

# The Role of Feature Emergence in Metaphor Appreciation

Akira Utsumi

*Department of Systems Engineering  
The University of Electro-Communications*

This study examined how emergent features, which are made salient in the interpretation of metaphor, are related to metaphor appreciation. According to an incongruity resolution model of poetic appreciation, the role of emergent features in metaphor appreciation is predicted to facilitate poetic appreciation by constituting a richer interpretation when topic–vehicle similarity is lower. Two experiments demonstrated that this prediction was supported for comprehensible metaphors. In Experiment 1, more emergent features were generated when comprehensible metaphors with lower topic–vehicle similarity were interpreted, and richer interpretations included more emergent features. In Experiment 2, poeticality rating of comprehensible metaphors was positively correlated with richness of interpretation. Furthermore, Experiment 2 found that conceptual aptness of metaphor also affected poeticality of comprehensible metaphors but, in contrast, that only emotive value of metaphor affected poeticality of less comprehensible metaphors. This finding suggests that the process of poetic appreciation may differ between comprehensible and less comprehensible metaphors.

Emergence of meaning, through which metaphor can create new similarities between a topic and a vehicle, has been recognized to be an important characteristic of metaphor comprehension process by many scholars, in particular by interaction theorists (e.g., Black, 1962; Indurkha, 1991; Ricoeur, 1977; Tourangeau & Sternberg, 1982). Along this line of thought, a number of empirical studies (Becker, 1997; Gineste, Indurkha, & Scart, 2000; Nueckles & Janetzko, 1997; Tourangeau & Rips, 1991) have recently examined *emergent features*, which are not salient either in the representation of the topic or in the representation of the

vehicle but are made salient in the interpretation of metaphor, and they have demonstrated that emergent features dominate the interpretation of metaphor. However, how emergent features are related to metaphor appreciation has never been examined empirically. This is not so surprising because the process of metaphor appreciation in general has received less attention from psychological researchers than the process of metaphor comprehension.

The aim of this study is to examine how emergent features contribute to metaphor appreciation by two experiments. I especially examined the relation between emergent features and people's poeticality rating in terms of an "incongruity resolution" model of poetic appreciation (Utsumi, 2002), which claims that poetic value is appreciated when processing effort caused by an incongruity involved in a metaphor is rewarded by a rich interpretation. In addition, I addressed a question of whether poeticality is explained only by richness of interpretation or it is affected by other metaphor properties, using six scales which were shown to be distinguishing between literary and journalistic metaphors by Steen (1994). In Experiment 1, I collected a list of features that participants thought were involved in the interpretation of metaphors. I then analyzed several characteristics of emergent features, especially the relation among emergent features, topic-vehicle similarity, and richness of interpretation. In Experiment 2, I collected poeticality ratings of the metaphors used in Experiment 1 as well as ratings of the six metaphor properties and analyzed how poeticality ratings are related to richness and other metaphor properties.

## METAPHOR APPRECIATION AND INCONGRUITY RESOLUTION

An *incongruity resolution model*, a general model of poetic appreciation, argues that poetic effect (or literality) is appreciated when a considerable incongruity is perceived in a verbal expression at the early stage of language comprehension and after that it is compensated by a rich interpretation (Utsumi, 2002). The underlying cognitive mechanism is that the perceived incongruity causes great processing effort and the interpretation process justifies such expended effort by yielding a rich interpretation consisting of diffuse meanings.

According to the incongruity resolution model, metaphor appreciation may proceed as follows: First, people perceive an incongruity in a metaphor that may be primarily caused by semantic dissimilarity between the topic and the vehicle, especially when a metaphor is presented without any extrasentential context. To receive a reasonable payoff for the efforts, people then try to search for a rich interpretation of the metaphor. If a rich interpretation is obtained, poetic or aesthetic value of the metaphor is appreciated. Poetic or novel metaphors may be often appreciated in this way. In contrast, conventional or familiar metaphors are less likely to be poetically appreciated, because such metaphors involve little incongruity and thereby

comprehending them requires no special processing effort. (This is empirically supported by the time-course studies of conventional or familiar metaphor; see, e.g., Blasko & Connine, 1993; Giora, 2003; Glucksberg, Gildea, & Bookin, 1982). Furthermore, when an incongruity is perceived but cannot be compensated at all, as in the case of anomalous sentences or incomprehensible metaphors like “*Life is a wooden desk,*” no poetic value may be appreciated.

One question that arises here is how emergence of features is related to the process of metaphor appreciation described previously. A reasonable answer would be that emergence of features may primarily bring about a rich interpretation; in other words, a richer interpretation may include more emergent features. This assumption is motivated by the previous research (Nueckles & Janetzko, 1997) showing that the proportion of emergent features people generated for the interpretation of metaphors is largest but they involve many different features. If this assumption is correct, the incongruity resolution model makes the following predictions:

- Prediction 1:* More emergent features are generated for a metaphor with less similarity between the topic and the vehicle to be comprehended coherently.
- Prediction 2:* An interpretation involving more emergent features is richer.
- Prediction 3:* A metaphor is appreciated poetically to the extent that its interpretation is rich.

My study examined the validity of these predictions on metaphor appreciation. In Experiment 1, I tested the first two predictions by assessing topic–vehicle similarity using the generated lists of features for the topic and for the vehicle. I also devised a measure of richness of interpretation by employing Shannon’s entropy in information theory (Shannon, 1948). In Experiment 2, I tested the last prediction about the relation between richness of interpretation and poeticity.

The incongruity resolution model, especially incongruity, has a great deal to do with the notion of “defamiliarization” or “foregrounding” in literary studies. *Defamiliarization* is the notion that was first employed in the works of Russian Formalists. According to Shklovsky (1965), the function of literature or art is to make people aware of the world in a creative and fresh way by the device of defamiliarization such as “to make objects ‘unfamiliar’, to make forms difficult, to increase the difficulty and length of perception” (p.12). As a result, “A work is created ‘artistically’ so that its perception is impeded and the greatest possible effect is produced through the slowness of the perception” (p. 12). *Foregrounding*, a notion similar to defamiliarization and with the roots in Prague Structuralism, is to highlight or to make prominent linguistic features within the literary text against the background of the rest of the text (Wales, 2001). It can be achieved by deviations from the expected or ordinary use of language including poetic metaphors and other figures of speech. These views of literary or aesthetic function were empiri-

cally examined by a pioneering work by van Peer (1986) and they have been passed on to, and developed by, recent studies of cognitive poetics (e.g., Gavins & Steen, 2003; Semino & Culpeper, 2002; Stockwell, 2003; Tsur, 1992).

