Similes on the internet have explanations

Carlos Roncero, John M. Kennedy and Ron Smyth

University of Toronto

Running head: Internet similes and explanations

Contact author: John M. Kennedy

Department of Life Sciences -- Psychology

University of Toronto at Scarborough

1265 Military Trail

Toronto ON M1C1A4 Canada
Abstract

We searched the internet for expressions linking topics such as crime and vehicles such as disease as similes (crime is like a disease) and metaphors (crime is a disease). We counted the number of times the expressions were accompanied by explanations (crime is like a disease because it spreads by direct personal influence). Similes were more likely than metaphors to be accompanied by explanations. Similes may be preferred if the writer wants to express an out-of-the-ordinary relation between the topic and the vehicle.
Here we will test a hypothesis about metaphors, similes and explanations, using sentences on the Internet.

A simile is a figurative comparison such as *highways are like snakes*. A metaphor is figurative without the term *like* as in *crime is a disease*. Similes imitate literal comparisons such as *Fords are like Chryslers*. Metaphors make claims about a category, as in *Fords are cars* (Glucksberg, 2001; Gentner and Bowdle, 2001). Comparisons and categorization are vital to human cognition, so figurative versions could be present in all cultures, but despite study for millennia the connection between similes and metaphors is still much debated (Chiappe and Kennedy, 1999).

A figurative relation usually can be expressed using the same word pairs as a metaphor or a simile. *Crime is like a disease* can drop *like* and have the same sense as *crime is a disease*. Literal comparisons cannot drop or add *like* with impunity. *Fords are like cars* is incorrect.

Chiappe, Kennedy and Smykowski (2003) argued metaphors are preferred when the relationship being expressed is quite apt, as in *cigarettes are time bombs*, similes if not, as in *trees are like straws*. Aptness is high if the vehicle (*time bombs*) points out what the reader takes to be significant features of the topic (*cigarettes*).

Of interest for present purposes, Chiappe, Kennedy and Chiappe (2003) found that aptness ratings correlated strongly with ease of comprehension. This suggests that metaphors would be used if a statement is easily comprehended, and similes if not. Similes are more challenging. Here we take up an intriguing implication of this argument. If writers thought comprehension were impeded, what would ensue? In everyday practice writers might add explanations to their expressions. That is, if a simile did indeed seem
unlikely to convey a key idea, the writer should expose its rationale to the light of day as in *trees are like straws in the way they suck up water and nutrients*. If they readily bring to mind what the writer wants the expression to specify, metaphors could often occur baldly, totally without explanation as in *life is a journey*, as compared to *life is like a box of chocolates – you never know what you are going to get*.

To check if compared to metaphors similes come more often with explanations, we turned to a very large corpus of sentences. We Googled the internet for metaphors and similes. We examined the products for accompanying explanations.

**Experiment 1**

**Method**

A set of 52 pairs of terms was selected from Chiappe, Kennedy and Chiappe (2003) e.g. rage paired with volcano, education with stairway. The pairs were written as sentences in metaphor form (*rage is a volcano*) and simile form (*rage is like a volcano*).

Each sentence was entered into a search engine, and an Internet search undertaken to discover the frequency of each sentence.

The sentences were sought via Google. When a sentence is written in quotation marks in the search box, Google displays a list of websites that contain each sentence, and shows the linguistic context. When *life is a journey* was entered into the search box, Google produced a list of websites containing *life is a journey*, and words adjacent to the sentence on each website. A count of the websites constitutes the frequency measure for that sentence. The order in which Google presents websites is determined by the number of links to that page by pages that have many links.
To ensure that the count included only relevant productions of metaphors and
similes, constraints were used, as follows:

**Constraints for the target sentences**

**Constraint 1:** The principle of 1 website = 1 production. Productions listed within the
same website were recorded as a single production. Thus no single website could
dominate the recorded frequency.

**Constraint 2:** The “no example” principle. Productions that were examples of figurative
claims were excluded. For example, a website would not be counted if it included the
sentence “The metaphor “life is a journey” can be rewritten as the simile “life is like a
journey.”” For this reason, productions from psychology articles, websites, and academic
discussion of figurative language were not included.

