The Role of Semantic Similarity in the Comprehension of Metaphor

Matthias Nueckles and **Dietmar Janetzko**

University of Muenster
Psychological Institute III
Fliednerstrasse 21
D 48149 Muenster, Germany
nueckles@psy.uni-muenster.de

University of Freiburg
Institute of Computer Science and Social Research
Friedrichstrasse 50
D 79098 Freiburg, Germany
dietmar@cognition.iig.uni-freiburg.de

Abstract

According to the comparison view, preexisting similarities between the constituent terms of a metaphorical sentence are an important source of information for generating a figurative meaning. The interaction approach, by contrast, claims that similarity is not an antecedent but a product of comprehension. We shall argue, however, that each of these approaches is too narrow to provide a complete and exhaustive account of metaphor comprehension. Instead, both theories point out to two different but complementary cognitive processes. We present three experiments that support the theoretical distinction between *analysis*-based vs. *synthesis*-based processes in the comprehension of metaphor.

Introduction

What information do people use to generate metaphorical meanings? In recent years, several psychological models have been proposed to provide an answer to this question (Ortony, 1979, Malgady & Johnson, 1980; Tourangeau & Sternberg, 1982; Gentner, 1988; Glucksberg & Keysar, 1990). Despite this theoretical abundance, only few empirical studies have been carried out. Therefore, the very question which of the proposed comprehension models is correct and which have to be abandoned is still a matter of debate. The oldest and hitherto most influential approach to the comprehension of metaphor is provided by the *similarity* or *comparison* view (Ortony, 1979, Malgady & Johnson, 1980). According to this theory, metaphors are implicit comparisons. In order to generate an interpretation, a metaphorical sentence such as *'the FAMILY is a NEST'* first has to be translated into an explicit comparison statement: *'the FAMILY is* like *a NEST'*. Then, a feature-matching process is applied to the representations of the noun concepts involved in the metaphor. The features of the *vehicle*-concept NEST are compared to the features of the *tenor*-concept FAMILIY in order to identify common features. In the example mentioned, possible candidates for the interpretation may be features such as 'security', 'warmth', or 'well-being'. Usually, only few such matching features can be identified whereas most of the features are exclusively part either of the vehicle- or the tenor-concept (Malgady & Johnson, 1980). Comparison models of metaphor comprehension mainly differ over the additional constraints they require for the matching features to be used in the interpretation. From Gentner's systematicity principle, e.g., follows that the matching features have to be relations, i.e., higher-order predicates rather than object attributes (Gentner, 1988). In Ortony's salience imbalance model (Ortony, 1979) metaphorical comparisons are distinguished from literal comparion statements according to differential feature salience. In literal comparisons, the matching features of tenor and vehicle are equally salient. By contrast, in metaphorical comparisons, the matching features are highly salient for the vehicle but show only low salience for the tenor-concept. Regardless of the specific constraints placed on matching features by different comparison models, the same assumption is underlying these accounts. There are semantic similarities *antecedently* existing between the represention of the tenor- and vehicle-concept on which the interpretation is based (Indurkhya, 1992).

The position that the interpretation of metaphors is based on preexisting similarities has been been questioned by the so-called *interaction* approach (Black, 1979; Tourangeau & Sternberg, 1982; Indurkhya, 1992). As Tourangeau & Sternberg (1982) argue, often a feature apparently shared by tenor and vehicle can be shared only metaphorically. Consider the metaphor *'MAN is a WOLF.'* In the context of the vehicle-concept WOLF , a feature such as 'predator' refers to how wolfes behave toward other animals. However, when this feature is applied to the tenor it assumes a slightly different meaning. Here, 'predator' may be referred to men's beahvior toward other men. Hence, in order to generate an interpretation, the feature has to be reinstantiated in the semantic domain of the tenor. As a result of such a reinstantiation process, the meaning of the feature is transformed. One might infer, e.g., that men are 'insidious' and 'competitive' in their dealings with other men. Such features, however, are unlikely to be considered as part of the conventional representations of MAN and WOLF. Therefore, the interaction account predicts that metaphors are not interpreted on the grounds of shared features. Rather than this, metaphorical meanings are constructed by means of *emergent* features , i.e., features which emerge when the representations of tenor and vehicle as well as their corresponding domains are brought into *interaction* during comprehension. To put it in Black's words: "It would be more illuminating...to say that metaphor creates the similarity than to say

that it formulates some similarity antecendently existing" (Black, 1979, p. 37).

