Slogans Are Not Forever: Adapting Linguistic Expressions to the News

Lorenzo Gatti
FBK-irst
Trento, Italy
l.gatti@fbk.eu

Gözde Özbal
FBK-irst
Trento, Italy
gozbalde@gmail.com

Marco Guerini
Trento RISE
Trento, Italy
marco.guerini@trentorise.eu

Oliviero Stock
FBK-irst
Trento, Italy
stock@fbk.eu

Carlo Strapparava
FBK-irst
Trento, Italy
strappa@fbk.eu

Abstract

Artistic creation is often based on the concept of blending. Linguistic creativity is no exception, as demonstrated for instance by the importance of metaphors in poetry. Blending can also be used to evoke a secondary concept while playing with an already given piece of language, either with the intention of making the secondary concept well perceivable to the reader, or instead, to subtly evoke something additional. Current language technology can do a lot in this connection, and automated language creativity can be useful in cases where input or target are to change continuously, making human production not feasible. In this work we present a system that takes existing well-known expressions and innovates them by bringing in a novel concept coming from evolving news. The technology is composed of several steps concerned with the selection of the sortable concepts and the production of novel expressions, largely relying on state of the art corpus-based methods. Proposed applications include: i) producing catchy news headlines by "parasitically" exploiting well known successful expressions and adapting them to the news at hand; ii) generating adaptive slogans that allude to news of the day and give life to the concept evoked by the slogan; iii) providing artists with an application for boosting their creativity.

1 Introduction

NAMING PRIVATE RYAN, headline of the Daily Mirror referring to football player Ryan Giggs, who was mentioned in the British House of Lords in connection with new legislation.

A given expression may be given novel life if adapted to ever changing input. For instance, the revised expression may evoke some of the news of the day, while keeping the original expression still perceivable. The news of the day, which obviously cannot be predicted in advance, can be promoted through a creative tagline based on the parasitic use of an automatically selected linguistic expression (for instance a slogan, movie title or well known quote). To achieve that, the expression can be slightly modified into a novel one that evokes the news by still winking to its origin.

While the example reported in the opening was the clever one-off creation of a professional, we can conceive frequent adaptation of given expressions to incoming news. However, if input (or target of the communication) are to change continuously, human production is not realistic, even for limited interventions and only a creative system can do the job.

Artistic creation is often based on the concept of blending. At the macro level, a story may integrate suspense and psychological theory, a film may speak about the story of an individual while emphasizing nature description. Looking closely at linguistic creativity, we can also consider a form of blending at the micro level. In this case the focus is just on combining a few words, or substituting words from a given expression, producing a correct expression, meaningful and effective for its desired impact on the audience. As Veale [2012] says "We use linguistic creativity to re-invent and re-imagine the familiar, so that everything old can be made new again." Poetry owes a lot to the introduction of metaphors and in general to innovative combination of predicate and arguments. Together with the euphonic properties of produced language the blended results are meant to produce an effect on the reader's cognition. In addition, if we look at "minor", circumscribed forms of creativity, we can observe how relevant this sense of blending is: take for instance ad-

This work has been partially supported by FBK and Trento RISE PerTe project.

vertisements or production of news headlines.

"Micro" blending can also be at the basis of related but slightly different concepts. Language creativity can have the goal of evoking a secondary concept in a subtle way while playing with an already given piece of language. This process can serve the goal of making the secondary concept well perceivable to the reader, and even dominating the scene, in a way exploiting the properties of the initial language to give visibility to the new concept. Or instead, it can be used for subtly evoking something additional while focusing on the initial concept.

