Exploring the Role of Word Associations in the Construction of Rhetorical Figures

Paloma Galván[1], Virginia Francisco[1], Raquel Hervás[1], Gonzalo Méndez[2], Pablo Gervás[2]
[1]Departamento de Ingeniera del Software e Inteligencia Artificial
[2]Instituto de Tecnología del Conocimiento
Universidad Complutense de Madrid, Spain
palomagalvan@ucm.es, {virginia, raquelhb, gmendez}@fdi.ucm.es, pgervas@sip.ucm.es

Abstract

Figurative language is a fundamental characteristic of elaborate forms of linguistic communication. We currently have very poor models of how figurative language may be constructed in computational terms. The overall aim is to identify possible regularities, intuitions or heuristics that may at a later stage be employed to drive a text generator that is capable of using this type of rhetorical figure.

Introduction

The use of figurative language is a fundamental tool in linguistic communication. One of the most easily identifiable characteristics of computer generated text is the tendency to stick to literal meanings. This is partly because literal meanings are unambiguous and have less risk of misinterpretation. But it is also in part due to the fact that we currently have very poor models of how figurative language may be constructed in computational terms. This paper explores the relationship between word associations as modelled in already available computational resources and the type of rhetorical figures that people employ regularly. The aim is to identify possible regularities, intuitions or heuristics that may at a later stage be employed to drive a text generator that is capable of using this type of rhetorical figure.

We consider three types of rhetorical figures or tropes. A *metaphor* is a widely-used literary mechanism which allows comparison between two disparate concepts. Metaphors transfer the qualities of one word to another, as in *Booger was a lion in the electoral arena*. Here, the qualities of *lion* (the source) are transferred to Booger (the target). A *simile* is a pointed, direct and explicit metaphor where two different things are compared to evolve a new meaning. A simile denotes the target to be like the source, and as such the target cannot totally be substituted by the source. A simile is a kind of metaphor where the comparison is made using the words “as” or “like”. For example, *Booger was like a lion*. An *analogy* links two disparate concepts by common properties, as in *Booger was as brave as a lion*. Here, the quality of be *brave* (the property) is used to link *lion* (the source) to *Booger* (the target).

Metaphors play an important role in communication, occurring as often as every third sentence (Shutova et al. 2012), so the generation of metaphors is essential for Natural Language Generation. The same occurs for analogies and similes.

Black (1955) made explicit that metaphors depend upon conceptual connections between networks of concepts. Inherent in this approach is the idea that metaphors are a matter of cross-domain mapping (Lakoff 1993). A metaphor is a cognitive process that builds or maps connections between networks of concepts as it occurs with similes and analogies. In consequence, to generate metaphors a conceptual structure is needed where every concept is placed not only taking into account its conventional usage but its diverse and unconventional usages (Veale 2014b). The best place to find this complex structure is the Web and that is where we are going to look for word associations in order to create our analogies, similes and metaphors.

This paper presents a new approach to finding word associations in the web using Thesaurus Rex (Veale and Li 2013). Then the potential of this system will be studied for the automatic generation of analogies, similes and metaphors.

This paper is organized as follows. The second section presents prior work on analogy, simile and metaphor generation. The third section explains our approach to finding word associations. In the fourth section the evaluation of our approach for rhetorical figures generation is presented. And finally, in the last section conclusions and future work are explained.

Related Work

Rhetorical figures have been the target of researchers in computational approaches to linguistics on and off for many years. However, only in recent years has the combination of available knowledge resources and accumulated insights allowed for the field to flourish. Metaphors have been widely studied in Natural Language Analysis but not so much in Natural Language Generation (NLG). There is a lot of work related to metaphor detection (Wilks et al. 2013), identification (Shutova, Sun, and Korhonen 2010), extraction and annotation (Wallington et al. 2003) but few related to metaphor generation. The reason can be that metaphor generation is as challenging as human creativity will allow. In this section the most important approaches for simile and metaphor generation are presented.

Approaches to Rhetorical Figures in NLG

In the field of natural language generation there have been a number of attempts to establish procedures for constructing rhetorical figures as important ingredients of generated spans of text. This has been attempted both in general terms (Hervás et al. 2006b) for different types of rhetorical figures, and for specific cases like analogies (Hervás et al. 2006a) or metaphors (Hervás et al. 2007). These attempts were all carried out before adequate sources of machine-readable knowledge were available and consequently suffered from a thirst of appropriate knowledge. The attempts considered the problem of rhetorical figure employment in text generation in general theoretical terms but lacked sufficient volume of explicit knowledge on the underlying semantics of words to be capable of practical generation.