Though the comparison and interaction approach have stimulated interesting theoretical speculations, few empirical studies have been carried out to exmanine both approaches. Comparison theorists typically present correlational data as a support of their position. For example, Johnson & Malgady (1979) had subjects rate the semantic similarity of tenor-vehicle pairs taken from a sample of poetic metaphors. Independent ratings of aptness were collected for the same metaphors. Johnson & Malgady found high positive correlations between the similarity ratings and metaphor aptness, irrespective of whether the tenor- and vehicle-concepts had been rated in the context of the metaphor sentences or in isolation. As support for the interaction approach, very often a lexical decision experiment by Camac & Glucksberg (1984) is cited in the literature. In this experiment, no semantic priming was found for word pairs made up by the tenor and the vehicle of apt and comprehensible metaphors. From this negative result, Camac & Glucksberg concluded that the tenor and the vehicle need not be semantically similar. Consequently, if no such similarities exist, then they cannot be used either for the interpretation process. Hence, the comparison theory must be wrong. Rather than being a prerequisite for comprehension, similarity might be its product. There are certainly more empirical data than the studies just reported. However, taken all available evidence together, it is impossible to decide which of the two theoretical approaches should be accepted and which rejected.

A closer examination of these approaches reveals that both the comparison and interaction view cannot adequately meet the complexity of the topic. Even on a descriptive level, they fail to provide a systematic and complete account of the different *semantic effects* that occur when interpreting metaphors. While the comparison view is limited to the investigation of features shared by tenor and vehicle, the interaction view focusses on the explanation of emergent features. However, as Tourangeau & Rips (1991) have recently argued, it is reasonable to assume that there are more semantic effects in metaphor interpretation than only shared and emergent features. According to *Figure 1* there are at least four types of effects to be distinguished. In our schema, a particular semantic effect is operationally defined by the *category of semantic feature* (segments 1-4, cf. *Figure 1*) that contributes to the ground of interpretation (Tourangeau & Rips, 1991). For the following, 'ground' refers to the features that are central to the interpretation. Consider the example *'HERCULES is a LION'*. The interpretation of this metaphor may in part be based on features shared by the tenor and the vehicle (*convergent features, 1,* e.g., *living organism*). In addition, the ground may incorporate features that belong to the vehicle but not to the tenor (*transferred features, 2,* e.g., *brave*). And also features of the tenor may be used once more to characterize the tenor within the metaphorical sentence (*redundant features, 3,* e.g., *divine*). Finally, the ground of interpretation may consist of features that are neither characteristic of the tenor nor of the vehicle but surface only in the interpretation (*emergent features, 4,* e.g., *mythical figure*). The classification schema of *Figure 1* shows that metaphorical meanings can adequately be described as *patterns* to which all four types of semantic effects contribute. From this perspective, the controversy between the comparison and interaction approach appears as somewhat misleading. We should not ask, is comprehension either based on a feature matching process or rather on a construction of emergent features. But the crucial question is under which conditions do people primarily resort to a feature matching process to identify convergents and under which conditions do they preferably construct emergents. To give an answer to this question, we have set up three studies.