The incongruity resolution model is most similar to the relevance–theoretic view of poetic function of language (Pilkington, 2000; Sperber & Wilson, 1995). Sperber and Wilson (1995) state, “Let us give the name *poetic effect* to the peculiar effect of an utterance that achieves most of its relevance through a wide array of weak implicatures” (p. 222). In relevance theory, utterance comprehension is seen as a process of optimizing relevance, that is, achieving the best possible balance of the cost of an utterance (i.e., processing effort) against its cognitive effect. Hence, when an utterance is foregrounded or defamiliarized by a deviant use of language which causes considerable processing effort, that utterance tends to be interpreted so as to achieve cognitive effect enough to offset the effort by a wide range of weak implicatures, resulting in evocation of poetic effect.

Miall and Kuiken’s (1999) empirical finding on literariness is also compatible with the view of incongruity resolution. They demonstrated that “literariness is constituted when stylistic or narrative variations defamiliarize conventionally understood referents and prompt reinterpretive transformations of a conventional feeling or concept” (p.123).

The notion of incongruity resolution is extensively accepted as a core component of humor appreciation (e.g., Attardo, 1997; Suls, 1972; Wyer and Collins, 1992; for the compatibility of relevance theory with the incongruity resolution model of humor, see Yus, 2003). According to the incongruity resolution view of humor, humor involves an incongruity between what was expected based on our conceptual pattern and what occurred in the humorous event, which is often expressed by a punch line in humorous texts. When such incongruity is resolved immediately by generating a reinterpretation of a humorous expression, humorous effect takes place. The optimal innovation hypothesis of pleasurability (Giora, 2003; Giora et al., 2004) is also associated with the incongruity resolution model. The optimally innovative stimulus, which would receive the highest pleasure rating according to the hypothesis, is a stimulus that induces a novel response but at the same time allows for an automatic recovery of a salient response so that an incongruity between them is recognized and resolved.

## EXPERIMENT 1

The main purpose of Experiment 1 is to examine the validity of Predictions 1 and 2 of the incongruity resolution model.

In addition, Experiment 1 serves to test a possibility that previous empirical findings on feature emergence showed only an illusion of importance of emergent features, although such an issue is not directly concerned with metaphor appreciation. Previous studies argued that emergent features play a central role in metaphor

comprehension primarily from the finding that they shared the largest proportion of listed features for all the metaphor materials. However, there is still a possibility that many irrelevant features, most of which might be emergent, were generated due to a limited experimental method for collecting features; in these studies participants were asked to express their own interpretations by simple phrases or words. In Experiment 1, therefore, participants were asked not only to list the features that they thought were involved in the interpretation of metaphors by words or phrases (feature listing task) but also to describe their own interpretations freely by sentences (free description task). If the results obtained by both tasks are consistent, it can be concluded that the importance of feature emergence is not an illusion created by an experimental method.

## Method

*Participants.* Eighty undergraduate students, in the 1st and 2nd year of arts and social sciences department at Japan Women's University, participated for a requirement of computer literacy course. All participants were native speakers of Japanese. None of them were familiar with metaphor research prior to the experiment.

*Materials.* Ten groups of four Japanese metaphorical sentences per group were constructed. The four metaphors in each group were created from all possible pairings of two topic words with two vehicle words. For example, from the two topics, "anger" ("ikari") and "sleep" ("nemuri"), and the two vehicles, "sea" ("umi") and "storm" ("arashi"), the following four metaphors were created: "Anger is the sea" ("Ikari ha umi da"), "Anger is a storm" ("Ikari ha arashi da"), "Sleep is the sea" ("Nemuri ha umi da") and "Sleep is a storm" ("Nemuri ha arashi da"). (In this article, the original Japanese materials used in the experiments are shown in parentheses preceded by their literal English translations.) The complete list of 10 metaphor groups is provided in the Appendix. Topic and vehicle words were selected from an experimental study on Japanese metaphor (Kusumi, 1987) or from a list of words frequently used for Japanese metaphors (Nakamura, 1995). This method of creating metaphor materials was identical to the method used by Becker (1997).

Each participant was assigned to one word (e.g., "sleep") and one metaphor (e.g., "Anger is the sea") which did not accompany the assigned word from each of the 10 metaphor groups. In other words, for example, when a participant was assigned to the word "sleep", that person did not see the "sleep" metaphors, either "Sleep is the sea" or "Sleep is a storm". As a result of this assignment, all participants each were assigned to 10 words and 10 metaphors. The order of the 10 words and the order of the 10 metaphors were randomly determined for each participant. Words and metaphors of each group were counterbalanced so that they were assigned to 20 participants.

*Procedure.* The experiment was carried out on a PC computer using a Common Gateway Interface (CGI) program I wrote for this experiment. All participants were familiar with the operation of a Web browser because this experiment was conducted in the last lesson of the computer literacy course. Participants, seated in front of a computer display, accessed the CGI page by a Web browser (Internet Explorer) and entered their user ID (student number) and password and began the experiment. The CGI program consisted of 25 pages: the first page explaining a general instruction of this program and the experiment, one page for the instruction on the feature listing task for single words with an example, 10 pages for the feature listing task of 10 single words, one page showing the end of that task, one page for the instruction on the metaphor interpretation task with an example, 10 pages for the interpretation task of 10 metaphorical sentences, and the last page showing the end of the experiment, in this order. These CGI pages were programmed so that participants could not move on the next page unless completing the task of this page.

Participants first carried out the feature listing task for 10 single words, followed by the metaphor interpretation task for 10 metaphors. In the feature listing task for single words, participants were asked to list at least three features or characteristics of (a concept expressed by) a single word displayed on the top of the screen and to rate the typicality of each listed feature. The ratings were made on a 3-point scale ranging from 1 (*slightly typical*) through 2 (*quite typical*) to 3 (*extremely typical*). The metaphor interpretation task consists of three subtasks: feature listing task, free description task, and comprehensibility rating task. In the feature listing task, participants were asked to consider the meaning of a metaphor displayed at the top of the screen, to list at least three features of the topic (presented with an underline) that were being described by the vehicle of that metaphor, and to rate the typicality of each listed feature on the same 3-point scale as used for single words. In the free description task, they were asked to describe their own interpretation of the metaphor freely by sentences. In the comprehensibility rating task, they were asked to rate the metaphor with respect to ease of interpretation on a 7-point scale ranging from 1 (*not at all comprehensible*) through 4 (*not sure whether comprehensible*) to 7 (*extremely comprehensible*).