Current language technology can do a lot in this connection, especially since it can exploit a number of linguistic resources and it proposes basic corpus-based and rule-based techniques that can be adapted and used in an original way. In the present work, we propose a framework for automatically integrating into a given expression words that evoke the secondary expression taken from unpredictable news. We use semantic similarity metrics to pair an expression with an appropriate news article. In addition, we use morpho-syntactic constraints and the dependency structure of the expression to obtain meaningful and grammatical sentences. As for aesthetics and the involved cognitive aspects, our "ideological" reference is the so called Optimal Innovation Theory, which implies that a given variation is more highly appreciated if the change is limited [Giora, 2003]. To the best of our knowledge, this is a novel attempt to blend well known expressions with recent news in a linguistically motivated framework that accounts for syntagmatic and paradigmatic aspects of language. With this study we also carry out a crowdsourced task to investigate the effectiveness of our approach for evoking a novel concept starting from a familiar expression.

To demonstrate the usefulness of the developed technology, it is worth mentioning three applied scenarios, all requiring some artistic creativity. Humans who perform such tasks (but without the challenge of continuous adaptation, as pointed out above) are normally selected for their "creativity" and their "artistic capabilities" in producing similar short linguistic expressions.

In the first scenario, as in our initial example, our system can be used by a newspaper to produce "catchy" headlines given a short description of a news article. It is well known that a catchy and sensational title in many cases is a fundamental prerequisite for the success of news [Schultz, 2009].

In the second scenario an advertisement company uses the system to adapt their slogan to the news of the day to attract attention. This is quite innovative with respect to current "stiff" technologies for contextual advertising, where the textual part of the ad is fixed and only matched against a relevant page via semantic similarity on a bunch of predefined set of keywords – see for example [Broder *et al.*, 2007; Chakrabarti *et al.*, 2008]. A very simple case of slogan renewal actually used in advertisement is *Absolut* ⟨*noun*⟩ (e.g., "Absolut Mexico", "Absolut mango", "Absolut chaos"), which is complemented by a picture of a bottle adapted to the new wording[1], so keeping *Absolut Vodka* well perceivable. As for our system, our ambition is to perform clearly more meaningful and sophisticated automated adaptation, evoking ever changing news (or, for that matter, concepts coming from any novel short text).

A third possible scenario can be envisaged for artists who could use the system for proposing variations of given expressions that evoke some external unpredictable news. Various experiments can be made, including letting novel words substitute or integrate words in the original expression and letting the original concept fade or reappear with new force. An example can be playing with a well known love-related quote, such as "love is a game where both players win", substituting the keyword *love* with appropriate words evoking the news, for instance "organising Olympics is a game where both players win".

The rest of the paper is organized as follows. We will first assess related research, then we will describe the system and the technology involved. Following that we will explain the system in use by following a detailed example. We will then summarize the results of a crowdsourcing-based evaluation study. Finally, we will draw our conclusions and outline possible future directions.

2 Related Work

Poetry generation systems face similar challenges to ours as they struggle to combine semantic, lexical and phonetic features in a unified framework. Greene et al. [2010] describe a model for poetry generation in which users can control meter and rhyme scheme. Generation is modeled as a cascade of weighted Finite State Transducers that only accept strings conforming to a user-provided desired rhyming and stress scheme. The model is applied to translation, making it possible to generate translations that conform to the desired meter. Toivanen et al. [2012] propose to generate novel poems by replacing words in existing poetry with morphologically compatible words that are semantically related to a target domain. Content control and the inclusion of phonetic features are left as future work and syntactic information is not taken into account. The Electronic Text Composition project[2] is a corpus based approach to poetry generation which recursively combines automatically generated linguistic constituents into grammatical sentences. Colton et al. [2012] present another data-driven approach to poetry generation based on simile transformation. The mood and theme of the poems are influenced by daily news. Constraints about phonetic properties of the selected words or their frequencies can be enforced during retrieval. After presenting a series of web services for novel simile generation, divergent categorization, affective metaphor generation and expansion, Veale [2014] introduces *Stereotype*, a service for metaphor-rich poem generation. Given a topic as input, *Stereotype* utilizes the previous services to obtain a master metaphor, its elaborations, and proposition-level world knowledge. All these ingredients are then packaged by the service into a complete poem.