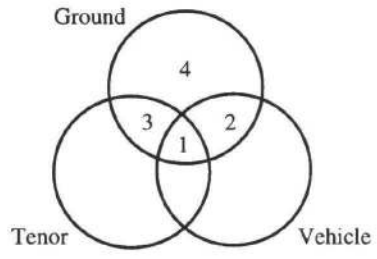

Figure 1: Semantic effects in metaphor interpretation

In *Study 1*, we will demonstrate that the grounds of metaphors can adequately be conceived of as patterns made up of different semantic effects. We will describe and analyze these patterns of effects and thereby elaborate on Tourangeau & Rips' results (1991). Another object of this study was to distinguish between two groups of metaphors according to the degree of semantic similarity between tenor and vehicle. This distinction between high similarity and low similarity metaphors will then be used to predict differences in reaction times in a lexical decision experiment (see *Study 3*). In *Study 2* we collected independent ratings of comprehensibility, propositional aptness and metaphoricity for the metaphor sample used in *Study 1*. This study was set up to extend and evaluate the results of the first study. In *Study 3* we tested the prediction that faster lexical decisions would result for tenor-vehicle pairs taken from high similarity metaphors than for tenor-vehicle pairs from low similarity metaphors. The aim of this experiment was to provide evidence against Camac & Glucksberg's conclusions. At the same time, if our distinction between high and low similarity metaphors was confirmed by such an independent test, this would enable us to shed new light on the controversy between the comparison and interaction approach.

Study 1: Feature Analysis of Metaphor Interpretations and Concept Descriptions

In this study, we examined interpretations of metaphors and compared them to independent descriptions of the tenor- and vehicle-concepts involved. Also, the concept of *semantic similarity* between the tenor and the vehicle will be introduced to empirically distinguish between high- and low-similarity metaphors.

Subjects, Materials and Procedure. 19 10th-grade highschool pupils (mean age = 16) participated in this study. All were native speakers of German. The study used 18 nominal metaphors of the form 'the *TENOR* is a *VEHICLE*'. The stimuli were selected from an item pool of

200 nominal metaphors. In a preliminary study, additional five subjects had been asked to produce nominal metaphors according to the instructions of a metaphor invention inventory, specifically designed for this purpose. The 18 metaphors selected for experimental investigation were perfectly interpretable even if no discourse context was provided to aid interpretation. This was established by collecting ratings of comprehensibility and propositional adequacy in another preliminary study. Besides these 18 nominal metaphors, also their tenors and vehicles served as stimuli. Subjects were asked to write extended interpretations of the selected 18 metaphors. They were also requested to produce descriptions of both the tenor and the vehicle concept of each metaphor. Each idea had to be expressed in a simple sentence or phrase. Metaphors and concepts were presented in separate blocks in a booklet so that the metaphors did not accompany the tenors or vehicles in the concept description task. Within the block of metaphors and the block of concepts, the order of items was randomized for each subject.

Results. Metaphor interpretations and concept descriptions were subjected to a content analysis to determine the proportions of different categories of semantic features in the interpretation. To this purpose, we compared the concept descriptions with the interpretations, classifying each feature in the interpretation on whether it characterized the tenor of the metaphor (*redundant* feature), the vehicle (*transferred* feature), both (*convergent* feature), or neither (*emergent* feature). We also calculated an estimate of the semantic similarity of the tenor and the vehicle for each metaphor (see below). *Figure 2* presents the proportions of the four categories of semantic features in the interpretations of the 18 metaphors. The

