BISLON: BISociative SLOgaN generation based on stylistic literary devices

Andraž Repar[1,3], **Matej Martinc**[1,2], **Martin Žnidaršič**[2] **Senja Pollak**[2,4]
[1] Jožef Stefan Postgraduate School and [2] Jožef Stefan Institute, Ljubljana, Slovenia
[3] Iolar d.o.o., Ljubljana, Slovenia
[4] Usher Institute, University of Edinburgh, UK
andraz.repar@iolar.com,{matej.martinc,martin.znidarsic,senja.pollak}@ijs.si

Abstract

We describe a novel slogan generator that employs bisociation in combination with the selection of stylistic literary devices. Advertising slogans are a key marketing tool for every company and a memorable slogan provides an advantage on the market. A good slogan is catchy and unique and projects the values of the company. To get an insight in construction of such slogans, we first analyze a large corpus of advertising slogans in terms of alliteration, assonance, consonance and rhyme. Then we develop an approach for constructing slogans that contain these stylistic devices which can help make the slogans easy to remember. At the same time, we use bisociation to imprint a unique message into the slogan by allowing the user to specify the original and bisociated domains from where the generator selects the words. These word sets are first expanded with the help of FastText embeddings and then used to fill in the empty slots in slogan skeletons generated from a database of existing slogans. We use a language model to increase semantical cohesion of generated slogans and a relevance evaluation system to score the slogans by their connectedness to the selected domains. The evaluation of generated slogans for two companies shows that even if slogan generation is a hard problem, we can find some generated slogans that are suitable for the use in production without any modification and a much larger number of slogans that are positively evaluated according to at least one criteria (e.g., humor, catchiness).

Introduction

A slogan is a key marketing asset for any company trying to sell its products and having a good slogan can make an enormous difference on their success. It can drive brand recognition and increase customer loyalty. Slogans are usually produced in brainstorming sessions that involve multiple people. Having a tool that would provide a large set of initial slogan candidates could potentially be of great benefit to marketers and advertisers.

Computational creativity is concerned with machines that exhibit behaviors that might reasonably be deemed creative (Colton and Wiggins, 2012), and our slogan generation system is conceived as a creative system supporting humans in their creative behavior. Closely related research areas include computational humor (Ritchie, 2009; Stock and Strapparava, 2003; Dybala et al., 2010) and poetry generation (see survey of Oliveira (2017)), with the most related to our approach being the lyrics generation system of Bay, Bodily, and Ventura (2017), which transforms an existing text based on certain parameters, including literary devices.

Several approaches to slogan generation have been developed in recent years. BRAINSUP (Özbal, Pighin, and Strapparava, 2013) is an extensible framework for generation of creative sentences. Users can select *target words* that have to appear in the final generated sentence and control the generation process across several dimensions, namely emotions, colors, domain relatedness and phonetic properties (rhymes, alliterations and plosives). The sentence generation process is based on sequences of morpho-syntactic patterns (skeletons) extracted from a corpus of existing marketing slogans. Tomašič, Papa, and Žnidaršič (2015) introduce genetic algorithms to mimic the brainstorming process, while in Žnidaršič, Tomašič, and Papa (2015) they propose a case-based reasoning approach.

We propose a new slogan generation system using literary stylistic devices—that we also analyze on a corpus of existing slogans—in a novel bisociative setting. Koestler (1964) argues that the essence of creativity lies in perceiving of an idea in two self-consistent but habitually incompatible contextual frames, and we use this cross-context approach as a principle in the design of our system. Compared to other slogan generation systems, the input to our system is neither individual words (as in Özbal, Pighin, and Strapparava (2013) or various online slogan generator systems[1]) nor a single document, as in Tomašič, Papa, and Žnidaršič (2015), but the documents from two distinct domains, resulting in bisociative slogans, bearing the marks of both domains.

Outline of the BISLON approach

The main aim of BISLON is to produce innovative slogan candidates of good quality, similar to the ones produced by marketing professionals. Our approach to slogan generation has two principal characteristics. First, it is based on literary stylistic devices (rhyming, alliteration, consonance and assonance) and second, it is related to the concept of bisociation (Koestler, 1964), which has not yet been explored in

[1] E.g., https://www.shopify.com/tools/slogan-maker

the context of slogan generation.