Recently, some attempt has been made to generate creative sentences for educational and advertising applications. Özbal et al. [2013] propose an extensible framework called BRAINSUP for the generation of creative sentences in which users

[1] Refer to www.absolutad.com/absolut_gallery/ for examples

[2] http://slought.org/content/11199

are able to force several words to appear in the sentences. BRAINSUP makes heavy use of syntactic information to enforce well-formed sentences and to constraint the search for a solution, and provides an extensible framework in which various forms of linguistic creativity can easily be incorporated. An extension of this framework is used in a more recent study [Özbal *et al.*, 2014] to automate and evaluate the keyword method, which is a common technique to teach second language vocabulary. The effectiveness of the proposed approach is validated by the extrinsic evaluation conducted by the authors.

As a notable study focusing on the modification of linguistic expressions, the system called Valentino [Guerini *et al.*, 2011] slants existing textual expressions to obtain more positively or negatively valenced versions by using WordNet [Miller, 1995] semantic relations and SentiWordNet [Esuli and Sebastiani, 2006]. The slanting is carried out by modifying, adding or deleting single words from existing sentences. Insertion and deletion of words is performed by utilizing Google Web 1T 5-Grams Corpus [Brants and Franz, 2006] to extract information about the modifiers of terms based on their part-of-speech. The modification is performed first based on the dependents from left to right and then possibly the head. Valentino has also been used to spoof existing ads by exaggerating them, as described in [Gatti *et al.*, 2014], which focuses on creating a graphic rendition of each parodied ad.

Lexical substitution has also been commonly used by various studies focusing on humor generation. Stock and Strapparava [2006] generate acronyms based on lexical substitution via semantic field opposition, rhyme, rhythm and semantic relations provided by WordNet. The proposed model is limited to the generation of noun phrases. Valitutti et al. [2009] present an interactive system which generates humorous puns obtained by modifying familiar expressions with word substitution. The modification takes place considering the phonetic distance between the replaced and candidate words, and semantic constraints such as semantic similarity, domain opposition and affective polarity difference. More recently, Valitutti et al. [2013] propose an approach based on lexical substitution to introduce adult humor in SMS texts. A "taboo" word is injected in an existing sentence to make it humorous. The substitute is required to have the same part-of-speech with the original word and to have a high cohesion with the previous word. The latter check is performed by using n-grams. Petrović and Matthews [2013] generate jokes by filling in a pre-defined template with nouns and adjectives mined from big data. The words used in the jokes are selected according to distributional criteria which, combined with the structure of the template, induce the humorous effect through incongruity.

Finally, it is worth mentioning that many creative systems are based on the Conceptual Blending Theory [Fauconnier and Turner, 2008], a framework for mapping two concepts from different domains into a new "blended space", which inherits properties from both starting domains. In our work, however, we insert a new concept into an existing expression by replacing a word, without modeling the starting domains and finding their shared properties.

3 System Description

An overview of the contextualization process performed by our system can be seen in Figure 1. The system mainly consists of four modules for (*i*) getting the news of the day from the web, (*ii*) extracting the most important keywords from the news and expanding the keyword list by using two resources (i.e., a lexical database and a knowledge base), (*iii*) pairing the news and the well known expressions in our database by using state of the art similarity metrics, (*iv*) contextualizing the well-known expression according to the news by satisfying the lexical and morpho-sytanctic constraints enforced by the original expression.

In the rest of this section, we provide the description of each module in more detail. We will consider only the second applied scenario (i.e. adapting a slogan to the news of the day), but the process is identical for the others.

3.1 Getting the news of the day

The important news of the day are retrieved from the RSS feed of BBC News[3]. Each entry of the feed is composed of a headline, a short (about 20 to 30 words) description of the article and other metadata, such as a link to the article and the publication date, which we currently do not use. News are sorted from the most recent to the oldest.

3.2 Extracting/expanding the keywords

To be able to extract the important words in a piece of news, we simply calculate the probabilities of each lemma in a news corpus (23,415 news documents from LDC Gigaword 5th Edition corpus[4]). Each probability of a lemma is calculated as the number of headlines that it appears divided by the total number of headlines occurring in the corpus.

