Automated blend naming based on human creativity examples

Senja Pollak[1], Pedro Martins[2], Amílcar Cardoso[2], and Tanja Urbančič[13]

[1] Jožef Stefan Institute, Ljubljana, Slovenia
[2] CISUC, DEI, University of Coimbra, Coimbra, Portugal
[3] University of Nova Gorica, Nova Gorica, Slovenia
senja.pollak@ijs.si, pjmm@dei.uc.pt, amilcar@dei.uc.pt
tanja.urbancic@ung.si

Abstract. In this paper we investigate which principles people use when they name new things as results of blending. The aim is to uncover patterns with high creative potential and to use them for automated generation of names for new creations or phenomena. We collected examples with a web survey in which participants were asked to evaluate pictures of animals with blended anatomies from two different animals, and to provide their own names for blended creatures on the pictures. The blended animals served as a trigger of human creativity manifested through imaginative, humorous, surprising names collected in the survey. We studied how the features from the pictures reflected in the names, what are different complexity levels of lexical blend formation and how far in other realms subjects "travelled" to search for associations and metaphors used in the names. We used the findings to guide automated generation of names for the blends.

Keywords: Computational creativity, human creativity examples, conceptual blending, lexical blend generation, creative naming, bisociation.

1 Introduction

Creativity is in the core of many human activities and has been studied for decades [9][2]. As a phenomenon challenging for being replicated with machines, it became also a topic of artificial intelligence research [21]. While creativity is an intriguing research question by itself, it is also a driving force of development and as such, it has an immense value for applications in countless areas, including scientific discovery, engineering inventions and design. One of the cognitive principles underlying such discoveries and inventions is conceptual blending [5] in which two mental spaces integrate into a new one, called blend. Conceptual blending has also been implemented and tested in computer systems to produce novel concepts [17]. However, there are still many open questions related to the choice of input mental spaces and the ways of projections that lead to blends, perceived as creative and inspiring. In our work we aim at providing guidance

for choosing input spaces and projections based on concrete findings about human creativity with elements of blending. More precisely, by investigating the patterns that we can find in the cases of human creations, we guide the blending process to the extent allowing for automated generation of blends.

Conceptual blending and case-based reasoning [10] can meet in a very fruitful way in areas such as design and architecture [4][6].In such domains, blends are not only a source of surprise, artistic satisfaction or inspiration, but have also their own functionality, bringing into the process additional constraints and priorities. Contexts and goals can also be used in computational approaches to conceptual blending and can beneficially affect the issues of efficiency [13]. Authors in [1] exploit a principle of creative transfer from one domain to another in the realm of design. Their IDEAL system abstracted patterns from design cases in one domain and applied them to design problems in another domain. connecting distant, self-consistent and usually not connected frames of reference has been recognised and used as an effective principle in the act of creation. Such connections of habitually incompatible domains through common patterns or bridging concepts are also referred to as bisociations [9].

In this paper, we address the issue of case-based reasoning and conceptual blending in the context of lexical creativity. While this might appear quite far from the discussion on design in the previous paragraph, the connection becomes evident based on an observation by Veale and Butnariu [20]: "Words are everyday things, as central to our daily lives as the clothes we wear, the tools we use and the vehicles we drive. As man-made objects, words and phrases are subject to many of the same design principles as the consumer artefacts that compete for our attention in the market-place.". The authors also draw attention to two basic principles of artefact design, as identified in [15], namely visibility and mapping. In the case of a well-designed product, the design should suggest a mental visualisation of a conceptually correct model of the product, and the mapping between appearance and function should be clear. Their Zeitgeist system [20] can automatically recognise neologisms produced as lexical blends, together with their semantic meaning. This is done based on seven different "design patterns" recognised in constructing neologisms as lexical blends. Types of lexical blends and how new lexical blends are formed is described and illustrated with many examples in [12]. An important issue of recognising and quantifying creativity in different combinations of words is studied in [11].