As a resource of slogan skeletons, we use a database of 5,287 English slogans.[2] Our slogan generation mechanism uses a large number of natural language processing (NLP) techniques, and was designed in order to correspond to the following properties of a good slogan:

- For a slogan to be *catchy* and *memorable*, the system uses literary stylistic devices (e.g., rhyme, alliteration).
- For a slogan to be *unique, interesting* and *surprising*, we propose a bisociative slogan generation mechanism. In a very simplified way, we can understand bisociations as cross-context associations and in our system we blend two matrices of thought (two input domains) into a new combined matrix. Word embeddings and Metaphor Magnet metaphores (Veale and Li, 2012) in candidate word generation process also contribute to surprising outputs.
- To address *relatedness* to the domain of interest, the system has a scoring function for weighting domain words.
- For *semantic* and *syntactic* cohesion, we use syntactic skeletons from existing slogans, perplexity computed in relation to the language model and a spell checker.

System Input

The system allows three types of inputs (the last two are optional but advised, in order to increase the variety and the relevance of generated slogans):

- A set of *original* and *bisociated text documents*: To support bisociation, the user is asked to input the documents from two domains. E.g., the user can select one domain as the domain describing the company for which the slogans are generated (*original* domain) and the *bisociated domain* can be selected based on some distant association.
- *Metaphor Magnet terms*: Users can define target and source concepts (corresponding to the original and bisociated input domains).
- *Domain specific terms*: The terms can be either manually defined as keywords of interest or extracted automatically from uploaded documents. We opted for automated term extraction on original documents, using the system from Pollak et al. (2012).

Literary stylistic devices

To enhance memorability and catchiness, slogans may contain various stylistic literary devices, such as rhyme, consonance, assonance and alliteration, which have roots in poetry. According to Baidick (2008):

- **Alliteration** is the repetition of the same sounds—usually initial consonants of words or of stressed syllables—in any sequence of neighboring words. For example, the initial sound L in: ***L**andscape-**l**over, **l**ord of **l**anguage*.[3]
- **Consonance** is the repetition of an identical or similar consonant in neighboring words whose vowel sounds are different, like the consonant K in this sentence: *Di**ck** li**k**es his new bi**k**e*.
- **Assonance** is the repetition of identical or similar vowels in the stressed syllables (and sometimes in the following unstressed syllables) of neighboring words: *The engin**ee**r held the st**ee**ring wh**ee**l to st**ee**r the v**e**hicle*.
- **Rhyme** is the identity of sound between syllables or paired groups of syllables, usually at the end of verse lines. For example: *A t**aste** too good to w**aste***.

For phonetic analysis of the words, we used the NLTK implementation of Carnegie Mellon Pronouncing Dictionary (CMPD) (Lenzo, 2007) which returns a list of phonemes for each word. Take as an example the word "house", for which the following output is obtained: [[HH, AW1, S]].

Analysis of literary devices in existing slogans

In this section we present the analysis of the usage of literary devices in a collection of real slogans. We focus only on nouns, verbs, adjectives and adverbs, since these are the parts-of-speech (POS) replaced in the generation step.

Rhyming

We calculated the rhyming level of two words, i.e. the number of ending phonemes that match, on the basis of the CMPD dictionary. According to Baidick (2008), a phoneme is a minimal unit of potentially meaningful sound within a given language's system of recognized sound distinctions. In general, two words rhyme if they have the same final stressed vowel and all the sounds following it to the end of the word. Example of a rhyme is: *Think about your **car** when you go to the **bar***.

No. of slogans	Precision
123	0.93

Table 1: Analysis of slogans containing rhymes.

As we can observe from Table 1, our system can recognize rhyming with a high level of precision. Our algorithm detects 123 potential rhyming slogans with more than 90% of them being considered true rhymes by a human evaluator. As explained above, we consider only rhymes between nouns, adjectives, verbs and adverbs.

Alliteration, consonance, assonance

For simplified alliteration analysis, we decided to focus only on initial consonants of words. In terms of neighboring words, we included two parameters:

- Strength S denotes the number of words included in the alliteration. If S is set to 3, we get only slogans with more than 3 words in the alliteration sequence.
- D denotes distance between words in the alliteration sequence. If D is set to 1, the distance between the words in the alliteration sequence cannot be greater than 1 (i.e. only one other word can be in between).

[2] We thank Polona Tomašič for her collection of slogans from Internet. Since we do not know how the copyright laws apply to slogans, they are not made publicly available.

[3] The examples in this section are made up due to potential copyright issues.

Consider the sentence: *It **takes** a **true** man to make a **tasty** pie.* If S is set to 3, then this slogan will not be returned by the algorithm. If D is set to 1, then only the first two words (takes, true) will be considered.