**Constraint 3:** The “identical syntax” principle. A production may have used the same
word order as the search sentence, but it was not counted if it was not syntactically the
same sentence. For example:

**Target sentence:**  
\textit{wisdom is like an ocean}

**Found sentence:**  
The person who doubts he will receive \textit{wisdom is like an ocean} wave that is driven and tossed.

**Constraint 4:** The principle of unclear context. The production was not used if it could
not be determined whether it was a use of figurative language or an example of figurative
language (as defined in constraint 2)

**Constraint 5:** The principle of 1 referent = 1 production. Productions with the same
referent were recorded as a single production. For example, several websites could
include the book title \textit{wisdom is an ocean}, but each refers to the same token.
Constraint 6: The principle of 1 context = 1 production. Metaphors and similes repeated with the same linguistic context were recorded as a single production. By this rule, all instances of the mind is an umbrella – best when open should be recorded as a single production. This prevents a few uses of an expression dominating the results.

Constraint 7: The principle of different semantics. Productions may match the target sentence in word order, but have a different meaning. For example,

Target sentence: Time is a thief

Found sentence: If time is a thief of memory, I’ve been royally fleeced.

The produced sentence still refers to time as a thief; however, the addition of the prepositional phrase of memory restricts the meaning of thief in a way not relevant to the target sentence.

Constraints for the explanations

Constraint 8: Principle of no repetition. Each explanation is counted once. Therefore, the number of explanations for a figurative claim is the number of different explanations found for the figurative claim. See also Constraint 6.

Constraint 9: Elaboration rather than explanation. Some productions are elaborations of the metaphor or simile, not explanations. Their relative clauses are introduced by complements other than because, and when replaced by the complement because they become ungrammatical. For example:

*Time is like a thief that steals everything away*

This is ungrammatical as *time is like a thief [because] steals everything away.*

A legitimate explanation is *Music is like medicine as it takes away the pain.*

This is grammatical as *Music is like medicine because it takes away the pain.*
52 topic-vehicle pairs used by Chiappe et al. (2003) were written into Google. Only 26 of these pairs, however, had corresponding productions on Google. A further 13 topic-vehicle pairs were removed from the analysis as there were fewer than 9 productions of metaphors and similes combined.

Using the constraints listed above, each production was judged for legitimacy. The productions and contexts were recorded and examined by all 3 authors independently, and only instances judged unanimously to be legitimate were kept. The final list of pairs of terms is in Table 1.

Because Google often produced a large number of instances of a particular topic-vehicle pair, only the first 30 legitimate productions of each metaphor and simile were retained if the number of productions exceeded 60. Otherwise, all productions of the target sentence were examined.

Results

The mean proportion of metaphors with explanations was 6% (SD = .07); of similes 31% (SD = .26), z-score 8.6, p<.01. For metaphors, proportions ranged from 0% for alcohol is a crutch, cities are jungles, lawyers are sharks, life is a journey, and time is money to 23% for music is medicine. For similes, proportions ranged from 0% for alcohol is like a crutch, cities are like jungles, and soldiers are like pawns to 73% for time is like money. In total, 79% of all explanations occurred with a simile (see Table 1).
The results were promising. However, the topic-vehicle pairs from Chiappe et al. were obtained from the psychological literature on metaphor, and the search was with these preselected pairs. Comparisons that are found with a less specific search would strengthen the claim. Also, only 13 topic-vehicle pairs were compared in Experiment 1. Additional comparisons would be useful.