In our work we investigate how humans approach the task of naming new things, and how based on human examples, a computer system could exhibit similar (and, why not, better) performance. We consider this principle of using past examples for revealing patterns to be used for new cases as a manifestation of case-based reasoning. The concrete task was to name creatures – animals with blended anatomies from two different animals. This was done in a web-based survey, designed primarily for a study of human perception of visual blends [14]. In this paper we continue using the material of the same study, but we examine it from a completely different angle, i.e. from the lexical creativity side by investigating creative naming of blends. Many offers for supporting naming of

snorse
(snake, horse)

chimporse
(chimpanzee, horse)

durse
(duck, horse)

guineabear
(guinea pig, bear)

hammerheadhorse
(hammerhead shark, horse)

pengwhale
(penguin, whale)

proboscis parrot
(proboscis monkey, bird)

chamelephant
(elephant,chameleon)

duckphant
(elephant, duck)

guinea lion
(guinea pig, lion)

horbit
(horse, rabbit)

hammerhead gull
(hammerhead shark, gull)

horduck
(horse, duck)

spider pig
(spider, guinea pig)

shark retriever
(shark, labrador retriever)

Fig. 1. Hybrid animals dataset used in the online questionnaire (available at http://animals.janez.me). Each sub-caption contains a name of the blend proposed by survey participants, as well as the input spaces. All blends were created by Arne Olav, with the exception of *shark retriever* and *camalephant*, whose authorship is unknown. For a better visualisation, some images were slightly cropped.

client's enterprises, products, etc. can be found on the web and show the application potential of creative naming. The task has already been approached with the goal of (semi-)automatic name generation and the results presented in [16] and [18] demonstrate a very big potential. While our work shares some of the ideas with above-mentioned related approaches, it differs from them by using visually triggered human examples as examples used for automatic lexical blend generation, and by using a novel categorisation of creativity level that guides construction of blends based on bisociation as one of the key principles inherent in many human creative processes.

After presenting the survey in which the names were collected in Section 2, we analyse different patterns and mechanisms used by people when coining names in form of lexical blends in Section 3. These patterns are used in Section 4 for automatically generating blends of different levels. In Section 5 we discuss the potential of our prototype and present further research perspectives.

2 Survey: Visual blends and their lexical counterparts

In [14], we introduced a survey consisting of an on-line questionnaire related to the quality of visual blends. Around 100 participants assessed 15 hybrid animals which were the result of blending anatomies from two different animals (Figure 1).The participants were asked to to rate criteria related to the coherence of blends as well as creativity.

Clearly in our questionnaire on animal blends the main focus was on visual blends. However, with the aim of getting more insight into potential connections, participants were also asked to provide a name (in English, Portuguese, Slovene, French or Spanish) to each of the hybrid creatures. By asking people to name the creatures we wanted to investigate the following questions: Would participants give names for all, for none, or for some of the creatures? How creative are they when naming the animals, how does the visual blended structure reflect in the lexical blend? Where the names provided by subjects mostly lexical blends or not? Do lexical blends use animal's "prototype" characteristics, or more sophisticated associations for which some background knowledge is needed (like titles of books, movies, history, etc.)? Does complexity of visual blends reflect in the names? The names given to the visual blends are the focus of our study.

In our survey we collected 1130 names for 15 animals. The general trend was that people gave more names at the beginning of the study and the trend of the number of given names was descending. However, some pictures triggered more generated names than expected by their position (e.g., *guinea lion* and *spider pig*). The guinea lion is also the blend for which the unpacking (recognising the input spaces) was the most difficult [14] and the one for which the highest number of very creative, bisociative lexical blends were formulated.

