the vehicle. After all participants completed the listing task, the features listed for each pair were subjected to the following feature combination process. First, closely related words and phrases were accepted as the same feature if they met any of the following four criteria: two words or phrases belonged to the same deepest category of a Japanese thesaurus BunruiGoiHyo (National Institute for Japanese Language, 1964); they shared the same root form; they differed only in degree because of an intensive modifier; or a dictionary description of one word included the other word or phrase. (These criteria were identical to the criteria used in Utsumi's (2005) experiment.) After that, any features listed by only one participant were dropped. As a result, a total of 233 features were selected as shared. (A topic–vehicle pair had 4 features at a minimum and up to 13 features at a maximum.) These features were used for the salience rating task. In the salience rating task, 30 participants were assigned to all the 30 pairs with the 233 shared features listed in the feature listing task. They were asked to read each pair and its shared features, and to judge how salient these fea-

tures were when they were seen as shared by the topic and the vehicle using a 7-point scale ranging from 0 (*not at all salient*) through 3 (*not sure whether salient or not*) to 6 (*extremely salient*).

In the aptness rating task, 36 participants were presented with 15 topic–vehicle pairs in the metaphor form and the other 15 pairs in the simile form. Two forms of each pair were counterbalanced so that they were rated an equal number of times. The participants were asked to rate how apt each metaphor or simile was on a 7-point scale ranging from 0 (*not at all apt*) through 3 (*not sure whether apt or not*) to 6 (*extremely apt*). In the vehicle conventionality rating task, 18 participants were assigned to the 30 vehicle words used in the topic–vehicle pairs, and asked to rate how conventional it is to use a vehicle term to convey a particular meaning, that is, the most salient feature of the pair determined in the salience rating task. For example, they rated how conventional it is to use the word "oven" to refer to something that is burning hot. The rating was done on a 7-point scale ranging from 0 (*very novel*) through 3 (*not sure whether conventional or novel*) to 6 (*very conventional*). In the familiarity rating task, 20 participants were asked to read topic–vehicle pairs presented in the mixed form such as "*Deserts are (like) ovens*" and to rate how familiar they were with the pairing on a 7-point scale ranging from 0 (*not at all familiar*) through 3 (*not sure whether familiar or not*) to 6 (*extremely familiar*).

Results and Discussion

For each pair, the similarity rating in the pilot study and the four ratings in the main study (i.e., metaphor preference, aptness, vehicle conventionality and familiarity) were averaged across participants. The mean similarity rating across 30 pairs was 2.96 (SD = 0.83), ranging from 1.45 (freedom–the dark) to 4.25 (hope–light). The mean metaphor preference rating across 30 pairs was 2.82 (SD = 0.69), ranging from 1.57 (tree–umbrella) to 4.20 (history–footprints). The mean metaphor preference rating was significantly lower than the midpoint 3 by participant analysis, $t_p(29) = 3.03$, $p < .01$, but it was not significant by item analysis, $t_i(29) = 1.39$, $p = .17$, indicating that there was somewhat a priori preference for the simile form. The mean aptness rating across 30 pairs was 3.60 (SD = 0.90), ranging from 1.81 (giraffe–skyscraper) to 5.44 (gamble–drug). Note that the mean aptness rating in the simile form (M = 3.77) was significantly higher than that in the metaphor form (M = 3.43), $t_i(29) = 3.02$, $p < .01$, $t_p(35) = 3.46$, $p < .01$. The mean conventionality rating across 30 pairs was 3.58 (SD = 1.18), ranging from to 1.06 (freedom–the dark) to 5.56 (gamble–drug). The mean familiarity rating across 30 pairs was 2.93 (SD = 1.27), ranging from 0.53 (cigarettes–time bombs) to 4.89 (life–journey). Interpretive diversity of each pair was calculated by Equation 1 as follows. First, the salience rating for each feature a_i of a topic-vehicle pair A was averaged across participants. Then $p(a_i)$ was assessed in such a way that the mean salience rating of the feature a_i was divided by the sum of all the mean salience ratings of the features a_1,

…, a_n for the pair A. The mean interpretive diversity across 30 pairs was 2.87 (SD = 0.42), ranging from 1.95 (deserts–ovens) to 3.68 (life–journey).

The simple correlation analysis revealed that, as predicted, the interpretive diversity of shared features was positively correlated with the metaphor preference rating, $r = .53$, $p = .002$. This finding is consistent with the prediction of the interpretive diversity view that as the interpretive diversity increases, the preference for the metaphor form increases as well. Furthermore, a one-sample t-test conducted on high-diversity pairs and low-diversity pairs showed results fully consistent with the predictions about difference in metaphor processing. The mean preference rating across the 10 topic–vehicle pairs with highest interpretive diversity ($M = 3.10$, $SD = 0.59$) was not significantly different from the midpoint 3, $t_i(9) = 0.54$, $p = .60$, $t_p(29) = 1.27$, $p = .21$, suggesting that in the case of diverse pairs there was no preference for one form over another. On the other hand, the mean preference rating across the 10 lowest-diversity pairs ($M = 2.41$, $SD = 0.61$) was significantly lower than the midpoint, $t_i(9) = 3.08$, $p < .05$, $t_p(29) = 8.23$, $p < .001$, showing that less diverse pairs were preferred in the simile form.