## Results and Discussion

Before classifying the listed features of metaphorical interpretations, I conducted the following preprocessing for each metaphor and each word. For the features listed in the feature listing task for single words and for metaphors, closely related words or phrases were accepted as the same feature if they met any of the following four criteria: (a) they belonged to the same deepest category of a Japanese thesaurus *Bunrui Goi Hyo* (National Institute for Japanese Language, 1964; e.g., “*kakasenai*” and “*hitsuyoufukaketsu*”, both of which mean being indispensable), (b) a dictionary description of one word included the other word or phrase (e.g.,

“lie” [“*uso*”] and “not true” [“*sinjitsu deha nai*”]), (c) they shared the same root form (e.g., “red” [“*akai*”] and “redness” [“*akasa*”]), and (d) they differed only in degree because of an intensive modifier (e.g., “frightened” [“*kowai*”] and “quite frightened” [“*totemo kowai*”]). Note that the first two criteria mean that two words are synonyms, alternative spellings or abbreviations. If two words or phrases met none of the criteria, they were regarded as expressing different features even though they seemed to intuitively share the same meaning. After this feature combination process, any feature that was mentioned by only one participant was eliminated from the list of features. The amended list of features by that preprocessing was used for the analysis.

The sentences written in the free description task, before being subjected to the preprocessing for feature combination just described, were divided into words or phrases in the following way. First, irrelevant sentences or phrases by which participants expressed their own internal process of interpretation (e.g., “I could not at all understand the meaning of that sentence,” “... by considering the sentence [X is Y] as [X is like Y] I understood that ...”) were eliminated. After that, each sentence was divided (as short as possible) into words and phrases that made sense. Vehicle words, topic words, and function words (e.g., conjunctions, particles) were eliminated from the generated list of words and phrases. For example, a description “Love is eternal as if a journey lasts forever” was fragmented into two phrases “being eternal” and “lasts forever” by this process.

After these preprocessings, the features listed for metaphorical interpretations were classified into four categories: Features were classified as shared features when they were also listed for both the topic and the vehicle, as vehicle features when listed only for the vehicle, as topic features when listed only for the topic, or as emergent features when listed for neither the topic nor the vehicle.

*Proportion of feature categories.* For each metaphor, frequency of each feature category was counted separately for the listing task condition and the description task condition. Each feature was counted as a type (i.e., counted only once no matter how many participants listed it) and as tokens (i.e., counted as many times as it is listed by different participants), as done in the previous studies. Moreover, the feature listed in the feature listing task was counted as graded tokens, that is, it is counted as many times as the sum of typicality ratings of the participants who listed it. For example, if a feature was listed by 2 participants whose typicality ratings were 3 and 2, it would be counted once as types, two times as tokens, or five times as graded tokens. Table 1 shows the percentage of four feature categories in the interpretations of the 40 metaphors and the mean number of four feature categories per metaphor.

In the feature listing task condition, regardless of counting method, emergent features had the highest percentage and shared features had the lowest. Although the differences between topic and vehicle features were relatively small, the percentage of topic features was higher than that of vehicle features in all the counting

TABLE 1  
Proportions and Mean Numbers of Four Feature Categories in the  
Interpretation of Metaphors

<i>Counting Method</i>	<i>Feature Category</i>			
	<i>Shared</i>	<i>Vehicle</i>	<i>Topic</i>	<i>Emergent</i>
Listing task				
Types				
Percentage (%)	7.71	21.34	25.71	45.24
Mean number	0.75	2.08	2.50	4.40
Tokens				
Percentage (%)	15.92	22.00	26.56	35.53
Mean number	5.50	7.60	9.18	12.28
Graded tokens				
Percentage (%)	17.41	20.59	28.10	33.90
Mean number	12.63	14.93	20.38	24.58
Description task				
Types				
Percentage (%)	8.80	18.00	18.40	54.80
Mean number	0.55	1.13	1.15	3.43
Tokens				
Percentage (%)	13.75	19.98	17.72	48.55
Mean number	3.20	4.65	4.13	11.30

methods. Analyses of variance (ANOVAs) of mean numbers were performed with feature category as the only within-subject factor for all the counting methods.

Each analysis was conducted with metaphor treated as a random factor. The analyses revealed significant differences across the four categories in all the counting methods,  $F(3, 117) = 29.87, p < .001$  for types;  $F(3, 117) = 5.78, p < .01$  for tokens; and  $F(3, 117) = 4.33, p < .01$  for graded tokens. Post hoc Tukey's Honestly Significant Difference (HSD) pairwise comparisons ( $p < .05$ ) indicated that significantly more emergent features were produced than shared and vehicle features in all the counting methods, and than topic features in the types analysis. These findings are almost consistent with the previous findings on emergent features (Becker, 1997; Gineste et al., 2000; Nueckles & Janetzko, 1997; Tourangeau & Rips, 1991).<sup>1</sup>

<sup>1</sup>One salient exception to this consistency is Becker's (1997) finding that interpretations of metaphor contained larger numbers of vehicle features than other categories of features. She argued that this finding favored the attributive category theory (Glucksberg, 2001; Glucksberg & Keysar, 1990), which predicts that more vehicle features may be generated than emergent features. However, this finding seems less convincing given that participants of her experiment were instructed to interpret metaphor materials by listing the features of the vehicle that were being used to describe the topic. This instruction may possibly induce an unnatural strategy for metaphor comprehension, because metaphor is an expression that describes the topic in terms of the vehicle.



More important is that the description task condition yielded the same result: emergent features had the highest proportion and shared features had the lowest. ANOVAs of mean numbers revealed significant differences across the four categories both in the types analysis,  $F(3, 117) = 42.37, p < .001$ , and in the tokens analysis,  $F(3, 117) = 16.73, p < .001$ . Tukey's HSD pairwise comparisons ( $p < .05$ ) also revealed that significantly more emergent features were produced than all the other categories of features. This result means that emergent features are not an undesirable by-product of the limited experimental method but that they are really relevant to metaphor interpretation. Moreover the proportion of emergent features produced in the description task condition was significantly higher than that in the listing task condition,  $\chi^2(3, N = 639) = 7.46, p = .059$  for types;  $\chi^2(3, N = 2,313) = 44.82, p < .001$  for tokens. It may suggest a possibility that the feature listing task of this study, as well as previous research, may underestimate feature emergence.