Then, we tokenize and PoS-tag the daily snippets with Stanford Parser [Klein and Manning, 2003] and lemmatize them with WordNet. We filter out the stop words and for each snippet, we extract the following ingredients:

1. n lemmas with the lowest probability values
2. Synonyms and derivationally related forms of the lemmas extracted in Step 1 (using WordNet)
3. Named entities by using Stanford Named Entity Recognizer [Finkel *et al.*, 2005].
4. For each named entity obtained in Step 3, we use FreeBase [Bollacker *et al.*, 2008], which is a large online knowledge base storing general human knowledge, to obtain *notable for*, *notable type* and *alias* information. To illustrate, for *Beyoncé Knowles* we retrieve *Bee* as her alias and we get the information that she is notable for *dance-pop* as a *celebrity*.

3.3 Blending slogans with news

In state of the art creative systems based on lexical substitution, the insertion of new "ingredients" in a given expression is generally easy, as it is normally limited to satisfying only

[3] http://feeds.bbci.co.uk/news/rss.xml

[4] http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2011T07

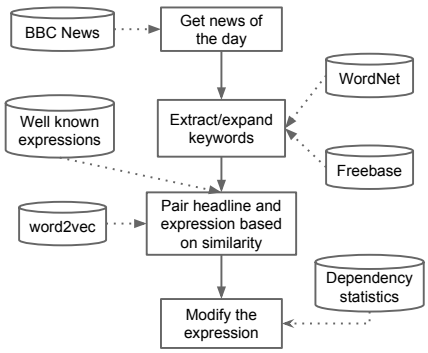

Figure 1: An overview of the contextualization process.

part-of-speech constraints, with n-gram counts being sometimes used for ensuring cohesion with adjacent words.

In our case we adopted a more refined mechanism. First, a similarity check is performed between the slogan and the news headline together with its description. This way we minimize the risk of generating nonsense output in which concepts with no relation are merged. Moreover, the cohesion between the replaced word and the rest of the sentence is based on all the syntactic relations that the word is involved in, not just on the adjacent tokens.

Sorting by similarity. The system calculates the similarity between a slogan and the news, using a skip-gram model [Mikolov *et al.*, 2013] trained on the lemmas of the GigaWord corpus. To compare the slogan with the news, we construct a vector representation of the former by summing the vectors of its lemmas (after filtering out stop words). The same is done for each news article, using lemmas present in the headline and description. Then, each article is sorted according to its cosine distance from the slogan. Moreover, articles that do not reach a certain similarity threshold are discarded. This not only allows us to get the n most similar news for any given slogan, but also ensures at least a minimum degree of relatedness between the two categories.

Dependency Statistics. To be able to satisfy the lexical and syntactic constraints imposed by the original expression, we follow a similar approach to Özbal et al. [2013], using a database of tuples that stores, for each relation in the dependency treebank of LDC GigaWord corpus, its occurrences with specific "governors" (heads) and "dependents" (modifiers). From the raw occurrences, the probability that word d and its governor g are connected by relation r is estimated as:

$$p_r(g,d) = \mathrm{c}_r(g,d)/(\sum_{g_i}\sum_{d_i}\mathrm{c}_r(g_i,d_i))$$

where $c_r(\cdot)$ is the number of times that d depends on g in the dependency treebank, and g_i, d_i are all the governor/dependent pairs observed in the treebank.

So, for each adjective, noun, adverb and verb w in a slogan, we determine all the words that are connected to w in a dependency relation. Then, we calculate how likely each keyword k, coming from the news articles that passed the similarity filter, can replace a w of the same part-of-speech. The dependency-likelihood of replacing w with k in a sentence $\vec{s}$ is calculated as:

$$\mathrm{f}(\vec{s},w,k) = \exp\left(\frac{\sum_{\langle g,d,r\rangle\in \mathrm{r}(\vec{s},w)}\log p_r(g,h)}{|r(\vec{s},w)|}\right)$$

with $r(\vec{s},w)$ being the set of dependency relations in $\vec{s}$ that involve w. We can then select the slot with the word w to be replaced, and the best keyword k for each news article, by simply maximizing this dependency likelihood.