Table 1 (the fifth column) shows correlations of the five properties with metaphor preference. The aptness, similarity, and conventionality of topic–vehicle pairs were all correlated positively with metaphor preference, as predicted by each of the three existing views. As expected, the familiarity of topic–vehicle pairs was also correlated positively with metaphor preference. Furthermore, a one-sample t-test yielded the same result on processing difference for these properties. None of the mean preferences of 10 most apt pairs ($M = 3.34$, $SD = 0.71$), 10 most similar pairs ($M = 3.29$, $SD = 0.71$), 10 most conventional pairs ($M = 3.14$, $SD = 0.71$), and 10 most familiar pairs ($M = 3.15$, $SD = 0.77$) was significantly different from 3, $t_i(9) = 1.53$, $p = .16$ (but $t_p(29) = 3.16$, $p < .01$) for apt pairs; $t_i(9) = 1.28$, $p = .23$ (but

TABLE 1
Regression Analysis of Five Metaphor Properties for Predicting the Metaphor Preference ($n = 30$)

	Regression			*Correlation*				
Variable	*B*	*SE B*	β	*Preference*	*2*	*3*	*4*	*5*
1. Interpretive diversity	0.81	0.27	.48*	.53**	.44*	.23	–.14	.39*
2. Aptness	0.28	0.19	.37	.63**	—	.72***	.42*	.80***
3. Topic–vehicle simiarity	0.04	0.21	.05	.45*		—	.47**	.82***
4. Vehicle conventionality	0.22	0.10	.38*	.43*			—	.32
5. Familiarity	–0.10	0.15	–.18	.46**				—

Note. $R^2 = .59$, $F(5, 24) = 6.86$, $p < .001$.
*$p < .05$ **$p < .01$. ***$p < .001$.

$t_p(29) = 3.45$, $p < .01$) for similar pairs; $t_i(9) = 0.61$, $p = .56$, $t_p(29) = 1.67$, $p = .11$ for conventional pairs; $t_i(9) = 0.60$, $p = .56$, $t_p(29) = 1.55$, $p = .13$ for familiar pairs. On the other hand, all the mean preferences of 10 least apt pairs ($M = 2.47$, $SD = 0.64$), 10 least similar pairs ($M = 2.41$, $SD = 0.59$), 10 least conventional pairs ($M = 2.55$, $SD = 0.59$), and 10 least familiar pairs ($M = 2.45$, $SD = 0.61$) were significantly lower than 3, $t_i(9) = 2.61$, $p < .05$, $t_p(29) = 7.41$, $p < .001$ for less aptness pairs; $t_i(9) = 3.17$, $p < .05$, $t_p(29) = 7.84$, $p < .001$ for less similar pairs; $t_i(9) = 2.41$, $p < .05$, $t_p(29) = 5.35$, $p < .001$ for less conventional pairs; $t_i(9) = 2.85$, $p < .05$, $t_p(29) = 7.00$, $p < .001$ for less familiar pairs.

To examine which of the five properties best explains metaphor preference, I conducted multiple regression analysis with preference rating as the dependent variable. Table 1 shows the result of the regression analysis (the second through the fourth columns), together with simple correlations among the independent variables (the sixth through the last columns). Only two variables, interpretive diversity and vehicle conventionality, accounted for a significant portion of the variance in metaphor preference. These standardized regression coefficients were positive; as the interpretive diversity or the conventionality of vehicles increased, the preference for the metaphor form increased as well. In addition, the standardized regression coefficient was larger for interpretive diversity than for vehicle conventionality. Other three variables—aptness, topic–vehicle similarity, familiarity—were not significantly related to metaphor preference. Since very strong mutual correlations among these three variables might cause a multicollinearity problem, I conducted another regression analysis with only interpretive diversity, vehicle conventionality, and aptness as the independent variables. Aptness was chosen because it had the highest simple correlation with metaphor preference among these three variables. The result was the same: interpretive diversity (ß = .47) and vehicle conventionality (ß = .38) accounted for a significant portion of the variance, $F(1,26) = 3.05$, $p < .01$ for interpretive diversity; $F(1,26) = 2.51$, $p < .05$ for conventionality. But aptness (ß = .26) was not related to metaphor preference, $F(1,26) = 1.54$, $p = .14$.

To further compare the importance of interpretive diversity, vehicle conventionality, and aptness, I also conducted commonality analysis with these three variables (Pedhazur, 1997). Commonality analysis is a method of variance partitioning by which we can calculate proportions of variance in the dependent variable associated uniquely with each of independent variables, as well as the proportions of variance attributed to various combinations of independent variables. Table 2 shows the result of the commonality analysis.[3] Interpretive diversity made the largest unique contribution to metaphor preference and vehicle conventionality also had a considerable unique contribution; 26.0% and 17.6% of the explained variance in metaphor preference were associated uniquely with interpretive diversity

[3]It must be noted that common contribution, i.e., the variance explained by a joint effect of multiple variables, may have a negative value, which means a lack of a joint effect.