In our consonance algorithm, we again made several simplifications: we only focus on identical phonemes relying on the CMPD dictionary and we disregard the second part of the definition about different vowels. Our algorithm essentially detects, within certain parameters, whether words contain the same consonants in non-initial positions (while initial positions are covered by alliteration). Just as with alliteration, the consonance algorithm has the parameters D and S. But unlike alliteration, where good results could be obtained even with a low S, consonance is a subtler device—with the same D, higher values of S are usually needed to produce a pronounced consonance effect.

E.g., consider the two sentences below, where consonance is relatively weak in the first example and very noticeable in the second one.

- A s**l**y and dead**l**y man.
- There is no righ**t** momen**t** to imi**t**a**t**e the beas**t**.

To analyze assonance, we took advantage of the stress annotation offered by the CMPD dictionary to detect only those vowels that have primary or secondary stress. Again, the same two parameters D and S are used. Just like consonance, assonance is also a subtler device than alliteration—for the maximum effect the vowel in question has to be present in several words closely together:

- The m**o**st important man has sp**o**ken.
- Hear the m**e**llow w**e**dding b**e**lls.

We can observe the relatively weak assonance effect in the first example and the comparatively stronger assonance in the second example.

We tested three different configurations of parameters D and S for alliteration, consonance and assonance. In terms of precision, all the results obtained are true examples of the respective literary devices. The only exception are rare cases, where the CMPD returns incorrect pronounciations. For results with different parameters settings, see Table 2.

Configuration	Conf1	Conf2	Conf3
Alliteration	339	11	172
Consonance	106	71	567
Assonance	33	20	222

Table 2: Number of slogans containing alliteration, consonance and assonance discovered with computational means (the actual number of the slogans could be higher). The following parameter configurations were used: D=0,S=1; D=1,S=3; D=10,S=2 for alliteration, D=0,S=2; D=1,S=3; D=2,S=2 for consonance and D=0,S=2; D=1,S=3; D=2,S=2 for assonance.

Compared to the total number of slogans in our database (5,247) the number of slogans found during the analysis is quite low. However, using less limiting settings of parameters would return a higher number of slogans.

Slogan skeleton generation

For slogan generation, the existing slogans in our database are used as the starting point. We converted the slogans to lowercase, tokenized and POS tagged them with the NLTK library (Bird, Klein, and Loper, 2009). We use the coarse-grained universal tagset, as we suppose that grammatical issues would be fixed by the language model. Every noun, adjective, verb and adverb is removed from the slogans, although we keep the information on their POS tags. Every other word type (prepositions, conjunctions etc.) is carried over to the new slogan. This leaves us with a slogan skeleton with empty slots ready to be filled in with appropriate word candidates.

For the replacement, we introduce a bisociation parameter B, which controls the percentage of original and bisociated domain replacement words. If B is 0.5, half of the words come from the original and the other half from the bisociated domain. If B is 0, then all of the words are from the original domain and if B is 1, then all are from the bisociated domain. According to B, the appropriate number of empty slots in the skeleton are randomly chosen and marked as original or bisociated positions.

The next step in the skeleton creation varies according to the chosen literary device described above. For alliteration, consonance and assonance, the user can control the final shape of the literary device with two parameters:

- Distance D controls the distance between words in the literary device sequence.
- Strength S controls the number of words in the literary device sequence.[4]

Let's consider the sentence *Any man looks extreme with XXX shaving cream*, select alliteration as the literary device, the following parameter configuration: B = 0.5, D = 2 and S = 0.5 and the following skeleton:

- Any NOUN VERB ADJ with NOUN VERB NOUN .

Based on D, S and B, the positions to be filled in with the literary device and original or bisociated replacement words are randomly set (replacement positions are numbered from left to right, starting with 1):

literary device positions [3, 5, 6]
original positions [1, 3, 5]
bisociated positions [2, 4, 6]

For rhyme, there are no D and S parameters. Instead, we use the results of the analysis. From the existing slogans we select the ones that contain rhymes and mark the positions of the rhyming words as literary device positions. If we take the same example as before, which contains a rhyme, the literary device positions would now be: [3, 6].

Candidate word pools

Candidate words generation

After generating slogan skeletons with empty slots marked with POS tags, domain and literary device positions, we

[4]Note that opposed to the analysis phase, this parameter has values between 0 and 1 (0.5 means that half of the words should use the literary device).

need to find appropriate candidate words to fill them in.

From the three types of input (documents, Metaphor Magnet terms and domain terms) we generate eight distinct word pools (noun, verb, adjective and adverb pools for original and bisociated domains). In order to do that, we first tokenize and POS tag the input documents and the list of domain terms. Target and source metaphors from the Metaphor Magnet web service are returned in a form of adjective-noun pairs, which are split and assigned appropriate POS tags. Next, we add the resulting words to their appropriate pools, according to their POS tag and domain.