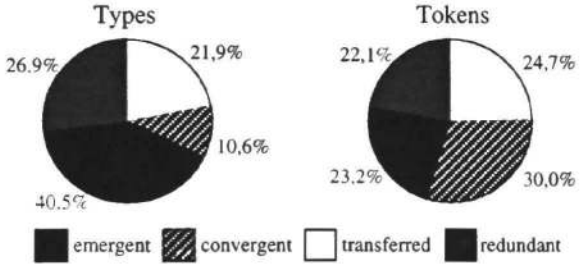

Figure 2: Proportions of categories of semantic features in the interpretations

results are presented in two different ways, with each feature counted once regardless of the number of subjects who listed it as part of their interpretation *(types)* and with each feature counted as many times as it was mentioned by different subjects *(tokens)*. A *1-dimensional* χ^2*-test* revealed highly significant differences across categories in *types* [$\chi^2(3) = 81.26$, $p < .001$] and also when *tokens* were analyzed [$\chi^2(3) = 16.47$, $p < .001$]. On *types*-level, *convergent* features are the smallest proportion compared to other categories [$\chi^2(1) = 40.363$, $p < .001$]. *Emergent* features, however, represented the largest proportion [$\chi^2(1) = 31.015$, $p < .0001$]. When *tokens* were analyzed, this picture was converted. Then, *convergent* features held the largest share [$\chi^2(1) = 9.359$, $p < .002$], while the other kinds of semantic categories showed proportions of comparable size. Apparently, the relatively small percentage of *convergent* features (10.6%) is considered relevant by many subjects (30%), while, on the other hand, *emergent* features obviously lack such a good stability (40.5% *types* vs. 23.4% *tokens*). On average, a *convergent* feature was elicited with a frequency of 7.2. Conversely, occurrence of a single *emergent* feature was 1.5 on average.

The semantic similarity of the tenor and the vehicle of a given metaphor was determined by computing a coefficient that basically expresses the percentage of features common to both concepts (cf. Jenkins & Cofer, 1957). That is, we determined the extent to which the representations of a given tenor- and vehicle-concept are semantically similar *independent of* their co-occurrence in the metaphor. Here, it was our intention to explore whether such bottom-up similarity would influence subjects' choices among feature categories when generating metaphor interpretations. To examine this possibility, we correlated tenor-vehicle similarity with the proportions of feature categories in the interpretations. There was a substantial positive correlation between the proportion of *convergent* features in the interpretation and tenor-vehicle similarity [*types:* $r = .62$, $p < .002$; *tokens:* $r = .69$, $p < .001$]. At the same time, the proportion of *emergent* features and tenor-vehicle similarity proved to be inversely related [*types:* $r = -.58$, $p < .005$; *tokens:* $r = -.46$, $p < .024$]. Furthermore, there was also a substantial correlation between tenor-vehicle similarity and the proportion of *transferred* features, when analysis concentrated on *tokens* [$r = -.45$, $p < .025$], but not when *types* were concerned [$r = -.23$, $p < .176$].

Discussion. In some respects, our results are similar to those obtained by Tourangeau & Rips (1991, Ex. 1). This holds in particular for the proportions of *emergent* and *convergent* features as well as for the changes from *types* to *tokens*. However, Tourangeau & Rips did not discuss these changes in proportions and they did not analyze interindividual agreement on feature categories either. The fact that few *types* of *convergent* features are considered relevant by a large number of subjetcs is perfectly in line with theoretical assumptions about the role of tenor-vehicle similarity in metaphor interpretation (Johnson & Malgady, 1979). Usually, outside the metaphorical context, the tenor and the vehicle are similar to each other only in few respects. But these partial similarities increase in salience when both terms are perceived together in the metaphor. Therefore, it is plausible that subjects applied only a few *types* of *convergent* features in their interpretations. However, because of the increased salience of these features, they resorted to them more often than to other feature categories. In contrast to the high stability of *convergent* features, agreement among subjects on *emergent* features was very low although subjects produced a large number of different types of *emergents*. This result can be explained if we assume that in order to generate an emergent feature a comprehender has to resort to knowledge *external* to the lexical meanings of tenor and vehicle (Murphy, 1988). Since this knowledge presumably varies much between individuals, the probability that different subjects infer the same *emergent* feature should indeed be low. Another important conclusion refers to the positive correlation of tenor-vehicle similarity with the

number of *convergent* features and its negative correlation with the number of *emergent* features. The more semantic similarities exist between the tenor and the vehicle, independent of any metaphorical context, the more these similarities are exploited for generating an interpretation. On the other hand, the less a tenor and a vehicle are similar, the more it is likely that the interpretation is based on *emergent* features. Though our results remain somewhat ambiguous concerning the transferred features, we may nonetheless speculate that features from the vehicle are likely to be tranferred onto the tenor when the similarity between these concepts is low.