TABLE 2
Unique and Common Contributions of Three Metaphor Properties in Accounting for the Variance in Metaphor Preference

Unique Contributions			*Common Contributions*				
ID	*AP*	*VC*	*ID & AP*	*ID & VC*	*AP & VC*	*ID, AP &VC*	*SUM*
.151	.038	.102	.209	–.070	.154	–.004	.580

Note. ID = Interpretive diversity; AP = Aptness; VC = Vehicle conventionality

and with vehicle conventionality, respectively. On the other hand, aptness uniquely explained only 6.6% of the explained variance. But instead, aptness had indirect effects on metaphor preference through what aptness shares with interpretive diversity or vehicle conventionality; 62.5% of the explained variance (36.0% with interpretive diversity and 26.5% with conventionality) was attributed to the joint effect of aptness.

All these findings indicate that interpretive diversity is the most important factor in explaining metaphor preference, and thus the interpretive diversity view is more plausible than other existing views on metaphor–simile distinction. However, vehicle conventionality does not lose its role as a supplementary source of metaphor preference; the conventionality view (i.e., the career of metaphor view) seems somewhat plausible. The aptness view may be much less plausible, but aptness functions as enhancement of diversity's or conventionality's effects on metaphor preference.[4]

Concerning familiarity, the result that familiarity was correlated with metaphor preference due to the strong correlation with aptness and topic–vehicle similarity replicates Chiappe and Kennedy's (2001) finding that familiarity has its effect on metaphor preference through enhancing aptness or similarity. Moreover, the lack of significant correlation between familiarity and vehicle conventionality strengthens the argument that vehicle conventionality is a different property from the conventionality of topic–vehicle pairing (Bowdle & Gentner, 2005; Chiappe & Kennedy, 2001).

[4]A joint effect of vehicle conventionality and aptness on metaphor preference was also demonstrated recently by Nakamoto and Kusumi (2004). They found that the effect of vehicle conventionality on metaphor preference was greater when topic–vehicle pairs were highly apt than when they were moderately apt.

EXPERIMENT 2

Experiment 1 used metaphor preference as a measure of metaphor-simile distinction and provided empirical evidence in favor of the interpretive diversity view. Experiment 2, on the other hand, addressed another measure, i.e., difference in comprehensibility between two figurative forms of a topic–vehicle pair. The difference in comprehensibility of a topic–vehicle pair was calculated as the comprehensibility rating of the metaphor form minus the comprehensibility rating of the simile form. i.e., the degree to which metaphor comprehensibility is higher than simile comprehensibility.

The interpretive diversity view predicts a positive correlation between interpretive diversity and difference in comprehensibility; an interpretively more diverse pair will achieve a greater difference in comprehensibility. Specifically, when topic–vehicle pairs are highly diverse, it is predicted that there should not be so much of difference in comprehensibility between both forms. When topic–vehicle pairs are less diverse, on the other hand, it is predicted that the metaphor form should be less comprehensible than the corresponding simile form. In Experiment 2, I tested these predictions by collecting comprehensibility ratings of metaphors and those of similes separately. As in Experiment 1, I also collected aptness, similarity and conventionality ratings and examined again whether the interpretive diversity view is more plausible than other competing views.

Method

Participants. Two hundreds and twenty-eight undergraduate students (144 students of the University of Electro-Communications and 84 students of Japan Women's University) participated in the experiment. Students at the University of Electro-Communications participated as volunteers and Japan Women's University students participated for a requirement of computer literacy course. All participants were native speakers of Japanese. None of them was familiar with metaphor research prior to the experiment.

Materials. Forty pairs of Japanese topic and vehicle words, which were used in the experimental study (Utsumi, 2005), were used for this experiment. These pairs consisted of 10 groups of four topic–vehicle pairs. Four pairs in each group were constructed from all possible pairings of two topic words with two vehicle words. For example, from the two topics "anger" ("ikari" in Japanese) and "sleep" ("nemuri" in Japanese), and the two vehicles "sea" ("umi" in Japanese) and "storm" ("arashi" in Japanese), the following four topic–vehicle pairs were created: *anger–sea*, *anger–storm*, *sleep–sea*, and *sleep–storm*. The complete list of 10 groups of pairs is provided in Appendix B. Topic and vehicle words were selected from an experimental study on Japanese metaphor (Kusumi, 1987) and from a list of words frequently used for Japanese metaphors (Nakamura, 1995). Topic–vehicle

pairs were presented in the metaphor form (e.g., "*Anger is the sea*" ["Ikari ha umi da" in Japanese]) or in the simile form (e.g., "*Anger is like the sea*" ["Ikari ha umi no you da" in Japanese]).

Procedure. This experiment was conducted separately for metaphor comprehension, simile comprehension, and rating of topic–vehicle pairs.

The experiment for metaphor comprehension was carried out on a PC computer using a CGI program designed for the experiment. Participants seated in front of a computer display, accessed the CGI page by a Web browser (Internet Explorer), entered their user ID (student number) and password, and started the experiment. The CGI pages were programmed so that participants could not move on the next page unless completing the task of the current page. Forty-two participants of Japan Women's University each were assigned to two metaphors that shared neither the topic nor the vehicle (i.e., "*Anger is the sea*" and "*Sleep is a storm*") from each of the 10 groups of pairs, and thereby a total of 20 metaphors. Metaphors of each group were counterbalanced so that they were assigned to 21 participants. Participants carried out three subtasks: feature listing task, free description task, and comprehensibility rating task. First, they were asked to consider the meaning of a metaphor and to list three or more features of the topic that were being described by the vehicle of that metaphor. Second, they were asked to describe their own interpretation of the metaphor freely by sentences. (In this paper, however, the data obtained in the free description task was not used in the analysis.) Finally, they were asked to rate the metaphor with respect to ease of comprehension on a 7-point scale ranging from 0 (*not at all comprehensible*) through 3 (*not sure whether comprehensible*) to 6 (*extremely comprehensible*).