3 Formation and complexity of lexical blends

Our previous investigations of relationship between conceptual blending and bisociation have drawn our attention to different levels of blend complexity. To deal with this issue in a more systematic way, we suggest the following categorisation regarding the input words used to form the name:

L1 each of the words appearing in the lexical blend is a commonly used word for one input animal (no mapping);
L2 both input words represent input animal in a rather common way, but are blended into one word by *portmanteau* principle, i.e. by using the prefix of one word and the suffix of the other word (possibly with some intersection);
L3 one word represents one input animal with a commonly used word for this animal, the other word represents a *visible characteristic* (part, colour etc.) of the other animal (variant L3*: both words use such characteristics);
L4 one word represents one input animal with a commonly used word for this animal, the other word represents a characteristic of the other animal for which *background knowledge* about this animal (habitat, way of moving, typical behaviour) is needed (variant L4*: both words use such characteristics);
L5 one word represents one input animal with a commonly used word for this animal, the other animal is represented with a more sophisticated association – *bisociation* – for which a creative discourse into another realm (e.g. from animals to literature) is needed (variant L5*: both words represented with such associations).

We illustrate the categories by the names actually given in the survey to the blended animal *guinea bear*:

L1 *mouse-bear* (input1: *mouse*, input 2: *bear*);
L2 *rabbear* (input1: *rabbit*, input 2: *bear*);
L3 *small-headed bear* (input1: *mouse* → *small head*, input 2: *bear*);
L4 *scared bear* (input1: *mouse* → *scared*, input 2: *bear*);
L5 *mickey the bear* (input1: *mouse* → *Mickey the mouse*, input 2: *bear*).

As seen from this example, while the bear was easily recognised as one of the constituting animals, there were different interpretations about the second animal, "contributing" the head to the blended creature. In fact, the variety in the whole dataset was even bigger as names given by different subjects suggested the second animal being a mouse, rabbit, hamster, guinea pig, rat, squirrel, wombat or opossum. The set of input words as used by the subjects is even bigger since it includes also diminutives, slang versions, etc.

The levels increasing indicate the increasing complexity (but not necessarily the quality) of the blends, but note that they do not build on just one criterion in a linear way and there might also be a combination of principles described at different levels present in one name. We illustrate this with a name *teddybbit*, generated as a portamanteau (L2), but using an association between bear and teddybear from the toys realm (L5).

However, we plan to improve this by introducing a creativity score in which not just the level of mappings used will be taken into account, but also the fact whether they were used for one or for both input animals, and how creative the combination was (e.g., by taking into account phonetic features or by recognizing references extrinsic to the two input animal spaces and their bisociations).

Note that not all of the names provided by the subjects in the survey were lexical blends. Here we do not analyse such names in more detail, but to study the potential for triggering creativity, they are important as well. Some examples collected in our survey for the *guinea bear* are *creepy*, *giant*, or *fluffy*.

4 Patterns from examples for automated name generation

We investigated how the above-mentioned categories of human-generated names could be used for automatic blend generation. Different categories represent different mechanisms. Names of Level 1 are very basic and easy to be automatically generated, their creativity level is low and the name can hardy be called a blend. On the other hand, higher levels (3-5) rely on human experience, background knowledge, associations and bisociations. To generate the names of levels 3 and 4, we use a large web corpus (the enTenTen corpus [7]) and the sketch grammar relations available in Sketch Engine [8]. For the last category (level 5), we used other resources of human knowledge (Wikipedia, imdb lists). For each category, we reveal the patterns in human given names and explain how they can be used in automatic generation. Our generated examples are all done by modifying only one animal name.

L1: In names given by humans, we found two different patterns at level 1. In each case, the two animals are used, the possible variations being either hyphen to indicate the combined meaning "animal1-animal2" (e.g. *dog-shark*) or creating a single word containing full names of both animals "animal1animal2" (e.g. *spiderrat*). The pattern with a premodifier of adjective can be recognised in the given name *mišasti medved*, where the first word is an adjective formed from the noun *miš* (Eng. *mouse*) and the second one is the noun medved (Eng. *bear*). Some word formations are language specific, e.g. in Slovene bare "noun-noun" word formation is not very productive.

To illustrate the automatic name generation, we took the animal names from each input space and concatenated them. Using these simple patterns resulted in names very much resembling those generated by humans, e.g. *duck-horse* or *duckhorse*. More examples are in the L1 row in Table 1.