For the optimal functioning of the slogan generation system, the size of all the word pools needs to be sufficiently large, so the system always has enough good candidates for the empty slots in the slogan skeletons. Large input documents would solve this problem but would also make the system user unfriendly, by requiring a lot of effort from the users to gather this large input corpora. Therefore, we expand our word pools by using FastText embeddings (Bojanowski et al., 2016). We loop through the vocabulary of the input text documents and find 15 most similar words for every word in the vocabulary according to its FastText vector.[5] These additional words are POS tagged[6] and also added to appropriate pools.

Candidate word weighting

Ideally, we want our slogan generator to output slogans as relevant to the specified original and bisociated domains as possible. In order to do that, we assign weights to all the words in the created word pools according to their relevance to either the original or the bisociated domain. The calculation of the word weight for a specific word depends on the source of the word. If a specific word does not appear in the input text documents nor in the list of domain specific terms, the weight is automatically 0, since the word is probably irrelevant to the original and bisociated domains. If the word appears in the input text documents, the relevance weight is calculated according to the following formula:

$$w = \frac{input\_document\_freq}{BNC\_corpus\_freq}$$

input_document_freq is the relative frequency of the word in the concatenation of either all the original or all the bisociated documents and *BNC_corpus_freq* is the relative frequency of the word in the BNC reference corpus (Leech, Rayson, and others, 2014). The comparison between the frequency of terms in a domain corpus and in the corpus of general language is a relatively standard approach[7] for keyness calculation, since it rewards domain specific words.

input_document_freq is not available for words from the list of domain specific terms. Therefore it is replaced by a terminology strength parameter *n* (with the default value of 10) in the relevance weight calculation. Finally, the normalized score is calculated for each weight:

$$normalized\_w_i = \frac{w_i - min(w)}{max(w) - min(w)}$$

We use weight normalization to make the relevance weight more easily interpretable by human readers. The final output of the weighting step are two lists of weighted words, one for the words from the original and the other from the bisociated domain pool.

Slogan generation

In the next step, we loop through a list of slogan skeletons and try to generate a new slogan for every skeleton by filling in the empty slots from left to right (this filling order is necessary for the language-model-based semantic and grammatical checks). For every empty slot, we first generate a list of possible word candidates that fit the following criteria:

- The POS tag and the domain is correct. This is done by using only the candidate words from the appropriate word pool. In order to generate meaningful slogans, a sufficient number of candidates is required for every empty slot. The minimum number of candidates that satisfy the criteria is set as a parameter, with the default value of 30. If there are less candidates than the limit for any empty slot, the slogan with the specific skeleton is not generated.
- If the empty slot is marked as a literary device position, the candidate word needs to have the correct phoneme structure for the production of the literary device. This criterion is not applied to the first literary device position in the slogan. Instead, the system remembers only the phoneme structure of the first chosen replacement word and uses it to select compatible replacement word candidates for the remaining positions. As with the previous criteria, here we also enforce the limit of at least 30 appropriate candidates for every literary device position, otherwise the slogan for the specific skeleton is not generated.
- The candidate is a semantically and grammatically appropriate continuation of the preceding word sequence. For this we use a character-aware deep neural language model (Kim et al., 2016) trained on 200,000 randomly chosen articles from the Wikipedia.[8] For training, the vocabulary size was 50,000 words, only character-level inputs were used and the model was trained for 21 epochs. The semantic and grammatical appropriateness criteria is not enforced, if the empty slot is in the position of the first word in the slogan. Otherwise, the language model takes the part of the already generated slogan left of the empty slot and returns probability for each word candidate, that it fills the next position in the sequence. Only five most probable candidates are chosen for the final list.[9]

[5] The 1 million word vector model trained on Wikipedia 2017, UMBC webbase corpus and statmt.org news dataset (16B tokens) https://fasttext.cc/docs/en/english-vectors.html is used for the word similarity calculation. The number of most similar words was determined experimentally and assures good functioning of the system even if the input texts are short.

[6] These words do not have any surrounding context, therefore POS tagging here is less reliable.

[7] It is also the underlying principle in the term extraction tool used in the system input phase.

[8] 90% of the data set was used for training, 10% for validation.

[9] Number five was chosen empirically and represents a balance between a variety and cohesion of the generated slogans.