Also in this case a threshold is enforced, so that sentences that do not reach a satisfactory level of grammaticality are removed, and the resulting outputs are sorted according to their dependency score.

Most of the ingredients derived from the news are nouns and verbs, thus most replacements affect these parts of speech, but a small percentage of adjectives and adverbs are also selected. Although we could also implement mechanisms for inserting these as new modifiers, we think the results would probably not be very interesting. Most of the meaning in a sentence is given by nouns and verbs, thus the insertion of a single modifier might not be enough to relate with the article. Looking at real-world examples we see that this is also what happens when newspapers or advertisers use lexical substitution for their work.

Final ranking. If the modification is successful (i.e., if not all the sentences were discarded by the similarity and dependency filters) the morphology of the replaced word w is applied to k by using MorphoPro [Pianta *et al.*, 2008] and the modified sentence is generated.

To select the final output, the system sorts each modified sentence according to its mean rank with respect to similarity and dependency scores, thus balancing the scores of grammaticality and relatedness to the news. The sentence with the lowest mean is chosen as the best one and presented to the user.

3.4 Data

In our experiments, we contextualize three types of data with news, namely cliches, and movie and song titles and advertisement slogans. We choose to experiment with these data as they are all likely to be rich in terms of creativity. In addition, this setting allows us to investigate the effectiveness of our system on expressions belonging to different domains and with varying length and syntax.

Cliches: We downloaded 2002 cliches from ClichéSite[5]. They were then ranked using the number of results they had in the Bing search engine, to maximize the likelihood that the original expression is known, and only the top 25% was kept. We also removed the cliches with less than 4 tokens, to avoid very short expressions which did not have enough content and could hardly be recognized after modification. The final dataset consists of 287 cliches.

Movie titles: The Top-250 titles collected from the Internet Movie Database[6]. Again, we removed very short expressions (i.e. less than 3 tokens), for a total of 64 movie titles.

[5] http://clichesite.com

[6] www.imdb.com/chart/top

Song titles: A dataset composed of 112 song titles downloaded from lists of "best selling" and very famous songs[7] and processed as the previous two lists.

Slogans: 2,046 slogans with an average token count of 6. These slogans were collected from online resources[8] for products from various domains (e.g., *beauty*, *hygiene*, *technology*).

3.5 A worked out example

Let us consider how the process works, from input selection to the generation of the final modified sentence, with an example taken from real news and a slogan.

We start from the slogan "Make a run for the border", used by Taco Bell, and parse it. A vector $\vec{s}$ is then created by adding the vectors of `make#v`, `run#n` and `border#n`, i.e., the lemmas and relative parts of speech that compose the sentence after excluding the stop words.

similarity	Headline
0.747	US air strike 'hits IS command'
0.710	Governor quits over Mexico missing
0.706	Mums on 'diabolical' holiday fines
0.705	Standing ovation for Kevin Vickers
0.702	The day UFOs stopped play
0.687	Mortar fire on Pakistan-Iran border

Table 1: News sorted according to their similarity with the slogan "Make a run for the border"

In the same way vectors are created from each news headline and description, and their cosine similarity to the slogan vector are calculated. The system takes up to 10 most similar articles among the ones with a similarity value greater than a given threshold[9]. Table 1 includes a subset of the news selected in this case.

We can notice how the skip-gram model picks up the similarity between `border#n` and countries that are often connected with this concept: US, Mexico and Canada (which is mentioned in the description of the fourth article).

Then, the system computes the dependency likelihood of modifying each word in the slogan with each keyword obtained from the news. For example, in the second headline the system attempts to modify the appropriate slots in the sentence with `governor#n`, `govern#v`, `gubernatorial#a`, `Mexico#n`, and many other keywords. With the help of the dependency likelihood scoring, the system favors replacing `border#n` with `governor#n` instead of `make#v` with the keyword `govern#v` (generating "Govern a run for the border") as the latter is less likely (i.e. fewer relations are satisfied).