Approaches to Conceptual Construction of Rhetorical Figures

The recent development of sources of knowledge that allow easy mining of large corpora of text for significant word associations has lead to the emergence of a number of systems that rely on these for constructing rhetorical figures of different types.

Jigsaw Bard Jigsaw Bard (Veale and Hao 2011) is a web service that exploits linguistic readymades to generate similes on demand. Jigsaw Bard scans Google n-grams to index potential readymades which are then re-purposed as a simile. For example, given the adjectival property *quiet* Jigsaw Bard returns the simile *"The peaceful life of a monastery"*. The Jigsaw Bard is best understood as a creative thesaurus: for any given property (or blend of properties) selected by the user, the Bard presents a range of apt similes, and users must decide which similes are most suited to their descriptive purposes.

Thesaurus Rex Thesaurus Rex (Veale and Li 2013) is a web service that given two concepts (for example, War and Divorce) returns a phase cloud of the nuanced categories that are shared by both concepts (in the given example it returns a cloud that contains *traumatic-event*, *stressful-event*, *unexpected-event*...).

Thesaurus Rex organizes concepts according to categories they are placed into by speakers in everyday language (*food*, *drink*, *beverage*...). These categories have an associated weight that represents their relative importance for the given concept. Thesaurus Rex can show different categories for each concept and allows in turn to consult the concepts in each category. For example for the concept *coffee*, some of its categories with more weight are *beverage* or *drink* and some with less weight are *leaf* or *apposition*. Concepts in Thesaurus Rex have associated properties or modifiers which are accompanied by a non-standard weight indicating how strong its relation to the concept is. For example, for *coffee* some of the modifiers with more weight are *hot*, *acidic* or *stimulating*, and modifiers with less weight are *smaller* or *adult*.

Metaphor Magnet Metaphor Magnet (Veale and Li 2012) is a Web service that allows users to enter queries with single terms (such as *leader*), compound terms with an affective spin (such as *good leader* or *+leader*), or copula statements (such as *"Steve Jobs is a +leader"*). For each input, the service marries its extensive knowledge of lexicalized stereotypes to the grand scale of the Google n-grams to generate the most appropriate affective elaborations and interpretations. In each case, Metaphor Magnet provides an explanation of its outputs. If *Steve Jobs* were to be viewed as a *master*, the properties *skilled*, *enlightened*, *free* and *demanding* are all highlighted as being most appropriate. Metaphor Magnet sees metaphor interpretation as a question of which properties are mapped from the source to the target.

Metaphor Magnet lacks a proposition level view of the world, in which stereotypes are linked to other stereotypes by arbitrary relations.

Metaphor Eyes Metaphor Eyes (Veale 2014a) employs a propositional model of the world that reasons with subject-relation-object triples rather than subject-attribute pairs (as Metaphor Magnet does). Metaphor Eyes acquires its world-model from a variety of sources and it views metaphor as a representational lever, allowing it to fill the holes in its weak understanding of one concept by importing relevant knowledge from a neighboring concept.

Metaphor Eyes metaphorize one concept (the source) as other concept (the target). Given *Scientist* and *Artist* it generates metaphors as *"Scientists develop ideas like artists"*.

Figure8 Figure8 (Harmon 2015) is a system that contains an underlying model for what defines creative and figurative comparisons, and evaluates its own output based on these rules. The system is provided with a model of the current world and an entity in the world to be described. A suitable vehicle is selected from the knowledge base, and the comparison between the two nouns is clarified by obtaining an understanding via corpora search of what these nouns can do and how they can be described. Sentence completion occurs by intelligent adaptation of a case library of valid grammar constructions. Finally, the comparison is ranked by the system based on semantic, prosodic, and knowledge-based qualities.

Word Association Generation

This section presents the proposed approach for the generation of word associations, which has been implemented as a web service. This service receives a common noun as an input, which is the target concept for the word association. Following the steps described in the Process section below, the system generates source concepts with similar properties to the target concept creating word associations.