After the filtering described above, we have a generated list of five candidates for each empty position in the skeleton (except for cases when the empty slot represents the first word, then the number of appropriate candidates is not limited). Out of these candidates we choose the final filler word according to the probabilities calculated from the relevance weights described in section Candidate word weighting. The probabilities for every candidate in the list are calculated by dividing the weight of the candidate with the sum of all the candidate weights in the list. Selection according to the computed selection probabilites was chosen, since it contributes to the variety of generated slogans by allowing that the most probable candidate is not always the one selected. On the other hand, the system keeps the slogans relevant by never selecting the irrelevant words (with 0 weight), if relevant candidates are in the list. If there are no such candidates each candidate has the same probability of being selected.

After the slogan is generated, it is first put through a spell checker[10] which tries to automatically remove possible grammatical mistakes. Finally, the semantic and syntactic cohesion check is performed by calculating the average perplexity of the whole slogan by summing the perplexities of all the words in the slogan and dividing the sum with the number of words. Perplexity is a measure of how well a probability model predicts a sample and represents a standard way of language model evaluation. We set the maximum perplexity score of a slogan to 50, which is slightly more than the perplexity of the language model evaluated on the Wikipedia validation set. If the score is above the selected threshold, we assume that this indicates semantic or syntactic inconsistency and the slogan is discarded.

When the empty slot that needs to be filled in is the first word of a slogan, there is no semantic and grammatical continuation or literary device filtering. This means that the only selection criterion is relevance, which causes that some very relevant candidates are selected very often. To avoid the too frequent repetition, we introduce a maximum repetition parameter with the default value of 10. The slogans with the same first word are grouped and if their count is higher than the repetition parameter, the extra slogans with the lowest perplexity get discarded. In this way, we only keep more "original slogans", which have higher perplexity.

Finally, the remaining generated slogans that passed all the tests described above are sorted by their relevance score, which is calculated as a sum of relevance weights of all the words in the slogan divided by the number of words.

Application

To test our generator in a real setting, we have tried to generate a slogan for two Slovenian companies Iolar and Elea IC. The former is in the translation business and the latter primarily deals with construction. For both firms original and bisociated domain documents were defined and terminology was extracted[11] from the original texts.

In the case of Iolar, we used Wikipedia articles on localization and translation memory and one of Iolar's marketing brochures as original domain texts and articles on eagles, Ireland, flight and aircraft for the bisociated domain, since the name of the company *Iolar* means eagle in Irish Gaelic (also a source of inspiration for the company's existing slogan *Flying over the borders*). For Metaphor Magnet generation, we used the phrase *translation is an +eagle* (the + sign limits the search space to positive connotations).

For Elea, the original domain text consisted of the promotional material describing the company, while for the bisociated domain the concept of Eleatics was selected (specifically, the Wikipedia article on this topic). The concept is related to the name of the company and denotes a pre-Socratic philosophy school. For the Metaphor Magnet, we used the phrase *building is a +philosophy*. Table 3 contains the final vocabulary size for individual domains.

	Iolar		Elea	
	Original	Bisociated	Original	Bisociated
Nouns	9,492	7,046	14,164	1,667
Verbs	2,191	1,409	2,884	423
Adjectives	1,926	1,331	2,796	291
Adverbs	572	518	678	161

Table 3: Number of word candidates (word pool sizes) for original and bisociated domains.

We generated slogans for each of the four literary devices. For alliteration, consonance and assonance, the settings D=2, S=0.8, B=0.3 were used, as we aimed for relatively strong literary device effects. D=2 means that the words considered for the specific effect should be relatively close together, and S=0.8 means that the majority of the words in the slogans will be considered for the effect. To simplify, the literary device effects in the resulting slogans will be very strong. The reason for the relatively low value of B is that we wanted to have the majority of the words coming from the original domain, while a smaller number of the words from the bisociated domain contributes to the variety and a unique character of the slogan.

The system generated altogether 1,400 slogans for Iolar (290 with alliteration, 457 with assonance, 413 with consonance and 240 with rhyme) and 811 slogans for Elea (174 with alliteration, 266 with assonance, 255 with consonance and 116 with rhyme). All the generated slogans, together with the human evaluation scores, are made available here: http://kt.ijs.si/data/cc/slogan_generation.zip

Evaluation

The resulting slogans were evaluated for each company. The evaluation criteria, adapted from Özbal, Pighin, and Strapparava (2013), are the following:

- **Catchiness**: is the slogan attractive, catchy or memorable? [Yes/No]
- **Humor**: is the slogan witty or humorous? [Yes/No];
- **Relatedness**: is the slogan semantically related to the company domain? [Yes/No];
- **Correctness**: is the slogan grammatically correct? [Yes/Minor editing/No];

[10] http://pypi.python.org/pypi/language-check

[11] http://clowdflows.org/workflow/5515/

Set	Catchiness Iolar	Catchiness Elea	Humor Iolar	Humor Elea	Relatedness Iolar	Relatedness Elea	Correctness Iolar	Correctness Elea	Usefulness Iolar	Usefulness Elea
Yes	0.152	0.263	0.085	0.218	0.220	0.282	0.345	0.443	0.035	0.068
No	0.848	0.737	0.915	0.782	0.780	0.718	0.527	0.415	0.880	0.877
Minor editing	-	-	-	-	-	-	0.128	0.142	0.085	0.055

Table 4: Evaluation results for top 400 slogans according to the relevance score.