Experiment 2

Statements to do with the terms “metaphor” and “simile” were Googled. Four search phrases were employed: “common metaphor,” “common simile,” “an example of a metaphor is,” and “an example of a simile is.” The same criteria as in Experiment 1 were used to deem examples to be relevant but without the no-example constraint, and one additional constraint:

Constraint 10: no synonym. An explanation was rejected if it simply replaced the vehicle with a synonym, for example, *God is a rock, that is to say, a stone.*

Results

The mean proportion of metaphors having explanations was 3%; of similes 31%, (z-score 10.17, p<.01). For metaphors, proportions ranged merely from 0% to 6%, for similes, from 0% to 56% (*love is like a rose*), see Table 2. Also while only 31% of similes were accompanied by explanations, 86% of the explanations were preceded by a simile (very much in accord with 79% in Experiment 1).
Discussion

Metaphors often appear in internet text without explanation. On the face of it, their writers deem them relatively intelligible. More frequently similes carry explanations. Seemingly, they are quite likely to require aid.

One factor could be the expression’s aptness. Metaphor is favored if the relevant features that the topic and vehicle share are highly significant in understanding the topic, i.e. it is apt. *Crime is a disease* points out contagion and decay are relevant, for example. *Education is a stairway* makes progress relevant. *Cities are jungles* argues danger and lack of control are relevant. *Life is a joke* suggests we are the butt.

In addition, writers may use metaphors when the key features are expected to come to mind without prompting (Gentner and Bowdle, 2001). *Love is a devil* alerts us to its perils. *Heaven is a treasure* emphasizes its value. If the topic and vehicle do not share a set of relevant and salient features (Giora, 2003), then interpretation is not easy. The topic and vehicle have many features to consider. Several interpretations might be possible. Of importance, the intended features may be so low in salience the unaided reader would never consider them.

No explanations were added to *time is money* as a metaphor. Its salient meaning is our time on the job costs someone money. The implication is our time is valuable and is not to be wasted.
Several expressions, including *time is money*, *Christians are salt* and *love is a rainbow* were given no explanations as metaphors and a large number as similes. Surely this means that the metaphor version of the time-money pair, for example, may have a specific meaning (Gibson, 1979), but the simile version is given creative or far-fetched accounts, abetted by explanations. *Time is like money*, a simile, was given explanations such as *the less we have of it to spare the further we make it go and once it’s spent it’s gone*. Accompanied by explanations, it specifies less salient aspects of money. It can refer to scarcity, investment or running out of money, for example.

The stock-in-trade of similes is what the prompting explanation can bring readily to mind. In effect, similes, via explanations, can license adroit narrow applications such as *time is like money – only retired executives have a lot*. A more upbeat example might be *time is like money and you can invest yours in your education*. This account can explain why similes are often used in riddles (as in *why is a raven like a writing desk?*) licensing highly specific referents (*because Poe wrote on both*) based on common features with a clever, unexpected, humorous side (playing on two meanings of *on*).

Low-apt expressions that subjects reject as metaphors can be rescued by expressing them as similes plus an explanation. *A canary is a wolf* might be rejected but readers may suspend judgment on *a canary is like a wolf* looking for a justification in the text. It might be that flocks of canaries in competition for territory in their natural habitat are unrelentingly vicious to some other birds.

*Conventionality*

Conventionality plays such a significant role in figurative language that highly conventional expressions have their own classification – cliché. Gentner and Bowdle
(2001) pointed out the vehicles of many metaphors are highly conventional, for example
*schools are a jungle* uses jungle in the same way as *cities are a jungle*. They argued that
one result may be we use a simile to express an unfamiliar comparison, a metaphor for a
familiar one. Interestingly however, explanation proportions on Google were strongly
correlated (r = +.70, p < 0.001) with the conventionality ratings of the vehicle provided
by Chiappe et al. (2003). For example, *thief, computers,* and *blueprints* are ranked first,
second, and third in terms of vehicle conventionality in Chiappe et al. (2003), and all
three have simile explanation ratings higher than 49%. That is, despite their high
conventionality, they are used as similes (albeit adorned with explanations) precisely the
reverse of what one might predict at first blush. Instructively, what this suggests is that
highly conventional vehicles are often qualified and repurposed! If *minds are computers*
does not spotlight the features the writer wants, they are specified explicitly. For
example, internet explanations redirected it with *in that brains are like hardware and
minds are like software,* and *as all information is stored there, and what we put in is what
we get out* and *garbage in, garbage out.*