*Correlation of comprehensibility with feature categories.* Table 2 (the second and third columns) shows correlations of comprehensibility with the number and the proportion of features in the four categories. For emergent features, neither the number nor the proportion was significantly correlated with comprehensibility. Furthermore, the same result was obtained for shared features. These findings suggest that people's judgment on metaphor comprehensibility was not affected by how many shared or emergent features were generated. Perhaps a surprising finding is that comprehensibility was negatively correlated with the number and the proportion of vehicle features; although not so high, four out of the eight correlations were significant. This finding seems to imply that people are likely to transfer salient features of the vehicle into the topic when interpreting less comprehensible metaphors.

*Correlation of similarity with feature categories.* The semantic similarity between the topic and the vehicle of a metaphor was determined so that the number of tokens of the features common to both the topic and the vehicle was divided by the total number of tokens listed in the vehicle or the topic. Note that the feature lists of the topic and of the vehicle used in calculation of similarity were generated in the feature listing task of single words. Therefore, these values represent to what degree these two concepts are similar when they are not presented together in the metaphor form.

Concerning emergent features, Table 2 (the fourth and the fifth columns) shows that, although some correlations between the proportion of emergent features and topic-vehicle similarity were significant, no clear evidence was found in favor of Prediction 1 that metaphors with lower topic-vehicle similarity provoke more emergent features.

There are at least two possible reasons for the result. First, anomalous or less comprehensible metaphors, which cannot be interpreted appropriately, may not provoke emergent features even though their topic-vehicle similarity is very low.

TABLE 2  
Correlations of the Number and the Proportion of Features in Four  
Categories with Comprehensibility, Similarity, and Interpretive Diversity

Counting Method/Task	Comprehensibility		Similarity		Diversity <sup>a</sup>	
	Number	Percentage	Number	Percentage	Number	Percentage
Shared						
Type/listing	.22	.17	.89***	.84***	-.28	-.37*
Token/listing	.24	.24	.92***	.92***	-.36*	-.38*
Type/description	.24	.26	.83***	.75***	.18	-.07
Token/description	.19	.22	.73***	.73***	.16	.01
Vehicle						
Type/listing	-.26	-.31*	-.43**	-.37*	.28	-.10
Token/listing	-.29	-.35*	-.35*	-.39*	.03	-.08
Type/description	-.23	-.24	-.20	-.24	.36*	-.00
Token/description	-.31*	-.33*	-.24	-.26	.17	-.13
Topic						
Type/listing	-.08	-.11	-.15	-.12	.15	-.19
Token/listing	.04	.00	-.09	-.17	-.04	-.09
Type/description	.10	.14	.19	.06	.29	-.11
Token/description	.01	.05	.11	.03	.23	.05
Emergent						
Type/listing	.07	.20	-.20	-.17	.74***	.47**
Token/listing	.03	.05	-.29	-.40*	.58***	.52***
Type/description	-.15	-.11	-.16	-.31	.65***	.14
Token/description	.05	.08	-.16	-.34*	.42**	.06

*Note.* Correlations in the graded tokens analysis are omitted for the sake of simplicity. These values did not substantially differ from the correlations in the tokens analysis.

<sup>a</sup>Interpretive diversity was calculated separately for the listing task condition and for the description task condition.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

Second, topic–vehicle similarity of this study may not be a proper measure of incongruity involved in metaphors; for example, high conventionality or familiarity of metaphors may reduce processing effort even when their topic–vehicle similarity is low. To test the first possibility that Prediction 1 would hold true of comprehensible metaphors, I divided the 40 metaphors into two groups: comprehensible metaphors (i.e., those whose comprehensibility rating had an average of 4.31 or higher for all the metaphors) and less comprehensible metaphors (i.e., those whose comprehensibility rating was less than the average). Table 3 lists correlations between similarity and the amount of emergent features for comprehensible and less comprehensible metaphors. As would be expected, most of the correlations for comprehensible metaphors were higher than those for less comprehensible metaphors and those for all metaphors. This result lends a little support to the plausibility of the modified version of Prediction 1 but cannot rule out the possibility of the second reason.

TABLE 3  
Correlations Between Topic–Vehicle Similarity and the Number and Proportion of Emergent Features for Comprehensible and Less Comprehensible Metaphors

<i>Counting Method/Task</i>	<i>Comprehensible Metaphors<sup>a</sup></i>		<i>Less Comprehensible Metaphors<sup>b</sup></i>	
	<i>Number</i>	<i>Percentage</i>	<i>Number</i>	<i>Percentage</i>
Type/listing	–.18	–.24	–.32	–.27
Token/listing	–.28	–.56*	–.33	–.33
Type/description	–.18	–.35	–.03	–.28
Token/description	–.22	–.48*	–.07	–.31

<sup>a</sup>( $n = 19$ ). <sup>b</sup>( $n = 21$ ).

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

An unsurprising finding shown in Table 2 was that the number and the proportion of shared features were positively correlated with topic–vehicle similarity. However, significant negative correlations were found between topic–vehicle similarity and the number, and the proportion, of vehicle features in the listing task condition. Given that the correlation between topic–vehicle similarity and comprehensibility was positive and significant ( $r = .33$ ,  $p < .05$ ), this result may also suggest that more vehicle features are generated when less comprehensible metaphors are interpreted.<sup>2</sup>

*Correlation of richness with feature categories.* Richness of interpretation of a metaphor was calculated using a concept of Shannon’s entropy in information theory (Shannon, 1948). Shannon’s entropy  $H(X)$ , defined by the following equation, expresses the average uncertainty about the realization of a single random variable  $X$ .

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (1)$$

A higher value of entropy means that probability  $p(x)$  is more uniformly distributed. The entropy is maximal,  $H(X) = \log_2 N$ , with  $N$  the number of states in  $X$ , when all states have the same probability  $1/N$ , that is, when the probability distri-

<sup>2</sup>Experiment 1 examining only metaphor products cannot make any claims about whether transferring, or carrying over, salient features of the vehicle into the topic is a process of metaphor comprehension which differs from the process based on shared and emergent features. There is, however, an undesirable possibility that this result is due to the experimental setting. Participants might think they had to write something as interpretation even when they could not interpret a given metaphor coherently. This possibility may be more serious for the listing task condition than for the description task condition, because the listing task forced participants to list at least three features. Indeed, as shown in Table 2, most of the significant correlations for vehicle features were observed in the data of the listing task condition.

bution  $p(x)$  of  $X$  is homogeneous. The entropy is minimal,  $H(X) = 0$ , when one state has probability 1 and all others have probability 0. Therefore, if we regard the distribution of the listed features for the interpretation of a metaphor as the probability distribution  $p(x)$ , the entropy becomes a reasonable measure of richness of that interpretation.