- **Usefulness**: could the slogan be a good slogan for your company? [Yes/Minor editing/No].

For evaluation in each company, the following slogans were selected, based on different criteria:

- Top 100 generated slogans from each literary device ranked according to the system's relevance score (400 slogans in total).
- An additional 12 slogans from each each literary device for inter-annotator agreement (IAA) calculation (48 in total), taken from top ranked slogans.
- 16 generated slogans from each each literary device with the lowest relevance score (64 slogans in total), aimed at evaluating the accuracy of the relatedness ranking score.
- For Iolar, we add 30 real-life slogans of other translation companies, which were available in our slogan database.

For each company, we split the evaluation dataset into 4 annotation sets, with proportionally and randomly selected slogans based on the above selection criteria. For example, each annotator received 124 top scored slogans (25 slogans from each literary device and 24 selected for IAA experiment) and 16 lowest ranked slogans, leading to 140 slogans in each evaluation set. For IAA, two sets of 24 slogans are annotated by a pair of annotators. Finally, three[12] Iolar evaluation sets also contained 10 real-life slogans used by translation companies today.

The sets were prepared for four employees from each company. While for Iolar four employees agreed to perform the evaluation, in Elea, due to time constraints, only two of our contacts were able to perform the task. The other two Elea evaluation datasets were annotated by five persons not employed in the company, but familiarized with the company's professional activities. This meant that we were unable to calculate IAA for one pair of the Elea datasets.

Results

The evaluation results (see Table 4) indicate that slogan generation remains a very difficult task. Nonetheless, more than 10% of the generated slogans can be considered at least partly useful, a similar number can be considered catchy or funny, while around 50% of slogans were evaluated as correct (categories Yes and Minor editing combined). In general, the results for Elea are a bit higher than for Iolar but the large majority of generated slogans are not considered good enough for actual use. However, our goal—to show that at least some actually useful slogans can be automatically generated—was achieved. In Figure 1 we provide a selection of the generated slogans evaluated as useful by the evaluators (for full list visit the url provided above).

[12]One set was evaluated by one of the paper's authors, who was aware of the difference between the generated and real-life slogans.

Iolar

Texts, translations, techniques.
Localization in central Europe.
Multilingual model of meaning.
Translating various cultures.
Localization of life and language.

Elea

Improve your move.
Highway. Holiday. Railway.
Your power is your tower.
Overpasses of creation and information.
Modernized world standards.

Figure 1: A selection of the best BISLON slogans. A total of 69 (Elea) and 76 (Iolar) slogans were deemed useful (Yes and Minor editing) by the annotators.

Comparing the results of the top 64 slogans against the bottom 64 slogans (Figure 2) confirms that the relevance score works as expected (the relatedness is higher among the top 64 generated slogans). However, there is no obvious qualitative difference in other criteria (in terms of catchiness and humor, the bottom 64 seem to be even better).

	Iolar1	Iolar2	Elea1
Catchiness	1	0.792	0.583
Humor	1	0.792	0.583
Relatedness	0.958	0.75	0.542
Correctness	0.792	0.667	0.583
Usefulness	0.958	0.667	0.917

Table 5: Observed agreement of three IAA sets.

Next, we observed the scores with regard to different literary devices. As it can be seen from Figures 3 and 4, for Elea the rhymes have the highest scores, while for Iolar, the best performing device is alliteration (in terms of usefulness).