Study 2: Assessment of Metaphors

Study 2 was set up to obtain assessments of metaphors on several dimensions. Additional data were required to answer questions still open when analyzing the results of *Study 1*. More specifically, the finding requires further analysis that the majority of subjects produces divergent *emergent* features but at the same time shows good agreement on *convergent* features. Does the hypothesis hold that *emergent* features surface only when making sense out of difficult metaphors? Is the good agreement on *convergent* features merely a consequence of their increased salience during metaphor comprehension?

Subjects, Materials and Procedure. 29 undergraduate students in psychology participated in *Study 2*. All of them were native German speakers. The metaphors of *Study 1* were rated on several dimensions: E.g., *comprehensibility* was described as ease of interpretation, *propositional adequacy* denoted the aptness of the proposition a metaphor communicates and *metaphoricity* referred to the perception of incongruence between the tenor and the vehicle. Subjects were asked to assess each metaphor with respect to these criteria on a unidimensional 7-point scale.

Results and Discussion. The ratings from each of the scales were averaged across subjects. Then rank-order correlations were calculated between the results from the rating study and the frequencies of different feature categories that occurred in the interpretations of metaphors. The *propositional adequacy* of a metaphor was negatively related to the number of *emergent* features in the interpretation. This correlation was on the level of types [$r = -.53$, $p < .01$] more clearly expressed than on the level of *tokens* [$r = -.38$, $p < .052$]. However, the correlation between *comprehensibility* and the number of *emergent* features remained below the level of significance [*types:* $r = -.08$, $p > .1$; *tokens:* $r = -.30$, $p < .106$]. Obviously, a metaphor that made a point subjects found difficult to accept provoked more *emergent* features than propositionally adequate metaphors. This relation was independent of the metaphor's comprehensibility. When taking a look at the *convergent* features, an inverted picture emerged. The number of those features correlated positively with *propositional adequacy* [*types:* $r = .44$, $p < .031$; *tokens:* $r = .39$, $p < .048$]. Again, the correlation with comprehensibility remained below the level of significance [*types:* $r = .37$, $p < .06$; *tokens:* $r = .30$, $p < .103$]. These results are complemented by the further findings of a positive correlation of *tenor-vehicle similarity* with *propositional adequacy* [$r = .51$, $p < .013$] and a negative correlation with *metaphoricity* [$r = -.46$, $p < .023$]. Our findings partly differ from Tourangeau & Rips' results (1991, Ex. 1). These authors also found a reliable positive correlation between the proportion of *convergent* features and metaphor aptness, i.e., *propositional adequacy*. But in contrast to our results, the proportion of *emergent* features was positively related to aptness ratings though it was admitted that this correlation remained below statistical significance. Together, the results of both studies support the claim that features shared by the tenor and the vehicle are particularly relevant for developing an interpretation. Especially such metaphors with a large percentage of *convergent* features received high ratings for *propositional adequacy*. Obviously, it it not only their increased salience which raises the probability of *convergent* features to be used in the interpretation. Rather, *convergents* seem to contribute to the plausibility of the proposition a metaphor communicates. On the other hand, our data show that if the commonalities between tenor and vehicle decrease the scores for *propositional adequacy* also decrease. The interpretation is then based on *emergent* features. However, there is no indication that such metaphors are particularly difficult to understand. One might speculate that the construction of new relations between the domains of tenor and vehicle is encouraged when processing metaphors which show low values for *tenor-vehicle similarity* and *propositional adequacy* but at the same time are markedly perceived as *metaphorical*.