Entry

The proposed approach receives a common noun as an input, which is the target concept for which the word association must be generated. Using Thesaurus Rex, the system unfolds a comparison between the target concept and another concept that acts as the source of the rhetorical figure, with similar properties to the target concept in order to create a word association.

Table 1: Examples of word associations obtained. Words in bold represent the choices made for each example.

Step	Target	snow	thunder	network
1	**Categories**	surface, elements, weather…	noise, sound, event…	system, structure, entity…
2	**Modifiers** soft, white…	natural, reflective, **slippery**, sudden, weather…	natural, **loud**, adaptive, physical…	**social**, complex,
3	**Categories for the selected modifier**	**surface**, ground, stuff…	**instrument**, thing…	**institution**, event activity, science…
4	**New query**	slippery surface	loud instrument	social institution
5	**Obtained concepts**	satin, silk, nylon, polyester… saxophone…	trumpet, drum, horn, religion…	family, government,

Process

Table 1 shows a few examples of target concepts and how Thesaurus Rex is used to obtain words associated to the target concepts. Taking the first concept, *snow*, as an example, the detailed process is the following:

1. **Target concept categories.** To obtain the filtered categories to which the target concept belongs, we first extract a list of all the general categories of the concept using a Thesaurus Rex query. From this list, only the $N\%$ of categories with the highest weights are considered as candidates. The value of N is configurable (in this example, N = 0.4). If a high N value is set, we will have in the list categories with lower weights, which are less relevant to the target concept. In the same way, we can set N to a low value, facing the risk of shortening the list to a single element. In the *snow* example, the categories with higher weights in Thesaurus Rex are *surface* and *weather*.

2. **Modifier extraction.** In addition to the categories, we also need a list of modifiers associated to the target concept, which is returned by a new query to Thesaurus Rex. From this list, the $N\%$ of attributes with the highest weights are considered as candidates (in this example, N = 0.6). For example, if our target concept is the noun *snow*, some of the most important properties extracted are: *natural*, *reflective*, *slippery*, *soft* and *white*.

   Modifier selection. One of the modifiers previously obtained is randomly selected. This random selection makes the system less repetitive, as the words associated to the same target concept are not always the same as if only the modifier with the highest weight were selected. For the current example, we suppose that the system has chosen the modifier *slippery*.

3. **Categories selection.** Using the modifier chosen in the previous step, a new query to Thesaurus Rex is performed in order to obtain categories that present this modifier as a highlighted property. In the *snow* example, the categories selected could be *surface*, *ground* and *stuff* which are categories that present the *slippery* property in Thesaurus Rex.

   Category selection. One of the categories obtained in the previous step is selected. The system could be parametrized to select a category which contains the target concept (a category that matches one obtained in step 1). It could also parametrized to choose a category in which the target concept is not included (discarding categories that match those obtained in step 1). For the current example, *surface* is supposed to be the selected category.

4. **New query composition.** A new query for Thesaurus Rex is then composed by using the category obtained in the previous step and the modifier selected in step 3. In the current example, we will assume this new query is *slippery surface*.

5. **Final concept selection.** With the query composed in the previous step, we obtain a list of concepts that belong to the category selected in step 5 (*surface*) and at the same time present the property selected in step 3 (*slippery*). This list is usually quite extensive, so the system randomly chooses among the results that have an associated weight among the $N\%$ of concepts with the highest weights (in this example, N = 0.1). In our example, the final concepts associated to the target concept are *satin*, *silk*, *nylon* or *polyester*

Output

The system output is the source concept that gives rise to the rhetorical figure, related through a shared property with the original target concept provided by the user. The shared property is significant in both concepts, which means that the property has a high weight for both of them. The resulting source concept is randomly chosen from the list of generated concepts, and is subsequently used to create a rhetorical figure.

Evaluation

The aim of this evaluation has been twofold. On the one hand, we intended to test the appropriateness of the analogies, similes and metaphors generated by our system, in order for us to be able to refine the process followed to generate them. On the other hand, we also expected to find out what kind of rhetorical figure is more enlightening for the evaluators and which one is closer to a rhetorical figure generated by humans.