The experiment for simile comprehension was also carried out on a PC computer using a CGI program in the same way as the metaphor comprehension experiment. Another 42 participants of Japan Women's University were assigned to two similes that shared neither the topic nor the vehicle (i.e., "*Anger is like the sea*" and "*Sleep is like a storm*") from each of the 10 groups of pairs, and thereby a total of 20 similes. Similes were counterbalanced across participants so that they were assigned to 21 participants. Each participant carried out the same three subtasks as those of the metaphor comprehension experiment, i.e., feature listing task, free description task, and comprehensibility rating task.

In the experiment for rating of topic–vehicle pairs, 72 Japanese students of the University of Electro-Communications were assigned to 20 topic–vehicle pairs that shared neither the topic nor the vehicle, half of which were presented in the metaphor form and the other half of which were presented in the simile form. The order of the metaphors and similes was randomized for each participant. Topic–vehicle pairs and the presentation form were counterbalanced across participants so that each pair was presented in the metaphor form to 18 participants and in the simile form to other 18 participants. The participants were given a booklet containing 10 metaphors and 10 similes. They were then asked to rate how apt each metaphor

or simile was on a 7-point scale ranging from 0 (*not at all apt*) through 3 (*not sure whether apt*) to 6 (*extremely apt*). Another 72 students of the University of Electro-Communications were assigned to 20 topic–vehicle pairs that shared neither the topic nor the vehicle, half of which were used for vehicle conventionality rating task and the other half of which were used for topic–vehicle similarity rating task. The order of two tasks was randomized so that half of the time the conventionality rating task was done first and half of the time the topic–vehicle similarity rating task was done first. In the conventionality rating task, the participants were asked to rate how conventional the canonical meaning of the figurative expression of each topic–vehicle pair was as an alternative sense of the vehicle term on a 7-point scale ranging from 0 (*very novel*) through 3 (*not sure whether conventional or novel*) to 6 (*very conventional*). As the canonical meaning, I chose the feature listed by the largest number of participants in the experiment for metaphor comprehension and simile comprehension. For example, the meaning "anger is fierce" was listed by the largest number of participants (17) for "*Anger is (like) the sea*," and thus the participants of this task were asked, "When we say that X is the sea, how conventional do we mean that X is fierce?" In the similarity rating task, the participants were presented with a topic–vehicle pair and asked to rate how similar the topic and the vehicle are on a 7-point scale ranging from 0 (*not at all similar*) through 3 (*not sure whether similar*) to 6 (*extremely similar*).

Results and Discussion

For each metaphor and simile, the comprehensibility rating was averaged across participants. The mean comprehensibility rating across 40 metaphors was 3.32 (SD = 1.05), ranging from 0.90 ("*A perfume is ice*") to 5.24 ("*Children are jewels*"). The mean comprehensibility rating across 40 similes was 3.52 (SD = 0.99), ranging from 1.14 ("*A perfume is like ice*") to 5.00 ("*Life is like a journey*"). For each topic–vehicle pair, I then calculated difference in comprehensibility between two forms by subtracting the mean comprehensibility rating for the simile form from the mean comprehensibility rating for the metaphor form. Hence the difference in comprehensibility of a topic–vehicle pair is positive when the metaphor form of that pair is more comprehensible than the corresponding simile form. The mean difference in comprehensibility across 40 pairs was –0.20 (SD = 0.49), ranging from –1.38 (character–fire) to 0.57 (love–journey). The mean difference in comprehensibility was significantly smaller than 0 by item analysis, $t_i(39) = 2.58$, $p = .01$ (although not significant by participant analysis, $t_p(82) = 1.45$, $p = .15$), thus indicating that, in general, similes were comprehended somewhat more easily than metaphors.

Concerning the other three properties, the ratings of each pair were averaged across participants. The mean aptness rating across 40 pairs was 2.91 (SD = 1.05), ranging from 0.94 (perfume–ice) to 5.17 (life–journey). Note that the mean aptness rating in the simile form (M = 3.13, SD = 1.07) was significantly higher than

that in the metaphor form (M = 2.70, SD = 1.07), $t_i(39) = 5.85$, $p < .0001$, $t_p(71) = 3.94$, $p < .0001$. The mean conventionality rating across 40 pairs was 3.36 (SD = 1.19), ranging from 0.83 (memory–arrow) to 5.28 (lover–sun). The mean similarity rating across 40 pairs was 2.85 (SD = 1.12), ranging from 1.06 (perfume–ice, voice–doll) to 4.94 (anger–storm).

The interpretive diversity of a topic–vehicle pair was calculated in the following way. First, a list of features listed either in the metaphor comprehension experiment or in the simile comprehension experiment was generated for each pair. Then closely related words or phrases in the generated list of features were accepted as the same feature if they met one of the four criteria used in Experiment 1. After this feature combination process, any features mentioned by only one participant were eliminated from the list of features. Finally, for each feature a_i in the amended list of features, $p(a_i)$ of Equation 1 was assessed as a ratio of the number of tokens of the feature a_i (i.e., the number of participants who listed that feature) to the total number of tokens involved in the list. The mean interpretive diversity across 40 pairs was 3.41 (SD = 0.36) ranging from 2.80 (sky–mirror) to 4.08 (love–journey).