Study 3: Semantic Priming of Tenor-Vehicle Pairs in a Double Lexical Decision Task

The object of this experiment was to provide evidence against Camac & Glucksberg's conclusions (1984). In their lexical decision task, word pairs that formed the tenor and the vehicle of good metaphors were *not* decided faster than randomly combined pairs of words. Camac & Glucksberg conluded that there are no semantic similarities between tenor and vehicle that could be used to aid the interpretation. If we succeeded in demonstrating semantic priming for at least one particular set of metaphors this would seriously question Camac & Glucksberg's claim. At the same time, such a result would limit the explanatory scope of the interaction approach. Based on our grouping in *Study 1*, we predicted that subjects would make faster lexical decisions about tenor-vehicle pairs with high semantic similarity than about tenor-vehicle pairs with a low degree of similarity. And we further expected that tenor-vehicle pairs would generally be responded to more quickly than word pairs formed by random combination. Such differences in reaction times should result from semantic priming which is a function of the similarity of prime and target.

Subjects and Materials. 27 undergraduate students served as subjects in the experiment. All of them were native German speakers. Subjects performed a lexical decision task in which they were to judge two strings of letters simultaneously presented on a screen. Subjects responded "yes" to three types of word-word pairs (a-c) and "no" to three types of nonword pairs (d-f). The 54-item test list contained (a) 9 tenor-vehicle pairs with high semantic similarities (HS-pairs), (b) 9 tenor-vehicle pairs with low seman-

tic similarities (LS-pairs), (c) 9 semantically unrelated pairs (UR-pairs), (d) 9 nonword-nonword pairs, (e) 9 word-nonword pairs, and (f) 9 nonword-word pairs. The last three types of pairs (d-f) were included to prevent subjects from using a strategy of looking at either the left or the right member of a pair to make quick nonword decisions. For these pair types merely served as fillers, they were not in the data-analysis.

Design, Apparatus and Procedure. Every subject saw every stimulus pair of each pair type, but no subject would see any string of letters more than once. Consequently, all treatments were *within* subjects. *Types of stimulus pairs* was treated as the independent measure, and *response time* was used as the dependent measure. Hence, we obtained a one-factorial design with *types of stimulus pairs* as a three-level fixed factor. Subjects were run individually during one session involving a series of discrete reaction time trials. Subjects were told that they would see pairs of letter strings, presented one above the other in the centre of a computer screen, and that half the pairs would both be words, while the other half would contain at least one non-word. The task was to decide as quickly and accurately as possible, whether or not both strings were words. If both strings were words, subjects pressed the 1-key on the right side of the computer keyboard to indicate "yes" and otherwise the 2-key to indicate "no".