To measure the richness of interpretation by Equation 1, I assessed  $p(x)$  of each feature  $x$  listed for the interpretation  $X$  of a metaphor by dividing the number of tokens of that feature (i.e., the number of participants who listed that feature) by the total number of tokens involved in that interpretation. For example, when three features “red”, “round” and “fresh” are listed for the metaphor “*Her cheeks are apples*” by 6, 6, and 4 participants, the entropy of such interpretation is calculated as  $-(6/16) \log(6/16) - (6/16) \log(6/16) - (4/16) \log(4/16) = 1.56$ . However, if the feature ‘red’ is listed by 12 participants and the other two features each are listed by 2 participants, its value decreases to 1.06. In the rest of this article, I refer to these values for richness by the term *interpretive diversity*.

The last two columns of Table 2 show correlations between interpretive diversity and the number and proportion of features of four categories. The substantial positive correlations with the number and the proportion of emergent features, together with the lack of correlation with topic and vehicle features, indicate that emergent features primarily constitute a rich interpretation. Significant negative correlations with the number and the proportions of shared features may suggest that an interpretation dominated by shared features is unlikely to be rich.<sup>3</sup> These results fully support Prediction 2 that more emergent features make an interpretation richer. Note that these results were replicated when comprehensible and less comprehensible metaphors were analyzed separately.

*Mean number of tokens for feature categories.* To further attest to the validity of Prediction 2, I calculated the mean number of tokens per feature (i.e., how many, on average, participants listed that feature) for each feature categories. The result was that a shared feature, a vehicle feature, a topic feature and an emergent feature were listed by 7.33, 3.66, 3.67, and 2.79 participants on average in the listing task condition and by 5.82, 4.13, 3.59, and 3.30 participants on average in the description task condition. One-way ANOVAs based on the mean token counts for only the metaphors for which all feature categories were present revealed that the difference between four categories was significant in the listing task condition,  $F(3, 39) = 13.52, p < .001$ , although not significant in the description task condition,  $F(3, 18) = 1.89, p > .1$ . These results show that a large number of different types of emergent features were judged to characterize the interpretation but agree-

---

<sup>3</sup>Comprehensibility was not correlated with either interpretive diversity,  $r = -.05$  for the listing task condition,  $r = -.08$  for the description task condition. Topic-vehicle similarity was not correlated with either interpretive diversity as well,  $r = -.28$  ( $p = .075$ ) for the listing task condition,  $r = .12$  for the description task condition.

ment on emergent features among participants was low, thus providing additional evidence in favor of Prediction 2.

## EXPERIMENT 2

The findings of Experiment 1 provided evidence in favor of Predictions 1 and 2 about metaphor appreciation. What we want to know next is the validity of Prediction 3 that a metaphor with a richer interpretation is perceived as more poetic. It is the main purpose of Experiment 2.

### Method

*Participants.* Forty-two undergraduate students, in the 1st, 2nd, and 3rd year of arts and social sciences department at Japan Women's University, participated for a requirement of computer literacy course. All participants were native speakers of Japanese. None of them participated in Experiment 1 and were familiar with metaphor research prior to the experiment.

*Materials.* The 40 metaphors used in Experiment 1 were used for this experiment. Each participant was assigned to two metaphors which share neither the topic nor the vehicle (i.e., "Anger is the sea" and "Sleep is a storm") from each of the 10 metaphor groups and thereby a total of 20 metaphors. The order of these materials was randomly determined for each participant. Metaphors of each group were counterbalanced across participants so that they were assigned to 21 participants.

*Procedure.* The experiment was carried out on a PC computer using a CGI program in the same way as Experiment 1. The CGI program used in Experiment 2 consisted of 23 pages: the first page explaining a general instruction of this program and the experiment, one page for the instruction on the rating task with an example, 20 pages for the rating task of 20 metaphors, and the last page showing the end of the experiment, in this order. Other experimental settings were identical to those of Experiment 1.

Participants were asked to read each metaphorical sentence, to consider its meaning and to rate it on a 7-point scale of poeticality ranging from 1 (*nonpoetic*) through 4 (*not sure whether poetic or nonpoetic*) to 7 (*poetic*). They were given the instruction that they can easily evaluate poeticality by regarding this scale as the degree of poetic effects evoked by that metaphor when used in poems or literary works. Furthermore, participants were asked to rate each metaphor on the following six 7-point scales: appropriateness ranging from 1 (*far-fetched*) to 7 (*appropriate*), beauty ranging from 1 (*ugly*) to 7 (*beautiful*), formality ranging from 1 (*informal*) to 7 (*formal*), political ranging from 1 (*apolitical*) to 7 (*political*), tastefulness

ranging from 1 (*tasteless*) to 7 (*tasteful*), and preciseness ranging from 1 (*vague*) to 7 (*precise*). These scales were selected from Steen's (1994) study on distinctive properties for distinguishing between literary metaphors and journalistic (i.e., non-literary) metaphors. In his study, literary and journalistic metaphors were rated on 22 scales and the obtained data were entered into a principal component analysis. The analysis revealed four factors: conceptual difficulty, emotive value, communicative manners and moral position, the first two of which were shown to be more distinctive properties. Appropriateness and preciseness scales had the two highest loadings on "conceptual difficulty" factor and beauty and tastefulness scales had the two highest loadings on "emotive value" factor. The other two scales, formality and political, had the highest loading on "communicative manners" and "moral position" factors, respectively.

## Results and Discussion

*Relation between poeticity and richness of interpretation.* Table 4 shows correlations between interpretive diversity in Experiment 1 and mean poeticity rating. When all metaphors were analyzed, although the interpretive diversity of the description task condition was positively correlated with poeticity, the correlation between the interpretive diversity of the listing task condition and poeticity was below the level of significance. A clear difference of correlation was observed between comprehensible metaphors and less comprehensible metaphors. For comprehensible metaphors, the interpretive diversity of both conditions was positively correlated with poeticity, thus supporting Prediction 3 that metaphors are poetically appreciated to the extent that their interpretations are rich. However, for less comprehensible metaphors, there were no correlations between interpretive diversity and poeticity, which is incompatible to Prediction 3.