Rhetorical Figures Generation using Word Associations

This approach uses the simplest and purest copula form for analogies, similes and metaphors:

- Analogy: *TARGET is as PROP as SOURCE.*

- Simile: *TARGET is like SOURCE.*
- Metaphor: *TARGET is SOURCE.*

Design of the Evaluation

The evaluation set was composed by 36 analogies, 36 similes and 36 metaphors. To create these elements, 36 different words were used as target concepts and one analogy, one simile and one metaphor were created for each of them. In order to avoid the possibility that one evaluator could evaluate several rhetorical figures related to the same target concept, the original data set was divided in three different subsets of 36 rhetorical figures. Each subset had 12 metaphors, 12 similes and 12 analogies, all of them created from a different target concept.

The evaluation was carried out as an online survey using Google Forms, where each evaluator received a link to one of the three surveys and was asked to score each of the figures using a Likert scale. Evaluators were asked to rate how appropriate or natural sounding each trope was, giving them a score from 1 to 7 (where 1 symbolizes a completely inappropriate trope and 7 represents a completely natural sounding trope). We chose to use the median and the mode because when working with the Likert scale, these are the most interesting metrics. Interpreting the average when managing categories such as "totally meaningful" or "totally meaningless", would not provide useful information. Adding the "totally meaningful" value (5) to two "meaningless" values (2) would result in an average of 4, but that is not a very rich interpretation. Traditional statisticians do not recommend using the average of the data in the Likert scale, which offers ordinal values.

In order to have two different baselines in our experiment to measure the quality of the figures generated by our system, we have used a set of commonly accepted rhetorical figures, together with a set of random manually generated ones, to compare them against the ones generated by our system.

The way in which the analogies, similes and metaphors were created was the following:

- Commonly accepted figures: 6 words (3 abstract and 3 concrete) were used as target concepts to obtain commonly accepted metaphors, similes and analogies:
  - TIME: Time is money / Time is like money / Time is as valuable as money
  - KNOWLEDGE: Knowledge is light / Knowledge is like light / Knowledge is as attractive as light
  - ARGUMENT: An argument is a war / An argument is like a war / An argument is as violent as a war
  - BALLERINA: A ballerina is a swan / A ballerina is like a swan / A ballerina is as graceful as a swan
  - STAR: A star is a diamond / A star is like a diamond / A star is as bright as a diamond
  - THUNDER: A thunder is a lion / A thunder is like a lion / A thunder is as mighty as a lion
- Randomly generated figures: 6 words (3 abstract and 3 concrete) were used as target concepts to obtain randomly generated metaphors, similes and analogies:
  - HUNGER: Hunger is knowledge / Hunger is like knowledge / Hunger is as mechanical as knowledge
  - SAILING: Sailing is boyhood / Sailing is like boyhood / Sailing is as allergenic as boyhood
  - SYLLOGISM: A syllogism is a nation / A syllogism is like a nation / A syllogism is as ungulate as a nation
  - ELEPHANT: An elephant is a napkin / An elephant is like a napkin / An elephant is as holy as a napkin
  - CORKSCREW: A corkscrew is a stamp / A corkscrew is like a stamp / A corkscrew is as furry as a stamp
  - TRAIN: A train is a violin / A train is like a violin / A train is as observational as a violin
- Automatically generated figures: 24 words (12 abstract and 12 concrete) were used as target concepts by our system to obtain metaphors, similes and analogies. Half of them were generated with the system configured to obtain the source concept from the same category as the target, and the other half to take the source concept from a different category.
  - Source and target from the same category:
    * WEDDING: A wedding is a party / A wedding is like a party / A wedding is as private as a party
    * WISH: A wish is a desire / A wish is like a desire / A wish is as mental as a desire
    * LIFE: Life is politics / Life is like politics / Life is as complex as politics
    * ANGEL: An angel is a fairy / An angel is like a fairy / An angel is as invisible as a fairy
    * DEVIL: Devil is love / Devil is like love / Devil is as spiritual as love
    * GOVERNMENT: Government is family / Government is like family / Government is as social as family
    * SNOW: Snow is a carpet / Snow is like a carpet / Snow is as soft as a carpet
    * NEEDLE: A needle is a knife / A needle is like a knife / A needle is as sharp as a knife
    * COTTON: Cotton is cashmere / Cotton is like cashmere / Cotton is as natural as cashmere
    * HONEY: Honey is sugar / Honey is like sugar / Honey is as sticky as sugar
    * BATTLE: A battle is a war / A battle is like a war / A battle is as historical as a war
    * WRITER: A writer is a designer / A writer is like a designer / A writer is as creative as a designer
  - Source and target from different categories:
    * SAVING: Saving is farming / Saving is like farming / Saving is as productive as farming
    * ACCIDENT: An accident is an electric shock / An accident is like an electric shock / An accident is as unexpected as an electric shock
    * NETWORK: Network is family / Network is like family / Network is as social as family
    * IDEA: Idea is colors / Idea is like colors / Idea is as abstract as colors

Table 2: Metaphor results.