The chasm that still needs to be overcome in slogan generation is obvious from comparison with evaluation of real life slogans in Table 6. Apart from humor, all other criteria exhibit much higher scores with usefulness exceeding 80%. However, one mitigating factor is that these slogans have

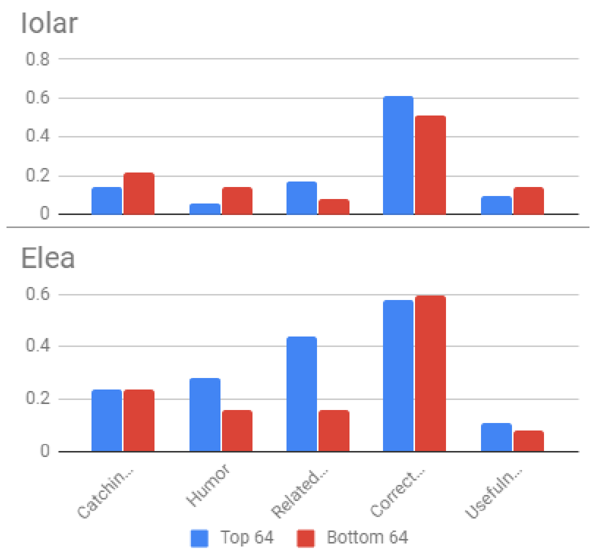

Figure 2: Ratio of positively evaluated slogans among the top and bottom 64 generated slogans (the positive (Yes) and partially positive (Minor editing) scores of the last two categories have been merged.)

most likely been perfected over the course of long brainstorming sessions. It may be more appropriate to compare the BISLON-generated slogans with draft slogans produced during a brainstorming session, to see how many useful slogans are actually generated during such sessions.

Finally, we calculated IAA on three annotation sets - each shared by two annotators. We can see that the overall agreement (OA) (presented in Table 5) is relatively high, with a mean value of 0.942 for the first pair of Iolar annotators, 0.734 for the second pair and 0.642 for the Elea annotators. We also calculated kappa values (Cohen, 1968). Overall, the values are low, but there are differences between the three pairs (average of 0.082 for the first, 0.419 for the second Iolar pair and 0.05 for the Elea pair). Note that the kappa score uses the expected agreement, which is extremely high for the first Iolar and the Elea annotator pairs. As the two Iolar annotators evaluated all the sentences with No for catchiness and humor, the overall agreement is 1, but so is the expected agreement, and consequently the kappa score is very low. According to Landis and Koch (1977), the agreement of the first Iolar annotator pair and of the Elea pair is *low* and the agreement of the second Iolar pair is *moderate*.

	Catch.	Humor	Related.	Correct.	Useful.
Yes	0.767	0.167	0.833	0.900	0.834
No	0.233	0.833	0.167	0.067	0.133
M. edits	-	-	-	0.033	0.033

Table 6: Evaluation results for real-life slogans.

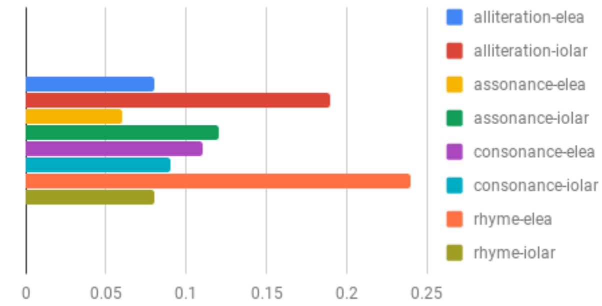

Figure 3: Ratio of useful slogans generated by BISLON (Yes and Minor editing are combined.)

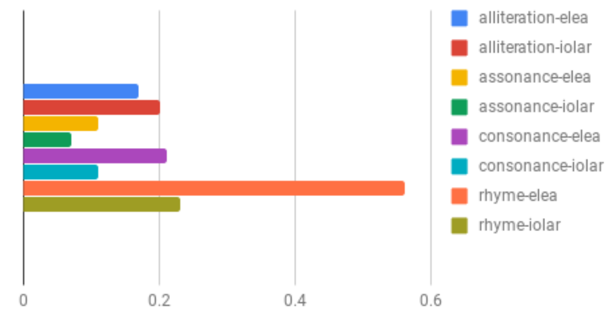

Figure 4: Ratio of catchy slogans generated by BISLON.

Conclusion and future work

This article contains an analysis of marketing slogans in terms of four stylistic literary devices: alliteration, consonance, assonance and rhymes and describes a novel bisociative slogan generator for the four devices. The system has several parameters allowing the user to tweak the strength of the literary device and bisociation and produces a score which ranks the slogans according to their relevance.

The system uses a large number of NLP techniques in order to produce interesting and unexpected slogans. In human evaluation cca. 10% (Yes and Minor editing) of slogans were evaluated as useful, which leaves room for improvement. However, despite the overall low number of positively evaluated generated slogans, the system still produces several good ones, which could actually be used. For example, for Iolar the slogans *Translating various cultures* or *Localization of life and language* are very good candidates. Similar for Elea, which is involved in transport infrastructure construction, a slogan like *Improve your move* could be perfectly applied.