*Internet productions*

Chiappe et al. (2003) examined preference for expressing a comparison between
two terms as a metaphor or simile. We correlated their metaphor preference ratings for a
given pair with the proportion of metaphors to similes for that pair found by Google in
Experiment 1. The 52 topic-vehicle pairs from Chiappe et al. (2003) were considered. We
rejected pairs that had 3 or fewer productions (metaphors and similes combined), leaving
21 pairs. We recorded all legitimate productions for the remaining topic-vehicle pairs
(not only the first 30). The Google proportions correlated with the Chiappe et al. (2003)
metaphor preferences $r = +0.57$, $p < 0.001$. Evidently, preferences expressed in an experimental context predict widespread use.

One might wonder if our results reflect written language but not spoken language. This is worth further examination. However, we note that messages on the Internet are often closer in form to actual speech than academic papers and published works. They include non-lexicalized sounds (e.g. that is waaaaaay too harsh, I SO disagree), and often include non-standard syntax (e.g. you wanna go?).

Internet search methods obtain information easily and reliably. The positive correlation between preferences in Chiappe et al. (2003) and Google’s productions suggests results can be generalized validly, paving a way for internet tests of other language theories. The close fit between Experiment 1 and 2’s results suggest different methods for searching the Internet reinforce each other’s conclusions.

We used the same word pairs in our searches and comparisons. This was to control for content. An interesting possibility is that some word pairs may occur only as similes or only as metaphors. An example is riddles, which overwhelmingly occur as similes. Another example is clichéd metaphors such as God is love. These fell outside of our constraints. We conjecture that the riddle-type are far fetched, and require explanation, the clichés have passed beyond an extreme measure of conventionalization, are often taken to be literal claims of fact, and are rarely followed by explanation. Another conjecture is that a sentence by sentence search of texts (or records of speech) for, say, 1000 metaphors and 1000 entirely unrelated similes should find explanations more frequent for similes.
We searched for explanatory text following a word pair. However text preparing the way for metaphors and similes with different levels of explanatory transparency may mirror the subsequent text. It may be terse before an apt metaphor that has no trailing explanation, and especially expository before a simile that has one.

We conclude that our internet study of metaphors and similes finds similes more often supplemented by explanations. Notably, similes are favored if a standard vehicle’s use has to be qualified.
References


Table 1: Metaphors and similes in Experiment 1, frequency of explanations

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Metaphors</th>
<th></th>
<th></th>
<th>Similes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With explanation</td>
<td>Without</td>
<td>With explanation</td>
<td>Without</td>
<td></td>
</tr>
<tr>
<td>Alcohol-crutch</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cities-jungles</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Crime-disease</td>
<td>3</td>
<td>27</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Genes-blueprints</td>
<td>1</td>
<td>35</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Lawyers-sharks</td>
<td>0</td>
<td>24</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Life-journey</td>
<td>0</td>
<td>30</td>
<td>5</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Love-drug</td>
<td>2</td>
<td>28</td>
<td>19</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Minds-computers</td>
<td>2</td>
<td>22</td>
<td>23</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Music-medicine</td>
<td>7</td>
<td>23</td>
<td>7</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Soldiers-pawns</td>
<td>1</td>
<td>19</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time-money</td>
<td>0</td>
<td>30</td>
<td>21</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Time-thief</td>
<td>5</td>
<td>24</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Words-daggers</td>
<td>1</td>
<td>24</td>
<td>1</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Metaphors and similes in Experiment 2, frequency of explanations

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Metaphors</th>
<th>Similes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With explanation</td>
<td>Without</td>
</tr>
<tr>
<td>Bible-sword</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Christ-door</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Christian-salt</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>God-fire</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>God-rock</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Hair-rainbow</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Heaven-treasure</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>Life-joke</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Life-river</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>Love-devil</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Love-flower</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>Love-gold</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Love-melody</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Love-oxygen</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>Love-rainbow</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Love-Rose</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Man-island</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>TV-drug</td>
<td>2</td>
<td>28</td>
</tr>
</tbody>
</table>