The simple correlation analysis revealed that, as shown in the last column of Table 3, difference in comprehensibility of topic–vehicle pairs was positively correlated with the interpretive diversity of these pairs, $r = .41$, $p < .01$. As interpretive diversity increased, relative comprehensibility of the metaphor form increased as well. On the other hand, none of the other properties—aptness, vehicle conventionality and topic–vehicle similarity—was correlated with difference in comprehensibility. Furthermore, to avoid an undesirable effect on interpretive diversity

TABLE 3
Regression Analysis of Four Variables for Predicting Difference in Comprehensibility Between Metaphor and Simile

	Four Variables			*Three Variables*			
	B	*SE B*	β	*B*	*SE B*	β	*r*
All pairs[a]							
Interpretive diversity	0.73	0.21	.57***	0.74	0.20	.54***	.41**
Aptness	0.01	0.16	.02	0.14	0.07	.31	.17
Vehicle conventionality	0.06	0.06	.14	0.06	0.06	.15	.16
Topic–vehicle similarity	0.14	0.16	.33	—	—	—	.17
Comprehensive pairs[b]							
Interpretive diversity	0.58	0.20	.54**	0.58	0.19	.54**	.57**
Aptness	0.10	0.15	.21	0.10	0.08	.21	.15
Vehicle conventionality	–0.09	0.08	–.20	–0.09	0.08	–.20	–.27
Topic–vehicle similarity	0.00	0.15	.01	—	—	—	–.05

Note. $R^2 = .31$, $F(4, 35) = 4.03$, $p < .01$ (four variables) and $R^2 = .30$, $F(3, 36) = 5.11$, $p < .01$ (three variables) for all pairs. $R^2 = .43$, $F(4, 20) = 3.72$, $p < .05$ (four variables) and $R^2 = .42$, $F(3, 21) = 5.15$, $p < .01$ (three variables) for comprehensible pairs.

[a]$n = 40$. [b]$n = 25$. *$p < .05$. **$p < .01$. ***$p < .001$.

caused by semantically irrelevant features that participants generated for less comprehensible metaphors possibly due to the experimental condition that they were forced to list at least three meanings, I restricted the correlation analysis to comprehensible pairs whose mean comprehensibility rating for both forms was the midpoint 3 or higher. The result was that, as shown in Table 3, a higher correlation ($r = .57$) was observed between difference in comprehensibility and interpretive diversity, but the other properties again showed no correlations with difference in comprehensibility. These findings clearly show that the metaphor form is relatively easy to comprehend as compared to the simile form to the extent that a potential figurative meaning is interpretively diverse, thus providing empirical evidence in favor of the interpretive diversity view and against the other views.

To further attest to the superiority of the interpretive diversity view, I conducted a multiple regression analysis with difference in comprehensibility as the dependent variable and with four properties as the independent variables. As shown in Table 3 (the second through the fourth columns), the result was again in favor of the interpretive diversity view. Regardless of whether all pairs or only comprehensible pairs were considered, only the interpretive diversity accounted for a significant portion of the variance in difference in comprehensibility. I also conducted an additional regression analysis in which topic–vehicle similarity was removed from the independent variables because of a high correlation between similarity and aptness ($r = .92$ for all pairs and $r = .82$ for comprehensible pairs). The result, which is shown in the fifth through the seventh columns of Table 3, was almost the same as in the regression analysis with four independent variables: only the interpretive diversity was significantly related to difference in comprehensibility. These findings lend further support to the interpretive diversity view.

I also examined which of the two forms was more comprehensible when topic–vehicle pairs were highly diverse and when they were less diverse by choosing the 10 highest-diversity pair and 10 lowest-diversity pairs from the 40 pairs or from the 25 comprehensible pairs. Table 4 lists these mean differences in comprehensibility as well as the mean comprehensibility ratings of metaphors and similes. For 10 highest-diversity pairs, there were no significant differences in comprehensibility between the metaphor and the simile forms (all t_is, t_ps< 1). This result is consistent with the prediction that both forms of diverse pairs do not differ in ease of comprehension. On the other hand, in the case of 10 lowest-diversity pairs, difference in comprehensibility was significantly less than 0, $t_i(9) = 5.13$, $p < .001$, $t_p(82) = 2.32$, $p < .05$ for all pairs and $t_i(9) = 4.18$, $p < .01$, $t_p(82) = 2.21$, $p < .05$ for comprehensible pairs. This result is also consistent with the prediction that less diverse pairs are more comprehensible in the simile form than in the metaphor form because they are too less diverse to be comprehended as categorizations.