One possible explanation of this incompatible result for less comprehensible metaphors would be that, as I described previously, incomprehensible or anomalous metaphors cannot be poetically appreciated, because yielded interpretations are not semantically coherent even though they are highly rich. However, this explanation is not correct; less comprehensible metaphors ( $M = 4.78$ ) were rated as

TABLE 4  
Correlations Between Poeticity and Interpretive Diversity for  
Comprehensible and Less Comprehensible Metaphors

	<i>All Metaphors</i> <sup>a</sup>	<i>Comprehensible</i> <sup>b</sup>	<i>Less Comprehensible</i> <sup>c</sup>
Diversity			
Listing	.13	.49*	-.20
Description	.34*	.46*	.16

<sup>a</sup>( $n = 40$ ). <sup>b</sup>( $n = 19$ ). <sup>c</sup>( $n = 21$ ).

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

TABLE 5  
Results of Principal Component Analysis of Interpretive Diversity and Six  
Metaphor Properties (Loadings < 0.40 Are Omitted)

Scale	Factor for Listing Task				Factor for Description Task			
	1	2	3	4	1	2	3	4
Appropriate–far-fetched	—	0.90	—	—	—	0.91	—	—
Beautiful–ugly	–0.97	—	—	—	–0.97	—	—	—
Formal–informal	–0.49	—	–0.79	—	–0.50	—	–0.76	—
Political–apolitical	—	0.52	–0.79	—	—	0.50	–0.81	—
Tasteful–tasteless	–0.92	—	—	—	–0.93	—	—	—
Precise–vague	—	–0.92	—	—	—	0.90	—	—
Diversity	—	—	—	1.00	—	—	—	0.99
Variance	2.19	2.07	1.30	1.02	2.20	2.06	1.29	1.05
Percentage explained variance	31.3	29.6	18.5	14.6	31.4	29.4	18.5	14.9

more, rather than less, poetic than comprehensible metaphors ( $M = 4.22$ ),  $t(38) = 2.02$ ,  $p < .05$ . This result suggests that less comprehensible metaphors could be poetically appreciated regardless of richness of interpretation. Then, what properties of less comprehensible metaphors affect poeticity ratings? Is richness of interpretation really a good predictor of metaphor poeticity? I address these questions in the following analysis.

**Regression analysis.** I conducted a regression analysis with poeticity rating as the dependent variable. Independent variables were factor scores from a principal component analysis with Varimax rotation conducted on the ratings of six distinctive properties and interpretive diversity. Principal component analysis was conducted twice, because interpretive diversity was obtained separately for the listing task condition and the description task condition.

The two principal component analyses yielded almost the same result and extracted four factors whose initial eigenvalue was greater than 1. These factors accounted for 94.1% of the total variance of the data for the listing task condition and 94.2% of the total variance of the data for the description task condition. Table 5 shows the factor loadings on each of the four factors. An important result is that interpretive diversity was related only to Factor 4 and Factor 4 had a high loading only from interpretive diversity. It means that interpretive diversity, and thus richness of interpretation, was an independent metaphor property which was not related to other metaphor properties.<sup>4</sup> The other three factors (i.e., Factors 1–3) are almost compatible with Steen's (1994) analysis. Factor 1, with high negative load-

<sup>4</sup>It must be noted that none of the six metaphor properties were correlated with the number and the proportion of emergent features ( $p < .05$ ), meaning that emergent features did not have an influence on these properties.

ings from two scales of beauty and tastefulness, exactly corresponds to Steen's "emotive value" factor. I therefore use the same label for Factor 1. Factor 2 had high positive loadings from appropriateness and preciseness which characterized the factor of conceptual difficulty in Steen's study. This factor also had considerable loading from the political scale. Because political metaphors are often used for an easy-to-understand explanation, all of these features indicate that this factor captures ease of comprehension or aptness of metaphors. Therefore, Factor 2 can be labeled as "conceptual aptness". Factor 3 had high negative loadings from the two scales of formality and political, which seem to indicate an informal or unofficial statement. It is thus labeled as "informality."

The regressions using these four factors as independent variables were calculated not only for all metaphors but also for comprehensible and less comprehensible metaphors in each task condition and thus six regression equations were obtained. Table 6 shows the regression equations for the description task condition and Table 7

TABLE 6  
Summary of Regression Analysis Predicting Metaphor Poeticity by the Four Factors of the Description Task Condition

Variable	All Metaphors			Comprehensible			Less Comprehensible		
	B	SE B	$\beta$	B	SE B	$\beta$	B	SE B	$\beta$
Emotive value	-0.22	0.05	-.46***	-0.11	0.08	-.23	-0.30	0.06	-.78***
Conceptual aptness	-0.28	0.05	-.58***	-0.37	0.10	-.58**	-0.09	0.12	-.12
Informality	-0.14	0.06	-.22*	-0.15	0.08	-.27	-0.11	0.08	-.21
Richness	0.20	0.07	.29**	0.21	0.10	.31*	0.09	0.09	.15

Note.  $R^2 = .68$ ,  $F(4, 35) = 18.62$ ,  $p < .0001$  for all metaphors;  $R^2 = .71$ ,  $F(4, 14) = 8.61$ ,  $p = .0010$  for comprehensible metaphors;  $R^2 = .66$ ,  $F(4, 16) = 7.74$ ,  $p = .0011$  for less comprehensible metaphors.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

TABLE 7  
Summary of Regression Analysis Predicting Metaphor Poeticity by the Four Factors of the Listing Task Condition

Variable	All Metaphors			Comprehensible			Less Comprehensible		
	B	SE B	$\beta$	B	SE B	$\beta$	B	SE B	$\beta$
Emotive value	-0.22	0.05	-.47***	-0.09	0.08	-.18	-0.31	0.06	-.81***
Conceptual aptness	-0.28	0.05	-.58***	-0.37	0.10	-.59**	-0.09	0.12	-.13
Informality	-0.17	0.06	-.28**	-0.18	0.08	-.32	-0.14	0.08	-.24
Richness	0.10	0.07	.15	0.17	0.13	.22	0.06	0.08	-.13

Note.  $R^2 = .66$ ,  $F(4, 35) = 16.93$ ,  $p < .0001$  for all metaphors;  $R^2 = .70$ ,  $F(4, 14) = 8.08$ ,  $p = .0014$  for comprehensible metaphors;  $R^2 = .66$ ,  $F(4, 16) = 7.93$ ,  $p = .0010$  for less comprehensible metaphors.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .



shows the regression equations for the listing task condition. The overall regression equations were significant for all six regression analyses. Concerning the relation between richness of interpretation and poeticality ratings, almost the same result was obtained as in the simple correlation analysis. Richness of the description task condition alone accounted for a significant portion of the variance in poeticality of all metaphors and poeticality of comprehensible metaphors, although richness of the listing task condition was not correlated significantly with poeticality ratings.