L2: Level two uses the *portmanteau* principle. In all the languages used in the survey this mechanism was used very frequently. For recognising these names from the list, we focused on words composed of the beginning of one animal word and ending of the other. Examples of basic portmanteau names given by the subjects are the names given in Figure 1. We automatically recognized L2 blends by combining pairs of animals and some simple heuristics.

In automatic generation, the starting point was to combine half of the each of the two input animal names. If the input word consists of two words, frequently

Table 1. Automatically generated names - examples for four fictional animals.

Input level	elephant & chameleon	snake & horse	horse & chimpanzee	duck & horse
L1	elephant-chameleon	snake-horse	horse-chimpanzee,	duck-horse
	elephantchameleon	snakehorse	horsechimpanzee	duckhorse
L2	elepheleon	snarse	horanzee	ducrse
L3	tusk chameleon	venom horse	hoof chimpanzee	beak horse
	trunk chameleon	fang horse	mane chimpanzee	arse horse
	graveyard chameleon	tail horse	bridle chimpanzee	back horse
	tail chameleon	poison horse	rump chimpanzee	feather horse
	ear chameleon	belly horse	withers chimpanzee	
L4	Asian chameleon	venomous horse	Trojan chimpanzee	Anaheim horse
	giraffe chameleon	poisonous horse	wild chimpanzee	lame horse
	captive chameleon	garter horse	Arabian chimpanzee	Peking horse
L5	Dumbo chameleon	Ser Hiss horse	Alfonso chimpanzee	Donald horse
			Daffy horse	Howard the horse

in the analysed examples one word is kept to from the blended name (which is not a proper portmanteau anymore). This pattern was used for generating examples like *guinea lion*, *hammerhead eagle*, *hammerhead goose*.

One could make different combinations based on different proportions of the input words or by using phonetic rules (vowels, consonants, rhymes), exact vs. inexact matching, pronunciation information, word's Greek or Latin origins, etc. as in many advanced existing systems proposing portmanteau name generation [19] [18] [3].

L3: In the next category of lexical blends, humans use visible characteristics of one animal and associate them to the other animal. The properties of the animal that gives the "head" to the new visual blend can be lexically expressed as prepositional phrase modifying the head noun, i.e. the name of the animal providing the body (*horse with snake head, elephant of the orange beak*), by adjective modifier (e.g. *nosy robin, duckbilled pachyderm, trunkheaded chameleon)*or in noun-noun constructions (e.g. *nosebird*). In some cases both animals are described by their characteristic visible parts (e.g. *tail-trunk*). Combinations with portmanteau structure is also possible (e.g. *grivasti kabod* [Eng. *mane horswan*]).

For automated blend generation of L3 we currently use only noun-noun constructions. We rely on the Skecth Engine tool by using word sketches constructed with Sketch grammar. Word sketches are automatic corpus-derived summaries of a word's grammatical and collocational behaviour [8]. From the word sketch of animal "contributing" the head to the visual blend (e.g. *elephant* in Figure 1), we use all the collocators (above selected frequency and salience threshold)

from the grammatical category *possessed*. This lists contains nouns that in the enTenTen corpus follow the search word and 's, e.g. for *elephant's* the list contains *tusk, trunk, ...* resulting from collocation *elephant's tusk* in the corpus. We construct then noun-noun blends, by adding the animal name of the animal providing the body (e.g. *chameleon*). As shown in Table 1, examples using this structure often correspond to parts of the body, (*tusk chameleon, trunk chameleon, tail chameleon, ear chameleon*), while *graveyard chameleon* does not represent the part of the body. Obviously, some of the compounds are irrelevant, e.g. *tail chameleon* – since chameleons have a tail themselves so this description does not contribute anything in terms of blending. Neither does the corpus provide the information if the "possessed" part is located on the animal's head and even less if it corresponds to the depicted picture (e.g. tusks are not depicted on the picture of elephant and chameleon from Fig. 1, even if they are prototypical part of elephant's head). More specific filters and knowledge bases will be used in future to narrow the choice to better candidates.