[7]Such as http://en.wikipedia.org/wiki/List_of_songs_considered_the_best, and the Top-20 for each decade after the 60's from www.billboard.com

[8]such as http://slogans.wikia.com/wiki and http://www.adslogans.co.uk/

[9]The similarity filter was set to $\cos > 0.6$. Threshold values were determined empirically.

Expression	What the world is coming to
Headline	UK anger at 1.7bn EU cash demand
Output	What the Euro is coming to
Expression	Unleash the power of the sun
Headline	Wood: Time for Wales to step up
Output	Unleash the power of the majority
Expression	They got the power to surprise!
Headline	The day UFOs stopped play
Output	They got the power to halt!
Expression	Make love not war
Headline	Women propose sex strike for peace
Output	Make peace not war
Expression	Welcome to the world
Headline	UK anger at 1.7bn EU cash demand
Output	Contribute to the world
Expression	Smells like teen spirit
Headline	Suicide mother gave girls acid drink
Output	Inquests like teen spirit
Expression	Raid kills bugs dead
Headline	Sinai bomb kills 25 soldiers
Output	Raid kills soldiers dead

Table 2: Output examples

Finally the articles are sorted according to their dependency scores and those that do not reach a threshold are discarded. This leaves us with "Make a run for the governor", and "Make a run for the parliament", which is generated by replacing `border#n` with `parliament#n`, a keyword coming from "Standing ovation for Kevin Vickers".

The article about the governor is the second most similar to the slogan, and is the one with the best dependency. Based on these ranks, the first article receives a score of $(2+1)/2 = 1.5$, while the other, being fourth for similarity and second for dependency, receives a score of $(4+2)/2 = 3$. Therefore, the system considers "Make a run for the governor" the best output for the given slogan.

It is interesting to notice how the dependencies exploit the existing syntax and change the literal phrase "running for ⟨*geographical place*⟩" into the metaphorical "running for ⟨*position*⟩" for an article concerning a resigning governor, creating a witticism "more funny and more resonant than its author had first intended" [Veale, 2012].

More examples of the system output can be seen in Table 2. Some of the results are particularly interesting. For instance, "What the Euro is coming to" is a potentially creative headline reflecting on the financial situation in Europe, in addition to being perfectly fit for a tabloid.

In the used configuration the system does not include a mechanism for filtering out inappropriate output, such as the last entry in the table, which is exposing us to the risk of inadvertently producing "black humor" or other inappropriate headlines. To prevent the generation of trivial expressions such as "Make peace, not war" from "Make love, not war", another filtering mechanism can be added with the help of a simple frequency check of the output on a large corpus. Parsing errors are another cause of inaccurate outputs as shown

Strategy 1	Strategy 2	Grammatical	Meaningful	Related
PoS	Dep	81%	88%	69%
PoS	Sim	88%	88%	88%
PoS	Dep+Sim	88%	81%	88%
Sim	Dep+Sim	75%	63%	50%

Table 3: Percentage of cases in which Strategy 2 improves over Strategy 1

Strategy 1	Strategy 2	Effectiveness 1	Effectiveness 2
PoS	Dep	13%	44%
PoS	Sim	6%	50%
PoS	Dep+Sim	13%	31%
Sim	Dep+Sim	25%	38%

Table 4: Pairwise strategy comparison in terms of effectiveness

by the example "Smells like teen spirit", where "smells" is identified as a plural noun instead of a verb and is replaced with the uncommon -but still uninteresting- word "inquests".

4 Evaluation

To compare the different strategies used to improve the output quality and to assess the overall effectiveness of the proposed approach, we carried out an evaluation of the system by using the CrowdFlower[10] crowdsourcing platform. All the annotators were selected from English speaking countries.

We presented the annotators a news headline, its description and two modified expressions generated by using different strategies on the sentences of in data . Each sentence was modified using either (*i*) simple PoS-replacement, (*ii*) PoS-replacement after filtering the expressions with a similarity threshold, (*iii*) dependency-based filtering with no similarity check, (*iv*) the full mechanism described in 3.3. We compared 16 pairs of sentences for each strategy, each one annotated by at least three "turkers" (average number of annotators: 4.5 for each pair), using majority voting.