The regression analysis also showed that poeticality rating could not be explained only by richness; emotive value and conceptual aptness accounted for more of the variance in poeticality when all metaphors were subjected to the regression analysis, as demonstrated by Steen (1994). These relations were negative so that emotively more negative (i.e., less beautiful and tasteful) and more apt (i.e., more precise and appropriate) metaphors were rated as less poetic. Informality was also related to poeticality but the amount of variance accounted for by informality was much less than by emotive value and conceptual aptness.

A more interesting finding is that a clear difference of regression equations was observed between comprehensible and less comprehensible metaphors. Although emotive value was not related to poeticality of comprehensible metaphors, conceptual aptness and richness accounted for a significant portion of the variance. In contrast, for less comprehensible metaphors only emotive value accounted for a significant portion of the variance in poeticality. This finding means that comprehensible metaphors are appreciated primarily on the basis of their semantic value (conceptual aptness and richness are both semantic properties) but that less comprehensible metaphors are appreciated on the basis of their emotive value, perhaps because coherent interpretations (and thus semantic values) cannot be obtained.

To sum up the findings, Prediction 3 of the incongruity resolution model is supported with respect to comprehensible metaphors; when the interpretation of comprehensible metaphors is rich, it is perceived as poetic. However, the incongruity resolution model has its limitation; metaphors are judged to be poetic by the effect of other properties, especially in the case of less comprehensible metaphors. In addition, conceptual aptness and emotive value, beyond the previous finding, simply showing that both properties affected metaphor appreciation (Steen, 1994), were found to play a differential role in metaphor appreciation according to metaphor comprehensibility.

## GENERAL DISCUSSION

It may be generally agreed that poetic appreciation of metaphor is deepened or facilitated by feature emergence and its consequent richness of interpretation but (at least to my knowledge) it is not always demonstrated empirically whether this is really the case. My study can be seen as providing empirical evidence that this argument is convincing for comprehensible metaphors. Furthermore, my study dem-

onstrated that emergent features are the principal source of richness and that in this respect feature emergence plays a substantial role in poetic appreciation of comprehensible metaphors, which has not been empirically demonstrated as well.

In the context of distinction between poetic (or expressive) and nonpoetic (or explanatory) metaphors, richness is often discussed with clarity of metaphor (Gentner, 1982; Stockwell, 2003). Gentner (1982) argued that expressive (often poetic) metaphors tend to be low in clarity but a high degree of richness, whereas explanatory (often scientific) metaphors tend not to be rich but are very clear. Given that conceptual aptness factor is a notion quite similar to clarity, it is obvious that my study may also be seen as providing empirical evidence for the argument in the case of comprehensible metaphors; these metaphors were found to be perceived as more poetic to the extent that their interpretations were richer and conceptually less apt.

My definition of richness as interpretive diversity differs from, and is intuitively more plausible than, Gentner's (1982) definition of richness. Gentner's definition of richness in her structure mapping theory of metaphor is based on the number of predicates that are mapped to the topic: "More precisely, the richness of an analogy is its predicate density: for a given set of nodes, the average number of predicates per node that can be plausibly mapped from base to target" (Gentner, 1982, p. 114). Therefore, for example, given that two metaphors *X* and *Y* include the equal number of features or predicates for their interpretations, they are judged to be equally rich regardless of how these features differ in strength or salience. However, imagine a situation where for the metaphor *X* one feature is much more salient or strong than all the other features, whereas where all the features for the metaphor *Y* are equally salient. In this case, it seems unreasonable to suppose that two metaphors are equally rich. The metaphor *X* may be less rich than the metaphor *Y*. My definition of richness by Equation 1 is based on both the number of features and the strength of each feature and, thereby, captures such intuitive difference in richness.

Other existing findings about metaphor appreciation, which addressed the relation between the process of metaphor appreciation and other processes for metaphor understanding, may be consistent with the findings on richness of interpretation obtained by this study. For example, Gibbs (2002) demonstrated that identification of poetic metaphors facilitated the appreciation of these metaphors so that participants were more aesthetically appreciative and felt emotionally more moved. One possible explanation of his finding, which is consistent with the incongruity resolution model, would be that, by consciously identifying expressions as metaphors, people might be more aware of the incongruity involved in these poetic metaphors and thus generate a richer interpretation to receive a reasonable payoff for the efforts. As a result, people who identified metaphors might appreciate them more and feel emotionally more moved than those who did not. Gerrig and Healy (1983) found that the order of metaphor and context within sentences affected reading times for metaphors (i.e., online comprehension) but did not affect their goodness ratings (i.e., appreciation). Although they interpreted this finding as suggesting that metaphor comprehension and appreciation are quite different processes, my findings suggest an alter-

native interpretation of their result; metaphor appreciation might not be affected by the presentation order, perhaps because degrees of the metaphor properties which were found to affect poetic appreciation in Experiment 2, such as richness, conceptual aptness, and emotive value, are obviously unchanged by the presentation order.

Another important contribution of my study to metaphor research is that it demonstrated that metaphor properties that affect poetic appreciation differ between comprehensible metaphors and less comprehensible metaphors; semantic properties affect appreciation of the former, although emotive properties affect appreciation of the latter. This finding has some implications about metaphor appreciation. First, the incongruity resolution model has its limitation in that it cannot explain all the ways in which poetic value is appreciated. To explore the mechanism of metaphor appreciation, we must not only modify or elaborate on the semantic-based model of incongruity resolution but also consider how people judge emotive values of metaphors. Second, semantic and emotive values of metaphor have a complementary, not additive, effect on metaphor appreciation. From this implication, we can consider at least two possible hypotheses for the mechanism of metaphor appreciation. One hypothesis is that metaphor appreciation starts with semantic-value-based processing, which may be closely connected to metaphor comprehension, and emotive-value-based processing is triggered only when the semantic-value-based processing does not work because a semantically appropriate interpretation cannot be derived as in the case of less comprehensible or anomalous metaphors. Another hypothesis is that metaphor appreciation starts with both semantic-value-based processing and emotive-value-based processing initially and emotive-value-based processing is suppressed later when semantic-value-based processing works properly. These hypotheses are, of course, speculative and must await further research but hopefully my empirical finding will open new research direction on metaphor appreciation.