Concerning aptness and vehicle conventionality, no expected differences in comprehensibility were observed between metaphors and similes. The aptness view predicts that the simile form of less apt pairs should be more comprehensible than the corresponding metaphor form, but there were no significant differences in

TABLE 4
Difference in Comprehensibility for 10 Most Diverse, Apt or Conventional Pairs and for 10 Least Diverse, Apt or Conventional Pairs

	Comprehensibility					
	All[a]			*Comprehensible*[b]		
	Difference	*Metaphor*	*Simile*	*Difference*	*Metaphor*	*Simile*
Interpretive diversity						
High	–0.02	3.88	3.90	–0.02	4.92	4.94
Low	–0.45***	4.42	4.87	–0.42**	4.66	5.06
Aptness						
High	–0.04	5.56	5.60	–0.04	5.56	5.60
Low	–0.37	3.20	3.57	–0.09	4.31	4.40
Vehicle conventionality						
High	–0.14	4.66	4.80	–0.28*	5.24	5.52
Low	–0.30	3.86	4.16	–0.07	4.65	4.72

[a]$n = 40$. [b]$n = 25$.
*$p < .05$. **$p < .01$. ***$p < .001$.

comprehensibility for the 10 lowest-aptness pairs $t_i(9) = 2.23$, $P = .053$, $t_p(82) = 1.78$, $p = .08$ (all pairs), $t_i(9) = 0.56$, $p = .59$, $t_p(82) = 0.28$, $p = .78$ (comprehensible pairs). In the same way, the conventionality view predicts that the simile form of low-conventionality pairs should be more comprehensible, but no significant differences were observed for the 10 lowest-conventionality pairs, $t_i(9) = 1.63$, $p = .14$, $t_p(82) = 0.82$, $p = .42$ (all pairs), $t_i(9) = 0.52$, $p = .62$, $t_p(82) = 0.33$, $p = .74$ (comprehensible pairs). What is worse for the conventionality view is that 10 highest-conventionality pairs chosen from comprehensible ones showed a significant difference such that similes were more comprehensible than metaphors, $t_i(9) = 2.30$, $p < .05$, $t_p(82) = 1.97$, $p = .052$, which is clearly opposite to the prediction of the conventionality view. These results make the aptness and conventionality views less plausible.

As an additional measure for whether metaphors are understood via a categorization process or a comparison process, we can think of the degree to which a metaphor and its corresponding simile share the same meanings. According to the dual reference property of metaphor vehicles (Glucksberg, 2001; Glucksberg & Haught, 2006b), the metaphor vehicle refers to two categories: a literal basic-level category and a metaphorical superordinate category. A superordinate category is used for comprehending figurative expressions as categorizations, whereas a literal basic-level category is used for comprehending them as comparisons. Hence it is predicted that metaphors and similes may share a smaller proportion of mean-

TABLE 5
Mean Coincidence Rate for 10 Most Diverse, Apt or Conventional Pairs and for 10 Least Diverse, Apt or conventional Pairs

	Coincidence Rate			
	All[a]		*Comprehensible*[b]	
	High	*Low*	*High*	*Low*
Interpretive diversity	.56	.78	.61	.79
Aptness	.66	.60	.66	.68
Vehicle conventionality	.64	.71	.70	.68

[a]$n = 40$. [b]$n = 25$.

ings when they are comprehended by different processes (i.e., metaphors are processed as categorizations and similes are processed as comparisons) than they are comprehended by the same process, i.e., comparison. As a measure of to what degree a metaphor and a simile share the same meaning, I assessed a concordance rate of both interpretations for each pair by generating a list of meanings separately for the metaphor and simile forms and then calculating the ratio of the number of tokens of common meanings included in both lists to the total number of tokens included in either list. The comparison between high-diversity and low-diversity pairs, which is shown in Table 5, revealed that the 10 highest-diversity pairs had a significantly lower concordance rate than the 10 lowest-diversity pairs, $t_i(9) = 4.46$, $p < .01$ (all pairs) and $t_i(9) = 4.42$, $p < .001$ (comprehensible pairs). However, there were no significant differences in concordance rate between 10 highest-aptness and 10 lowest-aptness pairs, $t_i(9) = 1.04$, $p = .31$ (all pairs), $t_i(9) = 0.34$, $p = .74$ (comprehensible pairs), and no significant differences between 10 highest-conventionality and 10 lowest-conventionality pairs, $t_i(9) = 1.21$, $p = .24$ (all pairs) and $t_i(9) = 0.45$, $p = .66$ (comprehensible pairs). These findings can also be seen as providing empirical evidence in favor of the interpretive diversity view; interpretive diversity, not aptness nor vehicle conventionality, determines whether metaphors are understood as categorizations or comparisons.

GENERAL DISCUSSION

The two experiments reported in this paper provided empirical evidence in favor of the argument that interpretive diversity is a major source of metaphor-simile distinction. It was demonstrated that, as the interpretive diversity of a figurative pair increased, so did the preference for and the relative comprehensibility of the metaphor form. This relation was found to be most crucial among the relations proposed by several views on metaphor-simile distinction.

The interpretive diversity view also posits that, as the attributive category theory (Glucksberg, 2001; Glucksberg & Keysar, 1990) claims, metaphors are basically processed as categorizations, but when the figurative meaning is less diverse a categorization process is discouraged or obstructed and thus metaphors are reinterpreted via a comparison process. Although the two experiments did not use online measures of processing and thus did not provide sufficient evidence for my argument about processing difference, the fact that high-diversity pairs and low-diversity pairs consistently yielded different results on metaphor preference, relative comprehensibility and concordance rate may indicate the possibility that interpretive diversity view is successful in explaining the processing difference of metaphors. To further justify the argument of the interpretive diversity view, further studies are needed that empirically explore more online aspects of processing and that computationally simulate human behavior of metaphor comprehension (e.g., Utsumi, 2006).