The annotators were asked to what extent each of the two sentences could be used as a headline for the article ("effectiveness" from now on), using a 5-point Likert scale. We also asked them to state which of the two expressions was more grammatically correct, which made more sense (we define this as the "meaningfulness" of the expression) and which one was more related to the article.

Each evaluator annotated 8 such news articles and expression tuples, and some test questions to filter out inaccurate workers. These test questions were created by manipulating the grammaticality and relatedness of a few regular questions in advance.

In Table 3, for each dimension we report the percentage of times the expression generated by a "smart" strategy (i.e., Strategy 2) was rated as being better than a simpler strategy (i.e., Strategy 1: either simple PoS-based replacement, or PoS-based replacement after sorting the expressions by similarity). For example, the first row tells us that Dep was found to be more grammatical than PoS in 81%, more meaningful in 88% and more related to the news in 69% of the cases. In Table 4, the strategies are compared in terms of effectiveness calculated as the percentage of times each strategy received a rating in the top-2 levels of the Likert scale.

Not surprisingly, the results show that all the strategies are better than simple PoS-replacement. The use of dependency scores improves on all the dimensions, although the effect is stronger in grammaticality and meaningfulness. The small improvement in relatedness could be due to the fact that using the context to restrict possible modifications helps to naturally select similar things. Although the simple use of similarity, as shown in the second row, improves on all the dimensions, the comparison between Sim and Dep+Sim clearly shows that dependencies ensure a higher level of grammaticality of the output and an improvement in the overall meaningfulness of the resulting sentences.

More in general, in Table 4 we can see that the effectiveness ratings are consistent with the other dimensions: both similarity filtering and dependency scoring give better results than simple PoS-replacement, and the combination of the two futher improves over similarity.

5 Conclusions

In this paper, we presented a system that utilizes various NLP techniques to innovate existing well-known expressions by bringing in a new concept coming from evolving news. Our study can be considered as a novel attempt to blend well known expressions with recent news in a linguistically motivated framework that accounts for syntagmatic and paradigmatic aspects of language. The evaluation that we carried out confirms the effectiveness of our approach for evoking a novel concept by modifying these well known expressions. While additional experiments and more data are needed to determine the real magnitude of these observations, the initial results are promising. We believe that effectiveness is correlated with the aesthetic pleasure involved in the appreciation of the modification. In any case, the automation of this kind of creative process can be of great use for innovative scenarios concerning applied arts.

As future work, we would like to experiment with the insertion of verbs coming from the news into the original expression. We expect this task to be quite challenging as verbs can take role in a relatively large number of dependency relations and modification of a verb might have a strong impact on the semantics of the whole sentence. In addition, for specific applications, we plan to exploit semantic incongruity during the blending process to produce humorous sentences. In a further evaluation we will ask annotators to indicate whether the original expression can be a better headline than the modified one. The data will then be used to identify the cases where the system should just "recycle" a well-known expression, instead of modifying it. We will also apply state-of-the-art sentiment analysis techniques to filter out very negative news to minimize the risk of generating insensitive or malicious expressions. Finally, we will improve the sorting mechanism with the addition of a memorability score for the modified expression.

[10] www.crowdflower.com

References

[Bollacker *et al.*, 2008] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of SIGMOD'08*, pages 1247–1250. ACM, 2008.

[Brants and Franz, 2006] T. Brants and A. Franz. Web 1T 5-gram version 1. Linguistic Data Consortium, 2006.

[Broder *et al.*, 2007] A. Broder, M. Fontoura, V. Josifovski, and L. Riedel. A semantic approach to contextual advertising. In *Proceedings of SIGIR'07*, pages 559–566, 2007.

[Chakrabarti *et al.*, 2008] D. Chakrabarti, D. Agarwal, and Vanja Josifovski. Contextual advertising by combining relevance with click feedback. In *Proceedings of WWW'08*, pages 417–426, 2008.