No contemporary metaphor theories can provide a sufficient explanation of the findings of this article. For example, the conceptual metaphor theory (Lakoff & Turner, 1989) has argued that a large number of metaphors used in poetry and literature are derived from novel uses of conceptual metaphors, image metaphors, and generic-level metaphors. Although their analysis is important and has been extended toward cognitive studies of poetics (e.g., Gavins & Steen, 2003; Stockwell, 2003), these studies do not explain the online mechanism of how certain metaphors are perceived as poetic; they only explain how metaphors in literature fit within the conceptual metaphor theory. Gentner's structure mapping theory (Gentner, 1983; Gentner & Markman, 1997) and Glucksberg's attributive category theory (Glucksberg, 2001; Glucksberg & Keysar, 1990) or Gentner's career of metaphor hypothesis (Gentner, Boedle, Wolff, & Boronat, 2001), which reconcile both theories, do not say anything about how metaphor is appreciated as well, although, as I described previously, Gentner (1982) stated that richness and clarity are properties for distinguishing scientific and literary metaphors. Relevance theory fundamentally involves the mechanism of incongruity resolution and thus is

consistent with the finding that richness of interpretation facilitates poetic appreciation but tells nothing about how less comprehensible metaphors are poetically appreciated according to their emotive value.

Although it is not certain at this point whether comprehension and appreciation are different processes or both ends of one continuous process, I believe that any adequate theory of metaphor should take both comprehension and appreciation into account equally. Hence I hope that my findings would contribute to empirical research exploring the mechanism of how metaphor properties affect metaphor appreciation and to theoretical research developing a comprehensive metaphor theory.

### ACKNOWLEDGMENTS

This research was supported by Grant-in-Aid for Encouragement of Young Scientists (No.14780263), The Ministry of Education, Culture, Sports, Science, and Technology, and by a grant from Nissan Science Foundation.

### REFERENCES

- Attardo, S. (1997). The semantic foundations of cognitive theories of humor. *Humor, 10*, 395–420.
- Becker, A. (1997). Emergent and common features influence metaphor interpretation. *Metaphor and Symbol, 12*, 243–259.
- Black, M. (1962). Metaphor. In M. Black (Ed.), *Models and metaphors* (pp. 25–47). Ithaca, NY: Cornell University Press.
- Blasko, D., & Connine, C. (1993). Effects of familiarity and aptness on metaphor understanding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 295–308.
- Gavins, J., & Steen, G. (Eds.). (2003). *Cognitive poetics in practice*. London: Routledge.
- Gentner, D. (1982). Are scientific analogies metaphors? In D. Miall (Ed.), *Metaphor: Problems and perspectives* (pp. 106–132). Sussex, England: Harvester.
- Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.
- Gentner, D., Boedle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *Analogical mind: Perspectives from cognitive science* (pp. 199–253). Cambridge, MA: MIT Press.
- Gentner, D., & Markman, A. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45–56.
- Gerrig, R., & Healy, A. (1983). Dual processing in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 667–675.
- Gibbs, R. (2002). Identifying and appreciating poetic metaphor. *Journal of Literary Semantics, 31*, 101–112.
- Gineste, M., Indurkha, B., & Scart, V. (2000). Emergence of features in metaphor comprehension. *Metaphor and Symbol, 15*, 117–135.
- Giora, R. (2003). *On our mind: Salience, context, and figurative language*. New York: Oxford University Press.
- Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N., & Zur, A. (2004). Weapons of mass distraction: Optimal innovation and pleasure ratings. *Metaphor and Symbol, 19*, 115–141.

- Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. New York: Oxford University Press.
- Glucksberg, S., Gildea, P., & Bookin, H. (1982). On understanding nonliteral speech: Can people ignore metaphors. *Journal of Verbal Learning and Verbal Behavior*, 21, 85–98.
- Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, 97, 3–18.
- Indurkha, B. (1991). *Metaphor and cognition: An interactionist approach*. Dordrecht, The Netherlands: Kluwer.
- 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., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago: University of Chicago Press.
- Miall, D., & Kuiken, D. (1999). What is literariness? Three components of literary reading. *Discourse Processes*, 28, 121–138.
- Nakamura, A. (1995). *Hiyu hyogen jiten* [Japanese dictionary of metaphorical expressions]. Tokyo: Kadokawa Shoten.
- National Institute for Japanese Language (Ed.). (1964). *Bunrui goi hyo* [Word list by semantic principles]. Tokyo: Shuei Shuppan.
- Nueckles, M., & Janetko, D. (1997). The role of semantic similarity in the comprehension of metaphor. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 578–583). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Pilkington, A. (2000). *Poetic effects: A relevance theory perspective*. Amsterdam: Benjamins.
- Ricoeur, P. (1977). *The rule of metaphor: Multi-disciplinary studies of the creation of meaning in language* (R. Czerny, Trans.). Toronto, Canada: University of Toronto Press.
- Semino, E., & Culpeper, J. (2002). *Cognitive stylistics: Language and cognition in text analysis*. Amsterdam: Benjamins.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
- Shklovsky, V. (1965). Art as technique. In L. Lemon & M. Reis (Eds. & Trans.), *Russian formalist criticism: Four essays* (pp. 3–57). Lincoln, NE: University of Nebraska Press.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford, England: Blackwell.
- Steen, G. (1994). *Understanding metaphor in literature: An empirical approach*. Harlow, England: Longman.
- Stockwell, P. (2003). *Cognitive poetics: An introduction*. London: Routledge.
- Suls, J. (1972). Two-stage model for the appreciation of jokes and cartoons: Information-processing analysis. In J. Goldstein & P. McGhee (Eds.), *The psychology of humor* (pp. 81–100). New York: Academic.
- Tourangeau, R., & Rips, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory and Language*, 30, 452–472.
- Tourangeau, R., & Sternberg, R. (1982). Understanding and appreciating metaphors. *Cognition*, 11, 203–244.
- Tsur, R. (1992). *Toward a theory of cognitive poetics*. Amsterdam: Elsevier.
- Utsumi, A. (2002). Toward a cognitive model of poetic effects in figurative language. In *Proceedings of 2002 IEEE International Conference on Systems, Man and Cybernetics (SMC2002)*. Hammamet, Tunisia: IEEE.
- van Peer, W. (1986). *Stylistics and psychology: Investigations of foregrounding*. London: Croom Helm.
- Wales, K. (2001). *A dictionary of stylistics* (2nd ed.). Harlow, England: Addison-Wesley.
- Wyer, R., & Collins, J. (1992). A theory of humor elicitation. *Psychological Review*, 99, 663–688.
- Yus, F. (2003). Humor and the search for relevance. *Journal of Pragmatics*, 35, 1295–1331.

APPENDIX  
Stimuli Used in the Experiments

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