Source	Mode			Median		
	Abstract	Concrete	**Total**	Abstract	Concrete	**Total**
Random	1	1	**1**	1	1	**1**
Commonly accepted	7	7	**7**	6	5	**5**
Generated (different category)	1	1	**1**	2	2	**2**
Generated (same category)	7	1	**7**	5	4	**5**
Generated	**1**	**1**	**1**	**3**	**3**	**3**

Table 3: Simile results.

Source	Mode			Median		
	Abstract	Concrete	**Total**	Abstract	Concrete	**Total**
Random	1	1	**1**	2	1	**1**
Commonly accepted	7	5	**7**	6	5	**5**
Generated (different category)	1	1	**1**	2	3	**3**
Generated (same category)	7	6	**6**	5	4	**5**
Generated	**1**	**1**	**1**	**4**	**3**	**4**

Table 4: Analogy results.

Source	Mode			Median		
	Abstract	Concrete	**Total**	Abstract	Concrete	**Total**
Random	1	1	**1**	1	1	**1**
Commonly accepted	7	7	**7**	6	6	**6**
Generated (different category)	1	7	**2**	3	4	**4**
Generated (same category)	5	7	**7**	4	4	**4**
Generated	**1**	**7**	**7**	**4**	**4**	**4**

Table 5: General results of the evaluation.

Source	Mode			Median		
	Abstract	Concrete	**Total**	Abstract	Concrete	**Total**
Random	1	1	**1**	1	1	**1**
Commonly accepted	7	7	**7**	6	5	**6**
Generated (different category)	1	1	**1**	2	3	**3**
Generated (same category)	7	7	**7**	5	4	**5**
Generated	**1**	**1**	**1**	**4**	**4**	**4**

∗ ASSEMBLY: An assembly is an aircraft / An assembly is like an aircraft / An assembly is as complex as an aircraft

∗ WINTER: Winter is salad / Winter is like salad / Winter is as cold as salad

∗ MOON: The moon is an halogen lamp / The moon is like an halogen lamp / The moon is as bright as an halogen lamp

∗ REFUGEE: A refugee is an elderly / A refugee is like an elderly / A refugee is as vulnerable as an elderly

∗ TEMPLE: A temple is a school / A temple is like a school / A temple is as public as a school

∗ ACID: Acid is a tiger / Acid is like a tiger / Acid is as dangerous as a tiger

∗ BULLET: A bullet is a bolt / A bullet is like a bolt / A bullet is as metal as a bolt

∗ DRAWER: A drawer is a chesnut / A drawer is like a chesnut / A drawer is as dark as a chesnut

Results of the Evaluation

The evaluation was carried out by 72 evaluators, so that each of the 3 subsets of rhetorical figures was assessed by 24 different evaluators.

The evaluation results for the metaphors are shown in Table 2. Overall the results obtained from the evaluation were as expected, random tropes turned out to be the ones with lower ratings, with a median of 1, and tropes with higher ratings were commonly accepted ones, with a median of 6. Regardless of the type of rhetorical figure and whether they represent specific or abstract concepts, the median of these figures is 4 (3 for those of different categories and 5 for those belonging to the same category).

Interestingly, the modes are the same for abstract and concrete concepts tropes, regardless of how they were generated. The mode of both the random tropes and the tropes generated by our system with different categories is 1, and the mode of the commonly accepted ones and the tropes generated in the same category also matches, with a value of 7.

If we take a closer look at the data subsets of the metaphors, similes and analogies, we can observe that the value of the medians of all figures generated randomly in the three data sets is 1. The results are more satisfactory for the commonly accepted figures, with median values between 5 and 6, proving that the evaluators did not take risks awarding the maximum score.

When we continue to analyze the subsets, we can see that the results obtained for the tropes belonging to different categories are less promising than those obtained for tropes with the same category, with variations between 2 and 4. In the case of the analogies, the median is the same for the ones generated in the same category or in different categories, with a value of 4. The difference between medians of random tropes and commonly accepted tropes fluctuates between 4 and 5.