A number of studies have shown empirical findings on the effects of the semantic richness on language comprehension in general, especially on lexical access. These findings are highly relevant to the interpretive diversity view. Among such findings, an ambiguity effect or an ambiguity advantage is the most well-known finding that lexical decisions were faster for semantically ambiguous words than for unambiguous words (e.g., Borowsky & Masson, 1996; Rodd, Gaskell, & Marslen-Wilson, 2002). Similarly, Pexman, Lupker, and Hino (2002) have found a number-of-feature effect, i.e., faster lexical decision responses for words with many semantic features than words with fewer semantic features. Furthermore, Pexman, Holyk, and Monfils (2003) demonstrated that the number-of-feature effect was also observed in the semantic categorization task and that such effect was greater when people made decisions about broader, and thus richer, semantic categories. These findings provide evidence for a general mechanism that the richness of the semantic representation of words facilitates word recognition, and the interpretive diversity view can be seen as based on such general mechanism. In other words, the findings presented in this paper suggest that the general mechanism of semantic richness may also underlie metaphor comprehension.

One may argue that the analysis of semantic features by which interpretive diversity or semantic richness is assessed cannot reveal how metaphorical meanings are understood. Especially cognitive linguistic researchers (e.g., Gibbs, 2006; Lakoff & Johnson, 1980, 1999) have argued that metaphors are comprehended on the basis of preexisting analogical mappings between different domains (or the so-called conceptual metaphors) and image schemata, which the feature analysis cannot explain. (Note also that some studies, e.g., Glucksberg and McGlone, 1999, showed that metaphorical meanings could not be fully derived from the cognitive linguistic analysis such as conceptual metaphors.) I agree with the argument of cognitive linguistics that the feature analysis is not enough to explore the mechanism of metaphor comprehension, but I also think that it is not a drawback of the interpretive diversity view. This study employed the feature analysis to measure

semantic richness quantitatively using the idea of Shannon's entropy, but if an appropriate method will be developed which measures semantic richness through the cognitive linguistic analysis, the interpretive diversity view can be applied to a wider array of metaphors and its applicability can be tested empirically. In other words, the interpretive diversity view may be able to bridge a gap between both analyses if semantic richness can be assessed from the integrated result of both analyses.

To conclude, the findings reported in this paper suggest that interpretive diversity has a potential ability to offer a comprehensive framework for metaphor-simile distinction and the process of metaphor comprehension in general. To be sure, the interpretive diversity view is not complete; as Experiment 1 showed, other properties such as vehicle conventionality had a unique effect on metaphor preference that could not be achieved by interpretive diversity. Hence it would be interesting and vital for further research to explore how the interaction between interpretive diversity and other properties is responsible for metaphor-simile distinction.

ACKNOWLEDGMENT

This research was supported by Grant-in-Aid for Scientific Research C (No. 17500171) and for Encouragement of Young Scientists (No.14780263), The Ministry of Education, Culture, Sports, Science and Technology, Japan. An earlier version of this paper was presented at the 27th Annual Meeting of the Cognitive Science Society (COGSCI05) in Stresa, Italy (July 2005).

REFERENCES

Borowsky, R., & Masson, M. (1996). Semantic ambituity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 63–85.

Bowdle, B., & Gentner, D. (1999). *Metaphor comprehension: From comparison to categorization*. In Proceedings of the 21st Annual Meeting of the Cognitive Science Society (pp. 90–95). Vancouver, BC: Cognitive Science Society.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*(1), 193–216.

Chiappe, D., & Kennedy, J. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin & Review*, *6*, 668–676.

Chiappe, D., & Kennedy, J. (2000). Are metaphors elliptical similes? *Journal of Psycholinguistic Research*, *29*(4), 371–398.

Chiappe, D., & Kennedy, J. (2001). Literal bases for metaphor and simile. *Metaphor and Symbol*, *16*(3&4), 249–276.

Chiappe, D., Kennedy, J., & Chiappe, P. (2003). Aptness is more important than comprehensibility in preference for metaphors and similes. *Poetics*, *31*(1), 51–68.

Chiappe, D., Kennedy, J., & Smykowski, T. (2003). Reversibility, aptness, and the conventionality of metaphors and similes. *Metaphor and Symbol*, *18*(2), 85–105.

Gentner, D., & Bowdle, B. (2001). Convention, form, and figurative language processing. *Metaphor and Symbol*, *16*(3–4), 223–247.

Gibbs, R. (2006). *Embodiment and cognitive science*. New York: Cambridge University Press.

Glucksberg, S. (2001). *Understanding Figurative language: From metaphors to idioms*. New York: Oxford University Press.

Glucksberg, S. (2003). The psycholinguisitics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96.

Glucksberg, S., & Haught, C. (2006a). Can Florida become like the next Florida? When metaphoric comparisons fail. *Psychological Science*, *17*(11), 935–938.

Glucksberg, S., & Haught, C. (2006b). On the relation between metaphor and simile: When comparison fails. *Mind & Language*, *21*(3), 360–378.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*, 3–18.