It is hard to judge the actual value of the system. On one hand the results can be compared to other systems for automated slogan generation. The BRAINSUP (Özbal, Pighin, and Strapparava, 2013) framework overall achieves higher results for all the evaluation criteria. We do however believe, that our system could be used more successfully in the early stage of the slogan production process (e.g, in a brainstorming session) since it does not require a very narrowly defined input and produces a large number of very diverse slogans for specified domains, out of which some could be used as out of the box slogans, while others could be used to broaden the space of possible final solutions and discover new meaningful associations. On the other hand, we provided the comparison to human generated slogans. The results for the human generated slogans were significantly higher, which is not surprising, since these slogans are used in production and were most likely already chosen from a list of human generated slogans as best candidates, thoroughly checked and finally approved. Comparing the output of our system to a list of human generated slogan candidates would therefore be a more reasonable comparison and that is something we plan to do in the future, in the context of brainstorming sessions in the advertising industry.

For further work, we plan to perform a detailed analysis of generated slogans, make a systematic evaluation of different parameter settings and, more specifically, analyze the role of bisociation. We also plan to improve several features of the system. First of all, we will analyze the existing slogan database for additional devices used in the slogan production and try to incorporate them into the system. In order to improve syntactic correctness, we will replace the POS tagger using universal tagset by more fine grained tagging or consider incorporating the information from the dependency parser. Semantic cohesion will be improved by training a larger language model and other kinds of semantic features. The system for measuring relevance could be improved by using a more recent corpus to calculate candidate word weight, since some newer words that are not part of the terminology of chosen domains (e.g., download, freeware...) were given very high weights because of their very low frequencies in the BNC. Next, we will allow combinations of different stylistic devices. Finally, a system for sentiment analysis of generated slogans will be implemented. This system will filter out the slogans with negative sentiment, further reducing the number of unuseful slogans. The evaluated slogans will also be considered as a training data for machine learning approaches.

Let's conclude this paper with what was initially a slogan generated for a construction company but is in fact a very wise advice for any situation in life:

The main work also includes the brain.

References

Baidick, C. 2008. Oxford dictionary of literary terms.

Bay, B.; Bodily, P.; and Ventura, D. 2017. Text transformation via constraints and word embedding. In *Proc. of the 8th International Conference on Computational Creativity*, 49–56.

Bird, S.; Klein, E.; and Loper, E. 2009. *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

Bojanowski, P.; Grave, E.; Joulin, A.; and Mikolov, T. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Cohen, J. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin* 70(4):213–220.

Colton, S., and Wiggins, G. 2012. Computational creativity: The final frontier? In *Proc. of the 20th European Conference on Artificial Intelligence*, 21–26.

Dybala, P.; Ptaszynski, M.; Maciejewski, J.; Takahashi, M.; Rzepka, R.; and Araki, K. 2010. Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments* 2:31–48.

Kim, Y.; Jernite, Y.; Sontag, D.; and Rush, A. M. 2016. Character-aware neural language models. In *Proc. of the 30th Conference on Artificial Intelligence*, AAAI'16, 2741–2749. AAAI Press.

Koestler, A. 1964. *The act of creation*. London Hutchinson.

Landis, J. R., and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 159–174.

Leech, G.; Rayson, P.; et al. 2014. *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.

Lenzo, K. 2007. The CMU pronouncing dictionary. http://www.nltk.org/_modules/nltk/corpus/reader/cmudict.html. [Online; accessed 20-April-2018].

Oliveira, H. G. 2017. A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In *Proc. of the 10th International Conference on Natural Language Generation*, 11–20.

Özbal, G.; Pighin, D.; and Strapparava, C. 2013. Brainsup: Brainstorming support for creative sentence generation. In *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics*, volume 1, 1446–1455.

Pollak, S.; Vavpetič, A.; Kranjc, J.; Lavrač, N.; and Vintar, S. 2012. NLP workflow for on-line definition extraction from English and Slovene text corpora. In *Proc. of the 11th Conference on Natural Language Processing KONVENS*, 53–60.

Ritchie, G. 2009. Can computers create humor? *AI Magazine* 30(3):71.

Stock, O., and Strapparava, C. 2003. Getting serious about the development of computational humor. In *Proc. of the 18th International Joint Conference on Artificial Intelligence*, volume 3, 59–64.

Tomašič, P.; Papa, G.; and Žnidaršič, M. 2015. Using a genetic algorithm to produce slogans. *Informatica* 39(2):125–133.

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.

Žnidaršič, M.; Tomašič, P.; and Papa, G. 2015. Case-based slogan production. In *Proc. of the Workshop on Experience and Creativity at the 23rd International Conference on Case-Based Reasoning, ICCBR 2015*. CEUR.