[Colton *et al.*, 2012] Simon Colton, Jacob Goodwin, and Tony Veale. Full-FACE poetry generation. In *Proceedings of ICCC'12*, pages 95–102, 2012.

[Esuli and Sebastiani, 2006] Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC'06*, pages 417–422, 2006.

[Fauconnier and Turner, 2008] Gilles Fauconnier and Mark Turner. *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books, 2008.

[Finkel *et al.*, 2005] Jenny Rose Finkel, Trond Grenager, and Christopher D. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of ACL'05*, pages 363–370, 2005.

[Gatti *et al.*, 2014] Lorenzo Gatti, Marco Guerini, Oliviero Stock, and Carlo Strapparava. Subvertiser: mocking ads through mobile phones. In *Proceedings of IUI'14*, pages 41–44, 2014.

[Giora, 2003] Rachel Giora. *On Our Mind: Salience, Context and Figurative Language*. Oxford University Press, New York, 2003.

[Greene *et al.*, 2010] Erica Greene, Tugba Bodrumlu, and Kevin Knight. Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of EMNLP'10*, pages 524–533, 2010.

[Guerini *et al.*, 2011] Marco Guerini, Carlo Strapparava, and Oliviero Stock. Slanting existing text with Valentino. In *Proceedings of IUI'11*, pages 439–440, 2011.

[Klein and Manning, 2003] Dan Klein and Christopher D Manning. Accurate unlexicalized parsing. In *Proceedings of ACL'03*, pages 423–430, 2003.

[Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS'13*, pages 3111–3119, 2013.

[Miller, 1995] George A. Miller. WordNet: A lexical database for English. *Communications of the ACM*, 38:39–41, 1995.

[Özbal *et al.*, 2013] Gözde Özbal, Daniele Pighin, and Carlo Strapparava. BRAINSUP: Brainstorming support for creative sentence generation. In *Proceedings of ACL'13*, pages 1446–1455, 2013.

[Özbal *et al.*, 2014] Gözde Özbal, Daniele Pighin, and Carlo Strapparava. Automation and evaluation of the keyword method for second language learning. In *Proceedings of ACL'14*, pages 352–357, 2014.

[Petrović and Matthews, 2013] Saša Petrović and David Matthews. Unsupervised joke generation from big data. In *Proceedings of ACL'13*, pages 228–232, 2013.

[Pianta *et al.*, 2008] Emanuale Pianta, Christian Girardi, and Roberto Zanoli. The TextPro tool suite. In *Proceedings of LREC'08*, pages 2603–2607, 2008.

[Schultz, 2009] David M. Schultz. Writing an effective title. In *Eloquent Science*, pages 21–27. American Meteorological Society, 2009.

[Stock and Strapparava, 2006] Oliviero Stock and Carlo Strapparava. Laughing with HAHAcronym, a computational humor system. In *Proceedings of AAAI'06*, pages 1675–1678, 2006.

[Toivanen *et al.*, 2012] J. M. Toivanen, H. Toivonen, A. Valitutti, and O. Gross. Corpus-based generation of content and form in poetry. In *Proceedings of ICCC'12*, pages 175–179, 2012.

[Valitutti *et al.*, 2009] A. Valitutti, C. Strapparava, and O. Stock. Graphlaugh: a tool for the interactive generation of humorous puns. In *Proceedings of ACII'09 Demo track*, pages 634–636, 2009.

[Valitutti *et al.*, 2013] Alessandro Valitutti, Hannu Toivonen, Antoine Doucet, and M. Jukka Toivanen. "Let everything turn well in your wife": Generation of adult humor using lexical constraints. In *Proceedings of ACL'13*, pages 243–248, 2013.

[Veale, 2012] Tony Veale. *Exploding the creativity myth: The computational foundations of linguistic creativity*. A&C Black, 2012.

[Veale, 2014] Tony Veale. A service-oriented architecture for metaphor processing. In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 52–60, 2014.