Results and Discussion. In this experiment, we are only interested in the effect on lexical decision time which can be attributed to the degree of semantic similarity between the words to be decided. However, we suspected that the tenor-vehicle pairs from the previous studies which we used as experimental stimuli would systematically differ with regard to psycholinguistic variables such as word length, frequency of occurrence, concreteness etc. To control for such possible influences on lexical decision time, we collected additional lexical decision latencies for every single word occurring in the present experiment. In the data-analysis, these single word decision latencies were treated as a covariate of the dependent variable, i.e., *response time for the word pairs*. A one-factorial ANCOVA based on item means was conducted for this dependent measure with *item means of single word decision latencies* as a covariate. The item analysis (F2) revealed a significant main effect for *types of stimulus pairs* [$F(2, 23) = 7, 75$, $p < .003$]. Orthogonal a-priori contrasts were calculated for the difference between HS- and LS-pairs and also for the difference between the whole set of tenor-vehicle pairs and UR-pairs. In ANCOVA models, such contrast tests are usually conducted for adjusted means, such that possible differences in the effect of the covariate for different levels of the factor are accounted for. The prediction that HS- and LS-pairs together should be faster than UR-pairs could be confirmed [$F(1, 23) = 13, 04$, $p < .002$]. However, the difference between HS- and LS-pairs just failed to reach statistical significance [$F(1, 23) = 3.13$, $p < .09$]. Rather than applying also a factorial design, analysis of subjects (F1) was carried out by using a multiple regression approach (Lorch & Myers, 1990). More specifically, seperate multiple regressions were conducted for each of the 27 subjects. The two theoretically interesting contrasts constituted two dummy-variables treated as nominally scaled predictors. Furthermore, *item means of the single word decision latencies* were a continuous predictor. The resulting regression coefficients for each predictor were averaged across subjects and tested for significance. In this subject analysis, both of our predictions were confirmed. Again, the difference between the whole set of tenor-vehicle pairs and UR-pairs proved to be reliable [$t(1, 26) = 2.53$, $p < .018$]. But also the prediction that HS-pairs should be decided faster than LS-pairs was statistically significant [$t(1, 26) = -2.13$, $p < .042$]. Taken together, these results show that subjects reliably responded faster to word pairs derived from metaphors than to random pairs of words (80,3 ms). At the same time, subjects' lexical decisions were faster to semantically similar tenor-vehicle pairs compared to more semantically dissimilar tenor-vehicle pairs (42,5 ms). Certainly, there remain some caveats regarding the generality of the latter conclusion with respect to stimulus material other than the present one. Nevertheless, our results are straightforward with regard to our main contention. Camac & Glucksberg's claim (1984) that there are no relevant similarities between the tenor-and vehicle-concepts of apt metaphors has successfully rebutted. Therefore, the theoretical assumption that similarity is fundamentally a *product* rather than a prerequisite of comprehension should be treated with some more caution.

Table: Results of the lexical decision task (F2)
(Reaction time in msec)

Pairtype	Mean (adj.)	Std. Dev. (adj.)
HS	753.505	51.074
LS	795.207	49.369
UR	851.929	51.544

General Discussion

The first study showed that metaphor interpretations were based to a large extent on emergent features. On the other hand, features shared by tenor and vehicle represented the smallest proportion compared to other feature categories. However, subjects agreed very well on these relatively few *convergents* whereas interindividual agreement was low on the *emergents*. *Study 1* further showed that the production of *emergent* features is encouraged especially when the similarity between tenor and vehicle is low. In contrast, if tenor-vehicle similarity was high then the ground of interpretation contained many *convergent* features. *Study 2* provided additional support for these conclusions. In particular, features shared by tenor and vehicle enhanced the plausibility of the proposition a metaphor communicates. On the other hand, metaphors that made a point subjects found difficult to accept (i.e., low ratings of propositional adequacy) provoked significantly more *emergent* features. In *Study 3* the prediction was confirmed that tenor-vehicle pairs with high similarities would produce shorter lexical decision latencies than tenor-vehicle pairs with low similarities.

What can be concluded from these results? Contrary to Camac & Glucksberg's claim at least one particular subset of nominal metaphors can be distinguished the tenor- and vehicle-concepts of which share significant semantic simila-

rities. And as can be concluded from the feature analysis and rating study, these similarities are also exploited in order to generate a coherent and plausible interpretation. However, Camac & Glucksberg are not completely wrong because there are also metaphors with relatively low tenor-vehicle-similarity. Such metaphors are predominantly interpreted via a construction of emergent features. Concerning the debate between the comparison and the interaction approach to metaphor, our results offer a clear answer. None of these approaches is broad enough to provide an exhaustive explanation of metaphor comprehension. Instead, both these approaches point out to two different but complementary cognitive processes underlying metaphor comprehension.