The graphs show comparative results for the different ways of generating the rhetorical figures: using concrete and abstract concepts, as well as the combined results. The first graph (see Figure 1), corresponds to the word association using abstract concepts and we can observe that the random tropes results are 1 except in the case of the similes median, which is 2. Commonly accepted rhetorical figures mode is 6 and the median is 7. The mode of generated tropes of different categories is always 1 while the median results range between 2 and 3. The mode for generated tropes of same category is between 5 and 7, and the median is between 4 and 5.

Concrete concepts results are shown in Figure 2. Similarly to the abstract concepts, both the results of the mode and the median of random tropes are 1. The mode and the median of commonly accepted rhetorical figures range between 5 and 7. Generated tropes mode of different categories is 1, except for the analogies, which is 7. The median in this case is between 2 and 4. Generated rhetorical figures median of the same category is always 4, while mode is 1 for metaphors, 6 for similes and 7 for analogies.

In Figure 3 the total results for all the rhetorical figures can be seen. Clearly, the result of the randomly generated rhetorical figures is 1. Commonly accepted tropes mode is 7, while the median is 5 for metaphors and similes, and 6 for analogies. The mode of generated tropes of different categories is 1 and 2, and the median is 2 for metaphors, 3 for similes and 4 for analogies. Generated tropes of the same category mode is 7 for metaphors and analogies, while simile mode is 6. The median is 5 for metaphors and similes, and 4 for analogies.

We can conclude that, although the process we have used to generate the rhetorical figures works quite well when concepts of the same category are used, according to the opinions of the evaluators, something different happens in the case of using concepts that belong to different categories, which, in general, obtain worse results. This fact points to the need of using additional properties or relationships in order to obtain concepts that can subsequently give rise to more meaningful rhetorical figures.

Discussion

As we can see, in all cases the randomly generated metaphors are rated as meaningless by the evaluators. In contrast, commonly accepted metaphors get the highest results, with a slight preference for the metaphors created using abstract concepts over the ones that are based on the use of concrete concepts. The automatically generated metaphors using concepts of different categories are also poorly rated, which points out that sharing only one property is not enough to generate a good metaphor. For the generated metaphors using concepts that belong to the same category, the difference that exists between the modes of the metaphors that use concrete and abstract concepts is remarkable. This suggests that abstract metaphors are more evocative and offer a wider range of interpretations than concrete ones. Finally, the overall median for the metaphors also suggests that more aspects need to be taken into consideration to increase the perceived quality of these rhetorical figures.

Table 3 shows the results for the evaluated similes. The ratings in this case are quite similar to the results obtained for the metaphors.

The results of the evaluation of the analogies can be seen in Table 4. The ratings in this case are slightly higher than in the two previous tropes, probably due to the fact that the aspect in which the two concepts are considered to be similar is explicitly stated. This same aspect may be the cause for the lower score obtained by the automatically generated analogies using abstract concepts that belong to the same category. In this case, the similarity perceived by the evaluators may be focused on a different characteristic than the one chosen by the system, which causes the score to be lower than the one granted to the previous figures. On the contrary, the analogies generated using concrete concepts belonging to different categories are much better rated than in the previous tropes. In this case, the reason seems to be the fact that the property used by the system to compare both concepts has been made explicit, so the evaluators can see the reason why the system considers the two concepts related to each other and they are more inclined to accept it as valid.

Finally, the overall results of the evaluation can be seen in Table 5. Although they don't differ much from the results obtained for the different tropes independently, the values of the modes are clearly shifted towards the limits of the scale. This effect suggests that human evaluators tend to accept or not accept a rhetorical figure as valid, but intermediate positions are less common. As for the value of the medians, the condensed results confirm the perception that, in terms of automatically generated tropes, the ones that use abstract concepts that belong to the same category are slightly better appreciated than the rest.

Conclusions and Future Work

We have proved that it is possible to evaluate the quality of rhetorical figures and get consistent results. One of the clearest conclusions is that in our system concepts generate tropes of the same category with significantly higher quality than the tropes based on concepts of different categories.