L4: Level 4 names are more diverse and require more background knowledge. As mentioned in Section 3, the observed categories are habitat, locomotion (*plavajoči konj* [Eng. *swimming horse*], typical behaviour (e.g.*elequack* using animal sounds) or usage (*saddleducks*. Again, also both animals can be represented by their properties, such as in the blended name *galloping quack*. For automated name generation at this level, we used again the word sketches, but we took the information from category *modifiers* (typical adjectival or noun collocators modifying the animal providing the head to the blended creature). E.g. adjectives *venomous* and *poisonous* are typical collocators of word *snake* and are used for forming blended names *venemous horse* and *poisonous horse*. Often breed names are used in modifier position; by selecting only lower case modifiers we can keep more general properties. For Level 4 , more background knowledge is needed. E.g., from automatically constructed names *Trojan chimpanzee, wild chimpanzee* or *Arabian chimpanzee*, the first one is referring to specific cultural reference Trojan horse and can be interpreted at level 5. Same goes for the *lame horse*, which is formed from the idiom *lame duck* (i.e.*an elected official who is approaching the end of his tenure, and esp. an official whose successor has already been elected* (Wikipedia)).

L5: In analysis of human lexical blends we manually classified in Level 5 the bisociative blends using characters from cartoons (*Spider Gonzalez*), children songs (*Slonček Raconček* refering to a Slovene song *Slonček Jaconček*), where slonček means small elephant and *raconcek* comes from *duck – raca*), movies (*My little mallard*), politicians (*Sharkozy*), legends (*Jezerski Pegasus* [Eng. *river Pegasus*]) and often combinations of several of them, e.g. character from movie and from comic strips *Jumbo Zvitorepec* (where Jumbo refers to the animal, while *Zvitorepec* is a character from Slovene comic strip by Miki Muster, but literally means curled tail which refers also to the visual representation of this animal (cf. picture elephant, chameleon in Fig. 1).

For automatically generating highly creative lexical blends inspired by the examples given by participants, we based the bisociative blend generation on

characters from the movies representing the input animal. We created a short list from Wikis, IMDB and Wikipedia pages about animal characters in movies where the last section covers cultural representations. In the name generation process, we first checked if character's name contains the name of the animal and if so we substituted this name with the name of the other input animal (e.g. horse substituting the duck in *Donald horse*). On the other side, if the animal does not appear explicitly we added the name of the second animal to the existing character name (*Dumbo chameleon*). In future, we will expand generation of names at this level by exploring other realms besides movies and books.

5 Discussion

We investigated the principles of creating lexical blends based on visual blends (blended animals). We revealed different mechanisms used in name formation and introduced a new categorisation of blend complexity (L1-concatenation blends; L2-portmanteaux; L3-blending based on visible characteristics; L4- blending using background knowledge and L5-bisociative blends). After the analysis of examples generated names by humans, we made a prototype system for automated generation of blends of different levels using word combinations, grammatical and collocational information and background knowledge resources. The most frequent mechanism used by humans was the portmanteau principle. But a portmanteau can vary from very basic ones to the bisociative ones, since blend strategies can easily be combined. For instance, the blend *shagull* can be interpreted as a simple portmanteau blend (shark+gull) or as bisociative blend refering to Chagall. This example shows that the bisociation can be used on the production level (e.g. creative blend but the reader cannot decompose it), on the interpretation level (e.g. even if there was no such intention when generating a name, the bisociation can be present at the reader's side) or both.

We like some names generated as lexical blends more than the others – what counts? Even if names are generated using similar principles, some of them are much more creative, achieving higher degree of creative duality, compressing multiple levels of meaning and perspective into a simple name [20]). It is the combination of simplicity and bisociation (in our case the switch from animal wor(l)d to cultural realm) that seems to be the most impressive. To verify this claim and to get a more thorough evaluation of automatically generated names, we plan to collect human subjects feedback as well as compare human-generated and automatically generated names. We will also further elaborate the automatic recognition of blend complexity and on the other side the blend generation part (e.g. including phonological criteria, rhymes, more background knowledge, etc.). Next, we will investigate the role of emotions: while some names were neutral, many had very strong emotional content (cf. negative emotions in *disgusoarse, horrabit* or the name given to the *hammerhead gull*, for which instead of naming it a user wrote "deserves death by fire, not a name") or positive emotions in *le trop joli*, name used for guinea lion. Another spectre of research is to investigate the generality of our blend categorisation by applying it to other domains.