Based on our results, we may speculate that metaphor comprehension proceeds in two successive stages. First an *analysis* of the lexical meanings of tenor and vehicle is attempted. If there are enough similarities to produce a coherent and plausible interpretation, the comprehension process will cease. In case the tenor-vehicle similarity is not sufficient a shift to a different processing mode will take place. In this case , the metaphor is not interpreted by just resorting to the lexical meanings of tenor and vehicle. Rather, there will be a *synthesis* of these terms. Such *synthesis* requires the activation of broader world knowledge about the semantic domains involved. A coherent interpretation is then achieved through a construction of new components of meaning. This is reflected in a high number of *emergent* features. The distinction between *analysis-based* and *synthesis-based* processing is closely akin to the distinction between similarity-based and similarity-creating metaphors (Indurkhya, 1992). However, Indurkhya's theory seems to suggest a conception of *discrete* metaphor types which is not confirmed by our results. In particular, the fact that metaphorical meanings can be empirically described as patterns of different semantic effects proposes to conceive of similarity-based and similarity-creating metaphor as opposite poles located on the same continuum. Certainly, our sketch of a process model is far from being specific enough to allow for implementation. Especially the hypothesis that metaphor comprehension proceeds in a cascaded fashion needs further investigation. Alternatively, it may be possible that the *analytic* and *synthetic* processing mode operate simultaneously during comprehenion. For example, a recent reaction time study by Gineste, Indurkhya & Scart-Lhomme (1997) shows that *emergent* features may be more quickly accessed during comprehension than transferred or redundant features. This result is further parallelled by work on the comprehension of contextual innovative expressions. These are expressions where a new meaning is assigned to a string of words instead of its conventional one. Gerrig (1989) contrasted two models of meaning construction and received experimental support for simultaneous processing rather than for an error recovery strategy. That is, selection of conventional word meanings and creation of a new meaning operated simultaneously rather than in successive stages. There are obviously some parallels between Gerrig's opposition of sense selection vs. sense creation and our distinction between *analysis* and *synthesis*. However, it must be stressed that with *analysis*-based processing we do not imply that a literal interpretation of the metaphor is attempted first before a figurative reading is created (cf. the *three-stage model of metaphor comprehension*, Searle, 1979). In contrast, it is our contention that *analysis* and *synthesis* refer to complementary cognitive processes controlling the generation of *metaphorical* meanings. Furthermore, the parallel between Gerrig's and our work indeed suggests that the principle of dual processes in meaning construction may even hold across the literal-figurative distinction. However, much work has yet to be done to fully explore these issues.

References

Black, M. (1979). More about metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 19-43). New York: Cambridge University Press.

Camac, M. K. & Glucksberg, S. (1984). Metaphors do not use associations between concepts, they are used to create them. *Journal of Psycholinguistic Research, 13*, 443-455.

Gentner, D. (1988). Metaphor as structure mapping: the relational shift. *Child Development, 59*, 47-59.

Gerrig, R. J. (1989). The time course of sense creation. *Memory & Cognition, 17*, 194-207.

Gineste, M.-D., Indurkhya, B. & Scart-Lhomme, V. (1997). *Mental representations in understanding metaphors* (Technical report No.97-02). Groupe Cognition Humaine, LIMSI-CNRS, Orsay Cedex.

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

Indurkhya, B. (1992). *Metaphor and cognition*. Dordrecht: Kluwer Academic Publishers.

Jenkins, P. M. & Cofer, C. N. (1957). An exploratory study of discrete free association to compound verbal stimuli. *Psychological Reports, 3*, 599-602.

Johnson, M. G. & Malgady, R. G. (1979). Some cognitive aspects of figurative language: Association and metaphor. *Journal of Psycholinguistic Research, 8*, 249-265.

Lorch, R. F. & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 149-157.

Malgady, R. & Johnson, M. (1980). Measurement of figurative language: Semantic feature models of comprehension and appreciation. In R. Honeck & R. Hoffman (Ed.), *Cognition and figurative language* (pp. 239-258). Hillsdale: NJ: Erlbaum.

Murphy, G. (1988). Comprehending complex concepts. *Cognitive Science, 12*, 529-562.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review, 86*, 161-180.

Searle, J. R. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 92-124). New York: Cambridge University Press.

Tourangeau, R. & Rips, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory and Language, 30*, 452-472.

Tourangeau, R. & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition, 11*, 203-244.