In view of the results, one of the paths we have to follow

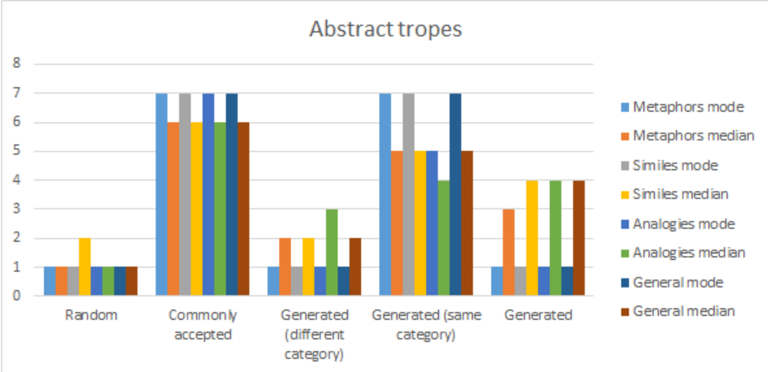

Figure 1: Abstract Tropes

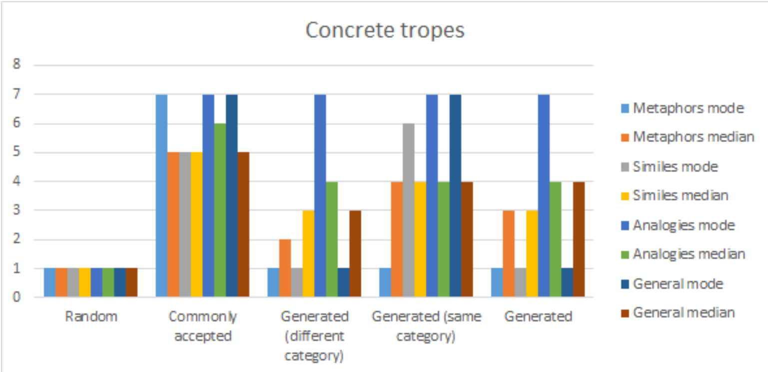

Figure 2: Concrete Tropes

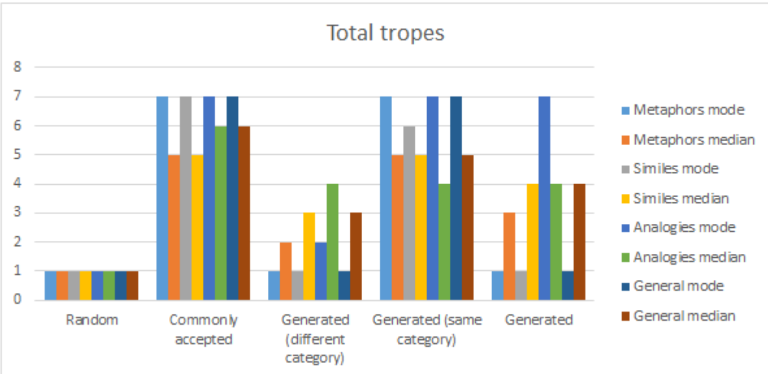

Figure 3: Total Tropes

is directed to find ways to generate good rhetorical figures from concepts of different categories, because in everyday life some of the best rhetorical figures are constructed from these kinds of terms, such as *time is money*. The categories with higher weights obtained for the concept *time* in Thesaurus Rex are *information, quantity and attribute*, while for the concept *money* they are *thing, property, value and assets*. As we have seen in the evaluation, this trope gets a good rating and we need more information about this type of rhetorical figures.

In order to generate appropriate figurative language depending on the content of a given text, it would be interesting to find sets of words grouped by topic. On the other hand, in order to adequate the figurative language to the goals of the reader, it would be helpful to have sets of concepts grouped by the complexity of their meaning.

Sometimes, constraints encountered arise from the web itself. This is because the information usually available on the web tends to be more literal than figurative. For the previous example, the attributes with more weight obtained from

Thesaurus Rex when searching for *time* are *physical, basic, measurable, relevant and abstract*. That suggests that it may be more appropriate for us to find or generate a specific knowledge resource that provide more evocative properties.

The highest mode for rhetorical figures generated by our system are obtained for analogies. In the case of the median of the total result tendencies are less clear. While the rhetorical figures in the same category produce better results in metaphors and similes, rhetorical figures with different category get better valuations in analogies.