This work has been supported by the projects ConCreTe (grant nb. 611733), WHIM (611560) and Prosecco (600653) funded by the European Commission, FP 7, the ICT theme, and the FET program. We thank also A. Fredriksen, the author of the pictures.

References

1. Bhatta, S.R., Goel, A.: Learning generic mechanisms for innovative strategies in adaptive design. The Journal of the Learning Sciences 6, 367–396 (1997)
2. Boden, M.: The Creative Mind: Myths and Mechanisms. Basic Books (1991)
3. Deri, A., Knight, K.: How to make a frenemy: Multitape FSTs for portmanteau generation. In: Proc. of the NAACL (2015)
4. Domeshek, E., Kolodner, J.: A case-based design aid for architecture. In: Gero, J., Sudweeks, F. (eds.) Artificial Intelligence in Design '92, pp. 497–516 (1992)
5. Fauconnier, G., Turner, M.: Conceptual integration networks. Cognitive Science 22(2), 133–187 (1998)
6. Goel, A.K., Craw, S.: Design, innovation and case-based reasoning. The Knowledge Engineering Review 20(3), 271–276 (2005)
7. Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlỳ, P., Suchomel, V., et al.: The tenten corpus family. In: 7th Int. Corpus Linguistics Conf. pp. 125–127 (2013)
8. Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D.: The sketch engine. In: Proc. EURALEX 2004. pp. 105–116 (2004)
9. Koestler, A.: The Act of Creation. New York:Macmillan (1964)
10. Kolodner, J.: Case-Based Reasoning. Morgan Kaufmann Publishers Inc. (1993)
11. Kuznetsova, P., Chen, J., Choi, Y.: Understanding and quantifying creativity in lexical composition. In: Proc. Conf. Empirical Methods in Natural Language (EMNLP). pp. 1246–1258 (2013)
12. Lehrer, A.: Blendalicious. Lexical creativity, texts and contexts pp. 115–133 (2007)
13. Li, B., Zook, A., Davis, N., Riedl, M.O.: Goal-driven conceptual blending: A computational approach for creativity. In: Proc. of the 3rd Int. Conf. on Computational Creativity. pp. 9–16 (2012)
14. Martins, P., Urbančič, T., Pollak, S., Lavrač, N., Cardoso, A.: The good, the bad, and the AHA! blends. In: Proc. of the 6th Int. Conf. on Computational Creativity (2015)
15. Norman, D.A.: The design of everyday things. Basic books, NY (1988)
16. Özbal, G., Strapparava, C.: A computational approach to the automation of creative naming. In: Proc. of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. pp. 703–711 (2012)
17. Pereira, F., Cardoso, A.: The horse-bird creature generation experiment. The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour(AISBJ) 1(3), 257–280 (2003)
18. Smith, M.R., Hintze, R.S., Ventura, D.: Nehovah: A neologism creator nomen ipsum. In: Proc. of the 5th Int. Conf. on Computational Creativity. pp. 173–181 (2014)
19. Veale, T.: Tracking the lexical zeitgeist with WordNet and Wikipedia. In: Proc. European Conf. on Artificial Intelligence (ECAI). pp. 56–60 (2006)
20. Veale, T., Butnariu, C.: Harvesting and understanding on-line neologisms. Cognitive Perspectives on Word Formation pp. 399–420 (2010)
21. Wiggins, G.A.: Towards a more precise characterization of creativity in AI. In: Proc. of Workshop Program of ICCBR – Creative Systems: Approaches to Creativity in AI and Cognitive Science (2001)