Glucksberg, S., & McGlone, M. (1999). When love is not a journey: What metaphors mean. *Journal of Pragmatics*, *31*(12), 1541–1558.

Gregory, M., & Mergler, N. (1990). Metaphor comprehension: In search of literal truth, possible sense, and metaphoricity. *Metaphor and Symbolic Activity*, *5*(3), 151–173.

Harris, R., Friel, B., & Mickelson, N. (2006). Attribution of discourse goals for using concrete- and abstract-tenor metaphors and similes with or without discourse context. *Journal of Pragmatics*, *38*(6), 863–879.

Johnson, A. (1996). Comprehension of metaphors and similes: A reaction time study. *Metaphor and Symbolic Activity*, *11*(2), 145–159.

Jones, L., & Estes, Z. (2005). Metaphor comprehension as attributive categorization. *Journal of Memory and Language*, *53*, 110–124.

Jones, L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, *55*, 18–32.

Kusumi, T. (1987). Effects of categorical dissimilarity and affective similarity between constituent words on metaphor appreciation. *Journal of Psycholinguistic Research*, *16*, 577–595.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Nakamoto, K., & Kusumi, T. (2004). *The effect of repeated presentation and aptness of figurative comparisons on preference for metaphor forms*. In Proceedings of the 26th Annual Meeting of the Cognitive Science Society (p. 1611). Chicago, IL: Cognitive Science Society.

Nakamura, A. (1995). *Hiyuhyogenjiten* [Japanese dictionary of metaphorical expressions]. Tokyo: Kadokawa Shoten.

National Institute for Japanese Language (Ed.). (1964). *Bunruigoihyo*[Word list by semantic principles]. Tokyo: Shuei Shuppan.

Pedhazur, E. (1997). *Multiple regression in behavioural research: Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace College Publishing.

Pexman, P., Holyk, G., & Monfils, M.-H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, *31*(6), 842–855.

Pexman, P., Lupker, S., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-feature effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, *9*(3), 542–549.

Pielou, E. (1975). *Ecological diversity.* New York: John Wiley and Sons.

Roberts, R., & Kreuz, R. (1994). Why do people use figurative language? *Psychological Science*, *5*(3), 159–163.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245–266.

Utsumi, A. (2005). The role of feature emergence in metaphor appreciation. *Metaphor and Symbol*, *20*(3), 151–172.

Utsumi, A. (2006). *Computational exploration of metaphor comprehension processes*. In Proceedings of the 28th Annual Meeting of the Cognitive Science Society (pp. 2281–2286). Vancouver, BC: Cognitive Science Society.

Zharikov, S., & Gentner, D. (2002). *Why do metaphors seem deeper than similes?* In Proceedings of the 24th Annual Meeting of the Cognitive Science Society (pp. 976–981). Fairfax, VA: Cognitive Science Society.

APPENDIX A
TOPIC–VEHICLE PAIRS USED IN EXPERIMENT 1

Topic	Vehicle
life (jinsei)	journey (tabi)
gene (idenshi)	blueprint (aojyashin)
crime (hanzai)	disease (byouki)
rooster (niwatori)	alarm clock (mezamashidokei)
mosquito (ka)	vampire (kyuuketsuki)
deserts (sabaku)	ovens (o-bun)
man (otoko)	wolf (ookami)
giraffe (kirin)	skyscraper (kousoubiru)
cigarettes (tabako)	time bombs (jigenbakudan)
education (kyouiku)	stairway (kaidan)
encyclopedia (hyakkajiten)	goldmine (kinmyaku)
idea (aidea)	diamond (daiamondo)
argument (giron)	war (sensou)
sermon (sekkyou)	sleeping pill (suiminyaku)
job (shigoto)	jail (kangoku)
tree (ki)	umbrella (kasa)
hope (kibou)	light (hikari)
gaze (shisen)	lightning (inazuma)
history (rekisi)	footprints (ashiato)
flower (hana)	jewel (houseki)
death (shi)	sleep (nemuri)
silence (chinmoku)	rock (iwa)
anxiety (fuan)	fog (kiri)
anger (ikari)	flame, fire (honoo)
mind (kokoro)	glass (garasu)
heart (sinzou)	engine (enjin)
woman (jyosei)	cat (neko)
freedom (jiyuu)	the dark (yami)
gamble (gyanburu)	drug (mayaku)
life (inochi)	fire (hi)

Appendix B
Groups of Topic–Vehicle Pairs Used in Experiment 2

Topic	Vehicle
life (jinsei)	journey (tabi)
love (ai)	game (ge-mu)
anger (ikari)	sea (umi)
sleep (nemuri)	storm (arashi)
perfume (ko-sui)	bouquet (hanataba)
star (hoshi)	ice (koori)
sky (sora)	mirror (kagami)
eye (me)	lake (mizuumi)
lover (koibito)	sun (taiyo)
hope (kibou)	rainbow (niji)
child (kodomo)	jewel (houseki)
words (kotoba)	water (mizu)
the aged (roujin)	deadwood (kareki)
voice (koe)	doll (ningyou)
character (seikaku)	fire (hi)
marriage (kekkon)	stone (ishi)
death (shi)	night (yoru)
anxiety (fuan)	fog (kiri)
time (jikan)	money (okane)
memory (omoide)	arrow (ya)