In the future, we would like to continue doing assessments to find patterns or similarities among the best rated rhetorical figures, and we wish to test this with larger datasets. Thus the evaluation findings could serve to improve the quality of the resources generated by our system.

We have used the terms ”concrete” and ”abstract” when categorising input concepts. It would be interesting to check whether it makes a difference to use a concrete word to describe an abstract concept (e.g. ”time is money”) and viceversa.

As future work we would also like to check the degree of similarity between the source and target concept. If the concepts are too similar, the resulting trope would be correct but not very practical.

With respect to the amount of information provided in the rhetorical figure, there are no significant differences between those that provide more or less information, because similar results are obtained for metaphors – in which only the original concept and the new concept are indicated – and analogies – in which the shared attribute is also shown.

The results obtained indicate that further attempts should be made to evolve our system and generate higher quality rhetorical figures, progressively evolving the quality of system results towards that of rhetorical figures generated by people. In the future, a useful feature that may improve our system is to relate the original concept with concepts that have more than one property in common. From now on another way that we should investigate is to generate rhetorical figures with concepts that are related through two or more attributes. In the example *A ballerina is a swan*, both concepts share properties as *pretty, graceful and stylized*.

Acknowledgements

This work is funded by ConCreTe. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733.

References

Black, M. 1955. XII.Metaphor. *Proceedings of the Aristotelian Society* 55(1):273–294.

Harmon, S. 2015. Figure8: A novel system for generating and evaluating figurative language. In *Proceedings of the Sixth International Conference on Computational Creativity*, 71–77.

Hervás, R.; Camara Pereira, F.; Gervás, P.; and Cardoso, A. 2006a. Cross-domain analogy in automated text generation. In *3rd Joint Workshop on Computational Creativity, dentro de la 17th European Conference on Artificial Intelligence*, 43–48.

Hervás, R.; Camara Pereira, F.; Gervás, P.; and Cardoso, A. 2006b. A text generation system that uses simple rhetorical figures. *Procesamiento de Lenguaje Natural* 37:199–206.

Hervás, R.; Costa, R.; Costa, H.; Gervás, P.; and Camara Pereira, F. 2007. Enrichment of automatically generated texts using metaphor. In *6th Mexican International Conference on Artificial Intelligence (MICAI-07)*, volume 4827, 944–954. Aguascalientes, Mexico: Springer Verlag, LNAI Series.

Lakoff, G. 1993. The Contemporary Theory of Metaphor. *Metaphor and Thought* 202–251.

Shutova, E.; de Cruys, V.; T.; and Korhonen, A. 2012. Unsupervised Metaphor Paraphrasing using a Vector Space Model. In *Proceedings of the 24th International Conference on Computational Linguistics*, 1121–1130.

Shutova, E.; Sun, L.; and Korhonen, A. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics*, 1002–1010. Beijing: Association for Computational Linguistics.

Veale, T., and Hao, Y. 2011. Exploiting Readymades in Linguistic Creativity: A System Demonstration of the Jigsaw Bard. In *ACL (System Demonstrations)*, 14–19.

Veale, T., and Li, G. 2012. Specifying Viewpoint and Information Need with Affective Metaphors: A System Demonstration of the Metaphor Magnet Web App/Service. In *Proceedings of the ACL 2012 System Demonstrations*, 7–12.

Veale, T., and Li, G. 2013. Creating Similarity: Lateral Thinking for Vertical Similarity Judgments. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 660–670.

Veale, T. 2014a. A Service-Oriented Architecture for Metaphor Processing. In *Proceedings of the Second Workshop on Metaphor in NLP*, 52–60.

Veale, T. 2014b. A service-oriented architecture for metaphor processing. In *Proceedings of the Second Workshop on Metaphor in NLP*, 52–60. Baltimore, MD: Association for Computational Linguistics.

Wallington, A.; Barnden, J.; Buchlovsky, P.; Fellows, L.; and Glasbey, S. 2003. Metaphor annotation: A systematic study. CSRP 03-04, School of Computer Science, The University of Birmingham, Birmingham.

Wilks, Y.; Galescu, L.; Allen, J.; and Dalton, A. 2013. Automatic metaphor detection using large-scale lexical resources and conventional metaphor extraction. In *Proceedings of the First Workshop on Metaphor in NLP*, 36–44. Atlanta, GA: Association for Computational Linguistics.